Energy Proceedings
Vol 47, 2024
ISSN 2004-2965

A Deep Reinforcement Learning Approach for Peak Shaving of Smart Buildings[#]

Yitao Deng[1], Yi Zhang[2*], Wai Kin Chan[1*], Yuming Zhao[3], Ilan Adler[4]

1 Tsinghua-Berkeley Shenzhen Institute, Tsinghua University
2 Institute of future human habitats, Shenzhen international Graduate School, Tsinghua University
3 Shenzhen Power Supply Bureau Co., Ltd
4 Industrial Engineering and Operations Research, University of California, Berkeley
(Corresponding Author: zy1214@sz.tsinghua.edu.cn/chanw@sz.tsinghua.edu.cn)

ABSTRACT

As an increasing number of areas are turning to renewable energy sources to meet the growing energy demand of buildings, the intermittent generation times of renewable energies present a significant challenge. These sources often fail to provide sufficient energy during peak consumption periods. Vehicle-to-Building systems (V2B), serving as flexible energy storage solutions within buildings, have the capability to overcome these intermittency issues. This work focuses on applying deep reinforcement learning (DRL) to control the complex building energy management system. Our algorithm leverages historical photovoltaic data, building energy consumption profiles, and the State of Charge (SOC) along with the entry and exit times of electric vehicles. It strategically sets the charging and discharging power of each EV in real-time to optimize energy usage and manage the unpredictability of renewable energy. We integrated the reinforcement learning framework with a city-level energy simulation platform and conducted on various urban forms in Shenzhen, China as case studies. A series of experiments were carried out, demonstrating the effectiveness and practicality of our approach compared with Model Predictive Control (MPC) methods in peak shaving and load leveling.

Keywords: deep reinforcement learning, peak shaving, vehicle-to-building, energy management system

NONMENCLATURE

Abbreviations	
EV	Electric vehicle
V2B	Vehicle-to-Building systems
SOC	State of Charge
DRL	deep reinforcement learning
MDP	Markov Decision Process
MPC	Model Predictive Control
PV	Photovoltaic power generation
PPO	Proximal Policy Optimization
QP	Quadratic Programming
Symbols	
P_{load}^{t}	Power consumption of the building
$B_{ev,i}$	Battery capacity of ev
$P_{ev,i}^{max}$	Maximum charging power
SOC_i	Battery condition
P_{PV}^{t}	PV power genenration
$Q_{ev,i}^{t}$	The charging rate of ev
P_{cap}	Total capacity of charging piles
P_{Grid}^{t}	Power got from grid at time t
$\eta_{EV,i}$	the efficiency of charging operations

1. INTRODUCTION

The growing demand for electricity during peak consumption periods in buildings presents substantial challenges to the stability and efficiency of power grids. This surge in demand not only places a considerable strain on the existing infrastructure but also hampers the effective integration of renewable energy sources, which often face inconsistency in power generation throughout the day. To tackle these challenges, it is imperative to incorporate flexible energy storage solutions into the building's energy management system. Advanced battery technologies and the utilization of electric vehicles serve as key components in such solutions. They have the potential to alleviate the impact of peak loads by storing surplus energy during periods of low demand and subsequently releasing it during times of high demand, thereby enhancing the overall efficiency and sustainability of the power grid. This integration of innovative energy storage solutions is a critical step towards achieving a more resilient and environmentally friendly energy system.

Intelligent control methods for energy storage devices, such as electric vehicles, have seen significant

[#] This is a paper for the 10th Applied Energy Symposium: Low Carbon Cities & Urban Energy Systems (CUE2024), May. 11-13, 2024, Shenzhen, China.

advancements over the years. Initially, rule-based approaches[1] were employed, which relied on simple and effective single decision variables for control. These were followed by the advent of control algorithms like Model Predictive Control (MPC) [2][3], which uses rolling optimization techniques to reformulate control problems into optimization challenges. In contemporary times, machine learning techniques[4], particularly reinforcement learning[5], are being harnessed to develop sophisticated control strategies. However, many of these methods are primarily focused on reducing electricity costs for consumers and decreasing grid dependency[6], thereby contributing indirectly to peak demand reduction without explicitly targeting peak values.

This paper presents a novel peak shaving algorithm based on reinforcement learning, which has been tested on a city-level simulator to ascertain its effectiveness and reliability. The experimental findings show the potential of our approach in effectively optimizing peak loads.

The paper is organized as follows: in the next section we describe the online optimization challenge and the reinforcement learning approach of dynamically adjusting the charging and discharging power of electric vehicles to optimize energy consumption profiles in buildings. In the section 3, we describe our experiments of urban forms in Shenzhen and analyze the results compared with the MPC method. In the section 4, we conclude the paper. In the last section, we propose some new problems to be solved in the future.

2. METHODOLOGY

2.1 Problem description

Real-time adjustment of electric vehicle (EV) charging and discharging power to optimize the building energy consumption profile is an online optimization problem. It requires consideration of multiple factors within the building, including building resilience and photovoltaic (PV) power generation. The optimization goal is to significantly reduce the maximum energy consumption of the building during peak times. However, there is a lower limit to the total amount of energy consumption, and reducing peak loads may create new peak areas elsewhere.

We optimize the current problem through modeling. The electric vehicle's battery charge is determined based on the remaining charge from the previous moment, the current charging or discharging power, and the battery capacity.

$$SOC_n(i+1) = SOC_n(i) + \eta_{EV,i}\frac{P_{ev,n}^{i} * \Delta t}{B_{ev,i}} \quad (1)$$

$$P_{ev,n}^{min} \le P_{ev,n}^{t} \le P_{ev,n}^{max} \quad (2)$$

$$SOC_n^{min} \le SOC_n(i) \le SOC_n^{max} \quad (3)$$

The variable $\eta_{EV,i}$ represents the efficiency of charging and discharging operations. Additionally, the power output of each charging station and the battery capacity of each electric vehicle are capped at their respective maximum values.

$$\sum_{n=1}^{N_t} P_{ev,n}^{t} \le P_{cap}, \forall t = 1,2,\cdots,T \quad (4)$$

The term N_t represents the number of charging stations within a building area, which corresponds to the maximum number of electric vehicles that can engage in Vehicle-to-Building[7] operations. In the model, at each moment t, the total power for charging and discharging must not exceed a predefined limit (P_{cap}). This constraint is designed to prevent the simultaneous high-power charging and discharging from causing a surge that could impact the electrical grid adversely

2.2 Reinforcement learning approach

In this study, we use deep reinforcement learning to address the challenge of peak shaving in smart building energy management[8] by regulating the charging and discharging power of electric vehicles connected within a Vehicle-to-Building system. The core of our approach involves formulating the energy management problem as a Markov Decision Process (MDP) problem, where decisions about charging and discharging EV batteries are made at discrete intervals, each spanning ten minutes. This time interval is selected to balance the complexity of decision-making with the practical responsiveness needed in dynamic building energy management.

State Space: The state space captures both the dynamic characteristics of the electric vehicles and the energy status of the building, enabling the model to make decisions that optimize energy usage and peak load management. It includes the state of charge (SOC_i), maximum charging and discharging power ($P_{ev,i}^{max}$), and battery capacity for each electric vehicle ($B_{ev,i}$) at time t. In addition to the electric vehicles, the state space also contains some basic information, including the building's energy consumption(P_{load}^{t}), the current time(t), and the photovoltaic power generation(P_{PV}^{t}) at the current time.

$$S = [ev_0, ev_1, \ldots, ev_n, t, P_{PV}^{t}, P_{load}^{t}] \quad (5)$$

$$ev_i = [SOC_i, B_{ev,i}, P_{ev,i}^{max}], i \in [1, n] \quad (6)$$

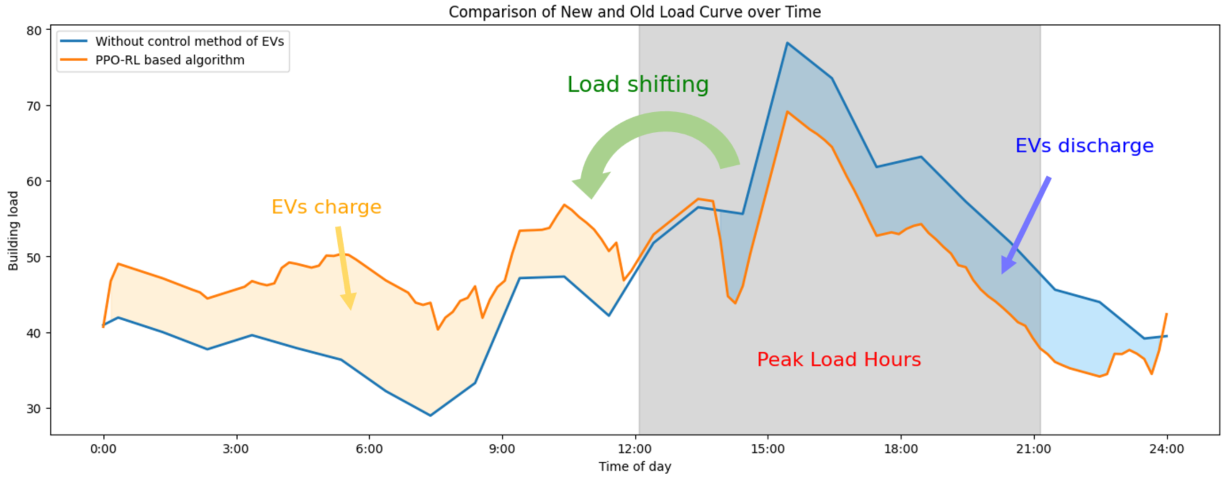

Fig. 1 Load shifting in one day in Minzhi Subdistrict of Longhua District,Shenzhen

Action Space: The action space consists of a continuous vector representing the charging and discharging rates of electric vehicles at each moment, with values ranging from -1 to 1. This approach of not directly selecting the power output but rather using a normalized action space facilitates faster and more stable convergence during training.

$$A_t = [Q^t_{ev,0}, Q^t_{ev,1}, \dots, Q^t_{ev,n}] \tag{7}$$

$$SOC_i(t) = SOC_i(t-1) + \frac{Q^t_{ev,i} * P^{max}_{ev,i}}{B_{ev,i}} \tag{8}$$

In this model, $Q^t_{ev,i}$ represents the rate of charging or discharging, where negative values indicate discharging and positive values indicate charging. The next state of each electric vehicle is determined based on this rate $Q^t_{ev,i}$ and the quantities defined in the state space.

Reward Function: We have creatively designed the reward function, as the ultimate goal is to reduce peak energy consumption. Considering the maximum value over a period in the reward function is challenging, so we employ a retrospective method. We retain historical data of building energy consumption for the past K moments and calculate the average energy consumption over this K-period. The penalty term in the reward function is the squared difference between the current moment's energy consumption and this average value. This design encourages the reinforcement learning training to smooth the energy consumption curve toward the historical average, effectively reducing peaks.

$$P^t_{Grid} = max\left(0, P^t_{load} - P^t_{PV} - \sum_{i=0}^{n} P^t_{ev,i}\right) \tag{9}$$

$$r_t = -\left(\frac{\sum_{i=t-k}^{t} P^i_{load}}{k} - P^t_{Grid}\right)^2 \tag{10}$$

The Algorithm 1 shows the process of running a PPO reinforcement learning method.

Algorithm 1. PPO-Based Approach

Input: $ev_i,\ t, P^t_{PV}, P^t_{load}$

Output: $A_t = [Q^t_{ev,0}, Q^t_{ev,1}, \dots, Q^t_{ev,n}]$

Initialization: Initial action, parameters θ and a storage buffer for trajectory memory

```
for each step of an episode do
    Get initial observation state $s_t$
    for t = 1……T do
        Select action $a_t$ from actor network
        Get the reward $r_t$ according to (10)
        Observe the next state $s_{t+1}$
        Store $\{s_t, a_t, r_t, s_{t+1}\}$ into replay buffer
        Sample from the replay buffer
        Compute the value of advantage function
        Update the network parameters
        $s = s_{t+1}$
    end for
end for
```

3. RESULTS AND DISCUSSION

3.1 Experimental setup

We conducted a case study in Shenzhen to validate the effectiveness of our algorithm. The entire experiment was based on the CityEnergyFlow Navigator, a city-level energy simulator. This system includes 480,000 electric vehicles moving in real-time within Shenzhen. At each time step, the state of each EV is updated according to a mobility model, and the building energy consumption state is updated based on a predictive model.

We interacted with the system using the Proximal Policy Optimization (PPO)[9] reinforcement learning algorithm, which allows real-time acquisition of state information at each time step. The parameter settings for training the network are shown in Table 1. This data is then processed and learned through a neural network to calculate the value of the reward function at each moment. The action vectors selected by the action network are returned to the simulator to precisely control the charging and discharging operations of the EVs within the buildings managed by the Navigator. The simulator then continues running, awaiting the return and results of the next time step. Through multiple rounds of training in interaction with the system, we fine-tuned and obtained the final parameters of the PPO model.

Tabel 1 PARAMETERS OF PPO TRAINING

Parameters	Value	Parameters	Value
learning rate	1e-3	clip epsilon	0.2
gamma	0.99	batch size	64
gae lambda	0.95	lambda entropy	0.01

In the platform, Shenzhen is divided into different physical urban forms. We treat each urban form as an agent for control purposes. Our primary training period is selected from a week in the summer, as the peak loads during this season pose the greatest threat to the electrical grid throughout the year. We imply distributed reinforcement learning training to each urban form agent. On average, each agent has thousands of electric vehicle entries per week, which provides enough flexible resources to control.

3.2 Performance evaluation

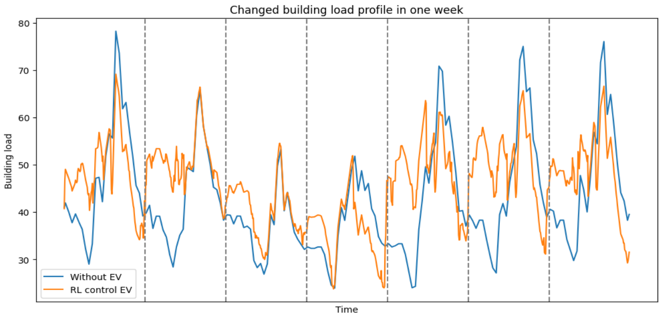

Fig. 2 Load shifting in one week

We initially conducted experiments in a physical region in the Minzhi Subdistrict of Longhua District where we trained and ran the PPO reinforcement learning network. The results, as shown in Figure 1, demonstrate that the new energy consumption curve for buildings, represented in yellow, greatly reduces peak values compared to the original curve. By utilizing electric vehicles for scheduling, the load during peak periods in the afternoon and evening was shifted to the off-peak periods at night and early morning.

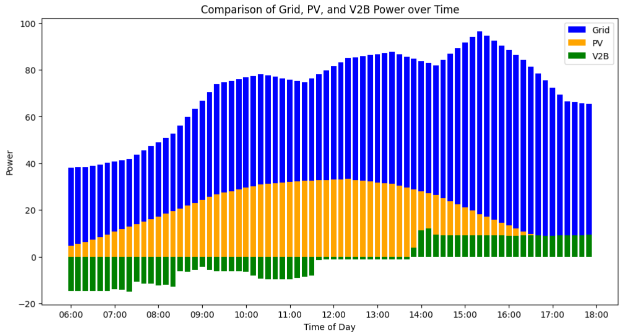

Fig. 3 Grid, PV and EV Power of the building

Figure 3 displays the power graphs for PV, EV, and building systems after optimization by our algorithm. To validate the effectiveness of our algorithm, we compared it with a baseline method, the Model Predictive Control Method[10]. MPC is a classical control algorithm that generally achieves good results under typical conditions.

Based on the optimization model described in section 2.1, we use the Model Predictive Control to obtain a receding horizon, transforming the control problem into an online Quadratic Programming (QP) optimization issue. The approach involves looking ahead several time slices at each time slice, aiming for the current algorithm to be locally optimal within this period. Only the outcome of the first time slice got from each calculation is applied, and this process is recursively rolled forward.

Our baseline MPC model is described as follows:

$$min\ S_t^T * Q * S_t + \sum_{n=1}^{N_t} \left(U_{t,n} - U_{t-1,n}\right)^T * R * \left(U_{t,n} - U_{t-1,n}\right) \quad (11)$$

$$S_t, U_{t,n} \in R^{T\times\mathbb{1}}, Q, R \in R^{T\times T} \quad (12)$$

$$S_t = \left[\xi_{0,t}, \xi_{1,t}, \xi_{2,t}, \cdots, \xi_{T-1,t}\right]^T \quad (13)$$

$$\xi_{i,t} = \left(P_{load}^{t+i} - P_{PV}^{t+i} - \sum_{n=1}^{N_t} P_{ev,n}^{t}\right) * \Delta t \quad (14)$$

In this baseline model, the second term represents the control loss of the MPC, which is the damage caused by frequent switching between charging and discharging to the batteries or charging stations.

Q, R are two coefficient matrices, and represents the importance you attached to each part in the model. We use Gurobi to solve the current model, obtaining the control vector for each moment under several constraints of the optimization model.

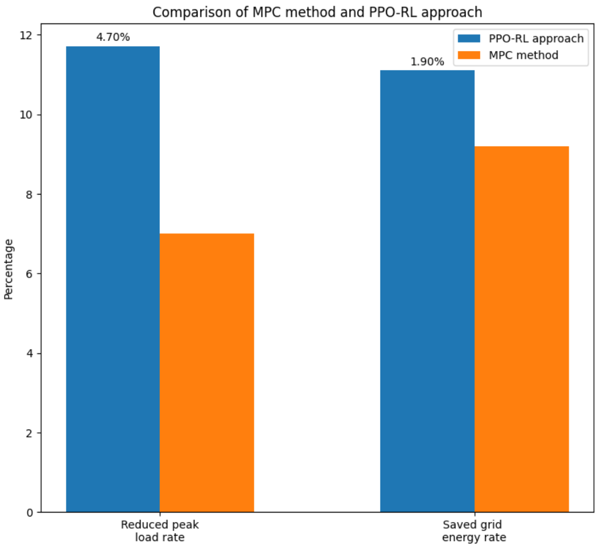

Fig. 4 Comparison of MPC method and RL method

We selected twenty physical areas as the subjects of our study, training Reinforcement Learning networks and running Model Predictive Control algorithms for each. The same time period was chosen for all areas, setting the total training timesteps at 100,000, with each region operating online for the duration of one week to collect data. The tests were conducted in a multi-threaded environment to evaluate the algorithms. After the experiments, we averaged the improvement ratios obtained for each area. Both algorithms showed improvements over the original unregulated charging and discharging states, with the RL algorithm achieving better results in reducing peak loads and overall electricity demand from the grid compared to the MPC algorithm. This superior performance of the RL algorithm shown in Figure 4 can be attributed to the limited foresight of the MPC algorithm. It has more probability to get trapped in local optima.

4. CONCLUSIONS

This paper introduces a deep reinforcement learning-based algorithm designed to optimize energy usage during peak demand times in urban forms. By integrating electric vehicles as flexible energy storage within building energy management systems, the study focuses on reducing peak energy loads to enhance grid stability, cut costs and improve PV self-consumption rate.

The algorithm was tested on the CityEnergyFlow Navigator platform which simulates the city-level mobility model. It effectively shifted energy consumption from peak to off-peak hours, easing the load on the power grid and improving the use of renewable energy. The deep reinforcement learning model's adaptability to real-time system changes allows for better decision-making in dynamic energy environments.

Compared to traditional Model Predictive Control methods, this approach offers a more flexible and efficient solution for complex energy management. The research emphasizes the benefits of combining advanced machine learning with energy management systems to tackle urban energy challenges.

We plan to conduct larger-scale experiments and analyses on the simulation platform of Shenzhen to gain a deeper understanding of energy regulation. Currently, each individual building area is treated as an isolated agent, overlooking the interactions between different agents in practice. In the future, we intend to employ algorithms such as multi-agent reinforcement learning[11] to enable cooperative control among building areas. This collaborative approach aims to achieve energy conservation and emission reduction goals collectively, working together to reduce the peak load across the entire region.

ACKNOWLEDGEMENT

This study is supported by the Tsinghua SIGS and Meituan through the International Collaboration Project; the Tsinghua SIGS-CCSTG Future City Joint Lab funding; Shenzhen Science and Technology Program under Grant KQTD20170810150821146; and Key Research Program of China Southern Power Grid (090000KK52210134).

REFERENCE

[1] Hofman, T., Steinbuch, M., Van Druten, R., & Serrarens, A. (2007). Rule-based energy management strategies for hybrid vehicles. International Journal of Electric and Hybrid Vehicles, 1(1), 71-94.

[2] Jinquan, G., Hongwen, H., Jiankun, P., & Nana, Z. (2019). A novel MPC-based adaptive energy management strategy in plug-in hybrid electric vehicles. Energy, 175, 378-392.

[3] He, H., Wang, Y., Han, R., Han, M., Bai, Y., & Liu, Q. (2021). An improved MPC-based energy management

strategy for hybrid vehicles using V2V and V2I communications. Energy, 225, 120273.

[4] Sarmas, E., Spiliotis, E., Marinakis, V., Tzanes, G., Kaldellis, J. K., & Doukas, H. (2022). ML-based energy management of water pumping systems for the application of peak shaving in small-scale islands. *Sustainable Cities and Society*, *82*, 103873.

[5] Liu, T., Hu, X., Li, S. E., & Cao, D. (2017). Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle. IEEE/ASME transactions on mechatronics, 22(4), 1497-1507.

[6] Pigott, A., Crozier, C., Baker, K., & Nagy, Z. (2022). Gridlearn: Multiagent reinforcement learning for grid-aware building energy management. Electric power systems research, 213, 108521.

[7] Barone, G., Buonomano, A., Calise, F., Forzano, C., & Palombo, A. (2019). Building to vehicle to building concept toward a novel zero energy paradigm: Modelling and case studies. *Renewable and Sustainable Energy Reviews*, *101*, 625-648.

[8] Uddin, M., Romlie, M. F., Abdullah, M. F., Abd Halim, S., & Kwang, T. C. (2018). A review on peak load shaving strategies. *Renewable and Sustainable Energy Reviews*, *82*, 3323-3332.

[9] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

[10] Kouvaritakis, B., & Cannon, M. (2016). Model predictive control. *Switzerland: Springer International Publishing*, *38*, 13-56.

[11] Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, *35*, 24611-24624.