Introduction to High Performance Computing

Shaohao Chen

Research Computing Services (RCS)

Boston University

Outline

- What is HPC? Why computer cluster?
- Basic structure of a computer cluster
- Computer performance and the top 500 list
- HPC for scientific research and parallel computing
- National-wide HPC resources: XSEDE
- BU SCC and RCS tutorials

What is HPC?

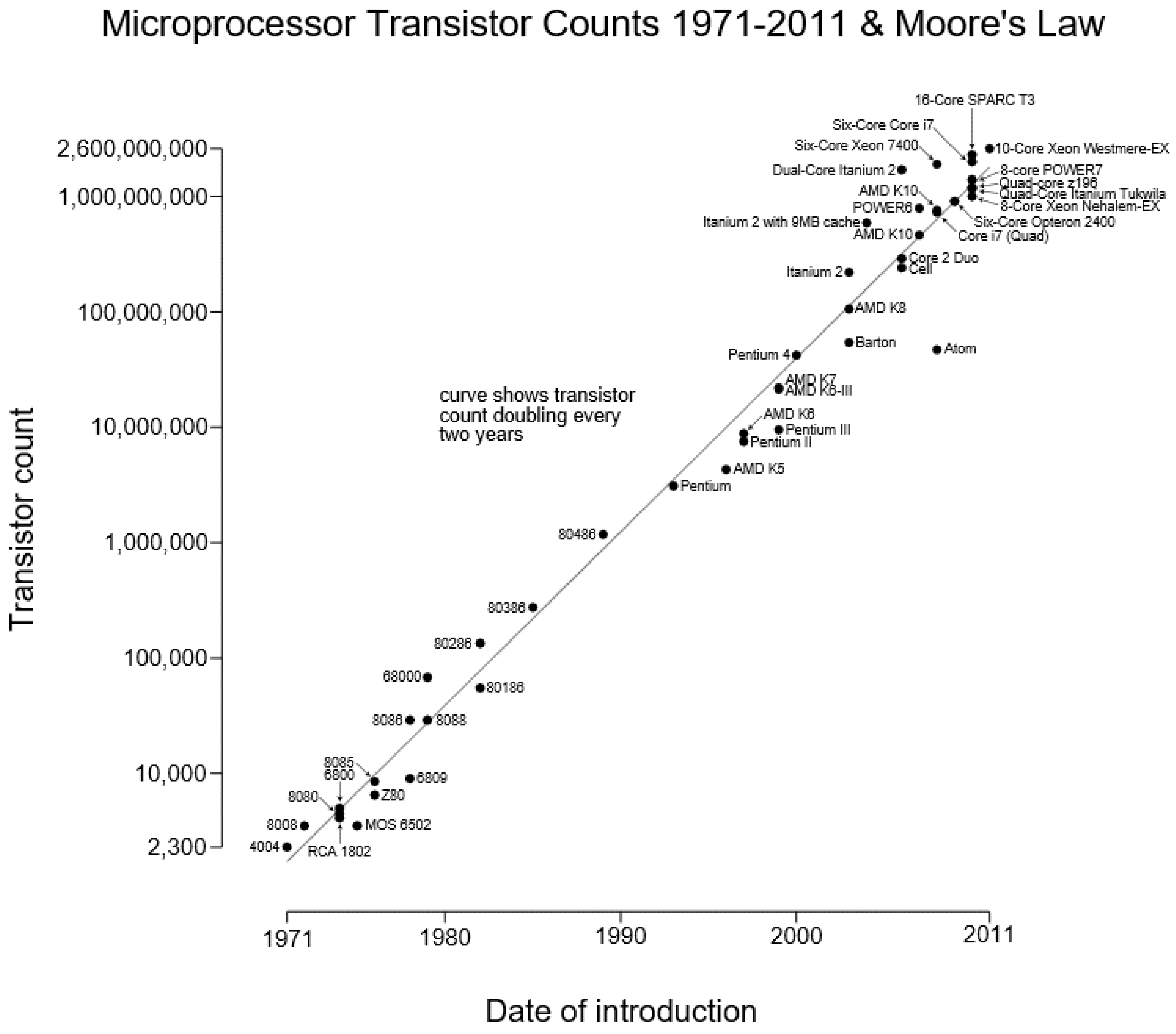
- High Performance Computing (HPC) refers to the practice of aggregating computing power in order to solve large problems in science, engineering, or business.
- Purpose of HPC: accelerates computer programs, and thus accelerates work process.
- Computer cluster: A set of connected computers that work together. They can be viewed as a single system.

- Similar terminologies: supercomputing, parallel computing.
- Parallel computing: many computations are carried out simultaneously, typically computed on a computer cluster.

- Related terminologies: grid computing, cloud computing.

Computing power of a single CPU chip

- Moore‘s law is the observation that the computing power of CPU doubles approximately every two years.
- Nowadays the multi-core technique is the key to keep up with Moore's law.



Why computer cluster?

- Drawbacks of increasing CPU clock frequency:

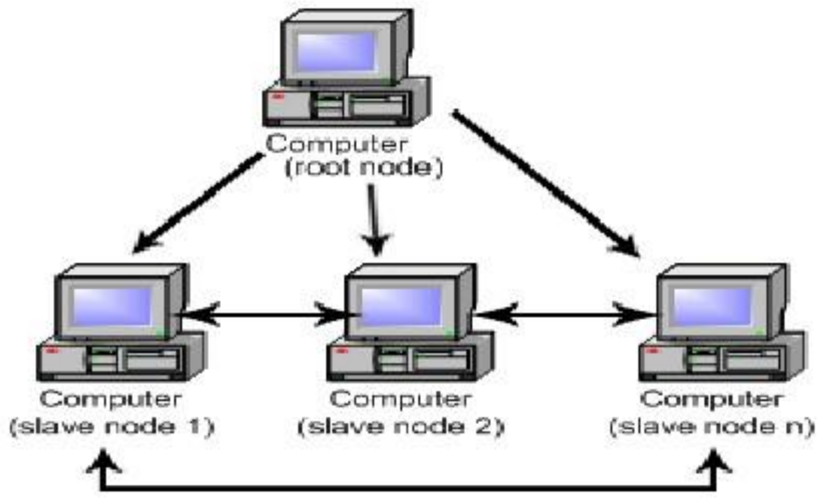
  --- Electric power consumption is proportional to the cubic of CPU clock frequency (ν^3).

  --- Generates more heat.

- A drawback of increasing the number of cores within one CPU chip:

  --- Difficult for heat dissipation.

- Computer cluster: connect many computers with high-speed networks.
- Currently computer cluster is the best solution to scale up computer power.
- Consequently software/programs need to be designed in the manner of parallel computing.



Basic structure of a computer cluster

- Cluster – a collection of many computers/nodes.
- Rack – a closet to hold a bunch of nodes.
- Node – a computer (with processors, memory, hard disk, etc.)
- Socket/processor – one multi-core processor.
- Core/processor – one actual processing unit.

- Network switch
- Storage system

- Power supply system
- Cooling system

- Figure: IBM Blue Gene supercomputer

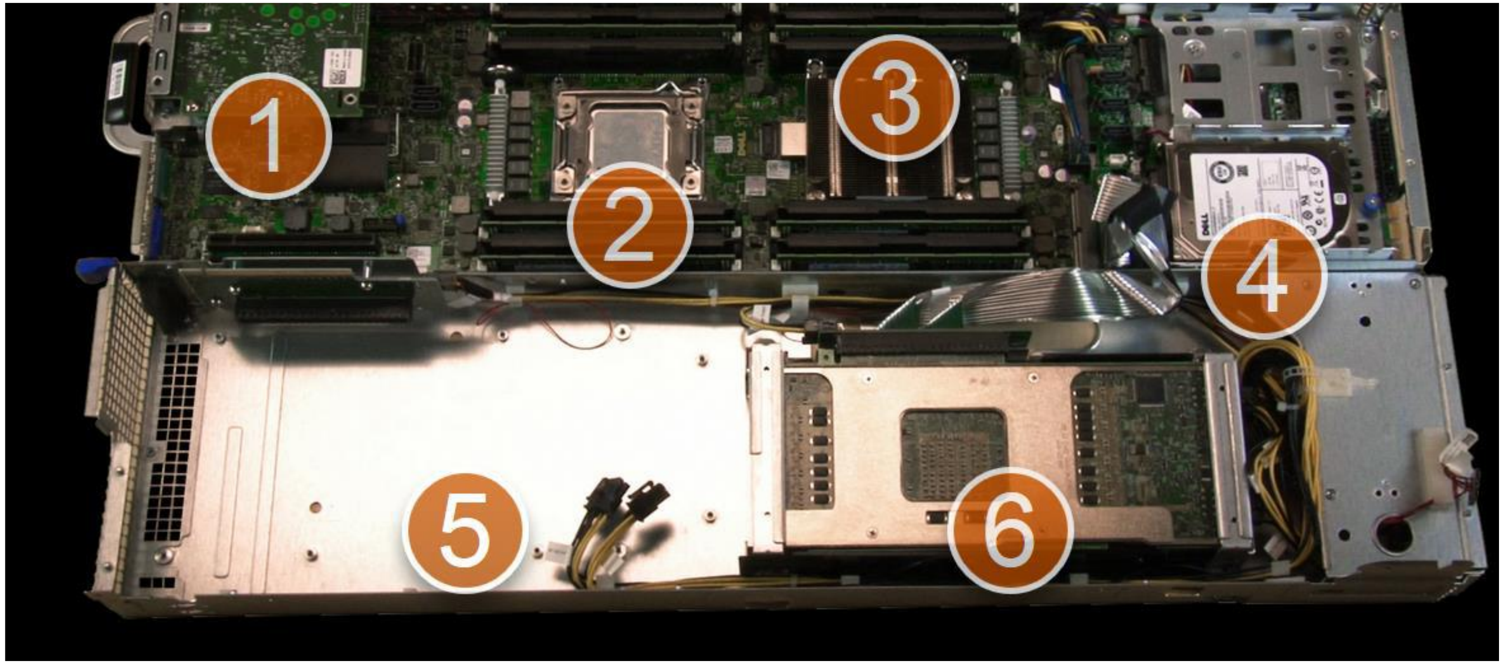
Inside a node

1. Network device --- Infiniband card:

   to transfers data between nodes.

2. CPU --- Xeon multi-core processors:

   to carry out the instructions of programs.



3. Memory:

fast and temporal storage,

to store data for immediate use.

4. Hard disk:

slow and permanent storage

to store data permanently.

5. Space for possible upgrade

6. Accelorator --- Intel Xeon Phi Coprocessor (Knights Corner):

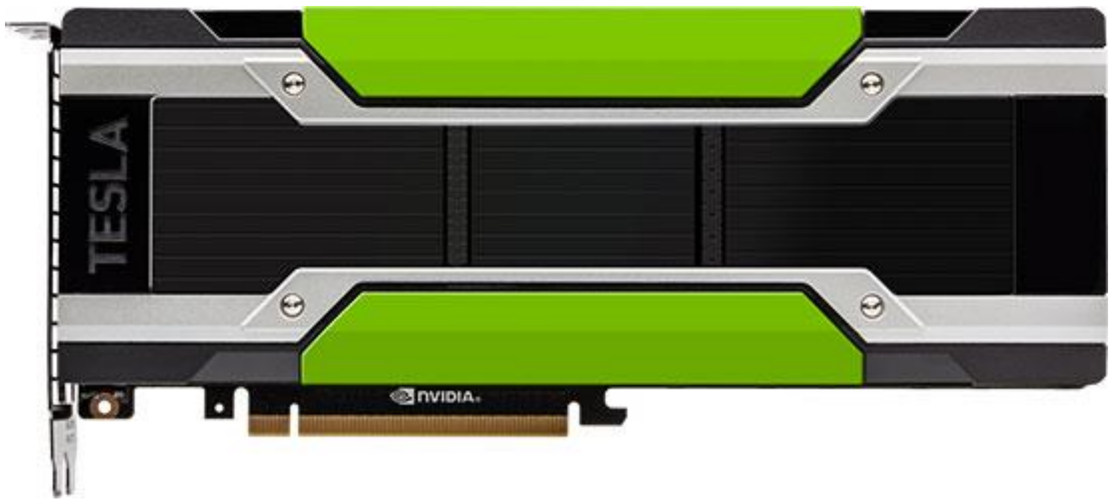
to accelerate programs.

- Figure: A node of the supercomputer Stampede at TACC.

Accelerators

❑ NVIDA GPU (Tesla P100):

- Multiprocessors: 56
- CUDA cores: 3584
- Memory: 12 GB
- PCI connection to host CPU
- Peak DP compute: 4036–4670 GFLOPS



❑ Intel Xeon Phi MIC processor (Knights Landing):

- Cores: 68; Threads: 272
- Frequency: 1.4 GHz; Two 512-bit VPUs
- Memory: 16 GB MCDRAM + external RAM
- Self-hosted
- Peak DP compute: 3046 GFLOPS

What resources does an HPC system provide?

- A large number of compute nodes and cores.
- Large-size (~ TB) and high-bandwidth memory.
- Large-size (~ PB) and fast-speed storage system; storage for parallel I/O.
- High-speed network: high-bandwidth Ethernet, Infiniband, Omni Path, etc.
- Graphic Processor Unit (GPU)
- Xeon Phi many-integrated-core (MIC) processor/coprocessor.

- A stable and efficient operation system.
- A large number of software or applications.
- User services.

How to measure computer performance?

- Floating-point operations per second (FLOPS):

$$FLOPS = nodes \times \frac{cores}{nodes} \times \frac{cycles}{second} \times \frac{FLOPs}{cycle}$$

- The 3rd term clock cycles per second is known as the clock frequency, typically 2 ~ 3 GHz.
- The 4th term FLOPs per cycle is how many floating-point operations are done in one clock cycle.

  Typical values for Intel Xeon CPUs are:

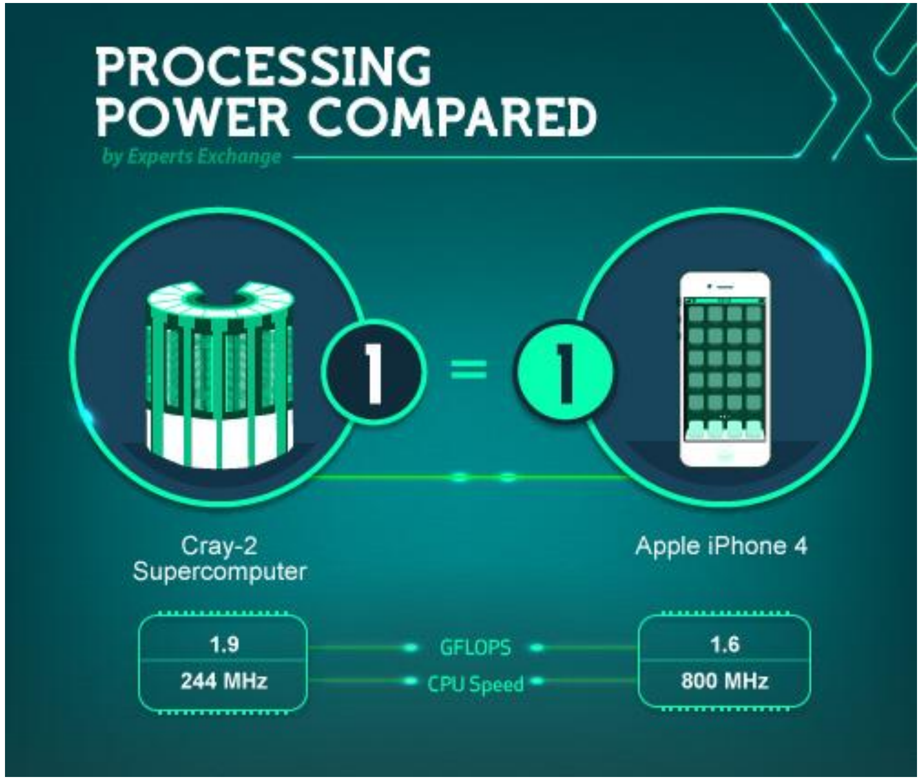
  --- Sandy Bridge and Ivy Bridge: 8 DP FLOPs/cycle, 16 SP FLOPs/cycle.

  --- Haswell and Broadwell : 16 DP FLOPs/cycle, 32 SP FLOPs/cycle.

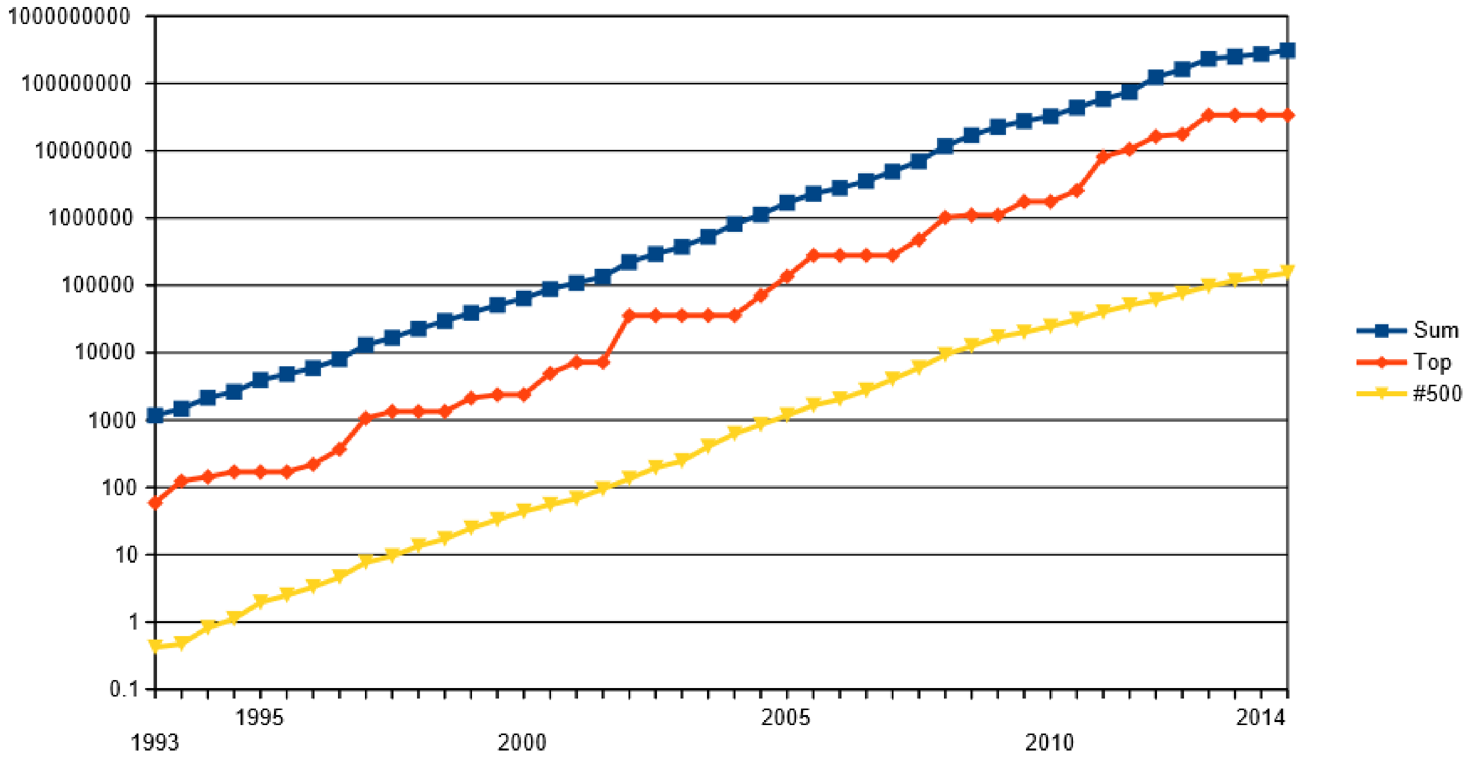
- GigaFLOPS – 10^9 FLOPS; TeraFLOPS – 10^{12} FLOPS; PetaFLOPS – 10^{15} FLOPS; ExaFLOPS – 10^{18} FLOPS.

Computer power grows rapidly

- Iphone 4 *vs.* 1985 Cray-2 supercomputer



- Rapid growth of the power of the top-500 supercomputers (logarithmic y-axis, in GFLOPS)

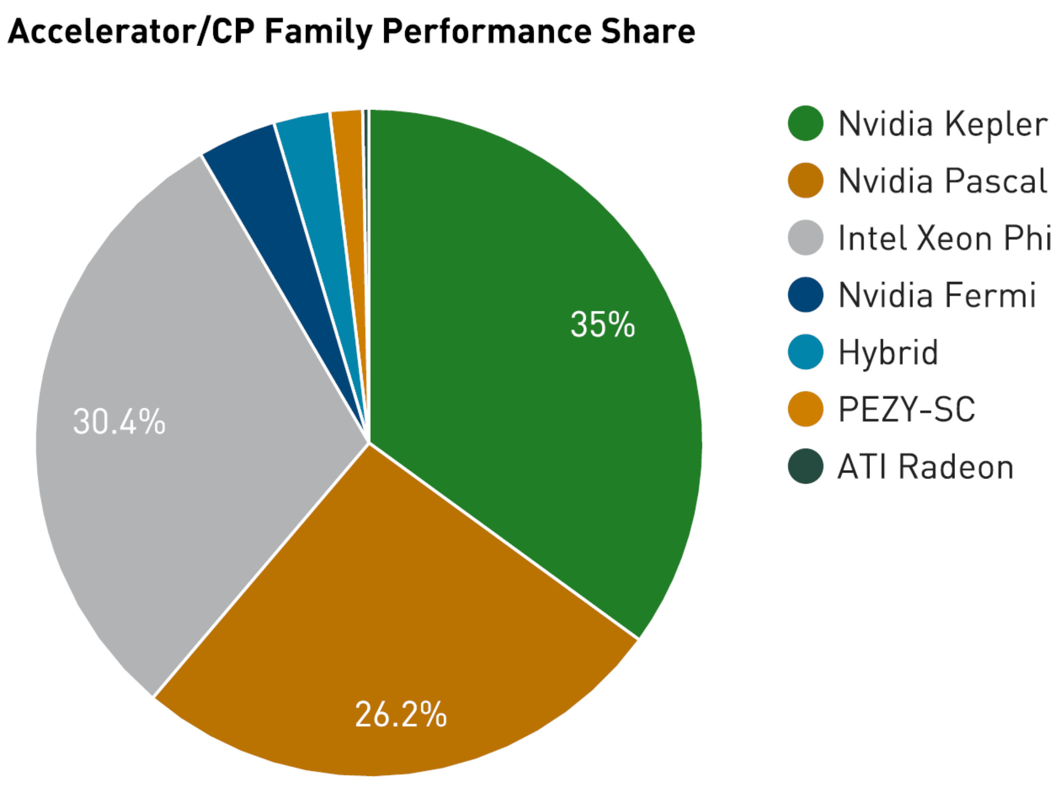


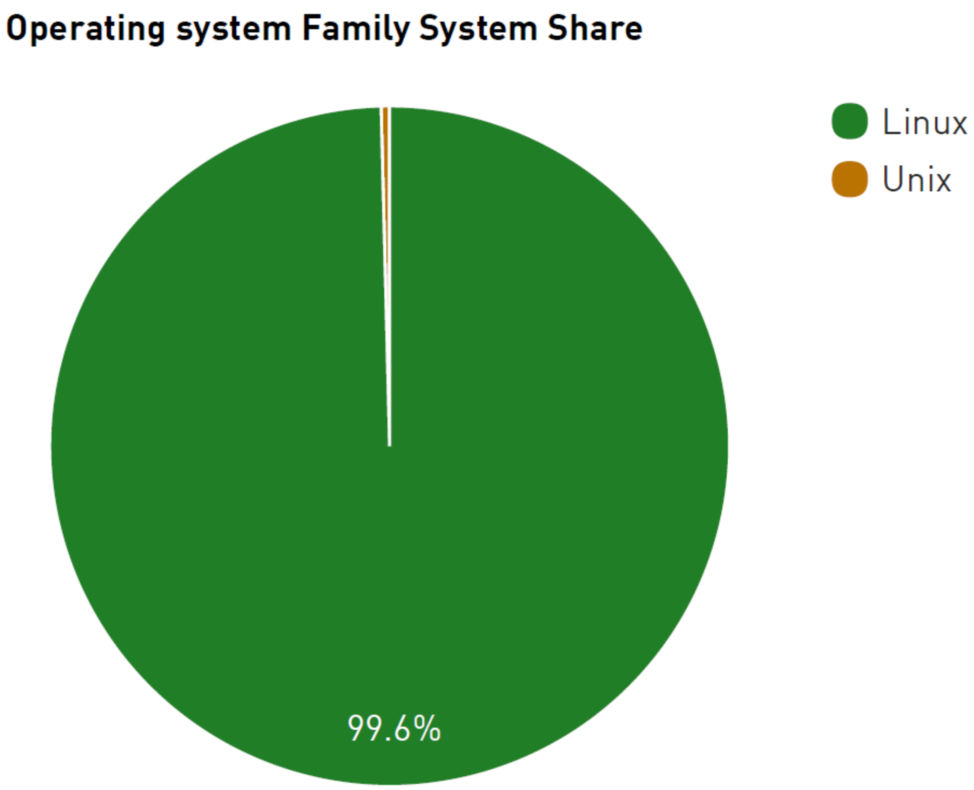
Top 500 Supercomputers

- The list of June 2017

Rank	Rmax Rpeak (PFLOPS)	Name	Model	Processor	Interconnect	Vendor	Site country, year
1	93.015 125.436	*Sunway TaihuLight*	Sunway MPP	SW26010	Sunway[13]	NRCPC	National Supercomputing Center in Wuxi China, 2016[13]
2	33.863 54.902	*Tianhe-2*	TH-IVB-FEP	Xeon E5–2692, Xeon Phi 31S1P	TH Express-2	NUDT	National Supercomputing Center in Guangzhou China, 2013
3	19.590 25.326	*Piz Daint*	Cray XC50	Xeon E5-2690v3, Tesla P100	Aries	Cray	Swiss National Supercomputing Centre Switzerland, 2016
4	17.590 27.113	*Titan*	Cray XK7	Opteron 6274, Tesla K20X	Gemini	Cray	Oak Ridge National Laboratory United States, 2012
5	17.173 20.133	*Sequoia*	Blue Gene/Q	A2	Custom	IBM	Lawrence Livermore National Laboratory United States, 2013
6	14.015 27.881	*Cori*	Cray XC40	Xeon Phi 7250	Aries	Cray	National Energy Research Scientific Computing Center United States, 2016
7	13.555 24.914	*Oakforest-PACS*	Fujitsu	Xeon Phi 7250	Intel Omni-Path	Fujitsu	Kashiwa, Joint Center for Advanced High Performance Computing Japan, 2016

Statistics of the Top 500





HPC user environment

- Operation system: Linux (Redhat/CentOS, Ubuntu, etc), Unix.
- Login: ssh, gsissh.
- File transfer: secure ftp (scp), grid ftp (globus).
- Job scheduler: Slurm, PBS, SGE, Loadleveler.
- Software management: module.
- Compilers: Intel, GNU, PGI.
- MPI implementations: OpenMPI, MPICH, MVAPICH, Intel MPI.
- Debugging and profiling tools: Totalview, Tau, DDT, Vtune.

- Programming Languages: C, C++, Fortran, Python, Perl, R, MATLAB, Julia

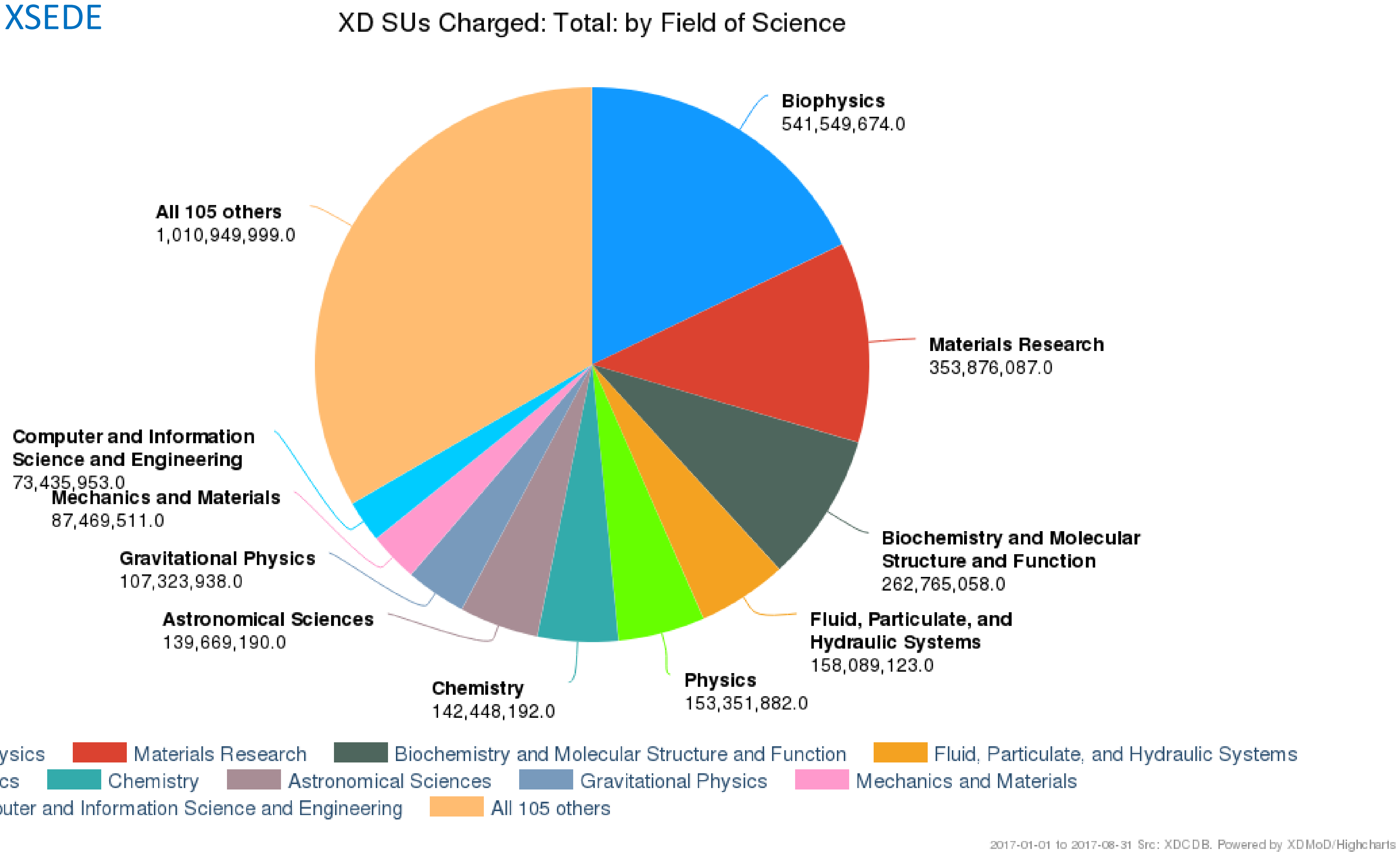
Scientific disciplines in HPC

❑ Typical scientific computing catalogs that can benefit from HPC:

- Computational Physics
- High-energy physics
- Astrophysics
- Geophysics
- Climate and weather science
- Computational fluid dynamics
- Computer aided engineering
- Material sciences
- Computational chemistry
- Molecular dynamics
- Linear algebra
- Computer science
- Data science
- Machine/deep learning
- Biophysics
- Bioinformatics
- Finance informatics
- Scientific Visualization
- Social sciences

CPU-hours by field of science

- Statistics from XSEDE



Scientific computing software

- Numerical Libraries: Lapack/Blas, FFTw, MKL, GSL, PETSc, Slepc, HDF5, NetCDF, Numpy, Scipy.
- Physics and Engineering: BerkeleyGW, Root, Gurobi, Abaqus, Openfoam, Fluent, Ansys, WRF
- Chemistry and material science: Gaussian, NWChem, VASP, QuantumEspresso, Gamess, Octopus
- Molecular dynamics: Lammps, Namd, Gromacs, Charmm, Amber
- Bioinformatics: Bowtie, BLAST, Bwa, Impute, Minimac, Picard, Plink, Solar, Tophat, Velvet.
- Data science and machine learning: Hadoop, Spark, Tensorflow, Caffe, Torch, cuDNN, Scikit-learn.

➢ XSEDE software: https://portal.xsede.org/software/

➢ BU SCC software: http://sccsvc.bu.edu/software/

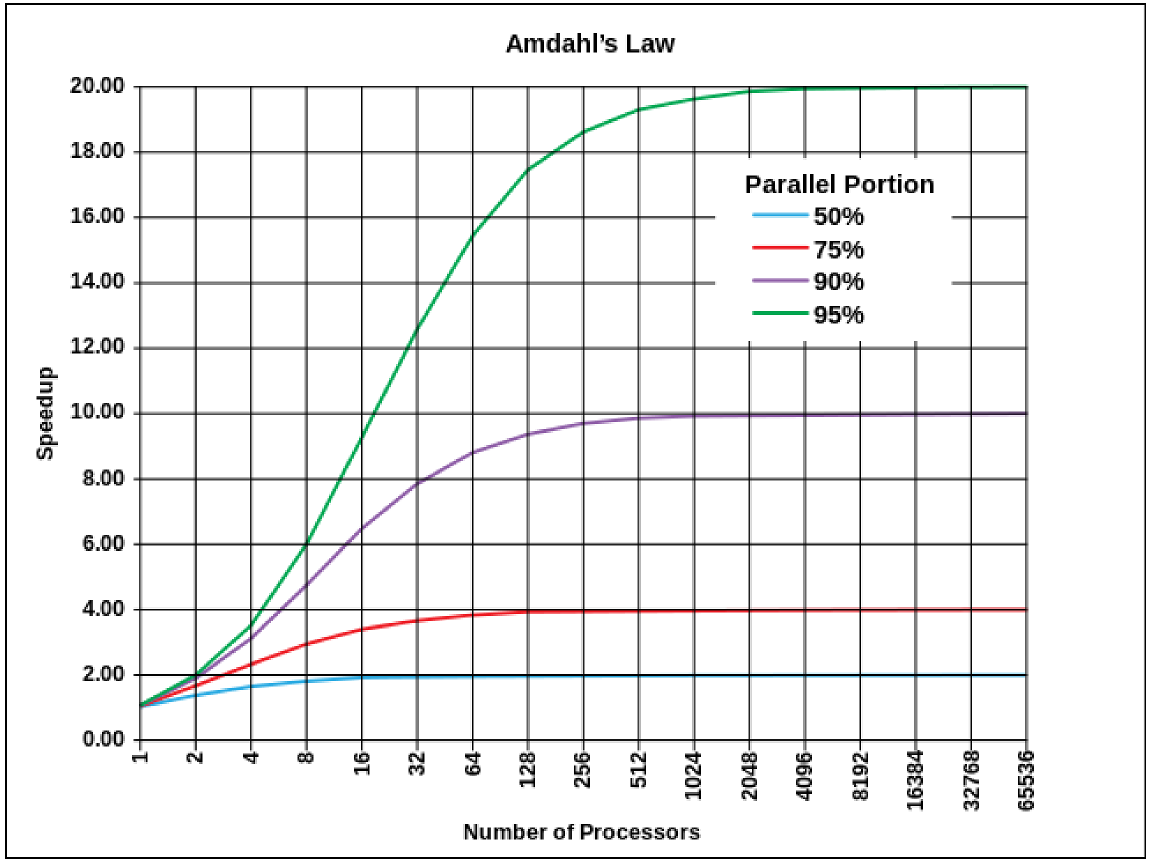
Parallel Computing

- Parallel computing is a type of computation in which many calculations are carried out simultaneously, based on the principle that large problems can often be divided into smaller ones, which are then solved at the same time.
- Speedup of a parallel program,

$$S(p) = {}^{T(1)}/_{T(p)} = \frac{1}{\alpha + {}^{1}/_{p}\,(1-\alpha)}$$

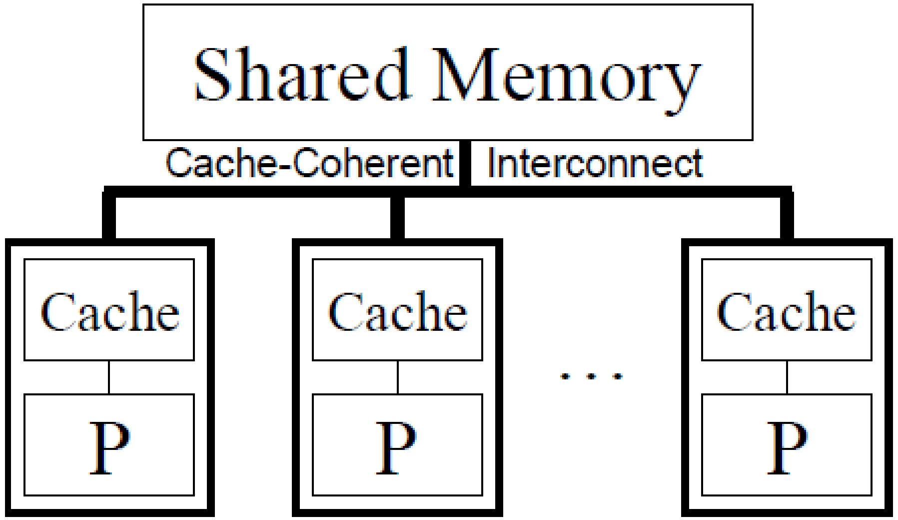
p: number of processors/cores,

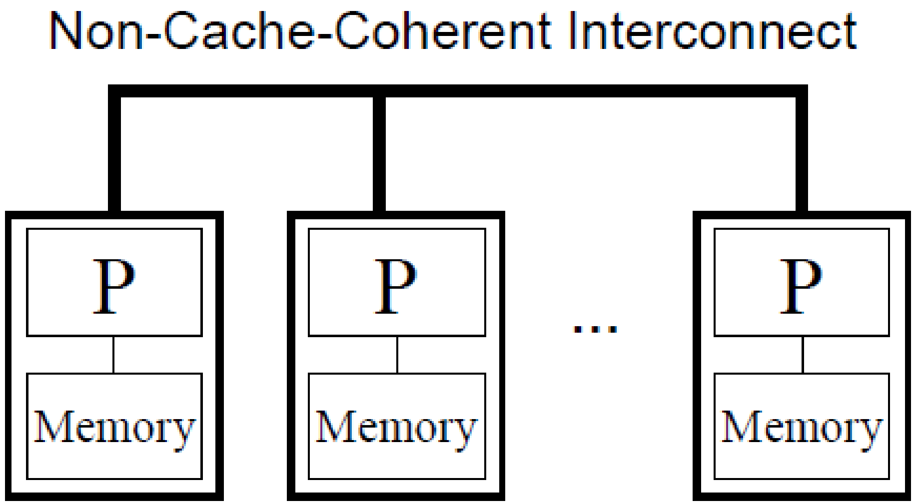
α: fraction of the program that is serial.



- The figure is from: https://en.wikipedia.org/wiki/Parallel_computing

Distributed or shared memory systems





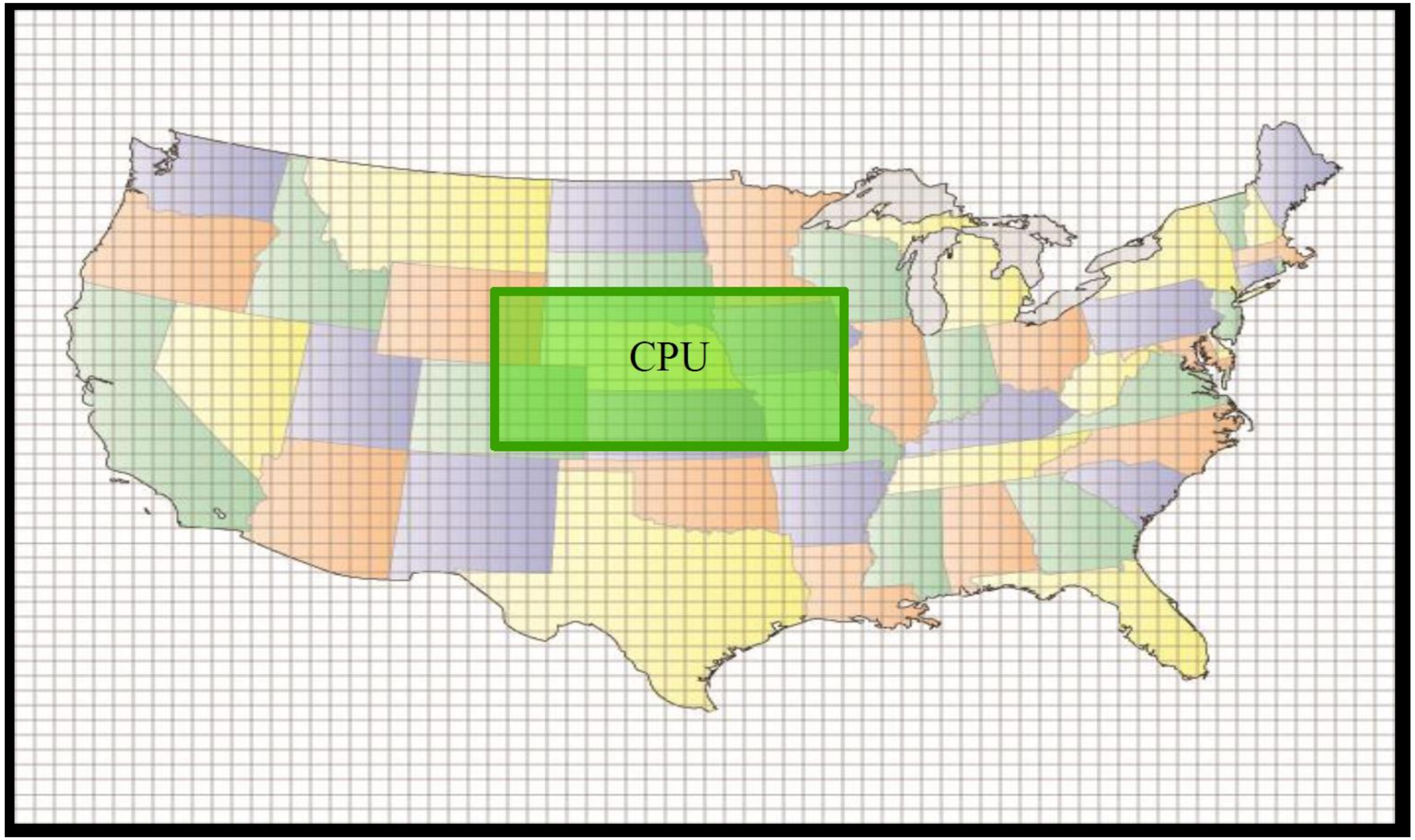
- Shared memory system
- For example, a single node on a cluster
- Open Multi-processing (OpenMP) or MPI

- Distributed memory system
- For example, multi-nodes on a cluster
- Message Passing Interface (MPI)

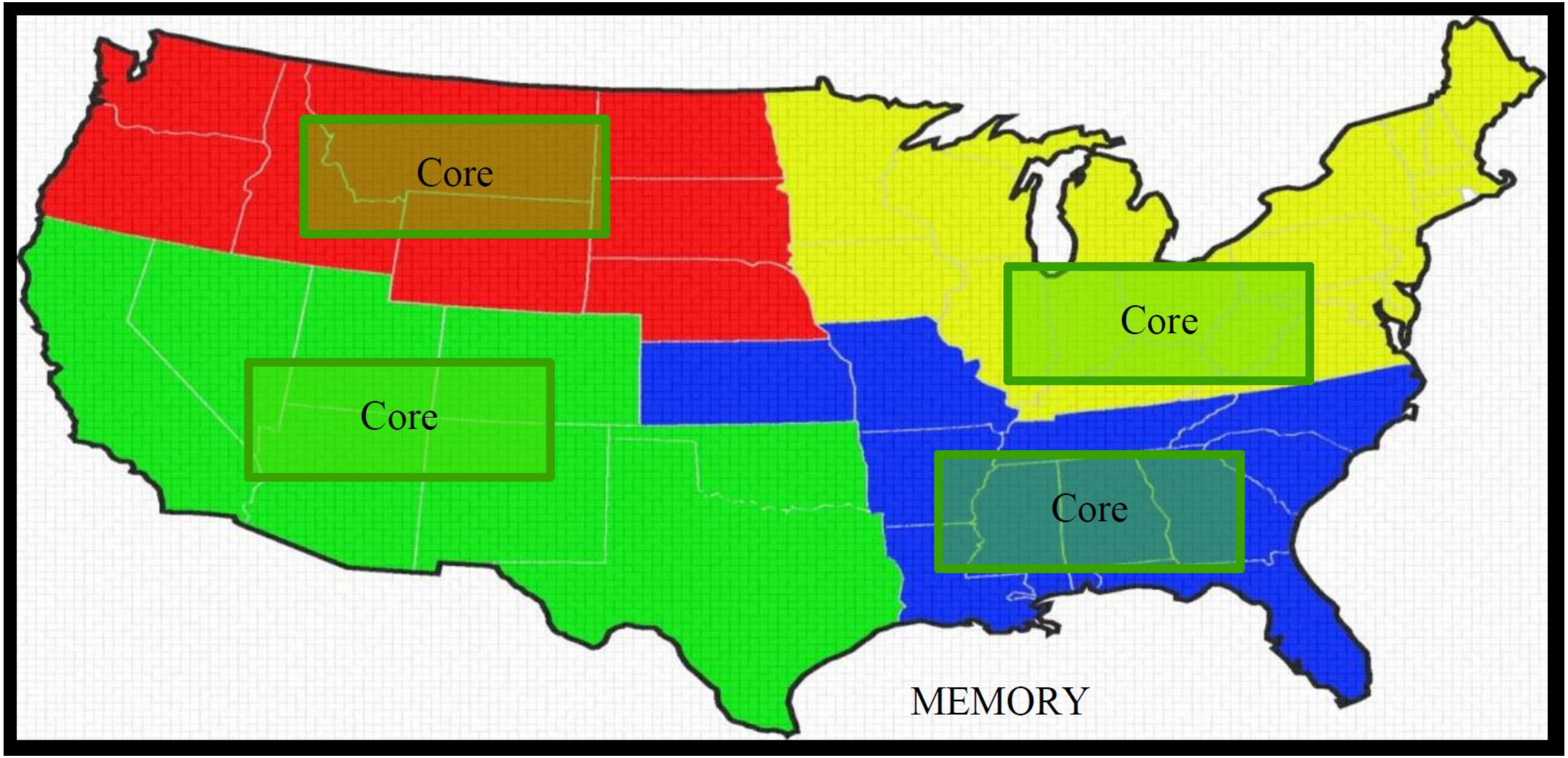
✓ Figures are from the book *Using OpenMP: Portable Shared Memory Parallel Programming*

An example: weather science

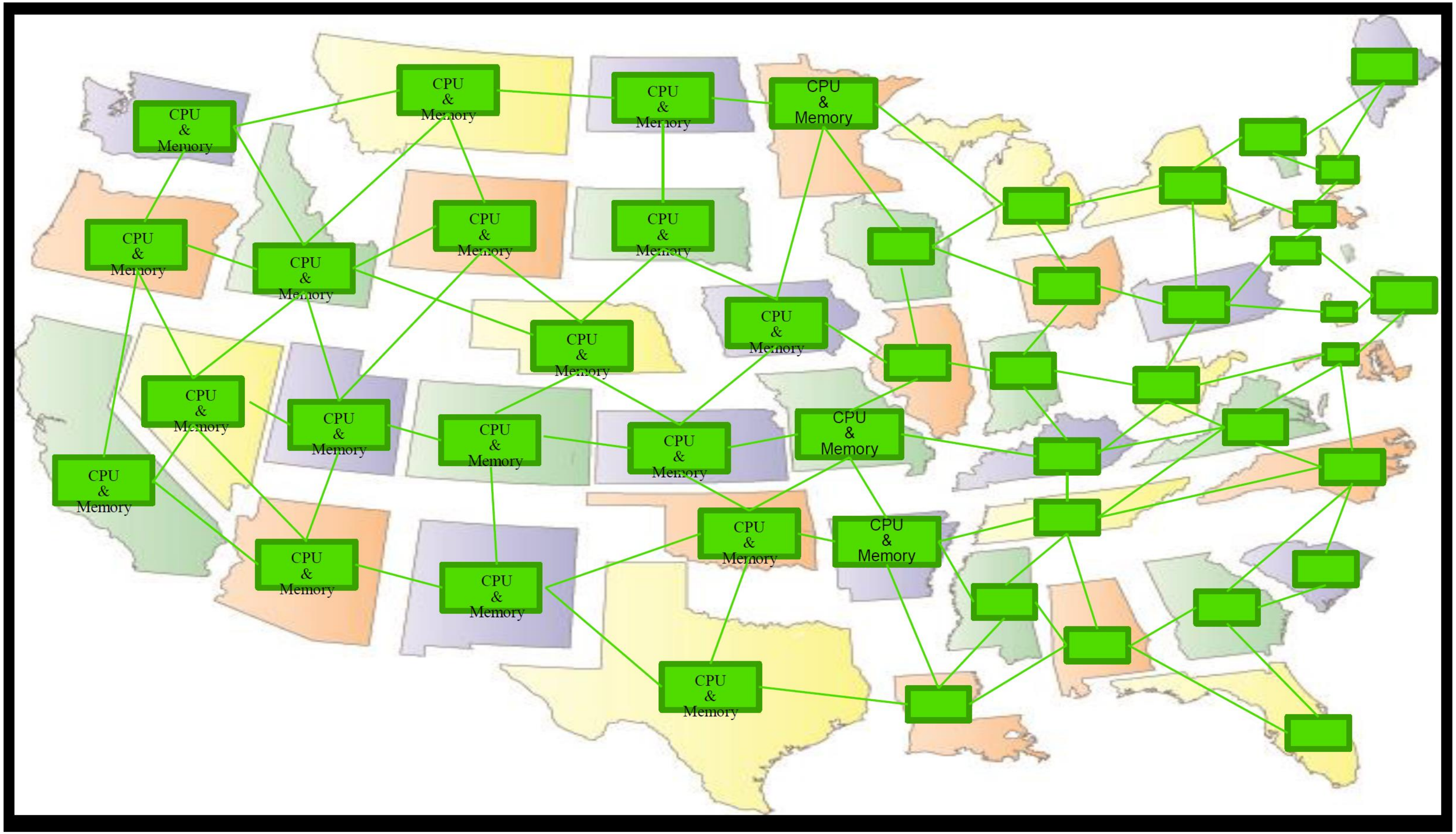
- Serial weather model



- Shared-memory weather model (for several cores within one node)



- Distributed-memory weather model (for many nodes within one cluster)

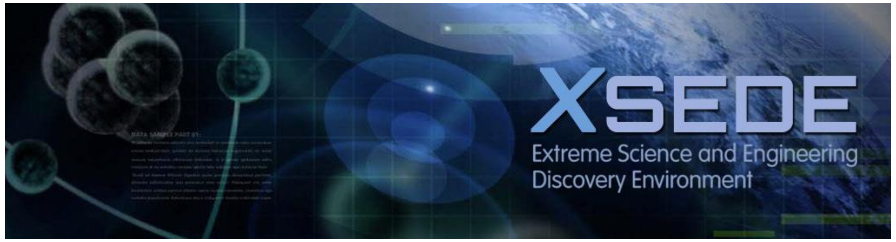


National-wide HPC resources: XSEDE

- XSEDE (eXtreme Science and Engineering Discovery Environment) is a virtual system that provides compute resources for scientists and researchers from all over the US.
- Its mission is to facilitate research collaboration among institutions, enhance research productivity, provide remote data transfer, and enable remote instrumentation.
- A combination of supercomputers in many institutions in the US.
- Available to BU users. How to apply for an XSEDE account and allocations? See details at http://www.bu.edu/tech/support/research/computing-resources/external/xsede/ .
- XSEDE provides regular HPC trainings and workshops:

--- online training: https://www.xsede.org/web/xup/online-training

---- monthly workshops: https://www.xsede.org/web/xup/course-calendar



XSEDE resources (1)

Machine Name	Resource Provider	Best Types of Computation	Resource Highlights
Bridges	Pittsburgh Supercomputing Center (PSC).	Good for MPI, OpenMP, or GPU jobs. Especially good for large-memory jobs.	Large-memory (3 TB) and extremely-large-memory (12 TB) nodes.
Comet	San Diego Supercomputing Center (SDSC).	Good for MPI, OpenMP, or GPU jobs. Supports virtual-machine jobs too.	Intel Haswell processors; GPU nodes; Virtual Machine repository.
Greenfield	Pittsburgh Supercomputing Center (PSC).	Good for shared-memory (such as OpenMP) jobs.	Giant compute nodes with around one hundred cores and around 10 TB memory on each.

XSEDE resources (2)

Machine Name	Resource Provider	Best Types of Computation	Resource Highlights
Jetstream	Indiana University (IU) and Texas Advanced Computing Center (TACC)	Particularly for cloud computing.	User-friendly cloud environment.
Maverick	Texas Advanced Computing Center (TACC).	Particularly for visualization jobs	VNC server; GPU and large memory nodes for visualization.
Stampede	Texas Advanced Computing Center (TACC).	The largest cluster among all XSEDE resources; Good for massive MPI or OpenMP jobs.	Thousands of compute nodes with 1 or 2 Intel Xeon-Phi/MIC coprocessors on each; GPU nodes; large-memory nodes.

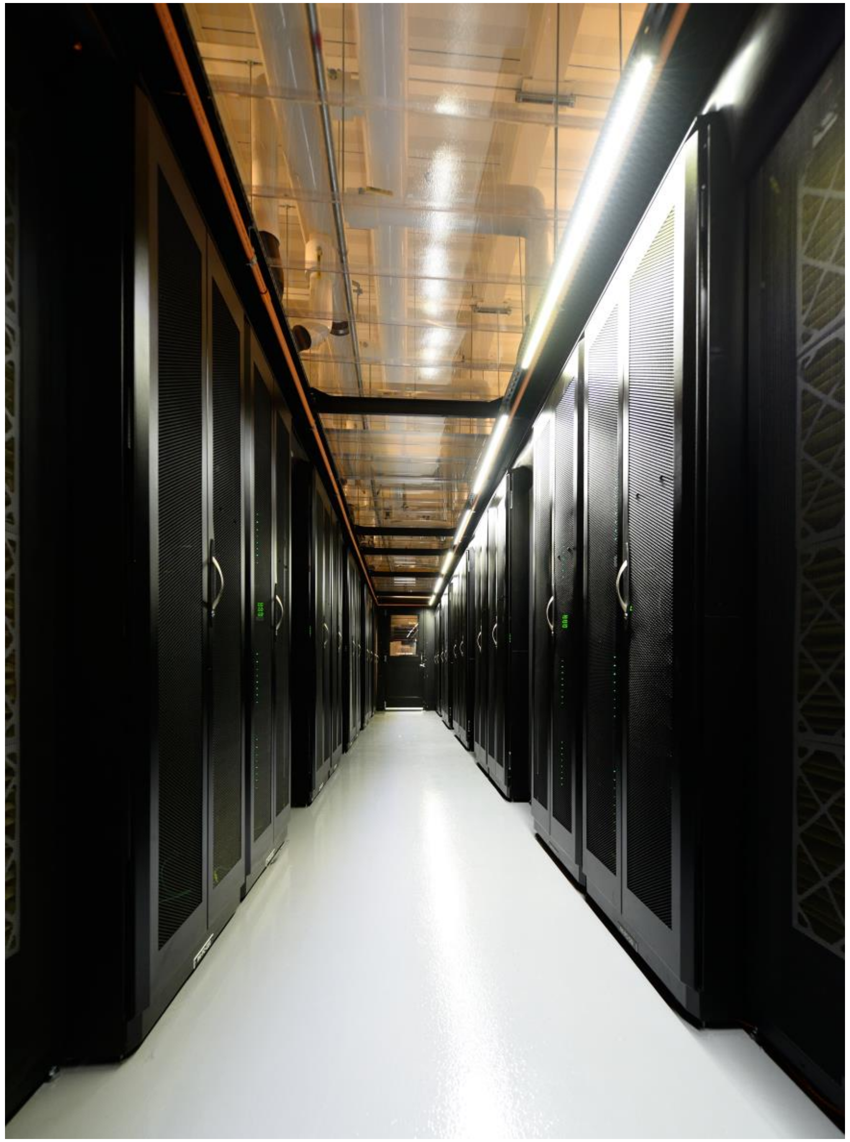
XSEDE resources (3)

Machine Name	Resource Provider	Best Types of Computation	Resource Highlights
SuperMIC	Louisiana State University (LSU).	Good for MPI or OpenMP jobs.	Hundreds of compute nodes with 2 Intel Xeon-Phi/MIC coprocessors on each; large-memory nodes.
XStream	Stanford Research Computing Center (SRCC)	Particularly for GPU jobs.	Tens of compute nodes with 8 K80 24GB GPU cards on each; A lot of machine/deep learning platforms.
Open Science Grid	Over 100 individual sites spanning the United States.	Good for distributed high throughput computing (DHTC).	Virtual cluster environment

BU Shared Computer Cluster (SCC)

- A Linux cluster with over 580 nodes, 11,000 processors, and 252 GPUs. Currently over 3 Petabytes of disk.
- Located in Holyoke, MA at the Massachusetts Green High Performance Computing Center (MGHPCC), a collaboration between 5 major universities and the Commonwealth of Massachusetts.
- Went into production in June, 2013 for Research Computing. Continues to be updated/expanded.

- Webpage: http://www.bu.edu/tech/support/research/computing-resources/scc/



BU RCS tutorials (1)

❑ Linux system:

- Introduction to Linux
- Build software from Source Codes in Linux

❑ BU SCC:

- Introduction to SCC
- Intermediate Usage of SCC
- Managing Projects on the SCC

❑ Visualization:

- Introduction to Maya
- Introduction to ImageJ

❑ Mathematics and Data Analysis:

- Introduction to R
- Graphics in R
- Programming in R
- R Code Optimization
- Introduction to MATLAB
- Introduction to SPSS
- Introduction to SAS
- Python for Data Analysis

BU RCS tutorials (2)

- Computer programming:
  - Introduction to C
  - Introduction to C++
  - Introduction to Python
  - Introduction to Python for Non-programmers
  - Introduction to Perl
  - Version Control and GIT.

- High-performance computing:
  - Introduction to MPI
  - Introduction to OpenMP
  - Introduction to GPU
  - Introduction to CUDA
  - Introduction to OpenACC
  - MATLAB for HPC
  - MATLAB Parallel Tool Box.

- Upcoming tutorials: http://www.bu.edu/tech/about/training/classroom/rcs-tutorials/
- Tutorial documents: http://www.bu.edu/tech/support/research/training-consulting/live-tutorials/