

Online Learning for Joint Energy Harvesting and Information Decoding Optimization in IoT-Enabled Smart City

Yongjae Kim¹, Member, IEEE, Bang Chul Jung², Senior Member, IEEE, and Yujae Song³, Member, IEEE

Abstract—In this study, we first present a framework that jointly optimizes energy harvesting and information decoding for Internet of Things (IoT) devices, which are capable of simultaneous wireless information and power reception, in a smart city. In particular, a generalized power-splitting receiver for IoT devices is designed, where each antenna in the receiver has an independent power splitter, unlike the existing works in which only one power splitter is employed regardless of the number of antennas in the receiver. Such a receiver design can provide a great degree of freedom to improve the network performance. Based on the presented framework, for each IoT device, we formulate an optimization problem whose objective is to maximize the harvested energy of each IoT device while satisfying its data rate requirement. To solve this problem, we propose a double-deep deterministic policy gradient-based online learning algorithm which enables each IoT device to jointly determine receive beamforming and power-splitting ratio vectors in real time. Furthermore, each IoT device can implement the proposed algorithm in a distributed manner using only its local channel state information. As such, cooperation and information exchange among the base stations and IoT devices are not necessary when performing the proposed algorithm at IoT devices. The extensive simulation results show the validity of the proposed algorithm.

Index Terms—Deep reinforcement learning (DRL), energy harvesting (EH), information decoding (ID), Internet of Things (IoT), smart city.

I. INTRODUCTION

ACCORDING to [1], 55% of the world's population now lives in cities, and it is expected to increase to 67% by the year 2050. The gradual increase in the urbanization causes city-related problems, such as energy depletion, traffic

congestion, safety security, etc. To address the growing challenges of urbanization, the Information and Communication Technologies (ICTs) are adopted, which makes the cities smart enough, through deploying and promoting sustainable development practices [2]. The key characteristics of a smart city are a high degree of information technology integration and a comprehensive application of information resources. The essential components of a smart city for urban development involve smart industry, smart technology, smart services, smart management, and smart life.

In particular, with an advancement in the field of smart cities' sensors, the new Internet of Things (IoT) applications are enabling smart city initiatives worldwide [3]. IoT aims at utilizing a variety of IoT devices equipped with sensors (e.g., radar sensors, temperature sensors, fire sensors, traffic cameras, humidity sensors, etc.), and connecting them to the Internet via specific protocols. It offers the ability to remotely monitor, manage, and control devices, and to create new insights and actionable information from massive streams of real-time data. To realize a smart city supporting diverse IoT applications, numerous battery-powered IoT devices should be installed in many places in the smart city according to the purpose of the applications. In this case, the main problem is that it might be impossible to manually replace the battery of numerous IoT devices [4]. As one way to solve this problem, we consider the simultaneous wireless information and power transfer (SWIPT) technique via radio frequency (RF), which allows the wireless battery charging of battery-powered IoT devices without the help of a lot of manpower. Hereafter, it is assumed that harvested energy via SWIPT at each IoT device is exploited for its battery charging. Unlike conventional energy harvesting (EH) from renewable energy sources, such as solar and wind, RF-based EH is not affected by weather, location, time, etc. However, it is difficult to receive the required energy for operating IoT devices by RF-based EH owing to the low energy transfer efficiency and implementation limitations. To address these problems, multiple-input multiple-output (MIMO) has been applied in many studies related to SWIPT. Thus, the concept of MIMO in SWIPT-enabled networks has been regarded as a promising research direction.

In conventional SWIPT-enabled MIMO networks, the works of [5] and [6] proposed three types of receivers, that is, separated, power splitting, and time-switching receivers, and analyzed the fundamental rate–energy tradeoff with respect to

Manuscript received 13 September 2022; revised 14 November 2022 and 6 January 2023; accepted 23 January 2023. Date of publication 1 February 2023; date of current version 7 June 2023. This work was supported in part by the “Development of Polar Region Communication Technology and Equipment for Internet of Extreme Things (IoET)” funded by the Ministry of Science and ICT (MSIT), and in part by the Basic Science Research Program through the National Research Foundation of Korea, Ministry of Education, Science and Technology under Grant NRF2020R1F1A1074175. (Corresponding author: Yujae Song.)

Yongjae Kim is with the Maritime ICT Research and Development Center, Korea Institute of Ocean Science and Technology, Busan 49111, South Korea (e-mail: yongjaekim@kiost.ac.kr).

Bang Chul Jung is with the Department of Electrical Engineering, Chungnam National University, Daejeon 34134, South Korea (e-mail: bcjung@cnu.ac.kr).

Yujae Song is with the Department of Robotics Engineering, Yeungnam University, Gyeongsan 38541, South Korea (e-mail: ednb1008@gmail.com). Digital Object Identifier 10.1109/JIOT.2023.3241577

the above receivers. In [7], the resource allocation algorithm for energy efficiency was studied in orthogonal frequency division multiple access (OFDMAs) SWIPT systems with power-splitting receivers. Ng et al. [7] considered data multiplexing of different users on different subcarriers and the discrete sets of power-splitting ratios for practical systems. In [8], the optimal information decoding (ID) and EH mode switching rules were provided to optimize the outage probability/ergodic capacity versus harvested energy tradeoffs for time-varying co-channel interference environment. Ng et al. [9] investigated the beamforming techniques for secure communication in multiple-input–single-output (MISO) downlink systems in which a single desired information receiver and other ID or EH receivers, which can be eavesdroppers, are existed. In [10] and [11], a joint subcarrier and power allocation were jointly optimized to maximize the data rate and energy efficiency, respectively, for a collaborative communication scenario in which relay sensor nodes support data transmission.

Furthermore, joint beamforming and power-splitting optimization problems were considered in many research in SWIPT-enabled networks [12], [13], [14], [15], [16]. Xu et al. [12] and Al-Obiedollah et al. [13] considered cooperative SWIPT nonorthogonal multiple access (NOMAs) systems. In [12], a user who has good channel condition harvested the energy from the downlink signal at the first time slot and then relayed by using the harvested energy to help another user to improve the communication reliability at the second time slot. In [13], the overall harvested energy was maximized subject to data rate requirement of each user and successive interference cancellation (SIC) requirement for NOMA. Xu et al. [12] and Al-Obiedollah et al. [13] applied a semidefinite relaxation (SDR) technique and a sequential convex approximation (SCA) technique for nonconvex problem, respectively, and proposed iterative algorithms to solve the above problems with low complexity. In [14] and [15], the transmit power minimization by joint beamforming and power-splitting ratios under the signal-to-interference-plus-noise ratio (SINR) and harvested energy constraints was investigated for SWIPT-enabled networks. The work of Shi et al. [14] derived the sufficient and necessary condition for the feasibility of formulated nonconvex optimization problem and solved the problem by applying an SDR technique. The authors showed that the proposed technique can achieve global optimum, and also proposed two suboptimal solutions with lower complexity where zero-forcing (ZF) and SINR-optimal-based transmit beamforming schemes are applied, respectively. In [15], a novel relaxation method based on second-order cone programming (SOCP), which can guarantee a feasible solution with low complexity, was proposed for a joint beamforming and power-splitting optimization problem. Also, a primal-decomposition-based distributed algorithm for the problem was developed and it showed that it can reach the optimal solution to the joint beamforming and power-splitting optimization problem under two sufficient conditions. In [16], the joint beamforming and power-splitting problem was handled to minimize transmit power subject to the individual SINR and harvested energy constraints. Liao et al. [16] proposed a reverse convex

nonsmooth optimization algorithm for imperfect channel state information (CSI) condition. However, the single-input–single-output (SISO) or MISO channel was considered, or no interference from other cells was assumed in [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], and [16].

Most joint optimization problems of beamforming and power-splitting ratios have nonconvexity properties for multicell SWIPT MIMO networks and, thus, the proposed iterative algorithms can be employed in practical networks [17], [18], [19]. In [17], a generalized triangular decomposition (GTD)-based method which allows the transmitter to use the strongest eigen-channel jointly for EH and information exchange was proposed for SWIPT-MIMO networks. The optimal GTD structure that maximizes the data rate for a given power allocation and EH constraint was derived, and also the optimal subchannel and power allocation algorithm was proposed to minimize the total transmitted power. Kwon et al. [18] developed a novel joint design of transmit power allocation, beamforming, and receive power-splitting strategy for SWIPT downlink for mmWave channel. To maximize the weighted sum of the rate and harvested energy, an iterative algorithm was proposed containing the procedures for the RF beam alignment, baseband beamforming, and joint power allocation and power splitting. Under the limited feedback of channel information, the rate–energy region of the proposed strategy was analyzed and the performance in terms of achievable rate and harvested energy was verified by extensive simulations. In [19], the optimal transmit strategy by the optimal distribution of the transmit symbol vector that maximizes the average harvested energy for MISO, SIMO, and multiuser MIMO systems. Shanin et al. [19] formulated a nonconvex optimization problem and provided an optimal solution based on monotonic optimization and a low-complexity iterative algorithm to obtain a suboptimal solution which can achieve near-optimal performance.

With the recent development of deep learning technologies, deep learning-based algorithms have recently been considered to handle multivariable optimization problems in SWIPT-enabled networks [20], [21], [22], [23], [24], [25], [26]. In [20], a weighted sum of the data rate and harvested energy maximization problem was formulated for a time-switching receiver structure based on the Markov decision process (MDP), and a time interval control algorithm was proposed based on the reinforcement learning technique. Luo et al. [21] investigated a total transmit power minimization problem with rate and energy requirements for a SWIPT-enabled multicarrier NOMA network, which was solved using a deep belief network (DBN)-based algorithm. In [22], a low-complexity deep neural network-based algorithm was proposed to jointly optimize the transmit power and power-splitting ratio for energy efficiency in multicell multiuser SWIPT-enabled SISO networks. In [23], a double-deep deterministic policy gradient (DDPG)-deep double- Q -network (DDQN)-based algorithm was presented to optimize beamforming in full-duplex MIMO SWIPT networks. In [24], a deep learning-based algorithm was developed with near-optimal data rate performance in spite of the low complexity of the joint optimization problem

of power allocation and power-splitting ratio in multicarrier NOMA systems with SISO networks. In [25], constrained MDP-based power and subcarrier allocation problems were studied to maximize the energy efficiency for pattern-division multiple access SWIPT-enabled networks with time-switching receivers. The constrained MDP problem was transformed into an unconstrained MDP by applying the Lagrangian duality, and then solved using the deep Q -network (DQN) technique. Lee et al. [26] derived suboptimal solutions for time-switching and power-splitting receivers by using an iterative algorithm based on the asymptotic strong duality based on the *harvest-then-transmit* protocol. In addition, a deep neural network framework based on a supervised and unsupervised hybrid training strategy was developed using the above iterative algorithm results.

In addition, SWIPT techniques have been collaborated with future wireless communications such an intelligent reflecting surface (IRS)-assisted network, 6G terahertz communication, unmanned aerial vehicle (UAV) communication, NOMAs [27], [28], [29], [30]. The work of Xu et al. [27] investigated the resource allocation design for large IRS-assisted SWIPT systems. Compared to existing works assuming an overly simplified system model, the work adopted a physics-based IRS model and a nonlinear EH model, which can better capture the properties of practical IRS-assisted SWIPT systems. In [28], the integration of SWIPT and hybrid-NOMA using THz frequency bands was presented. Specifically, an optimal SWIPT-pairing scheme was suggested for the multilateral proposed system, which represents a considerable enhancement in energy and spectral efficiencies while improving the system specifications. The work of [29] presented the sum-rate maximization problem of IRS-empowered UAV SWIPT networks. Under the constraints of EH threshold, multiple optimization parameters (e.g., UAV trajectory, SIC decoding order, UAV transmit power allocation, power-splitting ratio, and IRS reflection coefficient) were jointly optimized. In [30], the concept of Age of Information (AoI) was considered to quantify the freshness of the data packets at the information receiver in the IRS-assisted SWIPT network. Under such consideration, the sum AoI optimization was studied while ensuring that the power transferred to EH users is greater than the demanded value.

Through an extensive literature survey on SWIPT-enable MIMO networks, it is found that the existing studies adopted a simplified power-splitting receiver model for the SWIPT, where only one power splitter is used regardless of the number of antennas in a receiver. Such limited use of power splitters at the receiver may limit the network performance. The existing studies did not consider multiple power splitters at the receiver since it is very challenging to find the online algorithm that jointly optimizes receive beamforming and power-splitting ratio vectors. This motivates us to investigate the online algorithm that simultaneously optimizes receive beamforming and power-splitting ratio vectors in this article.

A. Contributions

The main contributions of this article are as follows.

TABLE I
LIST OF MAIN NOTATIONS

Notation	Meaning
Upper case boldface	Matrix
Lower case boldface	Vector
\mathbf{A}^H	Conjugate transpose of matrix \mathbf{A}
$\ \cdot\ $	Two-norm of matrix
$\mathbb{E}[\cdot]$	Expectation
$\mathbb{C}^{M \times L}$	Set of $M \times L$ complex matrices
$\mathcal{O}[\cdot]$	Big O for computational complexity
\circ	Hadamard product

- 1) The existing studies considered a simple power-splitting receiver model for SWIPT, where only one power splitter is used regardless of the number of antennas at the receiver. On the other hand, in this article, we consider a generalized power-splitting receiver model, where each antenna at the receiver is equipped with its power splitter. Such a receiver design can provide a greater degree of freedom to improve performance in terms of EH. This advantage contributes to increasing the operating time of IoT devices without battery exchange, which is crucial for a massive IoT scenario like a smart city.
- 2) For each IoT device, we formulate an optimization problem as a Markov game determining two different action sets simultaneously. One is for a receive beamforming vector, and the other is for a power-splitting ratio vector. The objective of the problem is to maximize the harvested energy of each IoT device while satisfying its data-rate requirement.
- 3) To solve the problem, a DDPG-based online learning algorithm is proposed. This enables each IoT device to jointly determine to appreciate beamforming and power-splitting ratio vectors in real time. Furthermore, each IoT device can implement the proposed algorithm in a distributed manner using only its local CSI. Therefore, cooperation and information exchange among the BSs and IoT devices are unnecessary when performing the proposed algorithm on IoT devices.

B. Organization

The remainder of this article is organized as follows. Section I explains existing works on SWIPT techniques in wireless networks and their limitations. In Section II, we formally present our system model. Section III presents the proposed power-splitting receiver design for IoT device in a smart city with SWIPT. In Section IV, we first formulate a joint EH and ID optimization problem for IoT-enabled smart city, and then propose a distributed online learning algorithm to solve the problem. The performance of the proposed algorithm is evaluated by extensive simulations in Section V. Finally, conclusions are drawn in Section VI.

C. Notations

Throughout this article, main notations in this article are summarized in Table I.

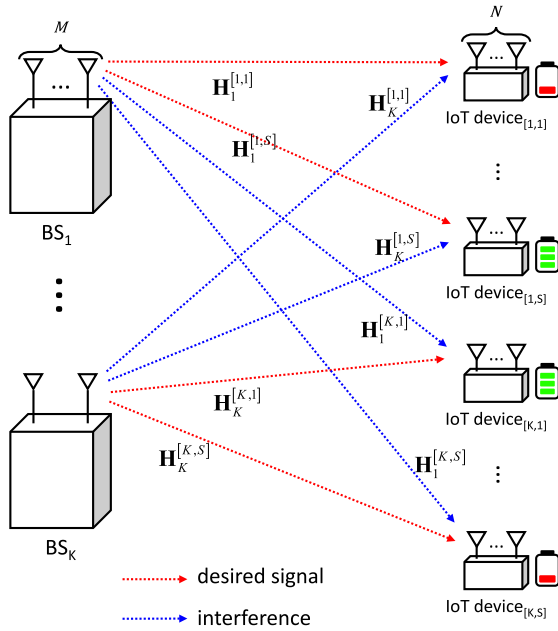


Fig. 1. Illustration of system model in a smart city.

II. SYSTEM MODEL

This study considers a smart city, as shown in Fig. 1, where a variety of IoT devices have been deployed to help solve the problems of urban pollution, traffic management, and environmental protection in an urban city. Depending on service types, each IoT device has a different scheduling interval to transmit. In this work, *IoT devices* refer to objects equipped with sensors (e.g., temperature sensors, fire sensors, traffic cameras, humidity sensors, etc.) and communication modules depending on their purposes, such as car, transmission line tower, house, etc. Depending on service types, each IoT device has a different scheduling interval. IoT devices are assumed to have three functionalities: 1) measurement and transmission of sensing data; 2) simultaneous wireless information and power reception; and 3) adaptive energy management of own battery system.

To support IoT devices in a smart city, we adopted K -cell SWIPT-enabled MIMO downlink networks. Each cell has a BS with M antennas and a set of IoT devices S ($S \leq M$) with L rectennas.¹ All IoT devices can harvest energy in their batteries from anonymous RF signals and decode information simultaneously by their rectennas. It is assumed that each BS transmits a single data stream to each IoT device, and the time-division duplex (TDD) is considered as in 5G new radio (NR) [31]. The channel between BS k to IoT device j associated to BS i is denoted by $\mathbf{H}_k^{[i,j]} \in \mathbb{C}^{L \times M}$, where $i, k \in \mathcal{K} = \{1, \dots, K\}$, and $j \in \mathcal{S} = \{1, \dots, S\}$. We reflect time-varying channel as well as time-invariant frequency-flat fading (i.e., the channel coefficients are constant during a transmission block). For time-varying channel, the first-order Gauss–Markov process is adopted to model the relationship

¹The rectenna is a special type of antenna, consisting of a diode and a low-pass filter, which converts RF signals into a direct current (dc) signal to recharge the battery. Hereafter, the rectennas and receive antennas can be used interchangeably.

between two successive time slots of small-scale Rayleigh fading as follows [32], [33]:

$$h_k^{[i,j]}(t) = \omega \cdot h_k^{[i,j]}(t-1) + \delta \quad (1)$$

where $h_k^{[i,j]}(t)$ denotes an element² of $\mathbf{H}_k^{[i,j]}$ at time t , and it is a complex Gaussian random variable with zero mean and unit variance. In addition, ω is the correlation coefficient of two successive small-scale Rayleigh fading realizations, and δ is a complex Gaussian random variable with zero mean and a variance of $1 - \omega^2$. We assumed that all elements of $\mathbf{H}_k^{[i,j]}$ had the same time correlation coefficient ω . The correlation coefficient was determined using Jake's statistical model for the fading channel [34] as follows:

$$\omega = J_0(2\pi f_D T) \quad (2)$$

where J_0 and T denote the zero-order Bessel function and time interval for data transmission, respectively. In addition, $f_D = \nu f_c / c$ denotes the maximum Doppler frequency, where ν is the user velocity, f_c is the carrier frequency, and $c = 3 \times 10^8$ m/s is the velocity of light. The CSI can be obtained by the transmitted CSI reference signals (CSI-RSs) for downlink channel estimation from all BSs. To estimate the CSI with neighboring BSs, multiple CSI-RS processes, which are specified in LTE Release 16 for coordinated multiple point (CoMP) operations [35], can be exploited.

The received signal at IoT device j in BS i is given as follows:

$$\begin{aligned} \mathbf{y}^{[i,j]} &= \sum_{k=1}^K \mathbf{H}_k^{[i,j]} \mathbf{V}_k \mathbf{x}_k + \mathbf{z}^{[i,j]} \\ &= \mathbf{H}_i^{[i,j]} \mathbf{V}_i \mathbf{x}_i + \sum_{k=1, k \neq i}^K \mathbf{H}_k^{[i,j]} \mathbf{V}_k \mathbf{x}_k + \mathbf{z}^{[i,j]} \\ &= \underbrace{\mathbf{H}_i^{[i,j]} \mathbf{V}_i^{[i,j]} \mathbf{x}^{[i,j]}}_{\text{desired signal}} + \underbrace{\mathbf{H}_i^{[i,j]} \sum_{s=1, s \neq j}^S \mathbf{V}_s^{[i,s]} \mathbf{x}^{[i,s]}}_{\text{intracell interference}} \\ &\quad + \underbrace{\sum_{k=1, k \neq i}^K \mathbf{H}_k^{[i,j]} \mathbf{V}_k \mathbf{x}_k}_{\text{intercell interference}} + \mathbf{z}^{[i,j]} \end{aligned} \quad (3)$$

where $\mathbf{V}_k \in \mathbb{C}^{M \times S}$ denotes the ZF filtering-based transmit beamforming matrix. Also, $\mathbf{x}_k \in \mathbb{C}^{S \times 1}$ and $\mathbf{z}^{[i,j]} \in \mathbb{C}^{L \times 1}$ are the transmit signal vector and additive noise consisting of independent and identically distributed (i.i.d.) complex Gaussian with zero mean and the variance of N_0 . By adopting ZF-based filtering at a transmitter side (i.e., each BS), intracell interference from adjacent IoT devices can be perfectly eliminated [36], [37]. As such, the considered network can be equivalent with a multicell and single-user network, assuming that ZF-based transmit beamforming is exploited without loss of generality. Therefore, the received signal at IoT device

²The indices for the channel matrix $\mathbf{H}_k^{[i,j]}$ elements are omitted for simplicity.

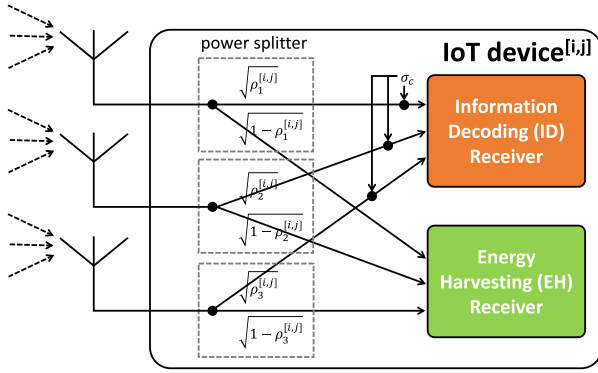


Fig. 2. Illustration of the concept of receiver hardware design for each IoT device: $L = 3$.

j in BS i (3) can be simplified as follows:

$$\mathbf{y}^{[i,j]} = \sum_{k=1}^K \mathbf{h}_k^{[i,j]} x_k + \mathbf{z}^{[i,j]} = \mathbf{h}_i^{[i,j]} x_i + \sum_{k=1, k \neq i}^K \mathbf{h}_k^{[i,j]} x_k + \mathbf{z}^{[i,j]} \quad (4)$$

where $\mathbf{h}_k^{[i,j]} \in \mathbb{C}^{L \times 1}$ is the channel vector between BS k and IoT device j associated to BS i , and x_k is the transmit symbol from BS k . Hereafter, our focus is to properly utilize intercell interference from neighbor BSs for maximizing the harvested energy of each IoT device while satisfying its data rate requirement.

III. POWER-SPLITTING RECEIVER DESIGN AND PROBLEM FORMULATION FOR IOT DEVICES

A. Design of Power-Splitting Receiver for Each IoT Device

To harvest energy from anonymous RF signals, there are several receiver structures, such as time switching, and power splitting [5], [6]. This work adopts a power-splitting receiver architecture with multiple receive antennas and multiple power splitters similar to [18] and [38]. Fig. 2 illustrates the basic concept of receiver hardware design for each IoT device, where $\rho_l^{[i,j]} \in [0, 1]$ stands for the power-splitting ratio of the l th antenna element for IoT device j in BS i and σ_c stands for the conversion noise power introduced by converting the RF passband signal into a baseband signal. As shown in Fig. 2, each independent power splitter is connected to each antenna. Under the presented receiver structure, the power of the received signals can be divided for EH and ID, according to the power-splitting ratio $\rho_l^{[i,j]}$ for each antenna. Therefore, the performance of the EH and ID depends on the power-splitting ratio as well as the receive beamforming strategy.

B. Problem Formulation

Each IoT device can simultaneously design its receive beamforming vector and power-splitting ratio vector to maximize the performance in terms of harvested energy or achievable data rate. Let $\mathbf{u}^{[i,j]} \in \mathbb{C}^{L \times 1}$ denote the receive beamforming vector of IoT device j in BS i with a unit-norm constraint, that is, $\|\mathbf{u}^{[i,j]}\|^2 = 1$. In addition, let $\boldsymbol{\rho}^{[i,j]} \in \mathbb{R}^{L \times 1}$ denote the power-splitting ratio vector of IoT device j in BS i

with constraints $0 \leq \rho_l^{[i,j]} \leq 1$ for $l \in \{1, \dots, L\}$, where $\rho_l^{[i,j]}$ is the l th element of $\boldsymbol{\rho}^{[i,j]}$. The objective of this study is to design the receive beamforming vector and power-splitting ratio vector that maximize the harvested energy from the downlink signals of all BSs, while guaranteeing the predetermined data rate constraint. In the downlink network scenario of a smart city with IoT devices, BSs require a small number of bits for downlink control signals, while IoT devices transmit several Mbps for the uplink case. Accordingly, we aim to maximize the harvested energy of IoT device j in BS i with data rate constraint by adjusting receive beamforming vector and power-splitting ratio vector for the considered system model. To this end, the receive beamforming vector and power-splitting ratio vector of IoT device j in BS i can be obtained by solving the following optimization problem:

$$(P1) \max_{\mathbf{u}, \boldsymbol{\rho}} Q^{[i,j]}(\boldsymbol{\rho}) = \max_{\mathbf{u}, \boldsymbol{\rho}} \zeta \left\| \mathbf{y}_{\text{EH}}^{[i,j]}(\boldsymbol{\rho}) \right\|^2 \quad (5)$$

$$\text{s.t. } R^{[i,j]}(\mathbf{u}, \boldsymbol{\rho}) \geq R_{\text{th},j} \quad (6)$$

$$\left\| \mathbf{u}^{[i,j]} \right\|^2 = 1 \quad (7)$$

$$0 \leq \rho_l^{[i,j]} \leq 1 \quad (8)$$

where ζ is the conversion efficiency, $\mathbf{y}_{\text{EH}}^{[i,j]}$ is the received signal for EH after the power-splitting operation, and $R_{\text{th},j}$ is the required data rate of IoT device j . In addition, $R^{[i,j]}$ denotes the achievable data rate of IoT device j in BS i , which will be described in Section III-C. To solve the optimization problem, i.e., P1, an online learning algorithm was developed, and the details have been explained in Section IV.

C. Downlink Data Transmission, ID and EH

After deciding the receive beamforming vector and power-splitting ratio at the IoT devices, each BS transmits the downlink data signal to its associated IoT devices. Based on the receive beamforming and power-splitting ratio vectors, each IoT device performs ID and EH at the ID and EH receivers, respectively, as shown in Fig. 1. From (4), the received signals are divided into ID and EH parts according to the power-splitting ratio as follows:

$$\begin{aligned} \mathbf{y}_{\text{ID}}^{[i,j]} &= \left(\boldsymbol{\rho}^{[i,j]} \right)^{\frac{1}{2}} \circ \mathbf{y}^{[i,j]}, \\ \mathbf{y}_{\text{EH}}^{[i,j]} &= \left(\mathbf{1} - \boldsymbol{\rho}^{[i,j]} \right)^{\frac{1}{2}} \circ \mathbf{y}^{[i,j]} \end{aligned} \quad (9)$$

where \circ denotes the Hadamard product. In the ID receiver, the received signal after receiving beamforming can be expressed as follows:

$$\begin{aligned} \tilde{\mathbf{y}}_{\text{ID}}^{[i,j]} &= \mathbf{u}^{[i,j]H} \mathbf{y}_{\text{ID}}^{[i,j]} \\ &= \mathbf{u}^{[i,j]H} \left[\left(\boldsymbol{\rho}^{[i,j]} \right)^{\frac{1}{2}} \circ \left(\sum_{k=1}^K \mathbf{h}_k^{[i,j]} x_k + \mathbf{z}^{[i,j]} \right) \right] \\ &= \underbrace{\tilde{\mathbf{u}}^{[i,j]H} \mathbf{h}_i^{[i,j]} x_i}_{\text{desired signal}} + \underbrace{\sum_{k=1, k \neq i}^K \tilde{\mathbf{u}}^{[i,j]H} \mathbf{h}_k^{[i,j]} x_k + \tilde{\mathbf{u}}^{[i,j]H} \mathbf{z}^{[i,j]}}_{\text{intercell interference}} \end{aligned} \quad (10)$$

where $\tilde{\mathbf{u}}^{[i,j]H} = (\rho^{[i,j]})^{(1/2)} \circ \mathbf{u}^{[i,j]}$. From (10), the achievable data rate of IoT device j in BS i can be computed as

$$R^{[i,j]} = \log_2 \left(1 + \frac{|\tilde{\mathbf{u}}^{[i,j]H} \mathbf{h}_i^{[i,j]} x_i|^2}{\sum_{k=1, k \neq i}^K |\tilde{\mathbf{u}}^{[i,j]H} \mathbf{h}_k^{[i,j]} x_k|^2 + |\tilde{\mathbf{u}}^{[i,j]H} \mathbf{z}_i|^2 + \sigma_c^2} \right). \quad (11)$$

In the EH receiver, the harvested energy of IoT device j in BS i can be obtained as follows:

$$Q^{[i,j]} = \zeta \left\| \left((\mathbf{1} - \rho_i)^{\frac{1}{2}} \right) \circ \left(\sum_{k=1}^K \mathbf{h}_{k,i} x_k \right) \right\|^2. \quad (12)$$

In (12), because the energy harvested from the background noise at the EH receiver is negligible, it can be ignored for simplicity. Based on (11) and (12), it is identified that there is a tradeoff relationship between harvested energy and achieved data rate. Thus, to maximize the data rate performance under the proposed framework, it is set to $\rho_i^{[i,j]} = 0$, which means no EH. Furthermore, to achieve more peak data rate, we can also adopt massive MIMO and high modulation techniques, which are introduced as enabling features to achieve 30 bps/Hz of a peak spectral efficiency in 5G White Paper [39]. These techniques can be implemented without any modification of the presented mathematical framework.

IV. JOINT EH AND ID OPTIMIZATION FOR IOT-ENABLED SMART CITY WITH SWIPT

Problem (P1) is a constrained optimization problem with two different types of action vectors, which makes it difficult to obtain optimal solutions to the problem in real time. Nevertheless, determining the two vectors at each IoT device in real time is very critical. This is because we consider a scenario that IoT devices are associated with a smart BS (i.e., cellular networks) which requires a very short scheduling interval (e.g., in the scale of millisecond). To interoperate IoT devices with cellular BSs, the proposed algorithm should be operated in real time. Therefore, we propose an online learning algorithm which enables each IoT device to jointly determine the appropriate receive beamforming and power-splitting ratio vectors by learning the relationship between an action and its reward.

From (11), it can be observed that receive beamforming vector \mathbf{u} and power-splitting ratio vector ρ affect the same function (i.e., achievable data rate), such that they should be determined at the same time. Moreover, the allowable ranges of the elements of two vectors are different. Specifically, the real and imaginary parts of the elements of the receive beamforming vector range from -1 to 1 , whereas those of the power-splitting vector range from 0 to 1 . For this reason, we formulate our problem as a Markov game which determines two different types of action sets as follows.

- 1) *First Action Set*: Receive beamforming vector \mathbf{u} with elements that are a complex number whose real and imaginary parts range from -1 to 1 .
- 2) *Second Action Set*: Power-splitting ratio vector ρ with elements that are a real number ranging from 0 to 1 .

That is, each IoT device should solve the Markov game to simultaneously determine different types of action sets. Since the proposed online strategy is based on a deep reinforcement learning (DRL) algorithm, we first describe our work as an MDP framework as follows.

A. MDP Definition

1) *State*: The state space for each IoT device includes channel qualities between IoT device j and all BSs (i.e., \mathcal{K}) in the considered networks

$$\mathbf{S}_j = \left\{ \mathbf{h}_1^{[i,j]}, \dots, \mathbf{h}_j^{[i,j]}, \dots, \mathbf{h}_K^{[i,j]} \right\}. \quad (13)$$

Note that each IoT device uses the same state space for determining the two action vectors.

2) *Action*: Each IoT device needs to determine two different types of action vectors. Specifically, the first action set, i.e., action $\hat{\mathbf{u}}$, is exploited to decide the real and imaginary parts of elements in receive beamforming vector \mathbf{u} . As such, action $\hat{\mathbf{u}}$ is a vector of size $2L$ whose elements are continuous and ranges from -1 to 1 . The second action set, i.e., action ρ , is to decide power-splitting ratio vector ρ of size L whose elements are continuous and ranges from 0 to 1 . Thus, the action space of IoT device j can be presented as

$$\mathbf{A}_j = \left\{ \begin{array}{l} \hat{\mathbf{u}} = \left\{ \begin{array}{l} u_1, \dots, u_L, u_{L+1}, \dots, u_{2L} \\ \text{real part} \quad \text{imaginary part} \end{array} \right\} \\ \rho = \{\rho_1, \dots, \rho_L\} \end{array} \right\} \quad (14)$$

where for all $u_l \in [-1, 1]$ and for all $\rho_l \in [0, 1]$. If action $\hat{\mathbf{u}}$ is achieved, the receive beamforming vector of IoT device j in BS i , i.e., $\mathbf{u}^{[i,j]}$, can be easily obtained by the normalization.

3) *Reward*: The objective of each IoT device is to maximize the amount of energy harvested from the downlink signals of all BSs without compromising its data rate requirement. Thus, we can define the reward of each IoT device, which is affected by the determined two action vectors (i.e., \mathbf{u} and ρ) as follows:

$$r_j^{[i,j]}(\mathbf{u}, \rho) = \begin{cases} Q^{[i,j]}, & \text{if } R^{[i,j]}(\mathbf{u}, \rho) \geq R_{\text{th}j} \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

According to (15), if the data rate constraint is satisfied, then reward equal to the harvested energy can be obtained; otherwise, there is no reward. With (15), we can present the accumulated discounted reward, as expressed by

$$\mathcal{R}_j^{[i,j]} = \sum_{t=1}^T \gamma^{t-1} r_z^{[i,j]}(t) \quad \forall z \quad (16)$$

where γ is the discount factor and T is the time horizon.

B. Proposed Algorithm

To obtain a solution to such a Markov game, we propose a double-DDPG-based online learning algorithm which enables each IoT device to achieve the receive beamforming and power-splitting ratio vectors in real time. The structure of the proposed double-DDPG-based DRL algorithm is illustrated in Fig. 3. Under the proposed algorithm, each IoT device

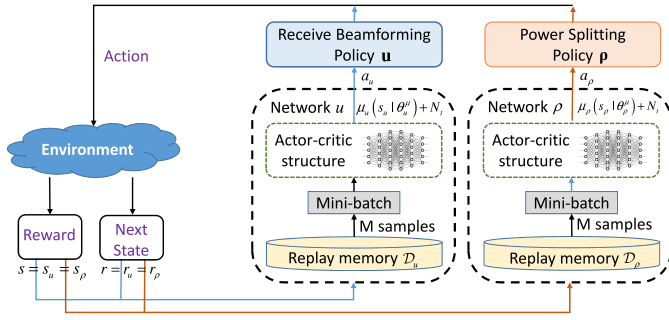


Fig. 3. Structure of the proposed double-DDPG-based online learning algorithm.

constructs two DDPG networks to decouple two action vectors instead of employing a single agent in the conventional DDPG network [40]. In particular, as shown in Fig. 3, the first and second networks are exploited for determining receive beamforming and power-splitting ratio vectors, respectively.

For ease of description on the proposed algorithm, we focus on an IoT device³ with two DDPG networks, and each DDPG network is denoted as parameter $z \in \mathcal{Z} = \{u, \rho\}$, where \mathcal{Z} is the set of the DDPG networks.

Consider a Markov game with two DDPG networks with deterministic policy μ_z with regard to parameter θ_z^μ , where $z \in \mathcal{Z}$. Thus, we can define the gradient of the accumulated discounted reward for network z , that is, $J_z = \mathbb{E}[\mathcal{R}_z]$ as follows:

$$\nabla_{\theta_z^\mu} J_z = \mathbb{E}_{s_z, a_z \sim \mathcal{D}_z} \left[\nabla_{\theta_z^\mu} \mu_z(s|\theta_z^\mu)|_{s=s_z} \nabla_a Q_z(s, a|\theta_z^Q)|_{s=s_z, a=\mu_z(s_z)} \right] \quad (17)$$

where $a_z \in \mathbf{A}$ and $s_z \in \mathbf{S}$, and \mathcal{D}_z is the experience replay buffer for network z that contains tuples (s_z, a_z, r_z, s_z') .

In addition, $Q_z(s_z, a_z|\theta_z^Q)$ refers to the action-value function that assumes the actions of two networks as input, that is, a_u and a_ρ , in addition to the state information s_z , and outputs the Q -value for network z . The action-value function Q_z is updated by minimizing the loss function L_z as follows:

$$\mathcal{L}_z = \mathbb{E}_{s_z, a_z, r_z, s_z' \sim \mathcal{D}_z} \left[\left(Q_z(s_z, a_z|\theta_z^Q) - y \right)^2 \right] \quad (18)$$

$$y = r_z + \gamma Q_z'(s_z', \mu_z'(s_z'|\theta_z^{\mu'})|\theta_z^{Q'})$$

where Q_z' is the target action-value function with respect to the set of target policies μ_z' with delayed parameters $\theta_z^{\mu'}$.

With (17) and (18), we propose a double-DDPG-based joint receive beamforming and a power-splitting vector decision algorithm, which is described in Algorithm 1. In Algorithm 1, \mathcal{N} is the noise process for constructing an exploration policy, φ is a predetermined value for the repetitive initialization of \mathcal{N} , and τ is the weight for soft target updates. Note that, as illustrated in Fig. 3, when deciding two action sets using two different DDPG networks, the same state information is used (i.e., $s = s_u = s_\rho$), and the determined two action vectors are used to evaluate the same reward function (i.e., $r = r_u = r_\rho$).

³The index for IoT device is omitted for simplicity.

Algorithm 1: Double-DDPG-Based Online Learning Algorithm for Joint Receive Beamforming and Power-Splitting Ratio Vectors

- 1 Randomly initialize critic network $Q_z(s_z, a_z|\theta_z^Q)$ and actor $\mu_z(s_z|\theta_z^\mu)$ with weights θ_z^Q and θ_z^μ .
 - 2 Initialize target network Q_z' and μ_z' with weights $\theta_z^{Q'} \leftarrow \theta_z^Q$ and $\theta_z^{\mu'} \leftarrow \theta_z^\mu$.
 - 3 Initialize replay buffer \mathcal{D}_z .
 - 4 Initialize a random process \mathcal{N} for action exploration.
 - 5 **for** $t = 1$ to T **do**
 - 6 For each agent, select action $a_z = \mu_z(s_z|\theta_z^\mu) + N_t$ w.r.t. the current policy and exploration.
 - 7 Execute a_z and observe reward r_z and new state s_z' .
 - 8 Store (s_z, a_z, r_z, s_z') in replay buffer \mathcal{D}_z .
 - 9 **for** $z \in \mathcal{Z}$ **do**
 - 10 Sample a random minibatch of M samples $(s^\phi, a^\phi, r^\phi, s'^\phi)$ in \mathcal{D}_z .
 - 11 Set $y^\phi = r_z^\phi + \gamma Q_z'(s'^\phi, \mu_z'(s'^\phi|\theta_z^{\mu'})|\theta_z^{Q'})$.
 - 12 Update critic by minimizing the loss:

$$L_z = \frac{1}{M} \sum_b (y^\phi - Q_z(s^\phi, a^\phi|\theta_z^Q))^2$$
 - 13 Update the actor policy using the sampled policy gradient: $\nabla_{\theta_z^\mu} J_z \approx \frac{1}{M} \sum_\phi \nabla_a Q_z(s, a|\theta_z^Q)|_{s=s^\phi, a=\mu_z(s^\phi)} \nabla_{\theta_z^\mu} \mu_z(s|\theta_z^\mu)|_{s^\phi}$.
 - 14 **end**
 - 15 Update the target network parameter for each agent z :

$$\theta_z^{Q'} \leftarrow \tau \theta_z^Q + (1 - \tau) \theta_z^{Q'}$$
 and $\theta_z^{\mu'} \leftarrow \theta_z^\mu + (1 - \tau) \theta_z^{\mu'}$.
 - 16 **if** $T \% \varphi = 0$ **then**
 - 17 Initialize a random process \mathcal{N} for action exploration.
 - 18 **end**
 - 19 **end**
-

Similar to existing learning research [20], [21], [22], [23], [24], [25], [26], the procedure of the proposed double DDPG-based online learning algorithm can be divided into training and execution stages. First, in the training stage, each IoT device learns its double DDPG networks, until the performances of the networks converge. Once the performances of the networks finally converge, each IoT device is ready to enter the execution state in which the trained DDPG networks are used to determine receive beamforming and power-splitting ratio vectors without the update of the networks, which enable each IoT device to implement the proposed algorithm in real time. In this sense, the *online* algorithm in this work means that the proposed algorithm can implement in real time in the execution stage. To identify the validity of real-time processing in the execution stage, time computational complexity analysis of training and execution stages in the proposed algorithm will be conducted in the following section.

C. Computational Complexity Analysis

We analyze time computational complexity of the proposed algorithm using big O notation denoted by $O[\cdot]$. Let N_i^z and N_m^z denote the number of the layers of DDPG network $z \in \{u, \rho\}$

TABLE II
LIST OF STATIC NETWORK PARAMETERS

Parameter	Value
Carrier frequency f_c	2 GHz
Time interval for data transmission T	10 ms
Conversion noise power σ_c^2	-32 dBm
Channel correlation coefficient values ω	0.93 and 0.1
Received signal-to-noise-ratio (SNR)	15 dB
Number of IoT devices	15
Number of rectennas for IoT device L	3

TABLE III
LIST OF HYPERPARAMETERS FOR DDPG NETWORKS

Parameter	Value
Minibatch size	128
Reply buffer size	10^6
Discount factor	0.99
Learning rate of actor	10^{-4}
Learning rate of critic	3×10^{-4}
Soft update rate of target parameters	2×10^{-1}

and the number of neurons in the m th layer of network z , respectively. In the training stage, the computational complexity of a single network for both evaluating and updating in a time slot can be presented by $O[N_b^z (\sum_{m=1}^{N_i^z-1} N_m^z N_{m+1}^z)]$, where N_b^z is the mini-batch size of network z [41]. If an IoT device operates double-DDPG networks serially, the total training computational complexity of the proposed algorithm is $O[T_{cv} \sum_{z \in \{u, \rho\}} N_b^z (\sum_{m=1}^{N_i^z-1} N_m^z N_{m+1}^z)]$, where T_{cv} is the number of time slots until performances of the two network converge. In the test stage, the computational complexity of the proposed algorithm in each time slot can be dramatically reduced to $O[\sum_{z \in \{u, \rho\}} (\sum_{m=1}^{N_i^z-1} N_m^z N_{m+1}^z)]$. This is because once the performances of the networks finally converge, we do not need iterations for the training of the networks.

V. SIMULATION RESULTS

In this section, we have evaluated the performances of the proposed algorithm described in Algorithm 1, and then compare them with those of the existing algorithms to identify the validity of the proposed algorithm.

A. Simulation Parameters

The static system parameters adapted for the numerical performance evaluations are summarized in Table II as follows. For performance evaluations, we considered a total of 15 IoT devices, and their performance were averaged out in figures. For simplicity, all IoT devices had the same data rate requirement, that is, $R_{th,j} = R_{th}$.

On learning environment, the actor and critic networks in the proposed double-DDPG-based algorithms are a fully connected neural network with two hidden, where first hidden and second hidden layers contain 512 and 256 neurons, respectively. Other learning hyperparameters are summarized in Table III.

B. Benchmark Algorithms

For performance comparison, the following existing algorithms are considered: a uniform power splitting with random beamforming (UPS-RBF) algorithm [42], an antenna selection technique with ZF (AS-ZF) algorithm [43], and a generalized downlink interference alignment with receive antenna partitioning (GDIA-RAP) algorithm [44]. In the UPS-RBF algorithm, the power-splitting ratios for a receiver are fixed across all antennas such that $\rho_l^{[i,j]} = \rho^{[i,j]}$ for all l , and the random beamforming strategy is exploited. It is known that multiuser diversity gain can be easily achieved by randomly adjusting transmit power and phase, i.e., random beamforming or opportunistic beamforming, in slow fading environment [45]. In the AS-ZF algorithm, receive antennas are selected for ID and EH to maximize the weighted sum of data rate and harvested energy, and the receive beamforming are constructed based on ZF filter. The GDIA-RAP algorithm, which is extended from AS-ZF, provides the optimal receive antenna configuration to maximize the weighted sum of data rate and harvested energy. In the GDIA-RAP, the receive beamforming vectors are constructed by interference alignment concept and ZF is exploited for transmit beamforming vectors.

C. Simulation Results

Fig. 4 illustrates the results of training stage for learning according to the iterations when $K = 2$ and $R_{th} = 0.5$ bps/Hz. Fig. 4(a) and (b) presents the moving averages of harvested energy and the corresponding data rate at IoT device, respectively. As illustrated in Fig. 4(a), it is identified that the harvested energy of the proposed algorithm converges as the number of iterations increases, and the proposed algorithm continuously performs better than the existing algorithms after a certain number of iteration. The reason why the harvested energy of the proposed algorithm converges while decreasing is to satisfy the data rate requirement while increasing the data rate which is shown in Fig. 4(b). Fig. 4(b) also shows that after a certain number of iterations, the achievable data rates of the proposed algorithm are higher than the data rate threshold, R_{th} . In particular, the maximal harvested energy while meeting the data rate requirement (i.e., 0.5 bps/Hz) is achieved between iterations $1.5 \sim 2 \times 10^4$, but the performances of the proposed algorithm converge when the moving average of the data rate approaches about 0.65 bps/Hz. This is because although the moving average value of data rate is above 0.5 bps/Hz, there might be a possibility that occurs cases where the data rate requirement is not satisfied in a time slot. To prevent this (that is, to satisfy the data rate requirement in every time slot as possible), the proposed algorithm learns to satisfy “ $R_{th} + \text{bias}$.” In this simulation environment, the bias is about 0.15 bps/Hz. Furthermore, in case of the FPS-RBF algorithm, the harvested energy for low $\rho^{[i,j]}$ value, i.e., $\rho^{[i,j]} = 0.3$, is greater than that for high $\rho^{[i,j]}$ value, i.e., $\rho^{[i,j]} = 0.6$. This is natural because a decrease in $\rho^{[i,j]}$ means a receiver allocates more power to harvest energy not decode information. Based on the results shown in Fig. 4, we demonstrate that the proposed algorithm can improve the performance in terms of harvested energy while satisfying the data rate constraint.

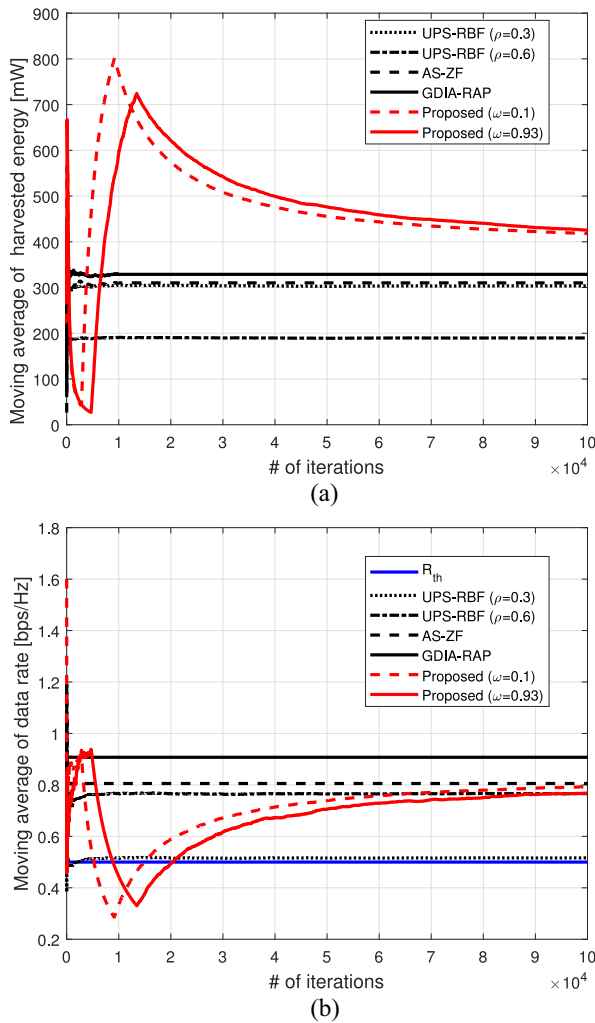


Fig. 4. Training results of the proposed and existing algorithms according to the number of iterations when $K = 2$ and $R_{th} = 0.5$ bps/Hz: moving averages of (a) harvested energy and (b) data rate.

Fig. 5 describes the average harvested energy and the corresponding average data rate according to the change of number of BSs. The total number of iterations for learning in the simulation is 10^5 . In Fig. 5(a), it is identified that in case of the existing algorithms, the harvested energy linearly increases as the number of BSs increases. This is because interference from the other BSs that can be used for EH increases linearly with the number of BSs. Likewise, the EH performance of the proposed algorithm improves as the number of BSs increases. It can be also observed that as the number of BSs increases, the EH performance gap between the proposed and existing algorithms increases regardless of the channel environment. For example, compared with the existing real-time algorithms, the proposed algorithm achieves minimum 50% harvested energy performance improvement in case of the number of BSs = 5 [as illustrated in Fig. 5(a)] while supporting the data rate requirement [as illustrated in Fig. 5(b)]. Unlike the case of harvested energy, Fig. 5(b) shows that the achievable data rate deteriorates as the number of BSs increases, because there is considerable intercell interference when there are numerous BSs. Moreover, it is identified that the data rate requirement

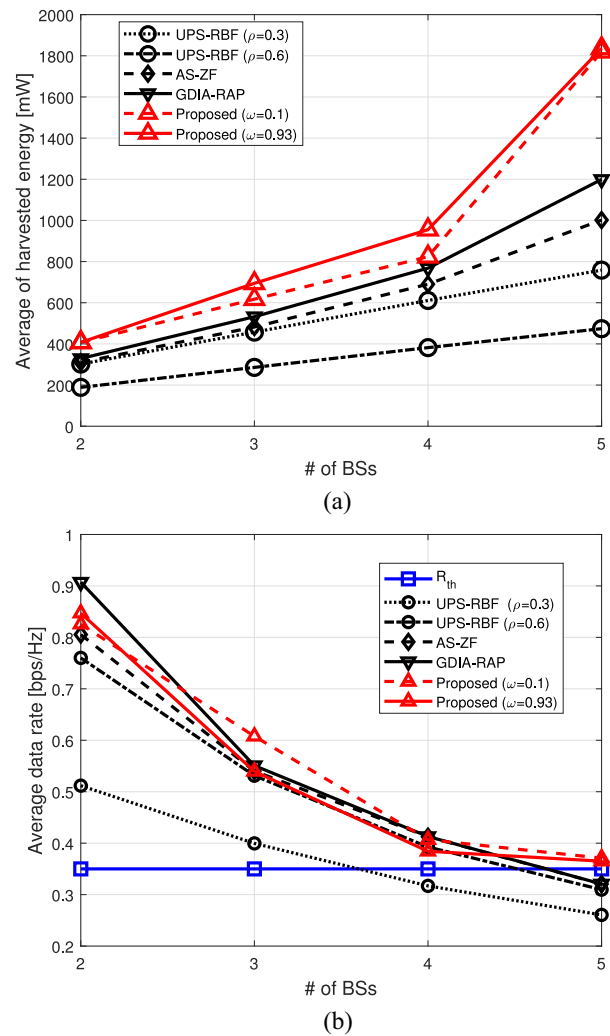


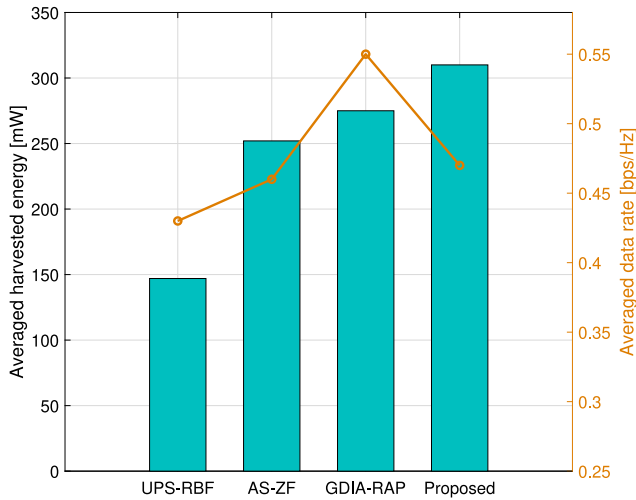
Fig. 5. Performance comparisons between the proposed and existing algorithms under a change in the number of BSs when $R_{th} = 0.35$ bps/Hz: averages of (a) harvested energy and (b) data rate.

cannot be achieved for all ρ values in the existing algorithms when the number of BSs is five. The results show that it is difficult to effectively manage interference from other BSs in the existing algorithm. On the other hand, the proposed algorithm always satisfies the data rate constraint, that is, $R^{[i,j]} \geq R_{th}$ in both correlated and uncorrelated channel environments. Therefore, based on the results from Fig. 5, we can conclude that the proposed algorithm outperforms the existing algorithms in terms of the EH performance while satisfying the data rate constraint regardless of the number of BSs in the network. In other words, the proposed algorithm can be well operated under a severe interference environment like a smart city where there are a lot of interferers such as IoT devices.

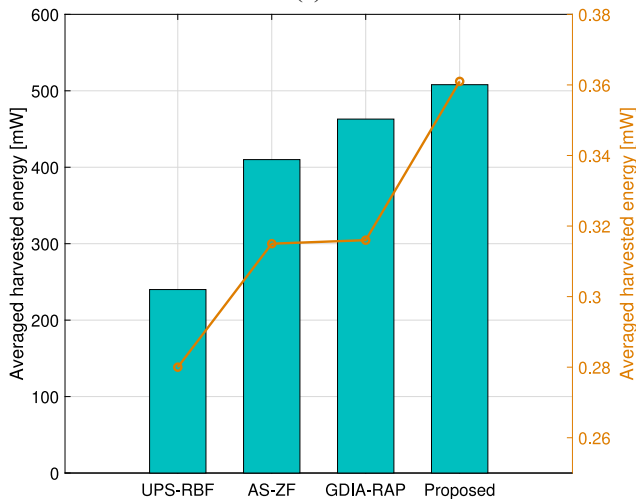
Table IV presents the harvested energy and the corresponding time computational complexity of the proposed algorithms under different power-splitting receiver frameworks (i.e., single power splitter versus multiple power splitters in a receiver). Table IV shows that the multiple power splitter framework increases the harvested energy (while satisfying data rate requirement) whereas the computational complexity is also

TABLE IV
PERFORMANCES OF THE PROPOSED ALGORITHM UNDER DIFFERENT POWER-SPLITTING RECEIVER FRAMEWORKS:
SINGLE POWER SPLITTER VERSUS MULTIPLE POWER SPLITTERS IN A RECEIVER

	Harvested energy [mW]	Computational complexity
Single power splitter	204	$O\left[\sum_{m=1}^3 N_m^u N_{m+1}^u\right] + O\left[N_1^\rho (N_2^\rho)^2 (N_3^\rho)^2\right]$
Multiple power splitters	238	$O\left[\sum_{m=1}^3 N_m^u N_{m+1}^u\right] + O\left[N_1^\rho (N_2^\rho)^2 (N_3^\rho)^2 N_4^\rho\right]$



(a)



(b)

Fig. 6. Performance comparisons between the proposed and existing algorithms under the uncorrelated channel scenario when $R_{th} = 0.35$ bps/Hz: (a) number of BSs = 3 and (b) number of BSs = 5.

increases, compared with the simple power splitter framework. This is a natural result because considering the more number of power splitters in a receiver offers more degree of freedom to improve performance, whereas the optimization problem becomes more complex.

To identify the validity of the proposed algorithm under various channel scenarios, we have conducted the performance evaluations under an uncorrelated channel scenario. Fig. 6 presents the performance comparisons between the proposed and existing algorithms under the uncorrelated channel

TABLE V
PERFORMANCE COMPARISON BETWEEN THE PROPOSED AND LINEAR-SEARCH ALGORITHMS UNDER THE UNCORRELATED CHANNEL SCENARIO WHEN $K = 2$ AND $R_{th} = 0.5$ bps/Hz

Algorithm	Harvested energy [mW]	Data rate [bps/Hz]
Linear-search	238	0.53
Proposed	213	0.67

scenario, when $L = 3$ and $R_{th} = 0.35$ bps/Hz. As shown Fig. 6, the proposed algorithm outperforms than other existing algorithms with respect to harvested energy while supporting the required data rate regardless of the number of BSs. In addition, similar to the results of correlated channel in Fig. 5(b), Fig. 6(b) shows that the proposed algorithm is the only algorithm that satisfies the data rate constraint in a severe interference scenario.

Table V shows the performance comparison between the proposed and linear-search algorithm with $L = 3$ and 10 for the step size of linear-search algorithm. In optimization problem (P1), it is identified that it is difficult to find the characteristics of function $R^{[i,j]}(\mathbf{u}, \boldsymbol{\rho})$ with respect to \mathbf{u} and $\boldsymbol{\rho}$ (e.g., convex or monotonic function). This means that to find an optimal solution of problem (P1), global optimization techniques should be considered [46]. Among them, we consider the linear-search algorithm for finding the optimal solution of (P1). To find an optimal solution of (P1) through the linear-search algorithm, its step size should be set as small as possible. For example, if we assume that each element of the two vectors is divided into total ten levels, total number of candidate solutions to check for obtaining the optimal solution is 10^{3L} . This means that it might be impossible to find an optimal solution of optimization problem with multiple continuous variables, even though the number of antennas is relatively small. From Table V, it is shown that the linear-search algorithm performs better than the proposed algorithm (i.e., about 10% performance degradation compared with the linear-search algorithm) with respect to harvested energy, and its performance plays a role in the upper bound of performance of the proposed algorithm. Whereas, to achieve the solution of the linear-search algorithm, it takes several hours unlike the proposed algorithm which can find the solution in real time, once the network performance converges. That is, the linear-search algorithm is impractical under the considered network scenario.

VI. CONCLUSION

In this work, we considered the joint EH and ID optimization for the smart city with IoT devices enabling

simultaneous wireless information and power reception. We proposed a distributed online learning algorithm to maximize the harvested energy with an achievable data rate constraint by jointly determining the receive beamforming and power-splitting ratio vectors for IoT devices. To implement the distributed online learning algorithm, we presented a double-DDPG-based learning algorithm, where two different types of action sets are determined in real time. The results of extensive simulations showed that the proposed algorithm can enhance the performance in terms of EH while satisfying the data rate requirement for IoT devices compared with the existing algorithms.

REFERENCES

- [1] P. Bocquier, "World urbanization prospects: An alternative to the UN model of projection compatible with the mobility transition theory," *Demograph. Res.*, vol. 12, no. 9, pp. 197–236, May 2005.
- [2] H. Zhang, M. Babar, M. U. Tariq, M. A. Jan, V. G. Menon, and X. Li, "SafeCity: Toward safe and secured data management design for IoT-enabled smart city planning," *IEEE Access*, vol. 8, pp. 145256–145267, 2020.
- [3] T.-H. Kim, C. Ramos, and S. Mohammed, "Smart city and IoT," *Future Gener. Comput. Syst.*, vol. 76, pp. 159–162, Nov. 2017.
- [4] G. Zhang, W. Zhang, Y. Cao, D. Li, and L. Wang, "Energy-delay tradeoff for dynamic offloading in mobile-edge computing system with energy harvesting devices," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4642–4655, Oct. 2018.
- [5] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.
- [6] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.
- [7] D. W. K. Ng, E. S. Lo, and R. Schober, "Wireless information and power transfer: Energy efficiency optimization in OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6352–6370, Dec. 2013.
- [8] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information transfer with opportunistic energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 288–300, Jan. 2013.
- [9] D. W. K. Ng, E. S. Lo, and R. Schober, "Robust beamforming for secure communication in systems with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4599–4615, Aug. 2014.
- [10] W. Lu, Y. Gong, X. Liu, J. Wu, and H. Peng, "Collaborative energy and information transfer in green wireless sensor networks for smart cities," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1585–1593, Apr. 2018.
- [11] W. Lu et al., "Energy efficiency optimization in SWIPT enabled WSNs for smart agriculture," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4335–4344, Jun. 2021.
- [12] Y. Xu et al., "Joint beamforming and power-splitting control in downlink cooperative SWIPT NOMA systems," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4874–4886, Sep. 2017.
- [13] H. Al-Obiedollah et al., "A joint beamforming and power-splitter optimization technique for SWIPT in MISO-NOMA system," *IEEE Access*, vol. 9, pp. 33018–33029, 2021.
- [14] Q. Shi, L. Liu, W. Xu, and R. Zhang, "Joint transmit beamforming and receive power splitting for MISO SWIPT systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3269–3280, Jun. 2014.
- [15] Q. Shi, W. Xu, T.-H. Chang, Y. Wang, and E. Song, "Joint beamforming and power splitting for MISO interference channel with SWIPT: An SOCP relaxation and decentralized algorithm," *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6194–6208, Dec. 2014.
- [16] J. Liao, M. R. A. Khandaker, and K.-K. Wong, "Robust power-splitting SWIPT beamforming for broadcast channels," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 181–184, Jan. 2016.
- [17] A. Al-Baidhani, M. Vehkaperä, and M. Benaissa, "Simultaneous wireless information and power transfer based on generalized triangular decomposition," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 751–764, Sep. 2019.
- [18] G. Kwon, H. Park, and M. Z. Win, "Joint beamforming and power splitting for wideband millimeter wave SWIPT systems," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 5, pp. 1211–1227, Aug. 2021.
- [19] N. Shanin, L. Cottatellucci, and R. Schober, "Optimal transmit strategy for multi-user MIMO WPT systems with non-linear energy harvesters," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1726–1741, Mar. 2022.
- [20] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Adaptive rate and energy harvesting interval control based on reinforcement learning for SWIPT," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2571–2574, Dec. 2018.
- [21] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT," *IEEE Access*, vol. 7, pp. 17450–17460, 2019.
- [22] K. Lee and W. Lee, "Learning-based resource management for SWIPT," *IEEE Syst. J.*, vol. 14, no. 4, pp. 4750–4753, Dec. 2020.
- [23] Y. Al-Eryani, M. Akrouf, and E. Hossain, "Simultaneous energy harvesting and information transmission in a MIMO full-duplex system: A machine learning-based design," 2020, *arXiv:2002.06193*.
- [24] J. Tang et al., "Decoupling or learning: Joint power splitting and allocation in MC-NOMA with SWIPT," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5834–5848, Sep. 2020.
- [25] L. Li et al., "Learning-aided resource allocation for pattern division multiple access-based SWIPT systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 131–135, Jan. 2021.
- [26] W. Lee, K. Lee, H.-H. Choi, and V. C. M. Leung, "Deep learning for SWIPT: Optimization of transmit-harvest-respond in wireless-powered interference channel," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5018–5033, Aug. 2021.
- [27] D. Xu, V. Jamali, X. Yu, D. W. K. Ng, and R. Schober, "Optimal resource allocation design for large IRS-assisted SWIPT systems: A scalable optimization framework," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1423–1441, Feb. 2022.
- [28] H. W. Oleiwi and H. Al-Raweshidy, "SWIPT-pairing mechanism for channel-aware cooperative H-NOMA in 6G terahertz communications," *Sensors*, vol. 22, no. 16, p. 6200, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/16/6200>
- [29] Z. Li, W. Chen, H. Cao, H. Tang, K. Wang, and J. Li, "Joint communication and trajectory design for intelligent reflecting surface empowered UAV SWIPT networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 12840–12855, Dec. 2022.
- [30] W. Lyu, Y. Xiu, J. Zhao, and Z. Zhang, "Optimizing the age of information in RIS-aided SWIPT networks," *IEEE Trans. Veh. Technol.*, early access, Sep. 22, 2022, doi: [10.1109/TVT.2022.3208612](https://doi.org/10.1109/TVT.2022.3208612).
- [31] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "A survey of 5G technologies: Regulatory, standardization and industrial perspectives," *Digit. Commun. Netw.*, vol. 4, no. 2, pp. 87–97, Apr. 2018.
- [32] T. Kim, D. J. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?" *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1654–1669, Dec. 2011.
- [33] L. Zhang, J. Tan, Y.-C. Liang, G. Feng, and D. Niyato, "Deep reinforcement learning-based modulation and coding scheme selection in cognitive heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3281–3294, Jun. 2019.
- [34] J. G. Proakis and M. Salehi, *Digital Communications*. Boston, MA, USA: McGraw-Hill, 2008.
- [35] *Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception (Release 16)*, 3GPP Standard TS 36.101, Dec. 2018.
- [36] B. C. Jung and W.-Y. Shin, "Opportunistic interference alignment for interference-limited cellular TDD uplink," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 148–150, Feb. 2011.
- [37] H. J. Yang, W.-Y. Shin, B. C. Jung, C. Suh, and A. Paulraj, "Opportunistic downlink interference alignment for multi-cell MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1533–1548, Mar. 2017.
- [38] G. Ma, J. Xu, Y. Zeng, and M. R. V. Moghadam, "A generic receiver architecture for MIMO wireless power transfer with nonlinear energy harvesting," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 312–316, Feb. 2019.
- [39] "5G wireless access: An overview," Ericsson, Stockholm, Sweden, White Paper, 2022. [Online]. Available: <https://www.ericsson.com/en/reports-and-papers/white-papers/5g-wireless-access-an-overview>
- [40] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in *Proc. ICLR*, 2016, pp. 1–14. [Online]. Available: <http://dblp.uni-trier.de/db/conf/iclr/iclr2016.html#LillicrapHPHETS15>
- [41] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

- [42] D. Mishra and G. C. Alexandropoulos, "Transmit precoding and receive power splitting for harvested power maximization in MIMO SWIPT systems," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 3, pp. 774–786, Sep. 2018.
- [43] B. Koo and D. Park, "Interference alignment and wireless energy transfer via antenna selection," *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 548–551, Apr. 2014.
- [44] Y. Kim, J. Youn, and B. C. Jung, "Interference alignment with receive antenna partitioning for SWIPT-enabled fog RANs," *ICT Exp.*, vol. 8, pp. 485–489, Dec. 2022.
- [45] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.
- [46] R. Horst and P. M. Pardalos, *Handbook of Global Optimization*. New York, NY, USA: Springer, 2002.



Yongjae Kim (Member, IEEE) received the B.S. degree (*summa cum laude*) in electronics engineering from Sejong University, Seoul, South Korea, in 2013, and the M.S. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2015 and 2019, respectively.

He was a Senior Researcher with the Korea–Russia Innovation Center, Korea Institute of Industrial Technology, Incheon, South Korea, from 2019 to 2020. He is currently a Senior Researcher with the Maritime ICT Research and Development Center, Korea Institute of Ocean Science and Technology, Busan, South Korea. His research interests include IoT networks, machine learning, simultaneous wireless information and power transfer, maritime wireless communication, and radio resource management for 5G and beyond 5G.



Bang Chul Jung (Senior Member, IEEE) received the B.S. degree in electronics engineering from Ajou University, Suwon, South Korea, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2004 and 2008, respectively.

He was a Senior Researcher/Research Professor with KAIST Institute for Information Technology Convergence, Daejeon, from January 2009 to February 2010. From March 2010 to August 2015, he was a Faculty Member with Gyeongsang National University, Tongyeong, South Korea. He is currently a Professor with the Department of Electrical Engineering, Chungnam National University, Daejeon. His research interests include 6G wireless communications, wireless IoT communications, statistical signal processing, information theory, wireless localization, interference management, radar signal processing, spectrum sharing, multiple antennas, multiple access techniques, radio resource management, machine learning, and GNSS receiver signal processing.

Prof. Jung was the recipient of the 5th IEEE COMMUNICATION SOCIETY ASIA–PACIFIC OUTSTANDING YOUNG RESEARCHER Award in 2011, the KICS Haedong Young Scholar Award in 2015, and the 29th KOFST Science and Technology Best Paper Award in 2019. He has been the Associate Editor of the IEEE VEHICULAR TECHNOLOGY MAGAZINE since May 2020. He was an Associate Editor of the IEICE TRANSACTIONS ON FUNDAMENTALS OF ELECTRONICS, COMMUNICATIONS, AND COMPUTER SCIENCES from 2018 to 2022.



Yujae Song (Member, IEEE) received the Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2016.

He was a Visiting Scholar of Communication Systems from KTH Royal Institute of Technology, Stockholm, Sweden, in 2015. He was a Senior Researcher with the Maritime ICT Research and Development Center, Korea Institute of Ocean Science and Technology, Busan, South Korea, from 2016 to 2022. He was an Assistant Professor with the Department of Computer Software Engineering, Kumoh National Institute of Technology, Gumi, South Korea, from 2022 to 2023. Since March 2023, he has been an Assistant Professor with the Department of Robotics Engineering, Yeungnam University, Gyeongsan, South Korea. His research interests include design, analysis, and optimization of various wireless communication systems, including 5G, maritime/underwater, and smart grid communications.