Carbon Footprint Reduction for Sustainable Data Centers in Real-Time

Soumyendu Sarkar[*†], **Avisek Naug**[†], **Ricardo Luna**[†], **Antonio Guillen**[†], **Vineet Gundecha**[†], **Sahand Ghorbanpour, Sajad Mousavi, Dejan Markovikj, Ashwin Ramesh Babu**

Hewelett Packard Labs @ Hewlett Packard Enterprise
820 N McCarthy Blvd, Milpitas, CA 95035, USA
soumyendu.sarkar, avisek.naug, rluna, antonio.guillen, vineet.gundecha, sahand.ghorbanpour, sajad.mousavi, dejan.markovikj, ashwin.ramesh-babu@hpe.com

Abstract

As machine learning workloads significantly increase energy consumption, sustainable data centers with low carbon emissions are becoming a top priority for governments and corporations worldwide. This requires a paradigm shift in optimizing power consumption in cooling and IT loads, shifting flexible loads based on the availability of renewable energy in the power grid, and leveraging battery storage from the uninterrupted power supply in data centers, using collaborative agents. The complex association between these optimization strategies and their dependencies on variable external factors like weather and the power grid carbon intensity makes this a hard problem. Currently, a real-time controller to optimize all these goals simultaneously in a dynamic real-world setting is lacking. We propose a Data Center Carbon Footprint Reduction (DC-CFR) multi-agent Reinforcement Learning (MARL) framework that optimizes data centers for the multiple objectives of carbon footprint reduction, energy consumption, and energy cost. The results show that the DC-CFR MARL agents effectively resolved the complex interdependencies in optimizing cooling, load shifting, and energy storage in real-time for various locations under real-world dynamic weather and grid carbon intensity conditions. DC-CFR significantly outperformed the industry-standard ASHRAE controller with a considerable reduction in carbon emissions (14.5%), energy usage (14.4%), and energy cost (13.7%) when evaluated over one year across multiple geographical regions.

Introduction

In recent years, sustainability and carbon footprint reduction have emerged as critical factors driving the need for innovative optimization techniques in data center (DC) operations. While energy and cost optimization have been primary concerns in smart-grid problems, the increasing sustainability commitments of companies with large DCs have made carbon footprint reduction an essential target for the industry. Achieving significant carbon footprint savings requires reducing energy consumption and replacing carbon-intensive energy sources with those with a lower carbon footprint.

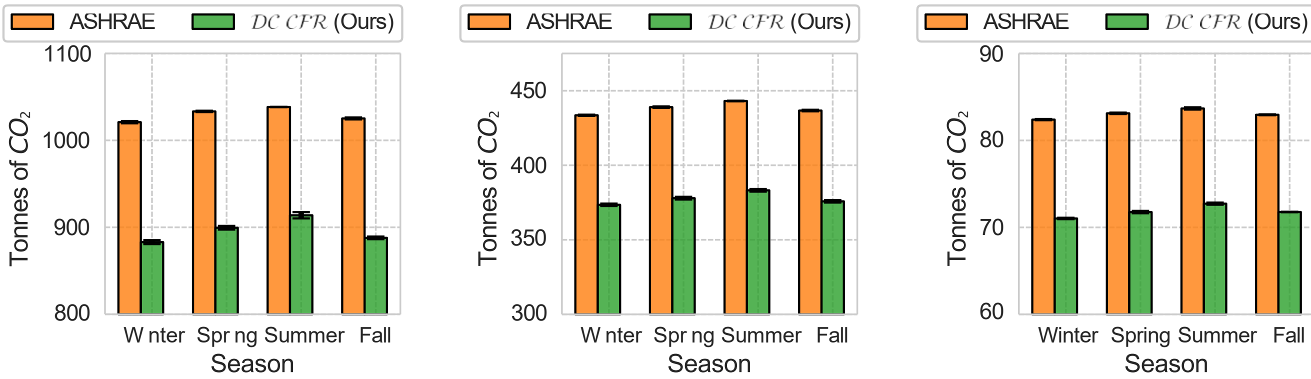
[*]Corresponding author.
[†]These authors contributed equally.

Static, isolated approaches for carbon footprint reduction, such as energy optimization, load shifting to less carbon-intensive hours, and battery usage for charging during low **power grid carbon intensity (CI)** hours to supplement load demand during high CI hours, are frequently used. However, achieving significant footprint savings with analytic pipeline-based planning has proven challenging due to the complexity of these individual problems and the reliance on long forecast horizons (24h) for static approaches. Furthermore, the dependencies between these approaches and the necessity of information exchange across separate problems have prevented the development of a cohesive strategy that can simultaneously reduce the carbon footprint using all three methods in real time.

In this paper, we introduce DC Carbon Footprint Reduction (DC-CFR), a novel framework that uses Multi-Agent Deep Reinforcement Learning (DRL) to optimize DC energy consumption, flexible load shifting, and battery operation decisions simultaneously in real time. The optimization is based on short-term weather and grid CI information. Grid CI refers to the amount of CO_2 emissions produced per unit of electricity consumed, which varies based on the source of the electricity (e.g., fossil fuels, renewable energy) at a given time. The lower the CI, the more renewable the energy source. Our approach effectively mitigates the drawbacks of existing, isolated methods. It does so by efficiently managing the complex interdependencies and information exchange among individual optimization strategies, a process that is illustrated in Figure 2 at a system level and in Figure 3 to show the dependencies.

The proposed contributions of the framework are as follows:

- A carbon emission-aware framework for controlling data centers by redistribution of server workloads, efficient cooling, and battery storage for auxiliary energy supply.
- Real-time control for the individual approaches under the framework, while coordinating between themselves using shared reward and state variables. The collaborative performance indicators help the agents self-adjust their operations.
- Implementation of the framework as a multi-agent reinforcement learning problem using industry-standard simulators for Load Shifting and Battery Supply from



(a) Arizona, USA.

(b) New York, USA.

(c) Washington, USA.

Figure 1: CO_2 generation (Tonnes) for data center (DC) control approaches in a 1.2 MWh DC in different locations. ASHRAE is an industry-standard controller for HVAC in DCs.

Meta (Acun et al. 2023) and Energy Plus software from NREL (Crawley et al. 2000) with a Sinergym Wrapper (Jime´nez-Raboso et al. 2021).

- Extensive evaluation of the approach across multiple geographical locations with different weather patterns and benchmark it against the current industry standard ASHRAE rule-based-controller (RBC).

Our approach can significantly decrease carbon emissions in data centers in the different tested locations (refer to Figure 1 for detailed results). Over a span of one year, and across different DC configurations and geographical locations, the DC-CFR framework demonstrated an average carbon footprint reduction of 14.46%.

In addition to reducing carbon emissions, our evaluations revealed the DC-CFR system's capacity to curtail energy consumption. On average, our tests showed a reduction in energy usage by 14.35%. Furthermore, the DC-CFR system also efficiently managed energy costs, reducing the average energy expenditure by 13.69%. These results underscore the potential of DC-CFR as a comprehensive and effective solution for achieving sustainability goals in DC operations.

Related Work

Energy Savings

Deep Reinforcement Learning (DRL) shows promise in dynamic thermal management in DCs, specifically for reducing energy consumption via Heating, Ventilation, and Air Conditioning (HVAC) system control (Zhang et al. 2023b,a; Mahbod et al. 2022; Biemann et al. 2021; Wang et al. 2022; Naug et al. 2023a,b). However, the real-world deployment of DRL-based systems is complicated by their sensitivity to hyperparameters, reward functions, and work scenarios (Zhang et al. 2023b; Mahbod et al. 2022; Biemann et al. 2021; Wang et al. 2022). Moreover, ensuring safety and satisfying operational constraints, especially for HVAC system control, is another challenge (Zhang et al. 2023b; Wang et al. 2022).

Despite the challenges, DRL has shown potential for energy savings in DCs. DRL-based strategies have achieved up to 12% savings compared to default controllers (Zhang et al. 2023b), and 8.84% compared to reference controllers. Additional savings of up to 5.5% have been noted in tropical climates (Mahbod et al. 2022), while in simulated environments, a reduction of at least 10% in energy consumption has been observed (Biemann et al. 2021).

Load Shifting and Battery Optimization

With DCs accounting for a significant portion of global energy consumption, Carbon-Aware Workload Scheduling (CAS) has emerged as a potential solution (Acun et al. 2023; Radovanovic´ et al. 2023). CAS uses delay-tolerant workloads to decrease carbon emissions by rescheduling them to times of lower CI. For instance, the Carbon Explorer framework (Acun et al. 2023) reduces the overall DC footprint by $\sim$ 4% on historical data by shifting the flexible part of the DC load to the lowest carbon-intensive hours.

DRL has been applied to optimize workload scheduling in DCs, improving energy efficiency (Ran et al. 2019, 2022a,b; Yi et al. 2019). One approach, GreenDRL, uses DRL for CAS, showing a reduction in the energy obtained from the main grid and an increase in the use of green energy (Zhang et al. 2022). However, GreenDRL primarily considers scenarios with on-site renewable energy resources.

Battery operation optimization is another area of focus, with strategies divided into static schedules based on day-ahead information and real-time control for when longer-term forecasts are unreliable (Acun et al. 2023; Radovanovic´ et al. 2023).

Real-time battery optimization strategies using DRL have been developed, but most overlook the degradation of battery charging and discharging rates across their instantaneous states of charge (Zhou, Zhou, and Yang 2022; Huang and Wang 2021; Abedi, Yoon, and Kwon 2022; Cao et al. 2020). For instance, a DRL agent for optimal battery operation assuming a battery degradation model has been developed, reducing net energy costs compared to a baseline battery operation algorithm (Cao et al. 2020).

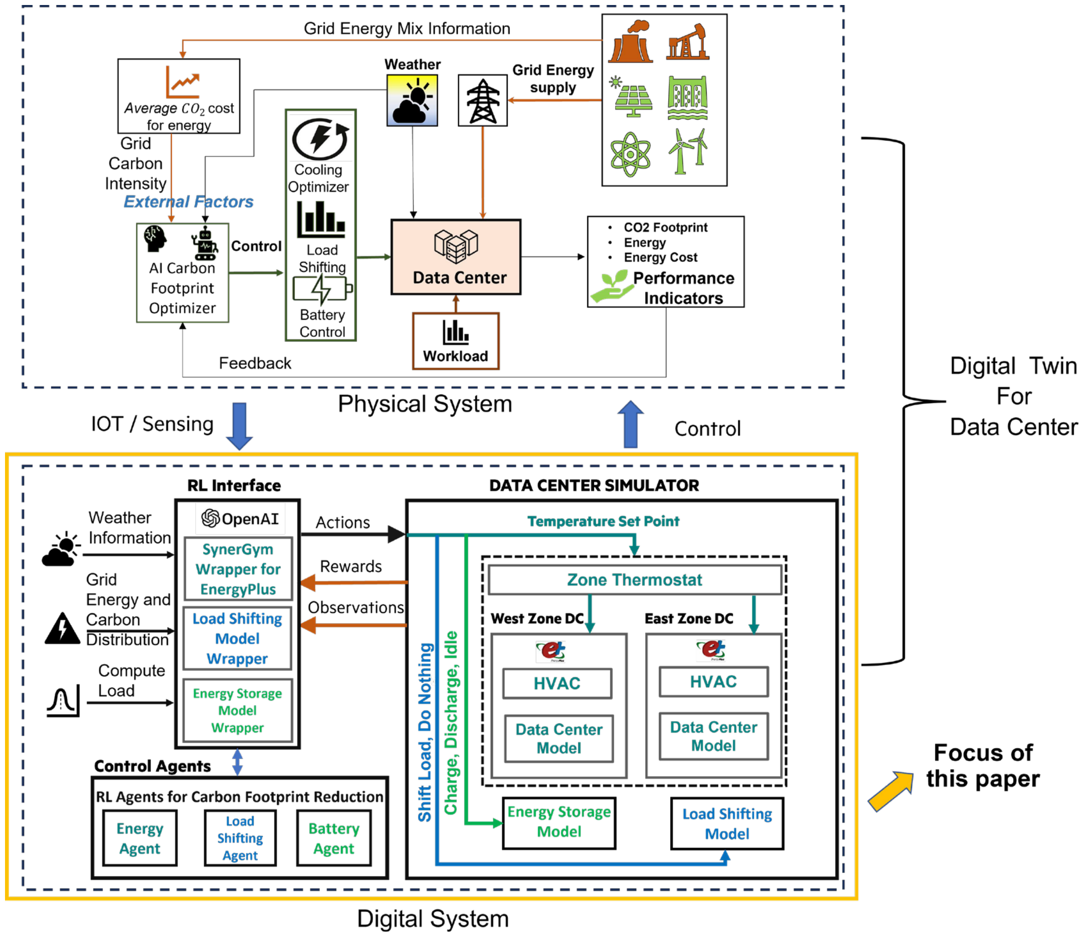


Figure 2: Overview of the physical and digital systems. For this work, we have used the simulation with EnergyPlus data center simulation from NREL, extended the RL interface with IBM's SinerGym, and used the battery model from Facebook.

Our Approach

While current carbon-reduction approaches show promise, they lack real-time operation capabilities and do not effectively combine multiple control strategies due to the complex interdependencies and balancing objectives. Our approach partially decouples the problem into sub-problems, each solved with an individual Markov decision process (MDP) formulation, the mathematical framework for RL, while the combined rewards and overlapping state variables in a collaborative multi-agent setting solve the dependencies in real-time. This design results in a comprehensive, real-time carbon footprint optimizer for sustainable societies, advancing beyond existing strategies to offer a more adaptable and robust solution.

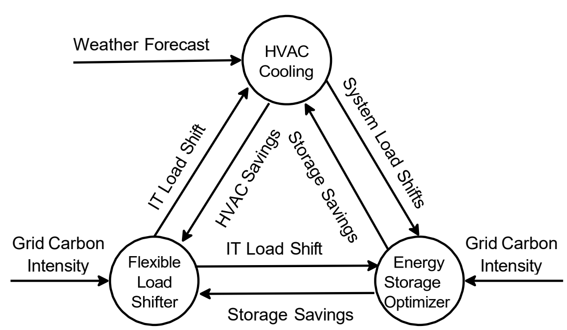


Figure 3: Internal and External Dependencies for the agents.

Problem Definition Using Markov Decision Processes

In this section, we present the problem formulation for reducing energy and carbon footprint in data centers. We do this by considering three MDPs that take into account DC workload shifting, energy reduction through cooling setpoint optimization, and auxiliary energy supply using energy storage systems. As mentioned earlier, these three problems have been tackled individually in the literature using static approaches (Acun et al. 2023; Radovanović et al. 2023) with day-ahead forecast information. We reformulate the problems so that they can be solved in real-time.

The three MDPs are described in Table 1 with their interdependencies summarized in Figure 3. Solving MDP_{LS} reduces the carbon footprint and energy consumption by shifting data center workloads to low-CI and low ambient temperature hours; MDP_E reduces data center energy use by optimizing the HVAC cooling setpoint; and MDP_{BAT} reduces the carbon footprint by charging and discharging the battery based on the grid CI.

As can be seen from the MDP description, the system involves a chain of dependencies starting with the *Flexible Load Shifter*, then proceeding through *HVAC Cooling*, and finally, Energy Storage Optimizer. The Load Shifter determines load shifts based on workload, grid CI, DC power usage, and thermal state. HVAC uses the resulting server workload and other factors like weather and battery charge to optimize energy with a surrogate model based on Energy Plus (Jime´nez-Raboso et al. 2021). This model estimates future energy use. The Energy Storage Optimizer uses this data and grid CI to decide battery actions. CO2Footprint rewards are calculated using CIt and TotalEnergyConsumptiont at each time step.

The interlinked dependencies offer potential energy and carbon savings but pose challenges for DRL agent training convergence. Challenges include changing state distributions with policy updates, different time constants for MDP changes, and state interdependencies. A collaborative reward framework is essential for agents to appropriately incorporate shared state variables from other MDPs into their rewards.

Multi-Agent Reinforcement Learning

Solving the challenges of sustainable data center operation requires addressing multiple interdependent sub-problems. Multi-Agent Reinforcement Learning (MARL) is a suitable approach due to the agents' cooperative nature. This allows individual agents to pursue individual objectives while considering other agents' actions through a reward mechanism, aiding collaboration. The study explores two MARL methods. We implemented Multi-Agent Deep Deterministic Policy Gradient (MADDPG) (Lowe et al. 2017), which involves decentralized agents learning from a centralized critic that incorporates all agents' behaviors. We also adapted the Independent Proximal Policy Algorithm (IPPO) Schulman et al. (2017) for independent yet collaborative agent actions. Both the RL algorithms converged to policies with similar performance.

Proposed Solution

Based on the above formulation for the MDPs for Load Shifting MDP_{LS}, Energy Reduction MDP_E, and Battery Operation MDP_{BAT} , we outline the DC Carbon Footprint Reduction (DC-CFR) multi-agent Reinforcement Learning algorithm (MARL) approach. The main goal is to efficiently reduce the total DC carbon footprint in real time by solving the MDPs simultaneously.

We undertake a systematic implementation of the multi-agent algorithm that accounts for the interdependency between individual agent actions and the next states.

Input: The operational/simulation diagram of the approach is described in detail in Figure 2. We simultaneously initialize the three DRL **control agents** A_{LS}, A_E and A_{BAT} for the optimal load shifting (MDP_{LS}), optimizing energy (MDP_E) and the optimal battery operation (MDP_{BAT}) respectively. Grid data, which includes grid energy and carbon distribution, weather conditions, and compute load, is configured to be queried from a database at every time step. On the other hand, the variables like DC temperature, IT Load, Unassigned flexible load, DC Energy, and Battery state of charge information are obtained via the exchange of information between the individual MDPs as a part of the **Data Center Simulator** which has the Load Shifting Model (Acun et al. 2023), Energy Plus model for the data center thermodynamics (Crawley et al. 2000) and Battery Storage Model (Acun et al. 2023; Sarkar et al. 2023b). The information exchange processes occur through the **RL Interface** with Open AI Gym wrappers. We shall provide detailed steps for this process when we describe the rollout stage for the agents.

Rollout: This is the real-time component of the approach. The agents A_{LS}, A_E and A_{BAT} are allowed to interact with their respective MDPs to collect rollout information $(S_t, A_t, S_t, R_{t+1}, \gamma, \text{done})$ in their respective memory buffers D. The different stages of the information exchange between MDPs are captured in the interdependency Fig. 3.

In the beginning, the agent A_{LS} considers the state variables time and CI. Unassigned flexible load is obtained from the MDP_{LS}, DC temperature, IT Load, and DC Energy are obtained from MDP_E, and Battery state of charge information is obtained from MDP_{BAT} . A_{LS} uses this information to decide its action on whether to reassign flexible load from this instant or to stay idle. The resulting IT load information is passed to the energy-optimizing agent A_E. It uses time as well as DC temperature, IT Load, DC Energy, and HVAC Setpoint obtained from the previous time step of MDP_E to decide the setpoint for the next time interval. The Energy Plus model of the DC calculates the resulting changes in energy consumption and DC temperature, and then communicates these back to MDP_E. Finally, the agent A_{BAT} considers the time, DC Energy from MDP_E, the current battery charge, and the CI information to decide on charging the battery from the grid or supplementing DC energy demand.

From an implementation perspective, the individual agents do not receive the rollout information tuple immediately after taking an action. They wait until it is their turn to take the action again. This incorporates the effect of its action in all other MDPs, making the reward more informed. We use a collaborative reward that considers the effects of actions from all agents. This formulation has been highlighted in Table 1. This enables the agents to look at the effects of their individual actions across dependent MDPs.

Concurrent Policy Update: At regular intervals, the buffer data is used to update the agent policies. For training the RL agents in this work, we are primarily using PPO (Schulman et al. 2017). Any other off-policy or on-policy agent may be used.

The overall DC-CFR approach is summarized in Algorithm 1.

	MDP_{LS} Flexible Load Shifting	**MDP_E** Energy HVAC Optimizer	**MDP_{BAT}** Battery Agent
State: S_t	Time, DC temperature, IT Load, Unassigned Flexible Load, DC Energy, Carbon Intensity, Battery Charge	Time, DC temperature, Weather, DC Energy, IT Load, HVAC Setpoint	Time, DC Energy, Battery Charge, Carbon Intensity
Action: A_t	Assign Flexible Load, Idle	HVAC Setpoint	Charge, Supply, Idle
Reward: $R_{t+1}(S_t, A_t)$	$0.8 * r_{LS} + 0.1 * r_E + 0.1 * r_{BAT}$	$0.1 * r_{LS} + 0.8 * r_E + 0.1 * r_{BAT}$	$0.1 * r_{LS} + 0.1 * r_E + 0.8 * r_{BAT}$

Table 1: MDPs for Load Shifting, HVAC Energy Optimization, and Battery Operation. Here $r_{LS} = -(CO_2 \text{ Footprint} + LS_{Penalty})$, $r_E = -(\text{Total Energy Consumption} \times \text{Cost per kWh})$, and $r_{BAT} = -(CO_2 \text{ Footprint})$, where $LS_{Penalty}$ is the scalar value of the unassigned flexible IT workload.

Algorithm 1: Data Center Carbon Footprint Reduction DC-CFR Multi-Agent Algorithm

Require: RL Agents A_{LS}, A_E and A_{BAT} ▷ RL Algorithm initialization
Require: CI ▷ Carbon Intensity data from the grid
Require: EW ▷ Weather data obtained from EnergyPlus
Require: Workload Model MDP_{LS} ▷ Model data center workload assignment
Require: Data Center Model MDP_E ▷ Model Data Center Thermodynamics in Energy Plus
Require: Battery Model MDP_{BAT} ▷ Model Battery operation
for i $\in 1, \ldots, L_b$ **do** ▷ L_b is the learning iterations budget
  Concurrent Rollout Phase
  while *episode not done* **do**
    State information is shared among the different MDPs
    Agent A_{LS} *sends action to* MDP_{LS} *and collects* $(s_t, a_t, s_{t'}, r_t, \gamma, \text{done})$ *in its replay buffer* D_{LS}
    Agent A_E *sends action to* MDP_E *and collects* $(s_t, a_t, s_{t'}, r_t, \gamma, \text{done})$ *in its replay buffer* D_E
    Agent A_{BAT} *sends action to* MDP_{BAT} *and collects* $(s_t, a_t, s_{t'}, r_t, \gamma, \text{done})$ *in its replay buffer* D_{BAT}
  end while
  Concurrent Policy Updates
  Update Agent Networks by training A_E *on* D_E, A_{LS} *on* D_{LS} *and* A_{BAT} *on* D_{BAT}
end for

Experiments

We conducted our experiments using EnergyPlus, an open-source building energy simulation software that can simulate the thermal performance of buildings and cooling systems. A two-zone DC HVAC with economizer model was used to simulate a DC consisting of two isolated zones with servers and HVAC cooling. We connected Python with EnergyPlus using the Sinergym framework (Jime´nez-Raboso et al. 2021), which wraps the EnergyPlus simulation engine following the OpenAI Gym interface to develop our control algorithms using DRL. This allows us to do step by step simulation of a DC and to dynamically change the cooling setpoint and the running workload of the DC. The load shifting and the battery models are based on work done in (Acun et al. 2023). These models are similarly wrapped via Open AI Gym interface.

For A_{LS}, as shown in (Acun et al. 2023), we set the flexible workload to constitute 10% of the DC's total daily workload. Moreover, the server capacity is at each time step is limited, preventing the assignment of all workloads in a single time slot. For A_{BAT}, we assume an installed battery capacity of 50% of DC max hourly energy consumption, as can be found in the uninterrupted power supply (UPS).

Our solution is designed with a reward signal that motivates the agents to reduce both energy consumption, carbon footprint and cost of energy. We have set the action interval at 15-minute time-step, which enables precise control of the system and to quickly respond to changes in the DC environment. We used IT load data of a large-scale real-world DC from the Alibaba (Cheng, Chai, and Anwar 2018) open source data set to improve the representativeness of our simulation.

We used New York weather and CI data to train our agents. To improve the generalizability of our solution, we employ an Ornstein-Uhlenbeck (OU) (Espen, Benth, and Šaltytė-Benth 2005) process to introduce noise into the weather data.

We tested the generality of our trained agents by evaluating their performance under diverse climatic and CI conditions. This was done using weather and CI data from three different locations: Arizona (AZ), New York (NY), and Washington (WA). These weather and CI files correspond to locations with distinct weather patterns, ranging from hot and arid to cold and humid. Additionally, we considered Time-of-Use rate plans for energy cost, where the cost vary with the hour.

By validating the model on various locations and weather conditions, we demonstrate the effectiveness of our ap-

	Percentage Reduction of Carbon Footprint with IPPO compared to ASHRAE Data Center Max Load 1.2MWh Experiment with EnergyPlus for a period of 1 year; Lookahead N = 4 hours						
	Algorithms						
	LS	EO	BAT	LS+EO	LS+BAT	EO+BAT	DC-CFR (Our proposal)
Arizona	7.72 ± 0.18	8.16 ± 0.05	0.25 ± 0.08	13.26 ± 0.07	7.98 ± 0.1	8.46 ± 0.05	14.36 ± 0.09
New York	7.13 ± 0.19	8.02 ± 0.06	0.41 ± 0.03	14.39 ± 0.08	7.68 ± 0.20	8.21 ± 0.07	15.08 ± 0.11
Washington	4.27 ± 0.20	7.54 ± 0.11	0.46 ± 0.05	13.62 ± 0.08	4.53 ± 0.17	7.78 ± 0.08	13.96 ± 0.06

Table 2: Carbon Footprint Reduction Percentages compared to industry standard ASHRAE: Performance of the individual approaches over a period of one year. We are ignoring embodied footprint for server and battery manufacturing.

	Percentage Reduction of Energy Consumption with IPPO compared to ASHRAE Data Center Max Load 1.2MWh Experiment with EnergyPlus for a period of 1 year; Lookahead N = 4 hours						
	Algorithms						
	LS	EO	BAT	LS+EO	LS+BAT	EO+BAT	DC-CFR (Our proposal)
Arizona	7.11 ± 0.17	8.32 ± 0.04	0.00 ± 0.00	14.28 ± 0.07	7.15 ± 0.09	8.41 ± 0.05	14.54 ± 0.33
New York	7.05 ± 0.18	8.07 ± 0.06	0.00 ± 0.00	14.35 ± 0.08	7.12 ± 0.20	8.28 ± 0.08	14.62 ± 0.07
Washington	4.38 ± 0.21	7.42 ± 0.11	0.00 ± 0.00	13.78 ± 0.06	4.46 ± 0.18	7.31 ± 0.04	13.85 ± 0.07

Table 3: Energy Reduction Percentages compared to industry standard ASHRAE evaluated over a period of one year.

	Percentage Reduction of Energy Cost with IPPO compared to ASHRAE Data Center Max Load 1.2MWh Experiment with EnergyPlus for a period of 1 year; Lookahead N = 4 hours						
	Algorithms						
	LS	EO	BAT	LS+EO	LS+BAT	EO+BAT	DC-CFR (Our proposal)
Arizona	7.38 ± 0.20	8.41 ± 0.04	0.23 ± 0.07	13.81 ± 0.11	7.59 ± 0.10	8.43 ± 0.05	14.07 ± 0.17
New York	6.74 ± 0.18	8.17 ± 0.06	0.31 ± 0.04	13.61 ± 0.09	7.66 ± 0.22	8.39 ± 0.07	14.16 ± 0.11
Washington	3.57 ± 0.17	7.52 ± 0.11	0.30 ± 0.02	12.81 ± 0.05	3.81 ± 0.14	7.32 ± 0.05	12.85 ± 0.05

Table 4: Energy Cost Reduction Percentages compared to industry standard ASHRAE evaluated over a period of one year.

proach in handling diverse environmental scenarios.

Experimental Setup

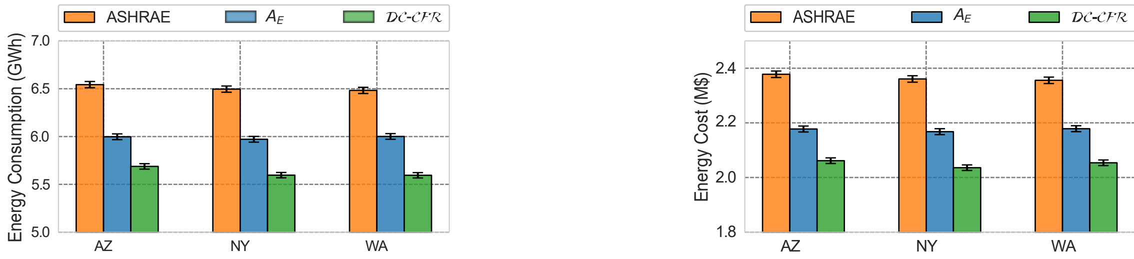
For training our agents, the Rllib (Liang et al. 2018) implementation of PPO (Schulman et al. 2017) was employed. The hyperparameters used for our experiments are the following: LR = 5×10^{-5}; Entropy Coefficient = 0.05; Clip Parameter = 0.05; $\gamma = 0.99$; λ (GAES Coefficient) = 0.95. The grid search function from Ray Tune (Liaw et al. 2018) was used to find the best learning rate, entropy coefficient and clip values. All agents use a neural network with 3 hidden layers of 128, 64 and 16 units each. The total computing budget for our experiments was approximately 1000 compute hours, utilizing 48 Intel(R) Xeon(R) Gold 6248 CPU @ 2.50GHz cores at an average utilization of less than 50%.

Results

In this section, we present the results of our approach evaluated against ASHRAE, the state-of-the-art controller used in DC cooling. To test the robustness of our approach and obtain more diverse results, we ran each experiment 20 times with different random seeds with varying noise in the weather data.

Tables 2, 3, and 4 present the annual reductions in Carbon Footprint, Energy Consumption, and Energy Cost, respectively, that can be achieved by using the DC-CFR framework and various combinations of the agents. These reductions were evaluated in three different locations. Importantly, the comprehensive DC-CFR approach outperforms the individual strategies. This superior performance is attributed to its ability to leverage the interdependencies among different aspects of data center operations and agents, thereby creating a more effective combined policy for energy and cost optimization.

The results obtained show that the proposed approach is able to achieve a high amount of savings in all the three metrics evaluated. On energy consumption, A_{BAT} has no effect since it cannot directly affect the power consumed by the



(a) Energy (GWh)

(b) Cost (M$).

Figure 4: Summary of results for data center (DC) control approaches in a 1.2 MWh DC in different locations. ASHRAE is an industry-standard controller for HVAC in DCs.

(Espen, Benth, and Šaltytė-Benth 2005).

Figures 1 and 4 provides a comprehensive summary of DC-CFR results, showcasing the optimization across various metrics and locations (Fig. 1 CO_2 footprint, Fig. 4 (a) Total energy consumption, Fig. 4 (b) Total energy cost) compared to ASHRAE and our standalone A_E agent. The figure shows how DC-CFR enhances performance on these metrics relative to the industry-standard ASHRAE.

Figure 6 illustrates how DC-CFR opportunistically increased energy expenditure on HVAC cooling, as shown by "spending". By enhancing cooling, DC-CFR effectively decreases the energy consumption of the IT infrastructure. This lowers the total energy consumption of the DC, as evidenced by the "Savings" in Figure 6.

Figure 5 (a) illustrates the actions of the battery agent (A_{BAT}) as it charges during periods of low CI and discharges supplying energy during periods of high CI. Figure 5(b) compares DC-CFR A_{LS}'s Carbon Aware flexible IT workload assignment against the default workload. A_{LS} shifts flexible IT load to low grid CI hours.

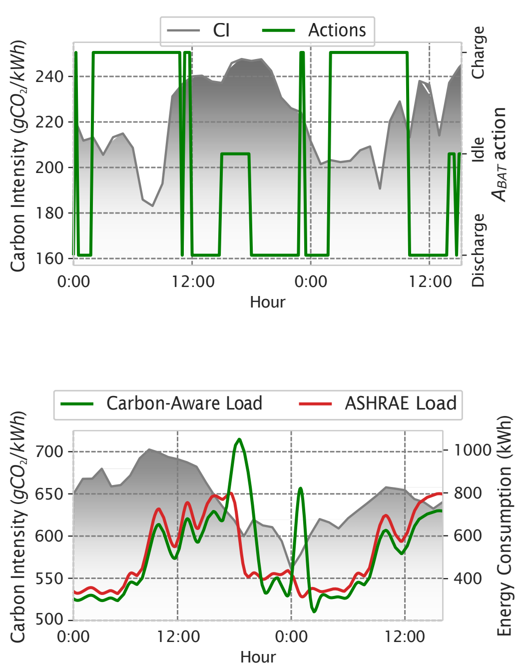


Figure 5: Snapshot of: (top) Actions taken by the A_{BAT} based on CI; (bottom) Carbon Aware Workload (Our proposal) against ASHRAE workload.

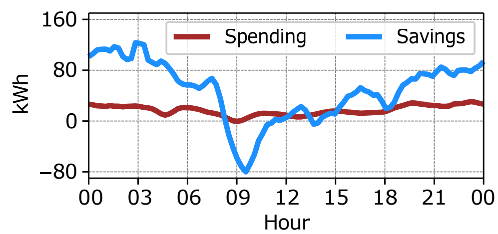


Figure 6: Energy Spending vs Savings over ASHRAE.

Conclusions

This paper presents a holistic framework called DC Carbon Footprint Reduction (DC-CFR), which optimizes data center energy consumption, load shifting, and battery operation decisions in real-time using Deep Reinforcement Learning (DRL). The framework employs three specialized agents working in concert to substantially reduce both carbon emissions and energy consumption.

The proposed DC-CFR methodology offers significant benefits over traditional static analysis methods. It effectively manages the complex interdependencies among various optimization strategies and uses short-term grid carbon intensity data to guide decision-making. Therefore, unlike static optimizations that rely on long forecast horizons and static seasonal models, our approach can deliver real-time optimization results in dynamic real-world applications.

We have evaluated our approach in multiple data center scenarios across various geographical locations, comparing it with industry-standard solutions such as the ASHRAE rule-based controller. Our method demonstrated significant improvements in carbon footprint reduction, energy efficiency, and cost of energy consumption. DC-CFR is effective for achieving sustainability goals in data center operations.

We plan to open-source the DC-CFR framework with the data center simulation, pluggable control agent abstraction, and OpenAI Gym RL interface, to democratize the carbon reduction efforts by the ecosystem. We plan to further enhance our DC-CFR framework by incorporating other optimization agents across the data center operations like dynamic heterogeneous computation resource allocation for CO_2 reduction, and achieving higher QoS with a lower hardware and carbon footprint. We plan to introduce ML CFD surrogates for better heat map estimates (Sarkar et al. 2023a). This will also enable digital twins for sustainable data centers, and the scalable architecture makes it applicable to supercomputing.

References

Abedi, S.; Yoon, S. W.; and Kwon, S. 2022. Battery energy storage control using a reinforcement learning approach with cyclic time-dependent Markov process. *International Journal of Electrical Power & Energy Systems*, 134: 107368.

Acun, B.; Lee, B.; Kazhamiaka, F.; Maeng, K.; Gupta, U.; Chakkaravarthy, M.; Brooks, D.; and Wu, C.-J. 2023. Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ACM.

Biemann, M.; Scheller, F.; Liu, X.; and Huang, L. 2021. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. *Applied Energy*, 298: 117164.

Cao, J.; Harrold, D.; Fan, Z.; Morstyn, T.; Healey, D.; and Li, K. 2020. Deep Reinforcement Learning-Based Energy Storage Arbitrage With Accurate Lithium-Ion Battery Degradation Model. *IEEE Transactions on Smart Grid*, 11(5): 4513–4521.

Cheng, Y.; Chai, Z.; and Anwar, A. 2018. Characterizing Co-Located Datacenter Workloads: An Alibaba Case Study. In *Proceedings of the 9th Asia-Pacific Workshop on Systems*, APSys ’18. New York, NY, USA: Association for Computing Machinery. ISBN 9781450360067.

Crawley, D. B.; Lawrie, L. K.; Pedersen, C. O.; and Winkelmann, F. C. 2000. Energy plus: energy simulation program. *ASHRAE journal*, 42(4): 49–56.

Espen, F.; Benth; and Šaltytė-Benth, J. 2005. Stochastic Modelling of Temperature Variations with a View Towards Weather Derivatives. .

Huang, B.; and Wang, J. 2021. Deep-Reinforcement-Learning-Based Capacity Scheduling for PV-Battery Storage System. *IEEE Transactions on Smart Grid*, 12(3): 2272–2283.

Jime´nez-Raboso, J.; Campoy-Nieves, A.; Manjavacas-Lucas, A.; Go´mez-Romero, J.; and Molina-Solana, M. 2021. Sinergym: A Building Simulation and Control Framework for Training Reinforcement Learning Agents. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 319–323. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391146.

Liang, E.; Liaw, R.; Nishihara, R.; Moritz, P.; Fox, R.; Goldberg, K.; Gonzalez, J.; Jordan, M.; and Stoica, I. 2018. RLlib: Abstractions for Distributed Reinforcement Learning. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 3053–3062. PMLR.

Liaw, R.; Liang, E.; Nishihara, R.; Moritz, P.; Gonzalez, J. E.; and Stoica, I. 2018. Tune: A Research Platform for Distributed Model Selection and Training. *CoRR*, abs/1807.05118.

Lowe, R.; Wu, Y. I.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; and Mordatch, I. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30.

Mahbod, M. H. B.; Chng, C. B.; Lee, P. S.; and Chui, C. K. 2022. Energy saving evaluation of an energy efficient data center using a model-free reinforcement learning approach. *Applied Energy*, 322: 119392.

Naug, A.; Guillen, A.; Luna Gutie´rrez, R.; Gundecha, V.; Ghorbanpour, S.; Dheeraj Kashyap, L.; Markovikj, D.; Krause, L.; Mousavi, S.; Babu, A. R.; et al. 2023a. 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*, 232–235.

Naug, A.; Guillen, A.; Luna Gutierrez, R.; Gundecha, V.; Ghorbanpour, S.; Mousavi, S.; Ramesh Babu, A.; and Sarkar, S. 2023b. A Configurable Pythonic Data Center Model for Sustainable Cooling and ML Integration. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*.

Radovanovic´, A.; Koningstein, R.; Schneider, I.; Chen, B.; Duarte, A.; Roy, B.; Xiao, D.; Haridasan, M.; Hung, P.; Care, N.; Talukdar, S.; Mullen, E.; Smith, K.; Cottman, M.; and Cirne, W. 2023. Carbon-Aware Computing for Datacenters. *IEEE Transactions on Power Systems*, 38(2): 1270–1280.

Ran, Y.; Hu, H.; Wen, Y.; and Zhou, X. 2022a. Optimizing Energy Efficiency for Data Center Via Parameterized Deep Reinforcement Learning. *IEEE Transactions on Services Computing*, 1–14.

Ran, Y.; Hu, H.; Zhou, X.; and Wen, Y. 2019. DeepEE: Joint Optimization of Job Scheduling and Cooling Control for Data Center Energy Efficiency Using Deep Reinforcement Learning. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE.

Ran, Y.; Zhou, X.; Hu, H.; and Wen, Y. 2022b. Optimizing Data Centre Energy Efficiency via Event Driven Deep Reinforcement Learning. In *2022 IEEE World Congress on Services (SERVICES)*. IEEE.

Sarkar, S.; Guillen, A.; Carmichael, Z.; Gundecha, V.; Naug, A.; Ramesh Babu, A.; and Luna Gutierrez, R. 2023a. Enhancing Data Center Sustainability with a 3D CNN-Based CFD Surrogate Model. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*.

Sarkar, S.; Naug, A.; Guillen, A.; Luna Gutierrez, R.; Gundecha, V.; Ghorbanpour, S.; Mousavi, S.; and Ramesh Babu, A. 2023b. Sustainable Data Center Modeling: A Multi-Agent Reinforcement Learning Benchmark. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms.

Wang, R.; Zhang, X.; Zhou, X.; Wen, Y.; and Tan, R. 2022. Toward Physics-Guided Safe Deep Reinforcement Learning for Green Data Center Cooling Control. In *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE.

Yi, D.; Zhou, X.; Wen, Y.; and Tan, R. 2019. Toward Efficient Compute-Intensive Job Allocation for Green Data Centers: A Deep Reinforcement Learning Approach. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE.

Zhang, K.; Wang, P.; Gu, N.; and Nguyen, T. D. 2022. GreenDRL: Managing Green Datacenters Using Deep Reinforcement Learning. In *Proceedings of the 13th Symposium on Cloud Computing*. ACM.

Zhang, Q.; Chng, C.-B.; Chen, K.; Lee, P.-S.; and Chui, C.-K. 2023a. DRL-S: Toward safe real-world learning of dynamic thermal management in data center. *Expert Systems with Applications*, 214: 119146.

Zhang, Q.; Zeng, W.; Lin, Q.; Chng, C.-B.; Chui, C.-K.; and Lee, P.-S. 2023b. Deep reinforcement learning towards real-world dynamic thermal management of data centers. *Applied Energy*, 333: 120561.

Zhou, K.; Zhou, K.; and Yang, S. 2022. Reinforcement learning-based scheduling strategy for energy storage in microgrid. *Journal of Energy Storage*, 51: 104379.