Deep-Learning-Based Resource Allocation for 6G NOMA-Assisted Backscatter Communications

Van Dat Tuong and Sungrae Cho

***Abstract*—The proliferation of Internet-of-Things applications has given rise to several challenges, including network congestion and high energy consumption. Among the promising technologies for beyond-5G networks, nonorthogonal multiple access (NOMA) and ambient backscatter communications stand out. These technologies enhance wireless access capacity and enable energy-efficient data sharing. In this study, we propose a novel energy-efficient resource allocation scheme for 6G NOMA-assisted backscatter communication networks. Our network model comprises a central reader (RD) and distributed backscatter devices (BDs) that harvest energy from incident signals to modulate useful data and reflect it toward the RD. To maximize energy efficiency, we formulated a joint optimization problem of channel resource allocation and BDs' reflection coefficients. However, solving this problem is challenging because of its nonconvexity and system dynamics. To address this issue, we developed a novel deep-learning-based algorithm that leverages the advantages of deep reinforcement learning. During training, we estimated the state components without relying on exact channel state information (CSI), which is computationally expensive. This estimation reduces communication overhead raised in collecting CSI data. Extensive simulations were conducted to demonstrate the superiority of the proposed scheme. Simulation results show that the proposed scheme notably enhances energy efficiency compared to existing benchmarks. Specifically, improvements of approximately 30.3%, 41.7%, 6.0%, and 4.4% were observed when compared to the greedy approach, random approach, Deep Q-Network, and successive convex approximation approach, respectively.**

***Index Terms*—Backscatter communications, deep reinforcement learning, energy-efficient communications, nonorthogonal multiple access (NOMA).**

I. Introduction

THE proliferation of IoT devices has brought forth numerous challenges for wireless networks, including connectivity congestion, high energy consumption, and throughput degradation. Nonorthogonal multiple access (NOMA) technologies have emerged as promising solutions, capable of enhancing wireless access capacity and data rate for beyond-5G networks [1]. NOMA facilitates access for a large number of devices by enabling multiple devices to share the same resource. A comprehensive review of recent NOMA advancements for the IoT was conducted that focused on grant-free connectivity [2]. Moreover, the superiority of NOMA over the conventional orthogonal multiple access (OMA) technique in reducing energy and delay costs was highlighted [3].

Several recent studies have focused on identifying effective solutions for improving energy efficiency (EE) toward sixth-generation (6G) networks [4]. Backscatter communication (BackCom) has emerged as a promising technology for enabling low-power communication in the Internet-of-Things (IoT) systems [5]. However, backscatter devices (BDs) pose certain challenges because of their lack of active radio components. BDs rely on modulating and reflecting incident radio frequency (RF) signals to facilitate network transmissions, in which signal reflection is implemented dynamically by varying antenna impedance. Notably, BDs can harness energy from RF signals to power their own circuits in wireless powered networks [6], [7] and Internet-of-Vehicles networks [8].

To meet the stringent requirements of massive IoT networks, such as ultra-low energy consumption and extremely high spectral efficiency, researchers have investigated the combination of BackCom and NOMA, which offers dual benefits to conventional systems [9]–[11]. Compared to the OMA technique, NOMA proves superior BackCom networks. Guo *et al.* [9] demonstrated that NOMA significantly improves the successful decoding rate performance. Yang *et al.* [10] integrated NOMA with dynamic time division multiple access to maximize the minimum throughput of BDs by jointly optimizing backscattering time and reflection rates. Xu *et al.* [11] extended the work of Ye *et al.* [6] and Yang *et al.* [7] to address the EE maximization problem in a NOMA-assisted BackCom network. To enhance the quality of service (QoS), several constraints were adopted, such as minimum signal-to-interference plus noise ratio (SINR), maximum transmit power, and NOMA decoding order.

Given the dynamic nature of wireless network environments with time-varying channels, conventional optimization approaches of the above studies often involve repeated computations to obtain the optimal solution, resulting in high computational overhead. This problem can be addressed using machine learning techniques that learn the optimal policy through training with the observed data input. Subsequently, the obtained policy enables optimal decision-making for each encountered state. Reinforcement learning (RL), which is among the advanced machine learning algorithms, has demonstrated its potential for BackCom networks in various optimization schemes. Examples include total throughput maximization using the Double Deep Q-Network (DDQN) algorithm [12], total UAV flight time minimization using the Multi-agent Deep Option Learning algorithm [13], and interference reduction

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information Communications) and in part by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (RS-2023-00209125). *(Corresponding author: Sungrae Cho.)*

Van Dat Tuong and Sungrae Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea. *(email: vdtuong@uclab.re.kr, srcho@cau.ac.kr)*

using a Q-learning model [14]. Therefore, it is necessary to investigate an efficient solution to enhance EE for 6G NOMA-assisted BackCom networks based on RL approach.

A. Literature Review

Several researchers have explored the performance of ambient BackCom in NOMA networks. Zhang *et al.* [15] proposed a downlink NOMA-assisted BackCom system and investigated the corresponding outage probabilities and ergodic rates. The system model involved a BD that transmits information to a cellular user using base station (BS) signals. Zeb *et al.* [16] explored the enhancement of outage probability and throughput in a wireless-powered BackCom network by incorporating a hybrid channel access mode with time division multiplexing access and NOMA. Ding *et al.* [17] emphasized the advantages of using NOMA as a multiple access technique in BackCom networks to improve throughput and connectivity over other OMA techniques. Farajzadeh *et al.* [18] integrated uplink NOMA and ambient backscatter technologies into aerial networks, enhancing successive decoded bit rates while minimizing flight time by optimizing the unmanned aerial vehicle (UAV) altitude. The provided numerical results indicated the potential for computing the optimal altitude of UAVs and revealed an intimate relationship between backscattering reflection coefficients (RCs) and network performance in terms of throughput. Chen *et al.* [19] investigated the expected rates, outage probability, and diversity-multiplexing trade-off in a cooperative NOMA-assisted BackCom network. Nazar *et al.* [20] derived closed-form expressions for the bit error rate in a NOMA-assisted BackCom system to evaluate the RCs needed to achieve the most favorable data rates. Li *et al.* [21] derived closed-form expressions for outage and intercept probabilities to analyze the secrecy of NOMA-assisted BackCom systems. Similar to the study of Zhang *et al.* [15], Raza *et al.* [22] proposed a massive machine-type communication framework based on a NOMA-assisted BackCom system and investigated the corresponding outage probabilities and ergodic rates. Li *et al.* [23] investigated the reliability and security of maritime transmission systems and derived analytical expressions for outage and intercept probabilities in a NOMA-assisted BackCom Internet-of-Vehicles network. The author of [24] proposed NOMA-assisted BackCom and wireless power transfer (WPT)-assisted NOMA systems to enhance energy and spectral efficiency. Li *et al.* [25] considered malicious eavesdroppers and analyzed the impact of channel estimation error, imperfect successive interference cancellation (SIC), and residual hardware impairments on the security and reliability of a NOMA-assisted BackCom system. By deriving analytical expressions for outage and intercept probabilities, insightful asymptotic analyses were conducted for the high signal-to-noise ratio and high main-to-eavesdropper ratio models.

In addition, within the realm of NOMA-assisted BackCom networks, researchers have explored diverse resource allocation and optimization frameworks. Notably, a study by [8] investigated a joint optimization framework that encompasses both cellular device association and power allocation, with the objective of maximizing the achievable EE while adhering to QoS constraints such as successive signal decoding and minimum rate requirements. Moreover, Xu *et al.* [11] proposed a joint optimization framework for maximizing EE in NOMA-assisted BackCom networks. This framework optimized the allocation of transmit power for the BS and the design of RCs for the BDs. The employed optimization relied on an iterative algorithm based on Dinkelbach's method and quadratic transformation. Liao *et al.* [26] jointly optimized the RC and power and time allocations to maximize the minimum user throughput for a full duplex NOMA-assisted BackCom network. The authors developed an iterative algorithm using block coordinated descent and successive convex optimization to address the formulated non-convex problem. Khan *et al.* [27] proposed a novel analysis for NOMA-assisted BackCom Vehicle-to-Everything networks that maximized the minimum achievable rate of all vehicles by jointly optimizing the transmit power allocation of BS and roadside units. Convex transformation was used to solve the formulated problem. Ding *et al.* [28] studied the application of NOMA-assisted BackCom to 6G ultra-massive machine-type communications and introduced an optimization framework for improving the uplink sum rate while mitigating the interference between downlink and uplink transmissions. A nonconvex optimization problem of the downlink transmit power and RCs of the BDs was formulated, which required a linear programming transformation to be solved. Ahmed *et al.* [29] investigated the EE maximization problem in a multi-cell NOMA-assisted BackCom IoT network that jointly optimized the total transmit power, power allocation, and RCs of the BDs. Applying Dinkelbach's method, the problem was transformed and decoupled into two subproblems of RC selection and power allocation, which were iteratively solved using Karush-Kuhn-Tucker conditions and the Lagrangian dual method.

B. Motivations and Contributions

State-of-the-art BackCom studies focus on improving energy efficiency, communication range, and secure and reliable communications for IoT systems. However, addressing the exponentially growing number of IoT devices necessitates a significant increase in access capacity, which requires advanced multiple access techniques. Therefore, this study is motivated to explore the effectiveness of integrating NOMA with BackCom for 6G IoT networks. Specifically, the study focuses on the energy problem, aiming to maximize EE by jointly optimizing the subcarrier allocations (SAs) and RCs of the BDs. The key contributions of this study can be summarized as follows:

- Model of a NOMA-assisted BackCom IoT system, which comprises a central reader and multiple distributed BDs as IoT nodes. The central reader sends probe signal and observes backscattering data from the BDs. The system allows multiple arbitrary BDs to be paired by NOMA to concurrently backscatter data using the same resource block. Each BD is assumed to be equipped with a battery to save harvested energy for its own operation, enhancing system realism.
- Formulation of a joint optimization problem of SAs and RCs to maximize EE. Owing to the deeply coupled

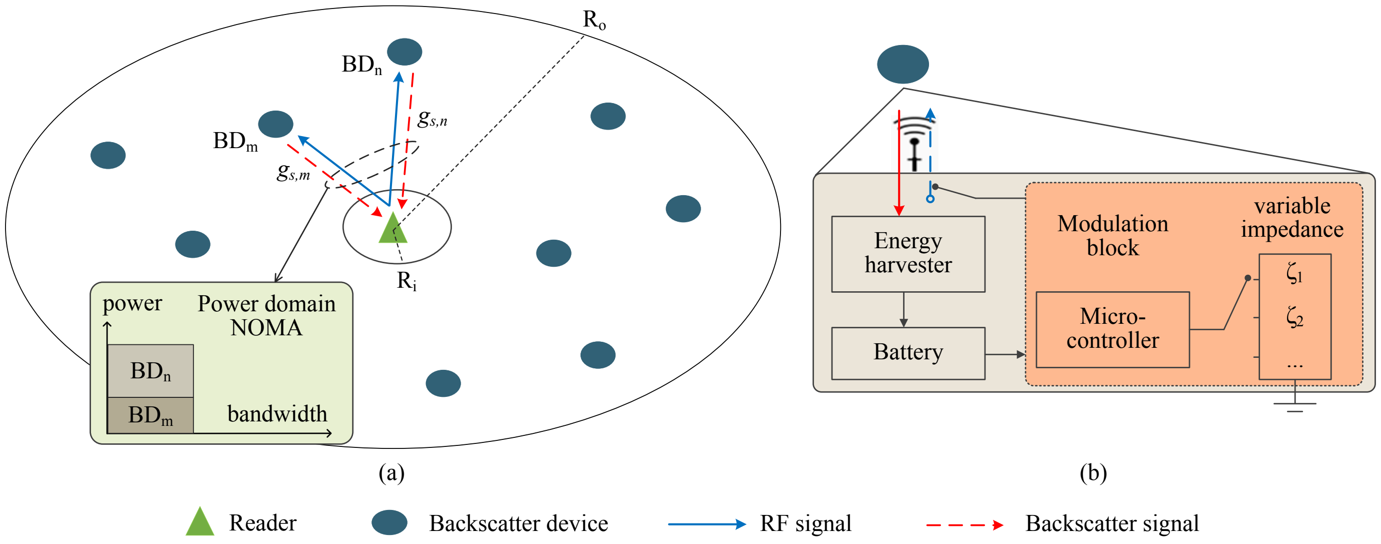

Fig. 1. Illustration of a NOMA-assisted BackCom network: (a) NOMA-paired backscatter devices simultaneously reflect signals, (b) Architecture of a backscatter device.

variables, the formulated problem is non-convex and challenging to solve directly. The study employs deep reinforcement learning (DRL), a promising tool in wireless communication problems, surpassing the complexity of iterative algorithms and successive convex approximation (SCA) techniques proposed in previous literature [11], [26]–[29].

- Development of a channel state information (CSI)-estimation technique to handle unknown CSI of sleep-state BDs. This technique computes the CSI based on historical data. Subsequently, a DRL framework is developed in which the estimated CSI and BD battery levels are incorporated into the system state. This inclusion notably minimizes communication overhead raised in collecting CSI data. To mitigate the impact of the CSI estimation error on the training results, the proposed DRL framework deducts the error of the estimated channel gain from the achievable EE in computing the step reward.
- Implementation of the proposed DRL framework based on the double deep Q-network (DDQN) algorithm, which is suitable for the discretized SA and RC spaces. Furthermore, an adaptive genetic algorithm (AGA) [30] is adopted in the action exploration process, which reduces training time and improves the action output.
- Conducting extensive simulations to demonstrate the superiority of the proposed scheme. Compared to greedy, random, DQN, and SCA approaches, the proposed algorithm improves the EE by approximately 30.3%, 41.7%, 6.0%, and 4.4%, respectively. Furthermore, the achievable EE of the proposed scheme closely resembles that of the optimum scheme based on the exhaustive search algorithm, with a difference of only 0.6%.

The remainder of this paper is organized as follows. Section II describes the system model and EE maximization problem formulation. Section III presents the proposed algorithm based on DRL. The performance evaluation is discussed in section IV. Section V presents the concluding remarks.

II. System Model and Problem Formulation

A. Network Model

As illustrated in Fig. 1-(a), we consider a NOMA-assisted BackCom network comprising a central reader (RD) and K single-antenna BDs. The BDs are uniformly and independently distributed within a network coverage area with inner and outer radii of, R_i and R_o, respectively. Let $\mathcal{K} = \{1, \ldots, K\}$ denote the set of BDs. Each BD represents an IoT device consisting of a receiver, transmitter, energy harvester, and micro controller, as shown in Fig. 1-(b). These BDs can harvest energy from the incident RF signals to power their circuit and reflect modulated signals that carry information to the RD. In addition, each BD is equipped with a battery to store the harvested energy ensuring sustained circuit operation in the long term. Furthermore, each BD controls its RC by varying the impedance. Let D be the number of impedance values, deriving D levels for RC selection.

The network model adopts a block fading model with flat fading. The channel gain between the RD and BD k over channel s can be defined as $g_{s,k} = |h_{s,k}|^2 r_k^{-\alpha}$, where α is the path-loss exponent, $h_{s,k}$ is the small-scale fading component, and r_k is the distance between the RD and BD k. Without loss of generality, the reciprocal RF and backscatter links are assumed to have the same channel condition owing to the proximity between the RD and BDs. The received signal at the RD is considered a double-path signal, which includes the RF path (RD→BD) and the backscatter path (BD→RD). The noise at the BD of the former path is negligible because each BD consists only of passive components [31]. Therefore, the signal received by the RD from BD k via channel s can be determined as follows:

$$\begin{aligned} y_{s,k} &= (g_{s,k}x_s)(\xi_k p_s g_{s,k} z_k) + N_s \\ &= \xi_k p_s g_{s,k}^2 x_s z_k + N_s, \end{aligned} \tag{1}$$

where x_s and p_s are the information symbol and transmit power of the RD, respectively; z_k and $\xi_k \in \Xi =$

$\{\varsigma_1, \ldots, \varsigma_D | 0 \le \varsigma_i \le 1, \forall i = 1, \ldots, D\}$ are the information symbol and RC of BD k, respectively; and N_s is the zero-mean additive white Gaussian noise arriving at the RD, with variance σ^2.

B. Uplink NOMA-assisted BackCom

The network operates using NOMA to enhance spectral efficiency (SE). According to NOMA standardization [32], two BDs can be multiplexed in the power domain, allowing them to share the same frequency resource. The RD receives the composite signal and then applies the SIC to decode it. In the case of uplink NOMA, the better signal, which considers the worse signal as noise, is detected and decoded first. Subsequently, the decoded signal is subtracted from the composite signal using SIC, allowing the remaining inferior signal to be decoded without NOMA interference.

Let $o_{s,k} \in \{0, 1\}$ denote the SA of BD k, where $o_{s,k} = 0$ and $o_{s,k} = 1$ represent the sleep and active states, respectively. The network model operates over T time slots indexed by $t \in \{0, \ldots, T-1\}$. At time t, two BDs, i.e., k_i and k_j ($k_i, k_j \in \mathcal{K}, k_i < k_j$), may be active on channel s with $o_{s,k_i}(t) = o_{s,k_j}(t) = 1$. Without loss of generality, all BDs are assumed to be sorted and indexed based on descending channel gain, i.e., $g_{s,k_i}(t) \ge g_{s,k_j}(t)$ if $k_i < k_j$. Moreover, the signal of BD k_i is decoded before that of BD k_j by performing NOMA. Therefore, the uplink SINRs considering transmissions from BDs k_i and k_j are computed as follows:

$$\varphi_{s,k_i}(t) = \frac{o_{s,k_j}(t)\xi_{k_i}(t) p_S g^4_{s,k_i}(t)}{o_{s,k_j}(t)\xi_{k_j}(t) p_S g^4_{s,k_j}(t) + \sigma^2} \tag{2}$$

and

$$\varphi_{s,k_j}(t) = \frac{o_{s,k_j}(t)\xi_{k_j}(t) p_S g^4_{s,k_j}(t)}{\sigma^2}. \tag{3}$$

C. Energy Harvesting Model

In the sleep state, most of the harvested energy of BD k is stored in its battery as follows:

$$E_k^{\text{sleep}}(t) = \left(1 - o_{s,k}(t)\right) \eta_k p_S g_{s,k}(t), \tag{4}$$

where η_k represents the energy harvesting efficiency coefficient of BD k. The harvested energy is defined without considering thermal noise because none of the BDs have any active RF components [33]. In the active state, the harvested energy is divided by ξ_k, in which a portion is used to reflect the modulated signal, $\xi_k p_S g_{s,k}(t)$. The remaining harvested energy is computed as follows:

$$E_k^{\text{active}}(t) = o_{s,k}(t)\eta_k(1 - \xi_k(t)) p_S g_{s,k}(t). \tag{5}$$

If $\xi_k(t)$ approaches 1, $E_k^{\text{active}}(t)$ becomes relatively close to 0, resulting in ineffective circuit operation. In this case, the accumulated energy is used to supplement the required power and maintain normal circuit operation. Let M_k denote the accumulated energy of BD k, which can be updated as follows:

$$M_k(t+1) = M_k(t) + E_k^{\text{sleep}}(t) + E_k^{\text{active}}(t) - P_k^c, \tag{6}$$

where E_k^{sleep} and E_k^{active} are concurrently controlled by $o_{s,k}$, and P_k^c denotes the constant circuit power of BD k.

D. EE Maximization Problem

Applying the Shannon formula, the overall SE at time t can be computed as follows:

$$\Psi(t) = \sum_{1 \le m \le n \le K} \left(\log_2(1 + \varphi_{s,m}(t)) + \log_2(1 + \varphi_{s,n}(t))\right). \tag{7}$$

The energy consumption consists of the (i) energy for RD transmissions, $E_0 = p_s/\varrho$, where $\varrho \in (0, 1]$ is the power amplifier efficiency; (ii) constant circuit power consumption of the RD, P_0^c; and (iii) constant circuit power consumption of the BDs, P_k^c. Each BD k can harvest energy as $\Delta M_k(t) = E_k^{\text{sleep}}(t) + E_k^{\text{active}}(t) - P_k^c$. Therefore, the EE of the complete system can be computed as follows:

$$\Upsilon(t) = \frac{\Psi(t)}{E_0 + P_0^c - \sum_{k \in \mathcal{K}} \Delta M_k(t)}. \tag{8}$$

The objective of this study is to maximize long-term EE by jointly optimizing SA and RCs. The mathematical formulation of the corresponding optimization problem is as follows:

$$\max_{\mathbf{o}(t), \mathbf{x}(t)} \sum_{t=0}^{T-1} \gamma^t \Upsilon(t), \tag{9}$$

$$\text{s.t.} \quad o_{s,k}(t) \in \{0, 1\}, \forall k \in \mathcal{K}, \tag{9a}$$

$$\sum_{k \in \mathcal{K}} o_{s,k}(t) \le 2, \tag{9b}$$

$$o_{s,k}(t) = 1 : \varphi_{s,k}(t) \ge \varphi^{\min}, \forall k \in \mathcal{K}, \tag{9c}$$

where $\mathbf{o}(t) = \{o_{s,k}(t) | k \in \mathcal{K}\}$ is the SA vector, $\mathbf{x}(t) = \{\xi_k(t) | k \in \mathcal{K}\}$ is the RC vector, and $\gamma \in (0, 1)$ is a discounting factor. Constraints (9a) and (9b) ensure that at most two BDs are allowed to share the channel in the context of SA. Constraint (9c) sets the lower bound of the SINRs ($\varphi^{\min}$) for SIC decoding.

III. Algorithm Design

The formulated problem poses challenges for conventional optimization tools because of two main reasons: First, the objective function is a long-term non-convex function with mixed-integer variables. Second, considering the dynamic channel conditions, the RD cannot determine the channel power gain with which sleep-state BDs harvest energy. Therefore, before the optimization process, the RD must send probing signals to all BDs and wait for the signals carrying the CSI. This method is highly complex and incurs significant communication overhead. To this end, DRL emerges as a promising solution that can maximize long-term reward for dynamic systems. We developed a DRL-based solution for the formulated problem. Specifically, a DRL agent is deployed at the RD, which follows a Markov decision process model. At each time t, the agent selects a joint action of the SA and RCs based on the observed states, computes EE result as a reward, and transitions to the subsequent state. This iterative process aims to learn the optimal policy that maximizes the achievable long-term EE. The state, action, and reward function are specified as follows:

1) *State space:* One of the main challenges in solving the formulated problem is the lack of knowledge regarding

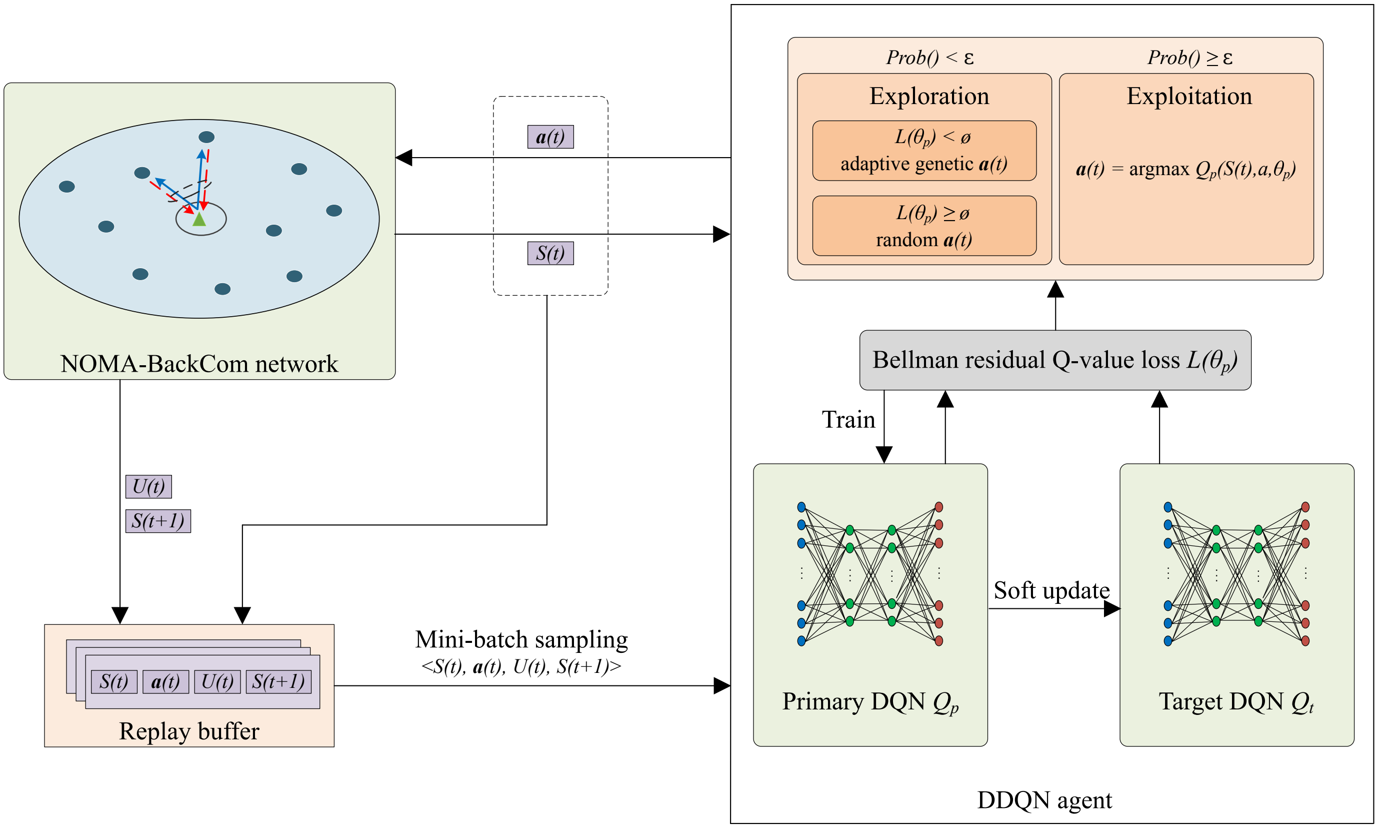

Fig. 2. Architecture of the DDQN algorithm for EE maximization.

the channel gain of sleep-state BDs. To address this challenge, we estimate the unknown channel gain using the historical CSI data collected during the active periods of BDs as follows:

$$\tilde{g}_{s,k}(t) = \mathbb{E}\left[g_{s,k}(t)|o_{s,k}(t) = 1\right], \qquad (10)$$

where the initial value can be initialized as $\tilde{g}_{s,k}(0) = r_k^{-\alpha}, \forall k \in \mathcal{K}$. Furthermore, the estimated channel gain can be incrementally updated as

$$\tilde{g}_{s,k}(t+1) = \frac{\delta_k \tilde{g}_{s,k}(t) + o_{s,k}(t) g_{s,k}(t)}{\delta_k + o_{s,k}(t)}, \qquad (11)$$

where $\delta_k \leftarrow \delta_k + o_{s,k}(t)$ is the total number of historical CSI data points of BD k. Then, at each time t, the agent obtains a system state, which is formulated based on the estimated channel gain and stored energy of the BDs:

$$S(t) \triangleq \{\tilde{g}_{s,k}(t), M_k(t) | k \in \mathcal{K}\}. \qquad (12)$$

2) *Action space:* The joint action of the agent is defined as

$$\begin{aligned}\mathbf{a}(t) &\triangleq \{\mathbf{o}(t), \mathbf{x}(t)\} \\ &= \{o_{s,k}(t), \xi_k(t) | k \in \mathcal{K}\}, \end{aligned} \qquad (13)$$

where $\xi_k(t) = 0$ if $o_{s,k}(t) = 0$ that simplifies the RC action in the sleep state. This consideration is reasonably practicable because signal reflection must be disabled in the sleep state.

3) *State transition and reward function:* When the selected action, $\mathbf{a}(t)$, is executed in the observed state, $S(t)$, the system transitions to the next state, $S(t+1)$, The state components are obtained based on $S(t)$ and the newly arrived CSI. In addition, a step reward is computed based on the achievable EE as

$$U(t) = \Upsilon(t) - \Delta\tilde{g}_{s,k}(t), \qquad (14)$$

where $\Delta\tilde{g}_{s,k}(t) = \sum_{k\in\mathcal{K}} o_{s,k}(t) \left|g_{s,k}(t) - \tilde{g}_{s,k}(t)\right|$ is the error of the estimated channel gain.

Based on the DQN algorithm, we developed a novel algorithm that combines the benefits of using a finite replay buffer memory to store recent experiences for training and an additional target Q-network. This design mitigates the overoptimistic estimation of Q-values [34], as shown in Fig. 2. The pseudo code is presented as Algorithm 1. Initially, the system parameters are initialized as follows: $\mathcal{K}$; $\mathcal{S}$; Z; T; replay memory $\mathcal{M}$; two DQNs Q_p (primary) and Q_t (target) with weights θ_p and θ_t, respectively. The training error $L(\theta_p)$; and error threshold value ϕ. Q_p and Q_t have the same neural network structure, each consisting of two fully connected hidden layers using the rectified linear unit as the activation function. The input and output layers are sized according to the dimensions of the state and action vectors, respectively. For each observed state $S(t)$, the agent selects action $\mathbf{a}(t)$ to collect experiences based on the ϵ-greedy strategy. Exploration is executed with probability ϵ, and exploitation is executed otherwise. To facilitate convergence, we partition the exploration based on a mini-batch Q-value loss. Specifically, if the loss is smaller than the threshold value ϕ, the adaptive genetic action

Algorithm 1 DRL-based algorithm for EE maximization

```
Initialize $\mathcal{K}, \mathcal{S}, Z, T, \mathcal{M}, \theta_p, \theta_t \leftarrow \theta_p, L(\theta_p), \phi$;
for episode $i = 1, \dots, Z$ do
    Reset $\tilde{g}_{s,k}(0)$ and $M_k(0)$, $\forall k \in \mathcal{K}$;
    for step $t = 1, \dots, T$ do
        Observe state $S(t)$;
        Select $\mathbf{a}(t)$ based on probability $\varepsilon$ and $\epsilon$-greedy:
            Adaptive genetic $\mathbf{a}(t)$ with AGA algorithm [30] if $\varepsilon < \epsilon$ and $L(\theta_p) < \phi$;
            $\mathbf{a}(t)$ is random if $\varepsilon < \epsilon$ and $L(\theta_p) \geq \phi$;
            $\mathbf{a}(t) = \mathrm{argmax}_{\mathbf{a}} Q_p(S(t), \mathbf{a}; \theta_p)$ o.w.;
        Execute $\mathbf{a}(t)$ to observe $U(t)$ and $S(t+1)$;
        Save tuple $\langle S(t), \mathbf{a}(t), U(t), S(t+1)) \rangle$ in $\mathcal{M}$;
        if number of tuples $\geq$ batch size $C$ then
            Sample $C$ tuples from $\mathcal{M}$ for learning;
            for $t = 1, \dots, C$ do
                Find $a(t) = \mathrm{argmax}_{\mathbf{a}} Q_t(S(t+1), \mathbf{a}; \theta_t)$;
                Find $y(t) = U(t) + \gamma Q_p(S(t+1), a(t); \theta_p)$;
            end for
            Perform gradient descent w.r.t. $\theta_p$ on Q-value loss: $L(\theta_p) = \mathbb{E}_t^C \left[ \left( y(t) - Q_p(S(t), \mathbf{a}(t); \theta_p) \right)^2 \right]$;
            Every $G$ steps, soft update $\theta_t \leftarrow \tau\theta_p + (1-\tau)\theta_t$;
        end if
    end for
end for
for each step $t$ in exploitation phase do
    Observe state $S(t)$;
    Output $\mathbf{a}^*(t) = \mathrm{argmax}_{\mathbf{a}} Q_p(S(t), \mathbf{a}; \theta_p)$;
end for
```

is selected from a candidate set using the AGA algorithm [30]. Otherwise, a random action is selected. The learning process employs the the stochastic gradient descent method to update the weight θ_p, minimizing the Bellman residual Q-value loss as follows:

$$L(\theta_p) = \mathbb{E}_{t=1}^C \left[\left(y(t) - Q_p(S(t), \mathbf{a}(t); \theta_p) \right)^2 \right], \tag{15}$$

where C is the batch size, $Q_p(S(t), \mathbf{a}(t); \theta_p)$ denotes the output of Q_p for the state-action pair $(S(t), \mathbf{a}(t))$, and $y(t) = U(t) + \gamma Q_p(S(t+1), a(t); \theta_p)$, $t \in \{1, \dots, C\}$, is a target Q-value with $a(t) = \mathrm{argmax}_{\mathbf{a}} Q_t(S(t+1), \mathbf{a}; \theta_t)$. The training process is terminated when all episodes are implemented or when the updated amount of θ_p becomes significantly small.

The following proposition defines the convergence of the proposed algorithm as the prerequisite condition for using the primary DQN to select adaptive genetic actions.

Proposition 1. *Algorithm 1 converges to the global optimal Q-function for EE maximization.*

Proof. In the training process of Algorithm 1, weight θ_p is updated at each iteration to minimize the Q-value loss, $L(\theta_p)$, such that the Q-value is gradually updated as follows:

$$Q_p(S(t), \mathbf{a}(t); \theta_p) \leftarrow (1-\upsilon) Q_p(S(t), \mathbf{a}(t); \theta_p) + \upsilon y(t), \tag{16}$$

where $\upsilon \in (0, 1)$ is the learning rate. Moreover, using the Bellman equation, the global optimal Q-value is obtained as

$$\begin{aligned} Q_p^*(S(t), \mathbf{a}(t)) = \sum \Theta(S^{'}(t) | S(t), \mathbf{a}(t)) \\ \times [U(t) + \gamma \max\nolimits_{\mathbf{a}^{'}(t)} Q_p(S^{'}(t), \mathbf{a}^{'}(t))], \end{aligned} \tag{17}$$

where $\Theta(S^{'}(t) | S(t), \mathbf{a}(t))$ is the state transition probability from $S(t)$ to $S^{'}(t)$ when action $\mathbf{a}(t)$ is executed. Without loss of generality, we assume that the state transition probabilities are stationary. For instance, the probabilities can be predefined or follow a Gaussian distribution. Therefore, a global optimal Q-value, $Q_p^*(S(t), \mathbf{a}(t))$ exists. Moreover, the difference between the training and optimal Q-values is

$$\begin{aligned} &\Delta Q_p(S(t), \mathbf{a}(t)) = \\ &(1-\upsilon) \Delta Q_p(S(t), \mathbf{a}(t)) + \upsilon \left(y(t) - Q_p^*(S(t), \mathbf{a}(t)) \right) \end{aligned} \tag{18}$$

Without loss of generality, we assume that the training process is sufficiently long to ensure that all possible state-action pairs are visited. This assumption is reasonably feasible because of the following reasons. First, the state-action space is shrunk remarkably owing to the quantization. Specifically, instead of using directly the estimated channel gain values for the state, we quantize them into four determined levels, such as excellent, good, normal, and poor channel gain states. As a result, the total number of state values is significantly reduced. Second, the probability of visiting each state-action pair is generally increased when implementing more episodes, which are based on a larger number of initial randomized generations of state values. Practically we conducted extensive simulations and achieved stable convergence based on 1000 episodes corresponding to 1000 initial randomized generations of state values. Therefore, the Q-value difference, $\Delta Q_p(S(t), \mathbf{a}(t))$, converges to zero because sufficient data is available for updating the estimated Q-value [35]. That completes the proof of Proposition 1. □

IV. Performance Evaluation

A. Simulation Settings

Extensive simulations were conducted based on a network area bounded by $R_i = 1$ m and $R_o = 15$ m. A set of $K = 10$ BDs was uniformly distributed in the coverage area. The transmit power of the RD was $p_s = 500$ mW. The path-loss exponent was $\alpha = 2.5$ for Rayleigh fading, and the noise power was $\sigma^2 = -100$ dBm. The other system-level parameters were set as follows: $\eta_k = 0.6$, $\varrho = 0.9$, $P_k^c = 1$ mW, $P_0^c = 110$ mW, and $\varphi^{\min} = 3$ dB. The DRL-based training process was implemented at a maximum of $Z = 1000$ episodes, each including $T = 50$ steps. The probability of exploring the action space (ϵ) was initialized as 0.9, and it was gradually reduced in each training step until it reached 0.1. The discounting factor was $\gamma = 0.9$. The replay memory was sized at $|\mathcal{M}| = 10^4$, which provided batches for learning, each with $C = 32$ samples. A learning rate of $\tau = 0.001$ was adopted to the soft update weight θ_t. Finally, each hidden layer of the DQNs had 128 nodes.

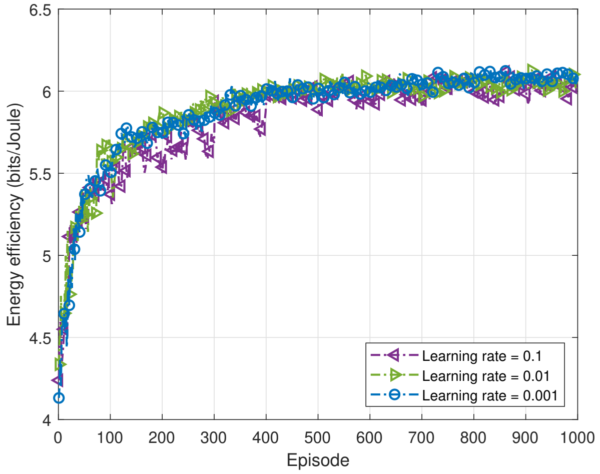

Fig. 3. Convergence of the proposed scheme in terms of achievable EE.

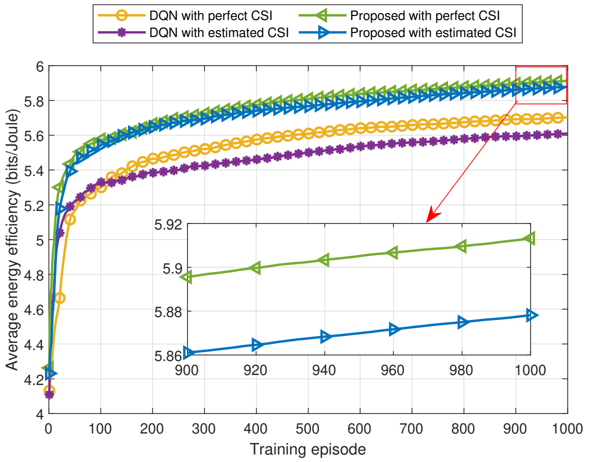

Fig. 4. Convergence of the proposed scheme when using perfect CSI and estimated CSI for training.

B. Convergence Analysis

Fig. 3 illustrates the convergence of the proposed scheme in terms of the achievable EE under various learning rates, i.e., $\upsilon = 0.1$, $\upsilon = 0.01$, and $\upsilon = 0.001$. Across all learning-rate settings, EE starts low and gradually increases as training progresses. We observe that during training, EE with learning rates of 0.01 and 0.001 exhibits slightly more superior performance than that with a learning rate of 0.1. Moreover, convergence is achieved after approximately 400 episodes. After convergence, the achievable EE remains consistent across different learning-rate settings. The slight difference in achievable EE between different learning-rate settings is attributed to the consecutive reduction in the exploration rate during training. This reduction leads to a decrease in the updating number of weight and decreases the impact of learning rates. Consequently, the effectiveness of the proposed DRL-based algorithm is observed to be relatively independent of the learning rate.

Fig. 4 illustrates the impact of using estimated CSI for training. We observe that both of the proposed scheme and DQN algorithm converge when using either the perfect or the estimated CSI. Furthermore, the performance gap of the proposed scheme when using the perfect and the estimated CSI is relatively small. For instance, after 1000 training episodes, the average achievable EE when using perfect CSI reaches to approximately 5.91 bits/J, which is 0.5% greater than that achieved when using estimated CSI. This small performance gap is explained because either of the perfect or estimated CSI can represent the system information, which is the input for the proposed DRL-based scheme to approximate the mapping from system information to the optimal SA and RC utilizing deep neural networks. In addition, the estimation error is deducted to step reward in each training iteration that also reduces the impact of estimation error on training performance.

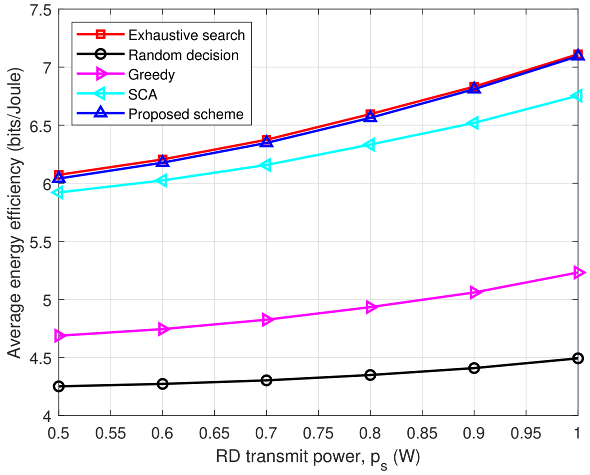

Fig. 5. EE comparison between existing approaches.

C. Performance Comparison

Fig. 5 depicts the EE comparison plotted against the RD transmit power, considering different approaches. Specifically, we compare the EE achieved in the proposed scheme with those of the following methods: greedy approach which maximizes the RCs of active BDs to enhance the SE; random decision; SCA approach; and the optimum strategy obtained through exhaustive search using the exact CSI. The achievable EE values are averaged from the results of 100 consecutive episodes. The results reveal that our proposed scheme outperforms the SCA, greedy, and random decision schemes. Remarkably, it is comparable to the optimum scheme with exhaustive search algorithm. For instance, at a transmit power of 0.5 W, the achievable EE in the proposed scheme is 6.07 bits/J, which is only 0.6% less than that of the optimum scheme and 30.3%, 41.7%, and 4% greater than those of the greedy, random, and SCA approaches, respectively. Moreover, as the transmit power increases, the EE increases across all schemes with different slopes. Notably, the most significant increase we observe in the proposed and optimum schemes. This result is attributed to the joint optimization of SA and

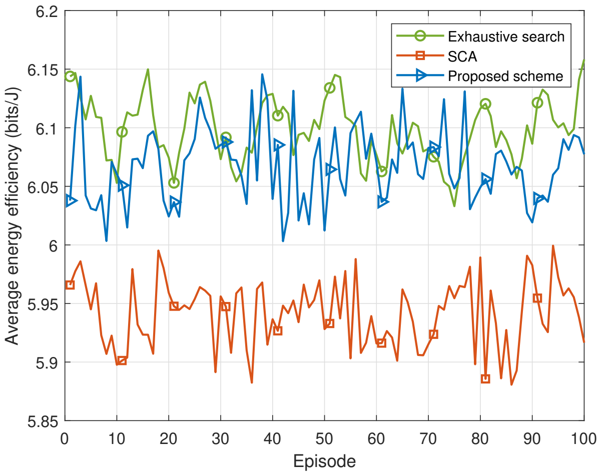

Fig. 6. Fluctuation of the average EE versus consecutive episodes.

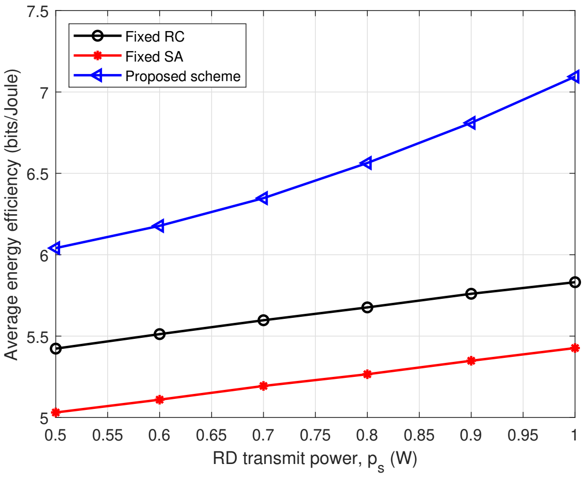

Fig. 8. EE achieved in joint and non-joint optimizations of SA and RCs.

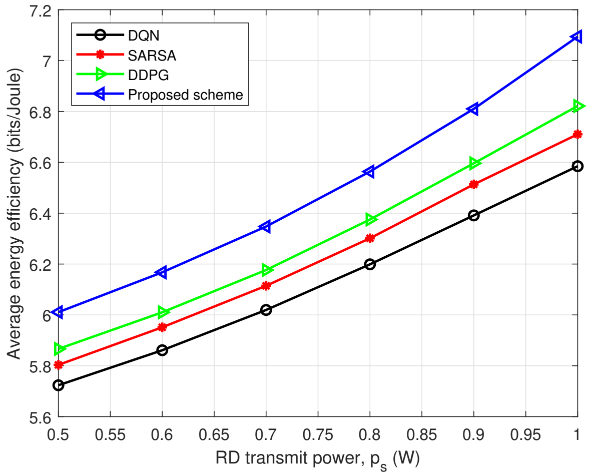

Fig. 7. EE comparison between various DRL frameworks.

RCs of the proposed scheme, which enhances utility at higher transmit power values.

We further compare the performance of 3 outstanding schemes: the SCA approach, exhaustive search, and proposed schemes. Fig. 6 shows the fluctuation in EE over 100 consecutive episodes. Notably, both the proposed and exhaustive search schemes exhibit similar average EE, which surpasses that of the SCA approach scheme. The gap between them falls within the range of 2% and 4%, demonstrating the superiority of the proposed scheme over the SCA approach. Moreover, the fluctuation range remains consistent across all three schemes and is influenced by the variation in channel conditions.

Fig. 7 shows the performance comparison of different DRL frameworks in terms of the achievable EE under various RD transmit power, p_s. The results reveal that increasing RD transmit power increases the achievable EE. For instance, the achievable EE increases from 6.0 bits/J with $p_s = 0.5$ W to 7.1 bits/J with $p_s = 1$ W in the proposed scheme. We observe similar behavior of the achievable EE in other DRL framework schemes. Compared to the other DRL schemes using DQN, SARSA, and DDPG algorithms, the proposed DRL framework significantly improves the achievable EE. For instance, when $p_s = 1$ W, the average EE in the proposed scheme reaches to 7.1 bits/J, which is approximately 4.4%, 6.0%, and 7.7% greater than those achieved using DDPG, SARSA, and DQN algorithms, respectively.

Fig. 8 depicts the achievable EE in joint and non-joint optimization schemes of SA and RC. Non-joint optimization schemes of SA and RC are implemented by fixing each of the variables and fully searching to determine the best remained variable. The results reveal that joint optimization of SA and RC significantly improves the achievable EE compared to single optimizations of SA and RC. Remarkably, as the transmit power increases, the performance gain in terms of the achievable EE becomes greater. For instance, at the transmit power of 0.5 W, the achievable EE gain of the proposed joint optimization scheme over the fixed RC and fixed SA schemes is approximately 11.1% and 20.0%, respectively, which increases to 22.4% and 31.5%, respectively at the transmit power of 1 W. These results reflect the effectiveness in terms of EE of the proposed joint optimization scheme applying for practical systems with high transmit power.

Fig. 9 depicts the stability of the achievable EE. The figure shows that the EE achieved in the proposed scheme significantly surpasses that of the existing algorithms, including the random decision, greedy, fixed SA, fixed RC, and SCA approaches. In addition, the proposed scheme exhibits the greatest stability with the fewest outliers. Notably, the achievable EE in the proposed scheme is approximately 6.0 bits/J, which is closely aligns with that of the optimum scheme obtained using exhaustive search algorithm.

Fig. 10 shows a comparison between NOMA and OMA in the context of a BackCom system under various RD transmit power. The results reveal that integrating NOMA with a BackCom system significantly increases data rate compared to the OMA scheme. Specifically, when the RD transmit power is 1 W, NOMA achieves a data rate of 5.35 bps, which is approximately 38.6% higher than that obtained using the

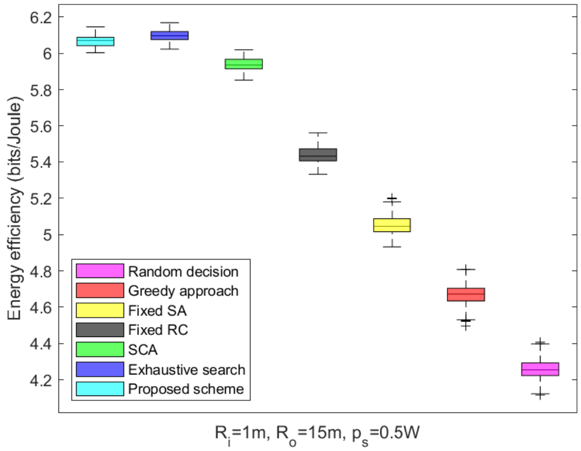

Fig. 9. Stability of the achievable EE.

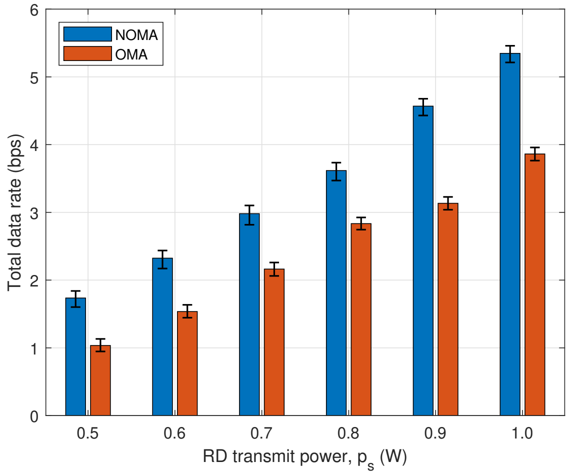

Fig. 10. Data rate comparison between NOMA and OMA in BackCom system.

OMA scheme. This improvement is attributed to NOMA's enhancement of the access capacity for BDs and the spectral efficiency of the BackCom system. The adoption of NOMA, which enhances data rate compared to OMA, leads to shorter transmission time and reduces energy consumption during data transmission.

D. Complexity and Practical Running Time

After Algorithm 1 converges, we can employ its trained weight to make the SA and RC action output. Specifically, for each step in the exploitation phase, the DRL agent delivers an optimal joint action of SA and RC as $\mathbf{a}^*(t) = \text{argmax}_{\mathbf{a}} Q_p(S(t), \mathbf{a}; \theta_p)$. The complexity of obtaining the optimal solution is calculated according to the size of the state vector, $S(t)$, action vector, $\mathbf{a}(t)$, and weight matrix, θ_p, as $O(K^2 MN)$. This term is polynomial where K is the number of BDs and M and N represent the number of nodes in the two hidden layers of the primary DQN. Moreover, communication overhead is significantly reduced because the RD agent does not require exact CSI data for training the optimal policy.

TABLE I
TIME CONSUMPTION FOR ACTION SELECTION

Scheme	Action time
Random decision	0.00186 s
Greedy approach	0.00535 s
SCA approach	0.03671 s
Exhaustive search	1.35685 s
Proposed scheme	0.01378 s

Moreover, we compute the action decision time of various schemes and present the results in Table I. The random decision and greedy approach schemes exhibit the shortest action decision time, with durations of 0.00186 and 0.00535 s, respectively. Regrettably, their performance does not match those of other schemes. Furthermore, the exhaustive search scheme requires the longest time to deliver action, which is up to 1.35685 s. Therefore, despite its ability to yield the best action, it is unsuitable for practical IoT systems. In addition, the proposed scheme outperforms the SCA approach scheme, delivering action more rapidly in only 0.01378 s, which is the inference time needed by the trained model to derive the SA and RC. Notably, the training time is not considered the running time of the proposed scheme because the training process involves collecting experience data and optimizing model weight parameters for the optimal policy of SA and RC. When the training is complete, we use model weight parameters to produce the SA and RC. Therefore, the proposed scheme is most suitable for practical IoT systems because of its ability to provide near-optimal action within a short time.

V. CONCLUSIONS

This paper proposes a novel energy-efficient resource allocation scheme for 6G NOMA-assisted BackCom networks. The scheme maximizes EE by jointly optimizing the SA and RCs of BDs. To address the practical energy harvesting scenario, where BDs are equipped with batteries, we formulated a challenging joint optimization problem of SA and RCs. This problem involved nonconvexity and system dynamics, making it difficult to solve straightforwardly. Therefore, we applied a DRL-based strategy to develop an efficient DDQN algorithm that learned the optimal joint action in the long term. During training, we estimated the CSI instead of relying on the computationally expensive perfect CSI that reduced the communication overhead raised in collecting exact CSI data. Extensive simulation results demonstrated reliable convergence. Compared to conventional approaches, such as the random decision, greedy, DQN, and SCA approaches, the proposed scheme significantly improved EE. The EE achieved in the proposed scheme was relatively close to that achieved in the optimum scheme adopted exhaustive search algorithm. Furthermore, the proposed scheme provides the most suitable solution for practical IoT systems because of its near-optimal result and rapid action decision time.

References

[1] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 2018.

[2] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 2020.

[3] V. D. Tuong, T. P. Truong, T.-V. Nguyen, W. Noh, and S. Cho, "Partial computation offloading in NOMA-assisted mobile-edge computing systems using deep reinforcement learning," *IEEE Internet Things J.*, vol. 8, no. 17, pp. 13 196–13 208, 2021.

[4] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, 2020.

[5] U. S. Toro, K. Wu, and V. C. M. Leung, "Backscatter wireless communications and sensing in green Internet of Things," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 37–55, 2022.

[6] Y. Ye, L. Shi, R. Q. Hu, and G. Lu, "Energy-efficient resource allocation for wirelessly powered backscatter communications," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1418–1422, 2019.

[7] H. Yang, Y. Ye, and X. Chu, "Max-min energy-efficient resource allocation for wireless powered backscatter networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 688–692, 2020.

[8] W. U. Khan, M. A. Javed, T. N. Nguyen, S. Khan, and B. M. Elhalawany, "Energy-efficient resource allocation for 6g backscatter-enabled noma iov networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9775–9785, 2021.

[9] J. Guo, X. Zhou, S. Durrani, and H. Yanikomeroglu, "Design of non-orthogonal multiple access enhanced backscatter communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6837–6852, 2018.

[10] G. Yang, X. Xu, Y.-C. Liang, "Resource allocation in NOMA-enhanced backscatter communication networks for wireless powered IoT," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 117–120, 2020.

[11] Y. Xu, Z. Qin, G. Gui, H. Gacanin, H. Sari, and F. Adachi, "Energy efficiency maximization in NOMA enabled backscatter communications with QoS guarantee," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 353–357, 2021.

[12] T. T. Anh, N. C. Luong, D. Niyato, Y.-C. Liang, and D. I. Kim, "Deep reinforcement learning for time scheduling in RF-powered backscatter cognitive radio networks," in *Proc. IEEE WCNC*, 2019, pp. 1–7.

[13] Y. Zhang, Z. Mou, F. Gao, L. Xing, J. Jiang, and Z. Han, "Hierarchical deep reinforcement learning for backscattering data collection with multiple UAVs," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3786–3800, 2021.

[14] F. Jameel, W. U. Khan, M. A. Jamshed, H. Pervaiz, Q. Abbasi, and R. Jäntti, "Reinforcement learning for scalable and reliable power allocation in SDN-based backscatter heterogeneous network," in *Proc. IEEE INFOCOM Workshops*, 2020, pp. 1069–1074.

[15] Q. Zhang, L. Zhang, Y.-C. Liang, and P.-Y. Kam, "Backscatter-NOMA: A Symbiotic System of Cellular and Internet-of-Things Networks," *IEEE Access*, vol. 7, pp. 20 000–20 013, 2019.

[16] S. Zeb, Q. Abbas, S. A. Hassan, A. Mahmood, R. Mumtaz, S. H. Zaidi, S. A. R. Zaidi, and M. Gidlund, "NOMA enhanced backscatter communication for green IoT networks," in *Proc. 16th Int. Sympos. Wireless Commun. Syst. (ISWCS)*. IEEE, 2019, pp. 640–644.

[17] Z. Ding and H. V. Poor, "Advantages of NOMA for multi-user backcom networks," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3408–3412, 2021.

[18] A. Farajzadeh, O. Ercetin, and H. Yanikomeroglu, "UAV data collection over NOMA backscatter networks: UAV altitude and trajectory optimization," in *ICC 2019-2019 IEEE Inter. Conf. Commun. (ICC)*. IEEE, 2019, pp. 1–7.

[19] W. Chen, H. Ding, S. Wang, D. B. da Costa, F. Gong, and P. H. J. Nardelli, "Backscatter Cooperation in NOMA Communications Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3458–3474, 2021.

[20] A. W. Nazar, S. A. Hassan, H. Jung, A. Mahmood, and M. Gidlund, "BER analysis of a backscatter communication system with non-orthogonal multiple access," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 2, pp. 574–586, 2021.

[21] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12 286–12 290, 2020.

[22] A. Raza, S. J. Nawaz, A. Ahmed, S. Wyne, B. Muhammad, M. N. Patwary, and R. Prasad, "A NOMA-enabled cellular symbiotic radio for mMTC," *Wireless Personal Commun.*, vol. 122, no. 4, pp. 3545–3571, 2022.

[23] X. Li, Y. Zheng, M. D. Alshehri, L. Hai, V. Balasubramanian, M. Zeng, and G. Nie, "Cognitive AmBC-NOMA IoV-MTS networks with IQI: reliability and security analysis," *IEEE Trans. Intell. Transp. Syst.*, 2021.

[24] Z. Ding, "Harvesting devices' heterogeneous energy profiles and QoS requirements in IoT: WPT-NOMA vs BAC-NOMA," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 2837–2850, 2021.

[25] X. Li, M. Zhao, M. Zeng, S. Mumtaz, V. G. Menon, Z. Ding, and O. A. Dobre, "Hardware impaired ambient backscatter NOMA systems: Reliability and security," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2723–2736, 2021.

[26] Y. Liao, G. Yang, and Y.-C. Liang, "Resource allocation in NOMA-enhanced full-duplex symbiotic radio networks," *IEEE Access*, vol. 8, pp. 22 709–22 720, 2020.

[27] W. U. Khan, F. Jameel, N. Kumar, R. Jäntti, and M. Guizani, "Backscatter-enabled efficient V2X communication with non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1724–1735, 2021.

[28] Z. Ding and H. V. Poor, "On the application of BAC-NOMA to 6G umMTC," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2678–2682, 2021.

[29] M. Ahmed, W. U. Khan, A. Ihsan, X. Li, J. Li, and T. A. Tsiftsis, "Backscatter sensors communication for 6G low-powered NOMA-enabled IoT networks under imperfect SIC," *IEEE Syst. J.*, vol. 16, no. 4, pp. 5883–5893, 2022.

[30] M. Srinivas and L. M. Patnaik, "Adaptive probabilities of crossover and mutation in genetic algorithms," *IEEE Trans. Syst. Man Cybernetics*, vol. 24, no. 4, pp. 656–667, 1994.

[31] S. Xiao, H. Guo, and Y.-C. Liang, "Resource allocation for full-duplex-enabled cognitive backscatter networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3222–3235, 2019.

[32] J. M. Meredith, "Study on Downlink Multiuser Superposition Transmission (MUST) for LTE," TR 36.859, 3rd Generation Partnership Project, 2015.

[33] D. T. Hoang, D. Niyato, P. Wang, D. I. Kim, and Z. Han, "Ambient backscatter: A new approach to improve network performance for RF-powered cognitive radio networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3659–3674, 2017.

[34] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proc. 30th AAAI Conf. Artif. Intell.*, vol. 30, no. 1, Phoenix, AZ, USA, 2016, pp. 2094–2100.

[35] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

Van Dat Tuong received the B.S. degree in mechatronics from Hanoi University of Science and Technology, Hanoi, Vietnam, in 2012, and the M.S. degree in system software, from Chung-Ang University, Seoul, South Korea, in 2021, where he is currently pursuing the Ph.D. degree in big data. From 2012 to 2018, he was a Software Engineer with the Viettel Software Center, Viettel Telecom and the Samsung Vietnam Mobile Center (SVMC), Samsung Electronics Vietnam, Hanoi, Vietnam. His research interests include machine learning, game theory, convex optimization, and their applications in wireless networking and ubiquitous computing.

Sungrae Cho is a professor with the school of computer sciences and engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an assistant professor with the department of computer sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a senior member of technical staff with the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a research staff member with electronics and telecommunications research institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a visiting professorship with the national institute of standards and technology (NIST), Gaithersburg, MD, USA. He received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002. He has been a KICS fellow since 2021. He received numerous awards including Haedong Best Researcher of 2022 in Telecommunications and Award of Korean Ministry of Science and ICT in 2021. His current research interests include wireless networking, network intelligence, and network optimization. He has been an editor-in-chief (EIC) of ICT Express (Elsevier) since 2024, a subject editor of IET Electronics Letter since 2018, an executive editor of Wiley Transactions on Emerging Telecommunications Technologies since 2023, and was an area editor of Ad Hoc Networks Journal (Elsevier) from 2012 to 2017. He has served numerous international conferences as a general chair, TPC chair, or an organizing committee chair, such as IEEE ICC, IEEE SECON, IEEE ICCE, ICOIN, ICTC, ICUFN, APCC, TridentCom, and the IEEE MASS.