A Simulation-Based Policy Improvement Method for Joint-Operation of Building Microgrids With Distributed Solar Power and Battery

Yuanming Zhang, *Student Member, IEEE*, and Qing-Shan Jia, *Senior Member, IEEE*

***Abstract*—This paper investigates the joint-operation of building microgrids with distributed solar power and storage battery. The purpose of this paper is to improve the energy efficiency of the building microgrids taking advantage of the complementary building load profiles. Three major contributions are made in this paper. First, in order to incorporate the uncertainties in both the building load and solar power as well as in the communication system, the operation problem is formulated as a finite-stage event-based optimization model. Second, an online simulation-based policy improvement method is developed to improve the given control policy and an action aggregation method is applied to accelerate the computational speed. Third, a balanced battery operation strategy considering the expected life cycles is proposed. Numerical examples based on the building microgrids in a university campus are used to demonstrate the effectiveness of the proposed method in improving the building energy efficiency. The computational time in a practical system is also analyzed.**

***Index Terms*—Building microgrids, joint-operation, event-based optimization, simulation-based policy improvement.**

NOMENCLATURE

$I_{\rm pv}, I_0, I_{\rm pv,n}$	Current generated by incident light, reverse saturation current, light-generated current in nominal condition.
V, I	Output voltage and current.
$R_{\rm s}, R_{\rm p}$	Series resistance and parallel resistance.
$N_{\rm s}, N_{\rm p}$	Number of pv cells connected in series and in parallel.
$T_{\rm a}, T_{\rm n}$	Actual and nominal (25°C) P-N junction temperature.
$G_{\rm a}, G_{\rm n}$	Actual and nominal ($1000{\rm W/m^2}$) solar irradiance.
q, k	Electron charge ($1.60217646 \times 10^{-19}$C) and Boltzmann constant ($1.30806503 \times 10^{-23}$J/K).
a, a_1	Diode ideality constants.
$V_{\rm oc,n}, I_{\rm sc,n}$	Open-circuit voltage and short-circuit current.
γ_i^t	SOC of battery i in stage t.
$Q_{{\rm b},i}$	Capacity of battery i.
$p_{{\rm b},i}^t$	Charging/discharging power in stage t.
$\underline{P}_{{\rm b},i}/\overline{P}_{{\rm b},i}$	Minimum/maximum value of $p_{{\rm b},i}^t$.
$\underline{\Gamma}_i/\overline{\Gamma}_i$	Minimum/maximum level of SOC.
$p_{{\rm pv},i}^t$	Solar power generation/prediction/observation of node i in stage t.
$\bar{p}_{{\rm pv},i}^t, \tilde{p}_{{\rm pv},i}^t$	Predicted/observed solar power of node i in stage t.
$p_{{\rm pv,l},i,j}^t$	Solar power delivered from node i to satisfy building load in node j.
$p_{{\rm pv,b},i,j}^t$	Solar power delivered from node i to charge battery in node j in stage t.
$p_{{\rm pv,g},i}^t$	Solar power fed into the grid in node i in stage t.
$p_{{\rm b,l},i,j}^t$	Discharging battery power in node i delivered to building load in node i in stage t.
$p_{{\rm g},i}^t$	Total power purchased from the grid in node i in stage t.
$p_{{\rm g,l},i}^t$	Power purchased in node i for building load in stage t.
$p_{{\rm g,b},i}^t$	Power purchased in node i for battery charging in stage t.
$p_{{\rm l},i}^t$	Building load in node i in stage t.
λ^t, ν	Time-of-use (TOU) electricity price and selling price of solar power.
$\varepsilon_{\rm pv}, \varepsilon_{\rm l}$	Prediction error of solar power and building load.
$\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{a}, \boldsymbol{\pi}$	System state, event, action and event-based belief state.
$I_{\rm d}(x)$	Indicator function, with $I_{\rm d}(x) = 1$ if $x > 0$ and $I_{\rm d}(x) = 0$ otherwise.
$\hat{c}^t(\boldsymbol{s}_j \vert \boldsymbol{s}_i, \boldsymbol{a}^t)$	One-step cost from state i to j after taking action $\boldsymbol{a}^t$.
$c^t(\boldsymbol{s}_i, \boldsymbol{a}^t)$	One-step cost for state i after taking action $\boldsymbol{a}^t$.
$C^t(\boldsymbol{\pi}^t, \boldsymbol{a}^t)$	One-step cost for belief state $\boldsymbol{\pi}^t$ after taking action $\boldsymbol{a}^t$.

Manuscript received July 30, 2016; revised January 16, 2017 and March 29, 2017; accepted May 8, 2017. Date of publication May 23, 2017; date of current version October 19, 2018. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0901905, in part by the National Natural Science Foundation of China under Grant 61673229, Grant 61174072, Grant 61222302, Grant 91224008, and Grant U1301254, in part by the Tsinghua National Laboratory for Information Science and Technology Funding for Excellent Young Scholar, in part by the Program for New Star in Science and Technology in Beijing under Grant xx2014B056, and in part by the 111 International Collaboration Program of China under Grant B06002. Paper no. TSG-01003-2016. *(Corresponding author: Qing-Shan Jia.)*

The authors are with the Department of Automation, Center for Intelligent and Networked Systems, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: zhang-ym12@mails.tsinghua.edu.cn; jiaqs@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2017.2707098

$\boldsymbol{d}, \boldsymbol{d}_{\text{opt}}$	operation policy and optimal policy.
$J(\boldsymbol{\pi}^1, \boldsymbol{d})$	Total cost with policy $\boldsymbol{d}$ from initial belief-state $\boldsymbol{\pi}^1$.
$Q^t, \widetilde{Q}^t, V^t$	Q-factor, estimated Q-factor and value function.
$\bar{P}^t_{\text{pv}}, \widetilde{P}^t_{\text{pv}}$	Predicted/observed solar power in stage t.
$\bar{P}^t_{\text{l}}, \widetilde{P}^t_{\text{l}}$	Predicted/observed load in stage t.
$P_{\text{pv,r}}$	Rated power of BIPV system.
$P_{\text{pv,d/c}}$	Maximum discharging/charging power.

I. Introduction

MOTIVATED by increasing demand for electricity and shortage of fossil fuels, improving the energy efficiency of large public buildings has been attracting people's attention. Technology advances has made a building operate as a smart microgrid with distributed renewable energy resources (DRER) (e.g., solar power and wind power) and storage battery, etc. Lots of research have focused on the operation of a single building microgrid [1], [2] and showed that an optimized operation policy can improve the building energy efficiency while satisfying the power demand and comfort requirement.

With the development of cyber-physical system (CPS) technology, communication extends from inside one single building to among a group of buildings as in a community or a university campus. With various functions, the buildings usually have different load profiles and some may even be complementary in a day. This operation characteristics bring the opportunity to jointly manage these buildings for further energy efficiency improvement.

While for joint operation, there exist some difficulties to obtain the optimal control policy. First, uncertainties exist in both DRER and building load, which are influenced by the environment (solar, wind, and temperature, etc.) and the occupants' behavior [3] separately. Second, there also exist measurement and communication errors [4], with which the descriptions of the energy systems are inaccurate. Third, there usually exist man-made rules on storage battery operation which make it a discrete-event dynamic system (DEDS) [2]. Fourth, the joint-operation of the building microgrids is spatiotemporally coupled, where the control decisions in one building not only influence the others' decisions but also influence all their following decisions. Fifth, the operation decisions are usually made for a finite-time horizon (as in one day) and obtaining the corresponding unstationary control policy is time consuming.

Considering the above difficulties and there usually exist rule-based or heuristic control policies in practice, in this paper we focus our investigation on improving the existent control policy and make the following major contributions. First, we formulate the joint-operation of building microgrids as a finite-stage event-based optimization (EBO) problem taking into account the uncertainties of distributed solar power and building load, as well as the randomness in the sensing and communication system. Second, we develop a simulation-based policy improvement (SBPI) method to solve the problem, where simulation method is used to evaluate the value function and action aggregation method is used to accelerate the policy improvement process. Third, based on the improved control policy, a balanced operation strategy for distributed batteries is further developed considering their expected life-cycles. Numerical examples based on the campus of Tsinghua University in Beijing, China are presented to demonstrate the performance of the proposed method and the test results show that the SBPI method can improve the performance of the given base policy and enhance the energy efficiency of the building microgrids. The application of the proposed method to a larger-scale practical system is also discussed.

The rest of this paper is organized as follows. We first make a brief literature review of energy management for building microgrids in Section II, then formulate the joint-operation problem in Section III, further present the solution methodology in Section IV, discuss the numerical results in Section V, and briefly conclude in Section VI.

II. Literature Review

Abundant research has been carried out on optimizing the operation of various kinds of microgrids. Our attention is focused on the building microgrids with renewables. The following representative literatures using several popular operation optimization methods are briefly reviewed and discussed.

The first category is based on mixed-integer linear programming (MILP). The MILP formulation is used by Stadler *et al.* [5] for distributed energy adoption of commercial buildings and by Khodaei and Shahidehpour [6] for optimal power scheduling of a community-based microgrid. In these applications, the uncertainties of power demand and supply are not considered, which decreases the performance of the acquired control policy and increases the risk [7] in practical operation. Then the second category as stochastic programming (SP) considering system uncertainties is introduced for the energy management of microgrid. An SP formulation is proposed by Guan *et al.* [1] for the operation optimization of an office building microgrid with various renewables and a two-stage SP is developed by Wang *et al.* [8] for the demand-side energy management of a commercial building. These stochastic formulations can better accommodate system uncertainties and can better adapt to practical situations in the operation compared to deterministic methods. While the scenario method is usually integrated with the SP to facilitate the solving process by transforming the stochastic formulation to a large-scale deterministic formulation. This solving process usually increases the computational complexity. The third category is intelligent optimization. The purpose of applying intelligent optimization to microgrid energy management is to find the global optimal control strategy, as Chen *et al.* [9] using genetic algorithm to optimize the load management problem and Wang *et al.* [10] applying particle swarm optimization to the energy and comfort management of smart buildings with plug-in electric vehicles. As the searching process of intelligent optimization makes use of stochastic heuristics, finding

the global and optimal control policy is usually probabilistic and cannot be ensured.

The aforementioned energy management methods are usually applied in an off-line manner, which makes the system operator unable to use the real-time or newly obtained information. Then different approaches that can deal with real-time information in an on-line manner are developed. Both Shi *et al.* [11] and Salinas *et al.* [12] developed an on-line energy management strategy for real-time operation of microgrids with Lyapunov optimization from the networked optimization theory for queueing systems. A detailed description of Lyapunov optimization can be found in [13]. Through a drift-plus-penalty process, Lyapunov optimization provides a good paradigm for addressing the problem of real-time control. When facing time-coupled constraints [12] that cannot be handled directly by standard Lyapunov optimization techniques, relaxation is usually applied which increases the complexity of the problem solving process [12], [14], [15]. Another method is model predictive control (MPC), which is preferable in the energy management of microgrid to deal with real-time information of uncertain load and renewables for its iterative sampling and optimizing process. Zong *et al.* [16] implemented MPC to the load management of an office building microgrid and Mayer *et al.* [17] presented a mixed-integer MPC method to optimize the management of hybrid energy supply systems in a building-complex of a university. In these applications, the key to assure the performance of the MPC is the sampling process which predicts the future system dynamics.

The SBPI method combines simulation-based optimization [18], [19] and rollout method [20], where simulation is used to evaluate the future influence of current action and rollout is used to improve the given control policy iteratively. For energy scheduling problem, SBPI method has been used for energy management of a commercial office building [2] and for matching wind power with EV charging load [21]–[23] under typical framework of Markov decision process (MDP). In this paper, the SBPI method is extended to solve the operation problem of building microgrids with an EBO formulation. The proposed SBPI method is related to the MPC method. First, they are both based on sampling. In the MPC, sampling is usually used to predict the future model states and in the SBPI, sampling is used to evaluate the future value function. Second, in a finite-time decision process the prediction horizon in the MPC and the evaluation horizon in the SBPI are both shifting forward which makes their control horizon recede with time going forward. Third, they are both applied on-line with real-time information. This makes the control process robust to system uncertainties. However, there also exist differences in their optimization processes. In the MPC, the control strategy is optimized for both the decision time and a future time horizon, whereas in the SBPI, the optimization is made only for the decision time. What's more, the MPC provides an optimal control strategy based on the current states and the future model prediction in one iteration, while the SBPI provides an improved control strategy based on some given base policy, which facilitates its using of the existent control policy that can be provided by the aforementioned various off-line methods and the performance improvement can be assured [24]. Note that the SBPI method and the MPC method share the same ideas of moving window and receding horizon in a finite-stage control process. The SBPI method is used for performance improvement of some intractable optimization problems where the optimal decision and even the sub-optimal decision are difficult to compute, whereas the MPC method is used to optimize the control decision based on the influence in a future moving window. Note also that the SBPI method focuses on improvement, whereas the aim of the MPC method is to compute the optimal decision.

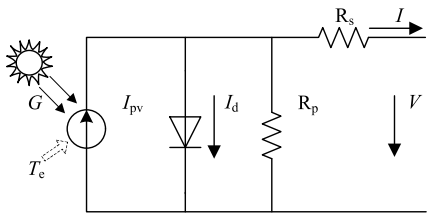


Fig. 1. Circuit diagram of a one-diode PV model for BIPV system.

III. Problem Formulation

Consider a building microgrid with N buildings. A building-integrated photovoltaic (BIPV) system and a storage battery (e.g., Tesla powerpack) are affiliated to each building. Buildings can purchase electricity from the grid for consumption and battery charging or consume their own solar power and battery power. Besides charging the battery, surplus solar power can be fed to the utility grid. In the joint-operation, surplus solar power in one building can be scheduled and utilized by other buildings instead of sold to the grid. What's more, considering the time-of-use (TOU) electricity price, battery in each building can be charged at low-rate hours and shares its power with other buildings when the rates are high. The control decisions for the joint-operation are made on an hourly basis in one day. Considering the aforementioned system uncertainties, the operation process is formulated as an EBO problem under the MDP framework.

A. System Description

1) BIPV: BIPV system is used to harvest solar power for each building. The one-diode PV model [25] based on the physics of P-N junction and the Thevenin's theorem for solar power generation process is shown in Fig. 1 and the corresponding mathematical formulations are from (1) to (4).

$$p_{\mathrm{pv}} = V\Big(I_{\mathrm{pv}} - I_{\mathrm{d}} - \frac{V + R_{\mathrm{s}} I}{R_{\mathrm{p}}}\Big) \tag{1}$$

$$I_{\mathrm{pv}} = N_{\mathrm{p}}\big(I_{\mathrm{pv,n}} + K_{\mathrm{I}}(T_{\mathrm{a}} - T_{\mathrm{n}})\big)\frac{G_{\mathrm{a}}}{G_{\mathrm{n}}} \tag{2}$$

$$I_{\mathrm{d}} = I_0\Big(\exp\Big\{\frac{q(V + R_{\mathrm{s}} I)}{a_1 \mathrm{k} N_{\mathrm{s}} T_{\mathrm{a}}}\Big\} - 1\Big) \tag{3}$$

$$I_0 = N_{\mathrm{p}} \frac{I_{\mathrm{sc,n}} + K_{\mathrm{I}}(T_{\mathrm{a}} - T_{\mathrm{n}})}{\exp\big\{\big(V_{\mathrm{oc,n}} + K_{\mathrm{V}}(T_{\mathrm{a}} - T_{\mathrm{n}})\big)/(aV_{\mathrm{t,n}})\big\} - 1} \tag{4}$$

Detailed descriptions of all the above parameters can be found in [25]. As the solar power output of this PV model is close

to that of the real products [25], it is used for solar power prediction with measured ground irradiance and temperature in the following numerical tests. Solar power generation is mainly influenced by solar irradiance G_{a} and P-N junction temperature T_{a}, which lead to the uncertainty of p_{pv}. The solar power prediction error $\varepsilon_{\mathrm{pv}}$ is assumed to be subject to Gaussian distribution as $\varepsilon_{\mathrm{pv}} \sim N(0, \sigma_{\mathrm{pv}}^2)$ and the relationship between the real generation p_{pv} and the prediction $\bar{p}_{\mathrm{pv}}$ is

$$p_{\mathrm{pv}} = (1 + \varepsilon_{\mathrm{pv}})\bar{p}_{\mathrm{pv}}. \tag{5}$$

As there also exists randomness in the measurement and communication, the observed (measured) solar power $\widetilde{p}_{\mathrm{pv}}$ by the coordinator is inaccurate. For solar power generation, as the observation $\widetilde{p}_{\mathrm{pv}}$ and the real-time prediction $\bar{p}_{\mathrm{pv}}$ are made before the decision instant, then instead of the real system states, the relationship between the observation and the prediction ($\widetilde{p}_{\mathrm{pv}} <$ or $\geq \bar{p}_{\mathrm{pv}}$) acquired by the system coordinator is used in the following event definition for the event-based optimization process.

2) Storage Battery: Besides storing solar power, battery can also shift the low-rate grid power to high-rate hours through charging/discharing control considering the TOU electricity price. In a discrete-time form, the dynamics of battery SOC (state of charge) between two decision stages is

$$\gamma_i^{t+1} = \gamma_i^t - \frac{p_{\mathrm{b},i}^t \delta t}{Q_{\mathrm{b},i}}, \tag{6}$$

with $\gamma_i^t \in [\underline{\Gamma}_i, \overline{\Gamma}_i] \subseteq (0, 1)$ and $p_{\mathrm{b},i}^t \in [\underline{P}_{\mathrm{b},i}, \overline{P}_{\mathrm{b},i}]$. δt is the operation time horizon. Considering the lifespan and efficiency, battery shouldn't be charged and discharged at the same time. Meanwhile, in order to avoid over-charging and deep-discharging, SOC is restricted between $[\underline{\Gamma}_i, \overline{\Gamma}_i]$.

3) Power Utilization: In the joint-operation, solar power and battery power in one building can be scheduled and utilized by other buildings besides local consumption. For building i, the power utilizations are formulated mathematically as (7)-(9).

a) Power from BIPV:

$$p_{\mathrm{pv},i}^t = \sum_{j=1}^{N} \left(p_{\mathrm{pv,l},i,j}^t + p_{\mathrm{pv,b},i,j}^t \right) + p_{\mathrm{pv,g},i}^t \tag{7}$$

b) Battery charging/discharging power:

$$\begin{cases} p_{\mathrm{b},i}^t = \sum_{j=1}^{N} p_{\mathrm{pv,b},j,i}^t + p_{\mathrm{g,b},i}^t, & \text{if } p_{\mathrm{b},i}^t \leq 0; \\ p_{\mathrm{b},i}^t = \sum_{j=1}^{N} p_{\mathrm{b,l},i,j}^t, & \text{if } p_{\mathrm{b},i}^t > 0. \end{cases} \tag{8}$$

c) Grid power:

$$p_{\mathrm{g},i}^t = p_{\mathrm{g,l},i}^t + p_{\mathrm{g,b},i}^t \tag{9}$$

Note that if the battery is being discharged, i.e., $p_{\mathrm{b},i}^t > 0$, power scheduling decision related to the charging decision become zero and vice versa.

4) Power Balance Constraints: Power balance is the coordination between power supply and demand. In the joint-operation, for building load $p_{\mathrm{l},i}^t$, we have

$$\begin{cases} \sum_{j=1}^{N} \left(p_{\mathrm{pv,l},j,i}^t + p_{\mathrm{b,l},j,i}^t \right) + p_{\mathrm{g,l},i}^t = p_{\mathrm{l},i}^t, & \text{if } p_{\mathrm{g,l},i}^t \geq 0; \\ \sum_{j=1}^{N} \left(p_{\mathrm{pv,l},j,i}^t + p_{\mathrm{b,l},j,i}^t \right) - p_{\mathrm{pv,g},i}^t = p_{\mathrm{l},i}^t, & \text{if } p_{\mathrm{pv,g},i}^t \geq 0. \end{cases} \tag{10}$$

Here $p_{\mathrm{g,l},i}^t \geq 0$ means additional power is purchased from the grid to satisfy the demand and $p_{\mathrm{pv,g},i}^t \geq 0$ means there exists surplus solar power that can be sold to the grid. The purchasing and selling decisions are incompatible. The load prediction error ε_{l} for each building is also assumed to be subject to Gaussian distribution as $\varepsilon_{\mathrm{l}} \sim N(0, \sigma_{\mathrm{l}}^2)$ and the real building load is $p_{\mathrm{l}} = (1+\varepsilon_{\mathrm{l}})\bar{p}_{\mathrm{l}}$. Like solar power, the relationship between the observed (measured) building load $\widetilde{p}_{\mathrm{l}}$ and the prediction $\bar{p}_{\mathrm{l}}$ as $\widetilde{p}_{\mathrm{l}} < (\geq)\bar{p}_{\mathrm{l}}$ is also used in the following event definition.

Note that in the above formulation of power scheduling and utilization, power loss in the transmission among different buildings is not considered. First, as the buildings are powered by the same distribution grid and the geographical span is relatively smaller for a microgrid compared to the power transmission in the utility grid, the power loss during transmission can be neglected. Second, the PV inverter efficiency has reached about 95% [26] and some products have even reached 98% [27]. Then the loss of power inversion can also be neglected. Third, for the power loss of battery charging/discharging, a further discussion is made considering practical application in the following solution methodology section.

B. EBO for Joint-Operation of Building Microgrids

MDP has been applied to various kinds of smart grids to capture the system randomness [2], [21], [28], where the system state is usually assumed to be completely observed. In our building microgrids, we can obtain the relationship between the observation and the prediction instead of the real system state and the decision is made based on the state estimation which is a probability distribution over the discretized state space. We formulate the joint-operation as a finite-stage EBO [29], where the decision is made based on the observed event and the notations are made as follows.

Define $\boldsymbol{s} = (\boldsymbol{s}_{b,1}, \boldsymbol{s}_{b,2}, \ldots, \boldsymbol{s}_{b,N})$ as the system state, where $\boldsymbol{s}_{b,i} = (p_{\mathrm{pv},i}, p_{\mathrm{l},i}, \gamma_i)$ is the state of the i-th building. Event is defined based on the observation and the prediction as $\boldsymbol{e} = (\boldsymbol{e}_{b,1}, \boldsymbol{e}_{b,2}, \ldots, \boldsymbol{e}_{b,N})$, where $\boldsymbol{e}_{b,i} = (e_{b,i}^1, e_{b,i}^2)$ and $e_{b,i}^1 \in \{\widetilde{p}_{\mathrm{pv},i} < \bar{p}_{\mathrm{pv},i}, \widetilde{p}_{\mathrm{pv},i} \geq \bar{p}_{\mathrm{pv},i}\}$, $e_{b,i}^2 \in \{\widetilde{p}_{\mathrm{l},i} < \bar{p}_{\mathrm{l},i}, \widetilde{p}_{\mathrm{l},i} \geq \bar{p}_{\mathrm{l},i}\}$. The action of building i is $\boldsymbol{a}_{b,i} = (p_{\mathrm{b,i}}, p_{\mathrm{g,i}}, p_{\mathrm{pv,g,i}})$, and the system action is $\boldsymbol{a} = (\boldsymbol{a}_{b,1}, \ldots, \boldsymbol{a}_{b,N})$. After taking action $\boldsymbol{a}$, the real system state transits from one to another according to the transition rule $p_r(\boldsymbol{a})$. Then the system can be described with a tuple $\langle \boldsymbol{s} \in \boldsymbol{S}, \boldsymbol{a} \in \boldsymbol{A}, \boldsymbol{e} \in \boldsymbol{ET}, p_r \in P_r \rangle$, where $\boldsymbol{S}$, $\boldsymbol{A}$, $\boldsymbol{ET}$ and P_r are the state space, action space, event space and transition probability space. Given the initial probability distribution on the system state as $\boldsymbol{\pi}^1$, the system evolves as follows.

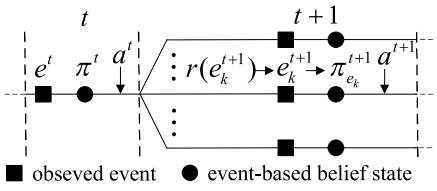


Fig. 2. Decision process from stage t to stage $t+1$ in event-based optimization of building microgrids operation.

In stage t, assume we have the event-based belief state [30] $\boldsymbol{\pi}^t = (\pi^t(\boldsymbol{s}_1), \ldots, \pi^t(\boldsymbol{s}_{|\boldsymbol{S}|}))$, where $\pi^t(\boldsymbol{s}_i)$ is the probability of being in state $\boldsymbol{s}_i$. After taking action $\boldsymbol{a}^t$, the prior probability of being in state $\boldsymbol{s}_j$ is

$$\pi_{\text{pri}}^{t+1}(\boldsymbol{s}_j) = \sum_{i=1}^{|\boldsymbol{S}|} \pi^t(\boldsymbol{s}_i) p_r\left(\boldsymbol{s}_j|\boldsymbol{s}_i, \boldsymbol{a}^t\right). \tag{11}$$

After transition to $\boldsymbol{s}_j$, the conditional probability [29] of observing event $\boldsymbol{e}_k$ is

$$r\left(\boldsymbol{e}_k|\boldsymbol{s}_j\right) = \phi_{\boldsymbol{e}_k}\left(\boldsymbol{s}_j\right), \tag{12}$$

where $\phi_{\boldsymbol{e}_k}(\boldsymbol{s}_j) = 1$ if $\boldsymbol{s}_j$ satisfies the necessary and sufficient condition of observing event $\boldsymbol{e}_k$ and otherwise 0. Then the posterior probability of being in state $\boldsymbol{s}_j$ with this observed event is

$$\pi_{\text{pos}}^{t+1}\left(\boldsymbol{s}_j|\boldsymbol{e}_k\right) = \frac{r\left(\boldsymbol{e}_k|\boldsymbol{s}_j\right)\pi_{\text{pri}}^{t+1}\left(\boldsymbol{s}_j\right)}{\sum_{\{\boldsymbol{s}_j|\phi_{\boldsymbol{e}_k}(\boldsymbol{s}_j)=1\}} r\left(\boldsymbol{e}_k|\boldsymbol{s}_j\right)\pi_{\text{pri}}^{t+1}\left(\boldsymbol{s}_j\right)}. \tag{13}$$

This probability is an update on the prior probability $\pi_{\text{pri}}^{t+1}(\boldsymbol{s}_j)$ and the decision process is shown in Fig. 2.

After taking action $\boldsymbol{a}^t$, the one-step cost of transition from $\boldsymbol{s}_i$ to $\boldsymbol{s}_j$ when observing $\boldsymbol{e}_k$ is

$$\hat{c}^t\left(\boldsymbol{s}_j|\boldsymbol{s}_i, \boldsymbol{e}_k, \boldsymbol{a}^t\right) = \sum_{n=1}^{N}\left[p_{\text{g},n}^t \lambda^t I_{\text{d}}\left(p_{\text{g},n}^t\right) - p_{\text{pv,g},n}^t \nu I_{\text{d}}\left(p_{\text{pv,g},n}^t\right)\right], \tag{14}$$

where the subscript n means the n-th building. Note that the power purchasing decision and selling decision are incompatible with each other and the action taken must satisfy the power utilization and power balance constraints in each building.

For $\boldsymbol{\pi}^t$, the one-step expected cost for taking action $\boldsymbol{a}^t$ is

$$C^t\left(\boldsymbol{\pi}^t, \boldsymbol{a}^t\right) = \sum_{i=1}^{|\boldsymbol{S}|} \pi^t(\boldsymbol{s}_i) c^t\left(\boldsymbol{s}_i, \boldsymbol{a}^t\right), \tag{15}$$

with

$$c^t\left(\boldsymbol{s}_i, \boldsymbol{a}^t\right) = \sum_{j=1}^{|\boldsymbol{S}|} p_r\left(\boldsymbol{s}_j|\boldsymbol{s}_i, \boldsymbol{a}^t\right) \sum_{k=1}^{|\boldsymbol{ET}|} r\left(\boldsymbol{e}_k|\boldsymbol{s}_j\right)\hat{c}^t\left(\boldsymbol{s}_j|\boldsymbol{s}_i, \boldsymbol{e}_k, \boldsymbol{a}^t\right). \tag{16}$$

Given a control policy $\boldsymbol{d} = (d^1(\boldsymbol{ET}, \boldsymbol{\Pi}), \ldots, d^T(\boldsymbol{ET}, \boldsymbol{\Pi}))$, where $\boldsymbol{\Pi}$ is the belief state space and $d^t(\boldsymbol{ET}, \boldsymbol{\Pi})$ is the decision rule in stage t as $d^t(\boldsymbol{ET}, \boldsymbol{\Pi})$: $\boldsymbol{ET} \times \boldsymbol{\Pi} \to \boldsymbol{A}$. The total T-stage cost by adopting policy $\boldsymbol{d}$ from the initial belief state $\boldsymbol{\pi}^1$ is

$$J\left(\boldsymbol{\pi}^1, \boldsymbol{d}\right) = \sum_{t=1}^{T} C^t\left(\boldsymbol{\pi}^t, d^t(\boldsymbol{ET}, \boldsymbol{\Pi})\right). \tag{17}$$

Our interest is to find the optimal policy $\boldsymbol{d}_{\text{opt}}$ in the policy space $\boldsymbol{D}$ which can minimize the total electricity cost as

$$J\left(\boldsymbol{\pi}^1, \boldsymbol{d}_{\text{opt}}\right) = \min_{\boldsymbol{d}\in\boldsymbol{D}} J\left(\boldsymbol{\pi}^1, \boldsymbol{d}\right). \tag{18}$$

IV. Solving Methodology

Considering the difficulties in obtaining the optimal control policy, the previous problem is first reformulated with action aggregation. Then, an on-line SBPI method is developed to improve the given base policy. Further, with the improved policy, a balanced battery operation method is developed considering the expected life-cycles.

A. Problem Reformulation

The electricity cost of the system is induced by the total power trade with the utility grid. Then the system coordinator need only care about the total power supply and demand. In decision stage t, denote the total building load and total solar power of the building microgrids as P_{l}^t and P_{pv}^t, with $P_{\text{l}}^t = \sum_{i=1}^N p_{\text{l},i}^t$ and $P_{\text{pv}}^t = \sum_{i=1}^N p_{\text{pv},i}^t$. Then the system power balance is

$$P_{\text{l}}^t = P_{\text{pv}}^t + \sum_{i=1}^{N} p_{\text{b},i}^t + P_{\text{td}}^t, \tag{19}$$

where P_{td}^t is the power trade decision that $P_{\text{td}}^t > 0$ means grid power is purchased and $P_{\text{td}}^t \le 0$ means power is sold to the grid. Assume that the prediction errors are independent, we have the following Propositions.

Proposition 1: The prediction errors of P_{pv}^t and P_{l}^t (η_{pv} and η_{l}) are both subject to Gaussian distribution as $\eta_{\text{pv}} \sim N(0, \sum_{i=1}^N \sigma_{\text{pv},i}^2)$ and $\eta_{\text{l}} \sim N(0, \sum_{i=1}^N \sigma_{\text{l},i}^2)$.

Proposition 2: In the joint-operation, the necessary conditions for optimal battery control are:

1. If $p_{\text{b},i} > 0$, then $p_{\text{b},\bar{i}} \ge 0$ for any $\bar{i} \in \{1, \ldots, N, \bar{i} \ne i\}$.
2. If $p_{\text{b},i} < 0$, then $p_{\text{b},\bar{i}} \le 0$ for any $\bar{i} \in \{1, \ldots, N, \bar{i} \ne i\}$.

Proposition 1 is obvious according to the elementary probability theory. For *Proposition 2*, if the charging power of battery i is $p_{\text{b},i} < 0$ and the discharging power of battery j is $p_{\text{b},j} > 0$, an equivalent better control policy is either charing battery i with power $p_{\text{b},i} + p_{\text{b},j}$ (when negative) or discharging battery j with power $p_{\text{b},i} + p_{\text{b},j}$ (when positive), especially considering the charing efficiency and battery lifespan in practical application. With *Proposition 2* the distributed batteries in the building microgrids can be viewed as an extended battery with capacity $\Phi_{\text{b}} = \sum_{i=1}^N Q_{\text{b},i}$. Define Γ as its SOC and the dynamics of the extended battery is

$$\Gamma^{t+1} = \Gamma^t - P_{\text{b}}^t \delta t / \Phi_{\text{b}}, \tag{20}$$

with P_{b}^t being the control decision and the system power balance becomes

$$P_{\text{l}}^t = P_{\text{pv}}^t + P_{\text{b}}^t + P_{\text{td}}^t. \tag{21}$$

For the system coordinator, an equivalent system description is: system state $\boldsymbol{s}_{\text{sys}} = (P_{\text{pv}}, P_{\text{l}}, \Gamma)$, event $\boldsymbol{e}_{\text{sys}} = (e_{\text{sys}}^1, e_{\text{sys}}^2)$ with $e_{\text{sys}}^1 \in \{\widetilde{P}_{\text{pv}} < \bar{P}_{\text{pv}}, \widetilde{P}_{\text{pv}} \ge \bar{P}_{\text{pv}}\}$ and $e_{\text{sys}}^2 \in \{\widetilde{P}_{\text{l}} < \bar{P}_{\text{l}}, \widetilde{P}_{\text{l}} \ge \bar{P}_{\text{l}}\}$ (tilde means observation and bar means prediction), action

$a_{\text{sys}} = P_{\text{b}}$ and state transition probability $p_{r,\text{sys}}$. The meaning of the above reformulation is that for a large-scale system the action space increases linearly instead of exponentially with the number of buildings. As for the system evolving process, equations (11)-(18) also apply.

Note first that *Proposition 1* is deduced from the previous Gaussian distribution assumption. For non-Gaussian distribution prediction errors, the EBO formulation and the following proposed SBPI method can also apply. The only difference lies in the sampling of $\widetilde{P}_{\text{pv}}$ and $\widetilde{P}_{\text{l}}$. When adopting Gaussian prediction errors, *Proposition 1* holds and samples of $\widetilde{P}_{\text{pv}}$ and $\widetilde{P}_{\text{l}}$ can be acquired according to the Gaussian distribution. While when adopting non-Gaussian prediction errors, Monte Carlo method needs to be applied in each building for solar power and building load samples, which then make up the samples of $\widetilde{P}_{\text{pv}}$ and $\widetilde{P}_{\text{l}}$. Note also that in the above formulation of battery operation, the charging and discharging efficiencies are neglected. In practical application, there exists power loss in the charging/discharging process and the charging/discharging efficiency varies with different statuses of battery [31], [32]. When the charging/discharging efficiency is considered, the joint-operation of decentralized batteries also follows the rules of *Proposition 2*. In decision stage t, if the real-time estimated charging/discharging efficiency for each battery is $\rho^t_{\text{c},i}/\rho^t_{\text{d},i}$ [31], equation (20) becomes

$$\Gamma^{t+1} = \Gamma^t - \sum_{i=1}^{N} p^t_{\text{b},i} * \rho^t_{\text{c},i}\left(\rho^t_{\text{d},i}\right)\delta t/\Phi_{\text{b}}. \tag{22}$$

Assume $P^t_{\text{b}} = \sum_{i=1}^{N} p^t_{\text{b},i} * \rho^t_{\text{c},i}(\rho^t_{\text{d},i})$, then the decentralized battery actions can still be aggregated to facilitate the following simulation-based policy improvement process. The only difference is that in the following balanced battery operation, $\rho^t_{\text{c},i}/\rho^t_{\text{d},i}$ needs to be considered in the update on the SOC of each battery.

B. On-Line SBPI for Joint-Operation

For a finite-stage EBO, finding the unstationary optimal control policy is intractable in practical application [30]. Facilitated by simple policies and heuristics, the following online SBPI method is developed for policy improvement.

In decision stage t, given the system belief state $\boldsymbol{\pi}^t_{\text{sys}}$ (probability distribution on $\boldsymbol{s}_{\text{sys}}$), denote the value function for adopting policy $\boldsymbol{d}_{\text{sys}} = (d^1_{\text{sys}}, \ldots, d^T_{\text{sys}})$ as $V^t(\boldsymbol{\pi}^t_{\text{sys}}, d^t_{\text{sys}})$. Then the performance of taking action a^t_{sys} is quantified using the following Q-factor

$$Q^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right) = C^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right) + E_{\boldsymbol{e}_{\text{sys}}}\left[V^{t+1}\left(\boldsymbol{\pi}^{t+1}_{\text{sys}}, d^{t+1}_{\text{sys,opt}}\left(\boldsymbol{e}^{t+1}_{\text{sys}}, \boldsymbol{\pi}^{t+1}_{\text{sys}}\right)\right)\right], \tag{23}$$

where $d^{t+1}_{\text{sys,opt}}(\boldsymbol{e}_{\text{sys}}, \boldsymbol{\pi}^{t+1})$ belongs to the optimal control policy $\boldsymbol{d}_{\text{sys,opt}}$ and the expectation $E_{\boldsymbol{e}_{\text{sys}}}$ is on all possibly observed events. Instead of using $\boldsymbol{d}_{\text{sys,opt}}$, we evaluate the Q-factor using the given base policy as

$$\widetilde{Q}^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right) = C^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right) + E_{\boldsymbol{e}_{\text{sys}}}\left[V^{t+1}\left(\boldsymbol{\pi}^{t+1}_{\text{sys}}, d^{t+1}_{\text{sys,b}}\left(\boldsymbol{e}^{t+1}_{\text{sys}}, \boldsymbol{\pi}^{t+1}_{\text{sys}}\right)\right)\right], \tag{24}$$

and improve this base policy by taking the most promising action as

$$\widetilde{a}^t_{\text{sys}} = \arg\min_{a^t_{\text{sys}} \in A_{\text{sys}}} \widetilde{Q}^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right), \tag{25}$$

It has been shown that (25) can assure the improvement of the given base policy [2], [24]. As the decision tree expands faster than exponentially with the decision stage moving on, an efficient way is using simulation [2], [30] in practice and the Q-factor can be further estimated with M event samples is as

$$\widetilde{Q}^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right) \approx C^t\left(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}\right) + \widetilde{E}_{\boldsymbol{e}_{\text{sys}}}\left[V^{t+1}\left(\boldsymbol{\pi}^{t+1}_{\text{sys}}, d^{t+1}_{\text{sys,b}}\left(\boldsymbol{e}^{t+1}_{\text{sys}}, \boldsymbol{\pi}^{t+1}_{\text{sys}}\right)\right)\right], \tag{26}$$

with

$$\widetilde{E}_{\boldsymbol{e}_{\text{sys}}}\left[V^{t+1}\left(\boldsymbol{\pi}^{t+1}_{\text{sys}}, d^{t+1}_{\text{sys,b}}\left(\boldsymbol{e}^{t+1}_{\text{sys}}, \boldsymbol{\pi}^{t+1}_{\text{sys}}\right)\right)\right] = \frac{1}{M}\sum_{m=1}^{M}\sum_{k=t+1}^{T} C^k\left(\boldsymbol{\pi}^k_{\text{sys}}, d^k_{\text{sys,b}}\left(\boldsymbol{e}^k_{\text{sys},m}, \boldsymbol{\pi}^k_{\text{sys}}\right)\right). \tag{27}$$

This on-line SBPI method is concluded in Algorithm 1.

Algorithm 1 On-Line SBPI for Joint-Operation

1. Given an base control policy $\boldsymbol{d}_{\text{sys},b}$;
2. In decision stage t, observing event $\boldsymbol{e}^t_{\text{sys}}$ and the event-based belief state is $\boldsymbol{\pi}^t_{\text{sys}}$;
3. Generate M samples of future observation based on solar power and building load estimation errors using Monte Carlo method;
4. For each feasible action $a^t_{\text{sys}} \in A$, evaluate its performance using the following simulation-based Q-factor as
$$\widetilde{Q}^t(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}) \approx C^t(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}}) + \frac{1}{M}\sum_{m=1}^{M}\sum_{k=t+1}^{T} C^k(\boldsymbol{\pi}^k_{\text{sys}}, d^k_{\text{sys,b}}(\boldsymbol{e}^k_{\text{sys},m}, \boldsymbol{\pi}^k_{\text{sys}}));$$
5. Take action $\widetilde{a}^t_{\text{sys}}$ which satisfies
$$\widetilde{a}^t_{\text{sys}} = \arg\min_{a^t_{\text{sys}} \in A_{\text{sys}}} \widetilde{Q}^t(\boldsymbol{\pi}^t_{\text{sys}}, a^t_{\text{sys}})$$
and improve the given base policy $\boldsymbol{d}_{\text{sys,b}}$.

C. Sub-Decision for Each Battery

After the improved action $\widetilde{a}^t_{\text{sys}}$ (P^*_{b}) is obtained, any battery control decision satisfying $\sum_{i=1}^{N} p^t_{\text{b},i} = \widetilde{a}^t_{\text{sys}}$ is feasible for the operation. As deeper discharge of the battery increases the capacity fade and causes higher shedding rates and rates of grain growth, the aging effects of cycling is modeled using an

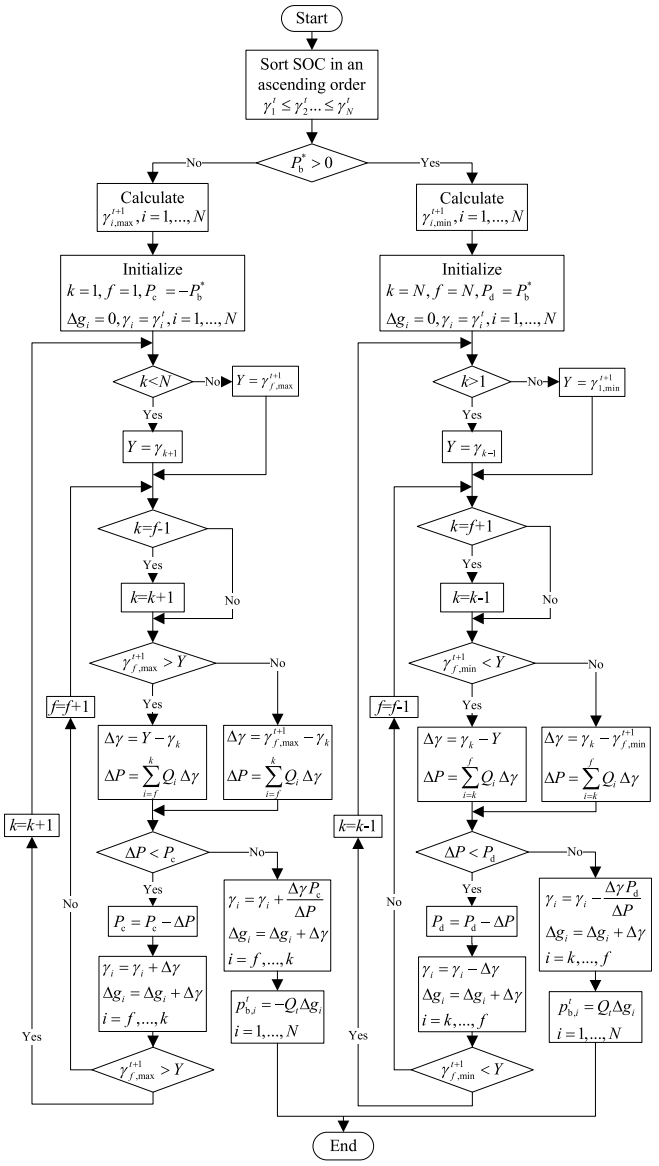


Fig. 3. Balanced operation of distributed batteries considering the expected life-cycles.

exponential function with the DOD suggested being smaller than 60% [33] and a semilogarithmic relationship between the life-cycles and the DOD is discussed in [34] with the most cost-effective DOD being between 25% and 33%. Considering the models and discussions in [33] and [34] and the experimental tests in [35], the operation range of SOC here is set between 0.5 and 1. The decision making process for the decentralized batteries are shown in Fig. 3. The battery SOCs are first sorted in an ascending order as $\gamma_1^t \le \gamma_2^t \le ... \le \gamma_N^t$. If $\widetilde{a}_{\text{sys}}^t > 0$, batteries are discharged from N to 1 with the discharging power constrained by both the maximum discharging power and the neighboring lower SOC. While if $\widetilde{a}_{\text{sys}}^t \le 0$, batteries are charged from 1 to N with the charging power constrained by both the maximum charging power and the neighboring higher SOC. This balanced battery operation not only avoids over-discharging but also provides more control alternatives for the next decision stage. When the charging/discharging

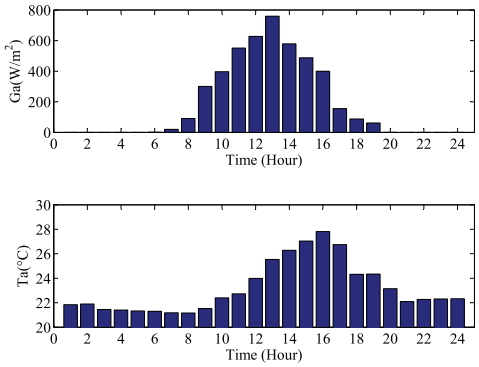


Fig. 4. Measured solar radiation and temperature.

TABLE I
TOU ELECTRICITY PRICE

Time periods	Price(RMB/kWh)
01:00 ∼ 06:00	0.3515
07:00 ∼ 10:00	0.8135
11:00 ∼ 18:00	0.4883
19:00 ∼ 24:00	0.8135

efficiency is considered, the decentralized battery operation decisions need to satisfy $\sum_{i=1}^{N} p_{\text{b},i}^t * \rho_{\text{c},i}^t(\rho_{\text{d},i}^t) = \widetilde{a}_{\text{sys}}^t$. The proposed balanced battery operation method can also apply with the update on the SOC of each battery considering the efficiency.

V. NUMERICAL RESULTS

In this section, two numerical examples based on the campus of Tsinghua University in Beijing, China are presented to demonstrate the performance of the proposed method.

A. 3-Building Microgrids

The first numerical example consists of a student dormitory, a dining hall and a teaching building. The weather information for a typical summer working day (Aug, 13, 2015) acquired from the weather station located inside the university campus is used for solar power prediction as in Fig. 4.

1) Parameter Settings: The day-ahead building electrical loads are estimated using EnergyPlus (version 7.2) with the comfort indoor temperature between 22°C and 26°C. The TOU price in Beijing area is shown in Table I and the price for selling electricity is 0.3 RMB/kWh, which is lower than the valley TOU price. Parameter settings for the 3-building microgrids are shown in Table II. The rated power of the BIPV system is determined by the available roof area of each building. The capacity of each battery is calculated according to that it can support the operation of local building for at least 5 hours at the peak demand. The initial SOCs for three batteries are set as 0.5, 0.7 and 0.8. The discretization interval of the action is 1 kW.

2) Policy Improvement With SBPI: A greedy policy $\boldsymbol{I}_g$ and two heuristic policies $\boldsymbol{I}_{h1}$ and $\boldsymbol{I}_{h2}$ are tested as the base policies and the test results in a day with 100 observation samples is shown in Fig. 5, Fig. 6, and Fig. 7 with $\boldsymbol{I}_g^*$, $\boldsymbol{I}_{h1}^*$ and $\boldsymbol{I}_{h2}^*$ being

TABLE II
PARAMETERS OF THE 3-BUILDING MICROGRIDS

No.	1	2	3
$P_{\mathrm{PV,r}}$(kW)	100	150	300
Q_{b}(kWH)	1000	1500	3000
$P_{\mathrm{b},d/c}$(kW)	$\pm$100	$\pm$150	$\pm$300

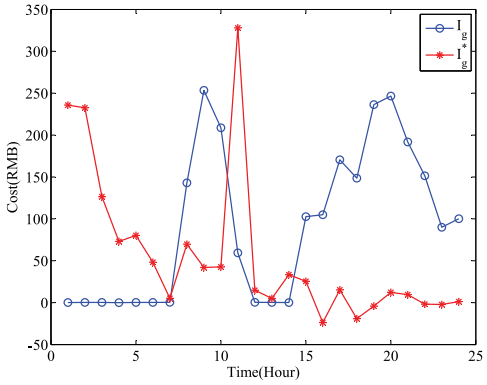


Fig. 5. Cost comparison between $\boldsymbol{I}_g$ and $\boldsymbol{I}_g^*$.

the improved policies. The three base policies $\boldsymbol{I}_g$, $\boldsymbol{I}_{h1}$ and $\boldsymbol{I}_{h2}$ are depicted as follows.

- $\boldsymbol{I}_g$: $\boldsymbol{I}_g$ is a greedy policy. When adopting $\boldsymbol{I}_g$, in each decision stage t, the action taken is to minimize the one-step expected cost as equation (28) without selling the discharged battery power to the utility grid. With this myopic policy, the system coordinator only cares about the current utility without considering the further influence of the action.

$$\widetilde{a}_{\mathrm{sys}}^t = \arg \min_{a_{\mathrm{sys}}^t \in A_{\mathrm{sys}}} C^t\left(\boldsymbol{\pi}_{\mathrm{sys}}^t, a_{\mathrm{sys}}^t\right) \tag{28}$$

- $\boldsymbol{I}_{h1}$: $\boldsymbol{I}_{h1}$ is a rule-based heuristic. When adopting $\boldsymbol{I}_{h1}$, in the periods of valley TOU price (from 01:00 to 06:00 and from 11:00 to 18:00), the action taken is to charge the battery with its maximum available value to store power from the utility grid as much as possible. While in the periods of peak TOU price (from 07:00 to 10:00 and from 19:00 to 24:00), the action taken is to discharge the battery that can minimize the one-step expected cost $C_t^{a_{\mathrm{sys}}^t}(\boldsymbol{\pi}_{\mathrm{sys}}^t)$ without selling the battery power to the utility grid. With this heuristic, the system coordinator can shift the low-rate grid power to high-rate hours through battery charging/discharging.
- $\boldsymbol{I}_{h2}$: $\boldsymbol{I}_{h2}$ is another rule-based heuristic. Similar to $\boldsymbol{I}_{h1}$, in the periods of valley TOU price, the action taken is the maximum available charging power and in the periods of peak TOU price, the action taken is the discharging power which minimizes the expected cost. The difference is that with $\boldsymbol{I}_{h2}$ the system coordinator can feed the battery power to the utility grid for profit, which in practice means that the decentralized battery can be scheduled as fast reserve by the utility grid.

For $\boldsymbol{I}_g$, this myopic policy discharges the battery in low-rate hours from 1:00 to 7:00 to minimize the one-step costs. While in the following high-rate hours from 8:00 to 11:00, power has to be purchased to satisfy the load. From 12:00 to 14:00, as solar power is abundant, there is no purchasing and solar power is charged to the battery. From 15:00 to 24:00, battery is discharged and high-rate power is purchased from 19:00 to 24:00. While in the improved policy $\boldsymbol{I}_g^*$, charging actions are taken in low-rate hours from 1:00 to 6:00 and from 11:00 to 14:00, which induces higher costs compared to $\boldsymbol{I}_g$. However, this shifts the low-rate power to the high-rate hours from 7:00 to 10:00 and from 19:00 to 24:00, which reduces the total electricity cost. Compared to $\boldsymbol{I}_g$, far-sighted actions are taken in the improved policy $\boldsymbol{I}_g^*$. The long-term effect of the action taken in each decision stage is evaluated through the value function based on simulation. Then low-rate power from the utility grid is shifted to high-rate hours for consumption and surplus solar power is stored for peak demand instead of fed to the grid for short-sighted profit.

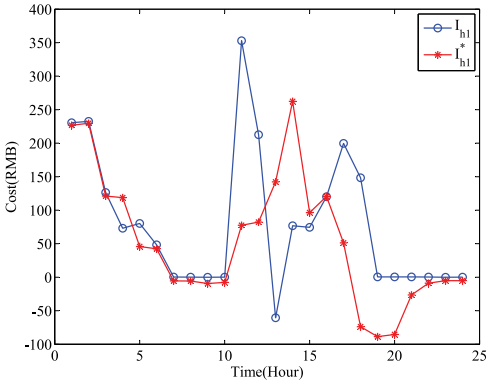


Fig. 6. Cost comparison between $\boldsymbol{I}_{h1}$ and $\boldsymbol{I}_{h1}^*$.

For $\boldsymbol{I}_{h1}$, this heuristic policy applies charing actions in low-rate hours from 1:00 to 6:00 and from 11:00 to 18:00, which shifts the low-rate power to high-rate hours from 7:00 to 10:00 and from 19:00 to 24:00. With the improved policy $\boldsymbol{I}_{h1}^*$, the operation is like $\boldsymbol{I}_{h1}$ from 1:00 to 10:00. In the following hours, $\boldsymbol{I}_{h1}^*$ charges the battery with higher power from 12:00 to 14:00. This improvement stores more solar power instead of feeding it to the grid. Although the costs of $\boldsymbol{I}_{h1}^*$ from 13:00 to 15:00 are higher than those of $\boldsymbol{I}_{h1}$, $\boldsymbol{I}_{h1}^*$ provides more cheap solar power in the following hours. From 17:00 to 22:00, more cheaper power (solar power and low-rate grid power) are discharged with $\boldsymbol{I}_{h1}^*$ to supply to buildings and less power is purchased from the utility grid. The improvement from policy $\boldsymbol{I}_{h1}$ to $\boldsymbol{I}_{h1}^*$ is mainly achieved by charging the battery with more solar power instead of discharging the battery for consumption and feeding the surplus solar power into the utility grid in low-rate hours from 12:00 to 14:00. Policy $\boldsymbol{I}_{h1}^*$ not only improves the efficiency of solar power utilization, but also provides more cheap power to high-rate hours in nighttime, which finally leads to the total cost reduction.

For $\boldsymbol{I}_{h2}$, this heuristic policy operates the building microgrids like $\boldsymbol{I}_{h1}$. The difference is that in high-rate hours the discharged battery power is allowed to fed into the grid in $\boldsymbol{I}_{h2}$. As shown in Fig. 7, from 7:00 to 10:00 and from 19:00 to 24:00, revenue is received for selling power. In the improved

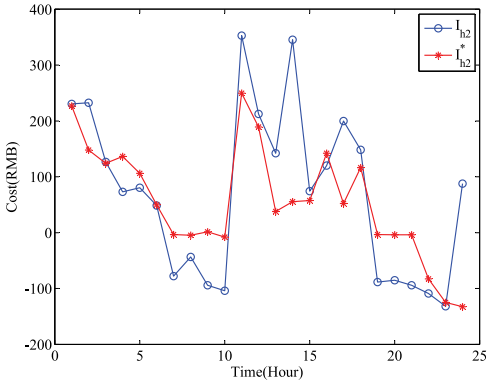


Fig. 7. Cost comparison between $\boldsymbol{I}_{h2}$ and $\boldsymbol{I}^*_{h2}$.

TABLE III
COST COMPARISON

	$\boldsymbol{I}_g/\boldsymbol{I}^*_g$	$\boldsymbol{I}_{h1}/\boldsymbol{I}^*_{h1}$	$\boldsymbol{I}_{h2}/\boldsymbol{I}^*_{h2}$
Cost(RMB)	2209.5/1349.9	1925.1/1295.1	1645.7/1328.3
Reduction	38.9%	32.7%	19.3%

policy $\boldsymbol{I}^*_{h2}$, from 7:00 to 10:00, battery power is discharged to satisfy the demand without fed into the grid, which provides more battery power and reduces the electricity cost in the following hours from 11:00 to 18:00. From 19:00 to 24:00, battery power is first discharged to satisfy the demand and then fed into the grid. The main reason for the improvement of $\boldsymbol{I}_{h2}$ is that in policy $\boldsymbol{I}^*_{h2}$ less power is discharged from the battery to be fed into the grid during discharging periods considering the high power demand afterwards. Then more solar power and low-rate grid power is stored and provided for the following peak demand and the total electricity cost is thus reduced.

The costs of three base policies and the improved policies are presented in Table III. Before the policy improvement, the cost of $\boldsymbol{I}_g$ is higher than that of $\boldsymbol{I}_{h1}$ and $\boldsymbol{I}_{h2}$, which means $\boldsymbol{I}_{h1}$ and $\boldsymbol{I}_{h2}$ are more economic than $\boldsymbol{I}_g$. While for the improved policies, their cost reductions are $38.9\% > 32.7\% > 19.3\%$. This implies that the extent of performance improvement is closely related to the given base policy, where a worse policy usually gains larger improvement. If the base policy is optimal, there will be no improvement. The advantage of the SBPI method is that it can employ the sub-optimal policy especially for those problems where obtaining the optimal policy is intractable.

3) Sensitivity Analysis: The test is performed 10 times for each policy. The mean costs and the variations of the improved policies are shown in Fig. 8. The average cost reductions for three policies are about -38%, -30%, and -20% and the variations are all less than 10% compared to their mean costs. We can see that the SBPI method provides a relatively stable improvement for the given base policy.

4) Effects of Moving Windows on Policy Improvement: In the MPC, the control strategy is optimized considering its effects in a future moving window, where a tradeoff is made between the computational time and the optimality. Given the base policy $\boldsymbol{I}_g$, the cost and total computational time of the SBPI method with different scales of moving window (2 hours

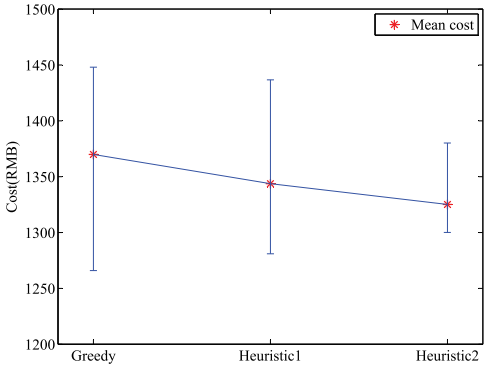


Fig. 8. Mean cost of the improved policies $\boldsymbol{I}^*_g$, $\boldsymbol{I}^*_{h1}$ and $\boldsymbol{I}^*_{h2}$.

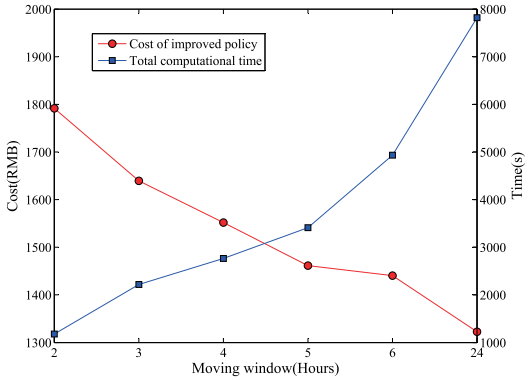


Fig. 9. Cost and computational time comparison for different moving windows.

to 6 hours) evaluating the future influence of the current action are investigated in Fig. 9. As shown in the figure, though the computational time increases with the moving window scale increasing, the cost of the improved policy decreases from 1791 RMB to 1322 RMB and the relative cost reductions are $10.2\% < 17.8\% < 22.2\% < 26.7\% < 27.8\% < 33.7\%$ compared to the cost using policy $\boldsymbol{I}_g$. We can see that a larger moving window leads to a greater policy improvement. This is mainly caused by the reason that the battery operation decision in the early morning may greatly influence the following decisions, even the system operation at late night. This temporally coupled decision process needs a larger moving window for performance evaluation of the current available actions as in the proposed SBPI method.

B. Large-Scale Building Microgrids

In order to demonstrate the potential of applying the proposed method to practical systems, a numerical example with 18-buildings located in Tsinghua university is presented. This 18-building microgrids consist of 5 research buildings (No.13∼17), 1 office building (No. 18), 4 teaching buildings (No. 3∼6), 2 dining halls (No.1∼2), 4 dormitories (No.7∼10), 1 cinema (No.11) and 1 concert hall (No.12). Parameters for this system are listed in Table IV. The discretization interval of the action for this large-scale system is 10 kW. Given $\boldsymbol{I}_g$ as the base policy, the computational time of SBPI for each stage in one day's operation is shown in Fig. 10. The maximum computational time is about 600 seconds with MATLAB (version

TABLE IV
PARAMETER SETTINGS OF THE 18-BUILDING MICROGRIDS

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$P_{\mathrm{PV,r}}$(100kW)	3	3	4	2.5	2.5	6	15	20	25	15	5	8	8	6	7	9	8	20
$P_{\mathrm{b,d/c}}$($\pm$100kW)	2	2	4	4	4.5	5	3.5	3.5	8	9	8.5	8	9.5	8.5	8.5	8	8	10
Q_{b}(100kWh)	20	20	40	40	45	50	35	35	80	90	85	80	95	85	85	80	80	100

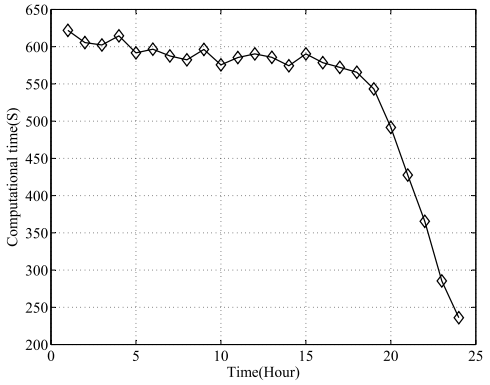


Fig. 10. Average one-step computational time for T stages with receding time horizon for the 18-building microgrids.

R2012a) on a PC with main frequency 2.9 Hz. This is applicable to this practical building microgrids system operated on an hourly basis. It is worth noting that the computational time for each decision stage is influenced by both the available action space and the future time horizon. Though the future decision horizon recedes with the decision time going forward and the computational time for value function evaluation process as in (27) is reduced, the evaluation of the large number of actions still accounts for the main part of the computational time before 19:00. While from 19:00 to 24:00 the computational time decreases for the dwindled action space and the fewer future decision horizons. There also exists the moving window method [36], [37], which looks ahead several steps instead of to the end to facilitate the computation of the value function. If the time-scale of the window is constant, the computational times for all decision stages are almost the same. However, in the proposed SBPI method, the receding horizon is used instead of the moving window for the reason that the moving window cannot ensure the improvement of the base policy for the whole operation horizon.

Note also that for a larger microgrid system with more buildings, the proposed policy improvement method can also apply and in order to further alleviate the computational burden to keep the one-step computational time in an appropriate scope for on-line application, the following two methods are suggested in practice. First, the action space can be discretized with coarse granularity. For higher power generation and consumption, a larger discretization interval is reasonable and can reduce the action space. Second, buildings can be aggregated according to their load profiles as the demand aggregation method in [21]. Buildings in the microgrid can be divided into several groups according to their different functions. Then buildings with similar load profiles can be aggregated to be treated as one building in the decision process. The operation decision can be first made for this aggregated building and then for each building. This aggregation method can greatly facilitate the decision process.

VI. CONCLUSION

In this paper the joint-operation optimization of building microgrids with distributed solar power and storage battery are investigated. The operation problem is first formulated into an EBO model considering the uncertainties of solar power, building load and communication. Then after aggregated battery operation, an on-line SBPI method is developed to improve the given base control policy and a balanced battery control method considering the expected life cycles is also proposed. At last, numerical results show that the electricity cost of the building microgrids is reduced with the improved policy. It is also shown through a larger-scale example that the proposed method can apply to practical systems.

REFERENCES

[1] X. Guan, Z. Xu, and Q.-S. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 243–252, Dec. 2010.

[2] Q.-S. Jia, J.-X. Shen, Z.-B. Xu, and X.-H. Guan, "Simulation-based policy improvement for energy management in commercial office buildings," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2211–2223, Dec. 2012.

[3] H.-T. Wang, Q.-S. Jia, C. Song, R. Yuan, and X. Guan, "Building occupant level estimation based on heterogeneous information fusion," *Inf. Sci.*, vol. 272, pp. 145–157, Jul. 2014.

[4] H. Wang, Q. Zhao, X. Guan, Q.-S. Jia, and L. Li, "Reconfiguring networked infrastructures by adding wireless communication capabilities to selected nodes," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4518–4528, Sep. 2013.

[5] M. Stadler *et al.*, "Electric storage in California's commercial buildings," *Appl. Energy*, vol. 104, pp. 711–722, Apr. 2013.

[6] A. Khodaei and M. Shahidehpour, "Optimal operation of a community-based microgrid," in *Proc. IEEE PES Innov. Smart Grid Technol.*, Medellín, Colombia, 2011, pp. 1–3.

[7] Y. Xiang, J. Liu, and Y. Liu, "Robust energy management of microgrid with uncertain renewable generation and load," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 1034–1043, Mar. 2016.

[8] Y. Wang, B. Wang, C.-C. Chu, H. Pota, and R. Gadh, "Energy management for a commercial building microgrid with stationary and mobile battery storage," *Energy Build.*, vol. 116, pp. 141–150, Mar. 2016.

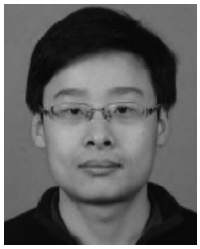
[9] C. Chen, S. Duan, T. Cai, B. Liu, and G. Hu, "Smart energy management system for optimal microgrid economic operation," *IET Renew. Power Gener.*, vol. 5, no. 3, pp. 258–267, May 2011.

[10] Z. Wang, L. Wang, A. I. Dounis, and R. Yang, "Integration of plug-in hybrid electric vehicles into energy and comfort management for smart building," *Energy Build.*, vol. 47, pp. 260–266, Apr. 2012.

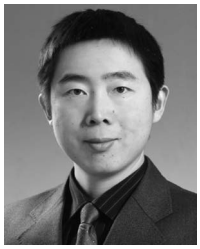
[11] W. Shi, N. Li, C.-C. Chu, and R. Gadh, "Real-time energy management in microgrids," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 228–238, Jan. 2017.

[12] S. Salinas, M. Li, P. Li, and Y. Fu, "Dynamic energy management for the smart grid with distributed energy resources," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2139–2151, Dec. 2013.

[13] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures Commun. Netw.*, vol. 3, no. 1, pp. 1–211, 2010.
[14] S. Sun, M. Dong, and B. Liang, "Joint supply, demand, and energy storage management towards microgrid cost minimization," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Venice, Italy, 2014, pp. 109–114.
[15] Y. Huang, S. Mao, and R. M. Nelms, "Adaptive electricity scheduling in microgrids," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 270–281, Jan. 2014.
[16] Y. Zong, D. Kullmann, A. Thavlov, O. Gehrke, and H. W. Bindner, "Application of model predictive control for active load management in a distributed power system with high wind penetration," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 1055–1062, Jun. 2012.
[17] B. Mayer, M. Killian, and M. Kozek, "Management of hybrid energy supply systems in buildings using mixed-integer model predictive control," *Energy Convers. Manag.*, vol. 98, pp. 470–483, Jul. 2015.
[18] L. H. Lee, E. P. Chew, P. I. Frazier, Q.-S. Jia, and C.-H. Chen, "Advances in simulation optimization and its applications," *IIE Trans.*, vol. 45, no. 7, pp. 683–684, 2013.
[19] Q.-S. Jia, "Efficient computing budget allocation for simulation-based policy improvement," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 2, pp. 342–352, Apr. 2012.
[20] D. P. Bertsekas, J. N. Tsitsiklis, and C. Wu, "Rollout algorithms for combinatorial optimization," *J. Heuristics*, vol. 3, no. 3, pp. 245–262, 1997.
[21] Q. Huang, Q.-S. Jia, Z. Qiu, X. Guan, and G. Deconinck, "Matching EV charging load with uncertain wind: A simulation-based policy improvement approach," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1425–1433, May 2015.
[22] Q. Huang, Q.-S. Jia, and X. Guan, "Robust scheduling of EV charging load with uncertain wind power integration," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1043–1054, Mar. 2018.
[23] Q. Huang, Q.-S. Jia, and X. Guan, "A multi-timescale and bilevel coordination approach for matching uncertain wind supply with EV charging demand," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 694–704, Apr. 2017.
[24] H. S. Chang, J. Hu, M. C. Fu, and S. I. Marcus, *Simulation-Based Algorithms for Markov Decision Processes*. London, U.K.: Springer-Verlag, 2013.
[25] M. G. Villalva, J. R. Gazoli, and E. R. Filho, "Comprehensive approach to modeling and simulation of photovoltaic arrays," *IEEE Trans. Power Electron.*, vol. 24, no. 5, pp. 1198–1208, May 2009.
[26] D. Zhang, N. Shah, and L. G. Papageorgiou, "Efficient energy consumption and operation management in a smart building with microgrid," *Energy Convers. Manag.*, vol. 74, pp. 209–222, Oct. 2013.
[27] B. Burger and D. Kranzer, "Extreme high efficiency PV-power converters," in *Proc. 13th Eur. Conf. Power Electron. Appl. (EPE)*, Barcelona, Spain, 2009, pp. 1–13.
[28] E. Walraven and M. T. Spaan, "Planning under uncertainty for aggregated electric vehicle charging using Markov decision processes," in *Proc. Int. Workshop Artif. Intell. Smart Grids Smart Build.*, Phoenix, Arizona, USA, Feb. 2016, pp. 904–912.
[29] X.-R. Cao, *Stochastic Learning and Optimization: A Sensitivity-Based Approach*. New York, NY, USA: Springer, 2007.
[30] Q.-S. Jia, "On solving optimal policies for finite-stage event-based optimization," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2195–2200, Sep. 2011.
[31] A. Jossen, J. Garche, and D. U. Sauer, "Operation conditions of batteries in PV applications," *Solar Energy*, vol. 76, no. 6, pp. 759–769, 2004.
[32] H. Yang, Z. Wei, and C. Lou, "Optimal design and techno-economic analysis of a hybrid solar–wind power generation system," *Appl. Energy*, vol. 86, no. 2, pp. 163–169, 2009.
[33] A. Millner, "Modeling lithium ion battery degradation in electric vehicles," in *Proc. IEEE Conf. Innov. Technol. Efficient Rel. Electricity Supply (CITRES)*, Waltham, MA, USA, 2010, pp. 349–356.
[34] L. H. Thaller, "Expected cycle life vs. depth of discharge relationships of well-behaved single cells and cell strings," *J. Electrochem. Soc.*, vol. 130, no. 5, pp. 986–990, 1983.
[35] T. Guena and P. Leblanc, "How depth of discharge affects the cycle life of lithium-metal-polymer batteries," in *Proc. 28th Annu. Int. Telecommun. Energy Conf. (INTELEC)*, Providence, RI, USA, 2006, pp. 1–8.
[36] Y. Ma, F. Borrelli, B. Hencey, A. Packard, and S. Bortoff, "Model predictive control of thermal energy storage in building cooling systems," in *Proc. 48th IEEE Conf. Decis. Control Held Jointly 28th Chin. Control Conf. (CDC/CCC)*, Shanghai, China, 2009, pp. 392–397.
[37] M. Beaudin, H. Zareipour, and A. Schellenberg, "Residential energy management using a moving window algorithm," in *Proc. 3rd IEEE PES Innov. Smart Grid Technol. Europe (ISGT Europe)*, Berlin, Germany, 2012, pp. 1–8.



Yuanming Zhang received the B.S. degree in detection, guidance, and control techniques from the Harbin Institute of Technology and the M.S. degree in control theory and engineering from the Automation Research and Design Institute of Metallurgical Industry in 2005 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Tsinghua University, Beijing, China. His research interests include power system optimization, optimization theory, and solar power system.



Qing-Shan Jia (S'02–M'06–SM'11) received the B.E. degree in automation and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2002 and 2006, respectively. He is an Associate Professor with the Department of Automation, Center for Intelligent and Networked Systems, TNLIST, Tsinghua University, Beijing. He was a Visiting Scholar with Harvard University in 2006, and a Visiting Scholar with the Hong Kong University of Science and Technology in 2010. His research interests include theories and applications of discrete event dynamic systems and simulation-based performance evaluation and optimization of complex systems.