electronics

Article

Joint Optimization of Energy Efficiency and User Outage Using Multi-Agent Reinforcement Learning in Ultra-Dense Small Cell Networks

Eunjin Kim [1], Bang Chul Jung [2,*], Chan Yi Park [3] and Howon Lee [1,*]

[1] School of Electronic and Electrical Engineering and IITC, Hankyong National University, Anseong 17579, Korea; gate1180@hknu.ac.kr
[2] Department of Electronic Engineering, Chungnam National University, Deajeon 34134, Korea
[3] Agency for Defense Development, Daejeon 34186, Korea; chyipark@add.re.kr
* Correspondence: bcjung@cnu.ac.kr (B.C.J.); hwlee@hknu.ac.kr (H.L.)

Abstract: With the substantial increase in spatio-temporal mobile traffic, reducing the network-level energy consumption while satisfying various quality-of-service (QoS) requirements has become one of the most important challenges facing six-generation (6G) wireless networks. We herein propose a novel multi-agent distributed Q-learning based outage-aware cell breathing (MAQ-OCB) framework to optimize energy efficiency (EE) and user outage jointly. Through extensive simulations, we demonstrate that the proposed MAQ-OCB can achieve the EE-optimal solution obtained by the exhaustive search algorithm. In addition, MAQ-OCB significantly outperforms conventional algorithms such as no transmission-power-control (No TPC), On-Off, centralized Q-learning based outage-aware cell breathing (C-OCB), and random-action algorithms.

Keywords: joint optimization; energy-efficiency; user outage; cell breathing; multi-agent distributed Q-learning; ultra-dense small cell network

Citation: Kim, E.; Jung, B.C.; Park, C.Y.; Lee, H. Joint Optimization of Energy Efficiency and User Outage Using Multi-Agent Reinforcement Learning in Ultra-Dense Small Cell Networks. *Electronics* **2022**, *11*, 599. https://doi.org/10.3390/electronics11040599

Academic Editor: Hirokazu Kobayashi

Received: 11 January 2022
Accepted: 14 February 2022
Published: 15 February 2022

1. Introduction

Recently, sixth-generation (6G) wireless networks have garnered significant attention from both industry and academia for supporting emerging novel mobile services such as high-fidelity holograms, immersive extended reality (XR), tactile internet, industry 4.0, smart home/city, and digital twins [1–3]. Various key performance indicators (KPIs) are being considered for 6G wireless networks, including peak data rate, user-experienced data rate, latency, mobility, connection density, spectral efficiency (SE), energy efficiency (EE), and reliability [4]. Among the KPIs, EE is expected to garner much more attention compared with the other KPIs because 6G wireless networks will integrate traditional terrestrial mobile networks with emerging space, aerial, and underwater networks to provide global and ubiquitous coverage [5,6].

1.1. Motivation and Related Works

Reducing the power consumption of base stations (BSs) is crucial because BSs consume approximately 80% or more of the total energy of cellular networks in general. Accordingly, various studies to maximize EE have been performed, particularly for heterogeneous ultra-dense networks. For example, the optimal frequency reuse factor that can maximize the SE or EE was investigated in ultra-dense networks [7]; it was demonstrated that the universal frequency reuse was optimal in terms of the SE for arbitrary BS/user density ratios, and that both the normalized spectral and energy efficiency (SEE) gains of the universal frequency reuse over the partial frequency reuse increased with the BS/user density ratio. In [8], a density clustering-based BS control algorithm was proposed for energy-efficient ultra-dense cellular Internet of Things (IoT) Networks, where each BS

switches to the awake/sleep mode based on user distribution to improve both the average area throughput and network EE. In addition, the authors of [9] proposed energy-efficient small cell networks using a smart on/off scheduling strategy where a certain fraction of small cell BSs (SBSs) are involved, which operate using less energy-consuming sleeping states to reduce the energy consumption.

Various machine-learning techniques have recently been applied to ultra-dense networks to improve EE. A deep Q-learning (DQL) algorithm to maximize the EE of ultra-dense networks has been proposed in [10,11]. In [10], a simple scenario involving only a single macro cell BS (MBS) and multiple femto BSs was considered to validate the performance, and the authors of [11] proposed the joint optimization of EE and throughput-adequacy of 5G heterogeneous cell. Here, the reward of DQL is designed on the basis of system-level EE. In addition, deep reinforcement learning (DRL) can be used to improve EE in heterogeneous cellular networks [12,13]. A DRL-based SBS activation strategy that activates the optimal number of SBSs to reduce cumulative energy consumption in heterogeneous cellular networks has been proposed in [12], where a deep neural network (DNN) was utilized to explicitly predict the arrival rates based on spatio–temporal correlations among SBS traffic. Moreover, in [13], three-kinds (centralized, multi-agent, and transfer learning) of Deep Q-Networks (DQNs) are utilized to improve the system-level EE of two-tier heterogeneous networks with multi-channel transmissions. Here, power control and user association were jointly optimized. The authors of [14] proposed a double DQN-based resource allocation framework to maximize the total EE by separating the selected action from the target Q-value generator in a cloud radio access network. In [15], a deep learning-based multiple-input multiple-output (MIMO) non-orthogonal multiple access (NOMA) framework was proposed to maximize both the sum rate and EE, where rapidly changing channel conditions and an extremely complex spatial structure were assumed for the MIMO–NOMA system. Furthermore, in [16], a joint optimization framework involving power ramping and preamble selection was considered to improve the EE of narrow-band Internet of Things (NB-IoT) systems, where users independently learn their own policies for preamble transmissions based on a distributed multi-agent reinforcement learning algorithm.

However, conventional techniques reported in the literature exhibit extremely high computational complexity and require repetitive and intensive computations because they are based on a DNN-based optimization framework. Therefore, we herein propose a novel multi-agent distributed Q-learning based outage-aware cell breathing (MAQ-OCB) technique to maximize EE while reducing the outage probability of users in ultra-dense small cell networks. In addition, we demonstrate the performance results of the proposed algorithm in accordance with the amount of SBS collaboration level.

1.2. Paper Organization

The remainder of this paper is organized as follows: The system model of our proposed reinforcement learning framework for jointly optimizing EE and user outage is described in Section 2. In Section 3, MAQ-OCB considering the level of SBS collaborations is presented. Via MATLAB simulations, we demonstrate the performance excellency of MAQ-OCB and present a comparison with several conventional algorithms. Finally, we present the conclusions in Section 5.

2. System Model

We herein take into account downlink ultra-dense small cell networks configured with several MBSs ($\mathbb{M}$), SBSs ($\mathbb{N}$), and users ($\mathbb{U}$). Assume that the users are randomly distributed within a cell radius, and MBSs are considered as interfering with users associated with each SBS. Users with a signal-to-interference-plus-noise ratio (SINR) less than the SINR threshold are referred to as outage users. Figure 1 shows the proposed multi-agent Q-learning framework for maximizing EE while minimizing user outage in ultra-dense small cell networks. In this framework, each agent considers its neighbor SBSs' state information, and the reward is shared with all SBSs.

Each user measures the channel quality of serving the SBS and the neighbor SBSs based on the measured reference-signal-received-power (RSRP). This is a general parameter used for deciding the user association in ultra-dense small cell networks. Further, an infinite-impulse response (IIR)-based averaging method is used so that we can reflect wireless channel dynamics caused by small-scale fading and noise [17], where reliable and stable RSRP values can be calculated. The RSRP value of user u from the SBS c can be calculated as $P_r(u,c) = \frac{P_t(c)}{d(u,c)^\rho}$. Here, $P_t(c)$ is the transmission power of SBS c, $d(u,c)$ is the distance between the user u and an SBS c, and ρ is the path loss exponent. User u is associated with an SBS that provides the highest RSRP value. Using this RSRP value, the SINR of the user u for the SBS c can be obtained as follows:

$$\gamma(u,c) = \frac{P_r(u,c)}{\sum_{i \neq c, i \in \mathbb{N}} P_r(u,i) + \sum_{j \in \mathbb{M}} P_r(u,j) + \sigma_u^2}. \quad (1)$$

Here, σ_u^2 is the thermal noise power of user u, '$\sum_{i \neq c, i \in \mathbb{N}} P_r(u,i)$' is the total amount of interference caused by all SBSs, and '$\sum_{j \in \mathbb{M}} P_r(u,j)$' is the total amount of interference caused by all MBSs. In this paper, if $\gamma(u,c) < \gamma_{th}, \forall c \in \mathbb{N}$, user u is treated as an outage user where γ_{th} is the SINR outage threshold.

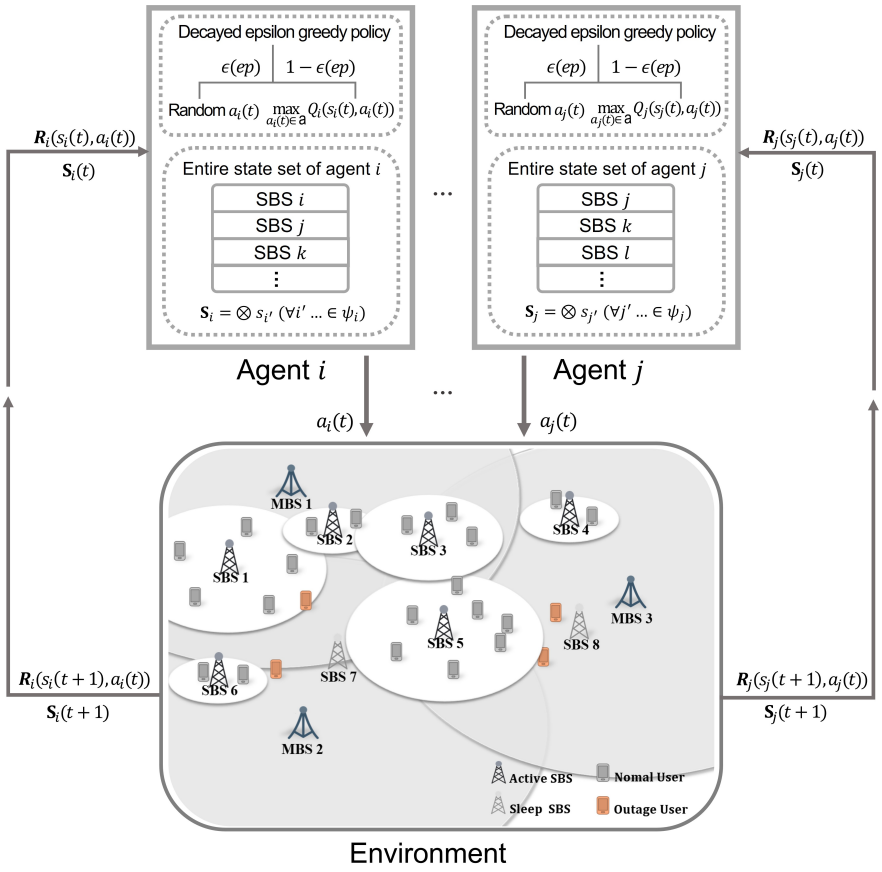


Figure 1. System model of proposed multi-agent Q-learning framework for maximizing EE while minimizing user outage in ultra-dense small cell networks.

Using Equation (1), the achievable data rate of user u included in SBS c ($\zeta(u,c)$) is expressed as

$$\zeta(u,c) = \frac{1}{|\mathbb{U}(c)|} \cdot W(c) \cdot \log_2\left(1 + \frac{P_r(u,c)}{\sum_{i \neq c, i \in \mathbb{N}} P_r(u,i) + \sum_{j \in \mathbb{M}} P_r(u,j) + \sigma_u^2}\right). \quad (2)$$

where $|\mathbb{U}(c)|$ is the number of users included in SBS c, and $W(c)$ denotes the total amount of bandwidth of SBS c. In the proposed MAQ-OCB, $W(c)$ is equally distributed to the users associated with SBS c. To calculate the EE of SBS c, we considered the following power consumption model for SBS c:

$$P_{tot}(c) = P_c(c) + \frac{1}{\delta} \cdot P_t(c). \tag{3}$$

Here, δ is the power amplifier efficiency; $P_c(c)$ and $P_t(c)$ express the amounts of fixed and transmission power consumed in SBS c, respectively. In particular, based on the SBS mode (active or sleep), the amount of power consumption may differ. Moreover, using Equations (2) and (3), the EE of SBS c ($EE(c)$) can be calculated as

$$EE(c) = \frac{\sum\limits_{u \in \mathbb{U}(c)} \left\{ \frac{1}{|\mathbb{U}(c)|} \cdot W(c) \cdot \log_2 \left(1 + \frac{P_r(u,c)}{\sum_{i \neq c, i \in \mathbb{N}} P_r(u,i) + \sum_{j \in \mathbb{M}} P_r(u,j) + \sigma_u^2} \right) \right\}}{P_{tot}(c)}. \tag{4}$$

Here, $\mathbb{U}(c)$ denotes a user set included in SBS c.

3. Joint Optimization of EE and User Outage Based on Multi-Agent Distributed Reinforcement Learning in Ultra-Dense Small Cell Networks

To obtain the Q-value, a centralized Q-learning algorithm takes into account the state information and reward of all agents. In this case, the computational complexity increases exponentially in proportion to the size of the Q-table, i.e., the learning time is significantly higher than that of the distributed Q-learning algorithm. In the proposed MAQ-OCB algorithm, agent i only considers the state information of the neighbor SBSs set ($\Im_i$), which is the union or intersection of active SBSs sets of users included in the serving SBS. With the neighbor SBSs set, the computational complexity can be reduced significantly, and the Q-value of agent i can be represented as

$$Q_i'(s_i(t), a_i(t)) = (1-\varsigma) Q_i(s_i(t), a_i(t)) + \varsigma [R_i(s_i(t+1), a_i(t)) + \eta \cdot \max_{a_i' \in \mathbf{a}} Q_i(s_i(t+1), a_i')]. \tag{5}$$

where ς is the learning rate, and η is the discount factor of the proposed Q-learning framework. $R_i(s_i(t+1), a_i(t)')$ describes the reward at the current time step t, and "$\eta \cdot \max_{a_i' \in \mathbf{a}} Q_i(s_i(t+1), a_i')$" expresses the maximum expected value of future reward. At the beginning of learning, the transmit power of each agent was set randomly, and the Q-values were also set as zero. Subsequently, each agent chooses one of the following actions: "transmission power up," "transmission power down," and "keep current transmission power". The action performed by each agent at the time step t is expressed as $\mathbf{a} = \{\Delta_{P_t}, -\Delta_{P_t}, 0\}$, and the state of each agent can be represented as $\mathbf{s}_i = \{P_{min}, P_{min} + \Delta_{P_t}, \ldots, P_{max}\}$. Furthermore, the state set based on the neighbor SBSs set can be obtained as a Cartesian product space, $\mathbf{S}_i = \otimes \mathbf{s}_{i'}, \forall i' \in \Im_i$ where $\otimes$ represents a set product. Thus, using the entire state set $\mathbf{S}_i$ and action set $\mathbf{a}$, the agent comprises its Q-table.

Assume that each SBS is as an agent in our multi-agent Q-learning framework. The proposed MAQ-OCB algorithm allows the agent to learn about the network environments using a decayed epsilon greedy policy so that the agent can explore more diverse states. The decayed epsilon greedy policy might be a good option to achieve and converge to the optimal solution by effectively adjusting the ratio between exploitation and exploration. That is, the decayed epsilon greedy policy gradually attenuates the value of ϵ considering the size of the action set ($|\mathbf{a}|$) and the decaying parameter (χ). In this policy, each agent performs a random action with the probability of $\epsilon(ep)$, and an optimal action with a probability of $1 - \epsilon(ep)$ to maximize the Q-value, i.e., $Q^* = \max_{a_t \in \mathbf{a}} Q(s_t, a_t)$ [18]. $\epsilon(ep)$ can be calculated as $\epsilon(ep) = \epsilon_{be} \times (1 - \epsilon_{be})^{\frac{ep}{\chi \times |\mathbf{a}|}}$.

3.1. MAQ-OCB with SBS Collaboration

MAQ-OCB with SBS collaboration (MAQ-OCB w/ SC) updates the Q-table using the states information of the SBSs included in its neighbor SBSs set. Briefly, by implementing the neighbor SBSs set, agent i adds SBS i' satisfying $d(i,i') \leq d_{th}$ in its neighbor SBSs set in MAQ-OCB. Here, d_{th} is the threshold value that determines whether SBS i' is included in the neighbor SBSs set. Therefore, the reward value of agent i in MAQ-OCB w/ SC is calculated as

$$R_i(s_t, a_t) = e^{-\frac{U_{\text{out}}}{|\mathbb{U}|}} \times \sum_{n \in \mathbb{N}} \left\{ \frac{\sum_{u \in \mathbb{U}(n)} \left\{ \frac{1}{|\mathbb{U}(n)|} \cdot W(n) \cdot \log_2(1 + \gamma(u,n)) \right\}}{P_{tot}(n)} \right\}. \quad (6)$$

where U_{out} and $|\mathbb{U}|$ denote the number of outage users not associated with any SBS and the total number of users, respectively. Based on Equation (6), as the number of users increases, the weighting factor decays exponentially. In addition, as the number of outage users decreases, the reward value of the proposed MAQ-OCB increases exponentially. MAQ-OCB with SBS collaboration can improve the EE of the SBSs while minimizing the number of outage users with the states and reward information of SBSs included in the neighbor SBSs set of the ultra-dense small cell networks.

3.2. MAQ-OCB without SBS Collaboration

In MAQ-OCB without SBS collaboration (MAQ-OCB w/o SC), each agent only considers its state and reward information. Therefore, it does not utilize the state and reward information of the neighbor SBSs. In the MAQ-OCB w/o SC, the reward of agent i can be represented as

$$R_i(s_t, a_t) = e^{-\frac{U_{\text{out}}(i)}{|\mathbb{U}(i)|}} \times \left\{ \frac{\sum_{u \in \mathbb{U}(i)} \left\{ \frac{1}{|\mathbb{U}(i)|} \cdot W(i) \cdot \log_2(1 + \gamma(u,i)) \right\}}{P_{tot}(i)} \right\}. \quad (7)$$

From Equation (7), each agent calculates its reward by considering only its own state information. Although this MAQ-OCB w/o SC algorithm is advantageous in terms of computational complexity, it is difficult to ensure its optimal performance compared with the MAQ-OCB w/ SC algorithm.

4. Simulation Results and Discussions

Simulation setups of the proposed algorithm were implemented in Matlab R2020a, and the training is conducted on a personal PC with a CPU i7-9750 at 2.6 GHz and a RAM of 16 GB. In this study, we considered an ultra-dense small cell network with three MBSs and a system bandwidth of 10 MHz. The detailed simulation parameters are shown in Table 1. To compare the performance of the proposed MAQ-OCB algorithm, we discuss the EE-Optimal algorithm. In addition, the no transmission-power-control (No TPC) algorithm determines the modes of SBSs on the basis of the initial user distribution, whereas a random action algorithm chooses SBSs' transmission power randomly. Moreover, it is assumed that a SBS to which no user is connected is in a sleep mode. The centralized reinforcement learning-based outage-aware cell breathing (C-OCB) algorithm considers all state and reward information of all SBSs. Moreover, "On-Off" implies that this algorithm contains only two actions, i.e., "On" (2W) and "Off" (0W).

Figure 2a,b show the simulation results with respect to the EE and reward when $|\mathbb{U}| = 20$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 4$. The users were randomly distributed within a cell radius of 400 m. As shown in Figure 2a, as the episodes increased, the EE of MAQ-OCB w/ SC converged rapidly to EE-Optimal, but the MAQ-OCB w/o SC converged slowly to low values. This result shows that determining the amount of collaboration is essential to achieve the objective of network operation and management. Because the MAQ-OCB w/ SC considers the state information of SBSs included in the neighbor SBSs set, it has a

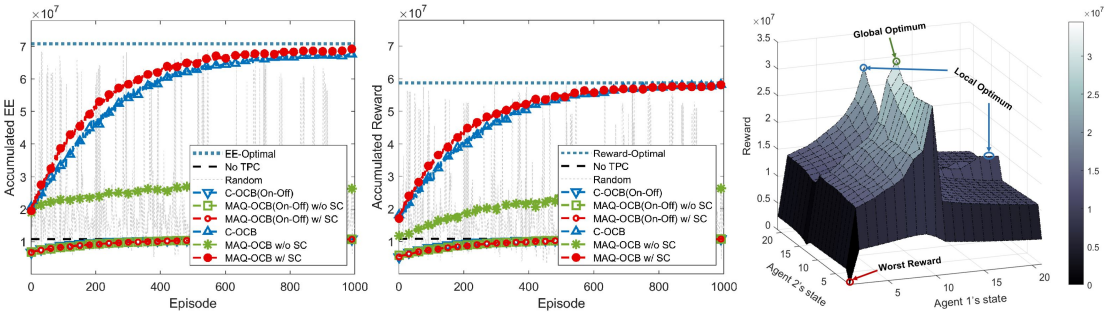
larger reward and converges faster than the MAQ-OCB w/o SC, which only considers its own state and reward information. Figure 2c shows the simulation result for the reward based on the states of agents 1 and agent 2 in the two-agent case of MAQ-OCB. Here, the global optimum implies the best reward among all states and the local optimum implies that it is less than the global optimum but higher than adjacent states. If we consider only each agent's own state, it is difficult for the agent to escape the local-optimal state. Although it acts randomly with the probability $\epsilon(ep)$, it is likely to return to the previous local optimum because it does not have sufficient actions to explore the global optimum. Because MAQ-OCB w/ SC considers the state information of neighboring SBSs, it may have relatively many opportunities to find the actions of leaving the local optimum and exploring the global optimum compared with MAQ-OCB w/o SC. As shown in Figure 2a,b, the C-OCB considers the states and reward information of all the SBSs, i.e., it has a higher computational complexity compared with MAQ-OCB w/ SC. Because this algorithm must explore more states than MAQ-OCB w/ SC, it converges relatively later. MAQ-OCB w/o SC is superior in terms of computational complexity compared to MAQ-OCB w/Sc and C-OCB, but shows low performance. On the other hand, in the corresponding scenario, C-OCB exhibits similar performance to MAQ-OCB w/SC, but computational complexity increases compared to MAQ-OCB w/SC. The computational complexity of each algorithm is described in Table 2. In the case of MAQ-OCB (On-Off) w/ SC, MAQ-OCB (On-Off) w/o SC, and C-OCB (On-Off), because each SBS involves only two actions, i.e., "On" and "Off", the rewards of these algorithms will eventually converge to that of the No TPC algorithm.

Table 1. Simulation parameters.

Parameter	Value	Parameter	Value
$\lvert\mathbb{M}\rvert$	3	$\lvert\mathbb{N}\rvert$	4, 6
$\lvert\mathbb{U}\rvert$	$20 \sim 60$	d_{th}	150 m $\sim$ 450 m
δ	0.5	W	10 MHz
σ^2	-174 dBm	P_c^a	0.25 W
P_c^s	0.025 W	ρ	3
ς	0.1	η	0.9
ϵ_{be}	0.99	χ	330
Δ_{P_t}	0.5 W	γ_{th}	0 dB

Table 2. Computational complexity analysis.

Algorithm	EE-Optimal, Reward-Optimal	C-OCB	MAQ-OCB w/o SC	MAQ-OCB w/ SC
$\mathcal{O}(\cdot)$	$\mathcal{O}(\lvert\mathbf{S}\rvert^{\lvert\mathbb{N}\rvert}\lvert\mathbf{A}\rvert^{\lvert\mathbb{N}\rvert})$	$\mathcal{O}(\lvert\mathbf{S}\rvert^{\lvert\mathbb{N}\rvert}\lvert\mathbf{A}\rvert^{\lvert\mathbb{N}\rvert})$	$\mathcal{O}(\lvert\mathbf{S}\rvert\lvert\mathbf{A}\rvert)$	$\mathcal{O}(\lvert\mathbf{S}\rvert^{\lvert\mathbb{3}\rvert}\lvert\mathbf{A}\rvert)$



(**a**) (**b**) (**c**)

Figure 2. Energy efficiency and reward vs. episode when $|\mathbb{U}| = 20$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 4$, and global optimum vs. local optimum in two-agent case. (**a**) Accumulated energy efficiency. (**b**) Accumulated reward. (**c**) Global optimum vs. local optimum.

Figures 3 and 4 show the simulation results with respect to the EE and reward when $|\mathbb{U}| = 60$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 6$. The users were randomly distributed within a cell radius of 400 m. This simulation scenario is extremely complicated compared with the scenario presented in Figure 2a,b. Hence, the convergence of the C-OCB occurs late and its speed is extremely low, i.e., a large number of episodes and iterations are required to obtain the global-optimal solution. Additionally, as mentioned before, MAQ-OCB w/ SC considers the state information of neighboring SBSs so that it has the greatest reward compared to No TPC, random action, C-OCB(On-Off), MAQ-OCB(On-Off), C-OCB, and MAQ-OCB w/o SC. Similar to Figure 3, the EE of MAQ-OCB w/ SC converged rapidly to EE-Optimal, but the MAQ-OCB w/o SC and C-OCB converged slowly to low values. After convergence, No TPC, C-OCB(On-Off), MAQ-OCB(On-Off) w/o SC, MAQ-OCB(On-Off) w/ SC, C-OCB, MAQ-OCB w/o SC, and MAQ-OCB w/ SC achieve 34.50%, 34.42%, 34.36%, 34.33%, 86.94%, 90.60%, and 93.23% in terms of EE, compared with EE-Optimal, respectively. In addition, Figure 5 shows the number of outage users based on the episode when $|\mathbb{U}| = 60$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 6$. The proposed MAQ-OCB technique converges to the Reward-Optimal, and the convergence speed of the C-OCB is relatively slower than that of the proposed algorithm because of its high computational complexity. Furthermore, after convergence, the numbers of outage users of EE-Optimal, Reward-Optimal, No TPC, C-OCB(On-Off), MAQ-OCB(On-Off) w/o SC, MAQ-OCB(On-Off) w/ SC, C-OCB, MAQ-OCB w/o SC, and MAQ-OCB w/ SC are 7, 1, 2, 2.07, 2.33, 2.25, 3.95, 2.24, and 1.61, respectively. That is, except Reward-Optimal, MAQ-OCB w/ SC has the smallest number of outage users among these algorithms.

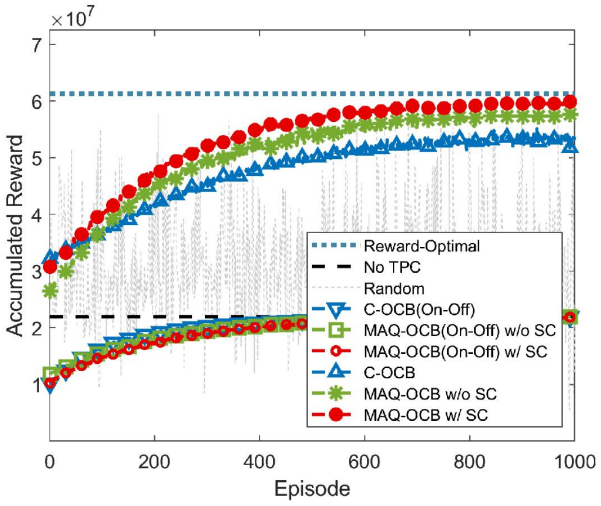


Figure 3. Reward vs. episode when $|\mathbb{U}| = 60$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 6$.

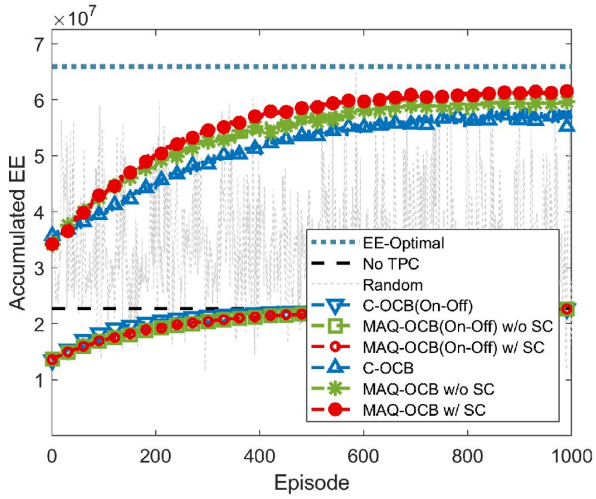


Figure 4. Energy efficiency vs. episode when $|\mathbb{U}| = 60$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 6$.

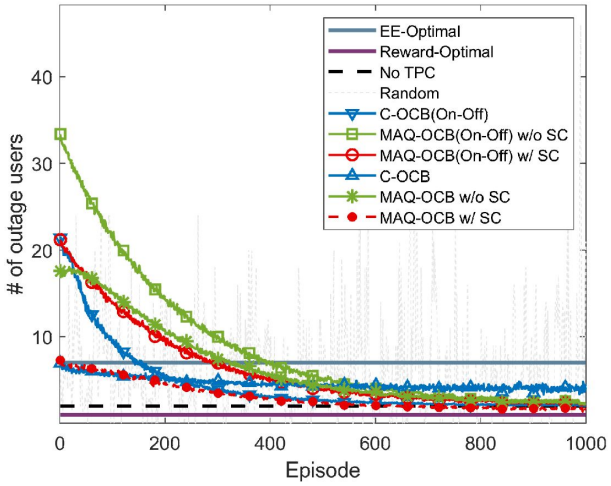


Figure 5. Number of outage users vs. episode when $|\mathbb{U}| = 60$, $|\mathbb{M}| = 3$, and $|\mathbb{N}| = 6$.

5. Conclusions

To maximize the network-wide EE while minimizing outage users in ultra-dense small cell networks, we proposed the MAQ-OCB algorithm based on a multi-agent reinforcement learning framework. To analyze the system performance based on the level of SBS collaborations, we introduced two outage-aware cell breathing algorithms in ultra-dense small cell networks, i.e., MAQ-OCB w/ SC, which considers the state information of SBSs included in its neighbor SBSs set, and MAQ-OCB w/o SC, which updates its Q-table considering only its own state and reward information. Through intensive simulations, we demonstrated that MAQ-OCB can achieve the EE-optimal solution and outperformed the conventional algorithms in terms of EE and the number of outage users.

Author Contributions: Conceptualization, H.L. and C.Y.P.; investigation, E.K., C.Y.P., B.C.J., and H.L.; methodology, B.C.J. and H.L.; supervision, B.C.J. and H.L.; writing—original draft, E.K. and H.L; writing—review and editing, E.K., B.C.J., C.Y.P., and H.L. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This research was supported by the Agency for Defense Development, Rep. of Korea.

Conflicts of Interest: We have no conflicts of interest to declare.

Abbreviations

6G	Sixth-generation
BS	Base station
C-OCB	Centralized Q-learning based outage-aware cell breathing
DNN	Deep neural network
DQL	Deep Q-learning
DQN	Deep Q-network
EE	Energy efficiency
IIR	Infinite impulse response
IoT	Internet of Things
KPI	Key performance indicator
MAQ-OCB	Multi-agent Q-learning based outage-aware cell breathing
MAQ-OCB w/ SC	MAQ-OCB with SBS collaboration
MAQ-OCB w/o SC	MAQ-OCB without SBS collaboration
MBS	Macro cell BS
MIMO	Multiple-input multiple-output
NB-IoT	Narrow-band Internet of Things
NOMA	Non-orthogonal multiple access
No TPC	No transmission power control
RSRP	Reference signal received power
SBS	Small cell BS
SE	Spectral efficiency
SEE	Spectral and energy efficiency
SINR	Signal-to-interference-plus-noise ratio
XR	Extended reality

References

1. Samsung Research. *6G: The Next Hyper-Connected Experience for All*; White Paper; July. 2020.
2. Tariq, F.; Khandaker, M.R.A.; Wong, K.-K.; Imran, M.A.; Bennis, M.; Debbah, M. A Speculative Study on 6G. *IEEE Wirel. Commun.* **2020**, *27*, 118–125.
3. Yu, H.; Lee, H.; Jeon, H. What is 5G? Emerging 5G Mobile Services and Network Requirements. *Sustainability* **2017**, *9*, 1–22.
4. Slalmi, A.; Chaibi, H.; Chehri, A.; Saadane, R.; Jeon, G. Toward 6G: Understanding Network Requirements and Key Performance Indicators. *Wiley Trans. Emerg. Telecommun. Technol.* **2020**, *32*, e4201.
5. Huang, T.; Yang, W.; Wu, J.; Ma, J.; Zhang, X.; Zhang, D. A Survey on Green 6G Network: Architecture and Technologies. *IEEE Access* **2019**, *7*, 175758–175768.
6. Lee, S.; Yu, H.; Lee, H. Multi-Agent Q-Learning Based Multi-UAV Wireless Networks for Maximizing Energy Efficiency: Deployment and Power Control Strategy Design. *IEEE Internet Things J.* **2021**, early access, 2020. https://doi.org/10.1109/JIOT.2021.3113128(Access on 10 January 2022).
7. Su, L.; Yang, C.; Chih-Lin, I. Energy Spectr. Effic. Freq. Reuse Ultra Dense Networks. *IEEE Tran. Wirel. Commun.* **2016**, *15*, 5384–5398.
8. Lee, W.; Jung, B.C.; Lee, H. DeCoNet: Density Clust.-Based Base Stn. Control Energy-Effic. Cell. IoT Networks. *IEEE Access* **2020**, *8*, 120881–120891.
9. Çelebi, H.; Yapıcı, Y.; Güvenç, İ.; Schulzrinne, H. Load-Based On/Off Scheduling for Energy-Efficient Delay-Tolerant 5G Networks. *IEEE Tran. Green Commun. Netw.* **2019**, *3*, 955–970.
10. Shi, D.; Tian, F.; Wu, S. Energy Efficiency Optimization in Heterogeneous Networks Based on Deep Reinforcement Learning. *IEEE Inter. Conf. Commun. Work.* **2020**, 1–6. https://sci-hub.se/10.1109/ICCWorkshops49005.2020.9145404(Accessed on 10 January 2022).
11. Spantideas, S.T.; Giannopoulos, A.E.; Kapsalis, N.C.; Kalafatelis, A.; Capsalis, C.N.; Trakadas, P. Joint Energy-efficient and Throughput-sufficient Transmissions in 5G Cells with Deep Q-Learning. In Proceedings of the 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), Athens, Greece, 7–10 September 2021; pp. 265–270.
12. Ye, J.; Zhang, Y.-J.A. DRAG: Deep Reinforcement Learning Based Base Station Activation in Heterogeneous Networks. *IEEE Tran. Mob. Comp.* **2020**, *19*, 2076–2087.
13. Giannopoulos, A.; Spantideas, S.; Kapsalis, N.; Karkazis, P.; Trakadas, P. Deep Reinforcement Learning for Energy-Efficient Multi-Channel Transmissions in 5G Cognitive HetNets: Centralized, Decentralized and Transfer Learning Based Solutions. *IEEE Access* **2021**, *9*, 129358–129374.
14. Iqbal, A.; Tham, M.-L.; Chang, Y.C. Double Deep Q-Network-Based Energy-Efficient Resource Allocation in Cloud Radio Access Network. *IEEE Access* **2021**, *9*, 20440–20449.

15. Huang, H.; Yang, Y.; Ding, Z.; Wang, H.; Sari, H.; Adachi, F. Deep Learning-Based Sum Data Rate and Energy Efficiency Optimization for MIMO-NOMA Systems. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 5373–5388.
16. Guo, Y.; Xiang, M. Multi-Agent Reinforcement Learning Based Energy Efficiency Optimization in NB-IoT Networks. In Proceedings of the 2019 IEEE Globecom Workshops (GC Wkshps), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
17. Tesema, F.B.; Awada, A.; Viering, I.; Simsek, M.; Fettweis, G.P. Evaluation of Adaptive Active Set Management for Multi-Connectivity in Intra-Frequency 5G Networks. *IEEE Wirel. Commun. Netw. Conf.* **2016**, 1–6. https://sci-hub.se/10.1109/WCNC.2016.7564823(10 January 2022).
18. Srinivasan, M.; Kotagi, V.J.; Murthy, C.S.R. A Q-Learning Framework for User QoE Enhanced Self-Organizing Spectrally Efficient Network Using a Novel Inter-Operator Proximal Spectrum Sharing. *IEEE Sel. Areas Commun.* **2016**, *34*, 2887–2901.