Holistic Sustainability for Geo-Distributed Data Centers using Hierarchical Optimization

Antonio Guillen-Perez, Avisek Naug, Vineet Gundecha,
Sahand Ghorbanpour, Ricardo Luna Gutiérrez, Soumyendu Sarkar
Hewlett Packard Enterprise

Motivation: The AI Carbon Problem

- AI workloads have a massive and growing energy demand.
- Data center **cooling (HVAC)** can be up to 40% of total energy use.
- **Challenge:** How can we co-optimize workloads and infrastructure?

Our Solution: A Digital Twin

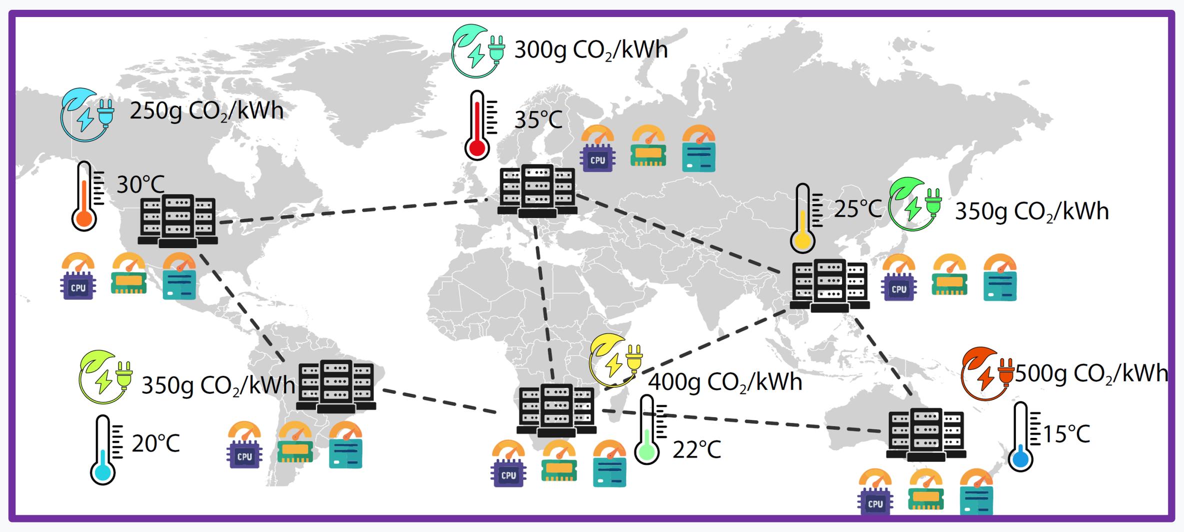


Fig. 1 Global figure.

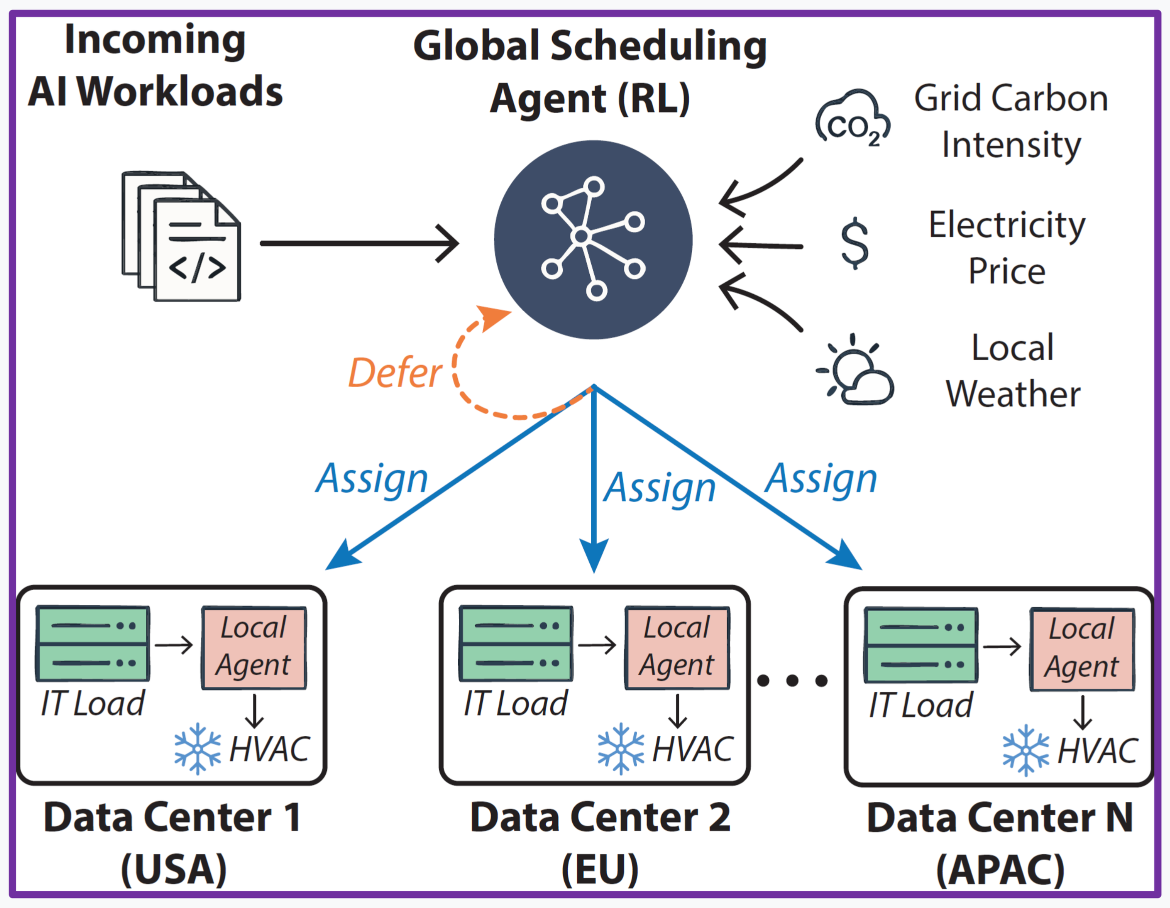


Fig. 2 Our solution

- Models a 5-node geo-distributed cluster.
- Integrates real-world data: Grid Carbon, Weather, Workloads.

Method: Hierarchical RL

Global Scheduler (MA-SAC)

Decides WHERE & WHEN to run tasks

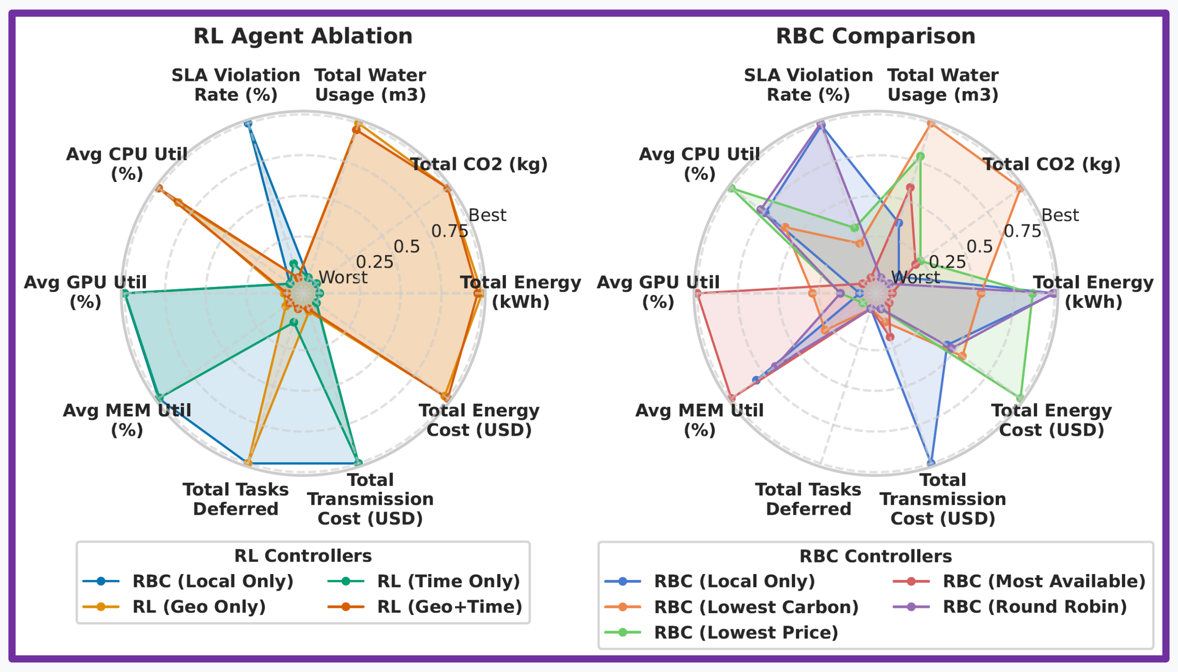
Local HVAC Controller (PPO)

Decides HOW to cool each DC

Maximized Carbon Reduction

Finding 1: RL Reveals a Critical Trade-Off

Controller	CO2 (t)	Cost ($)	SLA (%)
Heuristic	311	99,845	1.32
MA-RL	308.7	92,401	25.47



Finding 2: Hierarchy Unlocks Savings

Coordinating cooling with scheduling is essential for maximum climate impact.

Configuration	CO2 (t)
MA-SAC (Global Only)	308.7
MA-SAC + Local HVAC	273.1
MA-SAC + Local HVAC + HRU	268.9

+12.9% Additional Carbon Reduction!

Conclusion & Main Takeaway

- The future of sustainable computing is in **holistic, full-stack optimization**.
- Managing the **sustainability** vs. **performance trade-off** is the key challenge.
- Optimizing infrastructure (HVAC) and software (scheduling) **together** is essential.

Pathway to Impact & Future Work

- **Managing Trade-offs**: Developing **constrained RL agents** to let operators set an SLA budget.
- **Deployment**: Next steps involve validation on a **physical hardware testbed**.
- **Reproducibility**: Our simulator will be **open-sourced** to foster community research

Contact & Full Paper

Email:
antonio.guillen@hpe.com

Scan for Project page:

Scan for main paper in D&B track: