

# A Lightweight Calibrated Simulation Enabling Efficient Offline Learning for Optimal Control of Real Buildings

Judah Goldfeder  
Google, Columbia University  
United States  
judahg@google.com

John Sipple  
Google  
United States  
sipple@google.com

## ABSTRACT

Modern commercial Heating, Ventilation, and Air Conditioning (HVAC) devices form a complex and interconnected thermodynamic system with the building and outside weather conditions, and current setpoint control policies are not fully optimized for minimizing energy use and carbon emission. Given a suitable training environment, a Reinforcement Learning (RL) model is able to improve upon these policies, but training such a model, especially in a way that scales to thousands of buildings, presents many real world challenges. We propose a novel simulation-based approach, where a customized simulator is used to train the agent for each building. Our open-source simulator<sup>1</sup> is lightweight and calibrated via telemetry from the building to reach a higher level of fidelity. On a two-story, 68,000 square foot building, with 127 devices, we were able to calibrate our simulator to have just over half a degree of drift from the real world over a six-hour interval. This approach is an important step toward having a real-world RL control system that can be scaled to many buildings, allowing for greater efficiency and resulting in reduced energy consumption and carbon emissions.

## KEYWORDS

HVAC Optimization, Simulation, Reinforcement Learning

### ACM Reference Format:

Judah Goldfeder and John Sipple. 2023. A Lightweight Calibrated Simulation Enabling Efficient Offline Learning for Optimal Control of Real Buildings. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23)*, November 15–16, 2023, Istanbul, Turkey. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3600100.3625682>

## 1 INTRODUCTION

Energy optimization and management in commercial buildings is a very important problem, whose importance is only growing with time. Buildings account for 39% of all US carbon emissions [13]. Reducing those emissions by even a small percentage can have a significant effect, especially in more extreme climates. Our contributions include a highly customizable and scalable HVAC and building simulator, a rapid configuration method to customize the

simulator to a particular building, a calibration method to improve this fidelity using real world data, and an evaluation method to measure the simulator fidelity.

**Overview of the HVAC problem** Most office buildings are equipped with advanced HVAC devices, like Variable Air Volume (VAV) devices, Hot Water Systems (HWS), Air Conditioner (AC) and Air Handlers that are configured and tuned by the engineers, manufacturers, installers, and operators to run efficiently with the device’s local control loops [15]. However, integrating multiple HVAC devices from diverse vendors into a building “system” requires technicians to program fixed operating conditions for these units, which may not be optimal for every building and every potential weather condition. If existing setpoint control policies are not optimal under all conditions, the possibility exists that an ML model may be trained to continuously tune a small number of setpoints to achieve greater energy efficiency and reduced carbon emission [22].

**RL for HVAC systems** We define the state of the office building  $S_t$  at time  $t$  as a fixed length vector of measurements from sensors on the building’s devices, such as a specific VAV’s zone air temperature, gas meter’s flow rate, etc. The action on the building  $A_t$  is a fixed-length vector of device setpoints selected by the agent at time  $t$ , such as the boiler’s supply water temperature setpoint, etc. We also define a custom feedback signal, or reward, that indicates the quality of taking an action in the current state, as a weighted sum of negative cost functions for carbon emission, energy consumption, and zone-level setpoint deviation. The goal of RL is to maximize the expected long-term cumulative reward [18]. In order to converge to the optimal policy, the agent requires many training iterations, making training directly on the building from scratch inefficient and impracticable. Therefore, it is necessary to train in an offline sandbox environment that adequately emulates the dynamics of the building before being deployed on the actual building.

## 2 RELATED WORK

Considerable attention has been paid to HVAC control in recent years, and a growing portion of that has considered how RL and its various associated algorithms can be leveraged [8, 14, 20, 23, 24]. A central requirement in RL is the offline environment that trains the RL agent. Several methods for this have been proposed, largely falling under three broad categories.

**Data-driven Emulators:** Some works attempt to learn a dynamics model as a multivariate regression model from real world data [25, 29], often using recurrent neural network architectures, such as LSTMs [16, 19, 28]. The difficulty here is that data-driven models often do not generalize well to circumstances outside the training distribution, especially since they are not physics based.

<sup>1</sup>Available online: <https://github.com/google/sbsim>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*BuildSys '23, November 15–16, 2023, Istanbul, Turkey*

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0230-3/23/11.

<https://doi.org/10.1145/3600100.3625682>

**Physics-based Simulation:** EnergyPlus [7], a high-fidelity simulator developed by the Department of Energy, is commonly used [2, 21, 22, 27], but suffers from the scalability issues outlined above.

**Offline RL:** The second approach is to train the agent directly from the historical real world data, without ever producing an interactive environment [4–6]. While the real world data is obviously of high accuracy and quality, this presents a major challenge, since the agent cannot take actions in the real world and interact with any form of an environment, severely limiting its ability to improve over the baseline policy that produced the real world data [10].

To overcome the limitations of each of the above three methods, some work has proposed a hybrid approach [3, 26], and this is the category our work falls under. Our approach uses a physics-based simulator that achieves an ideal balance between speed and fidelity, which is sufficient to train an effective control agent off-line.

### 3 A SIMPLE CALIBRATED SIMULATION

**Design Considerations** A fundamental tradeoff when designing a simulator is speed versus fidelity. Fidelity is the simulator’s ability to reproduce the building’s true dynamics that affect the optimization process. Speed refers to both simulator configuration time, and agent training time, i.e., the time necessary for the agent to optimize its policy using the simulator. Every building is unique, due to its physical layout, equipment, and location. Fully customizing a high fidelity simulation to a specific target building requires nearly exhaustive knowledge of the building structure, materials, location, etc., some of which is unknowable. Also, the configuration time required for high-fidelity simulations limits their utility for deploying RL-based optimization to many buildings. Our goal is to develop a method for applying RL at scale to commercial buildings. To this end, we must have a simulated environment to train the agent, as real world training is not possible. In order to scale to many buildings it must be easy for our simulator to be configured to a new building, with enough fidelity to the real world to be useful. We note that while RL is the control method proposed in this work, our calibrated simulator is useful as well for other controller methods such as Model Predictive Control [9]. To meet these requirements we designed a lightweight simulator based on Finite Differences approximation of heat exchange, with a novel automated procedure to go from building floor plans to a custom simulator with little manual effort. For improved fidelity, we designed a calibration and evaluation pipeline based on real telemetry.

**Building Thermal Model** We developed our simulation around our abstract thermal model for office buildings, shown in Figure 1. A building consists of conditioned zones, where the measured zone temperature,  $T_z$ , should be within upper and lower setpoints,  $\hat{T}_{z,max}$  and  $\hat{T}_{z,min}$ . Thermal power for heating or cooling is supplied to each zone,  $\dot{Q}_s$ , and recirculated from the zone,  $\dot{Q}_r$  from the HVAC system, with additional thermal exchange  $\dot{Q}_z$  from walls, doors, etc. The Air Handler supplies the building with air at supply air temperature setpoint  $\hat{T}_s$  drawing fresh air,  $\dot{m}_{amb}$  at ambient temperatures,  $T_{amb}$  and returning exhaust air  $\dot{m}_{exhaust}$  at temperature  $T_{exhaust}$  to the outside using intake and exhaust fans:  $\dot{W}_{a,in}$  and  $\dot{W}_{a,out}$ . Some of the return air can be recirculated,  $\dot{m}_{recirc}$ . Central air conditioning is achieved with a chiller and pump that joins a refrigeration cycle to the supply air, consuming electrical energy for the AC compressor

$\dot{W}_c$  and coolant circulation,  $\dot{W}_{c,p}$ . The hot water cycle consists of a boiler that maintains the supply water temperature at  $T_b$  heated by natural gas power  $\dot{Q}_b$ , and a pump that circulates hot water through the building, with electrical power  $\dot{W}_{b,p}$ . Supply air is delivered to the zones through VAV devices.

We selected **water supply temperature**  $\hat{T}_b$  and the **air handler supply temperature**  $\hat{T}_s$  as agent actions because they affect the balance of electricity and natural gas consumption, they affect multiple device interactions, and they affect occupant comfort.

**Finite Differences Approximation** The diffusion of thermal energy in time and space of the building can be approximated using the method of Finite Differences (FD)[12, 17], and applying an energy balance. This method divides each floor of the building into a grid of three-dimensional control volumes and applies thermal diffusion equations to estimate the temperature of each control volume. By assuming each floor is thermally isolated, (i.e., no heat is transferred between floors), we can simplify the three-spatial dimensions into a spatial two-dimensional heat transfer problem. Each control volume is a narrow volume bounded horizontally, parameterized by  $\Delta x^2$ , and vertically by the height of the floor. The energy balance, shown below, is applied to each discrete control volume in the FD grid, and consists of the following components: (a) the thermal exchange across each face of the four participating faces control volume via conduction or convection  $Q_1, Q_2, Q_3, Q_4$ , (b) the change in internal energy over time in the control volume  $Mc \frac{\Delta T}{\Delta t}$ , and (c) an external energy source that enables applying local thermal energy from the HVAC model only for those control volumes that include an airflow diffuser,  $Q_{ext}$ . The equation is  $Q_{ext} + Q_1 + Q_2 + Q_3 + Q_4 = Mc \frac{\Delta T}{\Delta t}$ , where  $M$  is the mass and  $c$  is the heat capacity of the control volume,  $\Delta T$  is the temperature change from the prior timestep and  $\Delta t$  is the timestep interval.

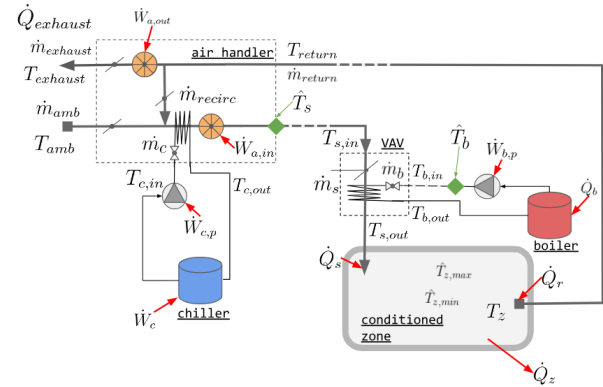


Figure 1: Building Thermal Model

The thermal exchange in (a) is calculated using Fourier’s law of steady conduction in the interior control volumes (walls and interior air), parameterized by the conductivity of the volume, and the exchange across the exterior faces of control volumes is calculated using the forced convection equation, parameterized by the convection coefficient, which approximates winds and currents surrounding the building. The change in internal energy (b) is parameterized by the density, and heat capacity of the control volume.

Finally, the thermal energy associated with the VAV (c) is equally distributed to all associated control volumes that have a diffuser.

Thermal diffusion within the building is mainly accomplished via forced or natural convection currents, which can be notoriously difficult to estimate accurately. We note that heat transfer using air circulation is effectively the exchange of air mass between control volumes, which we approximate by a randomized shuffling of air within zones, parameterized by a shuffle probability.

**HVAC Model** The HVAC system is modeled as an energy balance with a lossy thermal exchange between the hot water and air circuits, evaluating temperature differences, and air and water flow rates with efficiency parameters. Aggregate heating and cooling demands at the AC and HWS are governed by thermostats that create demand when its zone temperature is near zone air temperature setpoints. Special control volumes associated with air diffusers allow thermal energy exchange between the HVAC model and the FD grid,  $\dot{Q}_s$ .

**Simulator Configuration** For RL to scale to many buildings, it is critical to be able to easily and rapidly configure the simulator to any arbitrary building. To configure the simulator, we require floorplans, and HVAC metadata. We preprocess the detailed floorplan blueprints of the building, and extract a grid that gives us an approximate placement of walls. We also employ a user interface to label the location of each HVAC device on the floorplan grid. We tested this pipeline on our pilot building, which consisted of two floors with combined surface area of 68,000 ft<sup>2</sup>, and 127 HVAC devices. Given floorplans and HVAC layout information, a single technician was able to generate a fully specified simulation in under 3 hours that matched the real building in every device and room.

**Simulator Calibration and Evaluation** In order to calibrate the simulator to the real world using data, we must have a metric with which to evaluate our simulator, and an optimization method to improve our simulator on this metric. We proposed a novel evaluation procedure, based on  $N$ -step prediction. Each iteration step of our simulator was designed to represent a five minute interval, and our real world data is also sampled in five minute intervals. To evaluate the simulator, we take a chunk of real data of  $N$  observations. We then initialize the simulator so that its initial state matches that of the starting observation, and run the simulator for  $N$  steps, replaying the same HVAC policy as was used in the real world observations. Next, we calculate our simulation fidelity metric, which is the mean absolute error of the measured zone air temperatures, at the  $N$ th timestep. More formally, we define the spatial Mean Absolute Error (MAE) of  $Z$  zones at timestep  $t$  as  $\epsilon_t = \frac{1}{Z} \sum_{z=1}^Z |T_{real,t,z} - T_{sim,t,z}|$ , where  $T_{real,t,z}$  is the measured zone air temperature for zone  $z$  at timestamp  $t$ , and  $T_{sim,t,z} = \frac{1}{C_z} \sum_{c=1}^{C_z} T_{t,c}$  is the mean temperature of all control volumes  $C_z$  in zone  $z$  at time  $t$ . Thus, to evaluate the simulator on  $N$ -step prediction, we run the simulator for timesteps 0 to  $N - 1$ , and calculate the above metric for  $t = N - 1$ .

**Simulator Calibration** Once we defined our simulation fidelity metric, we can minimize the error by searching over the physical parameters that affect the simulation response. The variables tuned during the parameter search included the following: (a) the forced convection coefficient quantifying outside wind and air currents against the building exterior surfaces; (b) the thermal conductivity,

heat capacity, and density of exterior and interior walls; and (c) the shuffle probability that controls how likely an air volume will be exchanged with another within the zone to approximate the internal air circulation and interior forced convection.

## 4 EXPERIMENT RESULTS

We now demonstrate the results of how our simulator, when tuned and calibrated, is able to make real-world predictions.

**Experiment Setup** To test out our simulator, we obtained telemetry recordings from our pilot building, a commercial office building located in Mountain View, California. The building has two stories with a combined surface area of 68,000 square feet, and has 127 HVAC devices. We obtained floor plan blueprints and used them to configure a customized simulator for the building, a process that took a single human less than three hours to complete.

**Calibration Data** To tune and evaluate our simulator, we took three, non-overlapping intervals of telemetry: (a) Thursday, July 6 2023, from 1:40 AM to 7:40 AM ( $N = 72$ ), (b) Friday, July 7 2023, from 1:40 AM to 7:40 AM ( $N = 72$ ), and (c) from Friday, July 7 2023, from 11:40 AM to 5:40 PM ( $N = 72$ ). The first telemetry interval, (a), was used to tune the simulator for calibration. The second and third telemetry intervals, (b) and (c), were used for validation.

**Table 1: Hyperparameter ranges and chosen values**

Hyperparameter	range	best
exterior convection coefficient ( $W/m^2/K$ )	5 - 800	800
exterior wall conductivity ( $W/m/K$ )	0.01 - 1	0.01
exterior wall density ( $kg/m^3$ )	1 - 3000	2748
exterior wall heat capacity ( $J/Kg/K$ )	100 - 2500	2500
interior wall conductivity ( $W/m/K$ )	5 - 800	780
interior wall density ( $kg/m^3$ )	0.5 - 1500	0.5
interior wall heat capacity ( $J/Kg/K$ )	500 - 1500	500
interior air volume shuffle probability	0 - 1	1

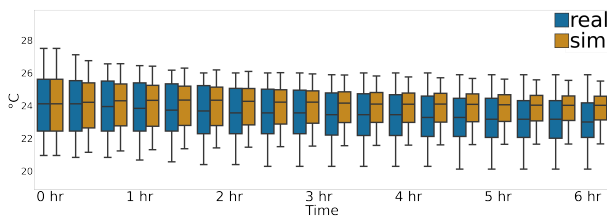
**Calibration Procedure** We tuned for 100 iterations on our simulator. The parameters varied, the sweep range limits, and the values found that minimized the calibration metric are shown in Table 1. Because of our focus on speed, we can simulate a five minute interval in about 15 seconds on one CPU, allowing for very rapid tuning. We reviewed the parameters that yielded the lowest simulation error from calibration. Densities, heat capacities, and conductivities plausibly matched common interior and exterior building materials. However, the external convection coefficient was higher than under normal weather conditions, and likely is compensating for radiative losses and gains, which were not directly simulated.

**Calibration Results** Table 2 shows the  $N$ -step prediction fidelity, as MAE, over a six-hour prediction window ( $N = 72$ ). We calculated the spatial mean absolute temperature error, as defined above. We also present a second metric, the median spatial temperature error. This was not used in the tuning process, but gives us some insight into how well the calibration process is performing. As indicated in Table 2, our tuning procedure drifts only 0.6° on average over a 6 hour period on the tuning set, and we get good

generalization: almost identical error for the same time period of a different day. It should be noted that an uncalibrated model had a much larger error of  $2.1^\circ$ . Furthermore, the median temperature in these two cases is near zero, indicating that even despite drift, our prediction remains a mostly unbiased estimator. For the different time of day scenario, we observe a larger drift of  $1.2^\circ$ , and the median is shifted as well, indicated that tuning on one time of day is not as useful when making predictions on another.

**Table 2: Mean and Median Absolute Error on  $N = 72$  prediction window**

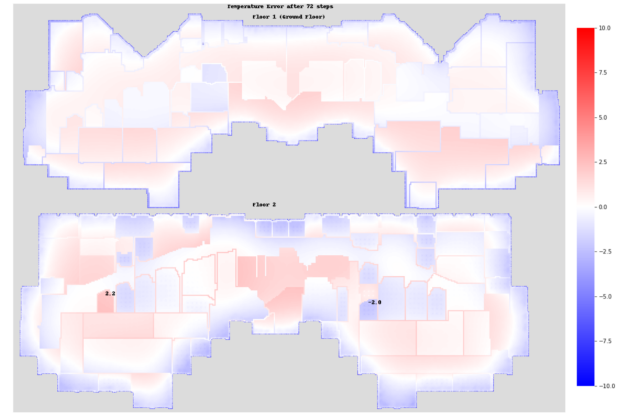
Metric	Tuning Data (a)	Validation (b)	Validation (c)
MAE	$0.64^\circ\text{C}$	$0.63^\circ\text{C}$	$1.18^\circ\text{C}$
Median	$0.01^\circ\text{C}$	$0.18^\circ\text{C}$	$0.98^\circ\text{C}$



**Figure 2: Temperature Drift Over 6 hours, from validation (b).**

**Visualizing Temperature Drift Over Time** Figure 2 illustrates temperature drift over time. At each time step, we plot the real world temperatures in a blue boxplot and the simulator ones in orange.

**Visualizing Spatial Errors** Figure 3 is a heatmap of the spatial temperature difference across both floors after six hours. Red regions indicate the simulator was warmer than the validation data, blue regions indicate cooler, and white regions had no temperature difference. The ring of blue around the building indicates that our simulator is too cold on the perimeter, which implies that the heat exchange with the outside is happening more rapidly than it would in the real world. The inside of the building remains red, which means that despite the simulator perimeter being cooler than the real world, the inside is warmer. The white band region running along the perimeter is where the temperature achieved a perfect fit with the validation data. The interior red regions are due to slight errors in modeling interior heat exchange, and the blue regions indicate errors due to high thermal losses with the outside. We suspect these errors are largely due to assuming thermal isolation between floors and assuming no radiative exchange with the outside. Consequently, the calibration process compensates with a high convection coefficient. Internal thermal exchange seems slower than the real world, and the calibration compensates with more rapid outside thermal exchange.



**Figure 3: Heatmap representing the temperature difference between the simulator and the real world after 6 hours, with red indicating the simulator is hotter, and blue colder.**

## 5 DISCUSSION AND FUTURE WORK

Developing a simulator is an important step toward achieving our objective of training and deploying an RL agent on the actual building. We already have a real building that we have the ability to control, and using our pipeline produced a calibrated simulator that perfectly matches all of its devices, setpoints and sensors. Our future goal, showing how an agent trained on this simulator is able to produce a useful RL policy on the real building, will require a few more components. In future live trials on the real building, we will train an RL agent on our simulator, and measure its performance in the real world. Care must be taken in this process, as RL agents are prone to learning false patterns from the simulator [1, 11].

We also have outlined several approaches that we intend to use to further improve our simulator and fine tuning process

- (1) Tuning on data from multiple seasons, various times of day, and longer time windows, to improve generalizability.
- (2) Updating our simulation by adding in a radiative heat transfer model that enables solar gains during the daytime and losses at night.
- (3) Expanding the action space to include AC static pressure and HWS differential pressure that govern the air and water flow rates.
- (4) Adding a data-driven predictor-corrector to improve fidelity.

We are optimistic that our novel simulation tuning process will allow us to develop a useful and transferable, HVAC control solution, and we hope our work will inspire further effort in this field.

## REFERENCES

- [1] Javier Arroyo, Carlo Manna, Fred Spiessens, and Lieve Helsen. 2022. Reinforced model predictive control (RL-MPC) for building energy management. *Applied Energy* 309 (2022), 118346.
- [2] Donald Azuatalam, Wee-Lih Lee, Frits de Nijs, and Ariel Liebman. 2020. Reinforcement learning for whole-building HVAC control and demand response. *Energy and AI 2* (2020), 100020.
- [3] Yasaman Balali, Adrian Chong, Andrew Busch, and Steven O’Keefe. 2023. Energy modelling and control of building heating and cooling systems with data-driven and hybrid models—A review. *Renewable and Sustainable Energy Reviews* 183 (2023), 113496.
- [4] Christian Blad, Simon Bøgh, and Carsten Skovmose Kallesøe. 2022. Data-driven offline reinforcement learning for HVAC-systems. *Energy* 261 (2022), 125290.

- [5] Bingqing Chen, Zicheng Cai, and Mario Bergés. 2020. Gnu-rl: A practical and scalable reinforcement learning solution for building hvac control using a differentiable mpc policy. *Frontiers in Built Environment* 6 (2020), 562239.
- [6] Liangliang Chen, Fei Meng, and Ying Zhang. 2023. Fast Human-in-the-loop Control for HVAC Systems via Meta-learning and Model-based Offline Reinforcement Learning. *IEEE Transactions on Sustainable Computing* (2023).
- [7] Drury B Crawley, Linda K Lawrie, Frederick C Winkelmann, Walter F Buhl, Y Joe Huang, Curtis O Pedersen, Richard K Strand, Richard J Liesen, Daniel E Fisher, Michael J Witte, et al. 2001. EnergyPlus: creating a new-generation building energy simulation program. *Energy and buildings* 33, 4 (2001), 319–331.
- [8] Cheng Gao and Dan Wang. 2023. Comparative study of model-based and model-free reinforcement learning control performance in HVAC systems. *Journal of Building Engineering* 74 (2023), 106852.
- [9] Basil Kouvaritakis and Mark Cannon. 2016. Model predictive control. *Switzerland: Springer International Publishing* 38 (2016).
- [10] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. 2020. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643* (2020).
- [11] Simeng Liu and Gregor P Henze. 2006. Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: Part 2: Results and analysis. *Energy and buildings* 38, 2 (2006), 148–161.
- [12] Harvard Lomax, Thomas H Pulliam, David W Zingg, and TA Kowalewski. 2002. Fundamentals of computational fluid dynamics. *Appl. Mech. Rev.* 55, 4 (2002), B61–B61.
- [13] Mengxue Lu and Joseph Lai. 2020. Review on carbon emissions of commercial buildings. *Renewable and Sustainable Energy Reviews* 119 (2020), 109545.
- [14] Karl Mason and Santiago Grijalva. 2019. A review of reinforcement learning for autonomous building energy management. *Computers & Electrical Engineering* 78 (2019), 300–312.
- [15] Faye C McQuiston, Jerald D Parker, Jeffrey D Spitler, and Hessam Taherian. 2023. *Heating, ventilating, and air conditioning: analysis and design*. John Wiley & Sons.
- [16] R Sendra-Arranz and A Gutiérrez. 2020. A long short-term memory artificial neural network to predict daily HVAC consumption in buildings. *Energy and Buildings* 216 (2020), 109952.
- [17] E M Sparrow. 1993. Heat Transfer: Conduction [Lecture Notes].
- [18] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [19] Kirubakaran Velswamy, Biao Huang, et al. 2017. A Long-Short Term Memory Recurrent Neural Network Based Reinforcement Learning Controller for Office Heating Ventilation and Air Conditioning Systems. (2017).
- [20] Marshall Wang, John Willes, Thomas Jiralerspong, and Martin Moezzi. 2023. A Comparison of Classical and Deep Reinforcement Learning Methods for HVAC Control. *arXiv preprint arXiv:2308.05711* (2023).
- [21] Mubashir Wani, Akshya Swain, and Abhisek Ukil. 2019. Control strategies for energy optimization of HVAC systems in small office buildings using energyplus tm. In *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*. IEEE, 2698–2703.
- [22] Tianshu Wei, Yanzhi Wang, and Qi Zhu. 2017. Deep reinforcement learning for building HVAC control. In *Proceedings of the 54th annual design automation conference 2017*. 1–6.
- [23] Liang Yu, Shuqi Qin, Meng Zhang, Chao Shen, Tao Jiang, and Xiaohong Guan. 2021. A review of deep reinforcement learning for smart building energy management. *IEEE Internet of Things Journal* 8, 15 (2021), 12046–12063.
- [24] Liang Yu, Yi Sun, Zhanbo Xu, Chao Shen, Dong Yue, Tao Jiang, and Xiaohong Guan. 2020. Multi-agent deep reinforcement learning for HVAC control in commercial buildings. *IEEE Transactions on Smart Grid* 12, 1 (2020), 407–419.
- [25] Chi Zhang, Sanmukh R Kuppannagari, Rajgopal Kannan, and Viktor K Prasanna. 2019. Building HVAC scheduling using reinforcement learning via neural network based model approximation. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation*. 287–296.
- [26] Huan Zhao, Junhua Zhao, Ting Shu, and Zibin Pan. 2021. Hybrid-model-based deep reinforcement learning for heating, ventilation, and air-conditioning control. *Frontiers in Energy Research* 8 (2021), 610518.
- [27] Jie Zhao, Khee Poh Lam, B Erik Ydstie, and Omer T Karaguzel. 2015. EnergyPlus model-based predictive control within design–build–operate energy information modelling infrastructure. *Journal of Building Performance Simulation* 8, 3 (2015), 121–134.
- [28] Dian Zhuang, Vincent JL Gan, Zeynep Duygu Tekler, Adrian Chong, Shuai Tian, and Xing Shi. 2023. Data-driven predictive control for smart HVAC system in IoT-integrated buildings with time-series forecasting and reinforcement learning. *Applied Energy* 338 (2023), 120936.
- [29] Zhengbo Zou, Xinran Yu, and Semiha Ergan. 2020. Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network. *Building and Environment* 168 (2020), 106535.