Energy-Efficient Power Allocation and Splitting for MmWave Beamspace MIMO-NOMA with SWIPT

Liangyu Chen, *Graduate Student Member, IEEE,* Bo Hu, *Member, IEEE,* Guixian Xu, and Shanzhi Chen, *Fellow, IEEE*

***Abstract*—Non-orthogonal multiple access (NOMA) has recently been employed into millimeter wave (mmWave) beamspace MIMO to further improve connection density and spectrum efficiency for the emerging cellular internet of things (IoT) with massive IoT sensors and devices. Meanwhile, to achieve energy-efficient green communication, an advanced energy harvesting technique named simultaneous wireless information and power transfer (SWIPT) has been regarded as a promising solution. In this paper, we study the integration of SWIPT in mmWave beamspace MIMO-NOMA system. To improve the energy efficiency while guaranteeing the minimum rate and harvested energy requirements at each IoT node, we analyse the energy efficiency maximization via power allocation at the base station and power splitting at each IoT node. The challenge is that this joint optimization problem is non-convex and intractable due to the presence of intra-beam and inter-beam interference. To tackle it, we propose a two-layer iterative algorithm by leveraging Dinkelbach method and alternating optimization (AO) framework. Specifically, the original fractional objective function is transformed to an equivalent subtractive one via Dinkelbach method at the outer layer, whereas in the inner layer, AO is adopted to address the transformed optimization problem. Numerical results indicate that: 1) the proposed joint power allocation and splitting algorithm can converge within five iterations; 2) the proposed scheme can achieve higher energy efficiency and achievable sum rate compared with the existing scheme and conventional beamspace MIMO with SWIPT.**

***Index Terms*—SWIPT, Internet of Things, NOMA, mmWave, beamspace MIMO, energy efficiency, power allocation, power splitting.**

I. Introduction

With the popularization of Internet of Things (IoT), sensor nodes and IoT devices experience explosive growth. It is estimated that the number of sensor nodes and IoT devices will reach 100 billion by 2025 [1]. In this context, wireless communication networks are envisioned to provide seamless access for such massive IoT. Hence, the cellular IoT has been considered as a key component in next-generation wireless communication networks to support large-scale connectivity of IoT [2]. As a matter of fact, information transmission and communication among IoT nodes consume a large amount of energy resource. However, it should be pointed out that these IoT sensors and devices are mostly battery-powered and thus have limited energy resource. Therefore, energy-efficient green communication has become an important issue for cellular IoT networks. This means that high data rate and low energy consumption should be satisfied simultaneously.

The combination of millimeter wave (mmWave) and massive multiple-input multiple-output (MIMO) is envisioned as an appealing solution for increasing the spectrum range of current networks and accommodating the explosive data traffic requirement. However, the use of massive antennas leads to exceedingly high energy consumption and hardware cost when fully digital signal processing is used in mmWave massive MIMO system, since the number of required RF chains is equal to that of antennas, and RF chains are power-hungry and costly at mmWave frequency [3]. To handle this problem, the recently proposed beamspace MIMO has become a promising solution [4]. By exploiting the lens antenna array (LAA), mmWave space channel can be converted into its sparse beamspace channel. On this basis, a portion of power-focused beams can be selected such that the number of required RF chains is significantly reduced [4]–[6], which constitutes the main feature of beamspace MIMO. Despite its advantages, each beam can support only one user at the same radio resource in conventional beamspace MIMO. This means that the maximum number of served users is no more than that of available beams [4], [5], [7], [8].

To further improve the number of supported users and achieve higher spectrum efficiency, non-orthogonal multiple access (NOMA) has been employed in mmWave massive MIMO [9]–[18]. It has shown that NOMA is more efficient than its orthogonal multiple access (OMA) counterpart, in terms of enhancing the power efficiency and reducing the spectrum usage. In [14], the authors introduced NOMA strategy into an unmanned aerial vehicle enabled mmWave MIMO system, and proposed a joint three-dimensional trajectory and power optimization scheme with the purpose of maximizing the downlink sum rate of the users. In [15], by integrating NOMA with intelligent reflecting surface (IRS) technology, the authors investigated a joint beamforming optimization scheme aiming at maximizing the sum rate of the users. It has shown that NOMA enhanced IRS system is capable of achieving higher system sum rate than orthogonal multiple access (OMA) enabled IRS system. Similar to the scenario in [15], the authors in [16] investigated energy efficiency (EE) maximization optimization problem, and the numerical results

This work was supported by the National Key R&D Program of China under Grant 2020YFB1807900 and the National Natural Science Foundation of China (NSFC) under Grant 61931005. (*Corresponding author* : *Bo Hu*)

L. Chen and B. Hu are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: chenliangyu@bupt.edu.cn; hubo@bupt.edu.cn).

G. Xu is with the Department of Electrical Engineering, Tampere University, Tampere 33720, Finland (e-mail: guixian.xu@tuni.fi).

S. Chen is with the State Key Laboratory of Wireless Mobile Communications, China Academy of Telecommunication Technology, Beijing 100191, China (e-mail: chensz@cict.com).

demonstrated that NOMA enhanced IRS system could also achieve higher EE than OMA based IRS system. Considering a 2-user mmWave NOMA system, the authors in [17] investigated a joint analog beamforming and power allocation scheme with the purpose of maximizing the sum rate of the users. In [18], the authors proposed an optimal user pairing scheme aiming at maximizing the sum rate of the users in a downlink 2-user NOMA network. In fact, compared with the conventional OMA in which at most one user is served by a beam, NOMA supports the communication of multiple users using the same time-frequency resource at the same beam, by exploiting superposition coding at transmitter and successive interference cancellation (SIC) at receiver [19], [20]. In this way, the number of supported users would not be restricted by that of beams. Particularly, NOMA is firstly introduced into mmWave beamspace MIMO in [9], where an iterative power optimization scheme is proposed to enhance the sum rate. Analogously, the authors in [21] investigate a beam selection algorithm to improve the sum rate.

Meanwhile, EE is another essential criterion in wireless communication systems. To this end, simultaneous wireless information and power transfer (SWIPT), which supports the parallel transmission of energy and information, has been regarded as a viable technique to effectively utilize the energy resource while guaranteeing the quality of service (QoS). One popular receiver structure for SWIPT is power splitting (PS) scheme [22], [23]. With the aid of the power splitter, the same RF signal can be simultaneously divided into two parts for energy harvesting (EH) and information detection (ID), which constitutes the essential idea of SWIPT. In fact, SWIPT technique has an immense potential for the multi-user systems. By employing SWIPT technique, the interference power can be converted into energy, which can be exploited to prolong the service life of energy-constrained IoT sensors and devices. However, interference channel reduces the transmission rate and thus has an adverse impact on ID. Hence, how to trade off EH and ID to improve the system performances is a challenge for a SWIPT-enabled system. To deal with this challenge, some recent research efforts have been conducted from different perspectives [24]–[29]. In [24], the authors propose a joint power allocation and splitting approach, aiming to maximize the minimum signal to interference plus noise ratio (SINR) of all users in a SWIPT-based multi-link interference channel system. In [25], the authors consider a MISO-SWIPT system, where beamforming and PS are jointly designed to maximize EE. Meanwhile, a zero-forcing beamforming scheme is adopted to reduce the computational complexity. In [26], the authors consider a multi-user MIMO-SWIPT system, where the joint optimization of beamforming and PS ratio is proposed to minimize the total transmit power, subject to the EH constraint and the SINR constraint. Besides, SWIPT technique has been investigated in NOMA systems. Considering a SWIPT-NOMA system with single antenna transceiver, the authors in [28] propose a joint power allocation and PS design with the consideration of the tradeoff between harvested energy and transmission rate. In [29], the authors study a SWIPT enabled mmWave Massive MIMO-NOMA system. They propose an effective user grouping approach and hybrid precoding scheme. Then, a joint power allocation and splitting scheme is designed to maximize the sum rate.

Observing from the existing literature, we notice that although SWIPT has the potential to achieve the energy-efficient green communication, the application of SWIPT in mmWave beamspace MIMO-NOMA is still an open problem to the best of our knowledge. [9] and [21] do not consider the SWIPT, and [24]–[29] do not consider the structure of the beamspace MIMO. In addition, it is worth noting that although the EE of mmWave beamspace MIMO-NOMA system has been considered in [9], their proposed power optimization approach is designed with sum rate maximization as metric. Therefore, the existing schemes can not be directly applied to the considered mmWave beamspace MIMO-NOMA with SWIPT system, which ought to be redesigned jointly with precoding, power allocation and power splitting.

From the perspective of practical application, we think that our considered system model has the potential to be applied in typical IoT networks with massive IoT sensors and devices. This can be explained by the following reasons. First, with the rapidly increasing of IoT sensors and devices, it is very difficult and costly to manually change the batteries of numerous IoT nodes. By utilizing SWIPT technique, IoT nodes can simultaneously harvest the energy and decode the information from received RF signal, and thus it is convenient to recharge the wireless-enabled IoT nodes. Second, mmWave communication in combination with beamspace MIMO structure can effectively improve the system capacity, and thus it is beneficial for data transmission and communication among IoT nodes. Third, by employing NOMA scheme, multiple IoT sensors and devices can be served by one beam at the same radio resource, and thus both spectrum efficiency and connection density can be further improved. On the other hand, the huge number of IoT nodes will lead to rapidly increasing energy consumption. Thus, it is necessary to provide novel energy-efficient optimization design for the considered system, which motivates us to develop this paper.

A. Main Contributions

Unlike the previous works, this paper studies the mmWave beamspace MIMO-NOMA with SWIPT system, where the power splitter is employed at each IoT node such that each IoT node can simultaneously obtain energy and information from the same received RF signals. A novel energy-efficient power allocation and splitting design is proposed to maximize EE while guaranteeing each IoT node' QoS demands in regard to energy harvesting and transmission rate. The primary contributions are summarized as follows:

1) We propose to integrate SWIPT in mmWave beamspace MIMO-NOMA system to achieve energy-efficient green communication. In order to suppress the inter-beam interference, a zero-forcing based precoding design is developed. On this basis, we formulate the EE maximization problem via power allocation at the BS and power splitting at each IoT node. Meanwhile, we consider the maximum transmit power limitation at the BS, the minimum rate limitation and harvested energy

TABLE I
TABLE OF NOTATIONS

Notations	Definition
N_{TX}	Number of BS antennas
N_{RF}	Number of BS RF chains
U	Number of users
$\mathcal{U}_g$	Set of NOMA users served by the g-th beam
$\mathbf{h}_u$	Spatial channel vector for user u
$\tilde{\mathbf{h}}_k$	Beamspace channel vector for user k
$\hat{\mathbf{h}}_{g,u}$	Beamspace channel vector for the u-th user in the g-th beam after beam selection
$x_{g,i}$	Transmitted signal with normalized power for the i-th user in the g-th beam
$z_{g,u}$	Noise term
$n_{g,u}$	Additive noise caused by ID
$p_{g,i}$	Transmitted power for the i-th user in the g-th beam
$\rho_{g,u}$	PS ratio for u-th user in g-th beam
$\mathbf{U}$	Spatial discrete fourier transformation matrix
$\hat{\mathbf{v}}_g$	Precoding vector for the g-th beam
$\mathbf{v}_g$	Normalized precoding vector for the g-th beam
$\mathbf{a}(\varphi_u)$	Array steering vector corresponding to the spatial direction φ_u
$\tau_u^{(0)}$	The complex gain of LoS component for user u
$\tau_u^{(i)}$	The i-th complex gain of NLoS component for user u

constraint at each IoT node. The formulated problem is presented as a non-convex and non-linear optimization problem.

2) We propose a two-layer iterative approach to handle this challenging optimization problem by applying Dinkelbach method and alternating optimization (AO) method. Specifically, in the outer layer, the EE-based fractional objective function is converted to a subtractive form by means of Dinkelbach method. In the inner layer, AO method is employed to address the remaining optimization problem.
3) Simulations results validate the convergence and effectiveness of the proposed energy-efficient power allocation and splitting method. It demonstrates that our proposed method can converge within five iterations. Meanwhile, considering the same system model, i.e., beamspace MIMO-NOMA with SWIPT system, our proposed optimization method can achieve significant enhancement of EE and achievable sum rate (ASR) performance than the existing optimization scheme. In addition, compared with the conventional beamspace MIMO with SWIPT, the proposed beamspace MIMO-NOMA with SWIPT can achieve higher EE and ASR performance.

B. Organization and Notations

The remainder of this paper is organized as follows. Section II presents the system model and formulates the joint power allocation and splitting optimization problem. Solutions to this joint optimization problem are provided in Section III. Our numerical results are presented in Section IV. This paper is concluded in Section V.

The following notations are employed throughout the paper. a, $\mathbf{a}$, $\mathbf{A}$, $\mathcal{A}$ denote a scalar, vector, matrix and set, respectively. $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ represent transpose, Hermitian transpose, and matrix inversion, respectively. $\|\cdot\|_n$ denotes n norm. $\mathbb{E}(\cdot)$ is expectation operator. $|\mathcal{A}|$ means the number of elements in set $\mathcal{A}$. $\mathbf{A}(b,:)$ and $\mathbf{A}(:,c)$ represent the bth row and cth column of matrix $\mathbf{A}$, respectively. $\mathbf{I}_K$ represents the $K \times K$ identity matrix.

II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1, we consider a downlink single-cell mmWave beamspace MIMO-NOMA with SWIPT system, in which one base station (BS) employs N_{TX}-element LAA, which is illustrated in Fig. 2, and N_{RF} RF chains to serve U single-antenna IoT nodes that equip power splitters. In the following, we first present mmWave channel model. Then, we introduce beamspace MIMO-NOMA with SWIPT system. Finally, we formulate the energy-efficient power allocation and splitting optimization problem.

A. MmWave Channel Model

We characterize mmWave spatial channel by applying the widely-adopted Saleh-Valenzuela model [4]:

$$\mathbf{h}_u = \tau_u^{(0)}\mathbf{a}\big(\varphi_u^{(0)}\big) + \sum_{i=1}^{I}\tau_u^{(i)}\mathbf{a}\big(\varphi_u^{(i)}\big), \tag{1}$$

where $\tau_u^{(0)}\mathbf{a}\big(\varphi_u^{(0)}\big)$ denotes the line-of-sight (LoS) component of user u. $\tau_u^{(0)}$ is the complex gain while $\mathbf{a}\big(\varphi_u^{(0)}\big) \in \mathbb{C}^{N_{TX}\times 1}$ is the array steering vector corresponding to the spatial direction $\varphi_u^{(0)}$. $\tau_u^{(i)}\mathbf{a}\big(\varphi_u^{(i)}\big)$ represents the ith non-line-of-sight (NLoS) component with complex gain $\tau_u^{(i)}$, and I represents the total number of NLoS components. For a typical uniform linear array (ULA), $\mathbf{a}\big(\varphi\big)$ can be denoted by

$$\mathbf{a}\big(\varphi\big) = \frac{1}{\sqrt{N_{TX}}}\left[1, e^{-j2\pi\varphi}, \cdots, e^{-j2\pi\varphi(N_{TX}-1)}\right]^T \tag{2}$$

where $\varphi = \frac{d}{\lambda}\sin\phi$ denotes the spatial direction, d denoting the antenna spacing, λ denoting the signal wavelength, and ϕ $(-\frac{\pi}{2} \leqslant \phi \leqslant \frac{\pi}{2})$ denoting the physical direction.

B. Beamspace MIMO-NOMA with SWIPT

In this work, we employ an LAA at the BS in order to reduce the number of required RF chains [4], [5]. Mathematically, the LAA as a $N_{TX} \times N_{TX}$ spatial discrete fourier transformation matrix $\mathbf{U}$, in which the N_{TX} rows of $\mathbf{U}$ contain the N_{TX} orthogonal steering vectors [30], [31], namely,

$$\mathbf{U} = \left[\mathbf{a}\big(\tilde{\theta}_0\big), \mathbf{a}\big(\tilde{\theta}_1\big), \cdots, \mathbf{a}\big(\tilde{\theta}_{N_{TX}-1}\big)\right], \tag{3}$$

where $\mathbf{a}\big(\tilde{\theta}\big) \in \mathbb{C}^{N_{TX}\times 1}$ represents the array steering vector for predefined spatial direction $\tilde{\theta}$. The spatial directions covering the entire space can be expressed as $\tilde{\theta}_k = \frac{1}{N_{TX}}\big(k - \frac{N_{TX}+1}{2}\big)$ for k varying from 0 to $N_{TX}-1$. By exploiting $\mathbf{U}$, the spatial channel matrix $\mathbf{H}$ can be converted into the following beamspace channel matrix $\tilde{\mathbf{H}}$:

$$\tilde{\mathbf{H}} = \mathbf{U}\mathbf{H} = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_U\right], \tag{4}$$

where $\tilde{\mathbf{h}}_k = \mathbf{U}\mathbf{h}_k \in \mathbb{C}^{N_{TX}} \times 1$ for k varying from 1 to U is beamspace channel vector. $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_U] \in \mathbb{C}^{N_{TX}\times U}$ is

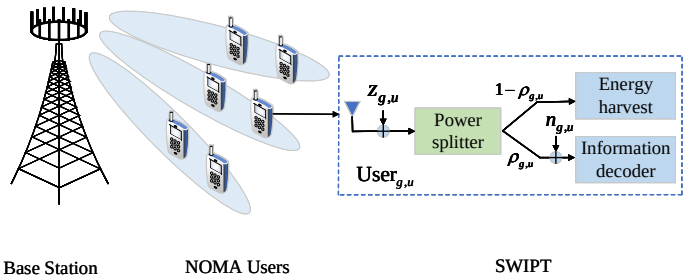


Fig. 1. The illustration of mmWave beamspace MIMO-NOMA with SWIPT system, where an LAA is deployed at the base station as shown in Fig. 2 and the power splitter is employed at each IoT node to support SWIPT.

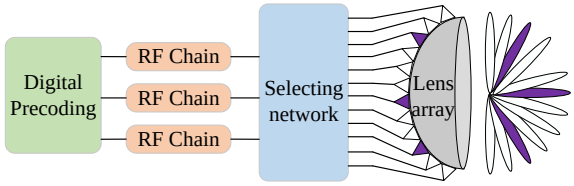


Fig. 2. LAA at the Base Station.

spatial channel matrix. $\mathbf{h}_k$ for k varying from 1 to U is spatial channel vector, which is represented by the formula (1).

Since there are a limited number of scatters in mmWave propagation, the number of NLoS components is much less than that of antenna elements [5]. This means that $\tilde{\mathbf{h}}_k$ has a sparse structure [7]. Thus, the dominant beams can be selected from $\tilde{\mathbf{h}}_k$ to design a dimension-reduced beamspace. In this way, the number of required RF chains is significantly reduced. Accordingly, N_{RF} maximum beams can be selected from $\overline{\mathbf{H}}$, i.e., $\overline{\mathbf{H}} = \tilde{\mathbf{H}}(q,:)_{q\in C}$ [5], [7], where $\overline{\mathbf{H}} \in N_{RF} \times U$ and C is the set of selected beams satisfying $|C| = N_{RF} < N_{TX}$.

In general, for a typical beamspace MIMO system, the received signal vector at users can be given by

$$\mathbf{y} = \overline{\mathbf{H}}^H \mathbf{V}\mathbf{P}\mathbf{x} + \mathbf{z}, \tag{5}$$

where $\mathbf{V} = \left[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_U\right]$ is the dimension-reduced precoding matrix of size $N_{RF} \times U$ with $\|\mathbf{v}_u\|_2 = 1$ for u varying from 1 to U. $\mathbf{P} = \mathrm{diag}\left[\sqrt{p_1}, \sqrt{p_2}, \cdots, \sqrt{p_U}\right]$ denotes the $U \times U$ power matrix satisfying $\sum_{u=1}^{U} p_u \le P_{max}$, where P_{max} is the maximum transmit power of the BS. $\mathbf{x} = \left[x_1, x_2, \cdots, x_U\right]$ is the $U \times 1$ transmitted signal vector for all U users with $\mathbb{E}\left(\mathbf{x}\mathbf{x}^H\right) = \mathbf{I}_K$. $\mathbf{z}$ denotes the $U \times 1$ noise vector following the distribution $\mathcal{CN}\left(\mathbf{0}, \delta_0^2 \mathbf{I}_K\right)$.

In this study, we focus on a structure with $U > N_{RF}$, which means that more than one user may be served by one beam. Inspired by [9], we adopt NOMA protocol in the beamspace MIMO system. A beam selection scheme based on maximum magnitude [5], [9] is employed to support multiple users, and each beam corresponds to a RF chain. For notational convenience, let $\mathcal{U}_g = \left\{1, 2, \cdots, \left|U_g\right|\right\}$ be the user set in the gth beam for g varying from 1 to G ($G = N_{RF}$), where $\mathcal{U}_g \cap \mathcal{U}_n = \varnothing$ ($g \neq n$) and $U = \sum_{g=1}^{G} \left|U_g\right|$. In the same beam, the beamspace channel vectors of the served users have some correlations [9]. Based on this idea, the beamspace channel vector with the largest norm can be selected as the representative channel vector for each beam. Therefore, we obtain a $N_{RF} \times N_{RF}$ representative channel matrix $\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \cdots, \hat{\mathbf{h}}_{N_{RF}}\right]$, where $\hat{\mathbf{h}}_g$ denotes the beamspace channel vector that has largest norm in gth beam. Then, with the aid of zero-forcing precoding technique, we derive the $N_{RF} \times N_{RF}$ precoding matrix $\hat{\mathbf{V}}$, which can be given by

$$\hat{\mathbf{V}} = \left[\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \cdots, \hat{\mathbf{v}}_{N_{RF}}\right] = \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^H \hat{\mathbf{H}}\right)^{-1}, \tag{6}$$

where $\hat{\mathbf{v}}_g$ denotes the $G \times 1$ precoding vector for the gth beam. Considering the transmit power has been separated from the precoding vector, we need to further normalize the precoding vectors. For the g-th beam, the normalized precoding vector can be written by $\mathbf{v}_g = \frac{\hat{\mathbf{v}}_g}{\|\hat{\mathbf{v}}_g\|_2}$, for g varying from 1 to N_{RF}. Then, for u-th user in g-th beam, its received signal can be expressed as

$$y_{g,u} = \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g s_g + \hat{\mathbf{h}}_{g,u}^H \sum_{j \neq g}^{N_{RF}} \mathbf{v}_j s_j + z_{g,u}, \tag{7}$$

where $s_g = \sum_{i=1}^{|U_g|} \sqrt{p_{g,i}} x_{g,i}$ is the superposed transmitted signals for $|U_g|$ users within the g-th beam and $x_{g,i}$ is transmitted signal with $\mathbb{E}\left(|x_{g,i}|^2\right) = 1$. $p_{g,i}$ is transmission power for i-th user in g-th beam. $\mathbf{v}_g \in \mathbb{C}^{N_{RF} \times 1}$ is the precoding vector for the gth beam with $\left\|\mathbf{v}_g\right\|_2 = 1$. $\hat{\mathbf{h}}_{g,u} \in \mathbb{C}^{N_{RF} \times 1}$ is the beamspace channel vector after beam selection. $z_{g,u}$ denotes noise term which follows the distribution $\mathcal{CN}\left(0, \delta_0^2\right)$.

In addition, we assume that each IoT user node can perform SWIPT by using PS technique, and thus the received signal can be divided into the EH and ID [25]. Let $\rho_{g,u}$ ($0 < \rho_{g,u} < 1$) be the PS ratio for u-th user in g-th beam, and its corresponding signal for EH can be given by

$$y_{g,u}^{EH} = \sqrt{1 - \rho_{g,u}}\, y_{g,u}, \tag{8}$$

and the harvested energy can be given by

$$E_{g,u}^{EH} = \varepsilon_{g,u}(1 - \rho_{g,u}) \left(\sum_{i=1}^{N_{RF}} \sum_{j=1}^{|U_g|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_i\right\|_2^2 p_{i,j} + \delta_0^2 \right), \tag{9}$$

where $\varepsilon_{g,u} \in (0, 1]$ represents energy conversion efficiency for EH. Moreover, we can express the signal for ID by the following form:

$$y_{g,u}^{ID} = \sqrt{\rho_{g,u}}\, y_{g,m} + n_{g,u}, \tag{10}$$

where $n_{g,u} \sim \mathcal{CN}(0, \delta_1^2)$ is the additional noise caused by ID.

According to the NOMA principle, superposition coding and SIC are performed at the transmitter and the receiver, respectively. For the users within the g-th beam, we assume that $\left\|\hat{\mathbf{h}}_{g,1}^H \mathbf{v}_g\right\|_2 \ge \left\|\hat{\mathbf{h}}_{g,2}^H \mathbf{v}_g\right\|_2 \ge \cdots \ge \left\|\hat{\mathbf{h}}_{g,|U_g|}^H \mathbf{v}_g\right\|_2$. By exploiting SIC at users in the g-th beam, the u-th user can decode the m-th user ($u + 1 \le m \le |U_g|$) and remove them from the received superposed signals. Accordingly, we can

adopt the following form to characterize the received signal for ID by u-th user in g-th beam:

$$y_{g,u}^{ID} = \sqrt{\rho_{g,u}}\Bigg(\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g \sqrt{p_{g,u}} x_{g,u} + \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g \sum_{m=1}^{u-1} \sqrt{p_{g,m}} x_{g,m} + \hat{\mathbf{h}}_{g,u}^H \sum_{n\neq g} \sum_{m=1}^{|U_n|} \mathbf{v}_n \sqrt{p_{n,m}} x_{n,m} + z_{g,u}\Bigg) + n_{g,u}. \quad (11)$$

Based on (11), the SINR for u-th user in g-th beam can be expressed by

$$\gamma_{g,u} = \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}}, \quad (12)$$

where $I_1 = \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m}$ is the intra-beam interference, and $I_2 = \sum_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m}$ is the inter-beam interference. Therefore, for u-th user in g-th beam, its achievable rate can be expressed as

$$R_{g,u} = \log_2\left(1 + \gamma_{g,u}\right). \quad (13)$$

C. Problem Formulation

In this subsection, we formulate the EE maximization problem via power allocation at the BS and power splitting at each user receiver. Meanwhile, we consider the maximum transmit power limitation at the BS, the minimum data rate limitation and the minimum harvested energy constraints at each user node. We define the EE as the ratio of achievable sum rate to the total power consumption [32], namely,

$$EE = \frac{\sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\{p_{g,u}\}, \{\rho_{g,u}\})}{P_t(\{p_{g,u}\}) + P_c} \text{ (bps/Hz/W)}. \quad (14)$$

where $\sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\{p_{g,u}\}, \{\rho_{g,u}\})$ is achievable sum rate of all users. $P_t(\{p_{g,u}\}) = \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} p_{g,u}$ is the total transmitted power. $P_c = P_{BB} + N_{RF}P_{SW} + N_{RF}P_{RF}$ is the circuit power consumption, where P_{BB}, P_{SW}, and P_{RF} represent the power consumption caused by baseband, beam switch and RF chain [5], respectively.

Therefore, the EE maximization optimization problem can be formulated as

$$\max_{\{\mathbf{P},\boldsymbol{\rho}\}} EE \quad (15a)$$

$$\text{s.t.} \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} p_{g,u} \leq P_{max}, \quad (15b)$$

$$p_{g,u} \geq 0, \forall g, u, \quad (15c)$$

$$0 < \rho_{g,u} < 1, \forall g, u, \quad (15d)$$

$$R_{g,u} \geq R_{min}, \forall g, u, \quad (15e)$$

$$E_{g,u}^{EH} \geq E_{min}, \forall g, u, \quad (15f)$$

where $\mathbf{P} = \{p_{g,u}\}$ and $\boldsymbol{\rho} = \{\rho_{g,u}\}$ denote the power matrix and PS ratio matrix, respectively. (15b) is the transmit power limitation of the BS. (15c) demonstrates the transmit power for each user node should be positive. (15d) is the PS ratio constraint. (15e) means that the minimum data rate for each user node should be guaranteed. (15f) describes the minimum energy harvesting requirement for each user node. Due to the presence of both intra- and inter-beam interference, the objective function (15a) and constraints in (15e)-(15f) are non-convex. As a result, problem (15) is a non-convex programming problem. In general, it is challenging to derive the optimal solution to (15) in polynomial time. Next, to find a feasible solution to problem (15), we propose an effective iterative approach.

D. Feasibility Analysis of Problem (15)

Before demonstrating our solution, we study the feasibility of problem (15). We first give the following lemma.

Lemma 1: Problem (15) is feasible if and only if the following problem is feasible.

$$\text{find} \quad \{\mathbf{P}, \boldsymbol{\rho}\} \quad (16a)$$

$$\text{s.t.} \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}} \geq 2^{R_{min}} - 1, \forall g, u, \quad (16b)$$

$$(15b) - (15d) \quad (16c)$$

Proof 1: Please refer to Appendix A.

Lemma 1 demonstrates that the feasibility of problem (15) is independent of the energy harvested constraint (15f). Furthermore, the following Lemma shows the feasibility of problem (15) is also independent of the PS ratio.

Lemma 2: Problem (16) is feasible if and only if the following problem is feasible.

$$\text{find} \quad \{\mathbf{P}\} \quad (17a)$$

$$\text{s.t.} \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1 + I_2 + \delta_0^2 + \delta_1^2} \geq 2^{R_{min}} - 1, \forall g, u, \quad (17b)$$

$$(15b) - (15c) \quad (17c)$$

Proof 2: Please refer to Appendix B.

Due to the total transmit power constraints of the BS and the SINR constraint of each user, problem (17) may not be always feasible. To elaborate the feasibility conditions of problem (17), we give the following proposition.

Proposition 1: The feasibility of problem (17) can be checked by solving the following problem

$$\min_{\{\mathbf{P}\}} \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} p_{g,u} \quad (18a)$$

$$\text{s.t.} \quad (17b) \quad (18b)$$

The detailed proof of Proposition 1 can refer to [33]. Here, (18a) is a power minimization optimization problem, which can be efficiently addressed by existing optimization tool. Therefore, the feasibility of problem (15) can be simply verified by checking the feasibility of problem (17). Furthermore, the feasibility conditions of problem (17) are given in Proposition 1. Without loss of generality, we assume that the problem (15) is feasible.

III. SOLUTION TO THE JOINT OPTIMIZATION PROBLEM

To make the problem (15) tractable, we first transform the original fractional objective function to a subtractive form by leveraging Dinkelbach approach. Then, for the inner-layer optimization problem with given energy efficiency parameter η, we propose an AO based iterative method. Finally, a two-layer iterative method is proposed to derive the feasible solution to problem (15).

A. Problem Transformation

We define the maximum EE of the system as

$$\begin{aligned}\eta^* &= \frac{\sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}^*, \boldsymbol{\rho}^*)}{P_t(\mathbf{P}^*) + P_c} \\ &= \max_{\{\mathbf{P},\boldsymbol{\rho}\}} \frac{\sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}, \boldsymbol{\rho})}{P_t(\mathbf{P}) + P_c}, \end{aligned} \tag{19}$$

where $\mathbf{P}^* = \{p_{g,u}^*\}$ and $\boldsymbol{\rho}^* = \{\rho_{g,u}^*\}$ represent power matrix and PS ratio matrix corresponding to η^*. Then, we introduce the following theorem.

Theorem 1: The maximum EE η^* is derived if and only if

$$\begin{aligned}&\max_{\{\mathbf{P},\boldsymbol{\rho}\}} \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}, \boldsymbol{\rho}) - \eta^* \big(P_t(\mathbf{P}) + P_c\big) \\ &= \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}^*, \boldsymbol{\rho}^*) - \eta^* \big(P_t(\mathbf{P}^*) + P_c\big) \\ &= 0. \end{aligned} \tag{20}$$

Proof 3: Please refer to [34].

According to Theorem 1, we can reformulate (15) as the following form

$$\max_{\{\mathbf{P},\boldsymbol{\rho}\}} \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}, \boldsymbol{\rho}) - \eta \big(P_t(\mathbf{P}) + P_c\big) \tag{21a}$$

$$\text{s.t. } (15b) - (15f). \tag{21b}$$

To obtain the maximum EE η^*, we propose a Dinkelbach-based iterative approach, which is summarized in **Algorithm 1**. After each iteration of **Algorithm 1**, the EE η can be further improved. The convergence proof of **Algorithm 1** is similar to [34]. Next, we handle the converted problem (21) with fixed η. However, since objective function (21a) and constraints in (15e)-(15f) are nonconvex, this problem is still intractable to address.

B. AO-Based Iterative Approach for Solving (21)

Based on a given energy efficiency factor η, we need to further tackle the inner-layer optimization problem. To this end, we propose an efficient iterative optimization approach to deal with the considered problem, which will be discussed in details in the following.

Algorithm 1: Dinkelbach-Based Iterative Approach

Initialize the maximum iterations $\triangle_1$, the maximum tolerance ε_1, iteration index $v = 0$, and the maximum EE $\eta = 0$.
repeat
  Solve problem (21) for a given η_v and derive power allocation and PS ratio $\{\mathbf{P}, \boldsymbol{\rho}\}$.
  Compute $\varepsilon_1^* = \sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}, \boldsymbol{\rho}) - \eta_v \big(P_t(\mathbf{P}) + P_c\big)$.
  if $\varepsilon_1^* < \varepsilon_1$ **then**
    Convergence=**true**.
    return $\{\mathbf{P}^*, \boldsymbol{\rho}^*\} = \{\mathbf{P}, \boldsymbol{\rho}\}$ and $\eta^* = \frac{\sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}, \boldsymbol{\rho})}{P_t(\mathbf{P}) + P_c}$.
  else
    Set $\eta_v = \frac{\sum_{g=1}^{N_{RF}} \sum_{u=1}^{|U_g|} R_{g,u}(\mathbf{P}, \boldsymbol{\rho})}{P_t(\mathbf{P}) + P_c}$ and $v = v + 1$.
    Convergence=**false**.
  end if
until Convergence=**true** or $v = \triangle_1$;

We first convert (21) into a more tractable and solvable form. Let

$$\begin{aligned}\tilde{y}_{g,u} = \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g \sqrt{p_{g,u}} x_{g,u} + \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g \sum_{m=1}^{u-1} \sqrt{p_{g,m}} x_{g,m} \\ + \hat{\mathbf{h}}_{g,u}^H \sum_{n \neq g} \sum_{m=1}^{|U_n|} \mathbf{v}_n \sqrt{p_{n,m}} x_{n,m} + z_{g,u} + \frac{n_{g,u}}{\sqrt{\rho_{g,u}}}. \end{aligned} \tag{22}$$

When mean square error (MSE) is utilized to detect $x_{g,u}$ from $\tilde{y}_{g,u}$ in (22), we can formulate the following detection problem

$$\xi_{g,u}^{opt} = \arg\min_{\{\xi_{g,u}\}} e_{g,u}, \quad \forall g, u, \tag{23}$$

where $\xi_{g,u}$ is the receiver filter at the u-th user within the g-th beam. $e_{g,u}$ is the corresponding MSE, which can be represented by

$$e_{g,u} = \mathbb{E}\big[|x_{g,u} - \xi_{g,u} \tilde{y}_{g,u}|^2\big], \quad \forall g, u. \tag{24}$$

Substituting (22) into (24), the MSE in (24), i.e., $e_{g,u}$, can be rewritten as

$$e_{g,u} = 1 + |\xi_{g,u}|^2 \Xi_{g,u} - 2\mathrm{Re}\left(\xi_{\mathrm{g,u}} \sqrt{\mathrm{p_{g,u}}} \hat{\mathbf{h}}_{\mathrm{g,u}}^{\mathrm{H}} \mathbf{v}_{\mathrm{g}}\right), \quad \forall \mathrm{g, u}, \tag{25}$$

where $\Xi_{g,u} = p_{g,u} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 + I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}$. Then, by solving problem (23), the optimal $\xi_{g,u}$ can be derived as follows:

$$\begin{aligned}\xi_{g,u}^{opt} &= \arg\min_{\{\xi_{g,u}\}} e_{g,u} \\ &= \left(\sqrt{p_{g,u}} \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right)^* \left(\Xi_{g,u}\right)^{-1}, \quad \forall g, u. \end{aligned} \tag{26}$$

Substituting (26) into (25), the minimum MSE corresponding to $\xi_{g,u}^{opt}$ can be given by

$$e_{g,u}^{opt} = 1 - p_{g,u} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \Xi_{g,u}^{-1}, \quad \forall g, u. \tag{27}$$

The detailed derivations to (26) and (27) are presented in Appendix C.

In addition, according to (12), we can derive the following equations

$$\left(1+\frac{\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1+I_2+\delta_0^2+\frac{\delta_1^2}{\rho_{g,u}}}\right)^{-1} = \left(I_1+I_2+\delta_0^2+\frac{\delta_1^2}{\rho_{g,u}}\right)\Xi_{g,u}^{-1} = 1-p_{g,u}\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2\Xi_{g,u}^{-1}. \quad (28)$$

One can observe that there are the equivalent relations between (27) and (28), namely,

$$\left(1+\frac{\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1+I_2+\delta_0^2+\frac{\delta_1^2}{\rho_{g,u}}}\right)^{-1} = \min_{\{\xi_{g,u}\}} e_{g,u}, \quad \forall g,u. \quad (29)$$

As a result, the achievable rate can be reformulated as

$$R_{g,u}(\mathbf{P},\boldsymbol{\rho}) = \log_2\left(1+\frac{\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1+I_2+\delta_0^2+\frac{\delta_1^2}{\rho_{g,u}}}\right) = \max_{\{\xi_{g,u}\}}\left(-\log_2\left(e_{g,u}\right)\right), \quad \forall g,u. \quad (30)$$

After removing the log function in (30), $R_{g,u}(\mathbf{P},\boldsymbol{\rho})$ can be further rewritten as

$$R_{g,u}(\mathbf{P},\boldsymbol{\rho}) = \max_{\{\xi_{g,u},\upsilon_{g,u}\}}\left(-\frac{\upsilon_{g,u}e_{g,u}}{\ln 2}+\log_2\upsilon_{g,u}+\frac{1}{\ln 2}\right), \forall g,u. \quad (31)$$

The detailed derivations to (31) are presented in Appendix D.

Replacing the $R_{g,u}(\mathbf{P},\boldsymbol{\rho})$ of objective function (21a) with (31), problem (21) is equivalently reformulated as

$$\begin{aligned}\max_{\{\mathbf{P},\boldsymbol{\rho},\xi_{g,u},\upsilon_{g,u}\}} & \sum_{g=1}^{N_{RF}}\sum_{u=1}^{|U_g|}\left(-\frac{\upsilon_{g,u}e_{g,u}}{\ln 2}+\log_2\upsilon_{g,u}+\frac{1}{\ln 2}\right) \\ & -\eta\left(P_t(\mathbf{P})+P_c\right) \\ \text{s.t. } & (15b)-(15f). \end{aligned} \quad (32)$$

To effectively solve problem (32), we propose an AO-based iterative method. Specifically, the variables $\{\mathbf{P}\}$, $\{\boldsymbol{\rho}\}$, $\{\xi_{g,u}\}$, and $\{\upsilon_{g,u}\}$ are alternately optimized at each iteration.

For given $\{\mathbf{P}^{r-1}\}$ and $\{\boldsymbol{\rho}^{r-1}\}$ in the $(r-1)$th iteration, the optimal $\{\xi_{g,u}^r\}$ can be obtained at the rth iteration according to (26), namely,

$$\xi_{g,u}^r = \left(\sqrt{p_{g,u}^{r-1}}\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right)^*\left(\Xi_{g,u}^{r-1}\right)^{-1}, \quad \forall g,u, \quad (33)$$

where

$$\Xi_{g,u}^{r-1} = p_{g,u}^{r-1}\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2+\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2\sum_{m=1}^{u-1}p_{g,m}^{r-1}+\delta_0^2+\frac{\delta_1^2}{\rho_{g,u}^{r-1}} + \sum_{n\neq g}\sum_{m=1}^{|U_n|}\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_n\right\|_2^2 p_{n,m}^{r-1}, \quad \forall g,u. \quad (34)$$

Meanwhile, the optimal $e_{g,u}^{opt,r}$ can be derived at the rth iteration according to (27), namely

$$e_{g,u}^{opt,r} = 1-p_{g,u}^{r-1}\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2\left(\Xi_{g,u}^{r-1}\right)^{-1}, \quad \forall g,u. \quad (35)$$

Algorithm 2: AO-Based Iterative Approach

1 **Initialize** the maximum iterations $\triangle_2$, the maximum tolerance ε_2, feasible solution $\{\mathbf{P}^0,\boldsymbol{\rho}^0\}$ and iteration index $r=1$.
2 **repeat**
3 Calculate $\{\xi_{g,u}^r\}$ according to (33).
4 Calculate $\{\upsilon_{g,u}^r\}$ according to (36).
5 Calculate $\{\mathbf{P}^r\}$ and $\{\boldsymbol{\rho}^r\}$ by solving problem (40).
6 **until** $\{\mathbf{P}^r,\boldsymbol{\rho}^r\}$ Convergence;

Based on (35), the optimal $\{\upsilon_{g,u}\}$ at the rth iteration can be given by

$$\upsilon_{g,u}^r = \frac{1}{e_{g,u}^{opt,r}}, \quad \forall g,u. \quad (36)$$

Based on the obtained $\{\xi_{g,u}^r\}$ and $\{\upsilon_{g,u}^r\}$, we can solve the following optimization problem:

$$\begin{aligned}\min_{\{\mathbf{P},\boldsymbol{\rho}\}} & \sum_{g=1}^{N_{RF}}\sum_{u=1}^{|U_g|}\frac{\upsilon_{g,u}^r e_{g,u}^r}{\ln 2}+\eta\sum_{g=1}^{N_{RF}}\sum_{u=1}^{|U_g|}p_{g,u}^r \\ \text{s.t. } & (15b)-(15f), \end{aligned} \quad (37)$$

to obtain $\{\mathbf{P}^r\}$ and $\{\boldsymbol{\rho}^r\}$ in the rth iteration, where $e_{g,u}^r = 1+|\xi_{g,u}^r|^2\Xi_{g,u}^r-2\mathrm{Re}\left(\xi_{g,u}^r\sqrt{p_{g,u}^r}\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right), \forall g,u.$

Note that (37) is non-convex optimization problem, since (15e)-(15f) are non-convex constraints. We first handle the non-convex constraint (15e). By substituting (12) and (13) into constraint (15e) in problem (37), the non-linear constraint (15e) can be rewritten by

$$\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}^r \geq \left(2^{R_{min}}-1\right)\left(I_1^r+T_2^r+\delta_0^2+\frac{\delta_1^2}{\rho_{g,u}^r}\right), \forall g,u, \quad (38)$$

where $I_1^r = \left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2\sum_{m=1}^{u-1}p_{g,m}^r$ and $I_2^r = \sum_{n\neq g}\sum_{m=1}^{|U_n|}\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_n\right\|_2^2 p_{n,m}^r$. In the converted constraint (38), the left-hand-side (LHS) is an affine function and the right-hand-side (RHS) is convex function with respect to (w.r.t.) the involved variables. Hence, the constraint (38) is convex.

Next, we focus on tackle constraint (15f). Substituting (9) into (15f) in problem (37), (15f) can be recast by

$$\left(\sum_{i=1}^{N_{RF}}\sum_{j=1}^{|U_g|}\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_i\right\|_2^2 p_{i,j}+\delta_0^2\right) \geq \frac{E_{min}}{\varepsilon_{g,u}(1-\rho_{g,u})}, \ \forall g,u. \quad (39)$$

It is evident that in (39), the LHS is an affine function and the RHS is convex function w.r.t. the involved variables, and thus, the constraint (39) is convex.

After a series of transformations based on (38) and (39), the problem (37) is equivalently reformulated as the following convex optimization problem

$$\begin{aligned}\min_{\{\mathbf{P},\boldsymbol{\rho}\}} & \sum_{g=1}^{N_{RF}}\sum_{u=1}^{|U_g|}\frac{\upsilon_{g,u}^r e_{g,u}^r}{\ln 2}+\eta\sum_{g=1}^{N_{RF}}\sum_{u=1}^{|U_g|}p_{g,u}^r \\ \text{s.t. } & (15b)-(15d),(38)-(39) \end{aligned} \quad (40)$$

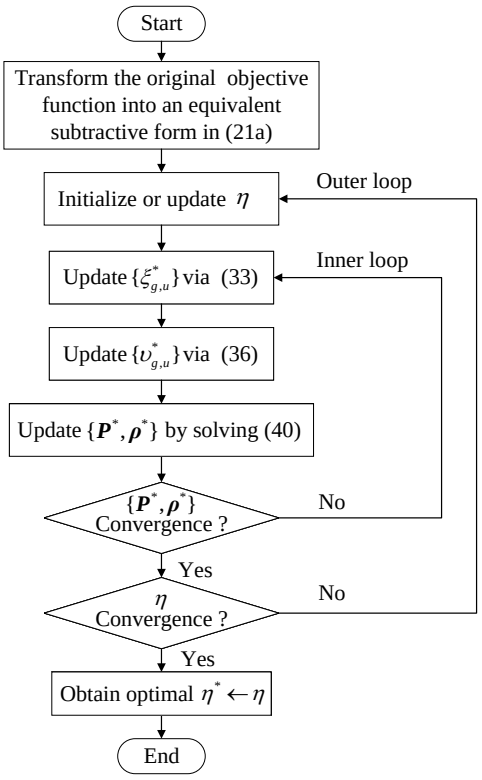


Fig. 3. The flow chart of the proposed two-loop iterative method.

which can be efficiently solved by means of advanced convex solvers, e.g., CVX [35].

The proposed AO-based iterative approach can be summarized as follows. We first initialize $\{\mathbf{P}^{r-1}\}$ and $\{\boldsymbol{\rho}^{r-1}\}$ in the $(r-1)$th iteration. Based on this, the optimal $\{\xi^r_{g,u}\}$ and $\{\upsilon^r_{g,u}\}$ are obtained at the rth iteration by exploiting (33) and (36), respectively. Then, the optimal $\{\mathbf{P}^{r-1}\}$ and $\{\boldsymbol{\rho}^{r-1}\}$ are derived at the rth iteration by solving the standard convex optimization problem (40). Since the optimal $\{\xi^r_{g,u}\}$, $\{\upsilon^r_{g,u}\}$, $\{\mathbf{P}^r\}$ and $\{\boldsymbol{\rho}^r\}$ can be obtained at each iteration, the objective value of problem (32) is non-decreasing when these variables are iteratively updated. Meanwhile, due to the transmit power and the PS radio constraints, the objective value of (32) has an upper bound. Hence, the obtained solution will be at least a local optimal solution. The above scheme is summarized in **Algorithm 2**.

C. Overall Algorithm

To make the non-convex and non-linear problem (15) tractable, a Dinkelbach-based iterative method (**Algorithm 1**) is proposed to convert the objective function in (15) to an equivalent subtractive form. With given energy efficiency parameter η, an AO-based iterative algorithm (**Algorithm 2**) is proposed to handle the converted problem (21) at the inner layer. Specifically, the optimal $\{\xi^r_{g,u}\}$ $\{\upsilon_{g,u}\}$, and $\{\mathbf{P}^r, \boldsymbol{\rho}^r\}$ are alternately updated by solving (33), (36) and (40) until convergence. After that, **Algorithm 1** and **Algorithm 2** are orderly executed. Then, η is updated and the above steps are iteratively repeated until convergence. The flow chart of overall optimization algorithm for solving problem (15) is illustrated in Fig. 3.

Next, we probe into the complexity of our proposed two-layer iterative method. Let l_1 and l_2 denote the iteration times of outer-layer (Dinkelbach approach) and inner-layer (AO approach), respectively. Notice that solving problem (40) dominates the overall complexity of the proposed method. When CVX toolbox is employed to solve the convex programming problem (40), it exploits GP with interior-point method (IPM). At each iteration of inner-layer, the complexity of solving problem (40) is $O(\log((4U+1)/\xi\varsigma)/\log(\pi))$ [36], where $4U+1$ is the total number of constraints in (40). ξ represents the initial point for approximating the accuracy of IPM, ς represents the stopping criterion for IPM, and π is adopted to update the accuracy of IPM. Accordingly, the overall complexity of proposed method is $O(l_1 \times l_2(\log((4U+1)/\xi\varsigma)/\log(\pi)))$.

IV. NUMERICAL RESULT

In this section, we evaluate the performance of our proposed energy-efficient power allocation and splitting scheme for mmWave beamspace MIMO-NOMA with SWIPT system, where the Monte-Carlo method is employed in the simulations. We consider a downlink single cell scenario, where one base station equips with an ULA with 64 antennas to serve U single antenna IoT sensor nodes that equip with PS receivers. We assume that there are one LOS and two NLOS paths for each node' channel. The EH threshold is $E_{min} = 0.1$. We consider that the maximum transmit power is $P_{max} = 15$ dBm, the signal-to-noise ratio is SNR=10 dB, the number of nodes is $U = 16$, if it is not specified. The minimum rate for all user nodes is set to $R_{min} = R_a/10$, where R_a is minimum rate and it is obtained according to the result from full digital MIMO with ZF precoding. We set $P_{BB} = 200$ mW, $P_{SW} = 5$ mW and $P_{NF} = 300$ mW [9]. Meanwhile, the maximum tolerance for Algorithm 1 and Algorithm 2 is set to $\varepsilon_1 = \varepsilon_2 = 10^{-4}$. The maximum iterations for Algorithm 1 and Algorithm 2 are set to 30.

We first evaluate the convergence performance of our proposed algorithms. Fig. 4 presents the convergence performance of Algorithm 1. It can be seen from Fig. 4 that Algorithm 1 converges to a stable point within four iterations. Fig. 5 presents the convergence performance of Algorithm 2. In this figure, we evaluate the ASR performance of the considered system, where η is set to 0. It is clear that the ASR tends to converge within five iterations.

Fig. 6 shows the ASR and EE performance against the maximum transmit power of BS (i.e., P_{max}) with different schemes. In this figure, 'Max-EE' represents the proposed EE maximization scheme, and 'Max-ASR' denotes the ASR maximization scheme, which refers to $\eta = 0$ in problem (21). It is pointed out that when addressing ASR maximization problem for considered beamspace MIMO-NOMA with SWIPT system, we only need to execute the inner layer of Algorithm 1, namely Algorithm 2. From Fig. 6(a), we can see that the ASR of 'Max-EE' is nearly same as that of 'Max-ASR' when P_{max} increases from 20dBm to 35 dBm. However, when P_{max} reaches a certain level, the ASR of

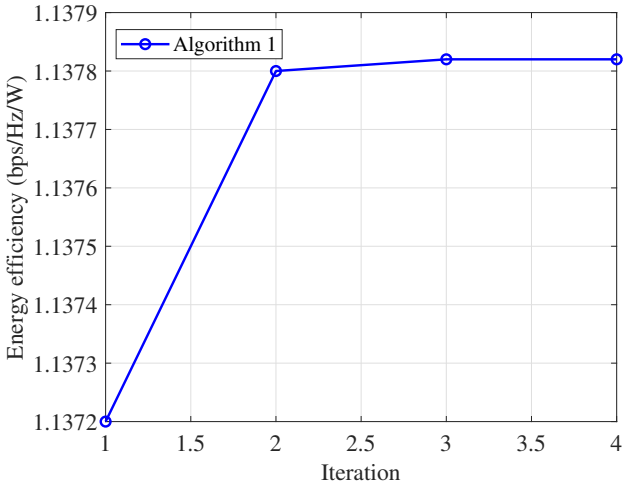


Fig. 4. Convergence of Algorithm 1.

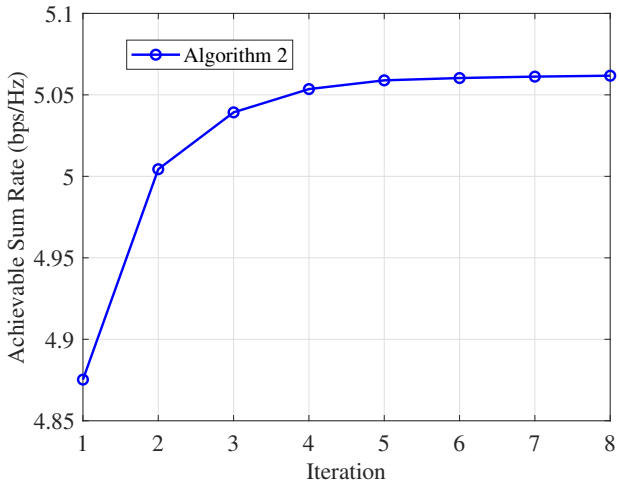


Fig. 5. Convergence of Algorithm 2.

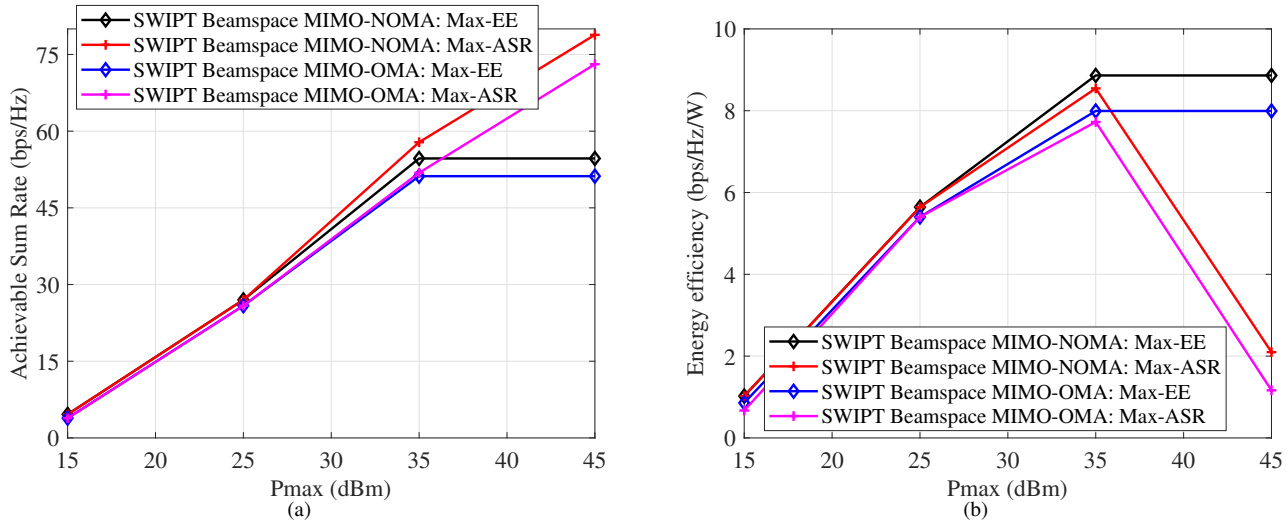


Fig. 6. ASR/EE versus the maximum transmit power of BS with different schemes.

'Max-EE' remains constant, whereas the ASR of 'Max-ASR' is still increasing. The reason is explained as follows. For the 'Max-ASR' schemes, the BS transmit power can be fully allocated to improve ASR performance. But as for the 'Max-EE' schemes, there is a tradeoff between ASR and power consumption, which means that only a portion of total transmit power is utilized in order to guarantee the EE performance. In fact, when P_{max} is lower than a power threshold, all the BS transmit power can be allocated in order to improve the ASR under the 'Max-EE' schemes. As a result, 'Max-EE' and 'Max-ASR' schemes obtain the same ASR performance. On the other hand, when P_{max} exceeds a power threshold, the balance between ASR and power consumption is achieved for the 'Max-EE' schemes, which means that an additional power budget will not contribute to an extra ASR gain. And Hence, the ASR realized by 'Max-EE' scheme remains constant. Meanwhile, this power threshold can be observed in Fig. 6(a). Also, we observe that NOMA schemes achieve higher ASR performance compared with the corresponding OMA schemes. This is because that NOMA schemes support multiple users to be served by a single beam at the same radio resource. As a result, the ASR performances are enhanced. Besides, from Fig. 6(b), we see that EE of 'Max-EE' increases significantly by a small power budget from 20 dBm to 35 dBm, and then approaches a peak value. This phenomenon can be explained by the reason analysed by Fig. 6(a). We also see that the EE of 'Max-ASR' first increases when P_{max} increases. After P_{max} reaches a power threshold, the increase in P_{max} will lead to the degradation of EE performance. This is because that the power consumption becomes higher when the P_{max} is large, and thus, the ASR performance of 'Max-ASR' scheme will decrease.

Fig. 7 presents the ASR and EE versus SNR with different schemes. For validating the effectiveness, we compare the proposed scheme with the following benchmark solutions: (i) 'SWIPT Full digital MIMO': In this scheme, the ZF precoding is utilized in order to mitigate the inter-user interference; (ii) Existing optimization scheme in [37]: The authors in

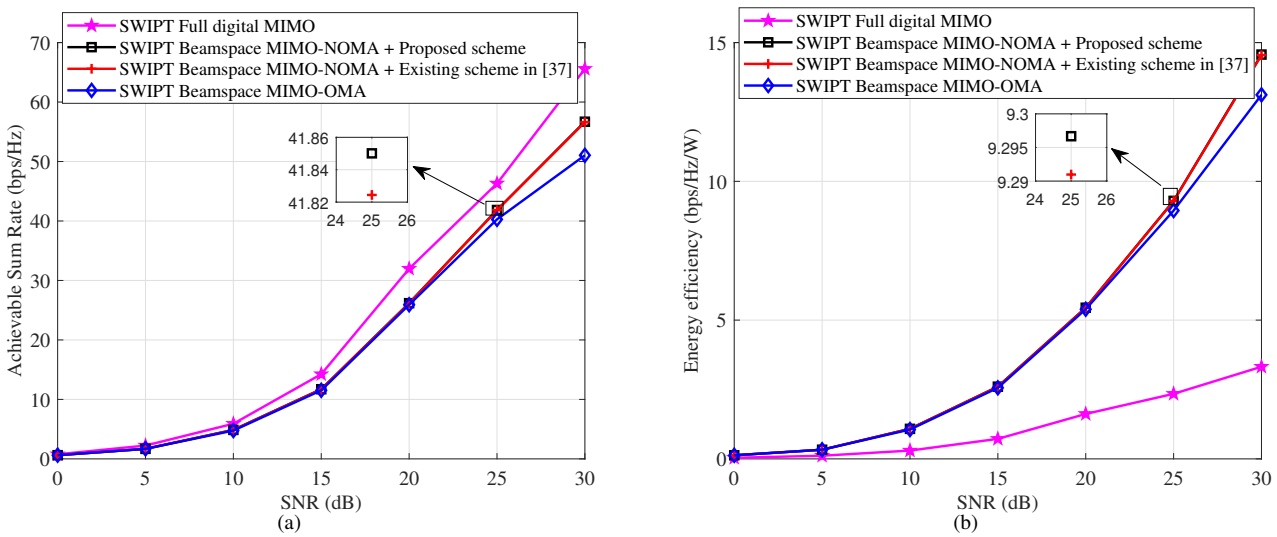


Fig. 7. ASR/EE versus SNR with different schemes.

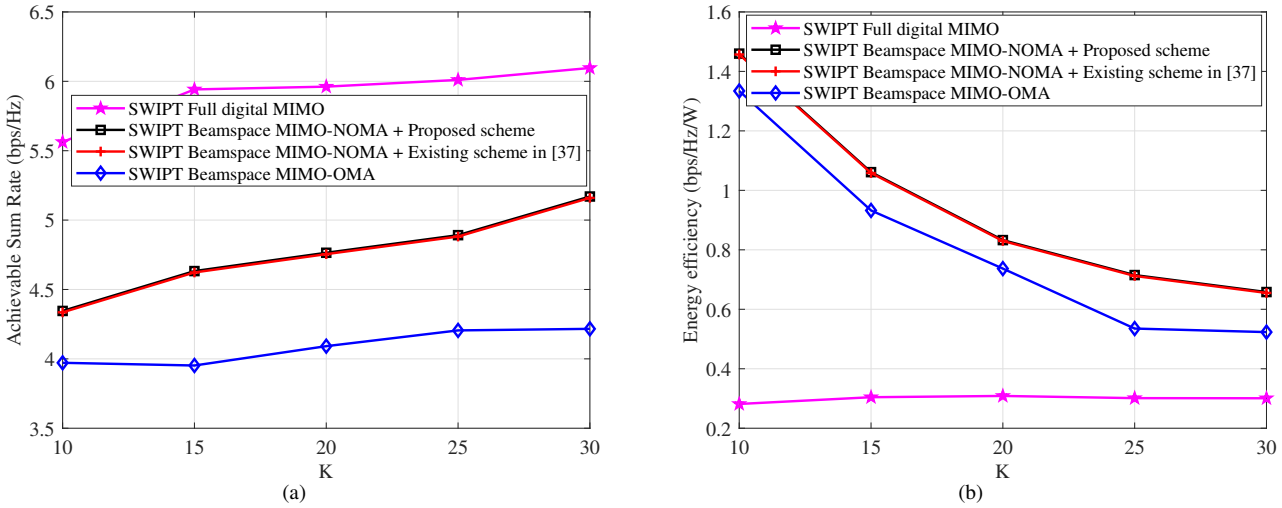


Fig. 8. ASR/EE versus the number of user nodes with different schemes.

[37] propose an energy-efficient power allocation approach by exploiting Dinkelbach method and Constrained Concave Convex Procedure (CCCP) technique. For comparison, we develop this method in beamspace MIMO-NOMA with SWIPT system and extend it to solve the optimization problem (15); and (iii) 'SWIPT Beamspace MIMO-OMA'. For the same system model, we can find that the proposed algorithm can achieve higher ASR and EE compared with the existing scheme in [37]. This result implies that our proposed algorithm is beneficial for enhancing the ASR and EE performances of the considered system. Besides, we observe that the EE performance of 'SWIPT beamspace MIMO-NOMA' is higher than that of 'SWIPT beamspace MIMO-OMA' and 'SWIPT Full digital MIMO'. Also, we see that although 'SWIPT Full digital MIMO' scheme achieves the highest ASR performance, this scheme obtains the worst EE performance. In fact, in a full digital MIMO system, the number of RF chains is equal to that of antennas. This means that the multiplexing gains obtained from RF chains can be fully exploited such that ASR is significantly improved. On the other hand, massive RF chains also bring huge energy consumption. As a result, the EE performance of 'SWIPT Full digital MIMO is dramatically degraded.

Fig. 8 makes the same comparison as Fig. 7, but from the number of user nodes (i.e., K) perspectives. From this figure, we can find that the ASR performance realized by the full digital MIMO structure is larger than the proposed scheme, the existing scheme in [37] and 'SWIPT beamspace MIMO-OMA' scheme, which shows similar trends with Fig. 7. In addition, we also see that the EE of the proposed scheme is better than that of the existing scheme in [37] and 'SWIPT beamspace MIMO-OMA', and it is significantly higher than the EE of the 'SWIPT Full digital MIMO' scheme. These results demonstrate that when the number of user nodes is very large, our proposed solution can still achieve higher EE compared with the benchmark schemes.

V. Conclusion

In this paper, we have proposed to integrate SWIPT with mmWave beamspace MIMO-NOMA to achieve energy-efficient green communication. Aiming to enhance EE of the considered system while guaranteeing the minimum rate and harvested energy requirements for each IoT node, we proposed an energy-efficient power allocation and splitting design. The proposed design was formulated as a non-convex and non-linear optimization problem with the target of maximizing EE. To handle the non-convexity of this problem, we convert the original problem into an equivalent and solvable form by a Dinkelbach-based iterative method at the outer layer. Next, we proposed an AO-based iterative approach to address the converted optimization problem at the inner layer. Simulation results have verified the convergence and effectiveness of the proposed two-layer iterative algorithm. Moreover, it has demonstrated that the EE and ASR of beamspace MIMO-NOMA with SWIPT system under the proposed algorithm outperform the existing scheme and conventional beamspace MIMO with SWIPT. For the future work, we will study advanced precoding design to suppress the inter-beam interference and improve the EE performance.

Appendix A

First, if problem (16) is not feasible, problem (15) cannot be feasible because problem (15) has an additional energy harvested constraint (15f). Second, we assume that problem (16) is feasible and let $\{p_{g,u}\}$ and $\{\rho_{g,u}\}$ be a feasible solution. One can see that another solution $\{\mu p_{g,u}\}$ and $\{\rho_{g,u}\}$, $\forall \mu > 1$, for the first constraint of problem (16) can be expressed as

$$\begin{aligned}
&\frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \mu p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} \mu p_{g,m} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 \mu p_{n,m} + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}} \\
&\geq \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m} + \frac{\delta_0^2}{\mu} + \frac{\delta_1^2}{\mu\rho_{g,u}}} \\
&\geq \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m} + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}} \\
&\geq 2^{R_{min}} - 1
\end{aligned} \tag{41}$$

This means that another solution $\{\mu p_{g,u}\}$ and $\{\rho_{g,u}\}$, $\forall \mu > 1$, is also a feasible solution for problem (16). Since there must exist a sufficiently large $\mu > 1$ such that the solution $\{\mu p_{g,u}\}$ and $\{\rho_{g,u}\}$ satisfies the energy harvested constraint (15f), namely, it is feasible for problem (15). Lemma 1 is thus proved.

Appendix B

First, we assume that problem (17) is feasible. Let $\{p_{g,u}\}$ and $\{\rho_{g,u}\}$ be a feasible solution of problem (17). For any given $0 < \rho_{g,u} < 1$, the solution $\overline{p}_{g,u} = \frac{p_{g,u}}{\rho_{g,u}}$ and $\overline{\rho}_{g,u} = \rho_{g,u}$ for the first constraint of problem (16), can be expressed as

$$\begin{aligned}
&\frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \overline{p}_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} \overline{p}_{g,u} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 \overline{p}_{g,u} + \delta_0^2 + \frac{\delta_1^2}{\overline{\rho}_{g,u}}} \\
&\geq \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m} + \rho_{g,u}\delta_0^2 + \delta_1^2} \\
&\geq \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m} + \delta_0^2 + \delta_1^2} \\
&\geq 2^{R_{min}} - 1
\end{aligned} \tag{42}$$

This means that the solution $\overline{p}_{g,u}$ and $\overline{\rho}_{g,u}$ is a feasible solution of problem (16). Therefore, if problem (17) is feasible, then problem (16) must be feasible.

Second, we consider the case that problem (17) is not feasible. Suppose problem (16) is feasible, and let $\{p_{g,u}\}$ and $\{\rho_{g,u}\}$ be a feasible solution of problem (16). Since $\rho_{g,u} < 1$, we have

$$\begin{aligned}
&2^{R_{min}} - 1 \\
&\leq \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,u} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{g,u} + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}} \\
&\leq \frac{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 p_{g,u}}{\left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m} + \sum\limits_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m} + \delta_0^2 + \delta_1^2}
\end{aligned} \tag{43}$$

This demonstrates that the feasible solution $\{p_{g,u}\}$ and $\{\rho_{g,u}\}$ of problem (16) is also feasible for problem (17), which contradicts the infeasibility of problem (17). Lemma 2 is thus proved.

Appendix C

The MSE $e_{g,u}$ in (24) can be rewritten as

$$\begin{aligned}
e_{g,u} &= 1 + |\xi_{g,u}|^2 \Xi_{g,u} - 2\mathrm{Re}\left(\xi_{\mathrm{g,u}} \sqrt{\mathrm{p_{g,u}}} \hat{\mathbf{h}}_{\mathrm{g,u}}^{\mathrm{H}} \mathbf{v}_{\mathrm{g}}\right) \\
&= \left|1 - \xi_{g,u} \sqrt{p_{g,u}} \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right|^2 + |\xi_{g,u}|^2 \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right\|_2^2 \sum_{m=1}^{u-1} p_{g,m}^{r-1} \\
&\quad + |\xi_{g,u}|^2 \sum_{n\neq g} \sum_{m=1}^{|U_n|} \left\|\hat{\mathbf{h}}_{g,u}^H \mathbf{v}_n\right\|_2^2 p_{n,m}^{r-1} \\
&\quad + |\xi_{g,u}|^2 \delta_0^2 + |\xi_{g,u}|^2 \frac{\delta_1^2}{\rho_{g,u}^{r-1}} \\
&= \left|1 - \xi_{g,u} \sqrt{p_{g,u}} \hat{\mathbf{h}}_{g,u}^H \mathbf{v}_g\right|^2 + |\xi_{g,u}|^2 \left(I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}\right).
\end{aligned} \tag{44}$$

To minimize the MSE $e_{g,u}$, we take the derivative of $e_{g,u}$, namely,

$$\frac{\partial e_{g,u}}{\partial \xi_{g,u}} = \xi_{g,u}^{*}\left(\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u} + I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}\right) - \sqrt{p_{g,u}}\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g = 0. \tag{45}$$

Thus, the optimal receiver filter $\xi_{g,u}^{opt}$ can be obtained as follows:

$$\begin{aligned}\xi_{g,u}^{opt} &= \left(\sqrt{p_{g,u}}\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right)^{*}\left(\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u} + I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}\right)^{-1} \\ &= \left(\sqrt{p_{g,u}}\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right)^{*}\left(\Xi_{g,u}\right)^{-1}. \end{aligned}\tag{46}$$

Substituting (46) into (25), the minimum MSE can be given by

$$\begin{aligned} e_{g,u}^{opt} &= 1 + \left|\xi_{g,u}^{opt}\right|^2 \Xi_{g,u} - 2\mathrm{Re}\left(\xi_{\mathrm{g,u}}^{\mathrm{opt}}\sqrt{\mathrm{p_{g,u}}}\hat{\mathbf{h}}_{\mathrm{g,u}}^{\mathrm{H}}\mathbf{v}_{\mathrm{g}}\right) \\ &= 1 + \left|\left(\sqrt{p_{g,u}}\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right)^{*}\left(\Xi_{g,u}\right)^{-1}\right|^2 \Xi_{g,u} \\ &\quad - 2\mathrm{Re}\left(\left(\sqrt{\mathrm{p_{g,u}}}\hat{\mathbf{h}}_{\mathrm{g,u}}^{\mathrm{H}}\mathbf{v}_{\mathrm{g}}\right)^{*}\left(\Xi_{\mathrm{g,u}}\right)^{-1}\sqrt{\mathrm{p_{g,u}}}\hat{\mathbf{h}}_{\mathrm{g,u}}^{\mathrm{H}}\mathbf{v}_{\mathrm{g}}\right) \\ &= 1 + \left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}\left(\Xi_{g,u}\right)^{-1} - 2\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}\left(\Xi_{g,u}\right)^{-1} \\ &= 1 - \left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}\left(\Xi_{g,u}\right)^{-1}. \end{aligned}\tag{47}$$

APPENDIX D

we first define the following function:

$$f(q) = -\frac{qv}{\ln 2} + \log_2 q + \frac{1}{\ln 2}, \tag{48}$$

where q is a positive scalar. Based on the function (48), we consider the following problem:

$$\max_{q>0} f(q) \tag{49}$$

One can be observed that the objective function $f(q)$ is concave w.r.t. q, and thus the optimal solution q^{opt} to problem (49) can be calculated by

$$\begin{aligned} &\left.\frac{\partial f(q)}{\partial q}\right|_{q=q^{opt}} = 0 \\ &\Rightarrow -\frac{v}{\ln 2} + \frac{1}{q\ln 2} = 0 \\ &\Rightarrow q^{opt} = \frac{1}{v} \end{aligned}\tag{50}$$

Substituting (50) into (48), we have

$$f\left(\frac{1}{v}\right) = \max_{q>0} f(q) \tag{51}$$

Based on the results in (51), (30) can be rewritten by

$$\begin{aligned} R_{g,u}(\mathbf{P},\boldsymbol{\rho}) &= \log_2\left(1 + \frac{\left\|\hat{\mathbf{h}}_{g,u}^{H}\mathbf{v}_g\right\|_2^2 p_{g,u}}{I_1 + I_2 + \delta_0^2 + \frac{\delta_1^2}{\rho_{g,u}}}\right) \\ &= \max_{\{\xi_{g,u}\}}\left(-\log_2\left(e_{g,u}\right)\right) \\ &= \max_{\{\xi_{g,u},\upsilon_{g,u}\}}\left(-\frac{\upsilon_{g,u}e_{g,u}}{\ln 2} + \log_2\upsilon_{g,u} + \frac{1}{\ln 2}\right). \end{aligned}\tag{52}$$

REFERENCES

[1] L. Han, R. Liu, Z. Wang, X. Yue, and J. S. Thompson, "Millimeter-wave MIMO-NOMA based positioning system for internet of things applications," *IEEE Internet Things J.*, May. 2020 (Early access).

[2] R. Jia, X. Chen, Q. Qi, and H. Lin, "Massive beam-division multiple access for B5G cellular internet of things," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2386–2396, Mar. 2020.

[3] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

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

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

[6] S. Chen, S. Sun, G. Xu, X. Su, and Y. Cai, "Beam-space multiplexing: Practice, theory, and trends, from 4G TD-LTE, 5G, to 6G and beyond," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 162–172, Apr. 2020.

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

[8] P. V. Amadori and C. Masouros, "Low rf-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, June. 2015.

[9] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.

[10] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.

[11] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X. Xia, "Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5065–5079, Nov. 2019.

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

[13] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, "Capacity analysis of NOMA with mmWave massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1606–1618, Jul. 2017.

[14] W. Feng, N. Zhao, S. Ao, J. Tang, X. Zhang, Y. Fu, D. K. C. So, and K. K. Wong, "Joint 3D trajectory and power optimization for UAV-aided mmwave MIMO-NOMA networks," *IEEE Trans. Commun.*, Dec. 2020 (Early access).

[15] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, Jul. 2020.

[16] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, Aug. 2020 (Early access).

[17] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May. 2018.

[18] L. Zhu, J. Zhang, Z. Xiao, X. Cao, and D. O. Wu, "Optimal user pairing for downlink non-orthogonal multiple access (NOMA)," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 328–331, Apr. 2019.

[19] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple accesss–A novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, Apr. 2017.

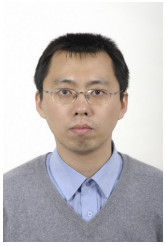
[20] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[21] K. Yang, X. Yan, K. Qin, and Q. Wang, "A uniform beam selection algorithm for beamspace MIMO-NOMA in millimeter-wave communication system," in *Proc. IEEE ICCWAMTIP'18*, Dec. 2018, pp. 166–169.

[22] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May. 2013.

[23] A. Rauniyar, P. E. Engelstad, and O. N. Østerbø, "Performance analysis of RF energy harvesting and information transmission based on NOMA with interfering signal for IoT relay systems," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7668–7682, May. 2019.
[24] B. Xu, Y. Zhu, and R. Zhang, "Optimized power allocation for interference channel with SWIPT," *IEEE Wireless Commun. Lett.*, vol. 5, no. 2, pp. 220–223, Apr. 2016.
[25] Q. Shi, C. Peng, W. Xu, M. Hong, and Y. Cai, "Energy efficiency optimization for MISO SWIPT systems with zero-forcing beamforming," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 842–854, Feb. 2016.
[26] Z. Zong, H. Feng, F. R. Yu, N. Zhao, T. Yang, and B. Hu, "Optimal transceiver design for SWIPT in k-user MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 430–445, Jan. 2016.
[27] J. Tang, J. Luo, J. Ou, X. Zhang, N. Zhao, D. K. C. So, and K.-K. Wong, "Decoupling or learning: Joint power splitting and allocation in MC-NOMA with SWIPT," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5834–5848, Sep. 2020.
[28] J. Tang, Y. Yu, M. Liu, D. K. C. So, X. Zhang, Z. Li, and K. Wong, "Joint power allocation and splitting control for SWIPT-Enabled NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 120–133, Jan. 2020.
[29] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.
[30] Z. Xiao, T. He, P. Xia, and X. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May. 2016.
[31] Z. Xiao, P. Xia, and X. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Oct. 2017.
[32] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
[33] G. Zheng, K. . Wong, and T. . Ng, "Energy-efficient multiuser SIMO: achieving probabilistic robustness with gaussian channel uncertainty," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1866–1878, Jun. 2009.
[34] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967.
[35] M. Grant and S. Boyd. (2014) CVX: Matlab software for disciplined convex programming, version 2.1. [Online]. Available: http://cvxr.com/cvx
[36] N. Mokari, F. Alavi, S. Parsaeefard, and T. Le-Ngoc, "Limited-feedback resource allocation in heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2509–2521, Apr. 2016.
[37] W. Hao, M. Zeng, Z. Chu, S. Yang, and G. Sun, "Energy-efficient resource allocation for mmWave massive MIMO HetNets with wireless backhaul," *IEEE Access*, vol. 6, no. 2457-2471, pp. 2457–2471, Dec. 2018.

Liangyu Chen (Graduate Student Member, IEEE) received the M.S. degree from Inner Mongolia University, Hohhot, China. He is currently pursuing a Ph.D. degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China. His current research interests include non-orthogonal multiple access, deep reinforcement learning, resource allocation, and optimization.



Bo Hu (S'04–M'07) received his PhD degree of communications and information systems in 2006, from Beijing University of Posts and Telecommunications (BUPT), China. Currently, he is a professor in the State Key Laboratory of Networking and Switching Technology, BUPT. He has published over 80 research papers in international journals & conferences and co-authored 3 books. He is active in the ITU-T SG13 for international standards and focus on the mobility management and machine learning for IMT-2020(5G) & beyond. His research interests include: integrated satellite and terrestrial mobile communication system for B5G and 6G, network artificial intelligence and mobility management & control.

Guixian Xu received his Ph.D. degree in communications and information systems from Beijing University of Posts and Telecommunications (BUPT), China, in 2017. He was a visiting Ph.D. student with the National Tsing Hua University, Hsinchu, Taiwan, from 2015 to 2016. He was a postdoc with Aalborg University and Aarhus University, Denmark, in 2018 and 2019, respectively. Since 2020, he is a postdoc research fellow with Tampere University, Finland. His research interests are in signal processing in 5G and beyond, machine learning, and convex optimization.

Shanzhi Chen [F'20] received the bachelor's degree from Xidian University in 1991 and the Ph.D. degree from the Beijing University of Posts and Telecommunications, China, in 1997. He joined the Datang Telecom Technology and Industry Group and the China Academy of Telecommunication Technology (CATT) in 1994, and has been serving as the EVP of Research and Development since 2008. He is currently the Director of the State Key Laboratory of Wireless Mobile Communications, CATT, where he conducted research and standardization on 4G TD-LTE and 5G. He has authored and co-authored four books [including the well-known textbook Mobility Management: Principle, Technology and Applications (Springer Press)], 17 book chapters, more than 100 journal papers, 50 conference papers, and over 50 patents in these areas. He has contributed to the design, standardization, and development of 4G TD-LTE and 5G mobile communication systems. His current research interests include 5G mobile communications, network architectures, vehicular communication networks, and Internet of Things. He served as a member and a TPC Chair of many international conferences. His achievements have received multiple top awards and honors by China central government, especially the Grand Prize of the National Award for Scientific and Technological Progress, China, in 2016 (the highest Prize in China). He is the Area Editor of the IEEE INTERNET OF THINGS, the Editor of the IEEE NETWORK, and the Guest Editor of the IEEE WIRELESS COMMUNICATIONS, the IEEE Communications Magazine, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.