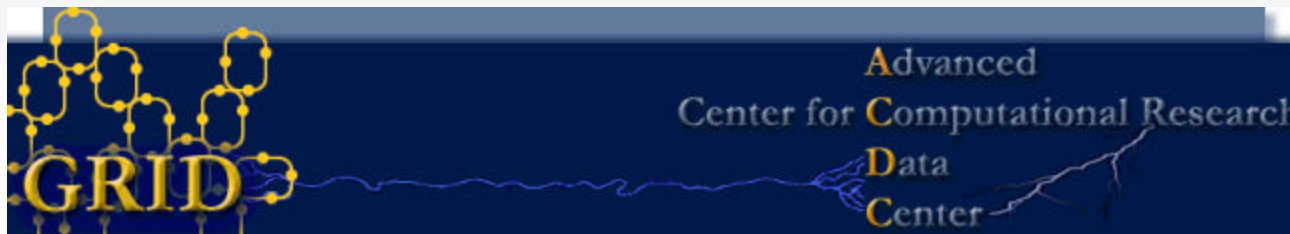
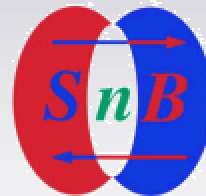


# The Center for Computational Research (CCR): An Overview

Russ Miller & Mark Green  
Center for Computational Research  
Computer Science & Engineering  
SUNY-Buffalo

Hauptman-Woodward Medical Inst

NSF, NIH, DOE,  
NIMA, NYS, HP



University at Buffalo

The State University of New York

# Biomedical Advances

■ **PSA Test (screen for Prostate Cancer)**

■ **Avonex: Interferon Treatment for Multiple Sclerosis**

■ **Artificial Blood**

■ **Nicorette Gum**

■ **Fetal Viability Test**

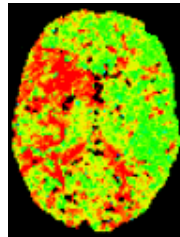
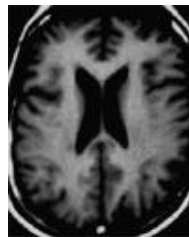
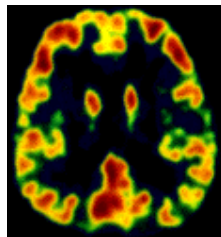
■ **Implantable Pacemaker**

■ **Edible Vaccine for Hepatitis C**

■ **Timed-Release Insulin Therapy**

■ **Anti-Arrhythmia Therapy**

□ **Tarantula venom**



■ **Direct Methods Structure Determination**

□ **Listed on “Top Ten Algorithms of the 20<sup>th</sup> Century”**

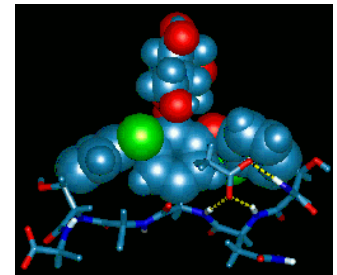
□ **Vancomycin**

□ **Gramacidin A**

■ **High Throughput Crystallization Method: Patented**

■ **NIH National Genomics Center: Northeast Consortium**

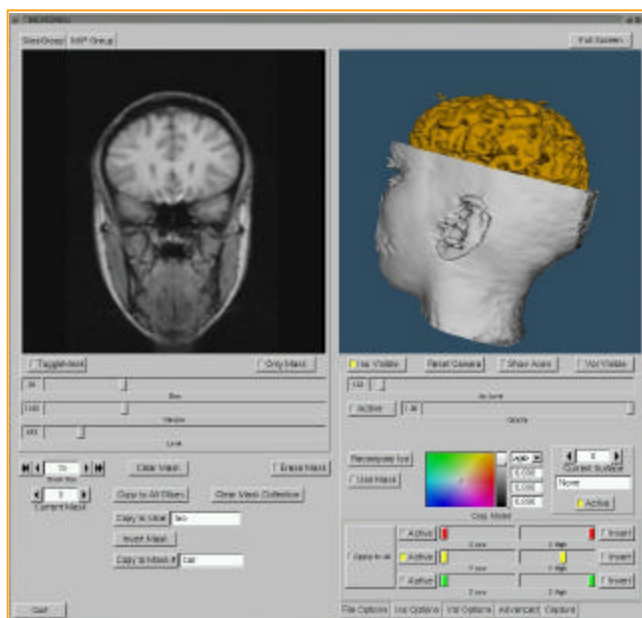
■ **Howard Hughes Medical Institute: Center for Genomics & Proteomics**



# Bioinformatics in Buffalo

## A \$360M Initiative

- **New York State: \$121M**
- **Federal Appropriations: \$13M**
- **Corporate: \$146**
- **Foundation: \$15M**
- **Grants & Contracts: \$64M**



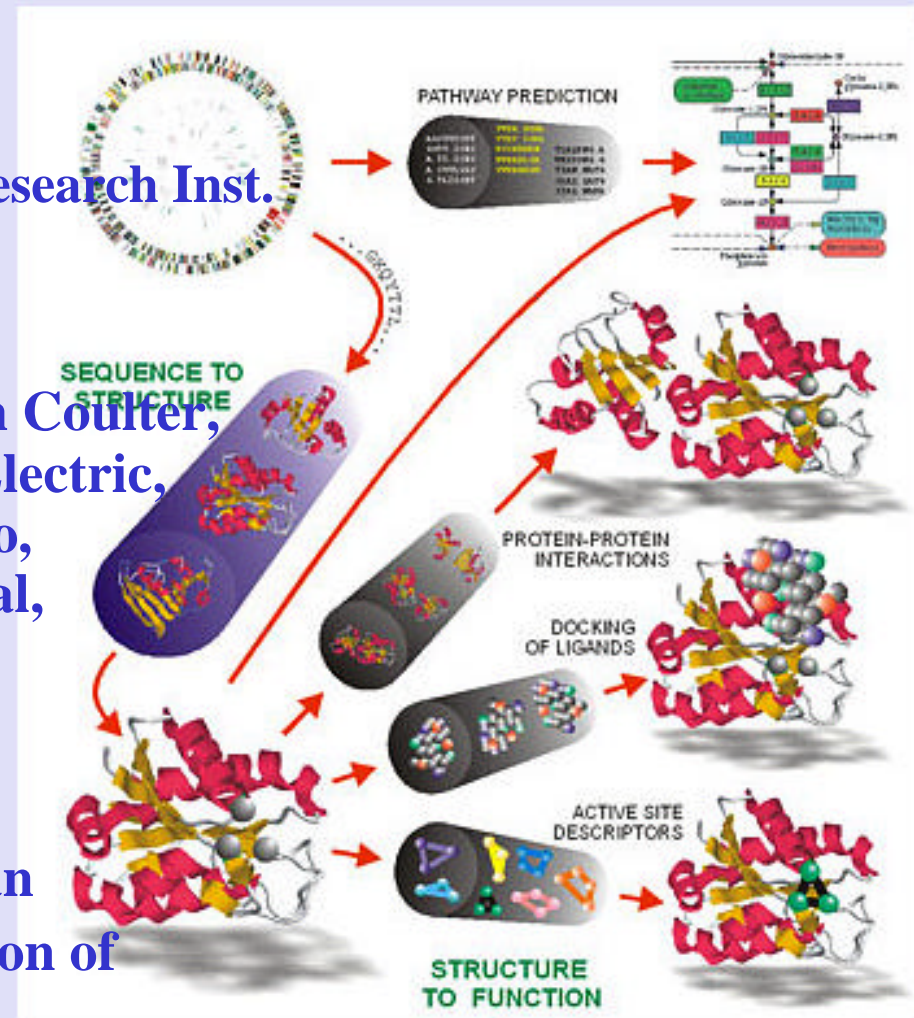
# Bioinformatics Partners

## ■ Lead Institutions

- ❑ University at Buffalo (UB)
- ❑ Hauptman-Woodward Medical Research Inst.
- ❑ Roswell Park Cancer Institute

## ■ Corporate Partners

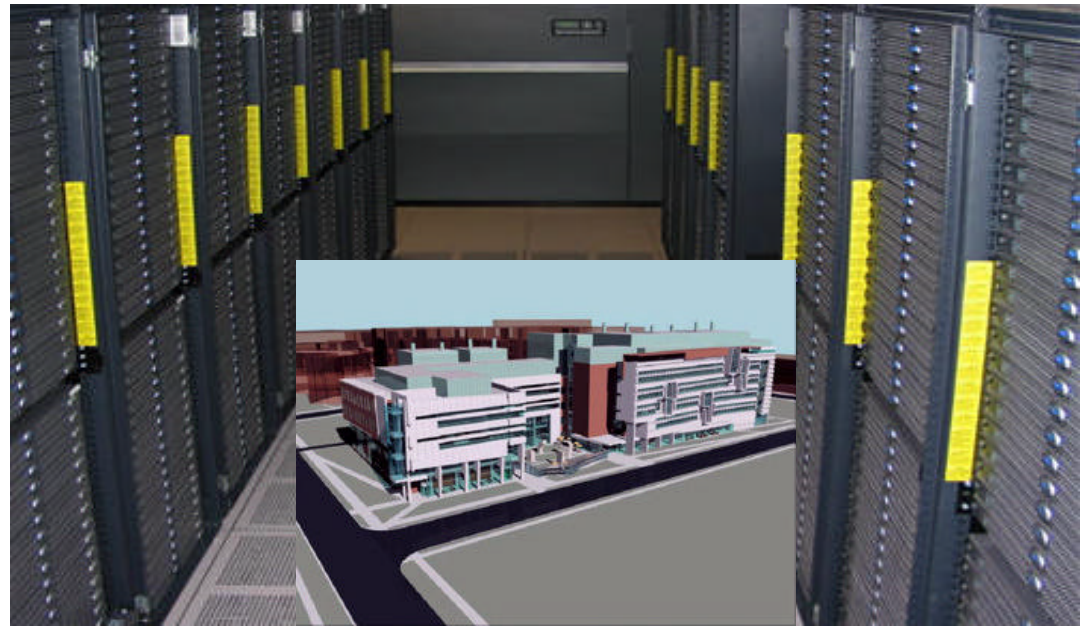
- ❑ Amersham Pharmacia, Beckman Coulter, Bristol Myers Squibb, General Electric, Human Genome Sciences, Immco, Invitrogen, Pfizer Pharmaceutical, Wyeth Lederle, Zeptomatrix
- ❑ Dell, HP, SGI, Stryker, Sun
- ❑ AT&T, Sloan Foundation
- ❑ InforMax, Q-Chem, 3M, Veridian
- ❑ BioPharma Ireland, Confederation of Indian Industries



# UB Bioinformatics Snapshot (2002-03)

- **7/02: Jeff Skolnick, Director**
  - ❑ Brought 13 addit'l staff with him
  - ❑ Authorized to hire 10 additional research groups
- **4/03: Norma Nowak, co-Dir**
  - ❑ Authorized to hire 10 additional research groups
- **9/03: Daniel Fischer, Dir of Ed**
- **Additional Members TBD**
- **External Funding (\$0)**
  - ❑ Applications submitted
- **Deliverables**
  - ❑ 12 scientific papers
- **5/04: Bruce Holm, Director**

- **Resources (Capaldi, Holm, Penksa, Miller, et al.)**
  - ❑ Building
  - ❑ 6TF ® 10TF Compute Cluster



# Experimental Facilities I

- **Molecular Targeting Laboratory**
  - ❑ Screen 30-50K compounds every 3 months
  - ❑ Apply compound to cell (different genes treated w fluor markers)
  - ❑ Rapidly identify effect on specific gene expression pathways
- **Gene Expression Laboratory**
  - ❑ High-throughput microarray and gene chip
  - ❑ Discover new genes, their functions, and pathways
- **Proteomics and Molecular Kinetics Lab**
  - ❑ Identify molecular targets found in Gene Expression Lab
- **Disease Modeling Laboratory**
  - ❑ In vivo testing (flies, mice, baboons,...)
  - ❑ Gene targeting and genetic mapping facilities

# Experimental Facilities II

- **Bioengineering Support Laboratory**
  - Capabilities in photonics and nano-tech research
  - E.g., handheld devices to test for diseases
- **Protein Scale-Up and Purification**
- **High-Throughput Robotic Combinatorial Chemistry/Parallel Synthetic Chemistry Capabilities**
  - Drugs created robotically; Tested for interaction with target protein
  - Rapid identification of a large number of potential drugs
- **Public Health and Molecular Pathology**
  - Tissue repositories; disease gene maps; medical informatics
- **High-Throughput Search Process for Structural Biology**
  - Tests 1536 “chemical cocktails” to determine effective parameters for crystallization

# Center for Computational Research 1999-2004 Snapshot

## ■ High-Performance Computing and High-End Visualization

- ❑ 110 Research Groups in 27 Depts
- ❑ 13 Local Companies
- ❑ 10 Local Institutions

## ■ External Funding

- ❑ \$111M External Funding
  - \$15.5M as lead
  - \$99.9M in support
- ❑ \$41.8M Vendor Donations
- ❑ Total Leveraged: \$0.5B

## ■ Deliverables

- ❑ 350+ Publications
- ❑ Software, Media, Algorithms, Consulting, Training, CPU Cycles...





# Major CCR Resources (12TF & 290TB)

- **Dell Linux Cluster: #22 ® #25 ® #38**
    - ❑ 600 P4 Processors (2.4 GHz)
    - ❑ 600 GB RAM; 40 TB Disk; Myrinet
  - **Dell Linux Cluster: #187 ® #368 ® off**
    - ❑ 4036 Processors (PIII 1.2 GHz)
    - ❑ 2TB RAM; 160TB Disk; 16TB SAN
  - **IBM BladeCenter Cluster**
    - ❑ 532 P4 Processors (2.8 GHz)
    - ❑ 5TB SAN
  - **SGI Origin3700 (Altix)**
    - ❑ 64 Processors (1.3GHz ITF2)
    - ❑ 256 GB RAM
    - ❑ 2.5 TB Disk
  - **SGI Origin3800**
    - ❑ 64 Processors (400 MHz)
    - ❑ 32 GB RAM; 400 GB Disk
  - **Apex Bioinformatics System**
    - ❑ Sun V880 (3), Sun 6800
    - ❑ Sun 280R (2)
    - ❑ Intel PIIIs
    - ❑ Sun 3960: 7 TB Disk Storage
  - **HP/Compaq SAN**
    - ❑ 75 TB Disk
    - ❑ 190 TB Tape
    - ❑ 64 Alpha Processors (400 MHz)
    - ❑ 32 GB RAM; 400 GB Disk
- 
- **IBM RS/6000 SP: 78 Processors**
  - **Sun Cluster: 80 Processors**
  - **SGI Intel Linux Cluster**
    - ❑ 150 PIII Processors (1 GHz)
    - ❑ Myrinet

# 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

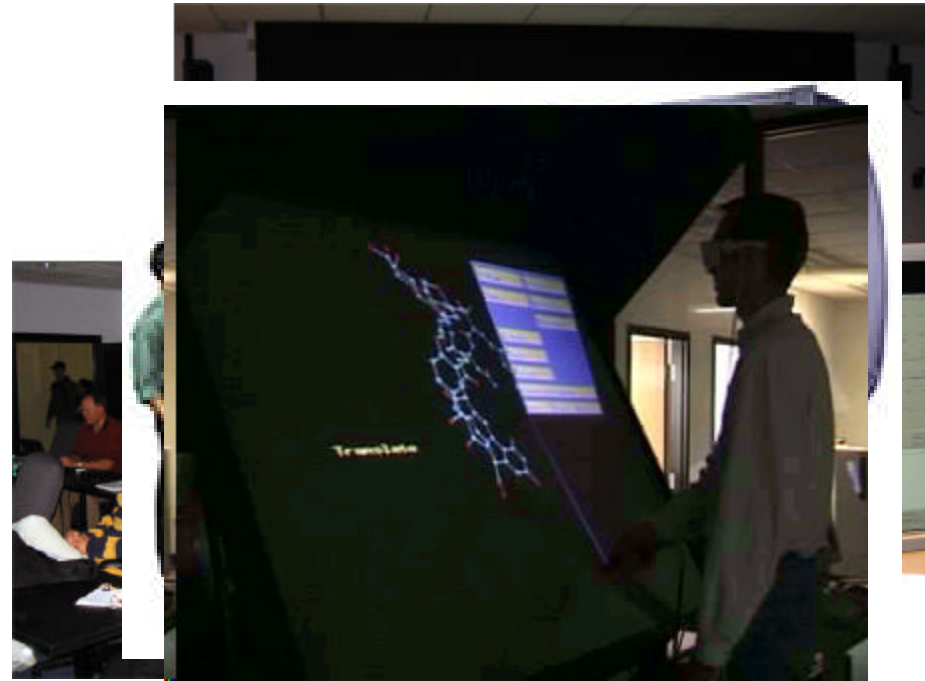
- ❑ 20 NEC projectors: 15.7M pixels
- ❑ Screen is 11' ´ 7'
- ❑ Dell PCs with Myrinet2000

## ■ Access Grid Node

- ❑ Group-to-Group Communication
- ❑ Commodity components

## ■ SGI Reality Center 3300W

- ❑ Dual Barco's on 8' ´ 4' screen



# 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

# Peace Bridge Visualization

## ■ Proposed Options

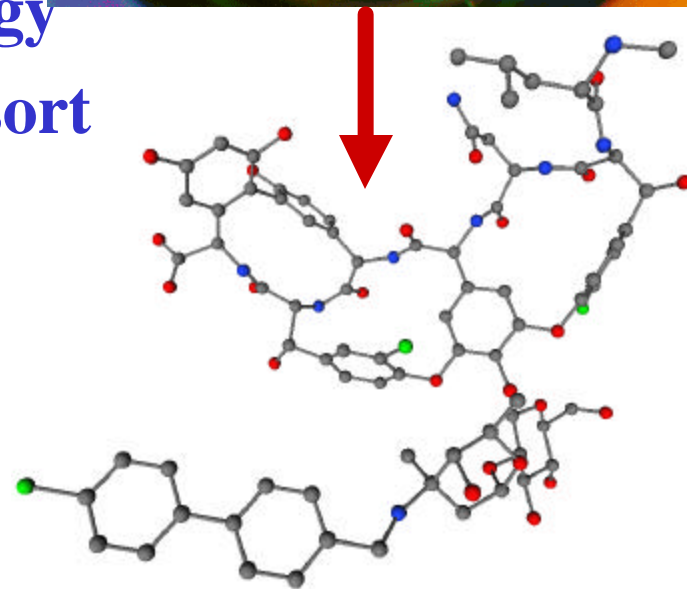
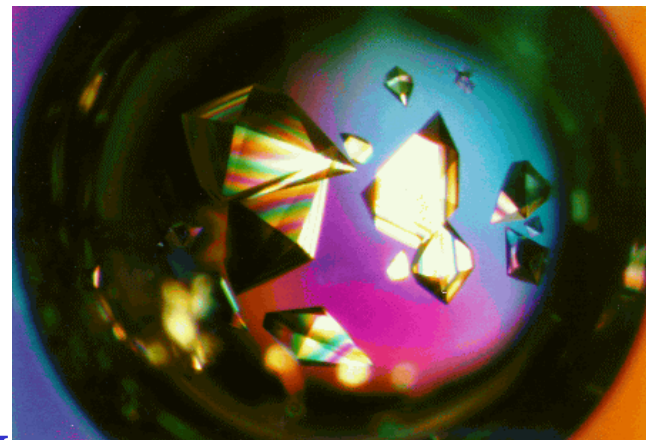
- Relocate US plaza
- Build a 3-lane companion span, rehab existing bridge
- Build a six lane signature span



PHOTO AND STORY BY BRUCE JACKSON

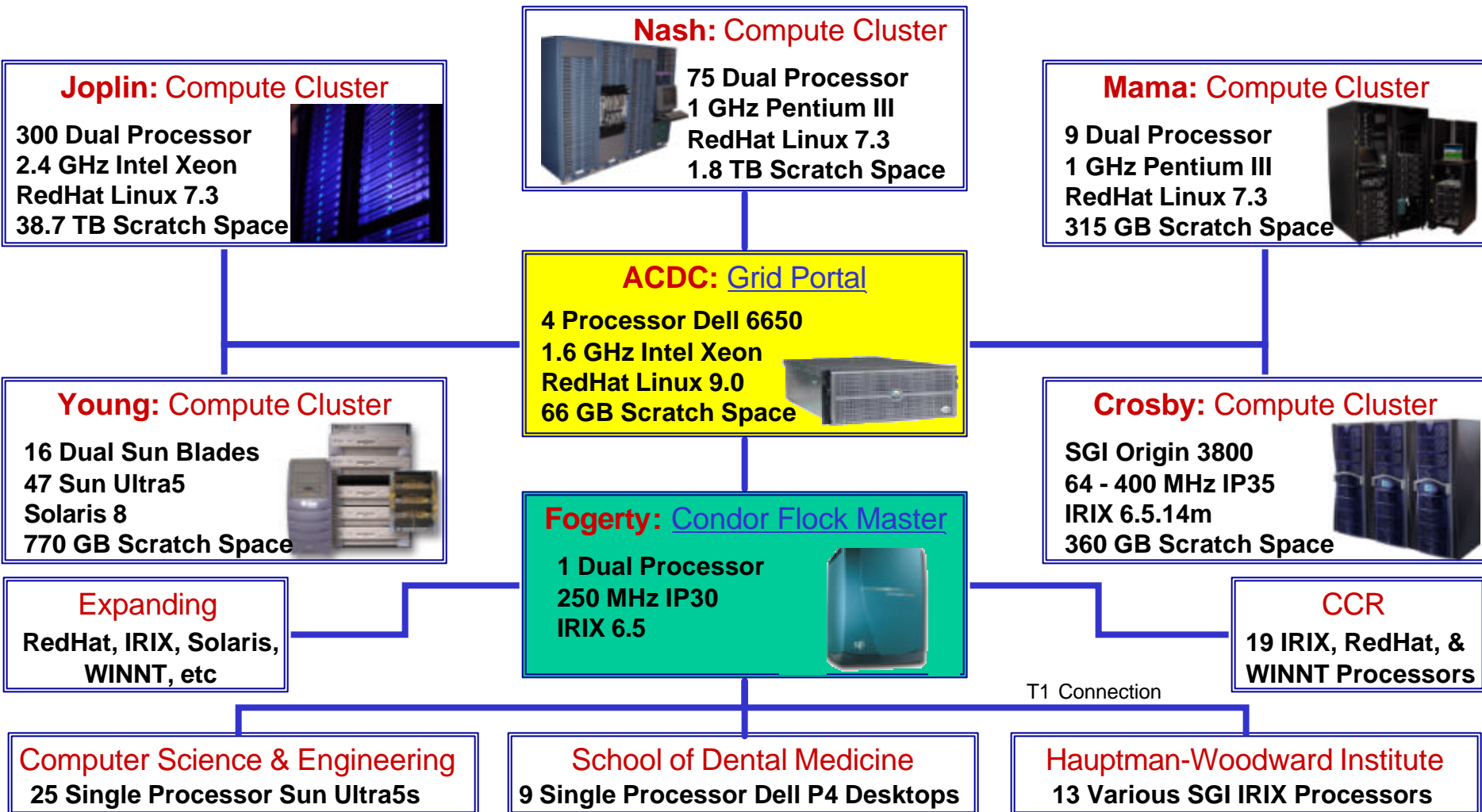
# Molecular Structure Determination via *Shake-and-Bake*

- *SnB* Software by UB/HWI
  - “Top Algorithms of the Century”
- Worldwide Utilization
- Critical to Rational Drug Design
- Important Link in Structural Biology
- Vancomycin: Antibiotic of Last Resort
- Current Effort
  - Grid
  - Collaboratory
  - Intelligent Learning



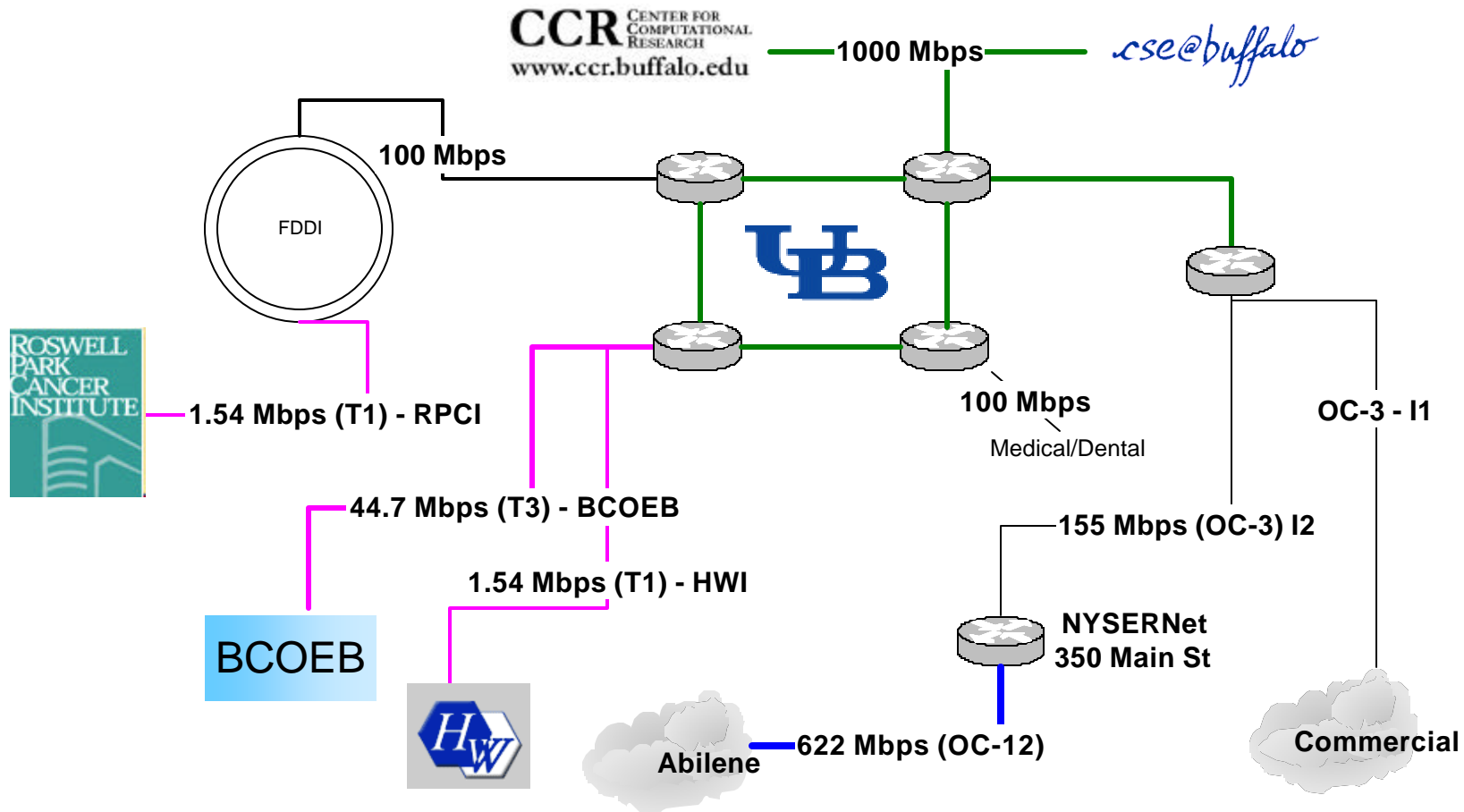
# Advanced Computational Data Center

## ACDC: Grid Overview



Note: Network connections are 100 Mbps unless otherwise noted.

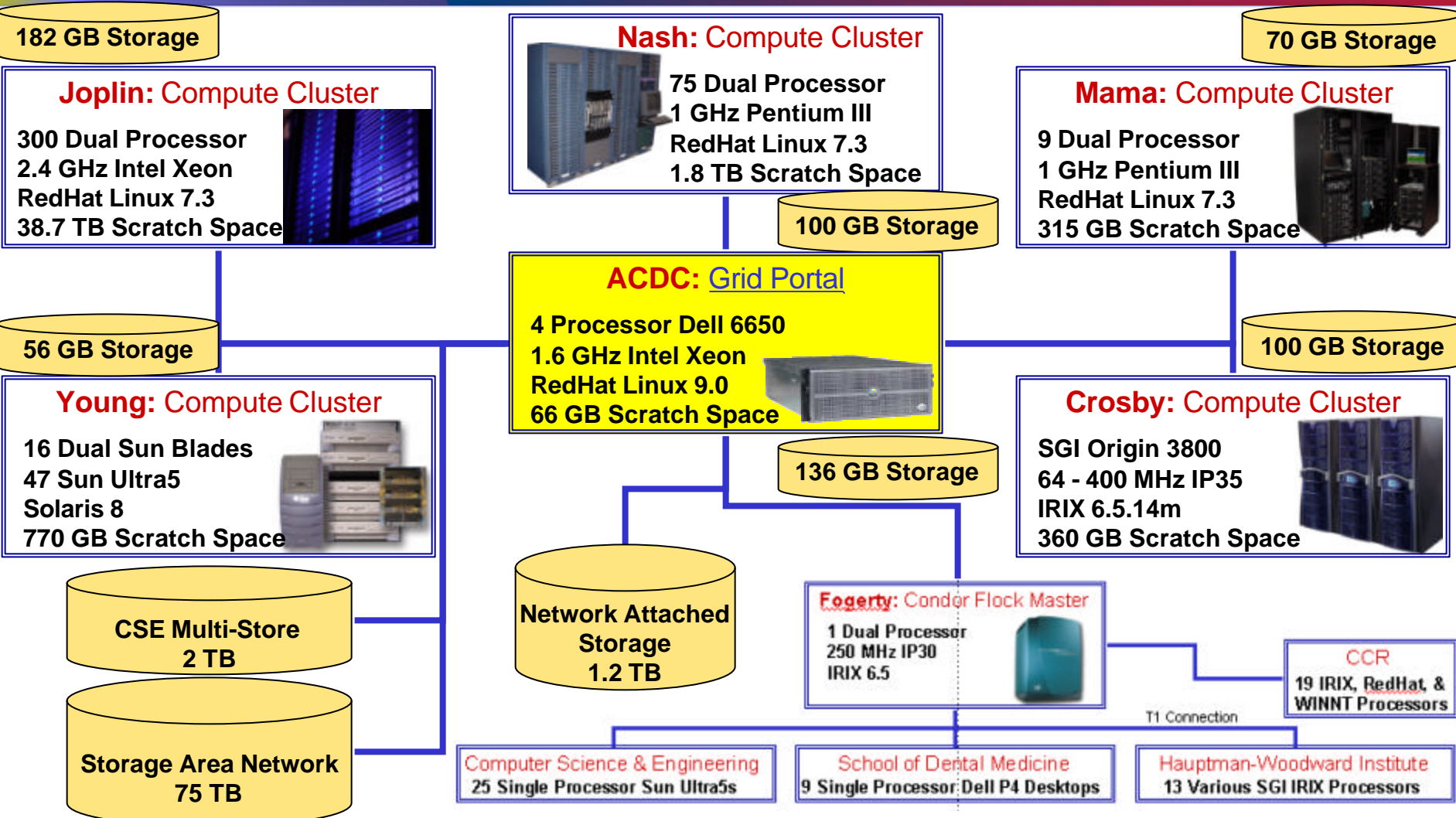
# Network Connections





# ACDC Data Grid Overview

## (Grid-Available Data Repositories)



Note: Network connections are 100 Mbps unless otherwise noted.

# ACDC-Grid

CCR Grid Computing Services - Microsoft Internet Explorer

Center for Computational Research GRID PORTAL  
High Performance Grid Computing

**PORTAL LOGIN**

- Grid General Info
- A Breadcrumb Trail
- Computational Grid
- Data Grid
- MakeGrid
- Technical Pages
- Presentations
- Contact Us
- Grid Account Request
- Grid Account Support
- Events
- News
- Projects
- Resources
- Education/Outreach
- Staff Only
- CCR HOME

**Welcome to Grid Computing Services**

University at Buffalo Center for Computational Research is currently forming the first Western New York computational grid. The computational grid consist of many supercomputers located at the Center and several other networked supercomputers throughout the Western New York region. These resources will be shared by many researchers from several departments working on a diverse suite of problems including Biomaterials, Computational Chemistry, and Medical Imaging to name a few.



We also provide grid computing support for the University's Center for Computational Research learning, teaching and research activities plus the infrastructure for both high performance computing and grid enabled software.

Get your "Grid Computing Guide"?

CCR Grid Computing Services Data Management - Microsoft Internet Explorer

Center for Computational Research GRID PORTAL  
High Performance Grid Computing

**PORTAL LOGOUT**

User Tools

- Manage Account
- Grid General Info
- Projects
- Resources
- Computational Grid
- Job Submissions
- Job/Queue Status
- Data Grid
- Network Status
- Running/Queued Jobs
- PDS Job History
- Grid Portal Statistics
- Leader/Flock Statistics
- User Info
- Education/Outreach
- Staff Only
- CCR HOME

VIEW: Group: **miller** UserList: rpplye

- repplye
  - KeyMaster
  - Morpheus
    - Tank
    - Agent
    - Rabbit
    - Tank
    - Morpheus
    - Oracle.m
    - Neo

Browser view of "miller" group files published by user

CCR Grid Computing Services Grid Admin - Microsoft Internet Explorer

Center for Computational Research GRID PORTAL  
High Performance Grid Computing

**PORTAL LOGOUT**

User Tools

- Manage Account
- Grid General Info
- Projects
- Resources
- Computational Grid
- Job Submissions
- Job/Queue Status
- Data Grid
- Network Status
- Running/Queued Jobs
- PDS Job History
- Grid Portal Statistics
- Leader/Flock Statistics
- User Info
- Education/Outreach
- Staff Only
- CCR HOME

View statistics for: **disk\_space**

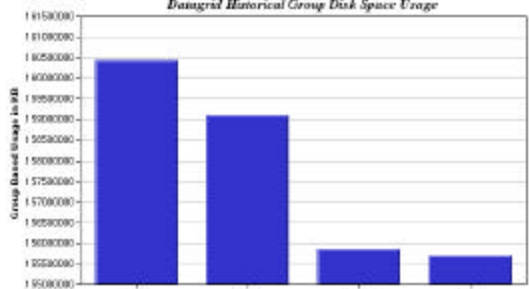
Data based on: **group**

from starting date: **January 1 2000**

to ending date: **September 13 2003** inclusive

for: **Grid Portal** resources: **OK**

**Daagrid Historical Group Disk Space Usage**



Group	Disk Space Usage (KB)
miller	~1,050,000
griddev	~1,000,000
ccrstaff	~1,000,000
mlgreen	~1,000,000

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

Center for Computational Research GRID PORTAL  
High Performance Grid Computing

**PORTAL LOGOUT**

User Tools

- Manage Account
- Grid General Info
- Projects
- Resources
- Computational Grid
- Job Submissions
- Job/Queue Status
- Data Grid
- Network Status
- Running/Queued Jobs
- PDS Job History
- Grid Portal Statistics
- Leader/Flock Statistics
- User Info
- Education/Outreach
- Staff Only
- CCR HOME

View statistics for: **disk\_space**

Data based on: **user**

from starting date: **January 1 2000**

to ending date: **September 13 2003** inclusive

for: **Grid Portal** resources: **OK**

File_num	File_ID	Filename	Dir_ID	Resource_ID	Owner	Groupname	Type
1	56033	Cypher.txt	52831	10	mlgreen	griddev	txt
2	56034	Cypher.sh	52858	10	mlgreen	griddev	sh
3	56035	Oracle.asc	52958	10	mlgreen	griddev	asc
4	56036	Cypher.sh	52634	10	mlgreen	miller	sh
5	56037	Rabbit.dat	52830	10	mlgreen	ccrstaff	dat
6	56038	Agent.exe	53064	10	mlgreen	griddev	exe
7	56039	Dozer.sh	52852	10	mlgreen	griddev	sh
8	56040	Neo.asc	52187	10	mlgreen	mlgreen	asc
9	56041	Agent.mpg	52833	10	mlgreen	mlgreen	mpg
10	56042	Tank.txt	52188	10	mlgreen	mlgreen	txt
11	56043	Smith.xls	52258	10	mlgreen	ccrstaff	xls
12	56044	KeyMaster.csh	52186	10	mlgreen	miller	csh
13	56045	Oracle.csh	52622	10	mlgreen	griddev	csh
14	56046	Dozer.xls	52808	10	mlgreen	mlgreen	xls
15	56047	Cypher.exe	52204	10	mlgreen	griddev	exe
16	56048	Rabbit.ppt	52861	10	mlgreen	miller	ppt
17	56049	Neo.dat	52217	10	mlgreen	ccrstaff	dat
18	56050	Cypher.asc	53086	10	mlgreen	griddev	asc

# ACDC-Grid Administration

**CCR Grid Computing Services: Grid Admin - Microsoft: Internet Explorer**

Center for Computational Research **GRID PORTAL**  
High Performance Grid Computing

**Grid Site Administration**

**PORTAL LOGOUT**  
User Tools  
Manage Account  
Grid General Info  
Projects  
Resources  
Computational Grid  
Job Submission  
Job/Queue Status  
Data Grid  
Data Grid Statistics  
Network Status  
Running/Queued Jobs  
PBS Job History  
Grid Portal Statistics  
Center Fleck Statistics  
User Information  
Education/Outreach  
Staff Only  
CCR HOME

**Users**  
Groups  
Portal Event Log  
Database Job List

Organizations (add, edit, delete)  
Resources (view, refresh, ping, delete, create host certificate)

**Globus Administration**  
Reports (machine usage, user access to machines, etc.)

**Generate Globus grid-mapfile**

Specifying an optional include file will cause the contents of this file to be included at the top of the generated grid-mapfile. If a grid-mapfile path is specified a copy of the generated file will be saved into this location. The generated file will be staged to the grid nodes unless the box is checked.

Optional include file:

Optional grid-mapfile path:

Do not stage this file to the grid nodes

**CCR Grid Computing Services: Database Job Admin - Microsoft: Internet Explorer**

Center for Computational Research **GRID PORTAL**  
High Performance Grid Computing

**Create New Database Job**

Create a new database job that can be run by the portal. Job scripts must reside in `home/griddev/www/jobscripts` prior to creating the database job entry.

Job Name:

Full Path To Script:

Accepts Arguments:

Run Script:

Run As User:

[Return to the Database Job Admin menu.](#)  
[Return to the Grid Admin menu.](#)

**CCR Grid Computing Services: Grid Admin - Resources - Microsoft: Internet Explorer**

Center for Computational Research **GRID PORTAL**  
High Performance Grid Computing

**MDS Resource Update Status**

Current Time: 16-September-2003 10:59:12

Resource	Last Updated	Next Update	Status
crasby.ccr.buffalo.edu	16-September-2003 09:15:30	2 minutes	OK
rogerty.ccr.buffalo.edu	16-September-2003 10:45:30	2 minutes	OK
joplin.ccr.buffalo.edu	16-September-2003 10:45:15	2 minutes	OK
mama.ccr.buffalo.edu	16-September-2003 10:45:15	2 minutes	OK
nash.ccr.buffalo.edu	16-September-2003 10:45:15	2 minutes	OK
newus.hwi.buffalo.edu	16-September-2003 10:45:20	2 minutes	OK
yardbirds.ccr.buffalo.edu	16-September-2003 10:45:13	2 minutes	OK
young.ccr.buffalo.edu	16-September-2003 10:45:27	2 minutes	OK

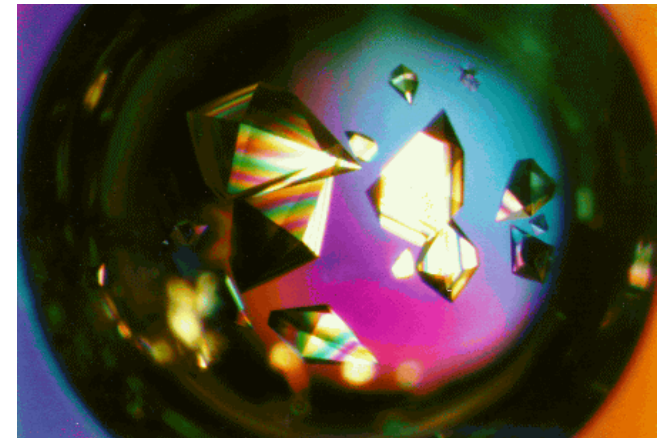
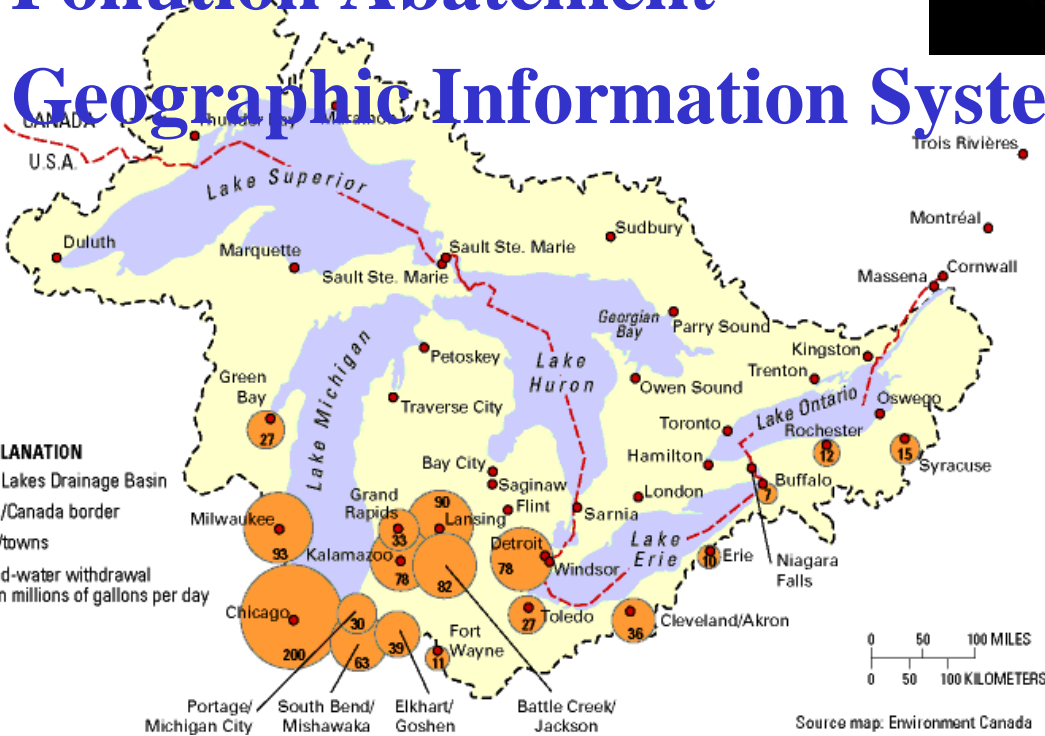
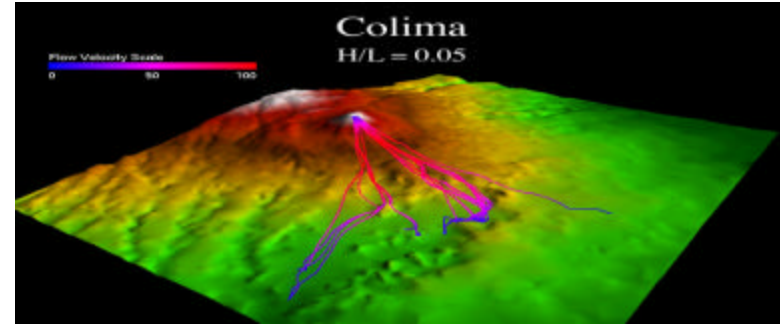
[Return to the Grid Resource Admin menu.](#)  
[Return to the Grid Admin menu.](#)

Advanced  
Center for Computational Research  
Data  
Center



# Grid-Enabling Application Templates

- Structural Biology
- Earthquake Engineering
- Pollution Abatement
- Geographic Information Systems & BioHazards



# ACDC-Grid Cyber-Infrastructure

## ■ Predictive Scheduler

- Define quality of service estimates of job completion, by better estimating job runtimes by profiling users.

## ■ Data Grid

- Automated Data File Migration based on profiling users.

## ■ High-performance Grid-enabled Data Repositories

- Develop automated procedures for dynamic data repository creation and deletion.

## ■ Dynamic Resource Allocation

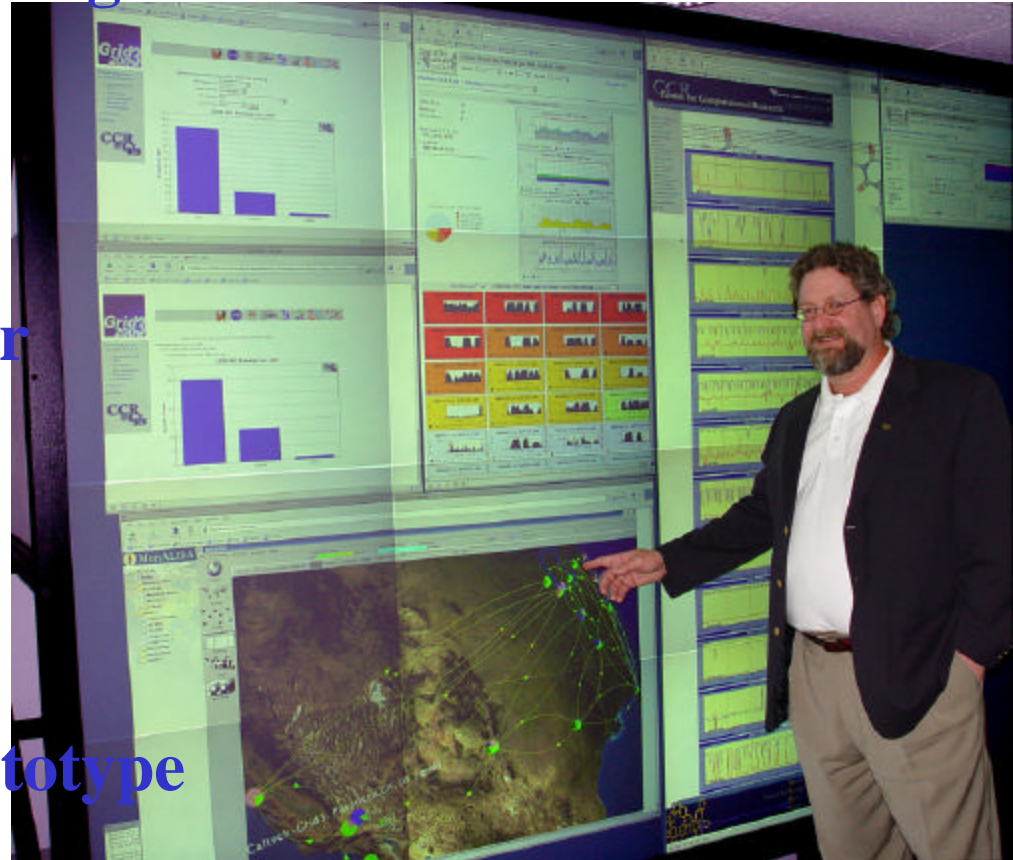
- Develop automated procedures for dynamic computational resource allocation.

# Middleware

- **Globus Toolkit 2.2.4 ® direct upgrade WSRF**
- **Condor 6.6.0**
- **Network Weather Service 2.6**
- **Apache2 HTTP Server**
- **PHP 4.3.0**
- **MySQL 3.23**
- **phpMyAdmin 2.5.1**

# ACDC-Grid Collaborations

- High-Performance Networking Infrastructure
- WNY Grid Initiative
- Grid3+ Collaboration
- iVDGL Member
- Open Science Grid Member
- Grid-Based Visualization
  - SGI Collaboration
- Grid-Lite
  - HP Labs Collaboration
- Innovative Laboratory Prototype
  - Dell Collaboration





[www.ccr.buffalo.edu](http://www.ccr.buffalo.edu)