Learning a Data Center Model for Efficient Demand Response

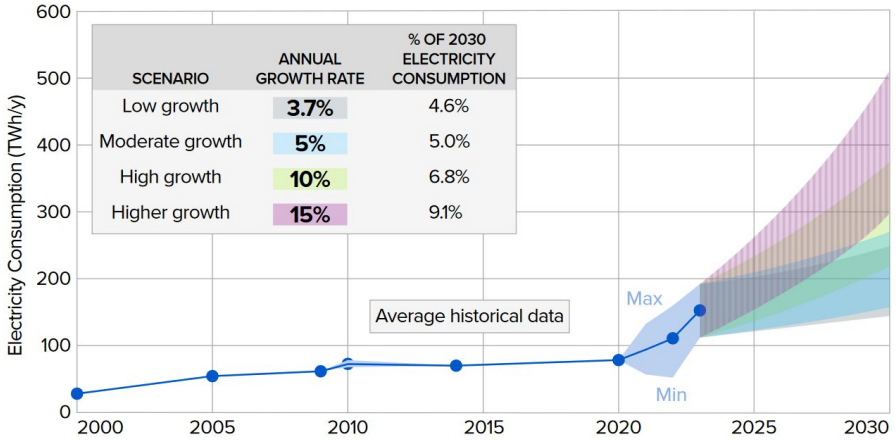
Quentin Clark, **Fatih Acun**, Ioannis Paschalidis, Ayse Coskun

Boston University

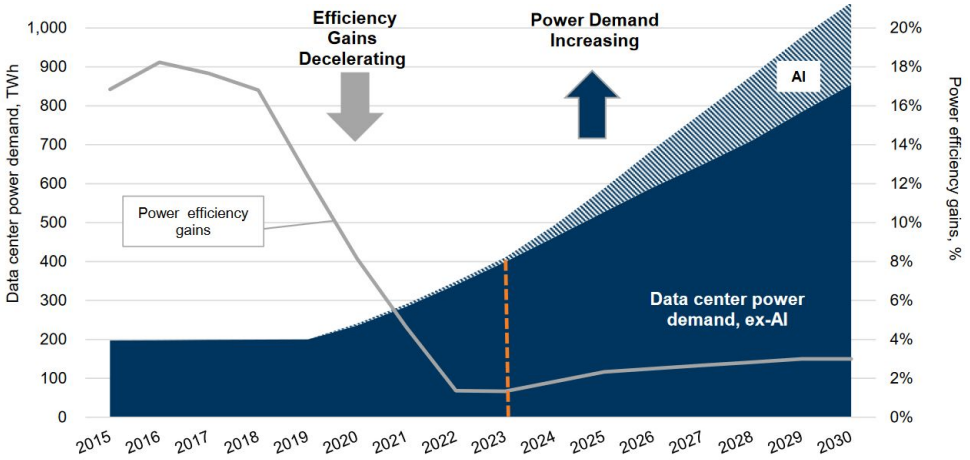
July 9th, HotCarbon 2024, Santa Cruz CA

BOSTON UNIVERSITY

The Future of Data Center Sustainability



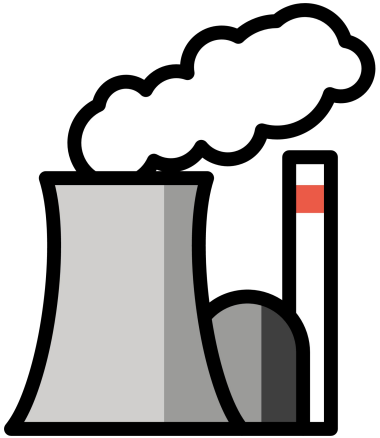
"Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." Electric Power Research Institute (EPRI). 28 May 2024, www.epri.com/research/products/3002028905



“AI, data centers and the coming US power demand surge”. Davenport et al. for Goldman Sachs Group, Inc. 28 April 2024, https://www.goldmansachs.com/intelligence/pages/gs-research/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf

Demand Response (DR)

Power Provider



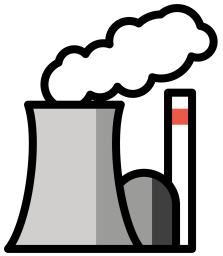
We need you to use less power due to low supply.

We will find ways on our end to reduce demand.

Data Center

Regulation Service Reserves DR

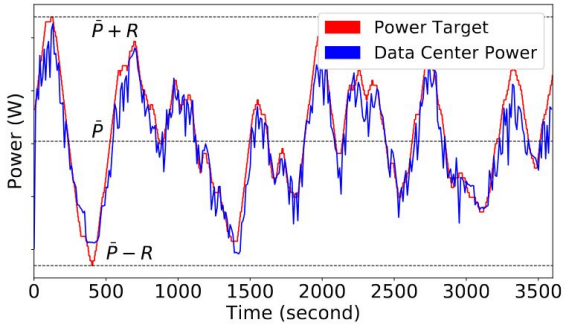
Power Provider



Average Power P
Reserve Capacity R

Reserve Request Signal

Data Center



BOSTON UNIVERSITY

Assuring Quality of Service (QoS) for Jobs

QoS: “Are our jobs completing as quickly as we’d like, most of the time?”

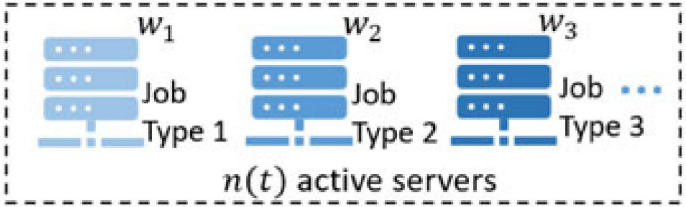
Job Type Examples:
AI Training Workload: takes a while, fine if it is slow
Search Query: takes not a lot of time, not fine if it is slow

BOSTON UNIVERSITY

Method: The Adaptive QoS-Assurance (AQA) Framework

[Zhang et al., TSUSC '20]

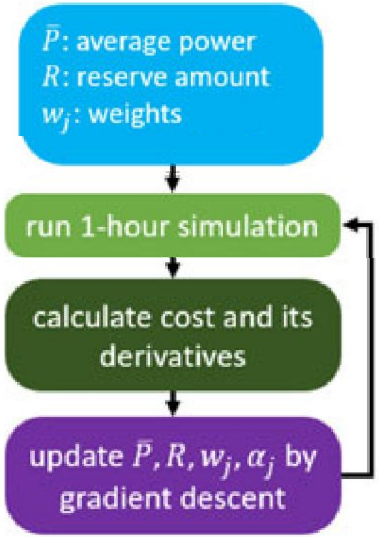
Data Center Simulator



Cost:

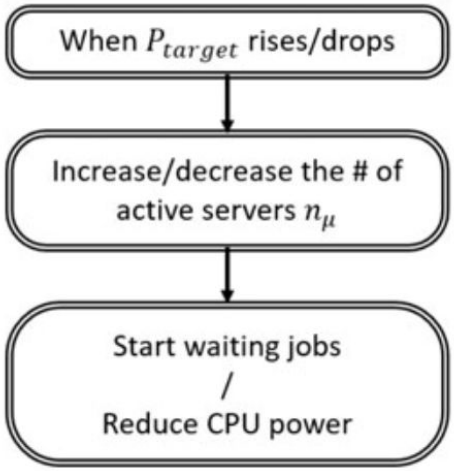
- Are we violating QoS?
- Are we meeting ISO signal?
- Are we saving money?

Parameter Selection



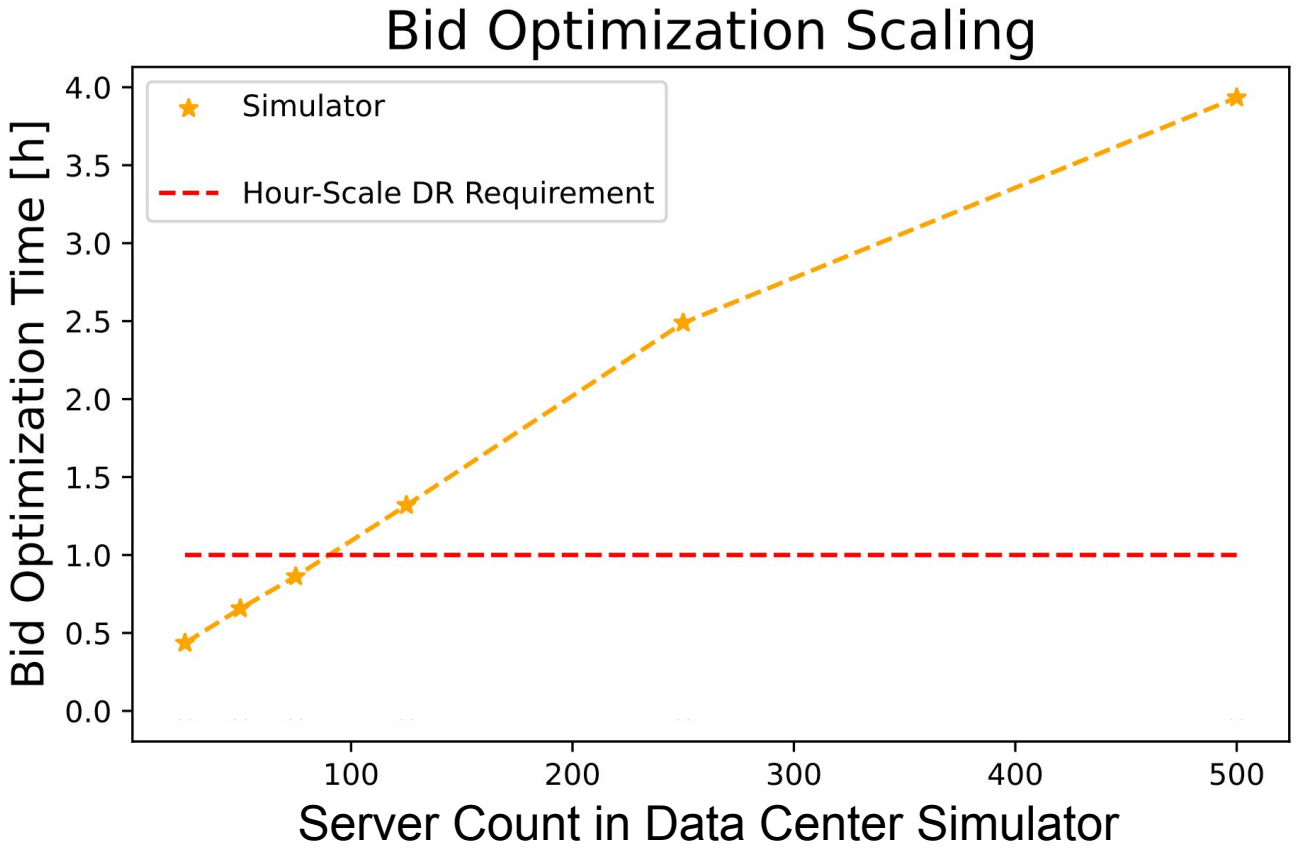
What we improve

Runtime Policy



Do Not Touch

Problem: Data Center Simulation is Slow

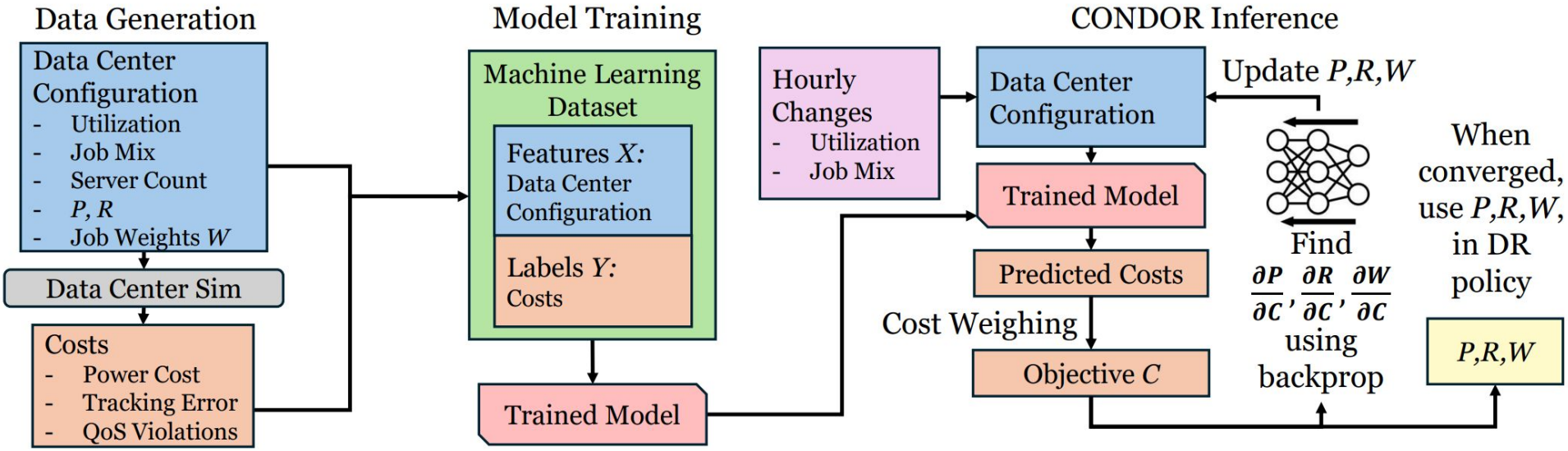


BOSTON UNIVERSITY

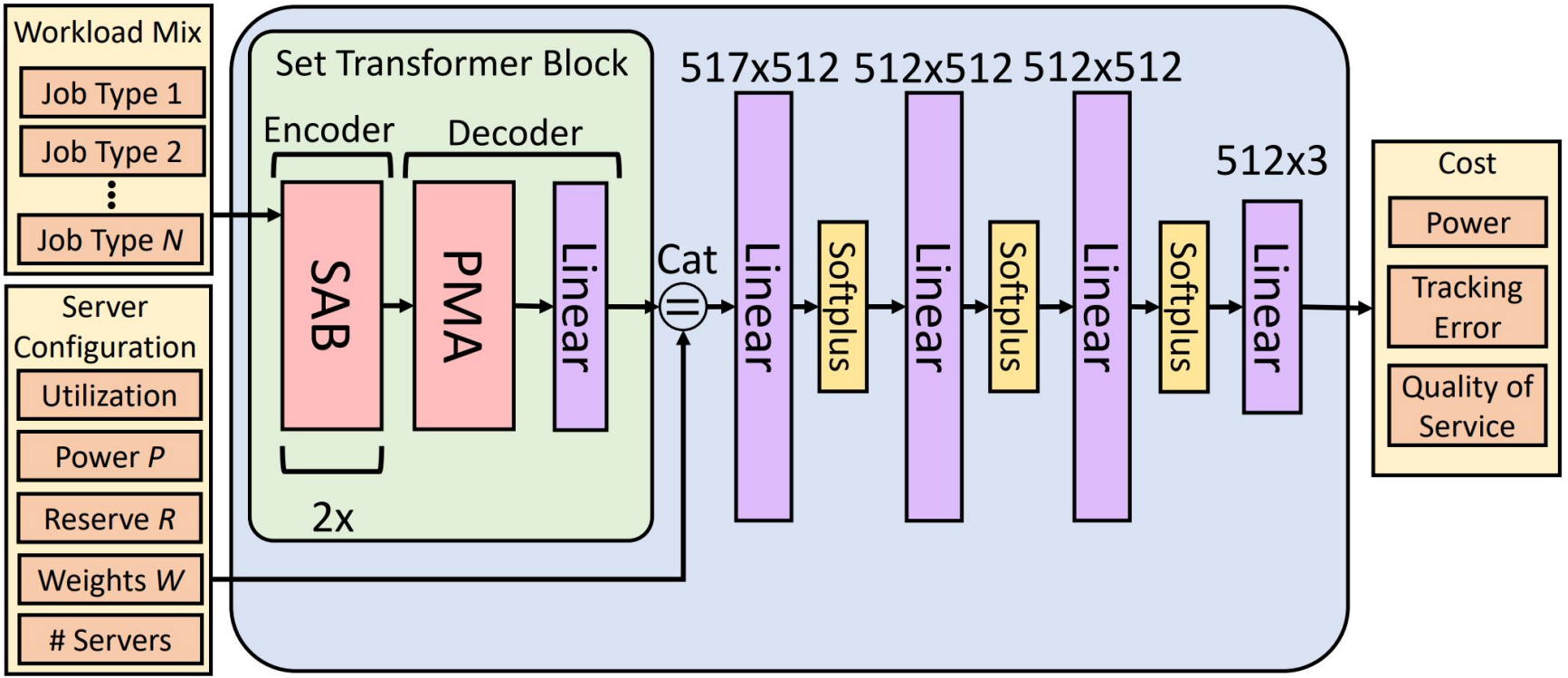
Solution: CONDOR Overview

Idea: Can we replace our slow simulator with a faster model?

Our model: CONDOR (**C**ost-**O**ptimization **N**eural Network for **D**ata Center **O**perational Demand **R**esponse)



Solution: Neural-Network Architecture

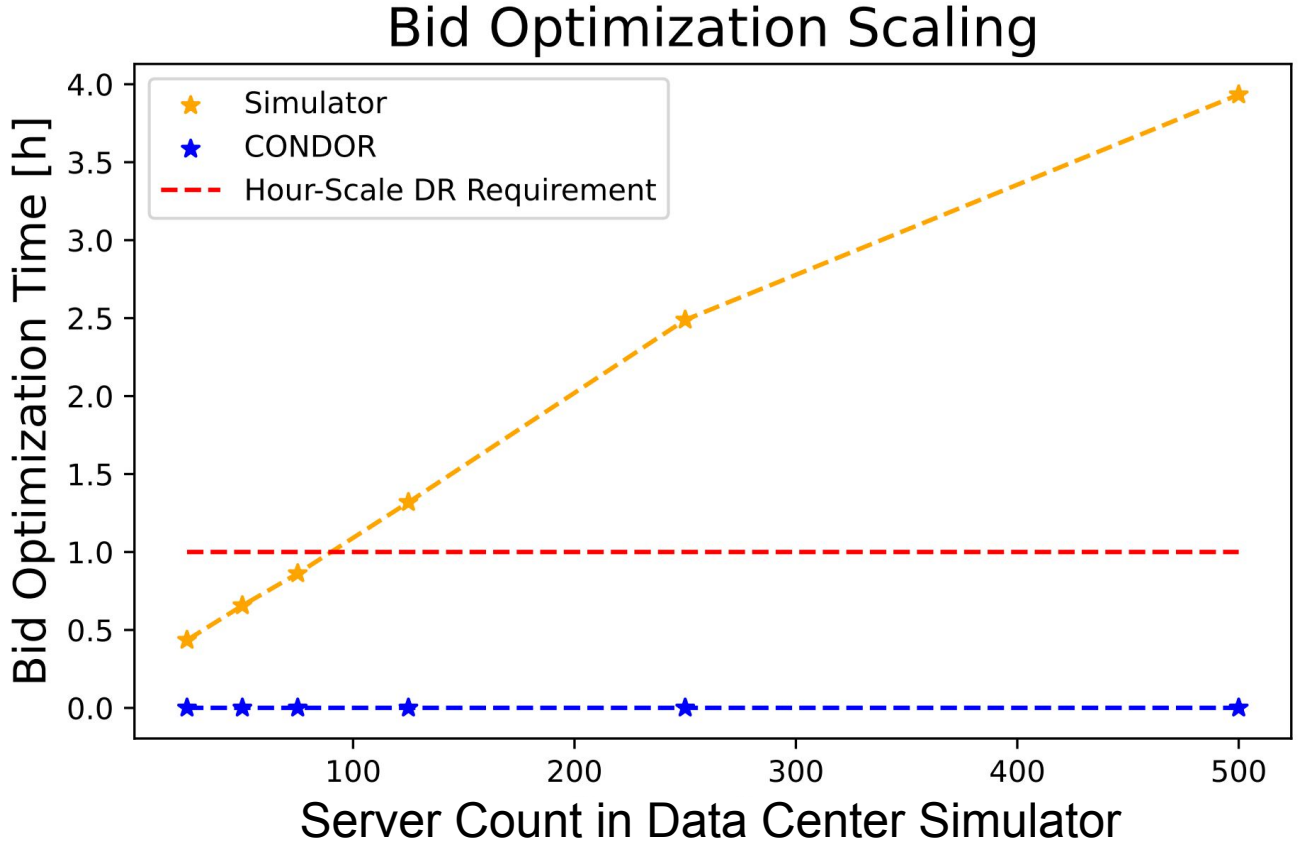


Results: ML Model vs AQA Simulator

Workload Mix \ Method	$\bar{P}$ (kW)		R (kW)		Execution Time		Norm. Cost	% Violation	
	Simulator	Model	Simulator	Model	Simulator	Model	Model	Simulator	Model
W3	160.7	175.4	26.3	31.2	236 m	0.911 s	1.171	12.5%	0%
W4	154.3	159.1	21.1	33.4	610 m	0.814 s	0.980	0%	0%
W5	154.4	147.3	23.5	26.4	531 m	0.790 s	0.920	0%	0%
W6	175.1	166.8	31.5	29.4	613 m	0.828 s	0.95	0%	0%
W7	159.5	171.6	23.9	29.9	547 m	0.841 s	1.119	0%	0%
W8	139.4	155.1	14.3	17.7	591 m	0.822 s	1.191	0%	0%

Punchline: CONDOR is comparable to the discrete simulator (average 5% cost penalty), but around 15,000 faster!

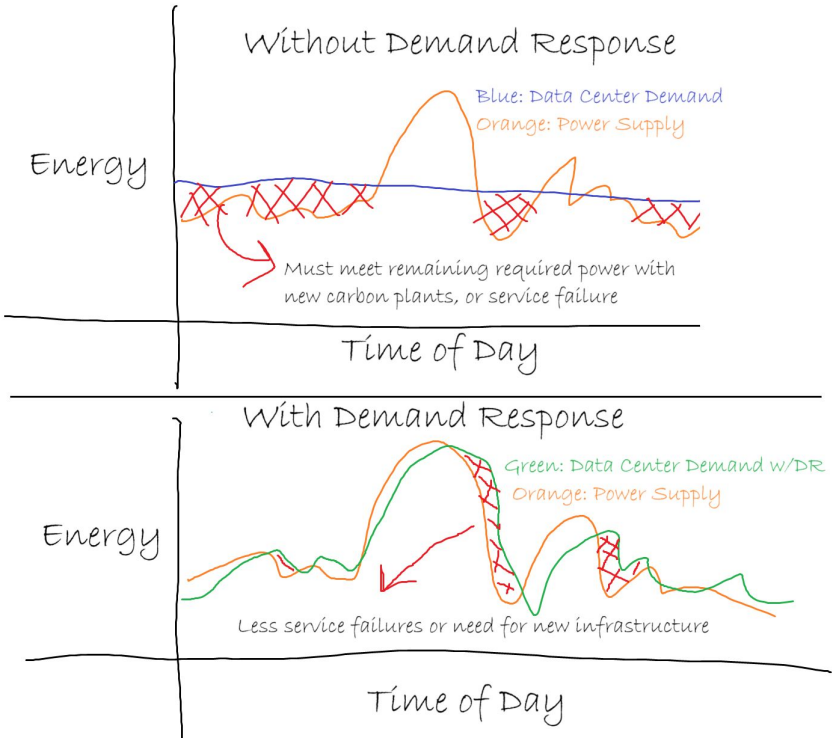
Problem: Data Center Simulation is Slow



BOSTON UNIVERSITY

Conclusion

- DR is a promising avenue for data centers to remain sustainable into the AI future
- We introduce a faster ML-based data center DR method to replace simulations
- Speedups enable previously computationally intractable DR methods to be brought to real data centers



Boston University College of Engineering

Fatih Acun
acun@bu.edu

BOSTON UNIVERSITY

References

- https://www.epri.com/research/products/3002028905
- https://www.goldmansachs.com/intelligence/pages/gs-research/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf
- Y. Zhang, D. C. Wilson, I. C. Paschalidis and A. K. Coskun, "HPC Data Center Participation in Demand Response: An Adaptive Policy With QoS Assurance," in *IEEE Transactions on Sustainable Computing*, vol. 7, no. 1, pp. 157-171, 1 Jan.-March 2022, doi: 10.1109/TSUSC.2021.3077254.