HEPiX Technology Working Group Report

HEPiX Spring 2024 Meeting

April 15-19, 2024

Andrea Chierici, Shigeki Misawa, Andrea Sciabà

Introduction

- HEP/NP experiments depend on advances in compute, network, and storage technologies
  - Changes in technology can affect current and future experiments
- The technology landscape has undergone very significant changes over the past few years
- The HEPiX Technology Watch WG was created in 2018 to monitor trends in technology and provide periodic reports to the HEPiX community on the state of technology
  - Outgrowth of previous work by Bernd Panzer and Helge Meinhard
- The group has been dormant since the start of the COVID-19 pandemic with the last meeting in Feb 2020 and the last major presentation at CHEP 2019

Restarting Techwatch

- New co-conveners: Andrea Chierici, Shigeki Misawa and Andrea Sciabà
- First meeting of the reconstituted Techwatch group on January 17
  - Introduce the purpose of the group and deliverables as originally envisioned by Bernd and Helge as well as the technologies of interest
  - Significant discussions on the nature of the working group
- Second meeting February 21
  - Identify the major areas of interest
  - Establish the structure of the organization
    - Create subgroups covering each major area of interest
  - Identify collaborative tools to be used by the subgroups
- Third meeting March 28
  - Approval of new mandate and organization
  - Preparation of the workshop presentation

Working Group Charge

The Working Group is tasked with the following duties:

- Understand the <u>trends</u> and the direction of the technology markets using <u>publicly available sources</u>
- Assist in making <u>cost predictions</u> and optimizing investments, taking also into account sustainability
- Provide technical, and where possible financial, <u>risk assessments</u> for technologies
- Leverage the <u>expertise</u> of the HEPiX community
- Inform the HEPiX board about technologies that may warrant a more in-depth investigation

Current Techwatch Subgroup Participants

- **General market trends**
  - Conveners: Shigeki Misawa
  - Other members: Xavier Espinal, Andrea Sciabà, James Walder
- **Server and data center infrastructure**
  - Conveners: Tristan Suerink
  - Other members: Andrea Chierici, Michele Michelotto, Eric Yen
- **Processing units, memory, buses and interconnects**
  - Conveners: Andrea Chierici
  - Other members: Erik Kooistra, Dhruva Kulkarni, Charles Leggett, Michele Michelotto, Andrea Sciabà, Andrea Valassi
- **Storage**
  - Conveners: TBD
  - Other members: Thomas Byrne, Martin Gasthuber, Alison Packer, Vladimir Sapunenko
- **Network**
  - Conveners: Harvey Newman
  - Other members: Dhruva Kulkarni, James Walder

Subgroups are open to everyone!

HEPiX

Deliverables and Audience

- Presentations at HEPiX workshops
  - One or more presentations from the working group at each meeting
- Presentations at other venues
  - WLCG GDB, WLCG workshops, conferences, …
- Reports
  - "Executive" summaries
    - Short documents for consumption by project planners
    - Identify technologies worthy of more in depth investigation to the HEPiX board
  - In-depth reports
    - Longer documents providing details on current state of technologies, major milestones and decision points in the evolution of technology
    - Targeted at the entire HEP/NP community
  - Live documents
    - Expectation is that all deliverables will be updated on a periodic basis

Call for Participants

- <u>Volunteers are needed</u> to lead subgroups or contribute information
  - Experts on each subject can contribute the knowledge they already have
  - Research information from public web-based sources
  - Can be an opportunity to learn more about technologies of interest, both in breadth and depth
  - But do not publish information covered by NDA!
- Minimal time commitment
  - Current meetings are monthly
  - Round tables among subgroups to report relevant developments
  - Expect contributions to be a minor extension of investigation that participants are already engaging in
  - Still, it is essential to keep that commitment and avoid concentrating effort close to deadlines!

Today's (very) preliminary report

- No documents ready yet
  - Still ramping up activities
- Only a few highlights from each area
  - No time to discuss any particular topic in detail!
- Main purpose for today is to present the "new" working group
  - And encourage more people to contribute!

- More information
  - Website: https://w3.hepix.org/techwatch.html
  - Indico category: https://indico.cern.ch/category/10621/
  - Mailing list: hepix-techwatch-wg@hepix.org

General market trends

(Shigeki, Andrea S.)

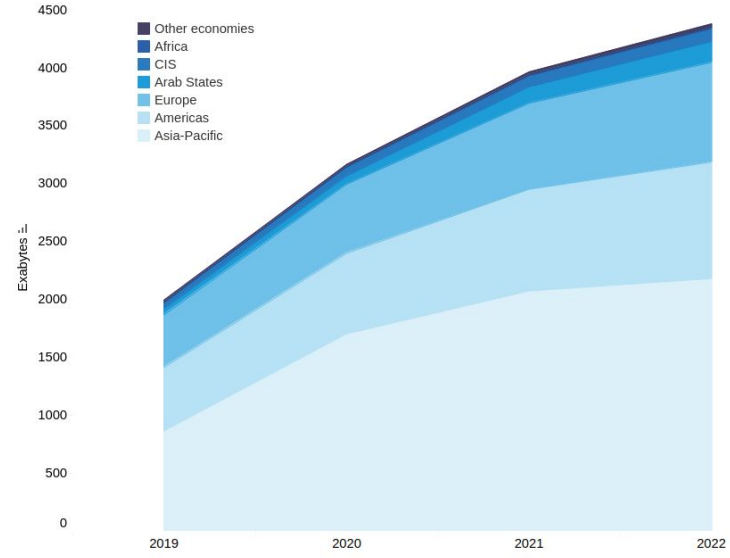# Global Internet Usage

- Internet usage continues to grow

https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/

Individuals using the Internet
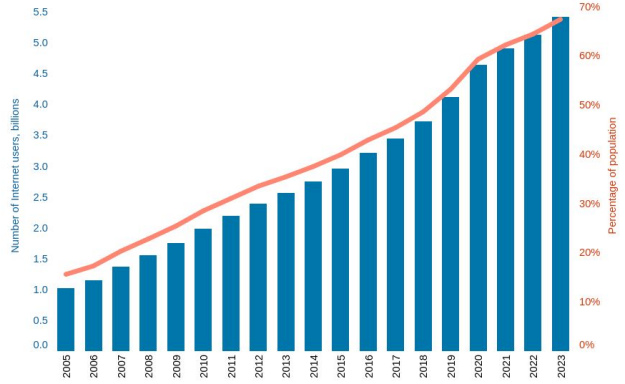
Source: ITU

Fixed-broadband traffic, 2019-2022

Legend:
- Other economies
- Africa
- CIS
- Arab States
- Europe
- Americas
- Asia-Pacific

○ Fixed-broadband traffic
○ Mobile-broadband traffic

Note: 1 Exabyte = 10^12 Megabytes. Mobile-broadband traffic refers to traffic within the country.
Source: ITU

- Internet traffic continues to grow

https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-traffic/
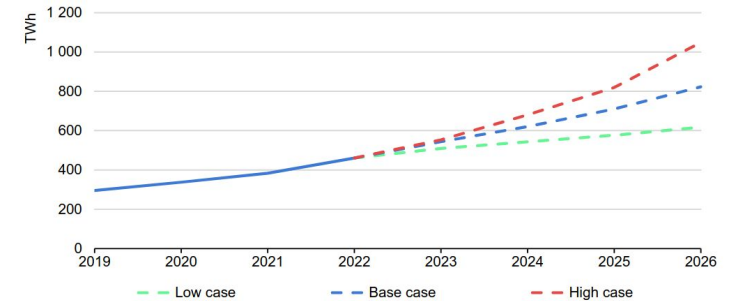
Global Data Center Power Consumption

- Data center power use was ~2% (460TWh) of global electricity use in CY2022 [1]
- AI and cryptocurrency are expected to add to power consumption [1]
- Inference currently more energy intensive higher than training [2]
- Power used by AI Google search estimated at 10x more power than standard search

[1]
https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf
[2] https://www.sciencedirect.com/science/article/pii/S2542435123003653?dgcid=author

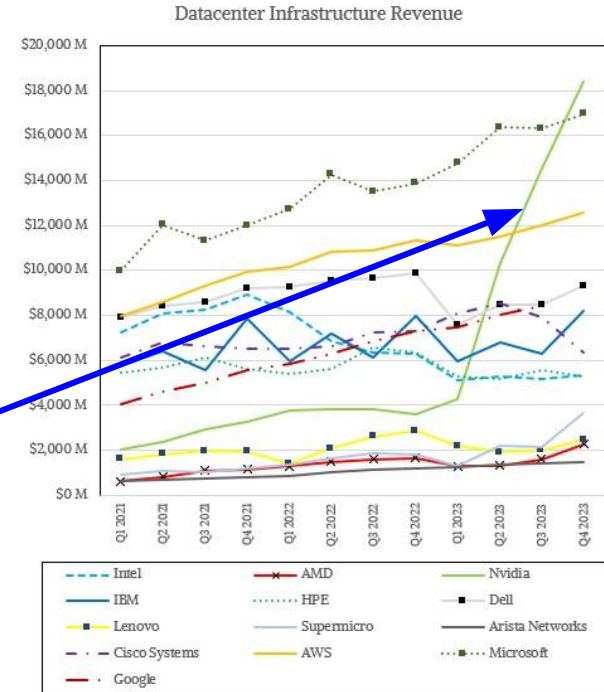Global electricity demand from data centres, AI, and cryptocurrencies, 2019-2026

Estimated electricity demand from traditional data centres, dedicated AI data centres and cryptocurrencies, 2022 and 2026, base case
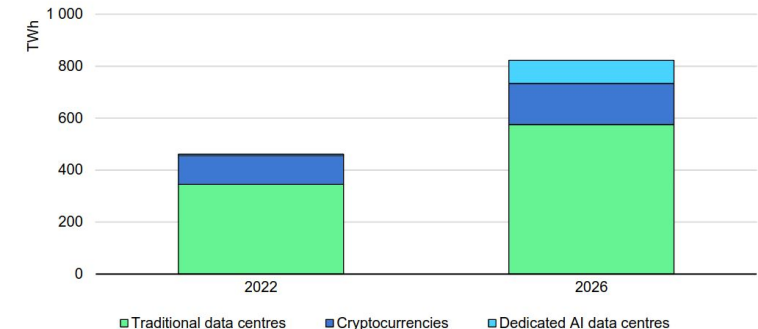
Computing Major Players at Scale

- Relative size of data center operators and equipment providers based on revenue estimates from The Next Platform [1]
  - The players influence the direction and speed of technology and product development
  - Nature of relationships define who has power
- Note the dramatic impact of AI in recent years as reflected in Nvidia revenues

[1] https://www.nextplatform.com/2024/04/08/mixed-results-for-the-datacenter-thundering-thirteen-in-q4/

Datacenter Infrastructure Revenue

Legend: Intel, AMD, Nvidia, IBM, HPE, Dell, Lenovo, Supermicro, Arista Networks, Cisco Systems, AWS, Microsoft, Google

General Semiconductor Market

- Worldwide revenues recovering from dramatic declines in late CY2022 and early CY2023
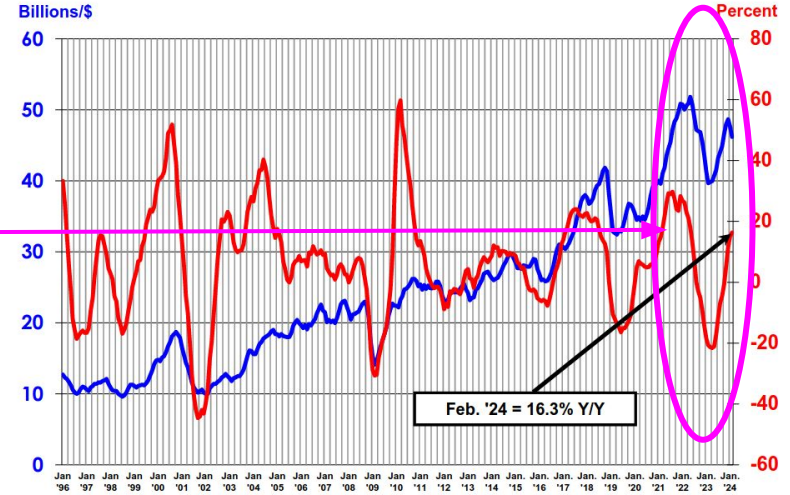
  https://www.semiconductors.org/wp-content/uploads/2024/04/February-2024-GSR-table-and-graph-for-press-release.pdf

2022 DEMAND BY END-USE

End-Use Category	Communication	Computer	Industrial	Consumer	Automotive	Government
Annual Growth	1.0	-13.9	24.2	17.7	13.4	56.5
Total Value ($B)	172.3	150.7	83.1	80.5	78.3	9.1

Worldwide Semiconductor Revenues
Year-to-Year Percent Change

Billions/$

Percent

Feb. '24 = 16.3% Y/Y

Source: WSTS — Revenue, Y/Y % Change

- Computing represents ~⅓ of the demand for semiconductors

Trends for fabrication processes

- Roadmap until 2036
  - Sub-1nm process nodes
  - Transition from FinFET transistors to Gate All Around nanosheet designs
- "CMOS 2.0"
  - Smaller nodes are more expensive
  - Breaking down chips into functional units using 3D designs helps bringing down costs
- Only three makers for leading edge chips - TSMC, Samsung, Intel
  - Huge investments planned on fabs in diversified regions

Potential logic scaling roadmap extension

	2018	2020	2022	2024	2026	2028	2030	2032	2034	2036
	N7	N5	N3	N2	A14	A10	A7	A5	A3	A2
Metal Pitch (nm)	40	28	21	21	18	18	16	16	16	16-12
Metal Tracks	6	6	6	6	5	5	5	4	<4T	<4T
	finFET	finFET	finFET	GAA NSFET	GAA NSFET	FSFET	FSFET	CFET	CFET	CFET atomic

Continued dimensional scaling

Device and material innovations

Context-aware interconnect

[Tom's Hardware](#)

SoC

Breaking down the complex System-on-Chip (SoC) into smaller, more manageable components

The blocks can be optimized separately at design stage but also at technology level

Reassembled using 2.5D (interposers) or 3D stacking

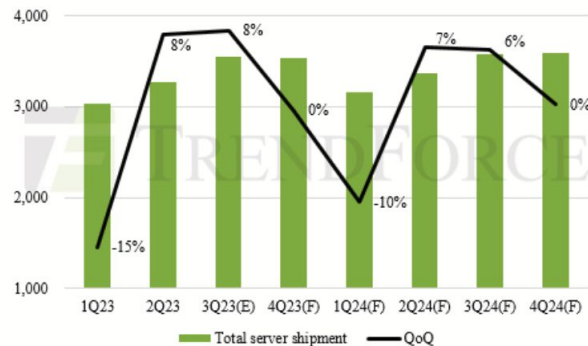Advanced packaging for high-bandwidth and low-latency connectivity

Server market

- Server shipments stagnant, all growth is in AI and long term shift to cloud

- ARM servers in the data center still very scarce but growing
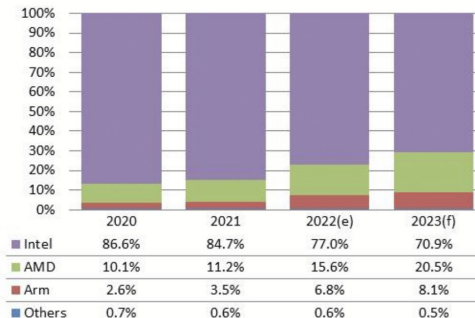 - AMD also gaining ground

Global server shipments for 2024

Tom's Hardware

Server and data center infrastructure
(Tristan)

More unique server designs on the market

- Motivated by fierce competition between AMD and Intel, but also from multiple ARM suppliers
- AMD's fully functional mono socket strategy (and high numbers of cores) gives OEM's the option to engineer non-standard designs, like the ASUS ESC4000A series or the Lenovo SR655 series
- Intel still doesn't want to give up on their dual socket strategy
    - … and hitting a lot of walls because of it (power, thermals, space)
- Need to see more ARM based systems in the near future

Liquid cooling will become mainstream

- Current last generation 1U systems need heatpipes with big radiators to cool down 400W+ CPU's in this form factor
- Expecting for next generation CPU's that 1U systems will become rare and 2U or even bigger will become the standard
- There is still no standard for liquid cooling and will probably take a few years for one to emerge (we hope so)
- NIKHEF for example is busy to setup an experimental environment for different liquid cooling solutions, that will fit in existing air cooled data centers

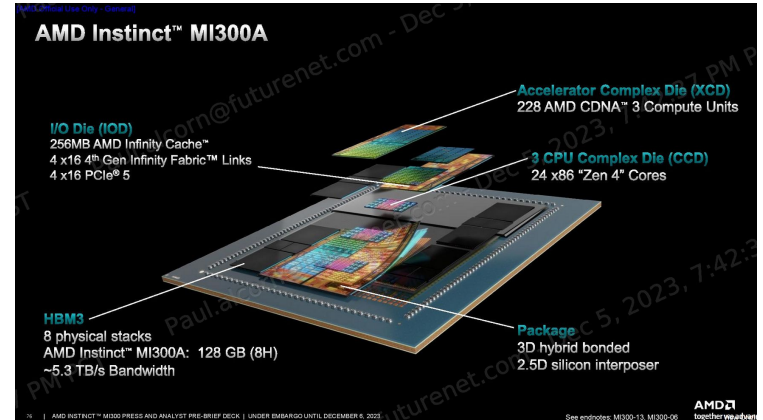# Processing units, memory, buses and interconnects

(Shigeki, Andrea C.)

Trends in CPUs

- Up to 128 cores today, 200+ announced
- TDP to reach 1 kW/socket in two generations
- Segmentation in CPU product lines
  - HPC (higher frequency)
    - Intel Granite Rapids (120 p-cores)
    - AMD Genoa (Zen 4, 96c)
  - Cloud (more cores)
    - Intel Sierra Forest (288 e-cores)
    - AMD Bergamo (Zen 4c, 128c)

- ARM resurgence
  - Ampere Altra (128c) / Ampere One (192c)
  - NVidia Grace CPU Superchip (144c)
  - Amazon Graviton 4
  - Microsoft Cobalt 100
- Single package GPU/CPU systems ("APUs")
  - AMD Mi300A, Nvidia Grace Hopper
- In-package memory
  - AMD 3D V-Cache (SRAM on die)
  - Intel Xeon Max (HBM2e DRAM)
  - Nvidia Grace (LPDDR5X DRAM)

Trends in Accelerators

- GPU accelerators and related technology
  - Nvidia H200 released in 2023 and Blackwell in 2024
  - AMD MI300X released in 2023
  - Intel Data Center Max in 2023
    - Major announcements expected in Q3
  - Broadcom announces support for AMD Infinity Fabric in its next-gen PCIe switch chips
    - To connect more AMD GPUs, similar to NVSwitch for Nvidia GPUs
- AI Accelerators
  - Captive processors from Amazon (Trainium2), Google (TPU v5) and Microsoft (Maia 100)
  - Intel Gaudi2, SambaNova SN40L, and others

GPU Codename	Ampere	Hopper	Blackwell	Rubin?	X
GPU SKU	A100	H100/H200	B100	R100	X100
Memory	HBM2e	HBM2e/HBM3/ HBM3e	HBM3e	HBM4?	HBM4e?
Launch	2020-2022	2022-2024	2024	2025	202X

AMD Instinct™ MI300A

Accelerator Complex Die (XCD)
228 AMD CDNA™ 3 Compute Units

I/O Die (IOD)
256MB AMD Infinity Cache™
4 x16 4th Gen Infinity Fabric™ Links
4 x16 PCIe® 5

3 CPU Complex Die (CCD)
24 x86 "Zen 4" Cores

HBM3
8 physical stacks
AMD Instinct™ MI300A: 128 GB (8H)
~5.3 TB/s Bandwidth

Package
3D hybrid bonded
2.5D silicon interposer

74 | AMD INSTINCT™ MI300 PRESS AND ANALYST PRE-BRIEF DECK | UNDER EMBARGO UNTIL DECEMBER 6, 2023   See endnotes: MI300-13, MI300-06

Previews in Logic and Memory

- All major logic foundries (TSMC/Samsung/Intel) now have EUV in production
  - Intel last with "Intel 4"
  - Next transition is High NA EUV or multi pattern EUV
- Advanced packaging increasingly important
  - Chiplets/die connectivity
  - Power delivery and cooling
- DRAM memory
  - CPU's transitioning from DDR4 to DDR5 memory (up to 8400?)
    - DDR6 in 2026
  - HBM3 introduced in 2022. Higher bandwidth "HBM3e" products introduced in 2023
  - All major manufacturers except Micron have transitioned to EUV
  - DRAM market recovering from collapse in late 2022/early 2023

DRAM Market Deflates

Source: WSTS, IC Insights
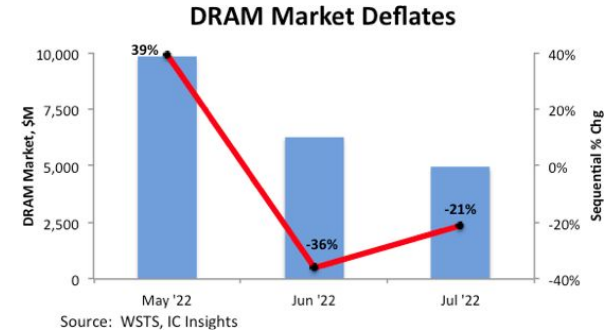
Dram Market Deflates Cyclical Downturn Looms (icinsights.com)

For further discussion

- Die size in CPUs/GPUs keeps growing leading to increased use of chiplet technology with ultrafast interconnects
- Manufacturers creating more specialized devices for domain specific applications, especially w/ AI specific designs
- Increasing proximity of GPU to CPU w/ high speed unified memory architectures

Storage
(Shigeki, Martin)

Preview of Updates in HDD Storage

- HAMR drives have finally arrived
  - Seagate Mosaic 3+ 30TB HAMR drives shipping in Q1 2024, will eventually be adopted by a wide range of products
  - Longer timescales for other players?
    - Latest WD CMR drive is 22TB
- SMR drives give a 20% capacity increase over PMR
  - ~50% of exabytes shipped by Western Digital are SMR drives
  - Not without issues when tested with EOS!
- Majority of exabytes shipped and revenue are nearline HDD
    - Market was down for 2023
    - But expected to increase by 22% (Gartner)
- Still, cost/GB gap with SSD destined to decrease in the long term
  - Nearline drives will be the last HDD holdout
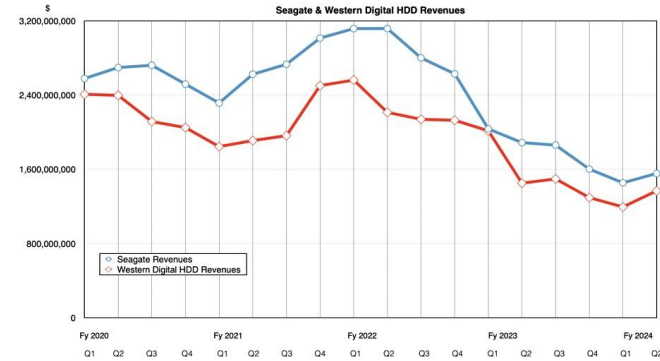
Western Digital roller coaster continues as Seagate brings down the HAMR – Blocks and Files
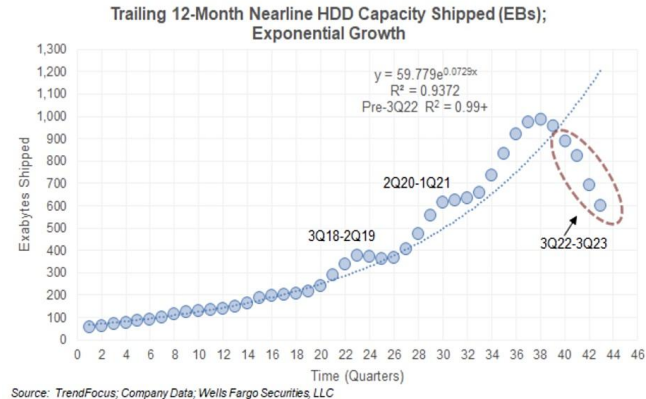
Nearline drives will be last HDD holdout by 2028 – Blocks and Files

25

Preview of Updates in Flash Storage

- SSD account for ~12% of enterprise storage capacity
- Samsung and SK Hynix dominate
  - Western Digital spinning off flash business
- PCIe Gen 5 SSDs now available
- ~ 200+ Layer 3D NAND flash chips from all five major vendors
  - ~1000 layers by 2030?
- Viability of penta level cells unclear
  - Exponentially more challenging to add more bits per cell
- Total revenues recovering from dip in late 2022/early 2023

3D Layer Cake

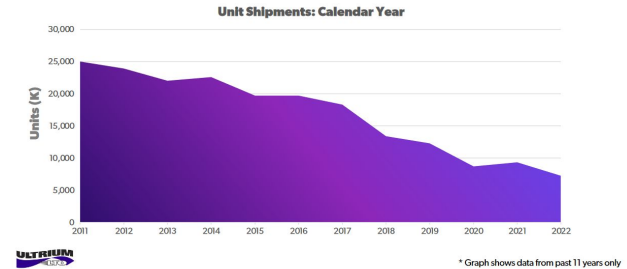| Micron		Samsung		SK hynix		SK hynix Solidigm		Western Digital/Kioxia		YMTC	
Generation	Layers	Generation	Layers	Generation	Layers	Generation	Layers	Generation	Layers	Generation	Layers
Gen 1	32	V3	48	V3	48	Gen 1	32	BiCS 2	48	Gen 1	32
Gen 2	64	V4	64	V4	72	Gen 2	64	BiCS 3	64	Xtacking 1 Gen 2	64
Gen 3	96	V5	96	V5	96	Gen 3	96	BiCS 4	96		
Gen 4	128	V6	128	V6	128	Gen 4	144	BiCS 5	112	Gen 3 Xtacking 2	128
Gen 5	176	V7	176	V7	176	Gen 5 (QLC 1H '23 & PLC?)	192	BiCS 6 (Q1 2023)	162	Gen ? 2022 2H	196
Gen 6 (End 2022)	232	V8 (2022)	236	V7(2022 Q3)	238	Gen 6?	238	BiCS 7 - will skip	212	Gen 4 2022 Xtacking 3.0	232
Gen 7	3xx	V9 (2024)	3xx	V8 (2023/4)	300			BiCS 8 (2023)	>212	Gen 5	5xx?
Gen 8	4xx	V10	4xx	V9 (2025/5)	500+			BiCS 9	300+	Gen 6	1,000?
Gen 9	5xx	V11	5xx	V10 (2030)	800+			BiCS 10	400+		
		V? (2030)	1,000								

SK hynix breezes past 300-layer 3D NAND mark – Blocks and Files

Preview of Updates in Archive Storage

- **Magnetic Tape**
  - Still a lot of room for scaling (unlike HDD)
  - Strategy change at IBM for enterprise drives
    - TS1170 - 50TB / cartridge. No backward compatibility
  - IBM Diamondback "library in a rack" targets cloud hyperscale and traditional enterprises
  - Total LTO cartridges shipped has been declining, but total exabytes shipped is flat

- **Optical disk dead**
  - Panasonic and Sony discontinued Archival Disc drives and libraries

- **On the horizon**
  - Cerabyte "ceramic nano-memory" - Data etched in material via laser or particle beam
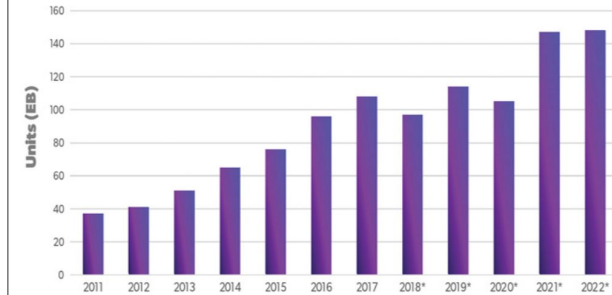 - Folio Photonics - No news since 2022

LTO MEDIA UNIT SHIPMENTS*

Unit Shipments: Calendar Year

https://www.lto.org/wp-content/uploads/2023/04/LTO-Ultirum-2022-Media-Shipment-Report-Slides.pdf

TOTAL CAPACITY BY CY** (EB COMPRESSED)

https://www.lto.org/wp-content/uploads/2023/04/LTO-Ultirum-2022-Media-Shipment-Report-Slides.pdf

Network

(Shigeki, Harvey)

Preview of Network Technology

- Transition to 400GbE (4x100Gbs) in progress
- 800GbE (8x100Gbs) specification released in 2020
- Cloud adoption of higher bandwidth Ethernet outpaces rate in the enterprise
- Ultra Ethernet Consortium formed to make Ethernet more competitive with Infiniband for AI workloads
- Co-packaged optics in the works
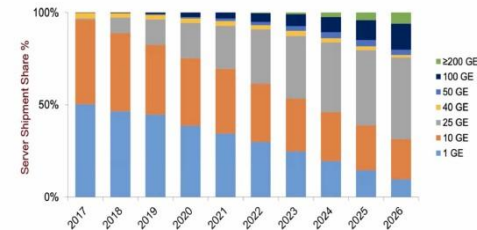 - Reduce power consumption

ETHERNET SPEEDS
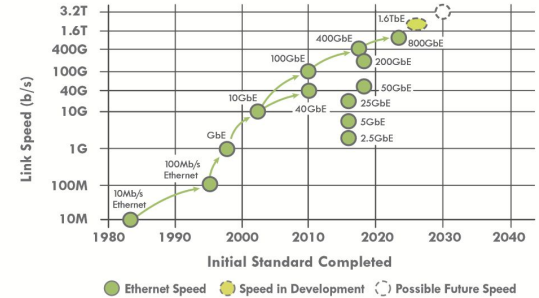
2023 Ethernet Roadmap - Ethernet Alliance
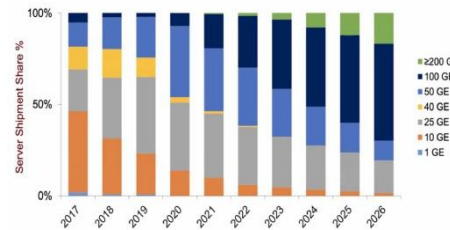
Cloud Link-Speed Forecast
Cloud: All about 100G+

Enterprise Link-Speed Forecast
Enterprise: Mix of 10G, 25G, 100G

Paving The Way For 800 Gb/sec Ethernet In The Enterprise (nextplatform.com)

29

Trends on WAN connectivity

- The LHC community is building on several R&D projects and the move to fully programmable ecosystems of networks and systems (SONiC P4, PolKA SRv6) and operations platforms (OSG, NRP, …)
  - Coordination by the GNA-G, WLCG, the worldwide R&E network community
- LHC network traffic exponentially increasing, will need Tb/s links on major routes by 2029
  - Aggregate network traffic from ATLAS + CMS will be O(10 Tb/s)
- R&D effort focusing on
  - Better estimates of required scale
  - Better models and well defined metrics for success
  - ML for system optimization
  - Better automation (monitoring, intelligence, network OSes and tools, controllability)

Some Technology Open Questions

- What is the future $/TB for HDD with the arrival of HAMR and will multi-actuator drives change the trajectory ?
- How healthy is the tape market and will the public clouds use of tape make things worse or better ?
- Are APUs, single package GPU + CPU, a better fit for HEP/NP compared to discrete GPU + CPU?
- Are CXL memory modules going to be interesting for us, e.g. when having hundreds of CPU cores in a servers?
- How viable are both fully custom and Neoverse derived ARM CPUs in the open market given the Cloud's use of internally sourced CPUs ?
- Is there a place for AI/ML processors in the open market?

HEPiX

Goal of the TechWatch WG is to answer these and other questions about technology that are of interest to the community

Conclusions

- The Techwatch working group is reborn
- It will provide the HEP/NP community with the information it needs to orient itself in the ever changing world of data center technology
- The ground to cover is vast!
- It can be successful only if a critical mass of motivated contributors is reached and maintained
- Please consider volunteering to the effort!