

Electricity Pricing aware Deep Reinforcement Learning based Intelligent HVAC Control

Kuldeep Kurte
Oak Ridge National Laboratory, USA
kurtekr@ornl.gov

Jeffrey Munk
Oak Ridge National Laboratory, USA
munkjd@ornl.gov

Kadir Amasyali
Oak Ridge National Laboratory, USA
amasyalik@ornl.gov

Olivera Kotevska
Oak Ridge National Laboratory, USA
kotevskao@ornl.gov

Borui Cui
Oak Ridge National Laboratory, USA
cuib@ornl.gov

Teja Kuruganti
Oak Ridge National Laboratory, USA
kurugantipv@ornl.gov

Helia Zandi
Oak Ridge National Laboratory, USA
zandih@ornl.gov

ABSTRACT

Recently, deep reinforcement learning (DRL) based intelligent control of Heating, Ventilation, and Air Conditioning (HVAC) has gained a lot of attention due to DRL's ability to optimally control HVAC for minimizing operational cost while maintaining resident's comfort. The success of such DRL-based techniques largely depends on the articulation of the problem in terms of states, actions, and reward function. Inclusion of the electricity pricing information in the problem formulation can play an important role in saving the cost of HVAC operation. However, less attention has been given in the literature on formulating well-crafted state features based on electricity pricing. In this work, we propose an approach for training the DRL model with a specific focus on feature engineering based on electricity pricing. During training, we generate random but sufficiently realistic electricity price signals so that the pre-trained DRL model is robust and adaptive to the dynamic and variable electricity prices. The validation results are encouraging and show the potential of $\approx 12\%$ – 15% savings in the one day cost of HVAC operation, proving the usefulness of including electricity pricing related features as state features.

CCS CONCEPTS

• **Computing methodologies** → **Reinforcement learning**; • **Hardware** → **Power and energy**.

KEYWORDS

deep reinforcement learning, intelligent HVAC control, demand response, building energy

ACM Reference Format:

Kuldeep Kurte, Jeffrey Munk, Kadir Amasyali, Olivera Kotevska, Borui Cui, Teja Kuruganti, and Helia Zandi. 2020. Electricity Pricing aware Deep Reinforcement Learning based Intelligent HVAC Control. In *The 1st International*

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

RLEM '20, November 17, 2020, Virtual Event, Japan

© 2020 Crown in Right of Canada. Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8193-2/20/11...\$15.00

<https://doi.org/10.1145/3427773.3427866>

Workshop on Reinforcement Learning for Energy Management in Buildings & Cities (RLEM '20), November 17, 2020, Virtual Event, Japan. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3427773.3427866>

1 INTRODUCTION

Buildings are the major consumers of energy in the United States, with over 40% of the total energy consumption [21]. The International Energy Agency (IEA) estimated a 30% growth in the building's energy demand by the year 2035 [9]. Thus, it is important to efficiently manage the energy usage in the buildings, both residential as well as commercial. According to a report by the Energy Information Administration (EIA) over 51% household's total annual energy consumption in the year 2015 was attributed to HVAC [7]. It is evident from these facts that intelligent HVAC control has the potential of significantly reducing the building's energy demand and cost as well as minimizing the environmental impact by reducing peak power usage.

Various approaches for intelligent HVAC control have been developed over the last few years to achieve better energy efficiency and to reduce the cost of HVAC operation while maintaining the resident's comfort. These approaches are broadly classified into three categories: 1) rule-based 2) model-based and 3) model-free approach. The rule-based approach is widely used today and includes a simple on/off strategy with a set of rules encoded by experts [21], proportional-integral-derivative (PID) [1, 22] and fuzzy logic-based approaches [2, 8, 10, 11]. These rule-based approaches need a detailed understanding of the building dynamics, which is difficult to capture in the form of static rules. The fuzzy logic based approaches have their challenges for multi-objective optimization, such as minimizing energy cost while maintaining the resident's comfort. Due to these reasons, the rule-based approaches may not provide the optimal solution and are useful for a specific house [3, 20].

As opposed to the rule-based approaches, the model-based approaches such as Model Predictive Control (MPC) use system models to predict building's thermal behaviour and energy performance over a finite horizon [1, 15, 19, 25]. The performance of such MPC-based approaches largely depends on the accuracy of building energy model and its calibration accuracy as well as accuracy on predicting the external disturbances such as weather. Moreover,

executing building energy models along with the optimization algorithm is computationally demanding task.

Due to these limitations of MPC and other approaches to controlling HVAC, model-free reinforcement learning (RL) based approaches such as Q-learning are getting a lot of attention from the research community [4, 5, 14]. The model-free approaches do not need an accurate model of the system as the RL agent interacts with the environment through state, action and reward. The Q-learning uses a lookup table approach, called Q-table, to store the value of any state-action pair. However, the continuous and time varying nature of the variables such as outdoor temperature, indoor temperature, and HVAC setpoints could lead a large state-action space which limits the Q-learning's applicability.

Recently, Mnih et al. used a deep neural network for approximating Q-table and achieved a human-level intelligence while playing Atari game [16]. Following this research, many recent works used deep q-network (DQN) for HVAC control and showed DQN's ability to work with the continuous state-action space [3, 17, 23, 24, 27]. Since DQN approximates the Q-table, it is designed such that it takes state features as an input and predicts Q-values for various discrete actions.

The performance of the DQN-based approach for HVAC control depends on the articulation of the problem in terms of state, action and rewards. Among these, the state represents a set of features that describes the state of the environment (e.g. building) at a given time. For instance, indoor temperature, outdoor temperature, solar condition, indoor humidity, occupancy, etc. could be used as features to represent the state of the environment. In this respect, the state defines the *observability* of the RL agent. Figure 1 depicts that the DQN networks use states (both previous and next states) as inputs and estimates the values of being in those states. Further, these value estimates from both the networks are used to train the DQN model. Therefore it is important to carefully craft the features of the state so that the RL agent could efficiently learn to estimate the optimal Q-value of a state. For brevity, we are not providing an in-depth discussion of the training process of DQN. More technical details of DQN training can be found in [12, 13, 16, 24].

The cost of operating HVAC depends on the electricity pricing and the energy used for HVAC operations, e.g. cooling/heating. This fact leads us to hypothesize that the features about electricity pricing such as time to next price change, type of price change (increasing or decreasing), etc. could play an important role by

aiding RL agent to learn the optimal policy that minimizes the cost of HVAC operation while respecting the resident's comfort. An example of such a policy is to pre-cool the house to avoid the peak period of electricity pricing. Most of the recent works on DQN-based HVAC control used the feature such as time of the day, outdoor temperature, indoor temperature, electricity price, etc. to describe the state of the environment [26]. We noticed that less attention has been given for crafting the electricity pricing related features (mentioned earlier) for formulating the state. We believe that such features would help RL agent in better learning the actions, e.g. when and how long to pre-cool the house. The objective of the work presented in this paper is to validate the usability of the electricity pricing related features in training DQN for intelligent HVAC control for minimizing the cost of operation while maintaining the resident's comfort.

In this paper, we propose to improve the *observability* of the RL agent by using electricity pricing related features along with other indoor and outdoor temperature related features for describing the state of the environment. Section 2 discusses the proposed approach in detail. This approach assumes the availability of the electricity pricing forecast from utilities for the next few hours. Specifically, along with the indoor and outdoor temperatures, we use current electricity price and few price lookahead features (PLA) describing the time-to-next-price-change, its duration, and the magnitude of the change as well as the average outdoor temperature during this lookahead period. In our earlier work, we showed the potential of the price lookahead [12], however, in this paper, we are particularly focusing on the feature engineering part of it. A similar approach used the time-to-occupancy as one of the state features [4].

While training the RL agent, we used randomly generated synthetic price signals that are inspired by various time-of-use price signals from different utilities. We used a separate validation price signal and compared two pre-trained RL models, one with the price lookahead features (PLA) and other without PLA, with the fixed setpoint baseline for few selected days which were not present during training. The preliminary results are encouraging and showed price saving in the range of $\approx 13\%$ – 15% for one-day operation of HVAC. The validation results are presented in section 3 followed by the conclusion and future work in section 4.

2 PROPOSED APPROACH

This section introduces electricity pricing related features and formulates the problem of HVAC optimization using RL and later discusses the training and validation approach.

2.1 Problem Formulation

The objective of the RL agent is to learn the optimal policy π^* that minimizes the cost of HVAC operation while satisfying the resident's comfort. We assume that the indoor temperature (T_t^{in}) at any time t depends only on the previous indoor temperature (T_{t-1}^{in}), previous outdoor temperature (T_{t-1}^{out}), and the previous action taken, i.e. a_{t-1} . We define state as $S_t = \{t, T_t^{in1}, T_t^{in2}, T_t^{out}, p_t, PLA_t\}$ where t is the time of the day, T_t^{in1} and T_t^{in2} represent the indoor temperatures of zone 1 and zone 2 of the building, T_t^{out} is the outdoor temperature, and p_t and PLA_t represent current electricity price and a set of price lookahead features, respectively. Since RL agent

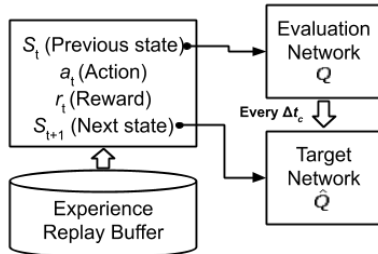


Figure 1: Structure of DQN. It consists of two networks, 1) evaluation network and 2) target network, and experience replay memory that stores the tuples $\langle S_t, a_t, r_t, S_{t+1} \rangle$.

interacts with the environment at every Δt_c interval, the T^{in} , T^{out} and p_t are averaged over the time interval of $[t - \Delta t_c, t]$.

We explain the price lookahead features PLA_t with the help of the Figure 2. At any time t_s , Δt_1 represents the time to the next price change, after which current price value P_1 changes to the new price value P_2 . Further, Δt_2 is the time duration during which the electricity price is P_2 . The duration Δt_1 is useful to inform RL agent about the time-to-next-price-change. So, if there is a peak price period in the near future then RL can act accordingly and pre-cool the house to minimize the work during peak period. We provide a 2°F of flexibility band below the cooling setpoint (for the summer period) so that, if needed, RL can pre-cool the house within this flexibility band. This flexibility band is defined as $[LT, UT]$ where UT is the cooling setpoint and $LT = UT - 2$. We also feel that Δt_2 is helpful for RL in determining how much to pre-cool. For instance, if the next peak price period has a longer duration then the RL agent can pre-cool the house for a longer period (depending on Δt_2) before the peak price to minimize the work during the peak period.

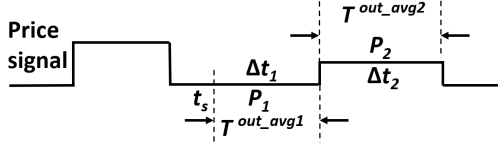


Figure 2: Sample price signal and price look ahead features.

Additionally, we calculate an index called the normalized differential price index ($NDPI_t$). It is a ratio of $(P_1 - P_2)$ to $(P_1 + P_2)$ and has a range of $[-1, +1]$. Its magnitude represents how big is the price difference and the sign represents the direction of the change, i.e. if the price will be increasing or decreasing. The value of $NDPI_t$ along with the values of P_1 and P_2 would collectively help the RL agent to guess how the electricity price will vary in the near future. Moreover, the values of Δt_1 and Δt_2 would help in knowing when this change is going to happen. The outdoor temperature during this lookahead period (i.e. Δt_1 and Δt_2) along with the current the indoor temperature could influence the RL's decision of pre-cooling. We include this information as average outdoor temperature during Δt_1 and Δt_2 and represented by T^{out_avg1} and T^{out_avg2} , respectively. Finally, the price lookahead features are represented as $PLA_t = \{P_1, \Delta t_1, T^{out_avg1}, P_2, \Delta t_2, T^{out_avg2}, NDPI_t\}$. Here, one important thing to note that Δt_1 and Δt_2 could take large values, e.g. a couple of hours. Instead of knowing the exact values of the time-to-next-price-change and the duration of the new price, knowing if the price is going to change within a fixed lookahead window, say one hour, could be beneficial. To realize this fact, we use a threshold of one hour for the lookahead window. This also reduces the state space the RL has to search and can avoid any unnecessary pre-cooling over a long period.

Since there are two zones in the building, the action space has four actions, $A = \{a_0, a_1, a_2, a_3\}$, where each action can take two values, on/off. These actions are further converted to the setpoint using the formalism shown in the Table 1. We use a negative reward strategy for reward (refer Eq. 1). The first term represents the cost of operating HVAC over the last control duration (Δt_c) with weight of

Table 1: Actions and their equivalent setpoints. T_{in} is indoor temperature and ΔT represents a small value (0.6°F).

Actions	Zone-1	Zone-2	Set point-1	Set point-2
a_0	0	0	$T_{in}^{in} + \Delta T$	$T_{in}^{in} + \Delta T$
a_1	0	1	$T_{in}^{in} + \Delta T$	$T_{in}^{in} - \Delta T$
a_2	1	0	$T_{in}^{in} - \Delta T$	$T_{in}^{in} + \Delta T$
a_3	1	1	$T_{in}^{in} - \Delta T$	$T_{in}^{in} - \Delta T$

4 and the second term represents the penalty for comfort violation with the weight value of 10.

$$r_t = -4 * cost_{t-\Delta t_c}^{avg} - 10 * [(T_{[t-\Delta t_c, t]}^{avg} - UT) + (LT - T_{[t-\Delta t_c, t]}^{avg})] \quad (1)$$

2.2 Training and Validation

The long learning period and likely poor comfort control during learning make online learning of DRL an unrealistic choice for intelligent HVAC control. In this work, we trained the DRL model in an offline mode using a building simulation environment. For this work, we developed a gray box building model using the data recorded in an unoccupied research house located in Knoxville, TN, USA. More information on the building simulation environment used in this work can be found in [6, 12, 13]. We used the DQN structure proposed by Mnih et al. [16] that consists of two neural networks with a memory called experience replay buffer. For the RL agent, the goal is to obtain the optimal policy π^* that maximizes the long term reward. The DQN training is an iterative procedure that uses a Q-value update equation (Eq. 2) to update DQN's weights. η represents a learning rate. The detailed description of the DQN training procedure and training parameters can be found in our prior work [13].

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \eta(r_{t+1} + \gamma \max_{a_{t+1}} \hat{Q}_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)) \quad (2)$$

During training, we generated synthetic price signals by randomly selecting the start time, the duration, and the magnitude of the price peak during a given day. This is inspired by various time-of-use price signals from different utilities. We trained the RL model for multiple episodes (50 episodes for this work) and for each episode we updated the price signal so that the RL agent can see sufficient variations in the price signal and outdoor and indoor temperatures. We used a separate synthetic validation price signal for which the domain expert provided us the expected behavior by the RL (refer Figure 3). The electricity price of this validation price signal varies between $[\$0.05, \$0.25]$ and has four peak price periods in a day, one in the morning and three in the afternoon time, with different amplitudes and duration. We trained two RL models one with PLA and other without PLA features and compared both them with the fixed setpoint baseline using the validation price signal.

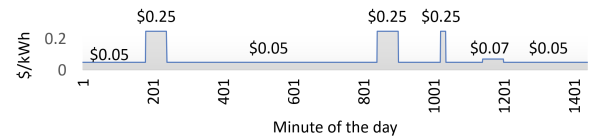


Figure 3: Validation price signal used for this study

3 RESULTS

3.1 Dataset Used

For training RL models, we used weather data from Typical Meteorological Year (TMY) data for Knoxville, Tennessee [18]. The price data was synthetically generated using a common time-of-use (TOU) electricity rate plans. The price peak and off-peak durations and the start of the peak were randomly chosen for each day in the dataset. We trained the RL model for 50 episodes for the month of July from TMY data. We validated the pre-trained RL model for the month of August using the price signal shown in Figure 3. We further selected three representative days from the month of August with the average outdoor temperatures of 24.6°C, 26.6°C, and 25.6°C respectively for detailed analysis.

3.2 Validation Results

We trained two RL models, one with price lookahead (PLA) features and another without PLA features, and compared them with the fixed setpoint baseline. The fixed cooling setpoint of 74°F (23.33°C) was used for this validation study. First, we validated the performance of pre-trained RL models for the month of August. We also selected three days from the beginning, mid and end period of the August month to validate the per-day cost. Table 2 shows the cost and energy consumption for the baseline with 74°F as a cooling setpoint and for the RL with and without PLA features. We can observe that the pre-trained RL model without PLA features showed a 6% cost reduction for the entire month of August, did not show a cost reduction for Day-1, showed 5% cost reduction for Day-2, and 7% cost reduction for Day-3 over the baseline cost. Whereas, the pre-trained RL model which used PLA features showed the cost reduction in the range of 12%–15% over the baseline cost. This cost comparison clearly demonstrates the importance of the price lookahead features for minimizing the cost of HVAC operation.

Table 2: % cost reduction (↓) using pre-trained RL model with and without PLA features over baseline.

		Baseline	w/o PLA	with PLA
August (TMY)	Cost (\$)	27.27	25.68 (6%↓)	24.02 (12%↓)
	Energy (kWh)	391.96	390.96	409.31
Day-1	Cost (\$)	0.83	0.83 (0%↓)	0.717 (13%↓)
	Energy (kWh)	11.84	12.03	12.77
Day-2	Cost (\$)	1.18	1.12 (5%↓)	1.01 (14%↓)
	Energy (kWh)	16.88	16.68	17.75
Day-3	Cost (\$)	0.88	0.82 (7%↓)	0.75 (15%↓)
	Energy (kWh)	12.22	12.47	12.73

For illustration, we present the temperature variation plots for Day-3 for HVAC control using baseline and pre-trained RL models with and without PLA features (refer Figure 4). We can observe that the pre-trained RL model with PLA features tried to cool down the house before the second and third peak price periods when the outdoor temperature is also at its peak. These are represented by the pre-cooling events in the Figure 4 (c). As expected, the RL model did not perform any pre-cooling before morning and evening peaks. The potential reason could be the lower outdoor temperature during these two time periods. Such pre-cooling events are not present in Figure 4 (b) where RL model did not use PLA features. The pre-cooling events shown in the Figure 4 (c) and the price reduction

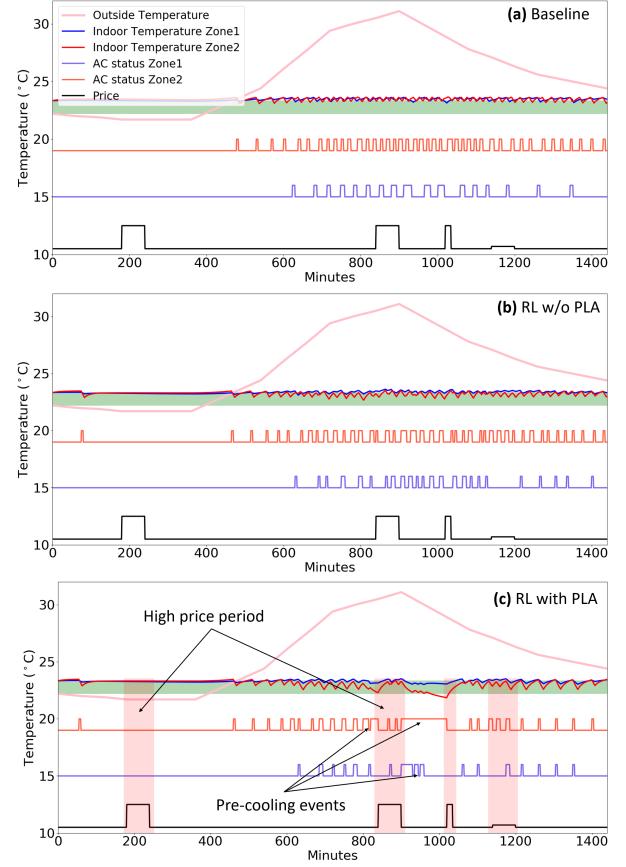


Figure 4: Validation plots for Day-3. The green band represents the 2° flexibility band below 74°F (i.e. 22.22°C–23.33°C).

shown in the Table 2 verify the significance of PLA features in reducing the cost of RL-based HVAC operation. Moreover, both RL-agents (with and w/o PLA features) learned to satisfy the resident’s comfort. Also, we have not observed significant variations in energy usage for all three cases.

4 CONCLUSION AND FUTURE WORK

In this paper, we presented our work on understanding the significance of using price lookahead (PLA) features while formulating the state for intelligent HVAC control using RL. We demonstrated that the RL model that was trained with PLA features achieved better cost reduction ($\approx 12\%$ – 15% over baseline with the fixed setpoint) than the RL model trained without PLA features while satisfying the resident’s comfort. These results are encouraging, however, we feel that this approach needs further investigation on understanding the relevance of each PLA features considered in this work. We also found that zone-1 has more temperature variations than zone-2. Also, the RL agent tried to pre-cool for a longer duration before the third peak price period (refer Figure 4 (c)). This observation leads us to think that, although the RL agent with PLA features reduced the cost, the solution could be sub-optimal. In the future, we will work on interpreting RL’s actions to achieve better optimization.

ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Building Technology Office under contract number DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

REFERENCES

- [1] Abdul Afram and Farrokh Janabi-Sharifi. 2014. Theory and applications of HVAC control systems—A review of model predictive control (MPC). *Building and Environment* 72 (2014), 343–355. <https://doi.org/10.1016/j.buildenv.2013.11.016>
- [2] Osman Ahmed. 2000. Method and apparatus for determining a thermal setpoint in a HVAC system. US Patent 6,145,751.
- [3] Ki Uhn Ahn and Cheol Soo Park. 2020. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. *Science and Technology for the Built Environment* 26, 1 (2020), 61–74.
- [4] Enda Barrett and Stephen Linder. 2015. Autonomous HVAC Control, A Reinforcement Learning Approach. In *Machine Learning and Knowledge Discovery in Databases*, Albert Bifet, Michael May, Bianca Zadrozny, Ricard Gavaldà, Dino Pedreschi, Francesco Bonchi, Jaime Cardoso, and Myra Spiliopoulou (Eds.). Springer International Publishing, Cham, 3–19.
- [5] Yujiao Chen, Leslie K Norford, Holly W Samuelson, and Ali Malkawi. 2018. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings* 169 (2018), 195–205.
- [6] Borui Cui, Jeffrey Munk, Roderick Jackson, David Fugate, and Michael Starke. 2017. Building thermal model development of typical house in US for virtual storage control of aggregated building loads based on limited available information. In *30th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS 2017)*, Vol. 5. 3179–3191.
- [7] EIA. 2015. Use of energy explained Energy use in homes (U.S. Energy Information Administration (EIA), 2015). <https://www.eia.gov/energyexplained/use-of-energy/homes.php>
- [8] MM Gouda, Sean Danaher, and CP Underwood. 2001. Thermal comfort based fuzzy logic controller. *Building services engineering research and technology* 22, 4 (2001), 237–253.
- [9] IEA. 2019. Policy pathways: Modernizing building energy codes. [url=https://www.iea.org/publications/freepublications/publication/](https://www.iea.org/publications/freepublications/publication/).
- [10] Fakhreddine Karray, Fakhreddine O Karray, and Clarence W De Silva. 2004. *Soft computing and intelligent systems design: theory, tools, and applications*. Pearson Education.
- [11] D Kolokotsa, D Tsiavos, GS Stavrakakis, K Kalaitzakis, and E Antonidakis. 2001. Advanced fuzzy logic controllers design and evaluation for buildings—occupants thermal—visual comfort and indoor air quality satisfaction. *Energy and buildings* 33, 6 (2001), 531–543.
- [12] Olivera Kotevska, Kuldeep Kurte, Jeffrey Munk, Travis Johnston, Evan Mckee, Kalyan Perumalla, and Helia Zandi. 2020. *RL-HEMS: Reinforcement Learning Based Home Energy Management System for HVAC Energy Optimization*. Technical Report. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).
- [13] Kuldeep Kurte, Jeffrey Munk, Olivera Kotevska, Kadir Amasyali, Robert Smith, Evan Mckee, Yan Du, Borui Cui, Teja Kuruganti, and Helia Zandi. 2020. Evaluating the Adaptability of Reinforcement Learning Based HVAC Control for Residential Houses. *Sustainability* 12, 18 (2020). <https://doi.org/10.3390/su12187727>
- [14] B. Li and L. Xia. 2015. A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. 444–449.
- [15] Mehdi Maasoumy, Alessandro Pinto, and Alberto Sangiovanni-Vincentelli. 2011. Model-based hierarchical optimal control design for HVAC systems. In *Dynamic Systems and Control Conference*, Vol. 54754. 271–278.
- [16] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.
- [17] Adam Nagy, Hussain Kazmi, Farah Cheaib, and Johan Driesen. 2018. Deep Reinforcement Learning for Optimal Control of Space heating. [arXiv:1805.03777 \[stat.AP\]](https://arxiv.org/abs/1805.03777)
- [18] National Renewable Energy Laboratory. 2015 (accessed May 17, 2019). *National Solar Radiation Data Base, 1991–2005 Update: Typical Meteorological Year 3*. https://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/data/tmy3/723260TYA.CSV.
- [19] Frauke Oldewurtel, Alessandra Parisio, Colin N Jones, Manfred Morari, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, and Katharina Wirth. 2010. Energy efficient building climate control using stochastic model predictive control and weather predictions. In *Proceedings of the 2010 American control conference*. IEEE, 5100–5105.
- [20] Germán Ramos Ruiz, Eva Lucas Segarra, and Carlos Fernández Bandera. 2019. Model predictive control optimization via genetic algorithm using a detailed building energy model. *Energies* 12, 1 (2019), 34.
- [21] Pervez Hameed Shaikh, Nursyarizal Bin Mohd Nor, Perumal Nallagownden, Irraivan Elamvazuthi, and Taib Ibrahim. 2014. A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews* 34 (2014), 409–429. <https://doi.org/10.1016/j.rser.2014.03.027>
- [22] Shengwei Wang and Zhenjun Ma. 2008. Supervisory and optimal control of building HVAC systems: A review. *Hvac&R Research* 14, 1 (2008), 3–32.
- [23] Yuan Wang, Kirubakaran Velswamy, and Biao Huang. 2017. A Long-Short Term Memory Recurrent Neural Network Based Reinforcement Learning Controller for Office Heating Ventilation and Air Conditioning Systems. *Processes* 5, 3 (2017). <https://doi.org/10.3390/pr5030046>
- [24] Tianshu Wei, Yanzi Wang, and Qi Zhu. 2017. Deep Reinforcement Learning for Building HVAC Control. In *Proceedings of the 54th Annual Design Automation Conference 2017 (Austin, TX, USA) (DAC â€ˆ17)*. Association for Computing Machinery, New York, NY, USA, Article 22, 6 pages. <https://doi.org/10.1145/3061639.3062224>
- [25] Tianshu Wei, Qi Zhu, and Mehdi Maasoumy. 2014. Co-scheduling of HVAC control, EV charging and battery usage for building energy efficiency. In *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE, 191–196.
- [26] L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, and X. Guan. 2020. Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings. *IEEE Transactions on Smart Grid* (2020), 1–1.
- [27] Zhiang Zhang, Adrian Chong, Yuqi Pan, Chenlu Zhang, and Khee Poh Lam. 2019. Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning. *Energy and Buildings* 199 (2019), 472–490.