

# 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
- Nation-wide HPC resources: XSEDE
- BU Shared Computing Cluster (SCC) and RCS tutorials

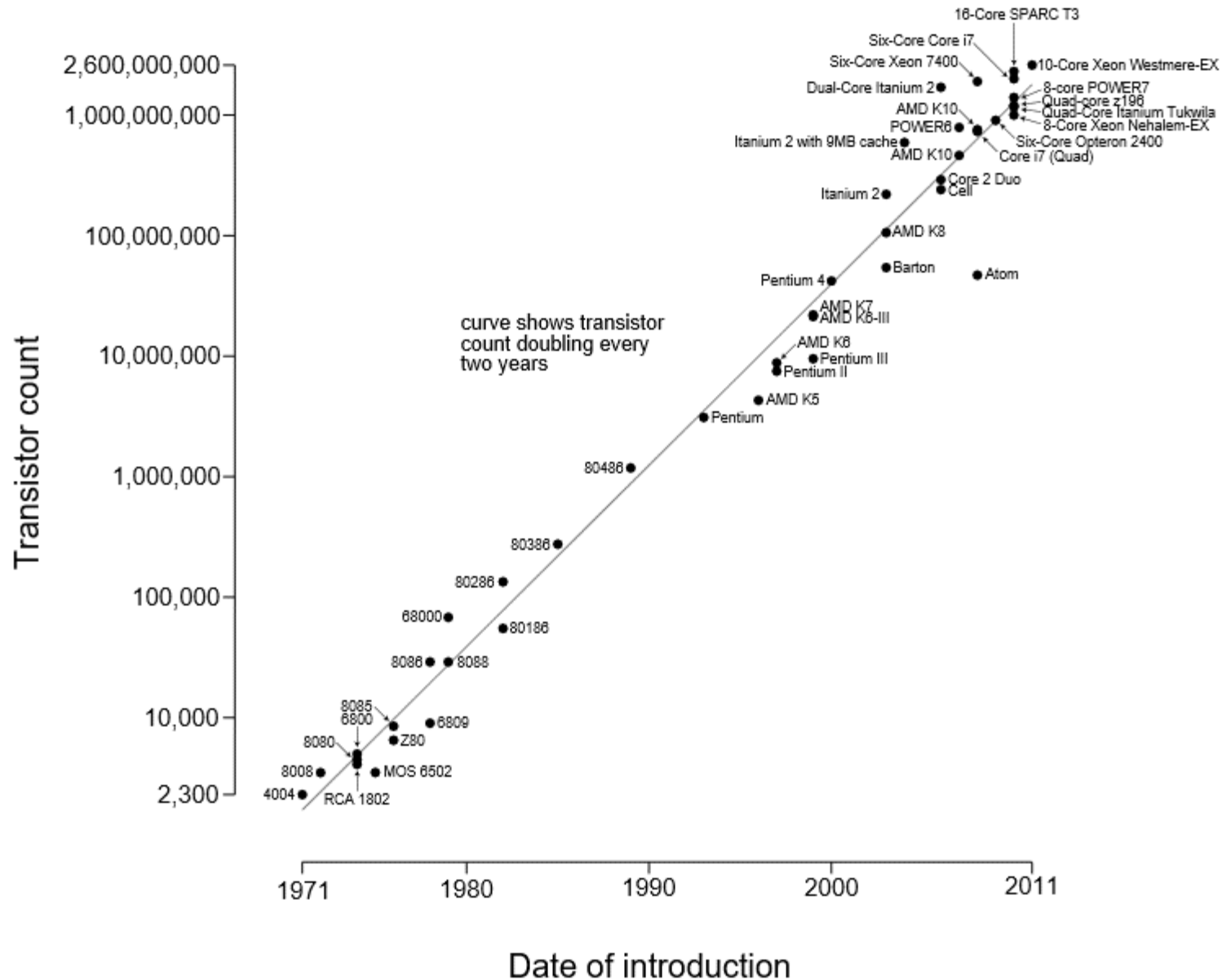
# What is HPC?

- High Performance Computing (HPC) refers to the practice of aggregating computing power in order to solve large problems in science, engineering, or business.
- The purpose of HPC: accelerate computer programs, and thus accelerate 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

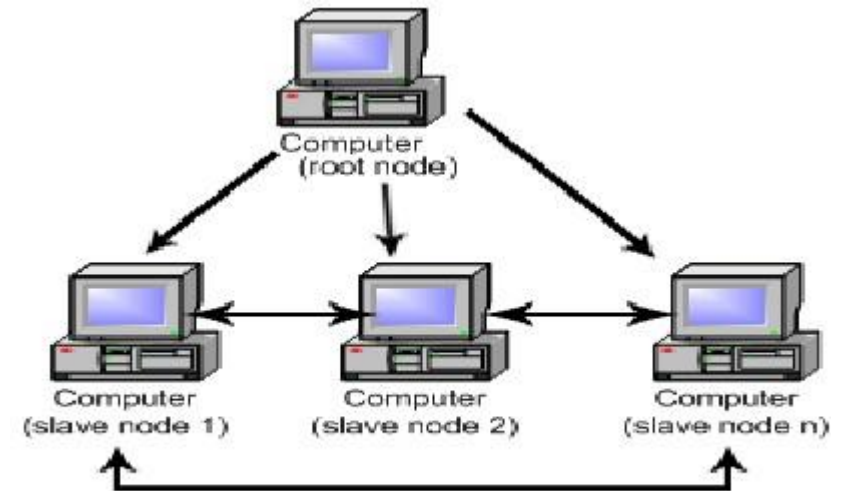
- **Moore's law** is the observation that the computing power of CPU doubles approximately every two years.
- Nowadays the **multi-core** technique is the key to keep up with Moore's law.

Microprocessor Transistor Counts 1971-2011 & Moore's Law



# Why computer cluster?

- Drawbacks of increasing CPU clock frequency:
  - Electric power consumption is proportional to the cubic of CPU clock frequency ( $v^3$ ).
  - 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 computing 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 processing unit.
  
- Network switch
- Storage system
- Power supply system
- Cooling system

■ Figure: IBM Blue Gene supercomputer





# Inside a node

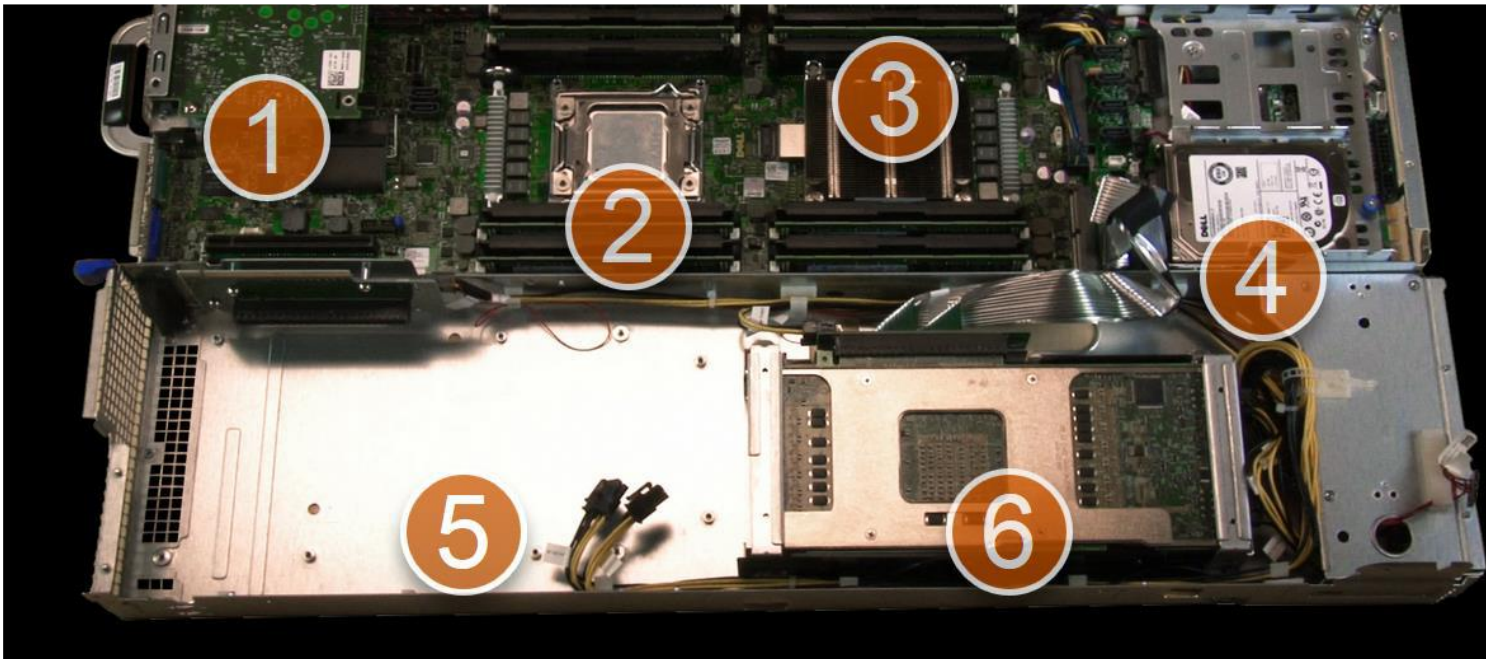
1. **Network device** --- Infiniband card:  
to transfers data between nodes.
2. **CPU** --- Xeon multi-core processors:  
to carry out the instructions of programs.

3. **Memory:**  
fast but temporary storage,  
to store data for immediate use.

4. **Hard disk:**  
slow but permanent storage  
to store data permanently.

5. **Space for possible upgrade**
6. **Accelerator** --- Intel Xeon Phi Coprocessor (Knights Corner):  
to accelerate programs.

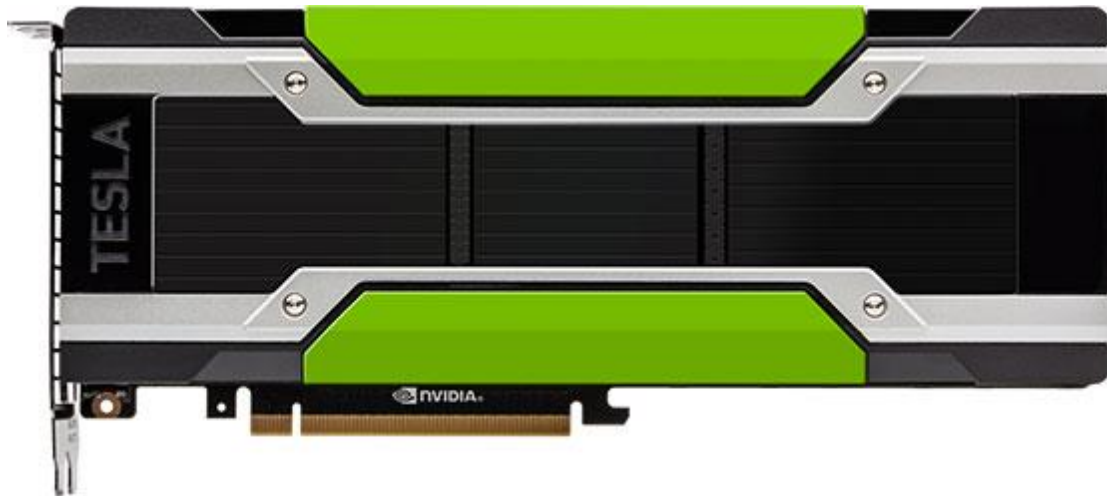
- Figure: A node of the supercomputer Stampede at TACC.



# Accelerators

## ❑ NVIDIA 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





# Typical resources in an HPC system

- 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 applications.
  - User services.

# How to measure computer performance?

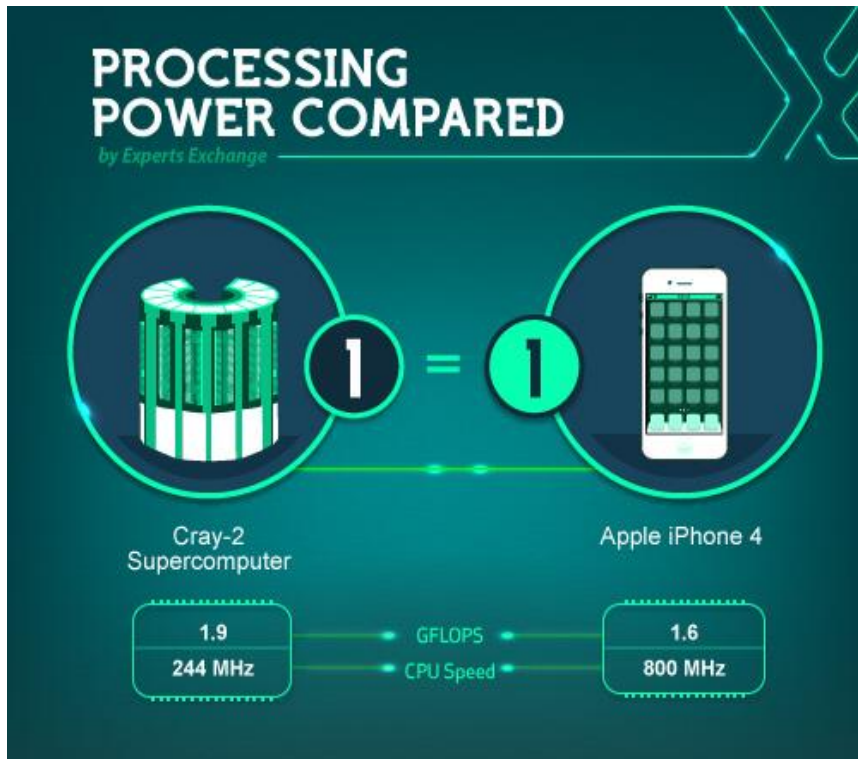
- Floating-point operations per second (FLOPS):

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

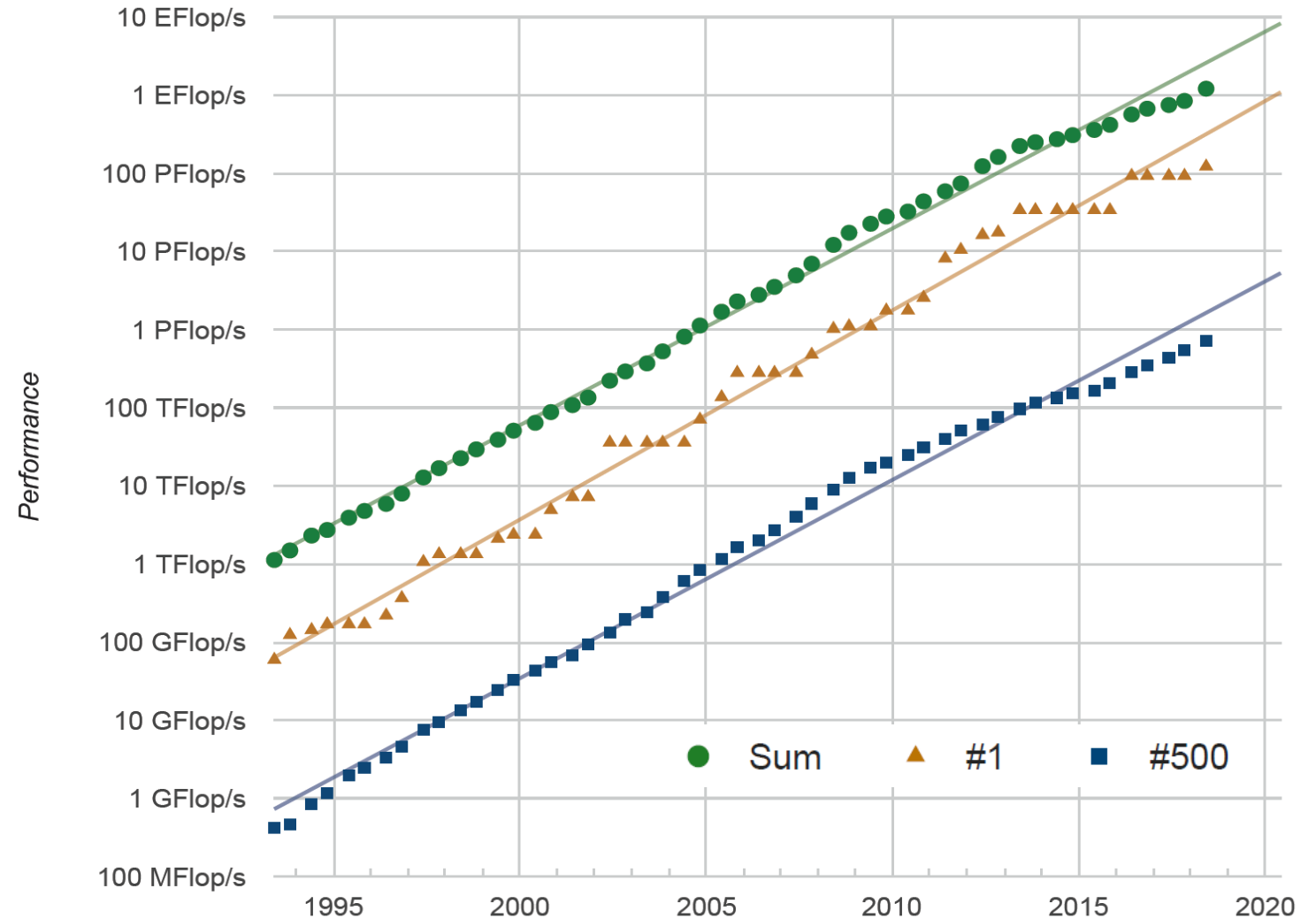
- The 3rd term clock cycles per second is known as the **clock frequency**, typically 2 ~ 3 GHz.
- The 4th term **FLOPs per cycle**: how many floating-point operations are done in one clock cycle.  
Typical values for Intel Xeon CPUs: 16 DP FLOPs/cycle, 32 SP FLOPs/cycle.
- **GigaFLOPS** –  $10^9$  FLOPS; **TeraFLOPS** –  $10^{12}$  FLOPS.
- **PetaFLOPS** –  $10^{15}$  FLOPS; **ExaFLOPS** –  $10^{18}$  FLOPS.

# Computer power grows rapidly

- Iphone 4 vs. 1985 Cray-2 supercomputer










Projected Performance Development



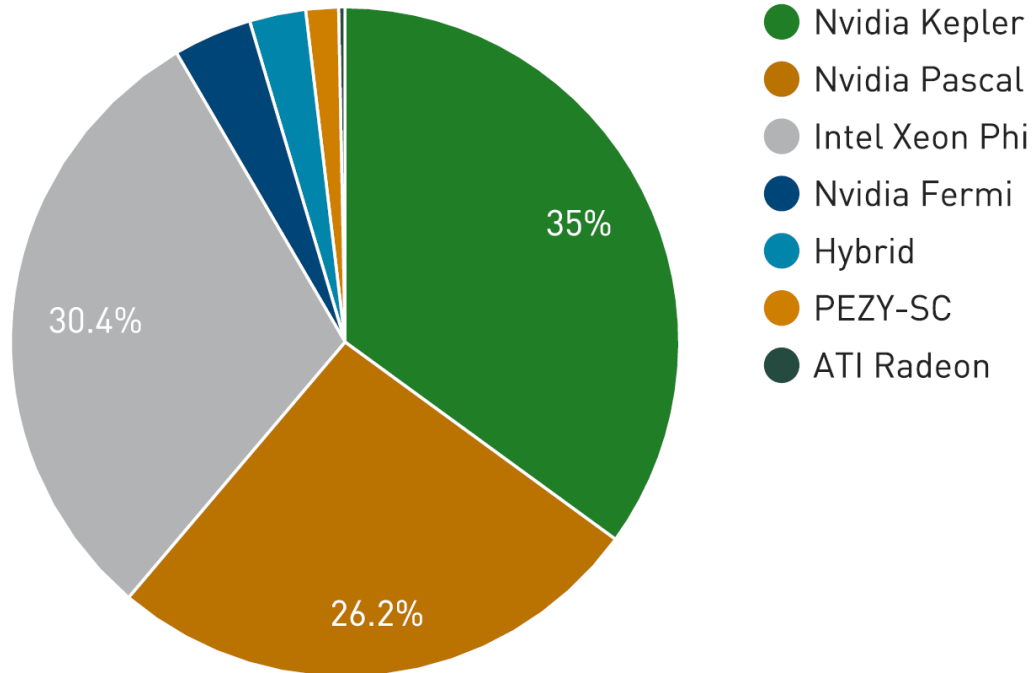
# Top 500 Supercomputers

- The list of June 2018

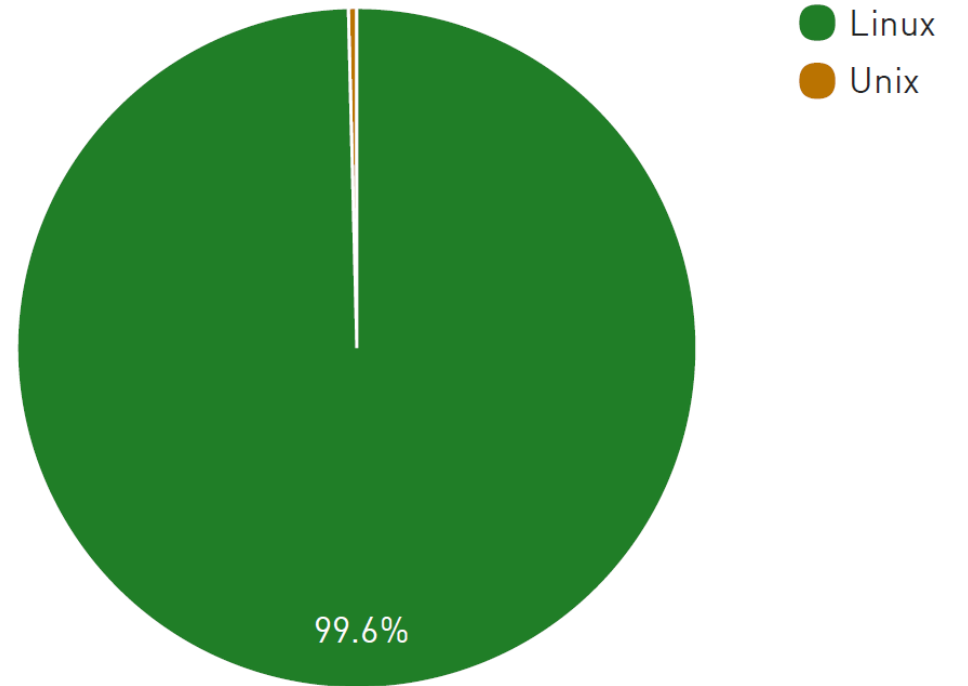
| Rank | Rmax<br>Rpeak<br>(PFLOPS) | Name  | Model               | Processor                                   | Interconnect           | Vendor         | Site<br>country, year  | Operating<br>system            |
|------|---------------------------|---|---------------------|---|------------------------|----------------|--|--------------------------------|
| 1 ▲  | 122.300<br>187.659        | <u>Summit</u>   | Power System AC922  | <u>POWER9</u> , <u>Tesla V100</u>           | Infiniband EDR         | <u>IBM</u>     | <u>Oak Ridge National Laboratory</u><br> <u>United States</u> , 2018                              | <u>Linux (RHEL)</u>            |
| 2 ▼  | 93.015<br>125.436         | <u>Sunway TaihuLight</u>                                | Sunway MPP          | <u>SW26010</u>                              | Sunway <sup>[18]</sup> | <u>NRCPC</u>   | <u>National Supercomputing Center in Wuxi</u><br> <u>China</u> , 2016 <sup>[18]</sup>             | <u>Linux (Raise)</u>           |
| 3 ▲  | 71.610<br>119.194         | <u>Sierra</u>   | Power System S922LC | <u>POWER9</u> , <u>Tesla V100</u>           | Infiniband EDR         | <u>IBM</u>     | <u>Lawrence Livermore National Laboratory</u><br> <u>United States</u> , 2018                     | <u>Linux (RHEL)</u>            |
| 4 ▼  | 61.445<br>100.679         | <u>Tianhe-2A</u>  | TH-IVB-FEP          | <u>Xeon E5-2692 v2</u> , <u>Matrix-2000</u> | TH Express-2           | <u>NUDT</u>    | <u>National Supercomputing Center in Guangzhou</u><br> <u>China</u> , 2013                        | <u>Linux (Kylin)</u>           |
| 5 ▲  | 19.880<br>32.577          | <u>AI Bridging Cloud Infrastructure</u> <sup>[19]</sup> | PRIMERGY CX2550 M4  | <u>Xeon Gold 6148</u> , <u>Tesla V100</u>   | Infiniband EDR         | <u>Fujitsu</u> | <u>National Institute of Advanced Industrial Science and Technology</u><br> <u>Japan</u> , 2018 | <u>Linux</u>                   |
| 6 ▼  | 19.590<br>25.326          | <u>Piz Daint</u>  | <u>Cray XC50</u>    | <u>Xeon E5-2690 v3</u> , <u>Tesla P100</u>  | Aries                  | <u>Cray</u>    | <u>Swiss National Supercomputing Centre</u><br> <u>Switzerland</u> , 2016                       | <u>Linux (CLE)</u>             |
| 7 ▼  | 17.590<br>27.113          | <u>Titan</u>  | <u>Cray XK7</u>     | <u>Opteron 6274</u> , <u>Tesla K20X</u>     | Gemini                 | <u>Cray</u>    | <u>Oak Ridge National Laboratory</u><br> <u>United States</u> , 2012                            | <u>Linux (CLE, SLES based)</u> |

# Statistics of the Top 500

Accelerator/CP Family Performance Share



Operating system Family System Share



# HPC user environment

- Operation system: Linux (Redhat/CentOS, Ubuntu), 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 or profiling tools: Totalview, Tau, DDT, Vtune.
- Programming Languages: C, C++, Fortran, Python, Perl, R, MATLAB, Mathematica, Julia



# Scientific disciplines in HPC

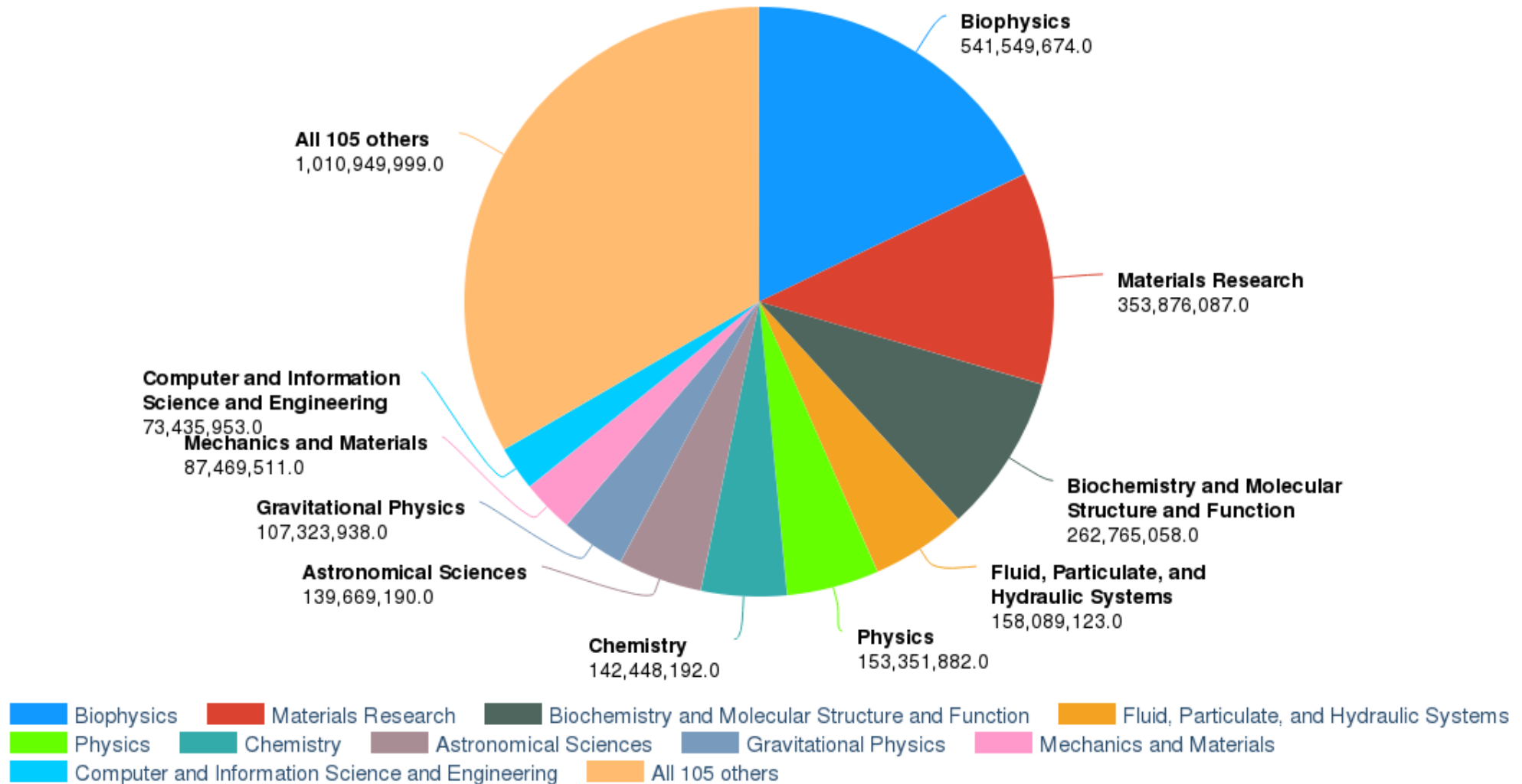
## □ Typical scientific computing catalogs:

- 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 (1)

- Statistics from XSEDE

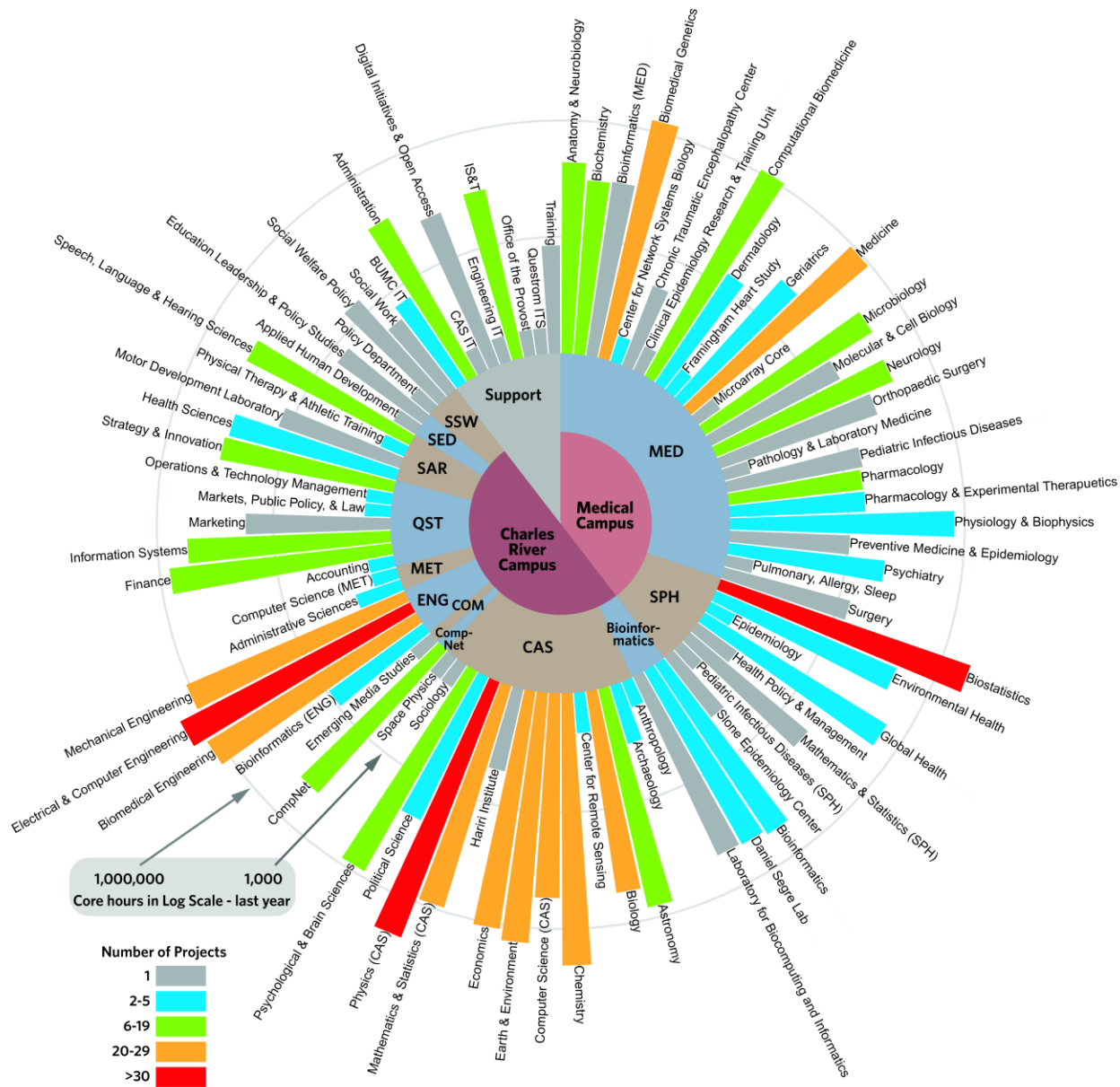
XD SUs Charged: Total: by Field of Science



## Breadth of Research on the Shared Computing Cluster (SCC)

# CPU-hours by field of science (2)

- Statistics from BU SCC



# 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

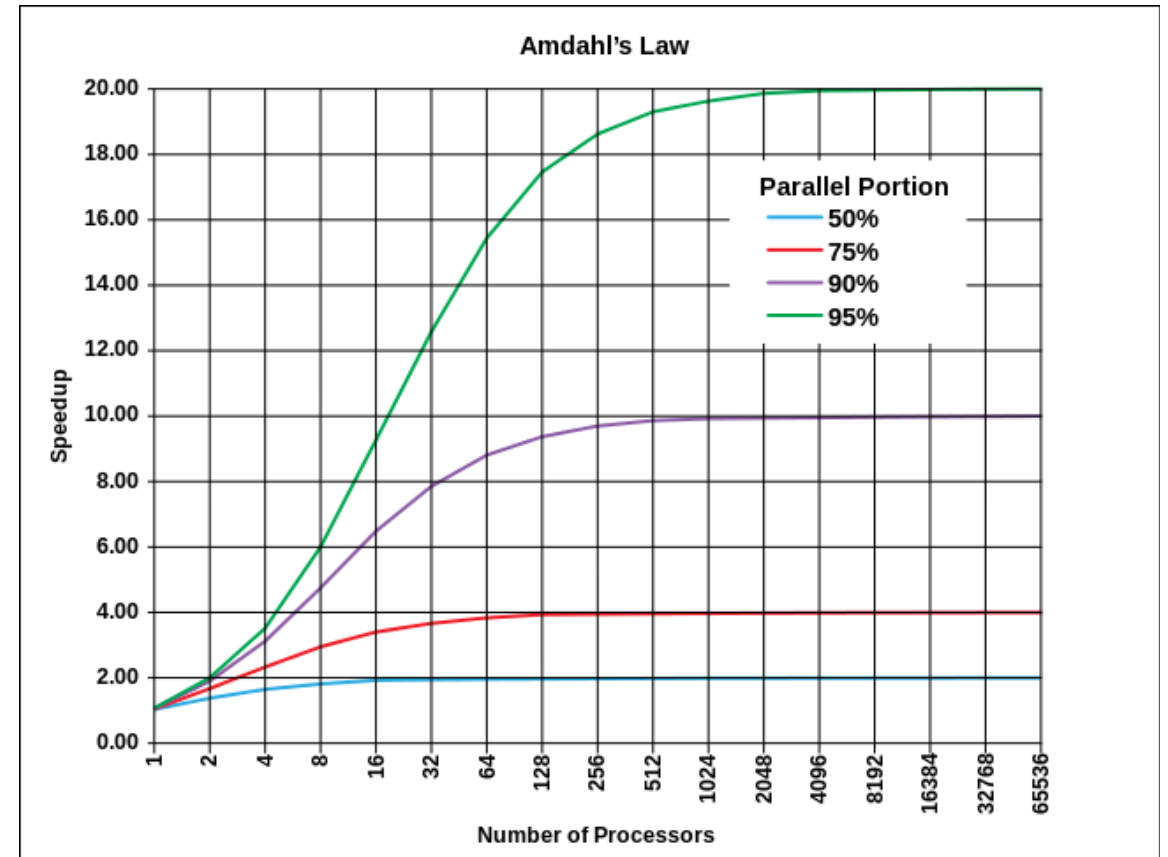
□ Parallel computing is a type of computation in which many calculations are carried out **simultaneously**, based on the principle that large problems can often be divided into smaller ones, which are then solved at the same time.

□ **Speedup** of a parallel program,

$$S(p) = \frac{T(1)}{T(p)} = \frac{1}{\alpha + \frac{1}{p}(1 - \alpha)}$$

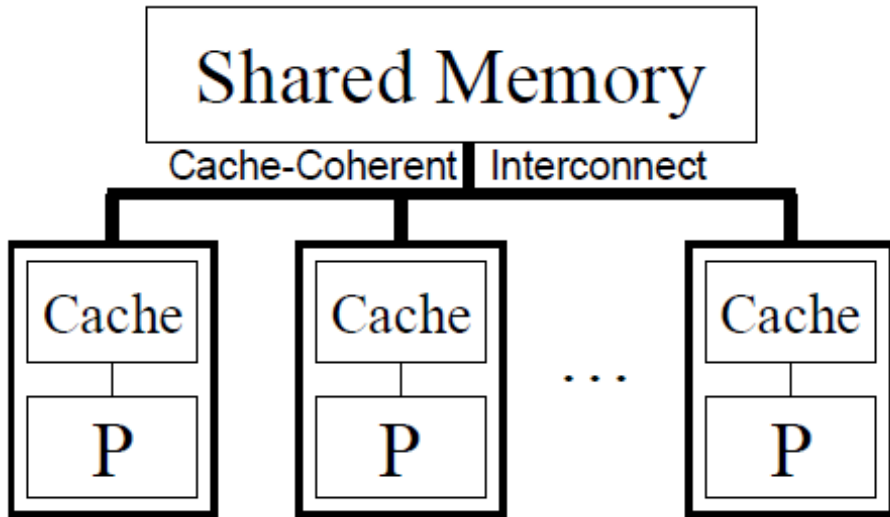
$p$ : number of processors/cores,

$\alpha$ : fraction of the program that is serial.

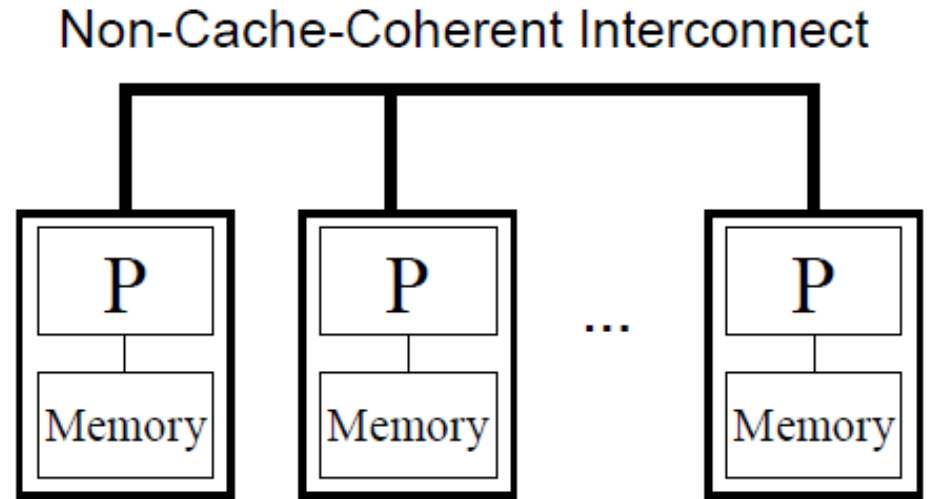


• The figure is from: [https://en.wikipedia.org/wiki/Parallel\\_computing](https://en.wikipedia.org/wiki/Parallel_computing)

# Distributed or shared memory systems



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



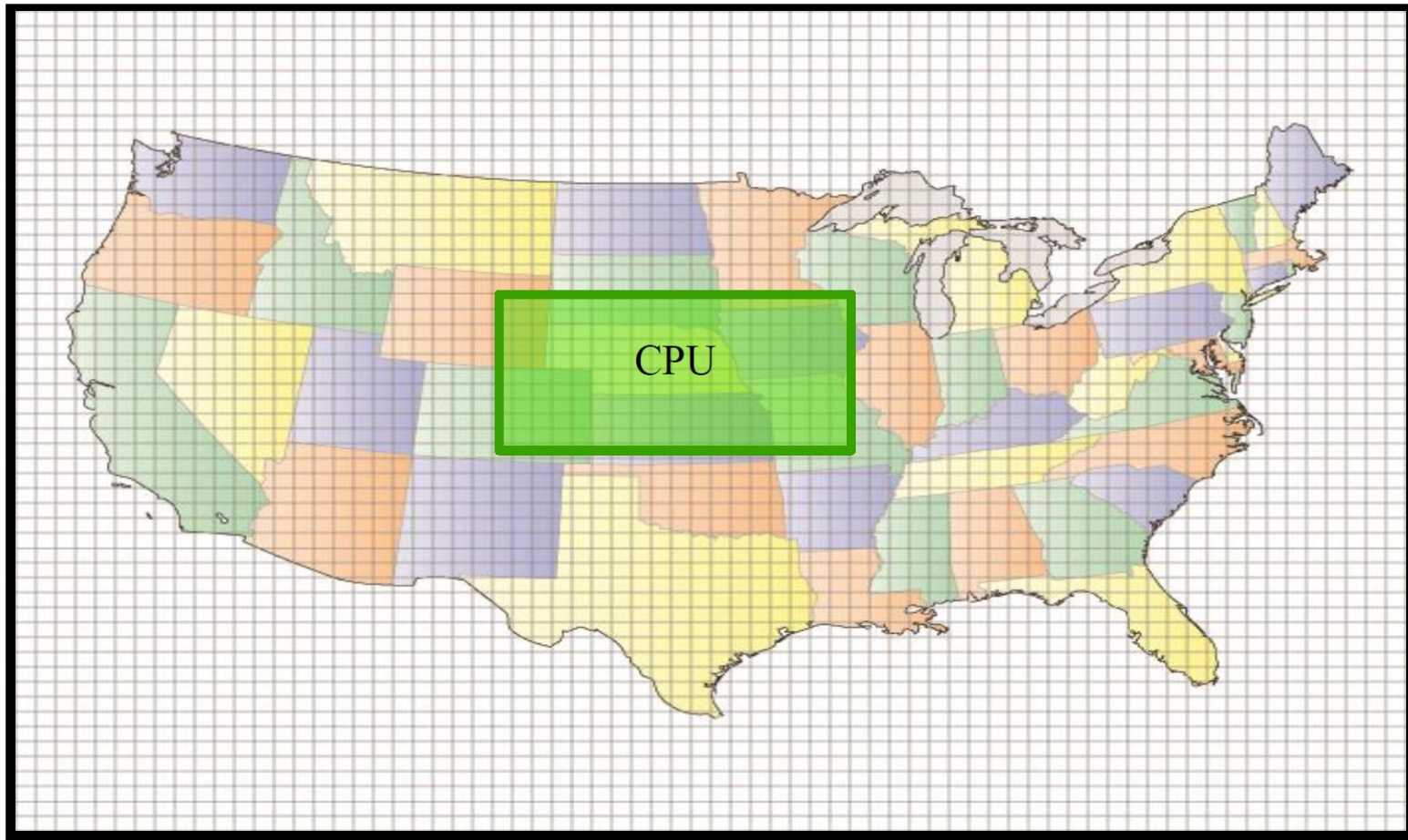
- Distributed memory system
- For example, multiple nodes on a cluster
- Message Passing Interface (MPI)

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

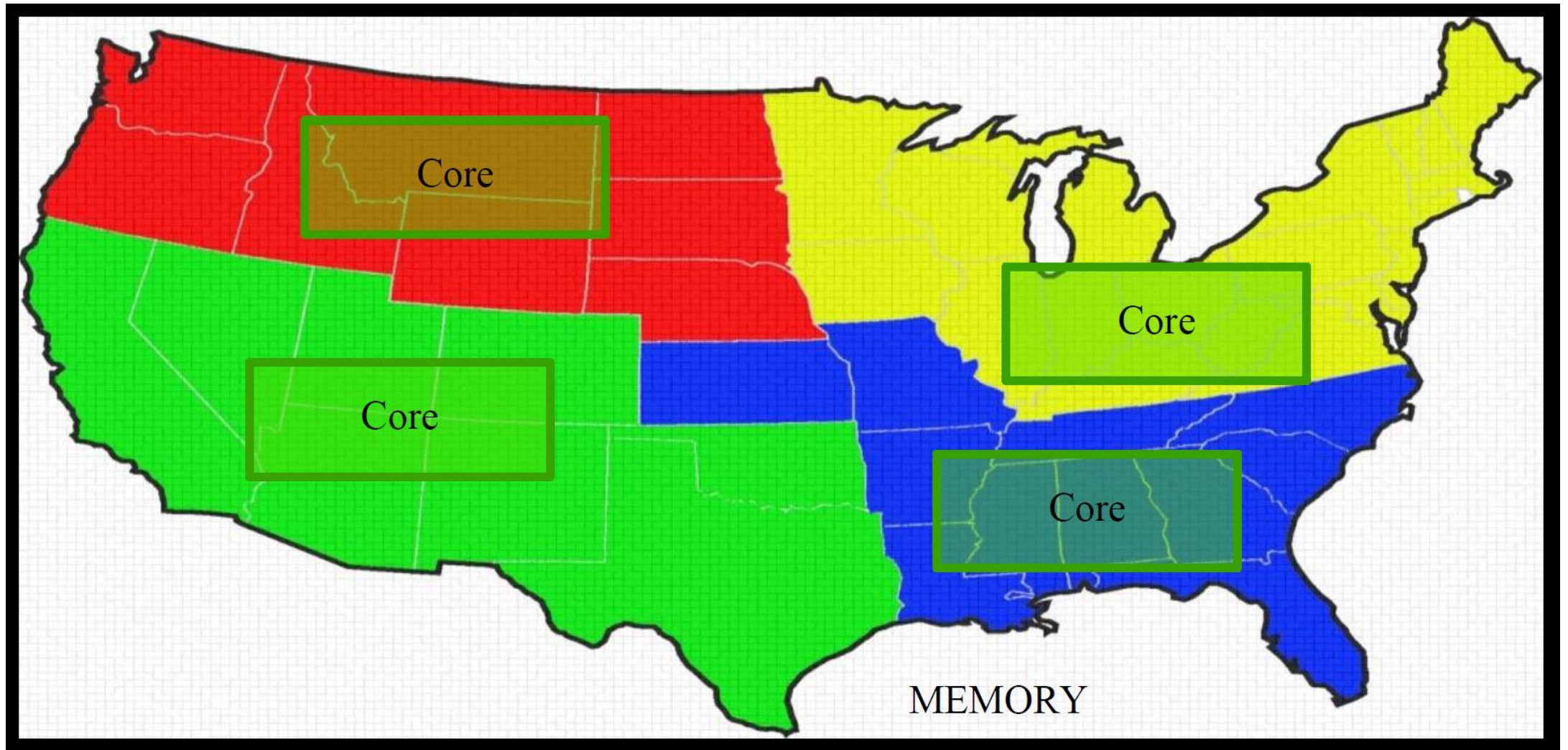


# An example: weather science

- Serial weather model

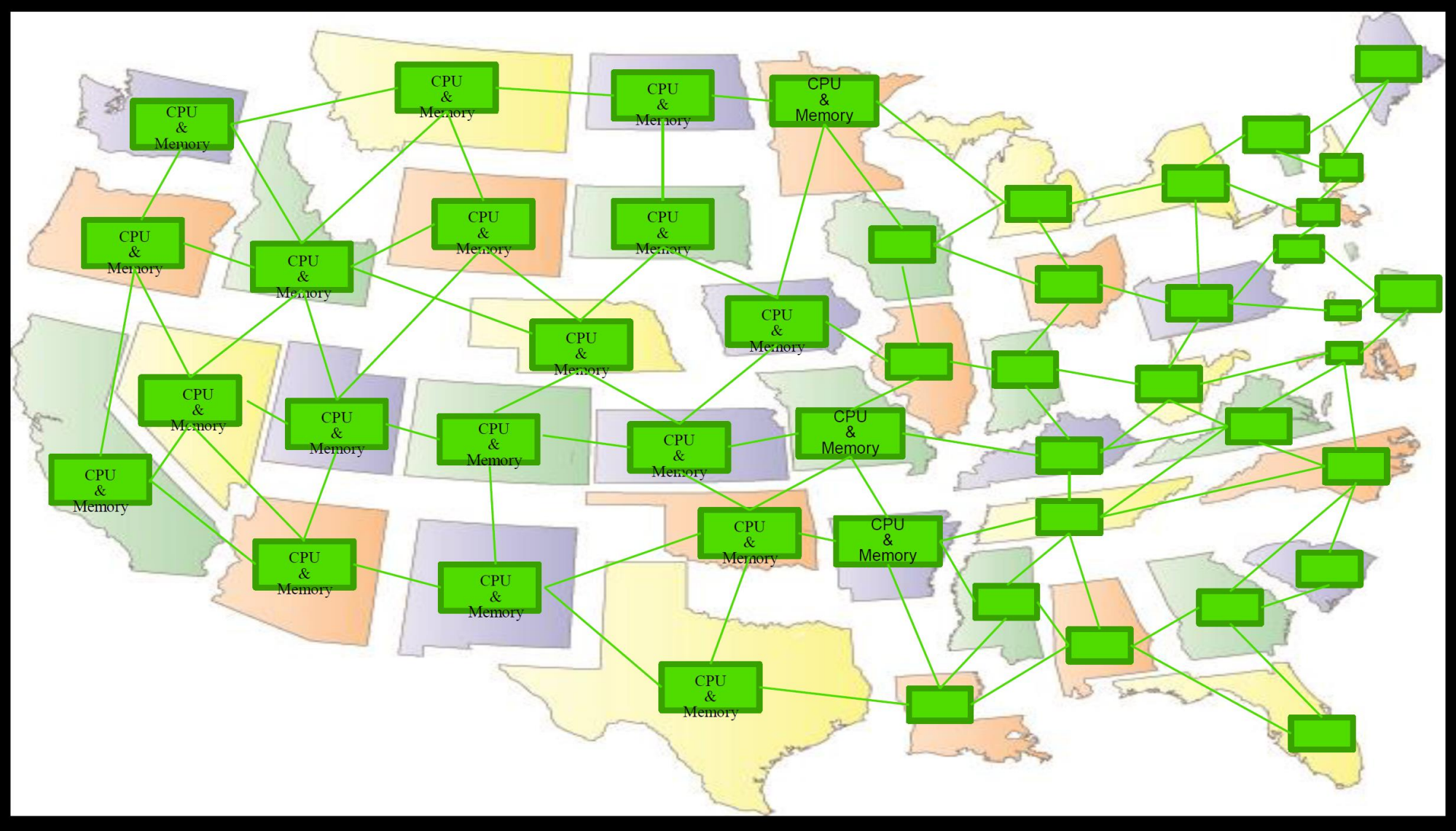


- Shared-memory weather model (for several cores within one node)





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



# Nation-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 country.
- 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)

| <b>Machine Name</b>               | <b>Resource Provider</b>                | <b>Best Types of Computation</b>  | <b>Resource Highlights</b>   |
|-----------------------------------|---|---|--|
| <a href="#"><u>Bridges</u></a>    | 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.                      |
| <a href="#"><u>Comet</u></a>      | 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.                   |
| <a href="#"><u>Greenfield</u></a> | 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)

| <b>Machine Name</b>              | <b>Resource Provider</b>   | <b>Best Types of Computation</b>  | <b>Resource Highlights</b>   |
|----------------------------------|--|---|--|
| <a href="#"><u>Jetstream</u></a> | Indiana University (IU) and Texas Advanced Computing Center (TACC) | Particularly for cloud computing.   | User-friendly cloud environment.   |
| <a href="#"><u>Maverick</u></a>  | Texas Advanced Computing Center (TACC).                            | Particularly for visualization jobs   | VNC server; GPU and large memory nodes for visualization.  |
| <a href="#"><u>Stampede2</u></a> | Texas Advanced Computing Center (TACC).                            | The largest cluster among all XSEDE resources; Good for massive MPI or OpenMP jobs. | Thousands of compute nodes; Intel Xeon Phi Knights Landing (KNL) processors; Intel Skylake processors. |

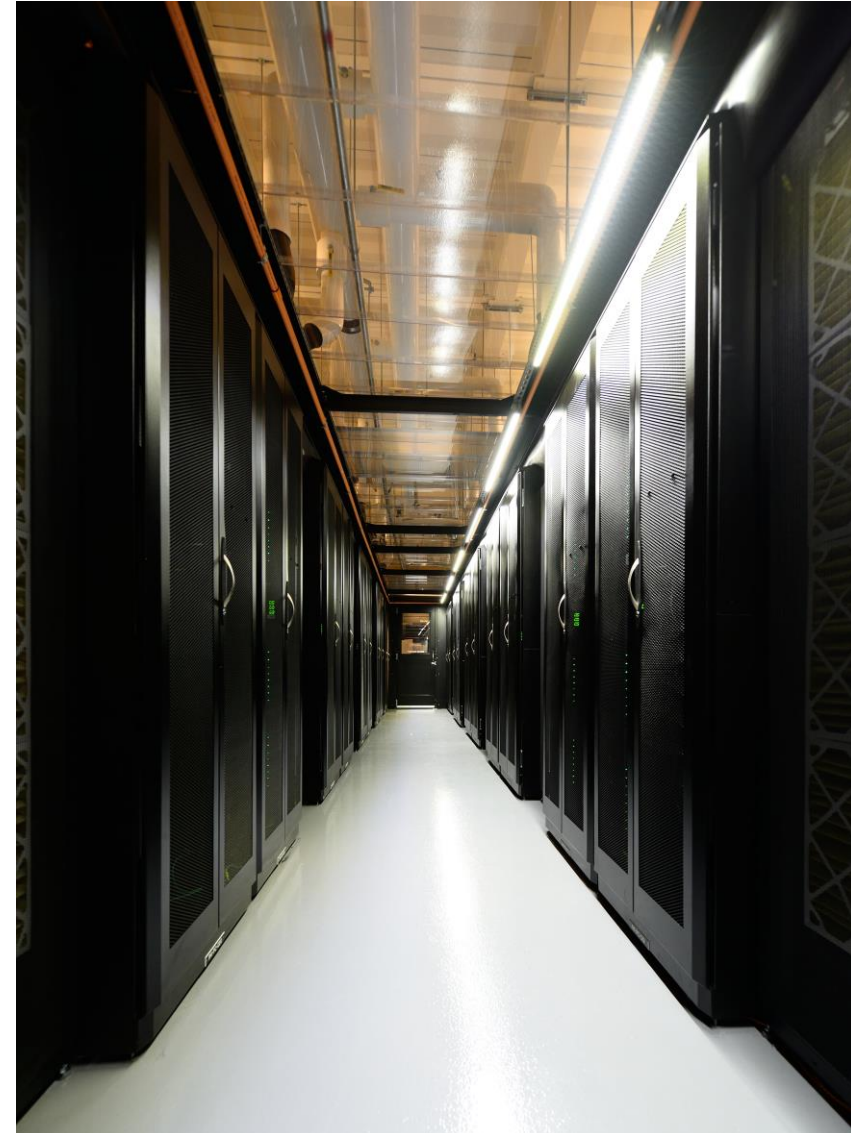


## XSEDE resources (3)

| <b>Machine Name</b>                      | <b>Resource Provider</b>                              | <b>Best Types of Computation</b>                       | <b>Resource Highlights</b>   |
|--|---|--|--|
| <a href="#"><u>SuperMIC</u></a>          | 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.      |
| <a href="#"><u>XStream</u></a>           | 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. |
| <a href="#"><u>Open Science Grid</u></a> | 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/>