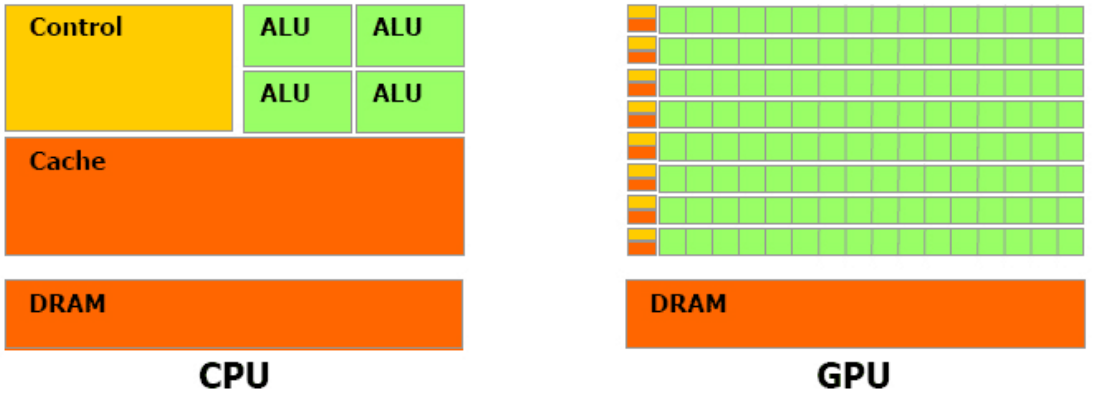
GPUs at CC-IN2P3:

Cluster overview

Workshop GPU @CC-IN2P3

Nicolas Fournials

Which differences with a CPU ?



- GPU : A lot of Arithmetic and Logical Units

  SIMD : identical operations on multiple datasets

- CPU : caches, advanced operations management

GPUs, differences with regular batch

- The GPU is a co-processor
  - Needs a CPU and its memory
  - The batch scheduler still relies on CPU attribution
  - Needs some additional work to be properly isolated
  - Needs some tuning to get proper performances
- The GPU has no advanced monitoring/accounting metrics
  - Rather basic metrics
  - Poorly integrated to the batch scheduler
- Architecture and performances redefined regularly
  - How to choose a reference benchmark?

GPUs available at CC-IN2P3

- 20 cards Nvidia Tesla K80
  - 40 GPUs Nvidia GK210, 12 GB DDR5

- 24 cards Nvidia Tesla V100
  - 24 GPUs Nvidia GV100, 32 GB HBM2, PCIe

Tesla K80-based servers

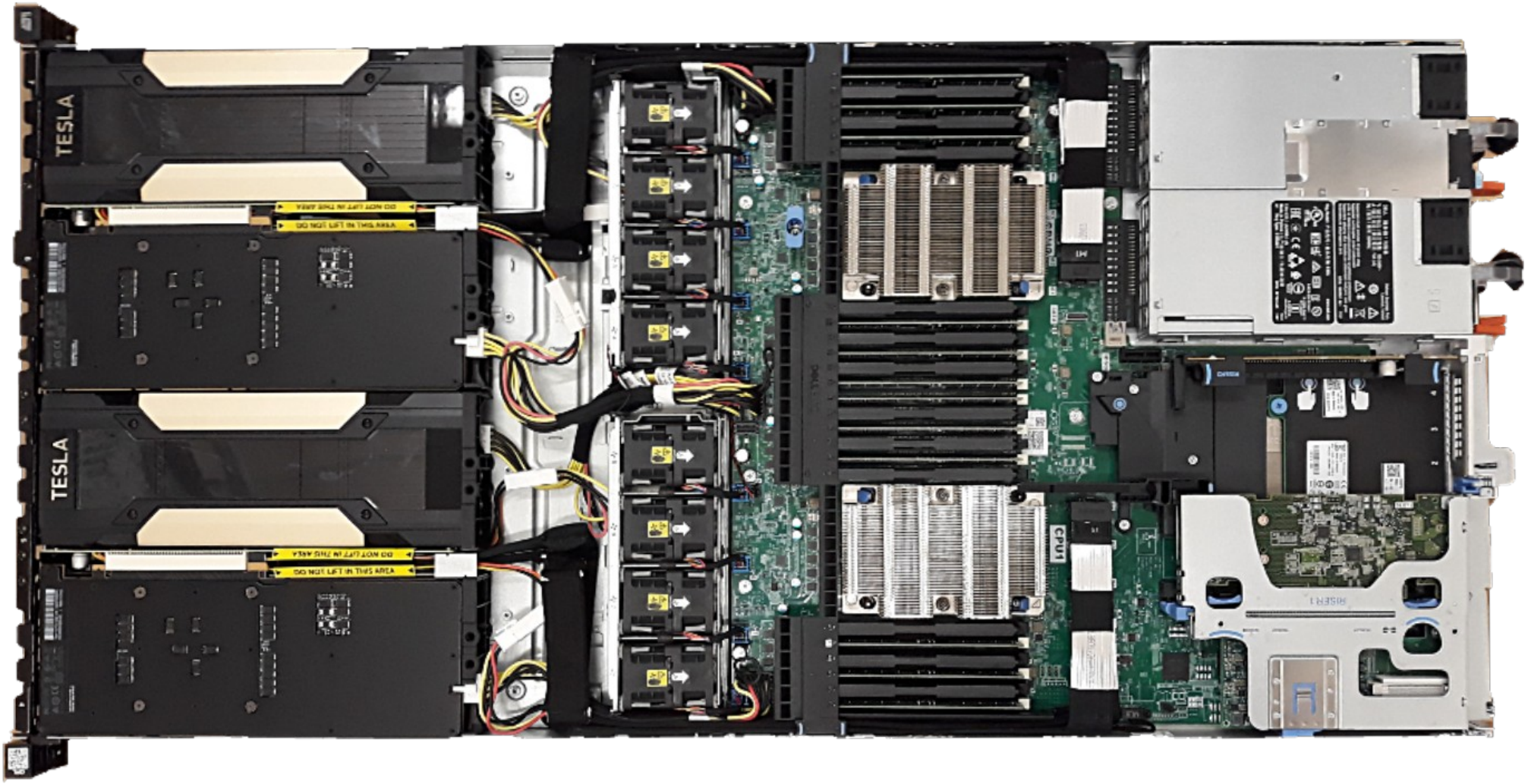
- Machines:
  - 10 Dell C4130
    - 2 Xeon E5-2640v3 @2,6 Ghz (8 cores)
    - 128 GB RAM
    - SSD
    - 2 Nvidia Tesla K80
      - ➡ 4 GPUs Nvidia GK210, 4x12 GB DDR5

- Network:
  - InfiniBand QDR 32 GB/s interconnect

Tesla V100-based servers

- 6 Dell C4140
  - 2 Xeon Silver 4114 @2,2 Ghz (10 cores)
  - 192 GB RAM
  - SSD M2
  - 4 Nvidia Tesla V100, 32 GB HBM2, PCIe

Tesla V100-based servers, how beautiful



GPUs available at CC-IN2P3: detailed view

- 20 cards Nvidia Tesla K80
  - 40 GPUs Nvidia GK210 with 12 GB DDR5
    - ALUs: 2496 FP32/832 FP64
    - 562/875 MHz
    - InfiniBand between nodes

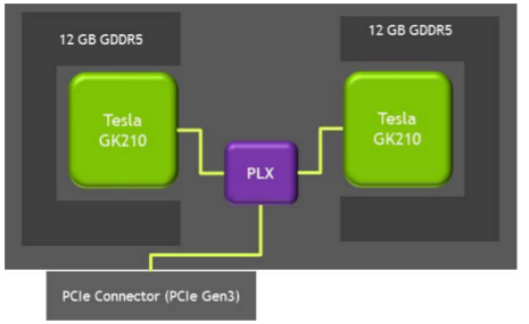
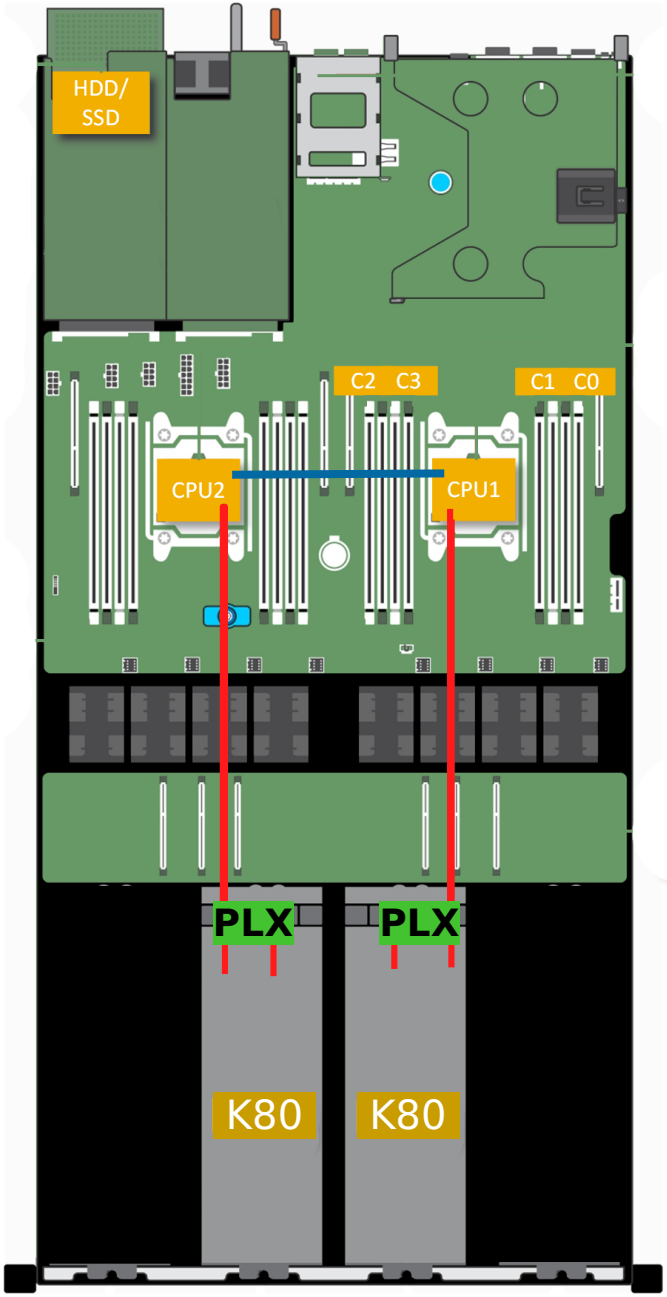


Figure 1. Tesla K80 Block Diagram

- 24 cards Nvidia Tesla V100
  - 24 GPUs Nvidia GV100 with 32 GB HBM2
    - ALUs: 5376 FP32/2688 FP64, Tensor cores
    - 1455 MHz
    - PCIe

GPU servers physical architecture

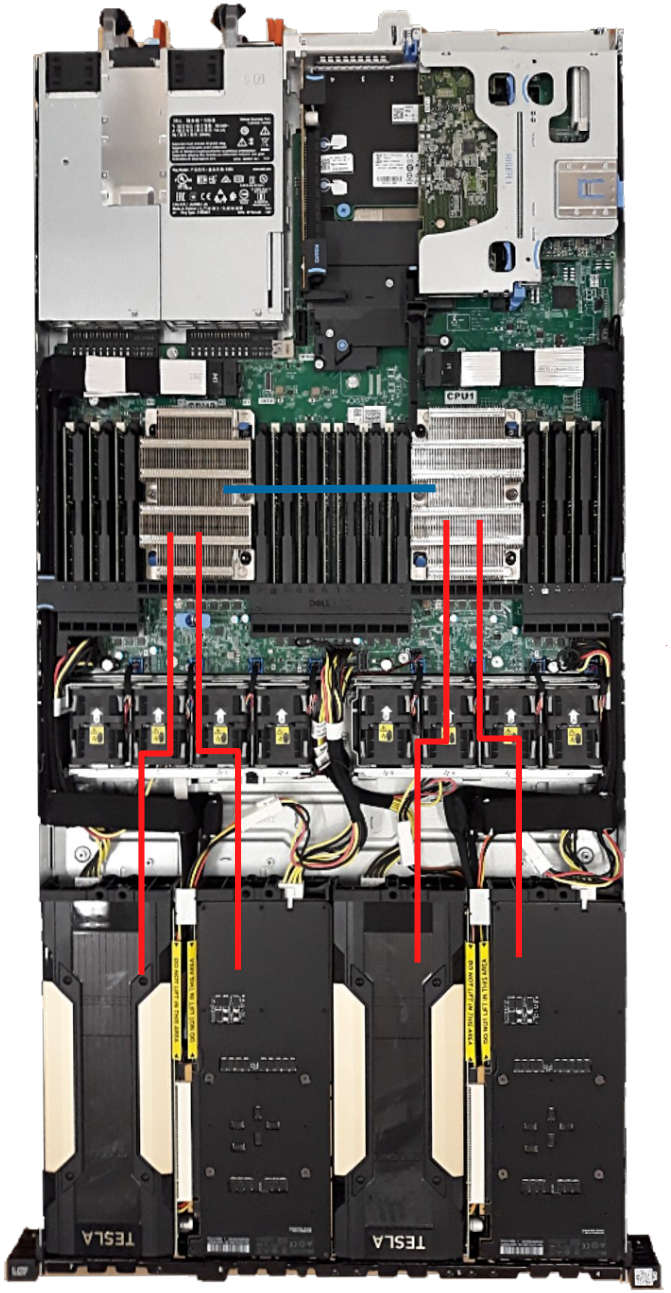
Dell C4130 (Tesla K80)



QPI

PCIe

Dell C4140 (Tesla V100)



- CUDA 10.1 (+ CUDA 9.2)
  - Relatively recent release
  - Keep the release n-1
  - **Singularity** for specific needs

- OpenCL 1.2 (OpenCL 2.0 partially)
  - Limited evolutions on Nvidia hardware
  - Open standard

OpenCL

Documentation

- https://doc.cc.in2p3.fr/jobs_gpu

Ideas, feedbacks? Let's talk!