Codebook Based Hybrid Precoding for Millimeter Wave Multiuser Systems

Shiwen He, *Member, IEEE*, Jiaheng Wang, *Senior Member, IEEE*, Yongming Huang, *Member, IEEE*, Björn Ottersten, *Fellow, IEEE*, and Wei Hong, *Fellow, IEEE*

Abstract—In millimeter wave (mmWave) systems, antenna architecture limitations make it difficult to apply conventional fully digital precoding techniques but call for low cost analog radio-frequency (RF) and digital baseband hybrid precoding methods. This paper investigates joint RF-baseband hybrid precoding for the downlink of multiuser multi-antenna mmWave systems with a limited number of RF chains. Two performance measures, maximizing the spectral efficiency and the energy efficiency of the system, are considered. We propose a codebook based RF precoding design and obtain the channel state information via a beam sweep procedure. Via the codebook based design, the original system is transformed into a virtual multiuser downlink system with the RF chain constraint. Consequently, we are able to simplify the complicated hybrid precoding optimization problems to joint codeword selection and precoder design (JWSPD) problems. Then, we propose efficient methods to address the JWSPD problems and jointly optimize the RF and baseband precoders under the two performance measures. Finally, extensive numerical results are provided to validate the effectiveness of the proposed hybrid precoders.

Index Terms—Hybrid precoding design, millimeter wave communication, energy efficient communication, successive convex approximation, power allocation.

I. INTRODUCTION

The proliferation of multimedia infotainment applications and high-end devices (e.g., smartphones, tablets, wearable devices, laptops, machine-to-machine communication devices)

S. He is with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 210096, China. (Email: shiwenhe@seu.edu.cn).

J. Wang is with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 210096, China. (Email: jhwang@seu.edu.cn).

Y. Huang (Corresponding author) is with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 210096, China. (Email: huangym@seu.edu.cn).

B. Ottersten is with the Interdisciplinary Centre for Security Reliability and Trust (SnT), University of Luxembourg, Luxembourg, and also with the Royal Institute of Technology (KTH), Stockholm, Sweden. (e-mail: bjorn.ottersten@uni.lu; bjorn.ottersten@ee.kth.seg).

W. Hong is with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 210096, China. (Email: weihong@seu.edu.cn).

causes an explosive demand for high-rate data services. Future wireless communication systems face significant challenges in improving system capacity and guaranteeing users' quality of service (QoS) experiences [1]. In the last few years, various physical layer enhancements, such as massive multiple-input multiple-output (MIMO) [2], cooperation communication [3], and network densification [4] have been proposed. Along with these technologies, there is a common agreement that exploiting higher frequency bands, such as the millimeter wave (mmWave) frequency bands, is a promising solution to increase network capacity for future wireless networks [5]–[7].

MmWave communication spans a wide frequency range from 30 GHz to 300 GHz and thus enjoys much wider bandwidth than today's cellular systems [8]. However, mmWave signals experience more severe path loss, penetration loss, and rain fading compared with signals in sub-6 GHz frequency bands. For example, the free space path loss (FSPL) at 60 GHz frequency bands is 35.6 dB higher than that at 1 GHz [9], [10]. Such a large FSPL must be compensated by the transceiver in mmWave communication systems. Fortunately, the very small wavelength of mmWave signals enables a large number of miniaturized antennas to be packed in small dimension, thus forming a large multi-antenna system potentially providing very large array gain. In conventional multi-antenna systems, each active transmit antenna is connected to a separate transmit radio frequency (RF) chain. Although physical antenna elements are cheap, transmit RF chains are not cheap. A large number of transmit RF chains not only increase the cost of RF circuits in terms of size and hardware but also consume additional energy in wireless communication systems [11]. Therefore, in practice, the number of RF chains is limited and much less than the number of antennas in mmWave systems.

For ease of implementation, fully analog beamforming was proposed in [12]–[14], where the phase of the signal sent by each antenna is manipulated via analog phase shifters. However, pure analog precoding (with only one RF chain) cannot provide multiplexing gains for transmitting parallel data streams. Hence, joint RF-baseband hybrid precoding, aiming to achieve both diversity and multiplexing gains, has attracted a great deal of interest in both academia and industry for mmWave communications [15]–[20]. El Ayach *et al* in [15] exploited the inherent sparsity of mmWave channels to design low-complexity hybrid precoders with perfect channel state information (CSI) at the receiver and partial CSI at the transmitter (CSIT). Alkhateeb *et al* further investigated channel estimation for multi-path mmWave channels and tried to improve the performance of hybrid precoding using full CSIT [16].

An energy-efficient hybrid precoding design was investigated by using successive interference cancelation (SIC) method for subconnected architecture [17]. Note that the hybrid precoding designs in [15]–[17] assume that either perfect or partial CSIT is available. In practice, while using partial CSIT may degrade system performance, perfect CSIT is often difficult to obtain in mmWave communication systems, especially when there are a large number of antennas. The RF-baseband hybrid precoders in [15]–[17] were designed to obtain the spatial diversity or multiplexing gain for point-to-point mmWave communication systems. It is well known that multiuser communications can further provide multiuser diversity [18]–[20]. In [19], the authors proposed a RF precoder for multiuser mmWave systems by matching the phase of the channel of each user also under the assumption of perfect CSIT. Later, a low-complexity codebook based RF-baseband hybrid precoder was proposed for a downlink multiuser mmWave system [20]. Note that both [19] and [20] assume that the number of users equals the number of RF chains. In mmWave multiuser systems, it is very likely that the number of the served users per subcarrier will be less than that of RF chains. Therefore, it is necessary to study more flexible hybrid precoding designs for multiuser mmWave communication systems.

The existing RF-baseband hybrid precoding designs focus on improving the spectral efficiency of mmWave communication systems [15]–[20]. On the other hand, accompanied by the growing energy demand and increasing energy price, the system energy efficiency (EE) becomes another critical performance measure for future wireless systems [21]–[23]. In mmWave communication systems, although reducing the number of RF chains can save power consumption, the RF-baseband hybrid architecture requires additional power to operate the phase shifting network, the splitter, and the mixer at the transceiver [24]. Therefore, it is also necessary to investigate the RF-baseband hybrid precoding for improving the system EE. Recently, following the idea in [15], an energy efficient hybrid precoding method was developed for 5G wireless communication systems with a large number of antennas and RF chains [25]. Differently, in this paper, we propose a codebook based hybrid precoding method that uses the effective CSIT to design the RF-baseband precoders.

In this paper, we study the RF-baseband hybrid precoding for the downlink of a multiuser multi-antenna mmWave communication system. The hybrid precoding design takes into account two hardware limitations: (i) the analog phase shifters have constant modulus and a finite number of phase choices, and (ii) the number of transmit RF chains is limited and less than the number of antennas. The design goal is to maximize the sum rate (SR) and the EE of the system. We introduce a codebook based RF precoding design along with a beam sweep procedure to reduce the complexity of the hybrid precoder and relieve the difficulty of obtaining CSIT. The contribution of this paper are summarized as follows.

- We investigate joint optimization of the RF-baseband precoders in multiuser mmWave systems under two common performance measures, i.e., maximizing the SR and the EE of the system.
- Considering the practical limitation of phase shifters, we propose a codebook based RF precoder, whose columns (i.e., RF beamforming vectors) are specified by RF codewords, and then transform the original mmWave system into a virtual multiuser downlink multiple input single output (MISO) system.
- We propose a beam sweep procedure to obtain effective CSIT with less signaling feedback by utilizing the beam-domain sparse property of mmWave channels.
- Based on the codebook based design, we are able to simplify the original RF-baseband hybrid precoding optimization problems into joint codeword selection and precoding design (JWSPD) problems.
- We propose an efficient method to address the JWSPD problem for maximizing the system SR.
- We also develop an efficient method to address the more difficult JWSPD problem for maximizing the system EE.
- Finally, extensive numerical results are provided to verify the effects of the proposed codebook based hybrid precoding design. It is shown that the proposed method outperforms the existing methods and achieves a satisfactory performance close to that of the fully digital precoder.

The remainder of this paper is organized as follows. The system model and optimization problem formulation are described in section II. Section III introduces a codebook based mmWave RF precoding design with beam sweep. An effective joint codewords selection and precoder design method is proposed for SRmax problem in section IV. In section V, an effective joint codewords selection and precoder design method is developed for EEmax problem. In section VI, numerical evaluations of these algorithms are carried out. Conclusions are finally drawn in section VII.

The following notations are used throughout this paper. Bold lowercase and uppercase letters represent column vectors and matrices, respectively. The superscripts $(\cdot)^T$, and $(\cdot)^H$ represent the transpose operator, and the conjugate transpose operator, respectively. $tr\,(\cdot)$, $\|\cdot\|_2$, $|\cdot|$, $\|\cdot\|_{\mathcal{F}}$, $\Re\,(\cdot)$ and $\Im\,(\cdot)$ denote the trace, the Euclidean norm, the absolute value (element-wise absolute if used with a matrix), Frobenius norm, the real and imaginary operators, respectively. $\boldsymbol{X} \geq \boldsymbol{Y}$ and $\boldsymbol{X} \leq \boldsymbol{Y}$ denote an element-wise inequality. $\boldsymbol{A} \succeq \boldsymbol{0}$ denotes matrix $\boldsymbol{A}$ is a semidefinite positive matrix. $\boldsymbol{1}_{N \times N}$ and $\mathbb{1}_N$ denote respectively $N \times N$ matrix with all one entries and $N \times 1$ all-one vector. $\boldsymbol{A}\,(m, n)$ represents the $(m\text{th}, n\text{th})$ element of matrix $\boldsymbol{A}$ and $diag\,(\boldsymbol{A})$ stands for a column vector whose elements are the diagonal element of the matrix $\boldsymbol{A}$. $\mathbb{R}$ and $\mathbb{C}$ are the real number field and the complex number field, respectively. $\log\,(\cdot)$ is the logarithm with base e. The function $floor\,(x)$ rounds the elements of x to the nearest integers less than x. $\mod\,(,)$ is the modulo operation. $\upsilon_{max}^{(d)}\,(\boldsymbol{A})$ is the set of right singular vectors corresponding to the d largest singular values of matrix $\boldsymbol{A}$.

II. PROBLEM STATEMENT

A. System Model

Consider the downlink of a mmWave multiuser multiple-input single-output (MISO) cellular system as shown in Fig. 1, where the BS is equipped with M transmit antennas and S

RF chains and serves $K \leq S$ single-antenna users. Different from conventional multi-antenna communication systems, e.g., [21]–[23], where the numbers of antennas and RF chains are equal, in mmWave systems the number of antennas could be very large and it is expensive and impractical to install an RF chain for each antenna, so in practice we often have $S \leq M$.
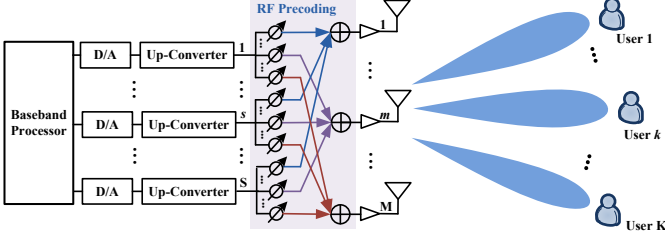

Fig. 1: Downlink mmWave system with hybrid RF-baseband precoding.

To exploit the full potential of mmWave system with a limited number of RF chains, we consider an RF-baseband hybrid precoding design, in which the transmitted signal is precoded in both the (digital) baseband domain and the (analog) RF domain. Specifically, the system model can be expressed as

$$\boldsymbol{y} = \boldsymbol{H}\underline{\boldsymbol{F}}\underline{\boldsymbol{G}}\boldsymbol{s} + \boldsymbol{n}, \qquad (1)$$

where $\boldsymbol{s}^T = [s_1, \cdots, s_K]$ with $s_k \sim \mathcal{CN}(0, 1)$ being the transmitted signal intended for the kth user, $\boldsymbol{y} = [y_1, \cdots, y_K]^T$ with y_k being the received signal of the kth user, $\boldsymbol{H}^H = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K]$ and $\boldsymbol{h}_k \in \mathbb{C}^M$ contains the channel coefficients between the BS and the kth user, and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_K)$ is an additive white gaussian noise (AWGN) vector with independent identically distributed (i.i.d.) entries of zero mean and variance σ^2. In (1), $\underline{\boldsymbol{G}} \in \mathbb{C}^{S \times K}$ is a baseband precoder that maps $\boldsymbol{s}$ to the S RF chains, and $\underline{\boldsymbol{F}} \in \mathbb{C}^{M \times S}$ is a RF precoder using analog circuitry, e.g., the analog phase shifting network. Due to the implementing limitation, the elements of $\underline{\boldsymbol{F}}$ are often required to have a constant modulus and only change their phases [12]. Then, given the RF precoder $\underline{\boldsymbol{F}}$, the baseband precoder $\underline{\boldsymbol{G}}$, and the instantaneous CSI $\boldsymbol{h}_k, \forall k \in \mathcal{K} \triangleq \{1, 2, \cdots, K\}$, the signal-to-interference-plus-noise ratio (SINR) of the kth user is

$$\underline{\text{SINR}}_k = \frac{\left| \boldsymbol{h}_k^H \underline{\boldsymbol{F}} \underline{\boldsymbol{g}}_k \right|^2}{\sum\limits_{l=1, l \neq k}^{K} \left| \boldsymbol{h}_k^H \underline{\boldsymbol{F}} \underline{\boldsymbol{g}}_l \right|^2 + \sigma^2}, \qquad (2)$$

where $\underline{\boldsymbol{g}}_k$ denotes the kth column of $\underline{\boldsymbol{G}}$.

B. Channel Model

In this paper, the channel between the BS and each user is modeled as a narrowband clustered channel based on the extended Saleh-Valenzuela model that has been widely used in mmWave communications [26], [27]. The channel coefficient vector $\boldsymbol{h}_k$ is assumed to be a sum of the contributions of N_{cl} scattering clusters, each of which includes N_{ray} propagation paths. Specifically, $\boldsymbol{h}_k$ can be written as [15]

$$\boldsymbol{h}_k = \sqrt{\frac{M}{N_{cl}N_{ray}}} \sum_{m_p=1}^{N_{cl}} \sum_{n_p=1}^{N_{ray}} \alpha_{m_p,n_p} \boldsymbol{a}\left(\phi_{m_p,n_p}, \theta_{m_p,n_p}\right), \qquad (3)$$

where α_{m_p,n_p} is a complex Gaussian random variable with zero mean and variance σ_{α,m_p}^2 for the n_pth ray in the m_pth scattering cluster, and $\phi_{m_p,n_p} \left(\theta_{m_p,n_p}\right)$ is its azimuth (elevation) angle of departure (AoD). $\boldsymbol{a}\left(\phi_{m_p,n_p}, \theta_{m_p,n_p}\right)$ is the normalized array response vector at an azimuth (elevation) angle of $\phi_{m_p,n_p} \left(\theta_{m_p,n_p}\right)$ and depends on the structure of the transmit antenna array only. The N_{ray} azimuth and elevation angles of departure ϕ_{m_p,n_p} and θ_{m_p,n_p} within the cluster m_p follow the Laplacian distributions with a uniformly-random mean cluster angle of ϕ_{m_p} and θ_{m_p}, respectively, and a constant angular spread (standard deviation) of σ_ϕ and σ_θ, respectively [28].

In particular, for an M-element uniform linear array (ULA), the array response vector is given by [15], [16]

$$\boldsymbol{a}_{\text{ULA}}(\phi) = \sqrt{\frac{1}{M}} \left[1, e^{j\frac{2\pi}{\lambda_s}d\sin(\phi)}, \cdots, e^{j(M-1)\frac{2\pi}{\lambda_s}d\sin(\phi)}\right]^T, \qquad (4)$$

where λ_s is the signal wavelength, and d is the inter-element spacing. For uniform planar array (UPA) in the yz-plane with M_1 and M_2 elements on the y and z axes respectively, the array response vector is given by [15], [16]

$$\boldsymbol{a}_{\text{UPA}}(\phi, \theta) = \sqrt{\frac{1}{M_1 M_2}}$$
$$\left[1, \cdots, e^{j\frac{2\pi}{\lambda_s}d(m_p \sin(\phi)\sin(\theta) + n_p \cos(\theta))}, \cdots, \atop e^{j\frac{2\pi}{\lambda_s}d((M_1-1)\sin(\phi)\sin(\theta) + (M_2-1)\cos(\theta))}\right]^T, \qquad (5)$$

where the antenna array size is $M_1 M_2$ and $0 \leq m_p < M_1$ $(0 \leq n_p < M_2)$ is the y (z) indices of an antenna element.

C. Problem Formulation

The goal of this paper is to design proper RF-baseband hybrid precoders for the mmWave communication system. For this purpose, we consider two common performance measures: the system sum rate (SR) and the system energy efficiency (EE). The problem of maximizing the system SR (SRmax) is formulated as:

$$\max_{\underline{\boldsymbol{F}}, \underline{\boldsymbol{G}}} \sum_{k=1}^{K} \underline{R}_k,$$
$$s.t. \underline{R}_k = \log\left(1 + \underline{\text{SINR}}_k\right) \geq \gamma_k, \forall k \in \mathcal{K}, \qquad (6)$$
$$\underline{\boldsymbol{F}} \in \mathcal{F}_{RF}, \quad \|\underline{\boldsymbol{F}}\underline{\boldsymbol{G}}\|_{\mathcal{F}}^2 \leq P.$$

The problem of maximizing the system EE (EEmax) is formulated as:

$$\max_{\underline{\boldsymbol{F}}, \underline{\boldsymbol{G}}} \frac{\sum\limits_{k=1}^{K} \underline{R}_k}{\epsilon \sum\limits_{k=1}^{K} \left\|\underline{\boldsymbol{F}}\underline{\boldsymbol{g}}_k\right\|_2^2 + Q_{dyn}}, \qquad (7a)$$
$$s.t. \underline{R}_k = \log\left(1 + \underline{\text{SINR}}_k\right) \geq \gamma_k, \forall k \in \mathcal{K}, \qquad (7b)$$

$$\underline{\boldsymbol{F}} \in \mathcal{F}_{RF}, \quad \|\underline{\boldsymbol{F}}\boldsymbol{G}\|_{\mathcal{F}}^2 \le P. \tag{7c}$$

In the above two problems, $\mathcal{F}_{RF}$ is the set of feasible RF precoders, i.e., the set of $M \times S$ matrices with constant-modulus entries, γ_k is the target rate of the kth user, P is the maximum allowable transmit power, $\epsilon \ge 1$ is a constant which accounts for the inefficiency of the power amplifier (PA) [29]. Q_{dyn} is the dynamic power consumption, including the power radiation of all circuit blocks in each active RF chain and transmit antenna, given by

$$Q_{dyn} = \|\ddot{\boldsymbol{g}}\|_0 \left(P_{RFC} + M P_{PS} + P_{DAC} \right) + P_{sta}, \tag{8}$$

where $\ddot{\boldsymbol{g}} = \left[\left\| \widetilde{\boldsymbol{g}}_1 \right\|_2, \cdots, \left\| \widetilde{\boldsymbol{g}}_S \right\|_2 \right]^T$ with $\widetilde{\boldsymbol{g}}_m$ denoting the mth row of $\underline{\boldsymbol{G}}$, and the ℓ_0-(quasi)norm $\|\ddot{\boldsymbol{g}}\|_0$ is the number of nonzero entries of $\ddot{\boldsymbol{g}}$, i.e., $\|\ddot{\boldsymbol{g}}\|_0 = \left| \left\{ t : \left\| \widetilde{\boldsymbol{g}}_t \right\|_2 \neq 0 \right\} \right|$. P_{RFC}, P_{PS}, and P_{DAC} denote the the power consumption of the RF chain, the phase shifter (PS), and the digital-to-analog converter (DAC) at the transmitter, respectively. $P_{sta} = M \left(P_{PA} + P_{mixer} \right) + P_{BB} + P_{cool}$, where P_{PA}, P_{mixer}, P_{BB}, and P_{cool} denote the power consumption of the PA, the mixer, the baseband signal processor, and the cooling system, respectively[1].

The formulated problems (6) and (7) are challenging due to several difficulties, including the constant-modulus requirement of $\underline{\boldsymbol{F}} \in \mathcal{F}_{RF}$, the coupling between $\underline{\boldsymbol{G}}$ and $\underline{\boldsymbol{F}}$, the nonconvex nature of the user rates and the QoS constraints, and the fractional form of the objective (in problem (7)). Another practical difficulty is the CSIT, which requires in general each user to estimate a large number of channels and feed them back to the BS. Throughout this paper, we assume that the set of user target rates is feasible. In the following, we will address these difficulties and propose efficient precoding designs.

III. CODEBOOK BASED MMWAVE PRECODING DESIGN WITH BEAM SWEEPING

In the mmWave system, the RF precoder is optimized in the analog domain and required to have a constant modulus. Unlike the digital baseband signal that can be precisely controlled, the RF signal is hard to manipulate and a precise shift for an arbitrary phase is prohibitively expensive in the analog domain. Therefore, in practice, each element of the RF precoder $\underline{\boldsymbol{F}}$ usually takes only several possible phase shifts, e.g., 8 to 16 choices (3 to 4 bits), while the amplitude change is usually not possible [12], [13]. To facilitate the low complexity implementation of the phase shifter, the RF precoder is often selected from a predefined codebook, which contains a limited number of phase shifts with a constant amplitude.

An RF codebook can be represented by a matrix, where each column specifies a transmit pattern or an RF beamforming vector. In particular, let $\boldsymbol{F} \in \mathcal{F}_{CB}$ be an $M \times N$ predesigned codebook matrix, where N is the number of codewords in the codebook $\boldsymbol{F}$, and $\mathcal{F}_{CB}$ denotes the space of all $M \times N$ constant-modulus RF precoding codewords. There are different RF codebooks, such as the general quantized beamforming codebooks and the beamsteering codebooks.

[1]The proposed framework in the paper can be readily extended to include the power consumption at the receivers.

A q-bit resolution beam codebook for an M-element ULA is defined by a codebook matrix $\boldsymbol{F}$, where each column corresponds to a phase rotation of the antenna elements and generates a specific beam. A q-bit resolution codebook that achieves the uniform maximum gain in all directions with the optimal beamforming weight vectors is expressed as [14]

$$\boldsymbol{F}(m,n) = \frac{1}{\sqrt{M}} j^{\frac{4(m-1)(n-1)-2N}{2^q}}, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \tag{9}$$

where j denotes the square root of -1, i.e., $j = \sqrt{-1}$, $\mathcal{M} = \{1, \cdots, M\}$, $\mathcal{N} = \{1, \cdots, N\}$.

The codebooks in IEEE 802.15.3c [26] and wireless personal area networks (WPAN) operating in 60 GHz frequency band [27] are designed to simplify hardware implementation. The codebooks are generated with a 90-degree phase resolution and without amplitude adjustment to reduce the power consumption. In this case, the (m,n)th element of the codebook $\boldsymbol{F}$ is given by (10), $\forall m \in \mathcal{M}, \forall n \in \mathcal{N}$.

$$\boldsymbol{F}(m,n) = \frac{1}{\sqrt{M}} j^{floor\left(\frac{4(m-1)\left(\mod\left((n-1)+\frac{N}{4}, N \right) \right)}{N} \right)}. \tag{10}$$

Note that when M or N is larger than 4, the codebooks obtained from (10) result in the beam gain loss in some beam directions, due to the quantized phase shifts per antenna element with a limited 2-bit codebook resolution.

In practice, discrete Fourier transform (DFT) codebooks are also widely used as they can achieve higher antenna gains at the beam directions than the codebooks in IEEE 802.15.3c. The entries of a DFT codebook are defined as

$$\boldsymbol{F}(m,n) \triangleq \frac{1}{\sqrt{M}} e^{-\frac{j2\pi(m-1)(n-1)}{M}}, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}. \tag{11}$$

The DFT codebooks generated in (11) do not suffer any beam gain loss in the given beam directions for any M and N. For mmWave systems, an efficient DFT codebook based MIMO beamforming training scheme was proposed in [30] to estimate the antenna weight vectors (AWVs).

In Fig. 2, we show the polar plots of array factor for two 3-bit resolution codebooks using (6) and (11), and a 2-bit resolution codebook using (10). It can be observed that compared to the 2-bit resolution codebook in IEEE 802.15.3c generated according to (10), the 3-bit resolution beam codebook generated according to (6) and the DFT codebook provide a better resolution and a symmetrical uniform maximum gain pattern with reduced side lobes.
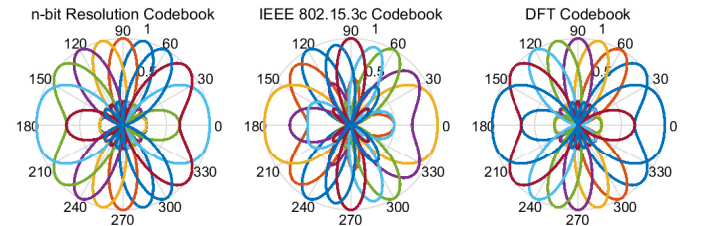

Fig. 2: Polar plots for array factor of 2-bit and 3-bit resolution codebooks with 8 patterns, $M = 4$, $N = 8$.

Adopting an RF codebook dramatically redeuces the complexity of computing the RF precoder. Indeed, given an RF

codebook $\boldsymbol{F}$, the optimization of the RF precoder $\underline{\boldsymbol{F}}$ in (6) and (7) is then equivalent to selecting S codewords (columns) from the RF codebook (matrix) $\boldsymbol{F}$. Moreover, instead of obtaining directly the exact CSIT, we can obtain the equivalent CSIT via a beam-sweep procedure [26], [27]. Specifically, during the beam-sweep procedure, the BS sends training packets from each direction defined in the RF codebook $\boldsymbol{F}$, and the users measure the received signal strength and estimate the effective channel across all directions. Then, each user provides the beam-sweep feedback to the BS, indicating the received signal strength and the effective channel of each direction, i.e., $\boldsymbol{h}_k^H \boldsymbol{f}_n$, where $\boldsymbol{f}_n$ is the nth codeword (column) of the RF codebook (matrix) $\boldsymbol{F}$. Such a beam-sweep procedure is shown in Fig. 3.
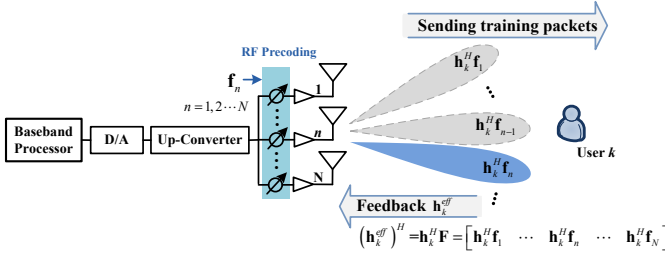

Fig. 3: Beam-sweep Procedure.

Remark 1. Through the beam sweeping, the original system can be viewed as a virtual multiuser MISO downlink system, as illustrated in Fig. 4, where the BS is equipped with N virtual antennas (i.e., codewords) and the channel coefficient between the BS and the kth user is $\boldsymbol{h}_k^{eff} = \boldsymbol{F}^H \boldsymbol{h}_k, \forall k \in \mathcal{K}$. It is well known that a mmWave channel equipped with a directional array usually admits a sparse property in the beam domain [15], [16]. That is, the effective channel may be near zero for most codewords $\boldsymbol{f}_n$ in the RF codebook $\boldsymbol{F}$. As a result, the effective channel coefficient vector $\boldsymbol{h}_k^{eff}$ is a sparse vector, implying that we only need to feedback a few nonzero effective channel coefficients to the BS. Therefore, by using a RF codebook along with the beam sweeping, the burden of obtaining CSIT in the mmWave system can be relieved.
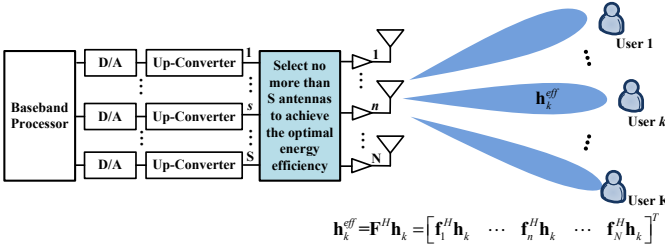

Fig. 4: Virtual Multiuser MISO Communication System.

Now, the hybrid precoding design becomes the joint optimization of the RF codeword selection and the baseband precoder. We show that this twofold task can be incorporated into the baseband precoder optimization. Specifically, instead of using the original $S \times K$ baseband precoder $\underline{\boldsymbol{G}}$, we introduce an expanded baseband precoder $\boldsymbol{G} \in \mathbb{C}^{N \times K}$ with size of

$N \times K$. Let $\widetilde{\boldsymbol{g}}_m$ denote the mth row of $\boldsymbol{G}$. Then, by multiplying the RF codebook $\boldsymbol{F}$ with $\boldsymbol{G}$, i.e., $\boldsymbol{FG}$, the mth codeword in the RF codebook $\boldsymbol{F}$ is selected if and only if $\widetilde{\boldsymbol{g}}_m$ is nonzero or equivalently $\|\widetilde{\boldsymbol{g}}_m\|_2 \neq 0$. Consequently, the original RF-baseband hybrid precoding design problem (6) can be reformulated into the following joint codeword selection and precoder design (JWSPD) SRmax problem:

$$\max_{\boldsymbol{G}} \ \sum_{k=1}^{K} R_k, \tag{12a}$$

$$s.t. \ R_k = \log\left(1 + \text{SINR}_k\right) \geq \gamma_k, \forall k \in \mathcal{K}, \tag{12b}$$

$$\sum_{k=1}^{K} \|\boldsymbol{Fg}_k\|_2^2 \leq P, \ \|\ddot{\boldsymbol{g}}\|_0 \leq S, \tag{12c}$$

where $\boldsymbol{g}_k$ denotes the kth column of $\boldsymbol{G}$, $\ddot{\boldsymbol{g}} = [\|\widetilde{\boldsymbol{g}}_1\|_2, \cdots, \|\widetilde{\boldsymbol{g}}_N\|_2]^T$, the SINR of the kth user is given by

$$\text{SINR}_k = \frac{\left|\boldsymbol{h}_k^H \boldsymbol{Fg}_k\right|^2}{\sum_{l=1, l \neq k}^{K} \left|\boldsymbol{h}_k^H \boldsymbol{Fg}_l\right|^2 + \sigma^2}. \tag{13}$$

In (12), the constraint $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ guarantees that the number of the selected codewords is no larger than the number of the available RF chains. Problem (12) represents a sparse formulation of the baseband precoder design as $\ddot{\boldsymbol{g}}$ has up to $S \leq N$ nonzero elements. It also implies that the baseband precoder $\boldsymbol{G}$ is a sparse matrix.

Similarly, problem (7) can be reformulated into the following JWSPD EEmax problem:

$$\max_{\boldsymbol{G}} \ \frac{\sum_{k=1}^{K} R_k}{\epsilon \sum_{k=1}^{K} \|\boldsymbol{Fg}_k\|_2^2 + P_{dyn}} \tag{14a}$$

$$s.t. R_k = \log\left(1 + \text{SINR}_k\right) \geq \gamma_k, \forall k \in \mathcal{K}, \tag{14b}$$

$$\sum_{k=1}^{K} \|\boldsymbol{Fg}_k\|_2^2 \leq P, \ \|\ddot{\boldsymbol{g}}\|_0 \leq S, \tag{14c}$$

where $P_{dyn} = \|\ddot{\boldsymbol{g}}\|_0 \left(P_{RFC} + MP_{PS} + P_{DAC}\right) + P_{sta}$. Let m_l be the row index of the lth nonzero row vector of $\boldsymbol{G}$ for $l = 1, \cdots, \|\ddot{\boldsymbol{g}}\|_0$ with $m_1 \leqslant \cdots \leqslant m_{\|\ddot{\boldsymbol{g}}\|_0}$. Without loss of generality, we can let the lth row vector of the baseband precoder be the $\widetilde{\boldsymbol{g}}_{m_l}$ and the lth phase shifter network steer vector be the m_lth codeword in the RF codebook $\boldsymbol{F}$ for the lth RF chain. Then, the remained $S - \|\ddot{\boldsymbol{g}}\|_0$ RF chains with the corresponding phase shifter networks can be turned off to save power.

So far, we have simplified the original RF-baseband hybrid precoding design into the JWSPD optimization problem. However, problems (12) and (14), although there is only one (matrix) variable $\boldsymbol{G}$, are still difficult, due to the nonconvex objective, the nonconvex QoS constraint, and the ℓ_0-(quasi)norm constraint $\|\ddot{\boldsymbol{g}}\|_0 \leqslant S$.

IV. JOINT CODEWORD SELECTION AND PRECODER OPTIMIZATION FOR SRMAX PROBLEM

In this section, we consider first the JWSPD SRmax problem (12), which, unfortunately, is NP-hard as a result of

the nonconvex (sum rate) objective and the ℓ_0-(quasi)norm constraint. Hence, finding its globally optimal solution requires prohibitive complexity, so in practice an efficient (probably suboptimal) solution is more preferred. In what follows, we will provide such an efficient solution.

A. Joint Codeword Selection and Precoder Design for SRmax problem

To address the joint codeword selection and precoder design (JWSPD) in (12), we first introduce some auxiliary variables $\alpha_k, \beta_k, \forall k \in \mathcal{K}, \tau, \kappa$, and χ. Let $\log(1 + \alpha_k) \geq \beta_k$ and $\mathrm{SINR}_k \geq \alpha_k, \forall k \in \mathcal{K}$. After some basic operations, (12) can be rewritten into the following equivalent form:

$$\min_{\{\boldsymbol{g}_k, \alpha_k, \beta_k\}} -\sum_{k=1}^{K} \beta_k \tag{15a}$$

$$s.t.\ 1 + \alpha_k \geq e^{\beta_k}, \forall k \in \mathcal{K}, \tag{15b}$$

$$\mathrm{SINR}_k \geq \alpha_k, \mathrm{SINR}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \tag{15c}$$

$$\sum_{k=1}^{K} \|\boldsymbol{F}\boldsymbol{g}_k\|_2^2 \leq P, \ \|\ddot{\boldsymbol{g}}\|_0 \leq S, \tag{15d}$$

where $\overline{\gamma}_k = e^{\gamma_k} - 1$. It can be easily proven that the constraints (15b) and $\mathrm{SINR}_k \geq \alpha_k, \forall k$ shall be activated at the optimal solution [31]. The difficulty lies in (15c) and (15d), as (15c) and $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ are nonconvex constraints. To overcome these difficulties, we first move the constraint $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ into the objective as follows:

$$\min_{\{\boldsymbol{g}_k, \alpha_k, \beta_k\}} -\sum_{k=1}^{K} \beta_k + \lambda\|\ddot{\boldsymbol{g}}\|_0 \tag{16a}$$

$$s.t.\ 1 + \alpha_k \geq e^{\beta_k}, \forall k \in \mathcal{K}, \sum_{k=1}^{K} \|\boldsymbol{F}\boldsymbol{g}_k\|_2^2 \leq P \tag{16b}$$

$$\mathrm{SINR}_k \geq \alpha_k, \mathrm{SINR}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \tag{16c}$$

where λ is a group-sparsity inducing regularization [32] to control the sparsity of the solution, i.e., the larger λ is, the more sparse solution of (16) is. Therefore, one can always choose a λ large enough such that the constraint $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ is satisfied.

Then, we use the convex $\ell_{1,\infty}$-norm squared to approximate the nonconvex ℓ_0-(quasi)norm[2]. In this way, problem (16) is approximated as:

$$\min_{\{\boldsymbol{g}_k, \alpha_k, \beta_k\}} -\sum_{k=1}^{K} \beta_k + \lambda\|\boldsymbol{G}\|_{1,\infty}^2 \tag{17a}$$

$$s.t.\ 1 + \alpha_k \geq e^{\beta_k}, \forall k \in \mathcal{K}, \sum_{k=1}^{K} \|\boldsymbol{F}\boldsymbol{g}_k\|_2^2 \leq P \tag{17b}$$

$$\mathrm{SINR}_k \geq \alpha_k, \mathrm{SINR}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \tag{17c}$$

[2]It is worth pointing out that the RF chain constraint $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ cannot be simply replaced by $\|\ddot{\boldsymbol{g}}\|_p \leq S$ with $p \geqslant 1$, since it is unknown whether ℓ_0-norm $\geq \ell_p$-norm or ℓ_0-norm $< \ell_p$-norm, which may result in a violation of the RF chain constraint.

where $\|\boldsymbol{G}\|_{1,\infty} = \sum\limits_{n=1}^{N} \max\limits_{k} |\boldsymbol{g}_k(n)|$ is as the $\ell_{1,\infty}$-norm of the matrix $\boldsymbol{G}$. Note that $\|\boldsymbol{G}\|_{1,\infty}^2$ in (17a) can be rewritten as follows:

$$\begin{aligned} \|\boldsymbol{G}\|_{1,\infty}^2 &= \left(\sum_{n=1}^{N} \max_{k} |\boldsymbol{g}_k(n)|\right)^2 \\ &= \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \left(\left(\max_{k} |\boldsymbol{g}_k(n_1)|\right)\left(\max_{k} |\boldsymbol{g}_k(n_2)|\right)\right) \\ &= \sum_{n=1}^{N} \sum_{m=1}^{N} \max_{i,j \in \{1,\cdots,K\}} |\boldsymbol{X}_{i,j}(n,m)|, \end{aligned} \tag{18}$$

where $\boldsymbol{X}_{i,j} = \boldsymbol{g}_i \boldsymbol{g}_j^H, \forall i, j$. Note that $\boldsymbol{X}_{i,j} = \boldsymbol{g}_i \boldsymbol{g}_j^H, \forall i, j$ if and only if $\boldsymbol{X}_{i,j} \succeq \boldsymbol{0}$ and $\mathrm{rank}(\boldsymbol{X}_{i,j}) = 1, \forall i, j$. Thus, problem (17) can be relaxed to

$$\min_{\{\boldsymbol{X}_{i,j}, \alpha_k, \beta_k\}} -\sum_{k=1}^{K} \beta_k + \lambda\|\boldsymbol{G}\|_{1,\infty}^2, \tag{19a}$$

$$s.t.\ 1 + \alpha_k \geq e^{\beta_k}, \forall k \in \mathcal{K}, \sum_{k=1}^{K} \mathrm{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_{k,k}\right) \leq P, \tag{19b}$$

$$\mathrm{SINR}_k \geq \alpha_k, \mathrm{SINR}_k \geq \overline{\gamma}_k, \boldsymbol{X}_{k,k} \succeq \boldsymbol{0}, \forall k \in \mathcal{K}, \tag{19c}$$

$$\mathrm{rank}(\boldsymbol{X}_{i,j}) = 1, \forall i, j, \tag{19d}$$

where $\widetilde{\boldsymbol{F}} = \boldsymbol{F}^H \boldsymbol{F}$, and

$$\mathrm{SINR}_k = \frac{\mathrm{tr}(\boldsymbol{H}_k \boldsymbol{X}_{k,k})}{\sum\limits_{l=1, l \neq k}^{K} \mathrm{tr}(\boldsymbol{H}_k \boldsymbol{X}_{l,l}) + \sigma^2}$$

where $\boldsymbol{H}_k = \boldsymbol{F}^H \boldsymbol{h}_k \boldsymbol{h}_k^H \boldsymbol{F}, \forall k \in \mathcal{K}$. The relaxed problem (19) is still difficult as it is still nonconvex. Nevertheless, note that $\boldsymbol{X}_{i,j}, \forall i \neq j$ only appear in the objective (19a), it is easy to have the following results which can help us simplify (19).

Theorem 1. *Let* $\{\breve{\boldsymbol{X}}_{i,j}, \breve{\alpha}_k, \breve{\beta}_k\}$ *be the optimal solution of (19), then the inequalities* $\left|\breve{\boldsymbol{X}}_{i,j}\right| \leqslant \left|\breve{\boldsymbol{X}}_{i,i}\right|, \forall i \neq j$ *hold.*

For brevity, let $\boldsymbol{X}_k = \boldsymbol{X}_{k,k}, \forall k \in \mathcal{K}$ and define $\boldsymbol{Z}(n,m) = \max\limits_{k \in \mathcal{K}} |\boldsymbol{X}_k(n,m)|, \forall m, n$. Considering that the rank one constraint is nonconvex [32], we obtain a tractable formulation form of problem (19) by dropping the nonconvex constraints $\mathrm{rank}(\boldsymbol{X}_k) = 1, \forall k \in \mathcal{K}$. According to Theorem 1, problem (19) can be relaxed to:

$$\min_{\{\boldsymbol{X}_k, \alpha_k, \beta_k\}, \boldsymbol{Z}} -\sum_{k=1}^{K} \beta_k + \lambda\mathrm{tr}(\mathbf{1}_{N \times N}\boldsymbol{Z}) \tag{20a}$$

$$s.t.\ 1 + \alpha_k \geq e^{\beta_k}, \forall k \in \mathcal{K}, \sum_{k=1}^{K} \mathrm{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k\right) \leq P, \tag{20b}$$

$$\mathrm{SINR}_k \geq \alpha_k, \mathrm{SINR}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \tag{20c}$$

$$\boldsymbol{X}_k \succeq \boldsymbol{0}, \boldsymbol{Z} \geq |\boldsymbol{X}_k|, \forall k \in \mathcal{K}. \tag{20d}$$

To address the nonconvex constraints (20c), we transform it into the following problem (21), at the top of this page, by introducing auxiliary variables $\psi_k, \phi_k, \forall k \in \mathcal{K}, \tau, \kappa$, and χ, The difficulty of solving (21) lies in (21d), as the constraints

$$\min_{\{\boldsymbol{X}_k,\alpha_k,\beta_k,\psi,\phi_k\},\boldsymbol{Z}} -\sum_{k=1}^{K}\beta_k + \lambda\mathrm{tr}\left(\boldsymbol{1}_{N\times N}\boldsymbol{Z}\right), \tag{21a}$$

$$s.t.\ \psi_k^2 \leqslant \mathrm{tr}\left(\boldsymbol{H}_k\boldsymbol{X}_k\right), \boldsymbol{X}_k \succeq \boldsymbol{0}, 1+\alpha_k \geq e^{\beta_k}, \forall k \in \mathcal{K} \tag{21b}$$

$$\sum_{l=1,l\neq k}^{K}\mathrm{tr}\left(\boldsymbol{H}_k\boldsymbol{X}_l\right) + \sigma^2 \leqslant \phi_k, \forall k \in \mathcal{K}, \sum_{k=1}^{K}\mathrm{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k\right) \leq P, \tag{21c}$$

$$\sum_{l=1,l\neq k}^{K}\overline{\gamma}_k\mathrm{tr}\left(\boldsymbol{H}_k\boldsymbol{X}_l\right) + \overline{\gamma}_k\sigma^2 \leqslant \mathrm{tr}\left(\boldsymbol{H}_k\boldsymbol{X}_k\right), \frac{\psi_k^2}{\phi_k} \geqslant \alpha_k, \forall k \in \mathcal{K}, \tag{21d}$$

$$\begin{bmatrix} \boldsymbol{Z}\left(n,m\right) - \Re\left(\boldsymbol{X}_k\left(n,m\right)\right) & \Im\left(\boldsymbol{X}_k\left(n,m\right)\right) \\ \Im\left(\boldsymbol{X}_k\left(n,m\right)\right) & \boldsymbol{Z}\left(n,m\right) + \Re\left(\boldsymbol{X}_k\left(n,m\right)\right) \end{bmatrix} \succeq \boldsymbol{0}, \forall k \in \mathcal{K}, m, n. \tag{21e}$$

$\frac{\psi_k^2}{\phi_k} \geqslant \alpha_k, \forall k$ are nonconvex. To overcome this difficulty, we exploit the SCA method [34] to approximate the inequality $\frac{\psi_k^2}{\phi_k} \geqslant \alpha_k, \forall k$ by its convex low boundary as

$$\frac{\psi_k^2}{\phi_k} \geq \Phi_k^{(I)}\left(\psi_k,\phi_k\right) \triangleq 2\frac{\psi_k^{(I)}}{\phi_k^{(I)}}\psi_k - \left(\frac{\psi_k^{(I)}}{\phi_k^{(I)}}\right)^2\phi_k, \forall k \in \mathcal{K}, \tag{22}$$

where the superscript I denotes the Ith iteration of the SCA method. Note that $\Phi_k^{(I)}\left(\psi_k,\phi_k\right)$ is in fact the first order of $\frac{\psi_k^2}{\phi_k}$ around the point $\left(\psi_k^{(I)},\phi_k^{(I)}\right)$. Thus, the approximate convex problem solved at iteration $I+1$ of (21) is given by:

$$\min_{\{\boldsymbol{X}_k,\alpha_k,\beta_k,\psi,\phi_k\},\boldsymbol{Z}} -\sum_{k=1}^{K}\beta_k + \lambda\mathrm{tr}\left(\boldsymbol{1}_{N\times N}\boldsymbol{Z}\right), \tag{23a}$$

$$s.t.\ (21b),(21c),(21e), \tag{23b}$$

$$\Phi_k^{(I)}\left(\psi_k,\phi_k\right) \geqslant \alpha_k, \forall k \in \mathcal{K}, \tag{23c}$$

$$\sum_{l=1,l\neq k}^{K}\overline{\gamma}_k\mathrm{tr}\left(\boldsymbol{H}_k\boldsymbol{X}_l\right) + \overline{\gamma}_k\sigma^2 \leqslant \mathrm{tr}\left(\boldsymbol{H}_k\boldsymbol{X}_k\right), \forall k \in \mathcal{K}, \tag{23d}$$

which can be solved efficiently via a modern convex solver such as MOSEK [22]. For conciseness, let $\boldsymbol{\Xi}^{(I)}$ denote the set of all variables in problem (23) at the Ith iteration. Algorithm 1 outlines an iterative procedure for finding a solution to problem (20) (or equivalently (21)) with a fixed λ, where τ denotes the objective of problem (20).

Algorithm 1 Joint Codeword Selection and Precoder Optimization with fixed λ

1: Let $I = 0$, generate initial points $\boldsymbol{\Xi}^{(I)}$ and compute $\tau^{(I)}$.
2: Solve (23) with $\boldsymbol{\Xi}^{(I)}$, then obtain $\boldsymbol{\Xi}^{(*)}$.
3: If $\left|\tau^{(*)} - \tau^{(I)}\right| \leq \zeta$, then output $\boldsymbol{\Xi}^{(*)}$, $\tau^{(*)}$, and stop iteration. Otherwise, $I \leftarrow I+1$, $\tau^{(I)} \leftarrow \tau^{(*)}$, $\boldsymbol{\Xi}^{(I)} \leftarrow \boldsymbol{\Xi}^{(*)}$, and go to step 2.

Problem (23) consists of a linear objective function, $K\left(M^2+1\right)$ positive-semidefinite constraints, $5K$ linear inequality constraints, and one convex constraint. It can be solved via convex optimization methods, such the interior point method [31]. The interior point method will take

$\mathcal{O}\left(\sqrt{KM}\log\left(\epsilon\right)\right)$ iterations, where the parameter ϵ represents the solution accuracy at the algorithm's termination. In each iteration, the complexity of solving (23) is $\left(\left(M^6+64\right)K^3 + 6K^2M^2\right)$ [33]. The optimal solution returned at the Ith iteration is also feasible for the problem at the $(I+1)$th iteration, as a result of the approximation in (23c). Hence, Algorithm 1 yields a nondecreasing sequence. Since the objective of problem (20) is bounded under the limited transmit power, the convergence of Algorithm 1 is guaranteed [35]. In addition, following the similar arguments in [34], it can be proved that Algorithm 1 converges to a Karush-Kunhn-Tuker (KKT) solution of problem (21) [36]. To obtain a good initial point $\boldsymbol{\Xi}^{(0)}$ for Algorithm 1, one can solve the problem (24) which was extensively studied in [32].

$$\min_{\{\boldsymbol{X}_k\},\boldsymbol{Z}} \mathrm{tr}\left(\boldsymbol{1}_{N\times N}\boldsymbol{Z}\right) s.t.(21e),(23d), \boldsymbol{X}_k \succeq \boldsymbol{0}, \forall k \in \mathcal{K}. \tag{24}$$

Remark 2. Let $\boldsymbol{\Xi}^\lambda$ denote the optimal solution to problem (19) with fixed λ. By definition, the nonzero diagonal entries of $\boldsymbol{Z}^\lambda$ correspond to the selected virtual antennas (codewords). If an entry of $\boldsymbol{Z}^\lambda$ is zero, then the corresponding entry in all $\boldsymbol{X}_k^\lambda, \forall k$ must be zero. Let L^λ be the number of nonzero diagonal entries of $\boldsymbol{Z}^\lambda$. Then, the effective channel of the kth user is an $L^\lambda \times 1$ vector $\overline{\boldsymbol{h}}_k = \widehat{\boldsymbol{F}}^H\boldsymbol{h}_k$ where the columns of $\widehat{\boldsymbol{F}}$ are the L^λ selected codewords from the RF codebook $\boldsymbol{F}$. Thus, the analog precoder $\underline{\boldsymbol{F}}$ is obtained as $\underline{\boldsymbol{F}} = \widehat{\boldsymbol{F}}^H$.

B. Sparse Parameter for SRmax problem

In the previous subsection, we have introduced a turnable sparse parameter λ to control the sparsity of the solution of the JWSPD optimization. In this subsection, we investigate how to choose a proper λ to satisfy the RF chain constraint $\|\ddot{\boldsymbol{g}}\|_0 \leq S$. Note that in (20), a larger λ makes the entries of $\boldsymbol{Z}$ (as well as $\boldsymbol{X}_k, \forall k \in \mathcal{K}$) more sparse, implying that less RF chains are used. On the other side, to maximize the system SR and guarantee the target rate requirement of each user, one cannot force all entries of $\boldsymbol{X}_k, \forall k \in \mathcal{K}$ to be zero. Thus, the sparse parameter λ has to be properly chosen to balance maximizing the system SR and minimizing the number of the selected virtual antennas (codewords).

It is not difficult to find that the system SR increases with the number of the RF chains. Therefore, the task of find the minimum λ such that the RF chain constraint $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ is satisfied can be accomplished by the classical one-dimension search methods, such as the bisection method [31]. For completeness, the algorithm used to find the proper sparse parameter λ such that $\|\ddot{\boldsymbol{g}}\|_0 \leq S$ is summarized in Algorithm 2, where $\boldsymbol{\Lambda}^\lambda$ and $\widetilde{\tau}^\lambda$ denote respectively the set of the solution of (20) and the value of $\sum_{k=1}^K \beta_k$ with λ, $\boldsymbol{\Lambda}^T$ and $\widetilde{\tau}^T$ denote respectively the set of the temporary solution of (20) and the temporary value of $\sum_{k=1}^K \beta_k$ with λ. Note that the initialization of Algorithm 2 can also be finished by solving (24).

Algorithm 2 SRmax Optimization for Hybrid JWSPD

1: Generate initial points $\boldsymbol{\Lambda}^0$ via solving (24), and computer $\widetilde{\tau}^T$. Let $flag = 1$.
2: **while** $flag$ **do**
3:     Let $\lambda = \frac{\lambda_L + \lambda_U}{2}$.
4:     Solve (20) with λ and Algorithm 1, then obtain $\boldsymbol{\Lambda}^\lambda$ and $\widetilde{\tau}^\lambda$.
5:     If $L^\lambda > S$, let $\lambda_L = \lambda$, otherwise, let $\lambda_U = \lambda$, $\widetilde{\tau}^\lambda \leftarrow \sum_{k=1}^K \beta_k^\lambda$.
6:     If $\left|\widetilde{\tau}^\lambda - \widetilde{\tau}^T\right| \leq \zeta$ and $L^\lambda \leq S$, then let $flag = 0$ and output $\boldsymbol{\Lambda}^\lambda$. Otherwise, $\boldsymbol{\Lambda}^T \leftarrow \boldsymbol{\Lambda}^\lambda$, $\widetilde{\tau}^T \leftarrow \widetilde{\tau}^\lambda$.
7: **end while**

C. Refined Solution for SRmax Problem

Recall that in the previous subsections, the ℓ_0(quasi)-norm has been approximated by the mixed $\ell_{1,\infty}$-norm squared to obtain a tractable solution. In addition, due to dropping the nonconvex rank constraint in (19), the solution $\boldsymbol{X}_k, \forall k \in \mathcal{K}$ obtained by solving (20) may not be rank one. Thus, the solution provided by (20) has to be refined to fit the original problem (15). For this purpose, after obtaining an approximate solution to (19), we propose to solve a size-reduced SRmax problem as the last step, omitting the antennas corresponding to the zero diagonal entries of the approximated sparse solution $\boldsymbol{Z}$. The size-reduced SRmax problem is given by:

$$\max_{\{\overline{\boldsymbol{g}}_k\}} \sum_{k=1}^K \overline{R}_k, \tag{25a}$$

$$s.t. \ \overline{\text{SINR}}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \tag{25b}$$

$$\sum_{k=1}^K \left\|\widehat{\boldsymbol{F}}\overline{\boldsymbol{g}}_k\right\|_2^2 \leq P, \tag{25c}$$

where $\overline{R}_k = \log\left(1 + \overline{\text{SINR}}_k\right)$, and $\overline{\text{SINR}}_k$ is given by

$$\overline{\text{SINR}}_k \triangleq \frac{\left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_k\right\|_2^2}{\sum_{l=1,l\neq k}^K \left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_l\right\|_2^2 + \sigma^2}. \tag{26}$$

Similarly, the size-reduced SRmax problem (25) can be equivalently reformulated as:

$$\max_{\{\overline{\boldsymbol{g}}_k, \overline{\alpha}_k, \overline{\beta}_k, \overline{\phi}_k\}} \sum_{k=1}^K \overline{\beta}_k, \tag{27a}$$

$$s.t. 1 + \overline{\alpha}_k \geqslant e^{\overline{\beta}_k}, \forall k \in \mathcal{K}, \sum_{k=1}^K \left\|\widehat{\boldsymbol{F}}\overline{\boldsymbol{g}}_k\right\|_2^2 \leq P, \tag{27b}$$

$$\frac{\left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_k\right\|_2^2}{\overline{\phi}_k} \geq \overline{\gamma}_k, \forall \frac{\left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_k\right\|_2^2}{\overline{\phi}_k} \geq \overline{\alpha}_k, \forall k \in \mathcal{K}, \tag{27c}$$

$$\sum_{l=1,l\neq k}^K \left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_l\right\|_2^2 + \sigma^2 \leqslant \overline{\phi}_k, \forall k \in \mathcal{K}. \tag{27d}$$

Similar to the problem (21), (27) is also a nonconvex problem due to the constraints in (27c). For (27c), we have the following the convex low boundary:

$$\frac{\left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_k\right\|_2^2}{\overline{\phi}_k} \geq \overline{\Phi}_k^{(I)}\left(\overline{\boldsymbol{g}}_k, \overline{\phi}_k\right) \triangleq$$
$$\frac{2\Re\left(\left(\overline{\boldsymbol{g}}_k^{(I)}\right)^H \overline{\boldsymbol{h}}_k \overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_k\right)}{\overline{\phi}_k^{(I)}} - \left(\frac{\left\|\overline{\boldsymbol{h}}_k^H \overline{\boldsymbol{g}}_k^{(I)}\right\|_2}{\overline{\phi}_k^{(I)}}\right)^2 \overline{\phi}_k, \forall k \in \mathcal{K}, \tag{28}$$

where I denotes the Ith iteration. Thus, the constraints in (27c) can be approximated as:

$$\overline{\Phi}_k^{(I)}\left(\overline{\boldsymbol{g}}_k, \overline{\phi}_k\right) \geq \overline{\gamma}_k, \overline{\Phi}_k^{(I)}\left(\overline{\boldsymbol{g}}_k, \overline{\phi}_k\right) \geq \overline{\alpha}_k, \forall k \in \mathcal{K}. \tag{29}$$

Consequently we can obtain a stationary solution to (27), by solving the following series of convex problems:

$$\max_{\{\overline{\boldsymbol{g}}_k, \overline{\alpha}_k, \overline{\beta}_k, \overline{\phi}_k\}} \sum_{k=1}^K \overline{\beta}_k, \ s.t. \ (27b), (27d), (29). \tag{30}$$

Such an iterative procedure is outlined in Algorithm 3, where $\overline{\boldsymbol{\Xi}}^{(I)}$ and $\overline{\tau}^{(I)}$ denote the set of the solution and the objective value of problem (30) at the Ith iteration, respectively. The convergence property of Algorithm 3 is similar with that of Algorithm 1. The computational complexity of Algorithm 3 is about $\mathcal{O}\left(M^4 K^4\right)$ [22].

Algorithm 3 Transmit Beamforming Optimization

1: Let $I = 0$, generate initial points $\overline{\boldsymbol{\Xi}}^{(I)}$ and compute $\overline{\tau}^{(I)}$.
2: Solve (30) with $\overline{\boldsymbol{\Xi}}^{(I)}$, then obtain $\overline{\boldsymbol{\Xi}}^{(*)}$.
3: If $\left|\overline{\tau}^{(*)} - \overline{\tau}^{(I)}\right| \leq \zeta$, then output $\overline{\boldsymbol{\Xi}}^{(*)}$ and stop iteration. Otherwise, $I \leftarrow I + 1$, $\overline{\tau}^{(I)} \leftarrow \overline{\tau}^{(*)}$, $\overline{\boldsymbol{\Xi}}^{(I)} \leftarrow \overline{\boldsymbol{\Xi}}^{(*)}$, and go to step 2.

In the next, we investigate how to obtain a good initial point for Algorithm 3. Let $\overline{\boldsymbol{g}}_k = \sqrt{q_k}\widehat{\boldsymbol{g}}_k, \forall k \in \mathcal{K}^3$. We propose to

---

[3]It is easy to find that (33) is a weighted sum power minimization problem which can be regarded as an extension of the conventional power minimization problem.

use the solution of the following problem as the initial point:

$$\min_{\{q_k, \widehat{\boldsymbol{g}}_k\}} \sum_{k=1}^{K} q_k \widehat{\boldsymbol{g}}_k^H \widehat{\boldsymbol{F}}^H \widehat{\boldsymbol{F}} \widehat{\boldsymbol{g}}_k \; s.t. \; \overline{\text{SINR}}_k \geq \overline{\gamma}_k, \|\widehat{\boldsymbol{g}}_k\|_2^2 = 1, \forall k \in \mathcal{K}. \tag{31}$$

We can show that problem (31) is dual to the following virtual uplink problem [37], [38]:

$$\min_{\{p_k, \widehat{\boldsymbol{g}}_k\}} \sigma^2 \sum_{k=1}^{K} p_k \; s.t. \; \overleftarrow{\text{SINR}}_k \geq \overline{\gamma}_k, \|\widehat{\boldsymbol{g}}_k\|_2^2 = 1, \forall k \in \mathcal{K}, \tag{32}$$

where $\widehat{\boldsymbol{g}}_k$ can be regarded as the combiner of the dual uplink channel, p_k has the interpretation of being the dual uplink power kth user in the virtual uplink, and $\overleftarrow{\text{SINR}}_k$ is given by

$$\overleftarrow{\text{SINR}}_k \triangleq \frac{p_k \left\| \overline{\boldsymbol{h}}_k^H \widehat{\boldsymbol{g}}_k \right\|_2^2}{\sum\limits_{l=1,l \neq k}^{K} p_l \left\| \overline{\boldsymbol{h}}_l^H \widehat{\boldsymbol{g}}_k \right\|_2^2 + \widehat{\boldsymbol{g}}_k^H \widehat{\boldsymbol{F}}^H \widehat{\boldsymbol{F}} \widehat{\boldsymbol{g}}_k}. \tag{33}$$

Furthermore, when the optimal solutions of problems (31) and (32) are obtained, we have $\sum\limits_{k=1}^{K} q_k \widehat{\boldsymbol{g}}_k^H \widehat{\boldsymbol{F}}^H \widehat{\boldsymbol{F}} \widehat{\boldsymbol{g}}_k = \sigma^2 \sum\limits_{k=1}^{K} p_k$. It was shown in [38] that the solution $\{\widehat{\boldsymbol{g}}_k\}$ of (32) is given by

$$\widehat{\boldsymbol{g}}_k^* \propto \max. \text{ eigenvector} \left(\left(\sum_{l=1,l \neq k}^{K} p_l \overline{\boldsymbol{H}}_l + \widehat{\boldsymbol{F}}^H \widehat{\boldsymbol{F}} \right)^{-1} \overline{\boldsymbol{H}}_k \right). \tag{34}$$

Thus, the algorithm used to solve (32) is summarized in Algorithm 4 with provable convergence [39].

Algorithm 4 Transmit Beamforming Initialization

1: Initialize beamforming vector $\{\widehat{\boldsymbol{g}}_k\}$.
2: Optimize $\{p_k\}$ by first finding the fixed-point p_k^* of the following equation by iterative function evaluation:

$$p_k^* = \overline{\gamma}_k \frac{\sum\limits_{l=1,l \neq k}^{K} p_l \left\| \overline{\boldsymbol{h}}_l^H \widehat{\boldsymbol{g}}_k \right\|_2^2 + \widehat{\boldsymbol{g}}_k^H \widehat{\boldsymbol{F}}^H \widehat{\boldsymbol{F}} \widehat{\boldsymbol{g}}_k}{\left\| \overline{\boldsymbol{h}}_k^H \widehat{\boldsymbol{g}}_k \right\|_2^2}$$

3: Find the optimal uplink beamformers based on the optimal uplink power allocation p_k^* with (35).
4: Repeat steps 2 and 3 until convergence.

To find $\{q_k\}$ in terms of $\{\widehat{\boldsymbol{g}}_k\}$ that is obtained from the virtual uplink channel, i.e., (34), we note that the SINR constraints in (31) must be all actived at the global optimum point. So

$$q_k = \sum_{l=1,l \neq k}^{K} q_l \frac{\overline{\gamma}_k}{\left\| \overline{\boldsymbol{h}}_k^H \widehat{\boldsymbol{g}}_k \right\|_2^2} \left\| \overline{\boldsymbol{h}}_k^H \widehat{\boldsymbol{g}}_l \right\|_2^2 + \sigma^2 \frac{\overline{\gamma}_k}{\left\| \overline{\boldsymbol{h}}_k^H \widehat{\boldsymbol{g}}_k \right\|_2^2}, \forall k \in \mathcal{K}. \tag{35}$$

Thus, we obtain a set of K linear equations with K unknowns $\{q_k\}$, which can be solved as

$$\boldsymbol{q} = \boldsymbol{\Psi} \boldsymbol{G} \boldsymbol{q} + \sigma^2 \boldsymbol{\Psi} \mathbb{1}_K, \tag{36}$$

where $\boldsymbol{q} = [q_1, \cdots, q_K]^T$, $\boldsymbol{\Psi} = diag \left\{ \frac{\overline{\gamma}_1}{\left\| \overline{\boldsymbol{h}}_1^H \widehat{\boldsymbol{g}}_1 \right\|_2^2}, \cdots, \frac{\overline{\gamma}_K}{\left\| \overline{\boldsymbol{h}}_K^H \widehat{\boldsymbol{g}}_K \right\|_2^2} \right\}$, $\boldsymbol{G}(k,k) = 0$ and $\boldsymbol{G}(k,l) = \left\| \overline{\boldsymbol{h}}_k^H \widehat{\boldsymbol{g}}_l \right\|_2^2$ for $k \neq l$. Defining an extended power vector $\widetilde{\boldsymbol{q}} = \left[\boldsymbol{q}^T, 1 \right]^T$ and an extended coupling matrix

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{\Psi} \boldsymbol{G} & \boldsymbol{\Psi} \mathbb{1}_K \\ \frac{1}{P_{max}} \boldsymbol{a}^T \boldsymbol{\Psi} \boldsymbol{G} & \frac{1}{P_{max}} \boldsymbol{a}^T \boldsymbol{\Psi} \mathbb{1}_K \end{bmatrix}. \tag{37}$$

where $P_{max} = \sigma^2 \sum\limits_{k=1}^{K} p_k$, $\boldsymbol{a}^T = [a_1, \cdots, a_K]$, $a_k = \widehat{\boldsymbol{g}}_k^H \widehat{\boldsymbol{F}}^H \widehat{\boldsymbol{F}} \widehat{\boldsymbol{g}}_k, \forall k$. According to the conclusions in [37], we can easily obtain the optimal power vector $\boldsymbol{q}$ as the first K components of the dominant eigenvector of $\boldsymbol{Q}$, which can be scaled such that its last component equals one. The solution for $\{q_k\}$, combined with that for $\{\widehat{\boldsymbol{g}}_k\}$, gives an explicit solution of the beamforming vector $\{\overline{\boldsymbol{g}}_k\}$ via an virtual uplink channel. Once the beamforming vector $\{\overline{\boldsymbol{g}}_k\}$ is obtained, the baseband beamforming vector $\boldsymbol{g}_k$ is obtained, as $\boldsymbol{g}_k = \left[\{\overline{\boldsymbol{g}}_k\}^T, \boldsymbol{0}_{(S-L^\lambda),1}^T \right]^T$. In fact, the remaining $S - L^\lambda$ RF chains with the corresponding phase shifter networks can be turned off to improve the system EE.

V. JOINT CODEWORD SELECTION AND PRECODER OPTIMIZATION FOR EEMAX PROBLEM

In this section, we consider the EEmax problem (14), which is more difficult than the SRmax problem. Indeed the objective in (14) is given by a more complex fractional form, and the ℓ_0-(quasi)norm appears not only in the constraint but also in the denominator of the objective. To find the globally optimal solution to (14) requires an exhaustive search over all $\sum\limits_{l=L_{Min}}^{S} \binom{N}{l}$ possible sparse patterns of $\ddot{\boldsymbol{g}}$, where $L_{Min} \leqslant S$ is the minimum number of the selected RF chains that can achieve the target rate requirement of each user under the power constraint. Unfortunately, for each pattern of $\ddot{\boldsymbol{g}}$, (14) is an NP-hard problem. Thus, we seek a practical and efficient method to address the EEmax problem (14).

A. Joint Codeword Selection and Precoder Design for EEmax problem

Similarly, we first use the convex squared $\ell_{1,\infty}$-norm to approximate the nonconvex ℓ_0-(quasi)norm in the power consumption term P_{dyn}. Then, we also introduce a turnable sparse parameter $\lambda \geq 0$ as a group-sparsity inducing regularization to control the sparsity of the solution so that the RF chain constraint (14c) can be temporarily omitted for fixed λ. By doing so, problem (14) can be relaxed as:

$$\max_{\{\boldsymbol{X}_{i,j}\}} \frac{\sum\limits_{k=1}^{K} R_k}{\epsilon \sum\limits_{k=1}^{K} \text{tr} \left(\widetilde{\boldsymbol{F}} \boldsymbol{X}_{k,k} \right) + P_{dyn}(\lambda)}, \tag{38a}$$

$$s.t. \; \text{SINR}_k \geq \overline{\gamma}_k, \forall k, \; \sum_{k=1}^{K} \text{tr} \left(\widetilde{\boldsymbol{F}} \boldsymbol{X}_{k,k} \right) \leq P, \tag{38b}$$

$$\boldsymbol{X}_{k,k} \succeq \boldsymbol{0}, \forall k \in \mathcal{K}, \qquad (38c)$$

where the nonconvex $\text{rank}\left(\boldsymbol{X}_{i,j}\right) = 1$, $\forall i, j$ constraints are dropped, and the dynamic power consumption is given by

$$P_{dyn}\left(\lambda\right) = f\left(\lambda\right) \sum_{n=1}^{N} \sum_{m=1}^{N} \max_{i,j \in \{1,\cdots,K\}} \left|\boldsymbol{X}_{i,j}\left(n,m\right)\right| + P_{sta}, \tag{39}$$

where $f\left(\lambda\right) = P_{RFC} + MP_{PS} + P_{DAC} + \lambda$. Note that $\boldsymbol{X}_{i,j}$, $\forall i \neq j$, only appear in the power consumption item $P_{dyn}\left(\lambda\right)$. Therefore, similar to Theorem 1, we have the following result.

Theorem 2. *Let* $\breve{\boldsymbol{X}}_{i,j}, \forall i, j \in \mathcal{K}$ *be the optimal solution of (38), then* $\breve{\boldsymbol{X}}_{i,i}, \forall i \in \mathcal{K}$ *with* $\boldsymbol{X}_{i,j} = \boldsymbol{0}$, $\forall i \neq j, i, j \in \mathcal{K}$ *is also the optimal solution of (38).*

Proof: First, we prove that the inequalities $\left|\breve{\boldsymbol{X}}_{i,j}\right| \leqslant \left|\breve{\boldsymbol{X}}_{i,i}\right|, \forall i \neq j$ hold. Suppose that there is one pair of indices $(i_0, j_0), i_0 \neq j_0$ and (n_0, m_0) such that $\left|\breve{\boldsymbol{X}}_{i,j}\left(n,m\right)\right| \leqslant \left|\breve{\boldsymbol{X}}_{i,i}\left(n,m\right)\right|, \forall i \neq j, n, m$ except for $\left|\breve{\boldsymbol{X}}_{i_0,j_0}\left(n_0,m_0\right)\right| \geq \left|\breve{\boldsymbol{X}}_{i,i}\left(n_0,m_0\right)\right|, \forall k$. Let $\overline{\boldsymbol{X}}_{i,j}, \forall i, j \in \mathcal{K}$ be another solution obtained by letting $\overline{\boldsymbol{X}}_{i,j}\left(n,m\right) = \breve{\boldsymbol{X}}_{i,j}\left(n,m\right), \forall i, j, n, m$ except for $\overline{\boldsymbol{X}}_{i_0,j_0}\left(n_0,m_0\right) = 0$. Note that $\boldsymbol{X}_{i,j}, \forall i \neq j$ only appear in the constraints (38b). Thus, $\overline{\boldsymbol{X}}_{i,j}, \forall i, j \in \mathcal{K}$ is a feasible solution to problem (38) and satisfies the following inequality (40), at the bottom of this page. Note that $\breve{\boldsymbol{X}}_{i,j}$ and $\overline{\boldsymbol{X}}_{i,j}$, $\forall i, j \in \mathcal{K}$ achieve the same user rate. Combining the objective of problem (38) and (40), we can obtain a better objective by using $\overline{\boldsymbol{X}}_{i,j}, \forall i, j \in \mathcal{K}$ than using $\breve{\boldsymbol{X}}_{i,j}, \forall i, j \in \mathcal{K}$, which is a contradiction. Therefore, we have $\left|\breve{\boldsymbol{X}}_{i,j}\right| \leqslant \left|\breve{\boldsymbol{X}}_{i,i}\right|, \forall i \neq j$.

Note that $\boldsymbol{X}_{i,j}$, $\forall i \neq j$, only appear in the power consumption item $P_{dyn}\left(\lambda\right)$. Combining $\left|\breve{\boldsymbol{X}}_{i,j}\right| \leqslant \left|\breve{\boldsymbol{X}}_{i,i}\right|, \forall i \neq j$ with (39), one can easily see that the power consumption item $P_{dyn}\left(\lambda\right)$ dose not change by setting $\boldsymbol{X}_{i,j} = \boldsymbol{0}$, $\forall i \neq j$. Consequently, $\breve{\boldsymbol{X}}_{i,i}, \forall i \in \mathcal{K}$ with $\boldsymbol{X}_{i,j} = \boldsymbol{0}$, $\forall i \neq j$ are still optimal. ∎

Theorem 2 also indicates that we can simplify problem (38) by setting $\boldsymbol{X}_{i,j} = \boldsymbol{0}$, $\forall i \neq j$ without any loss of optimality. Hence, similar to the transformation between (19) and (20), (38) is equivalent to

$$\max_{\{\boldsymbol{X}_k\},\boldsymbol{Z}} \frac{\sum_{k=1}^{K} R_k}{\epsilon \sum_{k=1}^{K} \text{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k\right) + P_{dyn}\left(\boldsymbol{Z},\lambda\right)}, \tag{41a}$$

$$s.t. \ \text{SINR}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \ \sum_{k=1}^{K} \text{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k\right) \leq P, \tag{41b}$$

$$\boldsymbol{X}_k \succeq \boldsymbol{0}, \boldsymbol{Z} \geq \left|\boldsymbol{X}_k\right|, \forall k \in \mathcal{K}, \tag{41c}$$

where $P_{dyn}\left(\boldsymbol{Z},\lambda\right) = f\left(\lambda\right)\text{tr}\left(\boldsymbol{1}_{N \times N}\boldsymbol{Z}\right) + P_{sta}$. Introducing auxiliary variables $\alpha_k, \beta_k, \psi_k, \phi_k, \forall k \in \mathcal{K}, \tau, \kappa,$ and χ, (41) can be equivalently rewritten as

$$\max_{\{\boldsymbol{X}_k,\alpha_k,\beta_k,\psi_k,\phi_k\},\boldsymbol{Z},\tau,\kappa,\chi} \chi, \tag{42a}$$

$$s.t. \ \frac{\tau^2}{\kappa} \geqslant \chi, \frac{\psi_k^2}{\phi_k} \geqslant \alpha_k, \forall k \in \mathcal{K} \tag{42b}$$

$$\sum_{k=1}^{K} \beta_k \geqslant \tau^2, (21b), (21c), (21e) \tag{42c}$$

$$\epsilon \sum_{k=1}^{K} \text{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k\right) + P_{dyn}\left(\boldsymbol{Z},\lambda\right) \leqslant \kappa \tag{42d}$$

$$\sum_{l=1,l\neq k}^{K} \overline{\gamma}_k \text{tr}\left(\boldsymbol{H}_k \boldsymbol{X}_l\right) + \overline{\gamma}_k \sigma^2 \leqslant \text{tr}\left(\boldsymbol{H}_k \boldsymbol{X}_k\right), \forall k \in \mathcal{K} \tag{42e}$$

Similarly, the difficulty of solving (42) lies in (42b), as the two constraints in (42b) are nonconvex. Thus, we exploit the SCA method [34] to approximate the two inequalities in (42b) by two convex constraints. By replacing (42b) with the convex lower bounds at the Ith iteration, problem (42) can be approximated by the following convex program:

$$\max_{\{\boldsymbol{X}_k,\alpha_k,\beta_k,\psi_k,\phi_k,\boldsymbol{\mu}_k\},\boldsymbol{Z},\tau,\kappa,\chi} \chi, \tag{43a}$$

$$s.t. \ (42c), (42d), (42e), \tag{43b}$$

$$\Psi^{(I)}\left(\tau,\kappa\right) \geqslant \chi, \Phi_k^{(I)}\left(\psi_k,\phi_k\right) \geqslant \alpha_k, \forall k \in \mathcal{K}, \tag{43c}$$

where $\Psi^{(I)}\left(\tau,\kappa\right) \triangleq 2\frac{\tau^{(I)}}{\kappa^{(I)}}\tau - \left(\frac{\tau^{(I)}}{\kappa^{(I)}}\right)^2 \kappa$. Thus, problem (43) can be solved via the similar procedure as described in Algorithm 1.

B. Sparse Parameter for EEmax problem

Similarly, a larger λ leads to a more sparse solution to the (approximated) EEmax problem (38), which corresponds to less RF chains used. On the other side, λ cannot be infinite, which would lead to a zero solution and contradict the task of maximizing the system EE. Hence, λ has to be properly chosen. However, unlike to the SRmax problem (20) or the total power minimization problem with RF chain constraints [32], the system EE is not monotonic with respect to the number of RF chains or the sparse parameter λ. Indeed, the system EE

$$\breve{P}_{dyn}\left(\lambda\right) = f\left(\lambda\right)\left(\left|\breve{\boldsymbol{X}}_{i_0,j_0}\left(n_0,m_0\right)\right| + \underbrace{\sum_{n=1}^{N}\sum_{m=1}^{N}}_{(n,m)\neq(n_0,m_0)} \max_i \left|\breve{\boldsymbol{X}}_{i,i}\left(n,m\right)\right|\right) + P_{sta}$$

$$> \overline{P}_{dyn}\left(\lambda\right) = f\left(\lambda\right)\sum_{n=1}^{N}\sum_{m=1}^{N}\max_i \left|\overline{\boldsymbol{X}}_{i,i}\left(n,m\right)\right| + P_{sta}. \tag{40}$$

$$\chi^{\lambda_{L+1}^{key}} = \frac{\left(\tau^{\lambda_{L+1}^{key}}\right)^2}{\kappa^{\lambda_{L+1}^{key}}} = \frac{\left(\tau^{\lambda_{L+1}^{key}}\right)^2}{\epsilon \sum_{k=1}^{K} \mathrm{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k^{\lambda_{L+1}^{key}}\right) + f\left(\lambda_{L+1}^{key}\right)\mathrm{tr}\left(\boldsymbol{1}_{N\times N}\boldsymbol{Z}^{\lambda_{L+1}^{key}}\right) + P_{sta}}$$

$$< \chi^{\lambda} = \frac{\left(\tau^{\lambda}\right)^2}{\kappa^{\lambda}} = \frac{\left(\tau^{\lambda}\right)^2}{\epsilon \sum_{k=1}^{K} \mathrm{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k^{\lambda}\right) + f\left(\lambda\right)\mathrm{tr}\left(\boldsymbol{1}_{N\times N}\boldsymbol{Z}^{\lambda}\right) + P_{sta}} \tag{44}$$

$$< \frac{\left(\tau^{\lambda}\right)^2}{\epsilon \sum_{k=1}^{K} \mathrm{tr}\left(\widetilde{\boldsymbol{F}}\boldsymbol{X}_k^{\lambda}\right) + f\left(\lambda_{L+1}^{key}\right)\mathrm{tr}\left(\boldsymbol{1}_{N\times N}\boldsymbol{Z}^{\lambda}\right) + P_{sts}},$$

is a piecewise function with respect to the sparse parameter λ, as illustrated in Fig. 7 and Table I. Consequently, the bisection method cannot be used to optimize λ [31].

To address the above issue, we devise a dynamic interval compression method to search a suitable λ. Specifically, let $\mathcal{A}_{Min}$ be the set of the indices of the L_{Min} selected virtual antennas (codewords). Let L_{Max} be the number of the virtual antennas (codewords) achieving the maximum EE by ignoring the available RF chain constraint, which correspondes to $\lambda = 0$, and $\mathcal{A}_{Max}$ be the set of the indices of the L_{Max} selected virtual antennas (codewords) in this case. Considering that the allowable number of RF chains is a discrete value but the sparse parameter λ is continuous, we introduce the following definition.

Definition 1. *For any small positive number ϵ and $\forall L \in \{L_{Min}, \cdots, L_{Max} - 1\}$, λ_L^{key} is called a breaking point if the optimal solutions $\boldsymbol{Z}^{\lambda_L^{key}-\varepsilon}$ and $\boldsymbol{Z}^{\lambda_L^{key}}$ of (41) have $L+1$ and L nonzero diagonal entries, respectively.*

Theorem 3. *Let $\boldsymbol{\Xi}^{\lambda}$ be the solution of (41) with $\forall \lambda \in \left[\lambda_{L+1}^{key}, \lambda_L^{key}\right)$ and $L, L+1 \in \{L_{Min}, L_{Min}+1, \cdots, L_{Max}\}$. Then, $\boldsymbol{Z}^{\lambda}$ has also $L+1$ nonzero diagonal entries and the inequality $\chi^{\lambda} \leqslant \chi^{\lambda_{L+1}^{key}}$ holds.*

Proof: Following the definition of the breaking point λ_{L+1}^{key}, it is easy to see that $\boldsymbol{Z}^{\lambda}$ has also $L+1$ nonzero diagonal entries. If $\chi^{\lambda} > \chi^{\lambda_{L+1}^{key}}$, recalling $\lambda_{L+1}^{key} \leqslant \lambda$, then we have (44), at the top of this page, which contradicts the fact that $\boldsymbol{\Xi}^{\lambda_{L+1}^{key}}$ is the optimal solution to problem (43) with fixed λ_{L+1}^{key}. Thus, the conclusions given in Theorem 3 are proven. $\blacksquare$

According the definition of the breaking point and the non-monotonic property of the system EE with respect to λ, one shall find the values of all breaking points. Let $\boldsymbol{Z}^{\lambda_i}$ be the solution of problem (41) with fixed λ_i, $i = 1, 2$. Let $\boldsymbol{Z}^{\lambda}$ be the solution of problem (41) for $\forall \lambda \in [\lambda_1, \lambda_2]$. Theorem 3 implies that if $\boldsymbol{Z}^{\lambda_1}$ and $\boldsymbol{Z}^{\lambda_2}$ have the same number of the nonzero diagonal entries, $\boldsymbol{Z}^{\lambda_1}$, $\boldsymbol{Z}^{\lambda_2}$, and $\boldsymbol{Z}^{\lambda}$ have the same number of the nonzero diagonal entries. Based on this result, we propose a one-dimension dynamic interval compression method, which is summarized in Algorithm 5, to find a suitable sparse parameter λ and obtain the corresponding codewords. Note that in Algorithm 5, $\mathcal{A}^{\lambda}$ denotes the set of the indices

of the selected virtual antennas (codewords) with fixed λ and ϱ^{λ} is calculated as

$$\frac{\sum_{k=1}^{K} \log\left(1 + \frac{tr(\boldsymbol{H}_k\boldsymbol{X}_k)}{\sum_{l=1,l\neq k}^{K} tr(\boldsymbol{H}_k\boldsymbol{X}_l) + \sigma_k^2}\right)}{\xi \sum_{k=1}^{K} tr(\boldsymbol{X}_k) + L^{\lambda}\left(P_{RCF} + M P_{PS} + P_{DAC}\right) + P_{sta}}.$$

The initialization of Algorithm 5 can also be obtained by solving (24) and letting other constraints to be activated. In addition, λ_U should be large enough such that the number of the active RF chains equals to L_{Min}. According to Theorem 3, if two intervals in the intervals set $\mathcal{I}$ have an intersection in the intervals set $\mathcal{I}$, they shall be combined into one interval, for example $[100, 150]$ and $[150, 200]$ are combined to $[100, 200]$.

Algorithm 5 RF Chain Set Generation: Part I

1: Initialize $\mathcal{L}_\lambda = \emptyset$, $\mathcal{A} = \emptyset$, $L_c = 0$, $L_0 = S$, $L_1 = L_{Min}$, $Flag = 1$.
2: Solve (43) with $\lambda_L = 0$, obtain $\boldsymbol{\Xi}^{\lambda_L}$, L^{λ_L}, and $\varrho^{(L)}$.
3: Let λ_U be a larger positive number, solve (43) with λ_U, obtain $\boldsymbol{\Xi}^{\lambda_U}$, L^{λ_U}, and $\varrho^{(U)}$.
4: Let $\lambda_{Temp} = 0$, $L_{Temp} = L^{\lambda_L}$, and $\mathcal{I} = \emptyset$ be a set of intervals.
5: If $L^{\lambda_L} \leqslant S$, then $\mathcal{L} = \mathcal{L} \cup \left\{\left(L^{\lambda_L}, \lambda_L\right)\right\}$, $L_c = L_c + 1$, $L_0 = L^{\lambda_L}$. Let $\mathcal{A}^{\lambda_L}$ be the set of the indices of the selected codewords, $\mathcal{A} = \mathcal{A} \cup \left\{\left\{\mathcal{A}^{\lambda_L}, \varrho^{\lambda_L}\right\}\right\}$.
6: If $L^{\lambda_U} \leqslant S$, then $\mathcal{L} = \mathcal{L} \cup \left\{\left(L^{\lambda_U}, \lambda_U\right)\right\}$, $L_c = L_c + 1$, $L_1 = L^{\lambda_U}$. Let $\mathcal{A}^{\lambda_U}$ be the set of the indices of the selected codewords indices, $\mathcal{A} = \mathcal{A} \cup \left\{\left\{\mathcal{A}^{\lambda_U}, \varrho^{\lambda_U}\right\}\right\}$.
7: If $L_0 == L_1 \,||\, L_c == L_0 - L_1 + 1$, then let $Flag = 0$.
8: **while** $Flag$ **do**
9: Running RF Chain Set Generation: Part II.
10: **end while**

C. Refined Solution for EEmax problem

Due to the introduction of the mixed $\ell_{1,\infty}$-norm squared for the selection of the RF chains in the previous subsections, the energy efficient beamforming vector cannot be directly extracted from the solution of (43), i.e., $\{\boldsymbol{X}_k\}$. Therefore, we need to construct the reduced-size channel $\overline{\boldsymbol{h}}_k = \widehat{\boldsymbol{F}}^H \boldsymbol{h}_k$

Algorithm 5 RF Chains Set Generation: Part II

1: **if** $L^{\lambda_L} - L^{\lambda_U} > 1$ **then**
2:   **for** $l = 1, \cdots, L^{\lambda_L} - L^{\lambda_U} - 1$ **do**
3:     $\lambda = \lambda_L + \frac{\lambda_U - \lambda_L}{L^{\lambda_L} - L^{\lambda_U}} l$.
4:     **if** λ is not in any interval of $\mathcal{I}$ **then**
5:       Solve (15) with Algorithm 1 and λ, obtain $\boldsymbol{\Xi}^\lambda$, L^λ, and ϱ^λ.
6:       **if** $L^\lambda > L_0$ **then**
7:         $\lambda_{Temp} = \lambda$, $L_{Temp} = L^\lambda$.
8:       **else**
9:         **if** $L^\lambda \notin \mathcal{L}$ **then**
10:           $\mathcal{L} = \mathcal{L} \cup \left\{ \left(L^\lambda, \lambda \right) \right\}$, $L_c = L_c + 1$.
11:           Let $\mathcal{A}_{L^\lambda}$ be the set of the selected codewords indices obtained by $\boldsymbol{\Xi}^\lambda$, $\mathcal{A} = \mathcal{A} \cup \{\mathcal{A}_{L^\lambda}\}$.
12:         **else**
13:           Sort the entries of $\mathcal{L}$ in ascending order with respect to λ.
14:           Find the set of the indices $\mathcal{T}$ such that $\mathcal{T} = \left\{ \iota : \mathcal{L}(\iota)(1,1) = L^\lambda \right\}$.
15:           **if** $|\mathcal{T}| == 1$ **then**
16:             $\mathcal{I} = \mathcal{I} \cup \left\{ \begin{bmatrix} \min(\mathcal{L}(\mathcal{T}(1))(1,2),\lambda, \\ \max(\mathcal{L}(\mathcal{T}(1))(1,2),\lambda) \end{bmatrix} \right\}$, $\mathcal{L} = \mathcal{L} \cup \left\{ \left(L^\lambda, \lambda \right) \right\}$.
17:           **else if** $|\mathcal{T}| == 2$ **then**
18:             $\mathcal{I} = \mathcal{I} \setminus \left\{ \begin{bmatrix} \mathcal{L}(\mathcal{T}(1))(1,2), \\ \mathcal{L}(\mathcal{T}(2))(1,2) \end{bmatrix} \right\}$.
19:             $\mathfrak{a} = \min\left(\mathcal{L}\left(\mathcal{T}\left(1 \right) \right)\left(1, 2 \right), \lambda \right)$,
$\mathfrak{b} = \max\left(\mathcal{L}\left(\mathcal{T}\left(2 \right) \right)\left(1, 2 \right), \lambda \right)$.
20:             $\mathcal{I} = \mathcal{I} \cup \{[\mathfrak{a}, \mathfrak{b}]\}$, $\mathcal{L}\left(\mathcal{T}\left(1 \right) \right)\left(1, 2 \right) = \mathfrak{a}$, $\mathcal{L}\left(\mathcal{T}\left(2 \right) \right)\left(1, 2 \right) = \mathfrak{b}$.
21:           **end if**
22:         **end if**
23:       **end if**
24:     **end if**
25:   **end for**
26: **end if**
27: **if** $L_c == L_0 - L_1 + 1$ **then**
28:   $Flag = 0$
29: **else**
30:   Sort $\mathcal{L}$ in ascending order with respect to λ.
31:   **if** $L_0 > \mathcal{L}(1)(1,1)$ **then**
32:     $\lambda_L = \lambda_{Temp}$, $L^{\lambda_L} = L_{Temp}$,
33:     $\lambda_U = \mathcal{L}(1)(1,2)$, $L^{\lambda_U} = \mathcal{L}(1)(1,1)$,
34:   **else**
35:     $Index = 1$, $flag = 1$;
36:     **while** $Index \leqslant |\mathcal{L}| - 1$ & $flag$ **do**
37:       **if** $\mathcal{L}(Index)(1,1) - \mathcal{L}(Index+1)(1,1) \geqslant 2$ **then**
38:         $\lambda_L = \mathcal{L}(Index)(1,2)$,
$L^{\lambda_L} = \mathcal{L}(Index)(1,1)$.
39:         $\lambda_U = \mathcal{L}(Index+1)(1,2)$.
40:         $L^{\lambda_U} = \mathcal{L}(Index+1)(1,1)$, $flag = 0$.
41:       **else**
42:         $Index = Index + 1$,
43:       **end if**
44:     **end while**
45:   **end if**
46: **end if**

according to the codewords selected by Algorithm 5. Thus, the reduce-sized EEmax problem is given by

$$\max_{\{\boldsymbol{g}_k\}} \frac{\sum_{k=1}^{K} \overline{R}_k}{\epsilon \sum_{k=1}^{K} \left\| \widehat{\boldsymbol{F}} \boldsymbol{g}_k \right\|_2^2 + P_{dyn}^*}, \tag{45a}$$

$$s.t. \ \overline{\mathrm{SINR}}_k \geq \overline{\gamma}_k, \forall k \in \mathcal{K}, \sum_{k=1}^{K} \left\| \widehat{\boldsymbol{F}} \boldsymbol{g}_k \right\|_2^2 \leq P, \tag{45b}$$

where $P_{dyn}^{\lambda_L^{key}} = L^{\lambda_L^{key}}(P_{RFC} + M P_{PS} + P_{DAC}) + P_{sta}$. Problem (45) can be formulated as:

$$\max_{\{\boldsymbol{g}_k, \overline{\alpha}_k, \overline{\beta}_k, \overline{\phi}_k\}, \overline{\tau}, \overline{\chi}, \overline{\kappa}} \overline{\chi}, \tag{46a}$$

$$s.t. \ \sum_{k=1}^{K} \overline{\beta}_k \geqslant \overline{\tau}^2, \frac{\overline{\tau}^2}{\overline{\kappa}} \geqslant \overline{\chi}, \tag{46b}$$

$$\epsilon \sum_{k=1}^{K} \left\| \widehat{\boldsymbol{F}} \boldsymbol{g}_k \right\|_2^2 + P_{dyn}^* \leqslant \overline{\kappa}, \tag{46c}$$

$$(27b), (27c), (27d). \tag{46d}$$

Similarly, instead of directly solving (46), we resort to solving the following convex approximated problem

$$\max_{\{\boldsymbol{g}_k, \overline{\alpha}_k, \overline{\beta}_k, \overline{\phi}_k\}, \overline{\tau}, \overline{\chi}, \overline{\kappa}} \overline{\chi}, \tag{47a}$$

$$s.t. \ \sum_{k=1}^{K} \overline{\beta}_k \geqslant \overline{\tau}^2, \overline{\Psi}^{(I)}\left(\overline{\tau}, \overline{\kappa} \right) \geqslant \overline{\chi}, (46c), (46d), \tag{47b}$$

where I denotes the Ith iteration, and $\overline{\Psi}^{(I)}\left(\overline{\tau}, \overline{\kappa} \right)$ is given by

$$\overline{\Psi}^{(I)}\left(\overline{\tau}, \overline{\kappa} \right) \triangleq 2 \frac{\overline{\tau}^{(I)}}{\overline{\kappa}^{(I)}} \overline{\tau} - \left(\frac{\overline{\tau}^{(I)}}{\overline{\kappa}^{(I)}} \right)^2 \overline{\kappa} \tag{48}$$

Thus, problem (47) can be solved in a similar manner as described in Algorithm 3.

VI. NUMERICAL RESULTS

In this section, we present numerical results to demonstrate the performance of our developed RF-baseband hybrid precoding design. A uniform linear array with antenna spacing equal to a half wavelength is adopted, and the RF phase shifters use quantized phases. The predesigned codebook $\boldsymbol{F}$ is the DFT codebook. The propagation environment is modeled as $N_{cl} = 6$ with $N_{ray} = 8$ for each cluster with Laplacian distributed angles of departure. For simplicity, we assume that all clusters are of equal power, i.e., $\sigma_{\alpha, m_p}^2 = \sigma_\alpha^2, \forall m_p$ [15]. The mean cluster angle of ϕ_{m_p} is uniformly distributed over $[-\pi, \pi)$, and the constant angular spread of AoD σ_ϕ is 7.5^o. $P_{RFC} = 43$ mW, $P_{PA} = 20$ mW, $P_{DAC} = 200$ mW, $P_{PS} = 30$ mW, $P_{mixer} = 19$ mW, $P_{BB} = 300$ mW, and $P_{cool} = 200$ mW [40]. The noise power spectrum density is $\sigma^2 = 1$. For fairness, all simulated precoding designs use the same total power constraint and the signal-to-noise ratio is defined as $\mathrm{SNR} = 10 \log_{10}\left(\frac{P}{\sigma^2} \right)$. The inefficiency factor of power amplifier ϵ is set to unit and the stop threshold is

$\zeta = 10^{-3}$. In all simulation figures, the simulated EE of the system is given by

$$\frac{\sum_{k=1}^{K} \log_2 \left(1 + \frac{\left\| \boldsymbol{h}_k^H \widehat{\boldsymbol{F}} \overline{\boldsymbol{g}}_k \right\|_2^2}{\sum_{l=1,l\neq k}^{K} \left\| \boldsymbol{h}_k^H \widehat{\boldsymbol{F}} \overline{\boldsymbol{g}}_l \right\|_2^2 + \sigma^2} \right)}{\epsilon \sum_{k=1}^{K} \left\| \widehat{\boldsymbol{F}} \overline{\boldsymbol{g}}_k \right\|_2^2 + P_{dyn}^*}. \tag{49}$$

We compare the performance of the proposed strategy to the optimal fully digital precoder with one RF chain per antenna, whose EE is calculated as

$$\frac{\sum_{k=1}^{K} \log_2 \left(1 + \frac{\left\| \boldsymbol{h}_k^H \overline{\boldsymbol{g}}_k \right\|_2^2}{\sum_{l=1,l\neq k}^{K} \left\| \boldsymbol{h}_k^H \overline{\boldsymbol{g}}_l \right\|_2^2 + \sigma^2} \right)}{\epsilon \sum_{k=1}^{K} \left\| \overline{\boldsymbol{g}}_k \right\|_2^2 + M \left(P_{RFC} + P_{DAC} + P_{PA} \right) + P_{BB} + P_{cool}}. \tag{50}$$

In our simulation scenario, fully digital precoding denotes using Algorithm 3 to solve the SRmax problem, where each antenna connects with an independent RF channel at the BS. Fully analog beamforming is achieved by selecting the best codeword from the codebook via beam training and setting the baseband precoder as an identity matrix with uniform power allocation between users. OMP SRmax Hybrid Precoding uses the orthogonal matching pursuit method [15] to obtain the RF-baseband precoders based on the solution of the fully digital SRmax problem.

Fig. 5 and Fig. 6 show the SR performance of various hybrid precoding designs as well as the fully digital precoder. The results are obtained by averaging over 1000 random channel realizations. The target rate of the kth user is set to be zero. Numerical results show that the proposed hybrid precoding design achieves the highest SR among several hybrid RF-baseband precoders. This is because the proposed hybrid RF-baseband precoding design provides a more flexible way to achieve the beam diversity gain. One can see that the fully analog beamforming method results in the worst SR performance, indicating that the inter-user interference cannot be effectively suppressed.
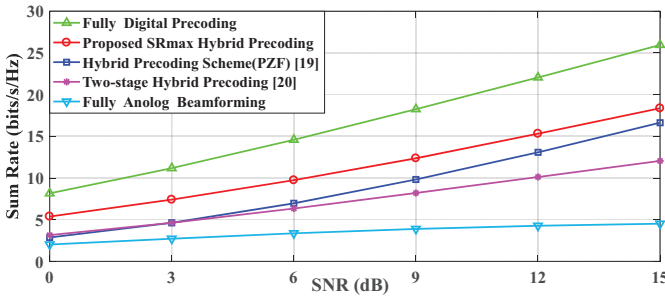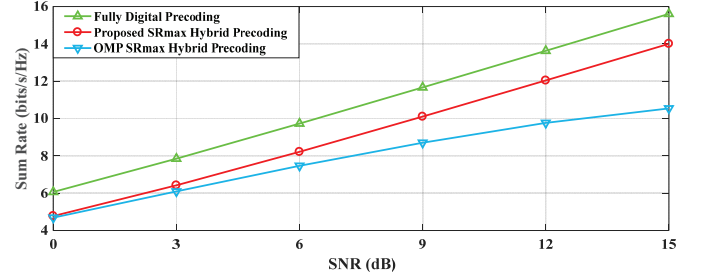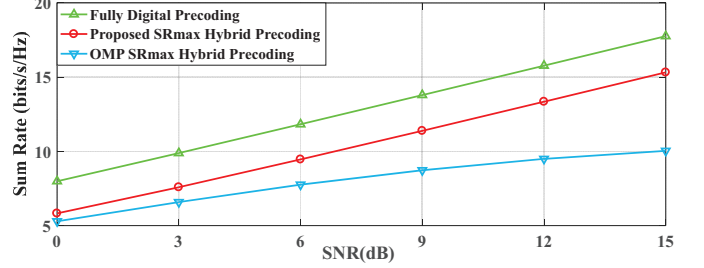

Fig. 5: Sum rate comparison of various hybrid precoders and fully digital precoder, $M = N = 16$, $S = 4$, $K = 4$.

Fig. 7 illustrates the change of the objective χ of (39) versus an increasing λ for two random channel realizations. The target rate of the kth user is set to be the rate achieved by randomly selecting S analog codeword from codebook $\boldsymbol{F}$

(a) $M = N = 16$, $S = 4$, $K = 2$
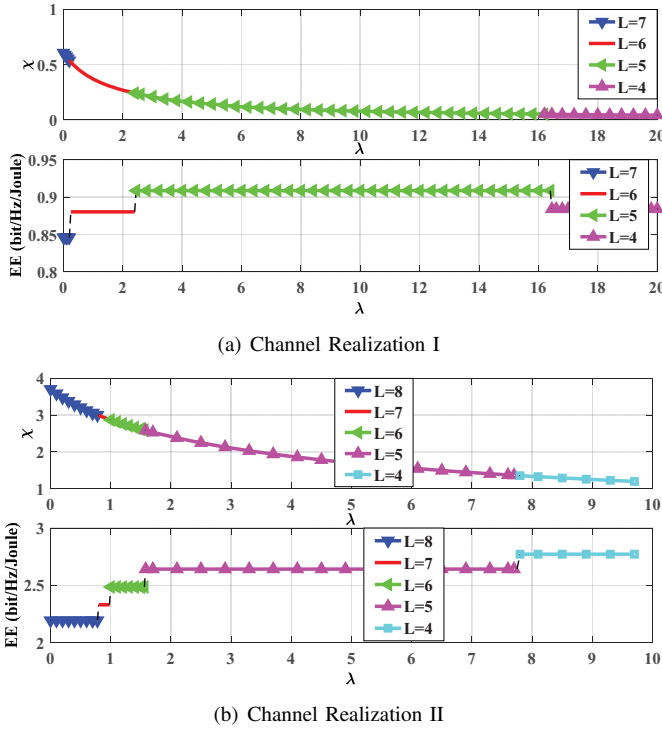

(b) $M = N = 32$, $S = 4$, $K = 2$

Fig. 6: Sum rate comparison of various hybrid precoders and fully digital precoder.

and using the baseband precoder as $\underline{\boldsymbol{G}} = \frac{P}{K} \upsilon_{max}^{(K)} \left(\boldsymbol{H} \underline{\boldsymbol{F}} \right)$. Simulation results show that there exist indeed breaking points of λ, i.e., the number of selected RF chains keep unchanged within a range of λ but suddenly changes at some points. One can observe that the number of selected RF chains decreases with an increasing value of λ. Within a certain interval of λ, the number of selected RF chains keep the same but the objective χ of (39) decreases when λ increases. Table I lists the set of the indices of the selected RF chains corresponding to the channel realization used in Fig. 7. One can observe that the same set of the RF chains (or codewords) are selected for any $\lambda \in \left[\lambda_{L+1}^{key}, \lambda_{L}^{key} \right)$. This observation is consistent with the result in Theorem 3, which has been used in Algorithm 5.

Fig. 8 and Fig. 9 illustrates the SR and EE of various mmWave precoding designs and the fully digital precoding design, respectively. The results are obtained by averaging over 1000 random channel realizations. The target rate of kth user is set to be the rate achieved by selecting randomly S analog codeword from codebook $\boldsymbol{F}$ and defining the baseband precoder as $\underline{\boldsymbol{G}} = \frac{P}{K} \upsilon_{max}^{(K)} \left(\boldsymbol{H} \underline{\boldsymbol{F}} \right)$. It is observed that using the DFT codebook in the proposed EEmax precoder is better than using the 802.15.3c codebook in terms of the EE performance, while the two codebooks lead to the similar SR performance. Compared to the fully digital precoder, the proposed SRmax/EEmax hybrid precoders have certain system SR performance loss as the RF-baseband hybrid architecture may not fully exploit the multi-path diversity gain. The circuit power consumption of the hybrid architecture increases with the number of phase shifter and the number of mixers, which are determined by the number of transmit antennas and the number of RF chains. Therefore, the system EE performance of the hybrid precoding is also determined by the number of transmit antennas and the number of RF chains. One can

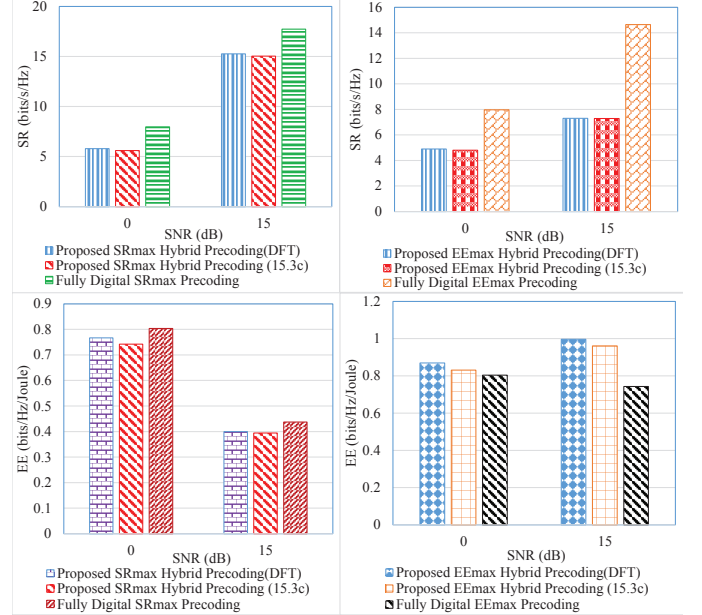TABLE I: Examples of λ, $\mathcal{A}_L$, and L with $M = N = 16$, $S = 8$, $K = 4$, SNR $= 15$dB.

λ	$\mathcal{A}_L$	L	EE (bits/Hz/Joule)	λ	$\mathcal{A}_L$	L	EE (bits/Hz/Joule)
0	$3, 4, 5, 6, 7, 8, 16$	7	0.8460	2.45	$4, 5, 6, 8, 16$	5	0.9087
0.2	$3, 4, 5, 6, 7, 8, 16$	7	0.8460	16.35	$4, 5, 6, 8, 16$	5	0.9087
.25	$4, 5, 6, 7, 8, 16$	6	0.8803	16.4	$4, 5, 6, 16$	4	0.8841
2.4	$4, 5, 6, 7, 8, 16$	6	0.8803	20	$4, 5, 6, 16$	4	0.8841

λ	$\mathcal{A}_L$	L	EE (bits/Hz/Joule)	λ	$\mathcal{A}_L$	L	EE (bits/Hz/Joule)
0	$1, 2, 3, 9, 10, 14, 15, 16$	8	2.1931	1.56	$1, 2, 3, 10, 15, 16$	6	2.4861
0.4	$1, 2, 3, 9, 10, 14, 15, 16$	8	2.1931	1.58	$1, 2, 3, 10, 15$	5	2.6415
0.78	$1, 2, 3, 9, 10, 14, 15, 16$	8	2.1931	5.9	$1, 2, 3, 10, 15$	5	2.6415
0.8	$1, 2, 3, 10, 14, 15, 16$	7	2.3299	7.75	$1, 2, 3, 10, 15$	5	2.6415
0.9	$1, 2, 3, 10, 14, 15, 16$	7	2.3299	7.8	$1, 3, 10, 15$	4	2.7720
0.98	$1, 2, 3, 10, 14, 15, 16$	7	2.3299	8.8	$1, 3, 10, 15$	4	2.7720
1	$1, 2, 3, 10, 15, 16$	6	2.4861	9.8	$1, 3, 10, 15$	4	2.7720
1.3	$1, 2, 3, 10, 15, 16$	6	2.4861	–	–	–	–

(a) Channel Realization I

(b) Channel Realization II

Fig. 7: χ versus value of λ, $M = N = 16$, $S = 8$, $K = 4$, SNR $= 15$dB.

Fig. 8: EE/SR comparison of the hybrid precoders and fully digital precoder, $M = N = 32$, $S = 4$, $K = 2$.

see that different configuration of the number of antennas, RF chains, and users leads to different system EE performance. For example, for the configuration of $M = N = 32$, $S = 4$, and $K = 2$, the EE performance of the hybrid precoders is better than that of the fully digital precoder. Besides, the fully digital precoder leads to a much higher hardware cost ($M = 32$ RF chains versus $S = 8$ RF chains), which is critical for mmWave communication systems using GHz bandwidth and even higher sampling rates.

VII. CONCLUSIONS

In this paper, we considered the design of the hybrid RF-baseband precoding for the downlink of multiuser multi-antenna systems with the aim to maximize the system SR and the system EE. We developed a codebook based RF precoding method and obtained the channel state information via a beam sweep procedure. Exploiting the codebook based design, we simplified the complicated hybrid precoders optimization problems to JWSPD problems. Then, efficient methods were developed to address the JWSPD problems for maximizing the SR and EE of the system. Finally, extensive numerical simulation results are provided to validate the effectiveness of the proposed hybrid precoding design.

REFERENCES

[1] S. Chen, Y. Wang, W. Ma, and J. Chen, "Technical innovations promoting standard evolution: From TD-SCDMA to TD-LTE and beyond," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 60-66, Feb. 2012.

[2] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: EE, estimation and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112-7139, Nov. 2014.

[3] X. Tao, X. Xu, and Q. Cu, "An overview of cooperative communications," *IEEE Commun. Mag.*, pp. 65-71, Jun. 2012.

[4] J. Wang, W. Guan, Y. Huang, R. Schober, and X. You, "Distributed optimization of hierarchical small cell networks: A GNEP framework," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 2, pp. 249-264, Feb. 2017.

[5] R. Heath Jr., N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave
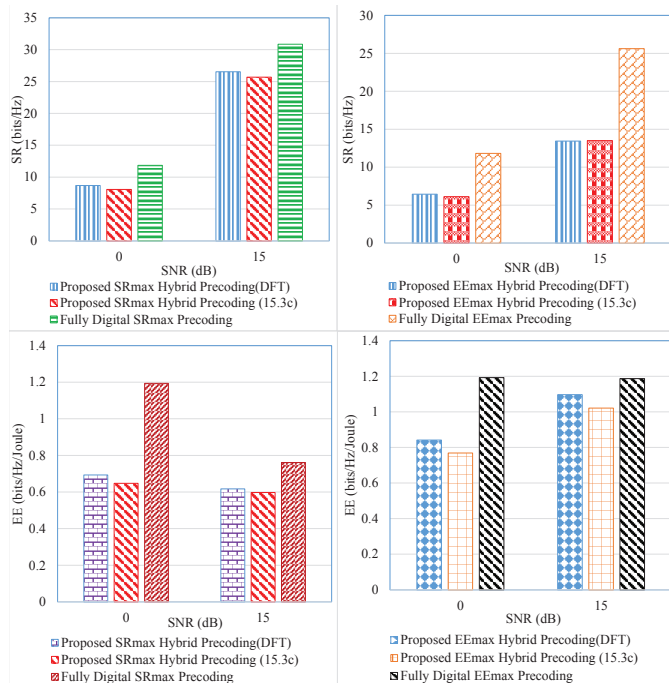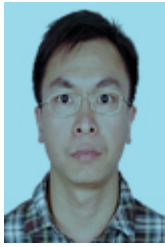
Fig. 9: EE/SR comparison of the hybrid precoders and fully digital precoder, $M = N = 32$, $S = 8$, $K = 4$.

MIMO systems," *IEEE J. Sel. Topics Signal. Process.*, vol. 10, no. 4, pp. 436-453, Apr. 2016.

[6] G. Yang, M. Xiao, J. Gross, H. Al-Zubaidy, Y. Huang, "Analysis of multi-hop outdoor 60 GHz wireless networks with full-duplex buffered relays", *Online: arxiv.org/abs/1608.00913, Aug.2016.*

[7] M. Xiao, S. Mumtaz, Y. Huang, *et al*. "Millimeter wave communications for future mobile networks," *arXiv:1705.06072.*

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

[9] G. MacCartney Jr. *et al.*, "Indoor office wideband millimeter-wave propagation measurements and models at 28 GHz and 73 GHz for ultra-dense 5G wireless networks (Invited Paper)," *IEEE Access*, vol, 3, pp. 2388-2424 2015.

[10] G. Yang, J. Du and M. Xiao, "Maximum throughput path selection with random blockage for indoor 60 GHz relay networks," *IEEE Trans. Commun.*, Vol. 63, No. 10, pp. 3511-3524, Oct. 2015.

[11] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089-1098, Aug. 2004.

[12] X. Zhang, A. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Sig. Proc.*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005.

[13] J. Wang et al., "Beam codebook based beamforming protocol for multi Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390-1399, Aug. 2009.

[14] S. Kutty, and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Suerveys & Tutorials*, vol. 18, no. 2, pp. 949-973, Second Quarter 2016.

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

[16] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal. Process.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.

[17] X. Gao, L. Dai, S. Han, C. I, and R. Heath Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, Apr. 2016.

[18] Q. Xue, X. Fang, M. Xiao, and Y. Li, "Multi-user millimeter wave communications with nonorthogonal beams," *IEEE Trans. on Veh. Tech., Special issue on Tera-Hz Commun.,* Oct. 2016.

[19] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in

massive multiuser MIMO systems," *IEEE Commun. Letter*, vol. 3, no. 6, pp. 653-656, Dec. 2014.

[20] A. Alkhateeb, G. Leus, and R. 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.

[21] S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems, " *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4961-4971, Dec. 2013.

[22] K. Nguyen, L. Tran, O. Tervo, Q. Vu, and M. Juntti, "Achieving energy efficiency fairness in multicell MISO downlink," *IEEE Commun. Lett.,* vol. 19. no. 8, pp. 1426-1429, Aug. 2015.

[23] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A Look at 5G wireless technologies," *IEEE Trans. on Sig. Proc.,* vol. 64, no. 7, pp. 1668-1683, Apr. 2016.

[24] R. Rial, C. Rusu, N.Prelcic, A. Alkhateeb, and R. Heath, Jr, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol, 4, pp. 247-267, 2016.

[25] R. Zi, X. Ge, J. Thompson, C. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chains systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.

[26] "Part 15.3: Wireless MAC and PHY specifications for high rate WPANs. Amendment 2: Millimeter-wave-based alternative physical layer extension," *IEEE Std 802.15.3c-2009*, Oct. 2009.

[27] "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 3: Enhancements for very high throughput in the 60 GHz band" *IEEE Std. 802.11ad-2012*, Dec. 2012.

[28] M. Zhang, P. Smith, and M. Shafi, "An extended one-ring MIMO channel model," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2759-2764, Aug. 2007.

[29] D. Kwan Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618-3631, Sep. 2012.

[30] L. Zhou and Y. Ohashi, "Efficient codebook-based MIMO beamforming for millimeter-wave WLANs," in Proc. *IEEE 23rd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, pp. 1885-1889, 2012.

[31] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[32] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Sig. Proc.*, vol. 61, no. 10, pp. 2660-2674, May. 2013.

[33] Y. Ye, Interior Point Algorithms: Theory and Analysis. New York: Wiley, 1997.

[34] B. Marks and G. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operat. Res.*, vol. 26, no. 4, pp. 681-683, Jul./Aug. 1978.

[35] J. Bibby, "Axiomatisations of the average and a further generalisation of monotonic sequences," *Glasgow Math. J.*, vol. 15, pp. 63-65, 1974.

[36] M. Hanif, L. Tran, A. Tölli, M. Juntti, and S. Glisic, "Efficient solutions for weighted sum rate maximization in multicellular networks with channel uncertainties," *IEEE Trans. Sig. Proc.*, vol. 61, no. 22, pp. 5659-5674, Nov. 2013.

[37] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18-28, Jan. 2004.

[38] S. He, Y. Huang, L. Yang, B. Ottersten, and W. Hong, "Energy efficient coordinated beamforming for multicell system: Duality based algorithm design and massive MIMO transition," *IEEE Trans. on Commun.*, vol. 63, no. 12, pp. 4920-4935, Dec., 2015.

[39] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Sig. Proc.*, vol. 54, no. 1, pp. 161-176, Jan. 2006.

[40] W. Li, Y. Chiang, J. Tsai, H. Yang, *et al*, "60-GHz 5-bit phase shifter with integrated VGA phase-error compensation," *IEEE Trans. on Micro. Theory and Tech.*, vol. 61, no. 3, pp. 1224-1235, Mar. 2013.

Shiwen He (M'14) received the M.S. degree from Chengdu University of Technology, Chengdu, China, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2009 and 2013, respectively. From 2013 to 2015, he was a Postdoctoral Researcher with the State Key Laboratory of Millimeter Waves, Department of Radio Engineering, Southeast University.

Since October 2015, he has been with the School of Information Science and Engineering, Southeast University. He has authored or coauthored over 80 technical publications, held over 25 granted invention patents, and submitted over 20 technical proposals to IEEE standards. From December 2015, he serves as a technical Associate Editor for IEEE 802.11aj. His research interests include multiuser MIMO communication, cooperative communications, energy efficient communications, millimeter wave communication, and optimization theory.

Jiaheng Wang (M'10-SM'14) received the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2010, and the B.E. and M.S. degrees from the Southeast University, Nanjing, China, in 2001 and 2006, respectively.

He is a Full Professor at the National Mobile Communications Research Laboratory (NCRL), Southeast University, Nanjing, China. From 2010 to 2011, he was with the Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. He also held visiting positions at the Friedrich Alexander University Erlangen-Nürnberg, Nürnberg, Germany, and the University of Macau, Macau. His research interests are mainly on optimization in communication systems and wireless networks.

Dr. Wang has published more than 80 articles on international journals and conferences. He serves as an Associate Editor for the IEEE Signal Processing Letters. He is a recipient of the Humboldt Fellowship for Experienced Researchers, and a recipient of the Best Paper Award in WCSP 2014.

Yongming Huang (M'10) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 2007.

Since March 2007, he has been a faculty member with the School of Information Science and Engineering, Southeast University, China, where he is currently a full professor. In 2008-2009, he visited the Signal Processing Laboratory, School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. His current research interests include MIMO wireless communications, cooperative wireless communications and millimeter wave wireless communications. He has published over 200 peer-reviewed papers, hold over 40 invention patents, and submitted over 10 technical contributions to IEEE standards. Since 2012, he has served as an Associate Editor for the IEEE Transactions on Signal Processing, EURASIP Journal on Advances in Signal Processing, and EURASIP Journal on Wireless Communications and Networking. He was elected as the Changjiang Young Scholar of Ministry of Education of China in 2015.

Björn Ottersten (S'87-M'89-SM'99-F'04) was born in Stockholm, Sweden, 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986. In 1989 he received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

Dr. Ottersten has held research positions at the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, and the University of Luxembourg. During 96/97 Dr. Ottersten was Director of Research at ArrayComm Inc, a start-up in San Jose, California based on Ottersten's patented technology. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, and 2013 and 3 IEEE conference papers receiving Best Paper Awards. In 1991 he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm. From 1992 to 2004 he was head of the department for Signals, Sensors, and Systems at KTH and from 2004 to 2008 he was dean of the School of Electrical Engineering at KTH. Currently, Dr. Ottersten is Director for the Interdisciplinary Centre for Security, Reliability and Trust at the University of Luxembourg. Dr. Ottersten is a board member of the Swedish Research Council and as Digital Champion of Luxembourg, he acts as an adviser to the European Commission. Dr. Ottersten has served as Associate Editor for the IEEE Transactions on Signal Processing and on the editorial board of IEEE Signal Processing Magazine. He is currently editor in chief of EURASIP Signal Processing Journal and a member of the editorial boards of EURASIP Journal of Applied Signal Processing and Foundations and Trends in Signal Processing. Dr. Ottersten is a Fellow of the IEEE and EURASIP and a member of the IEEE Signal Processing Society Board of Governors. In 2011 he received the IEEE Signal Processing Society Technical Achievement Award. He is a first recipient of the European Research Council advanced research grant. His research interests include security and trust, reliable wireless communications, and statistical signal processing.

Wei Hong (M'92-SM'07-F'12) received the B.S. degree from the University of Information Engineering, Zhengzhou, China, in 1982, and the M.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 1985 and 1988, respectively, all in radio engineering.

Since 1988, he has been with the State Key Laboratory of Millimeter Waves and serves for the director of the lab since 2003, and is currently a professor and the dean of the School of Information Science and Engineering, Southeast University. In 1993, 1995, 1996, 1997 and 1998, he was a short-term Visiting Scholar with the University of California at Berkeley and at Santa Cruz, respectively. He has been engaged in numerical methods for electromagnetic problems, millimeter wave theory and technology, antennas, RF technology for wireless communications etc. He has authored and co-authored over 300 technical publications and two books. He thrice awarded the first-class Science and Technology Progress Prizes issued by the Ministry of Education of China and Jiangsu Province Government etc. Besides, he also received the Foundations for China Distinguished Young Investigators and for "Innovation Group" issued by NSF of China.

Dr. Hong is a Fellow of IEEE, Fellow of CIE, MTT-S AdCom Member, Vice-Presidents of Microwave Society and Antenna Society of CIE, Chairperson of IEEE MTT-S/AP-S/EMC-S Joint Nanjing Chapter, and served for the associate editor of IEEE Trans. on MTT during 2007-2010, and the editor board members for IJAP, China Communications, and Chinese Science Bulletin etc.