



A novel approach to fairness-aware energy efficiency in green heterogeneous cellular networks

Fereshteh Atri Niasar¹ · Amir Reza Momen^{1,2} · Seyed Abolfazl Hosseini¹

Accepted: 4 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Heterogeneous cellular networks are a viable solution in response to the growing demand for broadband services in the new-generation wireless networks. The dense deployment of small cell networks is a key feature of next-generation heterogeneous networks aimed at providing the necessary capacity increase. However, the approach to apply green networks is very important especially in the downlink because uncontrolled deployment of too many small-cells may increase operational costs and emit more carbon dioxide. In addition, given the novel services and resource limitation of the user layer, energy efficiency and fairness assurance are critical issues in the uplink. Considering the uplink fairness criterion, this paper proposes a dynamic optimization model which maximizes the total uplink/downlink energy efficiency in addition to providing the essential coverage and capacity of heterogeneous cellular networks. Based on the non-convex characteristics of the energy efficiency maximization model, the mathematical model can be formulated to two sub-problems, i.e., resource optimization and user association. So that, a subgradient method is applied for fair resource management and also successive convex approximation and dual decomposition methods are adopted to solve the proportional fairness problem. The simulation results exhibit considerable throughput increase by 30% and 22% on average for random and hotspot user distributions, respectively. It also proved that the proposed approach managed to significantly improve the total network energy efficiency by up to 35%.

Keywords Green heterogeneous networks · Hierarchical resource management · Proportional fairness · Uplink/Downlink NOMA · User association

1 Introduction

1.1 Background and related works

In recent years, heterogeneous networks (HetNets) have become very popular among active key technologies employed for exponentially growing traffic requirements in the next-generation mobile networks [1–3]. The deployment of ultra-dense small-cells will probably be a major component of next-generation wireless networks, to manage the increasing traffic of mobile networks and to offload the traffic of high-utilized macro cells in order to improve

users' quality of service (QoS) [4–6]. One of the most considerable criteria in small-cell deployment is the backhauling technology applied to connect small-cells to the wireless core system, Millimeter-wave backhauling can speed up the deployment of new generation small-cell networks by mitigating deployment costs [7, 8]. Nonetheless, it is impossible to ignore the multihop technology in mesh backhauling communications because of applying directional beams and avoiding the severe fading effects of long-range backhauled. This approach needs impressive radio resource management (RRM) and effective routing strategies [9] in order to obtain optimal power utilization and decrease delay.

The hierarchical resource management relies on coallocation algorithm that jointly allocate computing and networking resources considering both data execution and data transfer times. The total energy consumption might also be very high in the access layer of HetNets due to the vast number of deployed small-cells and ineffective

✉ Amir Reza Momen
momen.scholar@gmail.com

¹ Department of Electrical Engineering, Islamic Azad University, Shahr-E-Ray Branch, Tehran, Iran

² Islamic Azad University, North Tehran Branch, Tehran, Iran

millimeter-wave boosters. Therefore, many researchers have tried to reduce the total energy consumed by small-cell networks through different methods such as intentionally turning off some of the network nodes. For instance, the authors in [10] suggested turning off macro cells or small-cells heuristically or changing the size of small-cells. In [11], the authors investigated the simple resource allocation (RA) problem with an assumption of perfect channel state information (CSI), for a single antenna scenario in Internet-of-Things (IoT) networks. The authors in [12] studied backhaul-aware joint user association (UA) and RA to obtain network utility reflecting proportional fairness in simultaneous wireless information and power transfer (SWIPT) with the total energy constraint. Considering the fairness index, the UA problem can be formulated as a utility maximization function. Most of the current research on UA for HetNets have focused on throughput maximization [13] assuming successive interference cancellation (SIC) capability, in such a way that the optimal capacity and throughput are achievable via jointly UA optimization and carrier scheduling (CS) within each cluster. Applying difference of convex functions (DC) programming, [14] presented a joint cell association and CS algorithm to find the maximum logarithmic network's data rate to obtain acceptable proportional fairness. Accordingly, joint resource allocation and data flow scheduling were analyzed in energy harvesting (EH) empowered sensor networks [15]. The authors of [16] integrated EH-empowered SWIPT with multihop HetNets to increase the network's throughput and efficiency, but the computational complexity of the scheme is considerably high. In [17], a precise EH-enabled model for hierarchical non-orthogonal multiple access (NOMA) HetNets is formulated as joint global power optimization and carrier allocation (CA) problems, which this model is transformed into a mixed-integer nonconvex problem to guarantee QoS constraints applying successive codebook ordering assignment (SOCA). The main target of this approach is the minimization of power consumption with lower time complexity which outperforms OFDMA from the EE perspective. The authors in [18] applied the NOMA technique in the cellular HetNets and studied the trade-off between the EH, proportional fairness, EE, and network throughput.

In order to guarantee QoS requirements in service-oriented hierarchical networks, fairness assurance is one of the most critical design criteria in heterogeneous networks [19]. As various services are provided on the platform of the next-generation mobile networks, the fairness issue in HetNets has drawn considerable attention. In [20], a fair user association scheme was proposed with load balancing capability for HetNets, and a logarithmic utility maximization problem was formulated at the network level to improve the resource allocation efficiency using successive

convex approximation (SCA) method. In [21], the researchers designed a scheme for user association and power allocation under different levels of load and transmission power limitations with respect to imperfect channel state information and proportional fairness in order to improve spectral efficiency. In [22], considering enhanced inter-cell interference coordination (eICIC), the authors formulated and solved a power allocation problem based on proportional fairness and sub-channel assignment for uplink multi-tier HetNets with the internetwork radio resource control capability. In [23] the authors analyzed the effects of user distribution patterns on the fair EE development for resource allocation and subcarrier scheduling in HetNets based on power domain-NOMA (PD-NOMA). The authors in [24] suggested an efficient intercellular algorithm with distributed interference coordination to improve EE and load balancing in ultra-dense HetNets. In [25], the authors formulated fair resource allocation and subcarrier scheduling with the CSI assumption for the uplink of heterogeneous backhauls considering internetwork interactions. Also, two EE-enhanced ICIC approaches were introduced for ultra-dense HetNets in [26] with respect to some applicable features of NGMN such as beamforming and carrier aggregation. In many of the previous studies, the researchers only analyzed the UA issue for guaranteeing proportional fairness but practically disregarded energy-aware hierarchical resource management [27]. Moreover, a few authors focused only on the wireless downlink heterogeneous networks without considering the adaptability of the proposed scheme to the reliability conditions and user fairness requirements of green heterogeneous cellular networks [28–30].

1.2 Motivations

To this end, most researchers who worked on energy efficiency (EE) and multi-layer resource management in heterogeneous systems have merely focused on joint UA and RA or EH and backhaul traffic analysis without considering user fairness criteria in uplink and reliability requirements of the practical networks. They fail to provide the necessary capability to control unstable transmission links and reliable strategies to simultaneously determine the ON/OFF switching frequency of small-cells for power-saving. Lack of backhaul traffic optimization compatible with ultra-dense networks is another drawback of most existing works. In the proposed backhauling model, each NOMA relay is able to be applied for dual-hop as well as multi-hop simultaneously. Most studies on backhaul traffic optimization have particularly assumed line of sight (LoS) connections for mmWave backhaul links without blockage handling strategy but such system assumptions are not feasible for areas with obstacles such as ultra-dense zones.

Unlike the conventional EE studies, which have not well considered the system reliability, we implemented the Crosswave propagation model with the optimized morphology indexes for accurate loss estimating and compatibility with the real ultra-dense urban areas which supports concurrent LoS and NLoS backhaul communications. This model also performs routing in the multi-hop mesh backhaul to efficiently use the existing infrastructure of small-cell networks for simultaneous dual-hop/multi-hop transmissions considering both uniform and hotspot user equipment (UE) distribution patterns. In practice, as small-cells interconnect to form a mesh topology, determination of the transmission strategy becomes one of the most effective aspects of the multi-hop networks' performance management, which should be taken into account in addition to determining the best cell for every user. These mentioned capabilities of the proposed approach verify the feasibility of the system assumptions and increase its reliability.

In this paper, the energy efficiency index is modeled as the total network EE maximization. The EE model is then formulated as maximization of the worse EE with respect to the uplink fairness index. The network KPI indices and optimal UE transmission power are guaranteed simultaneously based on the proportional fairness criteria. In addition, the dual decomposition method and the SCA method were applied to find the optimal solution for user association and resource allocation, respectively. According to the numerical analyses, the introduced optimal fairness approach yielded better energy efficiency results compared to the other existing algorithms. According to the literature review, the proposed approach is the first attempt to joint user association, fair power optimization, and dynamic UL/DL transmission control in dual-hop/multi-hop backhaul configuration of reliable HetNets with energy harvesting capability.

1.3 Paper organization

The paper is organized as follows: after briefly summarizing the related works and mentioning the motivation in the introduction part, Sect. 2 describes the system/signaling model and problem formulation for different fair resource allocation strategies. In Sect. 3, solutions and algorithms to solve the proposed optimization problems for carrier scheduling & resource allocation are investigated. Also, Sect. 4 introduced the fairness-based power optimization approach. In Sect. 5, simulation results are presented and interpreted in order to exhibit the effectiveness of different optimization algorithms. In this section, we described the simulation scenarios and environments and we tried to evaluate the performance of the proposed algorithms accurately. Finally, the concluding remarks are mentioned

in Sect. 6. Note that, the investigation on the different transmission strategies and the proposed backhauling method are presented in the appendix with focusing on the NOMA-based hybrid Dual-hop/multi-hop model for backhaul communications. Also, the main applied notations and symbols are presented in Table 1.

2 System model and problem formulation

2.1 System model

As shown in Fig. 1, in this NOMA-based hierarchical network, there is a single macro base station and K small-cells located in each cluster in such a way that each small-cell is empowered with the energy harvesting facilities [31]. All of the macro and small-cells are demonstrated by $k \in \{1, 2, \dots, K+1\}$, so that cell $k = K+1$ is the macro cell. $M_k \in \{M_1, M_2, \dots, M_{K+1}\}$ denotes the number of associated UE to cell k . BW indicates the bandwidth which is divided to N carriers, $n \in \{1, 2, \dots, N\}$, and each carrier's bandwidth is equal to $B_{sc} = BW/N$. In this notation, $h_{k,j,m,n}$ indicates carrier gain of UE m of BS k to BS j on subcarrier n where $m \in \{1, 2, \dots, M_K\}$. Also, $s_{k,m,n}$ is a validity factor of the carrier allocation process so that $s_{k,m,n} = 1$ if UE m is associated with cell k on subcarrier n , otherwise, $s_{k,m,n} = 0$. $p_{k,m,n}$ indicates the transmission power of UE m associated with cell k on carrier n . according to the presented notations, the carriers' vector and the set of allocated resources are denoted by $S = [s_{k,m,n}]$ and $P = [p_{k,m,n}]$ respectively. Based on the Shannon's ergodic capacity theorem as the upper bound of the capacity on the statistics channel (i.e. time-varying channel) which can be evaluated by averaging the capacity is obtained at a particular time instance on a fading channel over an infinite time interval, the maximum data rate of UE m associated with cell k on carrier n is obtained as

$$r_{k,m,n} = B_{sc} \log_2(1 + SINR_{k,m,n}). \quad (1)$$

In which $SINR_{k,m,n}$ indicates SINR of cell k on the n th carrier of UE m . In NOMA, a distinct carrier can be allocated to multiple UEs at the same time, so the demodulation and decoding should be done at the receiver's side by utilizing the SIC method, in which SIC removes the interference's effect based on the UEs' power level so that each UE can be distinctly distinguished by the base station (BS).

In this scenario, it is assumed that a distinct carrier can be dedicated to two different UEs, and each receiver first decodes the UE with the higher power level on the identical carrier. So, the received signal to noise/interference ratio of UE m relevant to carrier n of cell k can be expressed as

Table 1 The summary of notations

Notations	Descriptions
K & M_k ,	Number of small-cells & Associated UEs
$S = [s_{k,m,n}]$	Carriers' vector
$P = [p_{k,m,n}]$	Set of allocated resources
σ^2 & i	White noise & number of iterations
$P_{j,n}$	Total power of cell j on carrier n
$h_{k,k,m,n}$	Channel gain of UE m of BS k to BS j on carrier n
$r_{k,m,n}$	Throughput of UE m on carrier n
$Q(S, P)$ & $H_{k,m,n}$	Transmission power & harvested energy
$\lambda_{j,n}$	Energy harvesting factor
$P(S, P)$	Circuit power consumption
$R_{k,\min}$,	QoS threshold of cell k 's data rate
$\beta_{k,m,n}, \alpha_{k,m,n}$	Suxiliary variables of convex transformation
$\hat{R}(S, P)$ & $U(S, P)$	Modified total sum rate & total power consumption
$Z_k(n)$	Set of UEs dedicated to carrier n associated to cell k
μ, ν, ξ, a_n	Lagrangian coefficients
$D(\mu, \nu, \xi)$	Dual function
$\delta_1(i)$	The i 'th iteration's step-size
y_{R_j}	Received signal at j th intermediate relay
a_i	Resource allocation coefficient
$\gamma_{R_j}^{x_j}$	Signal to noise ratio for x_j and x_0 at R_j
ζ & g_{gS,R_j}	SIC perfection index & remaining interference
\mathcal{O}^{x_j} & \widetilde{C}_{x_j}	Outage probability of x_j & ergodic capacity for x_j
I_{k,m_0}	Total interference from the other UEs
Y_{m_0}	The received signal vector
∇_k	PPP-based initial position pattern of UEs
ξ	Uplink energy consumption coefficient
d_{k,m_0}	Distance between UE m_0 and the k th BS
E_{\max}	Upper bound of the number of iterations
ϕ_{blocking}	Obstacle distribution pattern

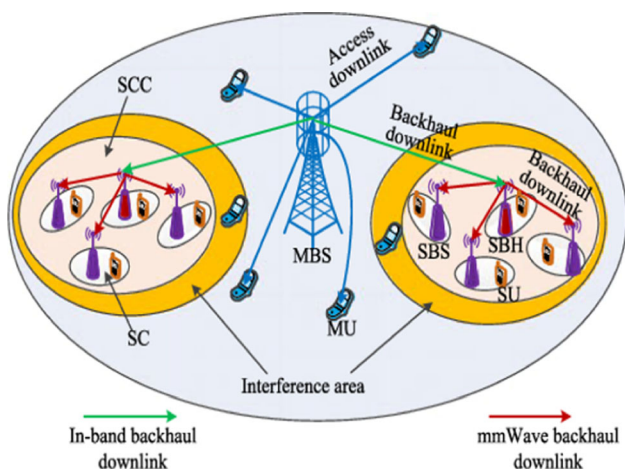


Fig. 1 Heterogeneous System model with multiple backhauls connections

$$SINR_{k,m,n} = \frac{s_{k,m,n} p_{k,m,n} |h_{k,k,m,n}|^2}{|h_{k,k,m,n}|^2 \sum_{r=m+1}^{M_k} s_{k,r,n} p_{k,r,n} + \sum_{j=1, j \neq k}^{K+1} p_{j,n} |h_{k,j,m,n}|^2 + \sigma^2} \quad (2)$$

In which σ^2 denotes white noise and $p_{j,n} = \sum_{m=1}^{M_j} s_{j,r,n} p_{j,r,n}$ represents the total resource of cell j on the carrier n . $s_{k,m,n}$ represents the subchannel allocation index. Actually, $s_{k,m,n}$ determines an exponent to the channel allocation. If UE n is allocated to BS k on the channel n , $s_{k,m,n}$ is equal to 1, otherwise, $s_{k,m,n}$ can be considered equal to zero. $h_{k,k,m,n}$ also indicates channel gain of UE m of BS_k to BS_j on channel n , in which $m \in \{1, 2, \dots, M_K\}$.

2.2 Signal model

This section describes the network and signaling model. According to Fig. 2, the proposed scenarios include a two-tier uplink MIMO-enabled HetNet consisting of certain clusters with a macro cell and several scattered small-cells. The relevant BSs of different tiers operate on orthogonal frequency bands to eliminate the effect of cross-tier interference. In this scheme, the operational bandwidth of a macrocell is assumed to be equal to those of small-cells; in other words, $B_m = B_s = B$. In this paper, the user distribution pattern is a combination of random and hotspot, and the random walk mobility model is used for moving UEs. The initial positions of UEs are determined with densities of ∇_k , through the Poisson point process. According to the proposed system model, UE m_0 is assumed to be connected to the k^{th} BS with N_k antenna. If s_{m_0} is the signal transmitted to the k^{th} BS from the UE m_0 , the received signal vector Y_{m_0} in this BS can be expressed as below:

$$Y_{m_0} = \sqrt{P_{m_0}^T d_{k,m_0}^{-\alpha}} \mathbf{h}_{k,m_0} s_{m_0} + \sum_{i \neq m_0} \sqrt{P_i^T d_{k,i}^{-\alpha}} \mathbf{h}_{k,i} s_i + n \quad (3)$$

where $E = [s_{m_0}] = 0$ and $E = [\|s_{m_0}\|^2] = 1$. In this scenario, d_{k,m_0} indicates the distance between UE m_0 and the k^{th} BS, whereas α refers to the power reduction of the path loss. Moreover, the value of α is always greater than 2. During the uplink transmission stage, all UEs send the orthogonal pilot sequences corresponding to $g_{m_0} = h_{k,m_0}^H$ through maximal ratio combining (MRC) simultaneously in the BS for channel estimation. In this case, the received signal of the k^{th} base station from user m_0 is achievable as below:

$$\begin{aligned} Z_{m_0} &= h_{k,m_0}^H Y_{m_0} \\ &= \sqrt{P_{m_0}^T d_{k,m_0}^{-\alpha}} \|h_{k,m_0}\|^2 s_{m_0} + \sum_{i \neq m_0} \sqrt{P_i^T d_{k,i}^{-\alpha}} h_{k,m_0}^H h_{k,i} s_i \\ &\quad + h_{k,m_0}^H n \end{aligned} \quad (4)$$

Therefore, the uplink signal to noise/interference ratio of the m^{th} UE associated to the k^{th} base station is modified as follows:

$$\gamma_{k,m_0} = \frac{P_{m_0}^T d_{k,m_0}^{-\alpha} \|h_{k,m_0}\|^2}{I_{k,m_0} + \sigma^2}, \quad (5)$$

where $I_{k,m_0} = \sum_{i=1, i \neq m_0}^{|M|} P_i^T \left| \frac{h_{k,m_0}^H h_{k,i}}{h_{k,m_0}^H} \right|^2 d_{k,i}^{-\alpha}$ refers to the total interference from the other UEs connected to the BSs of a similar tier, and $|M|$ denotes the number of all users allocated to the set of macrocells or small-cells.

2.3 Uplink energy consumption model

In this uplink power optimization approach, the energy consumption model of every user equipment consists of 2 major portions: static and dynamic. In other words, the static portion pertains to the functions of the hardware elements like converters, amplifiers, and processor units. However, the dynamic portion is referred to as the necessary energy for transmission. In this scenario, the static energy model of the user equipment is considered PCU, whereas the transmission energy consumption of the m^{th} user is denoted as P_m^T . The total energy consumption (P_m^{sum}) of the m^{th} UE can be defined as below: $P_m^{sum} = P_{CU} + \xi P_m^T$, which ξ denotes the energy utilization index. As discussed earlier, EE (bit/joule) pertains to the relationship between the m^{th} UE and the k^{th} BS. It is defined as (6):

$$\eta_{k,m} = r_{k,m} / P_m^{sum} \quad (6)$$

where $r_{k,m}$ refers to the accessible throughput of the m^{th} user connected to the k^{th} base station. Based on the Shannon theory, $r_{k,m}$ is calculated through the following equation:

$$r_{k,m} = B \log_2(1 + \gamma_{k,m}) \quad (7)$$

2.4 Problem formulation

In this section, the model is formulated jointly to solve user association and fair resource allocation with respect to the proportional fairness index. we try to maximize the entire network's energy efficiency and power utility by formulating the correlation between the network's throughput and total power consumption. So, the goal function can be

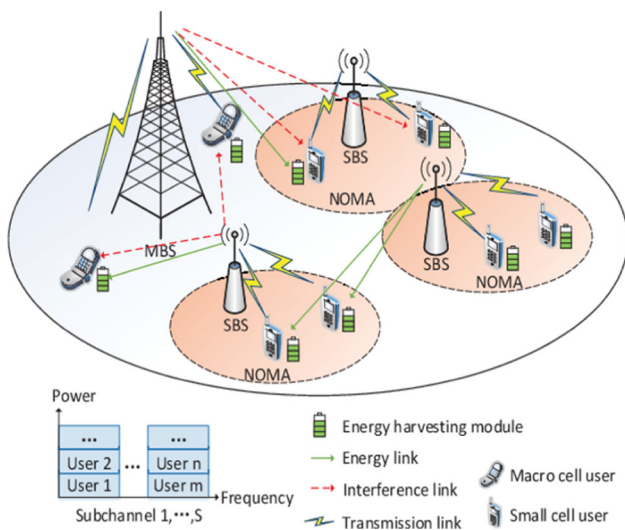


Fig. 2 Multi-tier NOMA based heterogeneous system

formulated based on obtaining the maximum data rate without violating the total power constraints. During the carrier and resource allocation process, base stations are empowered by energy harvesting capability. With these assumptions, the network's total sum rate is formulated as:

$$R(S, P) = \sum_{k=1}^{K+1} \sum_{m=1}^{M_k} \sum_{n=1}^N r_{k,m,n} \tag{8}$$

We can formulate the overall transmission power as

$$Q(S, P) = \sum_{k=1}^{K+1} \sum_{m=1}^{M_k} \sum_{n=1}^N s_{k,m,n} p_{k,m,n}. \tag{9}$$

We can also calculate the harvested energy of user m from cell k via Eq. (10). Which n indicates the carrier.

$$H_{k,m,n} = \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} s_{k,m,n} p_{k,m,n} |h_{k,j,m,n}|^2 \tag{10}$$

In this equation, $\lambda_{j,n}$ is a constant indicator of power harvesting efficiency. Considering the circumstances of the scenario and Eq. (10), the overall power harvested by network is calculated via Eq. (11)

$$H(S, P) = \sum_{k=1}^{K+1} \sum_{m=1}^{M_k} \sum_{n=1}^N \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} s_{k,m,n} p_{k,m,n} |h_{k,j,m,n}|^2. \tag{11}$$

In accordance that the circuit resource consumption of the cell is significantly less than the overall harvested power, hence, the final consumed resources can be exhibited as

$$U(S, P) = Q(S, P) - P(S, P) = \sum_{k=1}^{K+1} \sum_{m=1}^{M_k} \sum_{n=1}^N s_{k,m,n} p_{k,m,n} \left(1 - \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} |h_{k,j,m,n}|^2 \right) \tag{12}$$

The power efficiency index is computed by Eq. (13)

$$EE(S, P) = \frac{R(S, P)}{Q(S, P) - P(S, P)} = \frac{R(S, P)}{U(S, P)}. \tag{13}$$

So, based on the defined variables, the target optimization function can be mentioned as

$$\begin{aligned} \mathbf{P1} : \max_{S, P} EE(S, P) &= \max_{S, P} \frac{R(S, P)}{U(S, P)} \\ \text{s.t. } C1 : \sum_{m=1}^{M_k} \sum_{n=1}^N s_{k,m,n} p_{k,m,n} &\leq P_{k,max}, \quad \forall k \\ C2 : p_{k,m,n} &\geq 0, \quad \forall k, m, n \\ C3 : s_{k,m,n} &\in \{0, 1\}, \quad \forall k, m, n \\ C4 : \sum_{m=1}^{M_k} s_{k,m,n} &\leq 2, \quad \forall k, n \\ C5 : \sum_{m=1}^{M_k} \sum_{n=1}^N s_{k,m,n} r_{k,m,n} &\leq R_{k,min}, \quad \forall k \\ C6 : \sum_{k=1}^K \sum_{m=1}^{M_k} \sum_{n=1}^N s_{k,m,n} p_{k,m,n} |h_{k,k+1,n}|^2 &\leq I_{max} \end{aligned} \tag{14}$$

In which C1 and C2 are resource allocation constraints guaranteeing that the UE's resource is not negative and is less than the cell's total energy; the restrictions C3 and C4 ensure that the maximum number of UEs with the capability of using a special subchannel at the same time is 2; C5 represents QoS and data rate constraints; and C6 is related to cross-tier interference restriction in which $|h_{k,k+1,n}|^2$ denotes UE's gain from a small cell to the macro cell.

In addition to ensuring power utility, the goal is to guarantee proportional fairness for the maximization of the minimum EE in uplink communications of all users. Because the logarithmic goal function is nonconvex, it is effectively employed to develop the cost function based on the fairness criterion. The accessible efficiency value of the m th user equipment is obtained as the logarithmic form of EE, which is expressed as $\ln R_{k,m} / P_k^{sum}$. In this problem, power optimization is aimed at achieving the highest level of the total utility function for users. It is defined as Eq. (15).

$$\sum_{m=1}^M \ln \left(\sum_{k=1}^K r_{k,m} / P_m^{sum} \right) = \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \ln (r_{k,m} / P_m^{sum}). \tag{15}$$

The goal function of this optimization problem can be formulated as below by defining μ_m as the auxiliary with the value of $\mu_m = (r_{k,m} / P_m^{sum})$. So, The second target function as the proportional fairness problem is formulated as P2:

$$\begin{aligned}
 \mathbf{P2} : & \max_{X,P} \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \mu_{k,m} \\
 s.t. \text{ C1} : & 0 \leq P_m^T \leq P_m^{TMAX}, \quad m = 1, 2, \dots, M, \\
 \text{C2} : & \sum_k x_{k,m} \gamma_{k,m} \geq \gamma_{th}, \quad k = 1, 2, \dots, K, \quad m = 1, 2, \dots, M, \\
 \text{C3} : & x_{k,m} \in \{0, 1\}, \quad k = 1, 2, \dots, K, \quad m = 1, 2, \dots, M, \\
 \text{C4} : & \sum_k x_{k,m} = 1, \quad m = 1, 2, \dots, M.
 \end{aligned} \tag{16}$$

In the uplink goal function, C1 denotes that the transmission power of every user equipment cannot be more than the maximum allowed transmission power, Also, C2 guarantees the requirements of QoS indices for every user in a way that the minimum predefined $SINR_{\gamma_{th}}$ is guaranteed. Moreover, C3–C4 constraints help ensure that every UE can be connected only to one base station at a specific time.

3 Fairness-guaranteed carrier allocation and power optimization in NOMA HetNets

This section describes the proposed solution and then presents the process of solving the formulated optimization problems of the EE proportional fairness index. In this scenario, the energy-efficient problem is not a convex problem, and the described goal function (14) is a non-linear fraction. To find the optimal solution for this problem, we have to modify the formulation. At the beginning, the model presented in [32] was applied to estimate the convex alteration to find the lower bound of the UE’s throughput.

$$\hat{r}_{k,m,n} = B_{SC} \alpha_{k,m,n} \log_2(SINR_{k,m,n}) + \beta_{k,m,n} \tag{17}$$

where,

$$\alpha_{k,m,n} = \frac{\overline{SINR}_{k,m,n}}{\overline{SINR}_{k,m,n} + 1} \tag{18}$$

$$\begin{aligned}
 \beta_{k,m,n} = & \log_2(1 + \overline{SINR}_{k,m,n}) \\
 & - \frac{\overline{SINR}_{k,m,n}}{1 + \overline{SINR}_{k,m,n}} \log_2(\overline{SINR}_{k,m,n})
 \end{aligned} \tag{19}$$

In which, $\overline{SINR}_{k,m,n}$ shows the value of the final iteration’s signal to noise/interference ratio.

So, with considering the problem restrictions, the goal function was expressed as Eq. (14) will be reformulated as the following

$$\max_{S,P} EE(S,P) = \max_{S,P} \frac{\hat{R}(S,P)}{U(S,P)} \tag{20}$$

$$\begin{aligned}
 s.t. \text{ C5}' : & \sum_{m=1}^{M_k} \sum_{n=1}^N s_{k,m,n} \hat{r}_{k,m,n} \leq R_{k,min}, \quad \forall k \\
 & \text{C1} - \text{C4}, \text{C6}.
 \end{aligned} \tag{21}$$

However, because the goal function is not linear, its fractional formulation is altered to a subtraction form to reduce the computational complexity. Here, a parameter t , as the power efficiency index, is needed, which is defined as the following

$$t^* = \max_{S,P} \frac{\hat{R}(S,P)}{U(S,P)} = \frac{\hat{R}(S^*,P^*)}{U(S^*,P^*)} \tag{22}$$

Hence,

$$\hat{R}(S^*,P^*) - t^*U(S^*,P^*) = 0 \tag{23}$$

So, the goal function was formulated as (14) in addition to its six restrictions can be re-defined as below

$$\begin{aligned}
 & \max_{S,P} \hat{R}(S,P) - tU(S,P) \\
 s.t. \text{ C1} - \text{C4}, \text{C5}', \text{C6}.
 \end{aligned} \tag{24}$$

3.1 Carrier allocation algorithm

Based on the baseline idea of NOMA which is to serve multiple users using the same resource in terms of time, frequency, and space the carrier allocation process, UEs, cells, and carriers should be associated. In accordance with [33], a non-complex effective carrier matching algorithm has been presented for determining matrix S using DC programming and the multi-sided matching method [34]. The algorithm has two major stages. The first of which the UEs will be dedicated to the carriers based on the carriers’ quality, such that the UE with the best carrier quality is dedicated to the corresponding carrier. During the second stage, two UEs that are able to minimize power utilization are selected to achieve the goal, which is energy efficiency maximization. This procedure of the carrier matching is summarized step by step in Algorithm 1.

We consider $Z_k(n)$ as the set of UEs dedicated to carrier n associated to cell k and $\overline{Z}_k(n)$ as the UEs which have not been assigned to carriers on cell k . The EE of cell k and carrier n can be represented by the following

$$EE_{k,n} = \frac{\sum_{m=1}^{M_k} r_{k,m,n}}{\sum_{m=1}^{M_k} s_{k,m,n} P_{k,m,n} - \sum_{m=1}^{M_k} \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} s_{k,m,n} P_{k,m,n} |h_{k,j,m,n}|^2} \tag{25}$$

3.2 Power optimization algorithm

Considering the presented carrier allocation strategy, in this part of the paper we propose the power optimization approach to maximize the energy efficiency of the

problem. In this framework, we can assume $s_{k,m,n}$ as a constant parameter during the optimization process. So, if we consider $\tilde{P}_{k,m,n} = s_{k,m,n}P_{k,m,n}$, the new form of the goal function can be defined as (26)

$$\begin{aligned} & \max_{\tilde{P} > 0} \hat{R}(\tilde{P}) - tU(\tilde{P}) \\ \text{s. t. } & C1 : \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{p}_{k,m,n} \leq P_{k,max}, \quad \forall k \\ & C2 : \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{r}_{k,m,n} \leq R_{k,min}, \quad \forall k \\ & C3 : \sum_{k=1}^K \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{p}_{k,m,n} |h_{k,k+1,n}|^2 \leq I_{max}. \end{aligned} \tag{26}$$

Therefore, the goal function and signal to noise/interference ratio are represented as the following

$$\hat{R}(\tilde{P}) - tU(\tilde{P}) = \sum_{k=1}^{K+1} \sum_{m=1}^{M_k} \sum_{n=1}^N \left[\tilde{r}_{k,m,n} - t \left(1 - \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} |h_{k,j,m,n}|^2 \right) \tilde{p}_{k,m,n} \right] \tag{27}$$

$$\begin{aligned} \widetilde{SINR}_{k,m,n} &= \frac{\tilde{p}_{k,m,n} |h_{k,k,m,n}|^2}{|h_{k,k,m,n}|^2 \sum_{r=m+1}^{M_k} \tilde{p}_{k,r,n} + \sum_{j=1, j \neq k}^{K+1} \tilde{p}_{j,n} |h_{k,j,m,n}|^2 + \sigma^2} \\ \text{where } \tilde{p}_{j,n} &= \sum_{r=1}^{M_j} \tilde{p}_{j,r,n}. \end{aligned} \tag{28}$$

One of the best methods for solving this problem is the Lagrangian subgradient method in which the optimal solution of the primary problem is obtainable with solving its dual problem. The Lagrange equation of the goal function is formulated as the following in which $\mu = [\mu_1, \mu_2, \dots, \mu_{k+1}]$, $v = [v_1, v_2, \dots, v_{k+1}]$ and ξ represent the Lagrangian coefficients.

$$\begin{aligned} L(\tilde{P}, \mu, v, \xi) &= \sum_{k=1}^{K+1} \sum_{m=1}^{M_k} \sum_{n=1}^N \left[\tilde{r}_{k,m,n} - t \left(1 - \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} |h_{k,j,m,n}|^2 \right) \tilde{p}_{k,m,n} \right] \\ &+ \sum_{k=1}^{K+1} \mu_k \left(P_{k,max} - \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{p}_{k,m,n} \right) + \sum_{k=1}^{K+1} v_k \left(\sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{r}_{k,m,n} - R_{k,min} \right) + \xi \\ &\left(I_{max} - \sum_{k=1}^K \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{p}_{k,m,n} |h_{k,k+1,n}|^2 \right) \end{aligned} \tag{29}$$

Hence, the dual function is demonstrated as

$$D(\mu, v, \xi) = \max_{\tilde{P} > 0} L(\tilde{P}, \mu, v, \xi) \tag{30}$$

The dual function may also be indicated as

$$\min_{\mu, v, \xi} D(\mu, v, \xi). \tag{31}$$

The optimal solution for the power allocation algorithm ($\tilde{p}_{k,m,n}$) is achievable via derivatives of Eq. (30). We can also ignore the sixth constraint of the main problem when $k \neq K + 1$, because this constraint is relevant to the small cell to macro cell interference. Hence, in the condition that $k \neq K + 1$, Eq. (33) and subsequently Eq. (34) are used.

Where, $f(\tilde{p}_{k,r,n}) = \frac{B_{SC} \alpha_{k,r,n} (1 + v_k) \widetilde{SINR}_{k,r,n}}{\tilde{p}_{k,r,n}}$.

$$\begin{aligned} \frac{\partial L(\tilde{P}, \mu, v, \xi)}{\partial \tilde{p}_{k,m,n}} &= \frac{B_{SC} \alpha_{k,m,n} (1 + v_k)}{\tilde{p}_{k,m,n} \ln 2} - \sum_{r=1}^{m-1} \frac{B_{SC} \alpha_{k,r,n} (1 + v_k) \widetilde{SINR}_{k,r,n}}{\tilde{p}_{k,r,n} \ln 2} \\ &- \sum_{j=1, j \neq k}^{K+1} \sum_t^{M_j} \frac{B_{SC} \alpha_{j,t,n} (1 + v_j) \widetilde{SINR}_{j,t,n} |h_{j,k,t,n}|^2}{\tilde{p}_{j,t,n} \ln 2 |h_{j,j,t,n}|^2} - t \\ &\left(1 - \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} |h_{k,j,m,n}|^2 \right) - \mu_k - \xi |h_{k,k+1,n}|^2 = 0 \end{aligned} \tag{32}$$

On the contrary, when $k = K + 1$ we have Eq. (34).

$$\tilde{p}_{k,m,n} = \frac{B_{SC} \alpha_{k,m,n} (1 + v_k)}{\ln 2 \left[t \left(1 - \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} |h_{k,j,m,n}|^2 \right) + \mu_k + \xi |h_{k,k+1,n}|^2 \right] + \sum_{r=1}^{m-1} f(\tilde{p}_{k,r,n}) + \sum_{j=1, j \neq k}^{K+1} \sum_t^{M_j} f(\tilde{p}_{j,t,n}) \frac{|h_{j,k,t,n}|^2}{|h_{j,j,t,n}|^2}} \tag{33}$$

$$\tilde{p}_{k,m,n} = \frac{B_{SC} \alpha_{k,m,n} (1 + v_k)}{\ln 2 \left[t \left(1 - \sum_{j=1, j \neq k}^{K+1} \lambda_{j,n} |h_{k,j,m,n}|^2 \right) + \mu_k \right] + \sum_{r=1}^{m-1} f(\tilde{p}_{k,r,n}) + \sum_{j=1, j \neq k}^{K+1} \sum_t^{M_j} f(\tilde{p}_{j,t,n}) \frac{|h_{j,k,t,n}|^2}{|h_{j,j,t,n}|^2}} \tag{34}$$

After determining the resource allocation framework, the iterative sub-gradient method can be applied to update the Lagrange coefficients; during each iteration, the updated Lagrange coefficients are achievable as

$$\mu_k(i+1) = \mu_k(i) - \delta_1(i) \left(P_{k,max} - \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{p}_{k,m,n}(i) \right) \tag{35}$$

$$v_k(i+1) = v_k(i) - \delta_2(i) \left(\sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{r}_{k,m,n}(i) - R_{k,min} \right) \tag{36}$$

Algorithm 1: Carrier Allocation Algorithm

```

1: Initialize the carrier allocation set  $\mathcal{S}$  and all  $M$  units for resource allocation set  $\mathbf{P}$ ;
2: for  $k=1$  to  $K+1$  do
3: Determination of set  $Z_k(n)$  of UEs associated to carrier  $n$  and  $\overline{Z}_k$  for not be associated.
4: while  $\overline{Z}_k \neq \emptyset$  do
5: for  $m = 1$  to  $M_k$  do
6: obtain  $n^*$  meets  $|h_{k,k,m,n}|^2 \leq |h_{k,k,m,n^*}|^2$ ;
7: If  $|Z_k(n^*)|^2 < 2$  then
8: UE  $m$  is allocated to carrier  $n$ , and extracted from set  $\overline{Z}_k$ ,  $s_{k,m,n^*} = 1$ , ;
9: end if
10: If  $|Z_k(n^*)|^2 = 2$  then
11: two UEs are selected in such a way that makes  $EE_{k,n^*}$  maximum, and other UEs are rejected and are put in the set  $\overline{Z}_k$  with the Carrier allocation indicator equal to 0. The two selected UE with the same carrier allocation index, are also eliminated from the set  $\overline{Z}_k$ ; with the Carrier allocation indicator equal to 1;
13: end if
14: end for
15: end while
16: end for

```

$$\xi(i+1) = \xi(i) - \delta_3(i) \left(I_{max} - \sum_{k=1}^K \sum_{m=1}^{M_k} \sum_{n=1}^N \tilde{p}_{k,m,n}(i) |h_{k,k+1,n}|^2 \right) \tag{37}$$

In which, i represents the iteration number, and

$\delta_1(i), \delta_2(i)$, and $\delta_3(i)$ demonstrate the i 'th iteration's step-size. The step-by-step procedure of the resource allocation approach is exhibited as Algorithm 2.

Algorithm 2: Resource Allocation Algorithm

```

1: Initialize determination  $E_{max}$  as the upper bound of the number of iteration, number of iteration  $e = 0$ .
2: power efficiency factor  $t^{(0)}$ , and the tolerance index  $\varepsilon$ ;
3: while  $|\hat{R}(\mathbf{P}^{(e)}) - t^*U(\mathbf{P}^{(e)})| > \varepsilon$  or  $e < E_{max}$  do
4: Initialize  $I_{max}$ , consider the number of iterations  $i = 0$ , with the Lagrangian coefficients  $\mu_k, v_k, x_i$ ;
5: repeat
6: for  $k= 1$  to  $K+1$  do
7: for  $m= 1$  to  $M_k$  do
8: for  $n= 1$  to  $N$  do
9: a) update  $\tilde{p}_{k,m,n}$  via (33) and (34);
10: b) update  $\mu_k$  using (35)
11: c) update  $v_k$  using (36);
12: d) update  $x_i$  using (37);
13: end for
14: end for
15: end for
16:  $i = i + 1$ 
17: until convergence or  $i = i_{max}$ ;
18:  $e = e + 1, t^{(e)} = \frac{\hat{R}(\mathbf{P}^{(e-1)})}{U(\mathbf{P}^{(e-1)})}$ .
19: end while

```

4 Uplink EE based on proportional fairness

This section presents an optimal solution for the formulated optimal proportional fairness problem through fractional programming and Lagrangian dual decomposition methods. According to **P2**, it will evidently not be possible to obtain the optimal value of the utility function by solving either the UA or PC problems. In EE problems, the interaction between UA and PC should be taken into account. Hence, the final model is developed with respect to the non-convex mixed-integer framework. Also, achieving the global solution to this model is a NP-hard problem; thus, an alternative method is employed to provide a sub-optimal solution. Our proposed solution for solving the main problem is dividing **P2** to some sub-problems. In the UA sub-problem, a Lagrangian subgradient approach is adopted to create the UA matrix with respect to the constant transmission power of every user equipment. The optimized transmission power by Newton's method is then determined through the constant UA matrix for the PC sub-problem using the dual decomposition method. Subsequently, the UA(X) and transmission power (P) are continuously updated in each iteration until convergence is achieved. Given the proposed dual decomposition, the

main problem (P2) can be decomposed and rewritten as two sub-problems, i.e., UA (P2.1) and PC (P2.2):

$$P2.1 : \max_X \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \mu_{k,m} \tag{38}$$

$$P2.2 : \max_P \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \mu_{k,m} \tag{39}$$

s.t. C1.

Considering the primal–dual decomposition concept, in the analysis of P2.1, the iterative subgradient method was applied to achieve the optimal value of the logarithmic utility function for UE under the constant transmission power. Given the nature of the log function, the following equation is employed to develop the objective function:

$$\sum_{m=1}^M \sum_{k=1}^K x_{k,m} \mu_{k,m} = \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \ln(r_{k,m}) - \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \ln(P_m^{sum}) \tag{40}$$

Therefore, the Lagrangian function is defined as below:

$$L_1 = (\{x_{k,m}\}, \{P_m^{sum}\}, a) = \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \ln(r_{k,m}) - \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \ln(P_m^{sum}) + \sum_{m=1}^M a_m \left(\sum_{k=1}^K x_{k,m} \gamma_{k,m} - \gamma_{th} \right), \tag{41}$$

where $a = [a_1, a_2, \dots, a_M]$, and a_m is the positive Lagrangian coefficients. After that, the dual function and dual problem are formulated as Eqs. (42) and (43) respectively.

$$g_1(a) = \max_x L_1(\{x_{k,m}\}, \{P_m^{sum}\}, a), \tag{42}$$

$$\min_a g_1(a), \tag{43}$$

s.t. $a \geq 0$

The dual problem is then solved to obtain the best connection strategy of the BS to the m th UE:

$$k^* = \arg \max_k (\ln(r_{k,m}) + a_m(t_1) \gamma_{k,m} - \ln(P_m^{sum})). \tag{44}$$

In other words, the strategy for connecting m th user equipment to the k th base station is determined as:

$$x_{k,m} = \begin{cases} 1, & \text{if } k = k^* \\ 0 & \text{if } k \neq k^* \end{cases} \tag{45}$$

When using primal–dual decomposition method, since the dual isn't a differentiable function, it is impossible to obtain a closed-form expression based on the parameter a .

Moreover, the sub-gradient method is employed through Eq. (46) to update the Lagrangian coefficients $a = [a_1, a_2, \dots, a_M]$:

$$a_m(t_1 + 1) = a_m(t_1) - \delta_a(t_1) \left(\sum_{k=1}^K x_{k,m}(t_1) \gamma_{k,m} - \gamma_{th} \right), \tag{46}$$

$\forall m,$

where $\delta_a(t_1)$ denotes the step size of the updating process in every iteration. The optimal transmission power of every UE is obtained by solving P2.2 and creating the constant user association matrix. In P2.2, it is difficult to obtain an accurate value of p , and the following Newton's method is employed to find the sub-optimal points [35]. Moreover, the objective function is written as $f(P_m^T)$:

$$f(P_m^T) = \sum_{m=1}^M \sum_{k=1}^K x_{k,m} \ln \frac{\log_2 \left(1 + \frac{P_m^T \|h_{k,m}\|^2 d_{k,m}^{-\alpha}}{I_{k,m} + \sigma^2} \right)}{P_{CU} + \varepsilon P_m^T}. \tag{47}$$

In order to solve Problem (47), the value of SINR $\gamma_{k,m}$ for a non-objective application can be obtained from (48) according to the derivation of $f(P_m^T)$.

$$\gamma_{k,m} = \frac{P_m^T d_{k,m}^{-\alpha} \|h_{k,m}\|^2}{I_{k,m} + \sigma^2}, \tag{48}$$

where $I_{k,m}$ is defined as below:

$$I_{k,m} = P_m^T \frac{|h_{k,m}^H h_{k,m}|^2}{\|h_{k,m}\|^2} d_{k,m}^{-\alpha} + \sum_{i \neq m, \dot{m}}^{|M|} P_i^T \frac{|h_{k,m}^H h_{k,i}|^2}{\|h_{k,m}\|^2} d_{k,i}^{-\alpha}. \tag{49}$$

Accordingly, the 1th and 2th derivatives of the function f based on P_m^T d can be obtained from Eqs. (50) and (51) to determine the optimal power transmission direction strategy.

$$\frac{\partial f(a)}{\partial P_n^T} = \sum_{m=1}^M \frac{1}{r_{m,n}} \frac{\gamma_{m,n}}{\ln 2} \frac{x_{m,n}}{P_n^T} - \sum_{m=1}^M \sum_{n \neq n} \frac{x_{m,n}}{r_{m,n}} \frac{1}{\ln 2} \frac{1}{1 + \gamma_{m,n}} \gamma_{m,n}^2 \frac{\left| \frac{h_{m,n}^H h_{m,n}}{\|h_{m,n}\|} \right|^2 d_{m,n}^{-\alpha}}{(P_n^T d_{m,n}^{-\alpha} \|h_{m,n}\|)^2} - \sum_{m=1}^M \frac{x_{m,n} \xi}{P_{CU} + \xi P_n^T} \tag{50}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial^2 P_n^T} = & - \sum_{m=1}^M \frac{x_{m,n}}{(\ln 2 P_n^T)^2} \frac{\gamma_{m,n}^2 + \ln 2 \gamma_{m,n}^2 r_{m,n}}{[r_{m,n}(1 + \gamma_{m,n})]^2} \\ & - \sum_{m=1}^M \sum_{n \neq n} \frac{x_{m,i}}{\ln 2} C \frac{[2\gamma_{m,i} r_{m,i} + \gamma_{m,i}^2 r_{m,i} - (1/\ln 2)\gamma_{m,i}^2]}{[r_{m,i}(1 + \gamma_{m,i})]^2} \dot{\gamma}_{m,i} \\ & + \sum_{k=1}^K \frac{x_{k,m} \xi^2}{(PCU + \xi P_m^T)^2} \end{aligned} \tag{51}$$

In this formulation, parameters C and $\dot{\gamma}_{m,i}$ can be expressed as below:

$$\begin{aligned} C = & \left(\frac{|h_{k,m}^H h_{k,m}|^2}{\|h_{k,m}\|^2} d_{k,m}^{-\alpha} \right) / \left(P_m^T d_{k,m}^{-\alpha} \|h_{k,m}\|^2 \right) \text{ and } \gamma_{k,m} \\ = & -\gamma_{k,m}^2 \times C \end{aligned}$$

In this computational framework, a linear search method is utilized to update $P = [P_1^T, P_2^T, \dots, P_m^T]$. This method depends on ΔP_m^T and $\delta P(t)$; therefore, $P_m^T(t + 1)$ can be calculated via Eq. (52):

$$P_m^T(t + 1) = P_m^T(t) + \delta P(t) \Delta P_m^T \tag{52}$$

The back-tracking method is also employed in order to determine $\delta P(t)$. In this formulation, the step size is considered as $\Delta P_m^T = \left(\frac{\partial f}{\partial P_m^T} \right) / \left(\left| \frac{\partial^2 f}{\partial^2 P_m^T} \right| \right)$ to update the incremental Newton's step. Thus, the power control (PC) sub-problem can be solved by calculating the step size and

direction in the algorithm. After that, an effective iterative method is employed to integrate UA with PC in order to find the optimal solution for the proportional fairness power utility model. This algorithm gives a brief overview of steps in this method. Dual decomposition and Newton's method can be employed to solve $P2.1$ and $P2.2$ alternatively until the convergence.

The sub-gradient method converges to the optimal value of the dual problem $g_1(\cdot)$ by the determination of the constant power transmission vector. The first-order derivative of $g_1(\cdot)$ with respect to a_m is defined as below:

$$\frac{\partial g_1(a)}{\partial a_m} = \sum_{m=1}^M \sum_{k=1}^K x_{k,m}(a_m) \gamma_{k,m} - \gamma_{th}, \tag{53}$$

where $\sum_{m=1}^M \sum_{k=1}^K x_{k,m}(a_m)$ can be considered bounded and will be converged because $x_{k,m} \in \{0, 1\}$; therefore, Eq. (54) is obtained:

$$Sup \left\{ \left\| \frac{\partial g_1(a)}{\partial a_m} \right\| \right\} \leq s, \tag{54}$$

where S is a scalar. The convergence of this problem is proven through the above analysis based on the convergence conditions.

In this algorithm, the convergence of the UA with the constant transmission power p can be proven, whereas convergence is obvious for the power control sub-problem. Since $P2.1$ and $P2.2$ are aimed at obtaining optimal values

Table 2 The main simulation parameters

Parameter	Value
HetNet Configuration	Hexagonal network, 3-sector BSs
Small cell distribution pattern	uniform (U) and hotspot (Hs),
Operational BW (MHz)	2*20 Megahertz
Power backoff	3 dB
Max transmit power of macro-BS	43 dBm
Codec configuration	Adaptive multi-rate
Inter site spacing	250 m
Hopping method	Synthesized frequency hopping (RF)
Rx loss & Tx loss	5 dB
Propagation model	Standard propagation model
Scheduler	Fair
Maximum allowed iteration	500
L_m margin	13 dBm
Operational FREQ of BH link	6 GHz
Maximum weighting factor ω_{max}	0.95
Minimum weighting factor ω_{min}	0.55
Power consumption of macro-BS P_m^o	60 w
Power consumption of small-BS P_s^o	1.5 w
Small cell radius	150 m
Antenna Type	APE4517R0-0698X_CO-P45_03T

of the goal function $\sum_{m=1}^M \sum_{k=1}^K x_{k,m} \mu_{k,m}$, the convergence of the algorithm is guaranteed in accordance with [36].

5 Numerical results

In this section, the proposed EE algorithms are evaluated through different simulations. The simulation scenarios are implemented on a two-tier Uplink–Downlink decoupled NOMA HetNet in which the spatial distribution of macro and small-cells follow the Poisson process with various density patterns. In this regard, some of the main key performance indicators like total energy consumption, energy efficiency and network data rate are assessed under different scenarios.

5.1 Simulation scenarios

The simulation scenarios were implemented in Python so that the optimal solution can be achieved by applying CPLEX as a powerful flexible optimizer for high-performance mathematical programming. In the simulation scenarios, the macro base station was placed in the center of the macro cell, the radius of which was set to 700 m, whereas the radius of every small-cell was considered 150 m. The total network bandwidth was considered 2*20 MHz; therefore, the operational frequency was set to 2, 3.5, and 6 GHz, and the system consisted of 25 independent carriers. The small-cells were also randomly distributed in a 10 × 10 grid within the macro-cell range with a 250 m inter-site distance.

Table 2 illustrates the simulation characteristics in which all of the parameters are based on the 3GPP standard [37]. The user throughput threshold ($R_{Threshold}$) was defined considering $P^i(0) = P_p^{max}$, whereas $b^i(0) = 0$ and $\beta = 1$ were considered for ease of use. Although $R_{Threshold}$ can be relevant to the SCs' positions and user distribution, it is applied only to test the sensitivity of the proposed algorithm to the UE throughput limitations. The threshold can also be adjusted based on the circumstances.

To verify the proposed subchannel matching and power optimization algorithms in terms of EE, the simulations were done based on a HetNet layout comprising numerous small-cells, one macro base station, and 500 pieces of user equipment in addition to several backhaul links (for connecting baseband unit (BBU) and the core network) in line-of-sight communications. The length of every backhaul link ranges between 200 and 300 m. The small-cells were scattered in a 10 × 10 grid with a 250-m inter-site distance. In this model, small-cells were able to apply multiple-orthogonal carriers to avoid cross-tier interference,

although this presumption cannot be very precise considering the limited number of existing carriers in a dense environment with numerous small-cells. The frequency-reuse technique will be analyzed in future works, in which the cross-tier interference among small-cells can be calculated based on the on/off status of the cells. The users are distributed based on two distinct distribution patterns within the macro cell coverage area, 1- uniform (U) and 2-hotspot (Hs), in which 80% of the user equipment is concentrated within a 150-m radius surrounding the two random small-cells. In addition, regardless of the distribution pattern, 10 random drops are assumed for each scenario, without any severe blockage probability.

In this paper, the channel gain included line-of-sight path loss, log-normal shadowing, and fast Rayleigh fading. The propagation model is crosswave which is compatible with a dense urban profile. The Rayleigh fading model is appropriate for cellular radio propagation due to the existence of many reflection points of multipath connections. The channel gain is considered as an independent variable with unit-mean and exponential distribution. The required users' data rate ranged between 2.5 and 6 MB/s, whereas the bandwidth of every backhaul communication is 6 GHz. Other variables were initialized as: $R_{xloss} = T_{xloss} = 5dB$; $G_{TX} = G_{RX} = [V - band: 18dB, E - band: 20dB]$; $L_{margin} = 11dB$ and $NF = [V - band: 5dB, E - band: 5.5dB]$. In this scenario, the full beam width of each macro antenna is also assumed to be 60 degrees. Table 2 shows the rest of the simulation variables. The outcomes show the average results of less than 500 independent simulations. The performance indices used in these analyses include the system's EE, the total system's energy consumption, and the average user throughput. In addition, all of the assessment indexes have been analyzed under the maximum-transmission-power conditions.

The proposed approach called “Energy-Efficient hierarchical power optimization in Uplink–Downlink decoupled NOMA HetNets (NOMA-UD/EEHPO) was compared with some other energy-efficient flow control and resource optimization algorithms which we briefly introduce as the following:

Random Allocation (RA)- which is a discrete resource allocation scheme without any optimality in which the resource allocation is done merely based on the received demands [38]. Fixed Power Allocation (FPA)- in this scheme, all base stations equally share their maximum transmission power to all the associated user equipment and backhaul links and subsequently, each user equipment and backhaul link applies all the allocated resources [39]. User grouping and fractional transmit power control (UG-FTPC)- This algorithm acts upon user grouping and fractional MTP control. Based on the UG-FTPC scheme, the

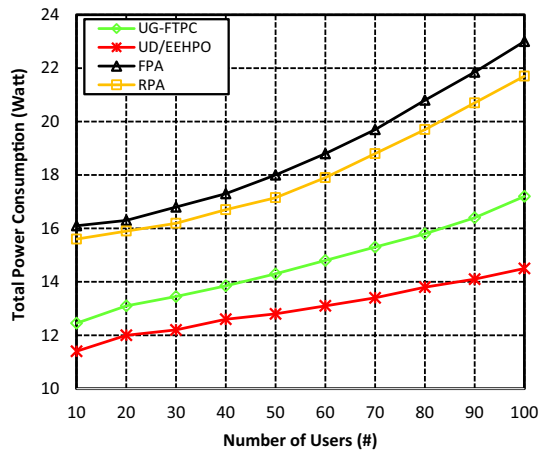


Fig. 3 Total power consumption (watt) versus number of users (#)

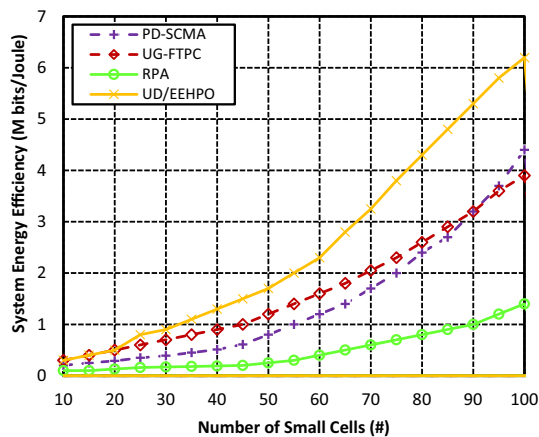


Fig. 4 System energy efficiency (M bits/joule) versus number of small-cells (#)

UEs are categorized as d_v groups in accordance with their channel gains, and each UE can only share subchannels with the UEs which are not in the same group with it. In this power allocation scheme, more power can be assigned to the UEs with poor channel quality to guarantee fairness consideration [40]. Successive Codebook Ordering Assignment (SCOA)- to better evaluate the performance of our proposed algorithms, the functionality of NOMA-UD/EEHPO was also compared to SCOA. The hybrid power domain sparse code NOMA (PD-SCMA) applies both power domain NOMA (PD-NOMA) and sparse code MA (SCMA) for an uplink hierarchical heterogeneous network. The functionality of SCOA is based on channel quality ordering criterion includes opportunistic MUE-SUE pairing (OMSP) for UE pairing (UP), and a QoS-aware resource allocation (QARA) for resource management (RM) [41].

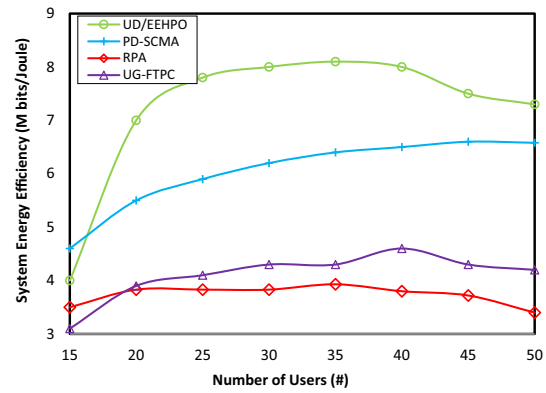


Fig. 5 System energy efficiency (M bits/Joule) versus Number of users (#)

5.2 Results and discussion

Figure 3 demonstrates the energy consumption level for different energy-efficient algorithms. In this figure, FPA has the worst possible performance, because every base station uses maximum-transmission-power for backhaul and access network. For this model, no significant saving was recorded in power optimization; it was implemented only to aim at flow control without finding an optimal solution for any EE function. Considering users' data rate constraints, UG-FTPC and NOMA-UD/EEHPO algorithms use the lowest level of energy, whereas NOMA-UD/EEHPO outperforms UG-FTPC by a small margin, it's why that UG-FTPC algorithm only minimizes the energy consumption and controls the flow; however, NOMA-UD/EEHPO algorithm also maximizes the users' data rate applying an effective carrier-matching algorithm with respect to the total resource allocation limitations.

Figure 4 illustrates the system's energy efficiency for different network configurations with 10 to 100 numbers of small-cells. The upper-bound transmission power of each macro base station was 40 dBm and 30 UEs were considered for every small-cell.

The energy efficiency of the proposed approach is compared with the User grouping and fractional power control algorithm, Random Power Allocation and the Successive Codebook Ordering Assignment algorithm simultaneously. In UG-FTPC and the Random Resource Allocation, the carriers are dedicated to UEs as random without considering the carrier status or the other channel quality conditions. So, we cannot expect acceptable energy efficiency for these two schemes. According to this figure, the proposed carrier matching power-optimized scheme (UD/EEHPO) has significantly premier energy efficiency compared to the random carrier allocation and PD-SCMA algorithms. This superiority is evident especially with the increase in the number of small-cells.

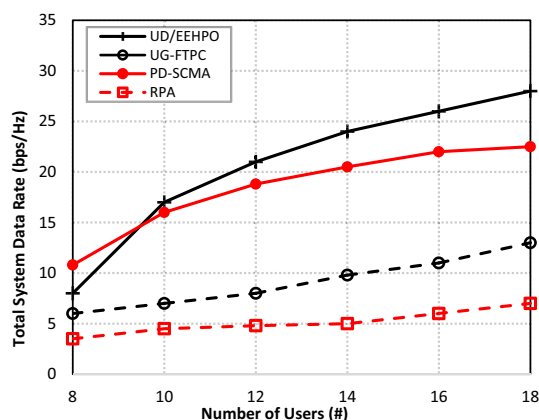


Fig. 6 Total system data rate versus number of users: with random user distribution pattern

Figure 5 demonstrates the energy efficiency performance of UG-FTPC, NOMA-UD/EEHPO and SCOA combined with OMSP and QARA algorithms considering different numbers of users. Accordingly, better energy efficiency performance was obtainable within broader constraint ranges due to an increase in the feasible range of solutions. This gives the model more freedom to not only guarantee the user throughput constraints but also simultaneously minimize the total energy consumption. As mentioned before, the UG-FTPC algorithm only minimizes the energy consumption and controls the data flow; however, SCOA and NOMA-UD/EEHPO algorithms also maximize the users' data rate to provide the required local capacity for UE demands in addition to maximizing the entire network's energy efficiency. Actually, QARA and UD/EEHPO try to minimize the network power utilization by formulating the correlation between the network's throughput and total power utilization.

If the upper bound of the demand rate is decreased, the feasible range becomes stricter, and energy efficiency is degraded. If the upper bound is decreased in a way that $y_{min} = y_{max} = C_{UE}$ (in which C_{UE} is the critical level of user demand that should be met under any circumstances), both PD-SCMA and UD/EEHPO algorithms have similar behavior and both obtain almost the identical energy efficiency level. Nonetheless, in general, UD/EEHPO with a variable user's data rate demand outperformed the PD-SCMA algorithm considering the different densities of user equipment regardless of the kind of the UE distribution

pattern. As Fig. 6 exhibits, NOMA-UD/EEHPO can stably enhance the performance of cell-edge UEs and provide a satisfactory data rate for the network regardless of the UE density as compared to the UG-FTPC, RPA and PD-SCMA. For example, with $N = 18$, for the uniform distribution pattern, the overall throughput is raised by 59% from 7.2 bps/Hz (by RPA) to 11.5 bps/Hz (by UG-FTPC) for cell-edge UEs and roughly 20% from 23 bps/Hz (by PD-SCMA) to 27 bps/Hz (by NOMA-UD/EEHPO) for cell-center UEs. For non-uniform user distribution pattern which includes ordinary mobile UEs and hotspot UEs, when $N = 18$, the total throughput is enhanced approximately 60% from 8.1 bps/Hz (by RPA + DPA) to 13 bps/Hz (by UG-FTPC) for cell-edge UEs and about 19% from 21 bps/Hz (by PD-SCMA) to 25 bps/Hz (by UD/EEHPO) for cell-center UEs.

The achieved results demonstrate the effectiveness of utilizing UD/EEHPO in increasing the capacity over the hierarchical HetNet. It is notable that we can even achieve somehow higher data rates by a combination of discrete power control and random resource allocation.

5.3 Overall effectiveness of the optimization approaches

Table 3 indicates an overall comparison among different optimization algorithms, from the sum rate, energy efficiency and service fairness viewpoints. As it is obvious, the UD/EEHPO approach provides the best possible energy efficiency among all of the investigated algorithms. Better fairness is also obtained by applying this algorithm compared to UG-FTPC and PD-SCMA. Based on the NOMA-UD/EEHPO approach, UEs with different channel quality are assigned to a subchannel and higher transmission power can be allocated to the UEs with worse channel quality. Also, for guaranteeing UE throughput demands, PD-SCMA and UD/EEHPO outperform UG-FTPC and the fixed resource allocation why that with these two schemes more than one UE can be allocated to a carrier with limited interference which improves the spectral efficiency in addition to applying an effective carrier-matching algorithm while UG-FTPC only minimizes the energy consumption and controls the flow.

Table 3 Comparison among different optimization algorithms

Optimization Algorithms	Sum Data Rate	Energy Efficiency (EE)	Fairness Index
UG-FTPC	Medium	High	Low
UD/EEHPO	High	Superlative	Medium
PD-SCMA	High	High	Low
RPA + DPA	Medium	Medium	Low

6 Conclusion

This paper proposes a dynamic optimization model which maximizes the total UL/DL energy efficiency in addition to providing the essential coverage and capacity of heterogeneous cellular networks considering the uplink fairness requirements. Based on the non-convex characteristics of the EE maximization model. The proposed model was divided to two sub-problems, UA and RA. So that, a sub-gradient approach was applied for the fair EE resource allocation and also successive convex approximation and dual composition are adopted to solve the proportional fairness problem. The cell selection process is also done according to the utmost average value of reference signal received power (RSRP) in which joint carrier allocation and power optimization criteria are considered for assigning UEs to each cell. The simulation results exhibit considerable energy efficiency improvement for various traffic models in addition to guaranteeing the fairness requirements. It also proved that the proposed approach managed to significantly improve the total network throughput. The simulation results exhibit considerable throughput increase by 30% and 22% on average for random and hotspot user distributions, respectively. It also proved that the proposed approach managed to significantly improve the total network energy efficiency by up to 35%. Furthermore, some other important aspects were introduced as the future works to develop this model. For example, online heuristics can be deployed to compute estimated solutions. However, there are still many challenges and open issues in energy-efficient 5G NOMA networks. NOMA with MIMO: Not only user scheduling and power allocation, but also beamforming can be considered in the energy efficient resource allocation. Also, the traffic pattern variations with respect to time and delay constraints can be modeled in the second step. The model is also expected to include different energy efficiency aspects for small-cells that can provide guidance on turning cells on or off when the demands' patterns change. Finally, the goal is to develop a robust model which can successfully cope with the incorrectly formulated parameters and link capacity fluctuations caused by backhaul link changes due to fast fading, atmospheric conditions, or temporal failures of links and nodes.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s11276-022-02987-x>.

References

- Yenihayat, G., & Karaşan, E. (2020). Downlink data rate, energy and spectral efficiency distribution in heterogeneous networks with cell-edge located small cells. *Wireless Networks*, 26(4), 2595–2608.
- Mishra, S., & Murthy, C. (2018). Efficient coverage management of pico cells in HetNets via spectrum slicing, cell biasing, and transmit power spreading. *Wireless Networks*, 24(8), 3099–3112.
- Kazi, B. U., & Wainer, G. A. (2019). Next generation wireless cellular networks: ultra-dense multi-tier and multi-cell cooperation perspective. *Wireless Networks*, 25(4), 2041–2064.
- Mohajer, A., Morteza, B., & Houman Z. (2016). QoS-aware coded multicast approach for wireless networks. *KSII Transactions on Internet and Information Systems (TIIS)*, 10(12), 5191–5211.
- Chai, G., Wu, W., Yang, Q., Liu, R., Qin, M., & Kwak, K. S. (2021). Energy-efficient resource allocation for multi-RAT networks under time average QoS constraint. *Wireless Networks*, 27(1), 323–338.
- Rangiseti, A. K., & Sathya, V. (2020). QoS aware and fault tolerant handovers in software defined LTE networks. *Wireless Networks*, 26(6), 4249–4267.
- Zhu, Y., Zheng, G., Wong, K.-K., Jin, S., & Lambotaran, S. (2018). Performance analysis of cache-enabled millimeter wave small cell networks. *IEEE Transactions on Vehicular Technology*, 67(7), 6695–6699.
- Ma, Z., Li, Bo., Yan, Z., & Yang, M. (2019). Remaining bandwidth based multipath routing in 5G millimeter wave self-backhauling network. *Wireless Networks*, 25(7), 3839–3855.
- Mohajer, A., Bavaghar, M., & Farrokhi, H. (2020). Mobility-aware load balancing for reliable self-organization networks: Multi-agent deep reinforcement learning. *Reliability Engineering & System Safety*, 202, 107056.
- Shuvo, M. S. A., Munna, M. A. R., Sarker, S., Adhikary, T., Razzaque, M. A., Hassan, M. M., Aloï, G., & Fortino, G. (2021). Energy-efficient scheduling of small cells in 5G: A meta-heuristic approach. *Journal of Network and Computer Applications*, 178, 102986.
- Ansere, J. A., Han, G., Liu, L., Peng, Y., & Kamal, M. (2020). Optimal resource allocation in energy-efficient Internet-of-Things networks with imperfect CSI. *IEEE Internet of Things Journal*, 7(6), 5401–5411.
- Li, H., Yang, S., Wen, Z., Cao, J., Luo, G., Wang, Z., & Wang, H. (2022). α -Fairness-based joint power allocation and power splitting for massive MIMO systems with SWIPT. *Digital Signal Processing*, 120, 103219.
- Goudos, S. K. (2019). Joint power allocation and user association in non-orthogonal multiple access networks: An evolutionary approach. *Physical Communication*, 37, 100841.
- Soleimani, B., & Sabbaghian, M. (2018). Cluster-based resource allocation and user association in mmWave femtocell networks. *IEEE Transactions on Communications*, 68(3), 1746–1759.
- Tran, H.-V., Kaddoum, G., & Truong, K. T. (2018). Resource allocation in SWIPT networks under a nonlinear energy harvesting model: Power efficiency, user fairness, and channel nonreciprocity. *IEEE Transactions on Vehicular Technology*, 67(9), 8466–8480.
- Li, X., Lan, X., Mirzaei, X., Aghdam, M. J. & I. E. E. Member. (2021) Reliability and robust resource allocation for cache-enabled HetNets: QoS-aware mobile edge computing. *Reliability Engineering & System Safety* 108272.
- Yin, W., Lei, Xu., Wang, H., Yang, Y., Wang, Y., & Chai, T. (2020). Resource management of video traffic over

- heterogeneous NOMA networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(9), 3643–3654.
18. Nikjoo, F., Mirzaei, A., & Mohajer, A. (2018). A novel approach to efficient resource allocation in NOMA heterogeneous networks: Multi-criteria green resource management. *Applied Artificial Intelligence*, 32(7–8), 583–612.
 19. Le, N.-T., Jayalath, D., & Coetzee, J. (2018). Spectral-efficient resource allocation for mixed services in OFDMA-based 5G heterogeneous networks. *Transactions on Emerging Telecommunications Technologies*, 29(1), e3267.
 20. Rahimi, P., Chrysostomou, C., Pervaiz, H., Vassiliou, V. & Ni, Q. (2021). dynamic resource allocation for SDN-based virtual Fog-RAN 5G-and-beyond networks. In *2021 IEEE Global Communications Conference (GLOBECOM)* (pp. 01–06). IEEE
 21. Wang, X., Yongjun, Xu., Wang, J., & Shuang, Fu. (2020). Joint user association and power allocation in heterogeneous NOMA networks with imperfect CSI. *IEEE Access*, 8, 47607–47618.
 22. Tang, W., Zhang, R., Liu, Y., & Feng, S. (2014). Joint resource allocation for eICIC in heterogeneous networks. In *2014 IEEE Global Communications Conference* (pp 2011–2016). IEEE
 23. Moltafet, M., Azmi, P., Mokari, N., Javan, M. R., & Mokdad, A. (2018). Optimal and fair energy efficient resource allocation for energy harvesting-enabled-PD-NOMA-based HetNets. *IEEE Transactions on Wireless Communications*, 17(3), 2054–2067.
 24. Li, B., Dai, Y., Dong, Z., Panayirci, E., Jiang, H., & Jiang, H. (2021). Energy-efficient resources allocation with millimeter-wave massive MIMO in ultra dense HetNets by SWIPT and CoMP. *IEEE Transactions on Wireless Communications*, 20(7), 4435–4451.
 25. Lai, J. Y., Wu-Hsiu, Wu., & Yu, TSu. (2020). Resource allocation and node placement in multi-hop heterogeneous integrated-access-and-backhaul networks. *IEEE Access*, 8, 122937–122958.
 26. Chaudhari, A., & Murthy, C. S. R. (2020). Efficient dynamic relay probing and concurrent backhaul link scheduling for mmWave cellular networks. *Computer Communications*, 149, 146–161.
 27. Javad-Kalbasi, M., & Valaee, S. (2021). Centralized and distributed algorithms for energy and spectrum efficient user association in small cell networks. *IEEE Transactions on Green Communications and Networking*, 5(4), 1781–1790.
 28. Mohajer, A., Sorouri, F., Mirzaei, A., Ziaeddini, A., Jalali Rad, K., & Bavaghar, M. (2022). Energy-aware hierarchical resource management and Backhaul traffic optimization in heterogeneous cellular networks. *IEEE Systems Journal*.
 29. Mohajer, A., Mazoochi, M., Niasar, F. A., Ghadikolayi, A. A. & Nabipour, M. (2013). Network coding-based QoS and security for dynamic interference-limited networks. In *International Conference on Computer Networks* (pp. 277–289). Springer: Berlin, Heidelberg.
 30. Akhtar, T., Tselios, C., & Politis, I. (2021). Radio resource management: Approaches and implementations from 4G to 5G and beyond. *Wireless Networks*, 27(1), 693–734.
 31. Temesgene, D. A., Miozzo, M., Gündüz, D., & Dini, P. (2020). Distributed deep reinforcement learning for functional split control in energy harvesting virtualized small cells. *IEEE Transactions on Sustainable Computing*, 6(4), 626–640.
 32. Bavaghar, M., Mohajer, A., & Motlagh, S. T. (2020). Energy efficient clustering algorithm for wireless sensor networks. *Journal of Information Systems and Telecommunication (JIST)*, 4(28), 238.
 33. Awad, M. K., Baidas, M. W., & Ahmad, A. (2021). Resource allocation for downlink non-orthogonal multiple access in joint transmission coordinated multi-point networks. *Computer Communications*, 173, 134–149.
 34. Wang, J., Song, X., Dong, L., Han, X. (2021). Power allocation for D2D aided cooperative NOMA system with imperfect CSI. *Wireless Networks* 1–14.
 35. Mohajer, A., Barari, M., & Zarrabi, H. (2016). Big data-based self optimization networking in multi carrier mobile networks. *Bulletin de la Société Royale des Sciences de Liège*, 85, 392–408.
 36. Mohajer, A., Yousefvand, M., Ghaleenoo, E. N., Mirzaei, P., & Zamani, A. (2014). Novel approach to sub-graph selection over coded wireless networks with QoS constraints. *IETE Journal of Research*, 60(3), 203–210.
 37. 3GPP TS 36.52–1, 2018. Evolved Universal Terrestrial Radio Access (EUTRA); User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Conformance testing. ETSI, pp. 1–130.
 38. Hsieh, C.-K., Chan, K.-L., & Chien, F.-T. (2021). Energy-efficient power allocation and user association in heterogeneous networks with deep reinforcement learning. *Applied Sciences*, 11(9), 4135.
 39. Xu, B., Chen, Y., Carrión, J. R., & Zhang, T. (2017). Resource allocation in energy-cooperation enabled two-tier NOMA HetNets toward green 5G. *IEEE Journal on Selected Areas in Communications*, 35(12), 2758–2770.
 40. Di, B., Song, L., & Li, Y. (2016). Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks. *IEEE Transactions on Wireless Communications*, 15(11), 7686–7698.
 41. Chege, S., & Walingo, T. (2021). Energy efficient resource allocation for uplink hybrid power domain sparse code nonorthogonal multiple access heterogeneous networks with statistical channel estimation. *Transactions on Emerging Telecommunications Technologies*, 32(1), e4185.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Fereshteh Atri Niasar received her B.E. degree in Electronics Engineering in 2010 and Senior degree in Telecommunication Engineering in 2015 from Islamic Azad University, Iran. Now she is Ph.D. Student in Yadegare-Imam Khomeini (RAH)-Shahre Rey Branch, Islamic Azad University, Tehran, Iran. She works in Big Tehran Power Distribution Company from 2011 up to now. Her research interests include Radio Resource Management, Heterogeneous Network, Mobility Management and Cognitive Radio Networks. As of today, she has published several papers in various journals and conferences and also she has written a Handbook.



Amir Reza Momen was born in Qom, Iran on 20 September 1976. He has received his B.Sc. degree in Electrical Engineering from Sahand University of Technology, Tabriz, Iran, and M.Sc. degree in Electrical and Telecommunication Engineering from Ferdowsi University, Mashhad, Iran and Ph.D. degree in Telecommunication Engineering from Islamic Azad University-Science & Research branch, Tehran, Iran. Since September 2003, he has been

with the Electrical and Computer Engineering Department of Islamic Azad University-Ray branch, Tehran, Iran where he is currently a member of the Communication Group. His main research interests are in wireless and vehicular sensor network, mobility modeling, and nonlinear signal processing.



Seyed Abolfazl Hosseini received his bachelor's degree in control engineering from Sharif University of Technology and his master's degree in telecommunication engineering from KN-Toosi University of Technology. He also completed his Ph.D. in Telecommunications Engineering at Tarbiat Modares University. He works as faculty member of the Islamic Azad University. His research interests include communication systems, signal processing, pattern recognition, remote sensing, and communication networks.