

Enabling a Grid for Molecular Crystal Structure Determination via *Shake-and-Bake*

Russ Miller

Director, Center for Computational Research

UB Distinguished Professor, Computer Science & Engineering

Senior Research Scientist, Hauptman-Woodward Medical Inst



**“Top 10 Worldwide
Supercomputing
Center”**

- www.gapcon.com



University at Buffalo

The State University of New York



Outline

- **Bioinformatics in Buffalo**
- **Supercomputing & Visualization**
- **Grid Computing Overview**
- **Grid Computing in Buffalo**
 - *Shake-and-Bake*: Computational Crystallography
 - ECCE: Computational Chemistry

Biomedical Advances

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

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

■ **Artificial Blood**

■ **Nicorette Gum**

■ **Fetal Viability Test**

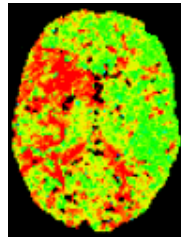
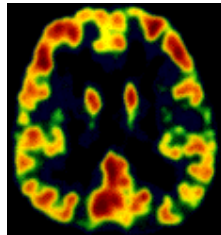
■ **Implantable Pacemaker**

■ **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 20th Century”**

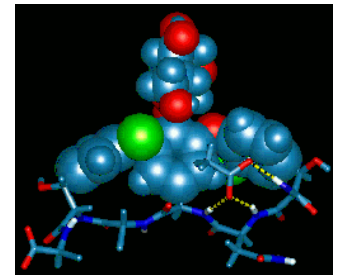
□ **Vancomycin**

□ **Gramacidin A**

■ **High Throughput Crystallization Method: Patented**

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

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



Bioinformatics in Buffalo

A \$290M Initiative

- **UB Center for Advanced Bioengineering & Biomedical Technologies**
 - NYS: \$1M/yr
 - Med Tech for Product Dev & Commer.
- **Center Disease Modeling & Therapy Discovery**
 - UB, HWI, RPCI, Kaleida
 - NYS: \$15.3M
 - Software, device development, and drug therapies
- **Buffalo Center of Excellence in Bioinformatics**
 - UB, HWI, RPCI
 - NYS: \$61M
 - Federal Government: \$10M
 - Corporate Funding: \$151
- **UB Faculty G&C Funding: \$64M**



Partnerships

■ **Lead Partners: SUNY-Buffalo, Hauptman-Woodward Medical Research Institute, Roswell Park Cancer Institute**



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

UB Bioinformatics Snapshot (2002-03)

■ Personnel

- ❑ Hired Jeff Skolnick as Director (7/02)
 - Brought 13 additional staff to Buffalo
 - Authorized to hire 10 additional research groups
- ❑ Hired Norma Nowak as co-Director (4/03)
 - Authorized to hire 10 additional research groups
- ❑ Hired Daniel Fischer as Dir. of Education (9/03)
- ❑ Additional members TBD

■ External Funding (\$0)

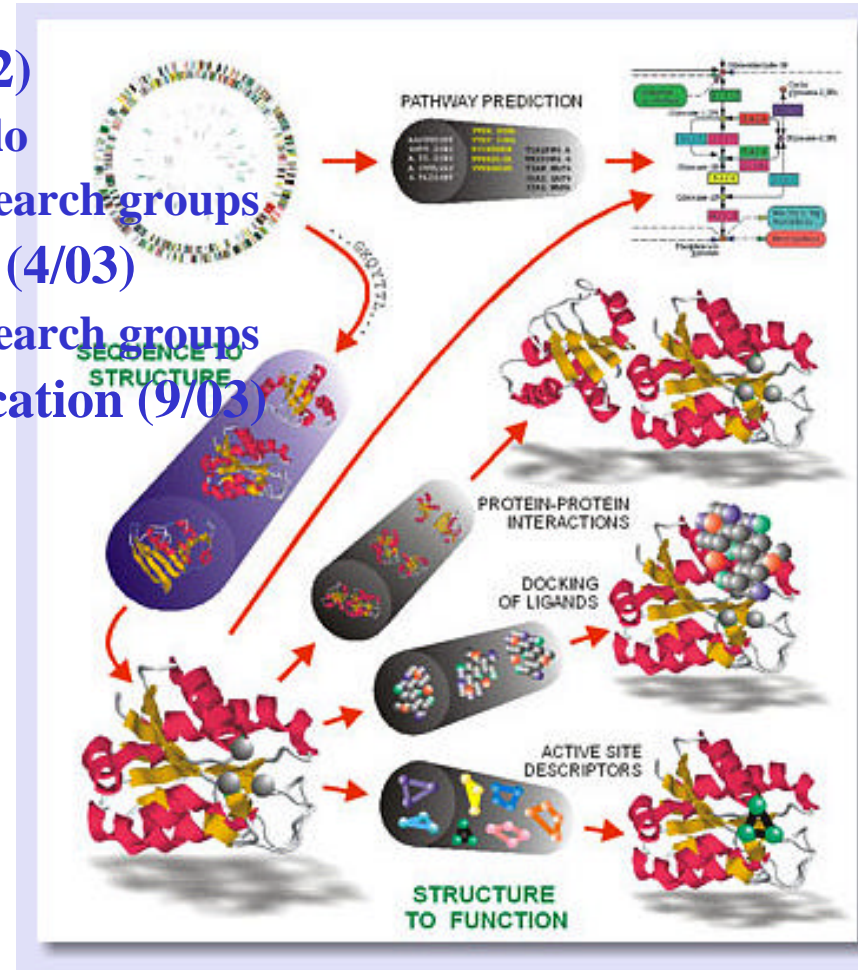
- ❑ Applications submitted

■ Deliverables

- ❑ Six (6) scientific papers

■ Resources

- ❑ Building
- ❑ 6TF ® 10TF Compute Cluster



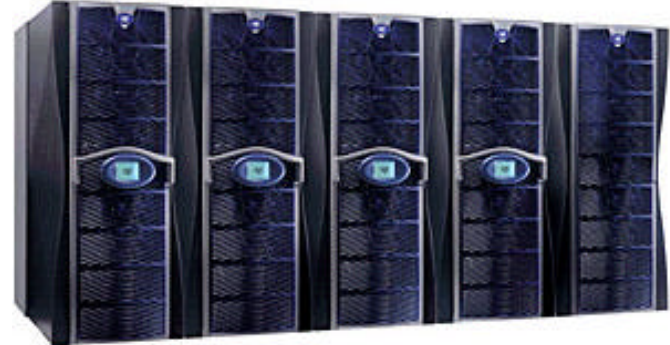
Related Academic Programs

- **Bachelor's & Master's Program in Bioinformatics**
- **Related Disciplines**
 - ❑ **Chemical Biology (Sloan Support)**
 - ❑ **Computational Chemistry (Sloan Support)**
 - ❑ **Environmental Analysis (Sloan Support)**
 - ❑ **Medical Informatics (Graduate Certificate)**
 - ❑ **Pharmacometrics**
 - ❑ **UB-HWI Department of Structural Biology (Ph.D.)**
- **Advanced Degrees under Development**
 - ❑ **Biophotonics**
- **Complementary Degrees in WNY**
 - ❑ **Canisius College & RIT**
 - ❑ **Niagara University NYS \$5M Center of Excellence in Bioinformatics (degrees in development)**

Center for Computational Research (World's 8th Largest Supercomputing Center)

■ High-Performance Computing and High-End Visualization

- ❑ 110 Research Groups in 27 Depts
- ❑ 25 Companies and Institutions

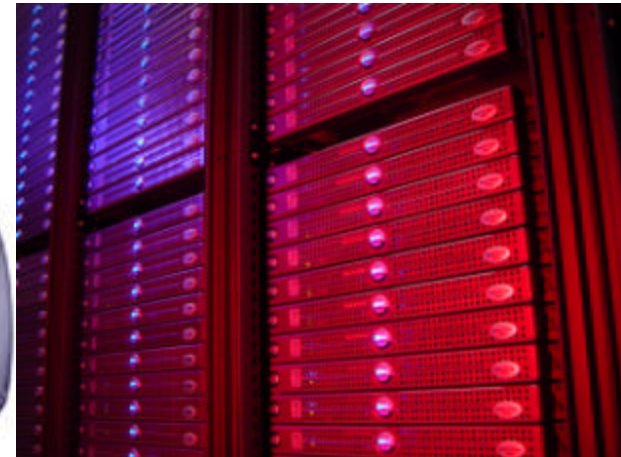


■ Sample Areas

- ❑ Urban Visualization and Simulation
- ❑ Computational Chemistry
- ❑ Ground Water Modeling
- ❑ Geophysical Mass Flows
- ❑ Networked Multimedia
- ❑ Medical Imaging

■ Training

- ❑ Workshops; Courses
- ❑ Degree Programs



CCR 1999-2003 Snapshot

■ Personnel

- ❑ 18 State-Supported Staff (15 Technical / 3 Clerical)
- ❑ 2 Grant-Supported Staff
- ❑ Undergraduates (REUs, Workstudy)

■ External Funding

- ❑ \$111M External Funding
 - \$13.5M as lead
 - \$97.5M in support
- ❑ \$41.8M Vendor Donations

■ Deliverables

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



Computational Resources (10TF)

- **Dell Linux Cluster - #25 on top500**
 - ❑ 600 P4 Processors (2.4 GHz)
 - ❑ 600 GB RAM; 40 TB Disk; Myrinet
- **Dell Linux Cluster - #368 on top500**
 - ❑ 4036 Processors (PIII 1.2 GHz)
 - ❑ 2TB RAM; 160TB Disk; 16TB SN



- **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 (4Q03)**
 - ❑ 75 TB Disk; 200 TB Tape

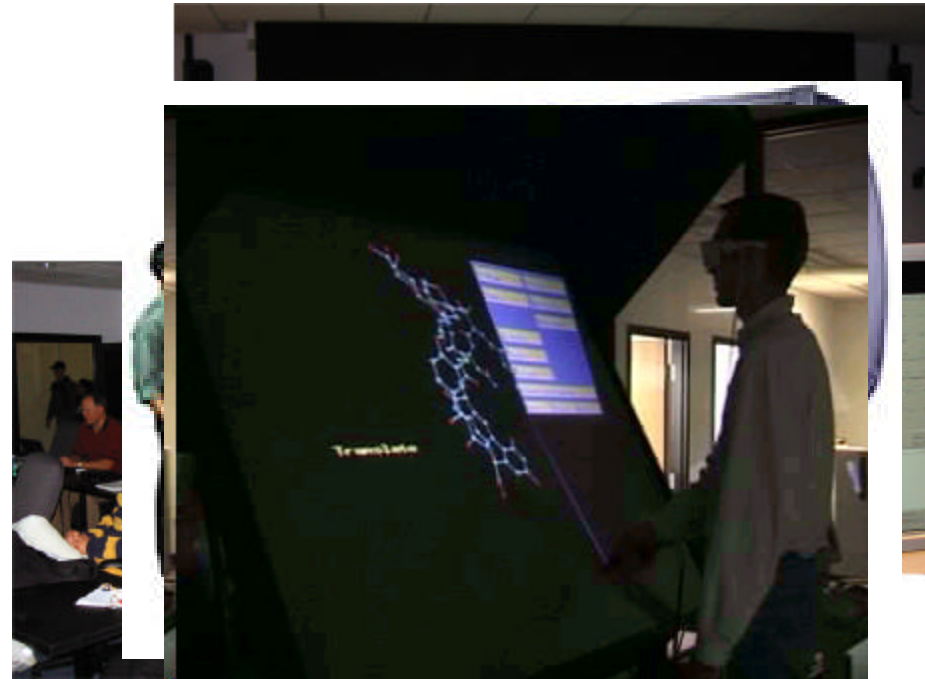


Sample Computational Research

- **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

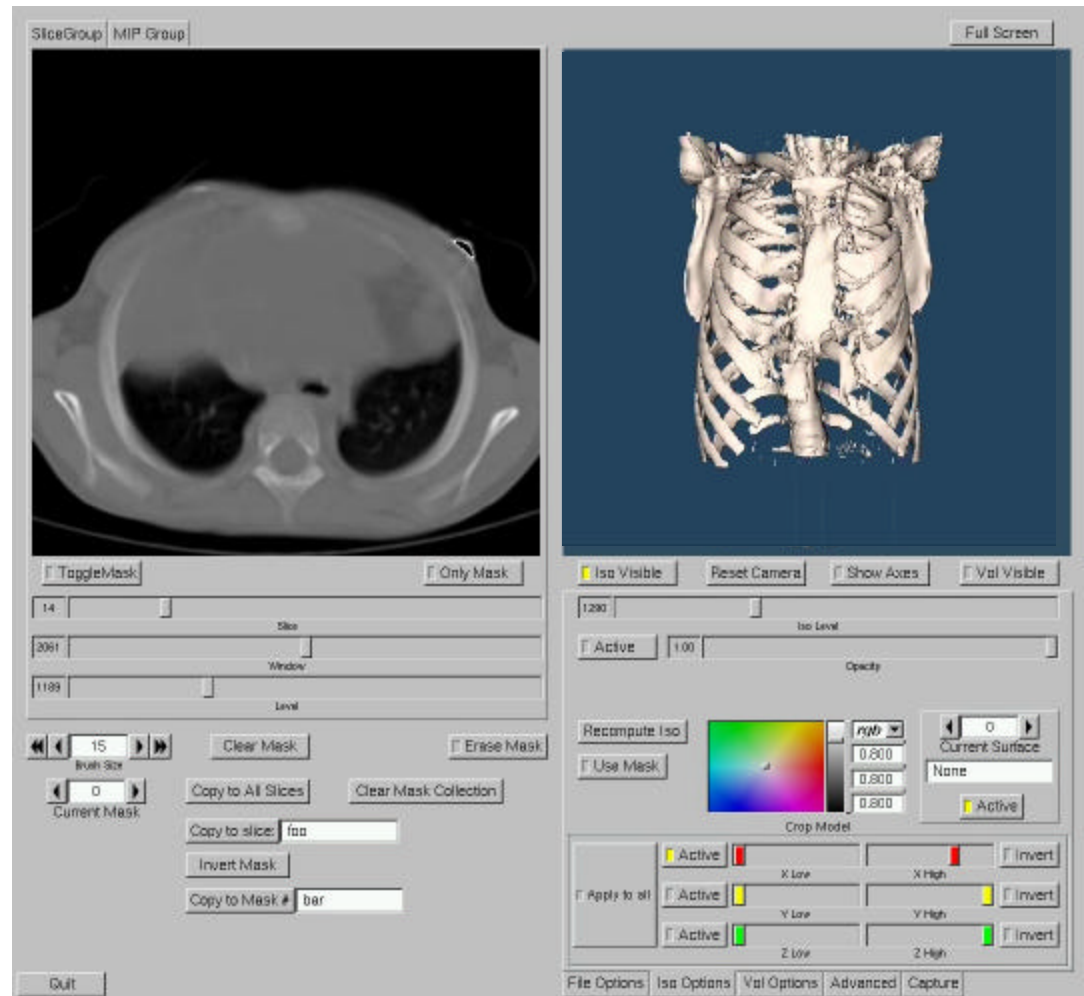


Sample Visualization Areas

- **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

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



StreetScenes[®] Demo

- *StreetScenes*[®] is a Virtual Reality (VR) software solution for 3D visualization of surface traffic
- 3D model of proposed soccer stadium in Rochester
- Used *StreetScenes*[®] to import output file from Synchro traffic simulation

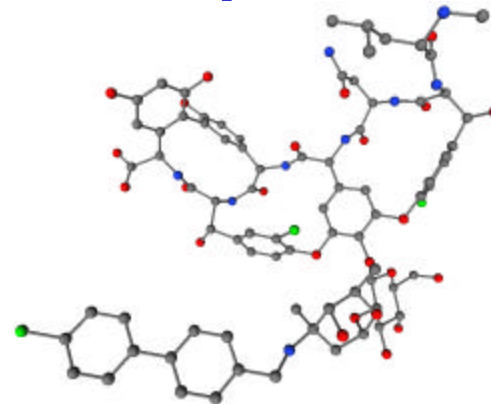
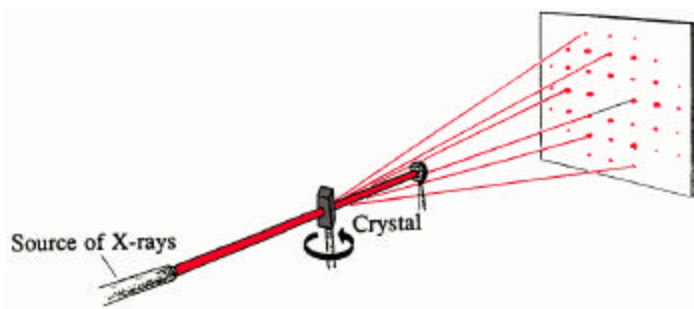
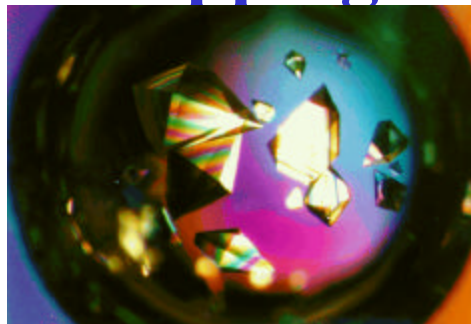


X-Ray Crystallography

- **Objective: Provide a 3-D mapping of the atoms in a crystal.**

- **Procedure:**

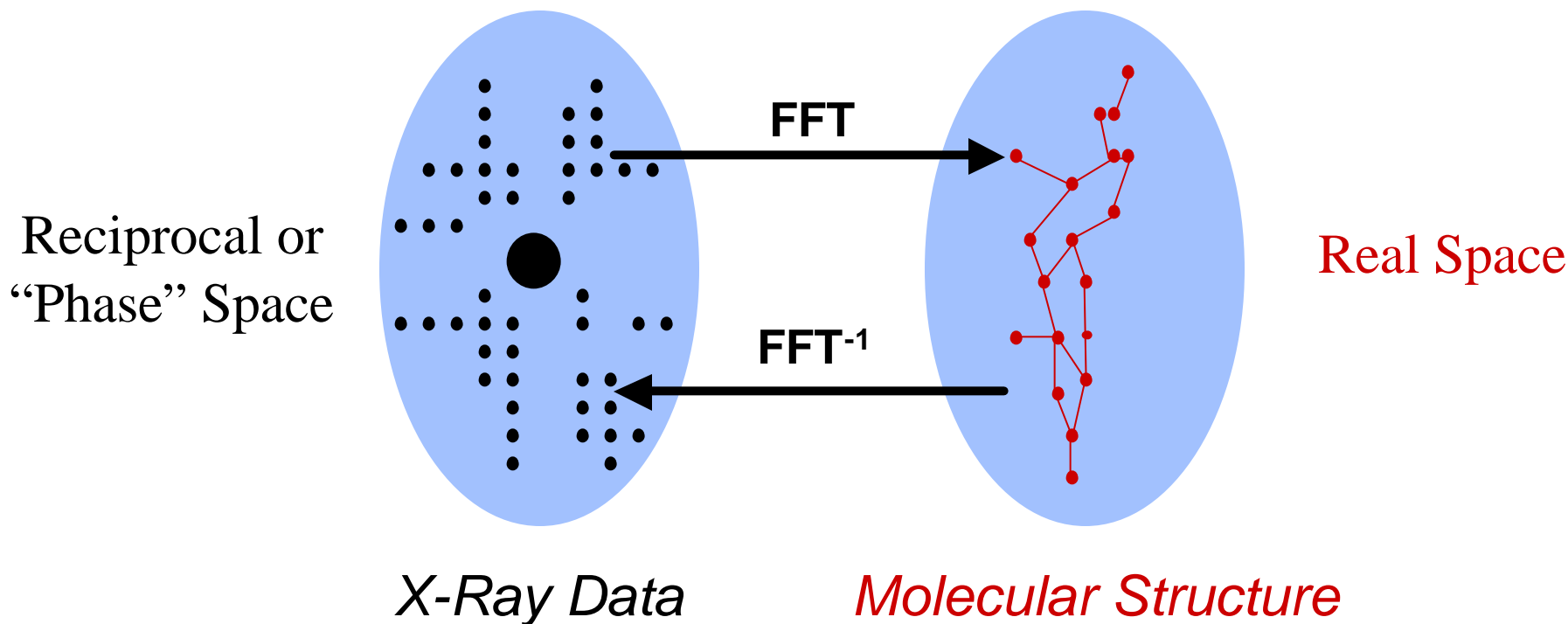
1. **Isolate a single crystal.**
2. **Perform the X-Ray diffraction experiment.**



3. **Determine molecular structure that agrees with diffraction data.**

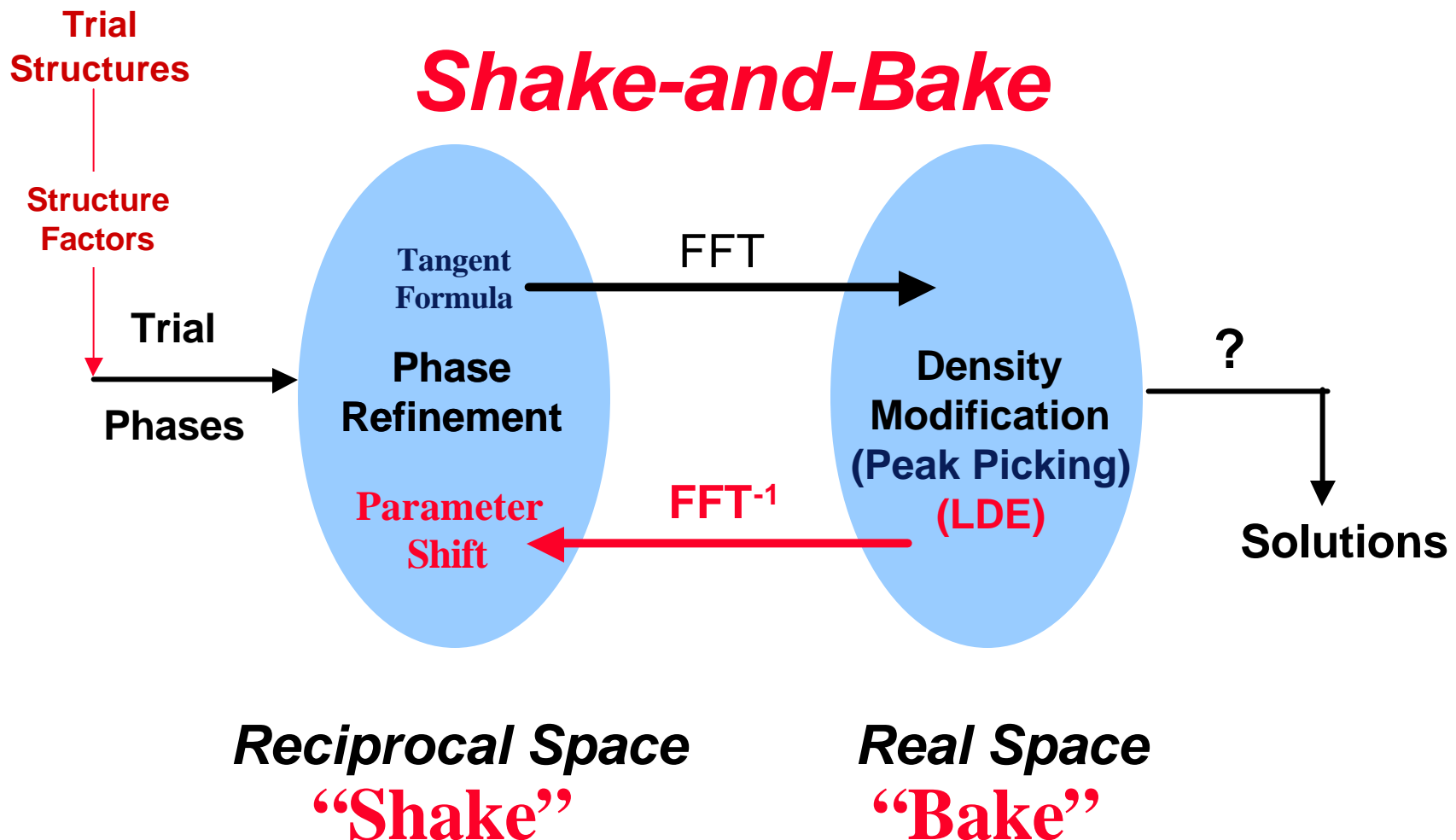
X-Ray Data & Corresponding Molecular Structure

Underlying atomic arrangement is related to the reflections by a 3-D Fourier transform.



- Phases lost during the crystallographic experiment.
- *Phase Problem*: Determine phases of the reflections.

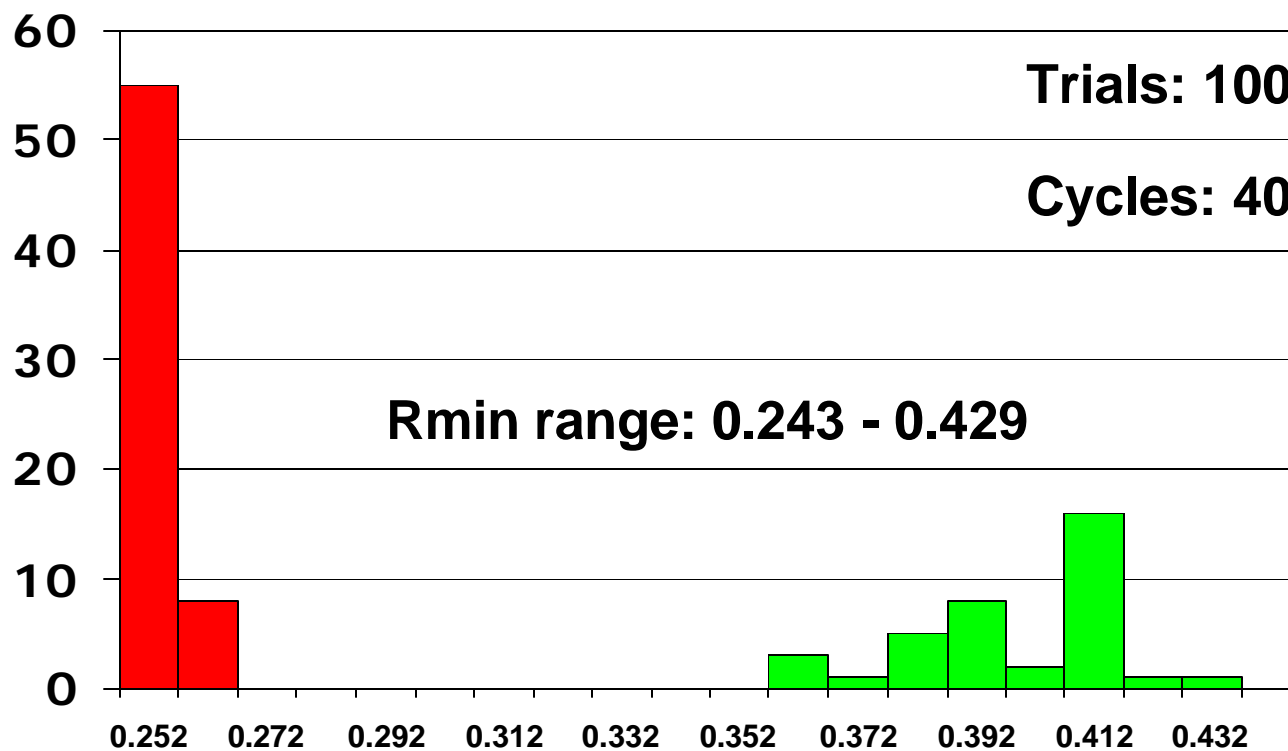
Shake-and-Bake Method: Dual-Space Refinement



Ph8755: *SnB* Histogram

Atoms: 74
Space Group: P1

Phases: 740
Triples: 7,400



Phasing and Structure Size

Se-Met with *Shake-and-Bake*?

Se-Met

190kDa

Multiple Isomorphous Replacement?

Shake-and-Bake

Conventional Direct Methods

Vancomycin



Number of Atoms in Structure

Grid Computing 2003



DISCOM

SinRG

APGrid

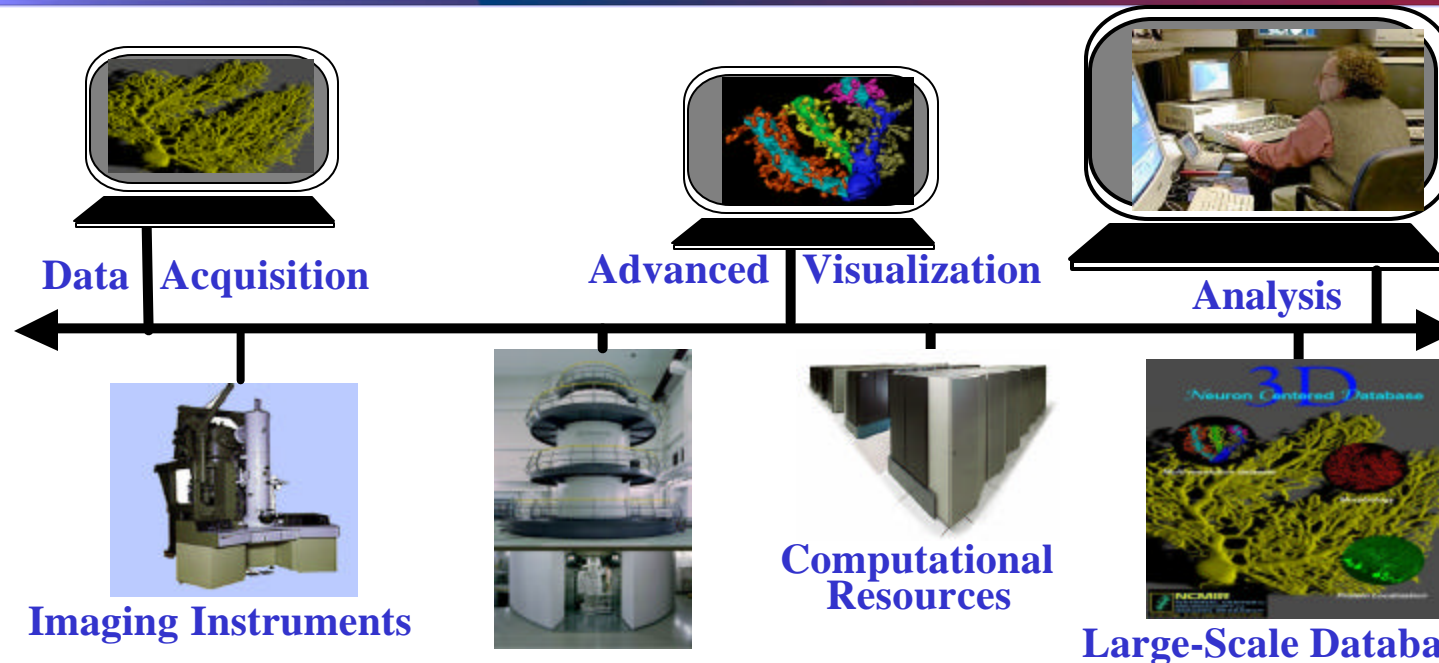
IPG ...



Asia-Pacific Advanced Network



Grid Computing Overview



Thanks to
Mark Ellisman

- Coordinate Computing Resources, People, Instruments in Dynamic Geographically-Distributed Multi-Institutional Environment
- Treat Computing Resources like Commodities
 - ❑ Compute cycles, data storage, instruments
 - ❑ Human communication environments
- No Central Control; No Trust

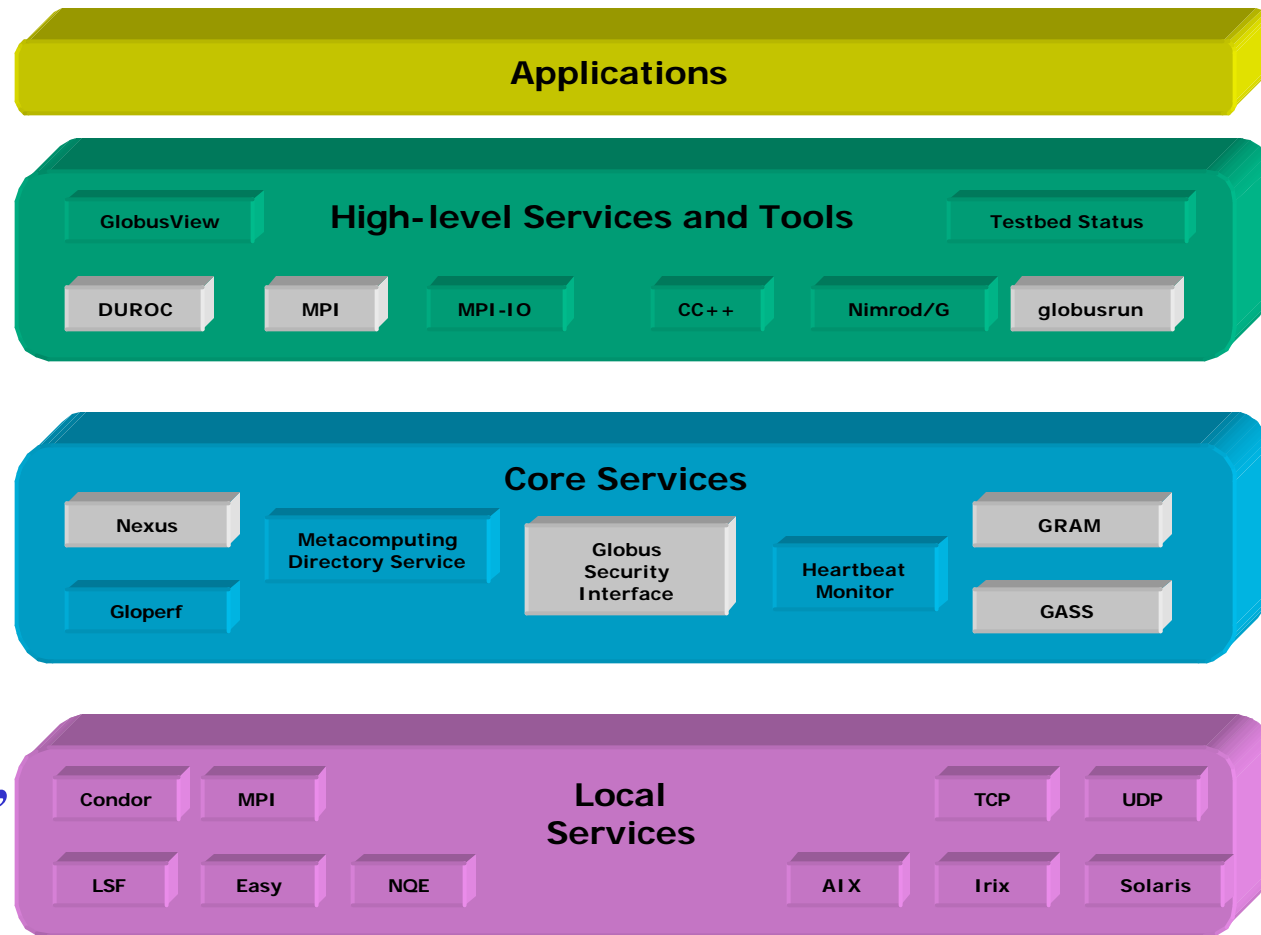
The Globus Project

(Ian Foster and Carl Kesselman)

■ Globus model focuses on providing key Grid services

- ❑ Resource access and management
- ❑ Grid FTP
- ❑ Information Service
- ❑ Security services
 - Authentication
 - Authorization
 - Policy
 - Delegation
- ❑ Network reservation, monitoring, control

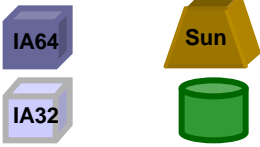
The Grid as a Layered Set of Services



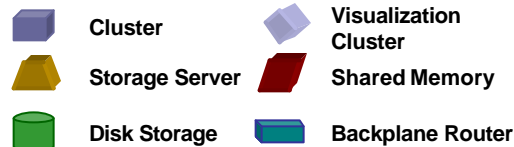
NSF Extensible TeraGrid Facility

Caltech: Data collection analysis

0.4 TF IA-64
IA32 Datawulf
80 TB Storage

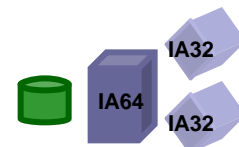


LEGEND

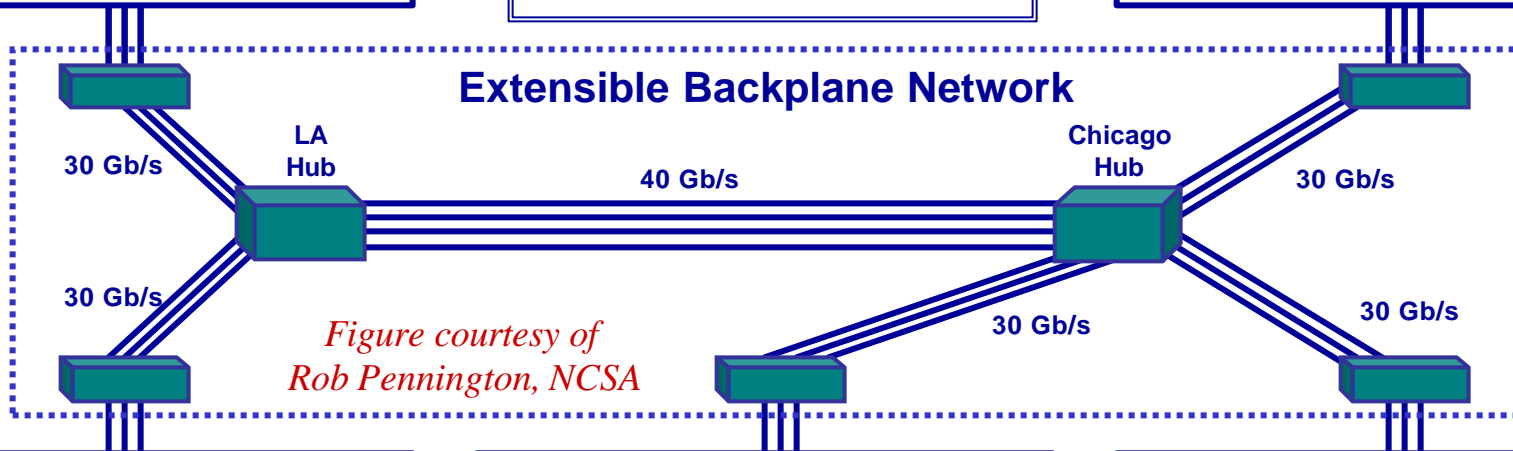


ANL: Visualization

1.25 TF IA-64
96 Viz nodes
20 TB Storage

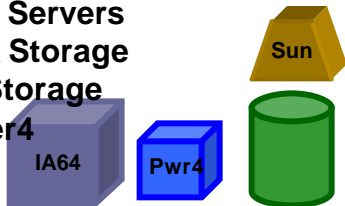


Extensible Backplane Network



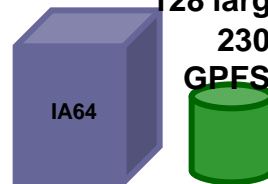
*Figure courtesy of
Rob Pennington, NCSA*

4 TF IA-64
DB2, Oracle Servers
500 TB Disk Storage
6 PB Tape Storage
1.1 TF Power4



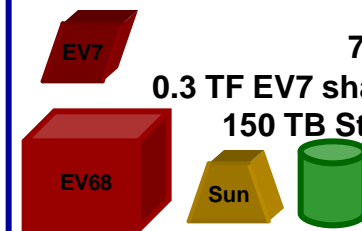
SDSC: Data Intensive

10 TF IA-64
128 large memory nodes
230 TB Disk Storage
GPFS and data mining



NCSA: Compute Intensive

6 TF EV68
71 TB Storage
0.3 TF EV7 shared-memory
150 TB Storage Server

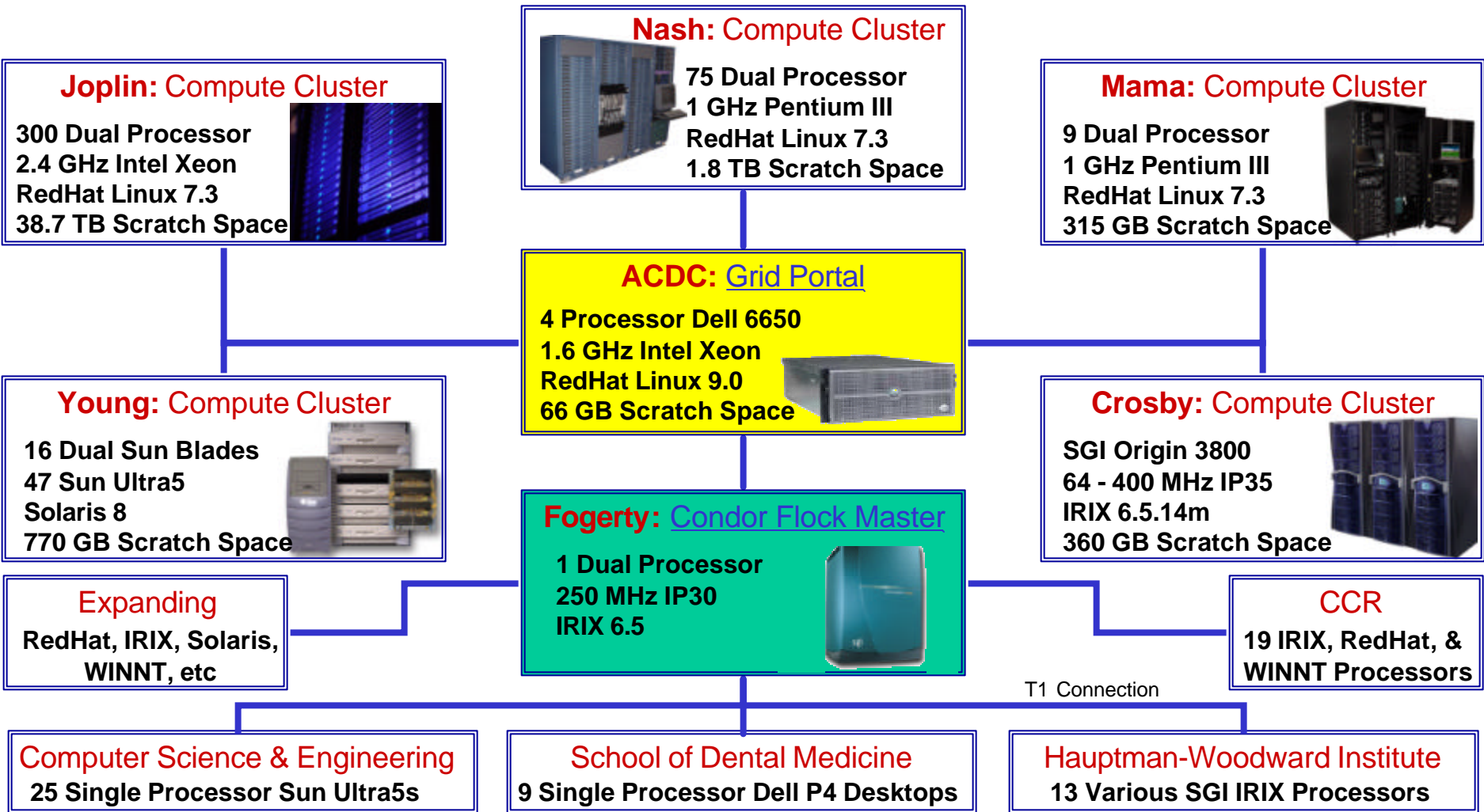


PSC: Compute Intensive

Factors Enabling the Grid

- **Internet is Infrastructure**
 - ❑ Increased network bandwidth and advanced services
- **Advances in Storage Capacity**
 - ❑ Terabyte costs less than \$5,000
- **Internet-Aware Instruments**
- **Increased Availability of Compute Resources**
 - ❑ Clusters, supercomputers, storage, visualization devices
- **Advances in Application Concepts**
 - ❑ Computational science: simulation and modeling
 - ❑ Collaborative environments ® large and varied teams
- **Grids Today**
 - ❑ Moving towards production; Focus on middleware

Advanced CCR Data Center (ACDC) Computational Grid Overview



Note: Network connections are 100 Mbps unless otherwise noted.

Data Grid Motivation & Goal

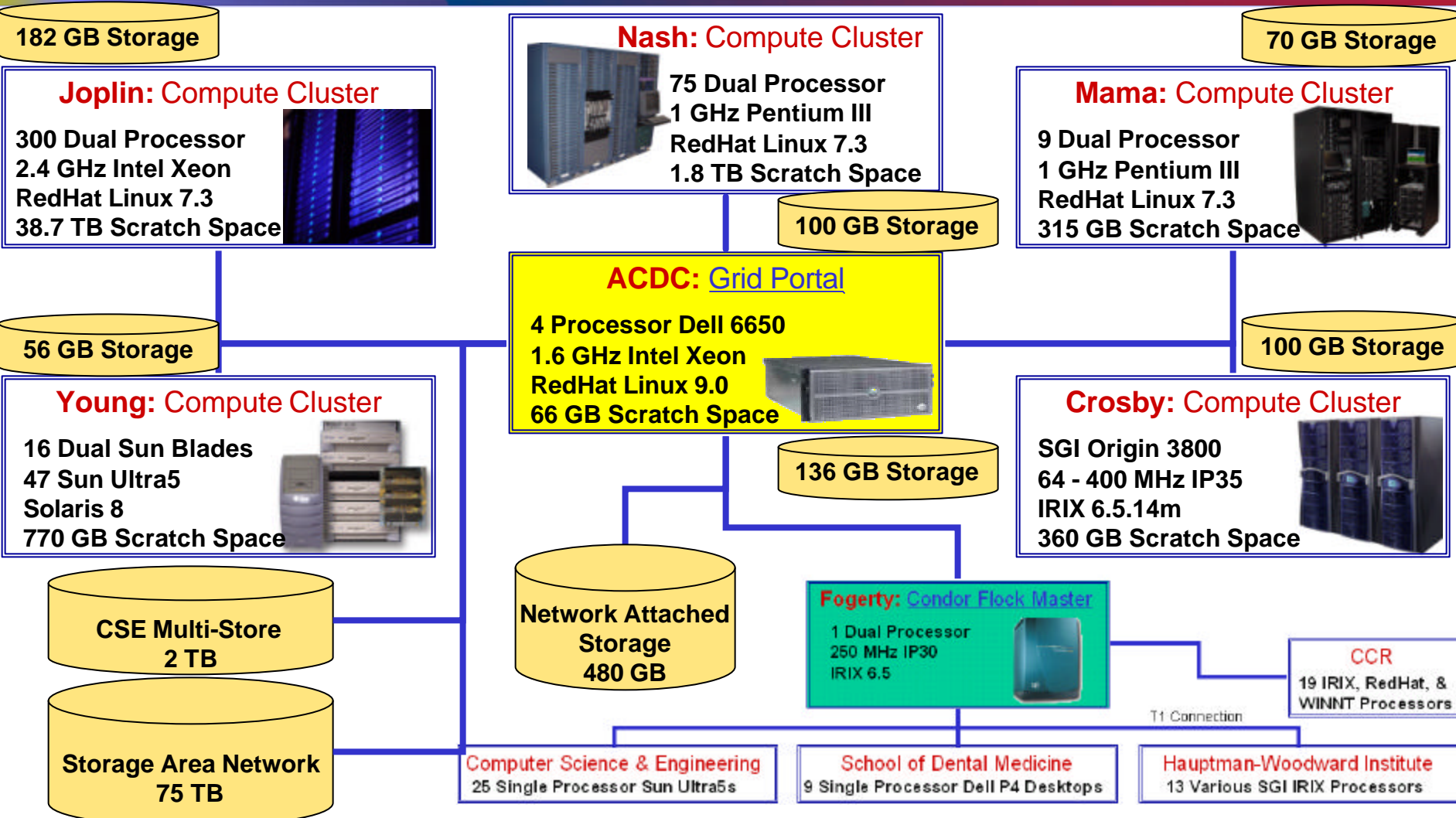
■ Motivation:

- ❑ Large data collections are emerging as important community resources.
- ❑ Data Grids inherently complements Computational Grids, which manipulate data.
- ❑ *A data grid denotes a large network of distributed storage resources such as archival systems, caches, and databases, which are linked logically to create a sense of global persistence.*

■ Goal:

- ❑ To design and implement transparent management of data distributed across heterogeneous resources, such that the data is accessible via a uniform web interface.

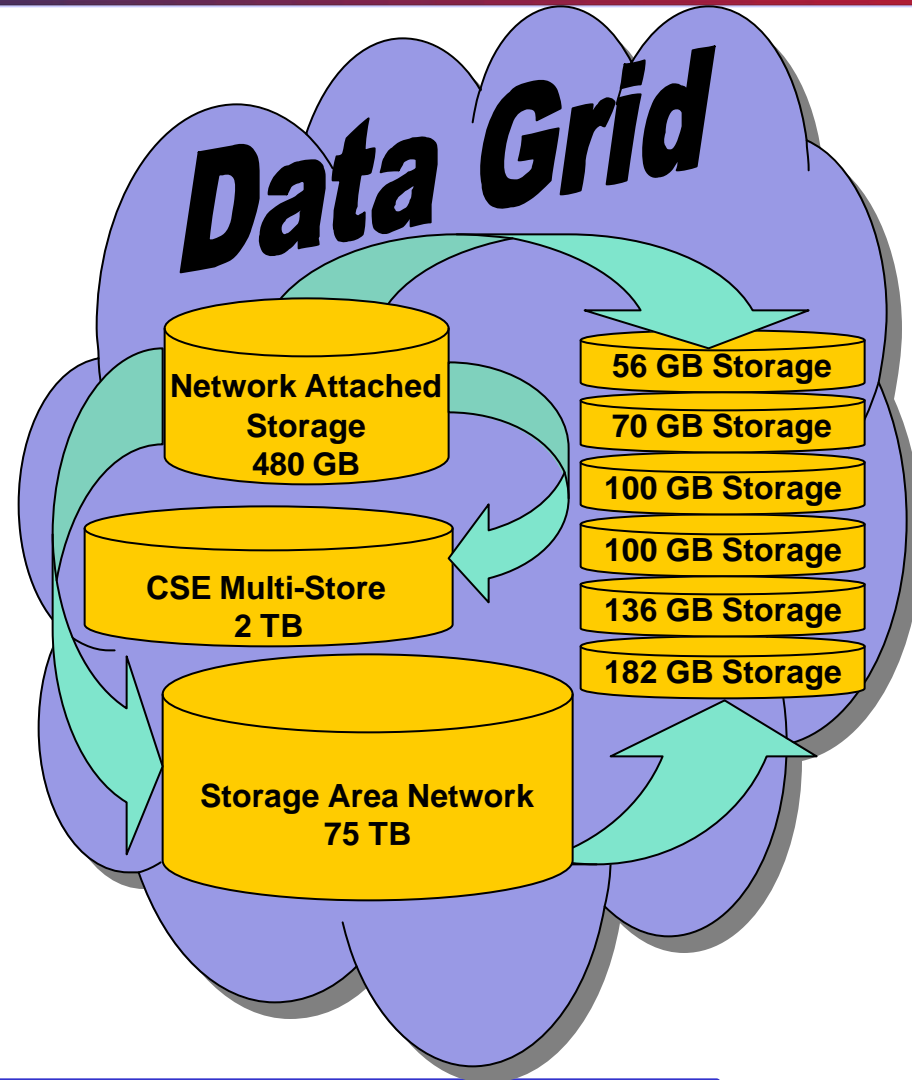
ACDC Data Grid Overview



Note: Network connections are 100 Mbps unless otherwise noted.

Data Grid Summary

- **544 GB Storage**
 - ❑ Located on 6 heterogeneous ACDC-Grid resources
- **480 GB Storage**
 - ❑ Located on 1 dual processor Dell PowerVault server
- **75,000 GB Storage (10/03)**
 - ❑ Served by 4 – 16 processor HP GS1280 servers
- **2,000 GB Storage**
 - ❑ Served by Sun Ultra-60 servers
- **78,024 GB Total Data Grid Storage available and accessible from the ACDC-Grid Portal**



Grid-Based *SnB* Objectives

- **Install Grid-Enabled Version of *SnB***
- **Job Submission and Monitoring over Internet**
- ***SnB* Output Stored in Database**
- ***SnB* Output Mined through Internet-Based Integrated Querying Tool**

- **Serve as Template for Chem-Grid & Bio-Grid**
- **Experience with Globus and Related Tools**

Grid Services and Applications

**ACDC-Grid
Computational
Resources**



Applications

Shake-and-Bake

Apache

MySQL

Oracle

High-level Services and Tools

Globus
Toolkit

NWS

MPI

MPI-IO

C, C++, Fortran, PHP

globusrun

Core Services

Metacomputing
Directory
Service

Globus
Security
Interface

GRAM

GASS

Local Services

Condor

Stork

MPI

RedHat Linux

WINNT

LSF

PBS

Maui Scheduler

TCP

UDP

Irix

Solaris

**ACDC-Grid
Data
Resources**



Adapted from Ian Foster and Carl Kesselman



Grid Enabled *SnB* Execution

□ User

- defines Grid-enabled *SnB* job using Grid Portal or *SnB*
- supplies location of data files from Data Grid
- supplies *SnB* mode of operation

□ Grid Portal

- assembles required *SnB* data and supporting files, execution scripts, database tables.
- determines available ACDC-Grid resources.

□ ACDC-Grid job management includes:

- automatic determination of appropriate execution times, number of trials, and number/location of processors,
- logging/status of concurrently executing resource jobs, &
- automatic incorporation of *SnB* trial results into the molecular structure database.

ACDC-Grid Portal

CCR Grid Computing Services: - Microsoft Internet Explorer

File Edit View Favorites Tools Help

UB University at Buffalo The State University of New York

CCR Center for Computational Research GRID PORTAL

High Performance Grid Computing

PORTAL LOGIN

- Grid General Info
 - » About ACDC Grid
 - » Computational Grid
 - » Data Grid
 - » Publications
 - » Technical Papers
 - » 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 Bioinformatics, 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.

Got your "Grid Computing Guide"?

Do you want to learn about 'Grid Computing'?

S n B

Advanced
Center for Computational Research
Data
Center



ACDC-Grid Portal Login

CCR Grid Computing Services: - Microsoft Internet Explorer

File Edit View Favorites Tools Help

UB University at Buffalo The State University of New York

CCR Center for Computational Research GRID PORTAL

High Performance Grid Computing

PORTAL LOGIN

Grid General Info

- » About ACDC Grid
 - » Computational Grid
 - » Data Grid
 - » Publications
 - » Technical Papers
 - » Presentations
- » Contact Us
- » Grid Account Request
- » Grid Account Support
- » Events
- » News

Projects
Resources
Education/Outreach
Staff Only
CCR HOME

Login

Username:

Password:

Grid Portal login screen

Advanced
Center for Computational Research
Data
Center

Data Grid Capabilities

The screenshot shows a web browser window titled "CCR Grid Computing Services: Data Management - Microsoft Internet Explorer". The page header includes the University at Buffalo logo and the text "Center for Computational Research GRID PORTAL High Performance Grid Computing". A sidebar on the left lists "PORTAL LOGOUT" and various user tools and resources. The main content area displays a "VIEW User" dropdown menu and a file tree for the "mlgreen" user. The tree includes folders like Dozer, Morpheus, Agent, Cypher, KeyMaster, Oracle, Neo, Oracle, Smith, and Tank. A file named "Morpheus.sh" is highlighted under the "Agent" folder. A yellow callout bubble points to this file with the text: "Browser view of 'mlgreen' user files stored in the Data Grid". The footer features a "GRID" logo and the text "Advanced Center for Computational Research Data Center".

Data Grid Capabilities

The screenshot shows a web browser window titled "CCR Grid Computing Services: Data Management - Microsoft Internet Explorer". The page header includes the University at Buffalo logo and the text "Center for Computational Research GRID PORTAL High Performance Grid Computing".

On the left side, there is a "PORTAL LOGOUT" menu with categories: "User Tools" (Manage Account), "Grid General Info", "Projects", "Resources" (Computational Grid, Job Submission, Job/Queue Status, Data Grid, Network Status, Running/Queued Jobs, PBS Job History, Grid Portal Statistics, Condor Flock Statistics, User Information), "Education/Outreach", "Staff Only", and "CCR HOME".

The main content area features a file tree view for the user "rappleye". At the top, there are controls: "VIEW" set to "Group", "GROUP" set to "miller", and "UserList" set to "rappleye". The file tree shows a hierarchy starting with "rappleye", containing folders like "KeyMaster", "Morpheus", "Tank", "Agent", "Rabbit", "Morpheus", "Neo", "Cypher", and "Oracle". A file named "Oracle.m" is highlighted under the "Morpheus" folder. A yellow callout bubble points to this file with the text: "Browser view of 'miller' group files published by user 'rappleye'".

At the bottom of the page, there is a decorative graphic of a molecular structure and the text: "Advanced Center for Computational Research Data".

Grid Portal Job Status

■ Grid-enabled jobs can be monitored using the Grid Portal web interface dynamically.

□ Charts are based on:

- total CPU hours, or
- total jobs, or
- total runtime.

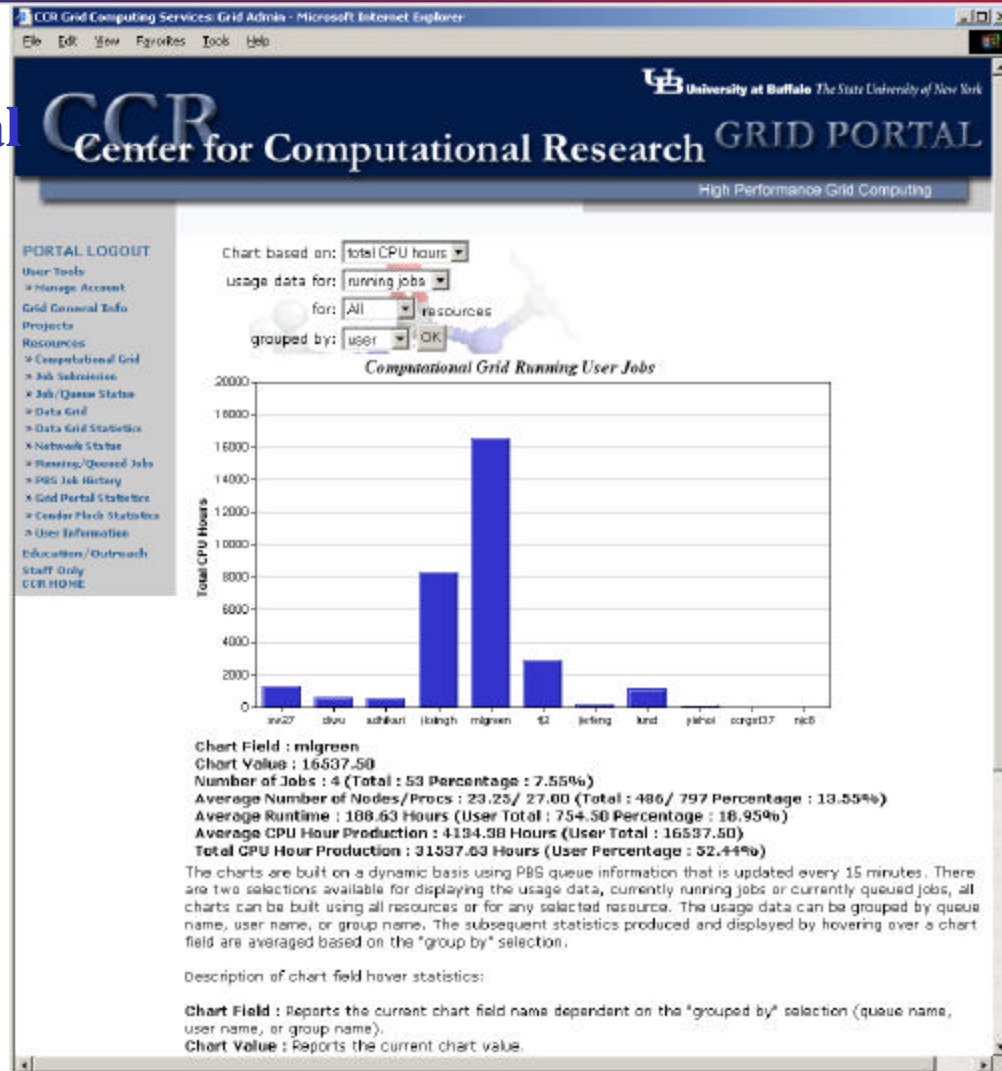
□ Usage data for:

- running jobs, or
- queued jobs.

□ Individual or all resources.

□ Grouped by:

- group, or
- user, or
- queue.



Grid Portal Job Status

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

CCR University at Buffalo The State University of New York 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
 Data Grid Statistics
 Network Status
 Running/Queued Jobs
 PBS Job History
 Grid Portal Statistics
 Center Fleet Statistics
 User Information
 Education/Outreach
 Staff Only
 CCR 8046

Chart based on: total CPU hours
 usage data for: queued jobs
 for: All resources
 grouped by: user

Computational Grid Queued User Jobs

User	Total CPU Hours
mlgreen	7623.17
ksingh	10800
sa27	500
dsingh	6500
ritdici	200
chen938	3000
janovi	1500
janng	800
libon	200

Chart Field : mlgreen
 Chart Value : 7623.17
 Number of Jobs : 27 (Total : 60 Percentage : 45.00%)
 Average Number of Nodes/Procs : 1.96 / 3.93 (Total : 1641 / 3206 Percentage : 9.31%)
 Average Runtime : 71.92 Hours (User Total : 1941.75 Percentage : 71.90%)
 Average CPU Hour Production : 282.34 Hours (User Total : 7623.17)
 Total CPU Hour Production : 31015.17 Hours (User Percentage : 24.58%)

The charts are built on a dynamic basis using PBS queue information that is updated every 15 minutes are two selections available for displaying the usage data, currently running jobs or currently queued charts can be built using all resources or for any selected resource. The usage data can be grouped name, user name, or group name. The subsequent statistics produced and displayed by hovering over field are averaged based on the "group by" selection.

Description of chart field hover statistics:

Chart Field : Reports the current chart field name dependent on the "grouped by" selection (queue user name, or group name).
Chart Value : Reports the current chart value.

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

CCR University at Buffalo The State University of New York 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
 Data Grid Statistics
 Network Status
 Running/Queued Jobs
 PBS Job History
 Grid Portal Statistics
 Center Fleet Statistics
 User Information
 Education/Outreach
 Staff Only
 CCR 8046

Chart based on: total CPU hours
 usage data for: queued jobs
 for: All resources
 grouped by: user mlgreen

Job_Num	Job_Id	Username	Groupname	Queue	Nodes	Cnt	Proc	Cnt	Rank	Request	Time(hr)	Start	Time(hr)
1	23364	mlgreen	miller	grid	1	2	1	71.916666666667		queued			
2	23365	mlgreen	miller	grid	1	2	2	71.916666666667		queued			
3	23366	mlgreen	miller	grid	2	4	3	71.916666666667		queued			
4	23367	mlgreen	miller	grid	2	4	4	71.916666666667		queued			
5	23368	mlgreen	miller	grid	2	4	5	71.916666666667		queued			
6	23369	mlgreen	miller	grid	3	6	6	71.916666666667		queued			
7	23370	mlgreen	miller	grid	2	4	7	71.916666666667		queued			
8	23371	mlgreen	miller	grid	2	4	8	71.916666666667		queued			
9	23372	mlgreen	miller	grid	2	4	9	71.916666666667		queued			
10	23373	mlgreen	miller	grid	2	4	10	71.916666666667		queued			
11	23374	mlgreen	miller	grid	2	4	11	71.916666666667		queued			
12	23375	mlgreen	miller	grid	2	4	12	71.916666666667		queued			
13	23376	mlgreen	miller	grid	2	4	13	71.916666666667		queued			
14	23377	mlgreen	miller	grid	2	4	14	71.916666666667		queued			
15	23378	mlgreen	miller	grid	2	4	15	71.916666666667		queued			
16	23379	mlgreen	miller	grid	2	4	16	71.916666666667		queued			
17	23380	mlgreen	miller	grid	2	4	17	71.916666666667		queued			
18	23381	mlgreen	miller	grid	2	4	18	71.916666666667		queued			
19	23382	mlgreen	miller	grid	2	4	19	71.916666666667		queued			
20	23383	mlgreen	miller	grid	2	4	20	71.916666666667		queued			
21	23384	mlgreen	miller	grid	2	4	21	71.916666666667		queued			
22	23385	mlgreen	miller	grid	2	4	22	71.916666666667		queued			
23	23386	mlgreen	miller	grid	2	4	23	71.916666666667		queued			
24	23387	mlgreen	miller	grid	2	4	24	71.916666666667		queued			
25	23388	mlgreen	miller	grid	2	4	25	71.916666666667		queued			
26	23389	mlgreen	miller	grid	2	4	26	71.916666666667		queued			
27	23393	mlgreen	miller	grid	2	4	27	71.916666666667		queued			

Description of table fields:

Job_Num : Sequential counter for row number.
Job_Id : PBS job identification number corresponding to local PBS queue job number.
Username : PBS job owner local username.
Groupname : PBS job owner local primary group name.

ACDC-Grid Portal User Management

CCR Grid Computing Services: User Admin: Manage Users - Microsoft Internet Explorer

CCR Center for Computational Research GRID PORTAL High Performance Grid Computing

PORTAL LOGOUT

User Tools

Manage Account

Grid General Info

Projects

Resources

Computational Grid

Job Queue Status

Job Queue Status

Data Grid

Data Grid Statistics

Network Status

Running/Queued Jobs

PBS Job History

Grid Portal Statistics

Consider Med. Statistics

New Information

Education / Outreach

Staff Only

CCR HD MC

Manage User Accounts

In order to select which user accounts to manage, you can select one or more usernames from the list below or search for users based on specified criteria. The "Last Name" and "Organization" fields are case sensitive. Selecting "Fuzzy Search" will search on fields containing the text entered. When entering search dates, if both start and end dates are entered then values falling within that range (inclusive) will be returned. Entering only a start date will search for all entries starting with that date while entries up to and including that date.

Username: Fuzzy Search

Account State: Fuzzy Search

Last Name: Fuzzy Search

Organization: Fuzzy Search

Date Added: / / through / /

Last Login: / / through / /

Sort by:

Return to the User Admin menu.

Advanced Center for Computational Research Data Center

Administrator based

CCR Grid Computing Services: User Admin: Edit User Information - Microsoft Internet Explorer

CCR Center for Computational Research GRID PORTAL High Performance Grid Computing

PORTAL LOGOUT

User Tools

Manage Account

Grid General Info

Projects

Resources

Computational Grid

Job Queue Status

Job Queue Status

Data Grid

Data Grid Statistics

Network Status

Running/Queued Jobs

PBS Job History

Grid Portal Statistics

Consider Med. Statistics

New Information

Education / Outreach

Staff Only

CCR HD MC

Edit information for user: mlgreen

Username: mlgreen

State: Active

Last Login: 2003-09-15 09:26:22

Last Logout: 2003-09-15 15:16:47

Date Added: 2003-03-22

Password:

First Name: Mark L.

Last Name: Green

Organization: Center for Computational Research

Address 1: University at Buffalo

Address 2: 0 Norton Hall

City: Buffalo

State: New York

Country: USA

Postal Code: 14260

Phone: 716-645-6500 x522

Fax: 716-645-6505

Email: mlgreen@ccr.buffalo.edu

Url: www.ccr.buffalo.edu

Return to the User Admin menu.

Advanced Center for Computational Research Data Center

user based

ACDC-Grid Portal Resource Management

Resource Access - Microsoft Internet Explorer

The **Grant All** button grants access to all resources in all site areas while **Remove All** removes access for all resources from all site areas. **Save** will save the new user access list and **Cancel** will close this window without saving any changes.

Available Grid Resources

- [All Resources](#)
- CCR-Buffalo-Dev
 - [All Resources](#)
 - Hardware
 - [young.ccr.buffalo.edu](#)
 - [yardbirds.ccr.buffalo.edu](#)
 - [fogerty.ccr.buffalo.edu](#)
 - [mama.ccr.buffalo.edu](#)
 - [joplin.ccr.buffalo.edu](#)
 - [crosby.ccr.buffalo.edu](#)
 - [nash.ccr.buffalo.edu](#)
 - Software
 - [BEAT](#)
 - [POM](#)
 - [SnB](#)
 - Data
 - All resources
- HWI
 - [All Resources](#)
 - Hardware
 - [nexus.hwi.buffalo.edu](#)
 - Software
 - Data
 - All resources

Mark I. Green

- [all](#)
 - Hardware
 - [young.ccr.buffalo.edu](#)
 - [mama.ccr.buffalo.edu](#)
 - [crosby.ccr.buffalo.edu](#)
 - [fogerty.ccr.buffalo.edu](#)
 - [joplin.ccr.buffalo.edu](#)
 - [nexus.hwi.buffalo.edu](#)
 - [yardbirds.ccr.buffalo.edu](#)
 - [nash.ccr.buffalo.edu](#)
 - Software
 - [POM](#)
 - [BEAT](#)
 - [SnB](#)
 - Data
 - All resources
 - [All Resources](#)
 - [general](#)
 - Hardware
 - Software
 - Data
 - All resources
 - [All Resources](#)
 - [support](#)

■ Administrator grants a user access to ACDC-Grid

- resources,
- software, and
- web pages.

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
neius.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 Enabled Data Mining

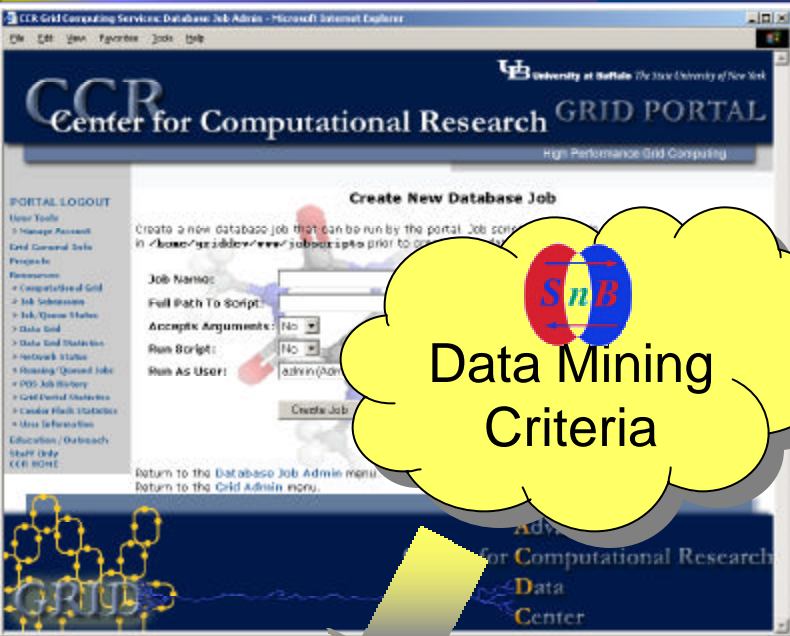
■ Problem Statement

- Use all available resources in the ACDC-Grid for executing a data mining genetic algorithm optimization of *SnB* parameters for molecular structures having the same space group.


■ Grid Enabling Criteria

- All heterogeneous resources in the ACDC-Grid are capable of executing the *SnB* application.
- All job results obtained from the ACDC-Grid resources are stored in a corresponding molecular structure databases.


Grid Enabled Data Mining




**Data Mining
Criteria**


**ACDC-Grid
Data Grid**

**ACDC-Grid Computational
Resources**


**Grid Portal
Workflow Job
Manager**

**Molecular
Structure
Database**

SnB Molecular Structure Database

domain_snb.evo_results running on Grid Portal - phpMyAdmin 2.5.1 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

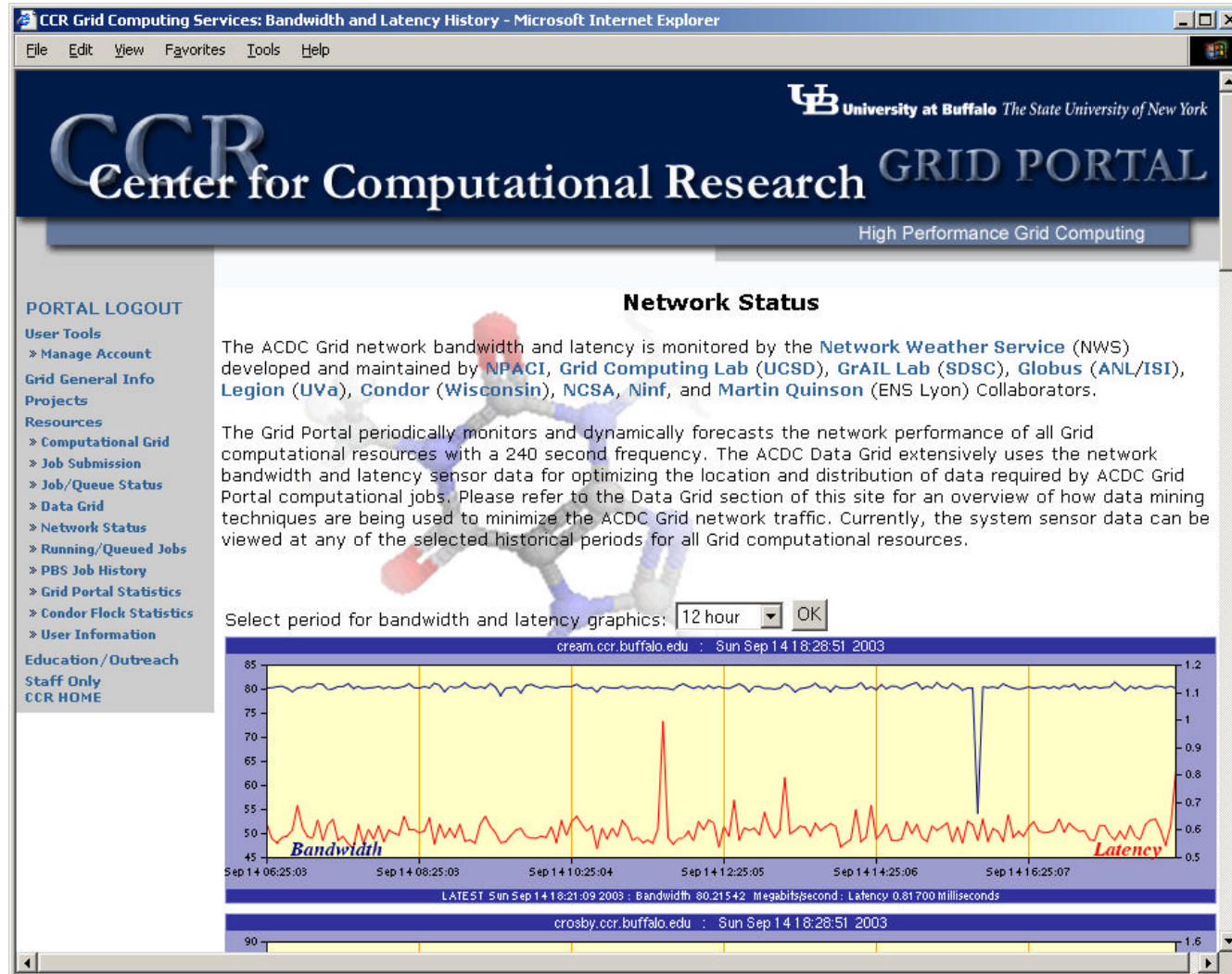
Structure Browse SQL Select Insert Export Operations Options Empty Drop

Field	Type	Attributes	Null	Default	Extra	Action
<input type="checkbox"/> DIR_LOC	varchar(255)		No			Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> PREFIX_OUT	varchar(255)		No			Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> ATOMSIZE	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> NUM_REF	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> RESO_MAX	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> E_SIG_CUT	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> NUM_INV	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> NUM_CYCLE	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> PH_REFINE_METHOD	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> PS_INIT_SHIFT	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> PS_NUM_SHIFT	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> PS_NUM_ITER	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> TAN_NUM_ITER	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> MIN_MAP_RESO	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> NUM_PEAKS_TO_OMIT	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> INTERPOLATE	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> C1	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> C2	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> P1	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> P2	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> NUM_TRIAL	int(11)		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> FUNC_VALUE	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> AVG_RMIN	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> RMIN_CUTOFF	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> RUNTIME	float		No	0		Change Drop Primary Index Unique Fulltext
<input type="checkbox"/> ID	bigint(20)	UNSIGNED	No		auto_increment	Change Drop Primary Index Unique Fulltext

Query window

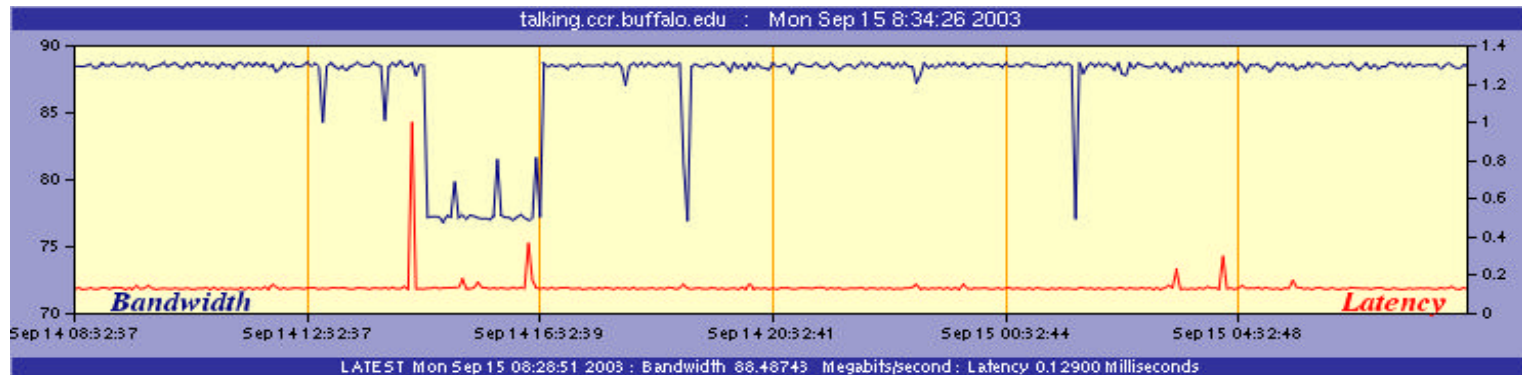
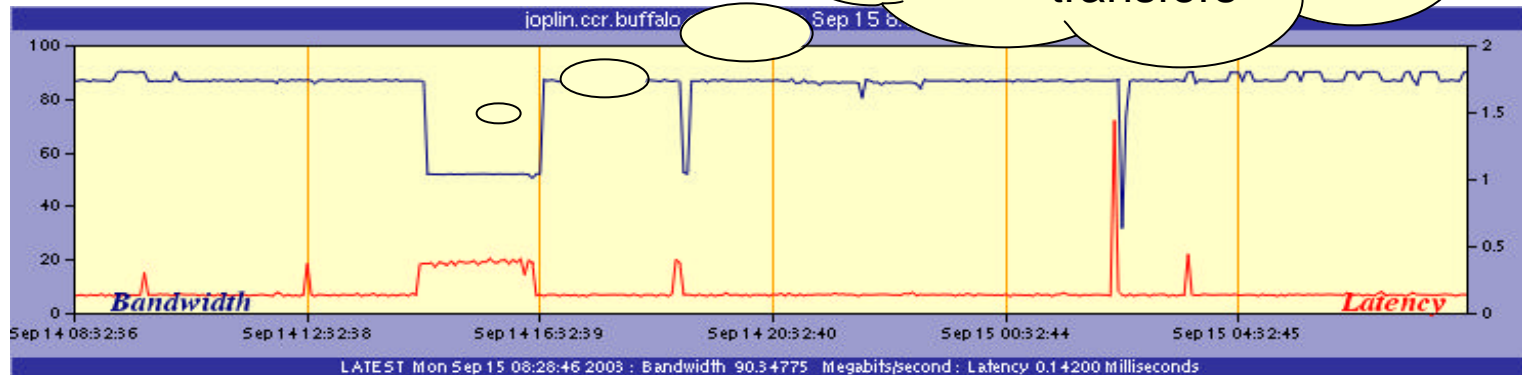
Molecular Structure Database

Data Grid Resource Info



Data Grid Resource Info

Both platforms have reduced bandwidth available for additional transfers



Data Grid File Age

Database *data_grid* - Table *file_management* running on *Grid Portal*

Structure Browse SQL Select Insert Export Operations Options Empty Drop

Showing rows 0 - 29 (13000 total, Query took 0.0022 sec)

SQL-query : [Edit] [Explain SQL] [Create PHP Code]
`SELECT 'File_ID', resource_id, filename, dir_id, access_time, file_age
FROM 'file_management'
WHERE 1 LIMIT 0, 30`

Show: 30 row(s) starting from record # 30
in horizontal mode and repeat headers after 100 cells Page number: 1

		File_ID <small>Unique File ID Value</small>	resource_id	filename	dir_id	access_time	file_age
Edit	Delete	62844	10	Dozer.csh	54030	2003-08-16 11:24:52	925832
Edit	Delete	57120	10	Tank.mpg	53304	2003-04-01 01:45:23	12837001
Edit	Delete	57121	10	Neo.ksh	53505	2003-06-08 06:41:05	6904459
Edit	Delete	57122	10	Trinity.m	53499	2003-07-30 11:34:54	2394030
Edit	Delete	57123	10	Rabbit.ksh	53541	2003-07-13 01:43:34	3898310
Edit	Delete	57124	10	Neo.ksh	53407	2003-06-22 06:19:03	5696181
Edit	Delete	57049	10	Agent.ppt	53928	2003-02-24 12:15:39	15909585
Edit	Delete	61710	10	Neo.txt	52724	2003-07-26 09:44:48	2746236
Edit	Delete	61711	10	Morpheus.sh	52710	2003-07-31 07:03:43	2367101
Edit	Delete	61712	10	Morpheus.ppt	52761	2003-08-26 08:04:38	117046
Edit	Delete	61713	10	Tank.jpg	52929	2003-06-26 09:59:37	5337347
Edit	Delete	61714	10	Rabbit.dat	52624	2003-08-26 05:57:43	124661
Edit	Delete	61715	10	KeyMaster.mpg	52770	2003-06-17 04:16:44	6178720

■ File age, access time, and resource id denote:

- the amount of time since a file was accessed,
- when the file was accessed, and
- where the file currently resides respectively.

ACDC-Grid

Development/Maintenance

■ Development Requirements

❑ 7 – Person months for Grid Services Coordinator

○ Including Grid and Database conceptual design and implementation

❑ 5 – Person months for Grid Services Programmer

○ Web portal programming

❑ 5 – Person months for System Administrator

○ Globus, NWS, MDS, etc. installations

❑ 3 – Person months for Database Administrator

○ Grid Portal Database implementation

■ Minimum Maintenance Requirements

❑ 1 – Grid Services Coordinator

○ 100% level of effort

❑ 1 – Grid Services Programmer

○ 100% level of effort

❑ 1 – System Administrator

○ 50% level of effort

❑ 1 – Database Administrator

○ 10% level of effort

Future ACDC Applications

- **Princeton Ocean Model (POM)**
- **Genetic Algorithms for Earthquake Structural Design**
- **Bioinformatics**
- **Computational Chemistry (Q-Chem)**
- **Environmental Engineering Applications**

ECCE "Grid" at CCR

- **Computational Chemistry**
 - Relativistic effects/Heavy elements
 - Algorithm development
 - Theoretical physical chemistry
- **Structural/Systems Biology**
 - Protein structure
 - Enzyme catalysis
- **Chemical Engineering**
 - Condensed phases/Mixed phase predictions
 - Catalysis
- **Geology, Pharmacology, Medical School**
- **Import Scientific Information**
 - Application independent input
 - ECCE automatically formats for target application (Gaussian98, NWChem)
- **Computing at CCR**
 - 881 available CPUs (>2.5TFlops)
 - (Xeon, P3, Power3, R12K)
 - Uniform access to all platforms via ECCE "job launcher"
- **Chemical Analysis**
 - Full complement of visual tools for understanding data/publication quality graphics



ECCE Periodic Table

File View Help

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc											
Cs	Ba	La	Hf	Ta	W	Re											
Fr	Ra	Ac	Rf	Db	Sg	Bh											
		Ce	Pr	Nd	Pm												
X	Nu	Th	Pa	U	Np												

ECCE - v3.0

exit calculation manager builder basis set tool calculation viewer machine browser periodic table help feedback preferences windows

ECCE Machine Browser

Machine

Configured Machines

coasters
drifters
joplin-production
joplin-short
nash
stills

ECCE Calculation Viewer

Calculation Display View Options Surface Run Mgmt

- Chemical System
- Basis Set: aug-cc-pVDZ
- Launch Info: joplin-short
- Setup Parameters
- Run Statistics
- Energies: -76.0418435622
- Geometry Trace
- Moments
- Normal Modes
- Mulliken Charges

0.051

Iso: [slider]

Queue: feed

ECCE Calculation Manager

Calculation Edit Options Run Mgmt Tools

Ecce Data Server--localhost

- share
- system
- users
 - ccrgst35
 - ecceadm
 - ishulgjn
 - jbednasz
 - jtilson
 - G94-test
 - Project

Type	Name	Reviewed	Creation Date	Modified Date	Application	Formula
Folder	Project		04/28/03 11:25			
File	HF-dimer-CCSD_1_1		05/30/03 13:47	05/30/03 14:35	NWChem	H3F3
File	HF-dimer-CCSD_1	✓	05/30/03 09:20	05/30/03 09:20	NWChem	H2F2
Folder	G94-test		05/30/03 16:02			
File	Calculation_9_1	✓	05/01/03 11:44	05/09/03 15:06	NWChem	H2O
File	Calculation_9		05/01/03 10:44	05/03/03 09:00	NWChem	H2O
File	Calculation_8		05/01/03 10:43	05/03/03 08:59	NWChem	H2O
File	Calculation_7	✓	05/01/03 10:32	05/01/03 10:34	NWChem	CF4

Taskbar with icons for applications and system clock showing 9:57 on 05/31/03.

“Genomics is powering the new biology, but Computing is in the driver’s seat.”

BioGrids

BioGrids provide scalable computing so that biologists can focus on biology.

■ EUROGRID BioGRID



■ Asia Pacific BioGRID



■ NC BioGrid



■ Bioinformatics Research Network



■ Osaka University Biogrid **Biogrid**

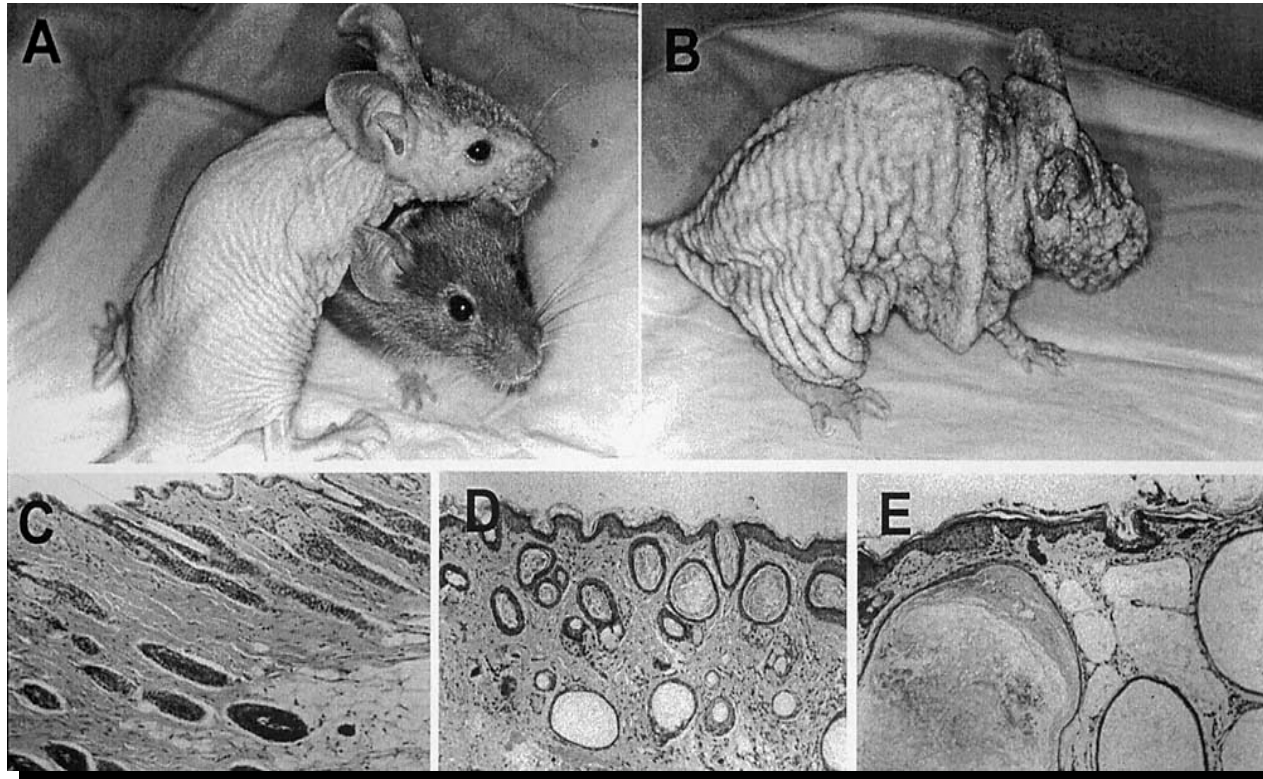
バイオグリッド研究会

■ Indiana University BioArchive BioGrid **IUBio-Archive**

Acknowledgments

- **Mark Green**
- **Steve Gallo**
- **Jason Rappleye**
- **Jeff Tilson**
- **Martins Innus**
- **Cynthia Cornelius**
- **George DeTitta**
- **Herb Hauptman**
- **Charles Weeks**
- **Steve Potter**
- **Philip Glick**
- **Betty Capaldi**
- **Bruce Holm**
- **Janet Penksa**

Contact Information



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