Trends in computing technologies and markets: The HEPiX TechWatch WG

Andrea Sciaba[1,*], *Helge* Meinhard[1,**], *Germán* Cancio[1,***], *Martin* Gasthuber[2,****], *Christopher* Hollowell[3,†], *Michele* Michelotto[4,‡], *Edoardo* Martelli[1,§], *Servesh* Muralidharan[1,¶], *Bernd* Panzer-Steindel[1,‖], *Rolf* Seuster[1,**], *Peter* Wegner[2,††], *Eric* Yen[5,‡‡], *Shigeki* Misawa[3,§§], and the TechWatch WG

[1]Conseil Européen pour la Recherche Nucléaire (CERN), Geneve, Switzerland
[2]Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany
[3]Brookhaven National Laboratory (BNL), Upton, NY, U.S.A.
[4]Istituto Nazionale di Fisica Nucleare (INFN), Rome, Italy
[5]Academia Sinica, Taipei, Taiwan

Abstract. Driven by the need to carefully plan and optimise the resources for the next data taking periods of Big Science projects, such as CERN's Large Hadron Collider and others, sites started a common activity, the HEPiX Technology Watch Working Group, tasked with tracking the evolution of technologies and markets of concern to the data centres. The talk will give an overview of general and semiconductor markets, server markets, CPUs and accelerators, memories, storage and networks; it will highlight important areas of uncertainties and risks.

1 Introduction

High energy physics (HEP) and other big science projects are highly dependent on significant advances in compute, storage, and network performance to store and analyze the expected near order of magnitude increases in collected data in the next few years. Any deviation from expected improvements in performance (or cost/performance ratios) will have an impact on future experiments. However, with a few exceptions, technical and economic factors have slowed the rate of improvement in recent years relative to historical rates.

*e-mail: andrea.sciaba@cern.ch
**e-mail: helge.meinhard@cern.ch
***e-mail: german.cancio@cern.ch
****e-mail: martin.gasthuber@desy.de
†e-mail: hollowec@bnl.gov
‡e-mail: michele.michelotto@pd.infn.it
§e-mail: edoardo.martelli@cern.ch
¶e-mail: servesh.muralidharan@cern.ch
‖e-mail: bernd.panzer-steindel@cern.ch
**e-mail: seuster@cern.ch
††e-mail: peter.wegner@desy.de
‡‡e-mail: Eric.Yen@twgrid.org
§§e-mail: misawa@bnl.gov

2 Semiconductors

The scientific community is dependent on the global semiconductor market to provide the advances in electronics that drives improvements in computing, detectors, and data acquisition. In FY2018, global demand for semiconductors exceeded 1 trillion in units [3]. In the first half (1H) of 2019, sales dropped 14% relative to 1H2018 due to a drop in memory and flash sales; however, the market is expected to recover in 2020 [3, 4]. From a technology perspective, leading edge semiconductor foundries are moving to smaller (7 nm) device sizes, requiring new and costly lithographic equipment. Future migration to 5 nm and beyond is expected to require another round of costly lithography changes. This reality has caused Global Foundries to stop 7 nm process development, leaving TSMC, Samsung, and Intel at the leading edge [1]. These and higher chip design costs are making it harder to justify development of new chips using these leading edge processes as only chips with high enough sales volume can justify the up front development costs [2, 6].

3 Servers

Compute servers are critical to HEP. Global revenues for the server market was $20 billion, down 11.6% year over year in 2Q2019, with 2.7 million servers shipped [7]. Average selling price (ASP) remained stable, as DRAM, flash and CPU costs dropped [8]. In FY2019, the hyperscalers (e.g. Amazon, Facebook, Google, Microsoft) purchased 35% of all Intel's server processors and the Original Device Manufacturers (ODMs), the suppliers of choice to the hyperscalers, captured a larger share of the server market than the traditional "Tier 1" servers brands [8, 9]. At this point in time the impact of the hyperscalers on HEP procurements is unclear.

4 Central Processing Unit (CPU)

The x86 instruction set architecture (ISA), the one implemented by Intel and AMD, continues to be the dominant CPU ISA in the server market, with IBM Power, ARM, and RISC-V attempting to grow market share or enter the market [16, 25].

x86 Architecture(Intel/AMD)

In the x86 market, AMD is starting to challenge Intel's hegemony in the server CPU market. AMD's Zen2 2nd generation "ROME" EPYC server CPUs promise competitive CPU cores, higher core counts (64 vs 28 cores), larger memory bandwidth (8 vs 6 DDR4 memory channels) and greater I/O connectivity (128 PCI-e Gen 4 lanes vs 48 Gen 3 lanes) compared to Intel CPUs [11]. Intel has responded with support for Optane persistent memory and Vector Neural Network instructions (VNNI) in their latest CPUs [10]. However, Intel has been hampered by problems with their 10 nm chip manufacturing process [12]. Meanwhile, AMD has moved to chiplets, small silicon die implementing specific functions, and multi-chip modules (MCMs), integrated circuit packaging containing multiple chiplets, to mitigate the cost of 7 nm processes. Chiplets and MCMs enable reduced die size and varied CPU configuration, by mixing and matching different chiplets, at the cost of more complex fabrication and potentially higher power consumption and latency [13, 14].

Non-86

IBM Power is the only non-x86 ISA that the HEP community likely to encounter at this time. The top two supercomputers in the November 2019 Top 500 list both use IBM Power 9

processors [28]. Unique features of Power 9 include 4 threads per core (SMT4), buffered memory (for large memory system), native NVLINK, and CAPI (coherent accelerator processor interface) for CPU connectivity to accelerators like GPUs. Power 9 was also an early adopter of PCI-e Gen 4 [15].

A new entrant in the data center market are CPUs from several vendors based on the ARM ISA, an ISA from ARM Holdings [16]. Marvell's ThunderX2 is available in systems from Cray, HPE, and Gigabyte [18, 19]. Fujitsu's A64FX is the CPU in the Fugaku (Post-K) supercomputer to be installed at Riken in Japan [20]. Other ARM based CPUs, including Ampere's eMAG and Amazon's Graviton, are also targeting the data center [21, 22]. ARM Holding has also announced their Neoverse N1 intellectual property blocks that can be used by licensees to quickly build custom CPUs that are designed for the data center [23]. The ARM ISA is also a core component of the European Union's European Processor Initiative [24] that aims to build a CPU for high performance computing (HPC). Given the breadth of industry activity with ARM, it is likely that the HEP community may encounter ARM in the data center in the near future.

RISC-V is an open source ISA that is attempting to follow the same path as ARM into the data center market place by starting in embedded/system on chip applications [25]. RISC-V has been targeted for inclusion in the European Processor Initiative's accelerator roadmap [26]. The EPI is part of the European Union's effort to build an HPC ecosystem [27].

5 Accelerators

The slowdown in the advancement of CPU performance in the past few years has lead computer designers to look to domain specific accelerators (DSA) for dramatic increases in performance and performance per watt. However, DSAs only work for specific types of applications, and typically require substantial modifications to software.

Graphics Processing Units (GPUs)

NVIDIA, leveraging development efforts in PC and console gaming, has made GPGPUs (General Purpose Graphics Processing Units) the DSA of choice in computing. Their flagship GPU is the V100 [29]. Nvidia GPUs are present in a large number of Top 500 HPC systems and provide a substantial fraction of the FLOPS (floating point operations per second) in these systems [28]. Both AMD with Radeon Instinct and Intel with X^e are attempting to capture some of the GPU market [30, 31]. Software libraries and tools remain a key to the adoption of GPUs, but the most mature packages tend to be platform specific. Recently, GPUs have added features to accommodate new applications (e.g., bfloat16 support for artificial intelligence) to thwart competition from dedicated AI accelerators.

AI Accelerator

Artificial Intelligence (AI) processors are another class of DSAs gaining traction, driven by the use of AI applications by the hyperscalers and other industries. Some AI processors are only available in the cloud, like Google's Tensor Processing Unit (TPU) while others are available as hardware that can be purchased in the open market (Intel Nervana, Habana Gaudi) [32–34]. The influence of the hyperscalers is evident in their ability to finance the development of Application Specific Integrated Circuits (ASICs) like Google's TPU and also to generated interest in AI accelerators in the broader market, to compel the development of commercial AI processors.

FPGA

DSAs can boost processing capability, but the cost of ASIC design is a significant barrier to their development. To avoid these costs, the electronics industry has turned to Field Programmable Gate Arrays (FPGAs) to implement hardware designs. FPGAs consist of logic blocks and interconnect that can be reconfigured in the field by the end user. Higher end FPGA's also can contain hardware implementations (hard IP blocks) of more common functions like DRAM controllers, network ports, and even CPUs. With sufficient skill and effort, it is possible to create DSAs with FPGAs at relatively low material cost. The open question is whether the benefits are worth the development effort.

6 Memory

Dynamic RAM (DRAM) is the working memory of choice for CPUs, but DRAM hasn't kept pace with CPUs performance for decades. Complex cache systems and DSAs have been developed to mitigate these problems, but cannot eliminate them. Advances in DRAMs have been limited to bandwidth and capacity, as access latencies have remained relatively unchanged. In addition, higher capacity and higher performance are inversely related, with High Bandwidth Memory 2 (HBM2) providing higher bandwidth at lower capacity compared to DDR5, which provides higher capacity at lower bandwidth, and with graphics memory (GDDR5) somewhere in between [35]. From a market perspective, DRAM sales are expected to decline by 38% in 2019 compared to 2018 [36].

7 Storage

With the increasing amounts of data being collected by HEP experiments, the advancements in storage are critical moving forward. The falling cost and increasing capacity of solid state memory (3D NAND flash) has resulted in a veritable renaissance in high performance storage. However, more traditional magnetic disks and tape continue to evolve slowly and are encountering technological and economic issues.

Flash

3D NAND flash is the technology of choice for non-volatile, high performance block storage. With more cell layers, currently at 128, and more bits per cell, 3 to 4 bits per cell in production, 5 bits in development, storage capacity continues to increase. Cost for flash memory dropped in 2019, due to production oversupply and relatively weak demand [36, 37, 40]. Flash revenue is expect to shrink by 32% in 2019, but recover in 2020 [37]. Flash has replaced magnetic disks for low latency and high bandwidth applications. To access these capabilities, legacy I/O buses like SCSI and block storage software stacks in operating systems have been replaced with NVMe (non-volatile memory express) and its associated NVMe storage stack. NAND flash is triggering the development of new form factors (e.g. M.2, EDSFF) while NVMe is enabling the development of network attached flash via NVMe over Fabric (NVMeoF) [38, 39].

Persistent Memory

The development of low latency flash, lower than "normal" capacity flash, and Intel's Optane non-volatile memory has spurred interest in Persistent or Storage Class memory (PM or SCM memory). PM is non-volatile memory on the CPU memory bus and is less expensive than DRAM, but higher latency. There are three different behavioral models for PM, defined in

the Storage Networking Industry Association (SNIA) Non-Volatile Memory Programming model [42]. These are DRAM-like but larger capacity (JEDEC NVDIMM-P), DRAM-like but persistent (JEDEC NVDIMM-N), and disks-like but faster [41]. When used in a computer system, the different modes require modification to different parts of the systems, this may include CPU support, BIOS/OS support, memory bus extensions (e.g. modifications to DDR4 for Optane), application support or a combination of these changes.

Hard Disk Drives (HDD)

Magnetic hard disks host the preponderance of online data for HEP. Historically, HDD capacity was doubling every 13 months and cost per TB was decreasing at an average rate of 40% per year, but more recently this has dropped to 20% per year [48]. The maximum areal bit density of Perpendicular Magnetic Recording (PMR) technology, used in current HDDs, and limits to the number of platters in the standard 3 1/2" form factor have been reached [43]. Near term solutions to increase drive capacities, notably shingled magnetic recording (SMR), significantly alter drive behavior, with some requiring explicit application support. Longer term solutions, like energy assisted HAMR (heat) and MAMR (microwave) recording, continue to be pushed back, with new estimates of drive availability in 2022 [43]. Performance limits, in the form of input and output operations per second (IOPS) have also plateaued. To rectify this situation, multi-actuator drives have been promised for 2020 that will double the IOPS per 3 1/2" drive [45]. In addition to these technical hurdles, the HDD market is also experiencing problems.

Over the past decade, the HDD market has been shrinking as SSDs continue to grab market share from HDDs. Near-line HDDs, the type used by HEP, is the only growth market for HDDs [44]. Also, hyperscalers purchase almost 50% of all HDDs, giving them significant influence in the market [46]. The contraction of the HDD market has lead to the consolidation of vendors, of which there are now three, and a reduction in the number of factories, with four plants between the top two vendors.

Tape

Historically, the archival media of choice for HEP has been magnetic tape. Its role has expanded over the years to near-line storage, with the deployment of robotic tape libraries. The technical road map for tape is clear, with 300-400MB/sec LTO-8 and IBM TS1160 tape drives in the market with uncompressed tape capacities at 12TB and 20TB respectively; LTO-9 technology at the horizon (2020?), and multi-100TB tapes demonstrated in the lab [47]. However, the economics of tape is less clear. Tape media revenue is currently estimated at $0.7 billion USD/year and has been flat or falling over the past few years [48, 49]. Use of large capacity disks as a backup media and high fixed costs for tape compared to disk have been factors in the decline. IBM is the only leading edge tape drive manufacturer and the dominant driver of tape R&D. On the media side, only two manufactures are left, and they only recently settled a legal dispute that disrupted sale of LTO-8 media [50]. Finally, the hyperscalers represent a substantial consumer of tape technology. These factors can lead to a distorted market that can have detrimental consequences.

8 Network

High performance networking is a fundamental requirement for HEP, as large volumes of data need to be moved from the detector to compute and storage systems located around the world. This is one area where HEP has benefited from the interest of hyperscalers as they

have driven the rapid transition to higher single lane bit rates, from 25 Gigabits per second (Gbps) to 50 Gbps and now 100 Gbps and concurrent reductions in cost [51, 53]. However, this rapid change is not without consequences. Shorter technology life cycles have increased the pressure to upgrade equipment earlier and reduce investments in "older" technology, e.g. 10/40 gigabit Ethernet (GbE). The pace of change has also exceeded the speed of the IEEE standardization process [52]. A side effect of this change is the proliferation of different optical interconnect options at 100, 200, and 400 GbE, with different transceiver types, distance limitations, single lane bit rates, and "break out" capabilities [54, 55].

9 System Architecture

New technologies like flash and domain specific accelerators, as well as limitations in existing technologies has resulted in a re-thinking of system architectures [59]. Chiplets (small, function specific integrated circuits), expanded use of multi-chip modules (e.g., AMD Zen, HBM2 memory), cache coherent interconnects (CCIX, CXL, NVLink, GenZ) for connectivity to CPUs, NVMe over Fabrics, and persistent memory are examples of new technologies that have been developed to allow radically different compute system architectures [56–58]. What new system architectures develop and which make lasting changes to computing is unclear at this time.

10 Summary

Next generation Big Science experiments like the High Luminosity LHC (HL-LHC), the Deep Underground Neutrino Experiment (DUNE), and the Square Kilometer Array (SKA), are highly dependent on the timely advancement of compute, storage, and network technology [60–62]. Cost, capacity, and performance need to improve at a reasonable pace so that network, compute and storage costs remain within reach. HEP directly benefits from a robust electronics industry and market as they finances R&D into more advance (faster, lower power, smaller) electronics, both analog and digital, and enables cost reduction from economies of scale and technology improvements. Healthy competition in the CPU business drives lower compute costs and higher performance, which benefits HEP greatly as data processing requirements increase due to more complex data extraction problems (like greater pile up) and larger volumes of data. The drive towards domain specific accelerators holds the potential to speed up data processing both at the counting house, enabling things like software triggers, and in the data center. Domain specific processors also may change HEP data analysis methods as machine learning techniques, for which AI processors are specifically designed, are being investigated by HEP experiments. The proliferation of these devices outside of HEP has fostered the development of software libraries that the HEP community has leveraged with minimal to no cost. Substantial increases in network performance, driven by the hyperscalers will enable the HEP community to move the data at the counting house, in the data center, and around the world at rates commensurate with the volume of data being generated. The migration towards chiplets and multi-chip modules, in combination with inter-chip interconnects like CXL, CCIX, and Gen-Z, may have an positive impact on HEP detectors and data acquisition systems by enabling the ability to compose customized hardware from standardized building blocks.

Advances are occurring, but there are definite technical and economic risks in the roadmap of several key technologies. In addition utilization of some of these advances entail a substantial amount of software development that must be started several years in advance of their availability (or use) as software development is a time consuming endeavor. Without

knowledge of the direction that technology is advancing, HEP will not be prepared to take advantage of these advances.

References

[1] Merritt, R. (2018, Aug. 27). *GlobalFoundries Halts 7nm Work*, EE Times, [Article] Retrieved from https://www.eetimes.com/globalfoundries-halts-7nm-work/

[2] LaPedus, M. (2018, Jun. 21). *Big Trouble At 3nm*, Semiconductor Engineering, [Article] Retrieved from https://semiengineering.com/big-trouble-at-3nm/

[3] Semiconductor Industry Association (SIA). (2019, Feb. 4). *Global Semiconductor Sales Increase 13.7 Percent to $ 468.8 Billion in 2018* [Press Release] Retrieved from https://www.semiconductors.org/global-semiconductor-sales-increase-13-7-percent-to-468-8-billion-in-2018/.

[4] Semiconductor Industry Association (SIA). (2019, Jul. 1). *Global Semiconductor Sales Decrease 14.6 Percent Year-to-Year in May* [Press Release] Retrieved from https://www.semiconductors.org/global-semiconductor-sales-decrease-14-6-percent-year-to-year-in-may/.

[5] World Semiconductor Trade Statistics (WSTS). (2019, Aug. 27). *Following 13.7 percent growth in 2018 the Worldwide Semiconductor Market is expected to be down 13.3 percent in 2019 returning to growth in 2020* [Press Release] Retrieved from https://www.wsts.org/esraCMS/extension/media/f/WST/4164/WSTS_nr-2019_08.pdf.

[6] Sperling, E. (2019, Jun. 3). *CEO Outlook: It Gets Much Harder From Here*, Semiconductor Engineering, [Article] Retrieved from https://semiengineering.com/ceo-outlook-the-easy-stuff-is-over/

[7] International Data Corporation (IDC). (2019, Sep. 4). *Worldwide Server Market Revenue Declined 11.6% Year Over Year in the Second Quarter of 2019, According to IDC* [Press Release] Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS45482519.

[8] Morgan, T. P. (2019, Sep. 5). *The Inevitable Slowdown In Server Spending*, The Next Platform, [Article] Retrieved from https://www.nextplatform.com/2019/09/05/the-inevitable-slowdown-in-server-spending/.

[9] Merritt, R. (2019, Jul. 25). *Semiconductors Swim with Seven Whales*, EE Times, [Article] Retrieved from https://www.eetimes.com/semiconductors-swim-with-seven-whales/.

[10] Cutress, I. (2019, Apr. 2). *The Intel Second Generation Xeon Scalable: Cascade Lake, Now with Up To 56-Cores and Optane*, AnandTech, [Article] Retrieved from https://www.anandtech.com/show/14146/intel-xeon-scalable-cascade-lake-deep-dive-now-with-optane

[11] Salter, J. (2019, Aug. 9). *A detailed look at AMD's new Epyc "ROME" 7nm server CPUs*, Ars Technica, [Article] Retrieved from https://arstechnica.com/gadgets/2019/08/a-detailed-look-at-amds-new-epyc-rome-7nm-server-cpus/

[12] Hruska, J. (2019, Jul. 18). *Intel Acknowledges It Was 'Too Aggressive' With Its 10nm Plans*, ExtremeTech, [Article] Retrieved from https://www.extremetech.com/computing/295159-intel-acknowledges-its-long-10nm-delay-caused-by-being-too-aggressive

[13] Sperling, E. (2019, Feb. 5). *Cheaper Packaging Options Ahead*, Semiconductor Engineering, [Article] Retrieved from https://semiengineering.com/cheaper-packaging-options-ahead/

[14] Simonite, T. (2019, Nov. 06). *To Keep Pace With Moore's Law, Chipmakers Turn to 'Chiplets'*, Wired, [Article] Retrieved from https://www.wired.com/story/keep-pace-moores-law-chipmakers-turn-chiplets/

[15] Stuecheli, J., Willenborg, S., Starke, W. (2019, Aug. 18). *IBM's Next Generation POWER Processor*, Hot Chips, Presentation at the Hot Chips 2019 Symposium. Retrieved from https://www.hotchips.org/hc31/HC31_1.3_IBM.v12.pdf

[16] ARM. *Armv8-A Instruction Set Architecture* [Product Information] Retrieved from https://developer.arm.com/architectures/learn-the-architecture/armv8-a-instruction-set-architecture

[17] Marvell. *ThunderX2 Arm-based Processors* [Product Information] Retrieved from https://www.marvell.com/products/server-processors/thunderx2-arm-processors.html

[18] Robinson, C. (2019, Dec. 1). *New Gigabyte Marvell ThunderX2 32 Core Servers Available*, Serve The Home, [Article] Retrieved from https://www.servethehome.com/new-gigabyte-marvell-thunderx2-32-core-servers-available

[19] Cray. (2018, Jun. 25). *Inside Isambard — the World's First Production Arm Supercomputer* [Blog Post] Retrieved from https://www.cray.com/blog/inside-isambard-worlds-first-production-arm-supercomputer/

[20] Fujitsu. *Supercomputer "Fugaku"* [Presentation] Retrieved from https://www.fujitsu.com/global/Images/supercomputer-fugaku.pdf

[21] Ampere Computing. *eMag* [Product Information] Retrieved from https://amperecomputing.com/emag/

[22] Williams, C. (2018 Nov. 27). *Amazon's homegrown 2.3GHz 64-bit Graviton processor was very nearly an AMD Arm CPU*, The Register, [Article] Retrieved from https://www.theregister.com/2018/11/27/amazon_aws_graviton_specs/

[23] Frumasanu, A. (2019, Feb. 20). *Arm Announces Neoverse N1 & E1 Platforms & CPUs: Enabling A Huge Jump in Infrastructure Performance*, AnandTech, [Article] Retrieved from https://www.anandtech.com/show/13959/arm-announces-neoverse-n1-platform

[24] European Processor Initiative (EPI). (2019). *European Processor Initiative* [Article] Retrieved from https://www.european-processor-initiative.eu/project/epi

[25] RISC-V. (2020). *RISC-V International*, Retrieved from https://riscv.org

[26] European Processor Initiative (EPI). (2019). *Accelerator European Processor Initiative* [Article] Retrieved from https://www.european-processor-initiative.eu/accelerator/

[27] European Processor Initiative (EPI). (2019). *European Processor Initiative* Retrieved from https://www.european-processor-initiative.eu/

[28] Top 500. (2019, Nov.). *November 2019 TOP500 Supercomputer Sites* [Press Release] Retrieved from https://www.top500.org/lists/2019/11/

[29] NVIDIA. *NVIDIA V100* [Product Information] Retrieved from https://www.nvidia.com/en-us/data-center/v100/

[30] AMD. (2019, May 7). *AMD EPYC CPUs, AMD Radeon Instinct GPUs and ROCm Open Source Software to Power World's Fastest Supercomputer at Oak Ridge National Laboratory* [Press Release] Retrieved from https://www.amd.com/en/press-releases/2019-05-07-amd-epyc-cpus-radeon-instinct-gpus-and-rocm-open-source-software-to-power

[31] Intel. (2019, Nov,17). *Intel Unveils New GPU Architecture with High-Performance Computing and AI Acceleration, and oneAPI Software Stack with Unified and Scalable Abstraction for Heterogeneous Architectures* [Press Release] Retrieved from https://newsroom.intel.com/news-releases/intel-unveils-new-gpu-architecture-optimized-for-hpc-ai-oneapi/

[32] Google. *Cloud TPU* [Product Information] Retrieved from https://cloud.google.com/tpu

[33] Cutress, I. (2019 Dec. 4). *Supermicro shows off Intel Nervana NNP-T Servers: 8-Way PCIe and OAM*, Anandtech, [Article] Retrieved from https://www.anandtech.com/show/15179/supermicro-shows-off-intel-nervana-nnpt-

servers-8way-pcie-and-oam

[34] Habana. *Training : Habana* [Product Information] Retrieved from https://habana.ai/training/

[35] Verma, A. (2020, Feb. 2). *GDDR5 vs GDDR5X vs HBM vs HBM2 vs GDDR6 Memory Comparison*, Graphics Card Hub, [Article] Retrieved from https://graphicscardhub.com/gddr5-vs-gddr5x-vs-hbm-vs-hbm2/

[36] IC Insights. (2019, Jul. 31). *Despite 38% Sales Decline, DRAM Expected to Remain Largest IC Market* [Press Release] Retrieved from https://www.icinsights.com/news/bulletins/Despite-38-Sales-Decline-DRAM-Expected-To-Remain-Largest-IC-Market/.

[37] Counterpoint Research. (2019, Nov. 8). *Expected Memory Market Rebound in 2020 and Continuous Investment Can Strengthen Samsung's Leading Position in Memory* [Press Release] Retrieved from https://www.counterpointresearch.com/expected-memory-market-rebound-2020-continuous-investment-can-strengthen-samsungs-leading-position-memory/

[38] Shaw, M. (2018, Aug. 8). *Datacenter Designs Using the EDSFF Form Factors* Paper presented the 2018 Flash Memory Summit. Retrieved from https://www.flashmemorysummit.com/English/Collaterals/Proceedings/2018/20180808_SSDS-201-1_Shaw.pd

[39] Mohan, R. (2019 Sept.). *NVMe over Fabric SAN for the next-gen Data Center? Why should I care?* Paper presented at the 2019 Storage Developer Conference. Retrieved from https://www.snia.org/sites/default/files/SDC/2019/presentations/NVMe-oF/Mohan_Rupin_NVMe_Over_Fabrics_SAN_for_the_Next-gen_Data_Center.pdf

[40] Handy, J. (2019 Aug. 6). *Where Are We in the Current Flash Cycle ?* Paper presented at the 2019 Flash Memory Summit. Retrieved from https://www.flashmemorysummit.com/Proceedings2019/08-06-Tuesday/20190806_FTEC-101A-1_Handy.pdf

[41] Gervasi, B. (2018, Aug. 7). *Non-Volatile Memory Modules [NVDIMM]*, Paper presented at the 2018 Flash Memory Summit. Retrieved from https://www.flashmemorysummit.com/English/Collaterals/Proceedings/2018/20180807_PMEM-102-1_Gervasi.pdf

[42] Storage Network Industry Association (SNIA). *NVM Programming Model (NPM)* [Standard] Retrieved from https://www.snia.org/tech_activities/standards/curr_standards/npm

[43] Mellor, C. (2019, Oct. 7). *WD and Seagate mull 10-platter HDDs: Stopgap or BFF?*, Blocks and Files, [Article] Retrieved from https://blocksandfiles.com/2019/10/07/10-platter-hard-disk-drives/

[44] Coughlin T., (2019, Nov 14). *CQ3 2019 Hard Disk Drive Results*, Forbes, [Article] Retrieved from https://www.forbes.com/sites/tomcoughlin/2019/11/14/cq3-2019-hard-disk-drive-results/

[45] Mellor, C. (2019, Jun. 18). *Double-headed Seagate disk drives? Yes, on their way*, Blocks and Files, [Article] Retrieved from https://blocksandfiles.com/2019/06/18/double-headed-seagate-disk-drives-coming/

[46] Mellor, C. (2019, Oct. 7). *Hyperscalers will save disk drive makers from death by SSD*, Blocks and Files, [Article] Retrieved from https://blocksandfiles.com/2019/10/07/ssds-will-kill-hard-disks-in-enterprise/

[47] Lantz, M. (2018, Aug 28). *Why the Future of Data Storage is (Still) Magnetic Tape*, IEEE Spectrum, [Article] Retrieved from

https://spectrum.ieee.org/computing/hardware/why-the-future-of-data-storage-is-still-magnetic-tape

[48] Fontana, R. (2018 Sep 17). *A Ten Year (2008-2017) Storage Landscape –$/GB and Exabytes LTO Tape Media, HDD, NAND* Paper presented at the Designing Storage Architectures for Digital Collections meeting 2018. Retrieved from http://www.digitalpreservation.gov/meetings/DSA2018/Day_1/4_TO_LOC%202018%20 Talk__Fontana%20Decad__09172018_03.pdf

[49] Fontana, R. (2019, Sep. 9). *The 2018 Storage Landscape – An 11 Year Perspective (2008-2019) LTO Tape Media, HDD, NAND* Paper presented at the Designing Storage Architectures for Digital Collections meeting 2019. Retrieved from http://www.digitalpreservation.gov/meetings/DSA2019/Day_1/07_fontana2_Cloud-Storage-and-Tape-09092019.pdf

[50] Mellor, C. (2019, Aug. 6). *Sony and Fujifilm settle LTO-8 tape media patent dispute*, Blocks and Files, [Article] Retrieved from https://blocksandfiles.com/2019/08/06/lto-8-tape-media-manufacturing-block-mysteriously-blown-away/

[51] Morgan, T. P. (2019 Aug. 18). *The Future Of Networks Depends On Hyperscalers And Big Clouds*, The Next Platform, [Article] Retrieved from https://www.nextplatform.com/2019/08/18/the-future-of-networks-depends-on-hyperscalers-and-big-clouds/

[52] Morgan, T. P. (2019 Feb. 21). *The Road To 400G Ethernet Is Paved With Bechtolsheim's Intentions*, The Next Platform [Article] Retrieved from https://www.nextplatform.com/2018/02/21/road-400g-ethernet-paved-bechtolsheims-intentions/

[53] Matsumoto, C. (2017, Feb. 8). *Arista Founder: Networking Standards Move Too Slowly for Cloud Giants*, SDxCentral, [Article] Retrieved from https://www.sdxcentral.com/articles/news/arista-founder-networking-standards-move-slowly-cloud-giants/2017/02/

[54] EthernetAlliance. (2019, Jan 1). *Ethernet Roadmap 2019 Side 2*, The Ethernet Alliance, [Poster] Retrieved from https://ethernetalliance.org/wp-content/uploads/2019/08/EthernetRoadmap-2019-Side2-ToPrint.pdf

[55] Arista Networks, *100G Transceivers and Cables: Q&A* [Article] Retrieved from https://www.arista.com/assets/data/pdf/Arista100G_TC_QA.pdf

[56] Bennet, J. (2020, Jan 30). *What's the Difference Between OpenCAPI, CXL, and Gen-Z*, Electronic Design, [Article] Retrieved from https://www.electronicdesign.com/industrial-automation/article/21121821/whats-the-difference-between-opencapi-cxl-and-genz

[57] NVidia, *NVidia NVLink High-Speed GPU Interconnect*, [Product Information] Retrieved from https://www.nvidia.com/en-us/design-visualization/nvlink-bridges/

[58] Mittal, M., Lawley, J. (2019). *Memory Expansion and Storage Acceleration with CCIX Technology* Paper presented at SuperComputing 2019, Retrieved from https://www.ccixconsortium.com/wp-content/uploads/2020/02/CCIX-SC19-Presentation-Storage-Acceleration.pdf

[59] Demir,Y., Pan, Y, Song, S., Hardavellas, N., Kim, J., Memik, G. (2014, Jun.). *Galaxy, a high-performance energy efficient multi-chip architecture using photonic interconnects* Paper presented at the 28th ACM International Conference on Supercomputing. Retrieved from https://dl.acm.org/doi/10.1145/2597652.2597664

[60] CERN. (2020). *High Luminosity LHC (HL-LHC)* Retrieved from https://home.cern/science/accelerators/high-luminosity-lhc

[61] Deep Underground Neutrino Experiment (DUNE). (2020). *Deep Underground Neutrino Experiment* Retrieved from https://www.dunescience.org/

[62] Square Kilometer Array (SKA). (2020). *Square Kilometer Array*, Retrieved from https://www.skatelescope.org/