HPI

HPI Hardware Update – June 2016

Markus Dreseler,
markus.dreseler@hpi.de

Summary

- Intel releases Broadwell E7 processors
- IBM refines Power roadmap
- Samsung first to produce sub-20nm DRAM
- Everspin announces new NVRAM product, attached to memory bus, but still order of magnitude smaller than DRAM
- IBM has breakthrough in another type of NVM, so-called Phase-Change Memory
- NVMe over Fabrics will allow for fast access to remote SSDs

New Xeon Broadwell E7 processors for Scale-Up

Intel has announced their new E7 v4 series. Doubling the available memory, they can now support 24 TB in an eight-socket system. Furthermore, they are socket-compatible to previous v3 processors, allowing an easy update. According to Intel, an average performance boost of 30% can be expected [XE1]. New hardware features of the Broadwell generation were discussed in the previous report.

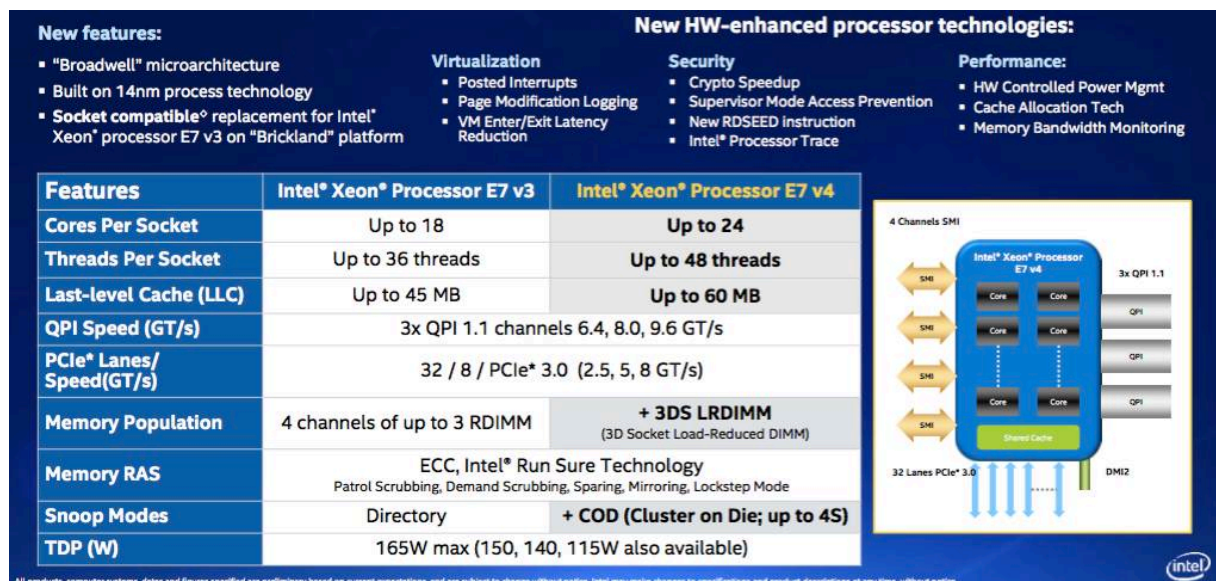The top-of-the-line E7-8890 v4 has 24 cores and a base frequency of 2.2 GHz [XE2].

Figure 1: Comparison of old (Haswell) and new (Broadwell) Xeon E7s [XE1]

More details about upcoming Power processors

At the OpenPower summit, IBM has released more information about their plans for the Power platform. Two things are to note: The upcoming Power8 refresh is no longer called Power8+, but Power8 with NVLink. This is significant as the plus-suffix previously denoted new processors on the same architecture, but with a shrunk process (e.g. 45nm to 32nm) and improved performance (compare to Intel's ticks) [PO1]. With the new generation of the Power8 processors, this shrink will not happen. Still, they are a significant improvement especially for the hyperscaler community, where the new NVLink interface allows for a tighter and faster (80 Gb/s with NVLink compared to 20 Gb/s) with PCIe coupling of accelerators such as Nvidia's Tesla series. These can help in machine learning or simulation applications [PO2].

Second, the upcoming Power9 architecture will be marketed in two flavors: Scale-Out (SO) and Scale-Up (SU). While processors optimized for both scaling models have been produced in the past, this is the first time IBM specifically labels them as such. While IBM announced that the P9 SO would have double the number of cores, information about the P9 SU is still rare. First guesses are that the Scale-Up variant will support more memory and have fewer cores at a higher clock speed [PO3].
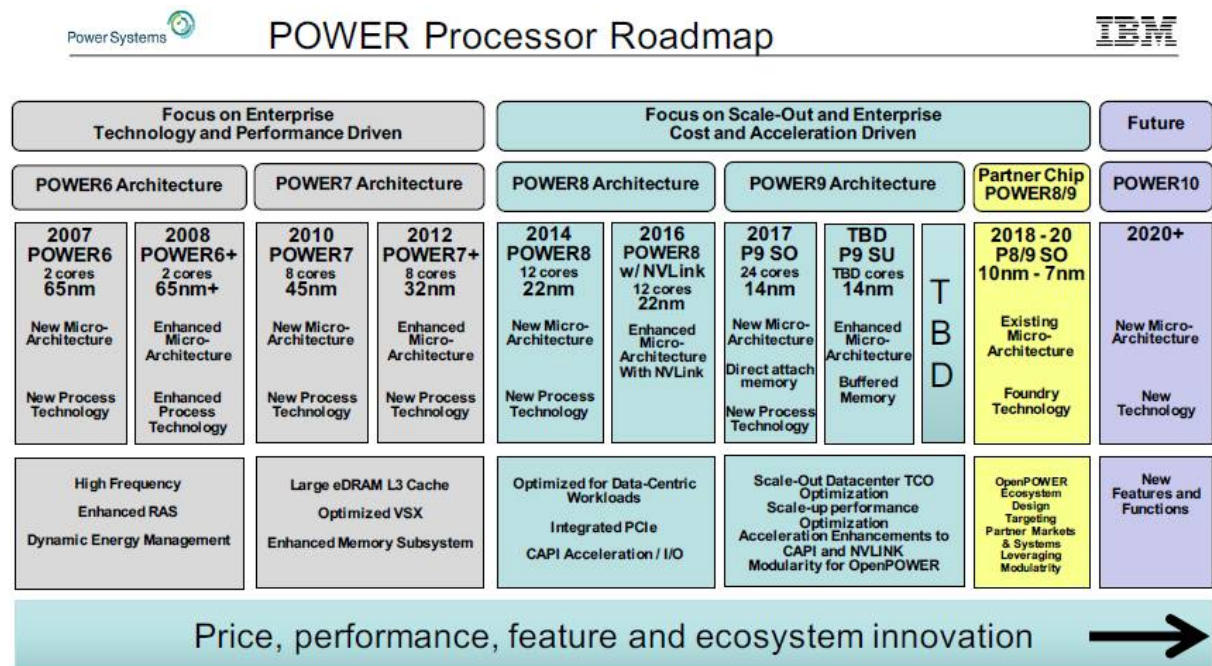

Figure 2: IBM's Power roadmap [PO1]

Google is already executing some of its applications on the Power8 platform, benefiting from what is said to be a better memory and I/O bandwidth [PO4]. Based on IBM's upcoming Power9, Google and Rackspace are now working together to build a new server design. The so-called "Zaius" platform will use two Power9 SO (Scale-Out) processors with 24 cores and 16 DDR4 memory sockets each. In addition to NVLink, the processors will

use IBM's Coherent Accelerator Processor Interface (CAPI) that allows CPUs and accelerators to coherently access the same memory [PO4].

According to their SVP of the technical infrastructure team, Google would switch to a Power architecture for its systems, even for a single generation, if it could get a 20 percent price/performance advantage [PO4].

Memory and Storage Technologies

While non-volatile memory continues to be considered a major change to the memory hierarchy, the development of traditional memory has not yet stopped.

Samsung produces 10nm-class DRAM

After presenting the first 128 GB DIMM earlier this year, Samsung announced another improvement. By producing DRAM that is smaller than 20nm, they can further improve the capacity and reduce both the production cost and the energy consumption [ME1].

Everspin announces 256 Mb ST-MRAM chips

One of the technologies for non-volatile memory is so-called Spin-Torque Magnetoresistive Memory (ST-MRAM). Different from other technologies, it uses magnetic effects instead of electrical effects to store information. While it is researched since the nineties, its commercial success has so far been limited. This is partly due to its limited capacity.

Everspin has now announced a new chip with a capacity of 256 Megabit, which is being sampled by selected customers. Later this year, a 1 Gigabit chip is said to be available. As the chips can be used on the DDR3 interface, this results in a 1 GB memory module [NV1] that is directly attached to the memory bus. On the one hand, Everspin claims "interface speeds comparable to DRAM" and "the highest endurance of currently available non-volatile memories" [NV2]. On the other hand, volatile memory has recently reached 128 GB per module, showing that MRAM is still far away from replacing DRAM and becoming "universal memory".

No prices are known yet. According to Everspin, it will be competitively priced compared to the cost of hardening a server to keep RAM electrified, hinting at a price that is significantly higher than volatile memory [NV3].

If the performance is really comparable to DRAM, this memory might serve as an efficient buffer in places such as a database log. Due to its limited capacity and high cost per byte, it will not be suitable for storing whole tables in the near future.

IBM stores 3 bit per PCM cell

Phase-Change Memory (PCM) is a technology that uses physical differences between the crystalline and non-crystalline (amorphous) phase of a material to store information. With electrical pulses, this material can be

changed from one phase to the other. IBM researchers have now stored three bits in one cell. This factor of three is said to increase the density and reduce the cost of this type of memory [NV4].

As of now, IBM's PCM is not yet ready for production. While IBM has not produced memory in a long time, this press release shows that they are still working on their own version of universal memory and looking to compete with Intel/Microns 3D XPoint [NV5].

Specification for NVMe over Fabrics released

NVM express (NVMe) is a protocol that specifies how fast SSDs can be attached via the PCIe bus. It optimizes for certain characteristics of the SSDs, such as their latency and parallelism advantages. A limitation up to now was that the PCIe connection limits both the number of storage devices per server and the flexibility in a data center.
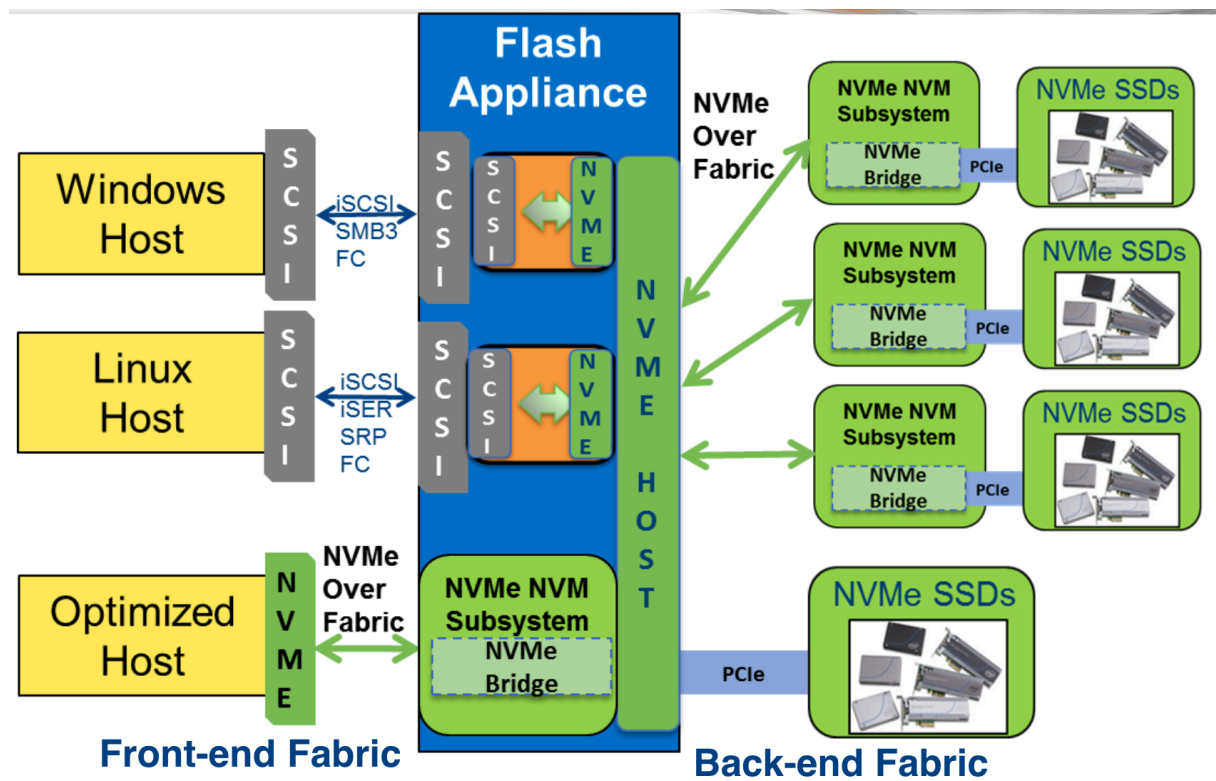

Figure 3: NVMe via Fabric (bottom) compared to current connections via SCSI (top) [NV7]

With the new NVMe over Fabrics standard, it will now be possible to access remote SSDs via Remote Direct Memory Access (RDMA) using fabrics such as Ethernet, Fibre Channel, or Infiniband [NV6]. With current technology, this adds 8 µs of latency compared to local PCIe access (~90 µs). With new memory technologies (3D XPoint, MRAM, …), the performance goal is to have remote data accesses via fabric with the same latency as local accesses via PCIe (~10 microseconds) [NV7].

Newsflash

- Joining the list of processor vendors that promise much for 2017, AppliedMicro has given more information about their performance goals for their X-Gene 3 ARM-based server chips. These are set to compete with Intel's current Broadwell Xeon E5 v4 processors. Features are 32 single-threaded cores that can issue four instructions at the same time and eight DDR4 memory controllers with a theoretical maximum memory bandwidth of 170 GB/s [NF1]. Compared to 76.8 GB/s with current Xeon E5 v4 and 102 GB/s with current E7 v4 processors [NF2], the memory subsystem can become a major selling point for the X-Gene 3 [NF1].

- Seven hardware vendors (AMD, ARM, Huawei, IBM, Mellanox, Qualcomm, and Xilinx) formed a consortium for Cache-Coherent Interconnect for Accelerators (CCIX). Their goal is to design a common, open architecture that makes it possible to coherently share memory between CPUs of different architectures, GPUs, and other types of accelerators [NF3]. This is interesting because it enables true hybrid computing in which different types of processors work on the same data, each doing the task they are best suited for. Notable is the membership of IBM, which already has such an interconnect, CAPI, and might tweak it towards a new CCIX standard [NF4].

- Google has been building its own processors, including one called "Tensor Processing Unit" (TPU) for use in machine learning applications. Compared to FPGAs and GPUs, these provide "an order of magnitude higher performance per Watt" [NF5].

References

[ME1] http://www.anandtech.com/show/10226/samsung-begins-to-produce-ddr4-memory-using-10-nmclass-process-tech

[NV1] http://www.eetimes.com/document.asp?doc_id=1329477
[NV2] https://www.everspin.com/file/965/download
[NV3] http://www.theregister.co.uk/2016/04/14/everspin_ships_256_megabyte_mram/
[NV4] http://www.enterprisetech.com/2016/05/18/memory-breakthrough-ibm-reports-3-bitscell-pcm/
[NV5] http://thememoryguy.com/ibm-jumps-on-the-new-memory-bandwagon/
[NV6] http://www.hpcwire.com/off-the-wire/nvm-express-fabrics-specification-released/
[NV6] https://www.openfabrics.org/images/eventpresos/workshops2015/DevWorkshop/Monday/monday_10.pdf

[NF1] https://semiaccurate.com/2016/04/25/appliedmicros-x-gene-3-aims-for-intels-e5-xeons/
[NF2] http://ark.intel.com/compare/91317,93793

[NF3] http://www.eetimes.com/document.asp?doc_id=1329734
[NF4] http://www.nextplatform.com/2016/05/23/chip-upstarts-get-coherent-hybrid-compute/
[NF5] http://www.eetimes.com/document.asp?doc_id=1329715

[XE1] http://www.theregister.co.uk/2016/06/06/intel_xeon_e7_v4/
[XE2] http://ark.intel.com/products/93790/Intel-Xeon-Processor-E7-8890-v4-60M-Cache-2_20-GHz

[PO1] http://www.nextplatform.com/2016/04/07/ibm-unfolds-power-chip-roadmap-past-2020/
[PO2] http://www.nextplatform.com/2016/04/18/power9-will-bring-competition-datacenter-compute/
[PO3] http://www.itjungle.com/tfh/tfh041216-story02.html
[PO4] http://www.nextplatform.com/2016/04/06/inside-future-google-rackspace-power9-system/