Computers, Materials & Continua CMC, vol.64, no.1, pp.501-514, 2020

Energy Efficient Resource Allocation Approach for Renewable Energy Powered Heterogeneous Cellular Networks

Yifei Wei[1, *], Yu Gong[1], Qiao Li[1], Mei Song[1, *] and Xiaojun Wang[2]

Abstract: In this paper, maximizing energy efficiency (EE) through radio resource allocation for renewable energy powered heterogeneous cellular networks (HetNet) with energy sharing, is investigated. Our goal is to maximize the network EE, conquer the instability of renewable energy sources and guarantee the fairness of users during allocating resources. We define the objective function as a sum weighted EE of all links in the HetNet. We formulate the resource allocation problem in terms of subcarrier assignment, power allocation and energy sharing, as a mixed combinatorial and non-convex optimization problem. We propose an energy efficient resource allocation scheme, including a centralized resource allocation algorithm for iterative subcarrier allocation and power allocation in which the power allocation problem is solved by analytically solving the Karush-Kuhn-Tucker (KKT) conditions of the problem and a water-filling problem thereafter and a low-complexity distributed resource allocation algorithm based on reinforcement learning (RL). Our numerical results show that both centralized and distributed algorithms converge with a few times of iterations. The numerical results also show that our proposed centralized and distributed resource allocation algorithms outperform the existing reference algorithms in terms of the network EE.

Keywords: Heterogeneous networks, energy harvesting, energy efficiency, resource allocation, energy sharing.

1 Introduction

After decades of intensive research on improving the energy efficiency (EE) of wireless networks driven by environmental and economic concerns, EE has become a main performance metric for the designing of current and future wireless mobile networks, e.g., 5 G NR [Buzzi, I, Klein et al. (2016)]. Many approaches have been investigated trying to improve the EE of wireless networks from three points of view: network plan, renewable energy supply and system design. First, heterogeneous cellular networks (HetNet) in which

[1] Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications, Beijing, 100867, China.

[2] School of Electronic Engineering, Dublin City University, Dublin, 9, Ireland.

[*] Corresponding Authors: Yifei Wei. Email: weiyifei@bupt.edu.cn; Mei Song. Email: songm@bupt.edu.cn.

Received: 11 February 2020; Accepted: 02 April 2020.

a macro cell is overlaid with small cells which have smaller coverage range, lower transmit power and thereafter lower energy consumption are promising network deployment for EE. Second, energy harvesting wireless networks with power transfer is the other approach to improving EE by reducing the amount of traditional energy consumption. Third, radio resource allocation in terms of spectrum assignment and power allocation can be designed with the goal of maximizing the network EE. Building on the above observation and combining these three aspects, this paper focuses on designing radio resource allocation schemes for renewable energy powered HetNet with the goal of maximizing EE.

Generally, resource allocation with the objective of maximizing the network EE of the HetNet with both intra-tier and inter-tier interference is formulated as a mixed combinatorial and non-convex optimization problem. By decomposing the original problem and leveraging on the quasi-concavity of the EE function, Tang et al. [Tang, So, Alsusa et al. (2015)] proposes a convex programming-based optimization scheme with extremely high complexity to solve the resource allocation problem. In order to reduce the complexity, zero forcing suboptimal algorithm which removes the inter-cell interference, is proposed thereafter. Xu et al. [Xu, Mao, Leng et al. (2017)] divides the resource allocation problem of maximizing EE of a ultra dense small-cell networks into two parts, i.e., sub-channel allocation and power allocation, in which the power allocation sub-problem is formulated as a two-stage Stackelberg game. Qu et al. [Qu, Wu, Wang et al. (2017)] investigates the effects of a particular type of noise in channels. Resource allocation in 5 G heterogeneous cloud radio access (H-CRAN) networks based on online learning is investigated in Alqerm et al. [Alqerm and Shihada (2018)] to mitigate inter-tier interference between macro and small cells and to maximize the network EE whilst maintaining the QoS requirements for all users in both centralized and distributed manners. Distributed resource allocation algorithm for self-organizing HetNet is proposed in Arani et al. [Arani, Mehbodniya, Omidi et al. (2017)] where each base station (BS) selects its channel and allocates power in a fully distributed manner. Channel allocation is modeled as a noncooperative game and solved by a no-regret learning algorithm. Transmit power is chosen based on an ON-OFF switching scheme.

When the HetNet is powered by renewable energy, the resource allocation problem with the goal of maximizing EE becomes more challenging and more important as well, due to the instability of the renewable energy sources. A fundamental study of resource allocation for EE optimization in CoMP-SWIPT HetNet is provided in Tang et al. [Tang, Shojaeifard, So et al. (2018)]. Joint beam forming and power allocation with intra-cell and inter-cell interference makes EE maximization problem non-convex and thereafter extremely hard to solve. To make the resource allocation problem easy to handle, it is separated into two sub-problems: beamforming design and power allocation. Linear zero-forcing beamforming is adopted to suppress the interference between users to reduce the complexity of the beamforming design sub-problem. Then the complexity of power allocation sub-problem is reduced as well and can be solved effectively. Qu et al. [Qu, Cheng, Liu et al. (2019); Qu, Li, Xu et al. (2019)] proposes the information can be embedded into the quantum carrier, and the receiver can decode the information to ensure the security of the information and provide guarantee for the effective allocation of resources. A mesh adaptive direct search algorithm is proposed in Chughtai et al. [Chughtai, Ali, Qaisar et al. (2018)] for EE maximization resource allocation in energy harvesting aided H-CRAN without the inter-cell interference.

In this paper, we investigated the resource allocation problem in renewable energy powered HetNet with intra-tier and inter-tier interference aiming at maximizing network EE. The contributions of this paper are threefold.

First, we define the objective function as the sum weighted EE to guarantee the priority of each user and constrain the power consumption of each BS with energy sharing to adapt to the uneven renewable energy supply.

Second, we develop an iterative subcarrier assignment and power allocation algorithm to solve the problem in a centralized manner.

Third, we also propose a distributed resource allocation algorithm which has a lower complexity based on reinforcement learning (RL).

The paper is organized as follows. We describe the system model and formulate the problem in Section II. The proposed centralized and distributed solutions to the formulated problem is presented in Section III. Section IV evaluates the performance of the proposed resource allocation algorithms. Section V concludes this paper.

2 System model and problem formulation

This section introduces the system scenario and energy sharing model, and formulates the HetNet EE optimization problem.

2.1 System scenario

Consider the downlink of a two-layer OFDMA HetNet in which M_m macro cells are overlaid with M_s small cells. Each macro (small) cell is serviced by a renewable energy powered macro (small) BS (MBS (SBS)). The set of MBSs and SBSs is denoted by $\mathcal{M}$ with $|\mathcal{M}| = M$where $|\cdot|$ is the cardinality of a set and $M = M_m + M_s$. All BSs are indexed from 1 to M in the set $\mathcal{M}$ with the first M_m indicating MBSs and the latter M_s SBSs. A total of K user equipment (UEs) are distributed in the HetNet with K_m UEs associating to BS m (m $\in \mathcal{M}$). The set of UEs associated with BSm is denoted by $\mathcal{K}_m$ with $|\mathcal{K}_m| = K_m$. For SBSs, open access model [Parkvall, Furuskar, Dahlman et al. (2011)] is adopted which means UEs are allowed to connect to either a SBS or a MBS. Total bandwidth B, which is divided into N subcarriers with $B_s = B/N$ the bandwidth of each subcarrier, is reused in all cells. Within one cell, in each execution period of resource allocation, each subcarrier can only be assigned to one UE and one UE can use multiple subcarriers. The set of subcarriers is denoted by $\mathcal{N}$ with $|\mathcal{N}| = M$. For subcarrier $n \in \mathcal{N}$, let k_{mn} indicates the UE chosen by BS m on subcarrier n, $\mathbf{k}^n = (k_1^n, \cdots, k_m^n, \cdots k_M^n)^T$. The subcarrier assignment is $\mathbf{k} = \{\mathbf{k}^1, \cdots, \mathbf{k}^{\mathrm{n}}, \cdots, \mathbf{k}^{\mathrm{N}}\}$.Let p_m^n indicate the transmit power of BS m on subcarrier n, and then the power allocation is $\mathbf{p} = \{\mathbf{p}^1, \cdots, \mathbf{p}^{\mathrm{n}}, \cdots, \mathbf{p}^{\mathrm{N}}\}$ where $\mathbf{p}^n = (p_1^n, \cdots, p_m^n, \cdots p_M^n)^T$. Assume perfect channel state information (CSI) is available at both the BS and UE. Indeed, FDD system can obtain CSI through feedback from UE while TDD system via uplink pilot signal. The signal to interference plus noise ratio (SINR) from BS m to UE k_m^n on subcarrier n can be expressed as

$$\gamma_{m,k_m^n}^n(\boldsymbol{p}^n) = \frac{p_m^n \left|h_{m,k_m^n}^n\right|^2}{B_S N_0 + \sum_{m' \in \mathcal{M} \not\supseteq \{m\}} p_{m'}^n \left|h_{m',k_m^n}^n\right|^2} \tag{1}$$

where N_0 is the spectral density of AWGN, and $h_{m,k_m^n}^n$ and $h_{m',k_m^n}^n$ are respectively the channel impulse response which reflects the joint effects of path loss, shadowing, and multi-path fading from BS m to UE k_m^n and from BS m′ to k_m^n on subcarrier n. The EE of the link from BS m to UE k_m^n on subcarrier n, denoted by $\eta_{m,k_m^n}^n$, can be expressed as

$$\eta_{m,k_m^n}^n = \frac{B_S log_2\left(1+\gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right)}{p_m^n + P_{C,m}^n} \tag{2}$$

where $P_{C,m}^n$ is the constant circuit power consumed by BS *m* for the transmission on subcarrier *n*. The EE of the HetNet, denoted by η_{EE}, is given by

$$\eta_{EE} = \sum_{n=1}^{N} \sum_{m=1}^{M} \omega_{m,k_m^n}^n \eta_{m,k_m^n}^n \tag{3}$$

where weight $\omega_{m,k_m^n}^n$ configures the priority of the EE of the link from BS *m* to UE k_m^n on subcarrier *n* to reduce the tendency of favoring links with better channel quality during the execution of resource allocation which is the case in global EE optimization in which the network EE is represented by the ratio of total amount of transmitted data and total amount of consumed power.

2.2 Energy sharing model

The resource allocation decision is made every execution period in a medium time scale which could be tens of seconds or minutes in order to control the signaling overhead and execution time. Each execution period, the energy budget for BSm is Pm which is determined by the renewable energy generation rate. The amount of energy transferred from BSm to BS $m' \in \mathcal{M} \not\supseteq \{m\}$ is denoted by e_m with $e_m > 0$, $e_m < 0$ and $e_m = 0$ successively meaning that BS m has sufficient energy and supplies e_m amount energy to other BSs, BSm is deficient in energy and demands e_m amount energy from other BSs, and BSm does not share energy with other BSs. In practice, energy can be transferred through wireless (wireless power transfer [Gurakan, Ozel, Yang et al. (2013)]) or wired (smart grid [Chia, Sun, Zhang et al. (2014); Tutuncuoglu and Yener (2015)]) method. The constraint of energy consumption of BSm with energy sharing can be expressed as

$$\sum_{n=1}^{N}\left(p_m^n + P_{C,m}^n\right) \le \bar{P}_m - e_m \tag{4}$$

The energy sharing policy of the HetNet is described as $\boldsymbol{e} = \{e_1, \cdots, e_m, \cdots, e_M\}$.

2.3 Problem formulation

Our goal is to maximize the HetNet EE by the energy sharing among BSs and the resource allocation of each BS in terms of subcarrier assignment and power allocation. This optimization problem can be formulated as

$$\max_{\mathrm{k,p,e}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{m=1}^{M} \omega_{m,k_m^n}^n \frac{B_S log_2\left(1+\gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right)}{p_m^n + P_{C,m}^n} \tag{5a}$$

Subject to $k_m^n \in \mathcal{K}_m, \forall n \in \mathcal{N}, m \in \mathcal{M}$ (5b)

$p_m^n \geq 0, \forall n \in \mathcal{N}, m \in \mathcal{M}$ (5c)

$\sum_{n=1}^{N}\left(p_m^n + P_{C,m}^n\right) \leq \bar{P}_m - e_m, \forall m \in \mathcal{M}$ (5d)

where Eq. (5(a)) is the problem of maximizing the weighted sum EE of the HetNet, Eq. (5(b)) constrains the subcarrier assignment that within one cell one subcarrier can be exclusively assigned to one UE, Eq. (5(c)) is the requirement on the transmit power on each subcarrier, and Eq. (5(d)) limits the energy consumption of each BS.

3 Proposed solution

In this section, we present the proposed solutions for problem Eq. (5) from centralized and distributed points of view.

3.1 Centralized approach

In this subsection, we analyze the property of problem (5) and propose a centralized resource allocation algorithm based on iterative subcarrier assignment and power allocation.

For any given feasible power allocation p, the optimal subcarrier assignment k only affects the numerator of $\eta_{m,k_m^n}^n$ in Eq. (5(a)). For any fixed p and $(m,\ n)$, based on Eq. (1) we see that $\gamma_{m,k_m^n}^n$ is totally determined by k_m^n, the chosen UE by BS m on subcarrier n, and not by the UEs other BSs assign on subcarrier n and the subcarrier assignment and power allocation of other subcarriers. Hence, for a given feasible power allocation p, the optimal subcarrier assignment can be obtained by

$$k_m^n = arg\max_{\mathrm{k_a} \in \mathcal{K}_m} \log_2\left(1 + \gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right), \forall n \in \mathcal{N}, m \in \mathcal{M} \quad (6)$$

Based on the above subcarrier assignment, Eq. (5) is reduced into a power allocation problem as follows:

$$\max_{\mathrm{p,e}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \sum_{m=1}^{M} \omega_{m,k_m^n}^n \frac{B_s log_2\left(1+\gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right)}{p_m^n + P_{C,m}^n} \quad (7a)$$

$Subject\ to\quad p_m^n \geq 0, \forall n \in \mathcal{N}, m \in \mathcal{M}$ (7b)

$\sum_{n=1}^{N}\sum_{m=1}^{M}\left(p_m^n + P_{C,m}^n\right) \leq \sum_{m=1}^{M}\bar{P}_m$ (7c)

where Eq. (7(c)) is the total power constraint. Problem Eq. (7) is a nonlinear fractional programming problem and the unconvexity of the numerator introduced by the interferences makes it extremely hard to find the optimal solution with limited complexity. An effective algorithm was proposed in Yu [Yu (2007)] for a similar nonconvex optimization problem to find at least a local optimal solution by iteratively solving the KKT conditions of the problem. Following the above problem solving methodology, first we analytically solve the KKT conditions of Eq. (7) and obtain a power allocation $\hat{\mathbf{p}}$, and then use $\hat{\mathbf{p}}$ to update subcarrier assignment Eq. (6). By iteratively solving Eqs. (7) and (6) we can achieve at least a local optimal solution for Eq. (5).

Proposition 1. The power allocation solution for Eq. (7), denoted by $\hat{\mathbf{p}}$, is the solution to the following water-filling problem:

$$\sum_{n=1}^{N}\sum_{m=1}^{M}\left[0,\frac{\omega_{m,k_m^n}^n}{\lambda X_m^n+Y_m^n+Z_m^n}-\frac{B_sN_0+\sum_{m'}p_{m'}^n\left|h_{m',k_m^n}^n\right|^2}{\left|h_{m,k_m^n}^n\right|^2}\right]^+\leq\sum_{m=1}^{M}\bar{P}_m-\sum_{n=1}^{N}\sum_{m=1}^{M}P_{C,m}^n$$

Proof. See Appendix 1

Finally, the energy sharing policy e is the solution to the following problem

$$\max_{\mathrm{e}}\ \sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{m=1}^{M}\omega_{m,k_m^n}^n\frac{B_s log_2\left(1+\gamma_{m,k_m^n}^n(\widehat{\boldsymbol{p}^n})\right)}{\widehat{p_m^n}+P_{C,m}^n} \tag{8a}$$

$$Subject\ to\ \ \sum_{n=1}^{N}\left(\widehat{p_m^n}+P_{C,m}^n\right)\leq\bar{P}_m-e_m,\forall m\in\mathcal{M} \tag{8b}$$

which can be solved effectively by linear programming techniques.

The procedure form of the proposed centralized resource allocation approach is summarized in Tab. 1. The general condition under which Tab. 1 will converge is difficult to establish yet the convergence can be observed in simulation experiments.

Table 1: The proposed centralized resource allocation algorithm

0: **repeat** (for each iteration):

0: Initialize **p**;

0: Calculate: $k_m^n=arg\ \max_{\mathrm{k_a}\in\mathcal{K}_m}\log_2\left(1+\gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right),\forall n\in\mathcal{N},m\in\mathcal{M}$;

0: Calculate:

$$\sum_{n=1}^{N}\sum_{m=1}^{M}\left[0,\ \frac{\omega_{m,k_m^n}^n}{\lambda X_m^n+Y_m^n+Z_m^n}-\frac{B_sN_0+\sum_{m'}p_{m'}^n\left|h_{m',k_m^n}^n\right|^2}{\left|h_{m,k_m^n}^n\right|^2}\right]^+$$
$$\leq\sum_{m=1}^{M}\bar{P}_m-\sum_{n=1}^{N}\sum_{m=1}^{M}P_{C,m}^n$$

and obtain $\widehat{\mathbf{p}}$;

0: **until** Convergence.

0: Achieve optimal resource allocation policy $\{k^*,p^*\}$;

0: Achieve resource sharing policy e^* by solving (8). = 0

3.2 Distributed approach

In this subsection, we propose a distributed resource allocation approach based on RL.

Unlike the centralized resource allocation approach which has a central controller to collect the network information and to execute resource allocation algorithm, in a distributed approach each BS autonomously decides its resource sharing and allocation policy based on its own optimization goal and its local observation of the network environment. In the context of our system scenario, BSs, which are RL agents learning their own resource allocation and sharing policies by interacting with the surrounding environment, constitute a multi-agent RL (MARL) system. The resource allocation decision of agent *m* (i.e., BS *m*) on subcarrier *n* in terms of the scheduled UE k_m^n and the allocated power p_m^n affects other agents' resource allocation decisions on subcarrier *n* due to the co-channel interference. Hence, the MARL system is characterized by *N* learning processes for all subcarriers. Among many RL algorithms, Q-Learning (QL) is suitable for our resource allocation task because it finds optimal decision policies without any prior knowledge of the environment which is the case in our system scenario.

For each agent *m* and subcarrier *n*, the learning process for the resource allocation policy is as follows. Let $S_{m,n}$ be the set of *S* possible states $S_{m,n} = \{s_{m,n}^1, s_{m,n}^2, \cdots, s_{m,n}^S\}$, and $A_{m,n}$ be the set of *A* possible actions $A_{m,n} = \{a_{m,n}^1, a_{m,n}^2, \cdots, a_{m,n}^A\}$. At each time step *t* agent *m*:

1) observes its state $s_{m,n}^t = s_{m,n}^s \in S_{m,n}$;

2) selects an action $a_{m,n}^t = a_{m,n}^a \in A_{m,n}$ based on $s_{m,n}^s$;

3) makes the state transition into the next state $s_{m,n}^{t+1} \in S_{m,n}$ and as a result receives an immediate reward $r_{m,n}^t$;

4) the reward $r_{m,n}^t$ is feedback to agent m and the process repeats. The goal of the learning process is to find the optimal action $\pi^*(s_{m,n}^s) \in \mathcal{A}$ for each state $s_{m,n}^s$ to maximize a long term value of the reward $r_{m,n}^t = r(s_{m,n}^s, a_{m,n}^a)$, denoted by $Q_{m,n}(s_{m,n}^s, a_{m,n}^a)$, named Q value. Over time the above repeated processes keep updating the Q value for each state-action pair (s, a), $Q(s, a)$, based on the following rule,

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \beta max_{a'}Q(s',a') - Q(s,a)] \tag{9}$$

where α is the learning rate and β is discount parameter that determines how much effect future reward has on the decisions of the current time step; s' is the next state transited from state s and a' is the action selected for s'. For more details about RL and QL, the readers are referred to Sutton et al. [Sutton and Barto (2015); Watkins and Dayan (1992)].

The state, action and reward for the learning process of agent *m* on subcarrier *n* at time step *t* are defined in detail as follows.

State: $s_{m,n}^t = \{P_m^t\}$, where $P_m^t = \{1, 0, -1\}$ indicates the state of renewable energy budget of BS *m* and is determined by the total amount of power allocated by BS *m* on all subcarriers at time step *t*:

$$P_m^t = \begin{cases} 1, \ \sum_{n=1}^N p_m^n < \bar{P}_m - \sum_{n=1}^N P_{C,m}^n \\ 0, \ \sum_{n=1}^N p_m^n = \bar{P}_m - \sum_{n=1}^N P_{C,m}^n \\ -1, \ \sum_{n=1}^N p_m^n > \bar{P}_m - \sum_{n=1}^N P_{C,m}^n \end{cases} \tag{10}$$

Action: $a_{m,n}^t = \{(k_{m,n}^t, p_{m,n}^t)\}$ where $k_{m,n}^t \in \emptyset \cup \{k_m^n\}$ is the UE scheduled by BS *m* on subcarrier *n*, which can be no UE, i.e., $\emptyset$ or any UE k_m^n associated with BS m. $p_{m,n}^t$ is the allocated power on subcarrier *n* which is chosen in a range from P_{min} to P_{max} with step Δp.

Reward:

$$r_{m,n}^{t} = \begin{cases} 0, if \ \sum_{m=1}^{M}\left(\bar{P}_m - \sum_{n=1}^{N} P_{C,m}^{n} - \sum_{n=1}^{N} p_{m,n}^{t}\right) < 0 \\ \sum_{n=1}^{N} \omega_{m,k_{m,n}^{t}}^{n} \eta_{m,k_{m,n}^{t}}^{n}, \ otherwise \end{cases} \tag{11}$$

The rationale behind the reward function is that agent *m* aims to maximize the EE of BS *m* while keeping the energy consumption under the limit of the renewable energy budget of the network because of the energy cooperation between BSs in the HetNet. The reward function indirectly combines the MARL system together with the total energy consumption limitation.

At each time step *t*, agent *m* needs to obtain the resource allocation information of other agents to calculate its immediate reward and update *Q* value based on:

$$Q\left(s_{m,n}^{t}, a_{m,n}^{t}\right) \leftarrow Q\left(s_{m,n}^{t}, a_{m,n}^{t}\right) + \alpha\left[r_{m,n}^{t}\left(s_{m,n}^{t}, \boldsymbol{a}\right) + \beta max_{t+1,a_{m,n}} Q\left(s_{m,n}^{t+1}, a_{m,n}^{t+1}\right) - Q\left(s_{m,n}^{t}, a_{m,n}^{t}\right)\right] \tag{12}$$

where a represents the actions taken by all agents at time step *t*. Hence, BSs need to exchange their learning knowledge in terms of the subcarrier assignment and power allocation policy during each step of the learning process. In practice, BSs can exchange information through intercell interface, e.g., the X2 interface. The procedure form of the QL based distributed resource allocation approach is summarized in Tab. 2. For each agent in the MARL system, the convergence can be proved by approximating other agents as environment and applying Bellman's criterion, although the proof does not hold as strictly as it does in single agent scenario it has been shown to correctly converge in many applications [Panait and Luke (2005)].

Table 2: The proposed distributed MARL algorithm

0: **Initialize:** $\boldsymbol{t = 0}$; $Q_{m,n}\left(s_{m,n}^{s}, a_{m,n}^{a}\right) = 0, \forall s_{m,n}^{s} \in S_{m,n}, a_{m,n}^{a} \in A_{m,n}; s_{m,n}^{0} = \{P_m^0\} = 1, \forall n \in \mathcal{N}, m \in \mathcal{M}$
0: For all subcarrier $n \in \mathcal{N}$, *M* agents execute the following learning process in parallel:
0: **repeat** (for each time step *t* of this episode):
0: **prepare:** Agent *m* sends its current state-action information to other *M*-1 agents and. receives the state-action information of other *M*-1 agents.
0: Agent *m* chooses action $a_{m,n}^{t} = \left\{\left(k_{m,n}^{t}, p_{m,n}^{t}\right)\right\}$ for state $s_{m,n}^{t}$ based on greedy policy;
0: Agent *m* takes action $a_{m,n}^{t}$ updates the Q value of state-action pair $\left(s_{m,n}^{t}, a_{m,n}^{t}\right)$ based on the rule (12), and transits into the next state $s_{m,n}^{t+1}$
0: **until** Convergence or Maximal number of iterations.
0: Achieve the optimal resource allocation and sharing policy by checking the Q-table and choosing the state-action pair which has the highest Q value. =0

4 Numerical results

4.1 Simulation settings

Assume one macro cell is overlaid with two small cells, where five and two uniformly-distributed UEs are served by MBS and SBS, respectively. The radius of the macro cell and small cell is set to 300 m and 50 m respectively. The path loss between BS and UE is modeled as $K\left[\frac{d_0}{d}\right]^{v}$, where K is path-loss factor, v is the path-loss exponent and d is the distance between a BS and a UE. All UEs are with identical and independent Rayleigh fading channels and Log-Normal shadowing with standard deviation of 8 dB [Tang, So, Alsusa et al. (2018)]. AWGN noise density is-174 dBm/Hz. The constant circuit power consumed by a MBS and a SBS for the transmission on subcarrier n are $25\,mW$ and $0.1\,mW$. Central frequency of HetNet is 1 GHz. The number and bandwidth of subcarriers are 5 and 15 kHz. The weight $\omega^{n}_{m,k^{t}_{m,n}}$ for each UE is chosen from [0, 1] based on the priority of the UE with a larger value of $\omega^{n}_{m,k^{t}_{m,n}}$ indicating a higher priority of the UE. Learning rate of QL is $\alpha = 0.5$ and the discount parameter is $\beta = 0.9$. Set $P_{min} = 0$ and $\Delta p = 4\,dBm$ for both MBS and SBS. Separately set $P_{max} = 44\,dBm$ and $P_{max} = 20\,dBm$ for MBS and SBS.

4.2 Simulation results

First, we verify the convergence of the proposed centralized resource allocation algorithm. The renewable energy budget of MBS 1, SBS 2 and SBS 3, and thereafter the total energy budget of the HetNet, denoted by $P-HetNet$, are respectively set (1) $\bar{P}_1 = 1W$, $\bar{P}_2 = 0.5W$, $\bar{P}_3 = 0.5W$, and $P-HetNet = 2W$; (2) $\bar{P}_1 = 2W$, $\bar{P}_2 = 0.5W$, $\bar{P}_3 = 0.5W$, and $P-HetNet = 3W$; (3) $\bar{P}_1 = 3W$, $\bar{P}_2 = 0.5W$, $\bar{P}_3 = 0.5W$, and $P-HetNet = 4W$ W;(4) $\bar{P}_1 = 3.5W$, $\bar{P}_2 = 1W$, $\bar{P}_3 = 0.5W$, and $P-HetNet = 5W$.

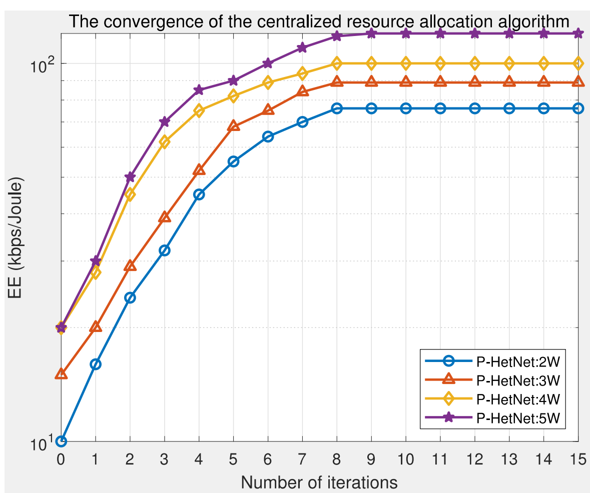

Figure 1: The convergence of the centralized resource allocation algorithm

As shown in Fig. 1, the centralized resource algorithm converges within 9 iterations for all level of energy budget of the HetNet. Furthermore, a higher EE of the HetNet can be achieved after the convergence of the resource allocation algorithm by a higher renewable energy budget of the HetNet. This is because the relatively lower amount of the renewable energy. Any increase of the transmit power can increase both the link EE and network EE when the total energy supply is limited.

We verify the convergence of the proposed distributed resource allocation algorithm in the second simulation. The renewable energy budget of MBS and SBS increase from 0.5 to 2.5 with step size 0.5. As shown in Fig. 2, the number of iterations for convergence for both MBS and SBS increase with the energy budget in exponential manner. This is because a higher energy budget leads to more valid actions in terms of power allocation which is constrained by the total energy budget. Hence, with more valid resource allocation actions, agent needs more iterations to find the optimal state-action decision policy. In addition, the number of iterations for MBS is much higher than that for SBS. It is because the MBS serves more UEs and has a larger action space produced by the subcarrier assignment of UEs.

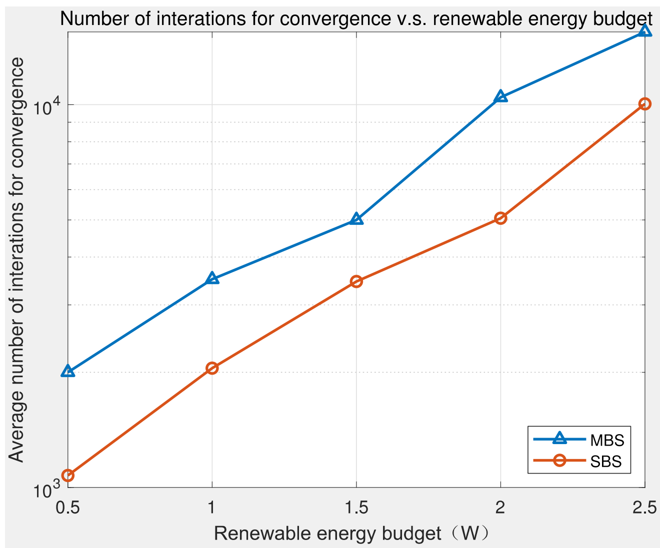

Figure 2: Number of iterations for convergence *vs.* renewable energy budget

The comparison of EE performance of the proposed centralized and distributed resource allocation algorithms and the reference resource allocation algorithms without energy sharing is illustrated in Fig. 3. Both the proposed centralized and distributed resource allocation algorithms outperform the reference resource allocation algorithms without energy sharing in terms of network EE when the energy supply is limited, i.e., when HetNet energy budget is lower than 9 W as shown in the figure. The EE performance gain is achieved by the energy sharing between those BSs with extra energy and energy deficiency. When the renewable energy supply is unstable and limited, energy sharing is promising for improving network EE. When the amount of renewable energy of the HetNet reaches a high level, i.e., 10 W or more, all resource allocation algorithms achieve the same performance of network EE. It can be explained by the definition of EE, when transmit power reached a certain level, the growth rate of the data rate starts decreasing with the increase of transmit power due to the interference between UEs. In addition, the proposed centralized resource allocation algorithm outperforms the proposed distributed resource allocation algorithms at the price of adding a central controller to execute the algorithm.

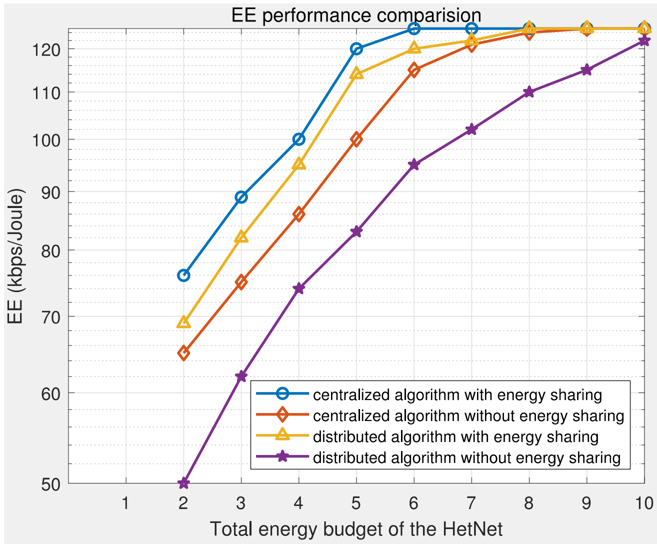

Figure 3: EE performance comparison

5 Conclusion

In this paper, we studied the resource allocation schemes aiming at maximizing network EE in renewable energy powered two-tier HetNet with energy sharing. We defined the problem as a mixed combinatorial and non-convex optimization problem and proposed both centralized and distributed algorithms to solve the resource allocation and sharing problem. The centralized algorithm consisted of three stages, i.e., subcarrier assignment, power allocation and energy sharing, and is implemented by iteratively computing each stage in order. In particular, the power allocation problem was solved by analytically solving the KKT conditions of the problem and then solving the subsequent water-filling problem. The distributed algorithm was based on the MARL with each BS acting as a QL agent to autonomously and dynamically learn its optimal resource allocation and sharing policies. Numerical results showed that both the centralized and distributed algorithms can converge with a few times of iterations. The distributed resource allocation algorithm achieved lower network EE than centralized one due to its partial knowledge of the network during making resource allocation decisions. Both centralized and distributed resource allocation algorithms achieved higher network EE than those reference algorithms that do not share energy.

Funding Statement: This work was supported by the National Natural Science Foundation of China (61871046 and 61871058).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Alqerm, I.; Shihada, B. (2018): Sophisticated online learning scheme for green resource allocation in 5 G heterogeneous cloud radio access networks. *IEEE Transactions on Mobile Computing*, vol. 17, no. 10, pp. 2423-2437.

Arani, A.; Mehbodniya, A.; Omidi, M.; Adachi, F.; Saad, W. et al. (2017): Distributed learning for energy-efficient resource management in self-organizing heterogeneous networks. *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9287-9303.

Buzzi, S.; I, C.; Klein, T.; Poor, H.; Yang, C.; Zappone, A. (2016): A survey of energy-efficient techniques for 5 G networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697-709.

Chia, Y.; Sun, S.; Zhang, R. (2014): Energy cooperation in cellular networks with renewable powered base stations. *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6996-7010.

Chughtai, N.; Ali, M.; Qaisar, S.; Imran, M.; Naeem, M. et al. (2018): Energy efficient resource allocation for energy harvesting aided H-CRAN. *IEEE Access*, vol. 6, pp. 43990-44001.

Gurakan, B.; Ozel, O.; Yang, J.; Ulukus, S. (2013): Energy cooperation in energy harvesting communications. *IEEE Transactions on Communications*, vol. 61, no. 12, pp. 4884-4898.

Panait, L.; Luke, S. (2005): Cooperative multi-agent learning: the state of the art. *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 11, pp. 383-434.

Parkvall, S.; Furuskar, A.; Dahlman, E. (2011): Evolution of LTE toward IMT-advanced. *IEEE Communications Magazine*, vol. 49, no. 2, pp. 84-91.

Qu, Z.; Cheng, Z.; Liu, W.; Wang, X. (2019): A novel quantum image steganography algorithm based on exploiting modification direction. *Multimedia Tools and Applications*, vol. 2019, no. 78, pp. 7981-8001.

Qu, Z.; Li, Z.; Xu, G.; Wu, S.; Wang, X. (2019): Quantum image steganography protocol based on quantum image expansion and grover search algorithm. *IEEE Access*, vol. 2019, no. 7, pp. 50849-50857.

Qu, Z.; Wu, S.; Wang, M.; Sun, L.; Wang, X. (2017): Effect of quantum noise on deterministic remote state preparation of an arbitrary two-particle state via various quantum entangled channels. *Quantum Information Processing*, vol. 16, no. 306, pp. 1-25.

Sutton, R.; Barto, A. (2015): Reinforcement learning: an introduction, 2nd ed. Cambridge MA, MIT press.

Tang, J.; So, D.; Alsusa, E.; Hamdi, K.; Shojaeifard, A. (2015): Resource allocation for energy efficiency optimization in heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2104-2117.

Tang, J.; Shojaeifard, A.; So, D.; Wong, K.; Zhao, N. (2018): Energy efficiency optimization for CoMP-SWIPT heterogeneous networks. *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6368-6383.

Tang, J.; So, D.; Alsusa, E.; Hamdi, K.; Shojaeifard, A. et al. (2018): Energy-efficient heterogeneous cellular networks with spectrum underlay and overlay access. *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2439-2453.

Tutuncuoglu, K.; Yener, A. (2015): Energy harvesting networks with energy cooperation: procrastinating policies. *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4525-4538.

Watkins, C.; Dayan, P. (1992): Technical note Q-learning. *Journal of Machine Learning*, vol. 8, pp. 279-292.

Xu, L.; Mao, Y.; Leng, S.; Qiao, G.; Zhao, Q. (2017): Energy-efficient resource allocation strategy in ultra dense small-cell networks: a stackelberg game approach. *IEEE International Conference on Communications*, pp. 1-6.

Yu, W. (2007): Multiuser water-filling in the presence of crosstalk. *Proceedings of Information Theory and Application Workshop*, pp. 414-420.

Appendix A.

The proof of Proposition 1 is given in this section. Since interferences only occur among co-channel UEs, (7a) can be decoupled from subcarriers. The KKT conditions of (7) are expressed as follows

$$-p_m^n \leq 0, \forall n \in \mathcal{N}, m \in \mathcal{M} \tag{13a}$$

$$\sum_{n=1}^{N}\sum_{m=1}^{M}\left(p_m^n + P_{C,m}^n\right) - \sum_{m=1}^{M}\bar{P}_m \leq 0 \tag{13b}$$

$$\lambda_m^n \geq 0, \forall n \in \mathcal{N}, m \in \mathcal{M} \tag{13c}$$

$$\lambda \geq 0 \tag{13d}$$

$$-\lambda_m^n p_m^n = 0, \forall n \in \mathcal{N}, m \in \mathcal{M} \tag{13e}$$

$$\lambda\left(\sum_{n=1}^{N}\sum_{m=1}^{M}\left(p_m^n + P_{C,m}^n\right) - \sum_{m=1}^{M}\bar{P}_m\right) = 0 \tag{13f}$$

$$-\frac{\partial\left(\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{m=1}^{M}\omega_{m,k_m^n}^n \frac{B_s log_2\left(1+\gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right)}{p_m^n + P_{C,m}^n}\right)}{\partial p_m^n} - \lambda_m^n + \lambda = 0 \tag{13g}$$

where λ_m^n and λ are the Lagrange multipliers associated with the inequality constraint (7b) and (7c). Using (13g), we can obtain

$$p_m^n = \frac{\omega_{m,k_m^n}^n}{(\lambda-\lambda_m^n)X_\mathrm{m}^\mathrm{n} + \mathrm{Y}_\mathrm{m}^\mathrm{n} + \mathrm{Z}_\mathrm{m}^\mathrm{n}} - \frac{B_s N_0 + \sum_{m\prime} p_{m\prime}^n \left|h_{m\prime,k_m^n}^n\right|^2}{\left|h_{m,k_m^n}^n\right|^2} \tag{14}$$

where

$$X_\mathrm{m}^\mathrm{n} = \frac{ln2\left(p_m^n + P_{C,m}^n\right)}{B_s}$$

$$\mathrm{Y}_\mathrm{m}^\mathrm{n} = \omega_{m,k_m^n}^n \frac{\ln 2 log_2\left(1+\gamma_{m,k_m^n}^n(\boldsymbol{p}^n)\right)}{p_m^n + P_{C,m}^n}$$

$$\mathrm{Z}_{\mathrm{m}}^{\mathrm{n}} = \sum_{m'}^{M} \left(\frac{p_m^n + P_{C,m}^n}{p_{m'}^n + P_{C,m'}^n}\right) \frac{\omega_{m',k_{m'}^n}^n \gamma_{m',k_{m'}^n}^n (\boldsymbol{p}^n) \left|h_{m,k_{m'}^n}^n\right|^2}{B_s N_0 + \sum_{i=1}^{M} p_i^n \left|h_{i,k_{m'}^n}^n\right|^2}$$

According to (13e), if $\lambda_m^n > 0, p_m^n = 0$; otherwise, $\lambda_m^n = 0, p_m^n > 0$. Hence the power allocation for problem (7) is as follows

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \left[0,\ \frac{\omega_{m,k_m^n}^n}{\lambda X_{\mathrm{m}}^{\mathrm{n}} + \mathrm{Y}_{\mathrm{m}}^{\mathrm{n}} + \mathrm{Z}_{\mathrm{m}}^{\mathrm{n}}} - \frac{B_s N_0 + \sum_{m'} p_{m'}^n \left|h_{m',k_m^n}^n\right|^2}{\left|h_{m,k_m^n}^n\right|^2}\right]^+ \le \sum_{m=1}^{M} \overline{\mathrm{P}}_{\mathrm{m}} - \sum_{n=1}^{N} \sum_{m=1}^{M} P_{C,m}^n \tag{15}$$

which is a water-filling problem that can be solved by bisection search.