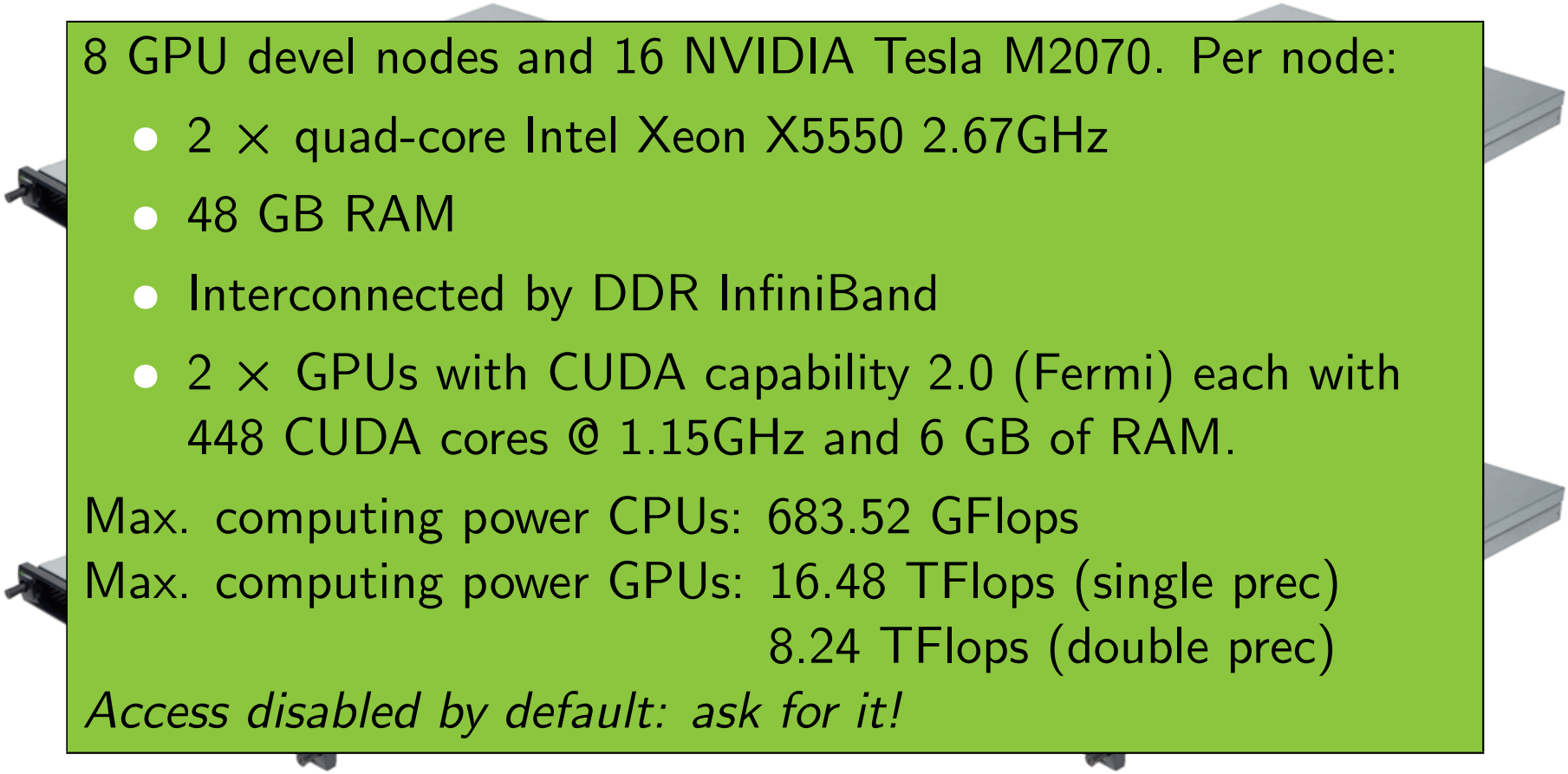
ARC: the GPU cluster at SciNet

SciNet HPC Consortium
Compute Canada

September 26, 2011

Accelerator Research Cluster (ARC)

Accelerator Research Cluster (ARC)



8 GPU devel nodes and 16 NVIDIA Tesla M2070. Per node:

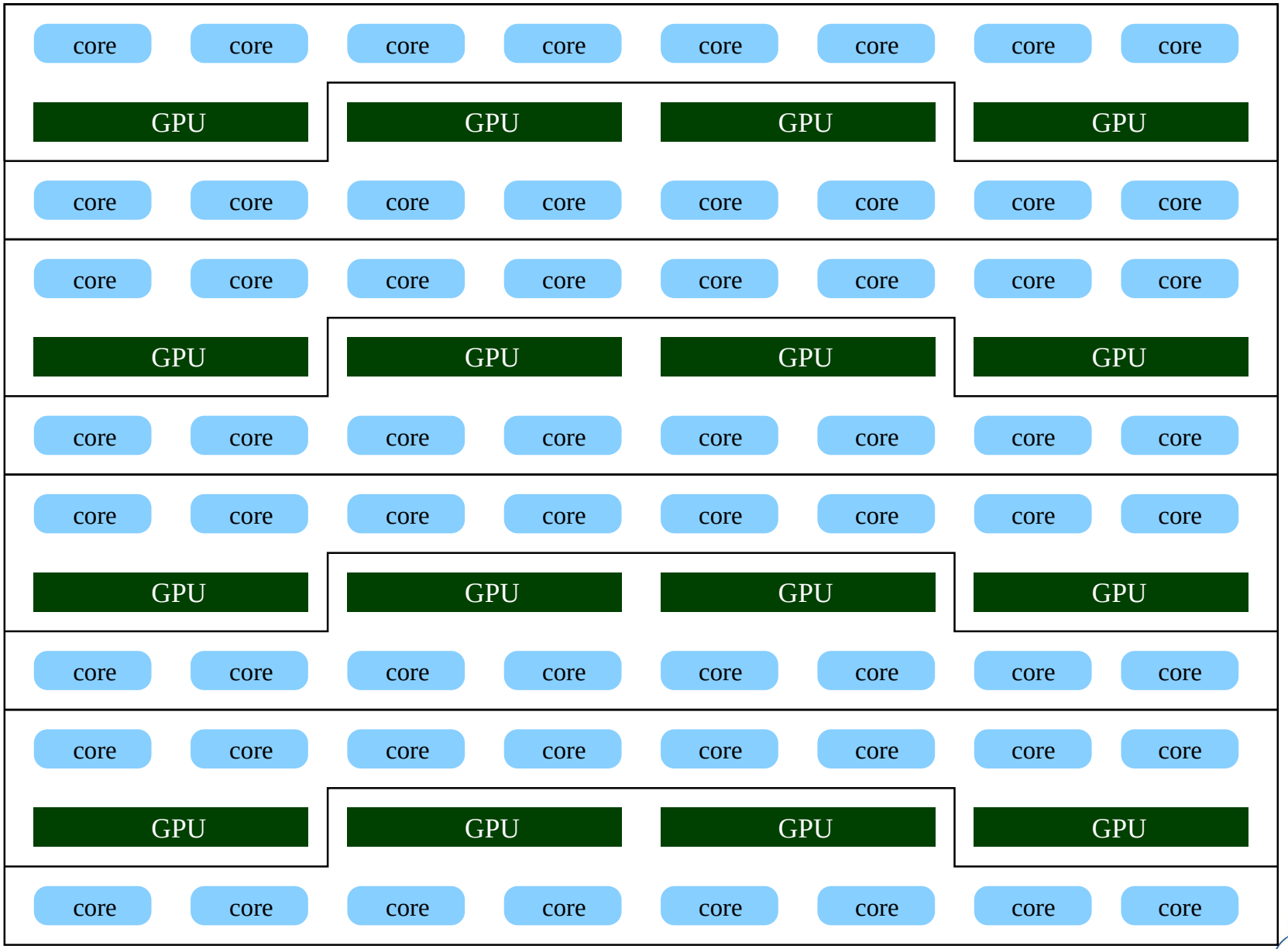
- 2 × quad-core Intel Xeon X5550 2.67GHz
- 48 GB RAM
- Interconnected by DDR InfiniBand
- 2 × GPUs with CUDA capability 2.0 (Fermi) each with 448 CUDA cores @ 1.15GHz and 6 GB of RAM.

Max. computing power CPUs: 683.52 GFlops
Max. computing power GPUs: 16.48 TFlops (single prec)
8.24 TFlops (double prec)

Access disabled by default: ask for it!

SciNet compute • calcul CANADA

Accelerator Research Cluster (ARC)



Part I

USING GPUS

Get on the system

1. Access systems: login.scinet.utoronto.ca

First ssh to login (not part of clusters):

```
ssh -l <username> login.scinet.utoronto.ca [-Y]
```

The login nodes are gateways, they are not part of any of the clusters and they are only to be used for small data transfer and to proceed logging into one of the devel nodes.

2. Go to the right cluster: ssh to the devel node arc01

SciNet compute • calcul CANADA

Software and Libraries

Once you log into devel nodes, what software is already installed?

- Other than essentials, all software installed as modules.
- modules set environment variables (`LD_LIBRARY_PATH`, `PATH`, ...)
- Allows multiple, conflicting versions of package to be available.
- `module load cuda/3.2`
  or
  `module load cuda/4.0`

```
arc01-$ module avail

------ /scinet/gpc/Modules6/version_indepen
3.2.8

------ /scinet/gpc/Modules6/3.2.8/modulefiles
dot                 modules             use.own
null                use.deprecated
module-info         use.experimental

------- /scinet/gpc/Modules6/modulefiles ---
R/2.13.1
Xlibraries/X11-64
...
cuda/3.2
cuda/4.0(default)
ddt/3.0(default)
intelmpi/4.0.2.003
openmpi/1.4.3-gcc-v4.6.1
...
vmd/1.9
```

http://wiki.scinet.utoronto.ca

SciNet compute • calcul CANADA

Compiling on the ARC

- The NVIDIA cuda compiler is available (3.2 and 4.0).
  You can `module load cuda/3.2` or `cuda/4.0`.
  The compiler is called `nvcc`.
- Optimize for the Tesla GPUs using the compiler flags

  ```
  -arch=sm_13 -O3
  ```

- OpenCL: Included as of version 3.0 in the CUDA Toolkit.
- Debuggers: cuda-gdb or ddt (`module load ddt`)
  To use the debuggers, compile with:

  ```
  -g -G
  ```

- MPI: The GPC MPI packages, i.e., openmpi and intelmpi, can be used on this system.

SciNet compute • calcul CANADA

ARC job scheduler

Submitting a job

- To run a job, you must submit to a batch queue.

```
qsub [options] <script>
```

- Common options:

  `-l`: specifies requested nodes and time, e.g.

```
-l nodes=1:gpus=2:ppn=8,walltime=1:00:00
```

  `-I` specifies that you want an interactive session.

- See submitted jobs with `qstat`.
- Limits:

time(hrs)	max jobs	max cores
48	32	64cpu / 7168 gpu

- Limits on this system may change.

SciNet compute • calcul CANADA

Links

Wiki

http://wiki.scinet.utoronto.ca
http://wiki.scinet.utoronto.ca/wiki/index.php/GPU_Devel_Nodes

Courses

https://support.scinet.utoronto.ca/courses
(12/12/11: Introduction to GPGPU with CUDA)

Gpgpu mailing list

https://support.scinet.utoronto.ca/mailman/listinfo/scinet-gpgpu

Technical support

support@scinet.utoronto.ca

SciNet compute • calcul CANADA