Research Avenues Towards Net-Zero Cloud Platforms

Daniel S. Berger, Fiodar Kazhamiaka, Esha Choukse, Íñigo Goiri, Celine Irvene, Pulkit A. Misra, Alok Kumbhare, Rodrigo Fonseca, Ricardo Bianchini

Azure Systems Research

Workshop on NetZero Carbon Computing, Montreal, February 2023

Why Focus on Cloud?


THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Podcasts | Latest Headlines | More

Home World U.S. Politics Economy Business Tech Markets Opinion Books & Arts Real Estate Life & Work Style Sports

CIO JOURNAL

Cloud Growth Continues Amid Tough Time for Tech

Companies will bring more discipline to cloud spending even as it claims an ever-larger share of IT budget


THE MIT PRESS READER

Search Articles

Subscribe

The Staggering Ecological Impacts of Computation and the Cloud

Anthropologist Steven Gonzalez Monserrate draws on five years of research and ethnographic fieldwork in server farms to illustrate some of the diverse environmental impacts of data storage.

The Cloud now has a greater carbon footprint than the airline industry. A single data center can consume the equivalent electricity of 50,000 homes.

[Microsoft. *Environmental Sustainability Report*] [Amazon. *Sustainability Report*] [Google, *Environmental Report*]

Systems Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design

Improve Resource Efficiency

Left-over Carbon

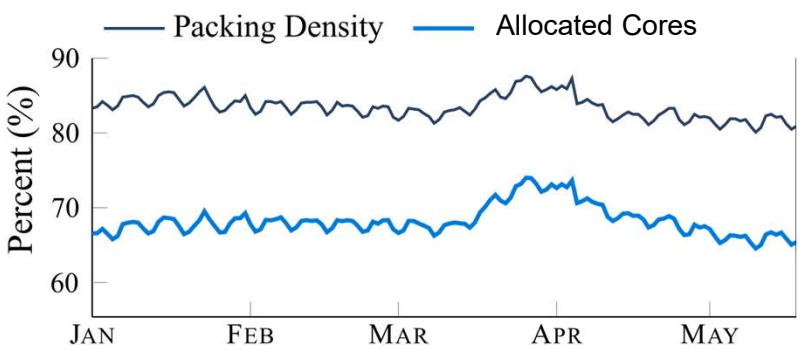
Scope 3

Scope 2

State of Practice

Insight: collocate workloads, reduce silos

➔ VMs allocated on most CPU cores



[Hadary, et al. *Protean.* OSDI 2020]

Research Avenues

VMs underutilize allocated resources

➔ Need to oversubscribe without impact

Example: harvesting allocated-but-unused CPUs

Key challenge: VMs are opaque!

[Wang, et al. *SmartHarvest. EuroSys 2021*] [Wang, et al. *SOL*. ASPLOS 2022]
[Parayil, et al. *Workload Characterization*. Submitted]

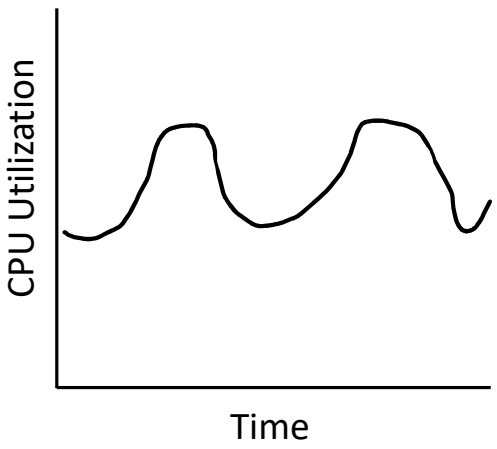
Our Research Avenues

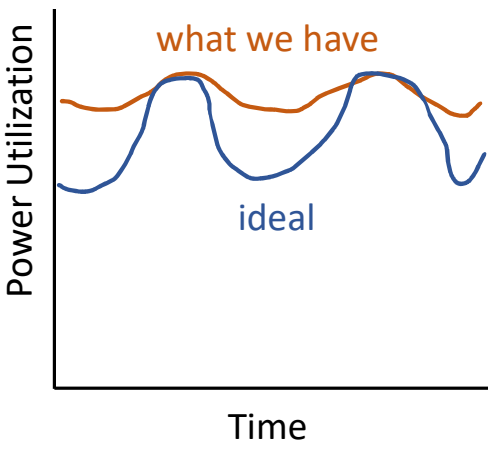
1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design

Improve Energy Efficiency

State of Practice

Long-standing challenge





Research Avenues

Many existing ideas

- Software awareness, software bloat
- Offload functionality
- Advanced C-states
- P-state or turn off idle servers
- Adaptive refresh for DRAM

➔Enable at scale: heterogeneous workloads, performance, load spreading

[Anderson, et al. *Treehouse. arXiv 2022*] [Yahya, et al. *AgileWatts*. MICRO 2022] [Li, et al. *LeapIO*. ASPLOS 2020]

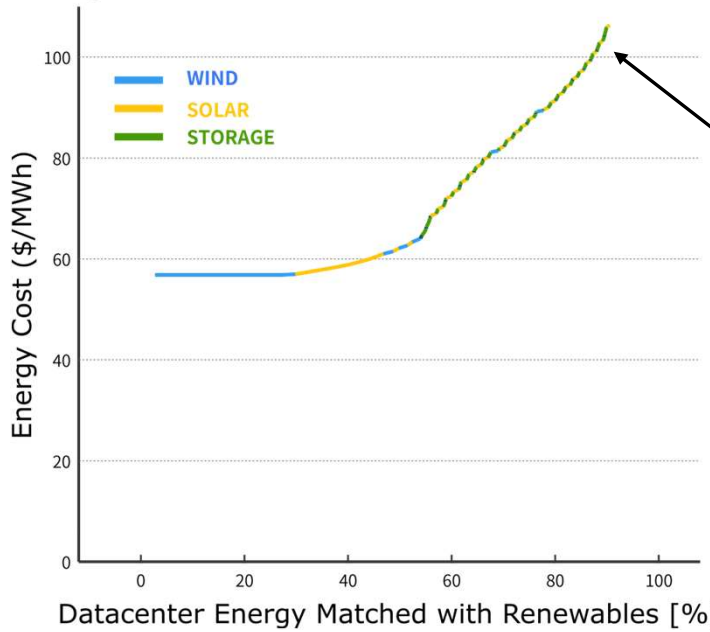
Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design

Enable Renewable Matching

State of Practice

Goal: renewable energy 100% of time at every DC



[Dysan, et al. *Clean Power by the Hour. RMI Report 2021*]

Research Avenues

Demand shifting

➔ Departure from current cloud model: customer specifies when & where

Grid-interactive UPS

➔ Requires datacenter-grid API

Opportunity: small shifts sufficient

[Zhang, et al. *Flex*. ISCA 2021] [James. *Grid-interactive UPS. 2022*]

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design

Extend Server Lifetime

State of Practice

Server, network carbon >> other infra

Recent lifetime extensions

	Minimum Lifetime	Date Changed
Microsoft	6 years	2022
AWS	5 years	2022
Google	6 years	2023
Meta	5 years	2023

Actual lifetimes often longer: ≈ 8 years

[Earnings releases in July 2022, February 2023]

Research Avenues

Pushing the envelope = lifetimes > 10 years

➔ Scheduling for high heterogeneity, performance-aware

➔ Repairs cost increasing

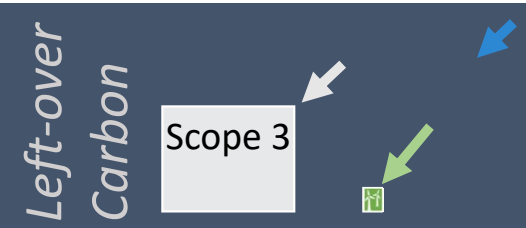
➔ Poor availability of repair parts

[Tomlinson and Porter. *Something old, something New*. HotCarbon 2022]
[Lyu, et al. Degraded Mode Operation. Submission]

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design

Datacenter (Re)Design

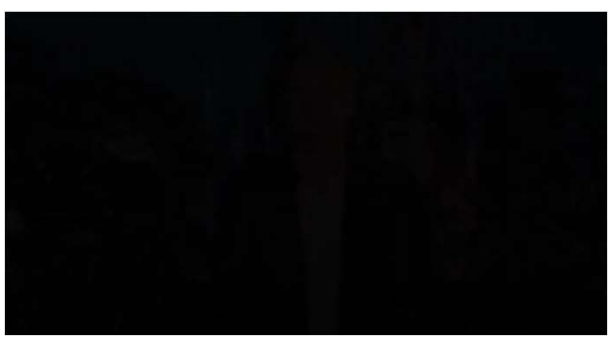


State of Practice

Overprovisioned redundant infra

Air-cooling

- 15-30% of energy
- Limits server lifetime: vibrations, oxidization, temperature fluctuation, …



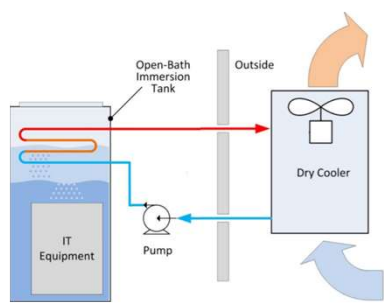
Research Avenues

Reduce redundancy

➔ Multiple availability offers

Immersion cooling

➔ Reduced failure rates

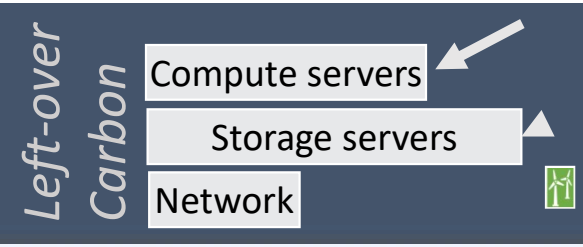


[Zhang, et al. *Flex*. ISCA 2021] [Majid, et al. *Immersion-cooled DCs*. ISCA 2021]

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design

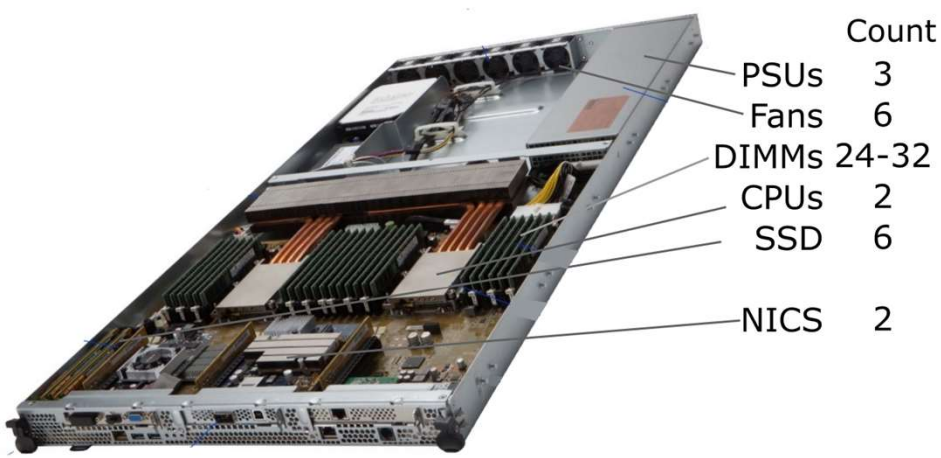
Server (Re)Design



State of Practice

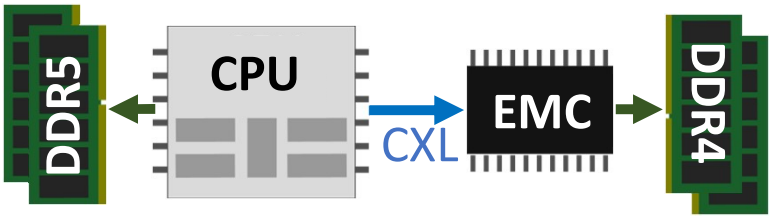
Majority of server carbon is chips

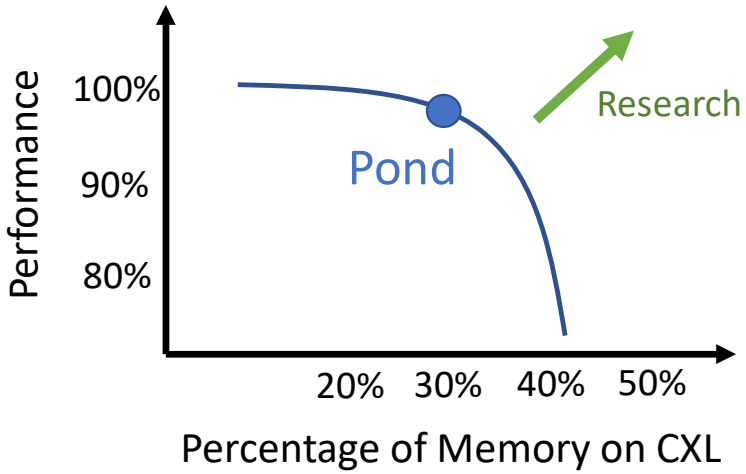
DRAM > SSD > CPU



Research Avenues

Reuse DRAM from old servers





[Li et al. *Pond*. ASPLOS 2023]

Microsoft Sustainability Commitments

Research Avenues

Renewable energy certificates for aggregate energy usage

Build new renewables that cover aggregate energy usage

#2, #3

Hourly/spatial demand matched by renewable energy

2012 2020 2025 2030 2050

#1

Baseline for scope 3 emissions

#4, #5, #6

Reduce scope 3 by at least 50%

Net-zero: removing all carbon emitted that year

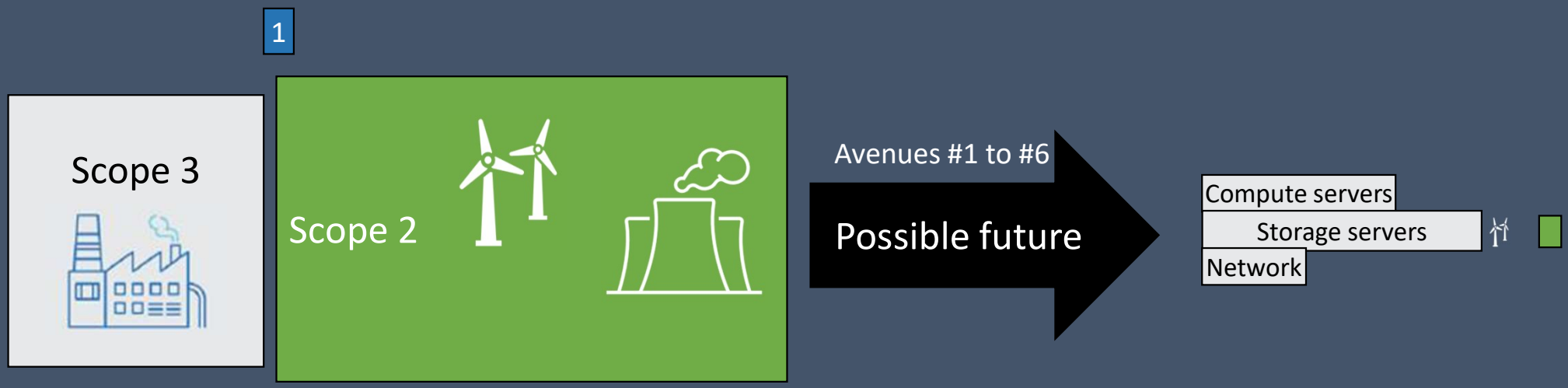
Removed historical carbon emissions (since 1975)

Scope 2 dominates

Scope 3 dominates

[Joppa, et al. *Lessons for Net Zero. Nature, 2021*]

Discussion



This outcome is not at all certain – research & transfer needed

Incorrect assumptions? More LCAs needed!

May want to think about: research avenues after #6?