IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 5, 2022, accepted February 22, 2022, date of publication February 28, 2022, date of current version March 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3155485

Decoupling Energy Efficient Approach for Hybrid Precoding-Based mmWave Massive MIMO-NOMA With SWIPT

AHLAM JAWARNEH[1], MICHEL KADOCH[1], (Life Senior Member, IEEE), AND ZAID ALBATAINEH[2], (Member, IEEE)

[1]Département de Génie Électrique, École de Technologie Supérieure, Montreal, QC H3T 1J4, Canada
[2]Department of Electronic Engineering, Yarmouk University, Irbid 21163, Jordan

Corresponding author: Ahlam Jawarneh (ahlam.jawarneh.1@ens.etsmtl.ca)

ABSTRACT In this paper, we address the problem of energy consumption associated with mixed signal components such as analog-to-digital components in millimeter-wave (mmWave) massive MIMO systems. We employ non-orthogonal multiple access (NOMA) in millimeter-wave (mmWave) massive MIMO systems to further enhance the spectrum efficiency. The simultaneous wireless information and power transmission technology (SWIPT) will be used in mmWave massive Multiple-Input multiple-Output MIMO systems. The utilization of SWIPT contributes to prolonging the battery life of mobile users (MUs) and enhances the system energy efficiency (EE), especially in the NOMA scenario where the inter-user interference can be reused for energy harvesting (EH). However, we initially designed a user grouping algorithm based on the affinity propagation clustering algorithm, which preferentially groups the user equipment (UE) based on their channel correlation and distance. Then, we design the analog RF precoder based on the selected user grouping for all beams, followed by a low-dimensional digital baseband precoder design to further mitigate inter-beam interference and maximize the achievable sum-rate for the considered system. Subsequently, we transform the original optimization problem into a joint power allocation and power-splitting maximization problem. The considered non-convex optimization problem is arduous to tackle, resulting from the presence of coupled variables and inter-user interference. To cope with this problem, a decoupled approach is adopted, in which the power allocation and power splitting are separated, and the corresponding sub-problems are solved using the Lagrangian duality method. Simulation results confirm the effectiveness of the proposed method and demonstrate that the proposed method is near-optimal and enjoys higher spectrum and energy efficiency compared with state-of-the-art designs and the conventional SWIPT-enabled mmWave MIMO-NOMA system.

INDEX TERMS SWIPT, mmWave, massive MIMO, NOMA, hybrid precoding, power allocation, power splitting.

I. INTRODUCTION

With 5G wireless communication networks, it is increasingly important to provide services with much higher quality, including enhancing the system capacity within the limited service power and spectrum resources [1]. Massive MIMO, utilizing millimeter waves (mmWave), is an emerging technology for 5G/6G wireless communications because it offers higher bandwidth and better spectrum efficiency [2].

Throughput and spectral efficiency are improved by orders of magnitude when the mmWave bandwidth is increased [3]–[5]. This makes 5G wireless communication an appealing technology for the future. Theoretically speaking, the capacity for multiuser MIMO (massive MIMO) to enhance spectral efficiency by order of magnitude has been proven to more significant multiuser gain [6]. However, the use of non-orthogonal multiple access (NOMA) in millimeter-wave large MIMO systems has recently been investigated to improve the spectrum efficiency [7]–[10]. Through the combination of multiple power levels on the same frequency

The associate editor coordinating the review of this manuscript and approving it for publication was Fan-Hsun Tseng.

resource block, NOMA can enhance the spectral efficiency across the entire system. This has led to the emergence of NOMA as a contender for 5G wireless communication technologies [7]. Overall, mmWave with higher frequencies is better suited for antenna arrays with a massive MIMO system due to small physical size of huge antenna array. In addition, a large antenna array can use precoding to avoid free space path loss of mmWave signals, thereby achieving significant array gain for connections with quality Signal-to-noise ratio (SNR) [8].

Massive MIMO systems employ a significant number of antennas, each of which has a single RF chain, resulting in higher costs and more energy usage. A solution to this problem has been offered in the form of hybrid precoding (HP), which helps to significantly reduce the number of required RF chains in mmWave massive MIMO systems without causing a visible drop in performance [9], [10]. HP focuses on developing completely digital precoders, which are composed of several analog and RF chains, to boost antenna gain and, as a result, reception quality [11]. It is often common to see HP networks with both fully connected and sub-connected topologies [12]. Sub-connected architectures are predicted to provide greater energy efficiency [13].

Although there are various ways to improve the system's energy efficiency, enhancing the endurance of numerous power-limited mobile devices and improving the energy efficiency of the system are also critical considerations for 5G networks, especially in the application scenarios of internet of things (IoT) and Massive Machine-Type Communications (mMTC). A revolutionary technology termed SWIPT was introduced in [14], [15] as a result of the advancement and development of wireless power transfer (WPT). Although SWIPT has certain advantages, the significant disparity in signal sensitivity between the information decoder and rectifier circuit causes this technology to be underutilized [7], [9], [16]–[18]. Two practical receiving methods, time switching (TS) and power splitting (PS), were developed in [19] to solve this problem. These schemes used time switching (TS) and power splitting (PS), with information decoding (ID) and EH, performed in separate time and power domains, respectively. As a result, SWIPT enables an improved system EE, a viable green communication option for future wireless networks. Therefore, it has been noticed by both academic and industrial people [20]–[22].

Precoding is done fully in the digital domain to eliminate interference between distinct data streams in the standard cellular frequency spectrum (e.g., 2–3 GHz) [23], [24]. Because of the higher energy demands, each antenna requires a specialized RF chain (including a digital-to-analog converter, up converter, etc.) with a total energy usage of approximately 250 mW per RF chain [25], [26]. A significant number of RF chains will be required for an mmWave massive MIMO system with 64 antennas because of the usual digital precoding method. A hybrid analog-digital precoding solution was developed to address this problem. Instead of using traditional digital precoding, RF chains are used to obtain these results, and an analog precoder is implemented using a large number of analog phase shifters (PSs) [27]. There is no performance difference between digital and hybrid precoding because hybrid precoding uses fewer RF chains while delivering equivalent energy efficiency [28], [29].

Two distinct classifications may be used for the current hybrid precoding strategies. the preliminary works [13], [30] that described the use of sparse precoding to hybrid precoding is called "precoding with sparse precoding." [31] presented an efficient method called orthogonal matching pursuit (OMP) to attain nearly optimum performance. In the second hybrid precoding method, which involves iterative searching among predefined codebooks [32]–[34], the best hybrid precoding matrix was found iteratively by sequentially passing through the codebooks. Each RF chain is linked to all base-station (BS) antennas through PSs. Under the assumption that there are a huge number of BS antennas (e.g., 256, as studied in [35]), the fully connected design will require thousands of PSs, which might introduce three new limitations: 1) in order to generate more energy, the larger phased array radar needs to absorb more energy for excitation; 2) in order to compensate for the insertion loss of PS, the larger phased array radar requires more energy; 3) because of the higher computational complexity, the larger phased array radar consumes more energy. While the hybrid precoding method with the sub-connected design uses fewer PSs, it requires all RF chains to be linked to each BS antenna. Because the sub-connected architecture is projected to be more energy efficient and simpler to implement for mmWave MIMO systems, it follows that the sub-connected architecture is expected to be more energy efficient and easier to implement for mmWave MIMO systems. The initial challenge of hybrid precoding with a fully connected architecture is difficult because of the new limitations imposed by the sub-connected architecture [36], [37].

The NOMA technique was previously used for beamspace MIMO for the first time in [38], which may be considered a straightforward realization of HP, and power allocation was adjusted to maximize the sum rate that could be achieved. Furthermore, in [23], the HP architecture employed NOMA overall, and digital precoding was implemented using digital block diagonalization (BD) precoding. In addition, more complex digital precoding was suggested in [24], known as minimization maximization (MM)-based precoding. Therefore, the power allocation for mmWave large MIMO-NOMA systems was adjusted to improve their energy efficiency, and an iterative technique was suggested to optimize the power allocation [25].

Improved spectrum efficiency, along with improvements in energy efficiency, are among the most key performance indicators (KPIs) for 5G, which are projected to result in an approximately 100-fold increase in spectral efficiency compared to present 4G wireless communications. Toward this end, SWIPT, presented for the first time in [39], has gained wide acceptance in the last few years [40], [41]. SWIPT proposes that the same received RF signals may include both information and energy, and that this may be accomplished using power-splitting receivers in practice. SWIPT is a tool used to increase the battery life of wireless communication devices by harvesting energy from RF signals. This can advance networks such as the Internet of Things, especially in IoT with many wireless devices. Careful consideration of the trade-off between information rate and harvested energy level is necessary when SWIPT is employed in multiuser systems because inter-user interferences might negatively impact the ID while supporting the EH [42]. Indeed, initiatives have been put out to address this issue. In addition, in [26], the transmit power was reduced under the signal-to-interference-plus-noise ratio (SINR) and Quality of service (QoS) requirements for multiuser MIMO systems to minimize interference and noise [43].

A further aspect of interest is the combined transceiver and power-splitting SWIPT downlink design, which also uses the mean squared error (MSE) criteria [44]. The combined transceiver and power splitting design was explored to enhance the energy efficiency in multicell multiuser downlink SWIPT systems. Even though SWIPT is capable of providing efficient wireless communications, it has only been tested on single-user systems, where future challenges to the joint transceiver and power splitting optimization will emerge.

In this paper, we are interested in a new system that can exist by combining the spectrum-efficient mmWave massive MIMO-NOMA systems with energy-efficient SWIPT. This work presents a new way to solve the joint power allocation, power splitting, and joint precoding problem in SWIPT-enabled mmWave MIMO-NOMA systems by incorporating user groupings.

Our contributions can be summarized as follows:

1. We explore hybrid analog/digital precoding and power splitting optimization to create SWIPT-enabled mmWave mMIMO-NOMA systems with hybrid analog-digital recording. To focus on the clustering process, we first propose a new affinity propagation clustering method for user grouping to help with the initial cluster formation process. The parameters for this algorithm include the channel correlation and channel distance values. In this case, we consider the hybrid analog-digital precoder, power allocation, and power slitting factor optimization problem as a sum-rate maximization problem. We seek to maximize the overall power and minimum rate values under the set power and rate restrictions for each UE.
2. We have now set out to build a hybrid mmWave MIMO-NOMA precoding matrix to overcome this challenge. In the first step, the analog precoder is intended to ensure that all beams acquire the maximum equivalent channel gain, depending on the user groupings. Finally, we construct the digital precoding vector for each UE, which prioritizes those users with the most substantial equivalent channel gain per beam to minimize inter-user interference. To simplify our total power and minimum rate restrictions at each UE, we frame the issue as a combined optimization of power allocation and power-splitting factors. The added requirement is that both variables are limited.
3. To optimize the attainable data rate of the system given the restrictions of transmit power and EH need, the combined power allocation and splitting control issue is mathematically modeled. Because of the interrelationship between the linked variables, non-convex and complicated issues emerge.
4. In contrast to [8], [10], we propose decoupling the joint power allocation and transmit power. Before attempting to optimize the PS ratio assignment with fixed power allocation, we address the subproblem of optimizing the PS ratio assignment with varying power allocation. The Lagrangian duality approach helps solve both the subproblems. Convergence is established when this technique is performed several times.

HP-based mmWave massive MIMO-NOMA systems with SWIPT were simulated to evaluate their performance in terms of both spectrum efficiency and energy efficiency. The results showed an enhancement in the spectrum and energy efficiency. The proposed method for mmWave massive MIMO-NOMA systems with SWIPT can outperform those of mmWave massive MIMO-OMA systems with SWIPT by achieving greater spectrum and energy efficiency.

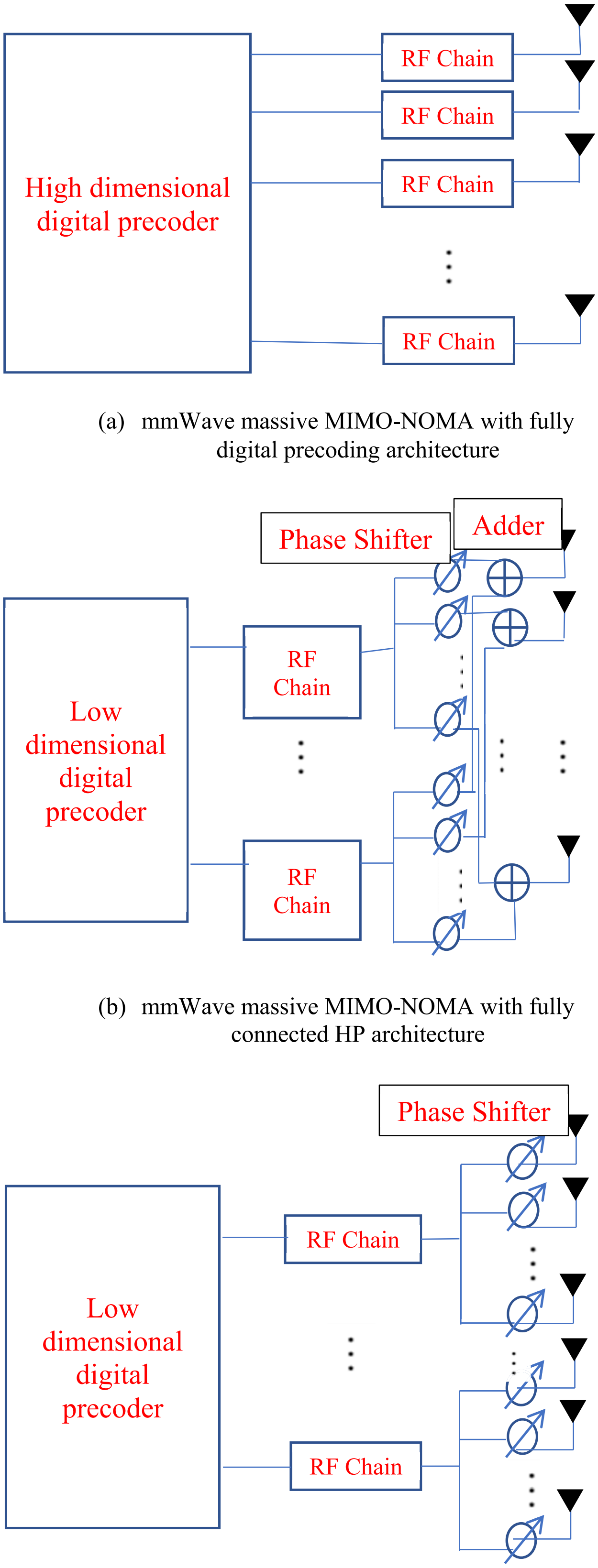
The remainder of this paper is structured as follows. Specifically, Section II describes the system model of the SWIPT-enabled mmWave mMIMO-NOMA system with hybrid analog-digital precoding as well as the sum-rate issue formulation. Section III describes the design of the user-grouping algorithm. Hybrid analog-digital precoder design is presented in Section IV. In Section V, the formulation of the problem itself and an iterative optimization technique to further simplify the solution of the non-convex issue, are presented. Section VI presents the results of the simulations for attainable rates and energy efficiency. Section VII concludes the paper with a summary of the findings.

Notation: In this paper, lower-case letters denote scalars, bold lower-case letters denote vectors, and bold uppercase letters denote matrices. $(.)^T$ denotes the transpose operator; $(.)^H$ represents the Hermitian transpose operator, $diag\,(\boldsymbol{v})$ represents the diagonal matrix with the vector $\boldsymbol{v}$; $\boldsymbol{v}_{\boldsymbol{\pi}}$ represents the sub-vector consisting of the elements of indexes $\boldsymbol{\pi}$; $\|.\|_p$ denotes the l_p-norm.

II. SYSTEM MODEL

Consider a single-cell downlink mmWave massive MIMO-NOMA system. The base station (BS) is equipped with N_{RF} RF chains and N_t transmitted antennas to serve K single antenna users. In this study, we assume that the user equipment is supplied with a power-splitting receiver for SWIPT.

Each antenna is connected to a dedicated RF chain in a fully digital MIMO system, as shown in Fig.1. (a). Moreover, the required number of RF chains is equal to the number of antennas, which causes high power consumption and expensive hardware costs. The fully hybrid precoding architecture


High dimensional digital precoder
RF Chain
RF Chain
RF Chain
RF Chain
Phase Shifter
Adder
Low dimensional digital precoder
RF Chain
RF Chain
Phase Shifter
Low dimensional digital precoder
RF Chain
RF Chain

(a) mmWave massive MIMO-NOMA with fully digital precoding architecture

(b) mmWave massive MIMO-NOMA with fully connected HP architecture

(c) mmWave massive MIMO-NOMA with Sub-connected HP architecture

FIGURE 1. **System models of massive mmWave MIMO architectures.**

is shown in Fig.1. (b). It is evident that the required number of RF chains in the hybrid precoding architecture is less than the number of antennas. Each of the N_{RF} RF chains in the fully hybrid precoding is linked to all N antennas owing to phase shifters. However, the required phase shifters are equal to NN_{RF} and each RF chain can employ the full array gain. In the subconnected hybrid precoding architecture Fig.1. (c), the required phase shifters are equal to N because each RF chain is linked to a subset of N base-station antennas.

In [7], it has been showed that the number of RF chains is larger than or equal to the number of beams, and each beam can only tolerate one user in hybrid precoding based on mmWave massive MIMO systems. However, we assume that the number of beams, G, equals the number of RF chains, N_{RF} to obtain the full multiplexing gain. Moreover, NOMA technology can be employed to make each beam tolerate more than one user. Consider $S_g \forall g = 1, 2, \ldots, G$ represents the set of users supported by the gth beam with $|S_g| \geq 1$, and we have $S_i \cap S_j = \emptyset \forall i \neq j$, thus $\sum_{g=1}^{G} |S_g| = K$. Then, the received signal at the mth user in the gth beam is given by:

$$y_{g,m} = h_{g,m}^H A \sum_{i=1}^{G} \sum_{j=1}^{|S_i|} d_i \sqrt{p_{i,j}} s_{i,j} + v_{g,m} \tag{1}$$

$$\begin{aligned} y_{g,m} &= h_{g,m}^H A d_g \sqrt{p_{g,m}} s_{g,m} \\ &\quad + h_{g,m}^H A d_g \left(\sum_{j=1}^{m-1} \sqrt{p_{g,j}} s_{g,j} + \sum_{j=m+1}^{|S_g|} \sqrt{p_{g,j}} s_{g,j} \right) \\ &\quad + h_{g,m}^H A \sum_{i \neq g} \sum_{j=1}^{|S_i|} d_i \sqrt{p_{i,j}} s_{i,j} + v_{g,m} \end{aligned} \tag{2}$$

In equation (2), the first, second, third, and last terms represent the desired signal, intra-beam interference, inter-beam interference, and noise, respectively. Where $s_{g,m}$ denotes the transmitted signal with $E\left\{ \left| S_{g,m} \right|^2 \right\} = 1$, $p_{g,m}$ represents the transmitted power of the mth user in the gth beam, $v_{g,m} \in \mathbb{CN}\left(0, \sigma_v^2\right)$ is the complex noise, $d_g \in \mathbb{C}^{N_{RF} \times 1}$ represents the digital precoding vector of the gth beam, and $A \in \mathbb{C}^{N \times N_{RF}}$ denotes the analog precoding matrix, where $\left\| A d_g \right\|_2 = 1 \forall g = 1, 2, \ldots, G$.

For the fully hybrid precoding architecture, the analog precoding matrix $A^{(full)}$ is given by:

$$A^{(full)} = \left[\bar{a}_1^{(full)}, \bar{a}_2^{(full)}, \ldots, \bar{a}_{N_{RF}}^{(full)} \right] \tag{3}$$

where $\bar{a}_n^{(full)} \in \mathbb{C}^{N \times 1} \forall n = 1, 2, \ldots, N_{RF}$ is the steering vector with the same amplitude of $\frac{1}{\sqrt{N}}$ and different phases [6].

For the sub-hybrid precoding architecture, the analog precoding matrix $A^{(sub)}$ is given by:

$$A^{(sub)} = \begin{bmatrix} \bar{a}_1^{(sub)}, 0, \ldots, 0 \\ 0, \bar{a}_2^{(sub)}, \ldots, 0 \\ 0, 0, \ldots, \bar{a}_{N_{RF}}^{(sub)} \end{bmatrix} \tag{4}$$

With no loss of generality, let us assume that $M = \frac{N}{N_{RF}}$ is an integer, and each RF chain is linked with M antennas in the sub-hybrid precoding architecture. $\bar{a}_n^{(sub)} \in \mathbb{C}^{M \times 1} \forall n = 1,$

$2, \ldots, N_{RF}$ is the steering vector with the same amplitude of $\frac{1}{\sqrt{M}}$ [7], [8].

Let us consider the mmWave MIMO channel model [6]–[8], where the $N \times 1$ channel vector $h_{g,m}$ of the mth user in the gth beam is given by

$$h_{g,m} = \sqrt{\frac{N}{L_{g,m}}} \sum_{l=1}^{L_{g,m}} \alpha_{g,m}^{(l)} a\left(\vartheta_{g,m}^{(l)}, \theta_{g,m}^{(l)}\right) \tag{5}$$

where $L_{g,m}$ represents the number of paths of the mth user in the gth beam. $\alpha_{g,m}^{(l)}$, $\vartheta_{g,m}^{(l)}$ and $\theta_{g,m}^{(l)}$ denote the complex gain, the azimuth angle of departure (AoD) and the elevation angle of departure of the lth path respectively. $a\left(\vartheta_{g,m}^{(l)}, \theta_{g,m}^{(l)}\right)$ is the $N \times 1$ steering vector.

For a uniform linear array (ULA) with N_1 elements in the horizon and N_2 elements in the vertical direction, the array steering vector $a\left(\vartheta_{g,m}^{(l)}, \theta_{g,m}^{(l)}\right)$ is given by

$$a\left(\vartheta_{g,m}^{(l)}, \theta_{g,m}^{(l)}\right) = a_{az}\left(\vartheta_{g,m}^{(l)}\right) \otimes a_{el}\left(\theta_{g,m}^{(l)}\right) \tag{6}$$

where

$$a_{az}\left(\vartheta_{g,m}^{(l)}\right) = \frac{1}{\sqrt{N_1}}\left[e^{j2\pi i\left(\frac{d_1}{\lambda}\right) sin\left(\vartheta_{g,m}^{(l)}\right)}\right]_{i \in J(N_1)} \tag{7}$$

and

$$a_{el}\left(\theta_{g,m}^{(l)}\right) = \frac{1}{\sqrt{N_2}}\left[e^{j2\pi i\left(\frac{d_2}{\lambda}\right) sin\left(\vartheta_{g,m}^{(l)}\right)}\right]_{i \in J(N_2)} \tag{8}$$

where $J(n) = \{0, 1, \ldots, n-1\}$, λ is the signal wavelength, d_1 and d_2 are the horizontal and antenna spacings, respectively. We usually assume that $d_1 = d_2 = \frac{\lambda}{2}$ for mmWave communication systems [7].

Power splitting receivers allow one to split the received signal into two parts. While some of the signals are used for information decoding (ID), others can be used for energy harvesting (EH) [27].

The signal for energy harvesting is expressed as:

$$y_{g,m}^{EH} = \sqrt{1-\beta_{g,m}} y_{g,m} \tag{9}$$

where $\beta_{g,m} \in [0, 1]$ is the power factor for the mth user in the gth beam, and the harvested energy is given by:

$$P_{g,m}^{EH} = \eta\left(1-\beta_{g,m}\right)\left(\sum_{i=1}^{G} \sum_{j=1}^{|S_i|} \left\|\bar{h}_{g,m}^{H} d_i\right\|_2^2 p_{i,j} + \sigma_v^2\right) \tag{10}$$

where $\bar{h}_{g,m}^{H} = h_{g,m}^{H} A$ represents the equivalent of the channel vector and $\eta \in [0, 1]$ denotes the energy conversion efficiency. However, the signal for information decoding is given by

$$y_{g,m}^{ID} = \sqrt{\beta_{g,m}} y_{g,m} + u_{g,m} \tag{11}$$

where $u_{g,m} \in \mathbb{CN}\left(0, \sigma_u^2\right)$ represents the noise of the power splitter.

Based on the NOMA at each beam, SIC at the receiver was performed as well as intra-beam superposition coding at the transmitter.

With no loss of generality, let us assume that $\left\|\bar{h}_{g,1}^{H} d_i\right\|_2^2 \geq \left\|\bar{h}_{g,2}^{H} d_i\right\|_2^2 \geq \ldots \geq \left\|\bar{h}_{g,|S_i|}^{H} d_i\right\|_2^2 \forall g = 1, 2, \ldots, G$. Then, the mth user in the gth beam can be diminished the interference from the jth user (for all $j > m$) in the gth beam using the SIC method [15]. The signal for information decoding at the mth user in the gth beam is as follows:

$$\begin{aligned} y_{g,m}^{ID} = \sqrt{\beta_{g,m}} \Bigg(& \bar{h}_{g,m}^{H} d_g \sqrt{p_{g,m}} s_{g,m} + \bar{h}_{g,m}^{H} d_g \sum_{j=1}^{m-1} \sqrt{p_{g,j}} s_{g,j} \\ & + \bar{h}_{g,m}^{H} \sum_{i \neq g} \sum_{j=1}^{|S_i|} d_i \sqrt{p_{i,j}} s_{i,j} + v_{g,m} \Bigg) + u_{g,m} \end{aligned} \tag{12}$$

Then, the SINR at the mth user in the gth beam is expressed as:

$$\gamma_{g,m} = \frac{\left\|\bar{h}_{g,m}^{H} d_g\right\|_2^2 p_{g,m}}{E_{g,m}} \tag{13}$$

where

$$\begin{aligned} E_{g,m} = & \left\|\bar{h}_{g,m}^{H} d_g\right\|_2^2 \sum_{j=1}^{m-1} p_{g,j} \\ & + \sum_{i \neq g} \left\|\bar{h}_{g,m}^{H} d_i\right\|_2^2 \sum_{j=1}^{|S_i|} p_{i,j} \\ & + \sigma_v^2 + \frac{\sigma_u^2}{\beta_{g,m}} \end{aligned} \tag{14}$$

Accordingly, the achievable rate is given by:

$$R_{g,m} = \log_2\left(1 + \gamma_{g,m}\right) \tag{15}$$

Lastly, the achievable sum rate is given by:

$$R_{sum} = \sum_{g=1}^{G} \sum_{m=1}^{|S_g|} R_{g,m} \tag{16}$$

Nevertheless, the achievable sum rate in (16) can be enhanced by designing user grouping, the analog RF precoder matrix A^{RF} digital baseband precoders d_g for the g-th UE, power allocation, and power splitting factors.

III. USER GROUPING

As the number of users (K) is larger than that of the RF chain N_{RF}, that is, $K > N_{RF}$, we need to schedule the user into G groups, that is, $G = N_{RF}$. To this end, we propose an intuitive algorithm for user grouping. Owing to the spatial directivity of the SWIPT-based mmWave Massive MIMO NOMA System, we use the affinity propagation clustering algorithm to implement the user grouping [9], [16], [17] For mmWave large MIMO systems, a clustering technique based on user multidimensional attributes is described in order to increase system performance by considering the similarity of users' characteristics.

Our solution, which uses mmWave Massive MIMO NOMA technology, calculates the relevance between users based on their characteristics to cluster them efficiently and precisely. We consider two types of features: h and p, which represent the user channel vector and the distance between users respectively and we define the feature vector $\boldsymbol{V} = (\boldsymbol{h}, \boldsymbol{p})$. Furthermore, prior to clustering, it is necessary to normalize the multidimensional aspects of the user characteristics. Here, the linear normalizing approach is modified to regulate the outcomes in the range of $[\mathbf{0}, \mathbf{1}]$ in order to achieve better control.

Our distance measures the similarity between user channels. We utilize the Euclidean distance to measure the similarity between users' relative locations, which is a vector, because the transmission channel is also a vector. The statement of the relevance between users $\boldsymbol{i}$ and $\boldsymbol{j}$, on the other hand, is defined as [17].

$$\boldsymbol{S_1} = \boldsymbol{H_{ij}} = \boldsymbol{arccos} \frac{\left| \boldsymbol{h_i^H h_j h_j^H h_i} \right|}{\|\boldsymbol{h_i}\| \, \|\boldsymbol{h_j}\|} \tag{17}$$

$$\boldsymbol{S_2} = \boldsymbol{P_{ij}} = \boldsymbol{d_{ij}} \tag{18}$$

$$\boldsymbol{S_{i,j}} = -\sqrt{\boldsymbol{w_1 S_1^2 + w_2 S_2^2}} \tag{19}$$

where $\sum_{i=1}^{2} \boldsymbol{w_i} = \mathbf{1} \forall \boldsymbol{w_i} \in [\mathbf{0}, \mathbf{1}]$ is the weight factor associated with the characteristics that meet the criteria. The higher the similarity between two users, the closer the distance between them. Thus, we utilize the negative distance to make it positively linked.

The affinity propagation (AP) clustering algorithm [17] is a semi-supervised clustering algorithm that does not require the user to specify the initial cluster center or the number of clusters in advance. It has good clustering stability and a low error rate and is widely used. We utilize the idea of information transmission of the AP method [17] based on multidimensional similarity for grouping users, as described in detail below.

1) By calculating and assigning the median of the similarity matrix for each user K in the similarity matrix [S], the reference degree of user $\boldsymbol{K}$ may be determined and assigned to the vector $\boldsymbol{s(i,k)}$.
2) Create a $\mathbf{0}$ in the responsibility $\boldsymbol{r(i,k)}$ and availability $\boldsymbol{a(i,k)}$ matrix to represent the initial state. Calculate the right number of iterations, $\boldsymbol{Itr}$, as well as the damping factor $(\boldsymbol{\lambda})$;
3) Use the following procedure to repeatedly compute the responsibility and availability for each user $\boldsymbol{k}$ with respect to user $\boldsymbol{i}$ in $\boldsymbol{Itr}$ times:

$$\boldsymbol{r_{t+1}(i,k)} = \boldsymbol{s\,(i,k)} - \max_{\boldsymbol{k \neq k'}} \left\{\boldsymbol{a_t}\left(\boldsymbol{i,k'}\right) + \boldsymbol{s}\left(\boldsymbol{i,k'}\right)\right\}$$

$$\boldsymbol{a_{t+1}\,(i,k)} = \boldsymbol{min} \left\{ \boldsymbol{0,r_t\,(k,k)} + \sum_{\boldsymbol{i' \notin \{i,k\}}} \boldsymbol{max} \left\{\boldsymbol{0,r_t}\left(\boldsymbol{i',k}\right)\right\} \right\}, \quad \boldsymbol{i \neq k}$$

$$\boldsymbol{a_{t+1}\,(i,k)} = \sum_{\boldsymbol{i' \neq k}} \boldsymbol{max} \left\{\boldsymbol{0,r_t}\left(\boldsymbol{i',k}\right)\right\}$$

4) Calculate the responsibility $\boldsymbol{r_t(i,k)} \forall \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{n}$ and availability $\boldsymbol{a_t\,(i,k)} \forall \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{n}$. In order to update information in the AP method, one can incorporate the attenuation coefficient $(\boldsymbol{\gamma})$, which is a real number between $\mathbf{0}$ and $\mathbf{1}$, with a typical value of between $\mathbf{0.5}$ and $\mathbf{0.9}$:

$$\hat{r}_{\boldsymbol{t+1}}\boldsymbol{(i,k)} = \boldsymbol{(1-\gamma) r_{t+1}(i,k)} \; \mathcal{C} \; \boldsymbol{\gamma r_t(i,k)}$$
$$\hat{a}_{\boldsymbol{t+1}}\,\boldsymbol{(i,k)} = \boldsymbol{(1-\gamma) a_{t+1}\,(i,k)} \; \mathcal{C}\boldsymbol{\gamma} \; \boldsymbol{a_t\,(i,k)}$$

5) Update the responsibility $\boldsymbol{r_{t+1}(i,k)} \forall \boldsymbol{i} = 1, \ldots, \boldsymbol{n}$ and availability $\boldsymbol{a_{t+1}\,(i,k)} \forall \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{n}$.
6) Calculate $\boldsymbol{e(k,k)} = \boldsymbol{r(k,k)} + \boldsymbol{a(k+k)}$ $\forall \boldsymbol{k} = 1, 2, \ldots, \boldsymbol{K}$, and if $\boldsymbol{e(k,k)} > \mathbf{0}$, $\boldsymbol{k}$ is the center of the cluster. After that, the cluster center set of users is established. Each user is allocated to the appropriate cluster based on the concept of the minimal distance between the two clusters.

Algorithm 1 provides the pseudocode for the improved AP scheme, which is a mathematical representation of the code.

Algorithm 1: Presented User Grouping Method

```
Input:
    Number of UEs: $K > N^{RF}$
    Number of RF chains: $N^{RF}$
    Number of beams: $G$
    Channel Matrix: $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]$
    Number of BS antennas: $N$
    Initialization: $M = 0^G$
    Set predefined threshold: $0 \le \gamma \le 1$
Output:
    Optimized User Grouping: $\mho = \left\{\widehat{\mathscr{G}}_1, \widehat{\mathscr{G}}_1, \ldots, \widehat{\mathscr{G}}_G\right\}$
$\mathcal{K} = \{1, 2, \ldots, K\}$
Initialize $\boldsymbol{\Omega_m^{(1)}} = \boldsymbol{k_m} \in K \forall m = 1, 2, \ldots, G$
        $\boldsymbol{\Psi} = \left[\|\boldsymbol{h_1}\|_2, \|\boldsymbol{h_2}\|_2, \ldots, \|\boldsymbol{h_K}\|_2\right]$
$\overline{\boldsymbol{H}} = \left[\frac{\boldsymbol{h}_1}{\|\boldsymbol{h_1}\|_2}, \frac{\boldsymbol{h}_2}{\|\boldsymbol{h_2}\|_2}, \ldots, \frac{\boldsymbol{h}_K}{\|\boldsymbol{h_K}\|_2}\right]$
Calculate $S_{i,j}$
        $\boldsymbol{S_{i,j}} = -\sqrt{w_1 S_1^2 + w_2 S_2^2}$
$t = 1$.
While Loop
    Initialize $\widehat{\boldsymbol{\mathscr{G}}}_{\boldsymbol{m}} = \boldsymbol{\Omega_m^{(t)}}$
    For $k \in {}^{K}\!\Big/\!\left\{\boldsymbol{\Omega_m^{(t)}}\right\}$
    $\boldsymbol{g = arg} \max_{1 \le g \le M} \boldsymbol{S_{i,j}}$
    $\mho = \widehat{\boldsymbol{\mathscr{G}}}_{\boldsymbol{m}} \cup k$
        End for
        $\boldsymbol{t = t + 1}$.
Update $\boldsymbol{\Omega_m^{(t)}}$ for $m = 1, 2, \ldots, G$
If $\left\{\boldsymbol{\Omega_m^{(t)}} = \boldsymbol{\Omega_m^{(t-1)}}\right\}$
    End While loop
Return $\mho = \left\{\widehat{\mathscr{G}}_1, \widehat{\mathscr{G}}_1, \ldots, \widehat{\mathscr{G}}_G\right\}$
```

IV. HYBRID PRECODER DESIGN

To maximize (11) for each UE, we should reduce the inter-beam interference while simultaneously increasing the effective channel gain. Zero forcing (ZF) is a technique that may be used in conventional multiuser MIMO (MU-MIMO) systems [7], [10], and [34].

We propose to use phase-only array response adjustment to link the N_{RF} RF chain outputs with the N_{BS} BS antennas, using low-cost phase shifters, in order to decrease hardware restrictions while still realizing the full potential of mmWave huge MIMO-NOMA systems.

Unfortunately, because of the elementwise constant-magnitude limitation on the analog precoder, that is, $\left|\left[F^{RF}\right]_{i,j}\right| = \frac{1}{\sqrt{N_{\mathrm{BS}}}}, \forall i, j$, they cannot be used directly in the hybrid analog-digital precoding method [7], [10], and [34]. Because of the constant-magnitude restriction, the subsets of feasible areas are not convex; thus, the solution is non-convex. Consequently, we are considering creating the analog RF precoder and the digital baseband precoder in distinct phases of the development process. Based on [7] and [11], we present an efficient analog RF precoding algorithm to design F^{RF} and a low-dimensional digital baseband precoding algorithm to design F^{BB} for downlink multiuser mmWave massive MIMO-NOMA systems. As a first step, we designed the analog RF precoding matrix.

A. ANALOG RF PRECODING METHOD

Our goal with the analog RF precoder is for the phases of $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]$ to be aligned so that the high array gain delivered by the massive MIMO system can be harvested effectively. Using **Algorithm 2**, we can quickly review the analog RF precoder architecture. For simplicity, it is preferable to focus on the main element of the proposed algorithm rather than providing a redundant demonstration. Initially, we start the analog precoder as an all-zero matrix to ensure that it operates correctly. It is necessary to extract the phases of the conjugate transpose of the aggregate downlink mmWave massive MIMO-NOMA channel from the BS to numerous users in Step 4 to complete the computation. Phase alignment of channel components is performed in Step 10 to build the analog RF precoder in order to harvest a significant array gain. Subsequently, once the effective baseband channel has been coupled with the ideal analog RF precoder acquired, the digital baseband precoder design is carried out to minimize interference and maximize the sum rate that can be accomplished.

B. DIGITAL BASEBAND PRECODING METHOD

The digital baseband precoding matrix is designed such that only the UEs in each beam with strong channels are selected to eliminate inter-user interference. To avoid inter-beam interference, the design of digital precoding is transformed into a typical massive MIMO-NOMA precoding issue. As shown in [7] and [10], the low-complexity zero-forcing (ZF) precoding technique is used for digital precoding without sacrificing generality.

Algorithm 2: Presented Analog RF Precoding Method for mmWave Massive MIMO-NOMA Systems With SWIPT

```
Input:
        Number of UEs: $K > N^{RF}$
        Number of RF chains: $N^{RF}$
        Channel Matrix: $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]$
        Optimized User Grouping: $\left\{\widehat{\mathscr{G}}_1, \widehat{\mathscr{G}}_1, \ldots, \widehat{\mathscr{G}}_G\right\}$
        Number of BS antennas:$N$
        Initialization: $\boldsymbol{F}^{RF} = 0^{N \times N^{RF}}$
        Number of quantization bits: $B$
Output:
        Optimal analog RF precoding: $\boldsymbol{F}^{RF}$
1: Set the phase: $\Lambda = \left\{\frac{2\pi n}{2^B}, n = 0, 1, \ldots., 2^{B-1}\right\}$
2: For Loop: $g = 1$ to $G$
3: Recall the optimized user grouping: $\left\{\widehat{\mathscr{G}}_1, \widehat{\mathscr{G}}_1, \ldots, \widehat{\mathscr{G}}_G\right\}$
4:   Set the aggregate downlink channel: $\overline{H} = [\boldsymbol{H}]_{:,\widehat{\mathscr{G}}_d}$
5:   Extract phase of the $\overline{\boldsymbol{H}}$: $\mathfrak{G} = \angle\overline{\boldsymbol{H}}$
6:     Initialize angle: $\vartheta = \mathbf{0}^N$
7:       For $m = 1$ to $\left|S_g\right|$
             $[\sim, \boldsymbol{k}] = \boldsymbol{min}\, |[\mathfrak{G}]_{\boldsymbol{m}} - \Lambda|$
             $\vartheta(m) = [\Lambda]_k$
8:     End for
9: Compute the optimal analog RF precoding:
        $\boldsymbol{F}^{RF}(:, g) = exp(j\vartheta)$
10: End for
```

Specifically, we present an algorithmic solution based on the concepts of [7] and [10] after designing the analog RF precoder ($\boldsymbol{F}^{RF}$). The pseudocode for the digital baseband precoder is given in **Algorithm 3**. We first set the number of UEs (K), number of RF chains N^{RF}, number of BS antennas (N), channel matrix $\boldsymbol{H}$, the optimized analog RF precoder ($\hat{F}^{RF}$) from **Algorithm 2**, and the optimized user grouping from **Algorithm 1**. The precoding algorithm then employs a zero-force precoding algorithm to reduce inter-user interference. As a result, the digital baseband precoder can be represented as

$$\hat{F}^{BB} = \boldsymbol{H}^{\boldsymbol{H}} \left(\boldsymbol{H}\boldsymbol{H}^{\boldsymbol{H}}\right)^{\boldsymbol{-1}} \tag{20}$$

Then, we normalize the digital precoder as follows.

$$\hat{F}^{BB} = \left[\frac{\hat{f}_1^{BB}}{\boldsymbol{f}_1^{BB*}}, \frac{\hat{f}_2^{BB}}{\boldsymbol{f}_2^{BB*}}, \ldots, \frac{\hat{f}_{N^{RF}}^{BB}}{\boldsymbol{f}_{N^{RF}}^{BB*}},\right] \tag{21}$$

where $\boldsymbol{f}_n^{BB*} = repmat\left(\left\|\boldsymbol{F}^{RF}\boldsymbol{f}_n^{BB*}\right\|_2, N^{RF}, 1\right) \forall n = 1, \ldots, N^{RF}$ and $\left\|\boldsymbol{F}^{RF}\boldsymbol{f}_n^{BB*}\right\|_2 = \sqrt{\sum \left|\boldsymbol{F}^{RF}\boldsymbol{f}_n^{BB*}\right|^2}$.

V. JOINT OPTIMIZATION OF POWER ALLOCATION AND POWER SPLITTING

In this section, we have investigated the combined power allocation and power splitting optimization to achieve the highest possible data rate in mmWave Massive MIMO-NOMA systems with SWIPT. Because of the presence of bothinter -and intra-group interferences in MIMO-NOMA systems

Algorithm 3: Presented Digital Baseband Precoding Method for mmWave Massive MIMO-NOMA Systems With SWIPT

```
Input:
    Number of UEs: $K > N^{RF}$
    Number of RF chains: $N^{RF}$
    Channel Matrix: $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]$
    Optimized User Grouping: $\left\{\widehat{\mathcal{g}}_1, \widehat{\mathcal{g}}_1, \ldots, \widehat{\mathcal{g}}_G\right\}$
    Number of BS antennas: $N$
    Optimal analog RF precoding: $\boldsymbol{F}^{RF}$
    Number of quantization bits: $B$
Output:
  Optimal Baseband precoding: $\boldsymbol{F}^{BB}$
1: Set the phase: $\Lambda = \left\{\frac{2\pi n}{2^B}, n = 0, 1, \ldots., 2^{B-1}\right\}$
2: $\overline{\boldsymbol{H}} = \boldsymbol{H}^{\boldsymbol{H}} \boldsymbol{F}^{RF}$
3: $\tilde{\boldsymbol{H}} = \left[\overline{\boldsymbol{H}}\right]_{:,\widehat{\mathcal{g}}_1}$
4: $\hat{\boldsymbol{F}}^{BB} = \tilde{\boldsymbol{H}}^{\boldsymbol{H}}\left(\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^{\boldsymbol{H}}\right)^{-\mathbf{1}}$
5: $\hat{\boldsymbol{F}}^{BB} = \left[\frac{\hat{\boldsymbol{f}}_1^{BB}}{\boldsymbol{f}_1^{BB*}}, \frac{\hat{\boldsymbol{f}}_2^{BB}}{\boldsymbol{f}_2^{BB*}}, \ldots, \frac{\hat{\boldsymbol{f}}_{N^{RF}}^{BB}}{\boldsymbol{f}_{N^{RF}}^{BB*}},\right]$
    where
      $\boldsymbol{f}_n^{BB*} = repmat\left(\left\|\boldsymbol{F}^{RF}\boldsymbol{f}_n^{BB*}\right\|_2, N^{RF}, 1\right) \forall n$
                $= 1, \ldots, N^{RF}$
6: Initialize baseband precoding: $\boldsymbol{F}^{BB} = \mathbf{0}^{\boldsymbol{N}^{RF} \times \boldsymbol{K}}$
7: $\left[\boldsymbol{F}^{BB}\right]_{:,\widehat{\mathcal{g}}_1} = \hat{F}^{BB}$
8: For Loop: $g = 1$ to $G$
        $\Lambda = nonzeros\left([\Lambda]_{\hat{g}_g}\right)^T$
9:      For $m = 2$ to $|\Lambda|$
            $\boldsymbol{F}^{BB}(:, \Lambda_n) = \left[\boldsymbol{F}^{BB}\right]_{:,\widehat{\mathcal{g}}_g}$
10:     End for
11: End for
```

with SWIPT, the existing optimization methods for solving the joint optimization problem of power allocation and power splitting in MIMO systems with SWIPT cannot be directly applied in MIMO-NOMA systems with SWIPT, where there are multiple groups and multiple users in each group. As a result, obtaining optimal solutions is quite difficult. To address this intractable problem, an iterative optimization technique is created in this section, which allows for the generation of suboptimal solutions while fulfilling the intended EH restrictions and the transmit power constraint requirements. Furthermore, the following formulation may be used to precisely express the issue of combined power allocation and power-splitting optimization:

$$\max_{\{p_{g,m}\},\{\beta_m\}} R_{sum}\left(p_{g,m}, \beta_m\right) \tag{22}$$

$$s.t. \ \sum\nolimits_{g=1}^{G} \sum\nolimits_{m=1}^{|S_g|} P_{g,m} \leq P_T \tag{23}$$

$$0 \leq \beta_m \leq 1 \quad \forall m \tag{24}$$

$$p_{g,m} \geq 0 \quad \forall g, m \tag{25}$$

$$P_{g,m}^{EH} \geq p_{g,m}^{req} \tag{26}$$

Constraint (23) indicates that the transmitted power constraint, that is, $\sum_{g=1}^{G}\sum_{m=1}^{|S_g|} P_{g,m}$, cannot exceed the threshold of P_T being the maximum total transmission power of the BS. Constraint (24) limits the power splitting factor β_m for the mth user to be in the range of [0, 1]. Constraint (25) indicates the non-negativity of the power allocated to the mth user in the gth beam. Constraint (26) shows that each mth user in the gth beam is required to harvest at least $p_{g,m}^{req}$ W Power being the minimum harvested energy for each mth user in the gth beam.

As a consequence of the objective function and the coupling of the multiple variables, the optimization issue of the attainable data rate described in (22)–(26) is neither convex nor linear owing to the objective function. Furthermore, the optimization problem mentioned above is a well-known NP-hard problem, and as a result, the solution is complex and cannot be easily achieved. There is a possibility that an exhaustive search approach will provide a solution to this problem. The computational complexity of the exhaustive search technique, on the other hand, increases substantially as the number of users increase. As a result, this technique is far from feasible, particularly in the context of IoT, where there is a desire for massive MIMO systems. We will create an iterative strategy to tackle this problem based on the Lagrangian duality methodology in this section, which will be as follows:

It is feasible for any optimization issue containing many variables to deal with the sub-problem over a subset of variables while treating the remainder as constants and then dealing with the sub-problem over the remaining variables. This is supported by the literature [30] and [31]. This separation of $p_{g,m}$ and β_m allows us to create a realistic and effective solution for the studied optimization issue in (22)–(26).

First, we examine the scenario in which all the components of the power allocation, $p_{g,m}\forall g, m$, are constants. Here, we focus on optimizing the power splitting factors $\beta_m \forall m$ under the fixed power allocation $p_{g,m}\forall g, m$. Therefore, the optimization subproblem can be rewritten as follows:

$$\max_{\{p,g,m\},\{\beta_{g,m}\}} R_{sum}\left(\beta_m\right) \tag{27}$$

$$s.t. \ 0 \leq \beta_m \leq 1 \quad \forall m \tag{28}$$

$$P_{g,m}^{EH} \geq p_{g,m}^{req} \tag{29}$$

According to (10) and constraint (29), $\beta_m \forall m$ is required to satisfy the following condition:

$$\beta_m \leq 1 - \frac{p_{g,m}^{req}}{\eta\left(\sum_{i=1}^{G}\sum_{j=1}^{|S_i|}\left\|\overline{h}_{g,m}^{H}\boldsymbol{d}_i\right\|_2^2 p_{i,j} + \sigma_v^2\right)} \cong \beta_m^{UB} \tag{30}$$

Considering (28) and (30) together, the supposed optimization problem is infeasible unless the $\beta_m^{UB} > 0 \forall m$.

Proposition 1: Assume that the process of power splitting in the receiver is almost idealized, and the noise power for all users in the gth beam is equal, that is, $|\sigma_u|^2 \rightarrow 0$. The considered optimization problem in (27)-(29) is convex with respect to the power splitting factors $\beta_{,m}\forall m$.

Proof: First, we ensure that the viable power splitting factor area is not empty and convex to guarantee the convexity of the optimization issue in (27)–(29). Because of the limitation of $\beta_m^{UB} > 0 \forall m$, the feasible area of the power splitting factor is not empty, and its convexity can be determined using Equations (29) and (30), respectively. After that, we conclude that the objective function in (27) is concave on the power splitting factors $\beta_m \forall m$. Let us recall the equation in (15), (31) and (32), as shown at the bottom of the page.

Let us assume that

$$A_{g,m} = \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 p_{g,m} \tag{33}$$

$$B_{g,m} = \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 \sum\nolimits_{j=1}^{m-1} p_{g,j} + \sum\nolimits_{i \neq g} \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_i \right\|_2^2 \sum\nolimits_{j=1}^{|S_i|} p_{i,j} + \sigma_v^2 \tag{34}$$

Given

$$R_g = \sum\nolimits_{m=1}^{|S_g|} \log_2 \left(1 + \frac{A_{g,m}\beta_m}{B_{g,m}\beta_m + \sigma_u^2} \right) \tag{35}$$

Which represents the achievable data rate on the gth beam.

Thus, $R_{sum}(\beta_m)$ in (27) is given as

$$R_{sum}(\beta_m) = \sum\nolimits_{g=1}^{G} R_g \tag{36}$$

$$R_{sum}(\beta_m) = \sum\nolimits_{g=1}^{G} \sum\nolimits_{m=1}^{|S_g|} \log_2 \left(1 + \frac{A_{g,m}\beta_m}{B_{g,m}\beta_m + \sigma_u^2} \right) \tag{37}$$

Then the first derivative of R_g with respect to β_m is given by

$$\frac{\partial R_g}{\partial \beta_m} = \frac{1}{ln2} \cdot \frac{A_{g,m}\sigma_u^2}{\left(A_{g,m}\beta_m + B_{g,m}\beta_m + \sigma_u^2 \right) \left(B_{g,m}\beta_m + \sigma_u^2 \right)} \tag{38}$$

Moreover, the second derivative of R_g with respect to β_m is given by

$$\frac{\partial^2 R_g}{\partial \beta_m^2} = -\frac{1}{ln2} \cdot \frac{A_{g,m}\sigma_u^2 \left(2\left(A_{g,m} + B_{g,m} \right) B_{g,m}\beta_m + 2B_{g,m}\sigma_u^2 + A_{g,m}\sigma_u^2 \right)}{\left(A_{g,m}\beta_m + B_{g,m}\beta_m + \sigma_u^2 \right) \left(B_{g,m}\beta_m + \sigma_u^2 \right)} \tag{39}$$

And

$$\frac{\partial^2 R_g}{\partial \beta_n \partial \beta_m} = 0 \quad \forall n \neq m \tag{40}$$

According to the equation above, the corresponding Hessian matrix $\boldsymbol{H}$ is given by

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{H}_{|S_g|} \end{pmatrix} \tag{41}$$

where $\boldsymbol{H}_m = \frac{\partial^2 R_g}{\partial \beta_m^2} \leq 0 \forall m \in \left[1, |S_g|\right]$. Correspondingly, the Hessian matrix is negative or equal to zero for all values of $\mathrm{\beta_m \forall m} \in \left[1, |S_g|\right]$, then the $\mathrm{R_g}$ is concave with respect to $\mathrm{\beta_m}$. Therefore, the objective function in (27) is concave on the power splitting factors $\beta_m \forall m$ because it represents the finite summation of concave functions. To that end, one can obtain the near-optimal solution for the optimization problem in (27)–(29) by using the Lagrangian duality-based method [30]. The corresponding Lagrangian function is formulated as in (42), shown at the bottom of the page, where $\boldsymbol{\lambda} = \left[\lambda_1, \lambda_2, \ldots, \lambda_{|S_g|}\right]^T$ and $\boldsymbol{\mu} = \left[\mu_1, \mu_2, \ldots, \mu_{|S_g|}\right]^T$ are non-negative Lagrange multipliers, which correspond to constraint (28). $\boldsymbol{\upsilon} = \left[\nu_1, \nu_2, \ldots, \nu_{|S_g|}\right]^T$ is a non-negative Lagrange multiplier corresponding to constraint (29).

Accordingly, one can express the Lagrange dual objective function as follows

$$\Gamma(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\upsilon}) = \max_{\boldsymbol{\beta}} \boldsymbol{\Upsilon}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\upsilon}) \tag{43}$$

$$R_{g,m} = \log_2 \left(1 + \frac{\left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 p_{g,m}}{\left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 \sum\nolimits_{j=1}^{m-1} p_{g,j} + \sum\nolimits_{i \neq g} \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_i \right\|_2^2 \sum\nolimits_{j=1}^{|S_i|} p_{i,j} + \sigma_v^2 + \frac{\sigma_u^2}{\beta_m}} \right) \tag{31}$$

$$= \log_2 \left(1 + \frac{\left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 \beta_m p_{g,m}}{\beta_m \left(\left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 \sum\nolimits_{j=1}^{m-1} p_{g,j} + \sum\nolimits_{i \neq g} \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_i \right\|_2^2 \sum\nolimits_{j=1}^{|S_i|} p_{i,j} + \sigma_v^2 \right) + \sigma_u^2} \right) \tag{32}$$

$$\begin{aligned} \boldsymbol{\Upsilon}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\upsilon}) = {} & \sum_{g=1}^{G} \sum_{m=1}^{|S_g|} \log_2 \left(1 + \frac{\beta_m \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 p_{g,m}}{\beta_m \left(\left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_g \right\|_2^2 \sum\nolimits_{j=1}^{m-1} p_{g,j} + \sum\nolimits_{i \neq g} \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_i \right\|_2^2 \sum\nolimits_{j=1}^{|S_i|} p_{i,j} + \sigma_v^2 \right) + \sigma_u^2} \right) \\ & + \sum_{m=1}^{|S_g|} \lambda_m \beta_m + \sum_{m=1}^{|S_g|} \mu_m (1 - \beta_m) + \sum_{m=1}^{|S_g|} \nu_m \left(\eta (1 - \beta_m) \left(\sum_{i=1}^{G} \sum_{j=1}^{|S_i|} \left\| \overline{h}_{g,m}^{H} \boldsymbol{d}_i \right\|_2^2 p_{i,j} + \sigma_v^2 \right) - p_{g,m}^{req} \right) \end{aligned} \tag{42}$$

Then, one can model the Lagrange dual optimization problem as follows

$$\min_{\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}} \Gamma\left(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}\right) \tag{44}$$

$$s.t.\ \boldsymbol{\lambda} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\mu} \succcurlyeq \mathbf{0},\ \boldsymbol{\upsilon} \succcurlyeq \mathbf{0} \tag{45}$$

To solve the Lagrange dual issue mentioned earlier, we first optimize the PS factor $\boldsymbol{\beta}$ using the provided dual variables $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon})$ using the gradient ascent technique, and then update the dual variables $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon})$ with the optimized $\boldsymbol{\beta}$ using a well-known sub-gradient methodology [31] to obtain the optimal.

1) We find the gradient direction of the Lagrange objective function in (46) regarding to power splitting factor $\beta_m\forall$m to optimize the β_m with given variables $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon})$ as follows

$$\begin{aligned}\nabla_{\beta_m}\boldsymbol{\Upsilon} = &\sum_{g=1}^{G}\frac{1}{ln2}\\ &\cdot\frac{A_{g,m}\sigma_u^2}{\left(A_{g,m}\beta_m + B_{g,m}\beta_m + \sigma_u^2\right)\left(B_{g,m}\beta_m + \sigma_u^2\right)}\\ &+\lambda_m - \mu_m - \nu_m\\ &\times\left(\eta\left(\sum_{i=1}^{G}\sum_{j=1}^{|S_i|}\left\|\overline{h}_{g,m}^{H}\boldsymbol{d}_i\right\|_2^2 p_{i,j} + \sigma_v^2\right) - p_{g,m}^{req}\right)\end{aligned} \tag{46}$$

where $A_{g,m}$ and $B_{g,m}$ are defined in (33) and (34), respectively.

Particularly, β_m can be updated using the following formula

$$\beta_m\left(Itr+1\right) = \beta_m\left(Itr\right) + \varepsilon\left(Itr\right)\nabla_{\beta_m(Itr)}\boldsymbol{\Upsilon} \tag{47}$$

where $\beta_m\left(Itr\right)$ and $\beta_m\left(Itr+1\right)$ represent the β_m in the Itr-th and $(Itr+1)$-th iterations, respectively. $\varepsilon\left(Itr\right)$ defines the updated step size for the β_m in the Itr-th iteration and satisfies the following condition:

$$\begin{aligned}\varepsilon\left(Itr\right) = arg\max_{\varepsilon}&\ \boldsymbol{\Upsilon}\\ &\left(\boldsymbol{\beta}\left(Itr+1\right),\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}\right)|_{\beta_m(Itr+1)=\beta_m(Itr)+\varepsilon(Itr)\nabla_{\beta_m(Itr)}\boldsymbol{\Upsilon}}\end{aligned} \tag{48}$$

Process in (46) is repeated until $\left|\nabla_{\beta_m(Itr)}\boldsymbol{\Upsilon}\right| \le \epsilon_1\forall m$, and the optimal power splitting factor is denoted as $\boldsymbol{\beta}^*$. Therefore, the Lagrange dual-objective function in (43) is given by

$$\Gamma\left(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}\right) = \boldsymbol{\Upsilon}\left(\boldsymbol{\beta}^*,\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}\right) \tag{49}$$

2) We update and determine the optimal Lagrange multipliers $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon})$ by solving the Lagrange dual optimization problem in (50)–(51) as follows

$$\min_{\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}} \Gamma\left(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon}\right) \tag{50}$$

$$s.t.\ \boldsymbol{\lambda} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\mu} \succcurlyeq \mathbf{0},\ \boldsymbol{\upsilon} \succcurlyeq \mathbf{0} \tag{51}$$

To state it bluntly, the dual issue is convex on the set of Lagrange multipliers $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon})$. As a result, to maximize the dual variables, a one-dimensional search strategy can be used. Nonetheless, the objective function (44) is not always differentiable; therefore, this gradient-based method is not always possible in all situations. The dual variables $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\upsilon})$ are determined using the widely used sub-gradient approach (as shown below), with the sub-gradient directions being applied as follows:

$$\nabla_{\lambda_m}\Gamma = \beta_{\boldsymbol{m}}^* \tag{52}$$

$$\nabla_{\mu_m}\Gamma = 1 - \beta_{\boldsymbol{m}}^* \tag{53}$$

$$\nabla_{\nu_m}\Gamma = \left(\begin{array}{l}\eta\left(1-\beta_{\boldsymbol{m}}^*\right)\left(\sum_{i=1}^{G}\sum_{j=1}^{|S_i|}\left\|\overline{h}_{g,m}^{H}\boldsymbol{d}_i\right\|_2^2 p_{i,j}+\sigma_v^2\right)\\ -p_{g,m}^{req}\end{array}\right) \tag{54}$$

To that end, the value of λ_m decreases if the $\nabla_{\lambda_m}\Gamma > 0$, the value of μ_m decreases if the $\nabla_{\mu_m}\Gamma > \mathbf{0}$, and the value of ν_m decreases if the $\nabla_{\nu_m}\Gamma > \mathbf{0}$. Based on this remark, we employ the binary search method [32] with an error tolerance ϵ_2 to identify the best Lagrange multipliers $\left(\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*,\boldsymbol{\upsilon}^*\right)$ for the particular scenario. Thus, the algorithms developed in steps 1 and 2 operate alternately until the duality gap no longer changes, that is,

$$\left|R_{sum}\left(\boldsymbol{\beta}^*\right) - \Gamma\left(\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*,\boldsymbol{\upsilon}^*\right)\right| = Const \tag{55}$$

where *Const* represents a non-negative constant value.

Second, we optimize the power allocation with a fixed power splitting factor in the optimization problem (22)–(26). We aim to find the power allocation $p_{g,m}\forall g,m$ under the optimized power splitting factor $\boldsymbol{\beta}^*$. However, we can rewrite the optimization problem in (22)–(26) as follows:

$$\max_{\{p_{g,m}\}} R_{sum}\left(p_{g,m}\right) \tag{56}$$

$$s.t.\ \sum\nolimits_{g=1}^{G}\sum\nolimits_{m=1}^{|S_g|} P_{g,m} \le P_T \tag{57}$$

$$p_{g,m} \ge 0 \quad \forall g,m \tag{58}$$

$$P_{g,m}^{EH} \ge p_{g,m}^{req} \tag{59}$$

Proposition 2: Assume that the process of power splitting in the receiver is almost idealized, and the noise power for all users in the gth beam is equal, that is, $|\sigma_u|^2 \to 0$. In (56)–(59), the convexity of the sub-optimization issue is determined by whether or not the feasible domain is empty.

Proof: It should be noted that the feasible domain of the sub-problems (56)–(59) is assumed to be non-empty and its convexity can be easily deduced from the constraints in (57)–(59). Next, we will examine the concavity of the objective function (56) in relation to the power allocation $p_{g,m}\forall g,m$.

Based on the assumption above, the objective function can be written as (60) and (61), shown at the bottom of the next page, where, (62), as shown at the bottom of the next page. Let us define the relationship between the m-th user and its decoding order as $m = \psi(m)$. Because the process of power splitting in the receiver is almost idealized and the noise power for all users in the gth beam is equal, the objective function can be rewritten as follows:

$$R_g = \sum_{m=1}^{|S_g|} \log_2 \left(1 + \frac{\left\|\overline{h}^H_{g,\psi(m)}\boldsymbol{d}_g\right\|_2^2 p_{g,\psi(m)}}{\left(\left\|\overline{h}^H_{g,\psi(m)}\boldsymbol{d}_g\right\|_2^2 \sum_{j=m+1}^{|S_g|} p_{g,\psi(j)} + \sigma_v^2\right)}\right) \tag{63}$$

$$R_g = \sum_{m=1}^{|S_g|} \log_2 \left(\frac{\left\|\overline{h}^H_{g,\psi(m)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,m} + \sigma_v^2}{\left(\left\|\overline{h}^H_{g,\psi(m)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,m+1} + \sigma_v^2\right)}\right) \tag{64}$$

$$R_g = \sum_{m=1}^{|S_g|} \log_2 \left(\left\|\overline{h}^H_{g,\psi(m)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,m} + \sigma_v^2\right) - \sum_{m=1}^{|S_g|} \log_2 \left(\left\|\overline{h}^H_{g,\psi(m)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,m+1} + \sigma_v^2\right) \tag{65}$$

where $\Theta_{g,m} = \sum_{j=m}^{|S_g|} p_{g,\psi(j)}$ and $\Theta_{g,m+1} = \sum_{j=m+1}^{|S_g|} p_{g,\psi(j)}$.

Now, one can find the first derivative of R_g with respect to $p_{g,\psi(m)}$ as follows:

$$\frac{\partial R_g}{\partial p_{g,\psi(m)}} = \frac{1}{ln2} \cdot \frac{\left\|\overline{h}^H_{g,\psi\ (1)}\boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi\ (1)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,1} + \sigma_v^2\right)} \quad \forall m = 1 \tag{66}$$

And, (67) as shown at the bottom of the page. Moreover, the second derivative of R_g with respect to $p_{g,\psi(m)}$ is given by

$$\frac{\partial^2 R_g}{\partial p_{g,\psi(m)} \partial p_{g,\psi(n)}} = -\frac{1}{ln2} \cdot \frac{\left\|\overline{h}^H_{g,\psi\ (1)}\boldsymbol{d}_g\right\|_2^4}{\left(\left\|\overline{h}^H_{g,\psi\ (1)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,1} + \sigma_v^2\right)^2} - \frac{1}{ln2} \cdot \sum_{l=2}^{m} \left(\frac{\left\|\overline{h}^H_{g,\psi(l)}\boldsymbol{d}_g\right\|_2^4}{\left(\left\|\overline{h}^H_{g,\psi(l)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,l} + \sigma_v^2\right)^2} - \frac{\left\|\overline{h}^H_{g,\psi(l-1)}\boldsymbol{d}_g\right\|_2^4}{\left(\left\|\overline{h}^H_{g,\psi(l-1)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,l} + \sigma_v^2\right)^2}\right) \quad \forall m \tag{68}$$

According to (68), it can easily be inferred that the Hessian matrix of R_g with respect to $p_{g,m} \forall g, m$ is negative or equal to zero. Consequently, the $\mathrm{R_g}$ is concave with respect to $p_{g,m}$. Therefore, because the sum of a finite number of concave functions stays concave, the objective function in (56) is concave on the power allocations $p_{g,m} \forall g, m$. Additionally, this study uses the Lagrangian duality-based method to obtain the near-optimal power allocation [30].

The corresponding Lagrangian function for the sub-problem in (56)–(59) is formulated as (69), as shown at the bottom of the next page.

$$R_{sum}\left(p_{g,m}\right) = \sum_{g=1}^{G} R_g \tag{60}$$

$$R_{sum}\left(p_{g,m}\right) = \sum_{g=1}^{G} \sum_{m=1}^{|S_g|} \log_2 \left(1 + \frac{\left\|\overline{h}^H_{g,m}\boldsymbol{d}_g\right\|_2^2 p_{g,m}}{\left(\left\|\overline{h}^H_{g,m}\boldsymbol{d}_g\right\|_2^2 \sum_{j=1}^{m-1} p_{g,j} + \sum_{i\neq g} \left\|\overline{h}^H_{g,m}\boldsymbol{d}_i\right\|_2^2 \sum_{j=1}^{|S_i|} p_{i,j} + \sigma_v^2\right)}\right) \tag{61}$$

$$R_g = \sum_{m=1}^{|S_g|} \log_2 \left(1 + \frac{\left\|\overline{h}^H_{g,m}\boldsymbol{d}_g\right\|_2^2 p_{g,m}}{\left(\left\|\overline{h}^H_{g,m}\boldsymbol{d}_g\right\|_2^2 \sum_{j=1}^{m-1} p_{g,j} + \sum_{i\neq g} \left\|\overline{h}^H_{g,m}\boldsymbol{d}_i\right\|_2^2 \sum_{j=1}^{|S_i|} p_{i,j} + \sigma_v^2\right)}\right) \tag{62}$$

$$\frac{\partial R_g}{\partial p_{g,\psi(m)}} = \frac{1}{ln2} \cdot \left(\frac{\left\|\overline{h}^H_{g,\psi(1)}\boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi(1)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,1} + \sigma_v^2\right)} + \sum_{l=2}^{m} \left(\frac{\left\|\overline{h}^H_{g,\psi(l)}\boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi(l)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,l} + \sigma_v^2\right)} - \frac{\left\|\overline{h}^H_{g,\psi(l-1)}\boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi(l-1)}\boldsymbol{d}_g\right\|_2^2 \Theta_{g,l} + \sigma_v^2\right)}\right)\right) \quad \forall 2 \leq m \leq \left|S_g\right| \tag{67}$$

where $\boldsymbol{\alpha} = \left[\alpha_1, \alpha_2, \ldots, \alpha_{|S_g|}\right]^T$, $\boldsymbol{\eta} = \left[\eta_1, \eta_2, \ldots, \eta_{|S_g|}\right]^T$ and $\boldsymbol{\kappa} = \left[\kappa_1, \kappa_2, \ldots, \kappa_{|S_g|}\right]^T$ are non-negative Lagrange multipliers that correspond to the constraints in (57), (58), and (59), respectively. Notably, $\boldsymbol{\alpha_n} = \left[\alpha_{n,1}, \alpha_{n,2}, \ldots, \alpha_{n,|S_g|}\right]^T$ is a non-negative Lagrange multiplier corresponding to constraint (57).

Accordingly, one can express the Lagrange dual objective function as follows

$$\overline{\Gamma}\left(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}\right) = \max_{\boldsymbol{p}} \overline{\boldsymbol{\Upsilon}}\left(\boldsymbol{p}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}\right) \tag{70}$$

Then, one can model the Lagrange dual optimization problem as follows

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}} \overline{\Gamma}\left(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}\right) \tag{71}$$

$$s.t.\ \boldsymbol{\alpha} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\eta} \succcurlyeq \mathbf{0},\ \boldsymbol{\kappa} \succcurlyeq \mathbf{0} \tag{72}$$

The proposed algorithm to solve the corresponding optimization problems consists of the following two steps:

First, we employed the gradient ascent method to determine the optimal power allocation $\boldsymbol{p}^*$. The gradient direction of the Lagrangian function with respect to the power allocation is given as

$$\begin{aligned}\nabla_{p_{g,\psi(m)}} \overline{\boldsymbol{\Upsilon}} = \frac{1}{ln2} \cdot \Bigg(& \frac{\left\|\overline{h}^H_{g,\psi(1)} \boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi(1)} \boldsymbol{d}_g\right\|_2^2 \Theta_{g,1} + \sigma_v^2\right)} \\ & + \sum_{l=2}^{m} \Bigg(\frac{\left\|\overline{h}^H_{g,\psi(l)} \boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi(l)} \boldsymbol{d}_g\right\|_2^2 \Theta_{g,l} + \sigma_v^2\right)} \\ & - \frac{\left\|\overline{h}^H_{g,\psi(l-1)} \boldsymbol{d}_g\right\|_2^2}{\left(\left\|\overline{h}^H_{g,\psi(l-1)} \boldsymbol{d}_g\right\|_2^2 \Theta_{g,l} + \sigma_v^2\right)} \Bigg)\Bigg) \\ & + \alpha_{g,m} - \eta_m \\ & + \left(\sum_{j=1}^{|S_i|} \kappa_j \eta \left(1 - \beta_j^*\right) \left\|\overline{h}^H_{g,\psi(j)} \boldsymbol{d}_g\right\|_2^2 \right)\end{aligned} \tag{73}$$

In particular, the power allocation for each user on the g-th beam ($1 \leq g \leq G$) can be sequentially updated using the following expressions:

$$p_{g,\psi(m)}\left(Itr + 1\right) = p_{g,\psi(m)}\left(Itr\right) + \overline{\varepsilon}\left(Itr\right) \nabla_{p_{g,\psi(m)}(Itr)} \overline{\Upsilon} \tag{74}$$

where $p_{g,\psi(m)}\left(Itr\right)$ and $p_{g,\psi(m)}\left(Itr + 1\right)$ represents the $p_{g,\psi(m)}$ in the Itr-th and $(Itr + 1)$-th iterations, respectively. $\overline{\varepsilon}\left(Itr\right)$ defines the updated step size for the $p_{g,\psi(m)}$ in the Itr-th iteration and it satisfies the condition $\left|\nabla_{p_{g,\psi(m)}} \boldsymbol{\Upsilon}\right| \leq \epsilon_3 \forall 1 \leq$.

The updated process in (73) and (74) for the power allocation on the g-th beam is repeated until $\left|\nabla_{p_{g,\psi(m)}} \boldsymbol{\Upsilon}\right| \leq \epsilon_3 \forall 1 \leq m \leq \left|S_g\right|$ And the optimal power allocation is denoted as $\boldsymbol{p}^*$. Therefore, the Lagrange dual-objective function in (70) is given by

$$\overline{\Gamma}\left(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\upsilon}\right) = \overline{\boldsymbol{\Upsilon}}\left(\boldsymbol{p}^*, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}\right) \tag{75}$$

Next, we can update and determine the optimal Lagrange multipliers $(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa})$ by solving the Lagrange dual optimization problem in (71)–(72) as follows:

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\upsilon}} \overline{\Gamma}\left(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}\right) \tag{76}$$

$$s.t.\ \boldsymbol{\alpha}\mathbf{0}, \quad \boldsymbol{\eta}\mathbf{0},\ \boldsymbol{\kappa}\mathbf{0} \tag{77}$$

We utilize the commonly used sub-gradient technique to find the dual variables $(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa})$,, for which the sub-gradient directions are applied in the following ways:

$$\nabla_{\alpha_m} \Gamma = p_{g,\psi(m)} \tag{78}$$

$$\nabla_{\eta_m} \Gamma = P_T - \sum_{g=1}^{G} p_{g,\psi(m)} \tag{79}$$

$$\begin{aligned}\nabla_{\kappa_m} \Gamma = & \Big(\eta \left(1 - \beta_m^*\right) \\ & \times \left(\sum_{i=1}^{G} \sum_{j=1}^{|S_i|} \left\|\overline{h}^H_{g,\psi(j)} \boldsymbol{d}_{\psi(j)}\right\|_2^2 p_{i,\psi(j)} + \sigma_v^2\right) \\ & - p^{req}_{g,\psi(m)}\Big)\end{aligned} \tag{80}$$

In this paper, we apply the binary search technique with error tolerance ϵ_4 to find the optimal solution of the Lagrange

$$\begin{aligned}&\overline{\boldsymbol{\Upsilon}}\left(\boldsymbol{p}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\kappa}\right) \\ &\quad = \sum_{g=1}^{G} \sum_{m=1}^{|S_g|} \log_2 \left(1 + 1 + \frac{\left\|\overline{h}^H_{g,\psi(m)} \boldsymbol{d}_g\right\|_2^2 p_{g,\psi(m)}}{\left(\left\|\overline{h}^H_{g,\psi(m)} \boldsymbol{d}_g\right\|_2^2 \sum_{j=m+1}^{|S_g|} p_{g,\psi(j)} + \sigma_v^2\right)}\right) \\ &\qquad + \sum_{g=1}^{G} \sum_{m=1}^{|S_g|} \alpha_{g,m} p_{g,\psi(m)} + \sum_{m=1}^{|S_g|} \eta_m \left(P_T - \sum_{g=1}^{G} p_{g,\psi(m)}\right) \\ &\qquad + \sum_{m=1}^{|S_g|} \kappa_m \left(\eta \left(1 - \beta_m^*\right) \left(\sum_{i=1}^{G} \sum_{j=1}^{|S_i|} \left\|\overline{h}^H_{g,\psi(j)} \boldsymbol{d}_{\psi(j)}\right\|_2^2 p_{i,\psi(j)} + \sigma_v^2\right) - p^{req}_{g,\psi(m)}\right)\end{aligned} \tag{69}$$

multipliers in their many forms $\left(\boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\kappa}^*\right)$. As a result, the proposed algorithm runs alternatively until the duality gap no longer changes, that is,

$$\left|R_{sum}\left(\mathbf{p}^*\right) - \overline{\Gamma}\left(\boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\kappa}^*\right)\right| = Const \tag{81}$$

where $Const$ represents a non-negative constant value.

To that end, we have developed a solution to the sub-problems to optimize the power allocation and power splitting factor. Nevertheless, the algorithm developed for the joint optimization problem in (22)–(26) is presented in **Algorithm 4**. The computational complexity of the developed method is given as

$$\mathcal{O}\left(G\left|S_g\right|^2 \log\left(\frac{1}{\epsilon_1^2}\right) \log\left(\frac{1}{\epsilon_2^2}\right) \log\left(\frac{1}{\epsilon_3^2}\right) \log\left(\frac{1}{\epsilon_4^2}\right)\right) \tag{82}$$

VI. SIMULATION RESULTS

Spectral efficiency is defined as the sum rate attained when operating within a given spectrum (16). In contrast, energy efficiency refers to the ratio between the sum rate obtained and the total power consumed [18] i.e.

$$EE = \frac{Achievable\ sum\ rate}{Total\ power\ consumption} \tag{83}$$

$$EE = \frac{R_{Sum}}{P_t + N_{RF}P_{RF} + N_{phaseshift}P_{phaseshift} + P_{BB}} \tag{84}$$

where $P_t = \sum_{g=1}^{G}\sum_{m=1}^{|S_g|} p_{g,m}$ is the total transmitted power, P_{RF} is the power consumed by each RF chain, P_{BB} represents the baseband power consumption, and $P_{phaseshift}$ is the power consumption of each phase shift. In particular, $P_{RF} = 300\ mW$, $P_{phaseshift} = 40\ mW \forall B = 4$ bit phase shifter, and $P_{BB} = 200\ mW$ are adopted as the typical values. In addition, $N_{phaseshift}$ is the number of phase shifters and is equal to NN_{RF} for hybrid precoding. Moreover, all presented results are averaged over 100 random channel realizations.

To demonstrate the performance of the proposed technique, we present the simulation results to illustrate both the spectrum efficiency and energy efficiency of the hybrid precoding architecture. The following parameters are provided for the simulation: the system's bandwidth is defined as 1 Hz, corresponding to a rate as high as possible (15). The BS and UE are equipped with uniform linear antennas (ULAs) with half-wavelength spacing. The BS is equipped with $N = 64$ antennas and $N_{RF} = 4$ RF chains, and can serve up to $K \geq N_{RF}$ UEs simultaneously. All K UEs are clustered into G = N, RF = 4 beams, with each beam consisting of more than one user simultaneously. According to equation (5), a channel vector for the mth user in the gth beam is created by considering one line-of-sight (LoS) component as well as two non-line-of-sight (NLoS) components, that is, the number of routes that the mth user takes in the gth beam ($L_{g,m} = 3$). The complex gain of the LoS path is $\alpha_{g,m}^{(1)} \sim \mathbb{CN}(0, 1)$ and the complex gains of the NLoS paths are $\alpha_{g,m}^{(l)} \sim \mathbb{CN}(0, 0.1)\ \forall 2 \leq l \leq L_{g,m}$. The azimuth angle of departure (AoD) is $\vartheta_{g,m}^{(l)}$ and elevation angle of departure $\theta_{g,m}^{(l)}$ of the lth path is assumed to follow the uniform distribution $\mathcal{U}(-\pi, \pi)\ \forall 1 \leq l \leq L_{g,m}$. The bit resolution $B = 4$ is used to quantize the phase shifters. The SNR is defined as the ratio of signal to noise $\left({}^{p_t}\!/_{\sigma^2}\right)$, where the maximum transmitted power $p_t = 30\ mW$, the minimal achievable rate for each user, is $R_{g,m}^{min}/10$, where $R_{g,m}^{min}$ is the lowest possible

Algorithm 4: Proposed Method for mmWave Massive MIMO-NOMA Systems With SWIPT

```
Input:
    Channel vectors: $\mathbf{h}_{g,m}\forall g, m$
    Digital precoding vectors: $\mathbf{d_g}\forall g$
    Noise variance: $\sigma_v^2$
    Maximum iteration times: $Itr_{max}$
Output:
    Optimal power allocation: $\mathbf{p}^* = p_{g,m}^*\forall g, m$
    Optimal power splitting factors $\boldsymbol{\beta}^* = \beta_m^*\forall m$
Initialize p and stop criteria $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$
While Loop 1:
  Step 1: Optimize the power splitting factors $\beta_m\forall m$
    under fixed power allocationp.
  While Loop 2
    Initialize dual variables $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\upsilon})$
    Solve the problem in (27) to obtain the optimal power splitting factors $\boldsymbol{\beta}^*$ according to (46)-(48).
  Until $\left|\nabla_{\beta_m(Itr)}\boldsymbol{\Upsilon}\right| \leq \epsilon_1\forall m$
    Determine the optimal Lagrange dual multipliers
      $\left(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\upsilon}^*\right)$ according to equations in (52)-(53).
  Until
      $\left|R_{sum}\left(\boldsymbol{\beta}^*\right) - \Gamma\left(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\upsilon}^*\right)\right| = Const$
  Step 2: Optimize the power allocation with fixed
    the power splitting factors $\boldsymbol{\beta}^*$
    While Loop 3
      Initialize the power splitting factors $\boldsymbol{\beta}^*$
      Solve the problem in (56) to obtain the optimal power allocation $\mathbf{p}^*$ according to (70)-(74).
    Until $\left|\nabla_{p_{g,\psi(m)}}\boldsymbol{\Upsilon}\right| \leq \epsilon_3\forall 1 \leq m \leq \left|S_g\right|$
      Determine the optimal Lagrange dual
      multipliers $\left(\boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\kappa}^*\right)$ according to equations in (78)-(80).
  Until
      $\left|R_{sum}\left(\mathbf{p}^*\right) - \overline{\Gamma}\left(\boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\kappa}^*\right)\right| = Const$
Until
      $R_{sum}\left(\mathbf{p}^*\right) = R_{sum}\left(\boldsymbol{\beta}^*\right)$
```

rate among all users when completely digital ZF precoding is used, and the lowest amount of energy collected by each user is $p_{g,m}^{min} = 0.1\, mW$.

In the simulations, we consider the proposed method with the following four methods of mmWave massive MIMO systems with SWIPT for comparison: (1) "SWIPT-Fully digital ZF Precoding," (2) "SWIPT-Hybrid Precoding NOMA proposed in [7]," (3) "SWIPT-Hybrid Precoding NOMA proposed in [9]," and (4) "SWIPT-Hybrid Precoding OMA," where OMA is implemented for UEs in each beam. An Intel Core i5-2400S @ 1.6 GHz (4 cores) and 8 GB of RAM were used to run the simulations.

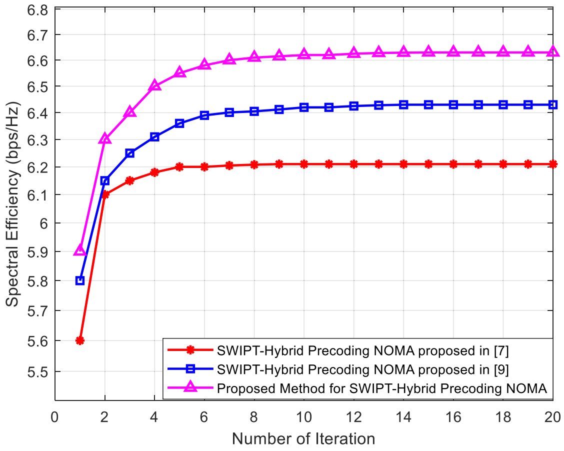


FIGURE 2. Spectrum efficiency of HP system versus the number of iterations for the joint power allocation and power splitting optimization.

Fig. 2 shows the spectrum efficiency as a function of the number of iterations, where the number of users is fixed at K = 6, and the SNR is set to 0 dB. The curves depicted in Fig. 2 illustrate the convergence of the proposed method described in Section IV, which addresses the problem of joint power allocation and power splitting for systems with fixed K users. From Fig. 2, the spectrum efficiency appears to have stabilized after the proposed method in Section IV has been iterated 13 times, which demonstrates the convergence of the proposed method. However, our proposed techniques require approximately 13 iterations for the combined power allocation and power splitting optimization to converge, whereas the SWIPT-Hybrid Precoding NOMA described in [7] requires approximately nine iterations for convergence. The SWIPT-Hybrid Precoding NOMA described in [7] converges to a greater spectrum efficiency than our proposed method. According to the SWIPT-Hybrid Precoding NOMA described in [9], the joint power allocation and power splitting optimization require approximately 12 iterations to converge. Therefore, to guarantee that each scheme can remain stable during the simulations, the number of iterations for the power allocation and power slitting optimization is set to 14.

The spectrum efficiency of the system is illustrated in Fig. 3. We consider the spectrum efficiency, SNR, and the number of users to determine which of the four signal-processing methods offers the best tradeoff between performance and cost. Because of NOMA's greater spectrum efficiency, we can say that the proposed mmWave massive MIMO-NOMA systems with SWIPT can achieve better spectrum efficiency than that of mmWave massive MIMO-OMA systems with SWIPT. As can be seen in Fig. 3, the spectrum efficiency increases as the SNR increases for all the methods being examined. SWIPT-Full-digital ZF Precoding performs better in increasing the overall spectral efficiency compared to all the precoding schemes, but it requires more processing than other methods.

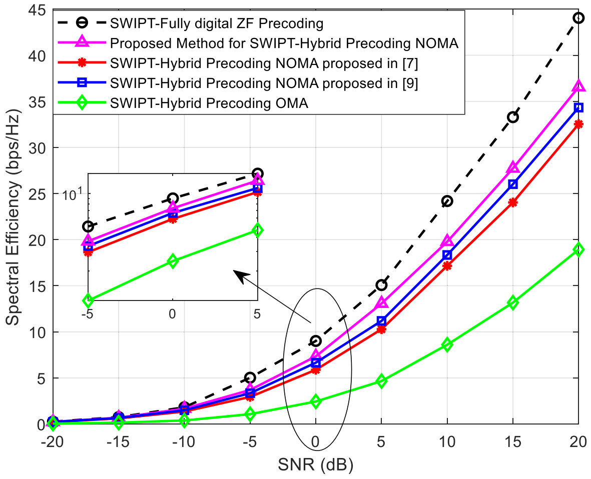


FIGURE 3. Spectrum efficiency against SNR.

Fig. 4 depicts the SNR-adjusted energy efficiency, which can accommodate up to six users. According to our findings in Fig. 4, the proposed mmWave massive MIMO-NOMA systems with SWIPT achieved greater energy efficiency than both mmWave massive MIMO-OMA systems with SWIPT and completely digital MIMO systems with SWIPT. With RF chains, as in fully digital MIMO systems, each RF chain needs 300 mW of power. Contrary to this statement, with SWIPT-Hybrid Precoding NOMA systems, the number of RF chains is significantly lower than the number of antennas.

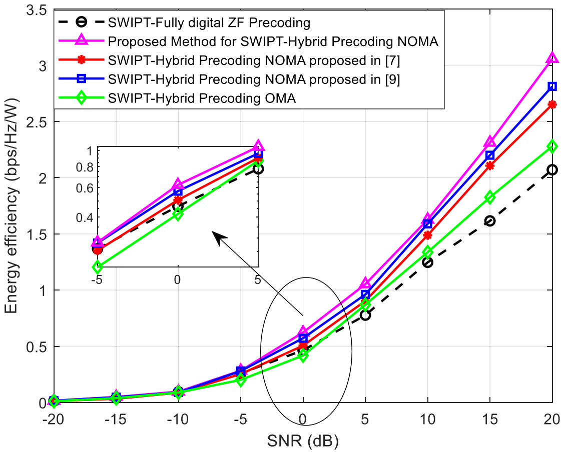


FIGURE 4. Energy efficiency against SNR.

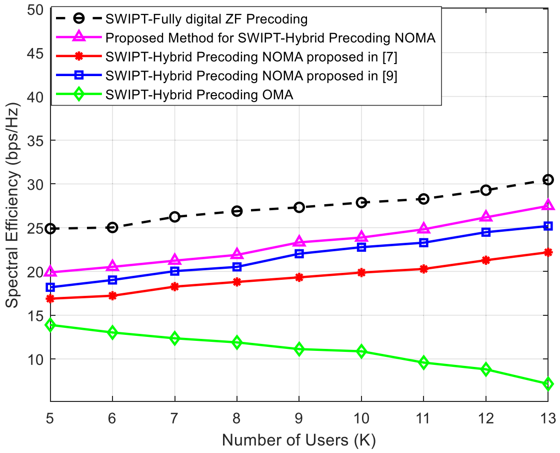


FIGURE 5. Spectrum efficiency of hybrid precoding system versus the number of users for the joint power allocation and power splitting optimization.

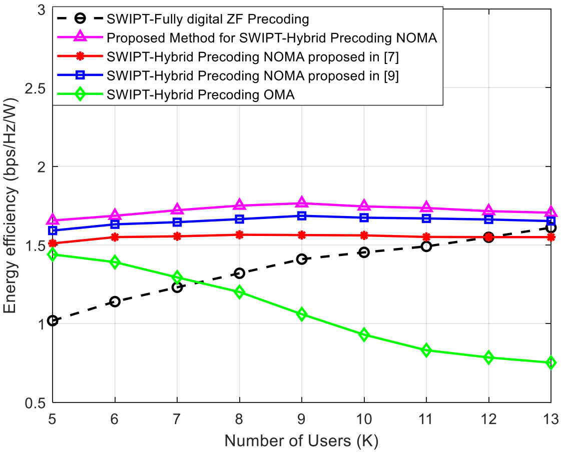


FIGURE 6. Energy efficiency of hybrid precoding system versus the number of users for the joint power allocation and power splitting optimization.

Therefore, compared to completely digital MIMO systems, RF chains generate much less energy. Furthermore, the SWIPT-enabled mmWave massive MIMO-NOMA system with hybrid precoding is shown to perform better than current systems in moderate to high SNR regimes because of the usage of NOMA.

Fig. 5 depicts a comparison of the spectrum efficiency vs. the number of UEs for all five schemes under discussion, with the SNR fixed at 10 dB for all five schemes considered. As shown in Fig. 5, the efficiency of the spectrum increases for all curves as the number of UEs increases. In this case, several UEs can share the same time-frequency resource block by utilizing intra-beam superposition coding at the base station and SIC at the receiver, which allows for greater efficiency. The proposed SWIPT-enabled mmWave huge MIMO-NOMA systems with hybrid precoding, on the other hand, outperforms the other methods and achieve performance that is comparable to the SWIPT-Full-digital ZF Precoding. As a result, using the suggested user grouping, analog RF precoder and digital baseband precoder design methods are helpful for interbeam interference cancellation while also enhancing the overall system performance and efficiency.

The energy efficiency versus the number of users is shown in Fig. 6. The SNR was adjusted to 10 dB. For illustration, Fig. 6 depicts several curves with various degrees of curvature. The energy efficiency of the SWIPT-Hybrid Precoding OMA system decreases with an increasing number of UEs. Another important observation is that the energy efficiency of the SWIPT-Full-digital ZF Precoding scheme increases with the number of UEs. Moreover, we have also noticed that the SWIPT-enabled mmWave mMIMO-NOMA system with SWIPT MMIMO-NOMA capability shows superior energy efficiency at a low and medium number of users. It increases efficiency as we go up with a number of users.

VII. CONCLUSION

In this study, hybrid precoding for SWIPT-enabled mmWave mMIMO-NOMA systems to enhance the attainable sum-rate and total energy efficiency. The optimization of user grouping is given first, followed by the creation of hybrid analog-digital precoders. Then, given the maximum transmit power budget restrictions and minimal EH need, we examined the feasible data rate maximization problem for SWIPT-enabled mmWave mMIMO-NOMA systems with PS receivers. Because of the coupling of many variables and the presence of inter-user interference, the maximization issue was non-convex, making it difficult to obtain the best solution directly. We used a decoupled strategy to solve this problem, in which the linked variables, such as power allocation and PS ratio assignment, were separated. The Lagrangian duality-based technique was then used to solve the associated subproblems. The proposed technique with hybrid precoding considerably increased the spectrum efficiency and energy efficiency of the studied system compared to existing state-of-the-art systems, demonstrating its efficacy. Furthermore, mmWave MIMO-NOMA continues to outperform mmWave MIMO-OMA.

REFERENCES

[1] N. Uwaechia, N. M. Mahyuddin, M. F. Ain, N. M. A. Latiff, and N. F. Za'bah, "On the spectral-efficiency of low-complexity and resolution hybrid precoding and combining transceivers for mmWave MIMO systems," *IEEE Access*, vol. 7, pp. 109259–109277, 2019.

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

[3] A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models design considerations antenna constructions and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, pp. 870–913, 2nd Quart. 2018.

[4] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[5] H. Xie, F. Gao, S. Zhang, and S. Jin, ''A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model,'' *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

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

[7] 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.

[8] N. Uwaechia and N. M. Mahyuddin, ''Spectrum and energy efficiency optimization for hybrid precoding-based SWIPT-enabled mmWave mMIMO-NOMA systems,'' *IEEE Access*, vol. 8, pp. 139994–140007, 2020.

[9] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. 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.

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

[11] J. Choi, G. Lee, and B. L. Evans, ''User scheduling for millimeter wave hybrid beamforming systems with low-resolution ADCs,'' *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2401–2414, Apr. 2019.

[12] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, ''Millimeter-wave massive MIMO: The next wireless revolution?'' *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[13] L. Cheng, G. Yue, D. Yu, Y. Liang, and S. Li, ''Millimeter wave time-varying channel estimation via exploiting block-sparse and low-rank structures,'' *IEEE Access*, vol. 7, pp. 123355–123366, 2019.

[14] T. Ding, Y. Zhao, L. Li, D. Hu, and L. Zhang, ''Hybrid precoding for beamspace MIMO systems with sub-connected switches: A machine learning approach,'' *IEEE Access*, vol. 7, pp. 143273–143281, 2019.

[15] X. Zhu, Z. Wang, L. Dai, and Q. Wang, ''Adaptive hybrid precoding for multiuser massive MIMO,'' *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.

[16] F. Talaei and X. Dong, ''Hybrid mmWave MIMO-OFDM channel estimation based on the multi-band sparse structure of channel,'' *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1018–1030, Feb. 2019.

[17] L. Huang, H. Lin, H. Zhang, and Y. Zhao, ''A multi-dimensional features-based clustering algorithm for massive MIMO system,'' in *Proc. 14th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Aug. 2019, pp. 423–427.

[18] 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.

[19] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. 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, pp. 998–1009, Apr. 2016.

[20] X. Gao, L. Dai, Y. Sun, and S. Han, ''Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems,'' in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[21] A. N. Uwaechia and N. M. Mahyuddin, ''A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges,'' *IEEE Access*, vol. 8, pp. 62367–62414, 2020.

[22] Q. Shi and M. Hong, ''Spectral efficiency optimization for millimeter wave multiuser MIMO systems,'' *IEEE J. Sel. Toptics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[23] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X. Xia, ''Joint Tx-Rx beamforming and power allocation for 5G millimeter-wave non-orthogonal multiple access networks,'' *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5114–5125, Jul. 2019.

[24] N. Zhao, W. Wang, J. Wang, Y. Chen, and Y. Lin, ''Joint beamforming and jamming optimization for secure transmission in MISO-NOMA networks,'' *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2294–2305, Mar. 2019.

[25] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, ''On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems,'' *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[26] Z. Wang, M. Li, Q. Liu, and A. Lee Swindlehurst, ''Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems,'' *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 256–269, May 2018.

[27] H. Zhang, A. Dong, S. Jin, and D. Yuan, ''Joint transceiver and power splitting optimization for multiuser MIMO SWIPT under MSE QoS constraints,'' *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7123–7135, Aug. 2017.

[28] T. Nguyen, M. Tran, P. Tran, and P. Tin, ''On the performance of power splitting energy harvested wireless full-duplex relaying network with imperfect CSI over dissimilar channels,'' *Secur. Commun. Netw.*, vol. 2018, pp. 1–11, Dec. 2018.

[29] Q. Qin, L. Gui, P. Cheng, and B. Gong, ''Time-varying channel estimation for millimeter wave multiuser MIMO systems,'' *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9435–9448, Oct. 2018.

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

[31] 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.

[32] L. Zhang, Y. Xin, and Y. C. Liang, ''Weighted sum rate optimization for cognitive radio MIMO broadcast channels,'' *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2950–2959, Jun. 2009.

[33] K. Singh, K. Wang, S. Biswas, Z. Ding, F. A. Khan, and T. Ratnarajah, ''Resource optimization in full duplex non-orthogonal multiple access systems,'' *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4312–4325, Sep. 2019.

[34] R. Pal, K. V. Srinivas, and A. K. Chaitanya, ''A beam selection algorithm for millimeter-wave multi-user MIMO systems,'' *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 852–855, Apr. 2018.

[35] D. W. K. Ng, E. S. Lo, and R. Schober, ''Wireless information and power transfer: Energy efficiency optimization in OFDMA systems,'' *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6352–6370, Dec. 2013.

[36] T. N. Do and B. An, ''Optimal sum-throughput analysis for downlink cooperative SWIPT NOMA systems,'' in *Proc. 2nd Int. Conf. Recent Adv. Signal Process., Telecommun. Comput. (SigTelCom)*, Jan. 2018, pp. 85–90.

[37] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, ''A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT,'' *IEEE Access*, vol. 7, pp. 17450–17460, 2019.

[38] H. M. Elmagzoub, ''On the MMSE-based multiuser millimeter wave MIMO hybrid precoding design,'' *Int. J. Commun. Syst.*, vol. 33, no. 11, p. e4409, Jul. 2020.

[39] P. Raviteja, Y. Hong, and E. Viterbo, ''Analog beamforming with low resolution phase shifters,'' *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 502–505, Aug. 2017.

[40] P.-H. Lee and Y.-P. Lin, ''Hybrid MIMO-OFDM for downlink multi-user communications over millimeter channels with no instantaneous feedback,'' in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–5.

[41] D. Zhang, Y. Wang, X. Li, and W. Xiang, ''Hybrid beamforming for downlink multiuser millimetre wave MIMO-OFDM systems,'' *IET Commun.*, vol. 13, no. 11, pp. 1557–1564, Jul. 2019.

[42] M. S. Aljumaily and H. Li, ''Machine learning aided hybrid beamforming in massive-MIMO millimeter wave systems,'' in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Nov. 2019, pp. 1–6.

[43] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, ''Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems,'' *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8315–8328, Dec. 2019.

[44] X. Zhao, Y. Zhang, S. Geng, F. Du, Z. Zhou, and L. Yang, ''Hybrid precoding for an adaptive interference decoding SWIPT system with full-duplex IoT devices,'' *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1164–1177, Feb. 2020.

AHLAM JAWARNEH was born in Irbid, Jordan, in 1981. She received the B.Sc. and M.Sc. degrees in electrical engineering from the Jordan University of Science and Technology (JUST), Irbid, in 2003 and 2008, respectively. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the Département de Génie Électrique, École de Technologie Supérieure, Montreal, QC, Canada. Her research interests include information and communications technologies, digital signal processing, nonlinear estimation and prediction, and wireless communications (5G).

MICHEL KADOCH (Life Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Université Concordia, Canada, the M.Sc. degree in electrical engineering from the Université de Carleton, and the Ph.D. degree in electrical engineering from the Université Concordia. He is currently a Professor with the Département de Génie Électrique, École de Technologie Supérieure, Montreal, QC, Canada. His research interests include information and communications technologies, IP telephony, performance analysis of telecommunication networks, simulation of telecommunication networks, telecommunication protocols, quality of service, local networks, MPLS networks and VPN, next generation networks, multicast services, wireless ad hoc networks, LTE, 4G and 5G networks, and network management.

ZAID ALBATAINEH (Member, IEEE) received the B.S. degree in electrical engineering from Yarmouk University, Irbid, Jordan, in 2006, the M.S. degree in the communication and electronic engineering from the Jordan University of Science and Technology (JUST), Irbid, in 2009, and the Ph.D. degree from the Electrical and Computer Engineering Department, Michigan State University (MSU), USA, in 2014. His research interests include blind source separation, independent component analysis, nonnegative matrix factorization, wireless communications, DSP implementation, VLSI, analog integrated circuit, and RF integrated circuit.