**Introduction**

Located at the Swiss National Supercomputing Centre, the Piz Daint is the flagship supercomputer for the Swiss HPC(high performance computing) service. In 2016, it was declared the fastest supercomputer in Europe, with a peak performance of up to 7.787 petaflops. Since then, Piz Daint has been combined with Piz Dora’s CX40 system into a hybrid XC50/XC40 Cray system that has speeds of up to 27 petaflops. The XC50 computation nodes contain a GPU and CPU while the XC40 nodes only have a CPU. All of these nodes are connected with a Aries routing and communications ASIC, with Dragonfly network topology [5.4].



Figure 1. Front end of Piz Daint [5.4]

**Architecture**

Piz Daint is two different supercomputer systems merged together into one. It contains 7517 total compute nodes. 5704 of these are CX50 nodes and 1813 are CX40 nodes. Each CX50 node is composed of one 12-core Intel Xeon E5-2690 v3 CPU and one NVIDIA P100 GPU. Conversely, each CX40 node contains two 18-core Intel Xeon E5-2695 v4 CPUs.

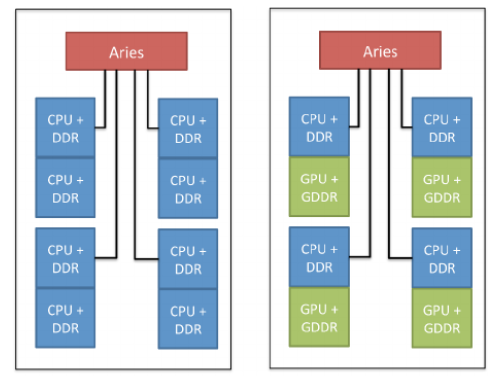


Figure 2. Diagram showing the Cray CX40 and CX50 nodes, respectively.[7]

The highest theoretical performance of the CX50 system is 27154 teraflops/s. For the CX40 the highest performance is at 2193 teraflops/s [5.3]. If we look at performance per compute node, the CX50 has 4.76 teraflops/s per node, while the CX40 system has 1.209 teraflops/s per node. Each CX50 node utilizes NVIDIA NVLink along with PCIe 3rd generation busses to maintain high speed connections between the CPU and GPU. This helps the Intel Xeon E5-2690 v3 CPU achieve a maximum memory bandwidth of 68GB/s and the Intel Xeon E5-2690 v4 CPU acheive 76.8 GB/s.

Piz Daint runs a Cray Linux environment along with the Nvidia CUDA toolkit v8.0. This includes various math libraries and optimization tools, along with support for a variety of languages such as C, C++, FORTRAN, and Python [9].

Piz Daint uses the Nvidia P100 GPU accelerator. It features 15.5 billion transistors and 54 streaming multiprocessors(SM) to help run multiple processes. It also has 3584KB of shared memory and a 4096KB L2 cache. Compared to the previous Nvidia M40 GPU, the P100 has over 500 more total CUDA cores and 7.3 billion more transistors [8.1].



Figure 3. A Nvidia P100 SM unit [8.1]

Additionally, each SM contains 32 double precision CUDA cores to process double precision arithmetic faster. This allows for more accurate results in HPS applications such as quantum chemistry [8.1].

**Parallelization**

Piz Daint was designed to analyze large volumes of data for use in data sciences and large scale high-resolution simulations. In order to return results in a reasonable time, the CX50 portion of the system utilizes parallelization. Every Intel Xeon E5-2690 v3 component in the CX50 portion has access to a Nvidia P100 GPU. This CPU has 12-cores, each of which can process 24 threads, for a maximum of 288 threads that can access up to 3584 CUDA cores on the GPU [12]. This allows the CPU to parallelize on a large scale, drastically reducing computational time. Additionally, Piz Daint has a 10-core Intel Xeon E5-2690 v3 CPU as a Login Node [5.4]. This node handles all user input and program compilation, allowing for the rest of the nodes on the CPU to focus on calculations.

Due to the large amount of data the Piz Daint processes, one important performance aspect is memory bandwidth speed. To address this, PIz Daint utilizes a high bandwidth memory(HBM) RAM interface[10]. HBM contains stacks, containing 4 DRAM dies, each with 2 independent memory channels. This means for each stack there are 8 independent memory channels to access DRAM[11]. These channels allow for multiple computational nodes to access the 8.7 PB of memory at once, reducing bottlenecks and possible stalls in the parallelization process.

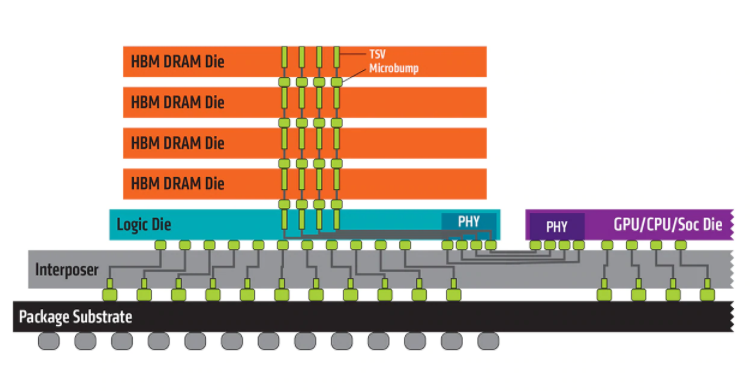


Figure 4. High Bandwidth Memory Diagram [11.1]

**Weaknesses**

Although the Intel Xeon E5-2690 v4 CPU components in the CX40 portion of the system are slightly better than the Intel Xeon E5-2690 v3 CPU components within the CX50, the CX40 contains only CPU components. Since CPU’s are primarily optimized for single threaded performance, this means CX40 system has far less parallelization potential compared to the CX50, making it far slower at performing large calculations. Since the Piz Daint is primarily used for large scale data analysis, most of it passes through the CX50 nodes with GPU accelerators, as they process much faster.

**Summit Comparison**

Piz Daint contains 7517 total compute nodes while Summit has 4608 nodes. However, only 5704 Piz Daint nodes are hybrid architecture. This means the rest of the nodes on Piz Daint cannot parallelize on a large scale. Each Piz Daint hybrid node has one 12-CPU and one GPU, on the other hand, each Summit compute node has 2 22-core CPUs and six GPU accelerators. Additionally, the GPU’s used by Summit are Nvidia V100’s, which are the successors to the Nvidia P100’s on the Piz Daint. One aspect of this is that the V100 has 24 more streaming multiprocessors, meaning it has 1536 more CUDA cores per GPU. Even so, both Summit and Piz Daint use NVLink. Summit has higher overall performance than Piz Daint.

Piz Daint utilizes PCIe 3rd generation data busses, which are slower than the 4th generation busses on the Summit [10], Piz Daint also has 241.3GB less storage than Summit. Since Piz Daint uses Pascal GPUs, they cannot page fault and any active processes must acquire the all of the memory space, which creates the potential for more bottlenecks in the parrelization. Additionally, the hybrid CPUs on Piz Daint have a maximum bandwidth capacity of 68 GB/s, while Summit has 100GB/s. Overall Summit has better memory performance and allocation the Piz Daint.

"Swiss National Supercomputing Centre (CSCS) | TOP500 Supercomputer Sites." *Top500.org*. Web. 2 Apr. 2019.

<https://www.top500.org/site/50422> [5.3]

"Piz Daint & Piz Dora." *CSCS*. N.p., 2019. Web. 2 Apr. 2019.

<https://www.cscs.ch/computers/dismissed/piz-daint-piz-dora/> [5.4]

"Systems-Level Configuration And Customisation Of Hybrid Cray XC30." *Research Gate*. N.p., 2014. Web. 2 Apr. 2019.

https://www.researchgate.net/publication/285583469\_Systems-level\_Configuration\_and\_Customisation\_of\_Hybrid\_Cray\_XC30 [7] (IN depth systems info)

*NVIDIA TESLA P100*. Web. 2 Apr. 2019.

<https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf> [8.1] (NVIDIA P100)

Marsella, Luca. *Best Practices For Building Software On Piz Daint*. 2019. Web. 2 Apr. 2019.

<https://www.cscs.ch/fileadmin/user_upload/contents_userLab/building_software_piz_daint.pdf> [9] (Software/OS)

Brueckner, Rich. "Creating Balance In HPC On The Piz Daint Supercomputer - Insidehpc." *insideHPC*. N.p., 2016. Web. 2 Apr. 2019.

<https://insidehpc.com/2016/08/creating-balance-in-hpc/> [10] (Datawarp Info)

O'Connor, Mike. *Highlights Of The High Bandwidth Memory (HBM) Standard*. 2014. Web. 3 Apr. 2019.

<http://www.cs.utah.edu/thememoryforum/mike.pdf> [11] (High bandwidth memory)

"High Bandwidth Memory | AMD." *Amd.com*. Web. 2 Apr. 2019.

<https://www.amd.com/en/technologies/hbm> [11.1] (High Bandwidth memory)

"Intel® Xeon® Processor E5-2690 V3 (30M Cache, 2.60 Ghz) Product Specifications." *Ark.intel.com*. Web. 3 Apr. 2019.

<https://ark.intel.com/content/www/us/en/ark/products/81713/intel-xeon-processor-e5-2690-v3-30m-cache-2-60-ghz.html> [12]