Learning Aided Joint Sensor Activation and Mobile Charging Vehicle Scheduling for Energy-Efficient WRSN-Based Industrial IoT

Jiayuan Chen, Changyan Yi, *Member, IEEE*, Ran Wang, *Member, IEEE*,
Kun Zhu, *Member, IEEE*, and Jun Cai, *Senior Member, IEEE*

***Abstract*—In this paper, the joint sensor activation and mobile charging vehicle scheduling for wireless rechargeable sensor network (WRSN) based industrial Internet of Things (IIoT) is studied. In the proposed framework, an optimal sensor set is selected to collaboratively execute a bundle of heterogeneous industrial tasks (e.g., production-line monitoring), meeting the quality-of-monitoring (QoM) of each individual task, and we consider that a mobile charging vehicle (MCV) is scheduled for recharging sensors before their charging deadlines, i.e., time instants of running out of their batteries, in order to prevent from any potential service interruptions (which is one of the key features of IIoT). Our goal is to jointly optimize the sensor activation and MCV charging scheduling for minimizing the system energy consumption, subject to tasks' QoM requirements, sensor charging deadlines and the energy capacity of the MCV. Unfortunately, solving this problem is nontrivial, because it involves solving two tightly coupled NP-hard optimization problems. To address this issue, we design a novel scheme integrating reinforcement learning and marginal product based approximation algorithms, and prove that it is not only computationally efficient but also theoretically bounded with a guaranteed performance in terms of the approximation ratio. Simulation results show the feasibility of the proposed scheme and demonstrate its superiority over counterparts.**

***Index Terms*—WRSN-based IIoT, sensor activation, mobile charging scheduling, joint optimization, reinforcement learning**

I. Introduction

It is widely known that wireless sensor network (WSN) is an essential component for enabling the industrial Internet of Things (IIoT) due to its capability of providing pervasive surveillance, control, maintenance and automation in intelligent industrial systems [2]. One one hand, the future IIoT applications may require tremendously high data rates and considerably large processing capacities [3], [4], which will accelerate the energy consumption of sensor devices. On the other hand, geographically distributed sensor devices are typically powered by energy-limited batteries, and once a sensor device runs out of the energy, its perception ability will be greatly reduced and the overall system may collapse. To tackle such sensor energy provisioning problem, researchers studied how to reduce the energy consumption by optimizing sensors deployment [5], wake-up and sleeping scheduling [6], [7], sensing radius adaption mechanism [8], and adapting sensors' sampling rate [9], etc. to prolong the system lifetime. However, these methods cannot fundamentally address the energy shortage, and thus recent advances of wireless power transfer (WPT) technology [10]–[12] has inspired the emergence of wireless rechargeable sensor networks (WRSNs) [13]. With the implementation of WRSNs, mobile charging vehicles (MCVs) equipped with powerful transceivers can be dispatched to travel around and replenish the energy of sensors via coupled magnetic resonance.

Although WRSN is envisioned to boost the energy capacity of the overall system thanks to the recharging capability of the MCV, it also suffers from several inherent restrictions. Particularly, since the energy capacity of the MCV is still relatively limited compared to the energy replenishment demands from the large number of sensors in the WRSN, the MCV has to frequently return to the depot for battery recharging or replacement if its charging route is not properly scheduled [14]. Obviously, this may significantly degrade the system efficiency. Therefore, to exploit the advantage of WRSNs, besides the sensor-end optimizations, the corresponding MCV charging scheduling should also be carefully addressed. Recent efforts in this area include energy-efficiency aware charging scheduling [15], [16], mobile data gathering and charging scheme [17]–[19], joint optimization of routing protocol and mobile charging [20], [21]. Nevertheless, there are some critical issues, especially those related to IIoT applications, which are of great importance, have not yet been well investigated.

a) As a typical networking system, WRSN-based IIoT is embarking on the era of massive machine-type Internet of Things communications (mMTIC) [22], where trillions of sensors will be connected and work collaboratively, and thereby a huge amount of energy will be consumed in sensing, computing and information exchanging. Additionally, sensors in IIoT may be heterogeneous in terms of sensing radius, types, battery capacities, and industrial tasks may be highly heterogeneous in terms of task requirements (i.e., quality-of-monitoring (QoM) requirements), locations, types, etc. These motivate us to study how to acti-

J. Chen, C. Yi, R. Wang and K. Zhu are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China. (E-mail: {jiayuan.chen, changyan.yi, wangran, zhukun}@nuaa.edu.cn)

J. Cai is with the Department of Electrical and Computer Engineering, Concordia University, Montréal, QC, H3G 1M8, Canada. (E-mail: jun.cai@concordia.ca).

A preliminary version [1] has been presented in IEEE ICC 2022.

(Corresponding author: Changyan Yi)

vate the optimal set of sensors for collaboratively executing on-purpose industrial tasks to minimize the system energy consumption while meeting the QoM of each individual task rather than activating all sensors. Furthermore, different from the traditional sensor activation problems [23], by enabling WRSNs, the rechargeability of sensor devices and the real-time energy replenishment of the MCV should be both taken into account.

b) For delay and reliability sensitive IIoT applications [24], [25], sensors must keep up high-intensity operations for long periods and continuously feed data back to controllers or actuators. For instance, while a cutting machine is working, industrial camera sensors must collaboratively monitor the positions of cutters in real-time and send out the data in a timely manner [26]. Any unpredictable sensor failure may cause serious consequences, e.g, unexpected damages and casualties. Hence, in order to guarantee that all activated sensors can work continuously without any potential interruptions during the monitoring period, all these sensors should be recharged by the MCV of the WRSN before their charging deadlines (i.e., the instant of running out of the battery). However, due to the limited moving velocity and the energy capacity constraint of the MCV, the charging route scheduling has to be well managed by considering the optimal set of activated sensors, their energy expenditures, charging time, locations, etc.

However, addressing the aforementioned features simultaneously is difficult because i) the optimal sensor set activation problem can be reduced to a generalized assignment problem (GAP) which is widely known as NP-hard [27]; ii) unlike the conventional travel salesman problem (TSP) which aims to minimize the total travel cost only [28], the MCV charging route scheduling should also take into account sensors' charging deadline constraints imposed by continuous operation requirement, resulting in a much more complicated combinatorial optimization issue; and iii) the above two problems (i.e., sensor set activation and MCV charging route scheduling) are tightly coupled, meaning that they cannot be solved independently.

To tackle these challenges and fill the gap in the literature, in this paper, we propose a novel scheme to jointly optimize the sensor activation and mobile charging scheduling for WRSN-based IIoT systems. Our goal is to minimize the total energy consumption of the entire system, subject to tasks' QoM requirements, sensor charging deadlines and energy capacity of the MCV. In the considered model, there is a central platform (industrial controller) who declares a bundle of tasks for industrial environment monitoring (e.g., production-line monitoring) at the beginning of a certain monitoring period, and is responsible to determine the optimal activated sensor set and charging route of the MCV. The MCV is required to start from the depot, travel along the scheduled route and return to the depot at the end of the trip. While traveling on its route, the MCV charges each activated sensor before its charging deadline expires. To solve such joint sensor activation and mobile charging scheduling problem, we design an efficient scheme integrating reinforcement learning (RL) and marginal product based approximation algorithms.

The main contributions of this paper are summarized in the following.

- With the objective of minimizing the total energy consumption of the entire system (including the energy consumption of all sensors and the MCV) under the constraints of satisfying the specific features and corresponding requirements for WRSN-based IIoT applications, a joint optimization of sensor activation and mobile charging scheduling problem is formulated.
- A novel solution scheme integrating reinforcement learning and marginal product based approximation algorithms is developed, and is prove to be not only computationally efficient but also theoretically bounded with a guaranteed performance in terms of the approximation ratio.
- Extensive simulations are conducted to demonstrate the superiority of the proposed solution scheme compared to counterparts.

The rest of this paper is organized as follows: Section II briefly reviews the related work and highlights the novelties of this paper. Section III introduces the system model and problem formulation of the joint optimization of sensor activation and mobile charging scheduling for WRSN-based IIoT systems. In Section IV, the complexity of the formulated optimization problem is first analyzed, and then a novel scheme integrating reinforcement learning and marginal product based approximation algorithms is proposed. Section V analyzes the theoretical performance of the proposed scheme. Simulation results are provided in Section VI, followed by the conclusion in Section VII.

II. Related Work

Due to the rapid development of IIoT, the scale and density of sensor devices are expected to increase dramatically, resulting in a substantial amount of energy consumption [29]. To support this new paradigm and overcome the energy bottleneck, WRSN has emerged as an alternative of WSN with a higher flexibility in energy replenishment [20].

It is worth noting that, compared to the energy management in WSNs (which commonly focused on sensor-end optimizations [5]–[8]), WRSNs further involve the energy expenditure and charging scheduling of the MCVs, and thus the corresponding energy management becomes much more complicated and has attracted explosive research attentions recently. For instance, Baek et al. in [15] developed a novel joint data collection and mobile charging scheme in the WRSN by using an unmanned aerial vehicle (UAV) which can execute the data collection and conduct the energy charging simultaneously for improving the energy efficiency of the entire network. Yang et al. in [16] proposed a dynamic charging scheme based on the actor-critic reinforcement learning algorithm to minimize the number of dead sensor devices while maximizing the energy efficiency. Han et al. in [20] jointly considered the routing protocol and mobile charging scheduling to achieve global energy balance. However, all of them were restricted to improving the energy efficiency only, but ignored the potential requirements of sensing qualities imposed by mission-critical applications, such as IIoT applications.

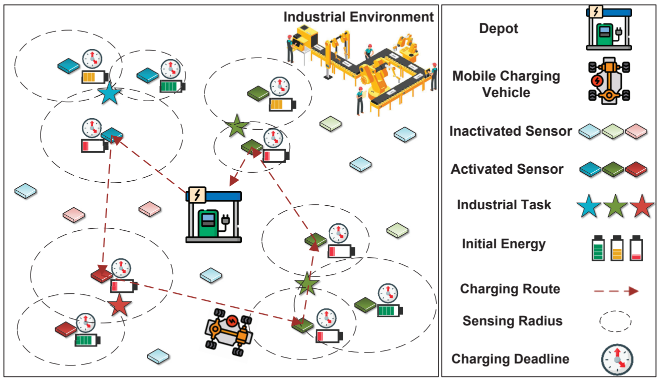

Fig. 1. An illustration of the considered WRSN-based IIoT system.

Some researches have been dedicated in studying how the sensing quality of various tasks and the network energy efficiency can be guaranteed simultaneously for WRSNs [30]–[32]. Specifically, Dai et al. in [31] designed a joint charging and scheduling scheme (i.e., choosing sensors to charge and scheduling sensors' activation) to maximize the QoM of stochastic events. Wu et al. in [30] solved a collaborated task-driven mobile charging and scheduling problem for maximizing the overall task utility, where the traveling energy consumption of the MCV was minimized. However, both works did not consider that there may exist stringent constraints on sensing qualities of heterogeneous tasks, and hence may not be suitable for IIoT applications. The most related research is [32], in which the authors proposed a joint energy replenishment and operation (active/sleep) scheduling mechanism for WRSN using an MCV with a given charging capacity, with the objective of maximizing the network lifetime while meeting a strict QoM. However, this paper assumed that all sensors could continuously operate for months or even years after fully recharging, and the charging and traveling time of the MCV were neglected. These do not fit the features of IIoT, where all sensors must keep high-intensity work with relatively large energy consumption rate and may encounter service interruptions if they cannot be recharged in time.

In summary, different from all existing work, this paper studies a joint optimization of sensor activation and mobile charging scheduling for WRSN-baesd IIoT systems, subject to tasks' QoM requirements, sensor charging deadlines and the energy capacity of the MCV.

III. System Model and Problem Description

In this section, the system model of WRSN-based IIoT is first described. Then, the corresponding joint optimization of sensor activation and mobile charging scheduling problem is formulated. For convenience, Table I lists some important notations used in this paper.

A. Network Model

TABLE I
Important Notations in This Paper

Symbol	Meaning
$\mathcal{Z}$	set of tasks
$\mathcal{N}$	set of sensor nodes
T	time duration of a monitoring period
M	total amount of tasks
J	total amount of task types
$\mathcal{H}$	set of activated sensors
$\mathcal{H}_\phi$	set of sensors that do not need to be charged
$\mathcal{H}_\psi$	set of sensors that need to be charged
$\mathcal{L}_{\mathcal{H}_\psi}$	charging sequence vector of the MCV
r_i	sensing radius of sensor i
$E_i^{initial}$	initial energy amount of sensor i
$E_i^{capacity}$	battery capacity of sensor i
E_i^{min}	minimum energy for sensor i to be operational
$\mathcal{C}_i$	energy consumption rate of sensor i
E_i^{demand}	charging demand of sensor i
ddl_i	charging deadline of sensor i
p_{i,z_j^m}	detection probability to task z_j^m of sensor i
α_i	decay factor of sensor i
$\mathcal{N}_j$	set of sensors specialized in task type j
$\eta_{z_j^m}$	QoM requirement of task z_j^m
$dist(i, z_j^m)$	Euclidean distance between sensor i and task z_j^m
E_{MCV}	energy capacity of the MCV
v	moving velocity of the MCV
δ	charging rate of the MCV
γ	traveling energy consumption rate of the MCV
a_g	g_{th} visiting target of the MCV
Λ_{a_g}	arrival time of the MCV at target a_g

Consider a WRSN-based IIoT system, as illustrated in Fig. 1, consisting of a group of heterogeneous industrial production-line monitoring tasks, a set of stationary rechargeable sensors $\mathcal{N}$ with cardinality of $|\mathcal{N}| = N$ randomly deployed in a certain area, and an MCV which is responsible for replenish energy of all sensors if necessary.

At the beginning of a monitoring period, the industrial controller declares a bundle of monitoring tasks $\mathcal{Z} = \{z_j^m | \forall m \in \{1, 2, ..., M\}, \forall j \in \{1, 2, \ldots, J\}\}$, where m and j stand for the index of the monitoring task and its corresponding type, respectively. For meeting the QoM requirements of all these tasks while avoiding excessive energy consumption, an appropriate group of sensors $\mathcal{H} \subseteq \mathcal{N}$ should be activated and working collaboratively. In practice, sensors' sensing radius are inherently limited, which can be denoted by r_i, $\forall i \in \mathcal{N}$. In addition, we assume that different types of sensors may only be able to execute tasks fitting their types. For example, noisy sensors can only detect noises, and vision sensors are particularly for capturing images of target objects [33]. Define $\mathcal{N}_j$ as the set of sensors specialized in task type j, and it is obvious that $\cup \mathcal{N}_j = N$, $\forall j \in \{1, ..., J\}$. Notice that, each sensor $i \in \mathcal{N}$ can execute task $z_j^m \in \mathcal{Z}$ only when the task z_j^m is located within its sensing radius r_i, and falls into its targeted type. Besides, in each monitoring period, each sensor is assumed to be capable of executing at most one task due to the hardware limitation [34]. In this paper, we adopt the widely used probabilistic sensing coverage (PSC) model [35] to describe sensors' contributions to certain monitoring tasks, and denote p_{i,z_j^m} as the detection probability of task z_j^m by sensor i, which can be calculated as

$$p_{i,z_j^m} = \begin{cases} e^{-\alpha_i \cdot dist\left(i, z_j^m\right)}, & \text{if } dist\left(i, z_j^m\right) \leq r_i, i \in \mathcal{N}_j, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where α_i, $\forall i \in \mathcal{N}$, is a factor that describes how fast the sensing capacity decays with the distance. We call α_i the decay factor of sensor i, which depends on sensor i's physical characteristics and its implementation environment. Besides, $dist\left(i, z_j^m\right)$ indicates the Euclidean distance between the locations of sensor i and task z_j^m.

To meet the QoM requirement of each individual task, the collaborative detection probability of sensor set $\mathcal{H}$ (i.e., these are activated) to the monitoring task z_j^m has to be larger or equal to a pre-determined threshold $\eta_{z_j^m}$, i.e.,

$$1 - \prod_{i \in \mathcal{H}} (1 - p_{i,z_j^m}) \geq \eta_{z_j^m}, \tag{2}$$

where $\eta_{z_j^m}$ measures the minimum QoM requested by each task z_j^m [32].

For IIoT applications, it is further required that if sensors are activated to execute tasks, they should work continuously during the monitoring period. Otherwise, service interruptions may occur, leading to serious consequences. However, the battery capacity of each sensor $E_i^{capacity}$ is relatively limited, and once the battery is completely consumed, the sensor stops working. To this end, the MCV is employed with an energy capacity E_{MCV} which travels starting at the depot located at the center of the area, charges dying sensors in $\mathcal{H}$ and returns to the depot for battery replenishment at the end.

We denote E_i^{intial} as the initial energy of each sensor $i \in \mathcal{N}$ at the beginning of the monitoring period. Without loss of generality, let us assume that for each sensor $i \in \mathcal{N}$, E_i^{intial} is sufficiently large to guarantee that $E_i^{intial} \geq E_i^{min}$, where E_i^{min} is the minimum energy for $i \in \mathcal{N}$ to be operational [36]. Here, we characterize the energy consumption rate of each sensor $i \in \mathcal{N}$ by $\mathcal{C}_i$. Note that it is possible that some sensors may have abundant amount of energy to support them working continuously during the monitoring period and are not necessary to be recharged by the MCV. We classify these sensors into set $\mathcal{H}_\phi \subseteq \mathcal{H}$, and categorize the others which have to be recharged by the MCV into set $\mathcal{H}_\psi = \mathcal{H} \backslash \mathcal{H}_\phi$. Then, the amount of energy that each sensor required to be recharged can be expressed as

$$E_i^{demand} = \begin{cases} T \cdot \mathcal{C}_i - (E_i^{initial} - E_i^{min}), & \forall i \in \mathcal{H}_\psi, \\ 0, & \forall i \in \mathcal{H}_\phi, \end{cases} \tag{3}$$

where T is the time duration of a monitoring period.

For ensuring that all activated sensors can execute tasks continuously, the MCV should charge the sensors in set $\mathcal{H}_\psi$ before each of its charging deadline ddl_i, $i \in \mathcal{H}_\psi$ expires, which can be calculated by

$$ddl_i = \frac{E_i^{initial} - E_i^{min}}{\mathcal{C}_i}, \forall i \in \mathcal{H}_\psi. \tag{4}$$

Besides, let us denote the charging route of the MCV by a vector that consists of its visiting targets $\mathcal{L}_{\mathcal{H}_\psi} = \{a_0, a_1, ..., a_g, ..., a_{|\mathcal{H}_\psi|}, a_{|\mathcal{H}_\psi|+1}\}$, where a_g signifies the g_{th} visiting target (i.e., the targeted sensor for recharging). Specifically, $a_0 = a_{|\mathcal{H}_\psi|+1} = 0$ indicates that the MCV travels starting from the depot and returns at the end, and $a_g \in \mathcal{H}_\psi$ for $g = 1, \ldots, |\mathcal{H}_\psi|$. Following the convention in the literature [16], [30], [31], we consider that each sensor $i \in \mathcal{H}_\psi$ can only be visited once, that is $a_g \neq a_{g'}$ for $g \neq g'$. Furthermore, denote the arrival time of the MCV at a visiting target a_g as Λ_{a_g}. Clearly, Λ_{a_g} depends on the arrival time of the previously visited target a_{g-1}, the service time (i.e., battery recharging time) for target a_{g-1}, and the traveling time of the MCV from a_{g-1} to a_g. Hence, Λ_{a_g} can be expressed as

$$\Lambda_{a_g} = \Lambda_{a_{g-1}} + \frac{E_{a_{g-1}}^{demand}}{\delta} + \frac{dist\left(a_{g-1}, a_g\right)}{v}, \forall a_g \in \mathcal{L}_{\mathcal{H}_\psi}, \tag{5}$$

where δ and v stand for the the charging rate and the velocity of the MCV, respectively. According to the definition in (3), $E_{a_g}^{demand}$ depicts the amount energy that the target a_g (or sensor a_g) demands for recharging. In particular, $E_{a_0}^{demand} = E_{a_{|\mathcal{H}_\psi|+1}}^{demand} = 0$, and $\Lambda_{a_0} = 0$, because both a_0 and $a_{|\mathcal{H}_\psi|+1}$ signify the depot.

B. Problem Formulation

The energy consumption of the entire WRSN-based IIoT system includes the energy consumption of the MCV and the energy consumption of all sensors in $\mathcal{H}$ for executing tasks. Although the energy cost of the MCV further consists of both the traveling energy cost and the recharging energy cost, all recharging energy will be consumed completely by sensors for a higher energy utilization efficiency, and thus such term is implied by the energy cost of sensors in $\mathcal{H}$. Therefore, the total energy consumption can be expressed as

$$E_{total}(\mathcal{H}, \mathcal{L}_{\mathcal{H}_\psi}) = \sum_{g=0}^{|\mathcal{H}_\psi|} \gamma \cdot dist\left(a_g, a_{g+1}\right) + \sum_{i \in \mathcal{H}} T \cdot \mathcal{C}_i,$$

where γ represents the energy consumption rate of MCV's traveling, so that the first term depicts the traveling energy consumption of the MCV and the second term is the energy consumption of all activated sensors.

Then, with the aim of minimizing the total energy consumption $E_{total}\left(\mathcal{H}, \mathcal{L}_{\mathcal{H}_\psi}\right)$, the joint optimization of sensor activation (i.e., selecting the optimal set of sensors to activate, denoted by $\mathcal{H}$) and mobile charging scheduling (i.e., determining the optimal charging route, denoted by $\mathcal{L}_{\mathcal{H}_\psi}$) can be formulated as

$$[\mathcal{P}1]: \min_{\mathcal{H}, \mathcal{L}_{\mathcal{H}_\psi}} E_{total}(\mathcal{H}, \mathcal{L}_{\mathcal{H}_\psi}) \tag{6}$$

$$s.t., \; 1 - \prod_{i \in \mathcal{H}} (1 - p_{i,z_j^m}) \geq \eta_{z_j^m}, \forall z_j^m \in \mathcal{Z}, \tag{6a}$$

$$\Lambda_{a_g} \leq ddl_{a_g}, g{=}1, \ldots |\mathcal{H}_\psi|, \tag{6b}$$

$$a_g \neq a_{g'}, g \neq g'; g{=}1, \ldots |\mathcal{H}_\psi|, g'{=}1, \ldots |\mathcal{H}_\psi|, \tag{6c}$$

$$\sum_{g=0}^{|\mathcal{H}_\psi|} \gamma \cdot dist\left(a_g, a_{g+1}\right) + \sum_{g=1}^{|\mathcal{H}_\psi|} E_{a_g}^{demand} \leq E_{MCV}, \tag{6d}$$

$$a_0 = 0, a_{|\mathcal{H}_\psi|+1} = 0, \tag{6e}$$

$$\mathcal{H} \subseteq \mathcal{N}, \tag{6f}$$

$$\mathcal{H} = \mathcal{H}_\phi \cup \mathcal{H}_\psi, \tag{6g}$$

$$\mathcal{L}_{\mathcal{H}_\psi} = \{a_0, a_1, ..., a_g, ..., a_{|\mathcal{H}_\psi|}, a_{|\mathcal{H}_\psi|+1}\}, \tag{6h}$$

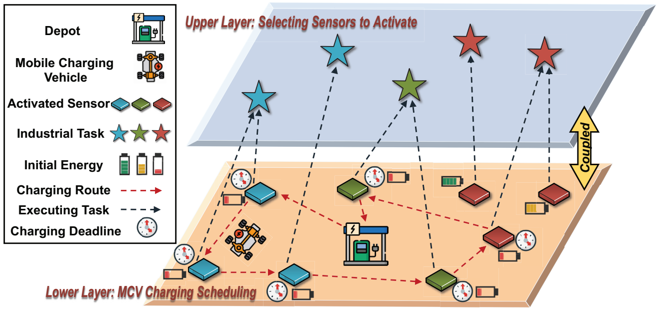

Fig. 2. An illustration of the considered two-layer coupled problem.

where constraint (6a) states that the QoM of each task should be guaranteed; constraint (6b) imposes a strict requirement that the MCV should always be scheduled to arrive before each sensor's charging deadline expires; constraint (6c) means that the MCV should not visit the same sensor more than once in the scheduled charging route; constraint (6d) indicates that the total energy consumption of the MCV should be less than or equal to its energy capacity E_{MCV}; and constraint (6e) illustrates that the MCV starts at the depot and returns to the depot at the end.

IV. Joint Sensor Activation and Mobile Charging Scheduling

In this section, we first prove that the formulated problem $[\mathcal{P}1]$ is NP-hard due to the combination of two coupling NP-hard problems. Then, we propose a novel solution scheme integrating reinforcement learning and marginal product based approximation algorithms, which is later shown to be not only computationally efficient but also performance guaranteed in terms of the approximation ratio.

A. NP-Hardness of the Joint Optimization Problem

From the formulation of problem $[\mathcal{P}1]$, we can observe that the joint optimization of sensor activation and mobile charging scheduling actually includes two-layer optimizations. The upper layer optimization mainly addresses the sensor set selection with tasks' QoM constraints, where the objective is to minimize the energy consumption of the activating sensor set $\mathcal{H}$. The lower layer optimization aims to determine the charging route of the MCV by taking into account sensors' charging deadlines and MCV's energy budget, where the objective is to minimize the energy consumption of the MCV. Obviously, these two optimization problems are tightly coupled, as illustrated in Fig. 2, because the input of the lower layer problem depends on the output of the upper layer one, while the optimization of the upper layer problem is in turn subject to the results of the lower layer one.

If the charging route $\mathcal{L}_{\mathcal{H}_\psi}$ of the MCV is given, we can get the subset of candidate sensors $\mathcal{N}^{'} \subseteq \mathcal{N}$, where all sensors in $\mathcal{N}^{'}$ have sufficient energy (namely with either sufficiently large initial energy or energy replenished by the MCV in route $\mathcal{L}_{\mathcal{H}_\psi}$) to execute tasks continuously during the monitoring period. Then, the upper layer sensor set activation problem becomes

$$[\mathcal{P}2] : \min_{\mathcal{H}} \sum_{i \in \mathcal{H}} T \cdot \mathcal{C}_i$$
$$s.t.,\ (6a),\ (6g) \text{ and } \mathcal{H} \subseteq \mathcal{N}^{'}.$$

***Lemma** 1:* Problem $[\mathcal{P}2]$ is NP-hard.

Proof: It is obvious that problem $[\mathcal{P}2]$ can be reduced to a generalized assignment problem (GAP). Specifically, given any instance of the GAP, the instance of problem $[\mathcal{P}2]$ can be easily constructed by considering that the detection probability of each sensor $p_{i,z_j^m}, i \in \mathcal{N}$ is analogous to the size of each item, QoM of each tasks $\eta_{z_j^m}, z_j^m \in \mathcal{Z}$ is analogous to the capacity of each bin, the energy consumption of each sensor $T \cdot \mathcal{C}_i, i \in \mathcal{N}$ is analogous to the profit of each item, and setting the values of $p_{i,z_j^m}, \eta_{z_j^m}$ and $T \cdot \mathcal{C}_i$ be negative. Since GAP is well-known as NP-hard [27], problem $[\mathcal{P}2]$ must also be NP-hard. ■

While if the set of activated sensors $\mathcal{H}$ is given, the set of sensors that require to be recharged, i.e., $\mathcal{H}_\psi$, can also be obtained. Then, the lower layer MCV charging route scheduling problem becomes

$$[\mathcal{P}3] : \min_{\mathcal{L}_{\mathcal{H}_\psi}} \sum_{g=0}^{|\mathcal{H}_\psi|} \gamma \cdot dist\left(a_g, a_{g+1}\right)$$
$$s.t.,\ (6b),\ (6c),\ (6d),\ (6e) \text{ and } (6h).$$

***Lemma** 2:* Problem $[\mathcal{P}3]$ is NP-hard.

Proof: Since the traveling salesman problem (TSP) is well-known as NP-hard [28], and problem $[\mathcal{P}3]$ is to find a TSP tour further constrained by the deadlines of visiting each target and the energy capacity of the MCV, meaning that problem $[\mathcal{P}3]$ is even more complicated than the conventional TSP, problem $[\mathcal{P}3]$ must be NP-hard. ■

With Lemmas 1 and 2, the following corollary is obvious.

***Corollary** 1:* Problem $[\mathcal{P}1]$ is NP-hard.

Since problem $[\mathcal{P}1]$ is NP-hard, it is impossible to solve it by brute force algorithms, which may result in a considerably high computational complexity. Thus, to circumvent this difficulty, we devise an efficient solution to problem $[\mathcal{P}1]$. Specifically, we first solve the MCV charging scheduling problem by applying a reinforcement learning (RL) approach, where the RL model only needs to be trained once at the beginning and can be used in later monitoring periods. Then, we develop a marginal product based approximation algorithm on top of the trained RL model, which can jointly optimize the sensor set selection and the MCV charging scheduling in an iterative manner.

B. MCV Charging Scheduling

Observed that, as a subproblem of $[\mathcal{P}1]$, $[\mathcal{P}3]$ is actually a sequential decision problem (where the charging route of the MCV can be determined in sequence), and thus the idea of reinforcement learning can be adopted in its solution design [37], [38]. Particularly, we can apply the actor-critic method [39] to train the corresponding neural network for eventually obtaining the solution.

1) Learning Model Construction: First of all, by considering that the g_{th} visiting target in the MCV's charging route is the output of the solution at g_{th} time step, we can formulate problem [$\mathcal{P}3$] in the form of a Markov decision process (MDP) [40], which is defined by a tuple $\{S_g, A_g, R, S_g^{'}\}$, where S_g is the state space at time step g (i.e., MCV's position, remaining energy, and targets that have already been visited by the MCV, etc.), A_g is the action space at time step g (i.e., targets that can be chosen to visit), R is the reward, and $S_g^{'}$ is the state transition after the agent executing the action at time step g. Then, at each time step g, the agent (i.e., the MCV in this paper) determines an action $a_g \in A_g$ (visiting target) according to the state S_g. When the termination state is reached (i.e., the MCV completes all charging tasks and returns to the depot) after executing a series of actions, the reward R can be calculated (i.e., in this paper, the reward can be calculated from the obtained charging route of the MCV). Hereafter, we describe the agent, state, action, policy, reward and system state transition for the considered MCV charging scheduling problem in detail.

Agent: The agent is the MCV that is responsible for making decisions to visit which target (i.e., which sensor or the depot) at each time step.

State: The state space includes locations, charging deadlines and energy demands of all sensors, and the position and remaining energy of the MCV. At each time step g, the state space can be defined as $S_g = \{s_g^0, s_g^1, \ldots, s_g^i, \ldots, s_g^{|\mathcal{H}_\psi|}\}$, where $|\mathcal{H}_\psi|$ indicates the number of sensors that need to be recharged. Each s_g^i is further represented by a sequence of tuples $\{s_g^i = (c^i, d_g^i)\}$, where c^i and d_g^i stand for the static and dynamic elements of target i, respectively. It is worth noting that dynamic elements are allowed to alter among different time steps, while static elements are invariant. To be more specific, c^i includes target i's 2-D coordinate and the charging deadline, d_g^i consists of the energy demand of target i (which becomes 0 after charging by the MCV) and the position and remaining energy of the MCV. Therefore, s_g^i can be viewed as a vector of features that depicts the state of target i at time step g. Particularly, s_g^0 represents attributes of the depot, which has a location at the center of the area, an infinite charging deadline, and no energy demand.

Action: For the MCV, its action signifies the next target to visit in its charging route.

Policy: The policy can be represented by a mapping function π which links an action a_g with a state S_g, i.e., $a_g = \pi(S_g)$.

Reward: Notice from the formulation of problem [$\mathcal{P}3$] that, besides the objective of minimizing the traveling energy consumption of the MCV, constraint (6b) imposes a strict requirement that the MCV should arrive at each sensor for recharging before its charging deadline expires. To tackle this, we first define the charging delay for each sensor $i \in \mathcal{H}_\psi$, as

$$del_i = \begin{cases} 0, & \text{if } \Lambda_i \leq ddl_i, \\ \Lambda_i - ddl_i, & \text{otherwise.} \end{cases} \tag{7}$$

Then, the reward (or actually the cost) of a charging route can be calculated by the sum of traveling energy consumption of the MCV and total charging delay, which can be expressed as

$$R = \sum_{g=0}^{|\mathcal{H}_\psi|} dist(a_g, a_{g+1}) + \sum_{g=1}^{|\mathcal{H}_\psi|} del_{a_g} \tag{8}$$

System State Transition: As mentioned in the definition of *State*, only the dynamic elements of a target are variant at every time step. The specific update process of the state is as follows. Recall that all sensors' energy demands are calculated according to (3) and the energy of the MCV is E_{MCV} at time step 0. At each time step g, MCV selects a sensor from the available set (i.e., $e_i^{demand}(g+1) \neq 0, \forall i \in \mathcal{H}_\psi$, where $e_i^{demand}(g+1)$ is the energy demand of sensor i at time step $g+1$) or the depot to visit in the next time step. After the MCV visit sensor $i \in \mathcal{H}_\psi$, the energy demands of sensors and residual energy of MCV are updated as

$$e_i^{demand}(g+1) = \begin{cases} 0, & \text{if } a_g = i, \\ e_i^{demand}(g), & \text{otherwise,} \end{cases}$$

$$e_{MCV}(g+1) = e_{MCV}(g) - e_{a_g}^{demand}(g).$$

Additionally, when the MCV visits the depot at the end of charging route, its energy is replenished, and thus we have

$$e_{MCV}(|\mathcal{H}_\psi| + 1) = E_{MCV}.$$

In summary, for solving problem [$\mathcal{P}3$], we can equivalently solve the above MDP, in which the objective is to minimize the reward (because it is actually the cost).

2) Reinforcement Learning based on Actor-Critic Method: To solve the formulated MDP, we construct two neural networks (NNs), i.e., actor NN and critic NN, which are associated with weight vectors θ and ξ, respectively. As shown in Fig. 3, firstly, the states (i.e., locations, charging deadlines and energy demands of sensors, and the position and remaining energy of the MCV) are inputted to the actor NN. Then, the actor NN calculates a probability distribution over the action space, and choose an action according the probability distribution. Afterward, the reward R can be obtained by calculating from the MCV implementing a series of actions (which forms the charging route of the MCV) in the environment (WRSN). Meanwhile, the reward approximation can be acquired from the critic NN according to the current state. Finally, the reward and its approximation are used together to update the actor NN and critic NN. The details of how the actor NN and the critic NN work are presented as follows.

Actor NN: As illustrated in Fig. 4, the actor NN consists of two main components, namely encoder and decoder. At each time step $g = 0, 1, \ldots, |\mathcal{H}_\psi| + 1$, the input is the state information, i.e., $S_g = \{s_g^0, s_g^1, \ldots s_g^{|\mathcal{H}_\psi|}\}$. Given this, the action a_g points to a sensor or the depot in S_g, determining the next visiting target. The states of sensors and the MCV in S_g are updated every time step after a target has been visited. When the energy demands of all sensors are satisfied, the process is terminated. The output of the model is a permutation of sensors and the depot, $\mathcal{L}_{\mathcal{H}_\psi} = \{a_0, a_1, ..., a_{|\mathcal{H}_\psi|}, ..., a_{|\mathcal{H}_\psi|+1}\}$.

To map input S_0 to output $\mathcal{L}_{\mathcal{H}_\psi}$, a probability chain rule

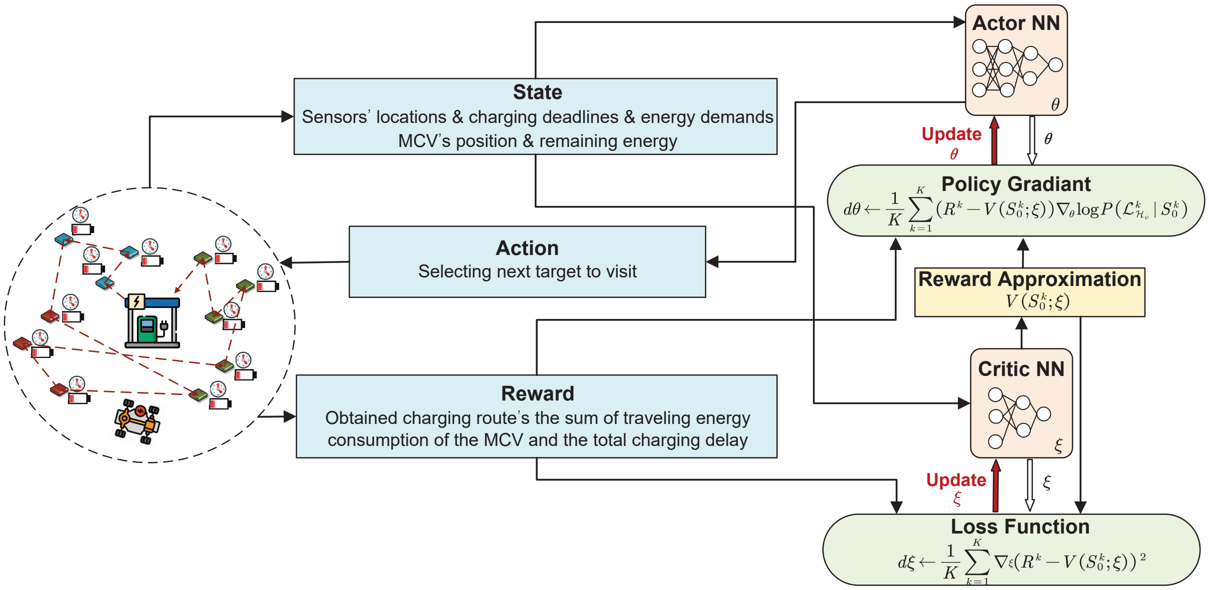
Actor NN
θ
Update θ
θ
State
Sensors' locations & charging deadlines & energy demands
MCV's position & remaining energy
Policy Gradiant
$d\theta \leftarrow \frac{1}{K}\sum_{k=1}^{K}(R^k - V(S_0^k;\xi))\nabla_\theta \log P(\mathcal{L}_{\mathcal{H}_\psi}^k \mid S_0^k)$
Action
Selecting next target to visit
Reward Approximation
$V(S_0^k;\xi)$
Critic NN
ξ
Reward
Obtained charging route's the sum of traveling energy consumption of the MCV and the total charging delay
Update ξ
ξ
Loss Function
$d\xi \leftarrow \frac{1}{K}\sum_{k=1}^{K}\nabla_\xi(R^k - V(S_0^k;\xi))^2$

Fig. 3. Actor-critic method for the MCV charging scheduling.

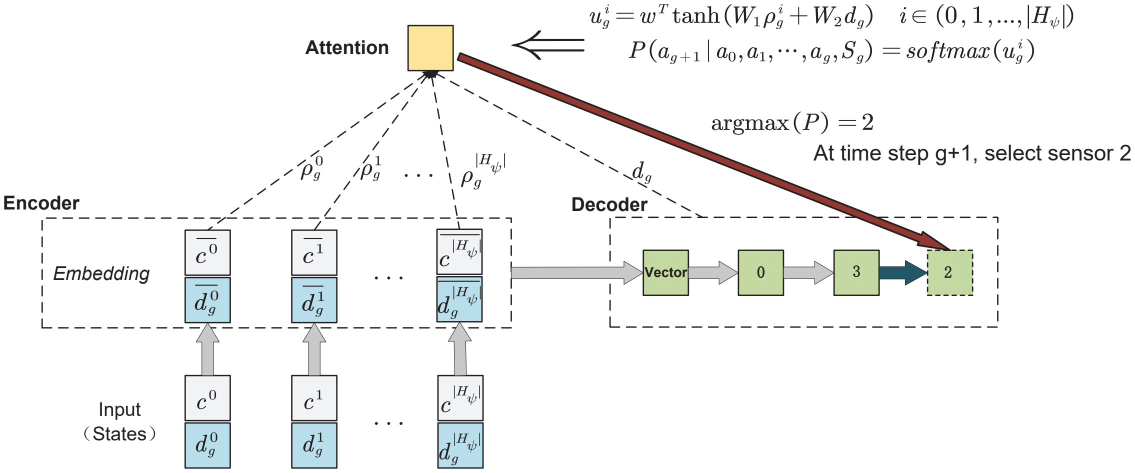

Fig. 4. Actor NN model for the MCV charging scheduling.

[41] is utilized:

$$P\left(\mathcal{L}_{\mathcal{H}_\psi} \mid X_0\right) = \prod_{g=0}^{|\mathcal{H}_\psi|} P\left(a_{g+1} \mid a_0, a_1, \cdots, a_g, S_g\right). \qquad (9)$$

Eq. (9) provides the probability of selecting the next visiting target conditioned on $a_0, a_1, \ldots, a_g$, i.e., the targets that have already been visited. A modified pointer network similar to that in [42] is used to further tackle (9). Its basic structure is the sequence-to-sequence model [43], a powerful model in the machine translation field, which can map one sequence to another. As a key enabler, both the encoder and decoder of this model have to be well designed.

Encoder is applied to encode the input sequence into a code vector containing the knowledge of the input. Since the attributes of the targets convey no sequential information and the order of targets in the input is trivial, unlike conventional works that commonly utilized the recurrent neural network (RNN), a simple embedding layer is adopted to encode the input which can significantly decrease the computational complexity without sacrificing the efficiency [42]. In this paper, we apply a 1-dimensional (1-D) convolution layer to encode the input to a high-dimensional vector.

For the decoder, since we need to store the knowledge of all previous steps $a_0, a_1, \cdots, a_g$ to assist for obtaining a_{g+1}, the RNN with gated recurrent unit (GRU) is adopted. The hidden state of RNN decoder d_g at any step g can memorize the previously selected visited targets. d_g is combined with $\rho_g^0, \rho_g^1, \ldots \rho_g^{|\mathcal{H}_\psi|}$ (which is the encoding of the inputs $s_g^0, s_g^1, \ldots s_g^{|\mathcal{H}_\psi|}$) to calculate the conditional probability $P\left(a_{g+1} \mid a_0, a_1, \cdots, a_g, S_g\right)$ over the next step of target selection. As shown in Fig. 4, to select the next target at step $g+1$, the hidden state d_g is obtained through the decoder. Afterward, in conjunction with $\rho_g^0, \rho_g^1, \ldots \rho_g^{|\mathcal{H}_\psi|}$, the conditional probability of determining the next target can be calculated by further applying the attention mechanism [44].

In general, the attention mechanism is utilized to calculate the degree of correlation of each input at the time step g. More $attention$ is given to the most relevant one which is more likely to be selected as the next target. Following such idea, in this work, its calculation can be expressed as

$$u_g^i = w^T \tanh\left(W_1 \rho_g^i + W_2 d_g\right), \quad i \in (0, 1, ..., |\mathcal{H}_\psi|),$$
$$P\left(a_{g+1} \mid a_0, a_1, \cdots, a_g, S_g\right) = \operatorname{softmax}\left(u_g^i\right),$$

where w, W_1, W_2 are $learnable$ parameters. For each target i, its u_g^i is computed by d_g and its encoder hidden state ρ_g^i, as shown in Fig. 4. The softmax operator is used to normalize $u_g^0, u_g^1, \ldots, u_g^{|\mathcal{H}_\psi|}$, and the probability for selecting each target i at time step g can then be obtained. For example, in Fig. 4, sensor 2 has the maximum conditional probability $P\left(a_{g+1} \mid a_0, a_1, \cdots, a_g, S_g\right)$ at step $g+1$, and hence it is selected as the next target.

Critic NN: Similar to [16], [42], the critic NN consists of four convolution layers as shown in Fig. 5, and all of them can be trained by a training process.

Here, we adopt the well-known actor-critic method to train the network. The policy π is parameterized by parameters θ, where θ is the vector of all learnable variables in the actor NN. It is worth noting that, during the training process, the actor NN selects the next target by sampling from the probability distribution instead of choosing the target with the highest probability in a greedy way. In addition, the critic NN with parameters ξ is used to evaluate the reward approximation given a specific state to iteratively improve the policy.

The training is conducted in an unsupervised way, and its detailed procedure is presented in Algorithm 1. During the training process, we generate problem instances from distributions $\{\Phi_{\Omega_1}, \Phi_{\Omega_2}, \ldots, \Phi_{\Omega_\mu}, \ldots\}$, where Ω signifies different input features of the targets, i.e., the targets' locations, sensing radii, charging deadlines, etc. K problem instances are sampled from $\{\Phi_{\Omega_1}, \Phi_{\Omega_2}, \ldots, \Phi_{\Omega_\mu}, \ldots\}$ for training the actor and critic neural networks. For each problem instance with different settings, the actor NN with current parameters θ produces the permutation of targets (charging route of the MCV) and obtains the corresponding reward. Then policy gradient is computed in line 11 to update the actor NN (refer to [39] for details of the formula of policy gradient). $V\left(X_0^k; \xi\right)$ here is the reward approximation of problem instance k calculated by the critic NN. After that, the critic NN is updated in line 12 by aiming to reduce the difference between the reward and its approximation.

By running Algorithm 1, all parameters of the actor NN and critic NN can be optimized, and the optimal charging route of the MCV can be obtained accordingly.

Algorithm 1: RL based on actor-critic training

```
   Output: The optimal model $\mathcal{M}^* = [\theta^*, \xi^*]$.
1  Initialize: Let the actor network with random weights $\theta$ and critic
     network with random weights $\xi$;
2  for episode $\leftarrow 1, 2, \ldots$ do
3      generate $K$ problem instances from $\{\Phi_{\Omega_1}, \Phi_{\Omega_2}, \ldots, \Phi_{\Omega_\mu,\ldots}\}$;
4      for $k \leftarrow 1, \ldots, K$ do
5          Initialize time step $g \leftarrow 0$; Get initial state $S_0^k$;
6          while not terminated do
7              select the next target $a_{g+1}^k$ according to
                 $P\left(a_{g+1}^k | a_1^k, \ldots, a_g^k, S_g^k\right)$;
8              Update $S_g^k$ to $S_{g+1}^k$ by leaving out the visited targets
                 and updating the residual energy of MCV;
9          compute reward $R^k$ ;
10     $d\theta \leftarrow \frac{1}{K}\sum_{k=1}^K \left(R^k - V\left(S_0^k; \xi\right)\right) \nabla_\theta \log P\left(\mathcal{L}_{\mathcal{H}_\psi}^k \mid S_0^k\right)$;
11     $d\xi \leftarrow \frac{1}{K}\sum_{k=1}^K \nabla_\xi \left(R^k - V\left(S_0^k; \xi\right)\right)^2$;
12     Update $\theta$ using $d\theta$ and $\xi$ using $d\xi$;
13 Determine $\theta^* = \theta, \xi^* = \xi$.
```

Algorithm 2: The implementation of JSACS

```
Output: $\mathcal{H}$, $\mathcal{L}_{\mathcal{H}_\psi}$.
Initialize: Let $\mathcal{N}_{z_j^m}^{can} = \{i | p_{i,z_j^m} \neq 0, \forall i \in \mathcal{N}\}$,
  $\mathcal{N}^{can} = \sum_{z_j^m \in \mathcal{Z}} \mathcal{N}_{z_j^m}^{can}$, $\mathcal{Z}^{un} = \mathcal{Z}$, $\mathcal{H}_\phi(0) = \emptyset$,
  $\mathcal{H}_\psi(0) = \emptyset, \mathcal{H}(0) = \emptyset$, $\mathcal{L}_{\mathcal{H}_\psi(0)} = \emptyset$, $\tau = 0$;
while $\mathcal{Z}^{un}$ is nonempty do
    $\tau = \tau + 1$;
    for each $i \in \mathcal{N}^{can}$ do
        if $E_i^{initial} - E_i^{min} \geq T \cdot \mathcal{C}_i$ then
            $h = \emptyset$;
        else
            $h = \{i\}$;
            Call the model $\mathcal{M}^* = [\theta^*, \xi^*]$ in Algorithm 1 to get
              a charging route $\mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup h}$ which meets each
              sensor's charging deadline (If there is no charging
              route that meets the sensor's charging deadline or the
              energy consumption of the MCV exceeds $E_{MCV}$,
              delete the sensor $i$ from $\mathcal{N}^{can}$);
    $i_{selected} =$
    $\arg\max_{i \in \mathcal{N}^{can}} \frac{\sigma(\mathcal{H}(\tau-1) \cup \{i\}) - \sigma(\mathcal{H}(\tau-1))}{E_{total}(\mathcal{H}(\tau-1) \cup \{i\}, \mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup h}) - E_{total}(\mathcal{H}(\tau-1), \mathcal{L}_{\mathcal{H}_\psi(\tau-1)})}$;
    Update $\mathcal{H}(\tau) = \mathcal{H}(\tau-1) \cup \{i_{selected}\}$;
    if $E_{i_{selected}}^{initial} - E_{i_{selected}}^{min} \geq T \cdot E_{i_{selected}}^{consume}$ then
        Update $\mathcal{H}_\phi(\tau) = \mathcal{H}_\phi(\tau-1) \cup \{i_{selected}\}$,
          $\mathcal{L}_{\mathcal{H}_\psi(\tau)} = \mathcal{L}_{\mathcal{H}_\psi(\tau-1)}$ ;
    else
        Update $\mathcal{H}_\psi(\tau) = \mathcal{H}_\psi(\tau-1) \cup \{i_{selected}\}$,
          $\mathcal{L}_{\mathcal{H}_\psi(\tau)} = \mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup\{i_{selected}\}}$;
    for each $z_j^m \in \mathcal{Z}^{un}$ do
        if $1 - \prod_{i \in \mathcal{H}(\tau)} (1 - p_{i,z_j^m}) \geq \eta_{z_j^m}$ then
            Update $\mathcal{N}^{can} = \mathcal{N}^{can} \backslash \{\mathcal{N}_{z_j^m}^{can}\}$,
              $\mathcal{Z}^{un} = \mathcal{Z}^{un} \backslash \{z_j^m\}$;
    Update $\mathcal{N}^{can} = \mathcal{N}^{can} \backslash \{i_{selected}\}$;
Determine $\mathcal{H} = \mathcal{H}(\tau)$, $\mathcal{L}_{\mathcal{H}_\psi} = \mathcal{L}_{\mathcal{H}_\psi(\tau)}$.
```

C. Joint Optimization Algorithm

Based on the solution to problem $[\mathcal{P}3]$, we are now ready to design a joint sensor activation and charging scheduling scheme (called JSACS) to the original two-layer optimization problem $[\mathcal{P}1]$. The core idea of JSACS is to iteratively add a sensor to $\mathcal{H}$ and calculate the corresponding charging route of the MCV until each task's QoM, i.e., $\eta_{z_j^m}, \forall z_j^m \in \mathcal{Z}$, is met. Denote the selected sensor set as $H(\tau) = H_\phi(\tau) \cup H_\psi(\tau)$ at the end of iteration τ, where $H_\phi(\tau)$ and $H_\psi(\tau)$ are the selected sensor sets that do not need to be charged and need to be charged, respectively. Initially, the candidate sensor set is denoted by $\mathcal{N}^{can} = \sum_{z_j^m \in \mathcal{Z}} \mathcal{N}_{z_j^m}^{can}$, where $\mathcal{N}_{z_j^m}^{can} =$

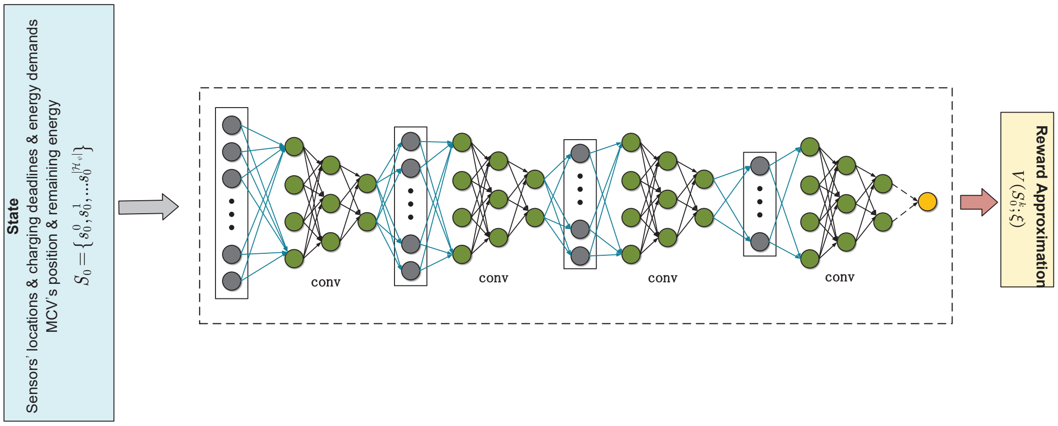
State
Sensors' locations & charging deadlines & energy demands
MCV's position & remaining energy
$S_0 = \{s_0^0, s_0^1, \ldots s_0^{|\mathcal{H}_\psi|}\}$
conv
conv
conv
conv
Reward Approximation
$V(S_0^k; \xi)$

Fig. 5. Critic NN model for the MCV charging scheduling.

$\{i|p_{i,z_j^m} \neq 0, \forall i \in \mathcal{N}\}$, and the set of QoM unsatisfied tasks is $\mathcal{Z}^{un} = \mathcal{Z}$. In addition, $\mathcal{H}(0) = \emptyset$, $\mathcal{H}_\phi(0) = \emptyset$ and $\mathcal{H}_\psi(0) = \emptyset$. Besides, the charging route $\mathcal{L}_{\mathcal{H}_\psi(0)} = \emptyset$ and $\tau = 0$. In each iteration τ, for each sensor $i \in \mathcal{N}^{can}$, we update the charging route by adding sensor i to $\mathcal{H}_\psi(\tau-1)$, i.e., $\mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup h}$, where h indicates whether this sensor has to be recharged or not:

$$h = \begin{cases} \emptyset, & \text{if } E_i^{initial} - E_i^{min} \geq T \cdot \mathcal{C}_i, \\ \{i\}, & \text{otherwise.} \end{cases}$$

After that, among all candidate sensors, the sensor with the largest marginal product[1] is selected in this iteration. Here, the marginal product is defined as the increase of sum of QoM of all tasks brought by consuming a unit of energy of the entire system. Mathematically, in each iteration τ, a new sensor is selected according to

$$\arg\max_{i \in \mathcal{N}^{can}} \frac{\sigma(\mathcal{H}(\tau-1) \cup \{i\}) - \sigma(\mathcal{H}(\tau-1))}{E_{total}(\mathcal{H}(\tau-1) \cup \{i\}, \mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup h}) - E_{total}(\mathcal{H}(\tau-1), \mathcal{L}_{\mathcal{H}_\psi(\tau-1)})},$$

where $\sigma(\mathcal{H}(\tau)) = \sum_{z_j^m \in \mathcal{Z}} \sigma_{z_j^m}(\mathcal{H}(\tau))$ represents the sum of QoM obtained by all tasks at the end of iteration τ and, $\sigma_{z_j^m}(\mathcal{H}(\tau)) = \min\{1 - \prod_{i' \in \mathcal{H}(\tau)}(1 - p_{i',z_j^m}), \eta_{z_j^m}\}$. Then, $\mathcal{H}(\tau)$ is updated. If the newly selected sensor $i_{selected}$ satisfies $E_{i_{selected}}^{initial} - E_{i_{selected}}^{min} \geq T \cdot E_{i_{selected}}^{consume}$, it is added to $\mathcal{H}_\phi(\tau)$, and the charging route is updated as $\mathcal{L}_{\mathcal{H}_\psi(\tau)} = \mathcal{L}_{\mathcal{H}_\psi(\tau-1)}$. Otherwise, $i_{selected}$ is added to $\mathcal{H}_\psi(\tau)$, and the charging route is updated as $\mathcal{L}_{\mathcal{H}_\psi(\tau)} = \mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup\{i_{selected}\}}$. Afterwards, we check whether each task's QoM $\eta_{z_j^m}, \forall z_j^m \in \mathcal{Z}^{un}$ is met or not, and exclude set $\mathcal{N}_{z_j^m}^{can}$ from set $\mathcal{N}^{can}$ if a task's QoM $\eta_{z_j^m}$ has already been met. If all tasks's QoM are satisfied, the iteration process stops. Otherwise, by excluding the selected sensor $i_{selected}$ from set $\mathcal{N}^{can}$, the iteration process continues. Algorithm 2 summarizes all detailed steps (pseudo code) of the proposed JSACS.

V. Performance Analysis

In this section, we first conduct a series of theoretical analysis on the monotonicity and submodularity of $\sigma(\mathcal{H})$, i.e., the sum of QoM of all tasks. Based on these, we then prove that the proposed scheme, i.e., JSACS, is not only computationally efficient but also theoretically bounded with a guaranteed performance, particularly for the upper-layer marginal product approximation.

A. Properties of $\sigma(\mathcal{H})$

Hereafter, we show that $\sigma(\mathcal{H})$ is not only monotone but also submodular. For clarity, we introduce two sets $\bar{\mathcal{A}} \subseteq \mathcal{N}^{can}$ and $\bar{\mathcal{B}} \subseteq \mathcal{N}^{can}$, which are both the subsets of activated sensors to execute tasks in $\mathcal{Z}$.

Definition 1: Given a finite ground set $\mathcal{N}^{can}$, a real-valued set function is defined as $\sigma(\mathcal{H}) : 2^{\mathcal{N}^{can}} \to R$, and $\sigma(\mathcal{H})$ is monotonically nondecreasing and submodular if only if it satisfies the following conditions, respectively [46].

1) $\sigma(\bar{\mathcal{A}} \cup \{i\}) \geq \sigma(\bar{\mathcal{A}})$, for $\bar{\mathcal{A}} \subseteq \mathcal{N}^{can}$ and $i \in \mathcal{N}^{can} \backslash \bar{\mathcal{A}}$ (monotone);
2) $\sigma(\bar{\mathcal{A}} \cup \{i\}) - \sigma(\bar{\mathcal{A}}) \geq \sigma(\bar{\mathcal{B}} \cup \{i\}) - \sigma(\bar{\mathcal{B}})$, for $\bar{\mathcal{A}} \subseteq \bar{\mathcal{B}} \subseteq \mathcal{N}^{can}$, $i \in \mathcal{N}^{can} \backslash \bar{\mathcal{B}}$ (submodular).

Theorem 1: $\sigma(\mathcal{H})$ is monotone and submodular.

Proof: After adding sensor $i \in \mathcal{N}^{can}$ into $\bar{\mathcal{A}} \subseteq \mathcal{N}^{can}$ to execute task $z_{j'}^{m'} \in \mathcal{Z}^{un}$, we have $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}} \cup \{i\}) = \min\{1 - \prod_{i' \in \bar{\mathcal{A}} \cup \{i\}}(1 - p_{i',z_{j'}^{m'}}), \eta_{z_{j'}^{m'}}\}$, while other tasks' QoM are constant. Additionally, $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}) = \min\{1 - \prod_{i' \in \bar{\mathcal{A}}}(1 - p_{i',z_{j'}^{m'}}), \eta_{z_{j'}^{m'}}\}$, and thus $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}} \cup \{i\}) \geq \sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}})$. According to the definition of $\sigma(\mathcal{H})$, we have the following inequality:

$$\sigma(\bar{\mathcal{A}} \cup \{i\}) = \textstyle\sum_{z_j^m \in \mathcal{Z}}(\sigma_{z_j^m}(\bar{\mathcal{A}} \cup \{i\})) \geq \sigma(\bar{\mathcal{A}}) = \sum_{z_j^m \in \mathcal{Z}}(\sigma_{z_j^m}(\bar{\mathcal{A}})), i \in \mathcal{N}^{can},$$

which implies that $\sigma(\mathcal{H})$ is monotone.

Then, our remaining issue is to prove that $\sigma(\mathcal{H})$ is also submodular by showing that $\sigma(\bar{\mathcal{A}} \cup \{i\}) - \sigma(\bar{\mathcal{A}}) \geq \sigma(\bar{\mathcal{B}} \cup \{i\}) - \sigma(\bar{\mathcal{B}})$, for $\bar{\mathcal{A}} \subseteq \bar{\mathcal{B}} \subseteq \mathcal{N}^{can}$, $i \in \mathcal{N}^{can} \backslash \bar{\mathcal{B}}$. Considering that adding sensor $i \in \mathcal{N}^{can} \backslash \bar{\mathcal{B}}$ into both sets $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ to execute task $z_{j'}^{m'} \in \mathcal{Z}^{un}$, it is equivalent to show that

$$\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}} \cup \{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}) \geq \sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}} \cup \{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}). \quad (10)$$

[1] Marginal product is a concept in economics, which refers to the increase in the total output brought by adding a unit of an input, assuming that the quantities of other inputs are maintained as constant [45].

Since the QoM of task $z_{j'}^{m'}$ is can not be met by sets $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$, and following from the above analysis (i.e., $\sigma(\mathcal{H})$ is monotone), we have

$$\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}) \le \sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}) < \eta_{z_{j'}^{m'}}. \tag{11}$$

Then, from (11), we can easily prove that

$$\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}}) \ge \prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}}). \tag{12}$$

Since there are three different relationships between $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\})$ and $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\})$, we analyze the submodularity of $\sigma(\mathcal{H})$ in three cases. For notation simplicity, let $\alpha = (1-p_{i,z_{j'}^{m'}})$.

Case 1 (If $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\}) < \eta_{z_{j'}^{m'}}$): In this case, we have $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\}) < \eta_{z_{j'}^{m'}}$.

Meanwhile, we can derive that

$$\begin{aligned}&\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}})\\ &= [1-\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})\cdot\alpha] - [1-\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})]\\ &= [\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})]\cdot[1-\alpha],\end{aligned} \tag{13}$$

and

$$\begin{aligned}&\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}})\\ &= [1-\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})\cdot\alpha] - [1-\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})]\\ &= [\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})]\cdot[1-\alpha].\end{aligned} \tag{14}$$

Then, according to (12), we can easily obtain (10).

Case 2 (If $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\}) = \eta_{z_{j'}^{m'}}$ and $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\}) < \eta_{z_{j'}^{m'}}$): In this case, we have

$$\begin{aligned}&\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}})\\ &= [1-\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})\cdot\alpha] - [1-\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})]\\ &= [\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})]\cdot[1-\alpha],\end{aligned} \tag{15}$$

and

$$\begin{aligned}&\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}})\\ &= \eta_{z_{j'}^{m'}} - [1-\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})]\\ &\le [1-\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})\cdot\alpha] - [1-\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})].\end{aligned} \tag{16}$$

Similar to Case 1, this obviously indicates (10).

Case 3 (If $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\}) = \eta_{z_{j'}^{m'}}$ and $\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\}) = \eta_{z_{j'}^{m'}}$): In this case, we have

$$\begin{aligned}&\sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}}\cup\{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{A}})\\ &= \eta_{z_{j'}^{m'}} - [1-\prod_{i'\in\bar{\mathcal{A}}}(1-p_{i',z_{j'}^{m'}})],\end{aligned} \tag{17}$$

and

$$\begin{aligned}&\sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}}\cup\{i\}) - \sigma_{z_{j'}^{m'}}(\bar{\mathcal{B}})\\ &= \eta_{z_{j'}^{m'}} - [1-\prod_{i'\in\bar{\mathcal{B}}}(1-p_{i',z_{j'}^{m'}})].\end{aligned} \tag{18}$$

Then, according to (11), we can also obtain (10).

All these together complete the proof. ■

B. Performance of JSACS

Based on the shown properties of $\sigma(\mathcal{H})$ (i.e., monotonicity and submodularity), we can theoretically analyze the worst-case performance of JSACS in terms of the approximation ratio. Besides, the computational complexity of JSACS is also analyzed. Note that, for notation simplicity, in the following, let $\eta = \sum_{z_j^m\in\mathcal{Z}}\eta_{z_j^m}$.

Lemma 3: In JSACS, at the beginning of any iteration τ, there must be a sensor $i\in\mathcal{N}^{can}\backslash\mathcal{H}(\tau-1)$ which satisfies

$$\frac{\sigma(\mathcal{H}(\tau-1)\cup\{i\}) - \sigma(\mathcal{H}(\tau-1))}{\bar{c}(i)} \ge \frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}, \tag{19}$$

where $\bar{c}(i) = E_{total}(\mathcal{H}(\tau-1)\cup\{i\},\mathcal{L}_{\mathcal{H}_\psi(\tau-1)\cup h}) - E_{total}(\mathcal{H}(\tau-1),\mathcal{L}_{\mathcal{H}_\psi(\tau-1)})$ indicates the energy consumption brought by adding sensor i into $\mathcal{H}(\tau-1)$, and $\eta_{\tau-1}$ depicts the shortfall in the total QoM requirement after $\tau-1$ iterations, which can be computed as

$$\eta_{\tau-1} = \eta - \sigma(\mathcal{H}(\tau-1)). \tag{20}$$

Besides, $E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})$ is the minimum energy consumption, where $\mathcal{H}^*$ and $\mathcal{L}^*_{\mathcal{H}^*_\psi}$ denote the theoretically optimal selected sensor set and the corresponding optimal charging route computing by our proposed RL based MCV charging scheduling algorithm, respectively.

Proof: Let $\mathcal{H}^*(\tau-1) = \mathcal{H}^* - \mathcal{H}(\tau-1)$, and $\mathcal{H}^*(\tau-1) = \{y_1,\dots,y_r\}$, where r is the amount of sensors in $\mathcal{H}^*(\tau-1)$. We start the proof by a hypothesis: suppose that for any sensor $i\in\mathcal{N}^{can}\backslash\mathcal{H}(\tau-1)$,

$$\frac{\sigma(\mathcal{H}(\tau-1)\cup\{i\}) - \sigma(\mathcal{H}(\tau-1))}{\bar{c}(i)} < \frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}. \tag{21}$$

Consider adding sensors in $\mathcal{H}^*(\tau-1)$ to $\mathcal{H}(\tau-1)$ one by one. At any step, $r' < r$, we have by submodularity of $\sigma(\mathcal{H})$ that $\sigma(\mathcal{H}(\tau-1)\cup\{y_1,\dots,y_{r'}\}) - \sigma(\mathcal{H}(\tau-1)\cup\{y_1,\dots,y_{r'-1}\}) \le \sigma(\mathcal{H}(\tau-1)\cup\{y_{r'}\}) - \sigma(\mathcal{H}(\tau-1))$. Additionally, according to the hypothesis (21), we have $\sigma(\mathcal{H}(\tau-1)\cup\{y_1,\dots,y_{r'}\}) - \sigma(\mathcal{H}(\tau-1)\cup\{y_1,\dots,y_{r'-1}\}) \le \sigma(\mathcal{H}(\tau-1)\cup\{y_{r'}\}) - \sigma(\mathcal{H}(\tau-1)) < \bar{c}(y_{r'})\cdot\frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}$. Adding all sensors from $\mathcal{H}^*(\tau-1)$ into $\mathcal{H}(\tau-1)$, this yields $\sigma(\mathcal{H}(\tau-1)\cup\{y_1,\dots y_r\}) - \sigma(\mathcal{H}(\tau-1)) \le \frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}\cdot(\bar{c}(y_1)+\cdots+\bar{c}(y_r))$ resulting in $\sigma(\mathcal{H}(\tau-1)\cup\{y_1,\dots y_r\}) <$

$\sigma(\mathcal{H}(\tau-1)) + \frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})} \cdot \sum_{1\le h\le r} \bar{c}(y_h) \le \eta$ which is contradicted with the fact that $\sigma(\mathcal{H}(\tau-1) \cup \{y_1, \dots y_r\}) \ge \eta$ in reality.

In conclusion, since the hypothesis does not hold, there must exist at least one sensor $i \in \mathcal{N}^{can} \backslash \mathcal{H}(\tau-1)$ which satisfies

$$\frac{\sigma(\mathcal{H}(\tau-1) \cup \{i\}) - \sigma(\mathcal{H}(\tau-1))}{\bar{c}(i)} \ge \frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi})}.$$

This completes the proof. ■

***Theorem** 2:* Given a candidate sensor set $\mathcal{N}^{can}$ with a cost function $E_{total}(\mathcal{H}, \mathcal{L}_{\mathcal{H}_\psi})$. Let $\varepsilon > 0$ be any shortfall and $\sigma(\mathcal{H}(l)) \ge \eta - \varepsilon$. Then, we must have $\frac{E_{total}(\mathcal{H}(l),\mathcal{L}_{\mathcal{H}_\psi(l)})}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})} < (1 + \ln(\eta/\varepsilon))$.

Proof: We add sensor $i \in \mathcal{N}^{can} \backslash \mathcal{H}(\tau-1)$ which provides the maximum marginal product into $\mathcal{H}(\tau-1)$ according to JSACS at iteration τ. Then, following Lemma 3, we have

$$\frac{\sigma(\mathcal{H}(\tau-1) \cup \{i\}) - \sigma(\mathcal{H}(\tau-1))}{\bar{c}_\tau} \ge \frac{\eta_{\tau-1}}{E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi})}, \tag{22}$$

where $\bar{c}_\tau$ is the energy consumption introduced by adding the selected sensor i within iteration τ. Recalling from (20), we have

$$\sigma(\mathcal{H}(\tau-1) \cup \{i\}) = \eta - \eta_\tau, \tag{23}$$

and

$$\sigma(\mathcal{H}(\tau-1)) = \eta - \eta_{\tau-1}. \tag{24}$$

Substituting (23) and (24) into (22), we have

$$\eta_\tau \le \eta_{\tau-1} \cdot (1 - \frac{\bar{c}_\tau}{E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi})}). \tag{25}$$

By applying the well known inequality $(1+\omega) \le e^\omega$, $\forall \omega \in R$, we can further convert inequality (25) to

$$\eta_\tau \le e^{\frac{-\bar{c}_\tau}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}} \cdot \eta_{\tau-1}, \tag{26}$$

which by expanding yields

$$\eta_\tau \le e^{\frac{-E_{total}(\mathcal{H}(\tau),\mathcal{L}_{\mathcal{H}_\psi(\tau)})}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}} \cdot \eta. \tag{27}$$

Assume that Algorithm 2 (JSACS) takes l iterations to achieve $\sigma(\mathcal{H}(l)) \ge \eta - \varepsilon$ such that

$$\sigma(\mathcal{H}(l-1)) < \eta - \varepsilon, \tag{28}$$

Then, according to (20) and (28), we have

$$\eta_{l-1} > \varepsilon. \tag{29}$$

Combining inequalities (27) and (29) yields

$$\eta \cdot e^{-\frac{E_{total}(\mathcal{H}(l-1),\mathcal{L}_{\mathcal{H}_\psi(l-1)})}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})}} > \varepsilon, \tag{30}$$

which implies that

$$E_{total}(\mathcal{H}(l-1), \mathcal{L}_{\mathcal{H}_\psi(l-1)}) < E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi}) \cdot \ln(\eta/\varepsilon). \tag{31}$$

TABLE II
MAIN SIMULATION PARAMETERS.

Parameter	Value
Area dimensions	40 m × 40 m
Task types j	[0,1,2,3]
Number of tasks M	100 (randomly chosen over [0,1,2,3])
QoM requirement $\eta_{z^m_j}$	0.5 ∼ 1.0
Time duration of monitoring period T	2 hours
Sensor types	[0,1,2,3]
Number of sensors N	1000 (number of each type: 250)
Sensing radius R_i	randomly chosen over [2,4,6,8] m
Energy capacity $E^{capacity}_i$	10.8 kJ
Energy consumption rate $E^{consume}_i$	0.5 J/s
Minimum energy E^{min}_i	250 J
Initial energy $E^{initial}_i$	3000 ∼ 4000 J
Intensity coefficient α_i	0.01 ∼ 0.05
Charging rate δ	30 W
Velocity v	5 m/s
Traveling energy consumption γ	50 J/m
Energy capacity of MCV E_{MCV}	200 kJ
Learning rate of Actor NN	5e-4
Learning rate of Critic NN	5e-4
Batch size	200
Size of training set	120000
Size of validation set	1000
Size of test set	1000

Besides, at any iteration, we always have $\sigma(\mathcal{H}(\tau)) - \sigma(\mathcal{H}(\tau-1)) \le \eta_{\tau-1}$. Thus, we can obtain $\bar{c}_\tau \le E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi})$ according to inequality (22), and we can transform it to

$$E_{total}(\mathcal{H}(l), \mathcal{L}_{\mathcal{H}_\psi(l)}) \le E_{total}(\mathcal{H}(l-1), \mathcal{L}_{\mathcal{H}_\psi(l-1)}) + E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi}). \tag{32}$$

Furthermore, combining inequalities (31) and (32) gives

$$\frac{E_{total}(\mathcal{H}(l), \mathcal{L}_{\mathcal{H}_\psi(l)})}{E^*_{total}(\mathcal{H}^*, \mathcal{L}^*_{\mathcal{H}^*_\psi})} < (1 + \ln(\eta/\varepsilon)). \tag{33}$$

This completes the proof. ■

***Corollary** 2:* By applying Algorithm 2 (JSACS), the approximation ratio of the solution can be derived as, $\frac{E_{total}(\mathcal{H}(\mathcal{T}),\mathcal{L}_{\mathcal{H}_\psi(\mathcal{T})})}{E^*_{total}(\mathcal{H}^*,\mathcal{L}^*_{\mathcal{H}^*_\psi})} < (1 + \ln(\eta/\varepsilon))$, where $\mathcal{T}$ is the last iteration of Algorithm 2 (JSACS).

***Theorem** 3:* The computational complexity of the proposed scheme, i.e., JSACS, is bounded by $O(|\mathcal{N}^{can}|^2)$, where $|\mathcal{N}^{can}|$ is the cardinality of the candidate sensor set.

Proof: From the proposed Algorithm 2, we can observe that the number of the overall iterations is bounded by $|\mathcal{N}^{can}|$. Moreover, within each iteration, it takes at most $O(|\mathcal{N}^{can}|)$ iterations to find a sensor $i \in \mathcal{N}^{can}$ with the maximum marginal product from $\mathcal{N}^{can}$. Thus, the total time complexity is bounded by $O(|\mathcal{N}^{can}|) \cdot |\mathcal{N}^{can}| = O(|\mathcal{N}^{can}|^2)$. ■

VI. SIMULATION RESULTS

In this section, simulations are conducted to numerically evaluate the performance of the proposed JSACS. Table II lists the values of main simulation parameters. Similar settings have also been employed in the literature [47]–[49]. Besides, the structure of neural networks (including actor and critic NNs) and hyper-parameter settings for the proposed learning based algorithm is presented in Table III. Note that some parameters may vary according to different evaluation scenarios.

For comparison purpose, we introduce a greedy algorithm (GRE) and an existing algorithm called reward-cost ratio

TABLE III
PARAMETER SETTINGS OF THE PROPOSED RL MODEL

	Actor Neural network (Pointer Network)
Encoder:	1D-Conv (input size = D_{input}, hidden size = 128, kernel size = 1, stride = 1)
Decoder:	GRU (hidden size = 128, number of layer = 1)
	Attention (No hyper parameters)
	Critic Neural Network
	1D-Conv (input size = D_{input}, hidden size = 128, kernel size = 1, stride = 1)
	1D-Conv (input size = 128, hidden size = 20, kernel size = 1, stride = 1)
	1D-Conv (input size = 20, hidden size = 20, kernel size = 1, stride = 1)
	1D-Conv (input size = 20, hidden size = 1, kernel size = 1, stride = 1)

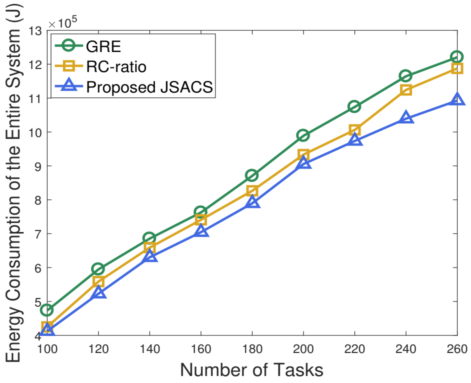

Fig. 6. Comparison of energy consumption of the entire system w.r.t. number of tasks.

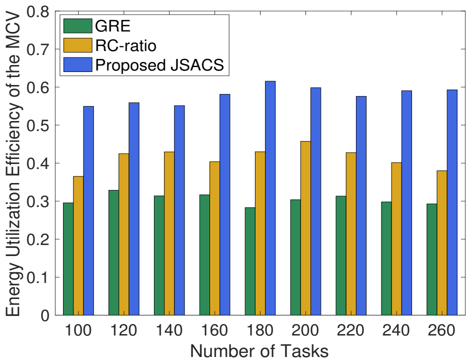

Fig. 7. Comparison of energy utilization efficiency of the MCV w.r.t. number of tasks.

algorithm (RC-ratio) [30] as benchmarks. GRE greedily selects sensors into $\mathcal{H}$ that have maximum detection probability until all tasks' QoM are satisfied, and then applies the earliest deadline first policy (EDF) [50] to derive the charging route of the MCV for $\mathcal{H}_{\psi}$. RC-ratio selects sensors into $\mathcal{H}$ for achieving the marginal product maximization (which is similar to our proposed JSACS) while its MCV's charging route is determined by EDF.

Fig. 6 demonstrates the superiority of the proposed JSACS in terms of the system energy consumption with respect to the number of tasks. It is shown that, the total energy consumption increases monotonically with the number of tasks. This is because with the growth of the number of tasks, more sensors need to be activated, leading to more energy consumption. Meanwhile, with more sensors being activated, a growing number of them need to be recharged, resulting in the increase of the MCV's traveling energy consumption. Additionally, it can be observed that the proposed JSACS outperforms both GRE and RC-ratio. The reason is that GRE iteratively select a sensor with the highest detection probability to activate while ignoring the impact on the total energy consumption. RC-ratio outperforms GRE because it selects a sensor with the largest marginal product to activate in each iteration, balancing the trade off of maximizing the detection probability and minimizing the energy consumption. The proposed JSACS achieves the best performance because it does not only select a sensor with the largest marginal product in each iteration, but also determines the charging route of the MCV by a well trained reinforcement learning model instead of the EDF.

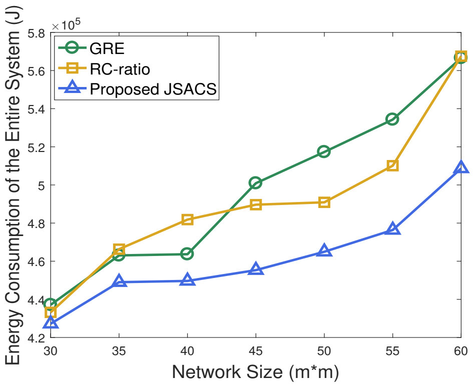

Fig. 8. Comparison of energy consumption of the entire system w.r.t. network size.

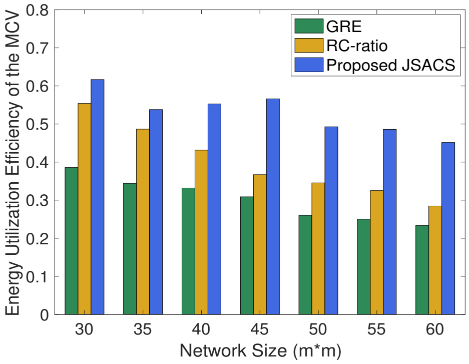

Fig. 9. Comparison of energy utilization efficiency of the MCV w.r.t. network size.

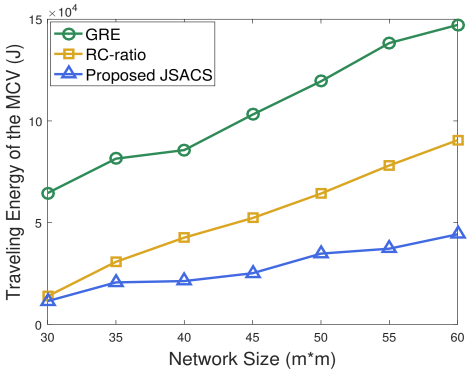

Fig. 10. Comparison of traveling energy of the MCV w.r.t. network size.

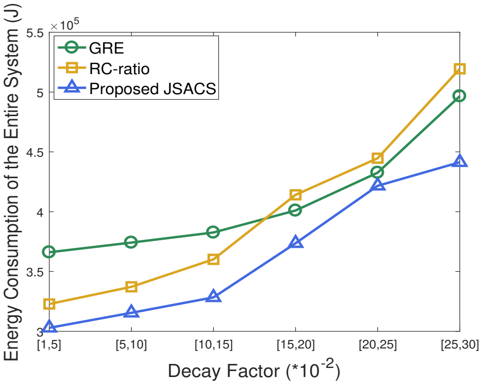

Fig. 11. Comparison of energy consumption of the entire system w.r.t. decay factor.

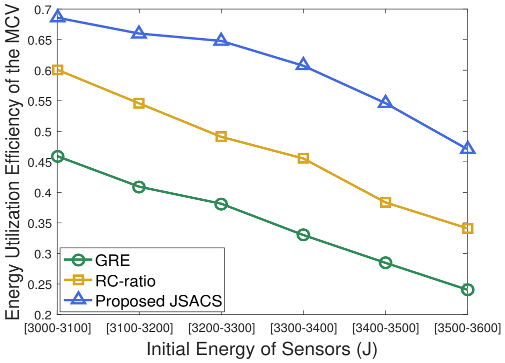

Fig. 12. Comparison of energy utilization efficiency of the MCV w.r.t. initial energy of sensors.

Fig. 7 compares the energy utilization efficiency of GRE, RC-ratio and the proposed JSACS with the increase of the number of tasks. The energy utilization efficiency refers to the proportion of the energy for recharging sensors to total MCV energy consumption. It is shown that the proposed JSACS performs better than GRE and RC-ratio. The reason is that the proposed JSACS considers the two-layer optimization simultaneously rather than in greedy or independent ways. In addition, the objective of the trained reinforcement learning model is to minimize the traveling energy consumption of the MCV while meeting the charging deadlines of all activated sensors. On the contrary, the EDF applied in GRE and RC-ratio does not consider the traveling length of the MCV, and it simply recharges sensors in a timely manner. Therefore, the proposed JSACS can prompt the MCV to utilize more energy for task execution to increase the QoM of tasks, instead of wasting the energy on traveling.

Fig. 8 shows that the energy consumption of the entire system for all these three schemes increases with the network size. Besides, the energy utilization efficiency of the MCV has a downward trend despite that the proposed JSACS is always superior to GRE and RC-ratio, as shown in Fig. 9. The reason is that the larger network size makes the sensor deployment more sparse, leading to more energy consumption on traveling, as later shown in Fig. 10. In addition, a larger network size also makes the distance between the sensor and its monitoring tasks larger, and the detection probabilities of sensors decrease, so that more sensors need to be activated to execute tasks, inducing more energy consumption on sensors. These imply that the proposed JSACS is more suitable to be applied in ultra-dense network scenarios, which is also of great importance in the future IIoT. Intuitively, the proposed JSACS outperforms both GRE and RC-ratio, benefiting from integrating reinforcement learning and marginal product based approximation algorithms to jointly solve the sensor activation and charging scheduling problem.

Fig. 11 illustrates the performance of three schemes in terms of the system energy consumption with respect to the decay factor. It can be observed that the total energy consumption increases with the decay factor. This is because, when QoM requirements of all tasks are fixed, a larger decay factor results in lower sensors' detection probabilities, so that more sensors are required to be activated for executing tasks. Besides, Fig. 12 shows that as the initial energy of sensors increases, the energy utilization efficiency of MCV decreases. The reason is that when sensors have more initial energies, their demands for recharging become less imperative. Moreover, Fig. 11 and 12 demonstrate that the proposed JSACS outperforms all benchmark schemes, and the explanations for this are similar to those for Fig. 6 and 7.

VII. Conclusion and Future Work

In this paper, a joint optimization of sensor activation and mobile charging scheduling for the WRSN-based IIoT system has been studied. By considering the objective of minimizing the total energy consumption of the entire system subject to tasks' QoM requirements, sensor charging deadlines and the energy capacity of the MCV, a novel scheme, called JSACS, is proposed integrating reinforcement learning and marginal product based approximation algorithms. Theoretical analyses show that, JSACS is not only computationally efficient but also theoretically bounded with a guaranteed performance in terms of the approximation ratio. Simulation results further demonstrate that, compared to counterparts, the proposed JSACS is superior in decreasing the system energy consumption and improving the energy utilization efficiency of the MCV.

In the future work, we will further investigate a similar problem in the scenario with a long-term objective over multiple time frames. Obviously, in this case, the interdependence between any two consecutive time frames has to be carefully studied (e.g., a sensor's remaining energy of the current time frame becomes this sensor's initial energy at the next time frame). This motivates us to derive a new objective function for characterizing the long-term energy consumption minimization and develop an online algorithm for addressing the system dynamics caused by the state variations among different time frames.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) under grants No. 62002164, No. 62176122, and No. 62171218, by the Industry-University-Research Cooperation Fund project of ZTE under grant

No. 2022ZTE02-07, and by the Postgraduate Research & Practice Innovation Program of NUAA under grant No. xcxjh20221614.

References

[1] J. Chen, C. Yi, R. Wang, K. Zhu, and J. Cai, "A joint optimization of sensor activation and mobile charging scheduling in industrial wireless rechargeable sensor networks," in *Proc. IEEE ICC*, 2022, pp. 3568–3573.

[2] L. D. Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Trans. Ind. Inform.*, vol. 10, no. 4, pp. 2233–2243, 2014.

[3] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, 2022.

[4] Y. Shi, C. Yi, B. Chen, C. Yang, K. Zhu, and J. Cai, "Joint online optimization of data sampling rate and preprocessing mode for edge–cloud collaboration-enabled industrial IoT," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16 402–16 417, 2022.

[5] Y. Liu, K.-W. Chin, C. Yang, and T. He, "Nodes deployment for coverage in rechargeable wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6064–6073, 2019.

[6] Q. Deng, Y. Ouyang, S. Tian, R. Ran, J. Gui, and H. Sekiya, "Early wake-up ahead node for fast code dissemination in wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3877–3890, 2021.

[7] Y. Zhao, Z. Li, B. Hao, and J. Shi, "Sensor selection for tdoa-based localization in wireless sensor networks with non-line-of-sight condition," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9935–9950, 2019.

[8] C.-T. Chang, C.-Y. Chang, S. Zhao, J.-C. Chen, and T.-L. Wang, "SRA: A sensing radius adaptation mechanism for maximizing network lifetime in WSNs," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9817–9833, 2016.

[9] Y. Shi, C. Yi, B. Chen, C. Yang, X. Zhai, and J. Cai, "Closed-loop control of edge-cloud collaboration enabled IIoT: An online optimization approach," in *Proc. IEEE ICC*, 2022, pp. 5682–5687.

[10] Z. Gao, D. Chen, and H.-C. Wu, "Energy loss minimization for wireless power transfer based energy redistribution in WSNs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12 271–12 285, 2019.

[11] H.-H. Choi and K. Lee, "Cooperative wireless power transfer for lifetime maximization in wireless multihop networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3984–3989, 2021.

[12] T. V. Nguyen, T.-N. Do, V. N. Q. Bao, D. B. d. Costa, and B. An, "On the performance of multihop cognitive wireless powered D2D communications in WSNs," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2684–2699, 2020.

[13] S. He, J. Chen, F. Jiang, D. K. Yau, G. Xing, and Y. Sun, "Energy provisioning in wireless rechargeable sensor networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 10, pp. 1931–1942, 2013.

[14] L. Mo, A. Kritikakou, and S. He, "Energy-aware multiple mobile chargers coordination for wireless rechargeable sensor networks," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8202–8214, 2019.

[15] J. Baek, S. I. Han, and Y. Han, "Energy-efficient UAV routing for wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1741–1750, 2020.

[16] M. Yang, N. Liu, L. Zuo, Y. Feng, M. Liu, H. Gong, and M. Liu, "Dynamic charging scheme problem with actor–critic reinforcement learning," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 370–380, 2021.

[17] J. Ji, K. Zhu, C. Yi, and D. Niyato, "Energy consumption minimization in UAV-assisted mobile-edge computing systems: Joint resource allocation and trajectory design," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 8570–8584, 2021.

[18] Q. Wu, P. Sun, and A. Boukerche, "A novel joint data gathering and wireless charging scheme for sustainable wireless sensor networks," in *Proc. IEEE ICC*, 2020, pp. 1–6.

[19] G. Han, X. Yang, L. Liu, and W. Zhang, "A joint energy replenishment and data collection algorithm in wireless rechargeable sensor networks," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2596–2604, 2018.

[20] G. Han, A. Qian, J. Jiang, N. Sun, and L. Liu, "A grid-based joint routing and charging algorithm for industrial wireless rechargeable sensor networks," *Comput. Netw.*, vol. 101, pp. 19–28, 2016.

[21] B. Li, X. Xiao, S. Tang, and W. Ning, "An energy efficiency-oriented routing protocol for wireless rechargeable sensor networks," in *Proc. IEEE IPEC*, 2021, pp. 1–5.

[22] S. R. Pokhrel, S. Verma, S. Garg, A. K. Sharma, and J. Choi, "An efficient clustering framework for massive sensor networking in industrial internet of things," *IEEE Trans. Ind. Inform.*, vol. 17, no. 7, pp. 4917–4924, 2021.

[23] M. Mukherjee, L. Shu, R. V. Prasad, D. Wang, and G. P. Hancke, "Sleep scheduling for unbalanced energy harvesting in industrial wireless sensor networks," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 108–115, 2019.

[24] W. Mao, Z. Zhao, Z. Chang, G. Min, and W. Gao, "Energy-efficient industrial internet of things: Overview and open issues," *IEEE Trans. Ind. Inform.*, vol. 17, no. 11, pp. 7225–7237, 2021.

[25] C. Yi, J. Cai, and Z. Su, "A multi-user mobile computation offloading and transmission scheduling mechanism for delay-sensitive applications," *IEEE Trans. Mobile Comput.*, vol. 19, no. 1, pp. 29–43, 2020.

[26] C. Zhang, G. Zhou, H. Li, and Y. Cao, "Manufacturing blockchain of things for the configuration of a data- and knowledge-driven digital twin manufacturing cell," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11 884–11 894, 2020.

[27] S. Martello and P. Toth, *Knapsack problems: algorithms and computer implementations*. John Wiley & Sons, Inc., 1990.

[28] S. Goyal, "A survey on travelling salesman problem," in *Midwest instruction and computing symposium*, 2010, pp. 1–9.

[29] K. Wang, Y. Wang, Y. Sun, S. Guo, and J. Wu, "Green industrial internet of things architecture: An energy-efficient perspective," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 48–54, 2016.

[30] T. Wu, P. Yang, H. Dai, C. Xiang, X. Rao, J. Huang, and T. Ma, "Joint sensor selection and energy allocation for tasks-driven mobile charging in wireless rechargeable sensor networks," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11 505–11 523, 2020.

[31] H. Dai, Q. Ma, X. Wu, G. Chen, D. K. Y. Yau, S. Tang, X.-Y. Li, and C. Tian, "CHASE: Charging and scheduling scheme for stochastic event capture in wireless rechargeable sensor networks," *IEEE. Trans. Mob. Comput.*, vol. 19, no. 1, pp. 44–59, 2020.

[32] Y. Shu, K. G. Shin, J. Chen, and Y. Sun, "Joint energy replenishment and operation scheduling in wireless rechargeable sensor networks," *IEEE Trans. Ind. Inform.*, vol. 13, no. 1, pp. 125–134, 2017.

[33] E. Regolin, A. Alatorre, M. Zambelli, A. Victorino, A. Charara, and A. Ferrara, "A sliding-mode virtual sensor for wheel forces estimation with accuracy enhancement via EKF," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3457–3471, 2019.

[34] M. Healy, T. Newe, and E. Lewis, "Wireless sensor node hardware: A review," in *IEEE Sensors 2008 Conf.*, pp. 621–624.

[35] H. P. Gupta, T. Venkatesh, S. V. Rao, T. Dutta, and R. R. Iyer, "Analysis of coverage under border effects in three-dimensional mobile sensor networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 9, pp. 2436–2449, 2017.

[36] S. Mao, M. H. Cheung, and V. W. S. Wong, "Joint energy allocation for sensing and transmission in rechargeable wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2862–2875, 2014.

[37] B. Zhu, E. Bedeer, H. H. Nguyen, R. Barton, and J. Henry, "UAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9540–9554, 2021.

[38] H. Wang, R. Wang, H. Xu, Z. Kun, C. Yi, and D. Niyato, "Multi-objective mobile charging scheduling on the internet of electric vehicles: a DRL approach," in *Proc. IEEE GLOBECOM*, 2021, pp. 01–06.

[39] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Proc. NeurIPS*, 2000, pp. 1008–1014.

[40] M. A. Wiering and M. Van Otterlo, *Reinforcement Learning: State-of-the art*. Springer, 2012, vol. 12, no. 3.

[41] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. NeurIPS*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014.

[42] M. Nazari, A. Oroojlooy, L. V. Snyder, and M. Takáč, "Reinforcement learning for solving the vehicle routing problem," in *Proc. NeurIPS*, 2018, pp. 9839–9849.

[43] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. NeurIPS*, 2014, pp. 3104–3112.

[44] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[45] A. Brewer, *The making of the classical theory of economic growth*. Routledge, 2010.

[46] A. Krause and D. Golovin, "Submodular function maximization," *Tractability*, vol. 3, pp. 71–104, 2014.

[47] T. Liu, B. Wu, S. Zhang, J. Peng, and W. Xu, “An effective multi-node charging scheme for wireless rechargeable sensor networks,” in *Proc. IEEE ICCC*, 2020.

[48] L. Xie, Y. Shi, Y. T. Hou, W. Lou, H. D. Sherali, and S. F. Midkiff, “Bundling mobile base station and wireless energy transfer: Modeling and optimization,” in *Proc. IEEE ICCC*, 2013, pp. 1636–1644.

[49] W. Ouyang, X. Liu, M. Obaidat, C. Lin, H. Zhou, T. Liu, and K.-F. Hsiao, “Utility-aware charging scheduling for multiple mobile chargers in large-scale wireless rechargeable sensor networks,” *IEEE Trans. Sustain. Comput.*, pp. 1–1, 2020.

[50] J. A. Stankovic, M. Spuri, K. Ramamritham, and G. C. Buttazzo, *Deadline scheduling for real-time systems: EDF and related algorithms*. Springer Science & Business Media, 2012, vol. 460.

Jiayuan Chen is currently pursuing the M.S. degree with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His research interests include machine learning (e.g., reinforcement learning) and mechanism design with applications in resource management and decision making for various wireless networks and mobile services including edge/fog computing, industrial IoT, vehicular/UAV systems and digital twin.

Changyan Yi (S’16-M’18) is currently a Professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), and is also affiliated with the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China. He received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Manitoba, MB, Canada, in 2018. From September 2018 to August 2019, he worked as a research associate in University of Manitoba, MB, Canada. He was awarded Changkong Scholor of NUAA in 2018, and Chinese Government Award for Outstanding Students Abroad in 2017. His research interests include game theory, queueing theory, machine learning and their applications in various wireless networks including edge/fog computing, IoT, 5G and beyond.

Ran Wang (M’18) is currently an Associate professor at College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), and Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, P.R. China. He received his B.E. in Electronic and Information Engineering from Honors School, Harbin Institute of Technology (HIT), P.R. China in July 2011 and Ph.D. in Computer Science and Engineering from Nanyang Technological University (NTU), Singapore in April 2016. He was a research fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore from October 2015 to August 2016. He has authored or coauthored over 50 papers in top-tier journals and conferences. He received the Nanyang Engineering Doctoral Scholarship (NEDS) Award, Singapore and innovative and entrepreneurial Ph.D. Award of Jiangsu Province, China in 2011 and 2017, respectively. His current research interests include intelligent management and control in smart grid, network performance analysis and internet of electric vehicles, etc.

Kun Zhu (M’16) received the Ph.D. degree from School of Computer Engineering, Nanyang Technological University, Singapore, in 2012. He was a research fellow with the Wireless Communications Networks and Services Research Group in University of Manitoba, Canada, from 2012 to 2015. He is currently a Professor in the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), and Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China. He is also a Jiangsu specially appointed professor. His research interests include resource allocation in 5G, wireless virtualization, and self-organizing networks. He has published more than fifty technical papers and has served as TPC for several conferences. He won several research awards including IEEE WCNC 2019 Best paper awards, ACM China rising star chapter award.

Jun Cai (M’04-SM’14) received the Ph.D. degree from the University of Waterloo, ON, Canada, in 2004. From June 2004 to April 2006, he was with McMaster University, Canada, as a Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellow. From July 2006 to December 2018, he has been with the Department of Electrical and Computer Engineering, University of Manitoba, Canada, where he was a full Professor and the NSERC Industrial Research Chair. Since January 2019, he has joined the Department of Electrical and Computer Engineering, Concordia University, Canada, as a full Professor and the PERFORM Centre Research Chair. His current research interests include edge/fog computing, ehealth, radio resource management in wireless communication networks, and performance analysis. Dr. Cai served as the Technical Program Committee (TPC) Co-Chair for IEEE GreenCom 2018; Track/Symposium TPC Co-Chair for the IEEE VTC-Fall 2019, IEEE CCECE 2017, IEEE VTC-Fall 2012, IEEE Globecom 2010, and IWCMC 2008; the Publicity Co-Chair for IWCMC 2010, 2011, 2013, 2014, 2015, 2017, 2020; and the Registration Chair for QShine 2005. He also served on the editorial board of IEEE Internet of Things Journal, IET Communications, and Wireless Communications and Mobile Computing. He received the Best Paper Award from Chinacom in 2013, the Rh Award for outstanding contributions to research in applied sciences in 2012 from the University of Manitoba, and the Outstanding Service Award from IEEE Globecom 2010.