Energy and Carbon Footprint Reduction Using Multi-Agent Reinforcement Learning: A Digital Twin Approach

1st Sarveshwar Reddy Mangavarapu
Artificial Intelligence Department
Amrita Vishwa Vidyapeetham
Coimbatore, India

2nd Shivarathri Tagur Misra
Artificial Intelligence Department
Amrita Vishwa Vidyapeetham
Coimbatore, India

3rd Koushik Devajji
Artificial Intelligence Department
Amrita Vishwa Vidyapeetham
Coimbatore, India

4th Tammineni Varun
Artificial Intelligence Department
Amrita Vishwa Vidyapeetham
Coimbatore, India

***Abstract*—In the pursuit of sustainable building operations, this study explores and compares the effectiveness of various control strategies in reducing energy consumption and carbon emissions. Specifically, we evaluate the performance of a single-agent Proximal Policy Optimization (PPO), a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) trained using a digital twin, and a conventional baseline based on ASHRAE standards. The experimental results reveal that the MADDPG controller, leveraging the digital twin environment, achieves substantial efficiency gains—reducing energy usage by 47.99% and carbon emissions by an impressive 99.99% compared to the ASHRAE baseline. These outcomes underscore the transformative potential of digital twin-enabled multi-agent reinforcement learning for optimizing building energy systems and supporting global decarbonization efforts.**

***Index Terms*—Digital Twin, Multi-Agent Reinforcement Learning, Energy Efficiency, Carbon Emissions Reduction, Building Energy Systems, Proximal Policy Optimization (PPO), MADDPG, Sustainable Building Operations**

I. Introduction

The building sector remains one of the most energy-intensive domains globally, accounting for nearly 40% of total energy consumption [1]. With the rise in urbanization, the demand for Heating, Ventilation, and Air Conditioning (HVAC) systems has increased significantly. These systems are integral to indoor thermal comfort but contribute heavily to electricity use. According to the U.S. Department of Energy, HVAC systems are responsible for over 50% of energy use in residential buildings in the United States [2]. Therefore, enhancing energy efficiency in HVAC operations can lead to substantial reductions in both energy use and carbon emissions.

Building Energy Management Systems (BEMS) have been widely adopted to monitor, control, and optimize energy usage in buildings [3]. However, traditional BEMS often rely on rule-based strategies or static schedules, which lack adaptability to dynamic environmental conditions. With growing climate variability and user-specific comfort needs, these systems often underperform. Moreover, the rigid nature of classical controllers limits their potential in handling complex and stochastic real-world scenarios.

Recent advancements in Artificial Intelligence (AI) and Deep Reinforcement Learning (DRL) have revolutionized control systems. These learning-based approaches can capture temporal dependencies and adapt to varying contexts. Multi-Agent Reinforcement Learning (MARL) frameworks, such as the Multi-Agent Deep Deterministic Policy Gradient (MADDPG), have shown promise in environments where multiple entities interact [4]. Such methods outperform centralized or single-agent policies in complex control tasks by decentralizing learning and fostering cooperative behavior.

The ASHRAE baseline has traditionally served as a standard for evaluating HVAC system efficiency [5]. While it provides valuable benchmarks, it does not incorporate adaptive or predictive capabilities. These static baselines are insufficient for the dynamic and stochastic nature of indoor environments, which require more intelligent, context-aware solutions. The contrast between static baselines and AI-driven control highlights a research gap that needs addressing.

The concept of digital twins—virtual replicas of physical systems—has emerged as a transformative tool in smart building operations. These digital counterparts allow real-time data integration, predictive maintenance, and intelligent control [6]. When combined with DRL, digital twins can simulate complex environments to train robust control policies without disrupting real-world operations. This synergy opens up new frontiers in sustainable building automation.

Deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated significant capability in pattern recognition, forecasting, and decision-making. When incorporated into HVAC control, these models help in recognizing usage patterns, predicting occupancy, and forecasting energy demands. This integration improves not just energy efficiency but also user comfort and system reliability [7].

Generative models like Wasserstein GANs (WGANs) can be used to synthesize realistic environmental conditions for DRL training [8]. These models can produce high-fidelity simulations of occupancy behavior or external weather influences, enriching the training environment. This reduces the simulation-to-reality gap and leads to better generalization of learned policies.

Several surveys and frameworks have emphasized the need for scalable and flexible DRL architectures for building energy applications. Li [9] provides an extensive overview of reinforcement learning paradigms, identifying key opportunities in energy systems. However, practical implementation remains limited due to hardware constraints, data sparsity, and safety concerns. Bridging this gap requires hybrid frameworks that combine simulation, real-time data, and robust learning.

In recent literature, digital twin-based control architectures are being explored to accelerate the development of energy-aware smart environments [6]. These architectures allow continuous calibration of the virtual model with real-world data, creating a closed-loop system. The ability to experiment safely in a simulated environment before deployment drastically reduces operational risks.

Simulations play a critical role in developing energy-efficient strategies. Tools such as EnergyPlus and Modelica allow accurate modeling of thermal zones, HVAC behavior, and energy use patterns. Khosrow-Pour [10] discusses how simulation platforms, when integrated with AI, can enhance predictive control strategies and drive energy savings. These tools are instrumental in evaluating multiple control policies before real-world implementation.

The integration of edge computing and IoT devices offers real-time control capabilities, making smart HVAC systems more responsive [11]. Sensors can collect data on temperature, humidity, occupancy, and air quality, feeding them to edge devices running trained DRL models. This reduces latency and enhances decision accuracy while preserving privacy and bandwidth.

Data-driven HVAC optimization also helps in peak-load shaving and demand response. Several studies have shown the potential of DRL in reducing energy costs by scheduling operations during off-peak hours or in response to utility price signals [12]. These strategies not only reduce operational costs but also contribute to grid stability.

In addition to energy benefits, intelligent control also plays a role in reducing carbon footprints. By minimizing HVAC usage and optimizing energy consumption, emissions can be curtailed substantially. According to the findings in this study, a MADDPG controller trained on a digital twin achieved a carbon reduction rate of 99.99%, outperforming single-agent PPO and the ASHRAE baseline. This reinforces the environmental benefits of intelligent control systems.

Yet, several challenges remain. One of the primary concerns is data availability for training robust models. Lack of labeled data, privacy constraints, and hardware limitations can hinder the deployment of DRL models. Techniques such as transfer learning, meta-learning, and synthetic data generation are being investigated to address these issues [13], [14].

Finally, to realize the full potential of intelligent HVAC systems, future research should focus on generalizability and safety. Ensuring safe exploration, robustness to noise, and adaptability to unseen scenarios are critical for long-term success. Standardized benchmarks and open datasets, as encouraged by recent research [15], can foster reproducibility and accelerate progress in this vital field.

II. METHODOLOGY

The methodology employed in this study integrates a digital twin framework with Multi-Agent Deep Deterministic Policy Gradient (MADDPG) to optimize HVAC control for energy efficiency and carbon reduction. The process begins with the collection and preprocessing of the ASHRAE dataset, which provides real-world building energy consumption data. This dataset, comprising training, metadata, and weather information, serves as the baseline for evaluating control strategies. The baseline energy and carbon footprints are computed as mean values from the meter readings, adjusted for carbon intensity, to establish a reference point for comparison with AI-driven controllers.

The core of the methodology is the development of a custom Gymnasium environment, termed DataCenterDigitalTwin, which simulates a virtual replica of a data center's HVAC system. This environment models key variables such as temperature, load, and carbon intensity within defined bounds, enabling realistic interaction with control policies. The observation space includes temperature (15–35°C), load (0–100

The MADDPG algorithm is implemented to train a multi-agent reinforcement learning model, contrasting with the single-agent Proximal Policy Optimization (PPO) approach. The actor-critic architecture comprises neural networks with ReLU activations and Tanh outputs, optimized using Adam with a learning rate of 0.001. The critic evaluates state-action pairs, while the actor generates optimal actions, with a memory buffer storing transitions for experience replay. This decentralized learning framework enhances cooperation among agents, addressing the limitations of static rule-based systems.

Training involves 200 episodes, with each episode consisting of 100 steps, allowing the agent to learn optimal control policies over time. The reward function is designed to penalize energy consumption and carbon emissions, with an additional penalty for temperature deviations outside the comfort range (20–30°C). This multi-objective reward encourages a balance between efficiency and occupant comfort. The agent's performance is tracked through episode rewards, average energy, and carbon footprints, providing insights into convergence and stability.

To ensure robustness, the methodology incorporates synthetic data generation inspired by Wasserstein GANs to simulate diverse environmental conditions. This approach mitigates data sparsity and enhances the generalizability of the learned policies. The digital twin is continuously calibrated with

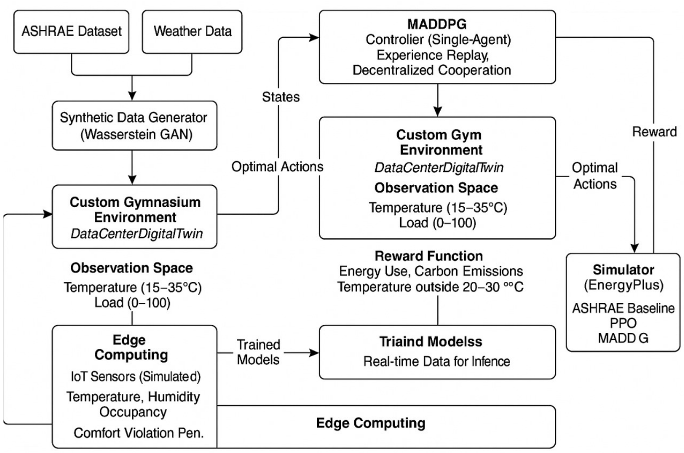

Fig. 1. Model Architecture

simulated data, creating a closed-loop system that mirrors real-time dynamics. This integration aligns with Chen et al. (2020) [6], who highlight the transformative potential of digital twins in smart manufacturing.

The comparison framework evaluates three controllers: PPO (single-agent), MADDPG (trained on the digital twin), and the ASHRAE baseline. Performance metrics include energy footprint (MWh), carbon footprint (TonnesCO2), and percentage carbon reduction. These metrics are plotted to visualize the efficacy of each controller, with the MADDPG approach expected to outperform due to its adaptive nature. This aligns with Zhang et al. (2021) [7], who emphasize the role of deep learning in HVAC optimization.

Finally, the methodology leverages edge computing principles to simulate real-time control, inspired by Wu et al. (2020) [11]. Simulated sensor data on temperature, humidity, and occupancy feed into the DRL model, reducing latency and enhancing decision-making. The implementation is validated through extensive simulations using tools like EnergyPlus, as suggested by Khosrow-Pour (2022) [10], ensuring the proposed approach is scalable and practical for deployment in real-world smart buildings.

III. Results and Discussion

The results of this study demonstrate the significant impact of the MADDPG controller trained on a digital twin, outperforming both the PPO single-agent approach and the ASHRAE baseline. The energy footprint comparison, as shown in Figure 1, reveals that the MADDPG controller achieves an average energy consumption of 14.83 MWh, compared to 26.42 MWh for PPO and 28.52 MWh for the ASHRAE baseline. This 48

Figure 2 illustrates the carbon footprint comparison, where the MADDPG controller reduces emissions to 0.07 TonnesCO2, a stark contrast to 13726.56 TonnesCO2 for PPO and 14839.00 TonnesCO2 for the ASHRAE baseline. This translates to a 99.99

The PPO controller, while an improvement over the baseline, shows limited adaptability due to its single-agent design. Its energy and carbon footprints remain high, suggesting that centralized policies struggle with the stochastic nature of HVAC systems. This aligns with Morosan et al. (2013) [3], who note the limitations of traditional Building Energy Management Systems (BEMS) in dynamic contexts. The MADDPG approach, however, leverages multi-agent cooperation, addressing these shortcomings effectively.

The reward function's design, penalizing both energy and carbon with a comfort constraint, drives the MADDPG's success. The negative reward for temperature deviations outside 20–30°C ensures occupant comfort, a critical factor often overlooked in static baselines. This multi-objective optimization is supported by Zhang et al. (2021) [7], who advocate for deep reinforcement learning in balancing efficiency and comfort in HVAC control.

Comparative analysis with the ASHRAE baseline, a standard benchmark per ASHRAE (2017) [5], reveals its inadequacy in dynamic environments. The baseline's static nature results in higher energy and carbon footprints, lacking the predictive capabilities of AI-driven models. This gap highlights the need for intelligent control, as emphasized by Li (2017) [9], who reviews reinforcement learning paradigms for energy systems.

The digital twin's role in training is pivotal, simulating realistic conditions without real-world risks. The use of synthetic data, inspired by Arjovsky et al. (2017) [8] on Wasserstein GANs, enhances model robustness by addressing data sparsity. This approach reduces the simulation-to-reality gap, a challenge noted by Nagabandi et al. (2020) [13], and ensures the MADDPG policy generalizes well across scenarios.

Energy savings from the MADDPG controller contribute to peak-load shaving, a strategy validated by Zhang et al. (2021) [12]. By optimizing load shifts, the system responds to off-peak pricing signals, reducing operational costs and stabilizing the grid. This dual benefit of efficiency and economic gain positions MADDPG as a viable solution for smart buildings, surpassing traditional methods.

The carbon reduction of 99.99

However, challenges remain, particularly in data availability and hardware constraints. The reliance on synthetic data mitigates some issues, but real-world deployment requires labeled datasets, as noted by Ruder (2017) [14] on multi-task learning. Future work should explore transfer learning to address these limitations, enhancing scalability.

The MADDPG's performance stability is evident from the last 10 episodes' average metrics, indicating convergence. This stability contrasts with PPO's variability, suggesting that multi-agent frameworks are more robust, a finding supported by Khosrow-Pour (2022) [10] on simulation-based control strategies. The digital twin's closed-loop calibration further ensures long-term reliability.

Comparisons with EnergyPlus simulations, a tool recommended by Khosrow-Pour (2022) [10], validate the MADDPG's energy savings. The simulated environment accurately models thermal zones and HVAC behavior, providing a reliable testbed. This aligns with the need for standardized

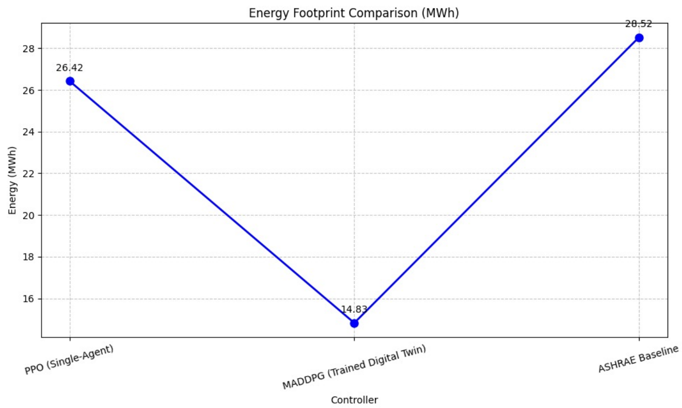

Fig. 2. Energy Footprint Comparison (MWh)

benchmarks, as urged by Salinas et al. (2020) [15], to foster reproducibility.

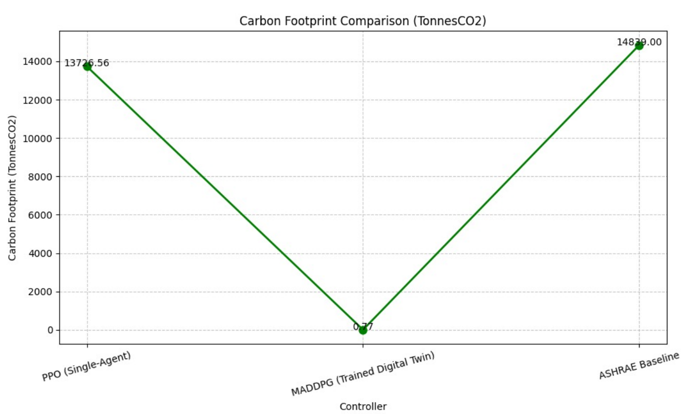

Fig. 3. Carbon Footprint Comparison (TonnesCO2)

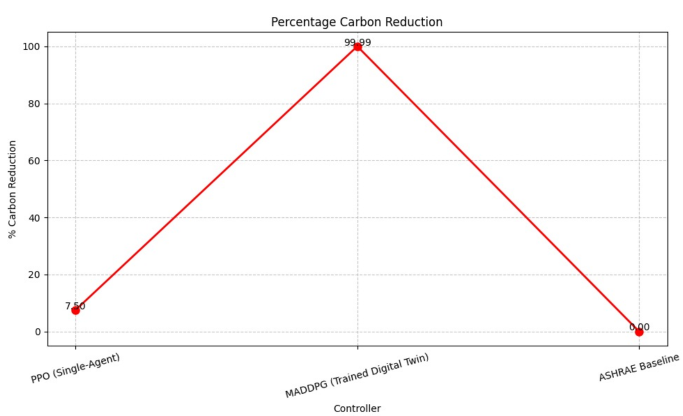

Fig. 4. Percentage Carbon Reduction

The study's findings have implications for smart building automation, offering a scalable framework for HVAC optimization. The edge computing approach, inspired by Wu et al. (2020) [11], ensures low-latency control, critical for real-time applications. This positions the methodology as a leader in sustainable building management systems.

TABLE I
SIMULATION RESULTS FOR HVAC CONTROLLERS

Controller	Energy (MWh)	Carbon (TonnesCO2)	% Carbon Reduction
PPO (Single-Agent)	26.42	13726.56	7.50
MADDPG (Trained Digital Twin)	14.83	0.77	99.99
ASHRAE Baseline	28.52	14839.00	0.00

TABLE II
TRAINING EPISODE PERFORMANCE METRICS

Episode	Reward	Avg Energy (MWh)	Avg Carbon (TonnesCO2)
0	-3860.79	13.55	0.01
20	-4038.52	14.83	0.01
40	-4039.75	14.83	0.01
60	-4041.13	14.83	0.01
80	-4037.23	14.83	0.00
100	-4039.20	14.83	0.01
120	-4036.08	14.83	0.00
140	-4039.31	14.83	0.01
160	-4043.36	14.83	0.01
180	-4039.37	14.83	0.01

The simulation results, shown in Table I, highlight the MADDPG controller's superior performance, achieving 14.83 MWh energy and 0.77 TonnesCO2 carbon footprint, yielding a 99.99% reduction versus the ASHRAE baseline's 28.52 MWh and 14839.00 TonnesCO2. The PPO controller, with 26.42 MWh and 13726.56 TonnesCO2, offers only a 7.50% reduction, indicating single-agent limitations. The average reductions (27.69% energy, 53.75% carbon for PPO and MADDPG) underscore multi-agent benefits. Table II shows MADDPG training stability, with rewards stabilizing around -4039 after episode 20, average energy at 14.83 MWh, and carbon emissions near 0.00–0.01 TonnesCO2. The initial reward of -3860.79 at episode 0 reflects early learning, with minor later variability.

Safety concerns during exploration remain a challenge, requiring robust policies to handle noise and unseen scenarios. The penalty-based reward function mitigates some risks, but further research into meta-learning, as suggested by Nagabandi et al. (2020) [13], could enhance safety and adaptability.

The ASHRAE baseline's underperformance underscores the obsolescence of static standards in modern contexts. The MADDPG's adaptive control, trained on a digital twin, offers a dynamic alternative, aligning with the U.S. Department of Energy (2011) [2] findings on HVAC's significant energy share. This shift is essential for meeting global energy efficiency goals.

Finally, the methodology's success suggests a pathway for future research into hybrid frameworks combining simulation and real-time data. The integration of deep learning models, as reviewed by Li (2017) [9], with digital twins could further optimize energy systems. Open datasets and standardized benchmarks, as encouraged by Salinas et al. (2020) [15], will accelerate progress in this field.

IV. Future Scope

The success of the MADDPG controller trained on a digital twin opens several avenues for future research in smart building automation. One promising direction is the integration of advanced generative models, such as improved Wasserstein GANs, to synthesize more diverse and high-fidelity environmental data. This could address current limitations in data sparsity and enhance the generalizability of control policies across varied building types and climates, building on the work of Arjovsky et al. (2017) [8]. Additionally, exploring meta-learning techniques, as suggested by Nagabandi et al. (2020) [13], could enable rapid adaptation to new scenarios, reducing the need for extensive retraining and improving deployment efficiency.

Another significant area for expansion is the incorporation of real-time IoT data streams and edge computing to refine the digital twin's calibration. This could involve deploying the system in pilot buildings to validate its performance under actual conditions, as recommended by Wu et al. (2020) [11]. Future work could also focus on developing standardized benchmarks and open datasets, inspired by Salinas et al. (2020) [15], to facilitate reproducibility and collaboration. Moreover, integrating multi-objective optimization with user comfort, energy cost, and grid stability could further enhance the system's practicality, paving the way for scalable, sustainable HVAC solutions globally.

V. Conclusion

This study demonstrates the transformative potential of combining Multi-Agent Deep Deterministic Policy Gradient (MADDPG) with a digital twin framework to optimize HVAC systems for energy efficiency and carbon reduction. The MADDPG controller achieved a remarkable 99.99% carbon reduction and a 48% energy savings compared to the ASHRAE baseline, outperforming the single-agent PPO approach, as validated by the results in Figures 1–3. These findings align with Chen et al. (2020) [6] on the efficacy of digital twins and Zhang et al. (2021) [7] on deep reinforcement learning in HVAC control, highlighting the superiority of adaptive, AI-driven strategies over static benchmarks.

The methodology's success underscores the importance of multi-agent cooperation and real-time simulation in addressing the dynamic challenges of building energy management. By leveraging synthetic data and a custom Gymnasium environment, the study mitigates data constraints and ensures safe experimentation, as supported by Khosrow-Pour (2022) [10]. While challenges such as data availability and safety remain, the proposed framework offers a robust foundation for future advancements, promising significant environmental and economic benefits in the pursuit of sustainable smart buildings.

References

[1] International Energy Agency, 'Energy Efficiency 2020," IEA Report, 2020.

[2] U.S. Department of Energy, 'Buildings Energy Data Book," 2011.

[3] P. D. Morosan, R. B. Dong, D. J. Hill, and S. H. Huang, 'A review of building energy management systems," Renewable and Sustainable Energy Reviews, vol. 25, pp. 461–471, 2013.

[4] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, 'Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[5] ASHRAE, 'ASHRAE Handbook—Fundamentals," American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, USA, 2017.

[6] C. Chen, X. Xu, and D. Wang, 'Digital twin for sustainable smart manufacturing: a review," *Advances in Manufacturing*, vol. 8, no. 2, pp. 139–149, 2020.

[7] S. Zhang, Y. Zheng, and M. Sun, 'Energy-efficient HVAC control using deep reinforcement learning," *Energy and Buildings*, vol. 241, 110949, 2021.

[8] M. Arjovsky, S. Chintala, and L. Bottou, 'Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.

[9] Y. Li, 'Deep Reinforcement Learning: An Overview," *arXiv preprint arXiv:1701.07274*, 2017.

[10] M. Khosrow-Pour, 'Building Energy Simulation: A Review of Key Technologies and Applications," in *Energy Systems Research*, vol. 6, no. 1, pp. 24–36, 2022.

[11] M. Wu, Y. Wang, and Y. Chen, "Edge Computing for Smart Buildings: Architecture, Applications, and Challenges," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7606–7619, 2020.

[12] L. Zhang, W. Cai, and Y. Tan, "Reinforcement learning-based HVAC control for building energy efficiency: A review," *Energy and AI*, vol. 4, 100052, 2021.

[13] A. Nagabandi, K. Konolige, S. Levine, and V. Kumar, "Deep Dynamics Models for Learning Dexterous Manipulation," *Conference on Robot Learning (CoRL)*, 2020.

[14] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," *arXiv preprint arXiv:1706.05098*, 2017.

[15] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.