BESS Aided Renewable Energy Supply Using Deep Reinforcement Learning for 5G and Beyond

Hao Yuan, Guoming Tang, *Member, IEEE*, Deke Guo, *Senior Member, IEEE*, Kui Wu, *Senior Member, IEEE*, Xun Shao, *Member, IEEE*, Keping Yu, *Member, IEEE*, and Wei Wei

***Abstract*—The year of 2020 has witnessed the unprecedented development of 5G networks, along with the widespread deployment of 5G base stations (BSs). Nevertheless, the enormous energy consumption of BSs and the incurred huge energy cost have become significant concerns for the mobile operators. As the continuous decline of the renewable energy cost, equipping the power-hungry BSs with renewable energy generators could be a sustainable solution. In this work, we propose an energy storage aided renewable energy supply solution for the BS, which could supply clean energy to the BS and store surplus energy for backup usage. Specifically, to flexibly regulate the battery's discharging/charging, we propose a deep reinforcement learning based regulating policy, which can adapt to the dynamical renewable energy generations as well as the varying power demands. Our experiments using the real-world data on renewable energy generations and power demands demonstrate that, our power supply solution can achieve an cost saving ratio of 77.9%, compared to the case with traditional power grid supply.**

***Index Terms*—5G base stations, BESS, renewable energy supply, deep reinforcement learning.**

I. Introduction

THE 5G network is considered as a promising technology to significantly improve the way how we live [1]. Compared to the 4G/LTE, it can ensure users with higher bandwidth and lower latency and thus enable various cutting-edge mobile services, such as the Internet of Vehicles [2], Virtual Reality [3], and Smart Medical Home [4]. Nevertheless, due to the adoption of high frequency bands by 5G base station (BS), its signal coverage range is much shorter than that of the 4G/LTE. Consequently, the mobile operators need to deploy a large number of 5G BSs to tackle the problem of poor signal coverage. This would result in an ultra-dense BS deployment, especially in "hotspot" areas, as illustrated in Fig. 1.

Building and operating such large-scale BSs require enormous investments and resources. According to a field survey in the cities of Guangzhou and Shenzhen, China, the full-load power consumption of a typical 5G BS is about $2 \sim 3$ times of that of a 4G one [5]. Considering the ultra-dense deployment of 5G BSs, it could lead to a tenfold increase in energy consumption. In addition, with the increasing emphasis on environmental protection, many governments have shut down some coal-fired power plants, resulting in severe power shortages in some areas. In this regard, how to effectively reduce energy consumption and improve the energy efficiency are critical problems.

Renewable energies like the solar and wind energies are eco-friendly with zero carbon emissions and become popular in more scenarios in recent years [6]. Owing to the continuing price decline in photovoltaic (PV) module and wind turbine, the installation cost of renewable energy has dramatically decreased over the past decade, e.g., it reports a 61% reduction of the solar equipment from 2010 to 2017 [7]. Such cost reductions lead to a rapid payback period for the renewable energy investment, from a couple of years to several months [8]. The above observations indicate the great potential of renewable energy on the market of fossil fuel replacement and carbon emission reduction.

It thus has inspired the mobile operators to utilize renewable energy as the auxiliary power supply to tackle the huge power demand at 5G BSs. In some developing countries, solar power has already been applied to supply the BSs, some of which occupies over 8% of the total electricity usage [9]. By installing the PV and wind turbine near the BSs, it shows that the maximum power from the solar and wind generators can reach up to 8.5kW and 6.0kW, respectively, which could remarkably cut down the communication energy supply from the traditional power gird.

To maximize the utilization of renewable energy, *energy storage* can be strategically utilized such that the energy can be continuously provided, as the renewable (like solar or wind) energy is intermittent and unstable. Meanwhile, most

Manuscript received April 16, 2021; revised November 1, 2021; accepted December 6, 2021. Date of publication December 17, 2021; date of current version May 20, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61802421 and Grant U19B2024; in part by the National Natural Science Foundation of Hunan Province under Grant 2019JJ30029; in part by the Telecommunications Advancement Foundation (Japan) Research Grant; in part by the RIEC Nationwide Cooperative Research Projects; in part by the Research Institute of Electrical Communication, Tohoku University, Japan, H31/B18; and in part by ROIS NII Open Collaborative Research 2021 under Grant 21FA03. *(Corresponding authors: Guoming Tang; Deke Guo.)*

Hao Yuan and Deke Guo are with the Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 564211, Hunan, China.

Guoming Tang is with the Department of Broadband Communication, Peng Cheng Laboratory, Shenzhen 518055, Guangdong, China.

Kui Wu is with the Department of Computer Science, University of Victoria, Victoria, BC V8P 5C2, Canada.

Xun Shao is with the School of Regional Innovation and Social Design Engineering, Kitami Institute of Technology, Kitami 090-0015, Japan.

Keping Yu is with the Global Information and Telecommunication Institute, Waseda University, Shinjuku 169-8555, Japan.

Wei Wei is with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710049, China.

Digital Object Identifier 10.1109/TGCN.2021.3136363

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

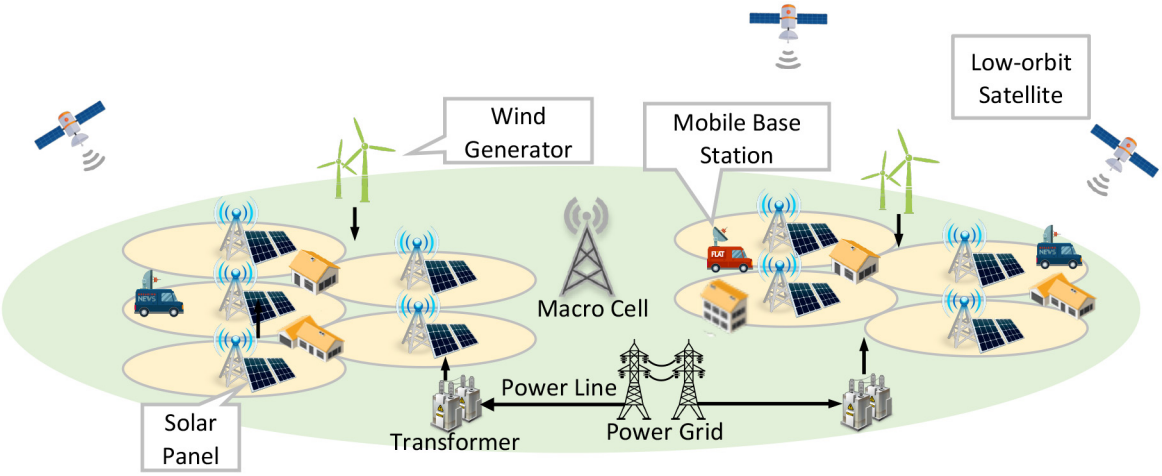


Fig. 1. A vision of the future radio access network (RAN) in 5G and beyond, which consists of macro and small cells, and also includes the mobile and space BSs. For the purpose of green communication, all the BSs could be supplied by both the renewable energy and power grid.

BSs are equipped with backup batteries to safeguard the BS's normal functioning against power outages, making it the natural energy storage. Besides, with the continuous price decline in battery storage these years [10], [11], combining the battery storage with renewable energy generators could offer even greater cost-reduction potential. Specifically, i) when the generated renewable power is less than the power demand (e.g., during the peak hours), the battery can be discharged to flatten the peak power demands, and ii) when the generated renewable power is more than the power demand (e.g., during the off-peak hours), the battery can be charged to store the surplus renewable energy.

In this paper, we propose a battery energy storage system (BESS) aided renewable energy supply solution for the 5G network and beyond. Aiming at energy cost reduction for mobile operators, our solution is to maximize the utilization of the renewable energy and thus minimize the utilization of power grid (i.e., fossil energy). Specifically, the energy charge can be continuously reduced by the generated renewable power, and the demand charge can be reshaped and flatten through strategic battery discharging/charging operations.

When designing the optimal control strategy in battery discharging/charging operations, we are faced with several challenges. Firstly, the renewable energy generation and power demand are highly varying in both spatial and temporal dimensions and thus hard to predict. Secondly, owing to the physical constraints of the battery discharging/charging operations (e.g., discharge/charge efficiency), it is complicated to design the optimal battery controlling policy. Thirdly, as the battery's capacity and lifetime are limited and shortened along with the discharge/charge cycles, it is necessary while non-trivial to trade-off between the cost of battery's degradation/replacement and the gain of renewable energy storage.

By tackling the above challenges, we make the following contributions in this work:

- We present the BESS aided renewable energy supply paradigm for 5G BS operations, in which the battery discharging/charging operation is modelled as an optimization problem. The model is comprehensive by taking into account the practical considerations of dynamic power demand and renewable energy generation, as well as battery specifications and physical constraints.
- To cope with the intermittent renewable energy generation and dynamic BS power demand, we propose a deep reinforcement learning (DRL) based battery discharging/charging operating policy. It can update the network parameters by interacting with the environment in real-time, so as to improve its decision-making efficiency.
- We conduct extensive evaluations using real-world BS deployment scenario and BS traffic load traces. The results show that the proposed DRL-based battery discharging/charging policy can effectively utilize the renewable energy and cut down the energy cost, with the cost saving ratio up to 77.9%.

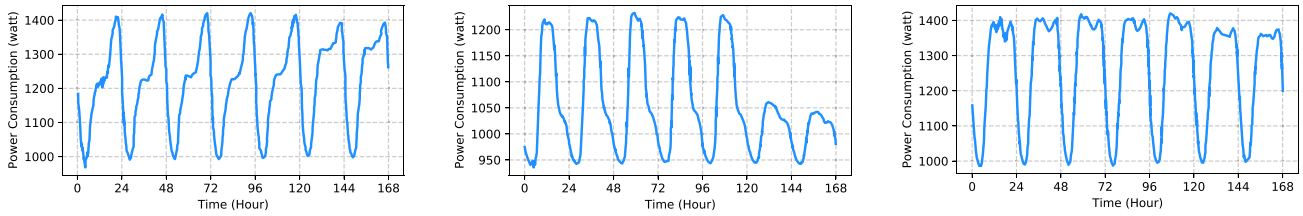
The rest of the paper is organized as follows. In Section II, we introduce the background of this paper. In Section III, we give the system models and formulations of the problem, and then propose the BESS aided renewable energy supply solution in Section IV. We develop a DRL-based battery discharging/charging controlling policy in Section V. We evaluate the proposed method by experiments with a real-world dataset in Section VI. We present the related work in Section VII and conclude the paper in Section VIII.

II. Background

A. Base Station Power Demand

The power demand pattern of a BS is mainly determined by its location and associated with the behavior of users there. Usually the demand also shows a periodic pattern (e.g., with a one-day or one-week period). As shown in Fig. 2, in this paper, we mainly consider three types of BSs at the areas of *resident*, *office*, and *comprehensive*, which account for nearly ninety percentage of the total demands [12]. To be detailed, the characteristics of these power demand patterns are as follows.

- *Power Demand of BSs at Resident Area:* The power demands of this type of BSs increase rapidly in the



(a) Power demand pattern of BSs at resident area. (b) Power demand pattern of BSs at office area. (c) Power demand pattern of BSs at comprehensive area.

Fig. 2. Power demand patterns of BSs at different area in one week period [12].

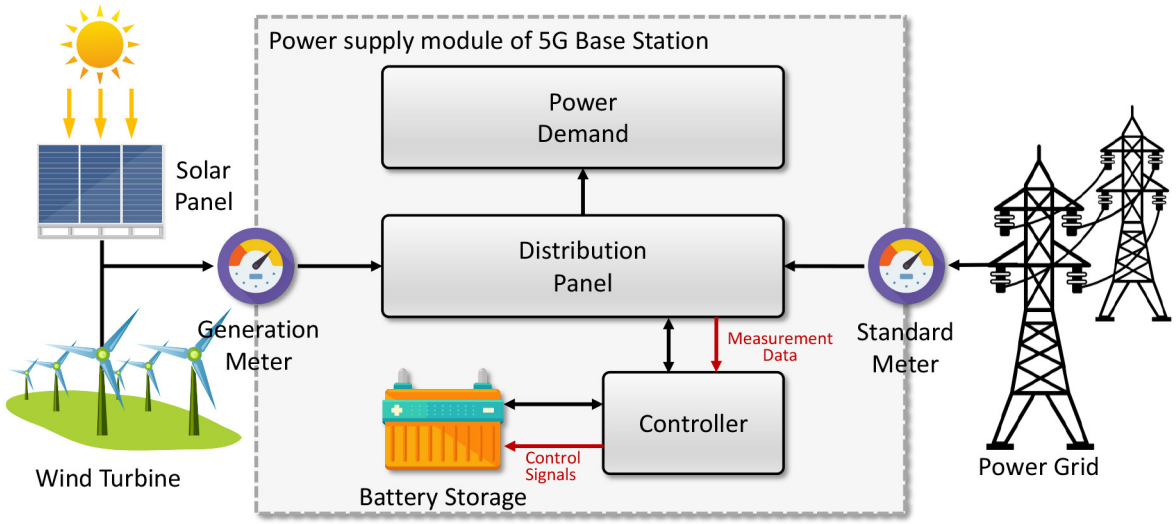


Fig. 3. An exemplified implementation of the BESS aided renewable energy supply solution for the 5G BS.

evening, as most people stay at home after work. Compared with those in weekdays, the power demands keep at high-levels in weekends.

- *Power Demand of BSs at Office Area:* The power demands of this type of BSs keep at the high-level in the day time, when most people work during the time. Besides, due to the fewer people work on the weekends, the weekend power demands are much lower than those on the weekdays.
- *Power Demand of BSs at Comprehensive Area:* Due to the diversity of the requests, compared to the above two BSs, the power demand patterns of this type of BSs are more stable: constantly keep at a high-level in the day time and evening and drop down to the valley in late night and early morning.

The first two types of power demand patterns change relatively dramatically, leading to a huge energy-saving potential, especially for the *demand charge*, which will be discussed in the next section.

B. Energy Cost of 5G BS

The energy cost of the mobile operator typically makes up of two components: i) *energy charge*, i.e., the total consumed electricity amount (in kWh) throughout the entire billing cycle, which is the interval of time from the end of one billing statement date to the next billing statement date for electricity (e.g., one month), and ii) *demand charge*, i.e., the peak power demand (in kW) during the billing cycle period. Specifically, the demand charge is regarded as a penalty due to the caused extra load burden to the power grid.

For example, for a commercial data center consuming 10 MW on peak and 6 MW on average, the monthly energy charge and demand charge amount to around $24,000 and $165,500, respectively [13]. The demand charge could be up to 8x the energy charge, therefore, effectively cutting down the demand charge could remarkably reduce the energy cost. However, there seems no practical way to flatten the peak power demands of 5G BSs, e.g., shifting the real-time demands from mobile users to the off-peak hours could lead to the long delay for some of the classes of jobs [14].

III. System Model

In this section, we present the system models and basic assumptions and problem formulation. For clarity, the major notations used in this paper are explained in Table I.

A. Scenario Overview

As illustrated in Fig. 3, the proposed BESS aided renewable energy supply solution deployed at each 5G BS mainly includes: i) a renewable energy generator, e.g., the PV panel and wind turbine, which is deployed near the 5G BS system and generates renewable energy for the system, ii) a battery

TABLE I
SUMMARY OF NOTATIONS

Notation	*Description*
$d(t)$	power demand of 5G BS in time slot t
$g(t)$	renewable energy generation in time slot t
$b(t)$	discharging/charging operations in time slot t
$\chi(t)$	battery state in time slot t
$p(t)$	power supplied by the power gird in time slot t
p_{max}	peak power consumption supplied by power gird
π	initial capacity of the battery
$\mathcal{C}^e(t)$	energy charge of 5G BS in time slot t
$\mathcal{C}^d(t)$	demand charge of 5G BS in time slot t
$\mathcal{C}^u(t)$	investment cost in time slot t
λ_e	prices of energy charge
λ_d	prices of demand charge
λ_u	prices of investment cost
α, β	discharging and charging efficiencies
$R+, R-$	max charge and discharge rates of battery
$s(t)$	environment state in time slot t
$a(t)$	action taken by the agent in time slot t
$r(t)$	reward of the action in time slot t
ψ	mapping policy from environment states to actions
$R(a(t), s(t))$	reward function of the DQN
$Q, \tilde{Q}$	Q-values of the main net and target net
$\theta, \tilde{\theta}$	parameters of the main net and target net

storage, which stores the surplus renewable energy and acts as the power source for the BS as needed, and iii) a controller, which can obtain the environment state (i.e., the measurement data) so as to control the battery discharging/charging operations through the control signals. In addition to the standard meter, as shown in Fig. 3, an additional generation meter is installed for the BS power supply system to measure the renewable energy generation. Furthermore, with commands from the controller, the distribution panel takes responsibility of power switch between the renewable energy and grid energy and ensures continuous and stable electricity supply for the BS.

As the essential component of the BESS aided renewable energy supply solution, the controller determines how efficient this paradigm is. Specifically, at each scheduling point, the controller needs to decide the amount of power supply from either the battery or the power grid. The scheduling operations should be made upon the power demands and battery states in real-time, so that the utilization of renewable energy can be enhanced and the total energy cost can be minimized.

Note that the feasibility of such an implementation as illustrated by Fig. 3 has been preliminarily verified in practice. According to [15], small integrated renewable energy generators have been provided by some commercial companies for the BS system, which are easily deployed in both open rural and crowded urban environments.

B. BS Power Supply and Demand

The power of each 5G BS is mainly supplied by three parts: power grid, generated renewable energy, and storage energy. In particular, i) when the generated renewable energy is more than the power demand (e.g., during the off-peak hours), each 5G BS is only supplied by the renewable energy (i.e., off-grid) and the surplus renewable energy is stored in the battery storage, ii) when the generated renewable energy is less than the power demand (e.g., during the peak hours), each 5G BS is supplied by all three parts in a cooperative way.

In this paper, we consider a discrete time model, where the entire billing cycle (e.g., one month) is equally spilt into T consecutive slots with length of Δt and denoted by $\mathcal{T} = \{1, 2, \ldots, T\}$. For an arbitrary 5G BS, the power demand during the entire billing cycle can be represented by a power demand vector:

$$d \coloneqq [d(1), d(2), \ldots, d(T)] \tag{1}$$

where $d(t)$ is the power demand in time slot t, which can be obtained by power meter readings at each BS.

C. Renewable Energy Generation

By harvesting energy from renewable energy resources, the BSs could be powered in an environmentally-friendly and cost-efficient way. In this paper, in order to make the model extensible, we denote the renewable energy generation vector as:

$$g \coloneqq [g(1), g(2), \ldots, g(T)] \tag{2}$$

In this work, we choose two typical renewable energy as the auxiliary way of power supply, i.e., solar energy (i.e., $g^s(t)$) and wind energy (i.e., $g^w(t)$). Accordingly, for an arbitrary time slot t, the renewable energy generation vector can be represented by:

$$g(t) = g^s(t) + g^w(t) \tag{3}$$

We assume that if the total generated renewable energy is beyond the power demand (i.e., $g(t) > d(t)$), the power is supplied in proportion to the renewable energy generated. The generation of both varies during a certain period (e.g., one day) and is affected by a some similar factors such as weather, temperature, wind speed, and so on.

1) Solar Energy Generation: Power generated by the solar PV system mainly depends on three factors: global horizontal irradiance (*GHI*(*t*)), outdoor temperature (*Temp*(*t*)), and time of day (*ToD*(*t*)). By arranging solar PV cells in series/parallel, solar PV could harvest energy and convert it into DC to charge the battery storage and supply the power demand. The generated power by the solar PV at time slot t can be measured by the following function:

$$g^s(t) = \mathbb{F}^S(GHI(t), Temp(t), ToD(t)) \tag{4}$$

where $\mathbb{F}^S(\cdot)$ is a known, non-linear function defined in PVLIB [16]. Accordingly, the solar energy generation during the entire billing cycle can be represented by a vector:

$$g^s \coloneqq [g^s(1), g^s(2), \ldots, g^s(T)]. \tag{5}$$

2) Wind Energy Generation: Power generated by the wind turbine generator fluctuates randomly with time and mainly depends on the wind velocity (*WV*(*t*)), weather system (*WS*(*t*)), and hub height (*HH*(*t*)). The wind turbine generate energy typically into two stages: first, it converts the wind power into

mechanical energy and then transforms into electricity. The amount of the power generated by the wind turbine at time slot t can be calculated by the following function:

$$g^w(t) = \mathbb{F}^W(WV(t), WS(t), HH(t)) \tag{6}$$

where $\mathbb{F}^W(\cdot)$ is a known, non-linear function defined in [17]. Accordingly, the wind energy generation during the entire billing cycle can be represented by a vector:

$$g^w \coloneqq [g^w(1), g^w(2), \ldots, g^w(T)]. \tag{7}$$

D. Battery Specification

At an arbitrary time slot *t*, the state of the battery is modeled as follows:

$$\chi(t) \coloneqq \langle SoE(t), SoC(t), DoD(t) \rangle \tag{8}$$

where the notations of SoE, SoC, and DoD represent the *state of effective capacity*, *state of charge*, and *depth of discharge* of the battery, respectively. Specifically, i) **SoE** indicates the current effective capacity of the battery, as a percentage of its initial capacity (denoted as π), ii) **SoC** indicates the current energy stored in the battery, as a percentage of the current effective capacity, and iii) **DoD** indicates how much energy the battery has released, as a percentage of the current effective capacity.

For simplicity to tackling the optimization problem, we discretize the SoC of a battery into *M* equal-spaced states (e.g., $M = 10$, i.e., $\{10\%, 20\%, \ldots, 100\%\}$). Accordingly, the DoD are also discretized (e.g., release 10% from 90%, i.e., 90% $\rightarrow$ 80%). Besides, for an arbitrary time slot *t*, in order to prevent the battery from over-discharging/charging, we use SoC_{max} and SoC_{min} to indicate the upper and lower bounds of SoCs, respectively, which is shown as follows.

$$SoC_{min} \leq SoC(t) \leq SoC_{max}. \tag{9}$$

IV. BESS Aided Renewable Energy Supply

The battery storage is deployed at 5G BS, and can charge by the surplus renewable energy (generated by solar PV and wind turbine system) and discharge to reshape the power demand, so as to maximize the utilization of renewable energy (or minimize the utilization of fossil fuel) and reduce the electricity bill.

We define the battery discharging/charging operations by a *battery operation vector*:

$$b \coloneqq [b(1), b(2), \ldots, b(T)] \tag{10}$$

where $b(t)$ is a real number variable and indicates the amount of discharging/charging operations. To be detail, i) **positive value** indicates discharging the power from the battery storage to the 5G BS during time slot *t*, ii) **negative value** indicates charging from the renewable energy to the battery storage, and iii) **zero value** indicates no discharging/charging operation performs.

Meanwhile, the discharging/charging operations is constrained by the maximum charging rate and maximum discharging rate, denoted as R^+ and R^-, respectively. It means the largest power that the battery can be recharged and supply with in a time slot, which is shown as follows.

$$-R^+ \leq b_n(t) \leq R^- \tag{11}$$

Besides, the battery storage need to meet the following conditions in discharging/charging operations:

$$b(t) \leq 0, \quad \text{if} \quad g(t) - d(t) \geq 0 \tag{12a}$$

$$b(t) > 0, \quad \text{if} \quad g(t) - d(t) < 0 \tag{12b}$$

which represents that the battery storage can only be charged when there exists surplus renewable energy after supplying to the 5G BS, and means that the battery storage cannot be simultaneously charged and discharged at any time slot.

Due to the power loss (e.g., AC-DC conversion and battery leakage [18]) occurred during discharging from battery storage to the power grid (or charging from renewable energy to the battery storage), we denote the actual discharging/charging operations from/to the battery by:

$$\tilde{b}(t) = \begin{cases} b(t)/\alpha, & \text{if} \quad b(t) \leq 0 \\ \beta \cdot b(t), & \text{if} \quad b(t) > 0 \end{cases} \tag{13}$$

where $\alpha \in (0, 1)$ and $\beta \in (0, 1)$ represent the charging and discharging efficiencies, respectively.

Given the power demand of the 5G BS (i.e., $d(t)$), the renewable energy generation (i.e., $g(t)$), and the battery discharging/charging operations (i.e., $b(t)$), we can derive the *power consumption vector* supplied by the power grid for an arbitrary time slot *t* by:

$$p \coloneqq [p(1), p(2), \ldots, p(T)] \tag{14}$$

where $p(t)$ is denoted as:

$$p(t) = max\Big\{0, d(t) - g(t) - \tilde{b}(t)\Big\}. \tag{15}$$

A. Energy Cost

The billing policy of the energy cost for the mobile operators throughout the entire billing cycle typically make up of two components, energy charge and demand charge, which is widely applied in previous [13], [14], [19]. And we will introduce them in detail as follows.

- Energy Charge: The total consumed electricity amount (in kWh) throughout the entire billing cycle (in the unit \$kWh and denoted by λ_e).
- Demand Charge: The peak power consumption supplied by power gird (in kW) during the entire billing cycle (in the unit \$kW and denoted by λ_d).

Therefore, the incurred cost of **energy charge** of the whole system in each time slot *t* can be represented by:

$$\mathcal{C}^e(t) = \lambda_e \cdot p(t) \cdot \Delta t \tag{16}$$

Accordingly, the incurred cost of **demand charge** of the whole system in each time slot *t* can be represented by:

$$\mathcal{C}^d(t) = \lambda_d \cdot max\{0, p(t) - p_{max}\} \tag{17}$$

where p_{max} records the peak power consumption during the past $t-1$ time slots. For any arbitrary time slot *t*, if $p(t) - p_{max} > 0$, p_{max} will be updated to $p(t)$ accordingly.

B. Investment Cost

Every usage of this equipment (solar PV, wind turbine, and battery storage) incurs a certain reduction of its lifetime, which is essential for the investor. Therefore, it is significant to understand, detail and quantify the various factors influencing the performance loss curves. For the accuracy of our model, we quantify the investment cost in every time slot as follows.

1) Renewable Energy Generator Cost: As modules of a renewable energy generated system age, they gradually lose some performance. In this paper, we assume the decline of the system is linear and positively related to its using time. We denote the lifetime of the renewable energy generator as *L*, which indicates the total time it can be used. For an arbitrary time slot *t*, the remaining lifetime of the renewable energy generator is denoted as $l(t)$, which is constrained by $0 \leq l(t) \leq L$. The renewable energy generator has to be discarded and replaced by a new one if $l(t) \leq 0$. Given the remaining lifetime of the renewable energy generator at time $t-1$, the remaining lifetime at time *t* is updated by:

$$l(t) = l(t-1) - \Delta t \cdot u(t) \tag{18}$$

where $u(t)$ is defined by:

$$u(t) = \begin{cases} 1, & \text{if using} \\ 0, & \text{if not using} \end{cases} \tag{19}$$

We formulate the using cost of the renewable energy generator in each time slot *t* as:

$$\mathcal{C}^u(t) = \lambda \cdot \frac{\Delta t \cdot u(t)}{L} \tag{20}$$

where λ is the investment cost of a new renewable energy generator.

We extend the model of renewable energy generator to specific system, i.e., the solar PV system and wind turbine system. To be detail, i) for the solar PV system, we denote the lifetime, the investment cost, and investment as $l^s(t)$, $\mathcal{C}^{u_s}(t)$, and λ_s, respectively, ii) for the wind turbine system, we denote the lifetime, the using cost, and investment as $l^w(t)$, $\mathcal{C}^{u_w}(t)$, and λ_w. Accordingly, we can derive the using cost of the solar PV system and wind turbine system by replacing the symbol in the Eq. (20).

2) Battery Storage Degradation Cost: Every cycle of discharge/charge operation does some "harm" to the battery (typically lead-acid) and reduces its capacity and lifetime. Especially, a deep discharging severely affect its internal structure, even can permanently damage the battery (e.g., an overdischarging). The battery has to be discarded and replaced by a new one, when the effective capacity drops down to the "ineffective" level, denoted by SoE_{ine} in this paper.

As illustrated in Fig. 4, each level of DoD has a corresponding number of discharge/charge cycles, thus, we can formulate the battery storage degradation cost by the relationship between both. Given a state of battery at time slot *t*, i.e., $\langle SoE(t), SoC(t), DoD(t)\rangle$, the SoE decrease of the battery during this time slot can be measured by:

$$\Delta SoE(t) = \begin{cases} 0, & \text{if } b(t) \leq 0 \\ \frac{1-SoE_{ine}}{h(DoD(t-1)+\Delta DoD(t))}, & \text{if } b(t) > 0 \end{cases} \tag{21}$$

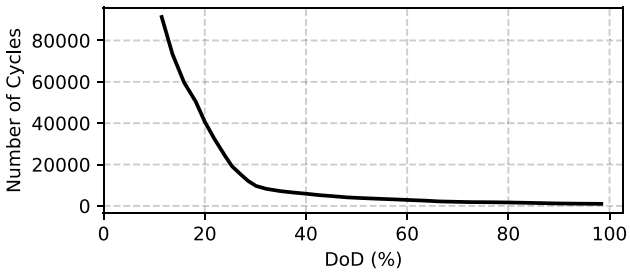


Fig. 4. Relationship between DoD levels and battery lifetime (in number of discharge/charge cycles) for LI battery, respectively [20].

where $h(\cdot)$ maps from an input DoD level to the total number of discharge/charge cycles (exemplified in Fig. 4), and $\Delta DoD(t)$ gives the increase of DoD and can be calculated by:

$$\Delta DoD(t) = \frac{b(t)\Delta t}{\pi} \tag{22}$$

With the above expression of SoE decrease in each time slot *t*, we can then formulate the **degradation cost** of the battery storage at each time slot *t* as:

$$\mathcal{C}^{u_b}(t) = \lambda_b \cdot \Delta SoE(t) \tag{23}$$

where λ_b is a coefficient converting the battery degradation to a monetary cost, with the unit of "$/SoE decrease."

To sum up, the total **investment cost** in each time slot *t* can be calculated as:

$$\mathcal{C}^u(t) = \mathcal{C}^{u_s}(t) + \mathcal{C}^{u_p}(t) + \mathcal{C}^{u_b}(t). \tag{24}$$

C. Optimization Formulation and Difficulty Analysis

The battery discharging/charging operations is controlled by the controller. Given the state (i.e., $\chi(t)$) of the battery storage in time slot $t-1$, the state in time slot *t* can be updated by:

$$\chi(t) \leftarrow \begin{cases} SoE(t) = SoE(t-1) - \Delta SoE(t) \\ SoC(t) = SoC(t-1) - b(t)\Delta t/\pi \\ DoD(t) = DoD(t-1) + \Delta DoD(t) \end{cases} \tag{25}$$

For the entire billing cycle $\mathcal{T}$, we need to find the optimal battery discharging/charging controlling policy to solve the optimization problem, so as to minimize the total electricity bill during the entire billing cycle, which is defined as follows.

$$\min_{b(t)} \sum_{t=1}^{T} \Big(\mathcal{C}^e(t) + \mathcal{C}^d(t) + \mathcal{C}^u(t) \Big) \tag{26a}$$

$$\text{s.t.} \quad (9),\ (11),\ (12),\ \text{and}\ (25),\ \forall t \in \mathcal{T}. \tag{26b}$$

When solving the above optimization problems, however, we are faced with the following three challenges.

1) Uncertainty of Renewable Energy: Renewable energy generation is affected by multiple factors such as outdoor temperature and wind velocity. It is hard to accurately forecast renewable energy generation (i.e., $g(t)$) and make the optimal discharging/charging operations (i.e., $b(t)$) of the battery storage without accurate information in advance, as the unpredictable and intermittent nature of these factors. Therefore, we need to propose a method to tackle the problem of the uncertainty of renewable energy generation.

2) Dynamic of Power Demand: In our modeled problem, we assume the power demand (i.e., $p(t)$) is known in advance and thus can essentially optimize in an offline way. However, such assumptions are unrealistic in practice. In fact, traditional offline optimization methods (e.g., dynamic programming [21], [22]) are hard to find the global optimal solution, as the power demand can be obtained only when the workload arrives at the 5G BS. Thus, an online method to deal with the dynamic power demands (i.e., $d(t)$), and make optimal discharging/charging operations (i.e., $b(t)$), is in great need.

3) High Computation Complexity: The optimization problem in Eq. (26) has embedded NP-hard subproblems. Firstly, in every time slot t, the controller needs to search the action space (mainly determined by M), so as to find the optimal discharging/charging operation (i.e., $b(t)$). For simplicity to solving the optimization problem, in this paper, we discretize the SoC of battery in to M equal-spaced states, however, in real scenario, the state of the battery is continuous, which leads to an enormous searching space. Secondly, during the entire billing cycle (i.e., $\mathcal{T}$), it is challenging for the controller to continuously make the optimal discharging/charging operation.

To tackle the above three challenges, we propose an online discharging/charging operation controlling method based on deep reinforcement learning (DRL) in the following section.

V. A DRL-Based Battery Operation Approach

Recent breakthrough of deep reinforcement learning (DRL) [23] provides a promising technique for enabling effective experience-driven control, which exploit the past experience (e.g., historical battery discharging/charging operations) for better decision-making by adapting to current state of environment. We consider DRL is particularly suitable for online discharging/charging operations because: i), it is capable of handling a high-dimensional state space (such as in AlphaGo [24]), which is more advantageous over traditional Reinforcement Learning (RL) [25], and ii) it is able to deal with highly dynamic time-variant environments such as time-varying power demand and renewable energy generation. Next, we will introduce the basic components and concepts of DRL and the proposed DRL-based battery discharging/charging controlling policy in detail.

A. Components & Concepts

A typical DRL framework consists of five key components: *agent*, *state*, *action*, *policy*, *and reward*. The concept and design of each component in our DRL-based battery discharging/charging controlling policy is explained as follows.

- *Agent:* The role of the agent is to make decisions in every episode by interacting with the environment. Specifically, at the beginning of each time slot, it determines the discharging/charging operations (i.e., $b(t)$) according the current state (e.g., $d(t)$, $g(t)$ and $\chi(t)$) of the environment. The objective is to find an optimal battery discharging/charging controlling policy to minimize the total electricity bill during the entire billing cycle.
- *State:* At each episode, the agent first observes the state of the current environment to take action. In order to take the optimal action at each episode, the current state should cover as much information as possible. In this paper, we define the *state vector* of the current environment as $s(t) = [d(t), g(t), \chi(t), p_{max}]$, which concludes the current information of the power demand, the renewable energy generation, the battery storage and the peak power consumption.
- *Action:* After observing the state of the environment, the agent will take an action accordingly. In our problem, the action is to control the battery discharging/charging operations in each time slot, i.e., $b(t)$, specifically, i) whether the battery should be discharged or charged, and ii) how much energy should be discharged or charged. We denote the action taken at time t by $a(t)$, which is equivalent to $b(t)$.
- *Policy:* The battery discharging/charging controlling policy $\psi(s(t)) : \mathcal{S} \rightarrow \mathcal{A}$ defines the mapping relationship from the state space to the action space, where $\mathcal{S}$ and $\mathcal{A}$ represent the state space and the action space, respectively. Specifically, the controlling policy can be represented by set of $a(t) = \psi(s(t))$, which maps the state of the environment to the action at time slot t.
- *Reward:* After interacting with the environment, the agent will receive a reward $r(t)$ (calculated by the reward function $R(s(t), a(t))$), which indicates the effect of the action in this episode, so as to update the controlling policy. The objective of the agent is to find a policy ψ to maximize the total reward through continuous interacting with the environment. The design of the reward function significantly affect the performance of the DRL-based algorithm, and we will introduce its detail in the next subsection.

To sum up, at each episode, the agent observes the state $s(t)$, takes an action $a(t)$ generated by the policy ψ, and receives a reward $r(t)$ calculated by the reward function $R(s(t), a(t))$. The objective of the proposed DRL-based battery discharging/charging controlling policy is to take the optimal action in every episode so as to maximize the total reward.

B. Reward Function Design

At the end of each time slot, the agent evaluates the performance of the action using a reward function, which transforms the performance statistics to a numerical utility value. For an arbitrary time t, the agent observes the state $s(t)$, takes the action $a(t)$ and adopts the following reward function to access the performance of the controlling action:

$$R(s(t), a(t)) = exp\Big(V^e(t) + V^d(t) + V^u(t) \Big) \quad (27)$$

in which:

- $V^e(t) = -\mathcal{C}^e(t)$, measures the reward of the incremental energy charge caused by the action in time slot t.
- $V^d(t) = -\mathcal{C}^d(t)$, measures the reward of the incremental demand charge caused by the action in time slot t.
- $V^u(t) = -\mathcal{C}^u(t)$, measures the reward of the investment cost caused by the action in time slot t.

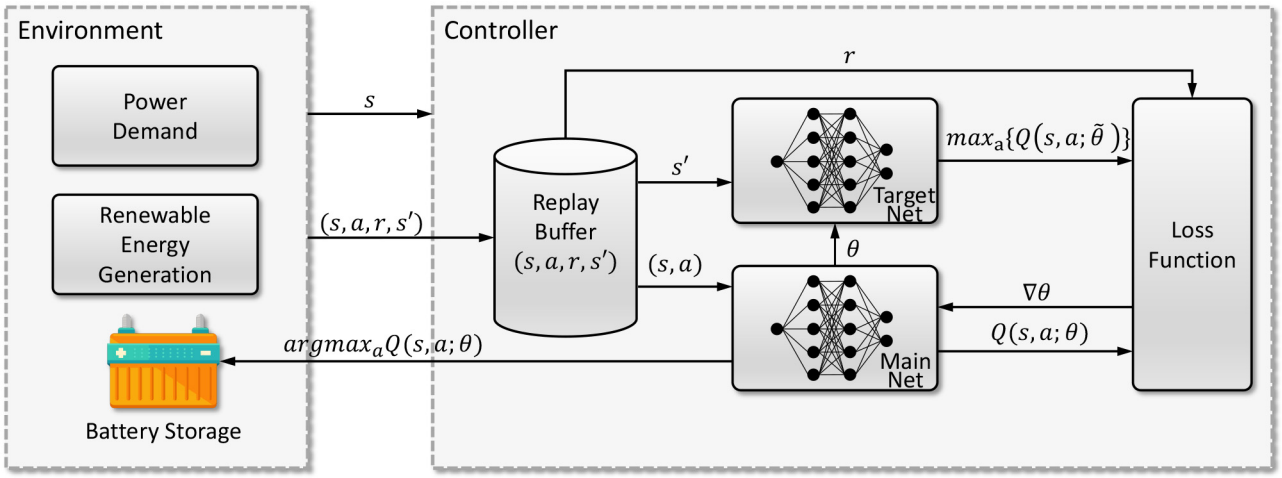


Fig. 5. The framework of the learning process in DQN. For simplicity, we denote $s(t+1)$ as s'. After interacting with the environment, the agent (i.e., controller) will determine the specific discharging/charging operation.

At the end of each time slot, the agent evaluates the performance of the action by the reward $r(t)$ calculated by the reward function $R(s(t), a(t))$. In the DRL-based framework, the objective is to maximize the expected cumulative discounted reward:

$$r(t) = \mathbb{E}\left[\sum_{k=t}^{\infty} \gamma^k R(s(t), a(t))\right] \tag{28}$$

where $\gamma \in (0, 1]$ is the discount accumulative factor indicating the degree of emphasis of future rewards, and the higher γ indicates a higher degree of emphasis on future rewards.

C. Learning Process Design

The learning process of the algorithm adopts a deep neural network (DNN) called Deep Q-Network (DQN) to derive the correlation between each state-action pair ($s(t)$, $a(t)$) and its value function $Q(s(t), a(t))$, which is the expected discounted cumulative reward. If the environment is in state $s(t)$ and follows action $a(t)$, the value function of the state-action ($s(t)$, $a(t)$) can be represented as:

$$Q(s(t), a(t)) = \mathbb{E}[r(t)|s(t), a(t)] \tag{29}$$

After obtaining the value of each state-action ($s(t)$, $a(t)$), the agent selects the action $a(t)$ with the ϵ-greedy policy ψ, that is, randomly selects the action with the probability of ϵ, and chooses the action with the maximum of $Q(s(t), a(t))$ with the probability of 1-ϵ, i.e., $\text{argmax}_{a(t)} Q(s(t), a(t))$.

As illustrated in Fig. 5, two effective techniques were introduced in [23] to improve stability: *replay buffer* and *target network*. Specifically,

- *Replay Buffer:* Unlike traditional reinforcement learning, DQN applies a replay buffer to store state transition samples in the form of $\langle s(t), a(t), r(t), s(t+1)\rangle$ collected during learning. Every κ time steps, the DRL-based agent updates the DNN with a mini-batch experiences from the replay buffer by means of stochastic gradient descent (SGD): $\theta_{i+1} = \theta_i + \sigma \bigtriangledown_\theta Loss(\theta)$, where σ is the learning rate. The higher learning rate will lead to the faster parameters updating speed. However, at the same time, the algorithm would be more affected by abnormal data, which is easy to diverge and difficult to converge. Compared with Q-learning (only using immediately collected samples), randomly sampling from the replay buffer allows the DRL-based agent to break the correlation between sequentially generated samples, and learn from a more independently and identically distributed past experiences. Thus, the replay buffer can smooth out learning and avoid oscillations or divergence.
- *Target Network:* There are two neural networks with the same structure but different parameters in DQN, the main net and the target net. $Q(s, a; \theta)$ and $Q(s, a; \tilde{\theta})$ represent the current Q-value and target Q-value generated by the main net and the target net, respectively. The DRL-based agent uses the target net to estimate the target Q-value $\tilde{Q}$ for training the DQN. Every τ time steps, the target net copies the parameters from the main net, whose parameters are updated in real-time. After introducing the target net, the target Q-value will remain unchanged for a period time, which reduces the correlation between the current Q-value and the target Q-value and improves the stability of the algorithm.

Accordingly, the DQN can be trained by the loss:

$$Loss(\theta) \leftarrow \mathbb{E}\left[\left(\tilde{Q} - Q(s(t), a(t); \theta)\right)^2\right] \tag{30}$$

where θ is the network parameters of the main net, and $\tilde{Q}$ is the target Q-value and calculated by:

$$\tilde{Q} \leftarrow r(t) + \gamma max_{a(t+1)} Q\left(s(t+1), a(t+1); \tilde{\theta}\right) \tag{31}$$

where $\tilde{\theta}$ is the network parameters of the target net and it updates every τ time slots by coping from the main net.

To sum up, the learning process is depicted by the pseudo-code in Algorithm 1. The controller first initializes the replay buffer and the parameters (i.e., θ and $\tilde{\theta}$) of the main net and target net, respectively (Line 1-3 in Algorithm 1). After obtaining the value of each state-action ($s(t)$, $a(t)$), the agent selects

Algorithm 1: Battery Controlling Algorithm With DRL

Input: Power demand of BS $d(t)$ and renewable energy generation $g(t)$, $1 \leq t \leq T$
Output: Discharging/charging actions $a(t)$, $1 \leq t \leq T$
1 Initialize replay buffer (RB) to capacity N;
2 Initialize main net Q with random weights θ;
3 Initialize target net $\tilde{Q}$ with weights $\tilde{\theta} = \theta$;
4 **for** $episode = 1 : MaxLoop$ **do**
5 **for** $t = 1 : T$ **do**
6 Get environment state $s(t)$;
7 $a(t) = \begin{cases} \text{argmax}_a Q(s(t), a(t); \theta), \text{ prob. } \epsilon \\ \text{random action, prob.} 1 - \epsilon \end{cases}$
8 Execute action $a(t)$ and receive $r(t)$ and $s(t+1)$;
9 Store $\langle (s(t), a(t), r(t), s(t+1) \rangle$ into RB;
10 Randomly sample a mini-batch of experience $\langle s(i), a(i), r(i), s(i+1) \rangle$ from RB by every κ steps;
11 $\tilde{Q} = \begin{cases} r(t), \text{terminates at step } t+1 \\ r(t) + \gamma \max_{a(t+1)} \{ Q(s(t+1), a(t+1); \tilde{\theta}) \}, \text{ else} \end{cases}$
12 Perform SGD on $(\tilde{Q} - Q(s, a; \theta))^2$ w.r.t. θ;
13 Set $\tilde{Q} = Q$ by every τ steps;
14 **end**
15 **end**

TABLE II
PARAMETER SETTINGS

	Parameter	*Setting*
Billing Policy	billing cycle window $\mathcal{W}$	one month (30 days)
	[1]energy charge price λ_e	US\$0.049/kWh
	[1]demand charge price λ_d	US\$16.08/kW
	[2]battery cost λ_b	US\$271/kWh
Battery Config.	discharge efficiency α	85%
	charge efficiency β	99.9%
	max charge rate $R+$	16 MW
	max discharge rate $R-$	8 MW
Solar PV	power rating g^s	4950 W
	price λ_s	US\$3950
	lifetime L^s	25 years
Wind Turbine	power rating g^w	6000 W
	price λ_w	US\$4500
	lifetime L^w	20 years

[1]Prices of energy/demand charges in 2018, referring to the contract in [26].
[2]Battery capacity costs in 2018, referring to the data in [27].

the action $a(t)$ with the ϵ-greedy policy ψ, and then performs the action $a(t)$ and interacts with the environment (Line 6-7 in Algorithm 1). Next, the agent will receive the reward $r(t)$ and observe the next state $s(t+1)$ of the environment, meanwhile store the state $\langle s(t), a(t), r(t), s(t+1) \rangle$ into the RB (Line 8-9 in Algorithm 1). Every κ time steps, the agent updates the DNN by Eq. (30) with a mini-batch experience from the replay buffer by means of stochastic gradient descent (SGD). The target net will copy the parameters of the main net by every τ time steps (Line 10-13 in Algorithm 1). During the learning process, we set the learning rate σ is 0.001, the ϵ in ϵ-greedy method is 0.9, the discount accumulative factor γ is 0.9, and the step parameters τ and κ are both 2000. For the whole battery discharging/charging scheduling process, the algorithm has an overall computational complexity of $O(C_{conv} \cdot T)$, where $C_{conv} = \sum_{i=1}^{n} C_{in}^i C_{out}^i$ represents the sum of the product of the input channel (neurons) and the output channel (neurons) of i-th linear layer, leading to the faster convergence speed compared to other DRL algorithm.

VI. PERFORMANCE EVALUATION

We evaluate the performance of the proposed DRL-based battery discharging/charging controlling policy through extensive numerical analysis.

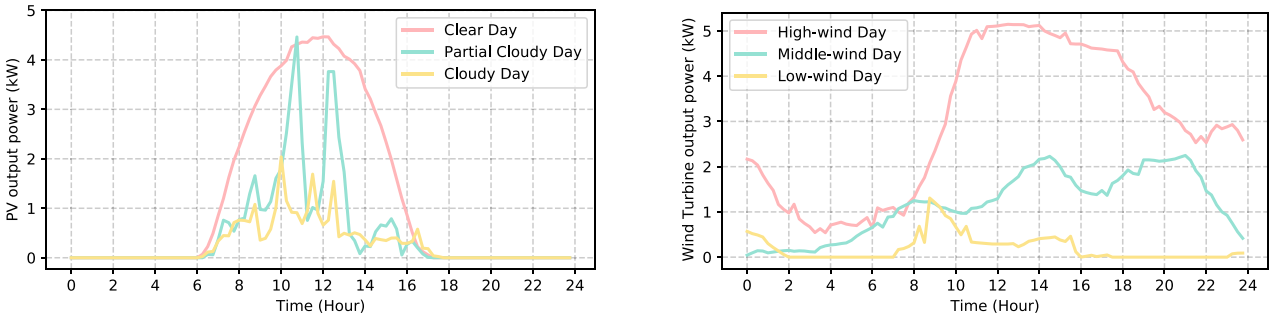
A. Experiment Setup

1) BS and Power Consumption Data: In order to show the performance of the proposed method, we mainly consider the 5G BS deployed at the three areas, i.e., *resident area*, *office area*, and *comprehensive area*, whose power consumption within one-week period are illustrated in Fig. 2, and we assume the power consumption of the same type BSs in different cities (e.g., Beijing, Shanghai and Guangzhou) is the same. For simplicity, we denote the BS deployed at the areas of resident, office, and comprehensive as type I, type II, and type III, respectively. We will apply the BESS aided renewable energy supply solution to different types of BSs in different cities under different weather conditions and evaluate its performance through massive simulation experiment.

2) Renewable Energy Generation Data: In Section III-C, we introduce the factors that impact the generation of renewable energy. For simplicity, we divide the weather conditions into three types. Accordingly, the output power pattern of the solar PV and wind turbine could be divided into three types. Specifically, for the solar PV, the weather conditions are divided into the *clear day*, *partial cloudy day*, and *cloudy day*; for the wind turbine, the weather conditions are divided into the *high wind velocity*, *middle wind velocity*, and *low wind velocity*. The output power patterns of the solar PV and wind turbine under different weather conditions are illustrated in Fig. 6 and the time slot Δt is 15 minutes in our experiment.

3) Equipment Parameter Settings: In this study, we use a quantity of 15 Panasonic Sc330 solar modules each with a power rating of 330W and JFNH-5kW wind turbine of Qingdao Jinfan Energy Science and Technology Co., Ltd. For the battery storage, we consider the mainstream lithium-ion (LI) battery on the current market. We then refer to [14], [26], [27] for parameter settings of electricity billing policy and battery configurations and the main parameter settings are summarized in Table II.

4) Scenario Settings: As the generation of the renewable energy is significantly affected by the weather conditions, we choose three representative cities in China for this paper, i.e., *Beijing*, *Shanghai*, and *Guangzhou*, which has different weather pattern during the billing cycle window (i.e., from 1st June 2020 to 30th June 2020). We compare and analyze the overall energy cost (including energy charge, demand charge and investment cost), detailed controlling results and return of investment (ROI) for three types of BSs (i.e., type I, type II, and type III BSs) in these cities, and the specific day of the



(a) The solar PV output power patterns under different weather conditions.

(b) The wind turbine output power patterns under different weather conditions.

Fig. 6. (a) The solar PV output power patterns under different weather conditions (i.e., $GHI(t)$, $Temp(t)$, and $ToD(t)$) in one day period. (b) The wind turbine output power patterns under different weather conditions (i.e., $WV(t)$, $WS(t)$, and $HH(t)$) in one day period.

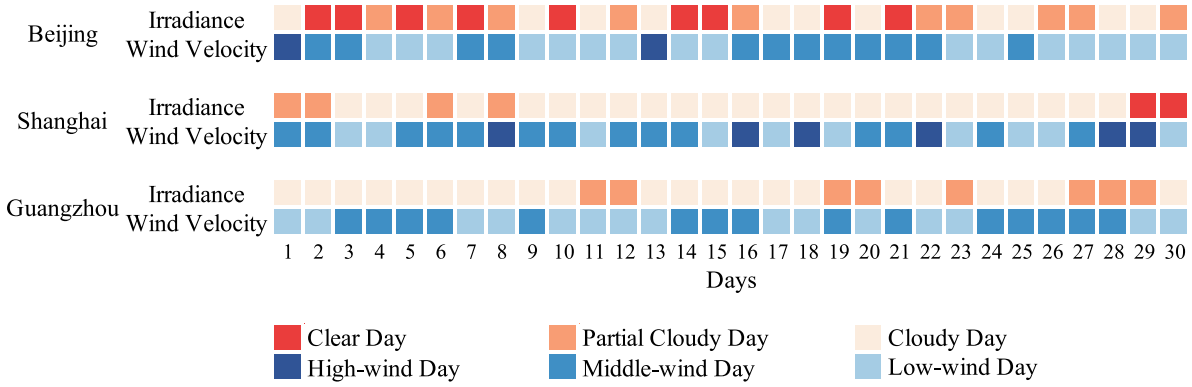


Fig. 7. The weather data is obtain from [28], and the billing cycle window is from 1st June 2020 to 30th June 2020.

weather conditions in these cities during the billing cycle window are shown in Fig. 7. Specifically, i) for *Beijing*, it has more clear days during the billing cycle window, ii) for *Shanghai*, it is in the plum rain season during the billing cycle window, thus it has more high-wind days but less clear days, and iii) for *Guangzhou*, the cloudy days and the low-wind days are relatively more than other two cities.

B. Performance Under Different Weather Conditions

As is shown in Fig. 6, the output power patterns of the solar PV and wind turbine are both divided into three types under different weather conditions. Accordingly, the weather pattern can be divided into nine types: *clear & high-wind day*, *clear & middle-wind day*, *clear & low-wind day*, *partial cloudy & high-wind day*, *partial cloudy & middle-wind day*, *partial cloudy & low-wind day*, *cloudy & high-wind day*, *cloudy & middle-wind day*, and *cloudy & low-wind day*.

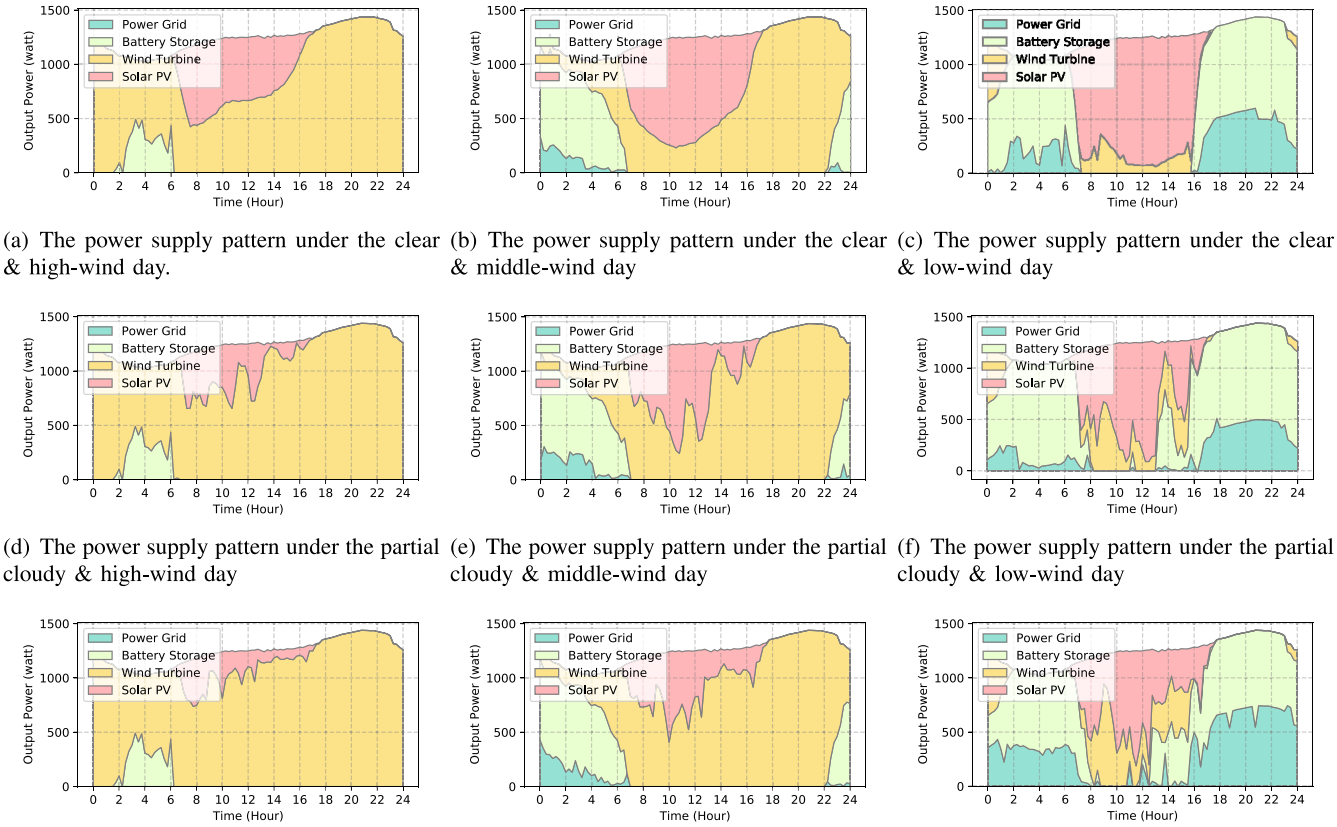
The power supply patterns under different weather conditions in one day period of 5G BS at the area of resident, office, and comprehensive are illustrated in Fig. 8, Fig. 10, and Fig. 11 (in the Appendix), respectively. As we can see, the BESS aided renewable energy supply solution could significantly reduce the power from the grid (i.e., *energy charge* and *demand charge*). Specifically, with the increase of radiation and wind velocity, renewable energy generation increased accordingly. It could cover most of the power demand and reduce the power supplied from the power grid. Especially, under high-wind days, the power demand could be totally supplied by the renewable energy and battery storage and need 0 power from the grid.

After calculating the power supply paradigm under different weather patterns, we can derive the electricity bill of these three types of BSs during the billing cycle in different cities (i.e., different weather patterns, which is illustrated in Fig. 7), and the results from all the set scenarios are summarized in Table III.

Specifically, for a single 5G BS without the proposed power supply paradigm, the energy charge and the demand charge are \$45.6 and \$22.8, respectively. However, after utilizing the BESS aided renewable energy supply solution on the 5G BS, the electricity bill is significantly reduced. Especially in Shanghai, which has relatively more clear and high-wind days, the energy charge and the demand charge can be reduced to \$3.8 and \$9.1, respectively. Although there exists equipment degradation during the discharge/charge cycles, the investment cost still keeps at a well accepted level. The highest cost saving for the BS which utilized the proposed power supply paradigm in Beijing, Shanghai, and Guangzhou in one billing cycle is \$50.4, \$50.7 and \$49.5, respectively. Accordingly, the saving ratio can be up to 74.4%, 74.8% and 73.2%, respectively.

C. Performance Under Different Types of BSs

As the different types of BSs has diverse power demand, resulting in different energy charge and demand charge, thus the performance of deployment of the BESS aided renewable energy supply solution could be different.



(a) The power supply pattern under the clear & high-wind day.

(b) The power supply pattern under the clear & middle-wind day

(c) The power supply pattern under the clear & low-wind day

(d) The power supply pattern under the partial cloudy & high-wind day

(e) The power supply pattern under the partial cloudy & middle-wind day

(f) The power supply pattern under the partial cloudy & low-wind day

(g) The power supply pattern under the cloudy & high-wind day

(h) The power supply pattern under the cloudy & middle-wind day

(i) The power supply pattern under the cloudy & low-wind day

Fig. 8. The power supply pattern of a single 5G BS at area of resident is supplied by different power supply methods under different weather conditions in one day period.

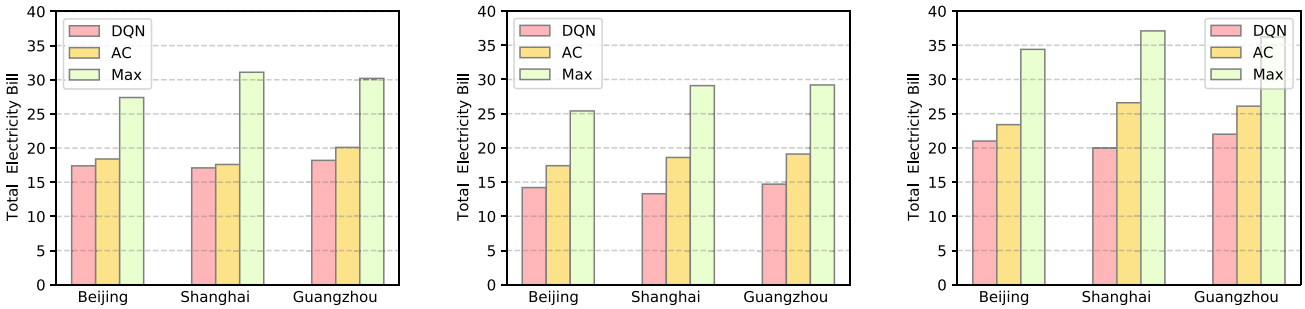
TABLE III
RESULTS SUMMARY (ONE BILLING CYCLE)

BS Type	Scenario	Energy Charge ($)	Demand Charge ($)	Investment Cost ($)	Cost Saving ($)	Saving Ratio (%)
Type I	No deployment	44.6	23.1	0	/	/
	Deployment in Beijing	5.0	12.0	0.4	50.4	74.4
	Deployment in Shanghai	4.7	12.0	0.4	50.7	74.8
	Deployment in Guangzhou	5.9	12.0	0.3	49.5	73.2
Type II	No deployment	40.1	20.2	0	/	/
	Deployment in Beijing	4.8	9.1	0.3	46.1	76.4
	Deployment in Shanghai	3.8	9.1	0.4	47.0	77.9
	Deployment in Guangzhou	5.3	9.1	0.3	45.6	75.6
Type III	No deployment	45.6	22.8	0	/	/
	Deployment in Beijing	6.8	13.9	0.3	47.4	69.3
	Deployment in Shanghai	5.7	13.9	0.4	48.4	70.8
	Deployment in Guangzhou	7.9	13.9	0.2	46.4	67.8

Specifically, as is shown in Table III, the type I BS has the highest cost saving compared to other two types of BSs, i.e., $50.4 in Beijing, $50.7, and $49.5. This is because that type I BS has the biggest power demand and peak value (near 1450 watts), making it has great potential in energy-saving and peak power shaving. Besides, as type II BS's power demands are relatively small, the generated and stored renewable energy can effectively reduce the power grid supply. Therefore it has the highest saving ratio, i.e., 76.4% in Beijing, 77.9% in Shanghai, and 75.6% in Guangzhou.

D. ROIs of Different Scenarios

The return of investment (ROI) is a financial metric defined by the benefit (cost saving in our case) divided by the total investment. It indicates the probability of gaining a return from an investment and has been widely used to evaluate the efficiency of an investment [29]. Typically, a bigger ROI value indicates a higher investment efficiency. With the costs of renewable energy generator and battery storage (given in Table II), the total investments can be calculated. Accordingly, the ROIs can thus be derived with the results in Table III.



(a) Total electricity bill of type I BS. (b) Total electricity bill of type II BS. (c) Total electricity bill of type III BS.

Fig. 9. Total electricity bills under different scheduling algorithms.

TABLE IV
PARAMETER SETTINGS

BS Type	Beijing	Shanghai	Guangzhou
Type I	5.43%	5.46%	5.33%
Type II	4.97%	5.06%	4.91%
Type III	5.11%	5.21%	5.00%

The ROIs of different types of BSs deployed in different cities are shown in Table IV. Specifically, type I BS has the highest ROI, which could reach to 5.43% in Beijing, 5.46% in Shanghai, and 5.33% in Guangzhou, respectively, indicating a relatively high investment efficiency for the operators. This is because that type I BS has the biggest cost saving.

As the equipment's cost is estimated to decrease dramatically in the future [30], and the ROI could rise significantly in 5G and beyond. Additionally, as we can see, the city with more clear and high-wind days will obtain a bigger ROI value, thus the proposed solution is more suitable for those cities with more sunny and windy days.

It is worth noting that, we assume the deployed renewable energy generator and the battery storage only supply power to one single 5G BS, and thus the surplus renewable energy (when the battery is full) will be discarded. This actually leads to a relatively low utilization, as given in this work. In practice, the generated renewable energy could supply to multiple BSs [5], so that the ROI and utilization of the renewable energy could be further improved.

E. Total Electricity Bill Under Different Algorithms

In order to reflect the performance of the proposed method, we mainly compare the total electricity bill with two baseline algorithms.

- *AC:* Which uses actor-critic (AC) method [31] to make the discharging/charging scheduling operations, one of the DRL methods.
- *Max:* Which satisfies the BS's current power demand to the greatest extent.

As shown in Fig. 9, because Max only meets the current power demand to the greatest extent and lacks predictability, it may consume too much power in the beginning and fail to discharge continuously when the power demand is high later. AC needs to train two networks (i.e., actor network and critic network), resulting in poor stability. Compared to AC, DQN can complete the operation in a shorter time, which can be converged under 300 iterations, proving the efficiency of the proposed method.

VII. RELATED WORK

The most involved related literatures can be divided into the following two categories.

A. Base Station Energy-Saving Method

With the increase of the BS power consumption, the energy-efficient cellular networks have recently received significant attention. One commonly used scheme is to switching-on/off the BS according to with the BS traffic load [32]–[34]. Intuitively, we can switch on the BS when the traffic load at the BS is high, and switch it off or turn it to sleep mode when the traffic load at the BS is low. In addition, by combining with AI, the accuracy of the traffic load prediction can be improved so that the corresponding energy-saving policies can be elaborately formulated. However, due to the shutdown of some BSs, the traffic latency may increase, degrading the QoS of wireless services.

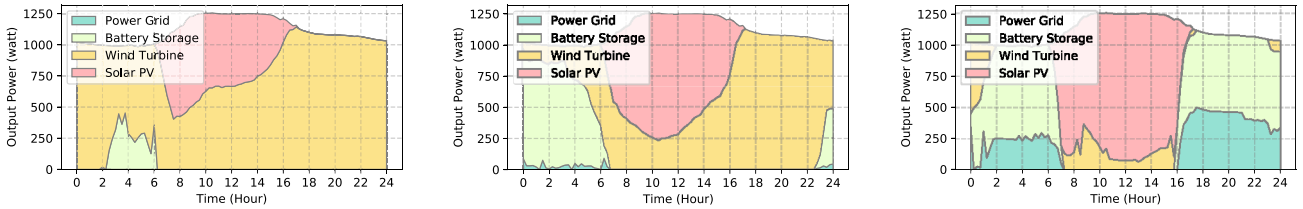
For energy management, the peak power shaving is a preferable approach to overcome the uneconomic and inefficiency of peak power supply, making the load curve flatten by reducing the peak amount of load and shifting it to times of lower load [35], [36]. Specifically, peak power shaving is achieved through charging energy storage system when demand is low (off-peak period) and discharging energy when demand is high (on-peak period). For task offloading, the total power consumption can also be reduced by dispatching tasks to BSs with lower loads [37], [38]. As shown in Table V, we have summarized the relevant literature.

B. Battery Storage Optimal Control

The optimal control of energy storage has been extensively studied in the past. Most related works formulate an optimization problem that aims to maximize the revenue generated by the battery storage co-located with renewable energy generator.

TABLE V
LITERATURE SUMMARY

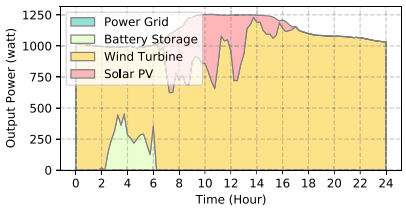
Literature	Objective	Approach	Solution
[32]	minimize the weight sum of energy and delay	BS ON-OFF switching	Decompose into two subproblems.
[33]	minimize energy over a period	BS ON-OFF switching	Evaluate on the network impact of turning ON/OFF.
[34]	maximize the energy efficiency	BS sleep mode design	Maximize a quasi-convex lower bound.
[35]	maximize the energy efficiency	peak shaving	Utilize a nonlinear programming method.
[36]	maximize the energy efficiency	peak shaving using EV	Convex optimization.
[37]	achieves load balancing	energy-aware resource allocation	min-max algorithm.
[38]	improve global resource utilization	energy-aware resource allocation	opportunistic scheduling algorithm.



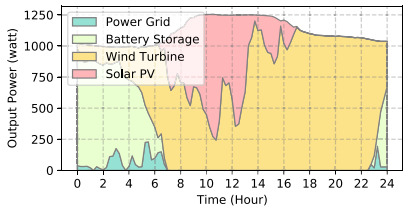
(a) The power supply pattern under the clear & high-wind day.

(b) The power supply pattern under the clear & middle-wind day

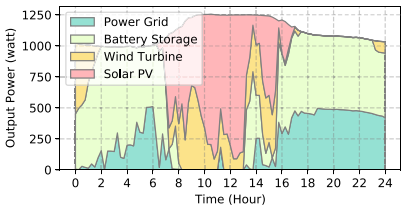
(c) The power supply pattern under the clear & low-wind day



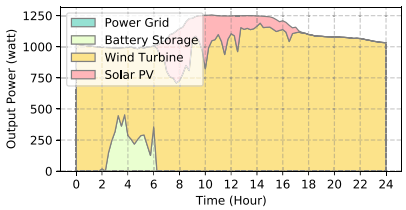
(d) The power supply pattern under the partial cloudy & high-wind day



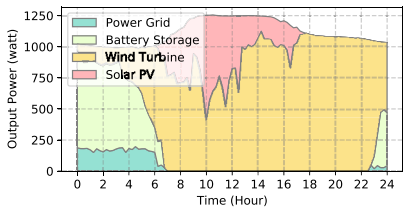
(e) The power supply pattern under the partial cloudy & middle-wind day



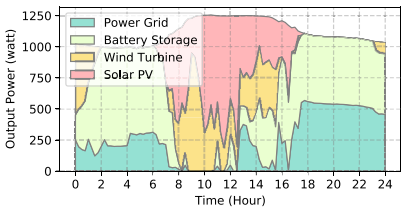
(f) The power supply pattern under the partial cloudy & low-wind day



(g) The power supply pattern under the cloudy & high-wind day



(h) The power supply pattern under the cloudy & middle-wind day



(i) The power supply pattern under the cloudy & low-wind day

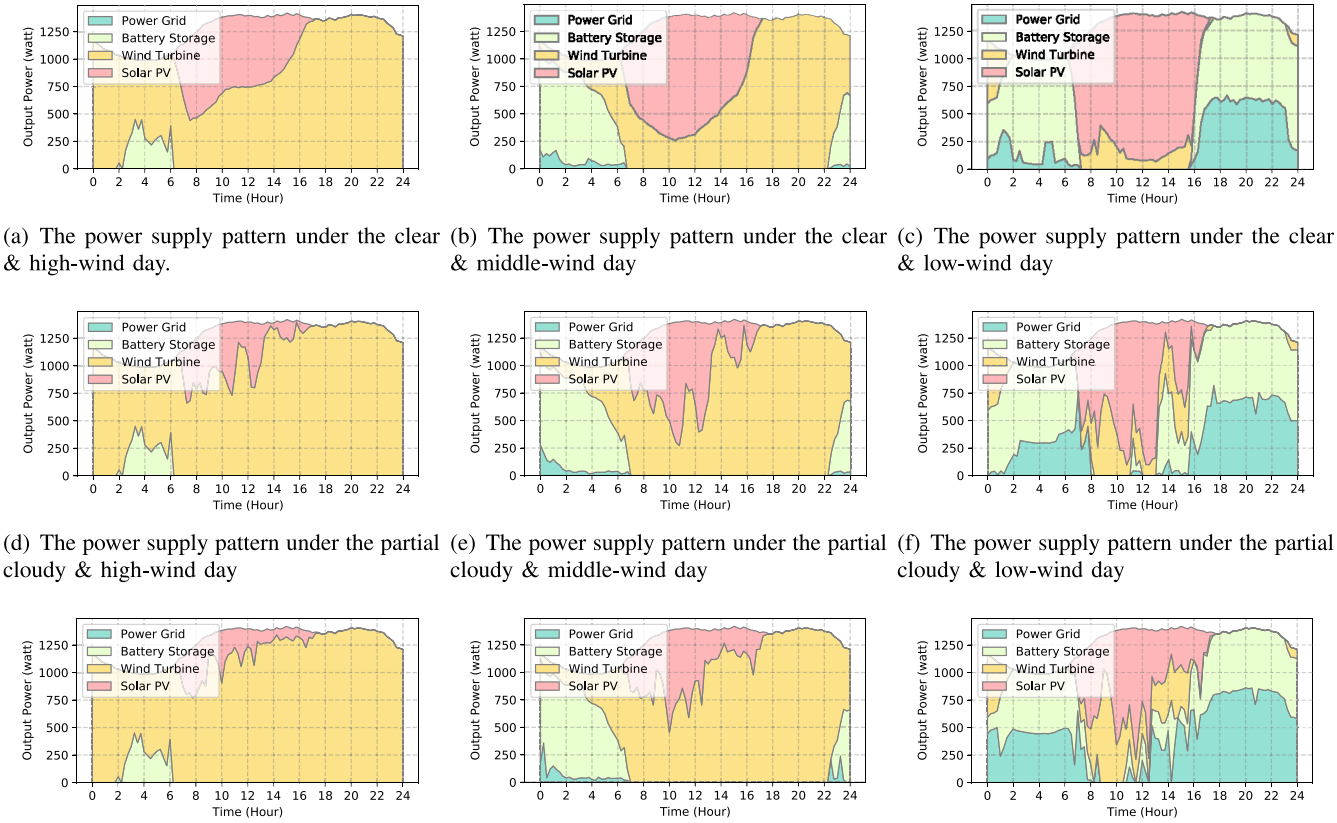
Fig. 10. The power supply pattern of a single 5G BS at area of office is supplied by different power supply methods under different weather conditions in one day period.

Babacan *et al.* [39] proposed a convex program to minimize the electricity bill of operators. Ratnam *et al.* [40] aimed to maximize the daily operational savings that accrue to customers while penalizing large voltage swings stemming from reverse power flow and peak load. Kazhamiaka *et al.* [41] studied the profitability of residential PV-storage systems in three jurisdictions and set up an integer linear program to determine the battery controlling policy. These works assume the generations of renewable energy and the power demand are known in advance and can be optimized in an offline way. However, these assumptions are unpractical in the real world.

Several papers study the optimal control of batteries under uncertainty and randomness. Guan *et al.* [42] utilized a reinforcement learning method to minimize the homeowner's cost by taking an action that yields the best expected reward. EnergyBoost [18] could provide a predictable ability of the renewable energy generation and power demand. However, these works are only applied in the home scenario, which generates a few demands compared to 5G BS. Therefore, we propose the DRL-based method to tackle the problem of large and constrained state- and action-space and the uncertainty of renewable energy generation and power demand.

VIII. CONCLUSION

To cope with the ever-increasing electricity bill for mobile operators in 5G era, we proposed a BESS aided energy supply



(a) The power supply pattern under the clear & high-wind day.

(b) The power supply pattern under the clear & middle-wind day

(c) The power supply pattern under the clear & low-wind day

(d) The power supply pattern under the partial cloudy & high-wind day

(e) The power supply pattern under the partial cloudy & middle-wind day

(f) The power supply pattern under the partial cloudy & low-wind day

(g) The power supply pattern under the cloudy & high-wind day

(h) The power supply pattern under the cloudy & middle-wind day

(i) The power supply pattern under the cloudy & low-wind day

Fig. 11. The power supply pattern of a single 5G BS at area of comprehensive is supplied by different power supply methods under different weather conditions in one day period.

solution for the 5G BS system, which models the battery discharging/charging as an optimization problem. With our proposed solution, besides the power grid, a BS can be powered by the renewable energy and the battery storage, to cut down the total energy cost. To solve the problem under the dynamic power demands and renewable energy generation, we developed a DRL-based approach to the BESS operation that accommodates all factors in the modeling phase and makes decisions in real-time. To evaluate the performance of our solution, we chose three cities with different weather patterns for experiments. The experimental results show that our power supply solution can achieve an cost saving ratio of 74.8% during the entire billing cycle and improve the renewable energy utilization.

In the future, with further development of the communication technology (e.g., B5G/6G), there will be more mobile BSs and air BSs equipped with more batteries, which could much rely on the renewable energy. Designing an effective battery discharging/charging policy to ensure the high QoS of mobile networks is also an interesting and challenging problem for future work.

APPENDIX
RESULTS FROM OFFICE & COMPREHENSIVE AREAS

See Figs. 10 and 11.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of Vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *Proc. IEEE World Forum Internet Things (WF-IoT)*, 2014, pp. 241–246.

[3] G. C. Burdea and P. Coiffet, *Virtual Reality Technology*. New York, NY, USA: Wiley, 2003.

[4] E. D. Muse, P. M. Barrett, S. R. Steinhubl, and E. J. Topol, "Towards a smart medical home," *Lancet*, vol. 389, no. 10067, p. 358, 2017.

[5] G. Tang, Y. Wang, and H. Lu, "ShiftGuard: Towards reliable 5G network by optimal backup power allocation," in *Proc. IEEE SmartGridComm*, 2020, pp. 1–6.

[6] H. Lund, "Renewable energy strategies for sustainable development," *Energy*, vol. 32, no. 6, pp. 912–919, 2007.

[7] R. Fu, D. Feldman, R. Margolis, M. Woodhouse, and K. Ardani, "U.S. solar photovoltaic system cost benchmark: Q1 2017," Nat. Renew. Energy Lab. (NREL), Golden, CO, USA, Rep. NREL/TP-6A20-68925, 2017.

[8] J. A. Turner, "A realizable renewable energy future," *Science*, vol. 285, no. 5428, pp. 687–689, 1999.

[9] X. Wang, A. V. Vasilakos, M. Chen, Y. Liu, and T. T. Kwon, "A survey of green mobile networks: Opportunities and challenges," *Mobile Netw. Appl.*, vol. 17, no. 1, pp. 4–20, 2012.

[10] B. Nykvist and M. Nilsson, "Rapidly falling costs of battery packs for electric vehicles," *Nat. Climate Change*, vol. 5, no. 4, pp. 329–332, 2015.

[11] A. Mondal, S. Misra, and M. S. Obaidat, "Distributed home energy management system with storage in smart grid using game theory," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1857–1866, Sep. 2017.

[12] H. Wang, F. Xu, Y. Li, P. Zhang, and D. Jin, "Understanding mobile traffic patterns of large scale cellular towers in urban environment," in *Proc. Internet Meas. Conf.*, 2015, pp. 225–238.

[13] H. Xu and B. Li, "Reducing electricity demand charge for data centers with partial execution," in *Proc. 5th Int. Conf. Future Energy Syst.*, 2014, pp. 51–61.
[14] M. Dabbagh, B. Hamdaoui, A. Rayes, and M. Guizani, "Shaving data center power demand peaks through energy storage and workload shifting control," *IEEE Trans. Cloud Comput.*, vol. 7, no. 4, pp. 1095–1108, Oct.–Dec. 2019.
[15] "Renewable Energy Generator." Qingdao Jinfan Energy Sci. Technol. 2019. [Online]. Available: http://www.jinfanenergy.cn
[16] W. F. Holmgren, R. W. Andrews, A. T. Lorenzo, and J. S. Stein, "PVLIB Python 2015," in *Proc. IEEE 42nd Photovolt. Specialist Conf. (PVSC)*, 2015, pp. 1–5.
[17] A. Jahid, M. S. Hossain, M. K. H. Monju, M. F. Rahman, and M. F. Hossain, "Techno-economic and energy efficiency analysis of optimal power supply solutions for green cellular base stations," *IEEE Access*, vol. 8, pp. 43776–43795, 2020.
[18] B. Qi, M. Rashedi, and O. Ardakanian, "EnergyBoost: Learning-based control of home batteries," in *Proc. 10th ACM Int. Conf. Future Energy Syst.*, 2019, pp. 239–250.
[19] Y. Shi, B. Xu, B. Zhang, and D. Wang, "Leveraging energy storage to optimize data center electricity cost in emerging power markets," in *Proc. 7th Int. Conf. Future Energy Syst.*, 2016, pp. 1–13.
[20] B. Aksanli, T. Rosing, and E. Pettis, "Distributed battery control for peak power shaving in datacenters," in *Proc. IEEE IGCC*, 2013, pp. 1–8.
[21] D. K. Maly and K.-S. Kwan, "Optimal battery energy storage system (BESS) charge scheduling with dynamic programming," *IEE Proc. Sci. Meas. Technol.*, vol. 142, no. 6, pp. 453–458, 1995.
[22] A. Oudalov, R. Cherkaoui, and A. Beguin, "Sizing and optimal operation of battery energy storage system for peak shaving application," in *Proc. IEEE Lausanne Power Tech.*, 2007, pp. 621–625.
[23] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[24] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
[25] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[26] "Rate 23—Industrial Power Service." Dominion Energy South Carolina. 2020. [Online]. Available: https://etariff.psc.sc.gov/Organization/TariffDetail/150?OrgId=411
[27] "Energy Storage Technology and Cost Characterization Report." U.S. Department of Energy. 2019. [Online]. Available: https://www.energy.gov/eere/water/downloads/energy-storage-technology-and-cost-characterization-report
[28] "Historical Weather Forecast." China Meteorological Administration. 2020. [Online]. Available: http://www.weather.com.cn/
[29] "Return on Investment." Wikipedia. 2020. [Online]. Available: https://en.wikipedia.org/wiki/Return_on_investment
[30] "Cost Projections for Utility-Scale Battery Storage." National Renewable Energy Laboratory (NREL). 2019. [Online]. Available: https://www.nrel.gov/docs/fy19osti/73222.pdf
[31] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Assoc., 2000, pp. 1008–1014.
[32] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, Sep. 2011.
[33] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.
[34] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1652–1661, Mar. 2016.
[35] E. Reihani, M. Motalleb, R. Ghorbani, and L. S. Saoud, "Load peak shaving and power smoothing of a distribution grid with high renewable energy penetration," *Renew. Energy*, vol. 86, pp. 1372–1379, Feb. 2016.
[36] C. G. Tse, B. A. Maples, and F. Kreith, "The use of plug-in hybrid electric vehicles for peak shaving," *J. Energy Resources Technol.*, vol. 138, no. 1, p. 7, 2016.
[37] Y. Bejerano and S.-J. Han, "Cell breathing techniques for load balancing in wireless LANs," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 735–749, Jun. 2009.
[38] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," in *Proc. 10th Annu. Int. Conf. Mobile Comput. Netw.*, 2004, pp. 302–314.
[39] O. Babacan, E. L. Ratnam, V. R. Disfani, and J. Kleissl, "Distributed energy storage system scheduling considering tariff structure, energy arbitrage and solar PV penetration," *Appl. Energy*, vol. 205, pp. 1384–1393, Nov. 2017.
[40] E. L. Ratnam, S. R. Weller, and C. M. Kellett, "An optimization-based approach to scheduling residential battery storage with solar PV: Assessing customer benefit," *Renew. Energy*, vol. 75, pp. 123–134, Mar. 2015.
[41] F. Kazhamiaka, P. Jochem, S. Keshav, and C. Rosenberg, "On the influence of jurisdiction on the profitability of residential photovoltaic-storage systems: A multi-national case study," *Energy Policy*, vol. 109, pp. 428–440, Oct. 2017.
[42] C. Guan, Y. Wang, X. Lin, S. Nazarian, and M. Pedram, "Reinforcement learning-based control of residential energy storage systems for electric bill minimization," in *Proc. 12th Annu. IEEE Consum. Commun. Netw. Conf. (CCNC)*, 2015, pp. 637–642.



Hao Yuan received the B.S. degree in management science and engineering from the National University of Defense Technology, Changsha, China, in 2019, where he is currently pursuing the M.S. degree. His main research interests include edge computing and green communication.



Guoming Tang (Member, IEEE) received the bachelor's and master's degrees from the National University of Defense Technology, China, in 2010 and 2012, respectively, and the Ph.D. degree in computer science from the University of Victoria, Canada, in 2017. He is a Research Fellow with the Peng Cheng Laboratory, Shenzhen, Guangdong, China. He was also a Visiting Research Scholar with the University of Waterloo, Canada, in 2016. His research mainly focuses on green computing, computing for green, and edge computing.



Deke Guo (Senior Member, IEEE) received the B.S. degree in industry engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in 2001, and the Ph.D. degree in management science and engineering from the National University of Defense Technology, Changsha, China, in 2008. He is currently a Professor with the College of System Engineering, National University of Defense Technology, and is also with the College of Intelligence and Computing, Tianjin University. His research interests include distributed systems, software-defined networking, data center networking, wireless and mobile systems, and interconnection networks. He is a member of the ACM.



Kui Wu (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from Wuhan University, China, in 1990 and 1993, respectively, and the Ph.D. degree in computing science from the University of Alberta, Canada, in 2002. He joined the Department of Computer Science, University of Victoria, Canada, in 2002, where he is currently a Full Professor. His research interests include smart grid, mobile and wireless networks, and network performance evaluation.



Xun Shao (Member, IEEE) received the Ph.D. degree in information science from the Graduate School of Information Science and Technology, Osaka University, Japan, in 2013. From 2013 to 2017, he was a Researcher with the National Institute of Information and Communications Technology (NICT), Japan. He is currently an Assistant Professor with the School of Regional Innovation and Social Design Engineering, Kitami Institute of Technology, Japan. His research interests include distributed systems and networking. He is a member of IEICE.



Keping Yu (Member, IEEE) received the M.E. and Ph.D. degrees from the Graduate School of Global Information and Telecommunication Studies, Waseda University, Tokyo, Japan, in 2012 and 2016, respectively. He was a Research Associate and a Junior Researcher with the Global Information and Telecommunication Institute, Waseda University, from 2015 to 2019 and from 2019 to 2020, where he is currently an Assistant Professor. His research interests include smart grids, information-centric networking, the Internet of Things, artificial intelligence, blockchain, and information security.



Wei Wei received the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2011, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an. He ran many funded research projects as a principal investigator and technical member. He has published over 100 research articles in international conferences and journals. His current research interests include the area of wireless networks, wireless sensor networks application, image processing, mobile computing, distributed computing, and pervasive computing, the Internet of Things, and sensor data clouds. He is an Editorial Board Member of the *Future Generation Computer System*, IEEE ACCESS, *Ad Hoc & Wireless Sensor Network*, Institute of Electronics, Information and Communication Engineers, and *KSII Transactions on Internet and Information Systems*. He is a Senior Member of the China Computer Federation.