The University of Michigan Center for Advanced Computing

Andy Caird

acaird@umich.edu

The CAC

What is the Center for Advanced Computing?

- we provide high-performance, cluster-based computing to the research community of the University of Michigan

The CAC

What is the Center for Advanced Computing?

- we provide high-performance, cluster-based computing to the research community of the University of Michigan
- we do this by making 4 clusters available to researchers at U of M.

The Hardware

Our Fast-Network Cluster: morpheus

- 104 nodes, 208 processors
- 38 nodes are faculty owned
- AMD Athlon CPUs (1.4GHz, 2.0GHz)
- 200 GB RAM (aggregate)
- 4.5 TB scratch disk space (aggregate)
- 0.5 TB home disk space (shared via NFS)
- Myrinet connected
- RedHat Linux 7.2

The Hardware

Our Biggest Cluster: nyx

- 224 nodes, 448 processors
- 64 nodes are faculty owned
- AMD Opteron CPUs (1.4GHz, 1.8GHz)
- 628 GB RAM (aggregate)
- 8.0 TB scratch disk space (aggregate)
- 2.0 TB home disk space (shared via NFS)
- Gigabit Ethernet (Force-10) connected
- SuSE Linux Enterprise Server 8.1

The Hardware

Our Apple Cluster: eliza

- 64 nodes, 128 processors
- all purchased by NPACI
- Apple G5 CPUs (2.0GHz)
- 128 GB RAM (aggregate)
- 4.5 TB scratch disk space (aggregate)
- 3.0 TB home disk space (shared via NFS)
- Gigabit Ethernet (Force-10) connected
- Mac OS X 10.3

The Hardware

Our Faculty-Owned Cluster: aon

- 160 nodes, 320 processors
- all 160 nodes are faculty owned
- Apple G5 CPUs (2.0GHz)
- 160 GB RAM (aggregate)
- 11 TB scratch disk space (aggregate)
- 1.5 TB home disk space (shared via NFS)
- HP ProCurve 10/100 connected
- Mac OS X 10.3

The Hardware

The Totals

- 552 nodes, 1104 processors
- 1.32 TB RAM
- 28 TB scratch disk space
- 7 TB home disk space shared via NFS



nyx and eliza: 288 nodes, 12 racks, 756 RAM, 17TB disk

The Networking

Myrinet

- very high-speed network; > 2Gbps
- very low-latency network; $\sim 8\mu$sec
- very expensive network
- requires PCI card in each node
- PCI card is fibre connected to switch

The Networking

Force-10

- high-speed network; 1Gbps
- normal Ethernet latency; $\sim 60\mu$sec
- full line rate between all ports
- uses standard gigabit network cards
- Using large network frames, we get about 940Mbps between nodes

The Networking

Other

- we have 10/100 switches around
- faculty-purchased, serial codes
- HP ProCurve, Linksys

The System Software

Operating Systems

- Linux
  - for us, pretty much based on cost
  - we have a bias towards RedHat, but will probably use SuSE because of UM's licensing

The System Software

Operating Systems

- Linux
  - for us, pretty much based on cost
  - we have a bias towards RedHat, but will probably use SuSE because of UM's licensing
- Mac OS X
  - we've been really happy with it
  - in clusters, it's just Unix with lots of instrumentation (I^2C), fast processors, and really pretty cases

The System Software

Queuing

- PBSPro
  - more stable than OpenPBS (but we haven't tried Torque)
  - good educational pricing
  - open enough to let us use the Maui scheduler
- Maui
  - better than PBSPro's scheduler

The System Software

Compilers

- Portland Group compilers on Linux (C and Fortran)
- IBM Compilers on Mac OS X (Fortran)
- gcc on Linux and Mac

The System Software

Parallel Libraries

- MPI
  - Message Passing Interface
  - we use LAM and MPI/CH

The Reason

Why do people use these systems?

- Fast individual processors
- Large amounts of memory in individual nodes
- Fast networks
- Dedicated access for days or weeks at a time
- “Fair” access managed by queuing system
- Certain problems can be parallelized

The Reason

What is “parallelization”?

- The “problem” is divided between multiple nodes
- Each node does its part, and the answer is gathered together at the end

Much of engineering and science is discrete representations of piece-wise continuous functions $\rightarrow$ matrix math

The Example

Given a matrix-vector multiply $\mathcal{M}\vec{v} = \vec{A}$, or

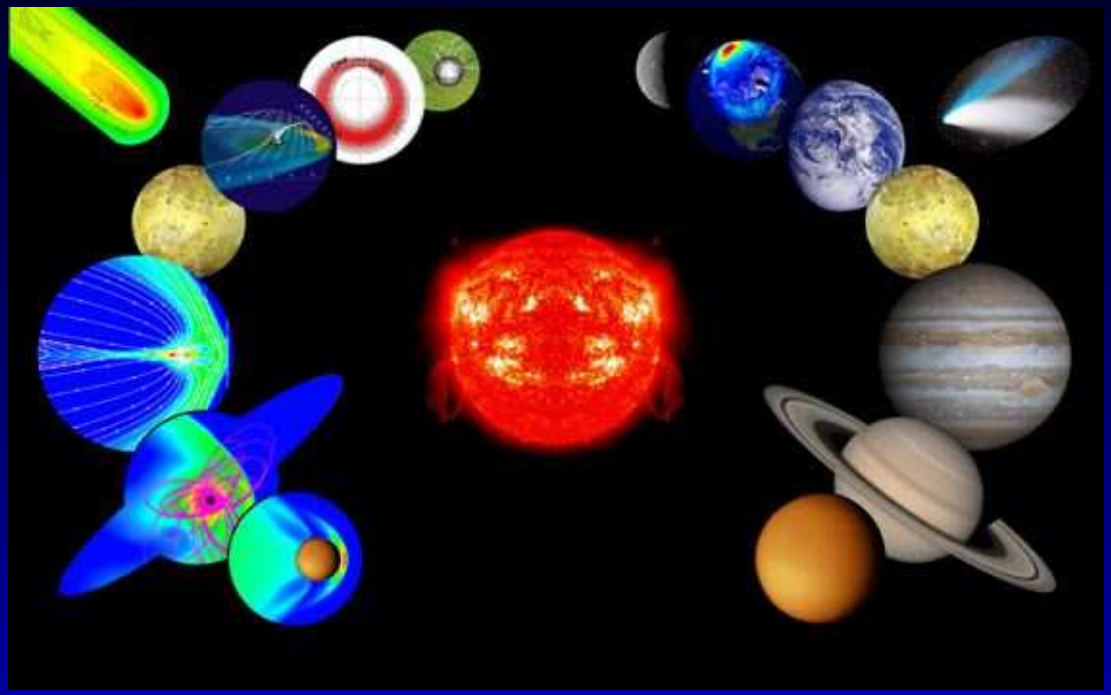
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \cdot e & + & b \cdot f \\ c \cdot e & + & d \cdot f \end{bmatrix}$$

It's clear that $\vec{A}$ depends only on the first row of $\mathcal{M}$ and $\vec{v}$.
A simplistic parallelization of a matrix multiply is to divide $\mathcal{M}$ into a series of row vectors $\vec{M}_r$ and give each processor P_i the vector $\vec{v}$ and one of the row vectors, $\vec{M}_{r=i}$.

The Usage

- Over 205 unique users of Linux clusters in 2004 and 2005 (up to 2/15)
- About 80% utilization of clusters (this is improving)
- The average wait time in the queue is about 7 hours, but this depends on what you can access

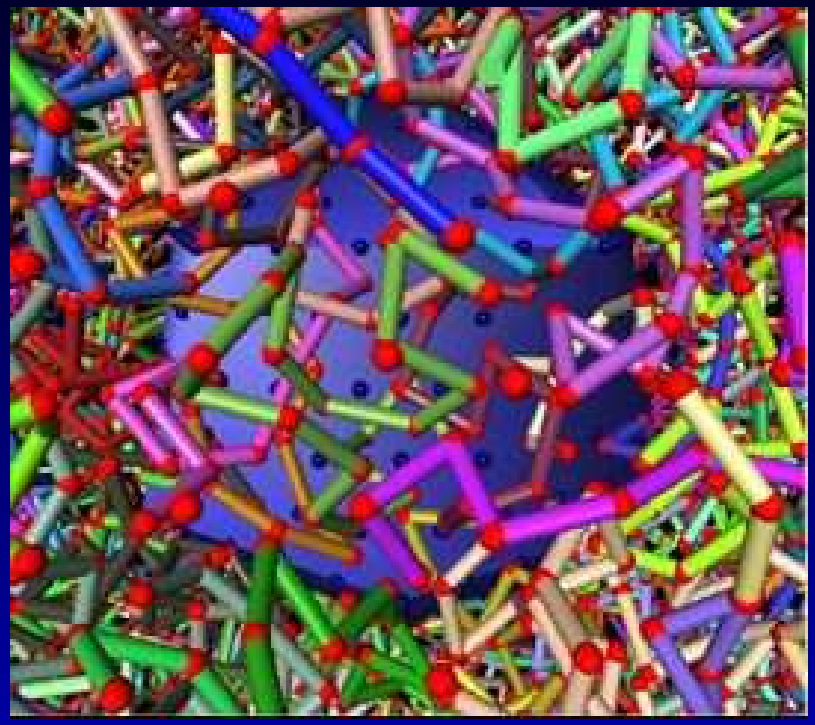
The Users: CSEM



CSEM: Center for Space Environment Modeling

The goal of CSEM is to develop high-performance computational models to describe and predict hazardous conditions in the near-earth space environment. The models must run considerably faster than real time on parallel computers.

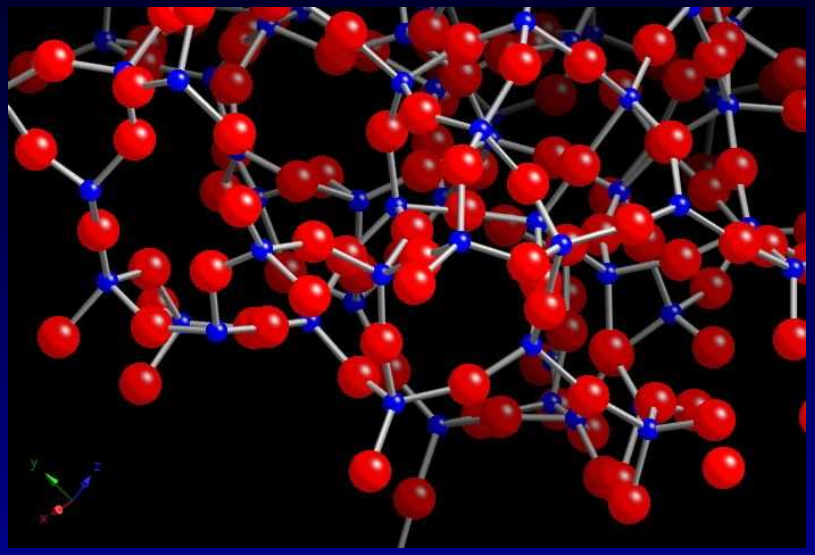
The Users: Glotzer



The Glotzer Group (ChemE)

The Glotzer group is discovering ways to control the dispersion of nanoparticles in polymer melts and blends by using molecular dynamics, Monte Carlo, and finite-difference and finite-element methods, among other techniques.

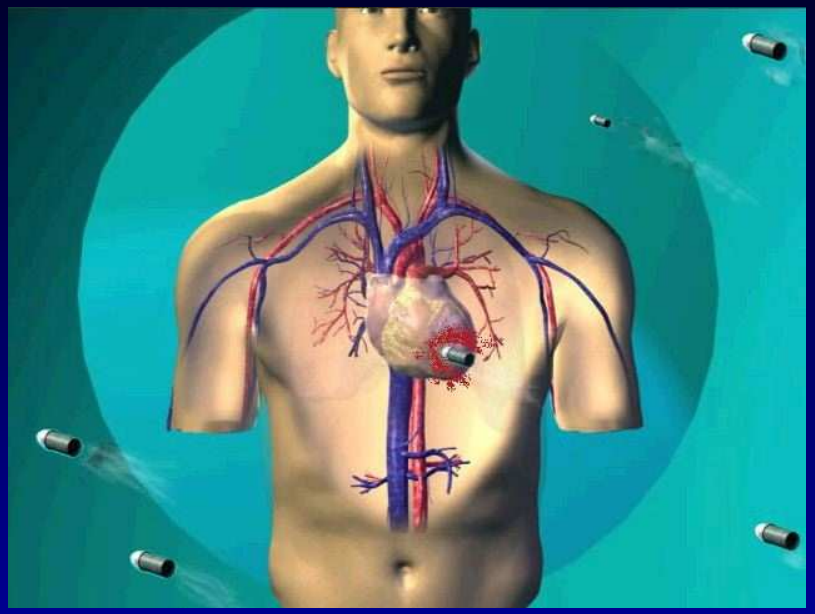
The Users: Keiffer



John Keiffer (MSE)

My objective is to derive an analytical framework for the characterization of structure-property relationships in porous materials using Molecular Dynamics simulations.

The Users: Virtual Soldier



Michigan Center for Biological Information

The goal of this project is the construction of a complete simulation of the human thorax to provide capabilities to diagnose battlefield injuries, specifically ballistic wounds to the heart.

The Administration

- PXE/DHCP
  - we use PXE, DHCP, and Kickstart/AutoYast to load Linux nodes
  - we use NetBoot to load the Apple nodes

The Administration

- PXE/DHCP
  - we use PXE, DHCP, and Kickstart/AutoYast to load Linux nodes
  - we use NetBoot to load the Apple nodes
- radmin
  - we use radmin on both Macs and Linux machines to keep them up to date

The Administration

- PXE/DHCP
  - we use PXE, DHCP, and Kickstart/AutoYast to load Linux nodes
  - we use NetBoot to load the Apple nodes
- radmin
  - we use radmin on both Macs and Linux machines to keep them up to date
- Kerberos
  - we run our own Kerberos realm to provide authentication
  - we needed this so we could support users from outside UM during the NPACI years

The Executives

The direction of the CAC is set by an executive committee made up of six members of the College of Engineering:

- Dr. Kenneth Powell, Aerospace Engineering, *Director of the CAC*
- Dr. Darren DeZeeuw, Research Scientist Space Physics Research Lab
- Dr. Michael Falk, Materials Science Engineering
- Dr. William Martin, Nuclear Engineering
- Dr. Kazuhiro Saitou, Mechanical Engineering
- Dr. Quentin Stout, Electrical Engineering and Computer Science

The Staff

Systems Staff

- Andy Caird, Systems Manager
- Matthew Britt, Systems Administrator
- Brock Palen, Student Systems Administrator

Applications and Programming Staff

- Dr. David Woodcock
- Dr. Abhijit Bose

The Staff



`nyx`, `eliza`, Andy, David, Matt, and Brock

The Reason this is Hard

Space

- we average 25 1U nodes per rack + 4 1U PDUs = 29U (of 42U/rack)
- 128 nodes ≈ 6 racks (compute nodes + head node + storage)

The Reason this is Hard

Space

- we average 25 1U nodes per rack + 4 1U PDUs = 29U (of 42U/rack)
- 128 nodes $\approx$ 6 racks (compute nodes + head node + storage)

Power

- we allow $\sim$ 3A per node (6 per 20A PDU)
- 128 nodes $\approx$ 400A (20 20A circuits)

The Reason this is Hard

Space

- we average 25 1U nodes per rack + 4 1U PDUs = 29U (of 42U/rack)
- 128 nodes $\approx$ 6 racks (compute nodes + head node + storage)

Power

- we allow $\sim$ 3A per node (6 per 20A PDU)
- 128 nodes $\approx$ 400A (20 20A circuits)

Cooling

- see power - most of that goes to heat

The Reason this is Hard

Breadth of Use

- We try to support people from all over U of M
- Different software needs (FEA to gene matching)
- Different hardware needs (CPU, memory, network)

The End

For more information about what the CAC does, if and how we can help you do your research or run your cluster:

`cac-info@umich.edu`

`http://cac.engin.umich.edu`