IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 12, 2020, accepted October 26, 2020, date of publication November 2, 2020, date of current version November 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035212

Fair Energy-Efficient Resource Allocation for Downlink NOMA Heterogeneous Networks

ZUHURA J. ALI, (Student Member, IEEE), NOR K. NOORDIN, (Member, IEEE), ADUWATI SALI, (Senior Member, IEEE), AND FAZIRULHISYAM HASHIM, (Member, IEEE)

Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia
Wireless and Photonic Networks Research Centre of Excellence (WiPNet), Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia

Corresponding authors: Nor K. Noordin (nknordin@upm.edu.my) and Zuhura J. Ali (zuhurathaputra@gmail.com)

This work was supported by UPM PUTRA BERIMPAK under Fundamental Research Grant 9584300.

ABSTRACT The increasing in energy consumptions of the current wireless networks, leads towards designing energy-efficient 5G networks. The application of non-orthogonal multiple access (NOMA) in the heterogeneous networks (HetNets) improves the spectrum utilization with the cost of efficient resource allocation. Hence, this article proposes optimal user-pairing and power allocation solutions towards achieving fair energy-efficient resource allocation in downlink femtocell NOMA-HetNets. In the proposed optimization process, the considered constraints are the user's transmission rate, transmit power budget at the base station (BS), and the interference. The energy consumption of both the transmitter and the receiver are considered to simulate the real system design. The Greedy Algorithm (GA) is used to achieve a low-complex optimal solution during the user-pairing process. Simultaneously, the max-min energy efficiency optimization approach is employed to maximize the minimum energy efficiency of the femtocell users to achieve the optimal power allocation solution. The mathematical formulation of the max-min energy efficiency is a non-convex fractional programming problem and is intractable. Thus, the fractional programming theory is adopted to transform the problem into a sequence of subtractive form, followed by the Sequential Convex Programming (SCP) approach to determine the optimal solution. Simulation results show that the proposed NOMA with optimal power allocation method using SCP and GA (NOMA-SCP-GA) achieves fair energy efficiency performance with lower complexity compared to the benchmark methods. Moreover, the minimum energy efficiency of the femtocell user is 38.22% higher than NOMA with Difference of Convex programming (NOMA-DC). The NOMA-SCP-GA method can assure 5G capability demands.

INDEX TERMS Greedy algorithm, heterogeneous network, non-orthogonal multiple access, power allocation, sequential convex programming.

I. INTRODUCTION

The growing demands of smart devices for high data rate creates new challenges in the current wireless cellular system, particularly spectrum allocation. The advent of 5G and other improved technological advancement is expected to address the higher data rate achievement and meeting the demand for massive wireless connectivity [1]. One of the key technologies to accomplish these objectives is by exploiting heterogeneous networks (HetNets). The HetNets are introduced to improve the network capacity growth, spectral efficiency improvement with low energy consumption, and consequently improve the overall network performance [2]–[4]. HetNets are composed of the macrocells and overlaid small-cells (e.g., microcells, picocells, and femtocells). The macrocells and small-cells are distinct in power transmission systems and various data processing capabilities, enabling them to support different radio access technologies. The main purpose of employing small-cells into the macrocells is to improve the coverage and increase the frequency reuse of the network. The macrocells have the most extensive coverage area, and uses High Powered Base Stations (HBSs), while the small-cells have smaller coverage areas with Low Powered Base Stations (LBSs) usage [2], [5].

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

Recently, non-orthogonal multiple access (NOMA) has emerged as a promising technology that can meet the 5G requirements in solving the spectrum scarcity problems [6]–[8]. The main operating principle of NOMA is to allow non-orthogonal spectrum usage by multiple users at the same time, frequency, and code. The superposition coding (SC) of the multiple user's signals is performed at the transmitter with allowable inter-user interference, while the successive interference cancellation (SIC) is performed at the receiver to separate the multiple user's signals accurately. However, the detection process at the receiver is highly complicated due to the multiple signals combination [9], [10]. The most popular type of NOMA is the power domain NOMA (PD-NOMA) in which powers are assigned to the users based on their channel conditions whereby strong channel gain users are assigned with lower power compared to weak channel gain users [11], [12].

Recently, NOMA has been employed in HetNets to harness the benefits of both technologies and improve on the spectral efficiency, and energy efficiency as well as coverage area. However, the combination of NOMA and HetNets faces the challenges of co-tier and cross-tier interferences as well as efficient resource allocation [12]. Thus, the efficient resource allocation and interference management are the fundamentals aspects to be considered in the design of the hybrid NOMA-HetNets system. In such a system, the overall energy efficiency is an important performance index to be considered [13]. It is also significant to consider the transmitter and receiver power consumption for energy efficiency resource allocation design [14]. The reason is that the Base Station (BS) transmission power and user circuit power consumptions are not equal even in short-range communication systems, such as in femtocells networks and wireless sensor networks.

The research studies in [15]–[18] are mainly focused on solving the resource allocation problem based on overall energy efficiency without considering the transmitting and receiving power consumption in the optimization process. This leads to unfairness among users because such designs favor the stronger channel gain users over the weaker channel gain users [19]. Fairness is an essential attribute to be considered in resource allocation to prevent resource deprivation and misuse in wireless systems [20]. The study on the fairness-based energy-efficient resource allocation in PD-NOMA-HetNets have earlier addressed in [21] by investigating the joint subcarrier and power allocation problem with the tradeoff between energy efficiency, fairness, and energy harvesting, while the authors in [22] considered the tradeoff between energy efficiency and user fairness in a multicarrier NOMA system. However, these works also did not consider the transmitter and receiver power consumption.

On the contrary, this article investigates the fairness of energy-efficient resource allocation problems in downlink femtocell NOMA-HetNets. The fairness-based energy efficiency optimization focused on maximizing the ratio of the individual user rate to its power consumption (bits/Joule) [23], [24]. This study focused on both transmitter and receiver power consumption, with attention centered on the presence of both co-tier and cross-tier interference within the network. These considerations provide more accurate modeling of the wireless communication system to achieve fairness in energy efficiency among users. The max-min energy efficiency adoption as the primary objective function helps to achieve system fairness and preserve integrity among users [19]. It ensures fairness for all users by maximizing their minimum energy efficiency for their assigned sub-channels.

The considered resource allocation problem is a joint user-pairing and power allocation problem in which attaining a system global optimal solution is very complex. Therefore, the joint problem was firstly decomposed into two subproblems. The first sub-problem is approached by deploying the Greedy Algorithm (GA) to solve the user-pairing problem [22]. On the other hand, the second sub-problem, which is the power allocation problem is addressed by formulating the fair energy efficiency as a maximization objective function problem, which is a mixed-integer non-convex fractional programming problem. Hence, it is then transformed into its subtractive form to obtain the optimal power allocation solutions by adopting the Sequential Convex Programming (SCP) approach [22].

Based on the literature, the investigation on energy efficiency maximization that ensures fairness among users using a max-min optimization approach, with considerations to interference, the transmitter, and the receiver power consumption has not been considered in NOMA-HetNets. Meanwhile, the fair energy-efficient resource allocation is of critical importance in achieving fair resources to each user in the 5G and upcoming future generations. With this motivation, this study proposes and demonstrates a novel approach of investigating the fair energy-efficient resource allocation problem for the downlink femtocell NOMA-HetNets, with the consideration to the co-tier, and cross-tier interference, the transmitter, and receiver power consumptions. The main contributions of this study are summarized as follows:

- A user-pairing approach based on GA is proposed by assigning the two users to a certain effectively utilized sub-channel. For each stage, two users are selected and allocated to an optimal sub-channel that maximized their energy efficiency. This approach achieves the sub-optimal performance with low computational complexity than the exhaustive search approach.
- A formulation of the fairness-based energy efficiency maximization problem is developed by using max-min approach with the following satisfied constraints: the user's transmission rate, transmit power budget at the BS, and interference from macrocell's users. The considered formulated problem is non-convex and intractable, so the fractional programming theory is adopted to transform the problem into a sequence of subtractive form to solve it with lower complexity and easily determine the sub-optimal solution.

- The SCP optimization approach is proposed to solve the sub-channel power allocation problem. On assuring the optimal sub-channel power allocation solution is achieved, the SCP optimization approach holds the convex part of the optimization problem correctly and keeps the non-convex part accurate. The obtained power allocation value is updated iteratively until the optimal power allocation solution is found.
- A new algorithm termed as fair energy-efficient power allocation (FEPA) is developed to find the optimal power that can assign to each of the paired users iteratively. Simulations verify that the proposed power allocation, alongside with user pairing, leads to more remarkable improvement in terms of energy efficiency and fairness compared to the benchmark methods.

The rest of this article is organized as follows. Section II explores the system model, the power consumption model, and the energy efficiency metric. Section III, presented the proposed resource allocation solutions. Section IV, describes the simulation results and discussions to show the significant improvement of the proposed solution. The paper concluded in Section V.

A. RELATED WORKS

The efficient resource allocation is essential to obtain high performance of the PD-NOMA system. In line with this objective, several studies proposed to address the resource allocation problems in single-cell scenarios. The power allocation solutions were introduced to maximize the sum-rate for PD-NOMA system in uplink transmission [25], [26], downlink transmission [27], [28], and both uplink and downlink transmission [29]. Besides, the power allocation solutions have also been proposed to maximize the energy efficiency for the downlink PD-NOMA system in [30]–[34].

On the other hand, to address the resource allocation problems in multi-cell scenarios, recent studies have been introduced for the sum-rate [35]–[41] and energy efficiency maximization [15]–[18]. The reviews from [35]–[41] have analyzed the effect of power allocation and user scheduling problems on maximizing sum-rate in NOMA-HetNets. The authors in [35] investigated the power allocation problem using a Stackelberg game approach by adopting a distributed power allocation algorithm to achieve the Stackelberg equilibrium. In [36], the Karush-Kuhn-Tucker (KKT) conditions for multi-cell multi-user based NOMA system was proposed under BS power and user's transmission rate constraints for efficient power allocation. However, these works did not tackle the issue of user-pairing.

To maximize the overall system throughput, the authors in [37] proposed user scheduling and iterative distributed power control algorithms to obtain the sub-optimal solution for both small-cells and macro-cells. The proposed algorithms achieved higher spectral efficiency and lower outage performance, but the system complexity is high. The Initialization Algorithm (IA) and Swap Operation Enabled Matching Algorithm (SOEMA) proposed to pair the users into the sub-channel to maximize the small-cells user's throughput [38]. However, the proposed algorithms have shown lower computational complexity, but the optimal power not assigned to the users.

Random and selective pairing solution with Fixed Power Allocation (FPA) was introduced for both the downlink and uplink multi-cell scenario in [39], where the average achievable data rate and outage probability evaluated. The proposed solution has low computational complexity from the fixed power allocation to the users, but this does not guarantee an optimal performance. The collaborative communication method was introduced in [40] to improve system throughput and reduce inter-cell interference. The sorting based user-pairing approach was introduced in [41] for cognitive radio in NOMA-HetNets system. The KKT condition and difference of convex (DC) algorithm were used to obtain the optimal power within and across the sub-channel, respectively.

The power and user scheduling optimization for energy efficiency maximization in downlink NOMA-HetNets have been presented in [15]–[18]. The fair power allocation with a given BS power constraint was proposed in [15], while the sub-channel and power allocation problems for small-cells were investigated in [16]. But these studies disregard the user-pairing issue. The performance of NOMA with non-uniform small-cells deployment has been evaluated with the consideration for both user-pairing and user association to obtain higher data rate and energy efficiency [17].

The perfect Channel State Information (CSI) and imperfect CSI has been considered in [18], where the matching theory was adopted to pair two users in a sub-channel, and the Fractional Transmitting Power Allocation (FTPA) was used to assign power for each user accordingly. The proposed approach solution has low computational complexity, but the performance is degraded due to the changing of users' channel conditions. The studies in [42]–[44] have considered the power control aspects in the NOMA system. The sum power minimization and sum-rate maximization were considered in the multi-cell NOMA system, while the sum power minimization in two cells NOMA networks was considered for downlink transmission in [42], [43] respectively, and uplink transmission in [44].

From the aforementioned studies, it can be deduced that, there is no comprehensive solution that addresses the fair energy-efficient resource allocation problem with consideration to the co-tier and cross-tier interference, alongside with transmitter and receiver energy power consumptions.

II. SYSTEM MODEL

This section describes the proposed system model and parameters of the NOMA-HetNets, alongside with the description of the applied power consumption model.

A. SYSTEM MODEL

The system model represents a downlink femtocell NOMA-HetNets consisting of one macro BS, overlaid with a

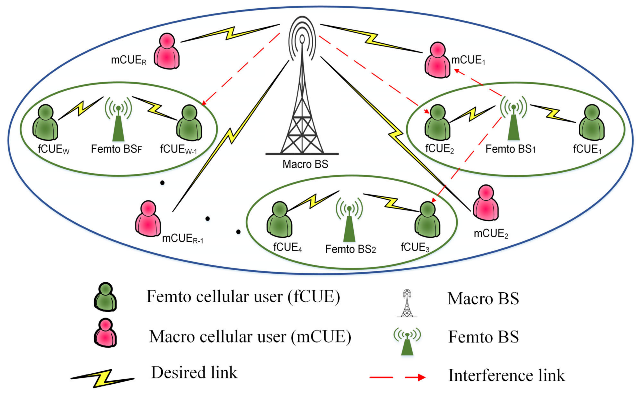

FIGURE 1. The system model of downlink femtocell NOMA-HetNets.

set of $\mathcal{F}$ Femto BSs. All BSs and all users are assumed to be equipped with a single antenna, which is single-input single output (SISO) system, as shown in Figure 1.

For simplicity, we denote macro BS as BS_m and femto BS as BS_f, where $m = 1$ and $f \triangleq \{1, 2, \ldots, \mathcal{F}\}$. The indices $\mathcal{U}_r \triangleq \{1, 2, \ldots, \mathcal{R}\}$ and $\mathcal{U}_w \triangleq \{1, 2, \ldots, \mathcal{W}\}$ represent the set of macrocell users (mCUEs) and femtocell users (fCUEs) served by BS_m and BS_f respectively. NOMA technology is adopted to serve the users in this HetNets using $\mathcal{S}$ sub-channels. Hence, mCUEs and fCUEs share the same resources through these $\mathcal{S}$ sub-channels. The total bandwidth of the system is represented by B_w. The bandwidth for each sub-channel is obtained by dividing B_w equally among $\mathcal{S}$ sub-channels. Where $B_{sch} = B_w/\mathcal{S}$. The $P^m_{\max}$ and $P^f_{\max}$ are defined parameters for the maximum transmitted power budget for BS_m and BS_f respectively. The p^m_s and p^f_s are the power assigned for each sub-channel from BS_m and BS_f respectively. The Rayleigh fading model which depends on distance and path loss attenuation, is employed.

The $\text{UE}^f_{w,s}$, $\text{UE}^q_{w,s}$, and $\text{UE}^m_{r,s}$ are used to indicate w-th fCUE in sub-channel-s served by BS_f, w-th fCUE in sub-channel-s served by BS_q, where $(f \neq q)$ and r-th mCUE in sub-channel-s is served by BS_m respectively. We assume that channel coefficients of all fCUEs from BS_f are arranged as seen in (1).

$$h^f_{1,s} \leq h^f_{2,s} \leq \cdots \leq h^f_{w,s} \cdots \leq h^f_{\mathcal{W},s} \tag{1}$$

where,

$h^f_{w,s}$ represents the channel fading coefficient of $\text{UE}^f_{w,s}$ defined as $h^f_{w,s} = g^f_{w,s}\chi^f_{w,s}$. The symbol $g^f_{w,s}$ is a parameter that represents Rayleigh fading, and $\chi^f_{w,s}$ is the shadowing and path loss between w-th fCUE with the associated BS_f.

According to the NOMA principle, the BS uses SC to transmit the multiple user's signals with different power levels. At the receiver, each user receives the desired signal from its served BS and the undesired signals (interfering signals) from other BSs. This proposed model considered the interfering signals that come from the same type of BSs (i.e., from BS_f to BS_q) which are typically called co-tier interference and from a different type of BSs (i.e., from BS_m to BS_f) which are known as cross-tier interference. We use this model to achieve realistic performance for both fCUEs as well as mCUEs. Hence, the received signal by $\text{UE}^f_{w,s}$ can be mathematically expressed as in (2).

$$\begin{aligned} y^f_{w,s} = {} & \underbrace{h^f_{w,s}\sqrt{p^f_{w,s}}x^f_{w,s}}_{\text{desired signal}} + \underbrace{\sum_{i=w+1}^{\mathcal{W}} h^f_{w,s}\sqrt{p^f_{i,s}}x^f_{i,s}}_{\text{intra-NOMA user interference}} \\ & + \underbrace{\sum_{q=1,q\neq f}^{\mathcal{F}} \alpha^q_{w,s}h^{f,q}_{w,s}\sqrt{p^q_{w,s}}x^q_{w,s}}_{\text{co-tier interference}} \\ & + \underbrace{\sum_{r=1}^{\mathcal{R}} \alpha^m_{r,s}h^{f,m}_{r,s}\sqrt{p^m_{r,s}}x^m_{r,s}}_{\text{cross-tier interference}} + \underbrace{z^f_{w,s}}_{\text{noise}} \end{aligned} \tag{2}$$

The desired signal term in (2) represents the desired transmitted signal of the $\text{UE}^f_{w,s}$, where $x^f_{w,s}$ and $p^f_{w,s}$ are the desired transmitted symbol and the power assigned to $\text{UE}^f_{w,s}$ respectively. The intra-NOMA user interference term represents the interference caused by other users. In the co-tier interference term, $x^q_{w,s}$, $h^{f,q}_{w,s}$, and $p^q_{w,s}$ are the desired transmitted symbol, the channel fading coefficient (between BS_f and $\text{UE}^q_{w,s}$), and the power assigned to $\text{UE}^q_{w,s}$, respectively. In cross-tier interference, the term $x^m_{r,s}$, $h^{f,m}_{r,s}$, and $p^m_{r,s}$ are the desired transmitted symbol, the channel fading coefficient (between $\text{UE}^m_{r,s}$ and BS_f), and the power assigned to $\text{UE}^m_{r,s}$ respectively. Also $h^{m,f}_{w,s}$ is the channel fading coefficient between $\text{UE}^f_{w,s}$ and BS_m. The noise term $z^f_{w,s} \sim \mathcal{CN}\left(0, \left(\sigma^f_{w,s}\right)^2\right)$ denotes the Additive White Gaussian Noise (AWGN) power at $\text{UE}^f_{w,s}$ with zero mean and the noise variance of $\left(\sigma^f_{w,s}\right)^2$. The parameters $\alpha^f_{w,s} \in \{0, 1\}$ and $\alpha^m_{r,s} \in \{0, 1\}$ are the binary variables that represent the sub-channel allocation indicator for the fCUEs and mCUEs in (3) and (4), respectively.

$$\alpha^f_{w,s} = \begin{cases} 1, & \text{if subchannel } s \text{ is assigned to fCUE } w \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

and,

$$\alpha^m_{r,s} = \begin{cases} 1, & \text{if subchannel } s \text{ is assigned to mCUE } r \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The NOMA technology allows more than one user in the same sub-channel and should apply the correct demodulation process at the receiver using the SIC technique. The SIC can remove interference by arranging all users according to their assigned power [38]. With this consideration in place, the $\text{UE}^f_{w,s}$ can decode all (w-1)-th fCUEs' signals efficiently and treat all ($w + 1$)-th fCUEs' signals as noise. Assuming that all BSs have perfect knowledge of CSI, then the Signal-to-Interference-plus-Noise-Ratio (SINR) received by $\text{UE}^f_{w,s}$ can be represented as in (5), shown at the bottom of the next page. Equation (5) can be simplified as expressed in (6), as shown at the bottom of the page, where, $\mathcal{I}_{w,s}^{f} = \left|h_{w,s}^{f}\right|^2 \sum_{i=w+1}^{\mathcal{W}} p_{i,s}^{f}$ is the intra-NOMA user interference, $\mathcal{I}_{co}^{f,q} = \alpha_{w,s}^{q} \left|h_{w,s}^{f,q}\right|^2 \sum_{q=1,q\neq f}^{\mathcal{F}} p_{w,s}^{q}$ is the co-tier interference, and $\mathcal{I}_{cr}^{f,m} = \alpha_{r,s}^{m} \left|h_{r,s}^{f,m}\right|^2 \sum_{r=1}^{\mathcal{R}} p_{r,s}^{m}$ is the cross-tier interference. Therefore, the $\mathrm{UE}_{w,s}^{f}$ achievable data rate measured in bits per second (bps) is given as expressed in (7).

$$R_{w,s}^{f}(\alpha, p) = \alpha_{w,s}^{f} B_{sch} \log_2\left(1 + \gamma_{w,s}^{f}\right) \text{ (bps)} \tag{7}$$

where,

$\alpha = (\alpha_{1,s}^{f}, \alpha_{2\ s}^{f}, \cdots, \alpha_{\mathcal{W},s}^{f})$ represent the set of sub-channel allocation parameters, while $p = (p_{1,s}^{f}, p_{2,s}^{f}, \cdots, p_{\mathcal{W},s}^{f})$ which is the power allocation values for each $\mathrm{UE}_{w,s}^{f}$. Accordingly, the total sum rate of all femtocells is calculated as in (8).

$$R_{sum}^{f}(\alpha, p) = \sum_{f=1}^{\mathcal{F}} \sum_{w=1}^{\mathcal{W}} \sum_{s=1}^{\mathcal{S}} \alpha_{w,s}^{f} R_{w,s}^{f} \tag{8}$$

B. POWER CONSUMPTION MODEL

The power consumption parameters model is executed based on [45], by considering the power consumption from the transmission side of the BSs, as well as from the users' receiver side. The power consumption of the femtocell is given by (9).

$$P_{T(w,s)}^{f}(\alpha, p) = \underbrace{\frac{1}{\xi_f} \sum_{w=1}^{\mathcal{W}} \sum_{s=1}^{\mathcal{S}} \alpha_{w,s}^{f} p_{w,s}^{f}}_{\text{dynamic power}} + \underbrace{P_{st}^{f} + \sum_{w=1}^{\mathcal{W}} P_{st}^{w}}_{\text{static power}} \tag{9}$$

where, the first component of (9) is the dynamic power, which is constituted by the transmission side parameters (i.e., the BS_f). This dynamic power corresponds to the power dissipated by the radio frequency (RF) signals in the power amplifiers [45], [46]. The term $\xi_f \in \{0, 1\}$ represents the efficiency of the power amplifier at the BS_f. The second component is the static power, which is made up of two constant parameters P_{st}^{f} and P_{st}^{w}. The P_{st}^{f} corresponds to the power consumed by operating circuits and transmitted signal systems (such as cooling system, filters etc.) at the BS_f. The P_{st}^{w} corresponds to operating circuits, which is responsible for the reception of the signal at the receiving side i.e., fCUE. The mathematical expression in (9) can be simply written as (10).

$$P_{T(w,s)}^{f}(\alpha, p) = \frac{1}{\xi_f} \sum_{w=1}^{\mathcal{W}} \sum_{s=1}^{\mathcal{S}} \alpha_{w,s}^{f} p_{w,s}^{f} + \tilde{P}_{st}^{f} \tag{10}$$

where,

$$\tilde{P}_{st}^{f} = P_{st}^{f} + \sum_{w=1}^{\mathcal{W}} P_{st}^{w} \tag{11}$$

Similarly, the power consumption for macrocell is given as displayed in (12). All corresponding parameters definitions are similar to the femtocell, except that they are related to BS_m and mCUE.

$$P_{T(r,s)}^{m}(\alpha, p) = \underbrace{\frac{1}{\xi_m} \sum_{r=1}^{\mathcal{R}} \sum_{s=1}^{\mathcal{S}} \alpha_{r,s}^{m} p_{r,s}^{m}}_{\text{dynamic power}} + \tilde{P}_{st}^{m} \tag{12}$$

where,

$$\tilde{P}_{st}^{m} = P_{st}^{m} + \sum_{r=1}^{\mathcal{R}} P_{st}^{r} \tag{13}$$

C. ENERGY EFFICIENCY METRIC

Energy efficiency is an important parameter in evaluating cellular system design performance targeting towards reducing the system energy consumption. By definition, energy efficiency $\left(\mathcal{Q}^{EE}\right)$ is the ratio of throughput (achievable data rate) to the total power consumption [47], [48]. It is measured in bits/Joule as expressed in (14).

$$\mathcal{Q}^{EE} = \frac{\mathit{Throughput\ (bps)}}{\mathit{Total\ Power\ Consumption}\ (\mathrm{Joule}/\mathrm{s})} \tag{14}$$

$$\gamma_{w,s}^{f} = \frac{p_{w,s}^{f} \left|h_{w,s}^{f}\right|^2}{\underbrace{\left|h_{w,s}^{f}\right|^2 \sum_{i=w+1}^{\mathcal{W}} p_{i,s}^{f}}_{\text{intra-NOMA user interference}} + \underbrace{\alpha_{w,s}^{q} \left|h_{w,s}^{f,q}\right|^2 \sum_{q=1,q\neq f}^{\mathcal{F}} p_{w,s}^{q}}_{\text{co-tier interference}} + \underbrace{\alpha_{r,s}^{m} \left|h_{r,s}^{f,m}\right|^2 \sum_{r=1}^{\mathcal{R}} p_{r,s}^{m}}_{\text{cross-tier interference}} + \underbrace{\left(\sigma_{w,s}^{f}\right)^2}_{\text{noise}}} \tag{5}$$

$$\gamma_{w,s}^{f} = \frac{p_{w,s}^{f} \left|h_{w,s}^{f}\right|^2}{\underbrace{\mathcal{I}_{w,s}^{f}}_{\text{intra-NOMA user interference}} + \underbrace{\mathcal{I}_{co}^{f,q}}_{\text{co-tier interference}} + \underbrace{\mathcal{I}_{cr}^{f,m}}_{\text{cross-tier interference}} + \underbrace{\left(\sigma_{w,s}^{f}\right)^2}_{\text{noise}}} \tag{6}$$

The energy efficiency of $UE^f_{w,s}$ can be obtained as in (15),

$$Q^{EE}_{w,s}(\alpha, p) = \frac{\sum_{s\in\mathcal{S}} R^f_{w,s}(\alpha, p)}{\frac{1}{\xi_f}\sum_{s\in\mathcal{S}} \alpha^f_{w,s} p^f_{w,s} + \tilde{P}^f_{st}} \tag{15}$$

Equation (15) can be simply written as in (16),

$$Q^{EE}_{w,s}(\alpha, p) = \frac{R^f_{w,s}(\alpha, p)}{P^f_{T(w,s)}(\alpha, p)} \tag{16}$$

The sum energy efficiency for the corresponding femtocell NOMA-HetNets can be determined as given in (17).

$$Q^{EE,f}_{sum}(\alpha, p) = \frac{\sum_{f=1}^{\mathcal{F}}\sum_{w=1}^{\mathcal{W}}\sum_{s=1}^{\mathcal{S}} \alpha^f_{w,s} R^f_{w,s}}{\frac{1}{\xi_f}\sum_{w=1}^{\mathcal{W}}\sum_{s=1}^{\mathcal{S}} p^f_{w,s} + \tilde{P}^f_{st}} \tag{17}$$

III. PROPOSED RESOURCE ALLOCATION SOLUTIONS

In the proposed system model, initially, the study assumes that the user association process has been performed by using state-of-the-art algorithms such as those discussed in [17], [49]. Note that our proposed user-pairing and power allocation solutions are independent of the used user association algorithm. The user-pairing and power allocation problem evaluation is then followed, which is the main focus of this study. The proposed solution is achieved by NOMA-SCP-GA which is referred to as applying SCP solution after pairing users based on GA to obtain the optimal and fair energy efficiency in femtocell NOMA-HetNets.

The user-pairing and power allocation form a joint optimization problem that affects the performance of NOMA systems. To decrease the computational complexity in solving this impending joint problem, it is decoupled into two sub-problems. The first section addresses the user-pairing problem solution using GA to assign two users to the same sub-channel. The second section uses the SCP approach to solve the power allocation problem to obtain the optimal and fair energy efficiency for each sub-channel and each user. The summary for the proposed resource allocation solution which is NOMA-SCP-GA is shown in Figure 2.

A. ENERGY EFFICIENT USER-PAIRING BASED ON GREEDY ALGORITHM (GA)

The user-pairing phenomenon highly affects the NOMA system performance. The designing of an optimal user-pairing solution is a requirement for enhancing the NOMA system performance. Exhaustive search approach can be used to achieve optimal system performance by allowing all possible combinations of user-paring. However, as the number of the users increase, the computational complexity also increases. Thus, to overcome this issue, this article employs a GA approach with lower computational complexity and with the capability of achieving a sub-optimal user-pairing solution for the femtocell NOMA-HetNets.

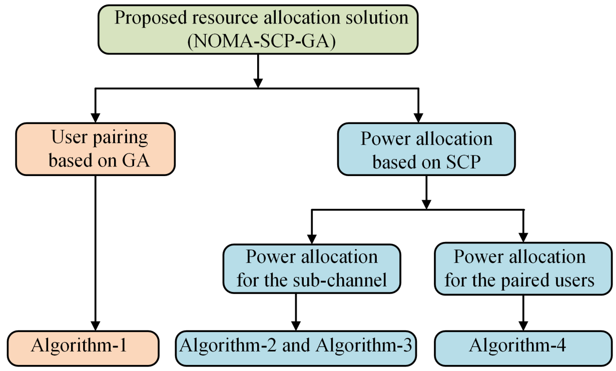

FIGURE 2. The proposed resource allocation solution (NOMA-SCP-GA).

With an assumption that only two users are assigned to each sub-channel, equal power is allocated for each sub-channel with fair power allocation for each paired users in each sub-channel as adopted in [15]. The fair power allocation assigned to the first user $\left(UE^f_{1,s}\right)$ is determined by (18).

$$p^f_{1,s} = \frac{p^f_s}{1 + \delta^f_{1,s} + \delta^f_{1,s}\delta^f_{2,s} + \cdots + \delta^f_{1,s}\delta^f_{2,s}\cdots\delta^f_{\mathcal{W}_s-1,s}} \tag{18}$$

where,

$\mathcal{W}_s$ is the number of paired users in the sub-channel and $\delta^f_{1,s}$ is the ratio of channel coefficient between the first and the second user expressed as (19).

$$\delta^f_{1,s} = \frac{G^f_{1,s}}{G^f_{i+1,s}} \tag{19}$$

where,

$$G^f_{1\ s} = \left|h^f_{1,s}\right|^2 \Big/ \left(\mathcal{I}^f_{w,s} + \mathcal{I}^{f,q}_{co} + \mathcal{I}^{f,m}_{cr} + \left(\sigma^f_{w,s}\right)^2\right) \tag{20}$$

The power assigned to the other paired user $\left(UE^f_{w_i,s}\right)$ is given as expressed in (21).

$$p^f_{w_i,s} = \frac{\left(\prod_{l=1}^{i-1}\delta^f_{w_l,s}\right)p^f_s}{1+\delta^f_{w_1,s}+\delta^f_{w_1,s}\delta^f_{w_2,s} + \cdots + \delta^f_{w_1,s}\delta^f_{w_2,s}\cdots\delta^f_{\mathcal{W}_s-1,s}} \tag{21}$$

The main idea for using GA is to select the best choice at each stage of the problem that possibly achieves the best optimal solution. The adoption of GA is focused on assigning two optimal users that can maximize their energy efficiency. Algorithm-1 describes the GA at assigning the two users to a certain effectively utilized sub-channel. For each stage, two users are selected and allocated to optimal sub-channel that maximized their energy efficiency. This process terminated after all users assigned to their best-utilized sub-channels. The procedures for algorithm-1 is divided into two processes; initialization process and sub-channel assignment process.

(a) *Initialization process*: Algorithm-1 starts by initializing sets of users, sub-channels, allocation variable, superimposed users, the empty data rate for each user, and power allocation for each sub-channel. Then, the algorithm obtains the SINR for each user and arranges them in ascending order (i.e., high to low SINR) to assign power to each user as defined in (18) and (21), then, the data rate for each user is updated.

(b) *Sub-channel assignment process*: The first user is assigned to a selected sub-channel if it has a higher SINR compared to other users and if it couldn't acquire its desired data rate. The first user should also satisfy the condition of having a co-channel interference that is lower than the threshold of the assigned sub-channel. After the first user selection, the superposition set and the user sub-channel allocation parameter are updated. Then this user is removed from the user set.

On the other hand, the second user is selected if it has a lower energy efficiency compared to other users and if its energy efficiency can improve by adding the energy efficiency of the first selected user. After selecting the second user, the superposition set, and the user sub-channel allocation parameter are all updated, the user then removed from the user set.

Consequently, the optimal pairing has been achieved, and the optimal sub-channel is removed from the sub-channel set. This process continues until all users are assigned to their best-utilized sub-channels. The summary of algorithm-1 is shown in Table 1.

TABLE 1. **Algorithm-1: Greedy Algorithm (GA) for User-Pairing.**

1: **Initialize**
- a) set of users $\mathcal{U}_w = \{1, 2, \ldots, \mathcal{W}\}$
- b) set of sub-channels $\mathcal{S} = \{1, 2, \ldots, \mathcal{S}\}$
- c) set the sub-channel allocation variable $\alpha^f_{w,s} = 0, \forall w, s$
- d) set the superimposed users $\Omega_s = \varnothing, \forall s$
- e) set $R^f_{w,s} = 0, \forall w, s$
- f) set $p^f_s = P^f_{\max}/S, \forall s \in \mathcal{S}$

2: Obtain $\gamma^f_{w,s}$, $\forall w, s$ and arrange it in ascending order $\gamma^f_{1,s} \geq \gamma^f_{2,s}, \ldots, \geq \gamma^f_{\mathcal{W},s}$

3: Obtain $p^f_{1,s}$ and $p^f_{w_i,s}$ using Equation (18) and (21) respectively $\forall w, s$

4: Obtain $R^f_{w,s} = B_{sch} \log_2\left(1 + \gamma^f_{w,s}\right)$, $\forall w, s$

5: **while** $\mathcal{U}_w \neq \varnothing$ **do**

6: **for** $w = 1$ to $\mathcal{W}$ **do**
- a) Find the user w_i^* with $R^f_{w_i^*,s} < R^{req}_{w_i^*,s}$
- b) For w_i^* find s^* satisfies $\mathcal{I}^{s^*}_{cr} \leq \mathcal{I}^{s^*}_{th}$ and $\gamma^f_{w_i^*,s^*} \geq \gamma^f_{w_i^*,s}$ $\forall s \in \mathcal{S}$
- c) Update $\Omega_s = \Omega_s \cup \{w_i^*\}$ and remove w_i^* from $\mathcal{U}_w = \mathcal{U}_w - \{w_i^*\}$ and $\alpha^f_{w_i^*,s} = 1$

if $(|\Omega_s|) = 2$ **then**
- a) Obtain $\mathcal{Q}^{EE}_{w,s} = \dfrac{\alpha^f_{w,s} R^f_{w,s}}{\frac{1}{\xi_f} p^f_{w,s} + \tilde{P}^f_{st}}$ for each multiplexed user
- b) Find w_j^* with $\mathcal{Q}^{EE}_{w_j^*} < \mathcal{Q}^{EE}_{w_j}$ and $\{\mathcal{Q}^{EE}_{w_j^*} + \mathcal{Q}^{EE}_{w_i^*}\} > \mathcal{Q}^{EE}_{w_j^*}$
- c) Update $\mathcal{S} = \mathcal{S} - \{s^*\}$, $\Omega_s = \Omega_s \cup \{w_i^* + w_j^*\}$ and remove w_j^* from $\mathcal{U}_w = \mathcal{U}_w - \{w_j^*\}$
- d) Assign $\alpha^f_{w_j^*,s} = 1$

7: **end if**

8: **end for**

9: Until $\mathcal{U}_w = \varnothing$

10: **end while**

B. ENERGY EFFICIENT POWER ALLOCATION SOLUTION BASED ON SEQUENTIAL CONVEX PROGRAMMING (SCP)

This section focuses on solving the power allocation optimization problem for the femtocell NOMA-HetNets system by mathematically formulating the optimization problem and then transforming the formulated problem into an equivalent problem. The equivalent problem is then iteratively solved to update the optimal energy efficiency parameter with the application of the SCP optimization approach to obtain the optimal power allocation solution.

1) PROBLEM FORMULATION

The max-min energy efficiency is formulated as the objective function to maximize the minimum energy efficiency of the fCUE by attaining the achievable maximum value to guarantee fairness for each fCUEs. In this problem formulation, consideration is given to the total transmitted power budget for each BS_f, the minimum required data rate for each fCUE, and the limitation of cross-tier interference caused by fCUEs to macrocell. This optimization problem can be expressed mathematically as objective function $P1$ with sets of formulated constraints as in (22).

$$
\begin{aligned}
P1:\ & \max_{\alpha,p} \min_{w\in\mathcal{W}} \mathcal{Q}^{EE,f}_{w,s}(\alpha,p) \\
\text{s.t. } C1:\ & \sum_{s\in\mathcal{S}} \alpha^f_{w,s} p^f_{w,s} \leq p^f_s, \quad \forall f, w \\
C2:\ & \sum_{s\in\mathcal{S}} p^f_s \leq P^f_{\max}, \forall f \\
C3:\ & \sum_{s\in\mathcal{S}} \alpha^f_{w,s} R^f_{w,s} \geq R^{req}_w, \quad \forall f, w \\
C4:\ & \sum_{f\in\mathcal{F}} \sum_{w\in\mathcal{W}} \alpha^f_{w,s} p^f_{w,s} \left|h^{m,f}_{w,s}\right|^2 \leq \mathcal{I}^s_{th}, \forall s \\
C5:\ & p^f_{w,s} \geq 0, \quad \forall f, w, s \\
C6:\ & \sum_{w\in\mathcal{W}} \alpha^f_{w,s} \leq 1, \quad \forall f, s \\
C7:\ & \alpha^f_{w,s} \in \{0, 1\}, \quad \forall f, w, s
\end{aligned} \tag{22}
$$

The constraints $C1$ and $C2$ limit the maximum transmitted power budget for each sub-channel and BS_f, respectively. $C3$ maintains the minimum QoS requirement for each $\mathrm{UE}^f_{w,s}$ which is limited by the data rate threshold R^{req}_w. The constraint $C4$ imposed to ensure that the maximum cross-tier interference caused by fCUEs to macrocell on each sub-channel does not exceed the threshold value denoted by $\mathcal{I}_{th}^{s}$. By satisfying the constraint $C4$, the mCUEs' QoS requirements are fulfilled significantly. The $C5$ ensures that the power assigned to $\mathrm{UE}_{w,s}^{f}$ is always positive, while the constraints $C6$ and $C7$ are associated with sub-channel assignment aspects that ensure at least one fCUE user is assigned on each sub-channel. It is challenging to solve the objective function $P1$ because it is a non-convex and non-deterministic polynomial (NP)-hard problem [19]. Thus, solving it in polynomial-time is impossible. Consider the following theorem:

Theorem 1: The objective function $P1$ is NP-hard problem, by cause of joint user pairing and power allocation problem. For the *Proof:* See **Appendix A**.

The objective function $P1$ is NP-hard, meaning that obtaining the optimal global solution is not guaranteed with polynomial-time algorithms [50]. Alternatively, efficient and approximate solutions with local optimum performance are most preferred in the practical wireless system design. Thus, we need to transform objective function $P1$ into subtractive form to solve it with lower complexity and easily determine the sub-optimal solution.

2) PROBLEM TRANSFORMATION AND ITERATIVE ALGORITHM DESIGN

Since the objective function $P1$ is a mixed-integer nonlinear programming problem due to $\alpha_{w,s}^{f}$ and $p_{w,s}^{f}$, it becomes hard to solve such nonlinear problem. After relaxing $\alpha_{w,s}^{f}$ to [0, 1], the function $P1$ is still a non-convex; hence, we assume that $R_{w,s}^{f}(\alpha, p) > 0$ and $P_{T(w,s)}^{f}(\alpha, p) > 0$. For simplicity, we denote $\mathcal{K} \in (\alpha, p)$ as a feasible solution set for formulated objective function $P1$. Therefore, the optimal energy efficiency is denoted as $\mathcal{Q}_{w}^{*}$, which is given as in (23).

$$\mathcal{Q}_{w}^{*} = \max_{(\alpha,p)\in\mathcal{K}} \min_{w} \frac{R_{w,s}^{f}(\alpha, p)}{P_{T(w,s)}^{f}(\alpha, p)} = \min_{w} \frac{R_{w,s}^{f}(\alpha^{*}, p^{*})}{P_{T(w,s)}^{f}(\alpha^{*}, p^{*})} \quad (23)$$

where (α^{*}, p^{*}) is the optimal solution. To solve the formulated problem function $P1$, the generalized fractional programming theory in [51], [52] is considered with the following theorem:

Theorem 2: The optimal solution $(\alpha^{*}, p^{*}) \in \mathcal{K}$ for the objective function $P1$ is obtained if and only if $P1 \equiv P2$ in (24).

$$\begin{aligned} P2: &\max_{(\alpha,p)\in\mathcal{K}} \min_{w} \left\{R_{w,s}^{f}(\alpha, p) - \mathcal{Q}_{w}^{*} P_{T(w,s)}^{f}(\alpha, p)\right\} \\ &= \min_{w} \left\{R_{w,s}^{f}(\alpha^{*}, p^{*}) - \mathcal{Q}_{w}^{*} P_{T(w,s)}^{f}(\alpha^{*}, p^{*})\right\} = 0 \end{aligned} \quad (24)$$

The optimum conditions stated in theorem 2 indicate that solving an optimization problem function $P1$, which is in fractional form is equivalent to solving problem function $P2$ in subtractive form. This entails that the optimal solution of the problem function $P1$ is the same as the optimal solution of its equivalent problem function $P2$ based on literature [53], [54]. For the *Proof:* See **Appendix B**.

The optimal solution to problem function $P1$ can be discovered by solving optimization function $P2$ to obtain $\mathcal{Q}_{w}^{*}$. However, $\mathcal{Q}_{w}^{*}$ cannot be obtained directly because it is challenging to solve objective function $P2$. Therefore, the proposed iterative algorithm-2 update $\mathcal{Q}_{w}$ and ensure that the optimal solution p^{*} remains feasible in each iteration. To design this algorithm, we define an equivalent function as shown in (25), followed by the proposed theorem 3.

$$F(\mathcal{Q}_{w}) = \max_{(\alpha,p)\in\mathcal{K}} \min_{w} \left\{R_{w,s}^{f}(\alpha, p) - \mathcal{Q}_{w} P_{T(w,s)}^{f}(\alpha, p)\right\} \quad (25)$$

Theorem 3: $F(\mathcal{Q}_{w})$ is a strictly monotonically decreasing function in $\mathcal{Q}_{w}$. For the *proof:* See **Appendix C**.

$$F(\mathcal{Q}_{w}) \geq 0 \text{ when } \mathcal{Q}_{w} = 0, \text{ and } F(\mathcal{Q}_{w}) < 0 \text{ when } \mathcal{Q}_{w} > 0.$$

Consequently, $\mathcal{Q}_{w}^{*}$ is achieved when $F(\mathcal{Q}_{w}) = 0$. Therefore, the bisection method-based iterative algorithm in [55] is applied to obtain the optimal solution $\mathcal{Q}_{w}^{*}$, which lies between two opposite signs intervals $\left[\mathcal{Q}_{w}^{\min}, \mathcal{Q}_{w}^{\max}\right]$. The detailed procedure of obtaining $\mathcal{Q}_{w}^{*}$ is given in algorithm-2 as shown in Table 2.

TABLE 2. Algorithm-2: Bisection Method-Based Iterative Algorithm for Obtaining $\mathcal{Q}_{w}^{*}$.

1: **Initialize**
  a) set the iteration index $k = 0$ and error tolerance $\varepsilon > 0$.
  b) set the maximum iteration $I_{\max}$.
  c) set $\mathcal{Q}_{w}^{\min}$ and $\mathcal{Q}_{w}^{\max}$, such that $\mathcal{Q}_{w}^{\min} \leq \mathcal{Q}_{w}^{*} \leq \mathcal{Q}_{w}^{\max}$
2: **repeat**
3: $\mathcal{Q}_{w}^{k} = \left(\mathcal{Q}_{w}^{\min} + \mathcal{Q}_{w}^{\max}\right)/2$
4: Solve (26) for a given $\mathcal{Q}_{w}^{k}$ and obtain fair transmit power p^{k}.
5: **if** $\left|F\left(\mathcal{Q}_{w}^{k}\right)\right| = \left|\min_{w}\left[R_{w,s}^{f}\left(p^{k}\right) - \mathcal{Q}_{w}^{k} P_{T(w,s)}^{f}\left(p^{k}\right)\right]\right| \leq \varepsilon$ **then**
6: $p^{*} = p^{k}$ and $\mathcal{Q}_{w}^{*} = \min_{w}\left[\frac{R_{w,s}^{f}\left(p^{k}\right)}{P_{T(w,s)}^{f}\left(p^{k}\right)}\right]$
7: **break**
8: **else**
9: **if** $\left|F\left(\mathcal{Q}_{w}^{k}\right)\right| < 0$ **then**
10: $\mathcal{Q}_{w}^{\max} = \mathcal{Q}_{w}^{k}$
11: **else**
12: $\mathcal{Q}_{w}^{\min} = \mathcal{Q}_{w}^{k}$
13: **end if**
14: **end if**
15: $k = k + 1$
16: **until** $k > I_{\max}$

The following optimization problem function $P3$ with stated constraints shown in (26) is solved at step 4 of the algorithm-2 with the considerable value of $\mathcal{Q}_{w}^{k}$.

$$\begin{aligned} P3: &\max_{(\alpha,p)\in\mathcal{K}} \min_{w} \left\{R_{w,s}^{f}(\alpha, p) - \mathcal{Q}_{w}^{k} P_{T(w,s)}^{f}(\alpha, p)\right\} \\ &\text{s.t. } C1\text{–}C7 \end{aligned} \quad (26)$$

3) SUB-CHANNEL POWER ALLOCATION USING SEQUENTIAL CONVEX PROGRAMMING (SCP)

To solve the power allocation problem across sub-channels, the fair energy efficiency is formulated as an objective function $P3.1$ subjected to the stated constraints in (27).

$$\begin{aligned} P3.1:\ &\max_{p_s^f}\min_{w}\left\{R_{w,s}^f(p)-\mathcal{Q}_w^k\left(P_{T(w,s)}^f\right)^T(p)\right\}\\ \text{s.t. } &C2:\ \sum_{s\in\mathcal{S}}p_s^f\le P_{\max}^f,\ \forall f\\ &C3:\sum_{s\in\mathcal{S}}R_{w,s}^f\ge R_w^{req},\quad\forall f,w\\ &C4:\sum_{f\in\mathcal{F}}\sum_{w\in\mathcal{W}}p_{w,s}^f\left|h_{w,s}^{m,f}\right|^2\le\mathcal{I}_{th}^s,\quad\forall s\\ &C8:p_s^f\ge0,\quad\forall f,s \end{aligned}\tag{27}$$

The formulation of the objective function problem $P3.1$ subjected to stated constraints in (27) is a non-convex problem with respect to p_s^f which makes it difficult to solve in polynomial-time. Thus, the SCP approach in [55], [56] is utilized as a local optimization method to hold the convex part of the optimization problem correctly and to formulate the non-convex part to an accurate convex. To solve this problem, the value of power allocation is updated until the optimal value is found. This proposed power allocation solution is called the NOMA power allocation with SCP (NOMA-SCP). The problem function in $P3.1$ can then be formulated as two difference concave functions with respect to p and expressed as in (28).

$$R_{w,s}^f(p)-\mathcal{Q}_w^kP_{T(w,s)}^f(p)=m_w(p)-v_w(p)\tag{28}$$

where,

$$\begin{aligned} m_w(p)&\\ =\sum_{s\in\mathcal{S}}B_{sch}\log_2&\left(1+\frac{p_{w,s}^f\left|h_{w,s}^f\right|^2}{\sum_{i=w+1}^{\mathcal{W}}\left|h_{w,s}^f\right|^2p_{i,s}^f+\sum_{q=1}^{\mathcal{F}}\left|h_{w,s}^{f,q}\right|^2p_{w,s}^q+\sum_{r=1}^{\mathcal{R}}\left|h_{r,s}^{f,m}\right|^2p_{r,s}^m+\left(\sigma_{w,s}^f\right)^2}\right)\\ -\mathcal{Q}_w^kP_{T(w,s)}^f& \end{aligned}\tag{29}$$

and,

$$v_w(p)=\sum_{s\in\mathcal{S}}\log_2\left(\sum_{i=w+1}^{\mathcal{W}}\left|h_{w,s}^f\right|^2p_{i,s}^f+\sum_{q=1,q\ne f}^{\mathcal{F}}\left|h_{w,s}^{f,q}\right|^2p_{w,s}^q+\sum_{r=1}^{\mathcal{R}}\left|h_{r,s}^{f,m}\right|^2p_{r,s}^m+\left(\sigma_{w,s}^f\right)^2\right)\tag{30}$$

By adopting the SCP procedure, the non-convex constraint $C3$ must be formulated into a convex as expressed in (31) by denoting this new constraint as $C3'$.

$$\begin{aligned} C3':\ &p_{w,s}^f\left|h_{w,s}^f\right|^2+\left(1-2^{R_w^{req}/B_{sch}}\right)\\ &\left(\mathcal{I}_{w,s}^f+\mathcal{I}_{co}^{f,q}+\mathcal{I}_{cr}^{f,m}+\left(\sigma_{w,s}^f\right)^2\right)\ge0,\quad\forall f,w \end{aligned}\tag{31}$$

Equation (31) is applied to reformulate the objective function problem $P3.1$ into an equivalent form in $P3.1.1$ with stated constraints in (32).

$$\begin{aligned} P3.1.1:\ &\max_{p_s^f}\min_{w}\{m_w(p)-v_w(p)\}\\ &\text{s.t. } C2,\ C3',\ C4,\ C5 \end{aligned}\tag{32}$$

The reformulated problem function $P3.1.1$ is a non-smooth optimization problem, with a new variable H introduced to smooth the optimization function. Hence, $P3.1.1$ is further re-written with an equivalent function form expressed in $P3.1.2$ with stated constraints in (33).

$$\begin{aligned} P3.1.2:\ &\max_{p_s^f,\mathcal{H}}\mathcal{H}\\ &\text{s.t. } C2,\ C3',\ C4,\ C5\\ &\quad C8:\{m_w(p)-v_w(p)\}\ge\mathcal{H},\quad\forall f,w \end{aligned}\tag{33}$$

Since the constraint *C8* is a combination of two different concave functions, then the SCP is effective in solving problem function $P3.1.2$ iteratively with stated constraints in (33). An iterative fair transmitted power allocation p^t is generated at step t and then used to approximate $v_w(p)$ using the first-order Taylor expansion as expressed in (34).

$$v_w(p)\approx v_w\left(p^t\right)+\nabla v_w^T\left(p^t\right)\left(p-p^t\right)\tag{34}$$

where $\nabla v_w(p)$ is the gradient of $v_w(p)$ at p and is expressed as in (35). The c_w is a W-dimensional column vector with $c_w(f)=0$ and $c_w(q)=\left|h_{w,s}^{f,q}\right|^2\Big/\ln2,\ q\ne f$.

$$\nabla v_w(p)=\frac{c_w}{\left(\sum_{i=w+1}^{\mathcal{W}}\left|h_{w,s}^f\right|^2p_{i,s}^f+\sum_{q=1,q\ne f}^{\mathcal{F}}\left|h_{w,s}^{f,q}\right|^2p_{w,s}^q+\sum_{r=1}^{\mathcal{R}}\left|h_{r,s}^{f,m}\right|^2p_{r,s}^m+\left(\sigma_{w,s}^f\right)^2\right)}\tag{35}$$

By substituting (34) into $P3.1.2$, the following convex optimization problem function $P3.1.3$ is generated with a stated sets of constraints in (36).

$$\begin{aligned} P3.1.3:\ &\max_{p_s^f,\mathcal{H}}\mathcal{H}\\ &\text{s.t. } C2,\ C3',\ C4,\ C5\\ &\quad C8:\ \left\{m_w(p)-v_w\left(p^t\right)+\nabla v_w^T\left(p^t\right)\left(p-p^t\right)\right\}\\ &\qquad\ge\mathcal{H},\quad\forall f,w \end{aligned}\tag{36}$$

The current formulated objective function $P3.1.3$ with new constraints in (36) is smoothened and standardized as the convex approximation of the formulated problem function $P3.1$, which can be effectively solved to obtain the objective function sub-optimal power allocation solution. The detailed procedure of finding the iterative power allocation is given in algorithm-3 as shown in Table 3 with consideration to theorem 4.

TABLE 3. Algorithm-3: Iterative Sub-channel Power Allocation for Obtaining p_s^*.

1: Initialize $t = 0$ and $\varepsilon > 0$
2: Set $p^{(0)}$ calculate $J^{(0)} = \min_w \{m_w(p^{(0)}) - v_w(p^{(0)})\}$
3: **while** $\lvert J^{(t+1)} - J^{(t)} \rvert > \varepsilon$ **do**
4: Solve Equation (36) to get the $p^{(*)}$ solution
5: Set $t = t+1$, $p^{(t)} = p^{(*)}$
6: $J^{(t)} = \min_w \{m_w(p^{(t)}) - v_w(p^{(t)})\}$
7: **end while**

Theorem 4: The iterative power allocation of algorithm-3 always converges, and with any feasible initial points, it reaches the optimal power allocation by converging to a stationary point. For the proof, refer to [22].

4) POWER ALLOCATION FOR THE PAIRED USERS USING FAIR ENERGY-EFFICIENT POWER ALLOCATION (FEPA) ALGORITHM

After solving the power allocation problem for the sub-channel assignment and obtaining the p_s^f, the power allocation for each user $\left(p_w^f\right)$ in each sub-channel is further processed. This is achieved from the formulated optimization problem function $P3.2$ with stated constraints in (37).

$$\begin{aligned} P3.2: \quad & \max_{p_{w,s}^f} \min_{w} \left\{ R_{w,s}^f(p) - \mathcal{Q}_w^k \left(P_{T(w,s)}^f \right)^T (p) \right\} \\ & s.t.\ C1: \sum_{w\in\mathcal{W}} p_{w,s}^f \le p_s^f, \quad \forall f, w \\ & C3: \sum_{s\in\mathcal{S}} R_{w,s}^f \ge R_w^{req}, \quad \forall f, w \\ & C4: \sum_{f\in\mathcal{F}} \sum_{w\in\mathcal{W}} p_{w,s}^f \left| h_{w,s}^{m,f} \right|^2 \le \mathcal{I}_{th}^s, \quad \forall s \\ & C5: p_{w,s}^f \ge 0, \quad \forall f, w, s \end{aligned} \tag{37}$$

Based on the optimization problem formulated function $P3.2$, the power is assigned to the paired users in each sub-channel according to their channel conditions. Meaning that the weak user is given more power than the strong user. According to the NOMA principle, the expression in (18) is used to calculate the power assigned to the weak user, and that in (21) is used to calculate the power assigned to the strong user as demonstrated in [15]. The power allocated to the weak user efficiently lies in the interval $\left[p_{1,s}^{\min}, p_{1,s}^{\max}\right]$, where,

$$p_{1,s}^{\min} = 0 \tag{38}$$

$$p_{1,s}^{\max} = p_s^f \Big/ \left(1 + \delta_{1,s}^f\right) \tag{39}$$

To simplify the power allocation process, the bisection method is applied to obtain the sub-optimal power allocation solutions for the paired users in their assigned sub-channel.

TABLE 4. Algorithm-4: Fair energy-efficient power allocation (FEPA) for the paired users.

1: Initialize $p_{1,s}^{\min} = 0$, $p_{1,s}^{\max} = p_s^f \big/ \left(1 + \delta_{1,s}^f\right)$ and $\varepsilon > 0$
2: **repeat**
3: Set $p_{1,s}^f = \left(p_{1,s}^{\min} + p_{1,s}^{\max}\right) \big/ 2$
4: Set $p_{2,s}^f = p_s^f - p_{1,s}^f$
4: Solve Equation (7) to get the $R_{w,s}^f$
5: **if** $\sum_{s\in\mathcal{S}} R_{w,s}^f \le R_w^{req}$ **then**
6: $p_{1,s}^{\max} = p_{1,s}^f$
7: **else**
8: $p_{1,s}^{\min} = p_{1,s}^f$
9: **end if**
10: **until** $\left(p_{1,s}^{\max} - p_{1,s}^{\min}\right) \le \varepsilon$
11: Output $p_{1,s}^* = p_{1,s}^f$ and $p_{2,s}^* = p_s^f - p_{1,s}^*$

The detailed procedure is given in algorithm-4 as shown in Table 4.

IV. RESULTS

The study considered the system model of femtocell NOMA-HetNets to evaluate the performance of the proposed solutions. The BS_m is located at the centre of the circular coverage area, with a radius of 500 m, and the maximum transmitted power of 46 dBm. The several BS_f are distributed uniformly inside the macro cell coverage area within a radius of 20 m and with the maximum transmitted power of 20 dBm. The total static power consumption in the network is 30 dBm, while the mCUEs and fCUEs are distributed randomly within the coverage area of their corresponding BSs. The path loss is modelled as in [45].

$$\text{Femtocell: } 127 + 30\log_{10}(d^f)\ \text{dB} \tag{40}$$

$$\text{Macrocell: } 128.1 + 37.6\log_{10}(d^m)\ \text{dB} \tag{41}$$

where d^m and d^f are the distances in km between BS_m to $UE_{r,s}^m$ and BS_f to $UE_{w,s}^f$, respectively. The shadowing standard deviation is set to 10 dB, the system bandwidth B_w is 10 MHz, and the noise of the power spectral density is -174 dBm/Hz. The summary for the default parameters is shown in Table 5.

A. COMPLEXITY ANALYSIS

The computational complexity analysis of the proposed user-pairing and power allocation methods are discussed with the assumption that $\mathcal{S}$ is the number of sub-channels and $\mathcal{L}$ represents the available users in the network (i.e., $\mathcal{L} = 2\mathcal{S}$). Applying the exhaustive search can help to achieve higher system performance, but needs the consideration of all possible user's combination that leads to the $\mathcal{O}\left(2\mathcal{S}!/2^{\mathcal{S}}\right)$ higher computational complexity, as seen in [30].

The computational complexity of the proposed solution based on the GA (algorithm-1) depends mainly on the

TABLE 5. **Parameters specifications.**

Parameter	Default Value
Macrocell radius	500 m
Femtocell radius	20 m
Number of BS_f	2
Number of fCUEs in each femtocell	2
Number of mCUEs in macrocell	2
Maximum transmit power for BS_m ($P^m_{\max}$)	46 dBm
Maximum transmit power for BS_f ($P^f_{\max}$)	20 dBm
Interference threshold for each mCUEs (I^s_{th})	– 101.2 dBm
Minimum rate requirement for each fCUE (R^{req}_w)	10 Kbps
Path-loss between BS_f to fCUEs	$127+30\log_{10}(d^f)$ dB
Path-loss between BS_m to mCUEs	$128.1+37.6\log_{10}(d^m)$ dB
Shadowing standard deviation	10 dB
User noise figure	9 dB
Antenna gain for macro cell/femto cell	14 dBi/5 dBi
Amplifier efficiency (ξ_f) and (ξ_m)	1
Static power consumption ($\tilde{P}^f_{st}$) and ($\tilde{P}^m_{st}$)	30 dBm
System bandwidth (B_w)	10 MHz
Error tolerance (ε)	10^{-3}
Noise power spectral density (N_0)	– 174 dBm/Hz
Decay factor (α) for FTPA	0.4

initialization process, which includes sorting of users based on their SINR and assignment process, which is based on assigning users to the sub-channels. The initialization process requires $\mathcal{L}(\mathcal{L}-1)/2$ operations, and assignment processes require $2\mathcal{L}\ln(\mathcal{L})$ operations, as used in [22]. Therefore, the total complexity of the GA can simply be written as in (43).

$$\begin{aligned}\text{GA complexity} &= \text{initialization process complexity} \\ &\quad + \text{assignment process complexity} \quad (42)\\ \text{GA complexity} &= \mathcal{O}\left(\mathcal{L}(\mathcal{L}-1)/2\right) + \mathcal{O}\left(2\mathcal{L}\ln(\mathcal{L})\right)\\ &= \mathcal{O}\left(\mathcal{L}^2\right) \quad (43)\end{aligned}$$

Therefore, the computational complexity of the GA (algorithm-1) is $\mathcal{O}\left(\mathcal{L}^2\right)$.

The power allocation problem is solved based on the SCP approach, using algorithm-3, since the solution is obtained through the transformation process; it first updates $\mathcal{Q}_w$ using algorithm-2 (bisection method) that requires I_1 iterations. Hence, based on the formulated optimization problem $P3.1$, the convergence of the sub-optimal power allocation solution is achieved with I_2 iterations. Thus, the computational complexity of algorithm-3 using the SCP approach is $\mathcal{O}\left(I_1 I_2 \mathcal{L}\mathcal{S}\right)$. Hence, the total complexity of the proposed resource allocation solutions (NOMA-SCP-GA) is given as in (44).

$$\begin{aligned}\text{NOMA-SCP-GA complexity} &= \mathcal{O}\left(\mathcal{L}^2\right) + \mathcal{O}\left(I_1 I_2 \mathcal{L}\mathcal{S}\right)\\ &= \mathcal{O}\left(\mathcal{L}^2 + I_1 I_2 \mathcal{L}\mathcal{S}\right) \quad (44)\end{aligned}$$

From the result observation, the computational complexity of the proposed solution is lower than the exhaustive search approach, and can be implemented with a polynomial-time.

TABLE 6. **The complexity comparison for the proposed and existing method.**

Method	Complexity
Proposed method (NOMA-SCP-GA)	$\mathcal{O}\left(\mathcal{L}^2 + I_1 I_2 \mathcal{L}\mathcal{S}\right)$
Exhaustive search method [30]	$\mathcal{O}\left(2\mathcal{S}!/2^{\mathcal{S}}\right)$

The summary of the computational complexity comparison of the proposed resource allocation solutions and the exhaustive search method is shown in Table 6.

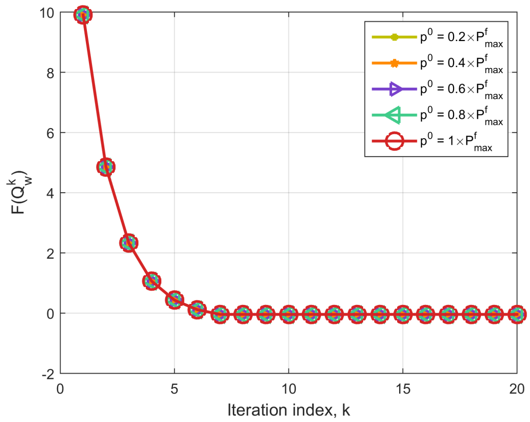

FIGURE 3. **The convergence of the proposed iterative algorithm-2 with $R^{req}_w = 10$Kbps, $\xi_f = 1$ and $\tilde{P}^f_{st} = 30$dBm.**

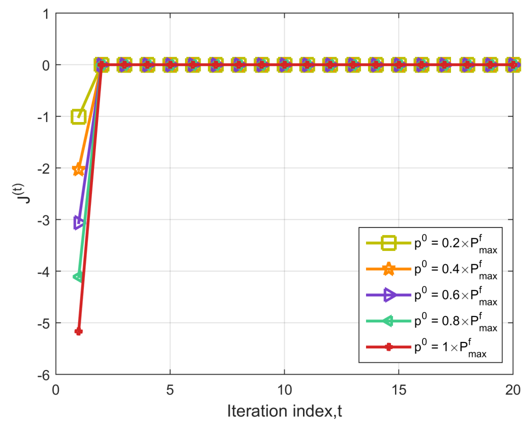

FIGURE 4. **The convergence of the proposed iterative algorithm-3 with $R^{req}_w = 10$Kbps, $\xi_f = 1$, $\tilde{P}^f_{st} = 30$dBm and maximum value of $Q^k_w = 5$bits/Joule.**

B. PERFORMANCE EVALUATION AND DISCUSSION

The performance of the proposed resource allocation solutions is evaluated and compared with the FTPA method [9], DC algorithm [30], and OFDMA technique for sufficient validation. Figure 3 and Figure 4 display the convergence of the proposed iterative algorithm-2 and algorithm-3, respectively in achieving the optimal value of estimated $\mathcal{Q}^*_w$ and p^*_s.

It is observed that after a few iterations, both algorithms converge to reach their respective optimal solutions when the initial transmitted power used are within the acceptable range values (i.e.,$0.2 \times P^f_{max} - 1 \times P^f_{max}$). This indicates that the initial transmitted power values do not have any effect on the obtained optimal solution. These results conclude that both deployed algorithms are effective and useful for practical situations.

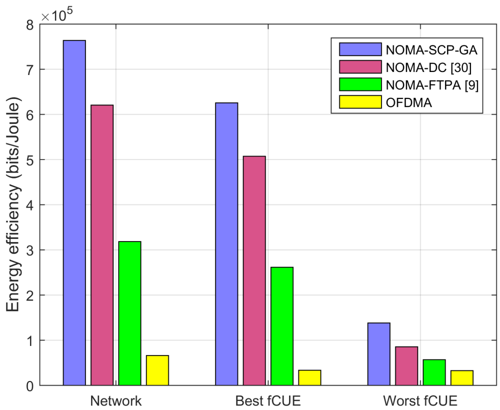

FIGURE 5. **Comparison of the energy efficiency of the network, best fCUE and worst fCUE with $P^f_{max} = 20dBm$, $\tilde{P}^f_{st} = 30dBm$, $\xi_f = 1$ and $R^{req}_w = 10$ Kbps for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

Figure 5 compares the proposed solution with the benchmarked methods in terms of energy efficiency of the network, best fCUE, and worst fCUE. As observed, the difference between the best fCUE and the worst fCUE is considerable in NOMA-DC and NOMA-FTPA, but with a slight deviation in the OFDMA method, because the OFDMA technique assigns equal powers to all users which subsequently leads to nearly equal energy efficiency among the served users. The network energy efficiency is high across all methods, but the proposed NOMA-SCP-GA achieving the highest performance value. The energy efficiency of the worst fCUE using NOMA-SCP-GA is higher by 38.22%, 58.84%, and 76.39% compared to NOMA-DC, NOMA-FTPA, and OFDMA methods, respectively.

Figure 6 presents the system energy efficiency of the different schemes with respect to the number of users, where the number of users was varied from 1 to 6. The energy efficiency for all schemes increased significantly when the number of users incremented from 1 to 2 and achieved saturated-like efficiency with the additional number of users. The proposed NOMA-SCP-GA shows the highest system energy efficiency compared to NOMA-DC, NOMA-FTPA, and OFDMA. The schemes with NOMA technology achieved better energy efficiency performance compared to OFDMA because more than one user is assigned in a single sub-channel in NOMA scheme, while only a single user is assigned in a single sub-channel in OFDMA scheme. When the number of users (UEs) is 6, the proposed

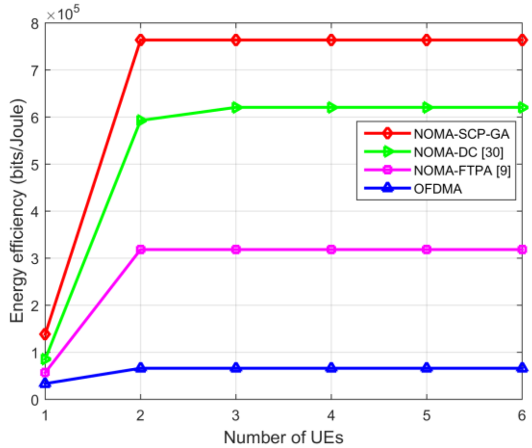

FIGURE 6. **The energy efficiency vs. number of UEs with $P^f_{max} = 20$ dBm, $\tilde{P}^f_{st} = 30$ dBm, $\xi_f = 1$ and $R^{req}_w = 10$ Kbps for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

NOMA-SCP-GA improves the system's energy efficiency by 18.75%, 58.31%, and 91.34% compared to NOMA-DC, NOMA-FTPA, and OFDMA, respectively.

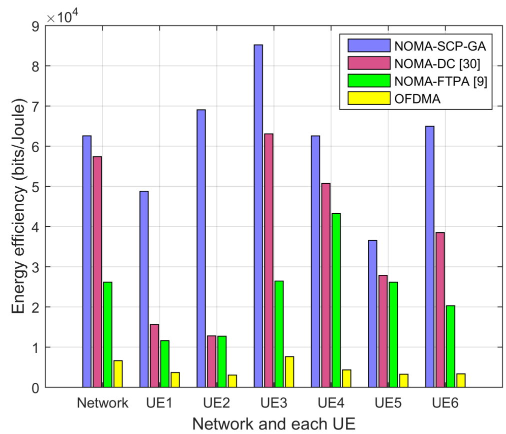

FIGURE 7. **The energy efficiency vs. network and each UE in the system with $P^f_{max} = 20$ dBm, $\tilde{P}^f_{st} = 30dBm$ and $\xi_f = 1$ for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

The overall energy efficiency performance of the whole system network and of each user (UE1 – UE6) is presented in Figure 7. The results show that the NOMA-SCP-GA and NOMA-DC schemes improve the network energy efficiency greatly compared to NOMA-FTPA and OFDMA schemes, because in NOMA schemes the users with worst channel conditions are prioritized during the resource allocation. In contrast, the resource allocation in NOMA-FTPA is unstable with consideration to the channel condition of all users, while the resource allocation in OFDMA is inefficient as it does not consider the channel condition of all users. Given these two polarizing approaches, the proposed NOMA-SCP-GA takes attention in considering the channel condition of users and is able to achieve better fairness among the served users

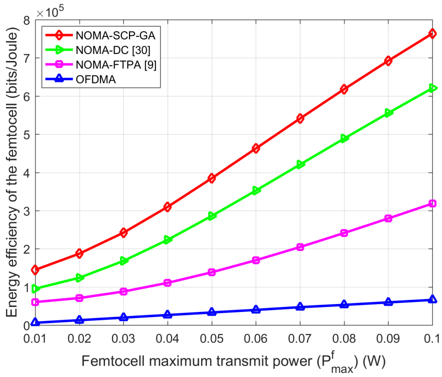

FIGURE 8. **The energy efficiency of the femtocell vs. femtocell maximum transmit power $\left(P^f_{max}\right)$ with $\tilde{P}^f_{st} = 30\ dBm$ and $\xi_f = 1$ for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

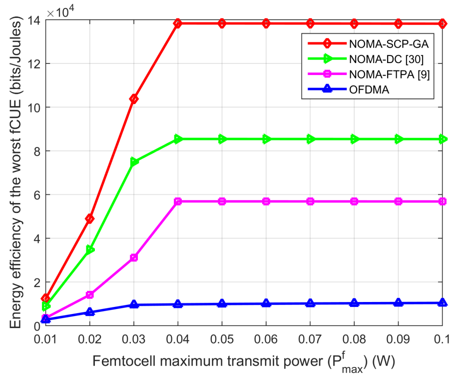

FIGURE 9. **The energy efficiency of the worst fCUE vs. femtocell maximum transmit power $\left(P^f_{max}\right)$ with $\tilde{P}^f_{st} = 30\ dBm$ and $\xi_f = 1$ for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

with the highest network energy efficiency compared to other benchmarked schemes.

Figure 8 demonstrates the effect of the femtocell BS's maximum transmission power $\left(P^f_{max}\right)$ on energy efficiency. The P^f_{max} is varied from 0.01 – 0.1 W with 4 users in the femtocell. It is observed that the energy efficiency of the femtocell increases progressively with the increase of P^f_{max} across all deployed methods. The highest P^f_{max} means additional value of optimal power is allocated to the BS_f through the optimization solution formulation in problem function $P3.1$. The received SINR for each $UE^f_{w,s}$ in each femtocell is improved and leads to higher improvement of the energy efficiency. The proposed NOMA-SCP-GA significantly outperformed all other validating benchmark methods with a remarkable high value of P^f_{max}. The energy efficiency gap between the proposed NOMA-SCP-GA and all other validating methods is notably high due to the efficient assignment of the additional optimal transmitted power compared to NOMA-DC, NOMA-FTPA, and OFDMA. When P^f_{max} is 0.05 W, the proposed NOMA-SCP-GA achieves a higher energy efficiency by 25.65%, 64.02%, and 91.37% than NOMA-DC, NOMA-FTPA, and OFDMA, respectively. These results conclude that P^f_{max} affects the femtocell energy efficiency performance.

In Figure 9, the influence of the femtocell BS's maximum transmission power (P^f_{max}) on the worst fCUE's energy efficiency is presented. The proposed NOMA-SCP-GA achieves the highest energy efficiency at the initial value of maximum transmission power and increases at a high rate up to 0.04 W, and remains unchanged beyond 0.04 W across all deployed methods. Meaning that, when P^f_{max} is low, the utilized constraints have a strong impact on the worst fCUE's assigned power, which leads to a high energy efficiency increment. However, when the P^f_{max} reaches a certain threshold value, the constraints no longer affect the assigned worst fCUE's power, and the additional power do not affect the energy efficiency. Therefore, it can be concluded that

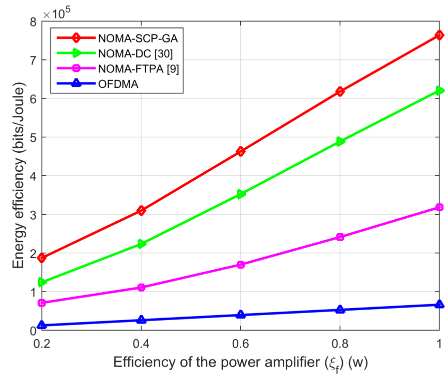

FIGURE 10. **The energy efficiency vs. efficiency of the power amplifier (ξ_f) at the BS_f with $\tilde{P}^f_{st} = 30\ dBm$ for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

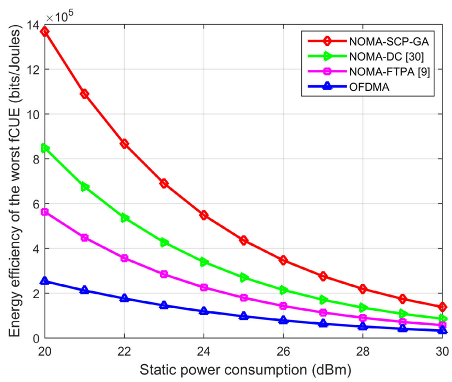

FIGURE 11. **The energy efficiency of the worst fCUE vs. static power consumption ($\tilde{P}^f_{st}$) with $P^f_{max} = 20dBm$ and $\xi_f = 1$ for all four schemes (NOMA-SCP-GA, NOMA-DC, NOMA-FTPA and OFDMA).**

the femtocell needs to operate below a certain level of the maximum transmitted power to improve the user's energy efficiency.

Figure 10 shows the impact of power amplifier efficiency on the network energy efficiency, where the P^f_{max} is varied from 0.02 W to 0.1 W, while ξ_f is varied from 0.2 W to 1 W, and the number of femtocell users $\mathcal{U}_w$ is set to 4. The network energy efficiency increases upon increasing the ξ_f across all deployed methods. The higher the value of ξ_f, the lower the obtained dynamic power consumption, as described in the mathematical formulation of (10). This results to lower signal dissipation and increases the obtained network energy efficiency of the proposed NOMA-SCP-GA that outperformed the NOMA-DC, NOMA-FTPA and OFDMA.

In Figure 11, the effect of static power consumption $\left(\tilde{P}_{st}^{f}\right)$ on the energy efficiency of the worst fCUE is demonstrated where the $\tilde{P}_{st}^{f}$ is varied from 20 – 30 dBm. The energy efficiency of the worst fCUE decreases with the increase in the $\tilde{P}_{st}^{f}$ which fulfils the mathematical formulation expressed in (15). Considering the observed trend in Figure 11, $\tilde{P}_{st}^{f}$ should be limited to a certain level to prevent the further decrease in energy efficiency. Furthermore, the energy efficiency of the worst fCUE using the proposed NOMA-SCP-GA is the highest compared to other benchmarked methods.

V. CONCLUSION

In this article, fair energy-efficient resource allocation solution with the proposed NOMA-SCP-GA for downlink femtocell NOMA-HetNets has been successfully presented. The study has achieved the objective of maximizing the energy efficiency for the user with the minimum energy efficiency performance. The considered maximization problem has been decoupled into user-pairing and power allocation problems to reduce the complexity and obtains the sub-optimal solutions. The GA was deployed to pair two users in each sub-channel and the power allocation problem was formulated as a fair energy efficiency maximization problem which is a mixed-integer non-convex fractional programming problem, with the constraints of BS transmission power, minimum user rate, and interference. Also, consideration focused on energy consumption from both transmitters and receivers to simulate the real system design. The problem was transformed into its subtractive form and then solved using the SCP approach to obtain the sub-channel power allocation solutions. The fair energy efficiency power allocation solution was proposed to assign power to the paired users. The results demonstrated that the proposed NOMA-SCP-GA has low complexity, and fast convergence with more fairness among users. Moreover, the proposed scheme achieved high energy efficiency performance compared to other deployed benchmarked methods. The proposed resource allocation solution is a promising solution for the 5G network. The study of fair energy-efficient resource allocation for NOMA-HetNets with multiple-input multiple-output (MIMO) system is considered as future research.

APPENDIX A
PROOF OF THEOREM 1

In the theory of complexity, to prove an optimization problem is NP-hard, we need to create the NP-hardness of its equivalent decision problem and show that the decision version is NP-hard. To do this, we need to consider the following three steps [50]. (i) select a relevant and known NP-complete decision problem C; (ii) establish a polynomial-time transformation from instance C to other instance D which is appropriate to the considered problem; (iii) after the transformation, then prove that the two instances (C and D) have the same objective solution. Based on these considerations, we now prove that the objective problem $P1$, as seen in (22) is NP-hard.

Proof: Two cases are considered for the proof, when (a) $\alpha_{w,s}^{f} = 1$ and (b) $\alpha_{w,s}^{f} > 1$.

(a) For $\alpha_{w,s}^{f} = 1$, the objective problem $P1$ is NP-hard as equivalent to the maximization problem with joint sub-channel and power allocation problem that has been proved as NP-hard for the OFDMA system [57].

(b) For $\alpha_{w,s}^{f} > 1$, we establish an instance of the objective problem $P1$ with known power allocation, followed by proving that $P1$ is NP-hard under this known power allocation solution. First, we consider an instance of the problem $P1$ corresponds to the multiple-choice knapsack problem (MKP), commonly known as the NP-hard problem. Then the instance of $\alpha_{w,s}^{f} = 2$ is considered, followed by proving that the joint sub-channel and power allocation problem in simplified form can be reduced to the knapsack problem. To understand the MKP, the following concept is considered as in [50].

Assuming that there are $B_1, B_2, \ldots, B_{\mathrm{N}}$ classes, such that each class i contains b_i items that need to be loaded in a knapsack, which corresponds to the total weight of P, and every item $j \in B_i$ acquire a weight $P_{i,j}$, and profit of $K_{i,j}$. Assuming that the item's weights, profits, and knapsack weights have non-negative values. The problem is to allocate the items to every class and ensure that the maximum profit is achieved without exceeding the knapsack's total weight.

Based on the MKP concept, now the problem $P1$ can be reduced to the MKP problem by assuming that each sub-channel is similar to the knapsack, and every user is similar to an item that is going to be loaded in the knapsack. The total number of items for every knapsack is limited to two (i.e., two users). The profit of every item loaded in the knapsack corresponds to the utility function ($K_{i,j}$), with the needed weight of $p_{i,j}$. The problem $P1$ intends to choose precisely two users (which corresponds to two items) along every sub-channel (which corresponds to class) in order to maximize the energy efficiency of the fCUEs with the stated power constraint p_s^f. The problem $P1$ can now be written in the form of $P1'$, as seen in (45).

$$\begin{aligned} P1': \ & \max_{\alpha,p} \min_{w\in\mathcal{W}} \mathcal{Q}_{w,s}^{EE,f}(\alpha,p) \\ & \text{s.t. } C1: \sum_{s\in\mathcal{S}} \alpha_{w,s}^{f} p_{w,s}^{f} \le p_s^f, \quad \forall f, w \\ & C6: \sum_{w\in\mathcal{W}} \alpha_{w,s}^{f} \le 1, \quad \forall f, s \\ & C7: \alpha_{w,s}^{f} \in \{0,1\}, \quad \forall f, w, s \end{aligned} \tag{45}$$

Therefore, the problem $P1'$ is NP-hard as being classified as the MKP, by the ordinary generalization of the knapsack problem. Observing that the problem $P1'$ is a specialized case of problem $P1$; therefore, the original formulation of the problem $P1$ is NP-hard.

APPENDIX B
PROOF OF THEOREM 2

Theorem 2 can be proved by considering both the necessary and sufficient conditions. Starting with the necessary condition and assuming that the feasible solution of $P2$ is (α, p), consider the expression in (46),

$$\min_{w}\left\{\frac{R^{f}_{w,s}(\alpha,p)}{P^{f}_{T(w,s)}(\alpha,p)}\right\} \leq \mathcal{Q}^{*}_{w}, \min_{w}\left\{\frac{R^{f}_{w,s}(\alpha^{*},p^{*})}{P^{f}_{T(w,s)}(\alpha^{*},p^{*})}\right\} = \mathcal{Q}^{*}_{w} \tag{46}$$

From (46), (47) is obtained as the following,

$$\begin{aligned}&\min_{w}\left\{R^{f}_{w,s}(\alpha,p) - \mathcal{Q}^{*}_{w}P^{f}_{T(w,s)}(\alpha,p)\right\} \leq 0,\\ &\min_{w}\left\{R^{f}_{w,s}\left(\alpha^{*},p^{*}\right) - \mathcal{Q}^{*}_{w}P^{f}_{T(w,s)}\left(\alpha^{*},p^{*}\right)\right\} = 0\end{aligned} \tag{47}$$

Therefore, (α^{*}, p^{*}) is the optimal solution of $P2$, and the necessary condition proof completed.

In proving the sufficient condition, it is assumed that, (α, p) is the feasible solution and (α^{*}, p^{*}) is the optimal solution of $P2$. Then (48) is given as the following,

$$\begin{aligned}&\min_{w}\left\{R^{f}_{w,s}(\alpha,p) - \mathcal{Q}^{*}_{w}P^{f}_{T(w,s)}(\alpha,p)\right\} \leq 0,\\ &\min_{w}\left\{R^{f}_{w,s}\left(\alpha^{*},p^{*}\right) - \mathcal{Q}^{*}_{w}P^{f}_{T(w,s)}\left(\alpha^{*},p^{*}\right)\right\} = 0\end{aligned} \tag{48}$$

Equation (49) is obtained by arranging (48),

$$\min_{w}\left\{\frac{R^{f}_{w,s}(\alpha,p)}{P^{f}_{T(w,s)}(\alpha,p)}\right\} \leq \mathcal{Q}^{*}_{w}, \min_{w}\left\{\frac{R^{f}_{w,s}(\alpha^{*},p^{*})}{P^{f}_{T(w,s)}(\alpha^{*},p^{*})}\right\} = \mathcal{Q}^{*}_{w} \tag{49}$$

Therefore, (α^{*}, p^{*}) is also the optimal solution of $P1$, and sufficient proof completed.

APPENDIX C
PROOF OF THEOREM 3

For any $\mathcal{Q}^{1}_{w}$ and $\mathcal{Q}^{2}_{w}$, assume that $\mathcal{Q}^{1}_{w} > \mathcal{Q}^{2}_{w}$ and the corresponding optimal solutions are $\left(\alpha^{1}, p^{1}\right)$ and $\left(\alpha^{2}, p^{2}\right)$. Therefore, the energy efficiency can be expressed as in (50).

$$\begin{aligned}F\left(\mathcal{Q}^{1}_{w}\right) &= \max_{(\alpha,p)\in\mathcal{K}}\min_{w}\left\{R^{f}_{w,s}(\alpha,p) - \mathcal{Q}^{1}_{w}P^{f}_{T(w,s)}(\alpha,p)\right\}\\ &= \min_{w}\left\{R^{f}_{w,s}\left(\alpha^{1},p^{1}\right) - \mathcal{Q}^{1}_{w}P^{f}_{T(w,s)}\left(\alpha^{1},p^{1}\right)\right\}\\ &< \min_{w}\left\{R^{f}_{w,s}\left(\alpha^{1},p^{1}\right) - \mathcal{Q}^{2}_{w}P^{f}_{T(w,s)}\left(\alpha^{1},p^{1}\right)\right\}\\ &\leq \min_{w}\left\{R^{f}_{w,s}\left(\alpha^{2},p^{2}\right) - \mathcal{Q}^{2}_{w}P^{f}_{T(w,s)}\left(\alpha^{2},p^{2}\right)\right\}\\ &= F\left(\mathcal{Q}^{2}_{w}\right)\end{aligned} \tag{50}$$

p^{2} is the optimal solution of $F\left(\mathcal{Q}^{2}_{w}\right)$. Hence, $F\left(\mathcal{Q}_{w}\right)$ is strictly a monotonically decreasing function in $\mathcal{Q}_{w}$, which completes the proof.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[3] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving sustainable ultra-dense heterogeneous networks for 5G," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 84–90, Dec. 2017.

[4] *LTE-Advanced: Heterogeneous Networks*, Qualcomm, San Diego, CA, USA, Jan. 2011, pp. 1–13.

[5] K. Bakht, F. Jameel, Z. Ali, W. U. Khan, I. Khan, G. A. S. Sidhu, and J. W. Lee, "Power allocation and user assignment scheme for beyond 5G heterogeneous networks," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–11, Nov. 2019.

[6] S. M. R. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Apr. 2018.

[7] L. Song, Y. Li, Z. Ding, and H. V. Poor, "Resource management in non-orthogonal multiple access networks for 5G and beyond," *IEEE Netw.*, vol. 31, no. 4, pp. 8–14, Jul. 2017.

[8] N. Al-Falahy and O. Y. Alani, "Technologies for 5G networks: Challenges and opportunities," *IT Prof.*, vol. 19, no. 1, pp. 12–20, Jan. 2017.

[9] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Dresden, Germany, Jun. 2013, pp. 1–5.

[10] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. E98.B, no. 3, pp. 403–414, Mar. 2015.

[11] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. 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.

[12] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[13] R. Gupta and E. C. Strinati, "Green scheduling to minimize base station transmit power and UE circuit power consumption," in *Proc. IEEE 22nd Int. Symp. Pers., Indoor Mobile Radio Commun.*, Toronto, ON, Canada, Sep. 2011, pp. 2424–2429.

[14] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[15] L. Xiang and H. Chen, "Energy-efficient and fair power allocation approach for NOMA in ultra-dense heterogeneous networks," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery (CyberC)*, Nanjing, China, Oct. 2017, pp. 89–94.

[16] F. Fang, J. Cheng, Z. Ding, and H. V. Poor, "Energy efficient resource optimization for a downlink NOMA heterogeneous small-cell network," in *Proc. IEEE 10th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Sheffield, U.K., Jul. 2018, pp. 51–55.

[17] T. Han, J. Gong, X. Liu, S. M. R. Islam, Q. Li, Z. Bai, and K. S. Kwak, "On downlink NOMA in heterogeneous networks with non-uniform small cell deployment," *IEEE Access*, vol. 6, pp. 31099–31109, 2018.

[18] H. Zhang, F. Fang, J. Cheng, K. Long, W. Wang, and V. C. M. Leung, "Energy-efficient resource allocation in NOMA heterogeneous networks," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 48–53, Apr. 2018.

[19] Y. Li, M. Sheng, X. Wang, Y. Zhang, and J. Wen, "Max–min energy-efficient power allocation in interference-limited wireless networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4321–4326, Sep. 2015.

[20] H. Shi, R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 5–24, Qua. 2014.

[21] M. Moltafet, P. Azmi, N. Mokari, M. R. Javan, and A. Mokdad, "Optimal and fair energy efficient resource allocation for energy Harvesting-enabled-PD-NOMA-based HetNets," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2054–2067, Mar. 2018.
[22] A. J. Muhammed, Z. Ma, P. D. Diamantoulakis, L. Li, and G. K. Karagiannidis, "Energy-efficient resource allocation in multicarrier NOMA systems with fairness," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8639–8654, Dec. 2019.
[23] S. He, Y. Huang, S. Jin, F. Yu, and L. Yang, "Max-min energy efficient beamforming for multicell multiuser joint transmission systems," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1956–1959, Oct. 2013.
[24] Y. Li, M. Sheng, C. W. Tan, Y. Zhang, Y. Sun, X. Wang, Y. Shi, and J. Li, "Energy-efficient subcarrier assignment and power allocation in OFDMA systems with max-min fairness guarantees," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3183–3195, Sep. 2015.
[25] H. Zuo and X. Tao, "Power allocation optimization for uplink non-orthogonal multiple access systems," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Oct. 2017, pp. 1–5.
[26] R. Ruby, S. Zhong, H. Yang, and K. Wu, "Enhanced uplink resource allocation in non-orthogonal multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1432–1444, Mar. 2018.
[27] Z. Yang, W. Xu, C. Pan, Y. Pan, and M. Chen, "On the optimality of power allocation for NOMA downlinks with individual QoS constraints," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1649–1652, Jul. 2017.
[28] I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "Sum-rate maximization of NOMA systems under imperfect successive interference cancellation," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 474–477, Mar. 2019.
[29] M. Shipon Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.
[30] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.
[31] M. R. Zamani, M. Eslami, M. Khorramizadeh, and Z. Ding, "Energy-efficient power allocation for NOMA with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1009–1013, Jan. 2019.
[32] G. Liu, R. Wang, H. Zhang, W. Kang, T. A. Tsiftsis, and V. C. M. Leung, "Super-modular game-based user scheduling and power allocation for energy-efficient NOMA network," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3877–3888, Jun. 2018.
[33] Z. J. Ali, N. K. Noordin, A. Sali, F. Hashim, and M. Balfaqih, "An efficient method for resource allocation and user pairing in downlink non-orthogonal multiple access system," in *Proc. IEEE 14th Malaysia Int. Conf. Commun. (MICC)*, Selangor, Malaysia, Dec. 2019, pp. 124–129.
[34] Z. J. Ali, N. K. Noordin, A. Sali, F. Hashim, and M. Balfaqih, "Novel resource allocation techniques for downlink non-orthogonal multiple access systems," *Appl. Sci.*, vol. 10, no. 17, p. 5892, Aug. 2020.
[35] Z. Song, Q. Ni, and X. Sun, "Distributed power allocation for nonorthogonal multiple access heterogeneous networks," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 622–625, Mar. 2018.
[36] W. U. Khan, Z. Yu, S. Yu, G. A. S. Sidhu, and J. Liu, "Efficient power allocation in downlink multi-cell multi-user NOMA networks," *IET Commun.*, vol. 13, no. 4, pp. 396–402, Mar. 2019.
[37] D. Ni, L. Hao, Q. T. Tran, and X. Qian, "Power allocation for downlink NOMA heterogeneous networks," *IEEE Access*, vol. 6, pp. 26742–26752, 2018.
[38] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, "Resource allocation for non-orthogonal multiple access in heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.
[39] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2771–2784, Dec. 2017.
[40] J. Zeng, B. Li, Y. Mao, X. Su, and Y. Kuang, "Research of heterogeneous networks convergence with NOMA," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Shenzhen, China, Nov. 2015, pp. 1–5.
[41] W. Xu, R. Qiu, and X.-Q. Jiang, "Resource allocation in heterogeneous cognitive radio network with non-orthogonal multiple access," *IEEE Access*, vol. 7, pp. 57488–57499, 2019.
[42] Z. Yang, C. Pan, W. Xu, Y. Pan, M. Chen, and M. Elkashlan, "Power control for multi-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 927–942, Feb. 2018.
[43] Y. Fu, Y. Chen, and C. W. Sung, "Distributed downlink power control for the non-orthogonal multiple access system with two interfering cells," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[44] C. W. Sung and Y. Fu, "A game-theoretic analysis of uplink power control for a non-orthogonal multiple access system with two interfering cells," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, Nanjing, China, May 2016, pp. 1–5.
[45] N.-T. Le, L.-N. Tran, Q.-D. Vu, and D. Jayalath, "Energy-efficient resource allocation for OFDMA heterogeneous networks," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7043–7057, Oct. 2019.
[46] X. Ge, J. Yang, H. Gharavi, and Y. Sun, "Energy efficiency challenges of 5G small cell networks," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 184–191, May 2017.
[47] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.
[48] E. Bjornson and E. G. Larsson, "How energy-efficient can a wireless communication system become?" in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 1252–1256.
[49] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2667–2680, Dec. 2017.
[50] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: WH Freeman, 1979.
[51] J.-P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Math. Program.*, vol. 52, nos. 1–3, pp. 191–207, May 1991.
[52] A. I. Barros, J. B. G. Frenk, S. Schaible, and S. Zhang, "A new algorithm for generalized fractional programs," *Math. Program.*, vol. 72, no. 2, pp. 147–175, Feb. 1996.
[53] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.
[54] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.
[55] S. P. Boyd, "Sequential convex programming," Stanford Univ., Stanford, CA, USA, Lect. Notes EE364b, 2008, pp. 1–19.
[56] A. Zappone and E. A. Jorswieck, "Energy-efficient resource allocation in future wireless networks by sequential fractional programming," *Digit. Signal Process.*, vol. 60, pp. 324–337, Jan. 2017.
[57] Y.-F. Liu and Y.-H. Dai, "On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 583–596, Feb. 2014.

ZUHURA J. ALI (Student Member, IEEE) received the B.Sc. degree in electronics and telecommunication engineering from Yıldız Technical University, İstanbul, Turkey, in 2007, and the M.Sc. degree in signal and information processing from the Tianjin University of Technology and Education, Tianjin, China, in 2014. She is currently pursuing the Ph.D. degree in wireless communication and network engineering with Universiti Putra Malaysia, Malaysia. She is also a Lecturer with the Karume Institute of Science and Technology, Zanzibar, Tanzania. Her research interests include radio resource allocation in wireless communication systems, non-orthogonal multiple access, and heterogeneous networks.

NOR K. NOORDIN (Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Alabama, USA, in 1987, the M.Eng. degree from Universiti Teknologi Malaysia, and the Ph.D. degree from Universiti Putra Malaysia. She is currently working as a Professor and the Dean of the Faculty of Engineering, Universiti Putra Malaysia. She has published more than 300 journals, book chapters, and conference papers. She has led many research projects. Her research interests include wireless communications and network systems.

ADUWATI SALI (Senior Member, IEEE) received the B.Eng. degree in electrical electronics engineering from the University of Edinburgh, U.K., in 1999, the M.Sc. degree in communications and network engineering from Universiti Putra Malaysia, in 2002, and the Ph.D. degree in mobile and satellite communications from the University of Surrey, U.K., in 2009. She was the Deputy Director of the Research Management Centre (RMC), Universiti Putra Malaysia, from 2016 to 2019. She is currently a Professor with the Department of Computer and Communication Systems Engineering. Her research interest includes mobile and satellite communication systems.

FAZIRULHISYAM HASHIM (Member, IEEE) received the M.Sc. degree from Universiti Sains Malaysia and the Ph.D. degree in telecommunication engineering from the University of Sydney, Australia. He is currently an Associate Professor with the Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia. His research interests include heterogeneous wireless communication systems and network security.