Modern (Computer) Storage Systems: Overview

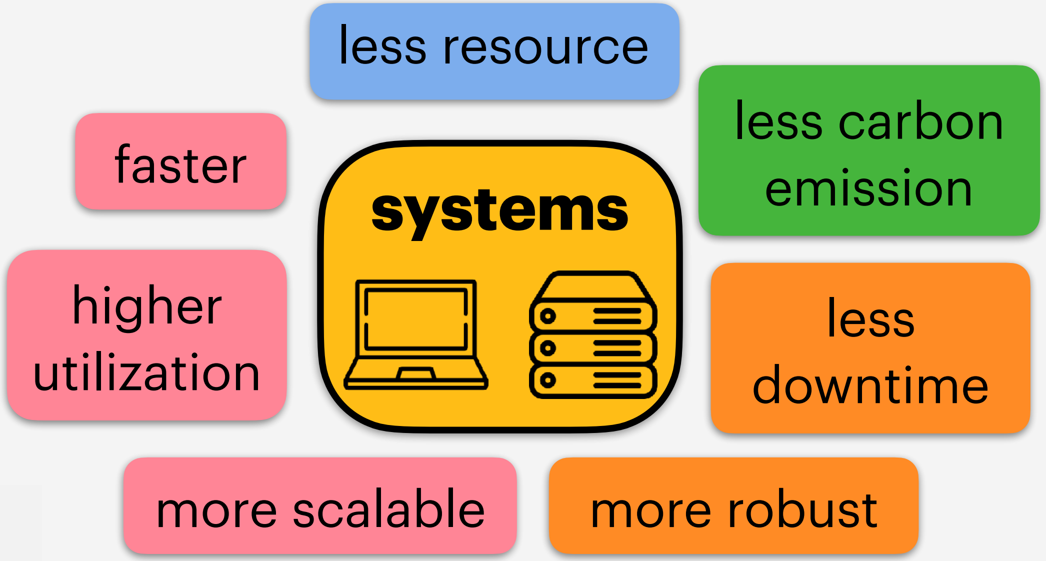
Juncheng Yang

Harvard John A. Paulson
School of Engineering
and Applied Sciences

HARVARD UNIVERSITY
MADSYS
Measurement and Design of Systems Lab

Harvard MadSys (Measurement and Design of Computer Systems) Lab

- Efficiency
- Performance
- Reliability and Robustness
- Sustainability



Me and TA

Juncheng Yang

Yao Xiao

Next, Your Turn

- Introduce yourself to others at the same table (5 min)
  - program and year
  - research interest
  - why you take this class and what you hope to learn from this course
  - anything you want to share
- Each table will select a leader to introduce everyone at the table

Plan for today

- Overview storage systems
- Logistics

Why should we learn about storage systems?

A small component in our laptop and servers

Why should we learn about storage systems?

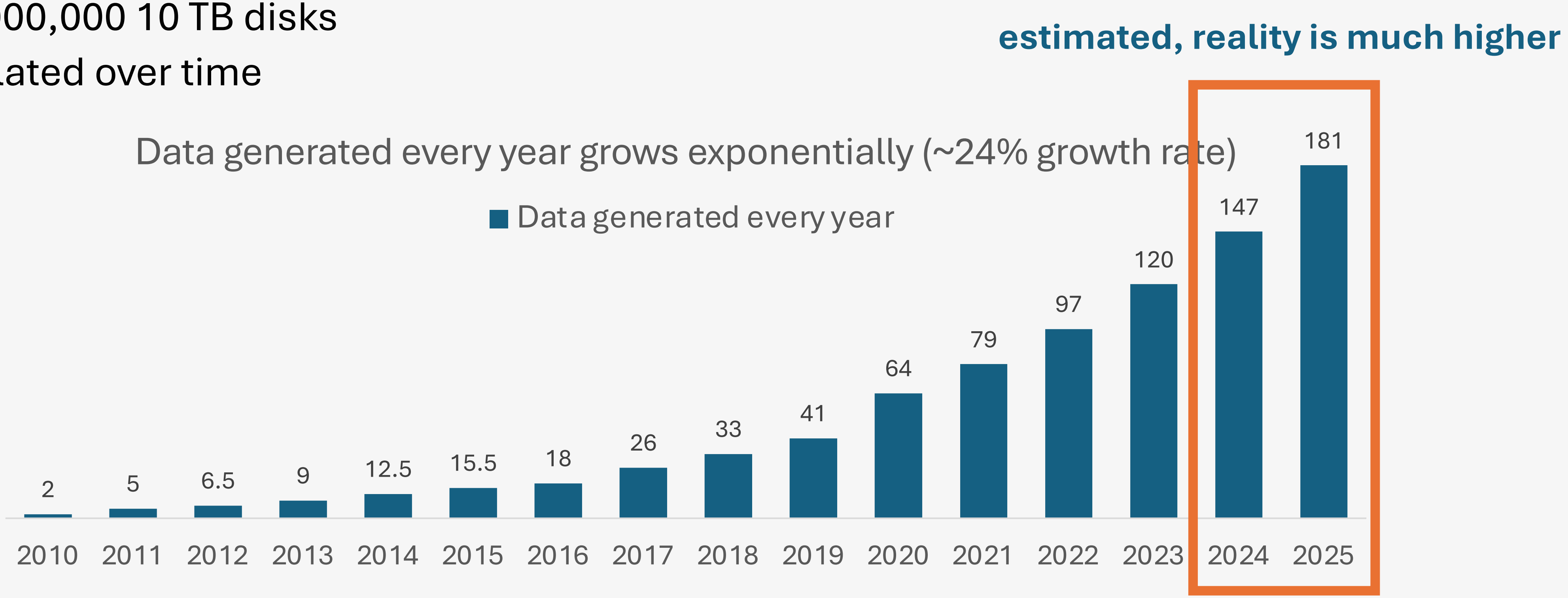
- A small component in our laptop and servers
- However, it is a critical component of almost every computer system, and often the performance and cost bottleneck of every large-scale system

- Capacity: too much data need to be stored
  - how much data do you think our society generates every second?

millions TB/s

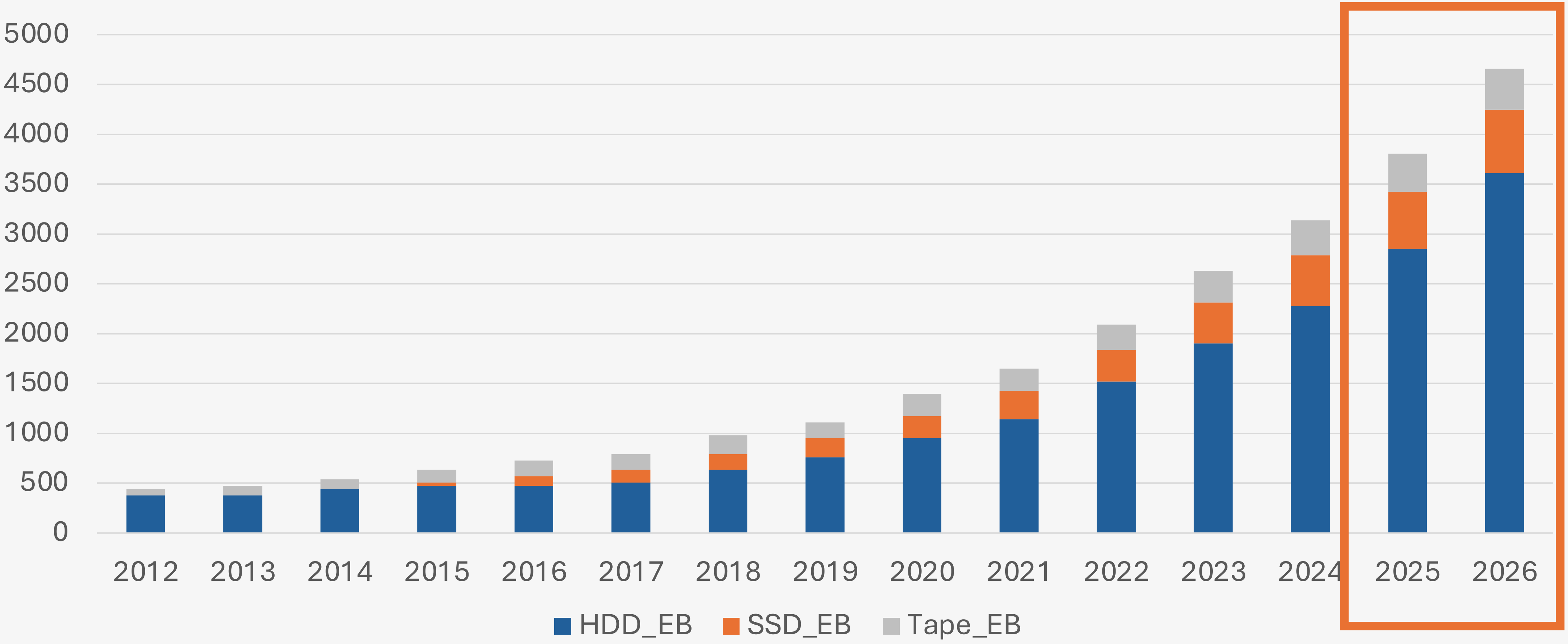
Data generation

- Data generated every year grows exponentially
  - 210+ ZB or 210,000,000,000,000,000,000,000 bytes (10^{24}) by 2026
  - => 6,659,000 TB/s
  - 21,000,000,000 10 TB disks
  - accumulated over time



https://www.statista.com/statistics/871513/worldwide-data-created/?srsltid=AfmBOornAApEbriWpNDChBgyrXYTeXcWmSwN9-kuE-rEXq_d_xzTsrWp

Annual storage shipped capacity



https://blocksandfiles.com/2025/05/17/disk-drive-capacity-shipment-out-to-2030-are-set-to-rocket-upwards/

Storage is the next victim of AI bubble

1TB SSD 2026 vs 2025 Prices Comparison

SSD	% Increase	January 2026	November 2025
Corsair MP700 Pro XT 1TB	38%	$219.99	$159.99
Samsung 990 Pro 1TB	83%	$199.99	$109.99
WD SN700 1TB	132%	$299.99	$129.99
Samsung 870 Evo 1TB	51%	$149.99	$99.99

Kioxia exec says the AI boom means the era of the cheap 1TB SSD is over —company's NAND supply is sold out for this year and likely through 2027

News By Zhiye Liu published January 21, 2026

Stock up on storage now before prices rise even more.

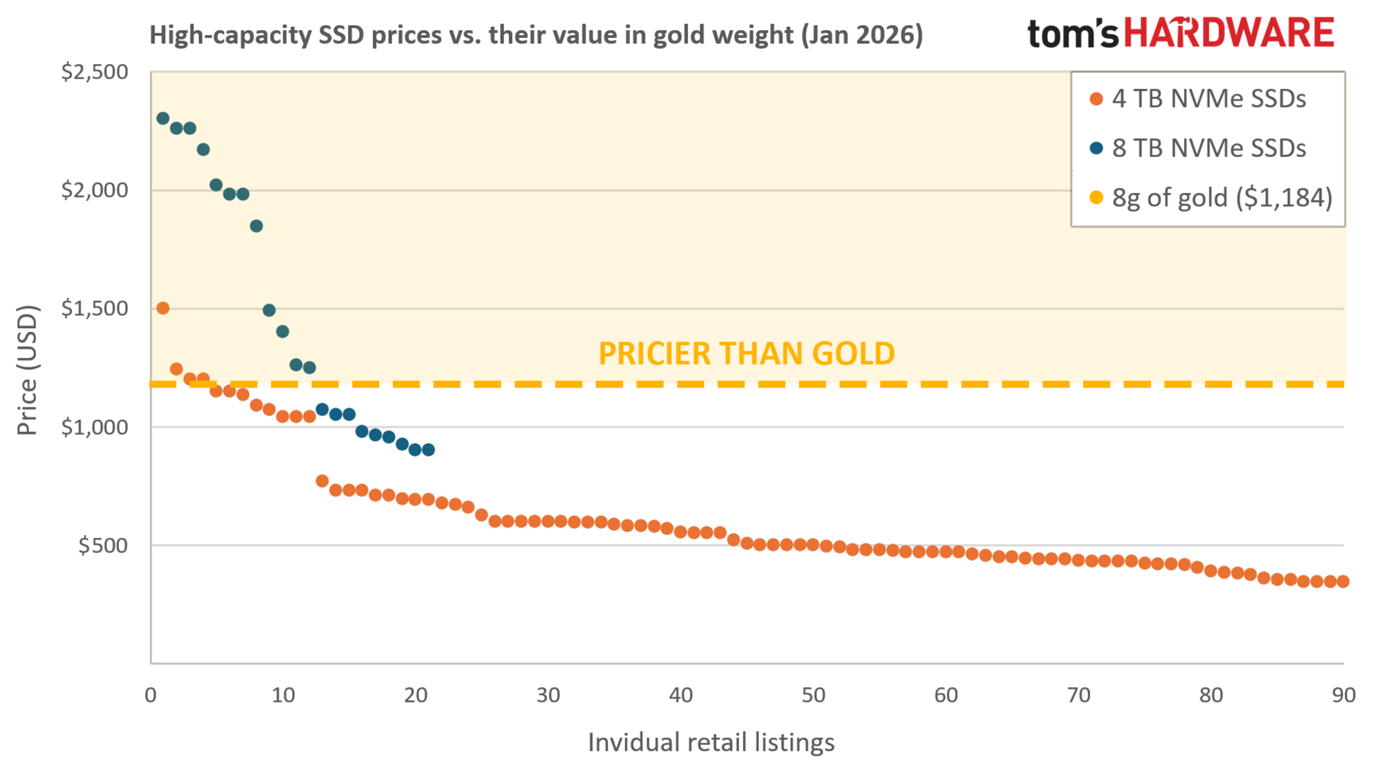
[News] Kingston Warns NAND Prices Have Surged 246% Since Early 2025, Signals More Price Hikes Ahead

2025-12-17 | Consumer Electronics / Semiconductors | editor

Storage is the next victim of AI bubble

Many high-capacity NVMe SSDs are now as expensive as gold by weight as shortage intensifies — we ran the numbers, here's what we found

News By Bruno Ferreira published January 16, 2026



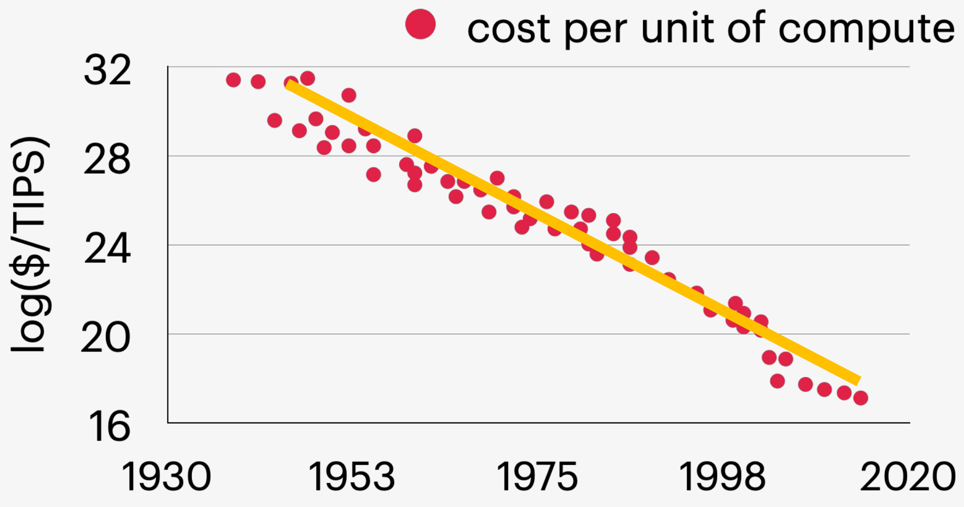
Why does AI need more storage?

- Model storage
  - llama4-maverick: 804 GB, GLM 4.7: 714 GB
  - checkpoints, fine-tuned variants
- Training data
  - companies: collect and store everything that they can
  - many variants of transformed and processed data
- AI generated data
  - images and videos
- Inference
  - LLM KV-cache

Storage hardware density improvement

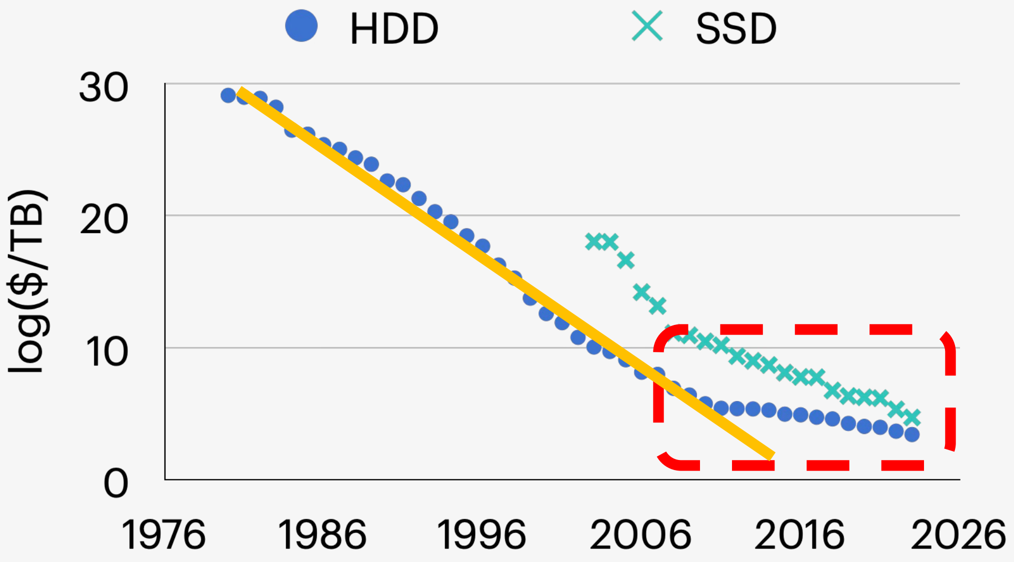
Moore’s Law (1965):

The number of transistors on an integrated circuit doubles roughly every 18 months.



Kryder’s L… (~2…):

The density … mation on hard drives doubles e… ths.



Why should we learn about storage systems?

- Capacity: too much data need to be stored
- Performance: storage is much slower than compute
  - HDD: 100 IOPS and 10ms, NVMe drive: 1 million IOPS and 10s μs latency
  - CPU: Ghz (1,000,000,000s operations)

Why should we learn about storage systems?

- Capacity: too much data need to be stored
- Performance: storage is much slower than compute
- Durability:
  - Do you think your data in the Google Drive and iCloud will never be lost?

Why should we learn about storage systems?

- Capacity: too much data need to be stored
- Performance: storage is much slower than compute
- Durability: 100% no data loss is impossible but data loss is not acceptable
  - fact: disk annual failure rate: 0.5–4%
  - if you have a million disks, 10-100 disk failure **every day**
    - "Amazon S3 provides the most durable storage in the cloud. Based on its unique architecture, S3 is designed to exceed 99.999999999% (11 nines) data **durability**."
    - S3 is "Designed to deliver 99.99% **availability** with an availability SLA of 99.9%"

Why should we learn about storage systems?

- Capacity: too much data
- Performance: storage is
- Durability: 100% no data
  acceptable


semianalysis

xAI's Colossus 2 - First Gigawatt Datacenter In The World, Unique RL Methodology, Capital Raise

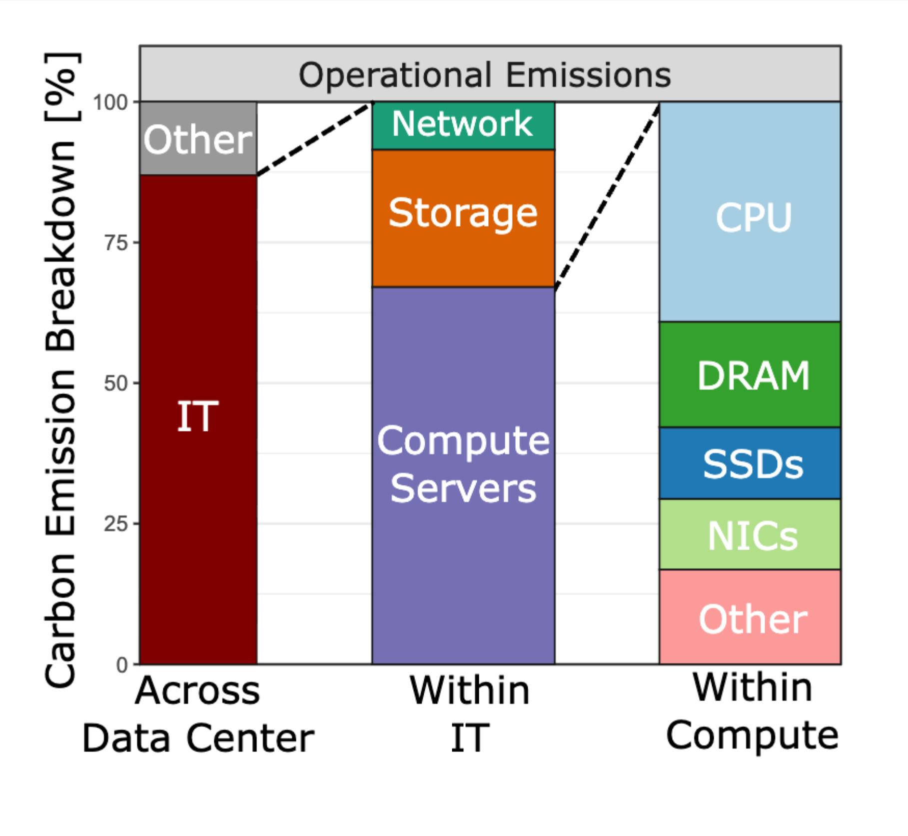
On Site Turbines, Mississippi Expansion, Solaris Energy, Can xAI afford it?, Middle East Funding, Tesla, Talent Exodus, API revenue, Consumer Growth, RL Environment

JEREMIE ELIAHOU ONTIVEROS, DYLAN PATEL, WEI ZHOU, AND 2 OTHERS
SEP 16, 2025 · PAID

- Sustainability
  - each disk consumes 10W, each year 100+ millions disks are manufactured
  - we need Gigawatts just for disks!
  - energy consumption occurs even when data are not being used (disproportionate to the load)

Why should we learn about storage systems?

Storage generate more embodied carbon than operational carbon



Why should we learn about storage systems?

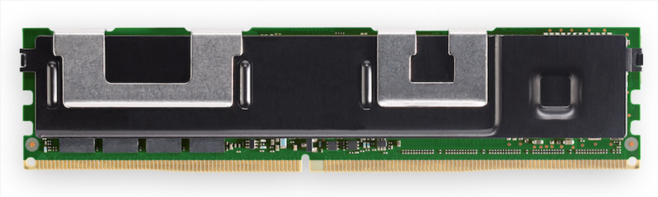
- **Capacity**: too much data need to be stored
- **Performance**: storage is much slower than compute
- **Durability**: 100% no data loss is impossible but data loss is not acceptable
- **Sustainability**: sustainable computing requires sustainable storage

What is a storage system?

More than just the storage devices, such as HDDs or SSDs!

What is a storage system?

- Hardware & software
  - media: HDD, SSD, NVRAM, Tape
  - controller and firmware
  - host interface and protocol
  - device driver
  - file system
  - client and libraries



What does a storage system do?

Providing access

Protection

Control

Persistence

Governance

Organization

What will we learn in this class?

- Storage hardware, e.g., hard disk drive, solid state drives
  - different components and how they work
  - the performance and reliability characteristics
  - historical and future hardware trend
- Storage software, e.g., file systems, distributed file systems
  - architecture
  - design choices and trade-offs
- Storage performance and reliability
  - storage device is slow and unreliable, how do we build fast and reliable systems

Sections

Part I The Only Component You Can Touch in a Storage System	Hard Disk and Solid-State Drives (three lectures)
Part II The Guy Sitting on Top of Disks	File Systems (two lectures)
Part III More Data, More Disks, What Should We Do	Distributed Storage Systems (two lectures)
Part IV Please Load My Data Fasterrrr	Storage System Performance Optimizations (two lectures)
Part V Leave No Byte Behind	Reliability of Storage Systems (two lectures)

This semester is the first offering of this course

Feedbacks are welcome!

- Teaching style
- Course content
- Can be anonymous if you prefer
- Send directly to the instructor
- Bonus is available!

Logistics

- Course website
  - https://modernComputerStorage.com
- Prerequisites: at least one system course
- Textbook
  - No
  - Background reading: Operating Systems: Three Easy Pieces (free)
- Format
  - First half: lectures
  - Second half: paper discussions
  - Two guest lectures

Logistics

- Office hour
  - Juncheng: Wed 2-3pm
  - TA: will update on website
- Contact
  - please send questions to cs2640-2026spring@googlegroups.com
    - **both** instructor and TA have access
    - **only** instructor and TA have access

Policy

- Late submission
  - up to eight days with 10% penalty each day
- Attendance and participation
  - not mandatory, but highly encouraged
  - no video recording
  - participation is 10% of your score
- AI-policy
  - encourage, be aware of AI hallucinations
  - important to learn how to use AI
  - **you are responsible for anything you submit**

Grading

- Mid-term exam (March 23): 20%
- Competition (Febuary – March): 10%
- Paper Presentation: 10%
- Project: 50%
- Class Participation: 10%
  - quiet roll call and in-class discussion
  - guest lecture attendance
- Bonus: 10%
  - creative use of AI
  - constructive feedback for instructor

Project

- Proposal: 10%
  - due March 09, one page **max**
- Midterm Checkpoint: 10%
  - due April 06, two pages **max**
- Project Presentation: 10%
  - April 27, 29, and May 04, peer-reviewed
- Project report: 20%
  - due May 06, three pages **max**, unlimited appendix, one page of AI use experience
  - code (with git history) should also be submitted for review

Project

- Individual project with AI coding agent
- Testbeds will be provided
  - sign up for cloudlab at https://cloudlab.us and join project cs2640
    - 100s of nodes with sudo access
  - contact the instructor and TA for other testbeds
    - specific hardware feature, e.g., CXL, FDP
    - physical hardware access
    - GPU related
- We provide free AI coding agents https://doc.freeinference.org
  - models: minimax m2, GLM 4.6 (soon 4.7 and 4.7 Flash), Qwen3-Coder-30B
  - other models: llama3-70B, llama4-Scout, llama4-Maverick
  - recommendation: VSCode + KiloCode (plugin) with Minimax or GLM (model)
  - you are welcome to use it outside of classroom, but DO NOT abuse it

Project ideas

- Research projects are welcome
  - integrating with your own research: ok, but should not use a finished project
  - e.g., building a more user-friendly or agent-friendly storage interface
- Measurement-heavy projects are encouraged
  - understand how existing systems work
  - e.g., how scalable are BeeGFS and 3FS? What are the bottlenecks?
- Crazy and engineering-focused projects are also valued
  - we all love to have fun, do we?
  - e.g., build a Dropbox using GitHub
- Reach out to instructor and TA for ideas if you need any
  - mention your interest and your goal in the email
- Not sure if it is a good idea? Run it by instructor and TA
  - key metric: satisfy your curiosity and try things no one has tried

History of storage systems: magnetic storage

- **~1950: magnetic tape adapted for data storage**
  - the start of high-capacity inexpensive "mass storage"
- **1956: IBM 350 RAMAC (Random Access Method of Accounting and Control)**
  - first random access (within milliseconds) storage
  - a massive unit, 50 platters, each 24 inches in diameter, total capacity ~3.75 MB
- **1971: IBM 3330**
  - tracking-following servo technology, tighter data packing (4x)
- **1973: IBM 3340 (template of modern HDDs)**
  - sealed Winchester heads, improved reliability and density
- **1980: Seagate ST-506**
  - the start of PC storage

History of storage systems: optical storage

- **Mechanism:** utilize light reflection to read and write data
- 1970-1980: compact disc (CD) from Philips ~700MB
- Later: DVD (4.7GB), Blu-ray (25GB) driven by optical physics, using shorter wavelength laser and larger numerical aperture
- eventually lose to flash memory due to speed and reliability

History of storage systems: solid-state storage

- **1976:** Dataram's "Bulk Core" (2MB)
- **1991**: SanDisk makes the first commercial SSD (20MB)

Today

- Consumer device
  - flash drives (dying)
  - mostly solid-state drives (SSDs)
- Enterprise
  - shifting from hard disk drives (HDDs) to SSDs
- Data center
  - majority storage capacity is still HDD
  - SSDs are frequently used for boot drives, caches, databases
  - tape is also widely used
- Future: Glass and DNA





Units and metrics

Metrics	Unit	Meaning
Latency	µs, ms	duration between sending a request and receiving the data
Bandwidth (~~throughput~~)	MB/s, GB/s	data transfer rate
IOPS		I/O requests per second
Hit ratio / miss ratio	N/A	#hits / #requests

Confusing units

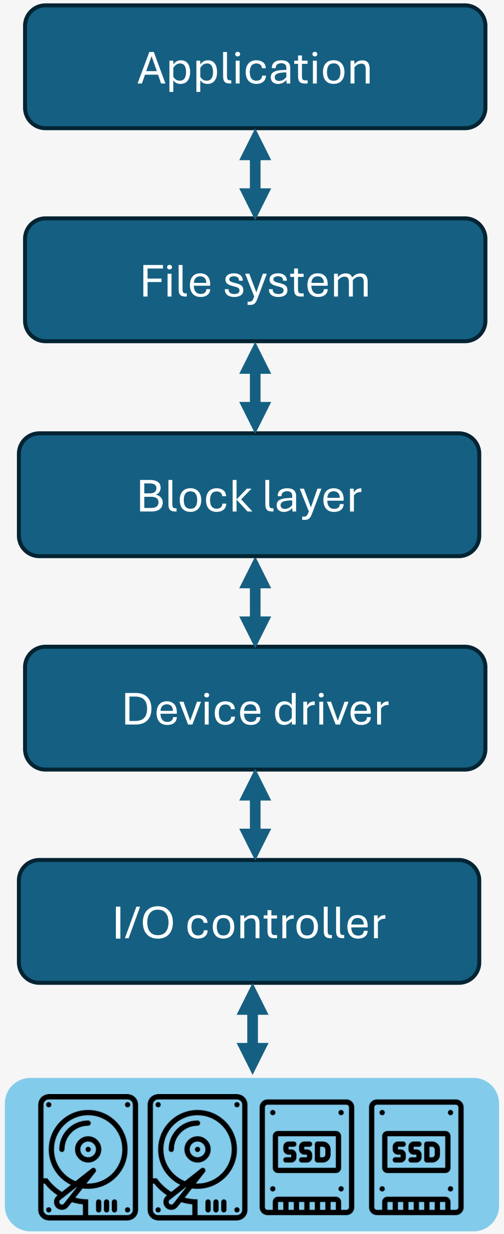
- MB (megabyte): 1,000,000 bytes (10^6)
- MiB (mebibyte): 1,048,576 bytes (2^{20})
- in operating systems and old low-level tools: MB=mebibyte
- in storage and networking MB most often refer to 1,000,000 bytes

Confusing units and terminology

- Block vs. page
- virtual memory
  - only **page** should be used
  - typically 4 KiB, but can be 2 MiB or 1 GiB
- storage and database
  - **block**: the unit of I/O operations
  - page: the same as block (but less common)
  - typically 4 KiB, but can be 512 B or larger
- SSD
  - **block**: erase unit (typically MBs)
  - **page**: read and write unit (typically 4-16 KiB)

Layers involved when serving a request

- **Application**
  - open, read, write
- **File system**
  - translate file operations into logical block addresses
  - manage directories, permissions, metadata
- **Block layer** (scheduling)
  - queue, re-order and merge requests, enforce priority and QoS
- **Device driver**
  - translate generic block commands to hardware-specific commands
- **Controller**
  - hardware that implements the AHCI/NVMe protocol, manages command queues, DMA transfers, generates interrupt
- **Physical layer**
  - electrical signals, e.g., SATA, PCIe



Life of a request

- **Application**: `read(file, offset, length)`
- **Filesystem**: translate to LBA 12345, 8 blocks
- **Block layer**: queue and re-order request
- **Driver**: build AHCI/NVMe command structure
- **Controller**: DMA setup, send command to disk
- **Disk**: execute, DMA data to memory
- **Controller**: interrupt CPU
- **Driver**: handle completion, wake application

Next time

- Hard disk drive
- Before end of next week
  - take a look at the schedule and topics
  - pick a topic and date for your presentation