



AGENDA

1. A brief introduction to BIOS IT

2. Technology Update: Covering

- 4TH & 5TH Gen AMD EPYC Processors
- H13 Turin drop-in support Roadmap
- 5th Gen Intel Processors
- X14 Supermicro Roadmap
- Ampere Processors & Solutions
- IBM Power10
- Storage Technologies
- Liquid Cooling

Presenters :

Glenn Rosenberg : glenn.Rosenberg@bios-it.co.uk

Shanal Tiwari : Shanal.Tiwari@bios-it.co.uk



BIOS IT is a design house, systems builder and integrated solution provider for high performance computing, storage & Networking.

Our engineering resources and BIOS IT Labs allow us to select the best components and newest technologies based on specific customer goals and budgets.

Access to our Labs, both remotely and on-site at our facilities is available.



Case study: UK company BIOS-IT outfitting new CERN data centre

Policy paper

UK strategy for engagement with CERN: unlocking the full potential of UK membership of CERN

Published 4 October 2023

By The Numbers

ISO 9001:2015

Quality Management

ISO 14001:2015

Environmental Management System

32

YEARS OF
INNOVATION IN
HPC AND FINTECH

\$120m

OVER \$120M
ANNUAL
TURNOVER

21

GLOBAL SALES &
INTEGRATION
FACILITIES

253

EMPLOYEES
GLOBALLY

20k

CUSTOM BUILT
ENTERPRISE GRADE
SYSTEMS 2022



4TH & 5TH GENERATION AMD EPYC CPUs

Spec	Socket SP5 EPYC Turin Classic CPU	Socket SP5 EPYC Turin Dense CPU	Socket SP5 EPYC Genoa CPU	Socket SP5 EPYC Bergamo CPU
CPU TDP	200W ~ 500W	320W ~ 500W	200W ~ 400W	320W ~ 400W
Socket	Socket SP5 (LGA 6096)			
Scalability	1S, 2S			
Cores	16C up to 128C / 256 threads per socket	96C up to 192C / 384 threads per socket	16C up to 96C / 192 threads per socket	112C up to 128C / 256 threads per socket
Memory	12 channels DDR5 per socket RDIMM, 3DS RDIMM, MRDIMM		12 channels DDR5 per socket RDIMM, 3DS RDIMM	
	1DPC = up to 6000, 2DPC = up to 4400 MT/s (Single Rank) Up to 9TB (256GB 3DS RDIMM)		1DPC = up to 4800, 2DPC = up to 4000 MT/s Up to 6TB (256GB 3DS RDIMM)	
Cache	L3: 32MB per core Up to 512MB (12 CCD)	L2: 1MB per core L3: 32MB per 16 cores/CCD Up to 384MB (12 CCD)	L2: 1MB per core L3: 32MB per 8 cores/CCD Up to 384MB (12 CCD)	L2: 1MB per core L3: 16MB per 8 cores/CCX (2 CCX/CCD) Up to 256MB (8 CCD)
PCIe	PCIe 5.0 (32 GT/s)			
	2S: 128 lanes (16 w/ 3-link XGMI) PCIe5 + 12 lanes PCIe3 (6 per socket) CXL: 64 lanes support CXL 2.0 with bifurcations supported down to x4		1S: 128 lanes PCIe5 + up to 8 lanes PCIe3 2S: 128 lanes (160 w/ 3-link XGMI) PCIe5 + 12 lanes PCIe3 (6 per socket) CXL: 64 lanes support CXL 1.1+ with bifurcations supported down to x4	
SATA	1S: up to 32 SATAs 2S: up to 32 SATAs (16 per socket)			
Process	5nm+ (Zen5)	3nm (Zen5c)	5nm+ (Zen4)	5nm+ (Zen4c)

Expected ~22% improvement in power efficiency over zen4

A+ H13 System Turin drop-in support

RACKMOUNT

Multi-Node

ATX

H13DSH

H13SSH

H13SSW

H13SSF

H13SST-G/GC

H13SSL-N/NT



Hyper 2P

Hyper 1P

CloudDC

All Flash

GrandTwin

Mainstream



ATX Barebone

AS -1125HS-TNR *
AS -2025HS-TNR *
AS -2125HS-TNR

AS -1115HS-TNR
AS -2015HS-TNR *
AS -2115HS-TNR

AS -1015CS-TNR
AS -1115CS-TNR
AS -2015CS-TNR *

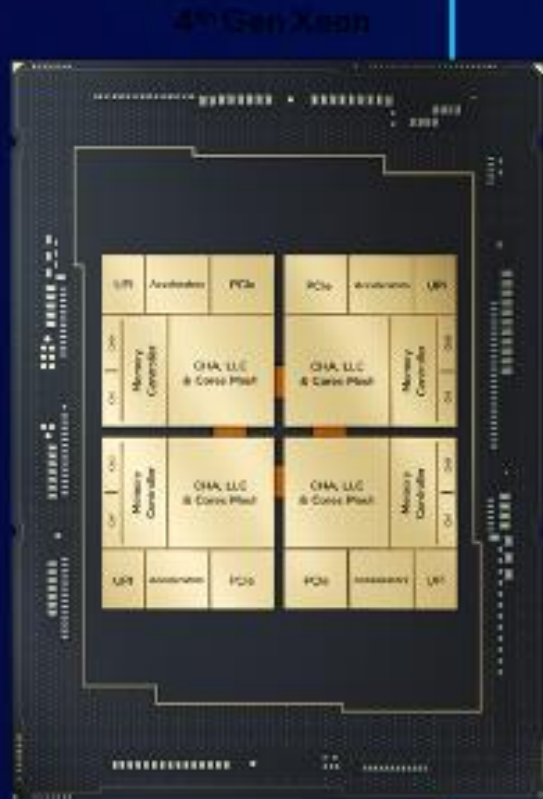
ASG-1115S-NE316R/X12R
ASG-2115S-NE332R

AS -2115GT-HNTR
(Rear IO)
AS -2115GT-HNTF
(Front IO)

AMD EPYC TURIN series drop-in support for H13 SP5 MB requires BIOS revision 2.x

5th Gen Intel® Xeon® Processors

Platform enhancements delivering significant gains at the same power envelope¹



Drop-in compatible with 4th Gen Xeon processors

Up to **64** cores per CPU

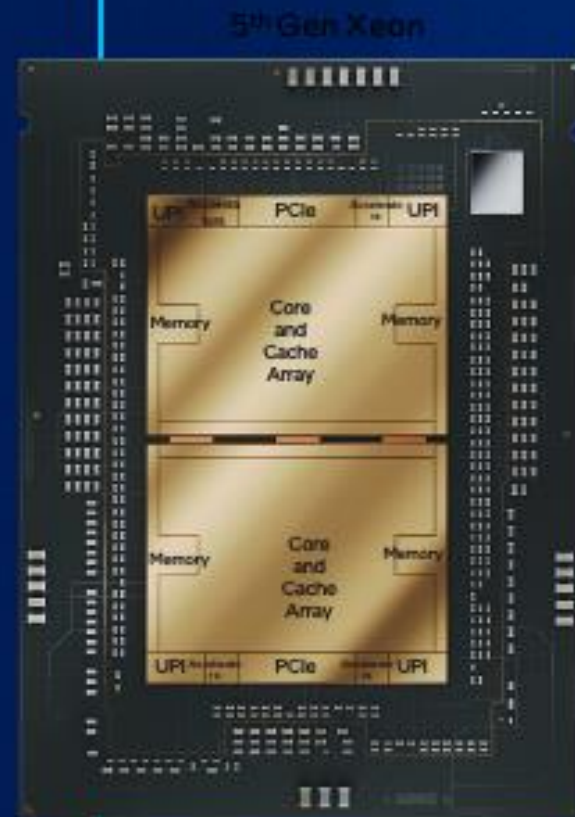
Up to **5600** MT/s memory speed

Up to **3x** shared Last Level Cache (up to **320 MB** total)

Up to **20** GT/s UPI 2.0 Speed

Type 3 memory support with Compute Express® Link 1.1*









Intel® Trust Domain Extension **broad support**



X14 Supermicro Roadmap



Confidential

	Q1 2024			Q2 2024			Q3 2024			Q4 2024			Q1 2025			Q2 2025				
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun		
Rackmounts				Sampling			 Hyper: 1U/2U Rackmount Enterprise. Consolidating compute, storage, expansion, network flexibility in one server line. (DP, 32DIMM, AIOM) (UP, 16DIMM, AIOM) – MP: May													
				Sampling			 DC-MHS: 1U/2U Rackmount Cloud. Cost effective DP solution for high volume cloud deployments. (DP, 32DIMM, Dual AIOM, DC-SCM) (UP, 16DIMM, AIOM) – MP: May													
				Sampling			 WIO: 1U/2U Rackmount. Mainstream/General Purpose with low TCO (UP, 8DIMM, SATA native). - MP: May													
Twin Series				Sampling			 BigTwin 2U Multi-node. Highest 2U node density with up to 4 nodes per enclosure. (DP, 16DIMM, AIOM) - MP: May													
				Sampling			 GrandTwin: 2U Flexible front or rear IO UP servers in 4 node or 2 node design stack up. (UP, 16DIMM, AIOM) – MP: May													
SuperBlade				Sampling			 SuperBlade: 8U 20-node, Highest density for Cloud, HPC, Enterprise (DP, 16DIMM, AIOM, E1.S) – MP: May													
				Sampling			 SuperBlade: 6U 10-node, Maximum Memory with GPU (DP 32DIMM/UP 16DIMM, PCIe, E3.S) – MP: May													
Storage				Sampling			 Petascale (DC-MHS): 1U/2U High Density All-Flash NVMe performance storage (DP, 32DIMM, EDSFF) MP: May													
			Intel Xeon 6 SRF-SP (QS)			Intel Xeon 6 SRF-SP (MP)														

AMPERE Processor Comparison

Specification	Ampere Altra	Ampere Altra Max	Ampere One
Architecture	ARM	ARM	ARM
Socket Compatible Packages	4926-Pin FCLGA	4926-Pin FCLGA	5964-Pin FCLGA
Cores	Up to 80	Up to 128	Up to 192
Max Frequency	3.3 GHz	3.0 GHz	3.3 GHz
Memory Support	DDR4-3200	DDR4-3200	DDR5-4800
Memory Capacity	Up to 4 TB	Up to 4 TB	Up to 8 TB
Memory Channels	8	8	8
TDP	Up to 250W	Up to 250W	Up to 300W
PCIe Gen	Gen4	Gen4	Gen5
Process Technology	7nm	7nm	5nm

AMPONE and AMPERE Ultra Max solution



8 x 3.5"/2.5" SATA + 4 x SATA/ NVMe Hybrid

2 x PCIe Gen5 x16 expansion slots
3 / 4 PCIe Gen5 x 8 expansion slots
2 x OCP 3.0 Gen5 x16 mezzanine slots

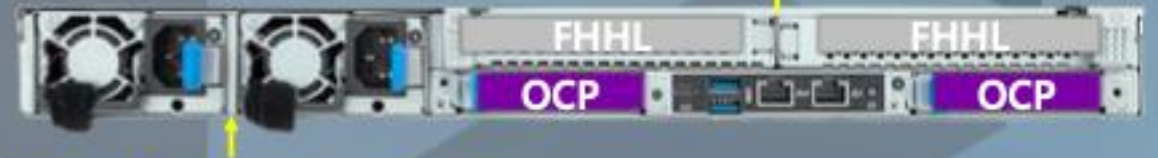


2 x 1300W 80+ Platinum Power Supply



4 x 3.5"/ 2.5" NVMe/SATA Hot-swap Bays

2 x PCIe Gen5 x16 expansion slots
2 x OCP 3.0 Gen5 x16 mezzanine slots



2 x 800W 80+ Platinum Power Supply

Processor Comparison - Power10 vs Xeon SP Platinum – AMD EPYC Genoa

Features	AMD EPYC 4 th Generation	Sapphire Rapids (4th Generation Intel® Xeon® Scalable Processors)	Power 10 DCM	Power 10 SCM
Processor	AMD EPYC 9654	Intel® Xeon® Platinum 8490H Processor	IBM Power 10 (Based on E1050 24C Processor)	IBM Power 10 (Based on E1080 15C Processor)
Lithography	5 nm	Intel 7 (aka 10nmESF)	7 nm	7 nm
Available Max Cores per Processor	96	60	24	15
Max Threads per Core	2	2	8	8
L3 Cache Per Core	4 MB	1.875 MB	8 MB	8 MB
Max Core Configuration	192	480 960 (16S HPE 3200 server)	96	240
Maximum Memory Speed	Up to DDR5 4800 MHz	Up to DDR5 4800 MHz Up to DDR5 4400 MHz(2 DPC)	2933 MHz	2933 MHz
Memory Channels Per Socket	12	8	16	16
Max Memory Configuration	12 TB	Up to 32TB (16S)	16TB	64 TB
Theoretical Memory Bandwidth/Processor	460 GB/s	307 GB/s (282 w/2 DPC)	409 GB/s	409 GB/s
IO Technology	PCIe 5.0	PCIe 5.0	PCIe4.0(PCIe 5.0 Supported)	PCIe4.0(PCIe 5.0 Supported)
Scalability	2S	Certified till 2S-8S* *HPE can go to 16S	2S-4S	16S

Power10 : no compromise between Scalability, Memory Size & Bandwidth , Core Performance

Why Power is different infrastructure ?

A complete design, all components are integrated single solution and unique architecture for data processing

- ✓ Processor, Memory Xbar, Microcode, Virtualization, OS, ...
- ✓ Infrastructure Management
- ✓ System Software, Cloud capabilities
- ✓ RAS and Security
- ✓ I/O connectivity (*Storage, Network*)
- ✓ Unique Flexibility features (*PEP, LUK*,
- ✓ Simple operations
- ✓ For Long term investment

Infrastructure matters for:

Security, Performance, Reliability, Flexibility, Agility, sustainability
and TCO

We design technology to help you drive business

2021: IBM World's First 2 nanometers Chip Technology  YouTube



**Infrastructure is critical
and it's only 25% of IT spending**

SuperStorage LFF Storage Family

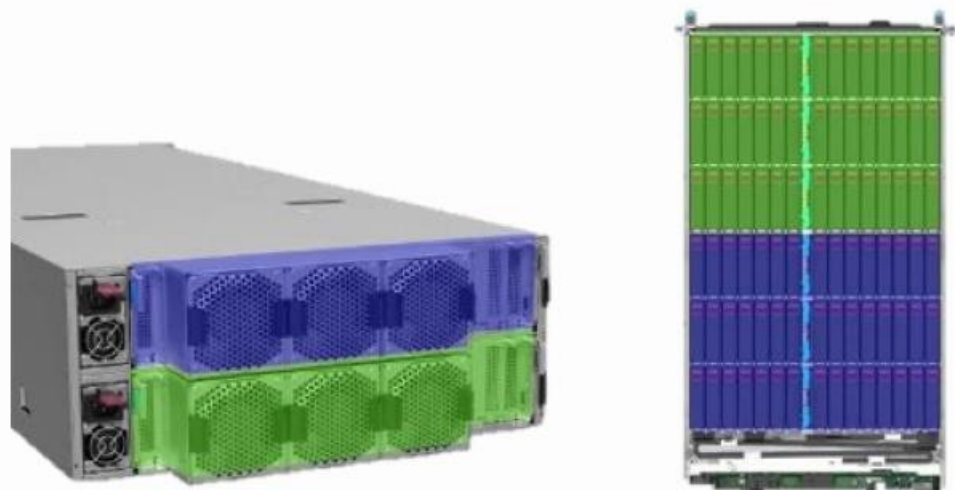
Completed LFF Storage Platform for Storage-Centric Data Center from Edge to Cloud



Intel X13 and AMD H13 Based Top Loading 90/60 Bay storage



Single Node Configuration



Dual Node Configuration

Specifications

Processor Support

- Dual Gen5 processors (Intel SPR)

Memory Capacity

- 16x DIMM slots, DDR5-4800

Expansion

- 3x PCI-E 5.0 x16

Networking & I/O

- 2x 10GBase-T LAN ports
- 1x RJ45 Dedicated IPMI LAN port and 2x USB 3.0 ports, 1x VGA ports and 1x Serial port

Drive Bays

- 60 or 90x Hot-swappable 2.5" or 3.5" drive bays (SAS3/SATA3)
- 2x rear Hot-swappable 2.5" SATA drive bays
- 2 onboard PCIe x2 M.2 slots and 2 internal slim SATA SSD slots
- 4 rear NVMe U.2 bays for cache support (optional)

Storage Controller

- Broadcom SAS 3816 AOM, 3816 AOC (IT mode)
- Broadcom SAS 3916, 3908 AOC (HW RAID mode)

System Cooling

- 6x Heavy duty counter-rotate redundant PWM cooling fans

Power Supply

- 2x 2000W or 2600W High-efficiency (Titanium level, 80%) redundant supply

System Management

- Built-in Server management tool (IPMI 2.0, KVM/media over LAN) with dedicated LAN port

Dimensions

- 60 Bay : 17.68" (W) x 6.9" (H) x 34.1" (D)
- 90 Bay : 17.68" (W) x 6.9" (H) x 42.9" (D)

Subject to change without notice

All Flash Storage server

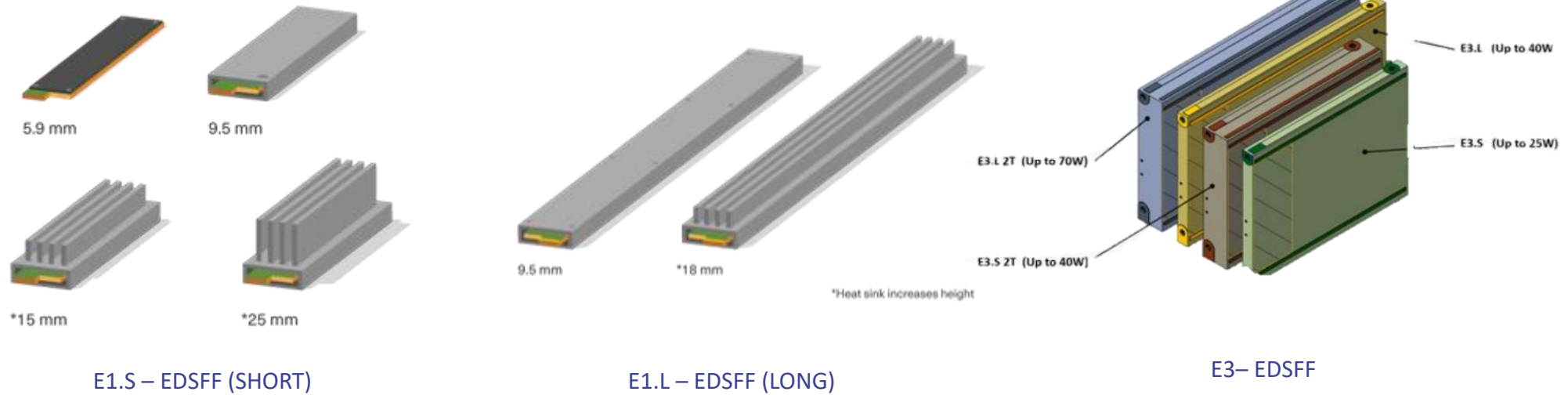
Revolutionary Petascale Nvme



Higher-Performance

- PCIe Gen 5 x4: ~15.8GB/s
 - Samsung PM1743: S/R 13GB/s, S/W 6.6GB/s
- PCIe Gen 4 x4: ~7.88GB/S
 - Samsung PM1733: S/R 7GB/s, S/W 3.5GB/s
- Higher Capacities
 - up to 15.36TB and 30TB
- Cost per GB parity vs SATA
 - 960GB, 1920GB, 3840GB, 7680GB
- EDSFF Form Factors
 - E3.S
- 1U front-loading all-flash storage server
- 24 E1.S NVMe drives and PCIe 5.0
- Peak performance of 4.44 million IOPS (random 4k reads)
- At 99.5 μ s in latency
- Writes 4.2 million IOPS (4.2 million IOPS at 64.8 μ s)

EDSFF FORM FACTORS



Why a customer would want an E1.S or an E1.L SSD

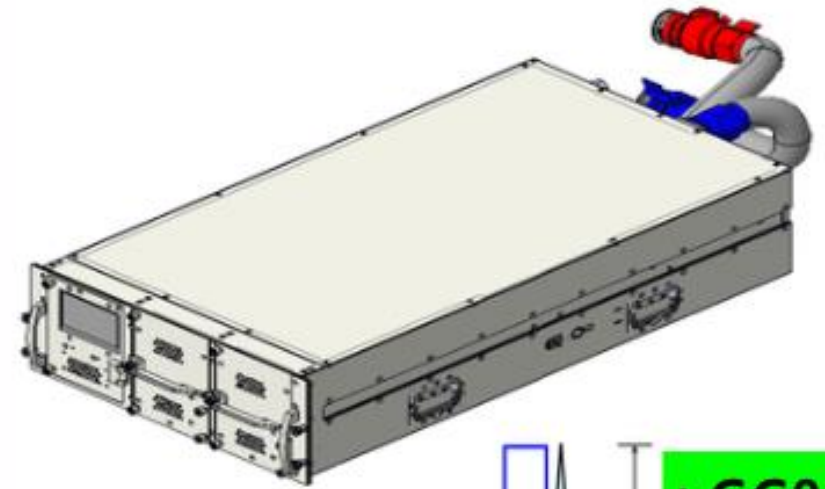
- The E1.S is space in a compact server
- The E1.L has a larger cooling surface, which will allow for higher operating temperatures.
- Tradeoff between the different widths available

Product	Width of Enclosure	Recommended sustained power	Length
E1.S	5.9mm	12 watts	111.49mm
E1.S	9.5mm	20 watts	118.75mm
E1.S	15mm	20+ watts	118.75mm
E1.S	25mm	25+ watts	118.75mm
E1.L	9.5mm	25 watts	318.75mm
E1.L	18mm	40 watts	318.75mm

Supermicro DLC Solution -- CDU

- CDU: **C**ooling **D**istribution **U**nit
 - High Performance
 - User Friendly
 - Robust and Reliable

- Replace pump in **2** mins
- Replace PSU in **1** mins
- More 66% pressure head supported
- User friendly maintains
- Pressure/flow rate/leakage monitor
- Control remotely or Touch panel
- Non-Condensation Design



+66% B
Performance



Save Money via Liquid Cooling

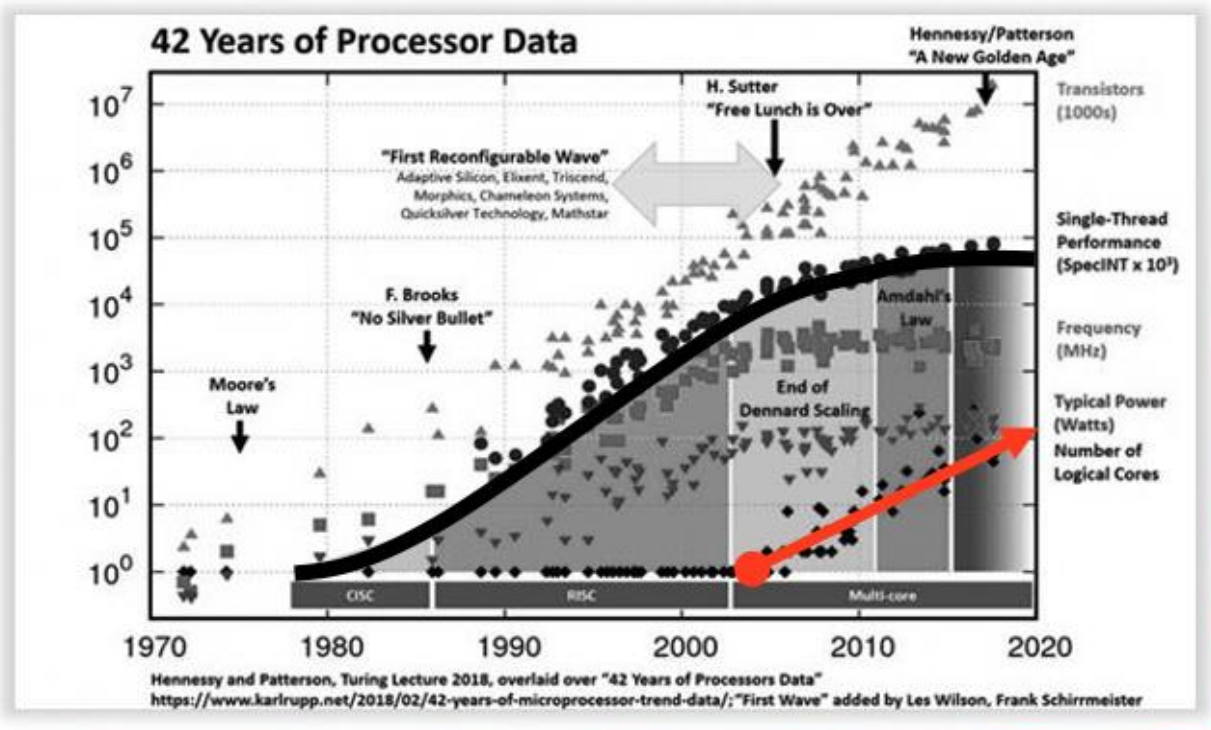
Item	Air Cooled - Traditional DC	Air Cooled - Modern DC	Cooling Tower with DLC (Cold Water)	Cooling Tower with DLC (Warm Water)
Supply-Return-WB (°C)	N/A	N/A	30-25-22	42-37-22
Process Fluid Flow (GPM)	Depends on DC design	Depends on DC design	1,865	1,865
Temperature Range (°C)	N/A	N/A	5	5
IT Total Power (MW)	2.45	2.45	2.21	2.21
Server Cooling Type	Air	Air	Liquid	Liquid
Facility Cooling Power Cons. (kW)	2450	1225	551	172
Gross PUE	2	1.5	1.25	1.08
United States Electricity Cost \$0.20 / kWh				
Cooling Electricity Cost (USD) per 5 years	21.5	10.7	4.8	1.5
IT Electricity Cost (USD) per 5 years	21.5	21.5	19.4	19.4
Total Cost (Cooling + IT)	42.9	32.2	24.2	20.9
Total Savings (Liquid vs Air) mil USD	N/A	N/A	18.7	22.1

Germany Electricity Cost \$0.52 / kWh				
Total Cost (Cooling + IT)	111.6	83.7	62.9	54.3
Total Savings (Liquid vs Air)	N/A	N/A	48.7	57.3

Japan Electricity Cost \$0.25 / kWh				
Total Cost (Cooling + IT)	53.7	40.2	30.2	26.1
Total Savings (Liquid vs Air)	N/A	N/A	23.4	27.6

UK Electricity Cost \$0.47 / kWh				
Total Cost (Cooling + IT)	100.9	75.7	56.8	49.0
Total Savings (Liquid vs Air)	N/A	N/A	44.0	51.8

Traditional Techniques No Longer Scale



Turbo Frequency

Hyperthreading

Scale Up Accelerators

Power Optimized Consistent Performance

Linear Core Scaling

High Performance General-Purpose Cores

Paradigm Shift

Get Started Now with Cloud Native Processors

Learn

Assess & Plan

Build

Test & Optimize

Deploy

- **Learn how to design with performance/rack metrics:**
<https://amperecomputing.com/home/increase-data-center-efficiency>
- **Browse Ampere Ready software references:**
<https://amperecomputing.com/solution>
- **Use our workload and solutions briefs:**
<https://amperecomputing.com/briefs>
- **Join the Ampere Developer Community:**
<https://community.amperecomputing.com>
- **Cloud Instances broadly available:**
<https://amperecomputing.com/where-to-try>