HPI Hardware Update - Oktober 2015

Markus Dreseler,
markus.dreseler@hpi.de

Summary

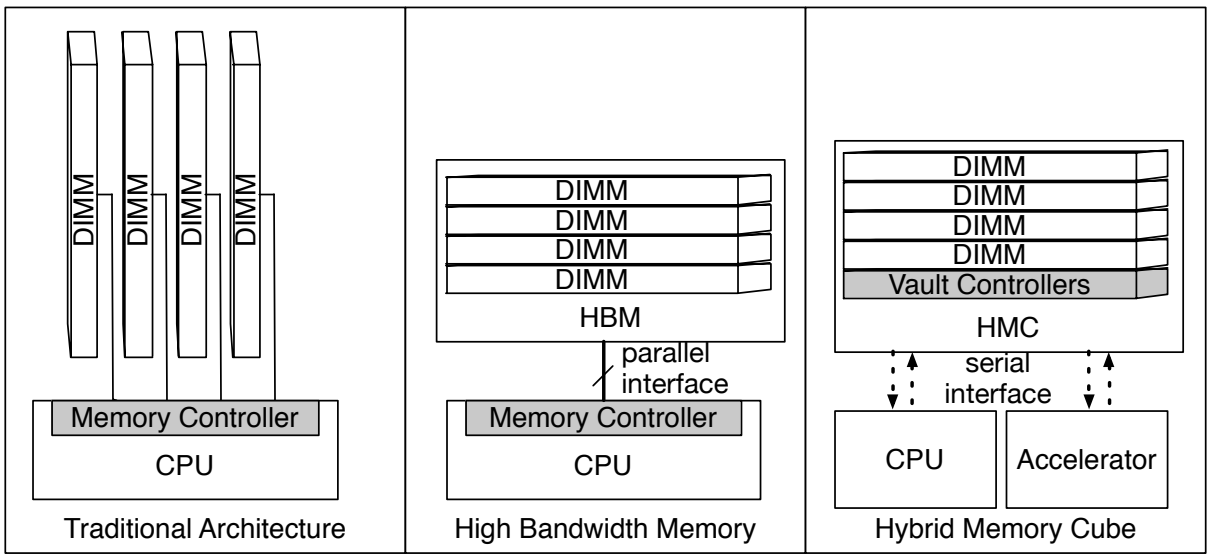
- New memory technologies stack existing DRAM cells to increase the bandwidth of RAM. This can be beneficial especially for applications with high memory-level parallelism (such as GPUs or accelerators), but comes with a higher cost than existing DRAM.
- Intel Broadwell server CPUs delayed to Q1/2 2016.
- Intel about to release Software Guard Extensions (SGX), protecting application data from foreign code and physical access to servers.
- Dell acquires EMC for $67B.

Memory Technologies

DRAM Pricing

As reported by DRAMeXchange, a website analyzing the memory market, the prices for both DDR3 and DDR4 continue to drop. From June to August, prices dropped by about 25%. For 2016, they forecast an even more substantial decrease [DD1, DD2].

High Bandwidth Memory (HBM)



Two competing memory technologies aim at stacking DRAM modules to achieve higher bandwidth. The first is High Bandwidth Memory Gen2 (HBM2), which is to be used in 2016 Nvidia GPUs. Produced by Samsung and SK Hynix, HBM2 allows for 4-8x higher bandwidth and consumes 40% less power [HB1] than DDR DRAM.

As HBM2 has a higher cost than DRAM, analysts predict that it will not be used for CPUs [HB2]. As far as accelerators go, opinions are divided and some sources claim that AMD's "Exascale Heterogenous Processor" will feature HBM2 [HB3]. However, AMD is said to have difficulties sourcing chips [HB4].

Hybrid Memory Cube (HMC)

The second competing technology is the Hybrid Memory Cube (HMC), developed by a consortium backed by Samsung (also involved in HBM2), Micron (also involved in 3D-XPoint), Intel, ARM, HP, and others. Other than HBM, which is primarily aimed at the graphics market, HMC also targets high-end servers and accelerators such as the Xeon Phi series [HM1].

With the current, second version of HMC, bandwidth up to 480 GB/s is reached [HM2]. The specifications for the third generation will be released later in 2016 [HM3].

Discussion of HBM and HMC (based on [DI1])

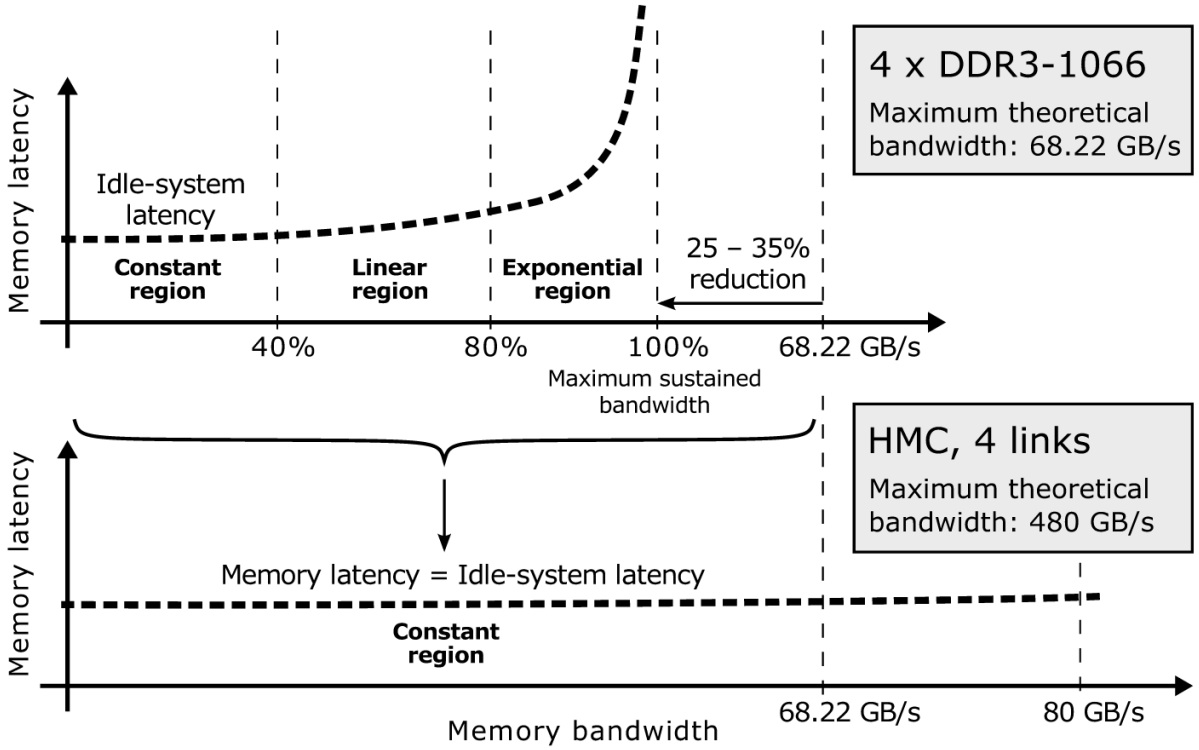
A main difference is that HBM keeps the traditional model with the CPU having a memory controller, while HMC moves the memory controller to the memory itself. Using a serial link, this allows HMC to connect to up to four "devices", e.g., CPUs, GPUs, or accelerators.

As both HBM and HMC are based on stacking existing DRAM technology, their main advantage is a higher achievable bandwidth. While the theoretically achievable latency remains the same, system load effects will also influence the latency when the system is under high load. This is shown in a paper from the Barcelona Supercomputing Center.

"If an application is in the constant-latency region [see graph below] of the DDR3 system (i.e., when memory latency corresponds to idle-system latency), upgrading the memory to the HMC will not reduce memory latency, nor will it improve overall performance. If the application is in the linear or exponential regions in the DDR3 system, a significant portion of its memory latency comes from collisions between concurrent memory requests. In this case, the bandwidth upgrade may reduce contention, which could reduce memory latency and improve performance".

Both HBM and HMC "support a logic layer at the bottom of the DRAM stack; this could support in-memory computation that reduces the amount of data transferred between memory and CPU"

These technologies are said to be unlikely to replace traditional DIMMs due to significantly higher costs.



Photonic Memory

An interesting technology further on the horizon is photonic Phase Change Memory [PM1]. Recently presented by researchers from Oxford and Münster, this technology uses light, instead of electricity, to change the phase of a cell. This is similar to the Phase Change Memory technology proposed for NVRAM, but is better suitable for upcoming architectures as it can be directly attached to a photonics-based fabric (rumored for Intel Cannonlake, successor to Skylake and targeted for 2017-2018 [PM2]).

Newsflash

- Intel has rescheduled the Broadwell server generation with single- and dual-socket SKUs (Broadwell-EP) planned for Q1 2016 and featuring up to 22 sockets. The Broadwell-EX for up to eight sockets (24 sockets, up to 60 MB L3-Cache) is said to be released in Q2 2016 [NF1, NF2].
- Following in Q3 2016 is the Xeon Phi x200 co-processor, codenamed Knights Landing (see previous hardware update) [NF1].
- Later in October, Intel will release the Software Guard Extensions (SGX) [NF3, NF4]. These allow applications to safeguard their data from access by other applications and to protect from attacks even "when an attacker has physical control of the platform and can conduct direct attacks on memory" [NF5].
- Dell acquired EMC for $67 billion. "The deal is being funded by [Michael Dell], his MSD Partners investment vehicle, Silver Lake and Singapore state-owned investment company Temasek Holdings, as well as debt financing, the VMware tracking stock and cash on hand"

[NF6]. EMC holds an 80 percent share in VMware [NF7]. Prior discussions with HP failed [NF8].

References

[DD1] http://www.kitguru.net/components/memory/anton-shilov/prices-of-ddr3-and-ddr4-memory-continue-to-drop-expected-to-decline-further/

[DD2] http://www.rambusblog.com/2015/09/16/mid-memory-price-dip-to-spur-ddr4-adoption/?utm_source=feedly&utm_medium=rss

[HB1] http://www.businesskorea.co.kr/news/industry/12151-hynix-and-nvidia-samsung-sk-hynix-supply-high-bandwidth-memory-nvidia

[HB2] http://www.fool.com/investing/general/2015/09/06/will-advanced-micro-devices-incs-upcoming-zen-base.aspx

[HB3] http://www.hpcwire.com/2015/09/22/chief-processor-architect-exits-amd/

[HB4] http://wccftech.com/amd-allegedly-facing-trouble-sourcing-hbm2-chips-arctic-islands-gpus-mainstream-lineup-stick-hbm1/

[HM1] http://www.extremetech.com/computing/197720-beyond-ddr4-understand-the-differences-between-wide-io-hbm-and-hybrid-memory-cube

[HM2] http://wccftech.com/open-silicon-announces-worlds-hmc-20-controller-ip-partners-include-nvidia-intel/

[HM3] http://www.kitguru.net/components/memory/anton-shilov/micron-to-unveil-third-gen-hybrid-memory-cube-in-2016/

[DI1] http://memsys.io/wp-content/uploads/2015/09/p31-radulovic.pdf

[PM1] https://techreport.com/news/28908/intel-reveals-details-of-its-omni-path-architecture-interconnect

[PM2] http://www.golem.de/news/purley-plattform-intels-cannonlake-ep-nutzt-ddr4-2933-und-silicon-photonics-1510-116691.html

[NF1] http://wccftech.com/intel-xeon-greenlow-broadwell-ep-ex-launch-schedule/

[NF2] http://www.kitguru.net/components/cpu/anton-shilov/intel-xeon-e5-e7-broadwell-processors-with-up-to-24-cores-due-in-2016/

[NF3] http://www.kitguru.net/components/cpu/anton-shilov/intel-alters-design-of-skylake-processors-to-enable-sgx-extensions/

[NF4] https://techreport.com/news/29146/intel-to-begin-shipping-skylake-cpus-with-sgx-enabled

[NF5] https://software.intel.com/en-us/blogs/2013/09/26/protecting-application-secrets-with-intel-sgx

[NF6] http://www.nytimes.com/2015/10/12/business/dealbook/dell-to-buy-emc-for-65-billion-a-record-takeover-in-technology.html

[NF7] http://www.reuters.com/article/2015/10/09/emc-us-ma-dell-idUSL3N12954920151009

[NF8] http://www.businessinsider.com/r-exclusive-tech-firms-hp-emc-call-off-merger-talks---sources-2014-10?IR=T