Optimizing Controls of IoT-based Manufacturing Buildings through Deep Reinforcement Learning

Dikai XU[1], Jaewoo SHIN[2], Lan ZHAO[2], Ming QU[1*]

[1] Lyles School of Civil Engineering, Purdue University,
West Lafayette, IN 47907, USA
[2] Rosen Center for Advanced Computing, Purdue University,
West Lafayette, IN 47907, USA
{dikaixu, shin152, lanzhao, mqu}@purdue.edu

* Corresponding Author

ABSTRACT

Maintaining an optimal operating environment within manufacturing facilities is crucial for enhancing energy efficiency, boosting manufacturing productivity, and ensuring occupant comfort and health. With the increasing adoption of Internet of Things (IoT) sensors and cloud-based data acquisition systems in manufacturing facilities, a wealth of IoT operational data is available. This abundance of data empowers data-driven energy analytics and facilitates intelligent control for building operations. Deep Reinforcement Learning (DRL) control is an emerging intelligent control that leverages building big data and artificial intelligence algorithms to optimize operational efficiency and environmental conditions. In this work, operational data is collected and streamed from a manufacturing building equipped with IoT sensors. A customized environment for the DRL is constructed using the operational data. Within the environment, building characteristics and heat transfer process are modeled by a Resistor-Capacitor network. The DRL model is subsequently trained with the Proximal Policy Optimization algorithm to find an optimal control policy. Results show that the DRL building control framework effectively maintains desired indoor conditions in conditioned zones with reduced fluctuation. Moreover, there is a notable decrease in energy consumption, with a demonstrated 33.8% reduction in energy cost savings over a two-month testing period. The implementation of the DRL method also leads to an estimated annual reduction of 4.80 kg/m^2 in carbon emissions for the entire building, therefore contributing to environmental impact mitigation.

1. INTRODUCTION

1.1 Background

Building control is a crucial aspect of modern manufacturing facilities management, encompassing a wide range of systems and technologies to optimize the performance, comfort, and energy efficiency of buildings and the health, productivity, and well-being of occupants. As environmental concerns become more visible to the public view, the importance of effective building control strategies has grown significantly. This involves monitoring and regulating various building systems, including Heating, Ventilation, Air Conditioning (HVAC), and lighting. The emergence of smart building technologies, such as model predictive control and reinforcement learning control, has further enhanced the capabilities of building control systems for multiple objectives, especially for advanced manufacturing buildings. Those buildings leverage interconnected devices and IoT platforms to enable centralized monitoring, remote management, and predictive maintenance. IoT-generated data are analyzed and used to dynamically adjust building systems, thus empowering building operators to proactively identify issues, implement targeted interventions, and continuously improve building performance over time.

For controlling such smart manufacturing facilities, conventional rule-based feedback control, which relies highly on pre-determined schedules, has its limitations in determining the optimal control strategy and optimizing the overall building performance since the operations are inherently dynamic and stochastic (Pinto, Deltetto, & Capozzoli, 2021). Model Predictive Control (MPC), as another thoroughly studied method (Drgoňa et al., 2020), takes predictive information into account. MPC method involves developing complex mathematical models of the building dynamics, designing predictive controllers, and tuning various parameters to achieve desired performance objectives. However, it is difficult to develop a general model for various buildings. Using MPC can be labor-intensive and requires a

high level of expertise. The more accurate the model, the better the manufacturing operation performance. To tackle these challenges, deep reinforcement learning (DRL) has emerged as a promising method. DRL-based method is built from a less sophisticated building model, which requires less effort and expertise in development. It exploits the IoT-generated data and leverages machine learning techniques (Nagy et al., 2023) to extract valuable predictive information in a purely data-driven manner. Compared to traditional methods, DRL-based control shows more robustness and accuracy and stands out as a better solution for multi-objective optimization tasks (Gao, Shi, Miyata, & Akashi, 2024). With the capability of self-tuning and self-learning, they have the potential to be generally used in all similar applications.

1.2 Reinforcement Learning

Basic components of DRL include agent, environment, state, action, and reward. The agent perceives states and rewards from the environment, makes strategic decisions with that information, and executes actions that dynamically affect the environment (Wang & Hong, 2020). In the training period, the agent learns to interact with the environment and update the policy through trial and error. In the context of building control, the states include the running status of building equipment, indoor air conditions (S. Lee & Karava, 2020), occupant behavior (Han, Zhao, Zhang, Shen, & Li, 2021), etc. The action could be changing the HVAC setpoint, altering the schedules, tuning the lighting level, and so on. The reward reflects the goodness of the action taken by the agent. Typically, it is a weighted aggregation of controlled temperature (Coraci, Brandi, Hong, & Capozzoli, 2024), occupants' thermal comfort, and energy consumption or cost (Liu, Wu, & Wu, 2024). These metrics guide the agent to find the optimal policy that has maximized cumulative rewards over time. Researchers have extensively explored the application of various RL algorithms in building control tasks. Notably, algorithms such as Proximal Policy Optimization (PPO) (Nonaka et al., 2023), Soft Actor-Critic (SAC) (Cui, Yap, Prosper, Balaji, & Chen, 2023), Deep Deterministic Policy Gradient (DDPG) (D. Lee, Jeong, & Chae, 2024), and Deep Q-Network (DQN) (Amasyali, Liu, & Zandi, 2024) have demonstrated effectiveness in optimizing building energy management and control systems. Despite the significant advancements in RL-based control strategies for building control and energy management, there remains a gap in exploring the potential of DRL within manufacturing facilities. Hence, this work focuses on manufacturing buildings and how DRL control method can help decision-making processes, ensure the productivity of manufacturing processes, and minimize energy consumption and environmental impact.

Creating a customized DRL environment based solely on real building operational data presents practical challenges because an DRL environment that emulates the building dynamics requires operating the building in various types of conditions. For a building that is already in use, conducting such experimentation will harm the thermal comfort of occupants and interfere with the normal functionality of the building system. Thus, some common choices of DRL environment (Jiménez-Raboso, Campoy-Nieves, Manjavacas-Lucas, Gómez-Romero, & Molina-Solana, 2021; Scharnhorst et al., 2021; Dey & Henze, 2024) are created and configured based on physical models, i.e., EnergyPlus. These physical-model-based environments are building-specific and case-by-case. Setting up the environment requires extensive domain knowledge and skills, and scaling up the building control application can be extremely time-consuming. Thus, in this work, we use a Resistor-Capacitor (RC) network to configure the DRL environment instead (Silvestri, Coraci, Wu, Borkowski, & Schlueter, 2023). RC networks provide a simplified yet effective representation of building dynamics, making them more accessible for DRL environment configuration.

1.3 Contribution of This Work

In this work, we have developed a DRL building control framework that suits the needs of a smart manufacturing facility. With the concern of interfering with normal building operations, we refrained from performing long-term adjustments to HVAC schedules. Instead, we developed an RC network, utilizing streaming data from the building to simulate the environment and operation, and created DRL environment based on it. The agent of the DRL model is trained with PPO algorithm and manages to optimize indoor conditions and maximize energy savings. The main contribution of this work is as follows.

- Design the workflow of using IoT-generated data to facilitate control for advanced manufacturing facilities.
- Provide the solution of using RC model to build and configure DRL environment.
- Develop DRL-driven building control model to reduce energy cost and carbon emissions of manufacturing buildings.
- Demonstrate a real building test case in Indianapolis, IN, to validate the feasibility and potential of the proposed method.

2. METHODOLOGY

2.1 Overview

The structure of the building control framework is shown in Fig.1. Building metadata and IoT-generated data are collected and streamed in real time to a data lake at Purdue University. The RL environment queries and fetches useful features from the data lake using its API. Then the DRL environment queries useful features from the data lake. The DRL environment was built using the base of Gymnasium (Towers et al., 2024), a well-known library with standard APIs for building customized DRL environments. Part of the state variables are generated based on an RC model that lumps the building characteristics. RC model uses heat resistors and heat capacitors to simulate heat transfer processes. Except for the state and action spaces, the environment also contains the reward function, which decides how good an action is. The DRL agent interacts and exchanges information with the environment. Specifically, the agent makes decisions, sends the action predicted by the policy network to the environment, and gets the response back as states and rewards.

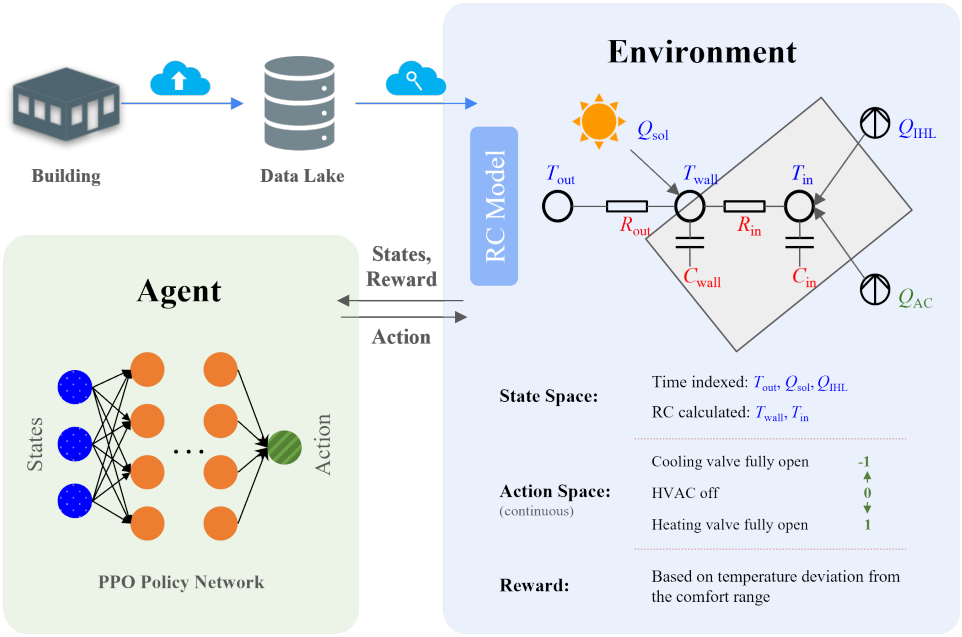

Figure 1: Schematic of the proposed RC-based DRL workflow.

2.2 Environment Configuration

State Space State space comprises two different sets of parameters. The first set of parameters comprises variables that change independently over time, such as outdoor temperature, solar radiation, and the building's operational schedule. These parameters can be directly extracted from a dataset indexed by time. The other set of parameters are those that need to be determined based on the environmental parameters, for example, indoor temperature and building power usage. These parameters are influenced by various factors including weather conditions, building characteristics, occupancy patterns, etc. Manually changing those conditions in a real building and gathering data to build an RL environment is impractical. To tackle this challenge, we use an RC model to simplify and represent building characteristics and heat transfer dynamics. Specifically, we developed a 2R2C model as shown in Fig. 1. There are three nodes in the 2R2C model: T_{out}, T_{wall}, and T_{in}. Heat transfer processes between these nodes are represented by Rs. The conduction and convection from the outdoor environment to the building exterior wall is represented by R_{out}. The wall is also affected by outdoor ambient temperature (T_{out}) and solar radiation (Q_{sol}). Meanwhile, the heat capacity of the exterior wall is represented by C_{wall}. R_{in} represents exterior wall to indoor air heat transfer. The indoor environment is represented by T_{in}, indoor air temperature. Based on the wall temperature node, we can use the following equation to describe the energy balance:

$$C_{wall}\frac{dT_{wall}}{dt} = \frac{T_{out} - T_{wall}}{R_{out}} + \frac{T_{in} - T_{wall}}{R_{in}} + co_{sol} \cdot Q_{sol} \quad (1)$$

where, $\frac{dT}{dt}$ represents the rate of change of temperature with respect to time (t), co_{sol} is the correction coefficient of solar radiation.

The indoor temperature is affected by the heat from the exterior wall, internal heat gain (Q_{IHL}), and input HVAC power (Q_{AC}). Internal heat gain changes with respect to the working schedule and comprises heat generated by the body of the occupants and heat generated by internal equipment and devices, such as manufacturing equipment, IoT gadgets, and computers. Q_{IHL} is estimated by the number of internal equipment and corrected by co_{IHL}. The heat capacity of the total indoor mass is represented by C_{in}. With the above information, the energy balance based on node T_{in} can be described as:

$$C_{in}\frac{dT_{in}}{dt} = \frac{T_{wall} - T_{in}}{R_{in}} + co_{IHL} \cdot Q_{IHL} + Q_{AC} \tag{2}$$

To solve derivative Eqs. (1) and (2), we convert them to the following differential equations.

$$T_{wall}^{t} = T_{wall}^{t-1} + \frac{\Delta t}{C_{wall}}\left(\frac{T_{out} - T_{wall}^{t-1}}{R_{wall}} + \frac{T_{in}^{t-1} - T_{wall}^{t-1}}{R_{in}} + co_{sol} \cdot Q_{sol}\right) \tag{3}$$

$$T_{in}^{t} = T_{in}^{t-1} + \frac{\Delta t}{C_{in}}\left(\frac{T_{wall}^{t-1} - T_{in}^{t-1}}{R_{in}} + co_{IHL} \cdot Q_{IHL} + Q_{AC}\right) \tag{4}$$

where, $t-1$ and t represent previous and current time step, Δt represents the time interval (s) in between.

We can then use curve-fitting solvers to solve these differential equations with initial guesses of R and C values. The solver iteratively adjusts the value until the equations are satisfied. It is important to note that the convergence and accuracy of the solution may depend on the choice of initial values and the precision of the solver.

Action Space Action space is simplified to one continuous action: valve opening percentage (denoted by θ), which spans from -1 to 1. A value of -1 means the cooling valve is fully open, while 1 means the heating valve is fully open. A value of 0 implies that the HVAC system is inactive, neither heating nor cooling the space. Thus, actions are sampled in a closed interval from -1 to 1. When the flow rates of chilled water and hot water circulating in the air handling unit (AHU) are known, Q_{AC} can be calculated using the following equations:

$$Q_{AC} = C_{water} \cdot q_m \cdot \alpha \cdot \theta \tag{5}$$

where, C_{water} is the heat capacity of water, q_m is the mass flow rate of water, α is the water-to-air heat exchange efficiency inside the AHU.

Energy consumption and energy cost serve as key performance indicators of this model. Thus, we also incorporated the calculation of the two parameters with respect to actions within the environment. Given that the building utilizes natural gas to power the boiler and electricity to power the chiller, Eqs.(6) and (7) calculate gas and electricity consumption, respectively. The equations are derived from the energy balance within the AHU.

$$G = \frac{\Delta t}{3600} \cdot C_{water} \cdot q_m^h \cdot (T_{supply}^h - T_{return}^h) \cdot \theta \cdot 0.0736/COP^h \tag{6}$$

$$E = \frac{\Delta t}{3600} \cdot C_{water} \cdot q_m^c \cdot (T_{return}^c - T_{supply}^c) \cdot (-\theta)/COP^c \tag{7}$$

where, G is the gas consumption (ccf), E is the electricity consumption (kWh), T_{supply} and T_{return} are supply water temperature and return water temperature in the AHU, respectively. COP is the coefficient of performance of the boiler and chiller. Superscripts h and c stand for heating and cooling scenarios, respectively. With G and E calculated, we can easily calculate the energy cost with the unit prices of natural gas and electricity, which are \$0.96/ccf and \$0.075/kWh, respectively, in this work.

Reward Function With the purpose of maintaining the indoor temperature in the desired range, we design the reward function to encompass both losses when the temperature deviates from the setpoint and penalties when it falls outside the specified range. For example, considering a setpoint of 23 °C and a deadband of ±1 °C, we compute the losses when the actual indoor temperature is within the range of 22-24 °C and apply penalties when the temperature exceeds this range. Specifically, penalties are incurred when the temperature falls below 22 °C or rises above 24 °C. It is worth noting that during the training process, the calculated reward is normalized to facilitate better convergence.

2.3 DRL Algorithm

The DRL model uses Proximal Policy Optimization (PPO), a policy gradient algorithm, as the agent. As illustrated in Fig. 1, the policy network receives the state variables as input and outputs the next optimal action. The training process starts with initializing the policy network with random weights. Then the agent will interact with the configured environment using the current policy to generate trajectories and collect experiences, namely states, actions, and rewards. The agent will calculate the advantages of the state-action pairs and assess how good or bad it is to take particular actions in a specific state. Once this iteration is completed, the agent will update the policy parameters to optimize the advantages. In each iteration, the model will monitor some key metrics, such as standard deviation, cumulative rewards, and training losses. The training will stop when the policy converges or when a stopping criterion is met. One advantage of PPO compared to other RL algorithms is that PPO aims to find policy updates that improve performance while ensuring that the policy changes are not too drastic, which ensures stability and reliability.

3. BUILDING CHARACTERISTICS

3.1 Building Description

The studied case in this work is a smart manufacturing building with an area of 5575 m^2. The operation of this building is automated according to predefined schedules. However, the building adopts IoT technology in all its equipment, including HVAC system and weather station. The IoT-generated data enables not only building status monitoring, but also data-driven applications like DRL-based building controls discussed in this work. The data is collected and streamed to a streaming data management and processing platform built based on the open-source StreamCI system (Shin et al., 2022) and Anvil high performance computing resources (Song et al., 2022). The data lake platform enables researchers to easily access, process, and connect IoT data into ML/AI modeling workflows. As a comparison baseline, the electricity usage of the building HVAC system from November 2022 to November 2023 is 44051 kWh, while the natural gas usage is 5009 ccf. Annual carbon emission is estimated to be 10.52 kg per conditioned area (US EPA, 2015).

3.2 Weather Conditions

The investigated building is located in Indianapolis, IN. The city has a hot-summer humid continental climate. Typical meteorological year weather data (Linda K & Drury B, 2022) is shown in Fig. 2. Based on the temperature and relative humidity profile, we divided the entire year into three seasons, namely heating season, cooling season, and shoulder season. The heating season spans from November to March the following year, while the cooling season spans from June to August. The remaining months are classified as shoulder season, representing transitional periods between heating and cooling seasons.

3.3 Building RC Model

As described in the environment configuration, the RC model is processed with a curve-fitting solver in Python environment in order to find the optimal values of Rs and Cs to match the simulated curve to the actual measured one. We chose to demonstrate the feasibility of this model during the shoulder season, a period characterized by transitions between heating and cooling modes in the HVAC system. This choice provides a diverse range of temperature variations, allowing for thorough testing and validation of the model's ability to adapt to changing environmental conditions.

With the fitted values, we can then simulate the indoor air temperature and wall temperature with Eqs. (3) and (4). The simulated results are shown in Fig. 3. We can see that the simulated curve aligns well with the actual measured data with less stochastic noise. The fitted results have a root mean squared error (RMSE) of 0.48, which is acceptable in a lumped 2R2C model. Within the testing period, there is some missing data. The RC model manages to mimic the operation pattern and predict the indoor temperature trend even in the presence of missing data. This ability is essential in the DRL environment to emulate the response of building HVAC system that represents real-world scenarios.

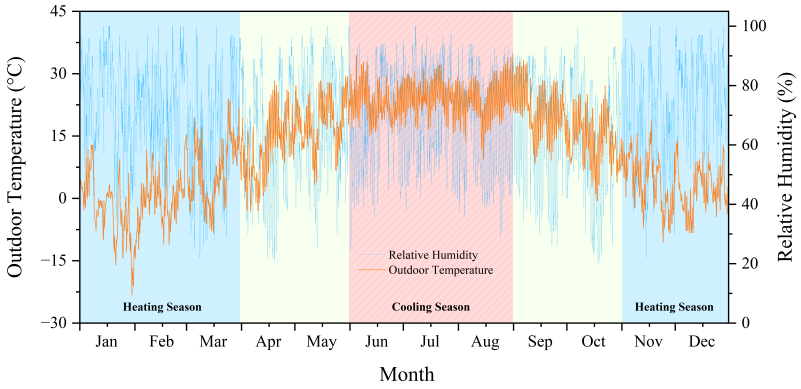

Figure 2: Typical meteorological year weather data of Indianapolis, IN.

Table 1: RC curve fit results

	Initial Value	**Fitted Value**
R_{out}	1.1274e-02	1.2086e-02
R_{in}	7.5039e-01	7.5102e-01
C_{wall}	1.2478e+05	1.2479e+05
C_{in}	1.2032e+06	1.2033e+06
co_{sol}	9.2674e-02	8.8053e-02
co_{IHL}	3.6320e-01	3.5977e-01

4. RESULTS AND DISCUSSION

We utilized Stable-Baseline3 (Raffin et al., 2021), a widely adopted library for implementing DRL algorithms, to develop and train our DRL model. This comprehensive solution offers robust functionalities, including model validation, training monitoring, and customizable callbacks. For the training process, we selected standard hyperparameters commonly used in DRL research. The learning rate was set to 0.0001, while the discount factor was set to 0.99. Additionally, we chose a value function coefficient of 0.99 to stabilize the training process by emphasizing the value function in loss calculation. To achieve convergence, we conducted training over 150 episodes.

With the configured RC-based DRL environment and GPU computing resources, we trained the DRL model in the time period from April 1st to May 30th (part of the shoulder season). During this period, the cumulative electricity usage of the building HVAC system was 5647 kWh, and the gas totaled 764 ccf. The total energy cost reached $1157. Fig. 4 shows part of the training results. We picked three periods to take a close look at the DRL control effectiveness. In the figure, the setpoint is 23°C (red dash) and the deadband is ±1°C (gray dash). the comfort range is shaded in light gray.

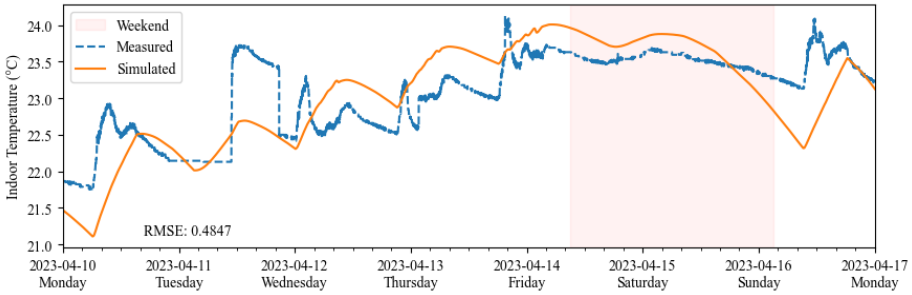

Figure 3: Simulated indoor temperature based on curve-fitted RC.

Along the time axis, the weekend spans are painted in light red, during which the HVAC system is not working.

During the heating-dominant period, we can see that the outdoor temperature is relatively low, which requires the HVAC system to run mostly in heating mode to keep a comfortable indoor built environment. During the working hours from 7:00 AM to 5:30 PM, the DRL control agent manages to maintain the indoor temperature in the gray range. Starting from 5:30 PM on Friday, the HVAC system shuts down and the indoor temperature changes according to outdoor conditions with thermal inertia. Once the time goes to 7:00 AM on Monday, the HVAC system runs at full capacity and tries to get the temperature back to the comfort range quickly. During the cooling-dominant period, the outdoor temperature along with the solar radiation brings excessive heat to the indoor environment. Similarly, the DRL agent manages to maintain the temperature within the comfort range. However, unlike the situation on April 10th where the temperature far deviates from the setpoint, the temperatures are mostly still in the comfort range. Instead of receiving a large penalty, the environment only returns a small loss. Thus, the HVAC system does not run at full cooling capacity. Due to this, we can see from the action histogram (Fig. 5) that most actions are minor, while only a few actions fall in the ±1 bins, leading to less temperature fluctuation than the current deployed control strategy. In the period where the actual measured data is missing (Fig. 4(c)), the DRL agent still manages to keep the indoor temperature within the comfort range during working hours, which proves the feasibility of robustness of using RC-based environment to simulate the building response.

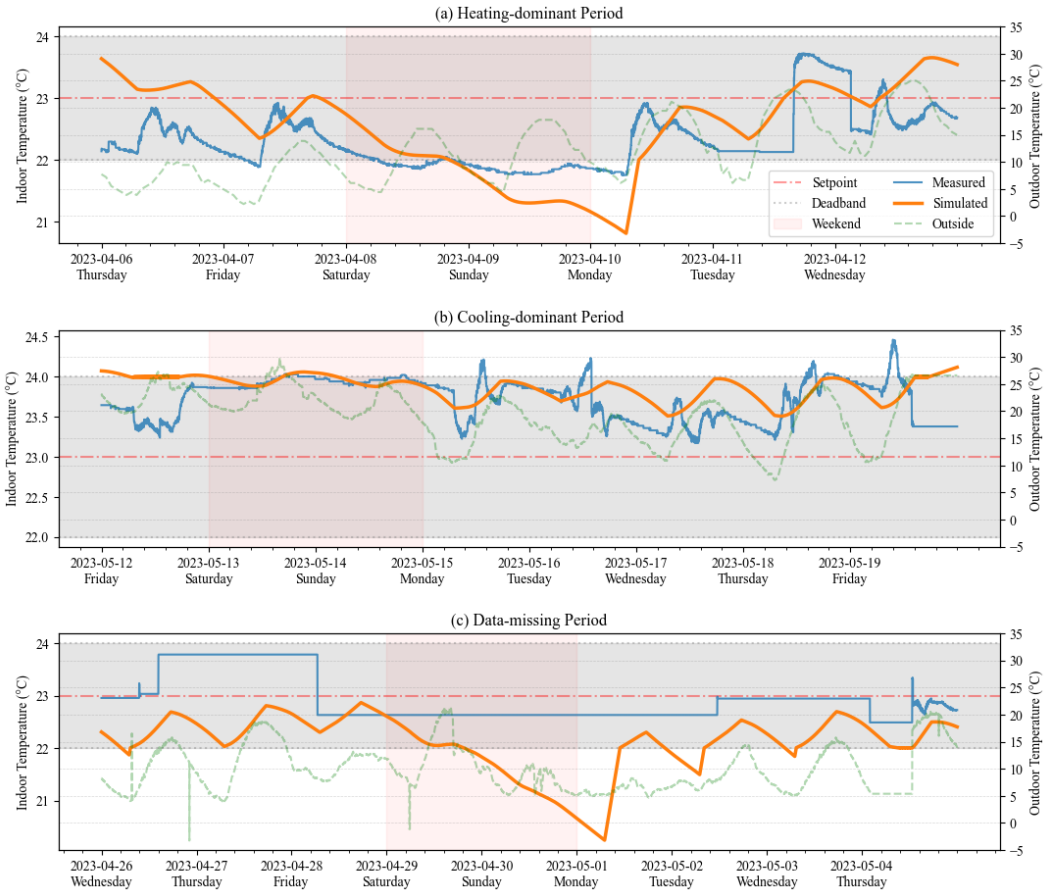

Figure 4: DRL controlled indoor air temperature in three different scenarios.

The corresponding energy consumption and energy cost is also given by the DRL environment as part of the results. With the DRL control model, the electricity consumption is significantly reduced in the shoulder season, especially in the cooling scenario. The main reason is that the DRL controller maintains the temperature in the comfort range with less fluctuation compared to rule-based control. The total cost accumulates to $766 in the two months period, while the actual cost is $1157, leading to a 33.8% of cost saving. The heat loss in the HVAC system is not considered when using Eqs. (6) and (7). Also, the actual building operation is controlled by several dynamic setpoints (thermostats manip-

ulated by occupants) rather than one. However, we can still see the prominent energy-saving potential of using DRL in building controls. Based on the simulation results, the estimated carbon emission reduction during the simulation period is 0.80 kg/m^2, and the annual carbon emission reduction is estimated to be 4.80 kg/m^2.

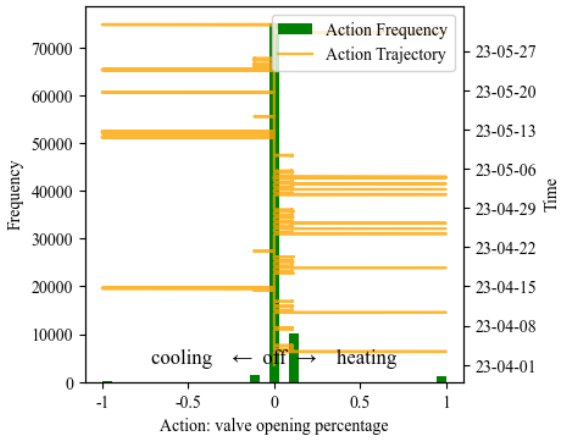

Figure 5: Actions over time and histogram in the control period.

As part of our ongoing efforts to advance the DRL-based control framework implemented in this work, we recognize the need for further improvement. First of all, the RC network in this work is simplified to 2R2C. It would be too lumped if we wanted to maintain different air conditions in different conditioned zones of a building. Therefore, in future iterations, we plan to include additional components to the RC-based DRL environment. A more comprehensive environment will enable the DRL agent to make more informed decisions and optimize a broader range of building systems. Secondly, the rewards function can also be refined to facilitate multi-objective optimization. For example, energy consumption and various conditions (temperature, humidity, and air quality) in different thermal zones. Moreover, while the current framework captures operational dynamics, incorporating more seasonal variation is essential for a robust performance across different weather conditions. Preferably, more years of data will be included for incremental training. This will improve the adaptability and efficiency of the control system under varying environmental conditions. Also, continuous improvement and optimization of the DRL algorithms and policies based on feedback from real-life deployment will be an ongoing focus. Our ultimate goal is to integrate the DRL-based control system into the manufacturing building management system seamlessly. By integrating with existing building automation and control infrastructure, the DRL agent can interact directly with building systems, making real-time adjustments based on environmental conditions, occupant behaviors, and energy demands. This integration will enable the DRL framework to become an integral part of the building management system, contributing to creating a comfortable, economical, and sustainable built environment.

5. CONCLUSIONS

In conclusion, this work addresses the critical need for maintaining optimal indoor environments in manufacturing facilities to ensure both productivity and occupants' well-being. Considering the growing adoption of IoT technologies in such facilities, this work utilizes operational data generated by IoT devices to facilitate data-driven building control workflow. A customized DRL environment is constructed using operational data collected from a real manufacturing building. The environment is modeled using an RC network, capturing building characteristics and heat transfer processes. PPO algorithm is then employed to train the DRL model and find an optimal control policy. Results of the test case demonstrate the efficacy of the proposed DRL framework in maintaining desired indoor conditions with reduced fluctuation, leading to 33.8% of energy cost savings. And the building annual carbon emission reduction is estimated to be 4.80 kg per conditioned area. The proposed DRL building control framework proves its potential in optimizing building operations, improving energy efficiency, and mitigating environmental impact in manufacturing facilities.

NOMENCLATURE

C	specific heat capacity	(J/kg K)
COP	coefficient of performance	(-)
co	correction coefficient	(-)
d	derivative	(-)
Q	heat flow rate	(W)
q	water flow rate	(kg/s)
R	heat resistance	(K/W)
T	temperature	(°C)
t	time	(-)
α	heat transfer efficiency	(%)
Δt	time interval	(s)
θ	valve opening percentage	(-)

Subscript / Superscript

AC	air conditioning
c	cooling
h	heating
IHL	internal heat load
in	indoor
m	mass
out	outdoor
return	return water
sol	solar
supply	supply water
wall	exterior wall
water	hot or chilled water

REFERENCES

Amasyali, K., Liu, Y., & Zandi, H. (2024, February). A Transfer Learning Strategy for Improving the Data Efficiency of Deep Reinforcement Learning Control in Smart Buildings. In *2024 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)* (pp. 1–5). (ISSN: 2472-8152) doi: 10.1109/ISGT59692.2024.10454120

Coraci, D., Brandi, S., Hong, T., & Capozzoli, A. (2024, February). An innovative heterogeneous transfer learning framework to enhance the scalability of deep reinforcement learning controllers in buildings with integrated energy systems. *Building Simulation*. doi: 10.1007/s12273-024-1109-6

Cui, J., Yap, W. Y., Prosper, C., Balaji, B., & Chen, J. (2023, November). Economizer Optimization with Reinforcement Learning: An Industry Perspective. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 366–369). Istanbul Turkey: ACM. doi: 10.1145/3600100.3625685

Dey, S., & Henze, G. P. (2024, March). Reinforcement Learning Building Control: An Online Approach With Guided Exploration Using Surrogate Models. *ASME Journal of Engineering for Sustainable Buildings and Cities*, *5*(011005). doi: 10.1115/1.4064842

Drgoňa, J., Arroyo, J., Cupeiro Figueroa, I., Blum, D., Arendt, K., Kim, D., … Helsen, L. (2020, January). All you need to know about model predictive control for buildings. *Annual Reviews in Control*, *50*, 190–232. doi: 10.1016/j.arcontrol.2020.09.001

Gao, Y., Shi, S., Miyata, S., & Akashi, Y. (2024, March). Successful application of predictive information in deep reinforcement learning control: A case study based on an office building HVAC system. *Energy*, *291*, 130344. doi: 10.1016/j.energy.2024.130344

Han, M., Zhao, J., Zhang, X., Shen, J., & Li, Y. (2021, April). The reinforcement learning method for occupant behavior in building control: A review. *Energy and Built Environment*, *2*(2), 137–148. doi: 10.1016/j.enbenv.2020.08.005

Jiménez-Raboso, J., Campoy-Nieves, A., Manjavacas-Lucas, A., Gómez-Romero, J., & Molina-Solana, M. (2021).

Sinergym: A Building Simulation and Control Framework for Training Reinforcement Learning Agents. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 319–323). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3486611.3488729

Lee, D., Jeong, J., & Chae, Y. T. (2024, January). Application of Deep Reinforcement Learning for Proportional–Integral–Derivative Controller Tuning on Air Handling Unit System in Existing Commercial Building. *Buildings*, *14*(1), 66. doi: 10.3390/buildings14010066

Lee, S., & Karava, P. (2020, October). Towards smart buildings with self-tuned indoor thermal environments – A critical review. *Energy and Buildings*, *224*, 110172. doi: 10.1016/j.enbuild.2020.110172

Linda K, L., & Drury B, C. (2022). *Development of Global Typical Meteorological Years (TMYx).* Retrieved from `http://climate.onebuilding.org`

Liu, X., Wu, Y., & Wu, H. (2024, January). Enhancing HVAC energy management through multi-zone occupant-centric approach: A multi-agent deep reinforcement learning solution. *Energy and Buildings*, *303*, 113770. doi: 10.1016/j.enbuild.2023.113770

Nagy, Z., Henze, G., Dey, S., Arroyo, J., Helsen, L., Zhang, X., … Bernstein, A. (2023, August). Ten questions concerning reinforcement learning for building energy management. *Building and Environment*, *241*, 110435. doi: 10.1016/j.buildenv.2023.110435

Nonaka, S., Taniguchi, I., Nishikawa, H., Zhao, D., Catthoor, F., & Onoye, T. (2023, November). Comfort-aware HVAC Aggregation Method based on Deep Reinforcement Learning. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 290–291). Istanbul Turkey: ACM. doi: 10.1145/3600100.3626265

Pinto, G., Deltetto, D., & Capozzoli, A. (2021, December). Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, *304*, 117642. doi: 10.1016/j.apenergy.2021.117642

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, *22*(268), 1–8. Retrieved from `http://jmlr.org/papers/v22/20-1364.html`

Scharnhorst, P., Schubnel, B., Fernández Bandera, C., Salom, J., Taddeo, P., Boegli, M., … Politi, C. (2021, January). Energym: A Building Model Library for Controller Benchmarking. *Applied Sciences*, *11*(8), 3518. (Number: 8 Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/app11083518

Shin, J., Zhao, L., Song, C. X., Kalyanam, R., Jin, J., Hosen, J. D., … Xu, D. (2022, October). *Enabling Scalable and Reliable Real Time Data Services for Sensors and Devices in StreamCI.* Retrieved from `https://zenodo.org/records/7186972`

Silvestri, A., Coraci, D., Wu, D., Borkowski, E., & Schlueter, A. (2023, November). Comparison of two deep reinforcement learning algorithms towards an optimal policy for smart building thermal control. *Journal of Physics: Conference Series*, *2600*(7), 072011. (Publisher: IOP Publishing) doi: 10.1088/1742-6596/2600/7/072011

Song, X. C., Smith, P., Kalyanam, R., Zhu, X., Adams, E., Colby, K., … St. John, J. (2022, July). Anvil - System Architecture and Experiences from Deployment and Early User Operations. In *Practice and Experience in Advanced Research Computing* (pp. 1–9). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3491418.3530766

Towers, M., Terry, J. K., Kwiatkowski, A., Balis, J. U., Cola, G., Deleu, T., … Younis, O. G. (2024, February). *Gymnasium.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.10655021`

US EPA, O. (2015, August). *Greenhouse Gases Equivalencies Calculator - Calculations and References* [Data and Tools]. Retrieved from `https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references`

Wang, Z., & Hong, T. (2020, July). Reinforcement learning for building controls: The opportunities and challenges. *Applied Energy*, *269*, 115036. doi: 10.1016/j.apenergy.2020.115036

ACKNOWLEDGMENT

This work was supported in part by a grant from the Central Indiana Corporate Partnership (CICP) under the AnalytiXIN Program.