A Call for Research on Storage Emissions

Sara McAllister*, Fiodar Kazhamiaka†, Daniel S. Berger†,
Rodrigo Fonseca†, Kali Frost†, Aaron Ogus†, Maneesh Sah†, Ricardo Bianchini†,
George Amvrosiadis*, Nathan Beckmann*, Gregory R. Ganger*

*Carnegie Mellon University †Microsoft Azure

PARALLEL DATA LABORATORY
Carnegie Mellon University

Storage in datacenters account for:

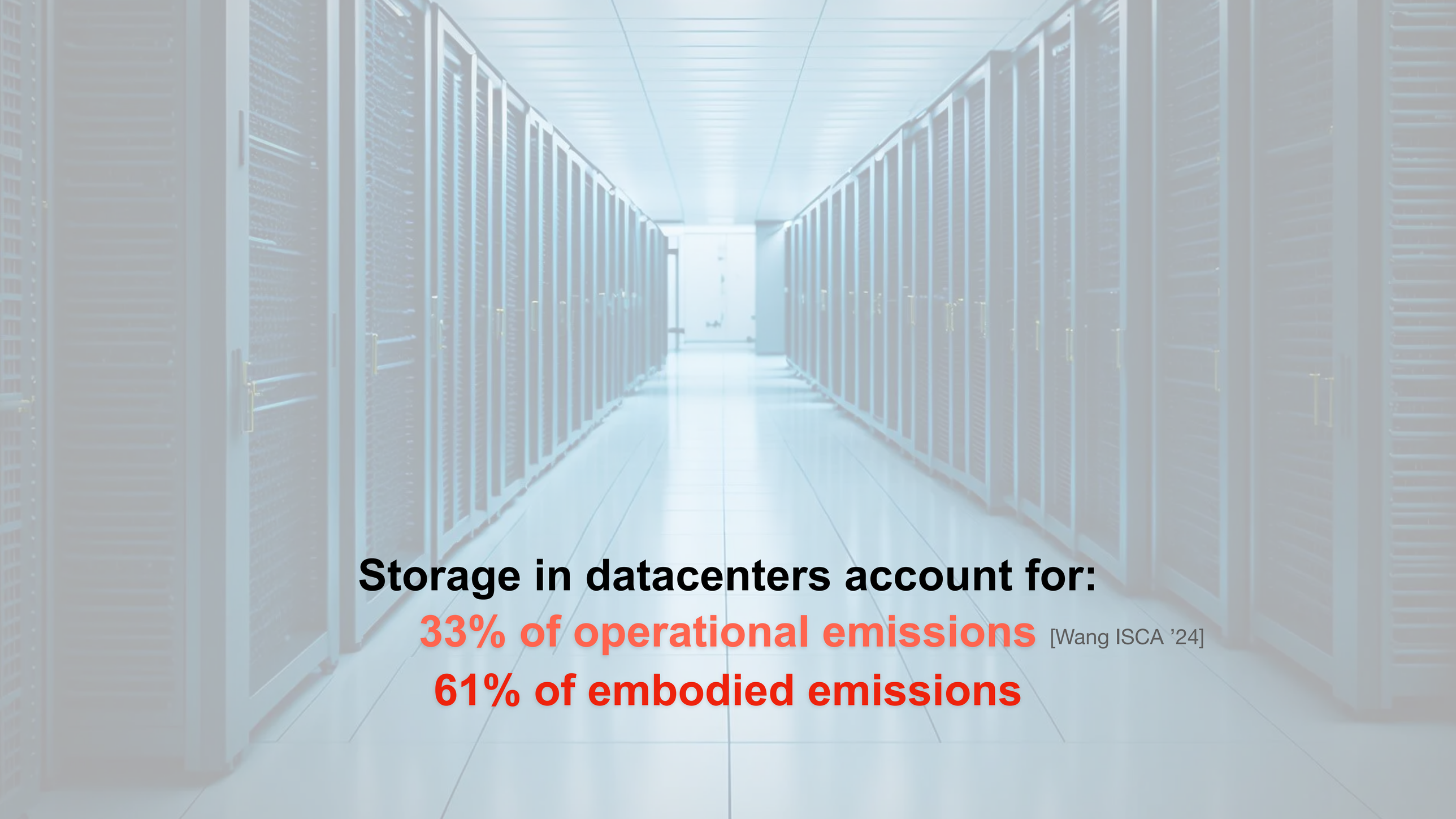
Storage in datacenters account for:

33% of operational emissions [Wang ISCA '24]

Storage in datacenters account for:

33% of operational emissions [Wang ISCA ’24]

Solution: Renewable energy



Storage in datacenters account for:

33% of operational emissions [Wang ISCA '24]

61% of embodied emissions

Storage in datacenters account for:

33% of operational emissions [Wang ISCA ’24]

61% of embodied emissions

So let’s talk about storage

But… what about AI?

But… what about AI?

Even given aggressive forecasts, **storage dominates emissions**

- AI = lots of operational emissions, offset by renewables
- **Embodied emissions: 2 CPUs ≈ 1 GPU ≈ 1.6-17 TB SSD**

What is datacenter storage?

What is datacenter storage?

8 HDD blades
1-2 CPUs, 88 HDDs ≈ 2.6 PB

What is datacenter storage?

8 HDD blades
1-2 CPUs, 88 HDDs ≈ 2.6 PB

Across cluster: 100,000s of disks

Multiple clusters per datacenter

Where are embodied emissions from?

HDD racks:

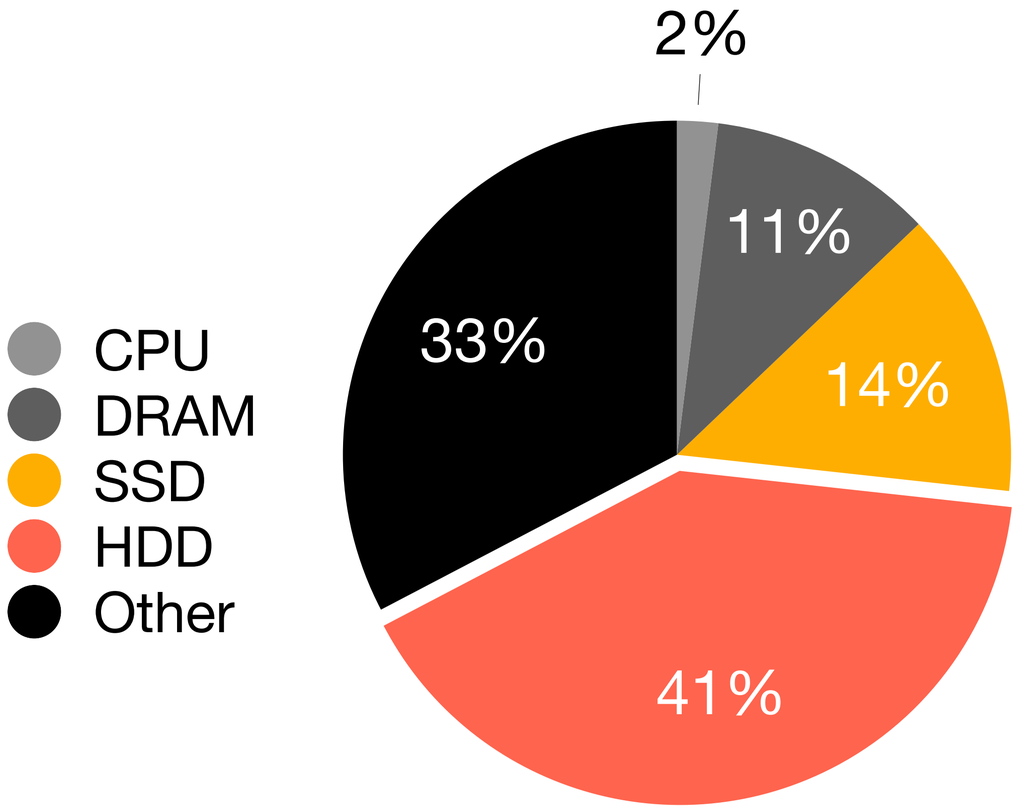
Capacity tier:
Less expensive per bit, large capacity

Other includes network switches, fans, power supplies, power delivery units, & passive materials

Where are embodied emissions from?

HDD racks:

Capacity tier:
Less expensive per bit, large capacity



Other includes network switches, fans, power supplies, power delivery units, & passive materials

Where are embodied emissions from?

SSD racks:

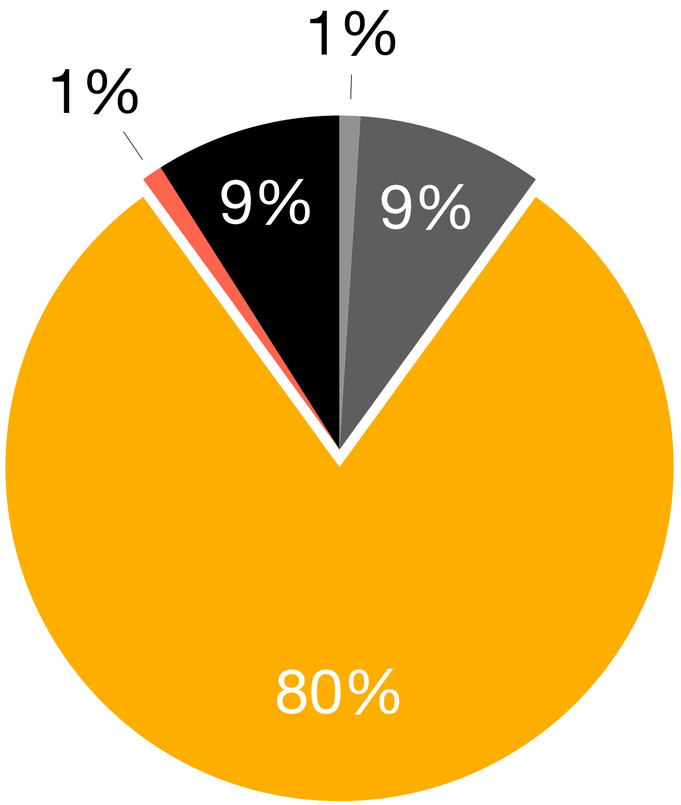
Performance tier:
More expensive per bit, lower capacity

Other includes network switches, fans, power supplies, power delivery units, & passive materials

Where are embodied emissions from?

SSD racks:

Performance tier:
More expensive per bit, lower capacity



Other includes network switches, fans, power supplies, power delivery units, & passive materials

Storage is different: Storage is stateful

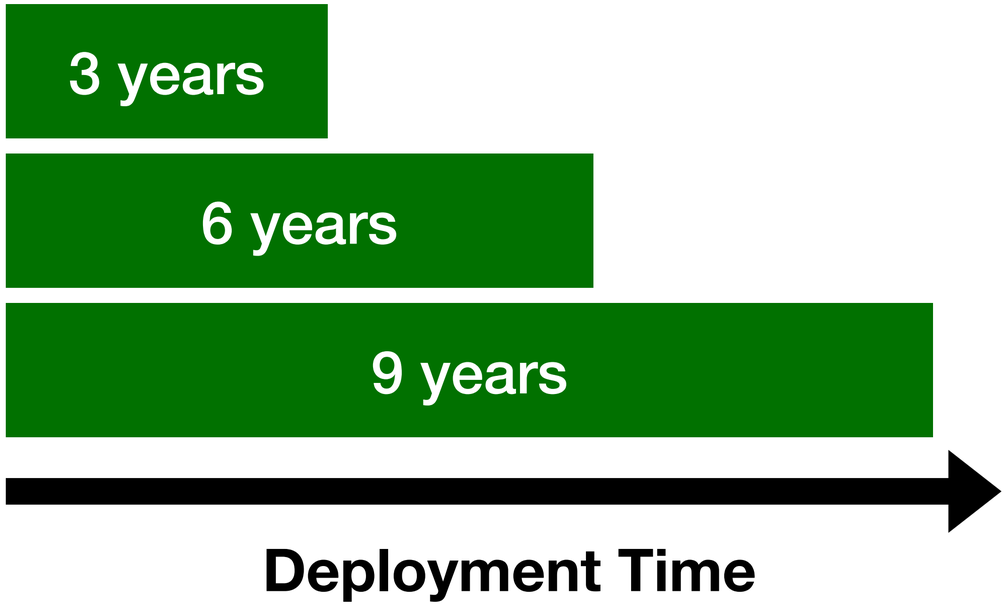
Longer lifetimes amortize embodied emissions

Extending lifetime causes **extra, correlated failures**

Storage is different: Storage wears out

Longer lifetimes amortize embodied emissions

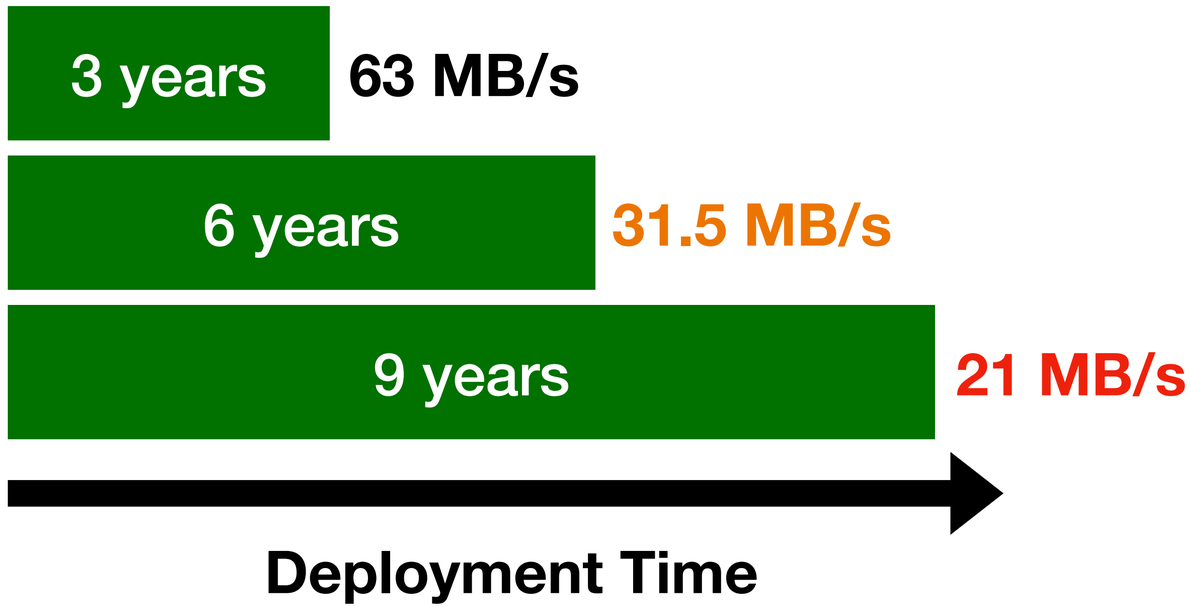
Extending lifetime causes **extra, correlated failures**



Storage is different: Storage wears out

Longer lifetimes amortize embodied emissions

Extending lifetime causes **extra, correlated failures**



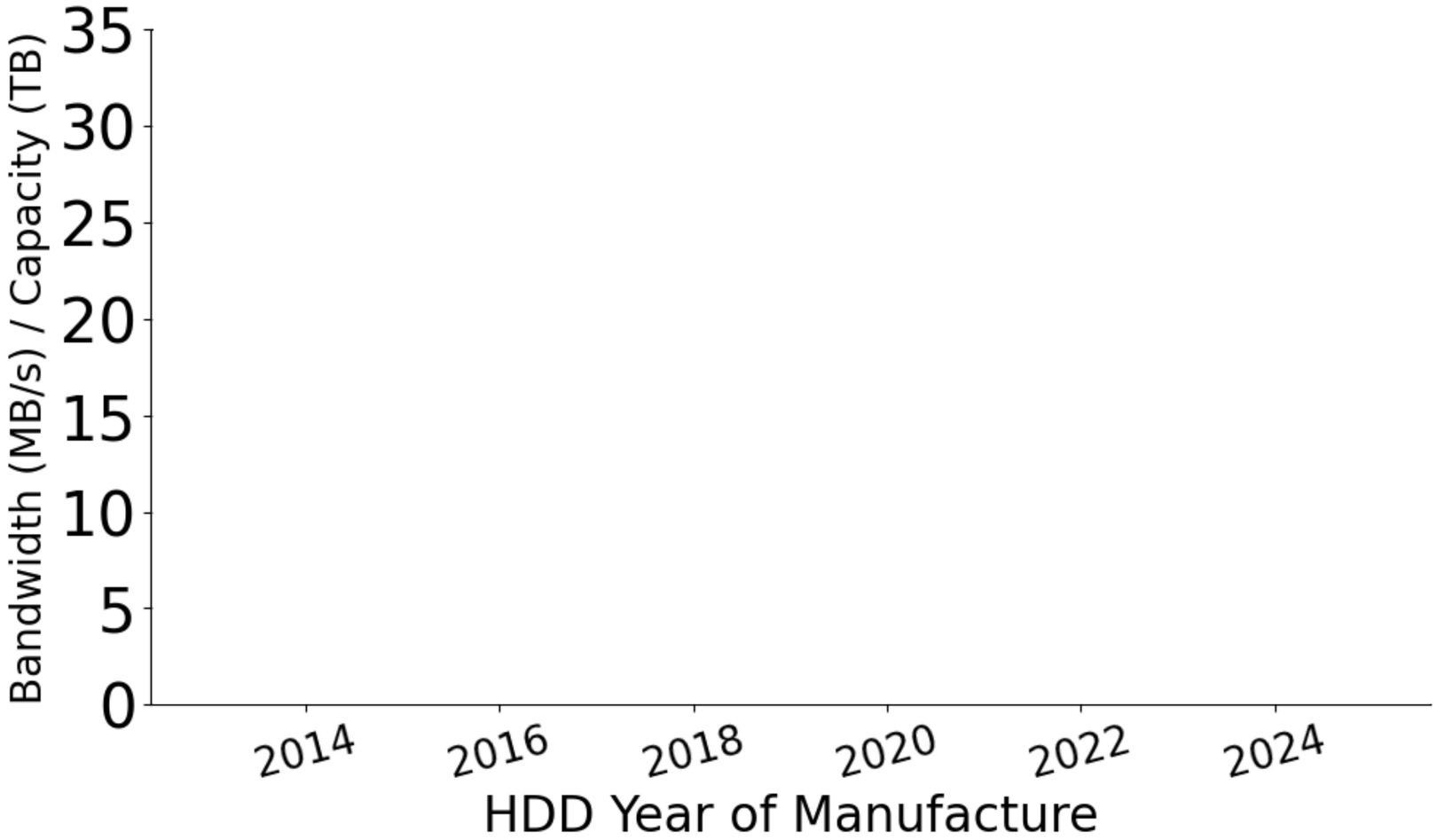
SSDs have limited writes

Storage is different: Denser drives → IO problems

Denser drives lead to fewer embodied emissions-per-bit

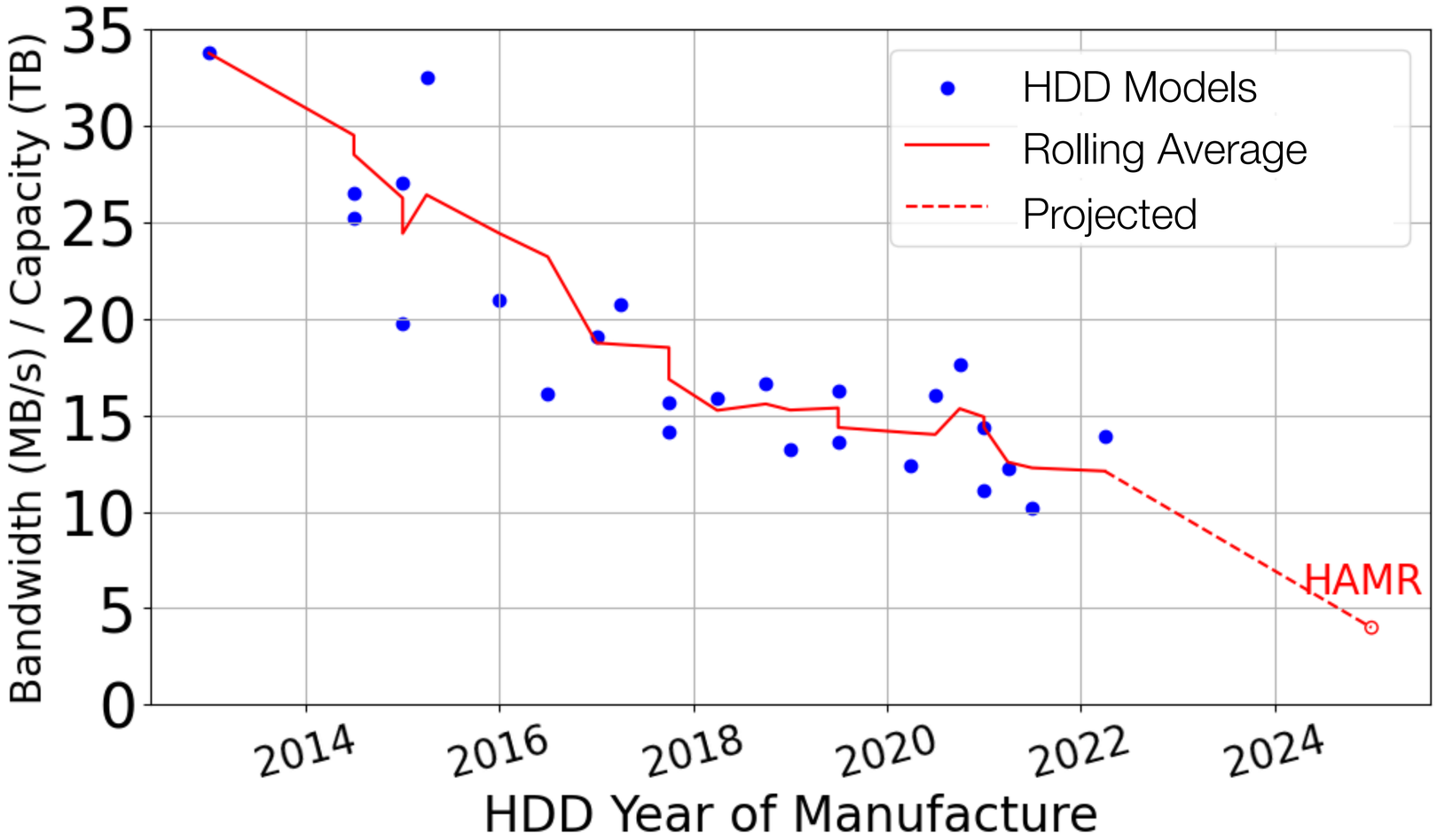
Storage is different: Denser drives → IO problems

Denser drives lead to fewer embodied emissions-per-bit



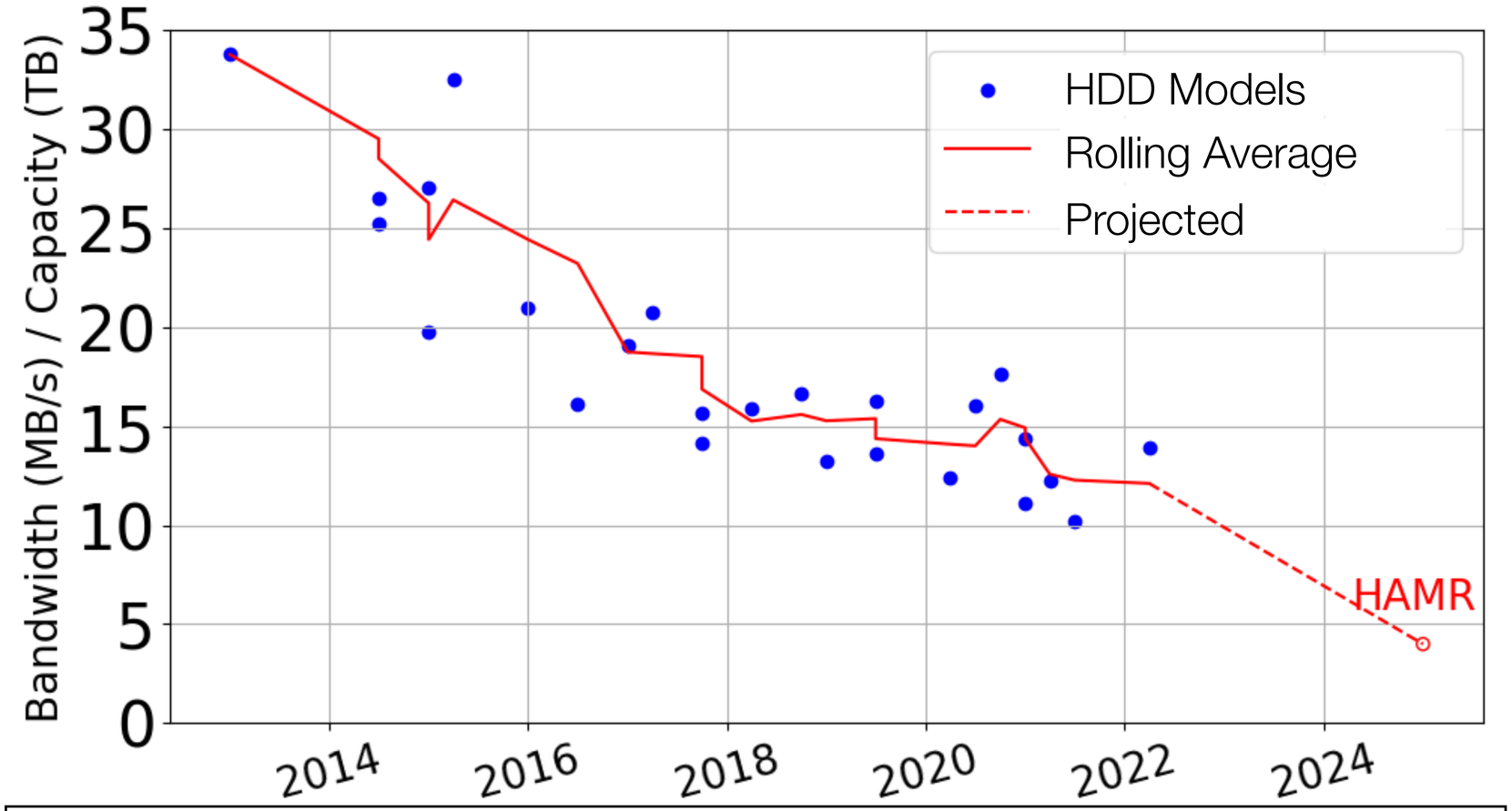
Storage is different: Denser drives → IO problems

Denser drives lead to fewer embodied emissions-per-bit



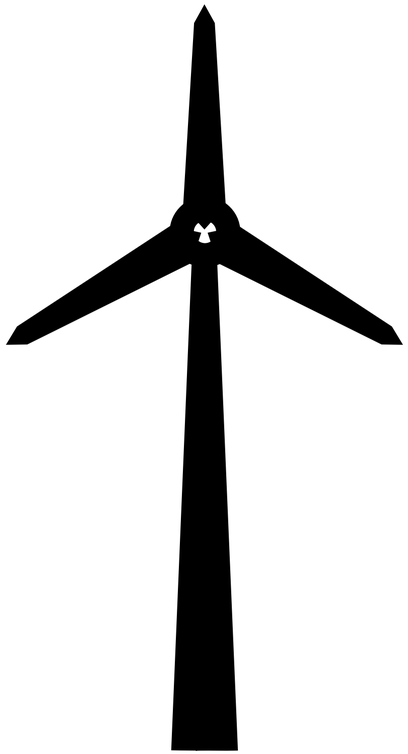
Storage is different: Denser drives → IO problems

Denser drives lead to fewer embodied emissions-per-bit



Bandwidth-per-capacity is falling (-8.5% per year)

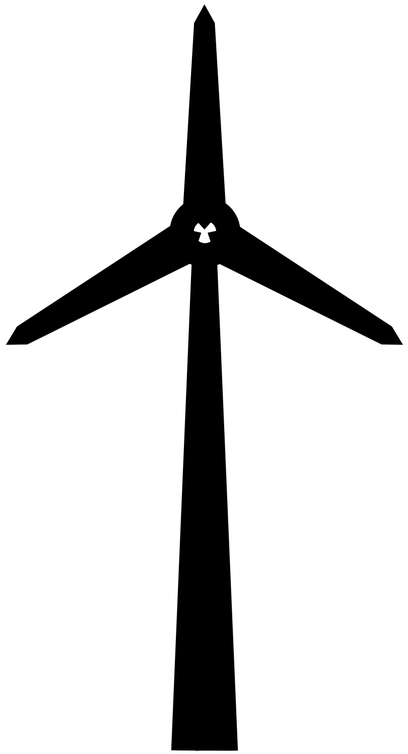
Storage is different: Embodied >>> Operational



Little dynamic power variation per day

Storage rack power changes <3% over a week

Storage is different: Embodied >>> Operational



Little dynamic power variation per day

Storage rack power changes <3% over a week

Need to focus on embodied emissions

Storage is different

We need solutions to reduce emissions in storage