MAX DRIVING THE EXASCALE TRANSITION

Introduction to Marconi100

Fabio Affinito
Cineca

Marconi100 system overview

Marconi 100 is an IBM AC922 (Whiterspoon) cluster

55 racks

980 nodes

Mellanox IB EDR DragonFly++

2 x Power9 CPU

16 cores each
4 HW threads each

4 x NVIDIA Volta V100 GPU
Nvlink 2.0, 16GB

256 GB/node

Marconi100 node architecture

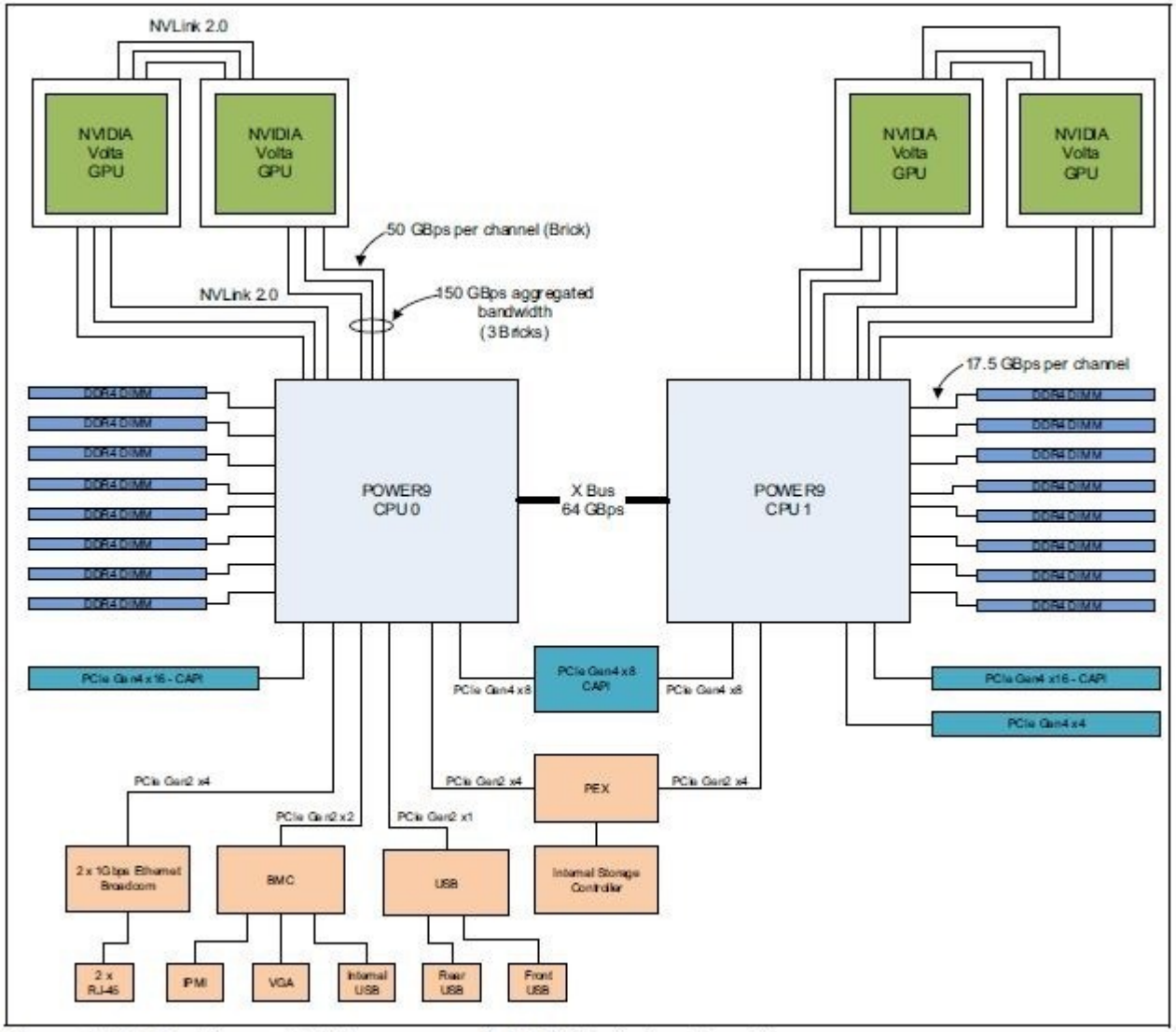


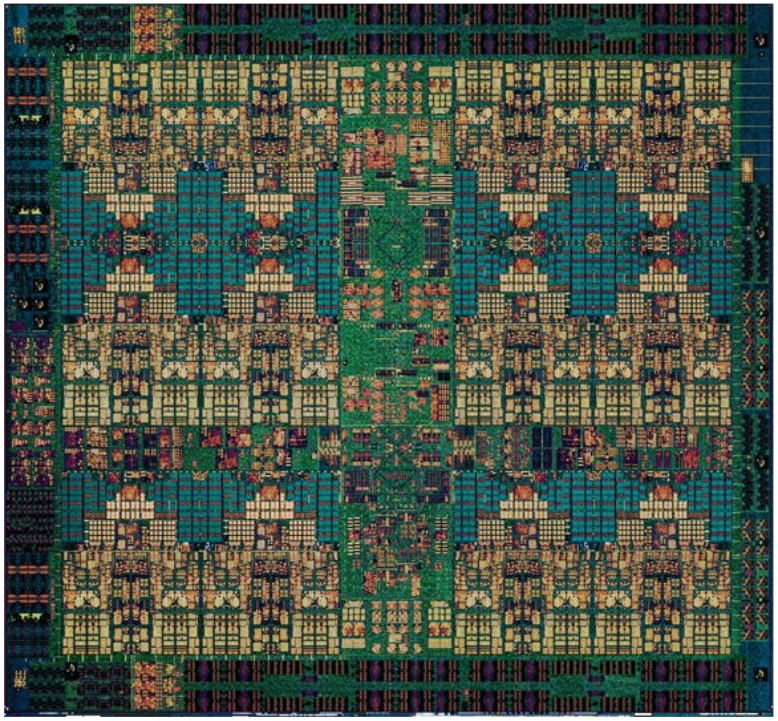
Figure 2-5 The Power AC922 server model GTH logical system diagram

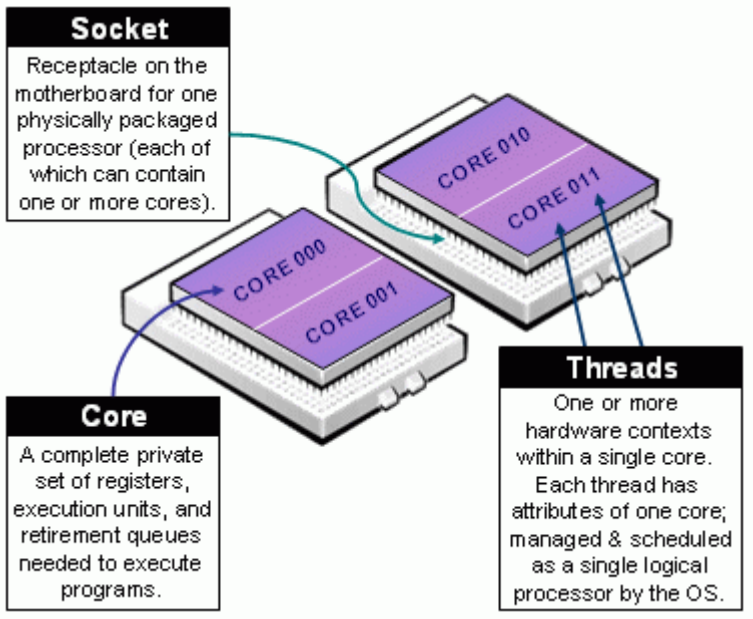
IBM Power9

Each node has two IBM Power9 sockets

2 (socket) x 16 (cores) x 4 HW threads

Total: 128 threads on the node





NVIDIA V100 GPU

Each Marconi100 node has four V100 GPUs

Each Tesla V100 GPU has:

- 150+150 GB/s total BW (**NVLink** v2.0)
- 5,120 CUDA cores (64 on each of 80 SMs)
- 640 Tensor cores (8 on each of 80 SMs)
- 20MB Registers | 16MB Cache | 16GB HBM2 @ 900 GB/s
- 7.5 DP TFLOPS | 15 SP TFLOPS | 120 FP16 TOPS

nVIDIA **GPUDirect** technology is fully supported (shared memory, peer-to-peer, RDMA, async), enabling the use of CUDA-aware MPI

Marconi100 Software Stack

- **Compilers**
  - **XL (IBM compilers: xlf90, xlC, etc.)**
  - GNU (gcc, gfortran)
  - **PGI**
  - CUDA
- **Communication Libraries**
  - **Spectrum_MPI**
  - OpenMPI
- **Libraries**
  - ESSL, BLAS, LAPACK, FFTW
  - HDF5, …
- **Other Modules Profiles**
  - **profile/chem**, profile/phys,

Support for OpenACC and CUDA Fortran

Fully optimized for M100 architecture

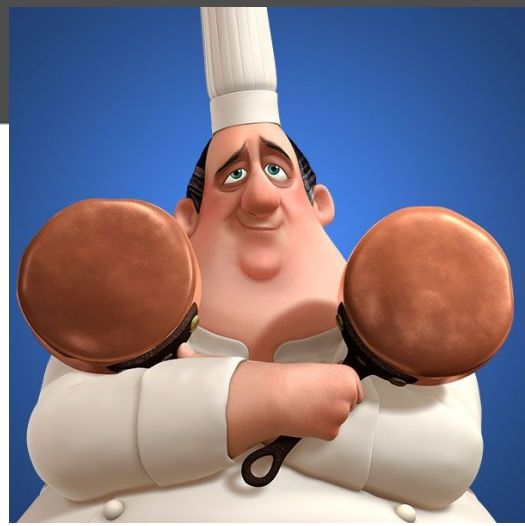
Here you can find a pre-built version of QE-GPU

Module environment – Compilers and libraries

```
[faffinit@login03 ~]$ ml
Currently Loaded Modulefiles:
 1) profile/base
[faffinit@login03 ~]$ ml av
------------------------------------------- /cineca/prod/opt/modulefiles/profiles --------------------------------------------
profile/advanced  profile/base  profile/candidate  profile/chem-phys  profile/deeplrn  profile/global  profile/lifesc

---------------------------------------- /cineca/prod/opt/modulefiles/base/environment ----------------------------------------
autoload

----------------------------------------- /cineca/prod/opt/modulefiles/base/libraries -----------------------------------------
blas/3.8.0--gnu--8.4.0                      lapack/3.9.0--pgi--19.10--binary              zlib/1.2.11--gnu--8.4.0
boost/1.72.0--spectrum_mpi--10.3.1--binary  nccl/2.6.4--cuda--10.1
essl/6.2.1--binary                          netcdf/4.7.3--gnu--8.4.0
fftw/3.3.8--gnu--8.4.0                      netcdf/4.7.3--spectrum_mpi--10.3.1--binary
fftw/3.3.8--spectrum_mpi--10.3.1--binary    netcdff/4.5.2--gnu--8.4.0
gsl/2.6--gnu--8.4.0                         netcdff/4.5.2--spectrum_mpi--10.3.1--binary
hdf5/1.12.0--gnu--8.4.0                     petsc/3.12.4--spectrum_mpi--10.3.1--binary
hdf5/1.12.0--spectrum_mpi--10.3.1--binary   scalapack/2.1.0--spectrum_mpi--10.3.1--binary
lapack/3.9.0--gnu--8.4.0                    szip/2.1.1--gnu--8.4.0

----------------------------------------- /cineca/prod/opt/modulefiles/base/compilers ------------------------------------------
cuda/10.1  gnu/8.4.0  pgi/19.10--binary  python/3.8.2  spectrum_mpi/10.3.1--binary  xl/16.1.1--binary

------------------------------------------- /cineca/prod/opt/modulefiles/base/tools --------------------------------------------
anaconda/2020.02  cmake/3.17.1  singularity/3.5.3  spack/0.14.2-prod  superc/2.0
[faffinit@login03 ~]$
```

For more detail, see the Marconi100 User Guide:
https://wiki.u-gov.it/confluence/display/SCAIUS/UG3.2%3A+MARCONI100+UserGuide

Building Quantum ESPRESSO



…if you don't want to use the one in the modules, you can build in it on your own…

Ingredients:
- **Source code (from:…)**
- **PGI compiler**
- **MPI (better Spectrum_MPI)**
- **Math libraries (better if you use OpenBLAS or ESSL)**
- **… some luck!**

The recipe…
https://gitlab.com/QEF/q-e-gpu/-/wikis/Marconi-100

Building Quantum ESPRESSO on Marconi100

```
module load profile/global
module load pgi/19.10--binary
module load cuda/10.1
module load spectrum_mpi/10.3.1--binary

wget https://gitlab.com/QEF/q-e-gpu/-/archive/gpu-develop/q-e-
gpu-gpu-develop.tar.bz2

tar xjf q-e-gpu-gpu-develop.tar.bz2

cd q-e-gpu-gpu-develop

./configure CC=pgcc F77=pgf90 FC=pgf90 F90=pgf90
MPIF90=mpipgifort --enable-openmp --with-cuda=$CUDA_ROOT
--with-cuda-runtime=10.1 --with-cuda-cc=70

make -j pw
```

Running Quantum ESPRESSO

```
#!/bin/bash
#SBATCH --nodes=16 # number of nodes
#SBATCH --ntasks-per-node=4 # number of tasks per node #SBATCH -
-ntasks-per-socket=2 # number of tasks per socket #SBATCH --
cpus-per-task=32 # number of HW threads per task #SBATCH --
gres=gpu:4 # gpus per node
#SBATCH --mem=230000MB
#SBATCH --time 01:00:00 # format: HH:MM:SS
#SBATCH -A YYYYYY
#SBATCH -p m100_usr_prod
#SBATCH –qos = …
export ….

mpirun --map-by ppr:${SLURM_NTASKS_PER_NODE}:node:PE=${OMP_NUM_THREADS}
       pw.x -npool 2 -ndiag 1 -inp file.in > file.out
```

How to get access to Marconi100

You can ask for computing hours on Marconi100 at Cineca with:

http://iscra.cineca.it

https://prace-ri.eu/hpc-access/

support@max-centre.eu

superc@cineca.it

DRIVING THE EXASCALE TRANSITION

THANKS