Facing Computing's Challenges at the End of Technology Scaling

Andrew A Chien (University of Chicago)

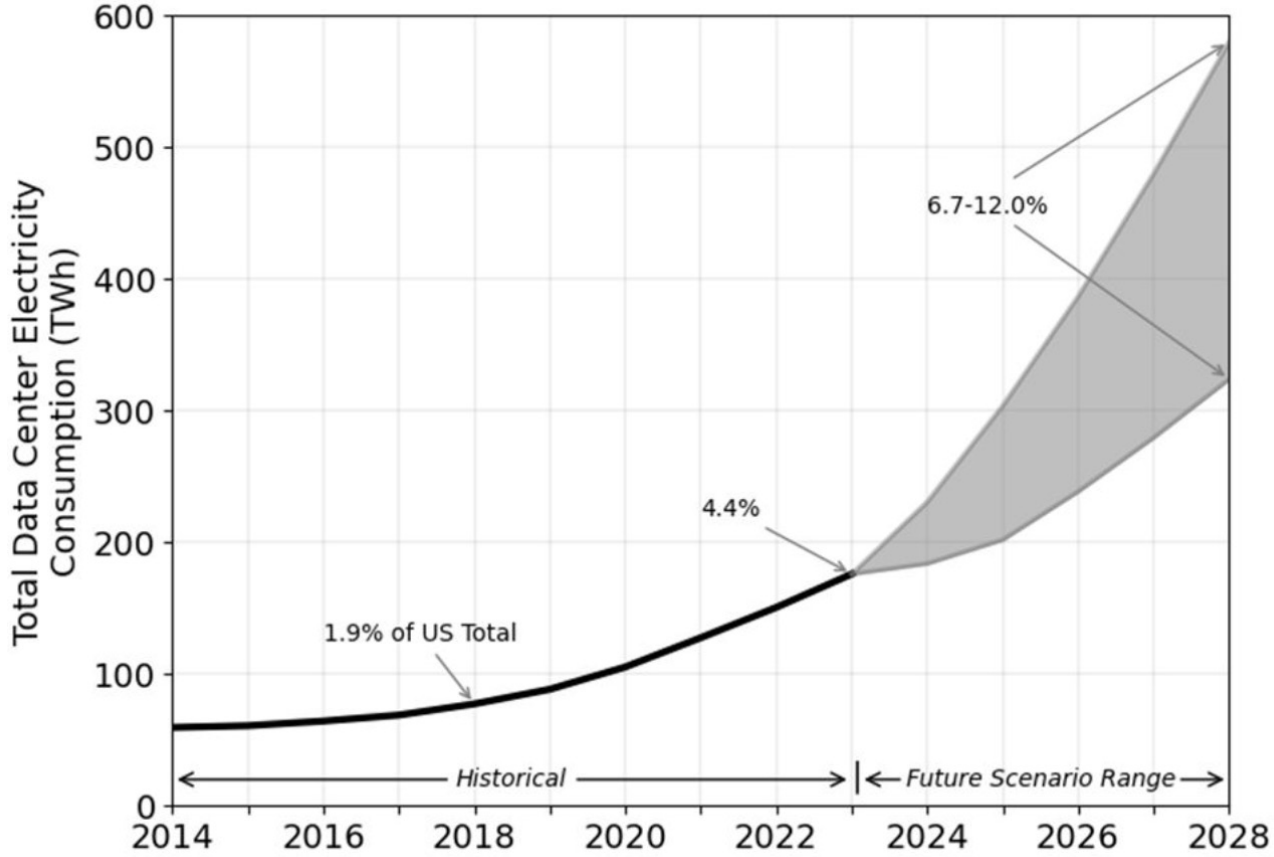
December 8, 2025

NSF Workshop on the Power/Energy-Compute Nexus

THE UNIVERSITY OF CHICAGO

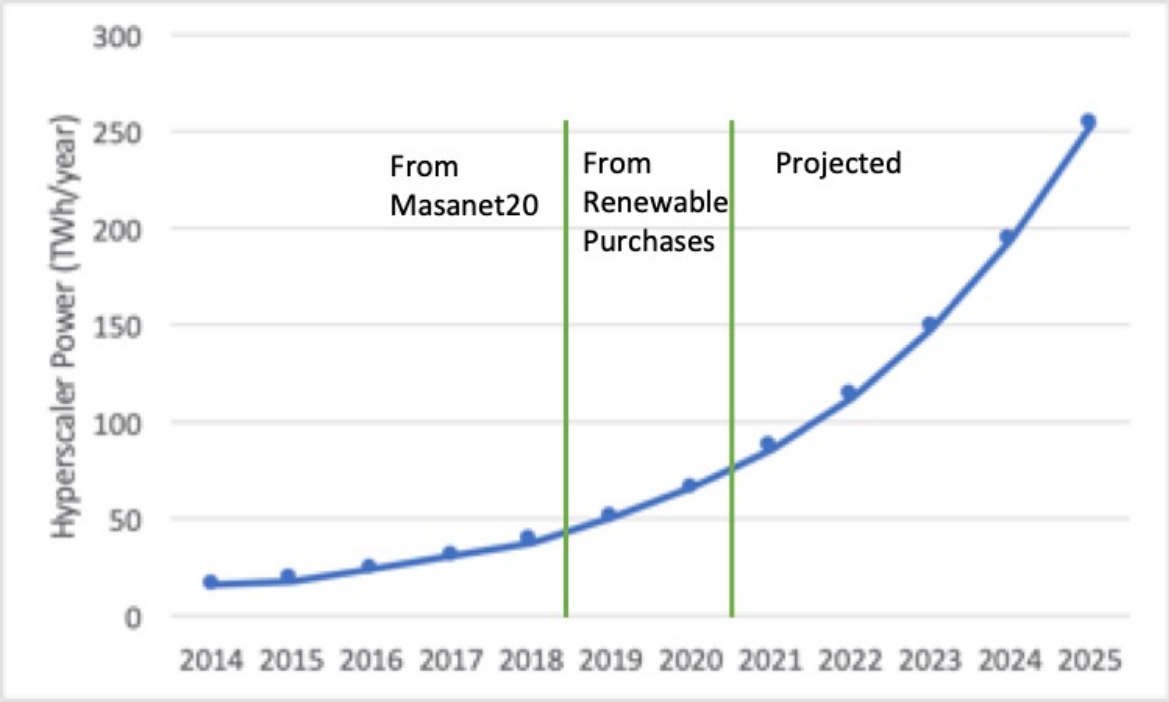
Argonne NATIONAL LABORATORY

Projected Rapid Growth of USA Datacenters



- 2023-28: Projected growth to 6.7-12% (and more beyond!)
- Ensuing Grid Challenges
  - Datacenter Power, Capacity Adequacy, Grid Stability (low and high frequency)
- => AI challenge: “We want to grow, but can’t get the power we need!”

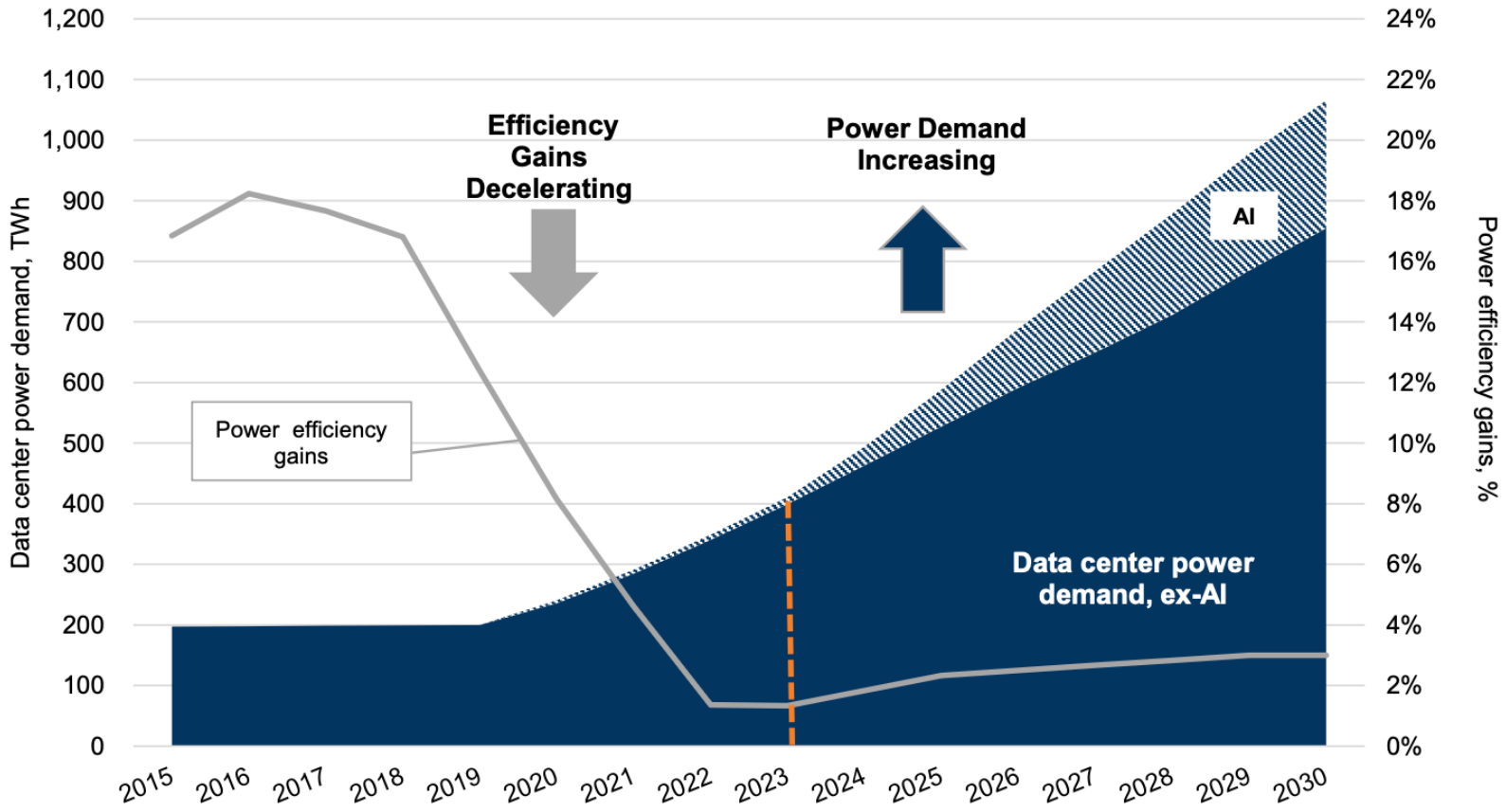
This problem has been with us a lot longer than GenAI



Mansanet, et. al. Science 2020 estimate – 30% CAGR (extrapolate rate for 2015-18)

- Cloud Power use scaling – End of Dennard (2005), End of Moore (2020)
  - Rapid growth in use
  - Overwhelms energy efficiency improvements

It was Cloud Computing, before it was AI



- Goldman Sachs Projection, March 2024

Datacenters drive Power Prices

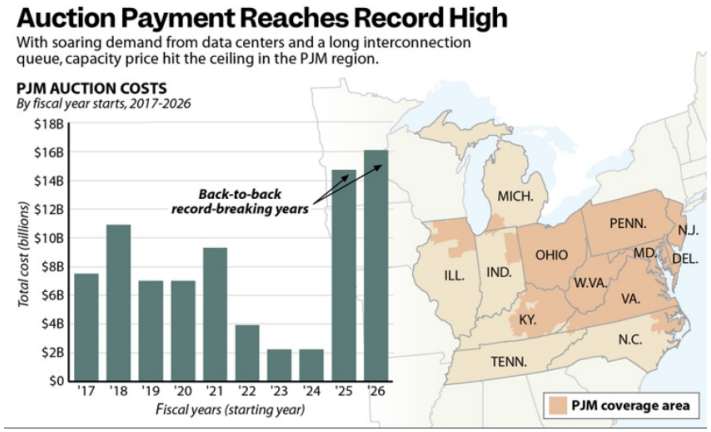
- Build more generation
- Build more transition
- Make it reliable

- PJM results
- Consumer Upset

- Datacenters are a large percentage of power
  - 1GW is Philadelphia, 80GW!
  - NoVa is 30%? And >50% in 2030's

Aug 29, 2025 - Energy & Climate

America's power bills surge as AI strains an aging grid



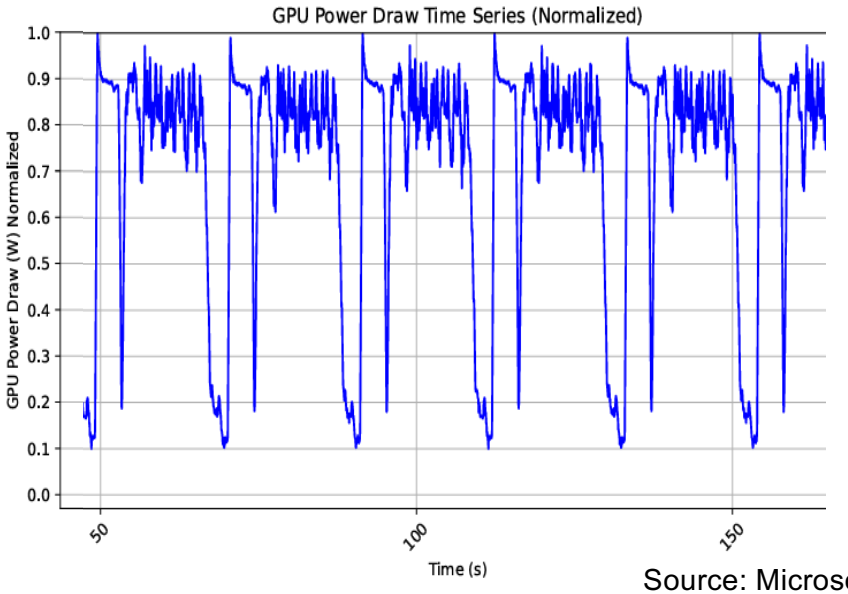
ARTIFICIAL INTELLIGENCE, BUSINESS, ECONOMY, NEWS

Microsoft scraps plans for Caledonia data center site, working to find an alternate location

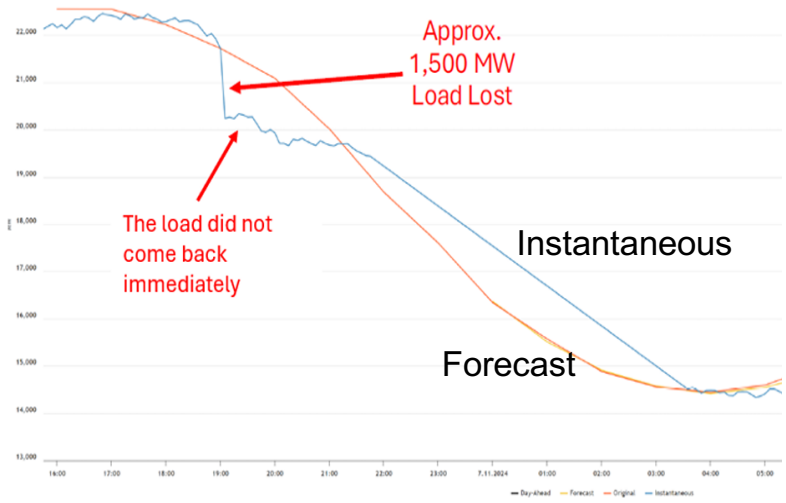
Microsoft cites 'community feedback' in its decision, as data center projects across Wisconsin receive increased local scrutiny

BY JOE SCHULZ • OCTOBER 8, 2025 • UPDATED OCTOBER 8, 2025 at 4:44 PM

Datacenters threaten Grid Stability



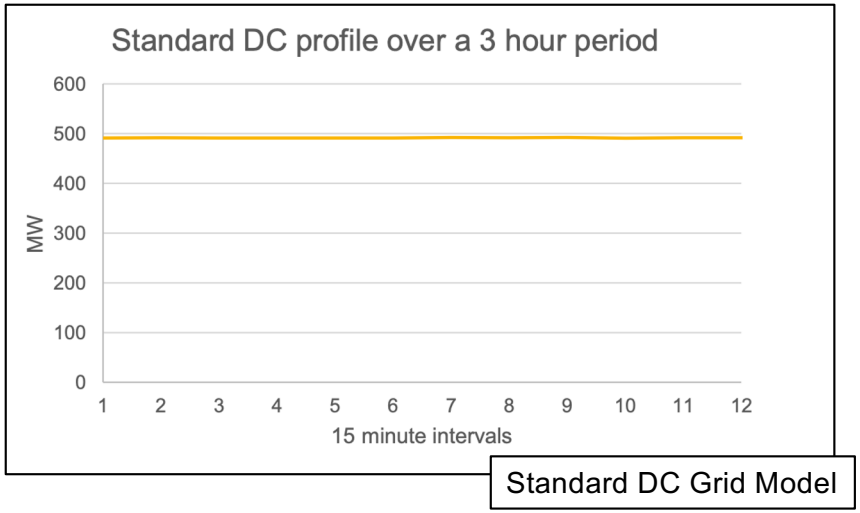
Source: Microsoft



Source: NERC, 2024

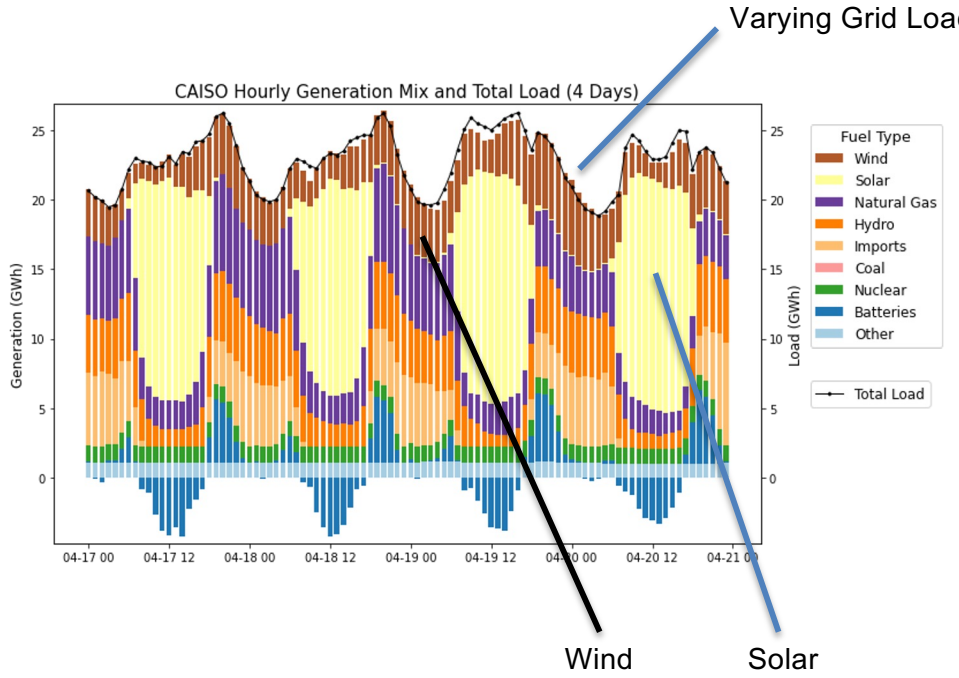
- AI Training Dynamics are Problematic (GPU swings, Regulation)
- Grid Dynamics are Problematic (LVRT) (DC tripoff, Grid Stability)

Balancing Challenge of Energy Transition



Standard DC Grid Model

vs



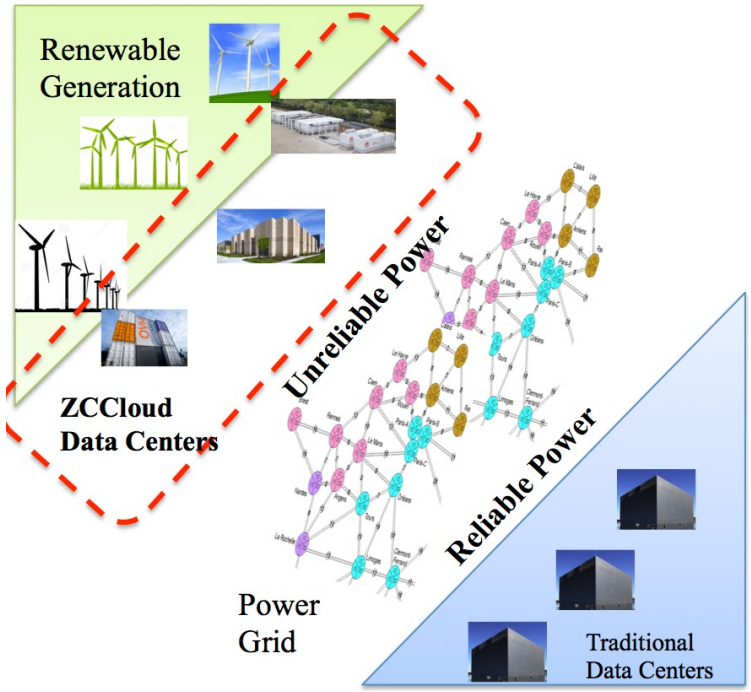
- Challenge: Matching nearly Constant AI Datacenter Load and Volatile Generation
- Reliability: “Peak” challenges occur ~4 days per year, but they determine grid reliability
- Efficiency: Better matching is possible daily, hourly. Reduce cost, carbon.
- Lin, et. al. Exploding AI Power Use: an Opportunity to Rethink Grid Planning and Management, E-energy ’24, 11/2024 New EPRI DCFlex Effort

Datacenters as Dynamic Loads

- Datacenters as 10% (and growing) of grid load
  - Fast moving, controllable
  - Sophisticated wealthy operators
  - Networked! Of course!

  - Datacenter load flexibility enables more reliable, efficient grids
  - Visibility into Load, planning
  - Flexibility can enable faster grid decarbonization (and the capacity the datacenter companies dearly want!)
  - Flexibility can reduce the massive energy storage needed to create reliability from wind/solar generation

Zero-Carbon Cloud: Computing on Excess Renewable (Stranded) Power



Uchicago Project, 2015-today

- Exploit excess renewable power that would go to waste
- Zero-marginal carbon power
- Help the Grid absorb more renewables
- => And, lower dispatch cost
- => Reality for AI in Texas, Lancium Company

Idea
Uchicago
2015

Reality
Ft Stockton
Lancium
2021

2025: its Stargate!
(Lancium=>Oracle)

"grid friendly" datacenter load

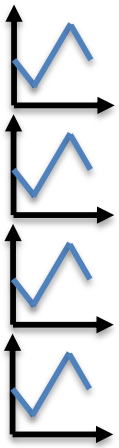
Cooperation Architectures Span Technologies, Markets, Regulatory Structure

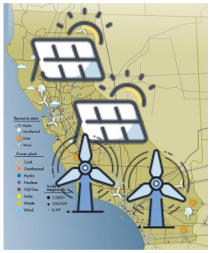
DC Flex Sources:
- Compute Load
- Cooling Power Flex
- Weather (temp)
- Power Middlebox

1-way Information Sharing (Planshare)

2-way Info Share Control (PS-Varscale)

2-way Multi-round Interaction

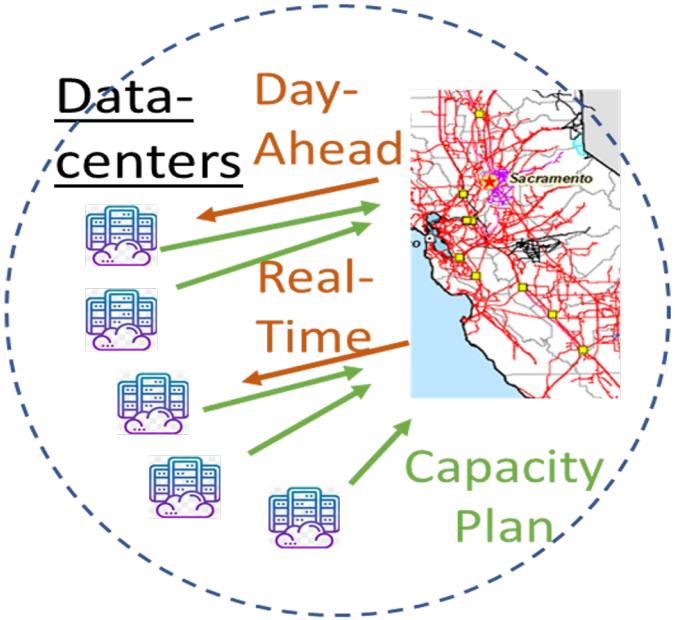




Grid Considerations:
- Reliability, Matching
- Social Welfare
- Optimal Power Flow
- Fairness
- Low prices
- Renewable Absorption
- …

- Cooperation: Exploiting AI Datacenter Flexibility
- Flexibility Sources: Flex Compute Load, Cooling Power flex, Weather, Power Middlebox
- Exploiting flexibility engages with markets (bid, price, products) and regulation (reliability, connection priority, cost recovery, ownership)
  - Cooperation Relaxes Reliability Constraints (Increase Load) and Increases Renewable Absorption (Decarbonize)

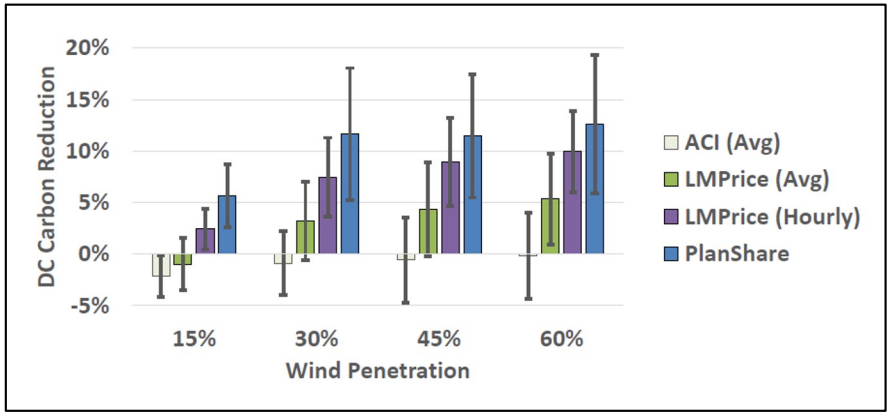
Types of Cooperation: Information Sharing, Negotiation, Shared Decoupling Mgmt



1-way

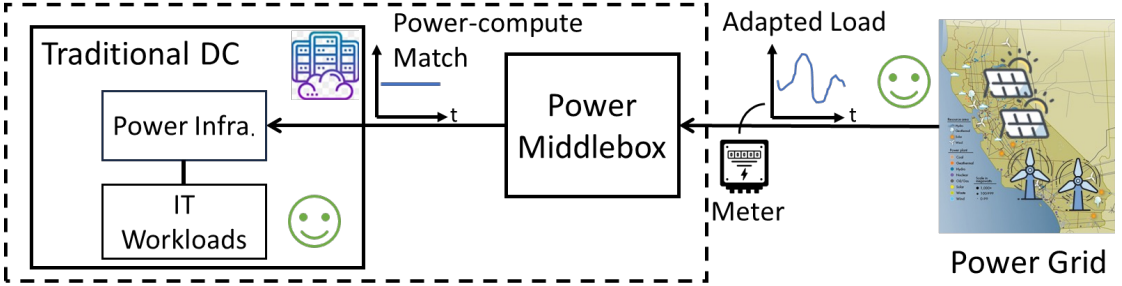
2-way (WIP)

Middlebox



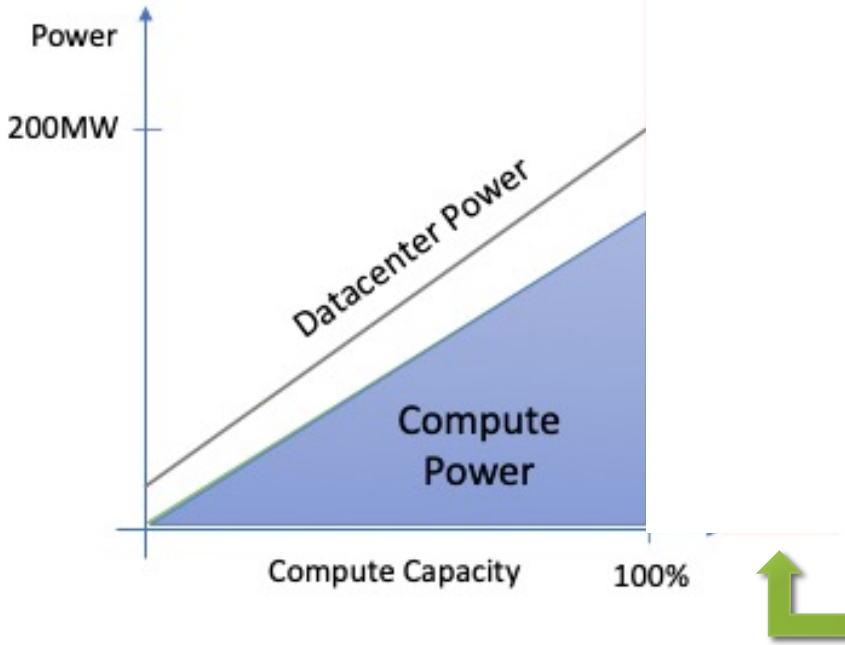
Lin and Chien, Greening Datacenter and Grid, ACM E-energy 2023.

Datacenter Power + Middlebox Power = Grid Load



- Richer engagement produces increasing benefits in capacity and decarbonization
- DCFlex and 1-way Coordination benefits = ~5 years of US Grid Decarbonization

Flexible Datacenters: Capacity Headroom, on the Cheap!

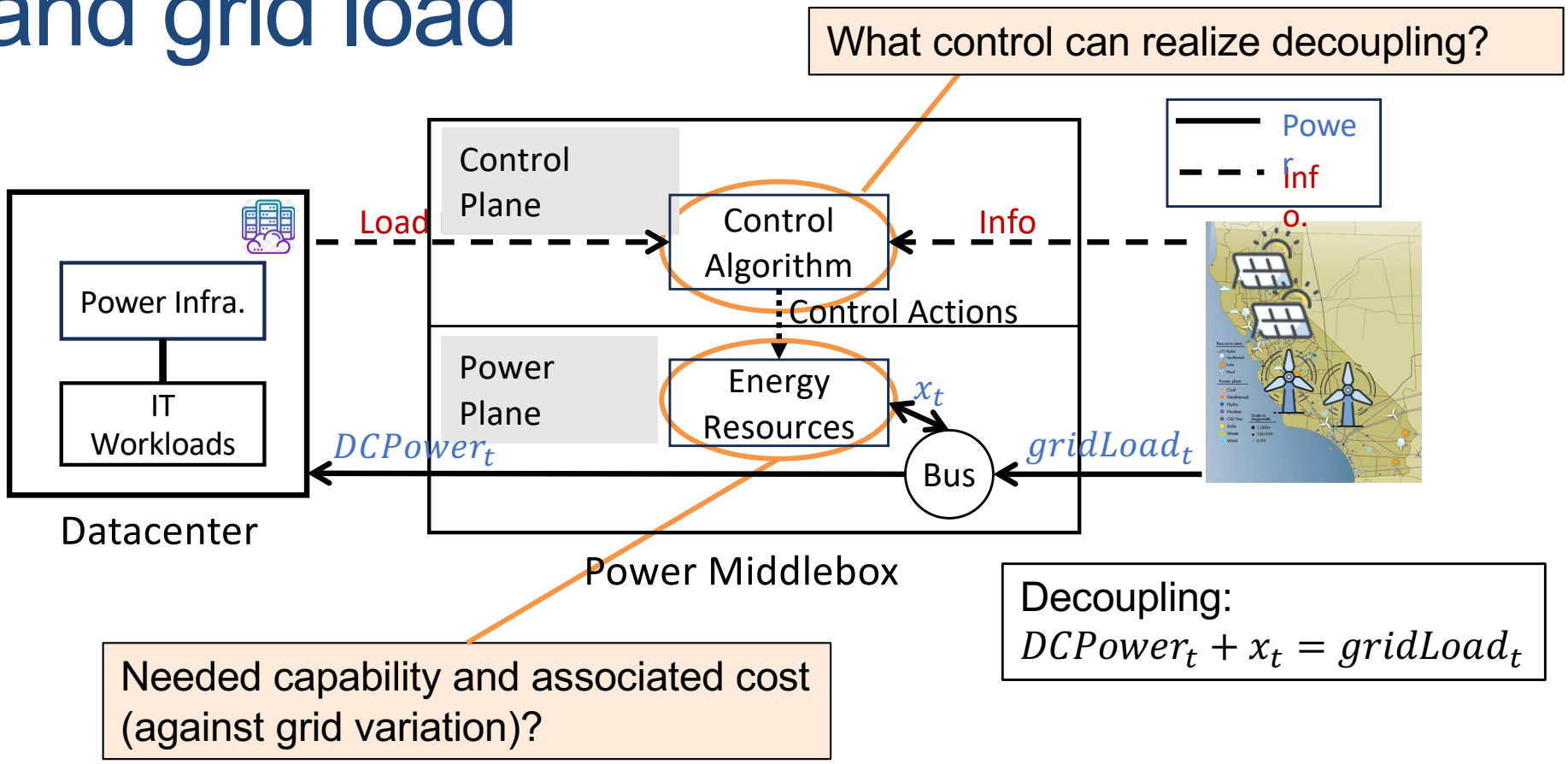


- Traditional PUE assumed linearity across an operating range.
- Optimize for multiple operating points
  - PUE critical for high-carbon, cost, power
  - PUE not critical for low-carbon, low-cost power
  - Capex critical always

Operate here when low-carbon or

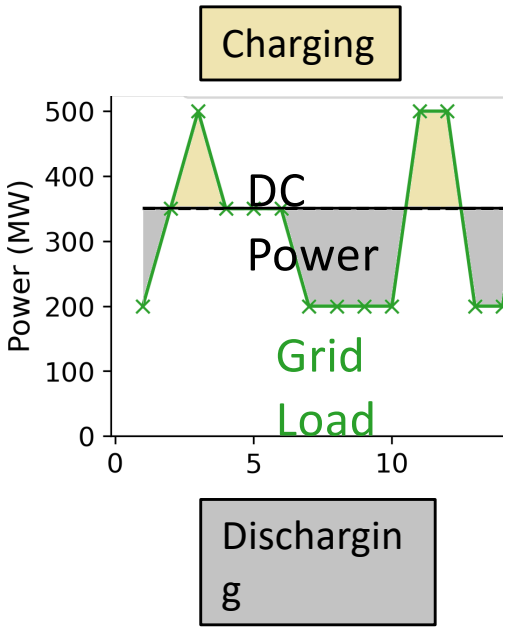
Beyond PUE: Flexible Datacenters Empowering the Cloud to Decarbonize, Hot Carbon Workshop, 2022.

Middlebox: decouple DC capacity and grid load

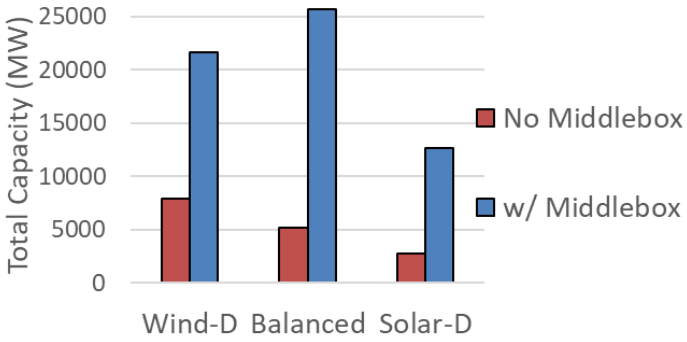


Middlebox meets BOTH grid and DC needs

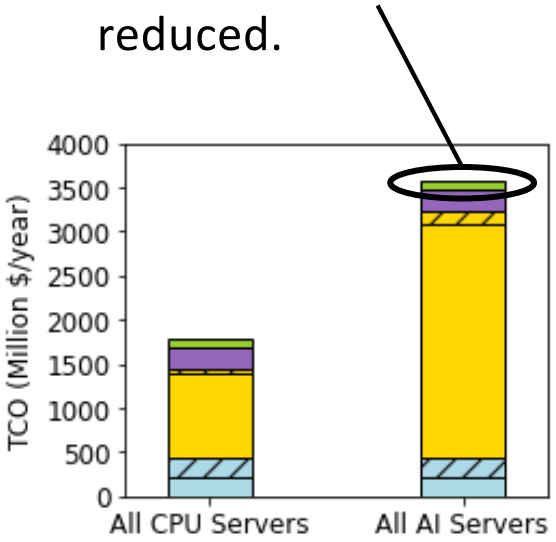
- Realizes decoupling
  - Storage example:



- Enables both DC growth and grid decarbonization
  - 2.7—5x grid capacity with same grid carbon intensity.



- Middlebox cost is small
  - 2.7% cost increase for an AI DC. Can be further reduced.



Datacenter TCO

Middlebox benefits = ~5 years of projected DC and Grid growth

Lin & Chien, Middlebox: Unlocking Datacenter Growth and Grid Decarbonization SoCC, November 2025

Power is increasingly an Economic Consideration for all Computing

$30/charge
$0.50/KWh

$6/hr
$500/KWh

Summary

- Computing continues explosive growth and efficiency is not going to save us! (Jevons!). But its still important.
- End of technology scaling (Dennard 2005, Moore 2020) is the real cause.
- Computing must deal with cooperation, economic, technological problems… at the energy-compute nexus
- Datacenter-Grid cooperation and decoupling is key, to getting large quantities of power, and helping the grid.
- At large and small scale, computing needs to be flexible and cooperative, [3-4 ways to do this!]
  - NOT “consume as much power as I want, when I want” anymore.

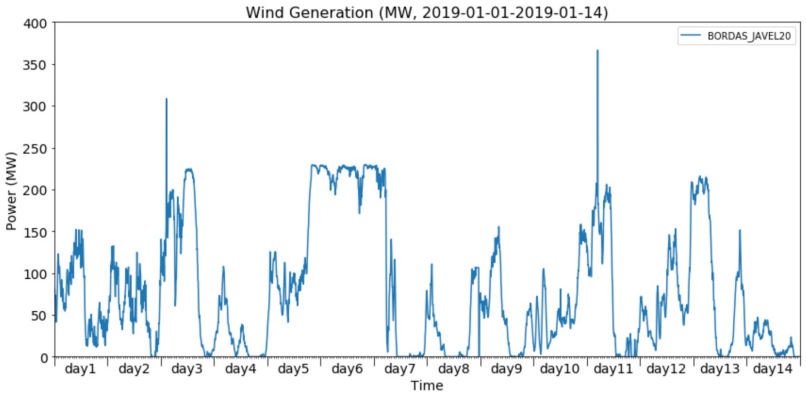
For More Information zccloud.cs.uchicago.edu

- Middlebox: Unlocking Datacenter Growth and Grid Decarbonization, ACM Symposium on Cloud Computing (SoCC), November 2025.
- FlexCoolDC: Datacenter Cooling Flexibility for Harmonizing Water, Energy, Carbon, and Cost Trade-offs, in E-energy '24, June 2024 (Singapore).
- As Grids Reach 100% Renewable at Peak, Growing Curtailment of 8 Gigawatts Looms as a Challenge to Decarbonization, Energy Informatics Review, January 2024.
- Can Datacenters Get the Power Needed to Meet the Explosive Demand for AI?, https://arxiv.org/abs/2311.11645, 11/2023 and E-energy ‘24 (6/2024)
- Adapting Datacenter Capacity for Greener Datacenters and Grid, in the ACM Symposium on Future Energy Systems (E-Energy 2023), June 2023
- Reducing the Carbon Impact of Generative AI Inference (today and in 2035). In Proceedings of the 2nd Workshop on Sustainable Computer Systems (HotCarbon '23).
- Beyond PUE: Flexible Datacenters Empowering the Cloud to Decarbonize, USENIX Hot Carbon ‘22
- "Evaluating Coupling Models for Cloud Datacenters and Power Grids", ACM E-energy Conference, July 2021
- Large-scale and Extreme-Scale Computing with Stranded Green Power: Opportunities and Costs, IEEE TPDS, (29) 5, Dec 2017.
- Data Centers as Dispatchable Loads to Harness Stranded Power, IEEE Transactions on Sustainable Energy, 8(1), January 2017

From Chien talk at CLEAR Workshop @ ISCA 2021; also CACM 12/2021

- **Good**

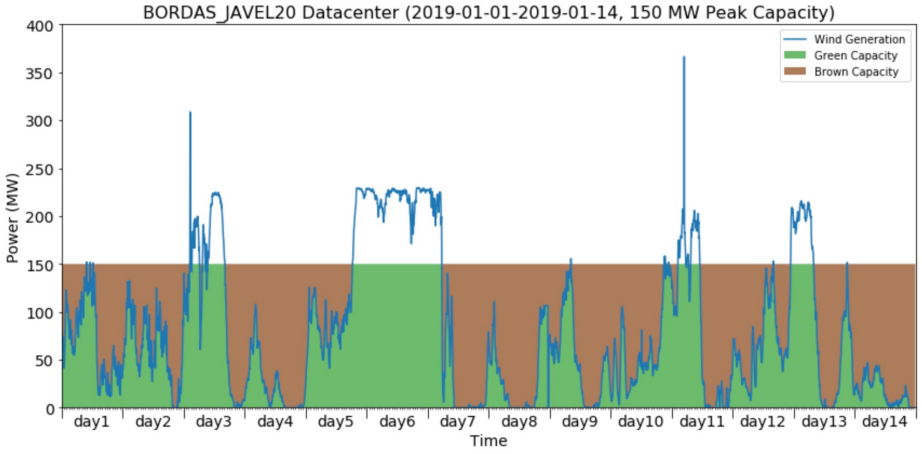
Offset (2007-2017)



vs.

- **Better**

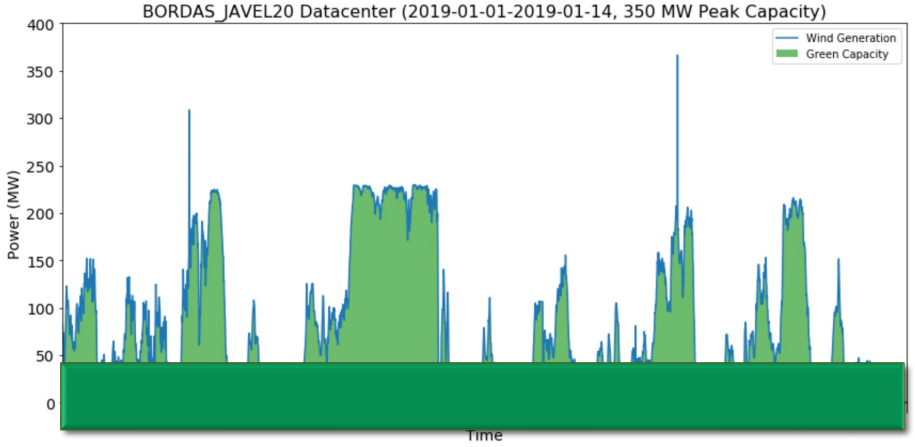
24x7 Matching (2018-2030)



Cover all the brown by 2030!

- **Best**

Flexible Load 2018-



Flex load UP to match renewable availability!

BACKUP