IT4Innovations national supercomputer center

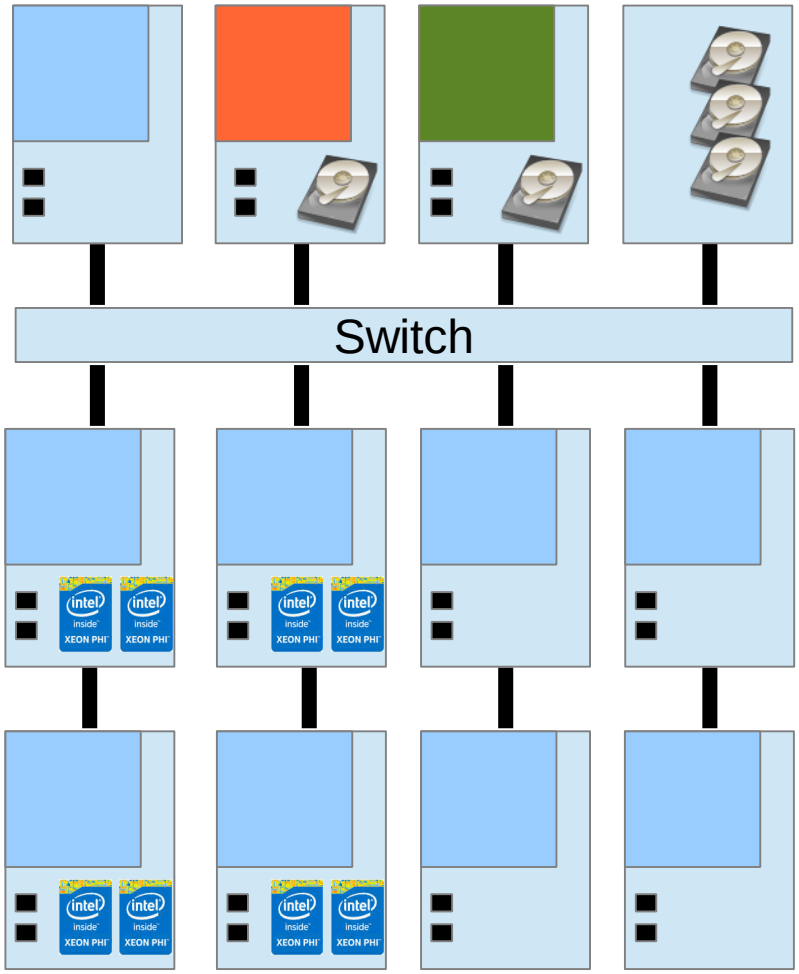
Branislav Jansík
branislav.jansik@vsb.cz

IT4Innovations national supercomputing center

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

2007-13
OP Research and Development for Innovation

Salomon cluster HPC infrastructure



Storage

- HOME: 500 TB
- SCRATCH 1700 TB

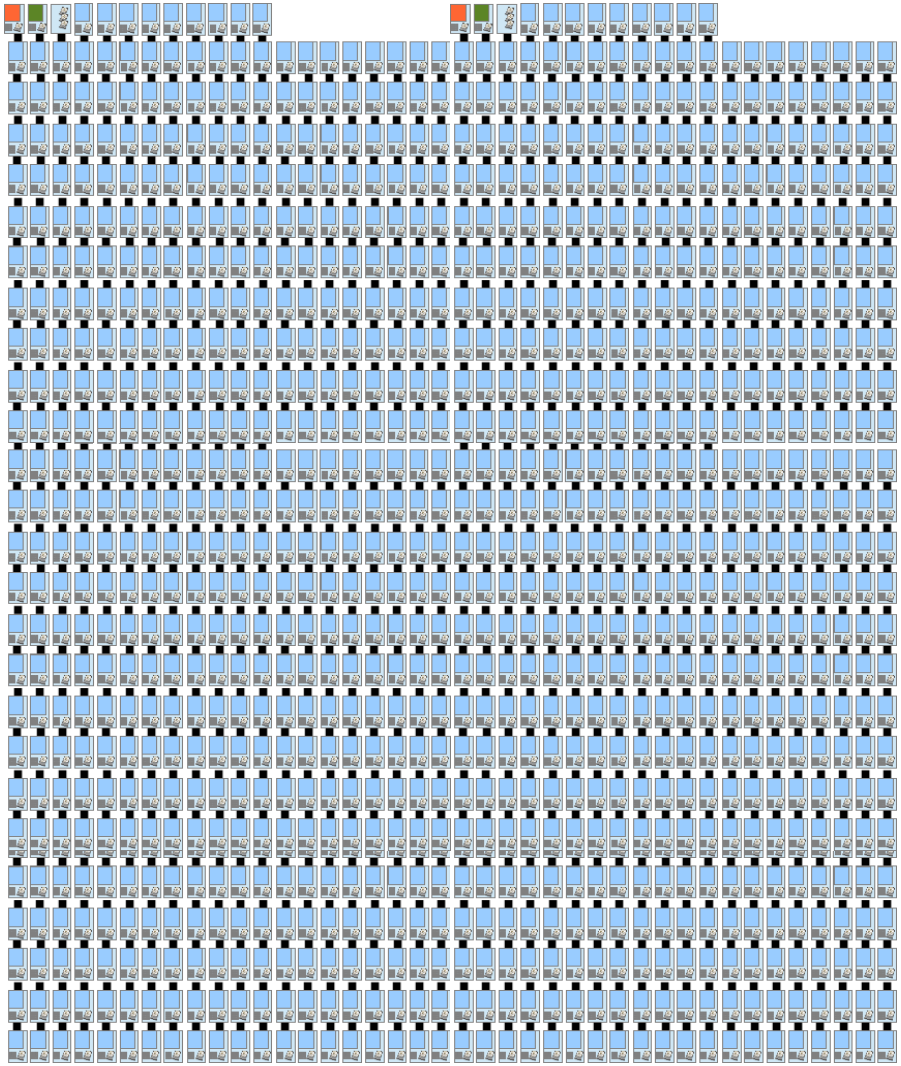
Interconnect

- Infiniband, 7D hypercube
- 56Gb/s

Compute

- 1008 nodes
- Haswell-EP 2.5GHz x86-64
- 24 cores, 256 bit FMA instr.
- 128 GB RAM
- 864x Intel Xeon Phi 7120P
- 61(244) cores, 512 bit FMA
- 16 GB RAM

Salomon cluster HPC infrastructure



Storage

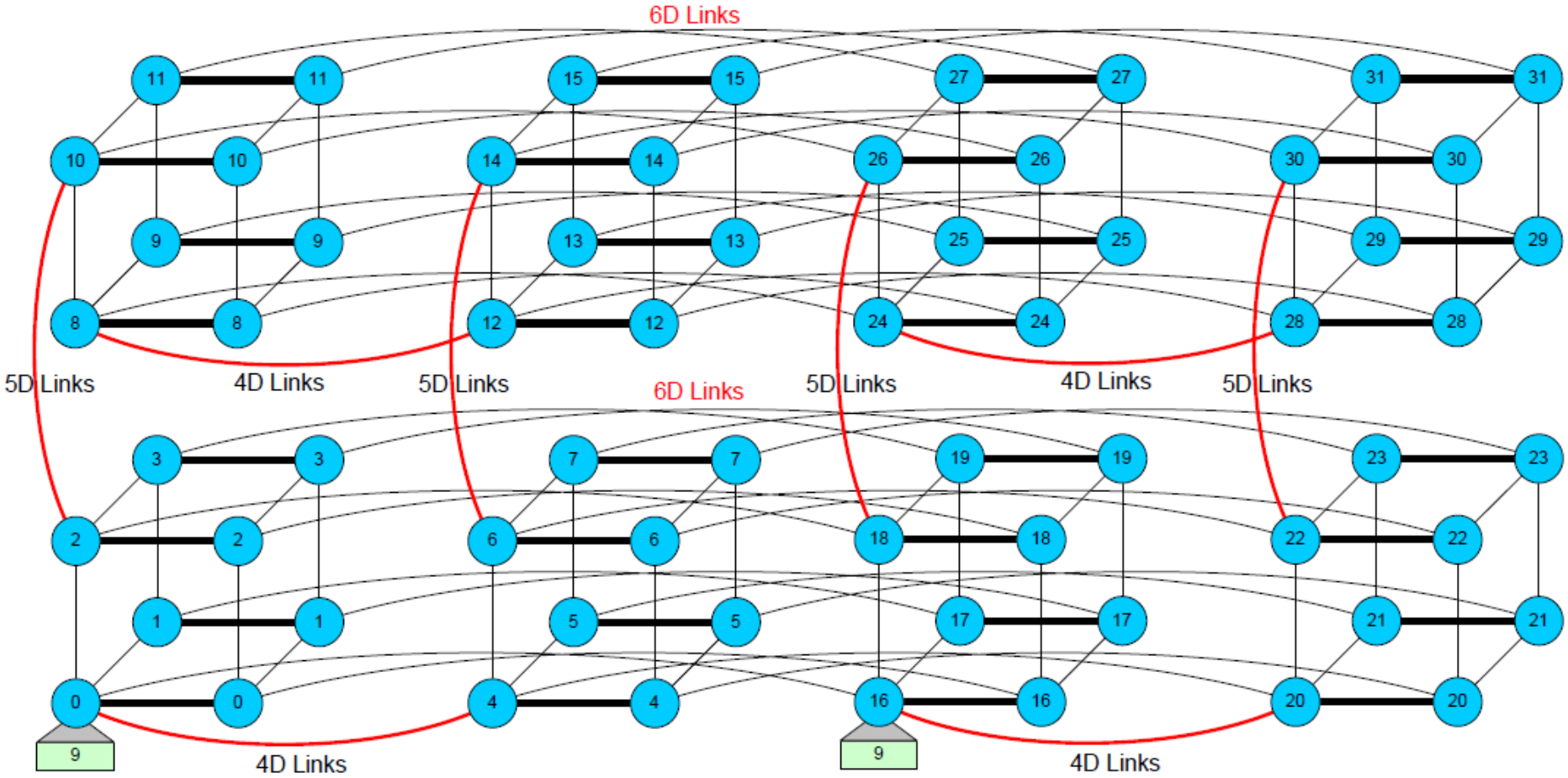
- HOME: 500 TB
- SCRATCH 1700 TB

Interconnect

- Infiniband, 7D hypercube
- 56Gb/s

Compute

- 1008 nodes
- Haswell-EP 2.5GHz x86-64
- 24 cores, 256 bit FMA instr.
- 128 GB RAM
- 864x Intel Xeon Phi 7120P
- 61(244) cores, 512 bit FMA
- 16 GB RAM

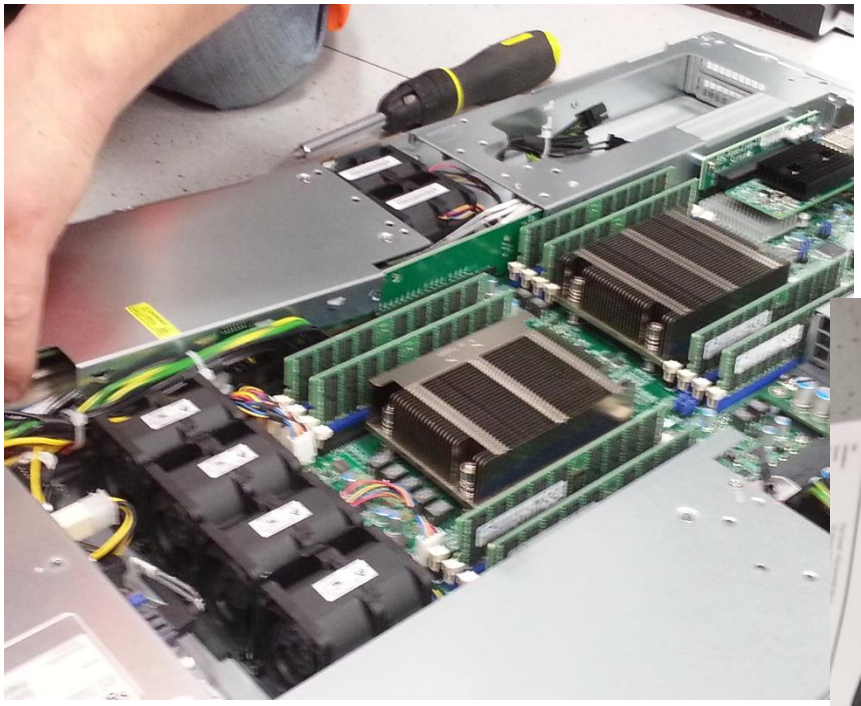


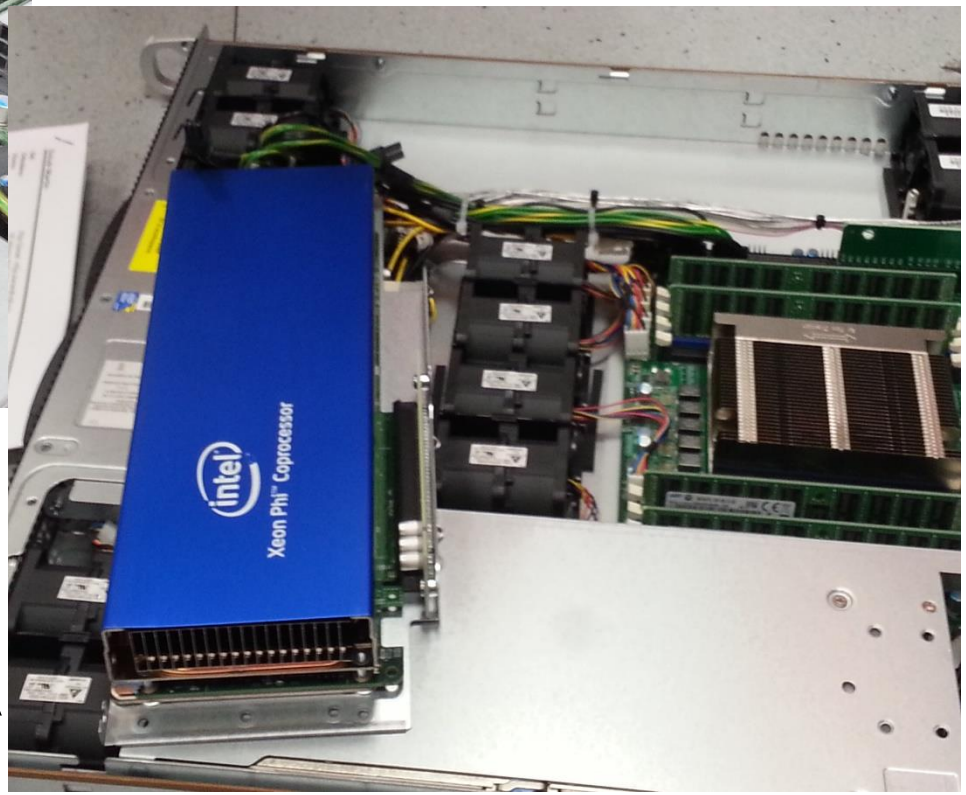
IT4Innovations
national supercomputing center

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

OP Research and Development for Innovation

Intel Xeon Phi accelerator cards

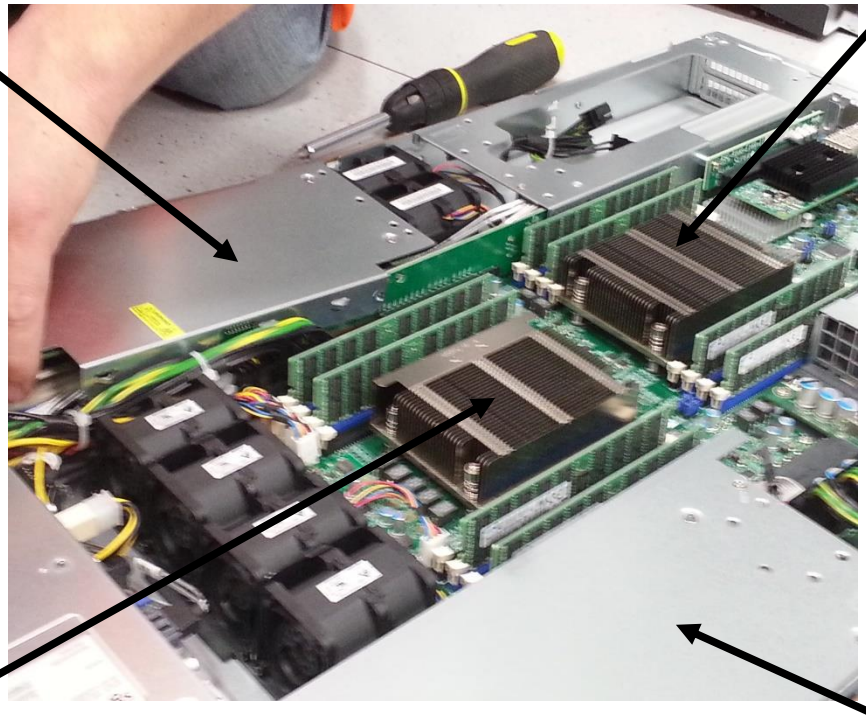




Xeon Phi 7120P

- x86 architecture, **1.2TF**
- 864x Intel Xeon Phi 7120P
- 61(244) cores, 512 bit FMA
- 16 GB RAM

3.36TF per accelerated node

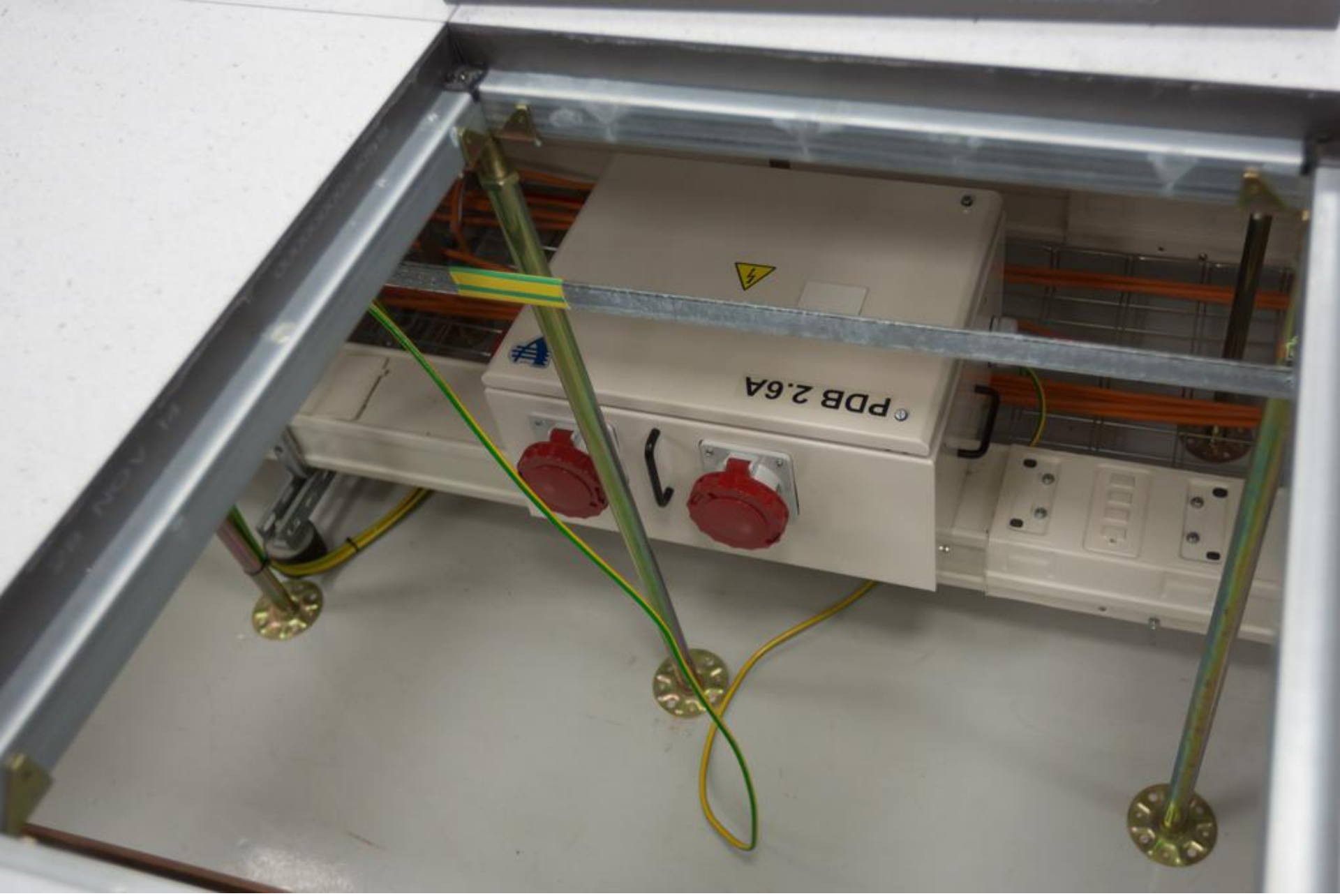


Los Alamos Blue Mountain (2000)
3.07TF


TRA B


DwT1 B
PRONIX


PRONIX
PRONIX


PDB 2.6A


SGI ICE X
sgi


OxyReduct®
PRONIX
WAGNER
A
OxyReduct®
PRONIX
WAGNER
B

IT4Innovations
national01$#&0
supercomputing
center@#01%101

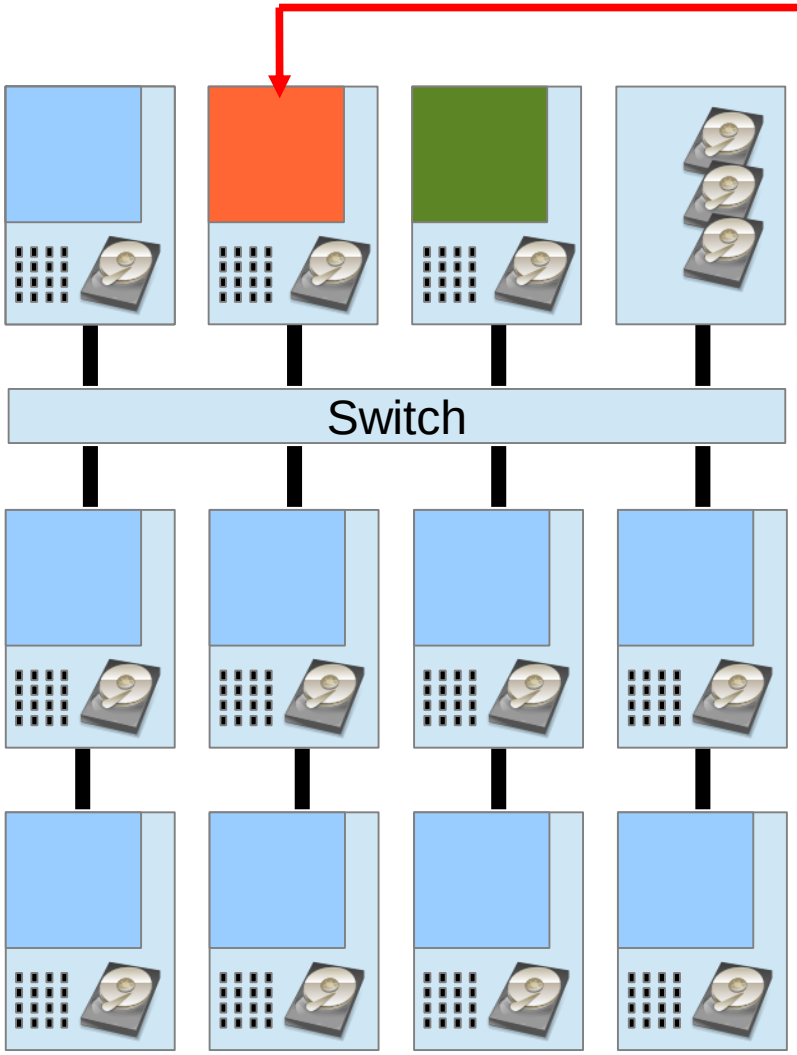
EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

2007-13
OP Research and
Development for Innovation


CIAT
PRONIX
CIAT
PRONIX


CIAT
CIAT
CIAT
CIAT
PROMAX

Logging in

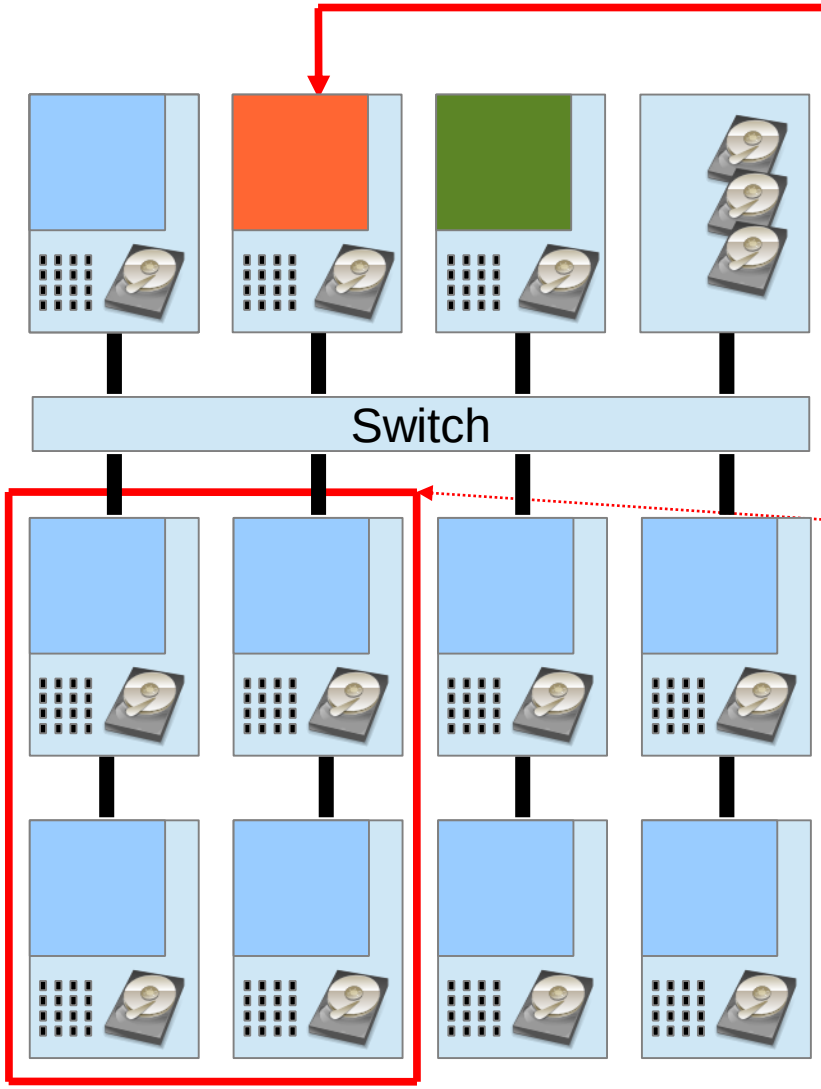


user0 login

Explore the login node

- Shell configuration
- Software modules

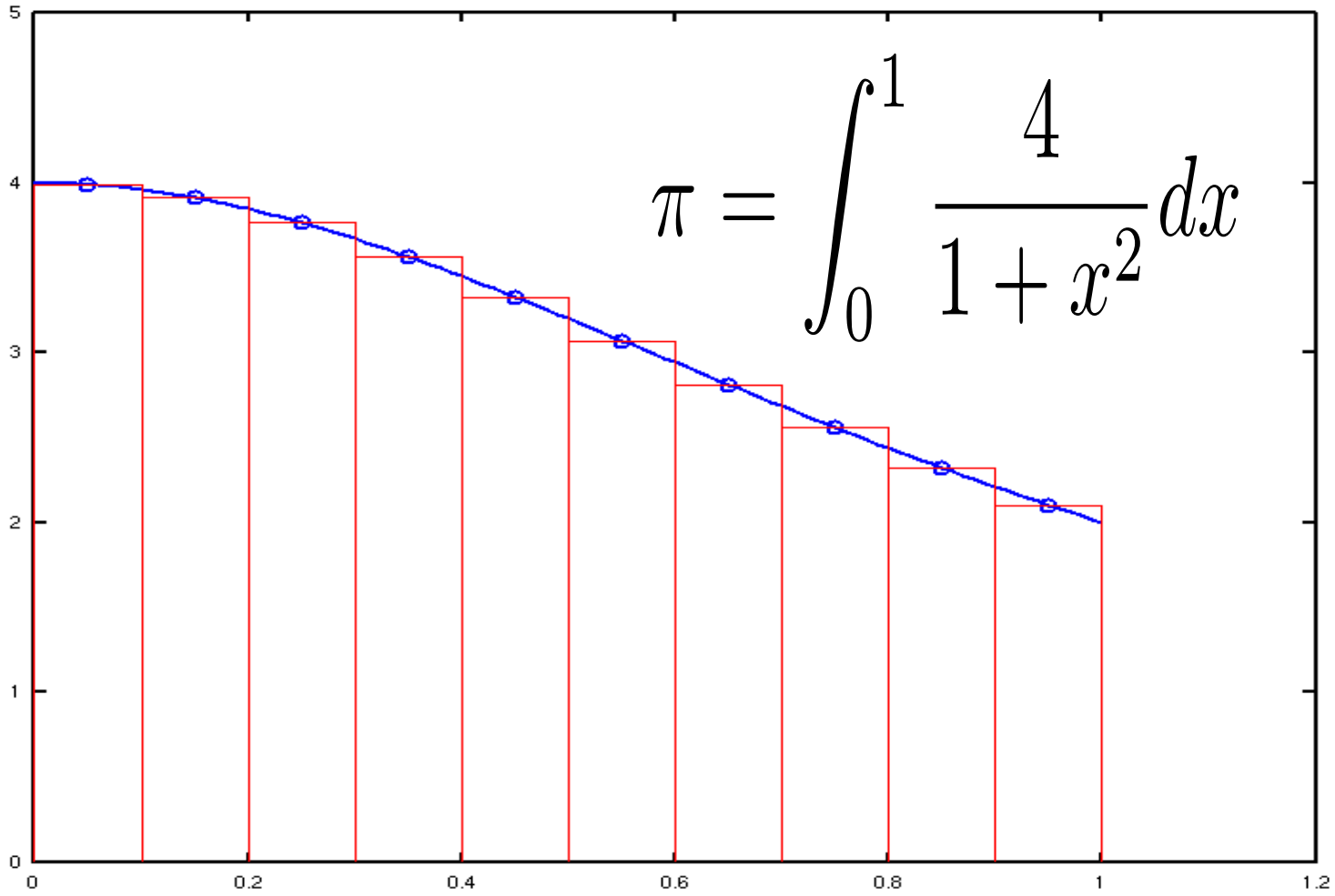
Allocation and execution



user0 login

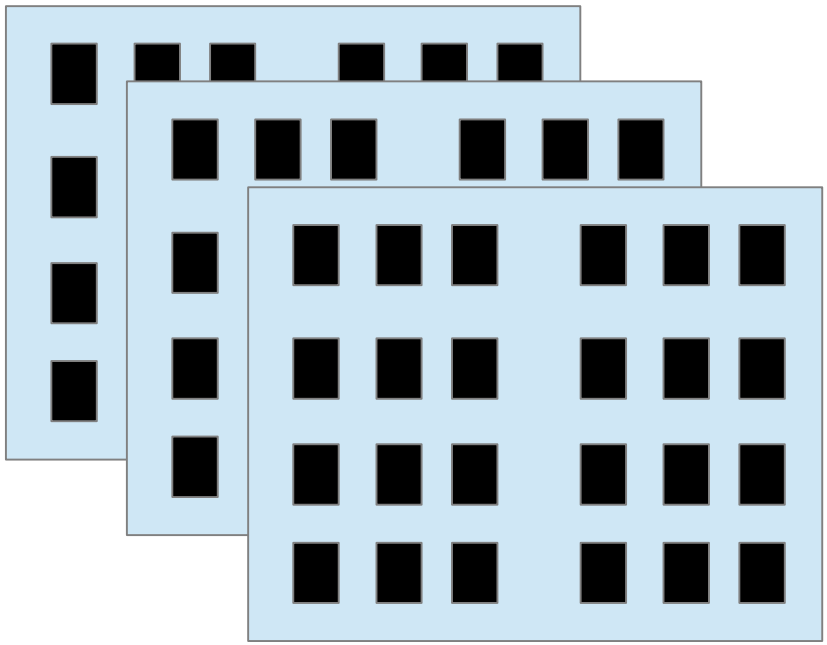
Resource allocation and execution via **PBS queue system**

A parallel job: pi integration



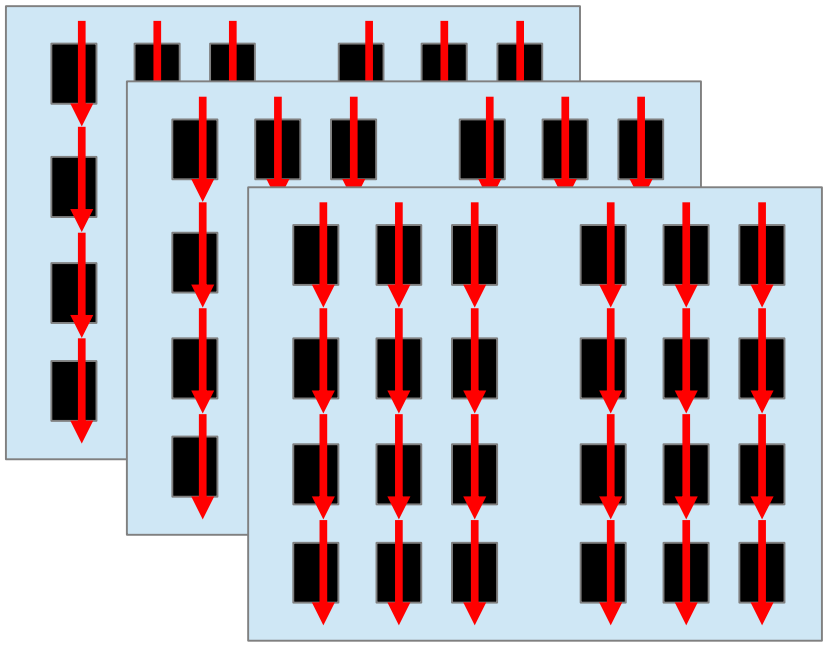
Running parallel jobs

- $ qsub -q qexp -l **select=3:ncpus=24 -I**



Running parallel jobs

- $ qsub -q qexp -l **select=3:ncpus=24 -I**
- module load impi
- mpicc pi3.c -o pi3.x
- mpirun -n 72 ./pi3.x



Molecular orbital localization

$$\xi^m_{\mathrm{FM}} = \sum_p (\mu^p_4)^m \quad \text{ir} \quad \mu^p_4 = \langle p|(\boldsymbol{r} - \bar{\boldsymbol{r}}_p)^4|p\rangle$$

Trust region Newton optimization

$$|\tilde{p}\rangle = \exp(-\hat{\kappa}) a^\dagger_{p\sigma}|\mathrm{vac}\rangle$$

$$Q_m(\boldsymbol{\kappa}) = \xi^{[0]}_m + \boldsymbol{\kappa}^T \boldsymbol{\xi}^{[1]}_m + \frac{1}{2}\boldsymbol{\kappa}^T \boldsymbol{\xi}^{[2]}_m \boldsymbol{\kappa}$$

Hessian linear transformation

$$\boldsymbol{\xi}_m^{[2]}\boldsymbol{\kappa} = m \sum_{x=x,y,z} \Big\lfloor -2\{(\boldsymbol{\kappa}\boldsymbol{x}^4 - \boldsymbol{x}^4\boldsymbol{\kappa})\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}]\}^o$$

- $$+8\{\boldsymbol{x}^3\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}(\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})]\}^o + 8\{(\boldsymbol{\kappa}\boldsymbol{x}^3 - \boldsymbol{x}^3\boldsymbol{\kappa})\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}\boldsymbol{x}]\}^o$$
$$+8\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}(\boldsymbol{\kappa}\boldsymbol{x}^3 - \boldsymbol{x}^3\boldsymbol{\kappa})\}^o - 24\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}(\boldsymbol{\kappa}\boldsymbol{x}^2 - \boldsymbol{x}^2\boldsymbol{\kappa})\boldsymbol{x}\}^o$$
$$-24\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}[\boldsymbol{x}^2](\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})\}^o + 72\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}[\boldsymbol{x}]^2(\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})]\}^o$$
$$+8\{(\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}[\boldsymbol{x}^3]\}^o - 24\{(\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}\boldsymbol{x}[\boldsymbol{x}^2]\}^o$$
$$+24(\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}[\boldsymbol{x}]^3]\}^o - 24\{\boldsymbol{x}^2\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}\boldsymbol{x}(\boldsymbol{\kappa}\boldsymbol{x} - \boldsymbol{x}\boldsymbol{\kappa})]\}^o$$
$$-12\{(\boldsymbol{\kappa}\boldsymbol{x}^2 - \boldsymbol{x}^2\boldsymbol{\kappa})\mathrm{dia}[(\boldsymbol{\mu}_4)^{m-1}[\boldsymbol{x}]^2]\}^o$$
$$-2\{\boldsymbol{x}^4\mathrm{dia}[\boldsymbol{\kappa}^T\boldsymbol{f}^{[1]}]\}^o + 8\{\boldsymbol{x}^3\mathrm{dia}[(\boldsymbol{\kappa}^T\boldsymbol{f}^{[1]})\boldsymbol{x}]\}^o$$
$$-12\{\boldsymbol{x}^2\mathrm{dia}[(\boldsymbol{\kappa}^T\boldsymbol{f}^{[1]})[\boldsymbol{x}]^2]\}^o + 8\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\kappa}^T\boldsymbol{f}^{[1]})[\boldsymbol{x}^3]]\}^o$$
$$-24\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\kappa}^T\boldsymbol{f}^{[1]})[\boldsymbol{x}^2]\boldsymbol{x}]\}^o + 24\{\boldsymbol{x}\mathrm{dia}[(\boldsymbol{\kappa}^T\boldsymbol{f}^{[1]})[\boldsymbol{x}]^3]\}^o\Big\rfloor$$

Hessian linear transformation

$$
\begin{aligned}
&+2m\sum_{i>j}\Big[-2\{(\boldsymbol{\kappa}\boldsymbol{x}_i^2\boldsymbol{x}_j^2-\boldsymbol{x}_i^2\boldsymbol{x}_j^2\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}\}^o\\
&-8\{(\boldsymbol{\kappa}\boldsymbol{x}_i\boldsymbol{x}_j-\boldsymbol{x}_i\boldsymbol{x}_j\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}\boldsymbol{x}_i\boldsymbol{x}_j]\}^o+(1+P_{ij})\{\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_i^2](\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})]\}^o\\
&-4(1+P_{ij})\{\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}(\boldsymbol{\kappa}\boldsymbol{x}_i^2-\boldsymbol{x}_i^2\boldsymbol{\kappa})\boldsymbol{x}_j]\}^o+4(1+P_{ij})\{\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}(\boldsymbol{\kappa}\boldsymbol{x}_i^2\boldsymbol{x}_j-\boldsymbol{x}_i^2\boldsymbol{x}_j\boldsymbol{\kappa})]\}^o\\
&+24(1+P_{ij})\{\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_j][\boldsymbol{x}_i](\boldsymbol{\kappa}\boldsymbol{x}_i-\boldsymbol{x}_i\boldsymbol{\kappa})]\}^o-8(1+P_{ij})\{\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}\boldsymbol{x}_i(\boldsymbol{\kappa}\boldsymbol{x}_i\boldsymbol{x}_j-\boldsymbol{x}_i\boldsymbol{x}_j\boldsymbol{\kappa})]\}^o\\
&-8(1+P_{ij})\{\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}\boldsymbol{x}_i(\boldsymbol{\kappa}\boldsymbol{x}_i-\boldsymbol{x}_i\boldsymbol{\kappa})[\boldsymbol{x}_i\boldsymbol{x}_j]]\}^o-4(1+P_{ij})\{(\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_i^2]\boldsymbol{x}_j]\}^o\\
&+4(1+P_{ij})\{(\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_i^2\boldsymbol{x}_j]\}^o+12(1+P_{ij})\{(\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_i]^2\boldsymbol{x}_j]\}^o\\
&-8(1+P_{ij})\{(\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_i\boldsymbol{x}_j]\boldsymbol{x}_i]\}^o-4(1+P_{ij})\{\boldsymbol{x}_i^2\mathrm{dia}[\boldsymbol{f}^{[0]}(\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})\boldsymbol{x}_j]\}^o\\
&-2(1+P_{ij})\{(\boldsymbol{\kappa}\boldsymbol{x}_j^2-\boldsymbol{x}_j^2\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}[\boldsymbol{x}_i]^2\}^o+4(1+P_{ij})\{(\boldsymbol{\kappa}\boldsymbol{x}_i^2\boldsymbol{x}_j-\boldsymbol{x}_i^2\boldsymbol{x}_j\boldsymbol{\kappa})\mathrm{dia}[\boldsymbol{f}^{[0]}\boldsymbol{x}_j]\}^o\\
&+4(1+P_{ij})\{\boldsymbol{x}_i^2\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}(\boldsymbol{\kappa}\boldsymbol{x}_j-\boldsymbol{x}_j\boldsymbol{\kappa})]\}^o-8(1+P_{ij})\{\boldsymbol{x}_i\boldsymbol{x}_j\mathrm{dia}[\boldsymbol{f}^{[0]}(\boldsymbol{\kappa}\boldsymbol{x}_i-\boldsymbol{x}_i\boldsymbol{\kappa})\boldsymbol{x}_j]\}^o
\end{aligned}
$$

Hessian linear transformation

- $$+ 2m \sum_{i>j} \Big[- 2\{\boldsymbol{x}_i^2 \boldsymbol{x}_j^2 \mathrm{dia}[\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]}]\}^o - 8\{\boldsymbol{x}_i \boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]}) \boldsymbol{x}_i \boldsymbol{x}_j]\}^o$$

$$- 4(1 + P_{ij})\{\boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]})[\boldsymbol{x}_i^2] \boldsymbol{x}_j]\}^o + 12(1 + P_{ij})\{\boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]})[\boldsymbol{x}_i]^2 \boldsymbol{x}_j]\}^o$$

$$4(1 + P_{ij})\{\boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]})[\boldsymbol{x}_i^2 \boldsymbol{x}_j]]\}^o - 2(1 + P_{ij})\{\boldsymbol{x}_i^2 \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]})[\boldsymbol{x}_j]^2]\}^o$$

$$+ 4(1 + P_{ij})\{\boldsymbol{x}_i^2 \boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]}) \boldsymbol{x}_j]\}^o \Big]$$

$$- \frac{1}{2}\{\boldsymbol{\kappa} \boldsymbol{\xi}_m^{[0]}\}^o$$

Hessian linear transformation

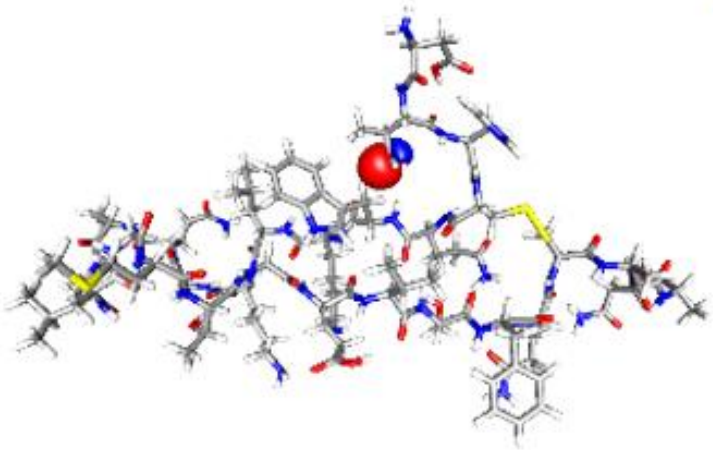
$$+ 2m \sum_{i>j} \Big[- 2\{\boldsymbol{x}_i^2 \boldsymbol{x}_j^2 \mathrm{dia}[\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]}]\}^o - 8\{\boldsymbol{x}_i \boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]}) \boldsymbol{x}_i \boldsymbol{x}_j]\}^o$$

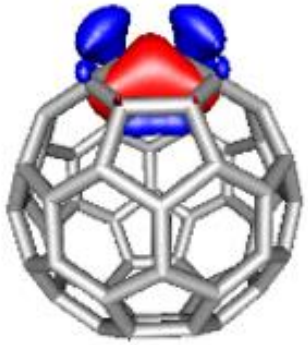
$$- 4(1 + P_{ij})\{\boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]})[\boldsymbol{x}_i^2] \boldsymbol{x}_j]\}^o + 12(1 + P_{ij})\{\boldsymbol{x}_j \mathrm{dia}[(\boldsymbol{\kappa}^T \boldsymbol{f}^{[1]})[\boldsymbol{x}_i]^2 \boldsymbol{x}_j]\}^o$$

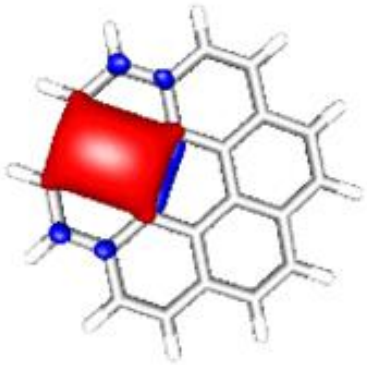
- 25 Matrix multiplications per micro iteration,
- 15000x15000, 1.6GB per matrix
- 20-100 micro iterations
- 20 Matrix multiplications per macro iteration
- 10-60 macro iterations
- **5200 to 151200 Matrix multiplications**

Molecular orbital localization

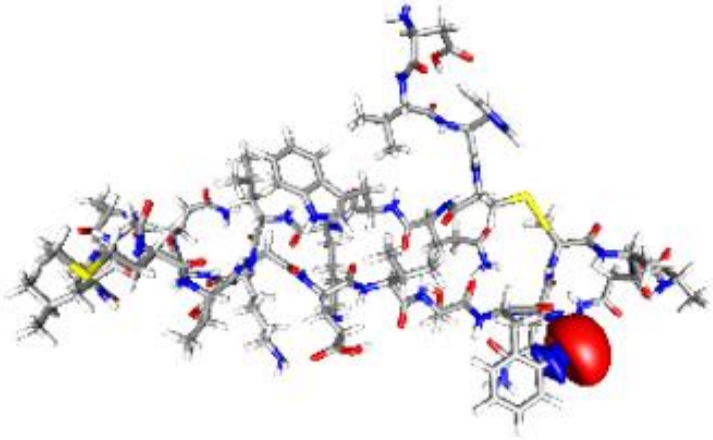
- Occupied orbitals

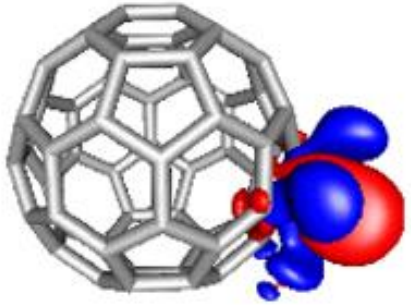






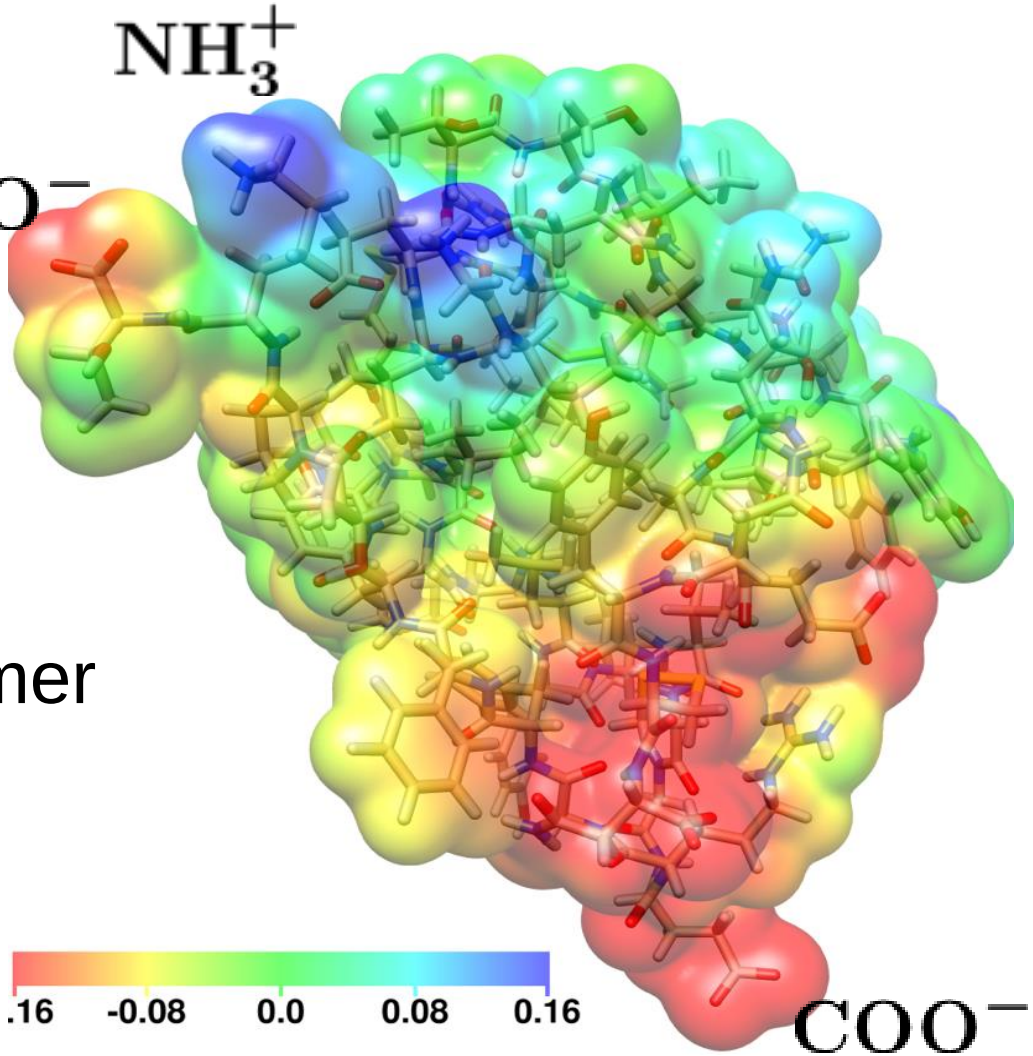
Virtual orbitals





MP2 Electrostatic potential

- Click on the
  2 level
  3 level
  4 level
  5 level



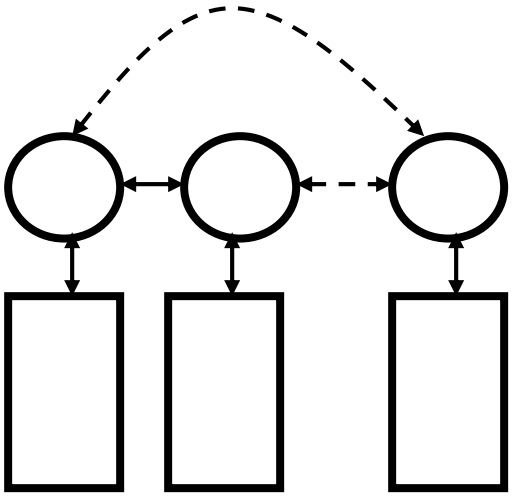
Insulin monomer

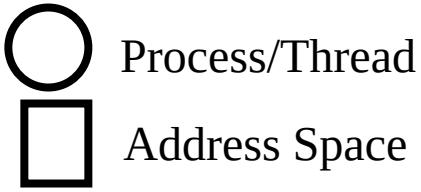
Programming Models

- Common Parallel Programming models
  - Data Parallel (e.g. HPF)
  - Message Passing (e.g. MPI)
  - Shared Memory (e.g. OpemMP)
    - GAS (Global Address Space)
  - PGAS or Distributed Shared Memory
    - Partitioned Global Address Space (e.g. UPC)

- Hybrid models
  - GAS/Shared Memory under Message Passing
    - e.g. MPI + OpenMP

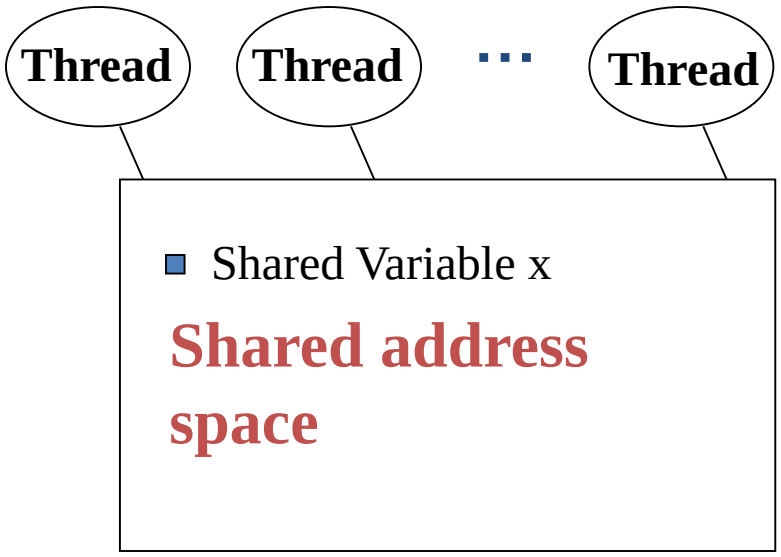
The Message Passing Model





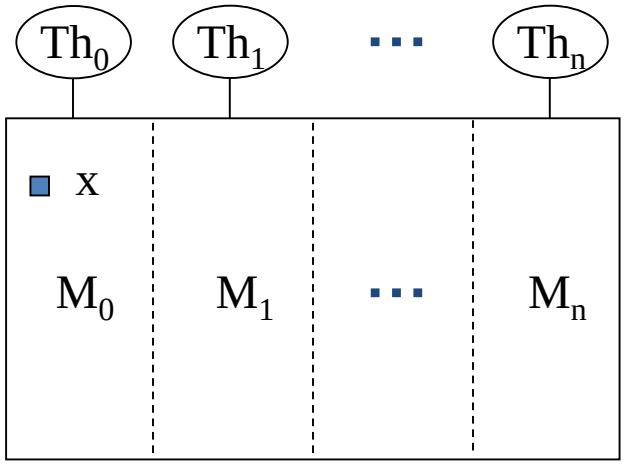
- Communicating sequential processes
- Typically SPMD
- Programmers control data and work distribution
- Explicit communication, two-sided
- Library-based
- Excessive buffering
- Significant communication overhead for small transactions
- Example: MPI

The Global Address Space/ Shared Memory Model



- Simple statements
  - » read remote memory via an expression
  - » write remote memory through assignment
- Manipulating shared data may require synchronization
- **Does not allow locality exploitation – No locality awareness!**
- Example: OpenMP

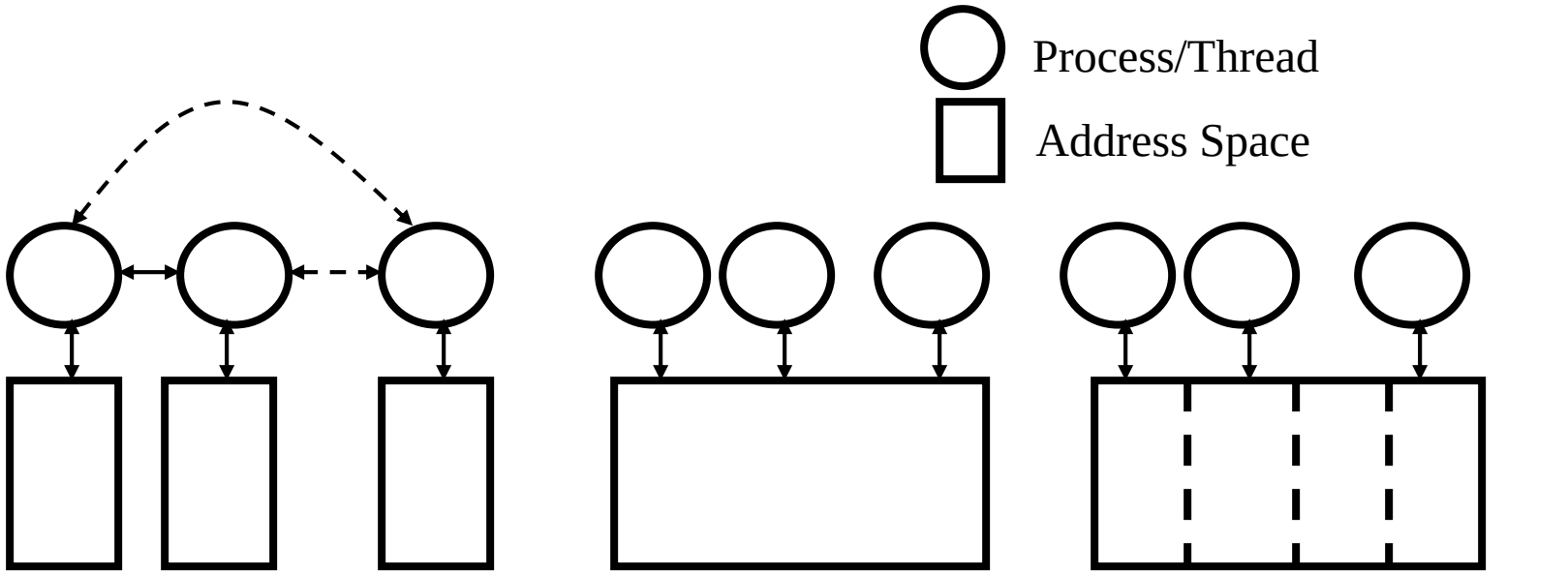
The Partitioned Global Address Space Model (PGAS)



Partitioned Global Address Space

- Similar to the shared memory paradigm
- Memory M_i has affinity to thread Th_i
  - » Locality awareness
- Helps exploiting locality of references
- Simple statements
- Examples: UPC, CAF, and Titanium

Programming Models

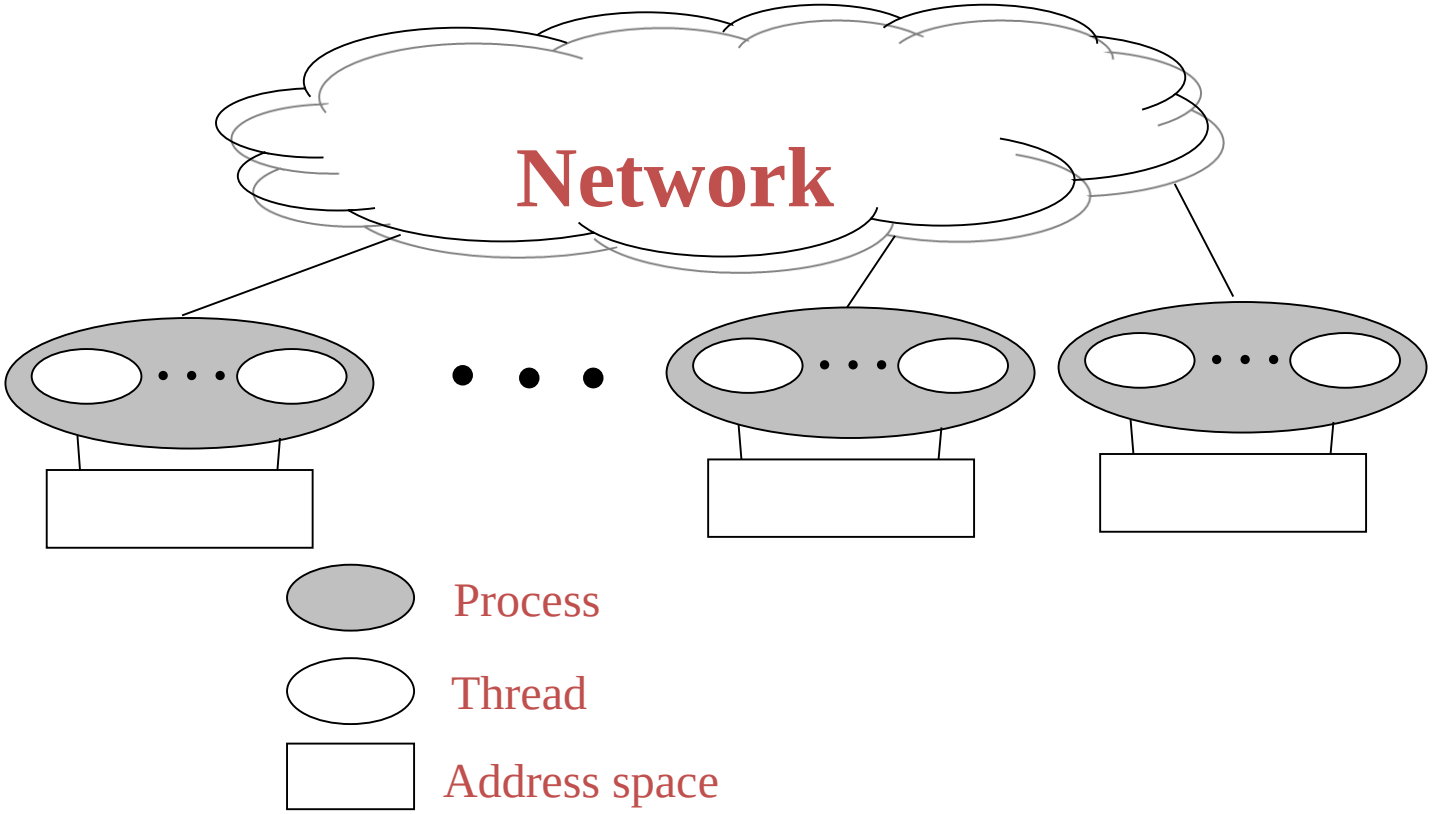


Message Passing
- MPI

Shared Memory
- OpenMP

Distribute Shared Memory (DSM)
- PGAS
- UPC

Hybrid Model(s)
Example: Shared + Message Passing



- Example: OpenMP at the node (SMP), and MPI in between

Amdahl's Law

Amdahl's Law - ratio of sequential part to parallel part in the application

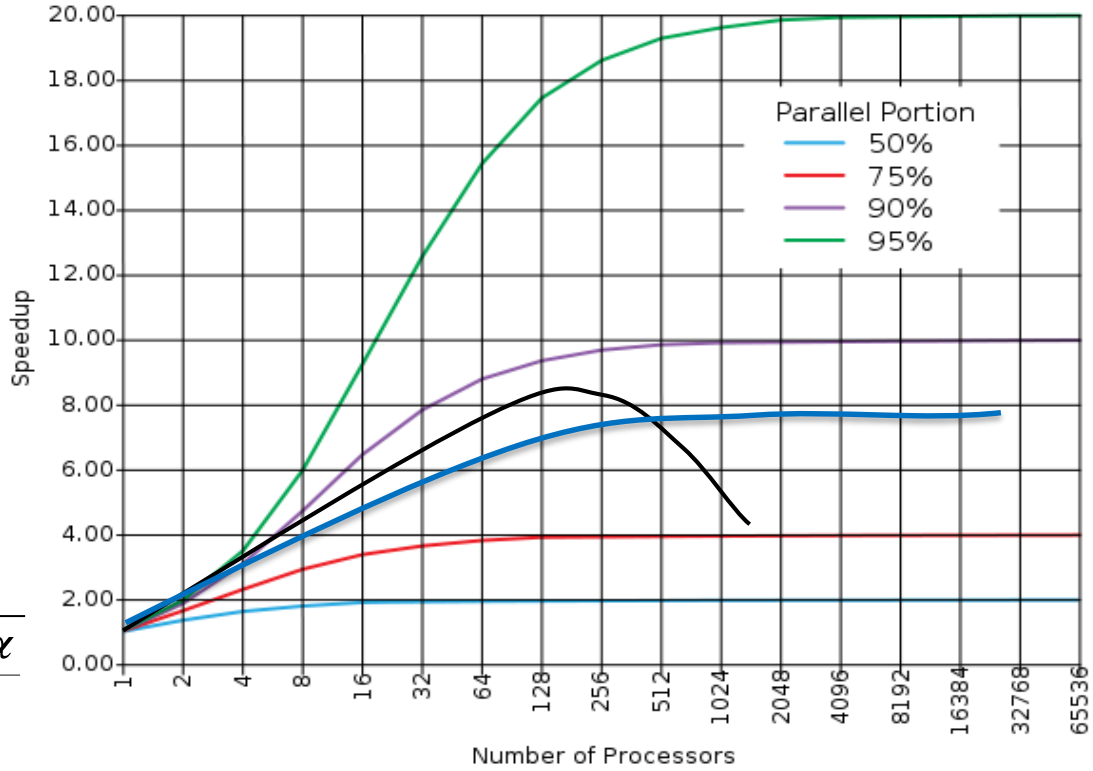
- The performance improvement that can be gained by a parallel implementation is limited by the fraction of time parallelism can actually be used in an application

- Let α = fraction of program (algorithm) that is serial and cannot be parallelized. For instance:
  - Loop initialization
  - Reading/writing to a single disk
  - Procedure call overhead

- Parallel run time is given by

$$T_p = \alpha T_s + \frac{(1-\alpha)T_s}{p}$$

$$S_p = \frac{T_s}{\alpha T_s + \frac{(1-\alpha)T_s}{p}} = \frac{1}{\alpha + \frac{1-\alpha}{p}}$$



Note: Communication overhead grows with # of processors

X% (parallel)	Y% (seq.)

$$S = \frac{original\ time}{improved\ time} = \frac{1}{\left(\frac{X}{n}\right) + Y} \rightarrow \frac{1}{Y}\ for\ n \rightarrow \infty$$

n – number of processors
Y is the bottle neck – the effect as comm. overhead

Amdahl's Law

- The speedup that is achievable on p processors is:

$$S_p = \frac{T_s}{T_p} = \frac{1}{\alpha + \frac{1-\alpha}{p}}$$

- If we assume that the serial fraction is fixed, then the speedup for infinite processors is limited by $1/\alpha$

$$\lim_{p->\infty} S_p = \frac{1}{\alpha}$$

- For example, if α=10%, then the maximum speedup is 10, even if we use an infinite number of processors

Comments on Amdahl's Law

- The Amdahl's fraction α in practice depends on the problem size n and the number of processors p
- An effective parallel algorithm has:

$$\alpha(n, p) \to 0 \text{ as } n \to \infty$$

- For such a case, even if one fixes p, we can get linear speedups by choosing a suitable large problem size

$$S_p = \frac{T_s}{T_p} = \frac{p}{1+(p-1)\alpha(n, p)} \to p \text{ as } n \to \infty$$

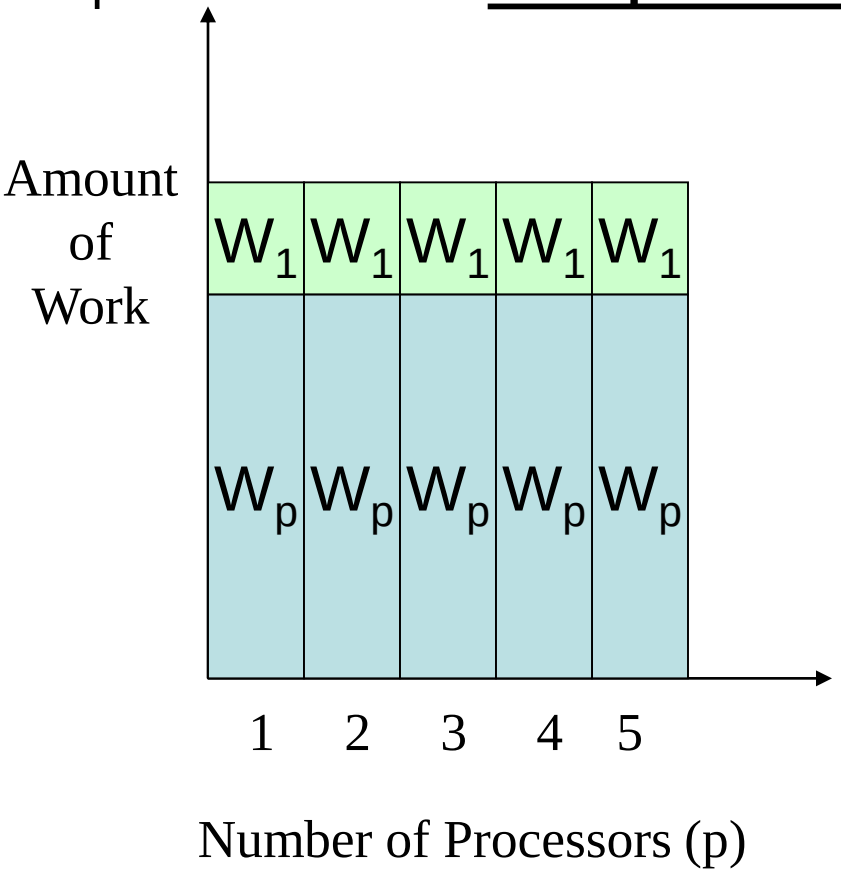
Scalable speedup

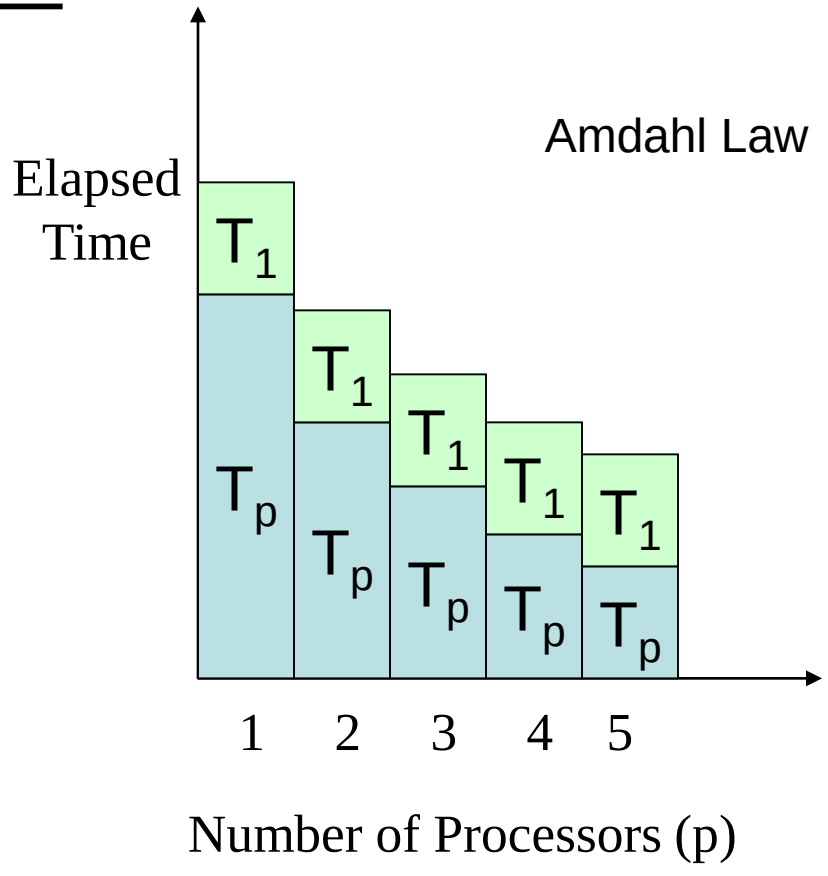
- Practically, the problem size that we can run for a particular problem is limited by the time and memory of the parallel computer

Scaling Characteristics of Parallel Programs

Scalability - the ability to maintain performance gains when system and/or problem size increase

- strong scaling - how the processing time varies with the number of processors for a **fixed problem size**

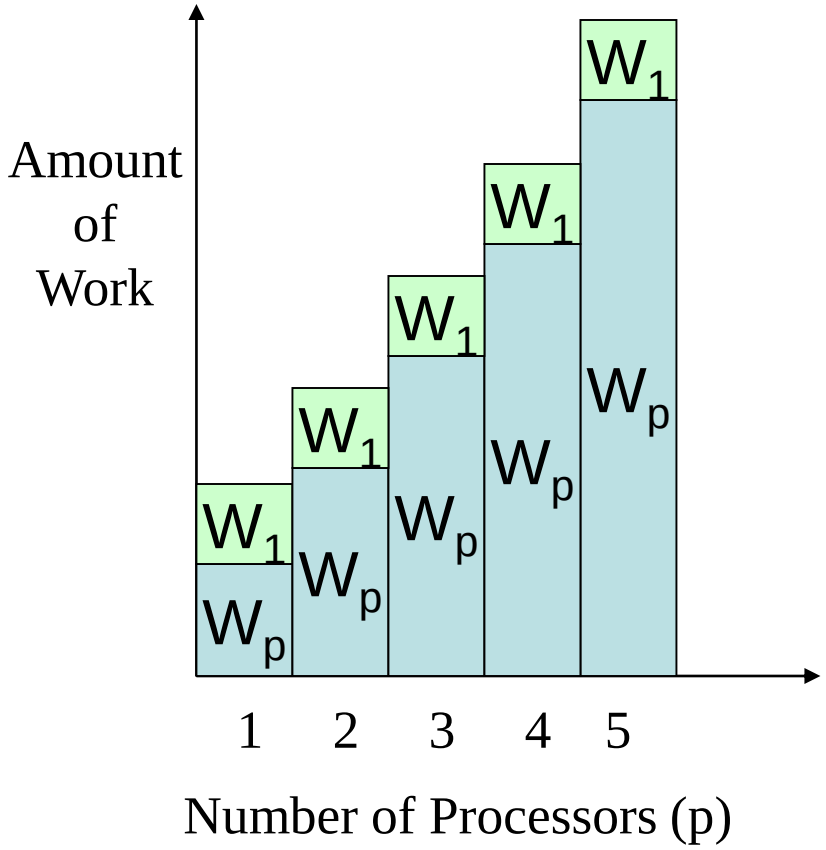


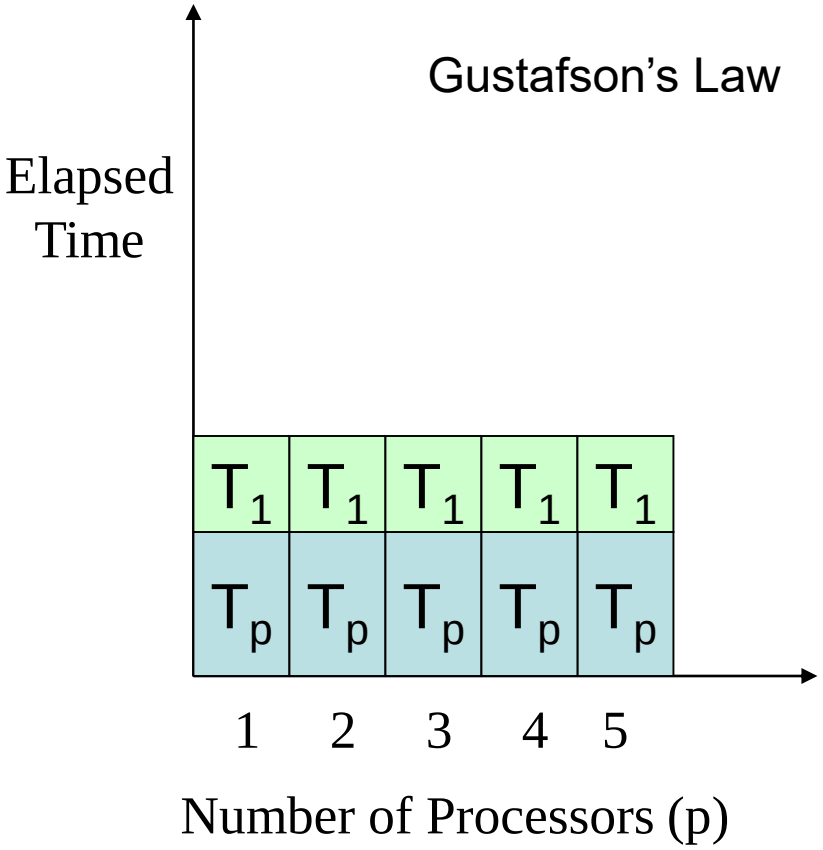


Scaling Characteristics of Parallel Programs

Scalability - the ability to maintain performance gains when system and/or problem size increase

- weak scaling - how the processing time varies with the number of processors for **a fixed problem size per processor**





Thank you!

Branislav Jansík