Efficient Reinforcement Learning Implementations for Sustainable Operation of Liquid Cooled HPC Data Centers

Avisek Naug[1], Antonio Guillen-Perez[1], Vineet Gundecha[1], Ashwin Ramesh Babu[1], Sahand Ghorbanpour[1], Ricardo Luna Gutierrez[1], Soumyendu Sarkar[1*]

[1] HPE Labs, Hewlett Packard Enterprise
{avisek.naug, antonio.guillen, vineet.gundecha, ashwin.ramesh-babu, sahand.ghorbanpour, rluna, soumyendu.sarkar}@hpe.com

Abstract

The rapid growth of data-intensive applications like AI has led to a significant increase in the energy consumption and carbon footprint of data centers. Liquid cooling has emerged as a crucial technology to manage the thermal loads of high-density servers more efficiently than traditional air cooling. However, optimizing the complex dynamics of liquid cooling systems to maximize energy efficiency remains a significant challenge. To accelerate research in this domain, we design a suite of highly scalable reinforcement learning (RL) control strategies for liquid-cooled data centers. We demonstrate our work on a digital twin of the Oak Ridge National Laboratory's Frontier supercomputer cooling system that provides a detailed, customization, and scalable platform for end-to-end liquid cooling control. We demonstrate the utility of our framework by developing and evaluating centralized and decentralized multi-agent RL controllers that optimize cooling tower and server-level operations. Our results show centralized RL-based control can significantly improve operational carbon footprint and thermal management compared to traditional RL applications in literature, thereby offering a promising path toward more sustainable data centers and mitigating their climate impact.

1 Introduction

Controlling the dynamic thermal environment in liquid-cooled HPC data centers is a challenging problem. Current industrial cooling often uses static strategies [1], failing to leverage the full energy-saving potential of liquid cooling [2]. While deep reinforcement learning (RL) has shown promise in optimizing cooling operations and achieving 10-14% energy and carbon reductions [3, 2, 4, 5, 6, 7, 8, 6], existing RL approaches struggle to scale to the complexity of modern HPC environments [9, 10, 11, 12] and face efficiency bottlenecks that hinder real-time application [3, 2].

This work addresses these gaps by introducing and validating a scalable, multi-agent RL architecture for end-to-end control of liquid-cooled HPC data centers. Grounded in a digital twin of the Oak Ridge National Laboratory's Frontier system, our approach introduces a novel batching and multi-head policy infrastructure to enable efficient, parallelized inference across thousands of actuators without compromising performance or safety.

*Corresponding author.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

2 System Overview

The testbed environment models an end-to-end liquid cooling system, from the site-level cooling towers to the data center cabinets and server blade groups. Figure 1 provides a system overview. The system consists of Cooling Towers, which reject heat to the environment, and Cooling Distribution Units (CDUs) that manage the coolant for the HPC server cabinets. The benchmark supports customizable data center setups, including the number of cooling towers, cabinets, and blade groups. It also includes a Heat Recovery Unit (HRU) model, which allows for the evaluation of strategies for reusing waste heat, further enhancing the system's sustainability.

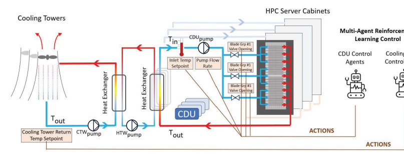

Figure 1: System Overview of end-to-end Control of Liquid Cooled Data Center. The CDU RL agents control the HPC server cabinets. The Cooling Towers are controlled by the CT RL agents.

3 Modeling and Control Interface

The digital twin environment uses Modelica-based models for all its components, ensuring a high-fidelity representation of the thermo-fluidic dynamics. To enable ML-based control, the model is exported to a Functional Mockup Unit (FMU), which integrates with Python frameworks like Gymnasium [13]. We focus on two primary control problems: **Cooling Tower Control:** Minimize the energy consumption of the cooling towers while ensuring adequate heat rejection. The corresponding Markov Decision Process (MDP) is detailed in Table 4 in the Appendix. Secondly, **Blade Group Level Control:** Maintain the operating temperatures of the server blade groups within optimal ranges to ensure reliability and performance. The MDP for this is shown in Table 5 in the Appendix.

4 RL Design for Efficient Control at HPC scale

We employ an efficient multi-agent reinforcement learning (MARL) approach to control the liquid cooling system. The Cooling Tower (CT) and Blade Group (BG) are controlled by independent multihead agents, addressing the scalability challenges of a centralized controller.

Centralized Action Execution in Multi-agent RL To improve the efficiency of the multi-agent setup, we implement a centralized action execution approach. This involves batching observations from similar entities (e.g., all cooling towers or all blade groups in a cabinet) and performing a single forward pass through the policy network. This approach, detailed in Figure 2 in the Appendix, significantly speeds up the inference process during rollouts without sacrificing the decentralized nature of the control policy.

Improved Reward Feedback via Multi-Head Policy For the Blade Group MDP, we utilize a multi-headed policy architecture. One head of the policy network determines the coolant supply temperature setpoint and pump flow rate, while the second head controls the valve openings for each blade group. This decomposition of the action space provides more direct reward feedback to each component of the policy, facilitating the discovery of more effective control strategies. The second head uses a Dirichlet distribution to ensure the valve openings are normalized, representing the proportional distribution of coolant.

5 Experiments and Results

We evaluate the performance of our RL-based control strategies against a baseline controller based on the industry standard ASHRAE Guideline 36 [1]. The experiments were conducted on a simulated data center with 2 cooling towers, 5 cabinets, and 3 blade groups per cabinet.

Ablation Study of RL Agent Design Table 1 presents an ablation study of different RL agent designs. The baseline control (Case 1) achieves a blade group temperature compliance of 76.92%. Introducing RL for the cooling tower alone (Case 2) slightly improves temperature compliance but increases power consumption. RL control at the blade group level (Cases 3 and 4) shows the potential for power savings. The multi-agent RL approaches (Cases 5-7) demonstrate significant improvements in both temperature compliance and energy efficiency. The multi-agent RL with centralized action and a multi-head policy (Case 7) achieves the best performance, with 95.63% temperature compliance and the lowest cooling tower power consumption. This performance improvement is further visually represented in Figure 3.

Table 1: **Ablation of RL Agent Design.** We incrementally replace the static baseline (Case 1) with RL controllers for: Cooling Tower (Case 2), CDU coolant setpoint/flow (Case 3), and Blade Group valves (Case 4). Case 5 introduces a single multi-agent RL controller, Case 6 adds batching for state space reduction, and Case 7 uses a multi-head policy. Experiments use N=2 towers, m=2 cells, C=5 cabinets with B=3 blade groups each, and are evaluated on an unseen exogenous trace. Blade-group temperature compliance $D_{blade,avg}$ is computed with $\mathcal{U}_T$=40°C and $\mathcal{L}_T$=20°C.

Metric →		$D_{blade,avg}$%	$\sum P_{ij}(kW)$	$\sum Q_i$	**Avg Episode Reward**	
Agent/Control Type ↓	**Control Details**	(% of time Temp within ideal range)	(Cooling Tower Avg Power)	(IT Level Avg Cooling Power)	**per** Cabinet	**per** Cooling Tower
1. Baseline Control	ASHRAE G36	76.92	237.31	235.28	1697.08	360.17
2. CT RL + BG Baseline	Only CT RL control	79.21	246.46	235.03	1702.16	352.28
3. CT Baseline + BG RL	No Valve Control	64.91	217.6	203.96	1638.48	372.97
4. CT Baseline + BG RL	With Valve Control	77.13	217.37	211.83	1698.36	373.52
5. Multiagent RL	Decentralized Action	78.24	218.11	212.94	1697.49	370.51
6. Multiagent RL	Centralized Action (CA)	90.46	207.37	208.69	1714.65	395.88
7. Multiagent RL	CA & Multihead policy	**95.63**	**206.52**	**197.18**	**1726.31**	**396.24**

Table 2: **Performance on Scale.** Evaluation of Rule-Based Control vs Multi-head Centralized Action Policy for Scaling of Cooling Tower Agent and Multi-head Blade-Group Agent with increasing Data Center sizes. Blade-Group Agent is trained on N=2 Cooling Towers, m=2 Cells per Tower, C=5 Cabinets, B=3 Blade Groups per Cabinet

Metric		N=2, m=2 C=10, B=3	N=2, m=2 C=15, B=3
$D_{\text{blade,avg}}$ %	ASHRAE G36	71.92	68.24
	CA Policy	96.28	86.19
$CFP(kgCO_2)$	ASHRAE G36	4448.96	6254.15
	CA Policy	4432.78	6392.21
Metric		**N=3, m=2 C=20, B=3**	**N=4, m=2 C=25, B=3**
$D_{\text{blade,avg}}$ %	ASHRAE G36	75.31	83.08
	CA Policy	94.07	92.61
$CFP(kgCO_2)$	ASHRAE G36	10518.78	15753.90
	CA Policy	9932.36	12652.18

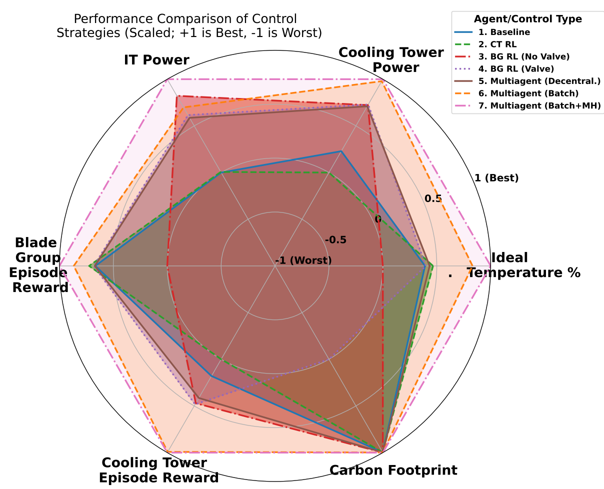

Table 3: Relative Performance of different RL approaches from Table 1 for $N = 2$ towers, $m = 2$ cells, $C = 5$ cabinets, with $B = 3$ blade groups in each cabinet

Sustainability on Scale We also evaluate the carbon footprint scalability of our approach by increasing the size of the data center. Table 2 shows that the multi-head centralized action RL policy consistently outperforms the ASHRAE G36 baseline in terms of temperature compliance across different data center scales. The RL policy also demonstrates significant carbon footprint savings, especially in larger configurations.

6 Conclusion and Climate Change Impact

We have introduced a suite of efficient RL agent implementations in PPO for developing and evaluating energy-efficient liquid cooling control strategies for data centers. Our experiments demonstrate the potential for significant energy and carbon footprint savings and improved thermal management compared to traditional RL methods. The significant carbon footprint reductions achieved by our RL agents offer a clear path toward more practical RL implementations for sustainable data centers. This work provides a valuable collection of RL-enabled solutions that can help tackle climate change by reducing the environmental footprint of our digital infrastructure.

References

[1] ASHRAE. Guideline 36: Best in Class HVAC Control Sequences, 2025. [Online; accessed 12. May 2025].

[2] Haoran Chen, Yong Han, Gongyue Tang, and Xiaowu Zhang. A dynamic control system for server processor direct liquid cooling. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 10(5):786–794, 2020.

[3] Jerry Luo, Cosmin Paduraru, Octavian Voicu, et al. Controlling commercial cooling systems using reinforcement learning. arXiv preprint arXiv:2211.07357, 2022.

[4] Avisek Naug, Antonio Guillen, Ricardo Luna, Vineet Gundecha, Cullen Bash, Sahand Ghorbanpour, Sajad Mousavi, Ashwin Ramesh Babu, Dejan Markovikj, Lekhapriya D Kashyap, Desik Rengarajan, and Soumyendu Sarkar. Sustaindc: Benchmarking for sustainable data center control. In *Advances in Neural Information Processing Systems*, volume 37, pages 100630–100669. Curran Associates, Inc., 2024.

[5] Soumyendu Sarkar, Avisek Naug, Ricardo Luna, Antonio Guillen, Vineet Gundecha, Sahand Ghorbanpour, Sajad Mousavi, Dejan Markovikj, and Ashwin Ramesh Babu. Carbon footprint reduction for sustainable data centers in real-time. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20):22322–22330, Mar. 2024.

[6] Antonio Guillen-Perez, Avisek Naug, Vineet Gundecha, Sahand Ghorbanpour, Ricardo Luna Gutierrez, Ashwin Ramesh Babu, Munther Salim, Shubhanker Banerjee, Eoin H. Oude Essink, Damien Fay, and Soumyendu Sarkar. Dccluster-opt: Benchmarking dynamic multi-objective optimization for geo-distributed data center workloads. *arXiv e-prints*, 2025.

[7] Soumyendu Sarkar, Avisek Naug, Antonio Guillen, Vineet Gundecha, Ricardo Luna Gutiérrez, Sahand Ghorbanpour, Sajad Mousavi, Ashwin Ramesh Babu, Desik Rengarajan, and Cullen Bash. Hierarchical multi-agent framework for carbon-efficient liquid-cooled data center clusters. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28):29694–29696, Apr. 2025.

[8] Avisek Naug, Antonio Guillen, Ricardo Luna Gutiérrez, Vineet Gundecha, Sahand Ghorbanpour, Lekhapriya Dheeraj Kashyap, Dejan Markovikj, Lorenz Krause, Sajad Mousavi, Ashwin Ramesh Babu, and Soumyendu Sarkar. Pydcm: Custom data center models with reinforcement learning for sustainability. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '23, page 232–235, New York, NY, USA, 2023. Association for Computing Machinery.

[9] Ali Habibi Khalaj and Saman K. Halgamuge. A Review on efficient thermal management of air- and liquid-cooled data centers: From chip to the cooling system. *Appl. Energy*, 205:1165–1188, November 2017.

[10] Soumyendu Sarkar, Avisek Naug, Antonio Guillen, Ricardo Luna, Vineet Gundecha, Ashwin Ramesh Babu, and Sajad Mousavi. Sustainability of data center digital twins with reinforcement learning. In *Proc. AAAI Conf. Artificial Intelligence*, volume 38, pages 22322–22330, 2024.

[11] Soumyendu Sarkar, Avisek Naug, Ricardo Luna Gutierrez, Antonio Guillen, Vineet Gundecha, Ashwin Ramesh Babu, and Cullen Bash. Real-time carbon footprint minimization in sustainable data centers with reinforcement learning. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023.

[12] Soumyendu Sarkar, Avisek Naug, Antonio Guillen, Ricardo Luna Gutierrez, Sahand Ghorbanpour, Sajad Mousavi, Ashwin Ramesh Babu, and Vineet Gundecha. Concurrent carbon footprint reduction (c2fr) reinforcement learning approach for sustainable data center digital twin. In *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, pages 1–8, 2023.

[13] Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, et al. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024.

Appendix: Centralized Action Execution

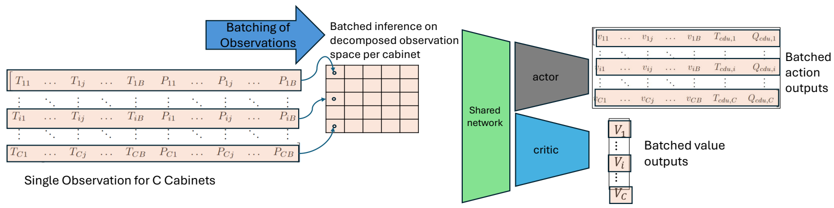

Figure 2: Centralized Action Execution Approach for scalable inference and rollouts at the CDU(s) and Blade Group(s) for HPC scale data center Digital Twins

Appendix: MDP Tables

Table 4: Cooling Tower MDP

MDP Attributes	Formulation	Remarks
State (s_t)	$\begin{bmatrix} P_{11} & \dots & P_{1j} & \dots & P_{1t}, & T_{ct,1} & \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \\ P_{i1} & \dots & P_{ij} & \dots & P_{it}, & T_{ct,i} & \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \\ P_{N1} & \dots & P_{Nj} & \dots & P_{Nt}, & T_{ct,N} & T_{wb} \end{bmatrix}$	P_{ij} refers to the power consumption of the j^{th} cell of the i^{th} cooling tower. $T_{ct,i}$ refers to the i^{th} cooling tower water return temperature. T_{wb} is the outside air wet bulb temperature
Action (a_t)	$\delta_1, \dots, \delta_i, \dots, \delta_N$	The agents sets the changes in cooling tower water return temperature setpoint $T_{ct,i}$ by δ_i across all the N cooling towers
Reward ($r_t(s_t, a_t, s_{t+1})$)	$-\sum_{i,j} P_{i,j}$	It is the sum total of the power consumption across all the cells for all the cooling towers

Table 5: Blade Group Level MDP

MDP Attributes	Formulation	Remarks
State (s_t)	$\begin{bmatrix} T_{11} & \dots & T_{1j} & \dots & T_{1B} & P_{11} & \dots & P_{1j} & \dots & P_{1B} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ T_{i1} & \dots & T_{ij} & \dots & T_{iB} & P_{i1} & \dots & P_{ij} & \dots & P_{iB} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ T_{C1} & \dots & T_{Cj} & \dots & T_{CB} & P_{C1} & \dots & P_{Cj} & \dots & P_{CB} \end{bmatrix}$	T_{ij} and P_{ij} refer to the temperature and thermal power input respectively of the j^{th} blade group of the i^{th} cabinet.
Action (a_t)	$\begin{bmatrix} v_{11} & \dots & v_{1j} & \dots & v_{1B} & T_{cdu,1} & Q_{cdu,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots \\ v_{i1} & \dots & v_{ij} & \dots & v_{iB} & T_{cdu,i} & Q_{cdu,i} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots \\ v_{C1} & \dots & v_{Cj} & \dots & v_{CB} & T_{cdu,C} & Q_{cdu,C} \end{bmatrix}$	$T_{cdu,i}$ and $Q_{cdu,i}$ refer to the liquid coolant supply temperature setpoint and pump flow rate of the i^{th} cabinet. v_{ij} refers to the valve actuation of the j^{th} blade group of the i^{th} cabinet
Reward ($r_t(s_t, a_t, s_{t+1})$)	$-\sum_{ij} T_{ij}$	It is the aggregate of the blade group operation temperatures

Appendix: Frontier Model

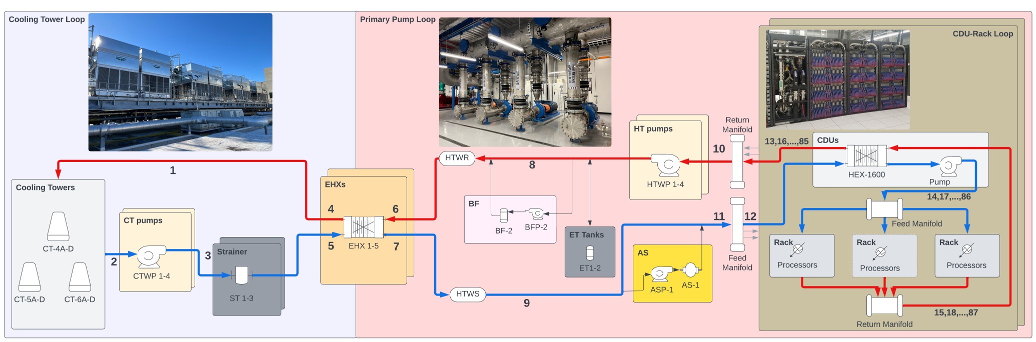

Figure 3: Frontier's Cooling System [Brewer et al. 2024]