

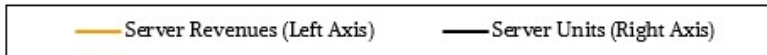
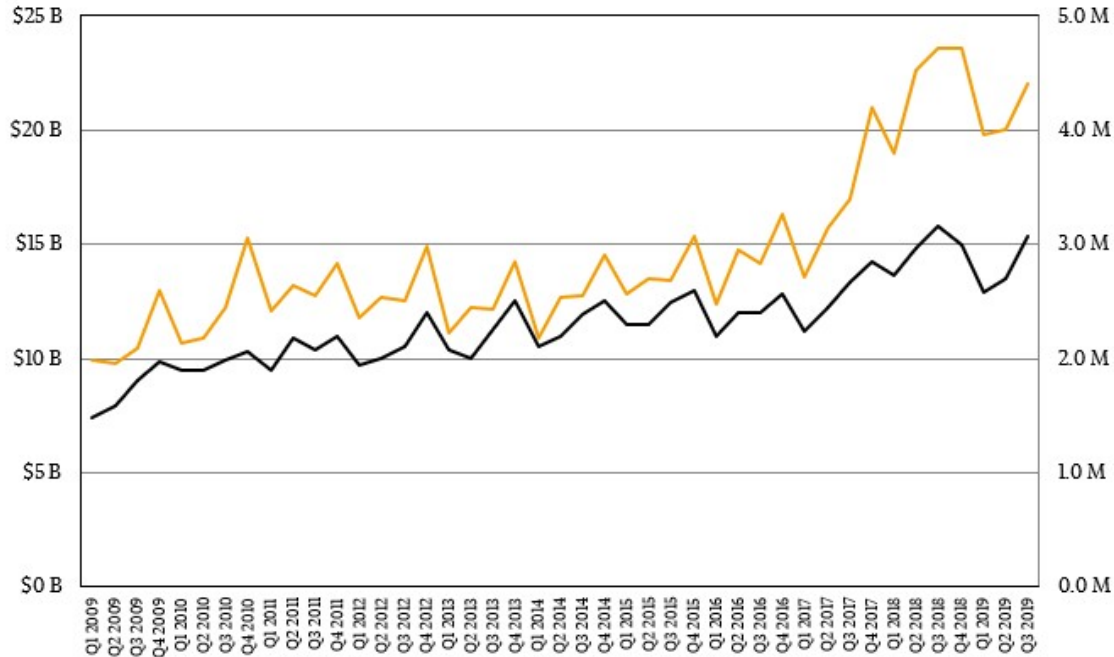
HEPiX TechWatch WG: Server

Chris Hollowell , Brookhaven National Laboratory
Michele Michelotto, INFN, Padova (IT)

On behalf of the HEPiX Technology Watch Server Market Subgroup

HEPiX Workshop
Taipei, March 2020

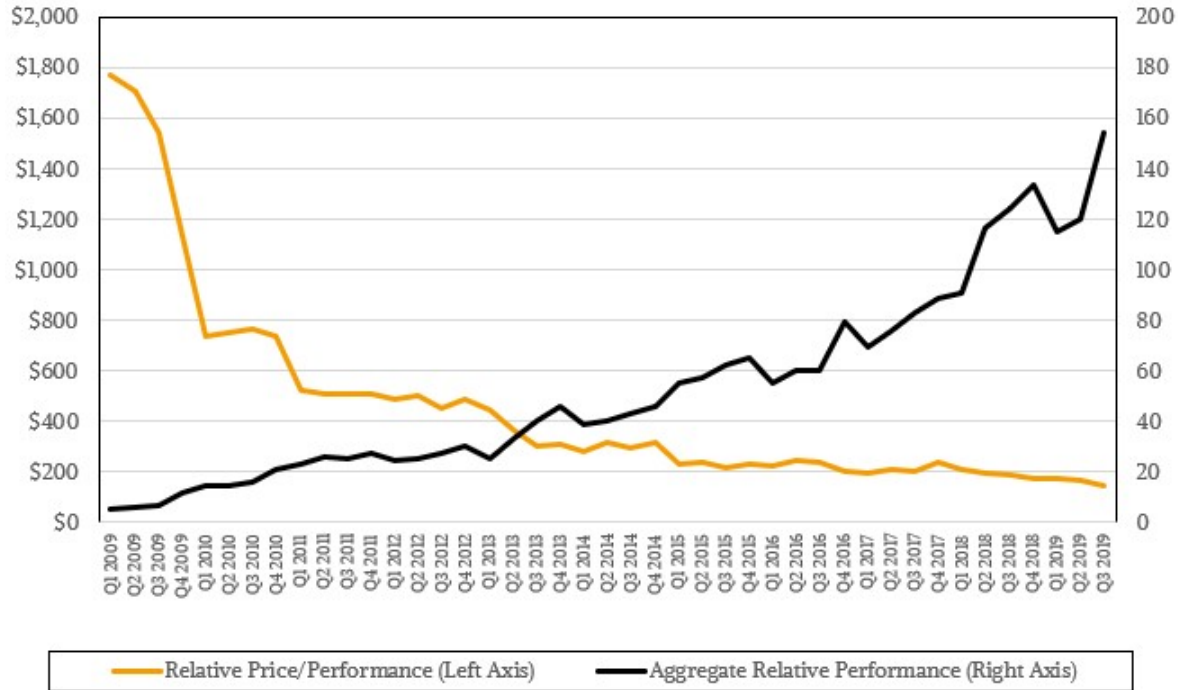
Server Market



Global server market revenue up substantially

- \$21.99 billion in Q3 2019 with 3.07 million servers shipped
- Primarily due to hyperscale cloud provider purchases
- 6.7% revenue decline and only 3% revenue declining due to increasing competition from AMD against Intel
- Memory and Flash prices down from the excessively high prices of the past

Price/Performances



Performances increased in the past due to:

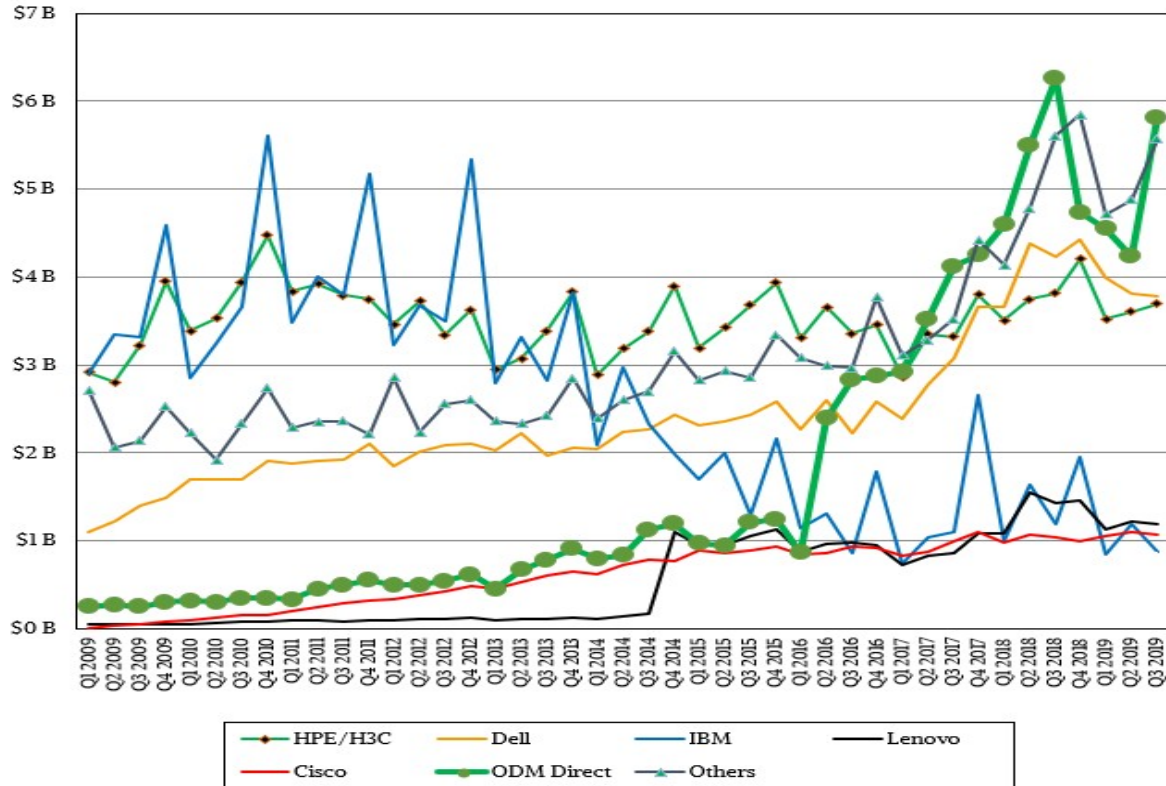
- Number of socket sold
- Average number of core per socket
- Architectural changes on each generation of server

On the left axis the relative price/performances

In the past decade the aggregate computer power has grown by 28x but the average cost per unit of CPU compute power has been decreased by 12x

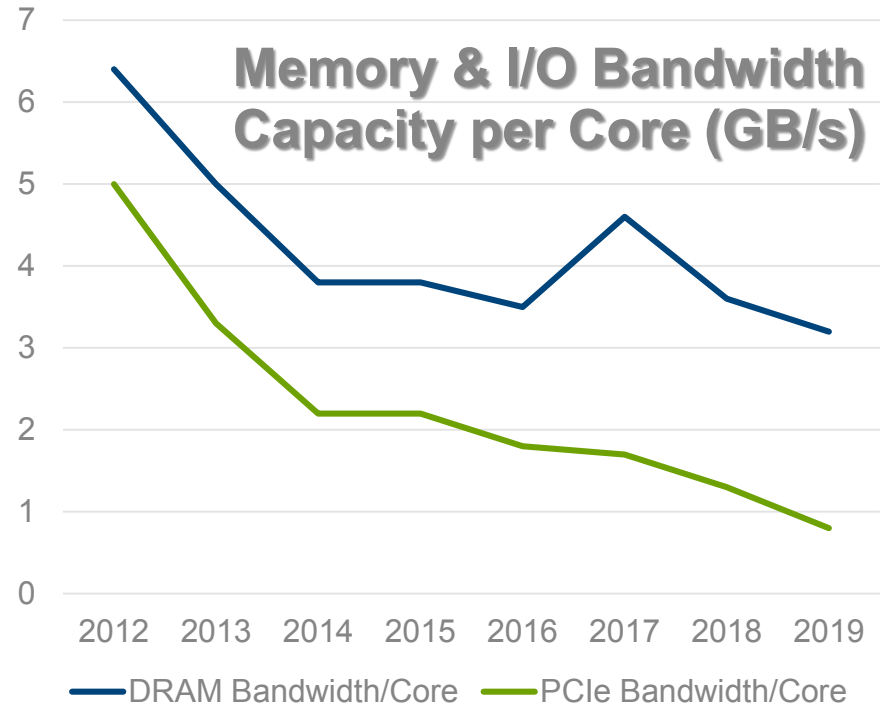
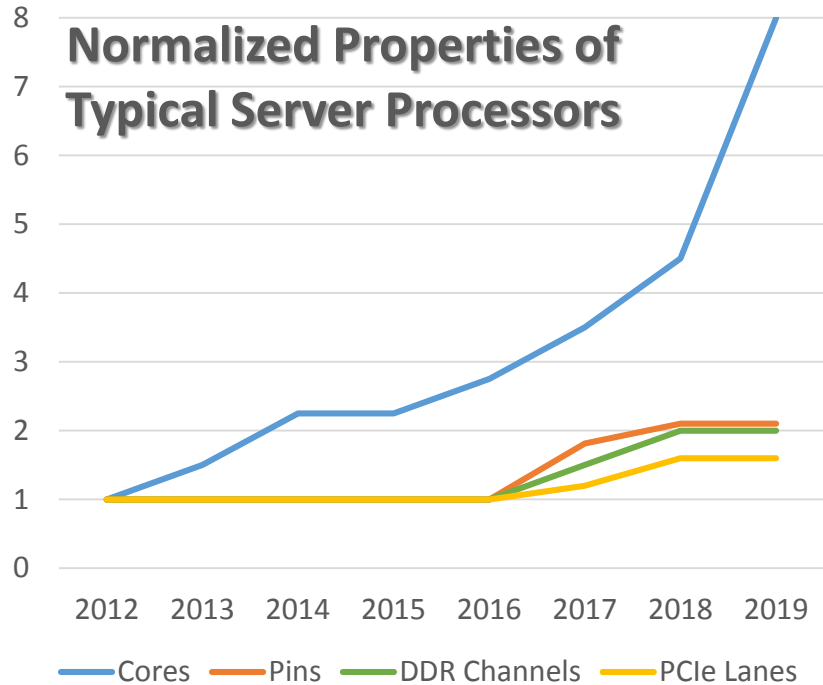
All of this at clock more or less constant

Server Market (Cont.)

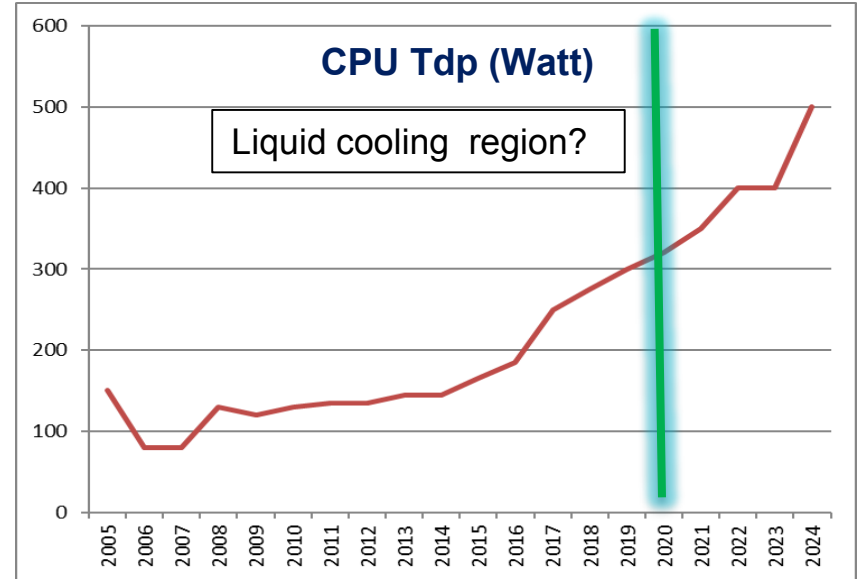
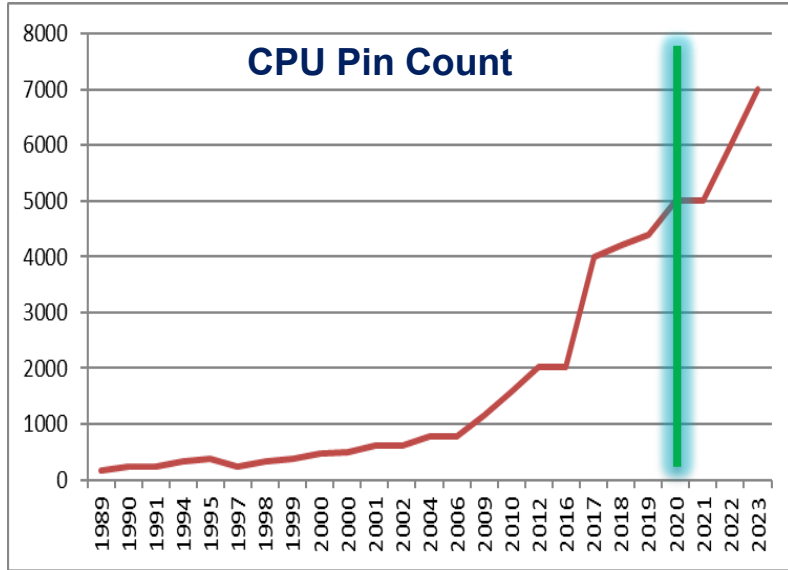


- No big differences from previous report
- 26.5% of total server revenues from all the ODM (Original Design Manufacturers)

Computer – Memory – I/O Balance



CPU-Mem-IO needs driving Power and Size





Data Center Liquid Cooling Market Increasing Demand with Key Players | Dynatron Corporation, ALFA LAVAL, Mitsubishi Electric Corporation IBM Corporation and More

BY DATA BRIDGE MARKET RESEARCH ON FEBRUARY 3, 2020

Data Center Liquid Cooling Market Growth Rate of more than 29.17% & worth 2969.8 Million USD by 2025

Mon Sep 24, 2018 - 11:00am UTC



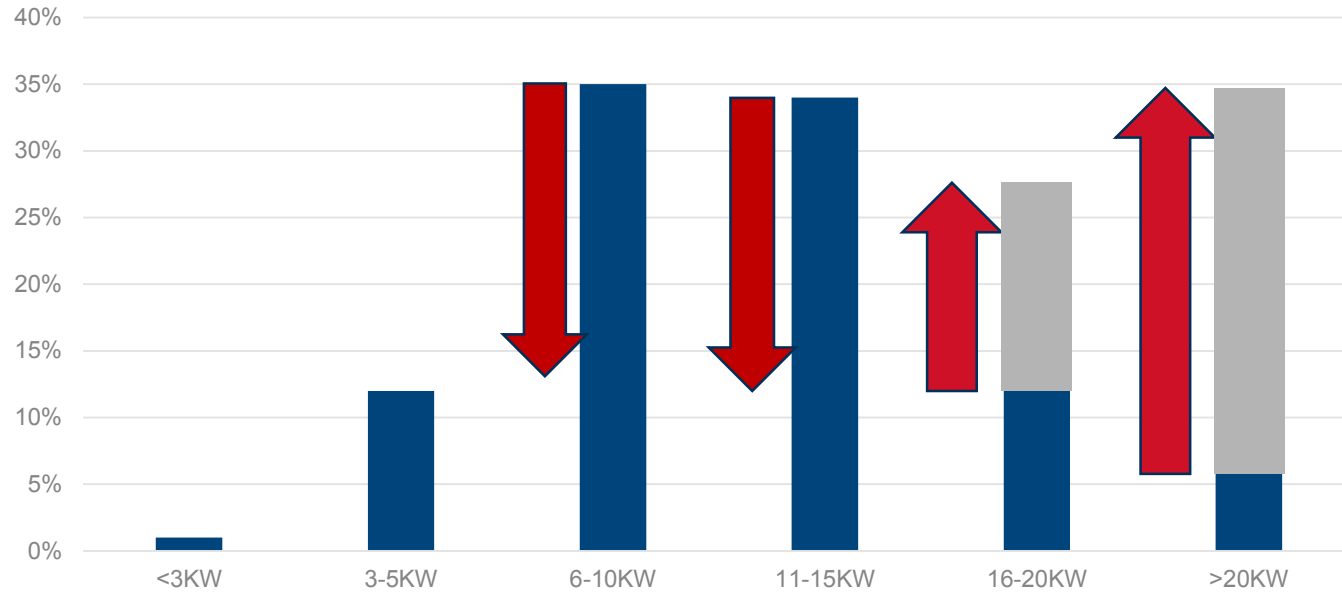
Arcognizance.com shared "Global Data Center Liquid Cooling market" size, industry status and forecast, competition landscape and growth opportunity. This research report categorizes the global Data Center Liquid Cooling market by companies, region, type and end use industry.

FEATURED NEWS

- 1 **Yi+ CEO Zhang Mo Shows NEW-TECH is Not Just for Guys at UN Banquet**
- 2 **Vietnamese logistics startup EcoTruck receives US\$1.7 million in pre-**

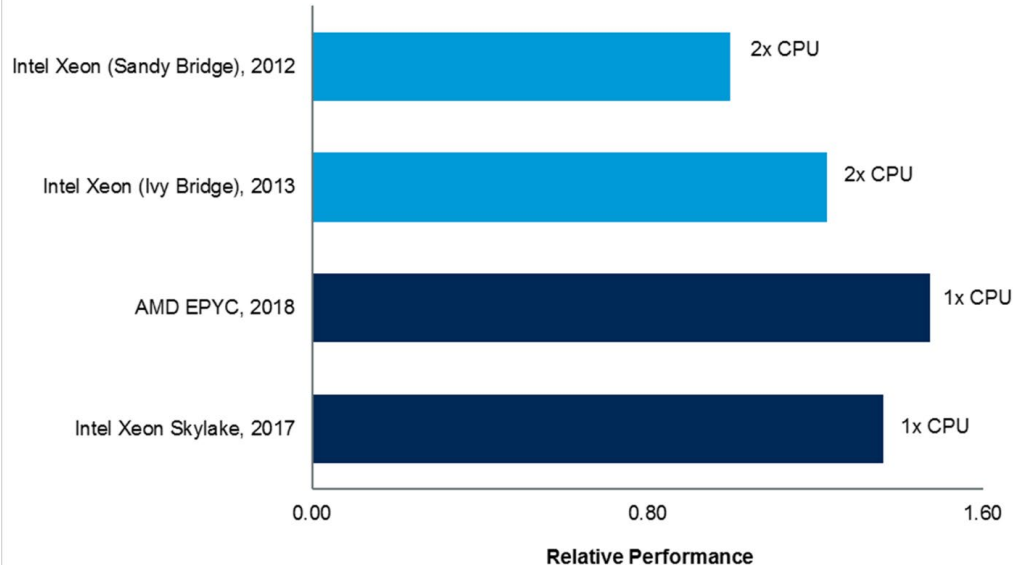
Rack Power

Rack Power Limit Trends



The rise of the single socket

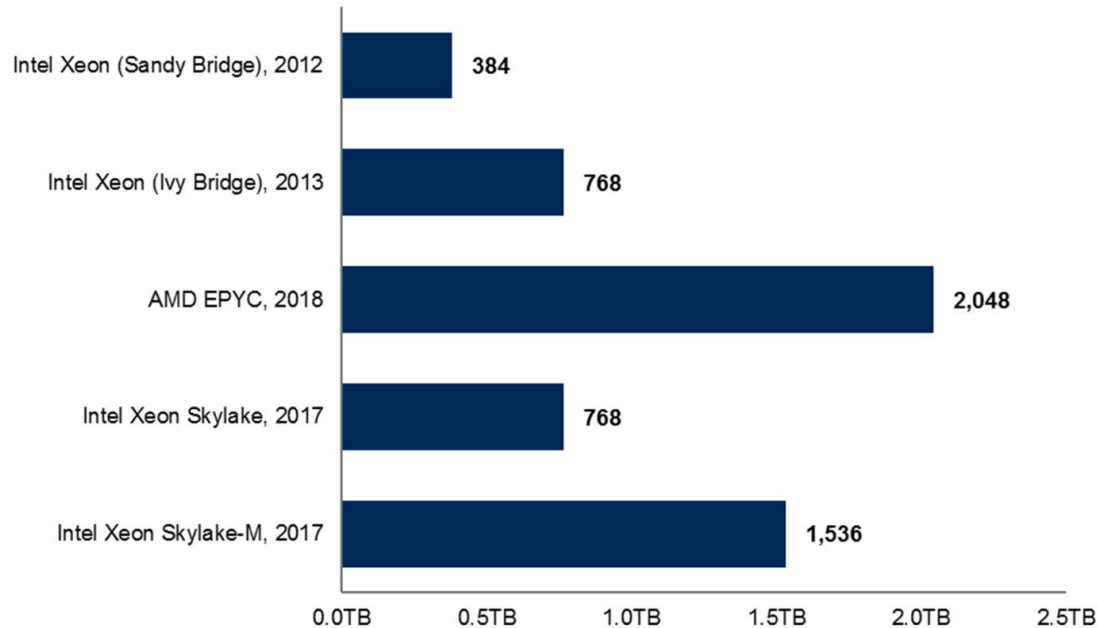
Relative Performance of 16 Cores



- At each generation of x86 processor, its additional cores and features over previous generation
- A single current-generation CPU server can compete with a previous generation dual CPU

Memory per socket

Maximum Memory per Socket

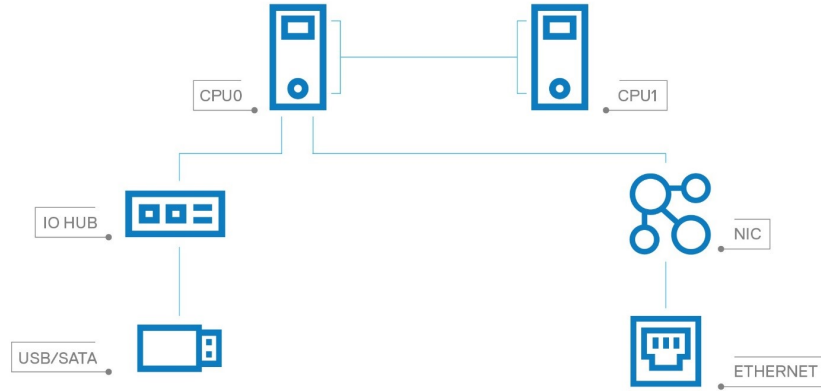


- Memory density is increasing: a single CPU server with 12 or 16 DIMM sockets can support memory densities previously supported by two sockets
- In most virtualized solutions memory capacity and bandwidth, rather than CPU performances, are the limiting factors

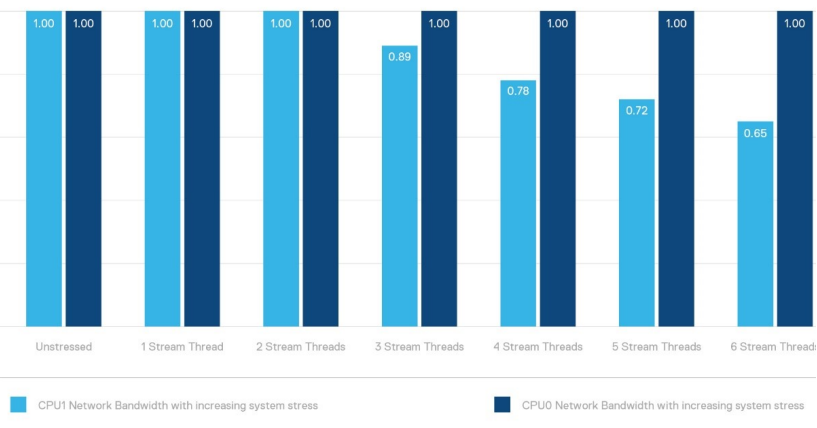
Single socket vs dual socket CPUs

	AMD EPYC (1 Socket)	Intel Xeon E Series (1 Socket Skylake)	Intel Xeon Processor E5 v2 Series (2 Sockets)	Intel Xeon Scalable (Skylake) (2 Sockets)
Launch Date	1Q17	2Q18	3Q13	3Q17
Cores	8 to 32	4 to 6	4 to 12	4 to 28
Memory Channels Per CPU	8	2	4	6
Max Memory	2TB	128GB	768GB	768GB or 1.5TB
PCIe Lanes	128	16	40	48
Cache Size	32MB to 64MB	12MB	10MB to 30MB	8.25MB to 38.5MB

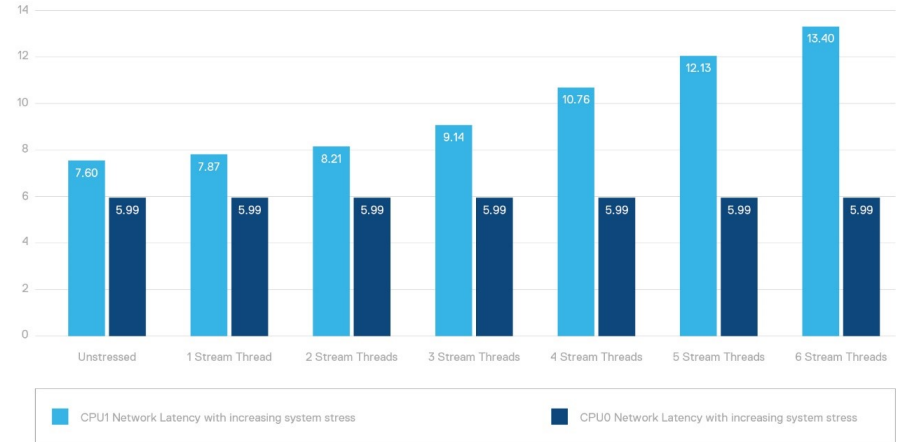
- Market has typically purchased dual-socket servers due to the historical limitations in terms of core count, addressable memory and I/O in CPUs designed for use in dedicated single-socket servers.
- This has changed with the recent introduction by AMD of its EPYC processors that are optimized for use in dedicated single-socket servers.
- A server with only one CPU socket does not incur the extra hardware cost (components, heat sinks, power supply capacity, fans, etc.) associated with the second socket, further reducing the cost of the hardware



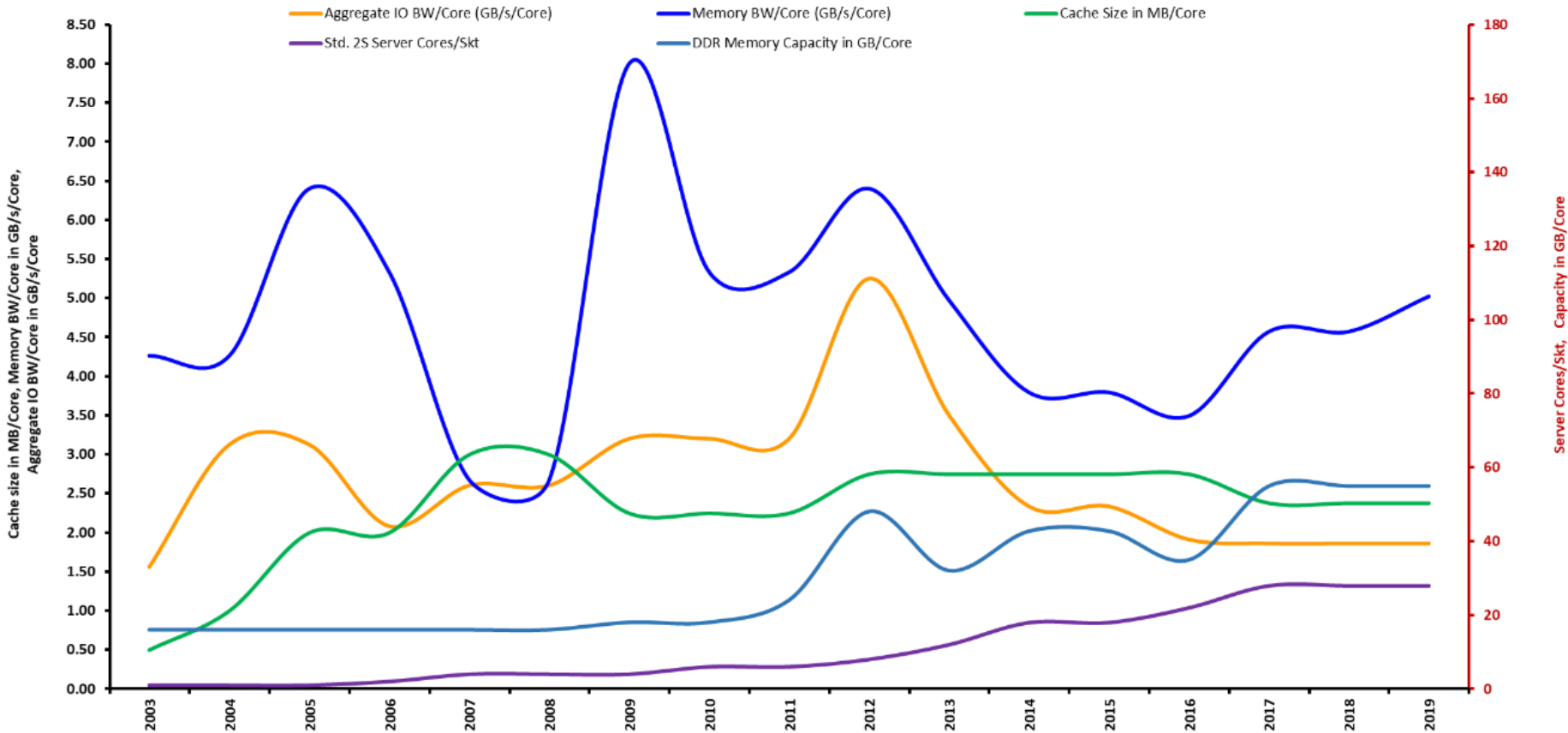
2S Server: CPU0 vs CPU1 Network Bandwidth to IO on CPU0 during stress



2S Server: CPU0 vs CPU1 Network Latency to IO on CPU0 during stress



Cores, Caches, Memory, & IO (Std. Servers)



From 2011 to 2020 per core metrics as memory bandwidth and IO bandwidth (**per core**) are falling since 2011. To fix this trend we need more pins and faster SERDES (PCIe Gen4/5, DDR5, Gen-Z) 1 socket enables us to make those pin trade-off at the system level

Why 1 socket server?

- More than enough cores per socket and trending higher
- Lower cost DRAM solutions per server (less required min DIMMs)
- Better software licensing cost for some models
- Avoid NUMA complexity and performances hits
- Permit deterministic performances
- Power density per server reduction, avoid rack power problems
- Repurpose NUMA pins for more MEM or IO channels (DDR_x, PCIe, Gen-Z)
- Enables better NVMe direct drive connect without PCIe switches

References

- <https://www.nextplatform.com/2019/12/09/datacenters-are-hungry-for-servers-again/>
- <https://www.idc.com/getdoc.jsp?containerId=prUS45482519>
- <https://rcpmag.com/blogs/scott-bekker/2018/12/server-market-hits-record-high.aspx>
- <https://technology.ihs.com/608048/the-canary-in-the-coal-mine-for-liquid-cooling-is-server-shipments-with-coprocessors>
- <https://www.scienceofchange.org/2020/02/03/data-center-liquid-cooling-market-increasing-demand-with-key-players-dynatron-corporation-alfa-laval-mitsubishi-electric-corporation-ibm-corporation-and-more/>
- <https://e.huawei.com/se/videos/global/older/201412301131/>
- <https://www.gartner.com/doc/reprints?id=1-680TWAA&ct=190212&st=sb>
- <https://www.nextplatform.com/2019/04/24/why-single-socket-servers-could-rule-the-future/>