Collaborative energy demand response with decentralized actor and centralized critic

Ruben Glatt
Felipe Leno da Silva
Braden Soper
William A. Dawson
Edward Rusu
Ryan A. Goldhahn
{glatt1,leno,soper3,dawson29,rusu1,goldhahn1}@llnl.gov
Lawrence Livermore National Laboratory
Livermore, California, USA

ABSTRACT

The ongoing industrialization and rising technology adoption around the world are leading to ever higher energy consumption. The benefits of electrification are enormous, but the growing demand also comes with challenges with respect to associated greenhouse gas emissions. Although continuing progress in energy research has brought up new technologies in energy generation, storage, and distribution, most of those technologies focus on increasing efficiency of individual components. Work on integration and coordination abilities between individual components in micro-grids will lead to further improvements and gains in efficiency that are necessary to reduce carbon footprints and slow down climate change. To this end, the CityLearn environment provides a simulation framework that allows the control of energy components in buildings that are organized in districts. In this paper, we propose an energy management system based on the decentralized actor-critic reinforcement learning algorithm *MARLISA* but integrate a centralized critic and call it $MARLISA_{DACC}$. In this way, we are training a model to autonomously control the energy storage of individual buildings in a CityLearn district to improve demand response guided by a better informed training signal. We show performance increases over baseline control techniques for a district but also discuss the resulting action selection for individual buildings.

CCS CONCEPTS

• **Applied computing** → *Environmental sciences*;

KEYWORDS

multi-agent, reinforcement learning, energy management, citylearn

ACM Reference Format:
Ruben Glatt, Felipe Leno da Silva, Braden Soper, William A. Dawson, Edward Rusu, and Ryan A. Goldhahn. 2021. Collaborative energy demand response with decentralized actor and centralized critic. In *The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3486611.3488732

1 INTRODUCTION

The energy grids of the past have been built based on the assumption of centralized energy generation and storage. The centralized approach has led to a grid utilization that closely follows the current energy demand of an area with high energy peaks during high demand times and very low utilization, for example, during the night. In this classical setup, a common approach to lower demand has been the introduction of multi-tariff charging schemes [2], which only slightly alleviated the problem. Despite these efforts, the overall increase in energy demand over the last years, a trend which will continue in the future, has led to rising stress on the existing energy infrastructure which in turn led to undesirable outcomes such as power outages through failure of infrastructure components.

On the other hand, technological advancements are leading the way to integrate more decentralized components where generation, storage, and consumption can happen in the same region of the grid. These new abilities and the advent of *smart grids* enable the development of power grids that can help reduce peak loads and better balance demand throughout the day. Smart grids leverage data about infrastructure and consumer behavior to improve efficiency, reliability, and sustainability of energy production and distribution [3]. Exploratory research already shows that the key to dealing with peak demand is directly related to the smart management of energy infrastructure components and consumer coordination [8].

Recently, we can see increasing interest in simulation frameworks that can deal with *distributed energy resources (DER)* such as *GridDyn* [11], *HELICS* [17], or *OptGrid* [4]. Those frameworks are high-fidelity simulations and require high performance computational resources making them unsuitable for lower budget research facilities or university research. In Figure 1, we show the training setup of a much simpler framework. It shows the general interactions in the CityLearn [26] environment, which offers an easy to use OpenAI Gym [5] interface for the implementation of Multi-Agent Reinforcement Learning (MARL) [6, 30]. CityLearn was created with the goal of supporting research and development of methods and approaches to optimize energy usage and reduce

BuildSys '21, November 17–18, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9114-6/21/11.
https://doi.org/10.1145/3486611.3488732

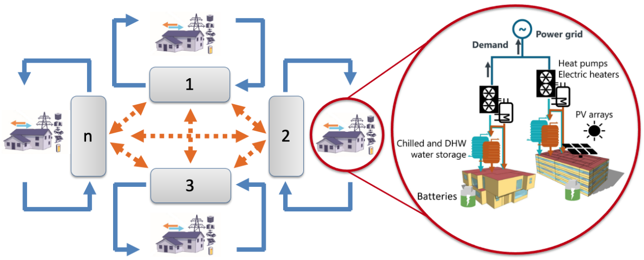

Figure 1: In the CityLearn environment [26], the user can control a number *n* of agents in centralized or decentralized manner and allow communication between agents. Each agent controls the energy storage devices of one building.

peak demand. To that end, it provides a realistic simulator with convenient entry-points to develop machine learning-based agents for controlling energy storage and utilization in buildings. Buildings are organized in districts and each agent controls the associated energy storage systems for a building such as electric batteries, domestic hot water (DHW) storage, and chilled water (for sensible cooling and dehumidification). The available scenarios include models of air-to-water heat pumps, electric heaters, solar photovoltaic arrays, and the pre-computed energy loads of the buildings, which include space cooling, dehumidification, appliances, DHW, and solar generation. The agents can observe current demand, state of charge of storage devices, time of day, and environmental variables such as forecasts for temperature, solar radiation, and humidity, to help predict demand and generation for the next hours.

The control problem in energy grids has been investigated from different optimization perspectives [1, 16, 22, 25] as well as a Reinforcement Learning (RL) task [19, 21, 29]. Here, we focus our efforts on the use of RL algorithms to coordinate the demand response of a given district and make comparisons between centralized and decentralized approaches. In realistic scenarios, apart from differences in building setup and available energy devices, demand and generation are often dependent on weather or other external factors, making it a non-trivial task to predict any future demand even for individual agents. Moreover, in practice, it is often not desirable or possible that agents share much information continuously with their neighbours (which in commercial settings could be competitors) so that the decision making is often based only on partial information about the system.

2 BACKGROUND

Control and decision making problems can often be represented as *Markov Decision Processes (MDP)* [20]. An MDP is defined by a tuple $\langle S, A, T, R, \gamma \rangle$ where S is the set of possible states, A is the set of available actions, $T : S \times A \times S \longrightarrow [0, 1]$ is a transition function providing a probability of observing a follow-up state s' after taking action a in state s, and $R : S \longrightarrow \mathbb{R}$ is a reward function providing a reward after reaching a state s, and $\gamma \longrightarrow [0, 1]$ is a discount factor for future rewards. In learning problems, T and R are unknown and agents need to learn from their experience collected as samples of $\langle s, a, r, s' \rangle$, where r is the reward observed after applying a in s and reaching follow-up state s'. Many approaches to solve MDPs are described in the field of RL [23], where an agent explores different strategies in a given environment, receives a feedback on its actions, and learns a behavior policy from its observations. The agent's goal is to find an optimal policy $\pi^*(s)$, that provides the best action in any state s. RL has been shown to successfully solve challenging domains such as Atari game play [9, 15], electric vehicle charging [18], and building energy management [14].

Even though it is possible to use single agent approaches in the CityLearn scenarios, real-world constraints often permit a centralized controller for a whole district because buildings have individual owners and are hesitant to share all available information. A more fitting approach would be to place it in a multi-agent scenario, modeling each building as an individual agent that has the ability to share some information with other agents through collaboration to achieve optimal results on a global level (in a given district).

In that case, the MDP becomes a *Stochastic Game (SG)* [12] with an extended set of actions A_i for all agents $i = 1, ..., n$ in the environment. Notably, all actions now form the joint action vector $\boldsymbol{u} = \{a_0, ..., a_n\}$ and the transition function is dependent on this joint action. Solution approaches to this kind of problems are often described in the field of Multi-Agent Reinforcement Learning (MARL) [7, 24]. Here, we leverage the MARLISA model [27] as a starting point for our approach.

MARLISA is a multi-agent algorithm specifically developed for the building energy management control problem modeled in City Learn, where each agent controls a single building in a given district. The coordination between agents is achieved through an interactive action selection process. At each turn a central controller randomly selects an order for all buildings in a district. In this order, each agent selects an action without performing it, predicts the expected energy demand of the building if that action is applied, and then transmits this information to the next agent. The next agent follows the same procedure and transmits the cumulative demand to the next agent and so on. This is repeated for several rounds before the joint action is applied to the system.

The learning process for the action selection is carried out by a Soft Actor-Critic (SAC) algorithm [10], where each agent learns their own actor and critic separately. SAC optimizes a stochastic policy in an off-policy way sampling from past experiences. The optimization works by following an objective that maximizes an expected reward but also the entropy of the policy. This incentivizes the policy to explore more widely and the stochastic nature offers more flexibility for acting when actions seem to have similar reward expectations leading to improved sample efficiency during learning while increasing robustness and stability of the learning curve.

3 DECENTRALIZED ACTOR, CENTRALIZED CRITIC

In earlier work, Lowe et al. [13] propose multi-agent deep deterministic policy gradient (MADDPG), an approach to the multi-agent domain based on the actor-critic framework for mixed cooperative-competitive environments. This algorithm is based on the assumption of a decentralized actor and a centralized critic and allows agents to learn under the condition that they can only use local information for the action selection and without the need for any

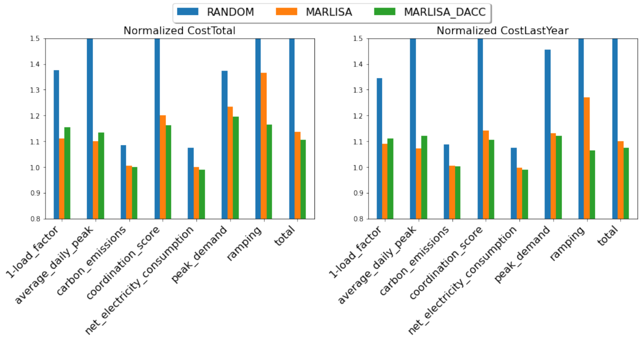

Figure 2: The graphs show the normalized cost for the total training period and for only the last year for a random approach, the original *MARLISA*, and our *MARLISA*$_{DACC}$.

particular communication method between agents. The decentralized actor is then able to act independently and does not need any communication after training. The centralized critic on the other hand is needed only during training but instead of local observations it uses the full state for training that is not available for individual agents (actors).

As we are allowing communication between agents, this setup is suitable for the CityLearn environment. Inspired by this approach, we here propose a slight change to the *MARLISA* algorithm by combining it with the idea of a decentralized actor and a centralized critic. For this paper, we keep *MARLISA*'s leader-follower approach to augment the agent's observation with information about the district demand, but future work might make this communication intensive step unnecessary. As the actor input remains the same, we change the input of the critic to reflect all buildings in the district. Specifically, we use the information about the non-shiftable load, the net electricity consumption, the cooling storage state of charge (SOC), the DHW storage SOC, the electrical storage SOC, and the solar generation rate for each building. We name the resulting algorithm *MARLISA*$_{DACC}$.

4 EXPERIMENTAL EVALUATION

We conduct our preliminary experiments in the CityLearn environment and provide a simple comparison between *MARLISA* and *MARLISA*$_{DACC}$ and a random agent as a baseline. The scenario is based on the task for the 2021 CityLearn Challenge [28] where agents control the energy storage of each building in a district of 9 buildings simulated on an hourly time-scale for a period of 4 years. The environmental data is modeled through a given climate zone and the setup of the buildings is predefined by the environment. Each agent controls the SOC for domestic hot water, chilled water, and electrical storage available in a building through 3 action variables. The objective of controlling the grid response is a coordination challenge, so a good solution depends on the joint actions of all agents. The reward function is based on the original *MARLISA* implementation [27] and considers the individual net electricity consumption and the total net electricity demand of the entire district.

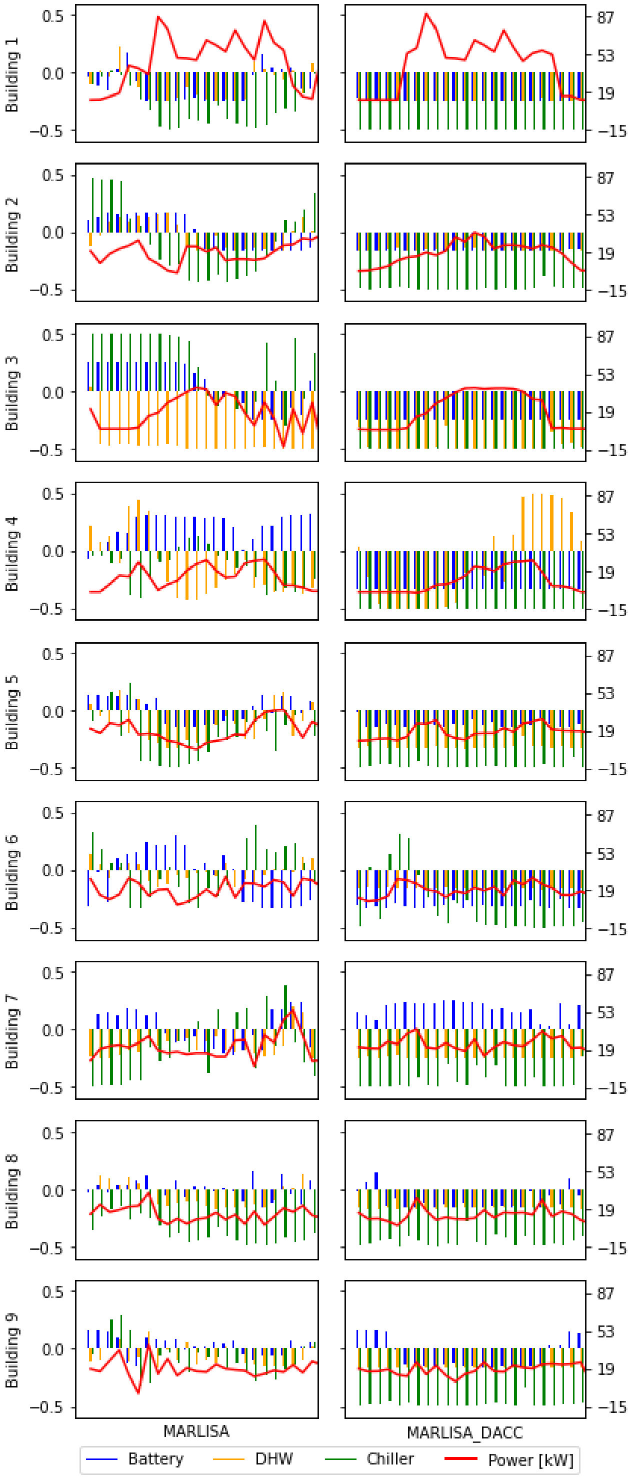

Figure 3: Learned action selection of each building for *MARLISA* (left) and *MARLISA*$_{DACC}$ (right) overlayed by the resulting demand curve for a typical work day.

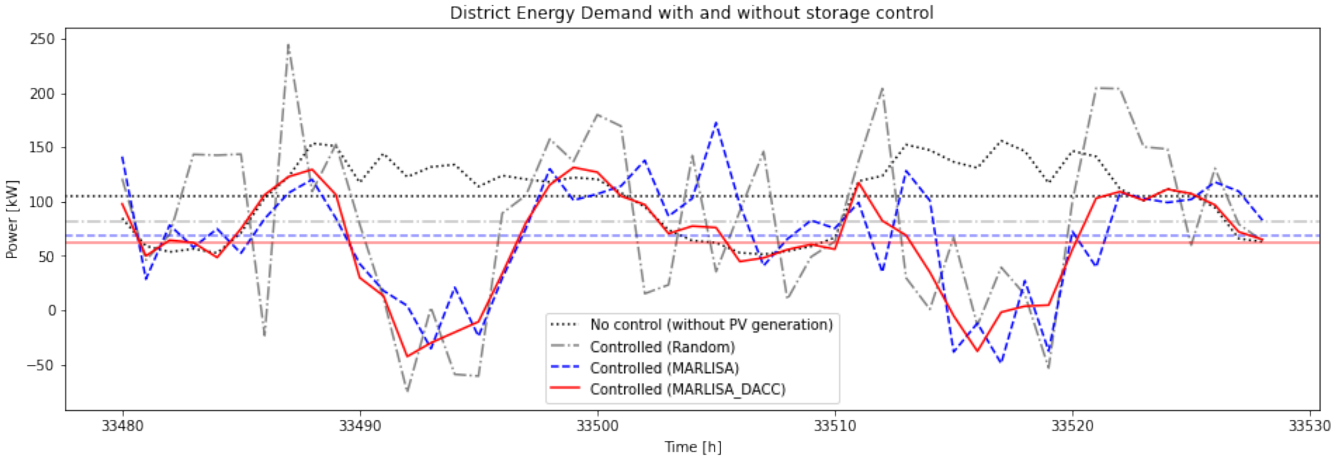

Figure 4: In this graph we can see the district demand for two consecutive days in a typical work week in the final year of the training process. The horizontal lines are the respective mean values for each approach.

In Figure 2 we show the results of a random agent, a *MARLISA* agent, and a $MARLISA_{DACC}$ agent in terms of a normalized cost with respect to a *rule-based controller (RBC)* over 8 metrics: load factor, average daily demand peak, carbon emissions, coordination score, net electricity consumption, overall peak demand, ramping factor, and a total metric combining the previous ones. The RBC and the metrics are provided by the CityLearn environment. Results are shown for the whole 4-year training process (left) and only for the last year (right), where a lower bar indicates a better performance. We can see that the random agent performs poorly in most metrics, but surprisingly this seems to have little impact on the carbon emissions and overall net electricity consumption. The other two agents both perform much better and have very similar scores across all metrics. *MARLISA* performs better for the load factor and the average daily peak demand while $MARLISA_{DACC}$ has a small advantage in the other metrics but a distinctive advantage in ramping. For both algorithms we can also see an improvement in the last year as the agents have learned better actions over time.

In Figure 3, we can see the actual learned policy of a *MARLISA* and a $MARLISA_{DACC}$ agent in the last year of training for a randomly selected day. The figure shows two graphs for each building, one for each algorithm. In each graph, we can see a barplot showing the action value for each of the 3 actions at 1 hour intervals in the range of -0.5 to 0.5 overlayed by the resulting building demand curve as a red line. As we can see, the demand curves are very similar but the ones for $MARLISA_{DACC}$ are smoother in general. The action values seem to differ more, as *MARLISA* seems to be more alternating between positive and negative values while $MARLISA_{DACC}$ has dominantly negative values. This last observation is somehow surprising as it seems there is little charging of the available storage involved on that particular day which will require further investigation in future work.

Figure 4 shows the energy demand of the whole district over two consecutive days. The curves present the actual values and the corresponding horizontal lines show the averages over the observation time. The dotted line provides the baseline of the energy demand of the district considering no storage and no power generation through solar energy. The second baseline is the demand curve if the agents were controlled by a random agent. Even though the agent uses the available storage randomly and has the highest peak loads, it still shows a reduced average demand compared to the no control baseline. We can then observe the curves for *MARLISA* and a $MARLISA_{DACC}$ that show very similar behavior as expected. The difference again is a smoother curve for $MARLISA_{DACC}$ and the higher peak for *MARLISA*. It also supports the findings from Figure 2 as we see the average demand is clearly lower for $MARLISA_{DACC}$, almost halving the demand compared to the original baseline without control.

5 CONCLUSION

Coordinating and optimizing energy grid utilization is a challenging task of high importance to deal with the continuing increase in energy demand and to reduce energy-related emissions. In this paper, we proposed a new algorithm, $MARLISA_{DACC}$, which leverages a decentralized actor, centralized critic architecture to facilitate learning coordinated energy demand response strategies and discussed results on a building as well as a district level in the CityLearn environment. In addition to outperforming existing baselines, we also took first steps to make learned control policies more transparent. To that end, we explored new evaluation aspects by visualizing individual action decisions on a building level to gain deeper insights into learned strategies and to support the development of new algorithms. Even though this is a work in progress and we report only preliminary results here, we are confident that this will serve as a base for future work to improve algorithm development and analysis.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC through the support of LDRD 20-SI-005. LLNL-CONF-826291.

REFERENCES

[1] Italo Atzeni, Luis G Ordóñez, Gesualdo Scutari, Daniel P Palomar, and Javier Rodríguez Fonollosa. 2012. Demand-side management via distributed energy generation and storage optimization. *IEEE Transactions on Smart Grid* 4, 2 (2012), 866–876.

[2] Galen Barbose, Charles Goldman, and Bernie Neenan. 2004. *A survey of utility experience with real time pricing.* Technical Report. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

[3] Heider Berlink, Nelson Kagan, and Anna Helena Reali Costa. 2015. Intelligent decision-making for smart home energy management. *Journal of Intelligent & Robotic Systems* 80, 1 (2015), 331–354.

[4] Andrey Bernstein and Emiliano Dall'Anese. 2019. Real-time feedback-based optimization of distribution grids: A unified approach. *IEEE Transactions on Control of Network Systems* 6, 3 (2019), 1197–1209.

[5] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. Openai gym. *arXiv preprint arXiv:1606.01540* (2016).

[6] Lucian Busoniu, Robert Babuska, and Bart De Schutter. 2008. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38, 2 (2008), 156–172.

[7] Lucian Buşoniu, Robert Babuška, and Bart De Schutter. 2010. Multi-agent reinforcement learning: An overview. *Innovations in multi-agent systems and applications-1* (2010), 183–221.

[8] Felipe Leno Da Silva, Cyntia EH Nishida, Diederik M Roijers, and Anna H Reali Costa. 2019. Coordination of electric vehicle charging through multiagent reinforcement learning. *IEEE Transactions on Smart Grid* 11, 3 (2019), 2347–2356.

[9] Ruben Glatt, Felipe Leno Da Silva, Reinaldo Augusto da Costa Bianchi, and Anna Helena Reali Costa. 2020. DECAF: deep Case-based Policy Inference for knowledge transfer in Reinforcement Learning. *Expert Systems with Applications* 156 (2020), 113420.

[10] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*. PMLR, 1861–1870.

[11] Brian M Kelley, Philip Top, Steven G Smith, Carol S Woodward, and Liang Min. 2015. A federated simulation toolkit for electric power grid and communication network co-simulation. In *2015 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*. IEEE, 1–6.

[12] Michael L Littman. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*. Elsevier, 157–163.

[13] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *arXiv preprint arXiv:1706.02275* (2017).

[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] 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.

[16] Daniel K Molzahn, Florian Dörfler, Henrik Sandberg, Steven H Low, Sambuddha Chakrabarti, Ross Baldick, and Javad Lavaei. 2017. A survey of distributed optimization and control algorithms for electric power systems. *IEEE Transactions on Smart Grid* 8, 6 (2017), 2941–2962.

[17] Bryan Palmintier, Dheepak Krishnamurthy, Philip Top, Steve Smith, Jeff Daily, and Jason Fuller. 2017. Design of the HELICS high-performance transmission-distribution-communication-market co-simulation framework. In *2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*. IEEE, 1–6.

[18] Jacob F Pettit, Ruben Glatt, Jonathan R Donadee, and Brenden K Petersen. 2019. Increasing performance of electric vehicles in ride-hailing services using deep reinforcement learning. *arXiv preprint arXiv:1912.03408* (2019).

[19] Giuseppe Pinto, Marco Savino Piscitelli, José Ramón Vázquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. 2021. Coordinated energy management for a cluster of buildings through deep reinforcement learning. *Energy* 229 (2021), 120725.

[20] Martin L Puterman. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[21] Dawei Qiu, Yujian Ye, Dimitrios Papadaskalopoulos, and Goran Strbac. 2021. Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach. *Applied Energy* 292 (2021), 116940.

[22] Gianluca Serale, Massimo Fiorentini, Alfonso Capozzoli, Daniele Bernardini, and Alberto Bemporad. 2018. Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. *Energies* 11, 3 (2018), 631.

[23] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction.* MIT press.

[24] Ming Tan. 1993. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*. 330–337.

[25] Ssennoga Twaha and Makbul AM Ramli. 2018. A review of optimization approaches for hybrid distributed energy generation systems: Off-grid and grid-connected systems. *Sustainable Cities and Society* 41 (2018), 320–331.

[26] José R. Vázquez-Canteli, Sourav Dey, Gregor Henze, and Zoltán Nagy. 2020. CityLearn: Standardizing Research in Multi-Agent Reinforcement Learning for Demand Response and Urban Energy Management. *CoRR* abs/2012.10504 (2020). arXiv:2012.10504 https://arxiv.org/abs/2012.10504

[27] Jose R Vazquez-Canteli, Gregor Henze, and Zoltan Nagy. 2020. MARLISA: Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 170–179.

[28] Jose R Vazquez-Canteli, Zoltan Nagy, Gregor Henze, and Sourav Dey. 2021. *The CityLearn Challenge.* https://sites.google.com/view/citylearnchallenge/home

[29] Zhe Wang and Tianzhen Hong. 2020. Reinforcement learning for building controls: The opportunities and challenges. *Applied Energy* 269 (2020), 115036.

[30] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control* (2021), 321–384.