Supercomputing and Visualization

Russ Miller, Director
Center for Computational Research

"Top 10 Worldwide Supercomputing Center"
- www.gapcon.com

University at Buffalo
*The State University of New York*

Outline

- **Pervasive Computing**
- **Computer Trends**
- **Definition of Supercomputer**
- **Overview of Center for Computational Research**
- **Sample CCR Projects**
- **Center of Excellence in Bioinformatics**
- **H.S. Workshop**

Take-Home Message

- **Computers play an important role in your life**
- **Currently ~10 processors per person**
- **Working with computers can be fun and exciting**

Computers Touch Every Aspect of Our Life....
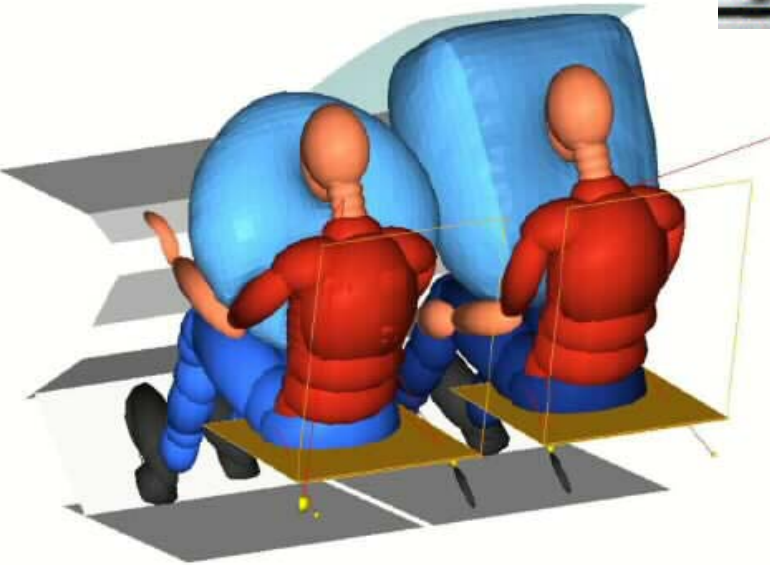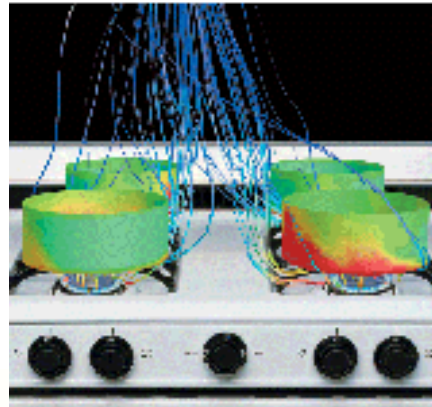
.... including entertainment

© 2002 Twentieth Century Fox. All rights reserved. Not For Sale Or Duplication.

▶ Even during an Ice Age, things can get hot for Sid the sloth.

Computers are used in Many Professions

- **Science and Engineering**
  - **Physics, Chemistry, Biology**
  - **Aerospace, Mechanical, Civil, Environmental**
- **Architecture**
  - **Building and Bridge Design**
- **Computer Animation**
  - **Cartoons, Movies, Advertising**
  - **Games (Playstation, Nintendo, PC games, etc)**
- **Graphic Arts/Design**
- **Computer programmers**

What is a CPU?

It's the computer's brain - it's the main *processor*

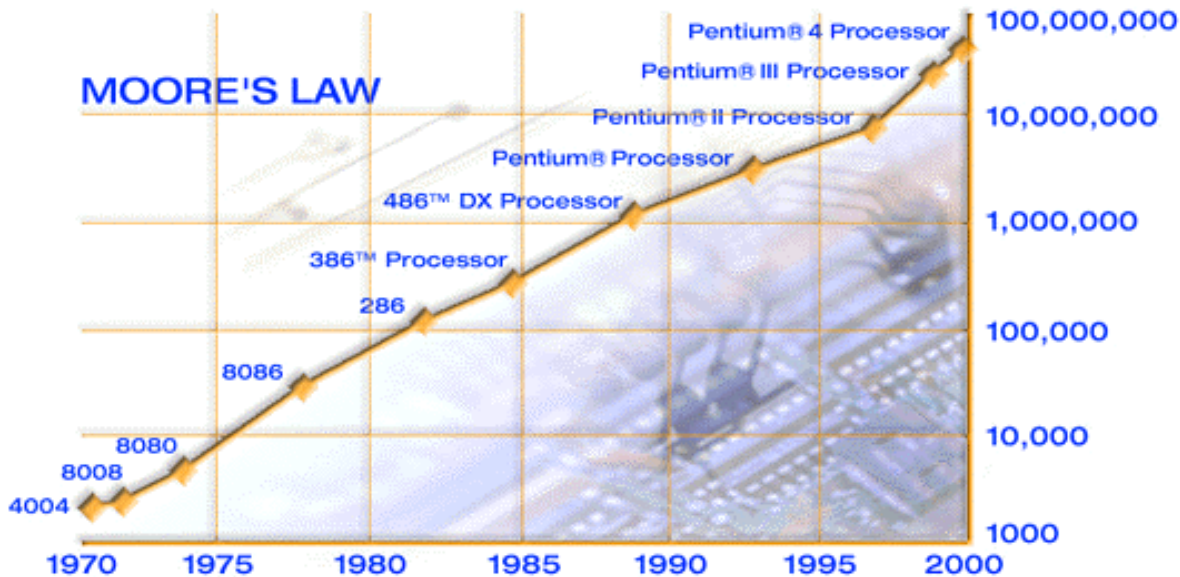**CPU stands for Central Processing Unit**

Gordon E. Moore

- **Co-Founder of Intel**
- **Predicted (1965/75) that transistor density would double every 12/18 months**
- **Processing speed doubling every 18 mos.**
- **Disk storage doubling every 12 mos.**
- **Aggregate bandwidth doubling every 9 mos.**

Gordon E. Moore

MOORE'S LAW

transistors

- Pentium® 4 Processor — 100,000,000
- Pentium® III Processor
- Pentium® II Processor — 10,000,000
- Pentium® Processor
- 486™ DX Processor — 1,000,000
- 386™ Processor
- 286 — 100,000
- 8086
- 8080 — 10,000
- 8008
- 4004 — 1000

1970 1975 1980 1985 1990 1995 2000

- **A computation that took 1 year to run on a PC in 1985 would only take 5 mins to run on a PC today!**
- **A computation that runs in 2 hours on a PC today would have taken 24 years to run on a PC in 1985!**
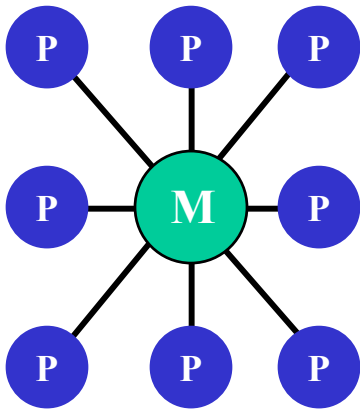
What is a Parallel Computer?

**A computer that contains
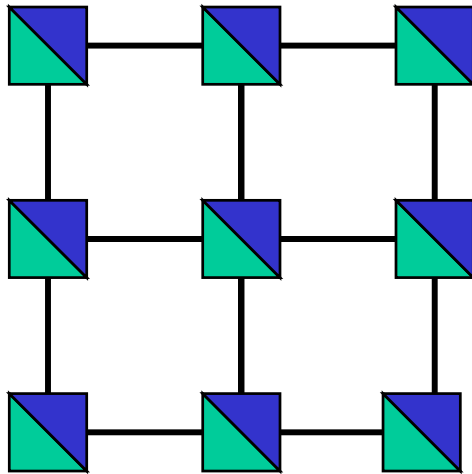more than 1 processor (CPU)**

Why are they used?

**To solve problems faster than they
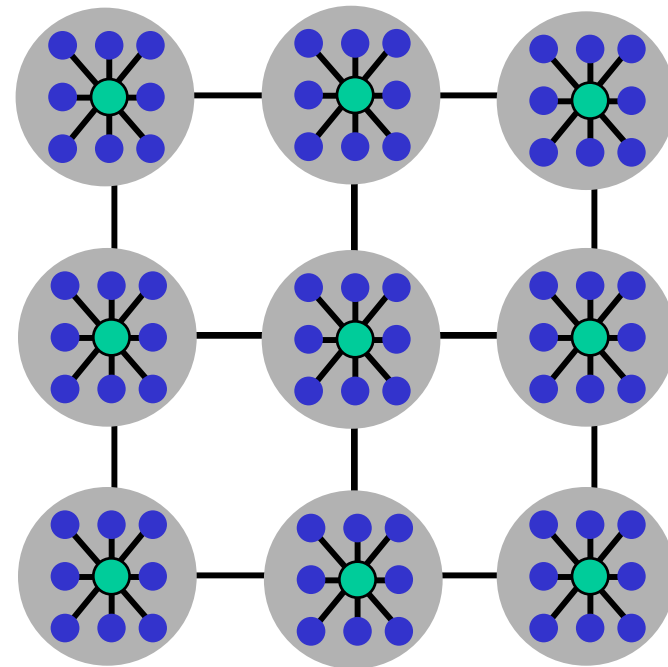could be solved using only 1 processor**

Parallel Computing Architectures

Shared Memory   **Distributed Memory**   **Distributed-Shared Memory**
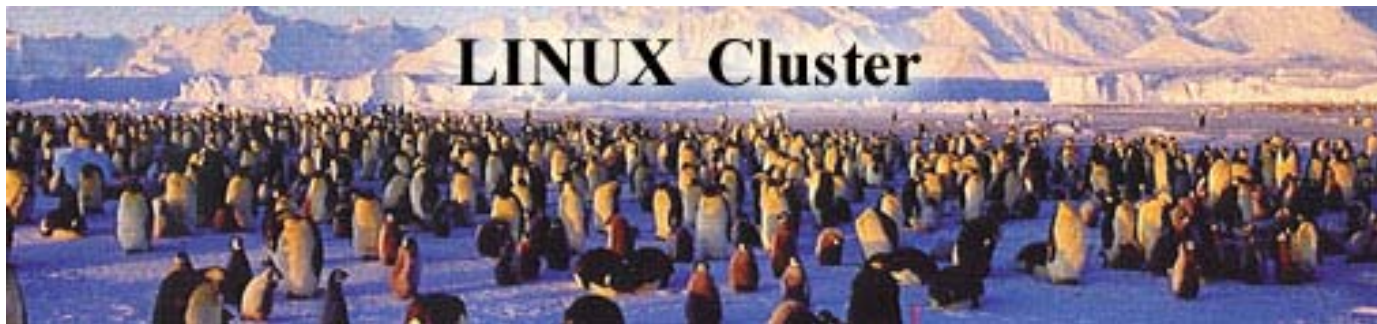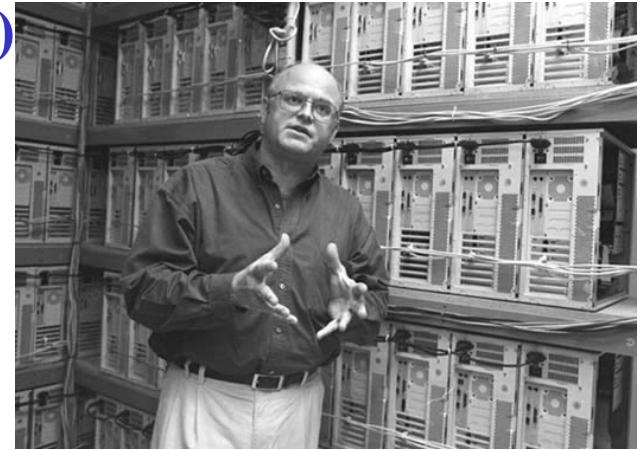
What is a (Beowulf) Cluster?

- **Industry Standard Hardware and Software**
  - ❑ **PC-Based Components (Intel or AMD)**
  - ❑ **Ethernet or Myrinet**
  - ❑ **Linux, PBS, MPI**
 - ❑ **"Commodity Off-The-Shelf" (COTS)**
- **Operates as a Single System**
- **Rivals Performance of Traditional Supercomputer at a Fraction of the Price**

Thomas Sterling Caltech

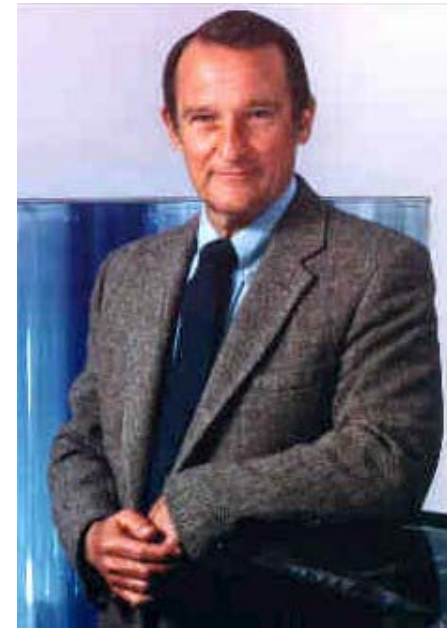

LINUX Cluster

What is a Supercomputer?

- **Fastest computers at any point in time**
- **Used to solve large and complex problems**
- **Machines 1000 times faster than a PC**
- **Machines 10 times slower than what you need to solve the most challenging problems**
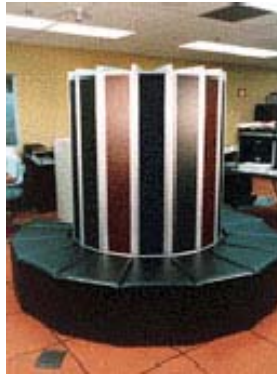
Cray1 - 1976

Seymour Cray
1925-1996

**"Seymour Cray is the Thomas Edison of the supercomputing industry"
- Larry L. Smarr**

Example

If you wanted to know what the weather will be like **tomorrow**, you could ...

Solve the problem at home on your PC and wait **one month** to get the answer

or

Solve the problem on a supercomputer and have the answer in **one hour!**

Fastest Computers

Year	Mach	Procs	GFlops		Year	Mach	Procs	GFlops
1976	Cray 1	1	0.1		1993	Cray T3D	1024	152
1982	Cray X-MP	4	0.9		1994	Fujitsu VPP	140	236
1986	Cray 2	4	2		1996	Hitachi SR2	2048	368
1989	Cray Y-MP	8	2.7		1997	Intel ASCI-R	9152	1830
1989	TMC CM-2	8192	28		1999	SGI ASCI-BM	6144	3072
1992	TMC CM-5	1024	131		2000	IBM ASCI-W	8192	12,288
A 1-year calc in 1980 = 5.4 sec today **A 1990 HPC = a laptop today**					2002	NEC E.S.	5120	40,960

Earth Simulator in Japan (NEC Vector Supercomputer)
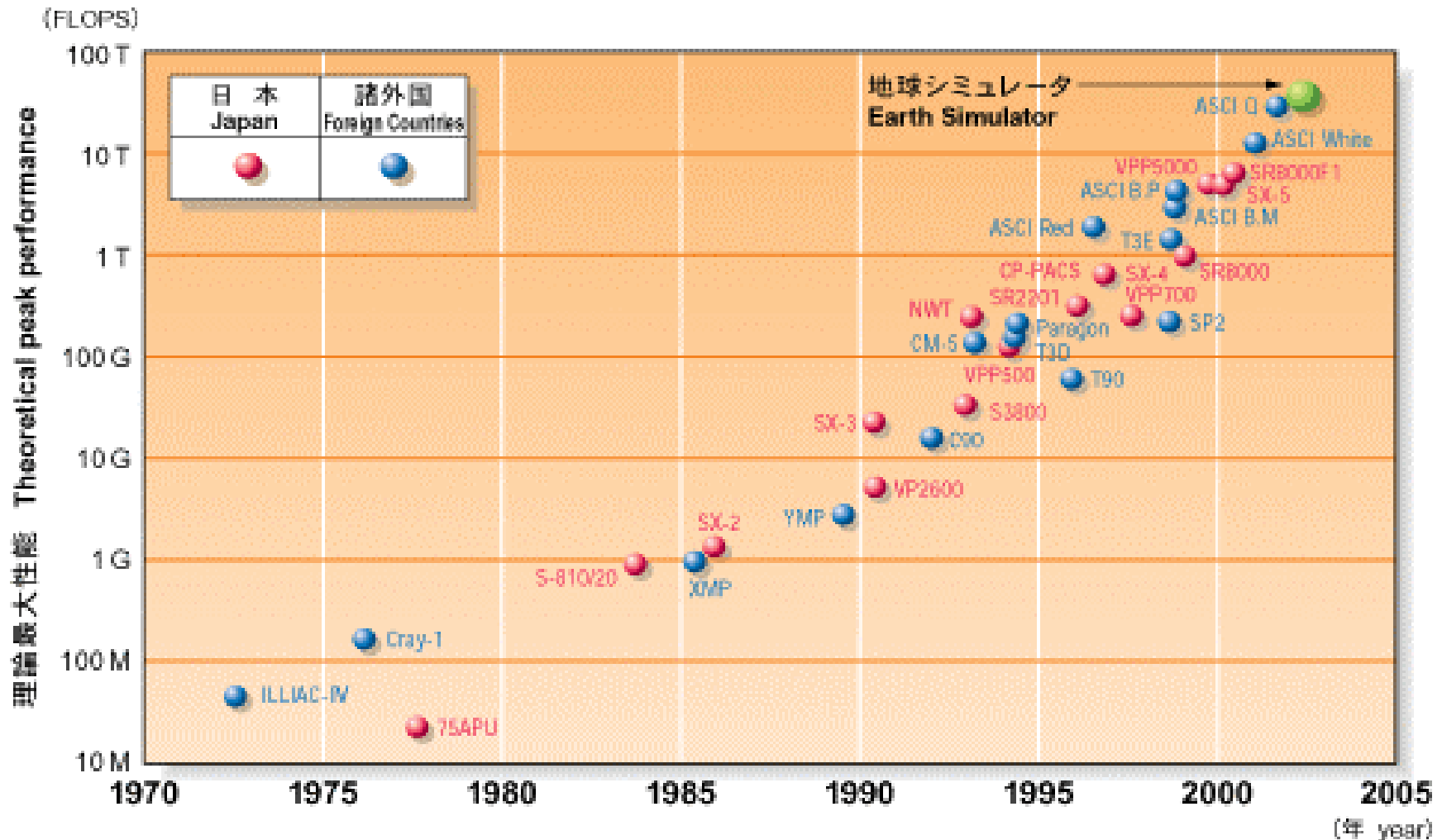
Earth Simulator

- **40TFlops Peak**
- **Homogeneous, Centralized, Proprietary, Vector**
- **Expensive!**
- **CFD-Weather, Climate, Earthquake**
- **640 NEC SX/6 Nodes (5120 CPUs)**
- **Footprint = 4 tennis courts**
- **$6M/year in power**

El Niño / La Niña

abnormal weather — Chilly summer, Warm winter, Heavy rain, Heavy snow, Drought

global warming — Burning of fossil fuel, Deforestation

crustal movement — Earthquake·Volcanic activity

acid rain — Air pollution

Center for Computational Research

- **High-Performance Computing and High-End Visualization**
  - ❑ **70 (40+ active) Research Groups in 27 Depts**
  - ❑ **13 Local Companies**
  - ❑ **10 Local Institutions**
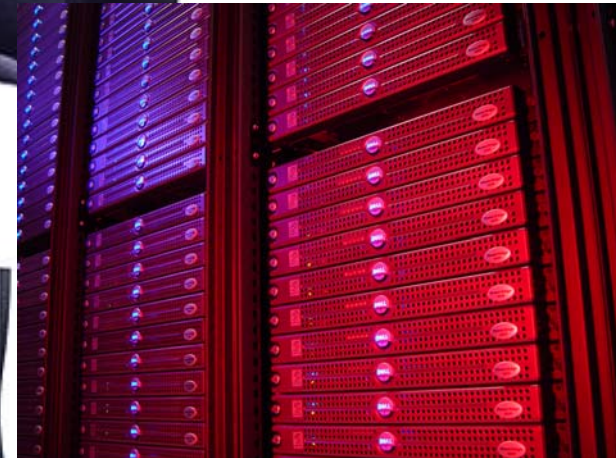 - ❑ **External Funds: $108M**
  - ❑ **Vendor Contributions: $41M**
- **Deliverables**
  - ❑ **350 Publications and Presentations**
  - ❑ **Hardware, Software, Algorithms, etc**
- **Training**
  - ❑ **Workshops**
  - ❑ **Courses**
  - ❑ **Degree Programs**

Computational Resources

Dell Linux Cluster - #22 in World
- 600 P4 Processors (2.4 GHz)
- 600 GB RAM; 40 TB Disk

Dell Linux Cluster - #187 in World
- 4036 Processors (PIII 1.2 GHz)
- 2TB RAM; 160TB Disk; 16TB RD
- Private Use

SGI Origin3800
- 64 Processors (400 MHz)
- 32 GB RAM; 400 GB Disk

IBM RS/6000 SP
- 78 Processors
- 26 GB RAM; 640 GB Disk

Sun Microsystems Cluster
- 48 Sun Ultra 5s (333MHz)
- 16 Dual Sunblades (750MHz)
- 30 GB RAM, Myrinet

SGI Intel Linux Cluster
- 150 PIII Processors (1 GHz)
- 75 GB RAM, 2.5 TB Disk Storage

Apex Bioinformatics System
- Sun V880 (3), 6800, 280R (2), PIIIs
- Sun 3960: 7 TB Disk Storage

HP/Compaq SAN (3/2003)
- 25 TB Disk; 250 TB Tape

Sample Computational Research

- **Computational Chemistry** (King, Kofke, Coppens, Furlani, Tilson, Lund, Swihart, Ruckenstein, Garvey)
  - ❑ **Algorithm development & simulations**
- **Groundwater Flow Modeling** (Rabideau, Jankovic, Becker, Flewelling)
  - ❑ **Predict contaminant flow in groundwater & possible migration into streams and lakes**
- **Geophysical Mass Flows** (Patra, Sheridan, Pitman, Bursik, Jones, Winer)
  - ❑ **Study of geophysical mass flows for risk assessment of lava flows and mudslides**
- **Bioinformatics** (Zhou, Miller, Hu, Szyperski – NIH Consortium, HWI)
  - ❑ **Protein Folding: computer simulations to understand the 3D structure of proteins**
  - ❑ **Structural Biology; Pharmacology**
- **Computational Fluid Dynamics** (Madnia, DesJardin, Lordi, Taulbee)
  - ❑ **Modeling turbulent flows and combustion to improve design of chemical reactors, turbine engines, and airplanes**
- **Physics** (Jones, Sen)
  - ❑ **Many-body phenomena in condensed matter physics**
- **Chemical Reactions** (Mountziaris)
- **Molecular Simulation** (Errington)

Visualization Resources

- **Fakespace ImmersaDesk R2**
  - ❑ **Portable 3D Device**
- **Tiled-Display Wall**
  - ❑ **20 NEC projectors: 15.7M pixels**
  - ❑ **Screen is 11'×7'**
  - ❑ **Dell PCs with Myrinet2000**
- **Access Grid Node**
  - ❑ **Group-to-Group Communication**
  - ❑ **Commodity components**
- **SGI Reality Center 3300W**
  - ❑ **Dual Barco's on 8'×4' screen**
- **VREX VR-4200 Stereo Imaging Projector**
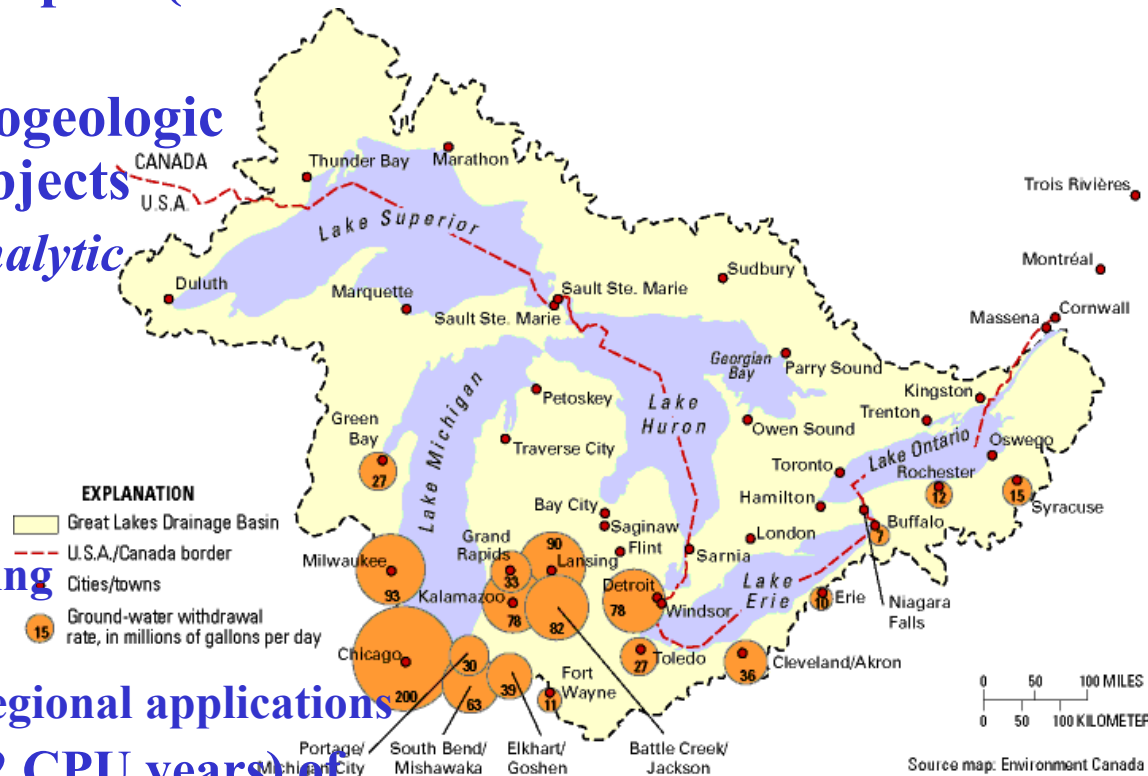 - ❑ **Portable projector works with PC**

Sample Visualization Areas

- **Computational Science** (Patra, Sheridan, Becker, Flewelling, Baker, Miller, Pitman)
  - ❑ **Simulation and modeling**
- **Urban Visualization and Simulation** (CCR)
  - ❑ **Public projects involving urban planning**
- **Medical Imaging** (Hoffmann, Bakshi, Glick, Miletich, Baker)
  - ❑ **Tools for pre-operative planning; predictive disease analysis**
- **Geographic Information Systems** (CCR, Bisantz, Llinas, Kesavadas, Green)
  - ❑ **Parallel data sourcing software**
- **Historical Reenactments** (Paley, Kesavadas, More)
  - ❑ **Faithful representations of previously existing scenarios**
- **Multimedia Presentations** (Anstey, Pape)
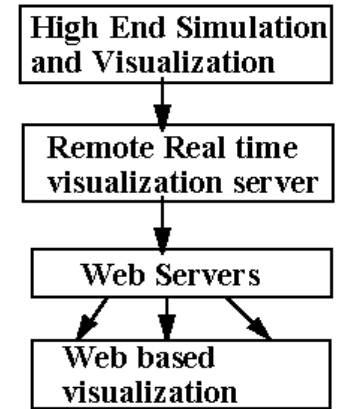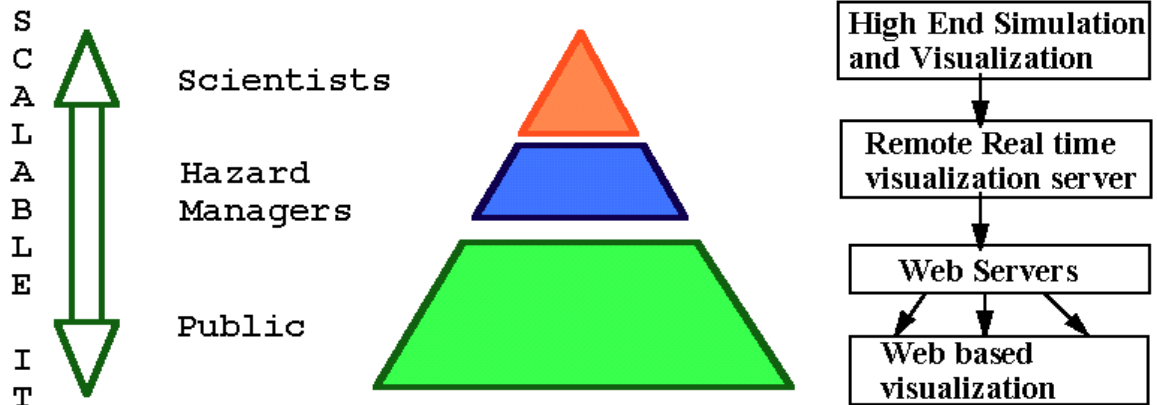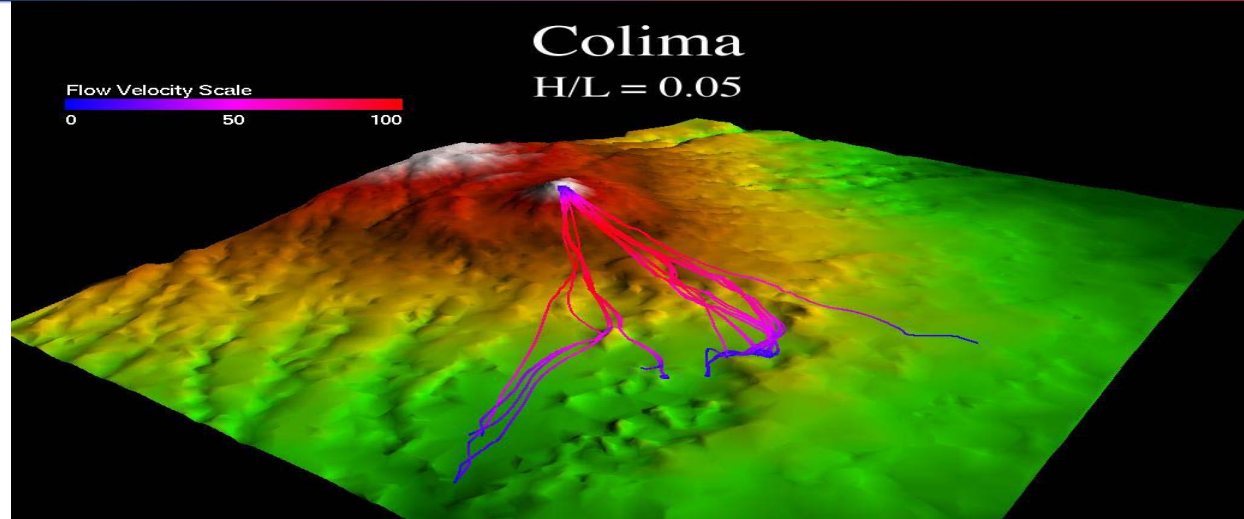 - ❑ **Networked, interactive, 3D activities**

Groundwater Flow Modeling

- **Regional-scale modeling of groundwater flow and contaminant transport (Great Lakes Region)**
- **Ability to include all hydrogeologic features as independent objects**
- **Current work is based on *Analytic Element Method***
- **Key features:**
  - **High precision**
  - **Highly parallel**
  - **Object-oriented programming**
  - **Intelligent user interface**
  - **GIS facilitates large-scale regional applications**
- **Utilized 10,661 CPU days (32 CPU years) of computing in past year on CCR's commodity clusters**

EXPLANATION

- Great Lakes Drainage Basin
- U.S.A./Canada border
- Cities/towns
- Ground-water withdrawal rate, in millions of gallons per day

Source map: Environment Canada

Risk Mitigation

Colima
H/L = 0.05

Flow Velocity Scale
0   50   100

- **Integrate information from several sources**
  - ❑ **Simulation results**
  - ❑ **Remote sensing**
  - ❑ **GIS data**
- **Develop realistic 3D models of geophysical mass flows**
- **Present information at user appropriate resolutions**

S C A L A B L E I T

Scientists

Hazard Managers

Public

High End Simulation and Visualization

Remote Real time visualization server

Web Servers

Web based visualization

Protein Folding

- **Ability of proteins to perform biological function is attributed to their 3-D structure.**

- **Protein folding problem refers to the challenge of predicting 3-D structure from amino-acid sequence.**

- **Solving the protein folding problem will impact drug design.**
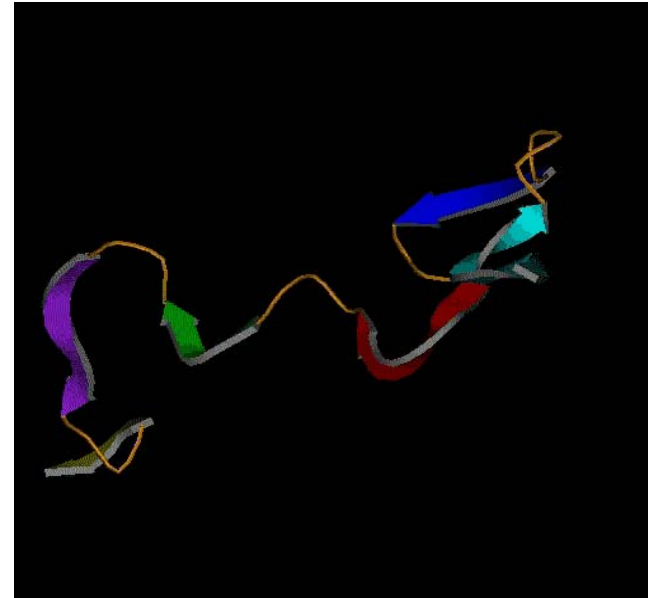
3D Medical Visualization App

- **Collaboration with Children's Hospital**
  - **Leading miniature access surgery center**
- **Application reads data output from a CT Scan**
- **Visualize multiple surfaces and volumes**
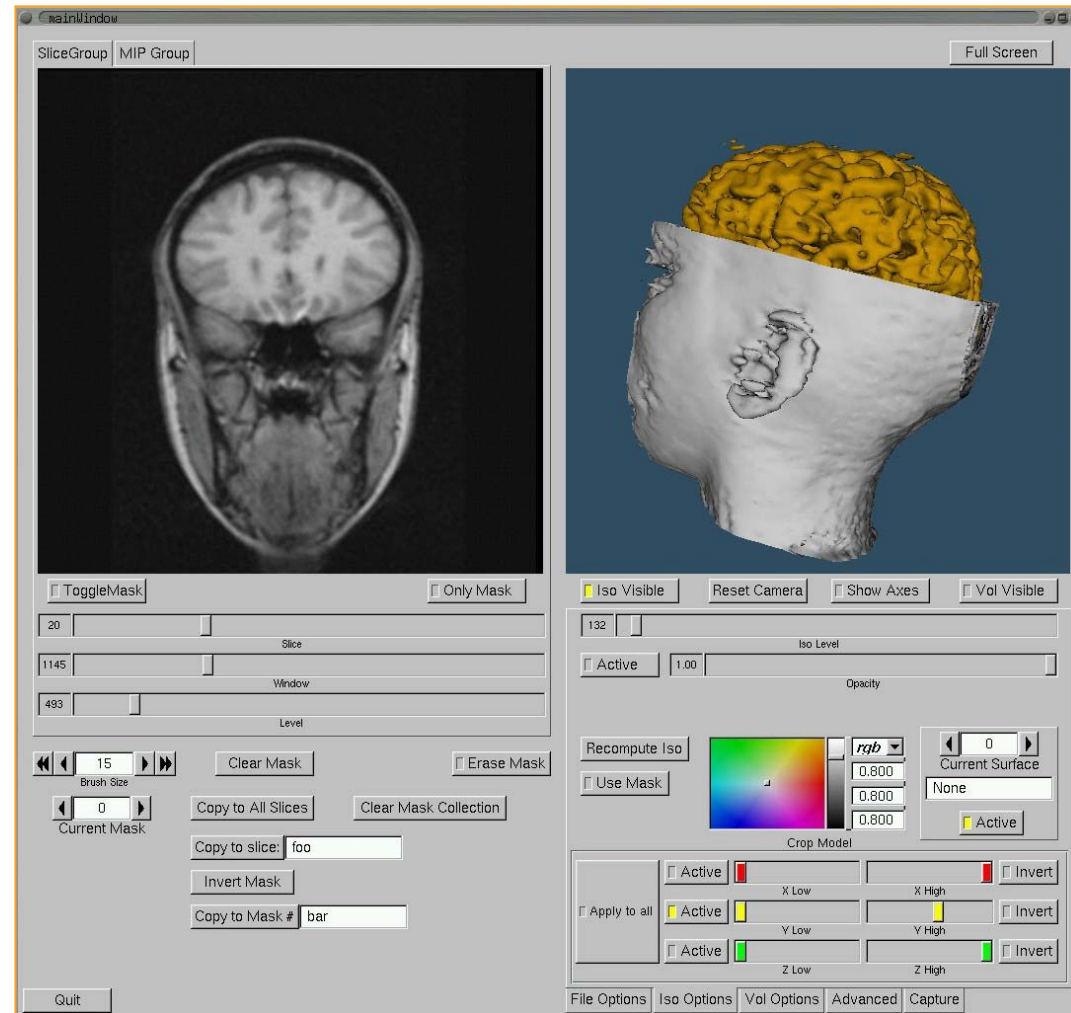- **Export images, movies or CAD representation of model**
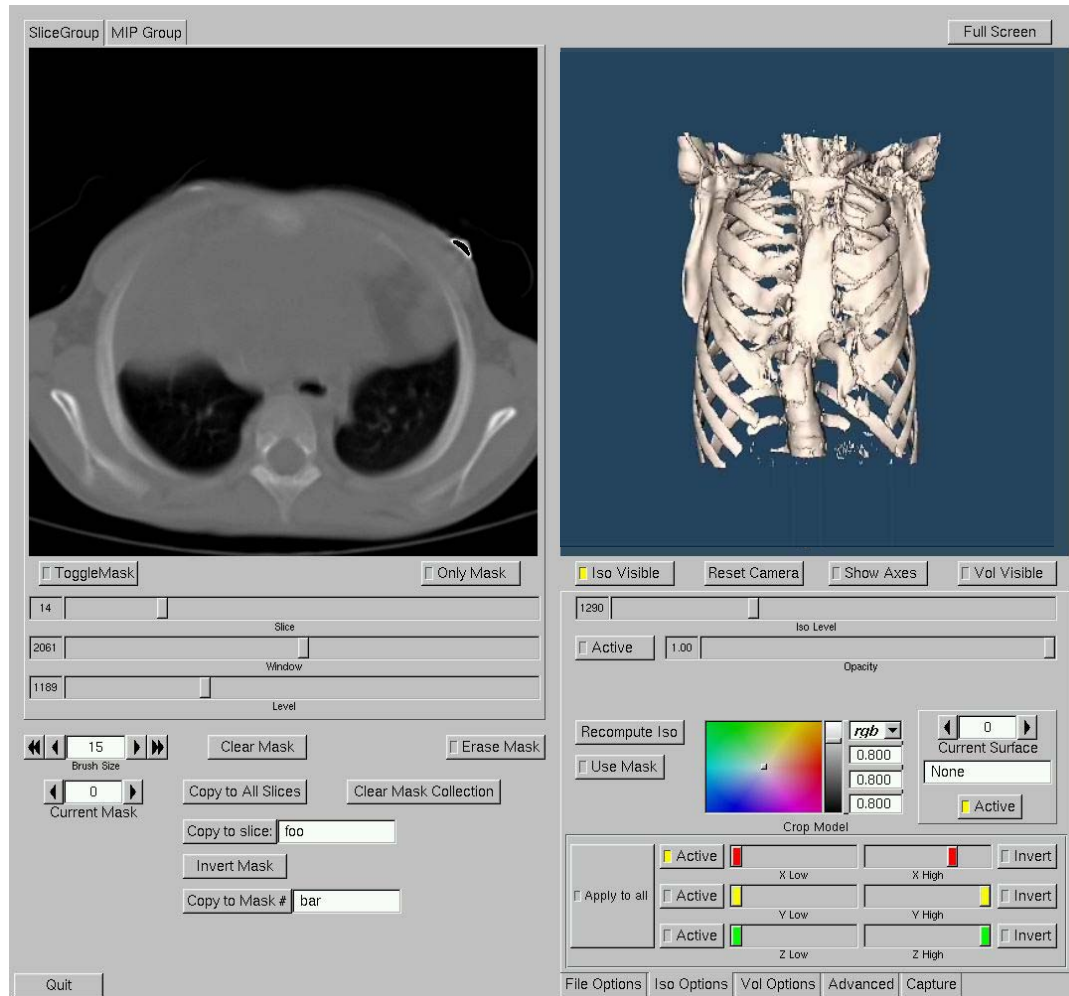
Multiple Sclerosis Project

- **Collaboration with Buffalo Neuroimaging Analysis Center (BNAC)**
  - **Developers of Avonex, drug of choice for treatment of MS**
- **MS Project examines patients and compares scans to healthy volunteers**

Peace Bridge Visualization

Proposed Options
- ❑ Relocate US plaza
- ❑ Build a 3-lane companion span, rehab existing bridge
- ❑ Build a six lane signature span

PHOTO AND STORY BY BRUCE JACKSON

Select WNY Synergies

- **IBC Digital**
  - Gov. Pataki Visit
  - Peace Bridge (Early & Current)
  - Buffalo-Niagara Medical Campus
  - Compute Cycles for Animation
- **Bergmann Associates**
  - Peace Bridge (Current)
  - NYS Thruway Toll Plaza
- **Azar & More**
  - Reenactment of 1901 Pan Am Exhibition
  - PHSCologram & Courses
  - Avid Digital Editing

- **Niagara College**
  - Start up
  - Peace Bridge (Current)
- **Hauptman-Woodward Medical Research Institute**
  - Computing
  - Collaboratory
- **The Children's Hospital of Buffalo**
  - Medical Visualization
- **Veridian**
  - Battlespace Management

Bioinformatics in Buffalo

"This Center [of Excellence in Bioinformatics] will, through the University of Buffalo's Center for Computational Research, create academic and industrial partnerships …"

- NYS Gov. George S. Pataki, January 2001

Gov. Pataki

Congressman Reynolds

Senator Clinton

WNY Biomedical Advances

- **PSA Test (screen for Prostate Cancer)**
- **Avonex: Interferon Treatment for Multiple Sclerosis**
- **Artificial Blood**
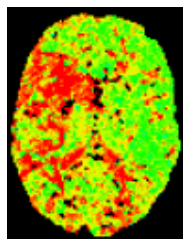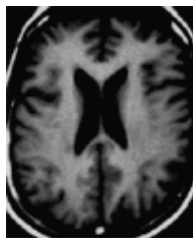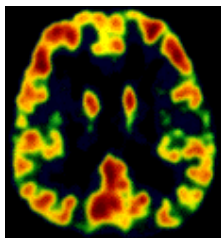- **Nicorette Gum**
- **Fetal Viability Test**
- **Implantable Pacemaker**
- **Edible Vaccine for Hepatitis C**
- **Timed-Release Insulin Therapy**
- **Anti-Arrythmia Therapy**
  - **Tarantula venom**

- **Direct Methods Structure Determination**
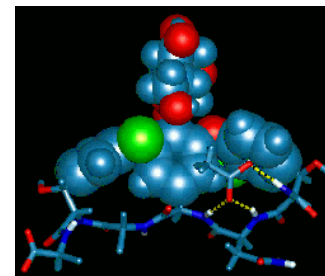 - **Listed on "Top Ten Algorithms of the 20th Century"**
  - **Vancomycin**
  - **Gramacidin A**
- **High Throughput Crystallization Method: Patented**
- **NIH National Genomics Center: Northeast Consortium**
- **Howard Hughes Medical Institute: Center for Genomics & Proteomics**

Bioinformatics in Buffalo

- **UB Center for Advanced Bioengineering & Biomedical Technologies**
  - ❑ **$1M/yr NYS**
  - ❑ **Med Tech for Product Dev & Commer.**
- **Center Disease Modeling & Therapy Discovery**
  - ❑ **UB, HWI, RPCI, Kaleida**
  - ❑ **$15.3M NYS**
  - ❑ **Software, device development, and drug therapies**
- **Buffalo Center of Excellence in Bioinformatics**
  - ❑ **UB, HWI, RPCI**
  - ❑ **$61M NYS**
  - ❑ **$10.6M Federal Government**
  - ❑ **$151 Corporate Funding**
  - ❑ **Significant Local Foundation Support**

Buffalo Center of Excellence in Bioinformatics

- Act as a *research*, *development*, *education*, and *economic resource* for industries based on bioinformatics, including information technology, biotech, and pharmaceuticals.

- Combine state-of-the-art *computational facilities* with *high-throughput experimental facilities* to enable the development of new medical treatments.

- Develop and exploit new algorithms for data acquisition, storage, management, and transmission.

Academic Programs

- **Bachelor's & Master's Program in Bioinformatics**
- **Related Disciplines**
  - ❑ Chemical Biology
  - ❑ Computational Chemistry
  - ❑ Environmental Analysis (Sloan Support)
  - ❑ Medical Informatics (Sloan Support)
- **Advanced Degrees under Development**
  - ❑ Pharmacometrics, Biophotonics
- **UB-HWI Department of Structural Biology**
- **Complementary Degrees**
  - ❑ Canisius College
  - ❑ Niagara University

Outreach

2003 H.S. Summer Workshop Bioinformatics

- **June 30 – July 11**
- **Perl Scripts**
- **Public Databases**
- **Filtering Results**
- **Graphics & Visualization**

- **Contact**
  - ☐ **Dr. Bruce Pitman (pitman@buffalo.edu)**

Lunch & Exhibition

miller@buffalo.edu
www.ccr.buffalo.edu

H.S. Program
pitman@buffalo.edu