This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3176306, IEEE Internet of Things Journal

An Iterative Optimization and Learning-based IoT System for Energy Management of Connected Buildings

Yixiang Gao, *Student Member, IEEE,* Shuhui Li, *Senior Member, IEEE,* Yang Xiao, *Fellow, IEEE,* Weizhen Dong, Michael Fairbank, and Bing Lu

***Abstract*—Buildings account for nearly 40% of primary energy and 36% of greenhouse emissions, which is one of the main factors driving climate change. Reducing energy consumption in buildings toward zero-energy buildings is a vital pillar to ensure that future climate and energy targets are reached. However, due to the high uncertainty of building loads and customer comfort demands, and extremely nonlinear building thermal characteristics, developing an effective zero-energy building energy management (BEM) technology is facing great challenges. This paper proposes a novel learning-based and iterative IoT system to address these challenges to achieve the zero-energy objective in BEM of connected buildings. Firstly, all buildings in the IoT-based BEM system share their operation data with an aggregator. Secondly, the aggregator uses these historical data to train a deep reinforcement learning model based on the Deep Deterministic Policy Gradient method. The learning model generates pre-cooling or pre-heating control actions to achieve zero-energy BEM for building heating ventilation and air conditioning (HVAC) systems. Thirdly, for solving the coupling problem between HVAC systems and building internal heat gain loads, an iterative optimization algorithm is developed to integrate physics-based and learning-based models to minimize the deviation between the on-site solar photovoltaic generated energy and the actual building energy consumption by properly scheduling building loads, electric vehicle charging cycles and the energy-storage system. Lastly, the optimal load operation scheduling is generated by considering customers' comfort requirements. All connected buildings then operate their loads based on the load operation schedule issued by the aggregator. The proposed learning-based and iterative IoT system is validated via simulation with real-world building data from the Pecan Street project.**

***Index Terms*—Internet of Things (IoT), building energy management (BEM), deep deterministic policy gradient (DDPG), deep reinforcement learning (DRL), zero energy building.**

NOMENCLATURE

Abbreviations

BEM	Building energy management
HVAC	Heating ventilation and air conditioning
DDPG	Deep deterministic policy gradient
DRL	Deep reinforcement learning
DR	Demand response
RES	Renewable energy source
EV	Electric vehicle
PV	Photovoltaic
RC	Resistance capacitance
ESS	Energy storage system
TESS	Thermal energy storage system
CL	Controllable load
NN	Neural network
SOC	State of charge.

Notation

t,T	Index and set for time steps
n,N	Index and set for buildings
m,NEV	Index and set for EVs
k,NL	Index and set for lighting systems
J,NCL	Index and set for CLs.

DDPG-based HVAC System Parameters and Variables

a_t	Actions at step *t*
s_t	Input state at step *t*
r_t	Reward at step *t*
w_A,w_C	Weights of main Actor NN and main Critic NN
$w_A^{'}$,$w_C^{'}$	Weights of target Actor NN, and target Critic NN
τ	Smoothing factor for updating target NN
γ	Discount factor
$\mathcal{B}$	Replay buffer
$\mathcal{N}_t$	Gaussian noise at step *t*
B	Minimum batch size for gradient descent updates
U	Minimum number of the tuples before starting to do gradient descent updates
D	Minimum number of BEM control interactions that should elapse between gradient descent updates
$P_{t,n}^{\mathrm{HVAC}}$	HVAC energy consumption at step *t* and building *n*
$P_{t,n}^{\mathrm{CD}}$	Building cooling demand at step *t* and building *n*
$P_{t,n}^{\mathrm{CS}}$	Cooling storage energy at step *t* and building *n*
$S_{t,n}^{\mathrm{CS}}$	Total cooling storage energy at step *t* and building *n*
T_t^{od}	Outdoor temperature at step *t*
ϕ_t^{od}	Related humidity at step *t*
m	Month index
h	Hour index
$P_{t,n}^{\mathrm{non\_HVAC}}$	Non-HVAC loads at step *t* and building *n*
$P_{t,n}^{\mathrm{PV}}$	PV generation energy at step *t* and building *n*

Physics-based System Parameters and Variables

$S_{t,n}^{\mathrm{ES}}$	Total ESS energy at step *t* and building *n*
$P_{t,n}^{\mathrm{ES\_C}}$	ESS charging powers at step *t* and building *n*

Y. Gao, S. Li, W. Dong, and L. Bing are with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL, 35401 (USA e-mail: ygao43@crimson.ua.edu; sli@eng.ua.edu; wdong5@crimson.ua.edu; blu2@crimson.ua.edu).

Y. Xiao is with the Department of Computer Science, University of Alabama, Tuscaloosa, AL, 35401 (USA e-mail: yangxiao@ieee.org).

M. Fairbank is with University of Essex, Colchester, CO4 3SQ, UK (e-mail: m.fairbank@essex.ac.uk).

$P^{\mathrm{ES\_D}}_{t,n}$ ESS discharging powers at step t and building n
$\eta^{\mathrm{ES\_C}}$ ESS charging efficiencies
$\eta^{\mathrm{ES\_D}}$ ESS discharging efficiencies
$I^{\mathrm{ES}}_{t,n}$ ESS operation mode at step t and building n
$S^{\mathrm{EV}}_{t,n,m}$ Total energy of EV's battery for mth EV at step t and building n
$P^{\mathrm{EV\_C}}_{t,n,m}$ EV charging power for mth EV at step t and building n
η^{EV} EV charging efficiency
$E^{\mathrm{Trip}}_{T_i,n,m}$ Minimum SOC demand for the next trip of the mth
$I^{\mathrm{EV}}_{t,n,m}$ EV charging mode for mth EV at step t and building n
$P^{\mathrm{L}}_{t,n,k}$ Energy for the kth lighting system at step t and building n
$A^{\mathrm{L}}_{t,n,m}$ Corresponding lighting area for the kth lighting system at step t and building n
$E^{\mathrm{L}}_{t,n,m}$ Illuminance for the kth lighting system at step t and building n
η^{L} Luminous efficiency
c Utilization factor
m^{L} Maintenance factor
$P^{\mathrm{CL}}_{t,n,j}$ Energy consumption for the jth CL at step t and building n
$E^{\mathrm{CL\_total}}_{n,j}$ Total energy assignments of the jth CL at building n
$\theta_{t,n,k}$ Workload parameter for the jth CL at step t and building n

Optimization Parameters and Variables

α Weighting factors between energy consumption from the electric utility and PV energy generation
β Weighting factor associated with the ESS depreciation cost
$P^{\mathrm{UCL}}_{t,n}$ Energy consumption for the uncontrollable loads at step t and building n
c^{L} Visual penalty price
E Best illuminance
c^{CL} CL overload penalty price
c^{EV} EV charging incompleteness penalty price
c^{T} Thermal penalty cost
$c^{\mathrm{ES\_life}}$ ESS battery depreciation cost
$n^{\mathrm{ES\_cycles}}$ Equivalent ESS charging/discharge cycles

I. Introduction

A. Motivation of the proposed research

SMART building energy management (BEM) is considered as an important Internet of Things (IoT) application which not only offers improvements for the quality of life of the inhabitants, but also greatly improves the efficiency and security of electric power systems [1]. In recent years, on-site rooftop photovoltaic (PV) systems have had the most installations over residential and commercial buildings [2]. However, high PV penetration leads to new technical issues for power system planning and operation too [3]. Fortunately, IoT-based technologies can integrate building's automation, control and communication systems into the building energy management, which allows end-users to reschedule their load operation processes to flatten the load profiles based on the electricity price [4]. For example, an IoT-based data-driven precooling method is proposed to determine HVAC (heating, ventilation, and air conditioning) operation strategies to reduce peak energy consumption and electricity costs [5]. [6] proposed a scheduling method for building energy supplies, which optimizes the overall cost of electricity for the building operation over a time horizon while satisfying the energy balance and complex constraints of individual energy supply equipment and devices within the building. A pool-based demand response (DR) methodology is presented in [7], which considers the variability of RES and supply-demand balance as options for scheduling customers' loads in the day-ahead electricity market. A decentralized EV-Based charging optimization model is proposed in [8] to coordinate EV charging with the RES power of buildings, which can improve the power supply reliability and potentially reduce the impacts of EV charging demand on the power grid.

However, the participation of large-scale, price-based BEM customers into the power system operation would cause the shift of peak load hours instead of solving the problems. In addition, the building control policies set by traditional DR customers have low efficiency and are difficult to achieve real-time BEM control. With the development of large-scale advanced metering infrastructure, it is easy to obtain data regarding building loads, PV generation, batteries, EVs charging cycles and human comfort [9]. Particularly, the revolution of IoT technologies has established a foundation and offered disruptive opportunities for the R&D of the building connectivity of a large swath of devices [10]. As shown in [11], IoT-based autonomous BEM technologies based upon deep reinforcement learning (DRL) are promising to address some of the above challenges. Driven by these challenges and opportunities, this paper proposes a learning-based and iterative optimization IoT system for day-ahead energy management of connected buildings, which can assist end-user customers to automatically schedule the operation of building appliances in a more efficient and comfortable way to achieve nearly-zero energy buildings.

B. Background and related works

Many previous research studies have focused on the energy storage and optimization for nearly-zero energy buildings. These studies try to develop an appropriate and accurate building simulation environment in order to achieve cost-efficient building energy management strategy without losing human comfort, and can be divided into three categories: 1) physics-based methods [12-17], 2) data-driven methods [18-21], and 3) model-free methods [22-26].

Physics-based methods usually develop mathematical models of building electrical equipment by considering their physics characteristics [12]. Furthermore, resistance and capacitance (RC) heat transfer models have been used to simulate building thermal system behaviors [13]. An RC building thermal dynamic model has been used to calculate indoor temperature in [14]. An IoT-based smart energy management system (SEMS) is presented to achieve the economic operation of combined cooling, heating, and power (CCHP) for a commercial building system in [15], in which physics-based methods are used to model an internal combustion engine, three-way valves, electric chillers, and batteries of IoT-based commercial buildings to improve the automation energy management efficiency and user comforts. In [16], a physics-based metaheuristic algorithm is proposed for IoT enabled smart homes to minimize building energy cost and peak-to-average ratio of building energy consumption. Hussain et al. presented a heuristic-based algorithm in [17] by considering DR, photovoltaic availability, and the state of charge (SOC) and charge/discharge rates of a storage battery for IoT-enabled smart homes to minimize the building electricity cost, in which the development of the algorithm used a physics model of the building. However, the physics-based models shown in all these studies are difficult to reflect highly complicated and nonlinear building thermal characteristics. In addition, considering uncertain customer comfort requirements in a physics-based model is still a challenge.

To reflect more accurate building thermal characteristics, data-driven building models were proposed in [18] to improve the accuracy for building room temperature regulation. For predicting electric energy consumption in an IoT-driven building context, an Elman recurrent neural network (RNN) model and an exponential model are developed in [19]. [20] proposed a fine-grained dynamic neural network approach to derive an accurate thermal comfort model for emerging IoT enabled smart building management and operation to improve building energy efficiency and occupant thermal comforts. Hu et al. proposed in [21] an intelligent thermal comfort neural network (ITCNN) model to evaluate the occupants' thermal comfort by integrating machine learning techniques and IoT-based pervasive sensing technologies. The main limitations of these data-driven models include 1) if the models are generated based on simulated building data, those models would be hard to reflect the complexity of real-world buildings, or 2) if the models are generated based on actually measured building data, adequate input and output relationship for measured building data is critical to develop correct and high-performance data-driven model in addition to a large amount of data that is typically needed to gain the models, making the data impractical to obtain in many situations.

To solve the model-based shortages, learning-based, model-free methods were proposed in many studies that could produce building energy management policy directly without requiring any system models [22]. A Q-value based reinforcement learning (RL) strategy considering end-users' priority is proposed in [23] for an optimal IoT-Enabled home appliances scheduling (HAS). [24] formulated an energy cost minimization problem as a Markov decision process and solved the problem based on a Deep Deterministic Policy Gradients method. A reinforcement learning mechanism was proposed to determine the building energy management policy to smooth the net building energy consumption curve in [25-26]. Apparently, these learning-based, model-free methods are still relatively new and need more research. Particularly, large-scale connected buildings with PV generations, loads, ESSs and TESSs will significantly increase the number of control states and actions, which would significantly decrease the deep reinforcement learning speed. In addition, the above learning-based, model-free methods focus on achieving global optimization policies by training data over a period of one year or several months. This would not be suitable for real-time building operation considering the day-ahead market nature of electric power systems.

C. Key contributions of this paper

In our previous work, we mainly focused on the economic energy management of a single appliance [27] or a single home [28]. However, for energy management that can benefit a larger-scale community, nearly zero-energy management technology is an emerging and important topic area pursued in the energy research field [29]. As such, the objective of this paper is to develop a BEM IoT system that can achieve the nearly zero energy goal for connected buildings via proper management of building energy consumption, generation, and storage. All connected buildings in service share their data with an aggregator who generates the next-day control actions for the energy management of the connected buildings. To overcome the challenges of complex and nonlinear properties of HVAC building loads on one aspect and a large number of other non-HVAC building loads on the other, this paper proposes a novel hybrid physics-based and learning-based IoT system for day-ahead BEM of connected buildings by combining the characteristics of physics-based methods, and data-driven and model-free methods together. In summary, the novel contributions of the paper are as follows:

1) an IoT system to achieve the zero-energy planning and management goal of connected buildings considering both electrical and thermal energy storage systems;
2) a deep reinforcement learning method to generate HVAC system pre-cooling and heating actions while overcoming the challenges of nonlinear and complex natures in HVAC system modeling;
3) a physics-based method to model energy consumption behaviors of all non-HVAC controllable loads based on real-world building data and human comfort requirements, which significantly reduces the size of the state space associated with the learning model;
4) an iterative optimization strategy by integrating physics-based non-HVAC models with data-driven-based HVAC models to achieve optimal scheduling of HVAC loads and non-HVAC loads in the building energy management of connected buildings.

D. Organization of the paper

The rest of the paper is structured as follows. Section II gives an overall view of the proposed BEM IoT system. Section III presents a learning-based data-driven model for building HAVC loads. Section IV gives physics-based model for building non-HAVC loads. An iterative optimization strategy is developed in Section V to generate optimal BEM control scheduling for combined HVAC and non-HAVC building loads. Section VI shows case studies and results. Finally, the paper concludes with summary remarks in Section VII.

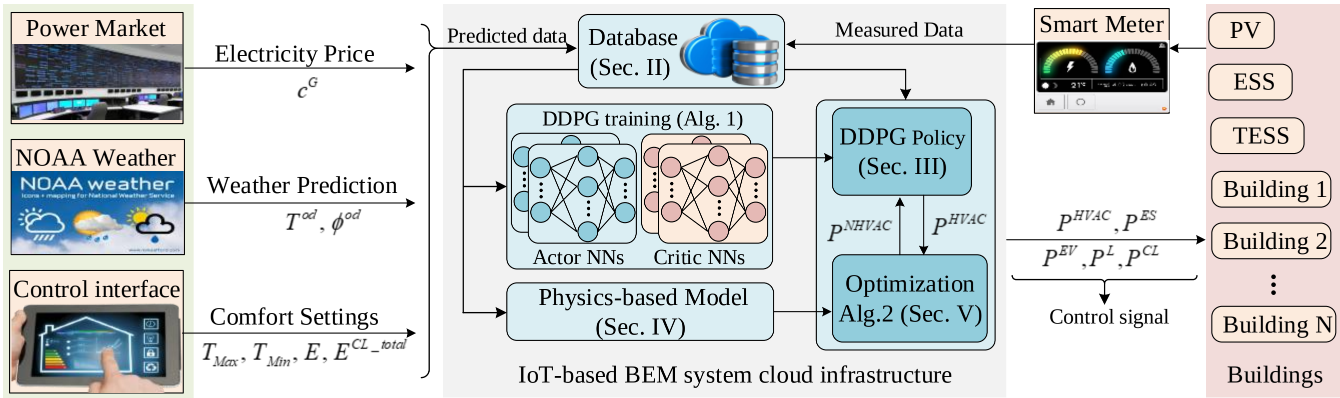

Fig. 1 Overall Structure of the proposed IoT BEM system

II. Overall Structure of the Proposed IoT BEM System

The proposed IoT BEM system, as shown in Fig. 1, is a hybrid system that integrates learning-based HVAC system models and physics-based non-HVAC system models for self-scheduling of connected buildings' energy sources, loads, ESSs, and TESS one day ahead to achieve the nearly zero building energy goal. Basically, at the end of each operating day *d*, the proposed IoT system collects all connected buildings' operation raw data, including electricity price data, and actual weather data, electricity usage data of building appliances. The collected data is used to obtain a learning-based model for HVAC systems and physics-based models for non-HVAC systems that are then used with an iterative optimization routine to generate next-day operation and scheduling signals for both HVAC and non-HVAC systems based on forecast data of next-day electricity price, weather, etc. Then, on the next day *d+1*, the energy management of the connected buildings is operated based on the operation and scheduling signals generated one day ahead. This proposed BEM IoT system mainly consists of four fundamental parts:

1) Database: The data collected and stored in the database can be divided into two categories: i) Predicted data and ii) Measured data. The predicted data involves day-ahead electricity prices from electric utility companies, weather predictions from national weather services, and customers' comfort settings of each building based on historical information. The measured data includes buildings' real-time operation data obtained from meters or sensors, involving PV generation power, ESS charging/discharging power, non-HVAC loads (e.g. dryer, EV…), HVAC thermal storage charging/discharging data, etc. The collected data should be cleaned, anonymized, and curated, and is saved in the Database for online or offline analysis. The measured data for all connected buildings is sent to an IoT aggregator.

2) DDPG-based BEM model for the HVAC systems of connected buildings: The DDPG is an actor-critic reinforcement learning model. The DDPG-based BEM model computes control actions $\mathcal{A}$ based on the states $\mathcal{S}$ of building HVAC systems to achieve the optimization goal of nearly zero energy management of the building HVAC systems. In our model, at each time step *t*, the state vector s_t for *N* connected buildings contains 2 weather-related states (temperature and humidity), 2 states indicating the month and hour information, *N* PV generation states, *N* non-HVAC load states, and *N* accumulated energy storage states related to building TESS systems; the control action vector a_t contains energy storage regulating actions of building TESS for all connected buildings. In the DDPG model, at each time step *t*, an Actor NN determines the control action vector a_t of building HVACs as shown in (1) and is updated to ensure the most efficient zero energy management of HVAC systems of connected building, and a Critic NN tries to approximate the optimal value function that is used to evaluate the performance of a control action and is updated to ensure that the optimal value function effectively captures the actual performance for zero-energy management.

$$a_t = \mathrm{A}_{\mathrm{NN}}\left(s_t, w_A\right) \tag{1}$$

$$C_{NN}\left(a_t, s_t, w_C\right) = r_t + \gamma \cdot C_{NN}\left(a_t, s_t, w_C\right) \tag{2}$$

where the HVAC energy consumptions of the connected buildings, $P_{t,1}^{HVAC}, P_{t,2}^{HVAC}, \cdots, P_{t,N}^{HVAC}$, are contained in the control action vector a_t, $\mathrm{A}_{\mathrm{NN}}(\bullet)$ represents the actor NN and w_A is the weight vector of the actor NN, $C_{NN}(\bullet)$ stands for the critic NN and w_C is the weight vector of the critic NN, r_t is the reward to measure the performance for nearly zero energy management of the building HVAC systems. Details about how the DDPG-based BEM model is developed for the HVAC systems of connected buildings, how to train the DDPG model, and how to use it to generate day-ahead scheduling policy of building HVAC systems are presented in Section III.

3) Physics-based models for building non-HVAC loads and systems: Buildings also contain many non-HVAC loads and systems. It is important that those non-HVAC loads and systems are considered in building the optimization problem to achieve the nearly-zero energy management. However, we do not include those non-HAC loads and systems into the DDPG-based actor-critic model, which could cause the DDPG model too complicated. In this paper, the energy consumptions or generations of the non-HVAC systems are conveniently modeled based on their physical properties and manufacturers' data, which would significantly reduce the complexity of the DDPG-based learning model in developing the proposed BEM IoT system. These devices include indoor lighting systems,

ESSs, EVs, washing machines, etc. In addition, the predicted customers' comfort constraints from the database are included in developing the physics-based models for non-HVAC loads. The physics-based model of a non-HVAC load or system, at each time step *t*, typically includes a mathematical model to describe the energy consumption or storage of the non-HVAC load or system and a set of constraint equations that specify the ranges allowed for the energy consumption or storage of the non-HVAC load or system. An example of the physics-based model for a building controllable load (CL) is shown below,

$$P_{t,n,j}^{CL} = \theta_{t,n,j} \cdot E_{n,j}^{CL\_total} \tag{3}$$

$$P_{Min}^{CL} \leq P_{t,n,j}^{CL} \leq P_{Max}^{CL} \tag{4}$$

Details about physics-based models of other building non-HVAC loads and systems are presented in Section IV.

4) Optimization routine: The optimization routine developed in this paper is an iterative optimization strategy. First, the energy consumption of HVAC systems obtained in 2), and the physics-based models for building non-HVAC loads and systems obtained in 3), together with the prediction data of next-day weather and electricity price, and customer comfort requirements from the database, are put together to formulate an overall optimization problem as shown in (5). Solving the optimization problem will give energy management scheduling for non-HVAC loads and systems. However, with the new energy consumption solution of the non-HVAC loads, the HVAC energy consumption obtained in 2) through the DDPG-based BEM model needs to be recalculated considering the thermal coupling between the HVAC and non-HVAC systems. With the new HVAC energy consumption obtained, the optimization problem of (5) is solved again. This process continues until a convergence is arrived.

Minimize:

$$\begin{aligned} C = {} & (1-\alpha)\cdot \mathrm{DeviationCost} + \alpha \cdot \mathrm{Electricity\ cost} + \beta \cdot \mathrm{ESS} \\ & \mathrm{DepreciationCost} + \mathrm{CL\ Overload\ Cost} \\ & + \mathrm{VisualDiscomfortCost} + \mathrm{Thermal\ Discomfort\ Cost} \\ & + \mathrm{EV\ ChargingIncompleteness\ Cost} \end{aligned} \tag{5}$$

Subject to:

$$P_t^G + \sum_{n=1}^{N} P_{t,n}^{PV} = \sum_{n}^{N} \left(P_{t,n}^{non-HVAC} + P_{t,n}^{HVAC}\right) \tag{6}$$

and, (4) and other constraints of non-HVAC systems.

In (5), 1) Deviation Cost term is used to address the nearly zero-energy management performance in terms of the deviation between PV energy generation and building energy consumption, 2) Electricity Cost term accounts for the energy usage from the utility companies, 3) ESS Depreciation Cost term is to address the depreciation associated with the building ESSs, 4) CL Overload Cost term is to avoid the scheduling of building controllable loads to be concentrated in certain time slots, 5) Visual Discomfort Cost term is to address the users' balance between the discomfort and electricity cost associated with building lighting system, 6) Thermal Discomfort Cost term is to address the users' balance between the discomfort and electricity cost related to building HVAC systems, and 7) EV Charging Incompleteness Cost term term is used to address the EV charging incompleteness. Also, in (5), α and $1-\alpha$ represent the weighting factors to emphasize energy consumption from the electric utility and PV energy generation, and β is a weighting factor associated with the ESS depreciation cost. In (5) and (6), at each time step t, the HVAC energy consumption $P_{t,i}^{HVAC}\,(i = 1, \cdots, N)$ are initially obtained from (1) in which the initial energy consumption of non-HVAC systems $P_{t,i}^{\mathrm{non\_HVAC}}$ are based upon the energy consumption of non-HVAC systems at previous time step. The $P_{t,i}^{\mathrm{HVAC}}$ obtained from (1) are then applied in (5) and (6). Then, by solving the optimization problem of (5) and (6), a new $P_{t,i}^{\mathrm{non\_HVAC}}$ are obtained. Since the HVAC energy consumption energy can be affected through internal heat gain of non-HVAC loads, the new $P_{t,i}^{\mathrm{non\_HVAC}}$ is then applied in (1) to get updated $P_{t,i}^{\mathrm{HVAC}}$. This iterative process continues until a convergence is arrived.

Details about the iterative optimization algorithm are presented in Section V. Then, the control signals are sent to each building for real-time operation of the connected buildings in the next day.

III. LEARNING-BASED BEM FOR BUILDING HVAC LOADS

As shown in Section II, the main challenge for the TESS management of an HVAC system is unknown parameters of the building construction materials and models related to complex building heat transfer. This section develops a DRL method to overcome the challenge and determine the optimal HVAC system control actions while non-HVAC building loads are managed based on physics-based models developed in Section IV. The primary reason to divide HVAC and non-HVAC loads via DRL and physics-based models, respectively, is to significantly reduce the size of the state space associated with the DRL learning model.

A. Actor-Critic based DRL Design for Building HVAC Loads

Actor-Critic algorithms are the foundation behind almost every modern reinforcement learning (RL) method. An Actor-Critic algorithm is a model-free learning algorithm, which means that when it is applied in building HVAC energy management, building thermal storage models are not needed. In an Actor-Critic RL system for HVACs, at each time step *t*, the Actor computes an action based on the state of building HVAC systems, and the Critic tries to find or approximate the optimal value function that is used to evaluate the performance of an action. In this paper, the control action of the building HVAC systems store energy into water-based chiller and boiler systems of buildings. Basically, during a period of high PV power production or low electricity price, the aim is to store surplus energy into a chiller or boiler; while during periods of low PV power production or high electricity price, release stored energy from a chiller or boiler to cool or heat the building. This is similar to the demand response management of a residential water heater system [30]. The advantage of this way is that the comfort demand of building users

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3176306, IEEE Internet of Things Journal

will not be affected during the building BEM. More details about the action, state, value function, and Actor and Critic neural networks (NNs) of the Actor-Critic RL system for the proposed HVAC energy management are explained as follows:

1) Action space: At each time step *t*, the building HVAC BEM for the *n*th building generates energy storage control actions of a chiller/boiler, $P_{t,n}^{\mathrm{CS}}/P_{t,n}^{\mathrm{HS}}$, in which a negative or positive value means extra discharge or charge power of a chiller/boiler. Since the energy storage control of a chiller is similar to that of a boiler, only the Actor-Critic RL model related to the building chiller energy storage control is presented in the rest of this paper. Considering the storage ability of a chiller or boiler, the total power consumption of a building can be represented by

$$P_{t,n}^{\mathrm{HVAC}} = P_{t,n}^{\mathrm{CD}} + P_{t,n}^{\mathrm{CS}} \tag{7}$$

The energy storage control actions of chillers for all connected buildings at time step *t* is described by a vector (8a) and the chiller's accumulated energy storage (8b) of the *n*th building is constrained by the maximum and minimum accumulated energy storage limits of the *n*th building, $S_{n,Max}^{\mathrm{CS}}$ and $S_{n,Min}^{\mathrm{CS}}$ (8c), corresponding to the highest and lowest chilled water temperatures in order to maintain the proper chiller's function that can meet customers' HVAC demand.

$$a_t = \left[P_{t,1}^{\mathrm{CS}}, P_{t,2}^{\mathrm{CS}}, ..., P_{t,n}^{\mathrm{CS}}\right] \tag{8a}$$
$$S_{t+1,n}^{\mathrm{CS}} = S_{t,n}^{\mathrm{CS}} + P_{t,n}^{\mathrm{HVAC\_C}} \tag{8b}$$
$$S_{n,Min}^{\mathrm{CS}} \le S_{t,n}^{\mathrm{CS}} \le S_{n,Max}^{\mathrm{CS}} \tag{8c}$$

2) State space: At each time step *t*, for *N* buildings, the state vector for HVAC BEM of the connected buildings is:

$$s_t = \begin{bmatrix} T_t^{\mathrm{od}}, \phi_t^{\mathrm{od}}, m, h, P_{t,1}^{\mathrm{PV}}, \cdots, P_{t,N}^{\mathrm{PV}}, \\ P_{t,1}^{\mathrm{non\_HVAC}}, \cdots, P_{t,N}^{\mathrm{non\_HVAC}}, S_{t,1}^{\mathrm{CS}}, \cdots, S_{t,N}^{\mathrm{CS}} \end{bmatrix} \tag{9}$$

where the first 4 variables in s_t are indicators of current and future weather conditions. Similarly P^{PV} is an indicator of the current sunshine levels. $P^{\mathrm{non\_HVAC}}$ indicates the current power consumed by the non-HVAC appliances, and S^{CS} represents the current levels of energy stored.

3) Action-Value function: At each time step *t*, for building HVAC BEM state s_t, the action-value function for action a_t is given by

$$Q_t(a_t, s_t) = r_t + \gamma \cdot Q_{t+1}(a_{t+1}, s_{t+1}) \tag{10}$$

where $Q_t(a_t, s_t)$ is the action-value function and $0 \le \gamma \le 1$ is a discount factor. r_t denotes a reward which is observed after the HVAC BEM agent takes an action a_t in each time *t*, and is expressed as:

$$r_t = -\sum_{n=1}^{N} \left(P_{t,n}^{\mathrm{PV}} - P_{t,n}^{\mathrm{HVAC}} - P_{t,n}^{\mathrm{non\_HVAC}}\right)^2 \tag{11}$$

which is a negative cost function based on the difference between the PV energy generated and the total energy consumed by each building. Note that the action vector in (8) affects only the second term in (11), via the choice of energy storage actions in (7). The goal of the RL actor-critic system is to intelligently choose the energy storage controls (8) to maximise (11) summed over all buildings and all time steps.

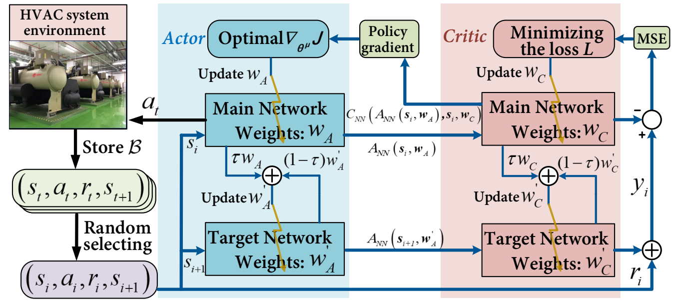

Fig. 2. DDPG Actor-Critic model

4) Actor and Critic NNs: Now, at a time step *t*, we will need a mechanism or function of Actor that can generate HVAC control action for building HVAC energy management and a mechanism or function of Critic that can provide performance evaluation for each control action issued by the Actor. Due to the complexity of HVAC BEM for connected buildings, the functions of the Actor and Critic are achieved through two separate NNs, the Actor NN and the Critic NN. The objective of the Actor NN is to generate a control action a_t for a given system state s_t that will maximize $Q_t(a_t, s_t)$ and the objective of the Critic NN to approximate the value function (10) so that a large memory space used to store the history information of $Q_t(a_t, s_t)$ is not needed. Based on the NN definitions and objectives, the HVAC's control action and the action-value at each time step *t* are mathematically represented as (12) and (13), respectively, as shown below:

$$a_t = \mathrm{A_{NN}}(s_t, w_A) \tag{12}$$

$$Q_t(a_t, s_t) = C_{NN}(a_t, s_t, w_C) \tag{13}$$

where $\mathrm{A_{NN}}(\cdot)$ denotes the Actor NN, and $C_{\mathrm{NN}}(\cdot)$ stands for the Critic NN.

B. DDPG Actor-Critic Model

Deep Deterministic Policy Gradients (DDPG) is a widely used actor-critic technique due to its advantages in many aspects [31] and is adopted in this paper to implement the actor-critic model for building HVAC energy management. DDPG uses four NNs to implement the actor-critic model [32] and includes a main Actor NN, a main Critic NN, a target Actor NN, and a target Critic NN with weights of w_A, w_C, w_A', and w_C' respectively.

At each time step, the main Actor NN takes observation s_t as inputs and updates the Actor NN weights w_A to maximize the expected sum of discounted future reward (10) with respect to control action a_t as shown by (14a). Assume a uniform probability distribution of all the rewards, the expected sum is then (14b) where the action a_t is replaced by the Actor NN.

$$J = \mathbb{E}\left[C_{NN}(a_t, s_t, w_C)\right] \tag{14a}$$
$$= \frac{1}{B}\sum_t \left[C_{NN}\left(\mathrm{A_{NN}}(s_t, w_A), s_t, w_C\right)\right] \tag{14b}$$

The expected sum of (14b) is the cost function of the Actor NN. The objective of the Actor NN training is to minimize the cost represented by (14b). Thus, using the policy gradient over the Actor weight vector w_A, the gradient $\partial J/\partial w_A$ is obtained as (15a)

$$\frac{\partial J}{\partial w_A}=\frac{1}{B}\sum_{T=1}^{B}\frac{\partial C_{NN}\left(\mathrm{A_{NN}}\left(s_t,w_A\right),s_t,w_C\right)}{\partial \mathrm{A_{NN}}\left(s_t,w_A\right)}\frac{\partial \mathrm{A_{NN}}\left(s_t,w_A\right)}{\partial w_A} \tag{15a}$$

$$w_{A\_update}=w_A+\alpha_A\cdot\partial J/\partial w_A \tag{15b}$$

Thus, the main Actor network weights w_A are updated using (15b) based on the gradient obtained in (15a), where $\alpha_A>0$ is the actor learning rate.

The main Critic NN takes observation s_t and action a_t as inputs and updates the Critic NN weights w_C to minimize the expected sum of the difference between the estimated action values obtained by the main Critic NN (13) and the actual action values y_t obtained by using equation (16b).

$$L=\frac{1}{B}\sum_{t=1}^{B}\left[y_t-C_{NN}\left(a_t,s_t,w_C\right)\right]^2 \tag{16a}$$

$$y_t=r_t+\gamma\cdot C_{NN}\left(a_{t+1},s_{t+1},w_C^{'}\right) \tag{16b}$$

The expected sum of (16a) is the cost function of the Critic NN. The objective of the Critic NN training is to minimize the cost represented by (16a). Then, using gradient decent over the main Critic NN weight vector w_C, the gradient $\partial L/\partial w_C$ is obtained as (17a)

$$\frac{\partial L}{\partial w_C}=-\frac{2}{B}\sum_{t=1}^{B}\frac{\partial C_{NN}\left(a_t,s_t,w_C\right)}{\partial w_C} \tag{17a}$$

$$w_{C\_update}=w_C+\alpha_C\cdot\partial L/\partial w_C \tag{17b}$$

Hence, the main Critic network weights w_C are updated using (17b) based on the gradient obtained in (17a), where $\alpha_C>0$ is the critic learning rate.

The DDPG target Actor and target Critic NNs weights are updated at every time step using a smoothing factor τ applied to the main Actor, target Actor, main Critic and target Critic NNs weights as shown by (18), where $0<\tau\leq 1$.

$$w_A'\leftarrow\tau w_A+(1-\tau)\,w_A' \tag{18a}$$

$$w_C'\leftarrow\tau w_C+(1-\tau)\,w_C' \tag{18b}$$

The DDPG is trained offline. After the training, the Actor NN can be applied to generate control action vector a_t of building HVACs that are then used in the overall BEM implementation including both HVAC and non-HVAC loads of connected buildings as shown in the conceptual view of the proposed method presented in Section II and detailed discussion of the proposed method in the following sections.

IV. PHYSICS-BASED MODEL FOR BUILDING NON-HVAC LOADS

A modern building involves multiple zones, which are used for a variety of purposes such as server rooms, office space, and common areas. Besides the HVAC loads, other building loads could consist of indoor lighting systems, ESSs, EVs, washing machines, renewable energy sources, etc. All these loads are classified as non-HVAC loads in this paper. Unlike a building HVAC load that is highly nonlinear and complex, the energy consumptions of non-HVAC loads are easy to model and/or predict based on the manufactures' data and specifications of these loads. Therefore, to significantly decrease the number of control states in a DRL problem, all the building non-HVAC loads and ESSs are modeled in this section based on their physics characteristics.

A. ESS Physics-Based Model

The ESSs can store PV energy and supply the stored energy to building loads during peak load hours through the charging and discharging of ESSs. The SOC transition of an ESS is described by (19a) and the associated ESS constraints are described by (19b)-(19d):

$$S_{t+1,n}^{\mathrm{ES}}=S_{t,n}^{\mathrm{ES}}+\eta^{\mathrm{ES\_C}}P_{t,n}^{\mathrm{ES\_C}}-\frac{P_{t,n}^{\mathrm{ES\_D}}}{\eta^{\mathrm{ES\_D}}} \tag{19a}$$

$$I_{t,n}^{\mathrm{ES}}\cdot P_{Min}^{\mathrm{ES\_C}}\leq P_{t,n}^{\mathrm{ES\_C}}\leq I_{t,n}^{\mathrm{ES}}\cdot P_{Max}^{\mathrm{ES\_C}} \tag{19b}$$

$$\left(1-I_{t,n}^{\mathrm{ES}}\right)\cdot P_{Min}^{\mathrm{ES\_D}}\leq P_{t,n}^{\mathrm{ES\_D}}\leq\left(1-I_{t,n}^{\mathrm{ES}}\right)\cdot P_{Max}^{\mathrm{ES\_D}} \tag{19c}$$

$$S_{Min}^{\mathrm{ES}}\leq S_{t,n}^{\mathrm{ES}}\leq S_{Max}^{\mathrm{ES}} \tag{19d}$$

where $P_{Max}^{\mathrm{ES\_C}}$, $P_{Max}^{\mathrm{ES\_D}}$, $P_{Min}^{\mathrm{ES\_C}}$ and $P_{Min}^{\mathrm{ES\_D}}$ are upper and lower limits of the ESS charging and discharging powers, S_{Max}^{ES} and S_{Min}^{ES} are maximum and minimum capacity of the ESS, and $I_{t,n}^{\mathrm{ES}}\in\{0,1\}$ represents the ESS operation mode. When $I_{t,n}^{\mathrm{ES}}$ is 1, the ESS works in charging mode; when $I_{t,n}^{\mathrm{ES}}$ is 0, the ESS works in discharging mode.

B. Physics-Based Model of Building Electric Vehicles

EVs account for more than 30% of the total building energy consumption according to the history data of the Pecan Street project [34]. The EV charging SOC transition and constraints are described by (20a) and (20b-20d), respectively:

$$S_{t+1,n,m}^{\mathrm{EV}}=S_{t,n,m}^{\mathrm{EV}}+\eta^{\mathrm{EV}}P_{t,n,m}^{\mathrm{EV\_C}} \tag{20a}$$

$$S_{t0,n,m}^{\mathrm{EV}}+\sum_{t0}^{\mathrm{T_i}}P_{t,n,m}^{\mathrm{EV\_C}}\geq E_{\mathrm{T}_i,n,m}^{\mathrm{Trip}} \tag{20b}$$

$$I_{t,n,m}^{\mathrm{EV}}\cdot P_{Min}^{\mathrm{EV\_C}}\leq P_{t,n,m}^{\mathrm{EV\_C}}\leq I_{t,n,m}^{\mathrm{EV}}\cdot P_{Max}^{\mathrm{EV\_C}} \tag{20c}$$

$$S_{Min}^{\mathrm{EV}}\leq S_{t,n,m}^{\mathrm{EV}}\leq S_{Max}^{\mathrm{EV}} \tag{20d}$$

where $P_{Max}^{\mathrm{EV\_C}}$ and $P_{Min}^{\mathrm{EV\_C}}$ are the upper and lower EV charging power limits, S_{Max}^{EV} and S_{Min}^{EV} represent the maximum and minimum EV battery capacities, and $I_{t,n,m}^{\mathrm{EV}}\in\{0,1\}$ is the predicted EV charging mode. When $I_{t,n,m}^{\mathrm{EV}}$ is 1, the EV works in charging mode; when it is 0, the EV works in trip mode. The accumulated EV charging energy should be larger than $E_{\mathrm{T_i},n,m}^{\mathrm{Trip}}$ during each charging cycle.

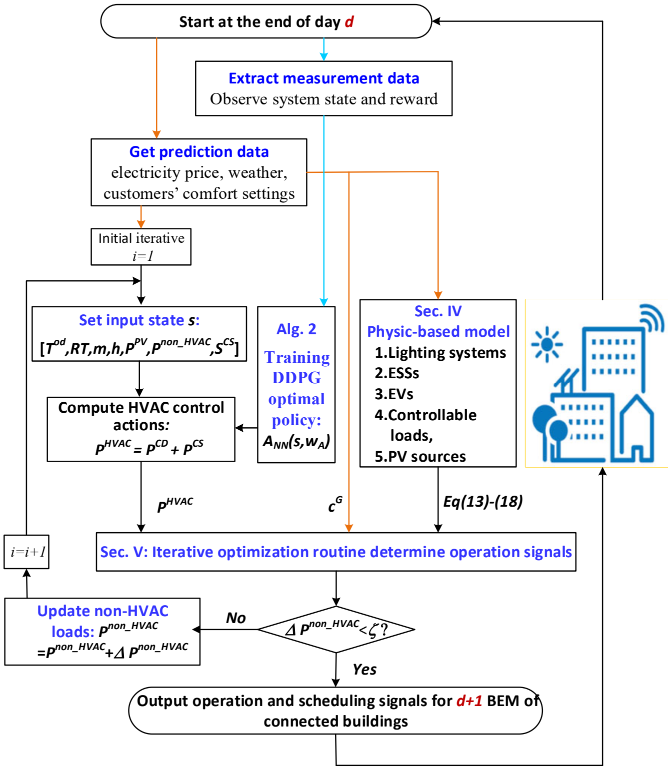

Fig. 3. BEM integration and implementation of connected buildings via the proposed iterative optimization algorithm

C. Physics-Based Model of Indoor Lighting Systems

An indoor lighting system is modeled as (21a) and the energy consumption constraints of the lighting systems is (21b).

$$P_{t,n,k}^{\mathrm{L}} = \frac{A_{t,n,k}^{\mathrm{L}} \cdot E_{t,n,k}^{\mathrm{L}}}{\eta^{\mathrm{L}} \cdot c \cdot m^{\mathrm{L}}} \tag{21a}$$

$$P_{Min}^{\mathrm{L}} \leq P_{t,n,k}^{\mathrm{L}} \leq P_{Max}^{\mathrm{L}} \tag{21b}$$

where P_{Max}^{L} and P_{Min}^{L} are upper and lower power limits of the lighting system.

D. Physics-Based Model of Building Controllable Loads

Controllable Loads (CLs) are flexible loads that can be rescheduled. CL assignments are assumed as constant in one day. These assignments can be finished in one time period or several time periods. The total energy consumption of CLs is modeled by (22a) and energy consumption of each CL assignment should be within the CL energy consumption constraints (22b):

$$P_{t,n,j}^{\mathrm{CL}} = \theta_{t,n,j} \cdot E_{n,j}^{\mathrm{CL\_total}} \tag{22a}$$

$$P_{Min}^{\mathrm{CL}} \leq P_{t,n,j}^{\mathrm{CL}} \leq P_{Max}^{\mathrm{CL}} \tag{22b}$$

where $\theta_{t,n,j} \in [0\ 1]$. When $\theta_{t,n,j}$ is close to 1, it means that the CL is mainly scheduled in that hour.

V. BEM IMPLEMENTATION OF CONNECTED BUILDINGS VIA DDPG ONLINE TRAINING AND ITERATIVE OPTIMIZATION

The BEM implementation involves 1) day-ahead scheduling of building HVAC and non-HVAC loads at the end of the day, and 2) BEM implementation at the operation day (next day). First, the day-ahead scheduling includes a) 24-hour-ahead HVAC control actions generated by the Actor as shown in Section III and Fig. 2, and b) the optimal 24-hour-ahead schedules of other non-HVAC controllable loads, ESSs and EVs of each building that are determined as shown later in this section. Second, the real-time BEM implementation includes a) the operation of connected buildings based on the day-ahead schedules generated above for HVAC and non-HVAC loads at the operation day, and b) at the end of the operation day, update the data storage with the newly obtained data during the operation day and then train the DDPG main and target NNs. Note: the data storage is a first-in, first-out queue so that it only keeps the most recent data of the past months. Finally, repeat the day-ahead scheduling and the online implementation as described above.

To determine the optimal schedule of other CLs, ESSs and EVs of each building, an optimization problem is formulated to minimize the deviation between the PV generated energy and the actual energy consumption of the connected buildings while considering the human comfort requirement as shown below:

Minimize:

$$\begin{aligned} C = \; & \underbrace{(1-\alpha)\sum_{t=1}^{T}\sum_{n=1}^{N}\left(P_{t,n}^{\mathrm{G}}\right)^2}_{\text{(1) Deviation Cost}} + \underbrace{\alpha\sum_{t=1}^{T} c_t^{\mathrm{G}} \cdot P_t^{\mathrm{G}}}_{\text{(2) Electricity cost}} \\ & + \underbrace{\beta\sum_{n=1}^{N} c_t^{\mathrm{ES\_life}} \cdot n^{\mathrm{ES\_cycles}}}_{\text{(3) ESS depreciation cost}} + \underbrace{\sum_{t=1}^{\mathrm{T}}\sum_{n=1}^{N}\sum_{j=1}^{NCL} c_t^{\mathrm{CL}} \cdot \theta_{t,n,j}}_{\text{(4) CL overload cost}} \\ & + \underbrace{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{k=1}^{NL} c_t^{\mathrm{L}} \cdot \left(E_{t,n,k}^{\mathrm{L}} - \bar{E}\right)^2}_{\text{(5) Visual discomfort cost}} + \underbrace{\sum_{n=1}^{N} c_t^{\mathrm{T}} \cdot \Delta T_{t,n}}_{\text{(6) Thermal discomfort}} \\ & + \underbrace{\sum_{n=1}^{N}\sum_{m=1}^{NEV} c_t^{\mathrm{EV}}\left(E_{T_i,n,m}^{\mathrm{Trip}} - S_{t0,n,m}^{\mathrm{EV}} + \sum_{t0}^{T_i} P_{t,n,m}^{\mathrm{EV\_C}}\right)}_{\text{(7) EV charging incompleteness cost}} \end{aligned} \tag{23}$$

Subject to:

1) Load balance constraints of connected buildings:

$$\begin{aligned} P_t^G + \sum_{n=1}^{N} P_{t,n}^{PV} = & \sum_{n}^{N}\big(P_{t,n}^{ES\_C} - P_{t,n}^{ES\_D} + \sum_{m}^{NEV} P_{t,n,m}^{EV\_C} \\ & + P_{t,n}^{UCL} + \sum_{k}^{NL} P_{t,n,k}^{L} + \sum_{j}^{NCL} P_{t,j,n}^{CL} + P_{t,n}^{HVAC}\big) \end{aligned} \tag{24}$$

2) Operation constraints of non-HVAC loads: (19)-(22).

In the above optimization formulation, (24) represents the total load and generation balance constraints of the connected buildings, and (19)-(22) are the operation models and constraints of non-HVAC loads as discussed in Section IV. Regarding the HVAC energy consumptions, they are determined by (7) in which the energy charge/discharge of TESSs is given by (12) based on the DDPG model discussed in Section III and trained continuously during the online implementation stage. In (23), $n^{\mathrm{ES\_cycles}}$ stands for the equivalent ESS charging/discharging cycles for 24 hours that can be described as [33-34]:

$$n^{\mathrm{ES\_cycles}} = \frac{\sum_{t=1}^{T} \left(P_{t,n}^{\mathrm{ES\_C}} + P_{t,n}^{\mathrm{ES\_D}} \right)}{S_{Max}^{\mathrm{ES}} - S_{Min}^{\mathrm{ES}}} \tag{25}$$

The objective of the optimization problem (23) consists of: 1) cost for the deviation between PV energy generation and building energy consumption which is a virtual cost with a unit of kW^2, 2) electricity cost from the utility companies, 3) ESS depreciation cost, 4) CL overload cost, 5) visual discomfort cost, 6) thermal discomfort cost of building users, and 7) cost to account for EV charging incompleteness. Also, in (23), α and $1-\alpha$ represent the weighting factors between energy consumption from the electric utility and PV energy generation, and β is a weighting factor associated with the ESS depreciation cost.

However, one challenge in the optimization problem is the thermal coupling between HVAC systems and the building non-HVAC internal heat gain (IHG) loads. To overcome the challenge, an iterative method is developed to solve the optimization problem. The method involves the following steps: 1) Generate 24-hour-ahead HVAC control actions by the Actor network generated by the DDPG algorithm as shown in Section III and Fig. 2, 2) non-HVAC optimization problem receive scheduled HVAC actions $P_{t,n}^{\mathrm{HVAC}}$ obtained in Step 1 and execute optimization process (23) to achieve 24-hour-ahead scheduling for lighting systems, CLs, EVs, and ESSs based on mixed-integer linear programming (MILP) method, and 3) Check IHG loads' mismatch $\Delta P_{t,n}^{\mathrm{non\_HVAC}} = P_{t,n}^{\mathrm{non\_HVAC}} - \hat{P}_{t,n}^{\mathrm{non\_HVAC}}$ with the previously estimated or calculated non-HVAC IHG energy consumption. If the mismatch $\Delta P_{t,n}^{\mathrm{non\_HVAC}}$ is less than ς, which is a very small number representing the maximum mismatch of the building non-HVAC loads, output the optimal schedule that will be used for the building control management at the next operation day. If not, update previous non-HVAC state values $\hat{P}_{t,n}^{\mathrm{non\_HVAC}} \leftarrow P_{t,n}^{\mathrm{non\_HVAC}}$ and then the process is repeated from Step 1) with the updated states applied to (12) to determine updated HVAC control actions. Fig. 3 summarizes this iterative process.

The formulated programming is implemented and solved in a co-simulation environment by using Python 3.6 and MATLAB 2019b. Basically, the DDPG-based HVAC model is trained and implemented by using Python. The HVAC system control actions generated from the DDPG model is sent to the optimization routine that is implemented by using MATLAB. The HVAC system control actions generated from the DDPG model are sent to the optimization routine that is implemented by using MATLAB. The optimization algorithm integrates the HVAC actions generated by the DDPG policy, physics-based non-HVAC models, electricity prices, and customer comfort constraints to generate the operation and scheduling signals for both HVAC and non-HAVC systems of all connected buildings. The optimization problem is developed and solved by using YALMIP toolbox and solved by CPLEX solver in MATLAB.

It is necessary to point out that by dividing the building appliances into HVAC and non-HVAC systems as proposed in this paper, it has greatly reduced not only the size of the state space and complexity associated with the proposed learning-based DDPG model but also the size and complexity of the optimization problem shown in this section. In future work, if the size N of the connected buildings is too large, the management of connected buildings can be divided into several sub BEM systems or aggregators via a hierarchical BEM structure to reduce the size and complexity of operational constraints.

VI. CASE STUDY AND RESULTS

A case study is conducted to evaluate the proposed BEM IoT system. Details about data characteristics of each connected building, hardware and software setups, and results from the case study are presented in this section.

A. Simulation Setup of Connected Buildings

1) Buildings data: We used data of real-world buildings from the Pecan Street project in Austin, Texas in our simulation model [35]. All buildings are interconnected and the power flow from buildings to buildings is bidirectional. Each building contains an HVAC system, a solar PV system, an ESS, an EV, a washer, a dryer, a dishwasher, a refrigerator, several plug-in fixed loads, and lighting systems for two bedrooms, a dining room, and two washing rooms. In our model, three buildings, with building IDs of 1714, 2470, and 3367, are managed via an aggregator that collects data about HVACs, ESSs, EVs, washers, dryers, dishwashers, lighting systems, FLs, and solar power production from each building. More detailed data characteristics of each appliance are shown in Table I. Note: HVACs are estimated based on the proposed DDPG model and solar power productions are estimated based on the weather forecast from the National Weather Services. Thus, HVACs and solar PVs are not included in Table I.

2) Iterative and Learning-based IoT System setup: The proposed learning-based IoT System for BEM is implemented through co-simulation strategy as discussed in Section V. The hardware environment for the co-simulation system is a computer with an Intel Core i7-8650U CPU at 1.90 GHz and 16 GB RAM.

The training and testing of the proposed DDPG-based HVAC model are implemented by using Python 3.6 based on the software system developed in [36]. We first trained the DDPG model offline, as shown in Fig. 2, by using building data from 05/01/2016 at 12:00 a.m. to 07/31/2016 at 23:00 p.m. We tested the DDPG model online by using building data from 08/01/2016 at 12:00 a.m. to 08/31/2016 at 23:00 p.m. In each training time step, the Actor receives a total of 10 states from the three buildings and takes a total of 6 control actions for the HVAC systems of the three buildings. The Actor and Critic NNs consist of two hidden layers. The numbers of neurons in hidden layers of the Actor NN are 10 and 5, and the numbers of neurons in hidden layers of the Critic NN are 20 and 10. We applied a self-adaptive Adam optimizer to the NNs with an initial learning rate of 0.001 for both Actor and Critic NNs. The activation function of all hidden layers is ReLU and the activation function of the output layer is linear. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3176306, IEEE Internet of Things Journal

TABLE I Data characteristics of the appliances in each of the three buildings

CL	P^{CL}_{Min} (kW)	P^{CL}_{Max} (kW)	Operation time (h)		
Washer	0	1.8	8:00 am-10:00 pm		
Dryer	0	7	8:00 am-9:00 pm		
Dishwasher	0	1.5	9:00 am-10:00 pm		
Lighting system	P^{L}_{Min} (kW)	P^{L}_{Max} (kW)	$\eta^{L}(lm/w)$	c	m
	0	2	90	0	0.7
EV	S^{EV}_{Min} (kWh)	S^{EV}_{Max} (kWh)	$P^{EV\_C}_{Max}$ (kW)	η^{EV}	
	0	30	7	0.85	
ESS	S^{ES}_{Min} (kWh)	S^{ES}_{Mac} (kWh)	$P^{ES\_C}_{Max}$ (kW)	$P^{ES\_D}_{Max}$ (kW)	η^{ES_C}/η^{ES_D}
	15	50	6	6	0.9/1.1
HVAC	S^{CS}_{Min} (kW)	S^{CS}_{Mac} (kW)	P^{CS}_{Max} (kW)	P^{CD}_{Max} (kW)	P^{HVAC}_{Max} (kW)
	0	50	5	5	5

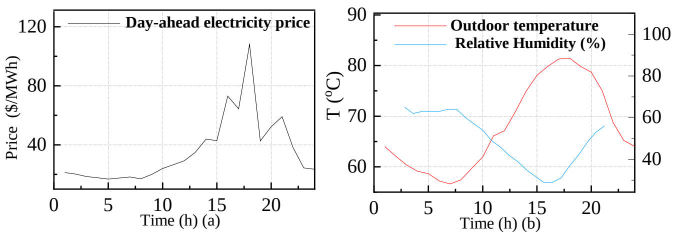

Fig. 4. Test day 24-hour: (a) day-ahead price, and (b) relative humidity, and building outside temperature.

discount factor is $\gamma = 0.99$ and the update factor of the target networks is $\tau = 0.005$. We used Gaussian distribution as the exploration noise process $\mathcal{N}(0, 0.2)$. The maximum episode for the training iteration is 2000.

The proposed iterative optimization algorithm is implemented by using MATLAB 2019b. The optimization formulation (23-24) is developed by using YALMIP toolbox and solved by using CPLEX solver in the MATLAB. As shown in TABLE I, the optimization problem needs to satisfy the operation limits of the buildings' non-HVAC loads and the customers' comfort requirements. The proposed iterative optimization algorithm is tested by using 24 hours building data from 09/01/2016 at 12:00 a.m. to 09/01/2016 at 23:00p.m. The 24-hour day-ahead electricity price data, shown in Fig. 4(a), is generated by the power market. The day-ahead weather data, shown in Fig. 4(b), is predicted by the weather station. Regarding the objective function (23), the electricity cost coefficient α is 0.02, and the battery cost coefficient β is 5.

B. Case Studies

In this section, five case studies were performed to analyze the real-time operation capacity and robustness of the proposed method over the Case 0. Case 0 stands for that no energy management is applied.

Case 1: In this case, only the HVAC energy management via the DDPG model is considered without BEM for non-HVAC loads. As shown in Section III, the reward function of our DDPG approach is to minimize the average peak load by

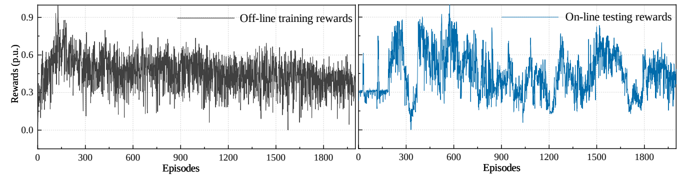

Fig. 5. Rewards of the proposed DDPG method: (a) off-line training rewards, and (b) on-line testing rewards.

considering the thermal storage capability of the chillers and boilers in the building energy management.

Case 2: In this case, both the HVAC and non-HVAC BEMs are considered. The proposed learning-based and iterative method is applied and its performance is compared with that of Case 1. We integrate both physics-based non-HVAC load scheduling and learning-based HVAC load scheduling to minimize both long-term and short-term connected load curve.

Case 3: In real-time building operation, the uncertainties of the weather prediction and changes of the customers' comfort requirements can significantly influence the performance of the BEM systems. To evaluate the robustness of the proposed method, we add an 15% noise process in weather prediction. In addition, a random 10% increase of EV charging, visual equipment and CLs are considered in this case.

Case 4: In this case, a pure physics-based method of the IoT-based smart energy management system (SEMS) shown in [15] is compared with the proposed method. The IoT-based smart SEMS uses physics-based energy storage and thermal storage models to minimize the energy costs of connected buildings. However, it is necessary to point out that due to the highly nonlinear nature of building thermal dynamics, it is very hard to get an accurate building thermal storage model using the physics-based modeling approach.

Case 5: A DRL method for the Smart Home Energy Management (SHEM) in [24] is compared with the proposed method. In [24], the pure learning-based SHEM to regulate the energy consumption of building ESS and HVAC systems only to minimize the energy cost of a single home. As it can be seen in [24], many other building controllable loads were not considered and included in the DRL method in [24] because this will significantly increase the complexity of the DRL method and make it impossible to implement.

C. Results and Evaluation

1) Evaluation of the DRL Model on HVAC BEM: We first evaluated the performance of the proposed DDPG by observing the rewards during 2000 episodes. Rewards present the sum of the deviation between the on-site PV power generation and the thermal and IHG loads of the three test buildings. Fig. 5 (a) shows normalized off-line training rewards based on the maximum and minimum values of the cumulated rewards and Fig. 5 (b) shows normalized on-line testing rewards. The value of the off-line training rewards is increasing fast in the first several episodes and gradually stable after 600 episodes. Similar to the off-line training, the on-line testing reward increases quickly during the first several episodes, then fluctuates a little bit because of the disturbances that are different from those seen in the off-line training data,

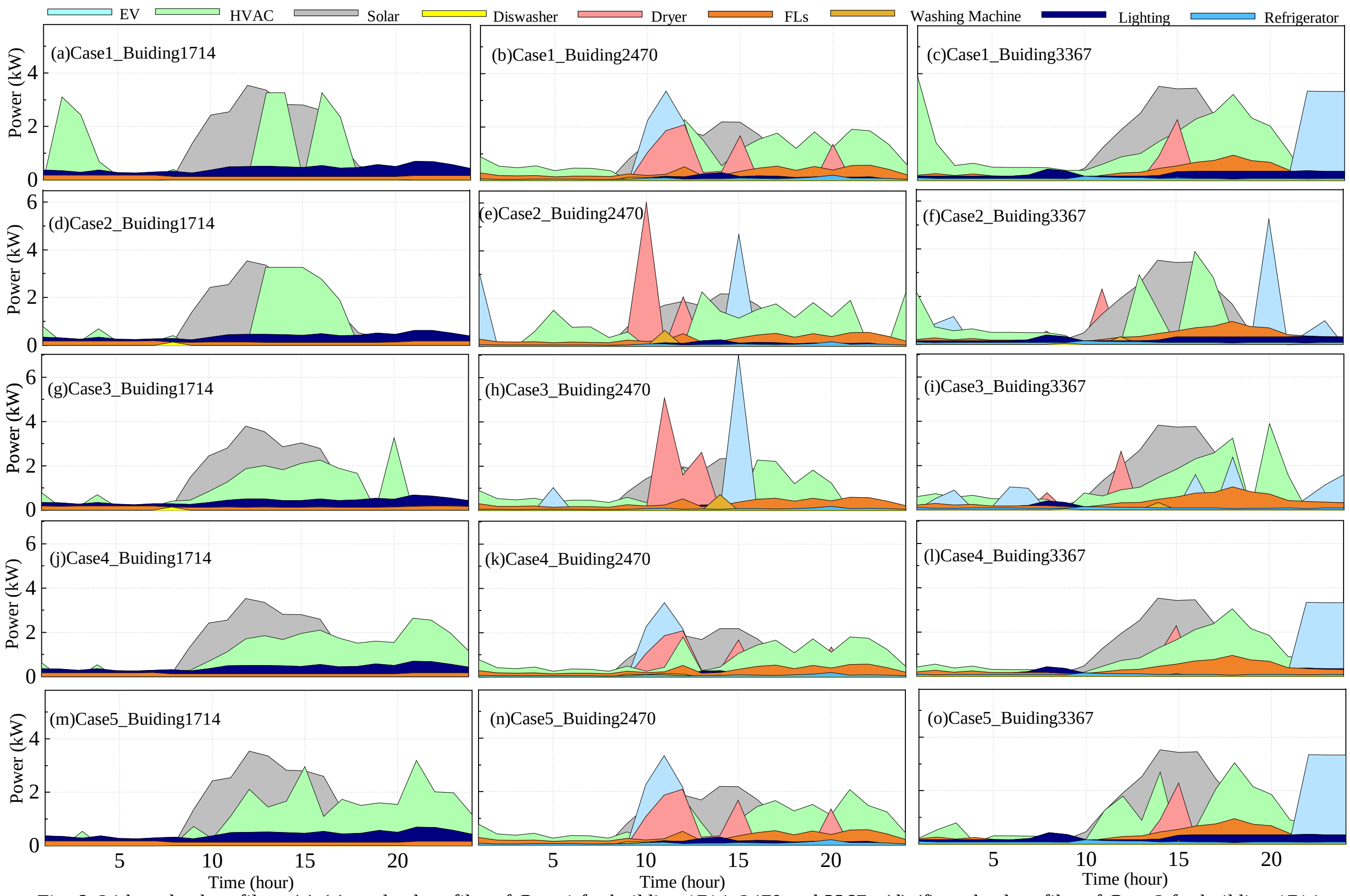

Fig. 6. 24-hour load profiles: (a)-(c) are load profiles of Case 1 for building 1714, 2470 and 3367; (d)-(f) are load profiles of Case 2 for building 1714, 2470 and 3367; (g)-(i) are load profiles of Case 3 for building 1714, 2470 and 3367; (j)-(l) are load profiles of Case 4 for building 1714, 2470 and 3367 and (m)-(0) are load profiles of Case 5 for building 1714, 2470 and 3367.

and finally reaches the region that meets the performance requirement. The DRL model is able to learn a better policy to minimize whole year peak load cost by controlling the TESS. The trained policy is applied to the optimization model for the on-line testing of the BEM schedule.

2) BEM for HVAC and non-HVAC: Fig. 6 presents the detailed control schedule for each building which includes EVs, washing machines, dryers, dishwashers, refrigerators, lighting systems, FLs, ESSs, and solar generators. Apparently, different buildings may have different electrical appliances and different frequency of using the appliances. Energy consumption of EVs and dryers in building 1714 is much less than that in building 2740 and 3367 and the solar energy in building 1714 is more than that of the other two buildings during on-line testing day.

Fig. 7 shows the scheduling of total loads, ESSs, TESSs, and solar generation of three connected buildings. From Fig. 7(a), the DRL model can control TESSs to charge and discharge thermal energy for cooling buildings based on the learned policy. Compared with uncontrolled model Case 0 as shown in Table II, the peak load of Case 1 decreases by 29.55% and the electricity cost decreases about \$205. However, the power consumption from the grid still fluctuates. In addition, it has an unexpected negative net load to the power system at 9:00 AM. The peak load hour of Case 1 appears at 10:00 PM and the peak loads value is 7.70 kW.

In Fig. 7(b), the results using the proposed method for both HVAC and non-HVAC loads show smoother net loads for energy consumption from the grid. The algorithm arranges each CL schedule by shifting controllable loads during PV generation hours or reducing EV charging power during peak load hours. By applying both TESSs and ESSs, Case 2 offers a flatter load profile. As shown in Table II, the maximum peak load value of Case 2 is 3.05 kW which is much lower than that of Case 1 and Case 0. During peak load hour, Case 2 has reduced 72% and 60% in peak load compared to that of Case 0 and Case 1, respectively. Furthermore, the electricity cost is much lower than that of the other methods by reducing total buildings' energy consumption during high electricity price hours.

3) Robustness of the Proposed Method: A mismatch with actual data for the predicted weather and customer comfort data is introduced in Case 3. The mismatch follows uniform distributions as follows: 1) outdoor temperature has $\varsigma^T = \mathcal{U}\left(-0.075\hat{T}_t^{\mathrm{od}}, 0.075\hat{T}_t^{\mathrm{od}}\right)$, 2) outdoor humidity has $\varsigma^{rh} = \mathcal{U}\left(-0.075r\hat{h}_t, 0.075r\hat{h}_t\right)$, and 3) customers' load requirements has $\varsigma^P = \mathcal{U}\left(0, 0.1\hat{P}_{t,n}^{\mathrm{UCL}}\right)$. Based on the results of Case 3, the mismatch of the weather prediction influenced the HVAC operation process. In addition, more load demand of EVs, dryers, lighting systems, and FLs by the customers

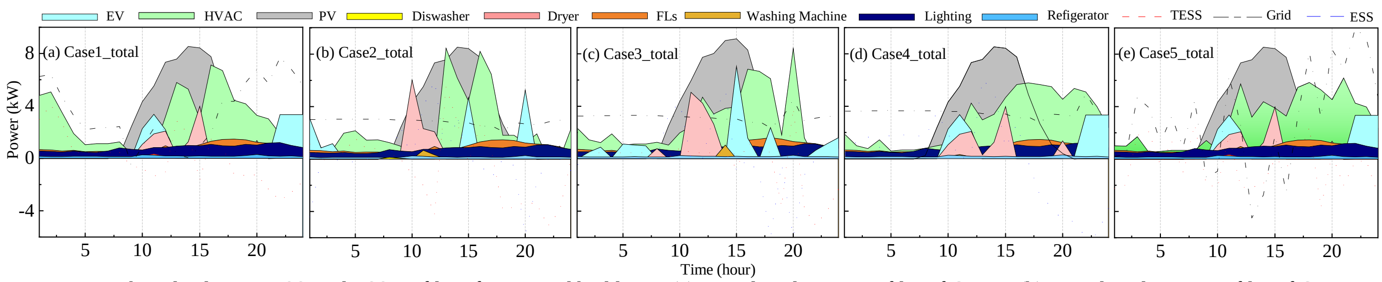

Fig. 7. 24-hour load, PV, TESS, and ESS profiles of connected buildings: (a) is total applicance profiles of Case 1; (b) is total applicance profiles of Case 2; (c) is total applicance profiles of Case 3; (d) is total applicance profiles of Case 4; and (e) is total applicance profiles of Case 5.

TABLE II Five Cases Simulation Costs

Case Studies	Case0	Case1	Case2	Case3	Case4	Case5
Peak load (kW)	10.93	7.70	3.05	4.51	4.93	9.92
Peak reduction (%)	0	29.55	72.10	58.74	55.93	9.1
Zero energy cost (kW^2)	569.92	436.33	196.06	244.18	304.44	567.88
Electricity cost ($)	3377.85	3172.61	2313.88	2613.37	2910.93	2925.65

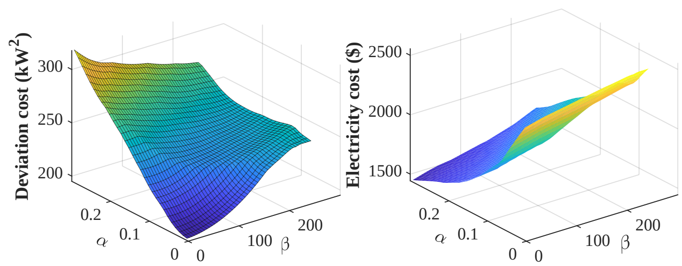

Fig. 8. Impact of weighting factors of electricity and battery depreciation costs on building cost

increased the usages of ESSs. The peak load is 1.46 kW which is more than that of Case 2. However, the effect of these disturbances is acceptable for the operation of connected buildings during the test day. The zero-energy cost of Case 3 is 244.18 kW^2 which is 325.74 kW^2 and 192.15 kW^2 lower than that of Case 0 and Case 1, respectively. The electricity cost is still lower than that of the DRL method. Therefore, the proposed algorithm has better robustness to handle uncertainty weather prediction and increasing customer load demand.

4) Proposed Method Compared with Literature Studies: In Case 4, the charging/discharging energy of the ESSs is larger than that of the proposed method and the net loads' pattern of the connected buildings fluctuates more (Fig. 7(b) and 7(d)). Also, as shown in Table II, the peak load reduction of Case 4 (55.93%) is less than the peak load reduction of the proposed method (72.10%). In addition, both the zero-energy cost and electricity cost are higher than that of Case 2 because of the lack of participation of all possible controllable loads into the BEM in Case 4.

In Case 5, the ESS power f_t and HVAC system power e_t are selected by the learned optimal policy using the DRL method in each time slot. As shown in Figures 6 and 7, the pure learning-based SHEM method didn't consider the non-HVAC loads and zero energy cost. Figure 7(e) shows a less peak load reduction compared with our proposed method. In addition, Case 5 has a higher energy cost because many building controllable loads are not considered and included in the DRL method.

5) Parameter Sensitive Analysis of the Proposed Method: In some areas, investment of ESSs and electricity cost are different from the simulated location in Austin, USA. To further explore the impact of electricity price and battery cost on the proposed method, we conducted a parameter sensitive analysis by setting α and β in (17a) as variables. We noticed that zero energy cost will increase when the electricity cost coefficient α and battery cost coefficient β increase (Fig. 8). The maximum zero energy cost is 317.72 kW^2. However, electricity cost is lower when α is high and β is low.

VII. CONCLUSION

This paper proposes an iterative optimization and learning-based IoT application for self-scheduling of ESSs, TESSs, and loads of connected-buildings to maximize usage of on-site PV energy, flatten building load profiles, and reduce electricity costs. The iterative optimization technology integrates the DRL method and physics-based method, which first learns a good thermal storage policy and then generates the optimal control commands through an iterative technique for the remaining loads of all connected buildings for each design day. Simulation results demonstrate that the proposed method has a better performance to cooperate and rearrange each building's electrical appliances to smooth the buildings' load and reduce the total energy consumption of the connected buildings. In addition, the method can operate well considering the weather prediction deviation and uncertainty of customers' load requirements. It is found that high electricity and battery costs can influence the cost to achieve the zero-energy goal. The customers can increase PV generation to minimize the cost of the electricity bill. In future work, we will continue to study the islanding mode of connected buildings by considering renewable energy generation and ESS cost.

In future work, we plan to extend the proposed BEM IoT system to nearly zero energy management of larger-scale systems, such as nearly zero energy communities and nearly zero energy cities. For example, to achieve nearly zero energy cities, future research works need to solve diverse challenges, such as integration of large-scale connected buildings to the grid, intelligent EV charging in cities, etc. These have been laid out in the research roadmap of the global research community in energy [37].

REFERENCES

[1] Minoli D, Sohraby K, Occhiogrosso B. IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems[J]. IEEE Internet of Things Journal, 2017, 4(1): 269-283.

[2] Susan S, Wardhani D. Building integrated photovoltaic as GREENSHIP'S on site renewable energy tool[J]. Results in Engineering, 2020, 7: 100153.

[3] Raugei M, Peluso A, et al. Life-Cycle Carbon Emissions and Energy Return on Investment for 80% Domestic Renewable Electricity with Battery Storage in California (USA)[J]. Energies, 2020, 13(15): 3934.

[4] Zhang X, Pipattanasomporn M, Chen T, et al. An IoT-based thermal model learning framework for smart buildings[J]. IEEE Internet of Things Journal, 2019, 7(1): 518-527.

[5] Vishwanath A, Chandan V, Saurav K. An IoT-based data driven precooling solution for electricity cost savings in commercial buildings[J]. IEEE Internet of Things Journal, 2019, 6(5): 7337-7347.

[6] Guan X, Xu Z, Jia Q S. Energy-efficient buildings facilitated by microgrid[J]. IEEE Transactions on smart grid, 2010, 1(3): 243-252.

[7] Wu H, Shahidehpour M, et al. Demand response exchange in the stochastic day-ahead scheduling with variable renewable generation[J]. IEEE Transactions on Sustainable Energy, 2015, 6(2): 516-525.

[8] Yang Y, Jia Q S, Guan X, et al. Decentralized EV-based charging optimization with building integrated wind energy[J]. IEEE Transactions on Automation Science and Engineering, 2018, 16(3): 1002-1017.

[9] Xia Z, Qahouq J A A. State-of-Charge Balancing of Lithium-Ion Batteries With State-of-Health Awareness Capability[J]. IEEE Transactions on Industry Applications, 2020, 57(1): 673-684.

[10] Tiwary A, Mahato M, et al. Internet of Things (IoT): Research, architectures and applications[J]. International Journal on Future Revolution in Computer Science & Communication Engineering, 2018, 4(3): 23-27.

[11] Yu, Liang, et al. "A review of deep reinforcement learning for smart building energy management." IEEE Internet of Things Journal (2021).

[12] Zhou D P, Hu Q, Tomlin C J. Quantitative comparison of data-driven and physics-based models for commercial building HVAC systems[C]//2017 American Control Conference (ACC). IEEE, 2017: 2900-2906.

[13] Wang Z, Chen Y, Li Y. Development of RC model for thermal dynamic analysis of buildings through model structure simplification[J]. Energy and Buildings, 2019, 195: 51-67.

[14] Metallidou, Chrysi K., Kostas E. Psannis, and Eugenia Alexandropoulou Egyptiadou. "Energy efficiency in smart buildings: IoT approaches." IEEE Access 8 (2020): 63679-63699.

[15] Liu, Chunming, Dingjun Wang, and Yujun Yin. "Two-stage optimal economic scheduling for commercial building multi-energy system through internet of things." IEEE Access 7 (2019): 174562-174572.

[16] Kazmi, Saqib, et al. "Towards optimization of metaheuristic algorithms for IoT enabled smart homes targeting balanced demand and supply of energy." IEEE access 7 (2017): 24267-24281.

[17] Hussain, Bilal, et al. "An innovative heuristic algorithm for IoT-enabled smart homes for developing countries." IEEE Access 6 (2018): 15550-15575.

[18] Minakais M, Mishra S, Wen J T. Database-driven iterative learning for building temperature control[J]. IEEE Transactions on Automation Science and Engineering, 2019, 16(4): 1896-1906.

[19] Bedi, Guneet, Ganesh Kumar Venayagamoorthy, and Rajendra Singh. "Development of an IoT-driven building environment for prediction of electric energy consumption." IEEE Internet of Things Journal 7.6 (2020): 4912-4921.

[20] Zhang W, Hu W, Wen Y. Thermal comfort modeling for smart buildings: A fine-grained deep learning approach[J]. IEEE Internet of Things Journal, 2018, 6(2): 2540-2549.

[21] Hu, Weizheng, et al. "itcm: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings." IEEE Internet of Things Journal 5.5 (2018): 4164-4177.

[22] Ye Y, Qiu D, Wu X, et al. Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3068-3082.

[23] Iqbal, Muhammad Muzaffar, et al. "IoT-enabled smart home energy management strategy for DR actions in smart grid paradigm." 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST). IEEE, 2021.

[24] Yu L, et al. Deep reinforcement learning for smart home energy management[J]. IEEE Internet of Things Journal, 2019, 7(4): 2751-2762.

[25] Vázquez-Canteli J R, Ulyanin S, Kämpf J, et al. Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities[J]. Sustainable cities and society, 2019, 45: 243-257.

[26] Vázquez-Canteli J, Kämpf J, Nagy Z. Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted Q-iteration[J]. Energy Procedia, 2017, 122: 415-420.

[27] Lin, Bo, Shuhui Li, and Yang Xiao. "Optimal and learning-based demand response mechanism for electric water heater system." Energies 10.11 (2017): 1722.

[28] Gao, Yixiang, et al. "Energy management and demand response with intelligent learning for multi-thermal-zone buildings." Energy 210 (2020): 118411.

[29] Saheb, Yamina, Sophie Shnapp, and Daniele Paci. From nearly-zero energy buildings to net-zero energy districts. JRC Technical Reports, 2019.

[30] Yuan J, Cui C, Xiao Z, et al. Performance analysis of thermal energy storage in distributed energy system under different load profiles[J]. Energy Conversion and Management, 2020, 208: 112596.

[31] Actor-critic learning for optimal building energy management with phase change materials[J]. Electric Power Systems Research, 2020, 188: 106543.

[32] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv:1509.02971, 2015.

[33] Onda K, Ohshima T, Nakayama M, et al. Thermal behavior of small lithium-ion battery during rapid charge and discharge cycles[J]. Journal of Power sources, 2006, 158(1): 535-542.

[34] Qahouq J A A, Xia Z. Single-perturbation-cycle online battery impedance spectrum measurement method with closed-loop control of power converter[J]. IEEE Transactions on Industrial Electronics, 2017, 64(9): 7019-7029.

[35] PECAN STREET – Pecan Street Inc. (2012, October 15) [Online]. Available: https://www.pecanstreet.org/

[36] OpenAI. Available at: https://openai.com/

[37] Villa-Arrieta, Manuel, and Andreas Sumper. "Economic evaluation of nearly zero energy cities." Applied energy 237 (2019): 404-416.