Data-Driven Computing:
Storage, Processing, Networking, and Visualization

Russ Miller
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
Embrace digital data-driven society
Empower students to compete in knowledge-based economy
Support research, scholarship, education, and outreach
Deliver high-end cyberinfrastructure to enable efficient
- Collection of data
- Management/Organization of data
- Distribution of data
- Analysis of data
- Visualization of data
Center for Computational Research
1998-2005 Snapshot

- High-End Computing, Storage, Networking, and Visualization
  - ~100 Research Groups in 37 Depts
    - Physical Sciences
    - Life Sciences
    - Engineering
    - Scientific Visualization, Medical Imaging, Virtual Reality
  - 13 Local Companies
  - 10 Local Institutions

- External Funding: $300M+
- Total Leveraged WNY: $500M+

- Deliverables
  - 1100+ Publications
  - Software, Media, Algorithms, Consulting, Training, CPU Cycles…

University at Buffalo  The State University of New York  Center for Computational Research
CCR by the Numbers

- **Technical Staff: 13**
  - Associate Director
  - Computational Scientist (3)
  - Database Administrator
  - Scientific Visualization
  - System Administration (5)
  - Storage Area Network Admin
  - Multimedia

- **Support Staff: 3 FTE**
  - Financial/Contracts (2)
  - Receptionist

- **Research Staff: 5 FTE**

- **Initial 7-Year Funding Model**
  - SUNY-Buffalo Contribution: $1.3M
    - Personnel: $1.2M
    - Operating: $0.1M
  - User’s Contributions: $0.4M
  - Annual Expend: ~$2.4M
  - Opportunistic Funding Model
    - Equipment, Maintenance, Licenses
  - ROI: $7M → $300M @ SUNY-B

- **Current (New) Funding Model**
  - Personnel + Operating: $677K (2007)
  - Increase Users Contributions
  - Maintain Opportunistic Funding
  - Move into Bioinformatics
  - Provides “Stability”
CCR Research & Projects

- Ground Water Modeling
- Computational Fluid Dynamics
- Molecular Structure Determination
- Protein Folding
- Digital Signal Processing
- Grid Computing
- Computational Chemistry
- Bioinformatics

- Real-time Simulations and Urban Visualization
- Accident Reconstruction
- Risk Mitigation (GIS)
- Medical Visualization
- High School Workshops
- Virtual Reality
CCR-Supported Research at UB

- Physical Sciences
  - Autschbach (Chem), Coppens (Chem), Errington (CBE), Furlani (CCR), Han (Physics), Jones (CCR), King (Chem), Kinney (Physics), Kofke (CBE), Lund (CBE), Markelz (Physics), Ruckenstei (CBE), Sen (Physics), Swihart (CBE), Szyperski (Chem)

- Life Sciences
  - Almon (Biology), Andreadis (Chem Eng), Beal (CSE), DeTitta (Structural Biology), Halfon (Biochemistry), Gaile (Biostatistics), Hu (CCR), Hauptman (Structural Biology), Koffas (Chem Eng), Miller (CSE), Murphy (Medicine), Nowak (RPCI), Sullivan (Ophthalmology), Szyperski (Chem), Trevisan (Public Health), Weeks (Structural Biology), Willsky (Biochemistry), Zhang (CSE), Zhou (Physiology and Biophysics)

- Engineering
  - Atkinson (CSEE), Aref (CSEE), Bisantz (IE), Becker (Geology), Bucher (CCR), Bursik (Geology), Cartwright (EE), Dargush (CSEE), DesJardin (MAE), Frewelling (Geography), Green (CCR), Jankovic (CSEE), Jayaraman (CSE), Jones (CCR), Llinas (IE), Madnia (MAE), Nagi (IE), Patra (MAE), Pitman (Math), Qiao (CSE), Rabideau (CSEE), Reinhorn (CSEE), Sheridan (Geology), Singh (MAE), Upadhyaya (CSE), Zubrow (Anthropology)

- Scientific Visualization, Medical Imaging, Virtual Reality
  - Ansty (Media), Baker (Nuclear Med), Evans (Oral Bio), Geffan (Oral Bio), Hoffmann (Nuclear Med), Innus (CCR), Jones (CCR), Kesavadas (MAE), Lockwood (Neurology, Nuclear Med), Miletich (Nuclear Med), Pape (Media), Paley (Classics), Yao (Nuclear Med)
Major Compute/Storage Resources

- Dell Linux Cluster (10TF)
  - 1600 Xeon EM64T Processors (3.2 GHz)
  - 2 TB RAM; 65 TB Disk
  - Myrinet / Force10
  - 30 TB EMC SAN

- Dell Linux Cluster (2.9TF)
  - 600 P4 Processors (2.4 GHz)
  - 600 GB RAM; 40 TB Disk; Myrinet

- Dell Linux Cluster (6TF)
  - 4036 Processors (PIII 1.2 GHz)
  - 2TB RAM; 160TB Disk; 16TB SAN

- IBM BladeCenter Cluster (3TF)
  - 532 P4 Processors (2.8 GHz)
  - 5TB SAN

- SGI Intel Linux Cluster (0.1TF)
  - 150 PIII Processors (1 GHz)
  - Myrinet

- SGI Altix3700 (0.4TF)
  - 64 Processors (1.3GHz ITF2)
  - 256 GB RAM
  - 2.5 TB 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
CCR Visualization Resources

- Fakespace ImmersaDesk R2
  - Portable 3D Device
  - Onyx2: 6 R10000 @ 250MHz
  - 2 IR2 Pipes; 3 64MB texture memory mgrs.

- Tiled-Display Wall
  - 20 NEC projectors: 15.7M pixels
  - Screen is 11’x7’
  - Dell PCs with Myrinet2000

- Access Grid Nodes (2)
  - Group-to-Group Communication
  - Commodity components

- SGI Reality Center 3300W
  - Dual Barco’s on 8’x4’ screen
  - Onyx300: 10 R14000 @ 500MHz
  - 2 IR4 Pipes; 1 GB texture mem per pipe
Peace Bridge Visualization

The Problem
- 75 year old bridge
- 3 lanes – poor capacity
- Existing US plaza: small and poor design

Proposed Options
- Relocate US plaza
- Build a 3-lane companion span, rehab existing bridge
- Build a six lane signature span
Williamsville Toll Barrier Improvement Project

Initial Photo Match incorporating real and computer-generated components
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.
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
Multiple Sclerosis Project

- Compare caudate nuclei between MS patients and healthy controls
- Looking for size as well as structure changes
  - Localized deformities
  - Spacing between halves
- Able to see correlation between disease progression and physical structure changes
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
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
Geophysical Mass Flow Modeling

- Modeling of Volcanic Flows, Mud flows (flash flooding), and Avalanches
- Integrate information from several sources
  - Simulation results
  - Remote sensing
  - GIS data
- Develop realistic 3D models of mass flows
- Present information at appropriate level
Bioinformatics in Buffalo
A $360M Initiative

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

Lead Institutions
- SUNY-Buffalo
- Roswell-Park Cancer Institute
- Hauptman-Woodward Medical Research Institute
Recent Biomedical Advances (Buffalo, NY)

- PSA Test (screen for Prostate Cancer)
- Avonex: Interferon Treatment for Multiple Sclerosis
- Artificial Blood
- Nicorette Gum
- Fetal Viability Test
- 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
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.
Northeast Structural Genomics Consortium

- **Consortium**
  - UB, Rutgers, Columbia, Cornell, PNNL, Yale, UTToronto, Robert Wood Johnson Medical Center, Hauptman-Woodward Medical Research Center

- **Mission**
  - Develop integrated technologies for high-throughput (htp) protein production and 3D structure determination
  - The goal is to determine 500 new protein structures over 5 years
  - Combination of strong parallel efforts in both X-ray crystallography and solution-state NMR spectroscopy
  - UB Professor Thomas Szyperski awarded Scientific American’s Top 50 Scientists in 2003 for novel work in high-throughput structure determination with NMR
Computational Chemistry

- UB Software Development in Quantum Chemistry
  - Q-Chem – development of parallel algorithms and combined QM/MM methods for large molecular systems
  - ADF – development of algorithms to calculate magnetic and optical properties of molecules

- Used to determine
  - Molecular Structure
  - Electronic Spectra
  - Chemical Reactivity

- Applications
  - Pharmaceutical Drug Design
  - Industrial Catalysis
  - Materials Science
  - Nanotechnology
  - Solution Phase Chemistry
  - Chemical Kinetics
Western New York
Health Information Project

Goals:
- Build a secure community-wide healthcare database
- Develop an electronic patient medical record that “follows the patient”
- Provide care providers with real-time patient information wherever they are
- Provide a tool to aid agencies in community safety, epidemiology, resource allocation, and bioterrorism response
- Improve the overall quality of healthcare while reducing costs

Selected Participants:
- SUNY-Buffalo (CCR, School of Informatics, School of Medicine, Health Science Library)
- Buffalo Academy of Medicine
- Erie County DoH
- New York State DoH
- WNY HealtheNet
- Involvement from Kaleida Health, ECMC, Catholic Health System, Independent Health, HealthNow, and Univera Healthcare
Molecular Structure Determination via Shake-and-Bake

- **SnB Software by UB/HWI**
  - “Top Algorithms of the Century”

- **Worldwide Utilization**

- **Critical Step**
  - Rational Drug Design
  - Structural Biology
  - Systems Biology

- **Vancomycin**
  - “Antibiotic of Last Resort”

- **Current Efforts**
  - Grid
  - Collaboratory
  - Intelligent Learning

1. Isolate a single crystal
2. Perform the X-Ray diffraction experiment
3. Determine the crystal structure
Experiment yields reflections and associated intensities.

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

Phase angles are lost in experiment.

Phase Problem: Determine the set of phases corresponding to the reflections.
Conventional Direct Methods

- Trial Phases
- Phase Refinement
- Tangent Formula
- FFT
- Density Modification (Peak Picking)
- Solutions

Reciprocal Space
Real Space
Shake-and-Bake Method: Dual-Space Refinement

Trial Structures → Structure Factors

Trial Phases → Phase Refinement

Phase Refinement → Tangent Formula

Tangent Formula → FFT

FFT → Density Modification (Peak Picking) (LDE)

Density Modification (Peak Picking) (LDE) → Parameter Shift

Parameter Shift → FFT⁻¹

FFT⁻¹ → Solutions

Reciprocal Space

“Shake”

Real Space

“Bake”
Useful Relationships for Multiple Trial Phasing

Tangent Formula
\[ \tan \phi_H = \frac{-\sum_K E_K E_{-H-K} \sin(\phi_K + \phi_{-H-K})}{\sum_K E_K E_{-H-K} \cos(\phi_K + \phi_{-H-K})} \]

Parameter Shift Optimization
\[ R(\phi) = \frac{1}{\sum_{H,K} W_{HK}} \sum_{H,K} W_{HK} \left( \cos \Phi_{HK} - \frac{I_1(W_{HK})}{I_0(W_{HK})} \right)^2 \]

where \( |E_H| \propto |F_H| \) normalized in resolution shells

Invariants: \( \Phi_{HK} = \phi_H + \phi_K + \phi_{-H-K} \approx 0 \)

Weights: \( W_{HK} = A_{HK} = 2N^{-1/2} |E_H E_K E_{-H-K}| \)
Ph8755: SnB Histogram

Atoms: 74  Phases: 740
Space Group: P1  Triples: 7,400

Trials: 100
Cycles: 40
Rmin range: 0.243 - 0.429
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

0  100  1,000  10,000  100,000
Vancomycin

- Interferes with formation of bacterial walls
- *Last line of defense* against deadly
  - streptococcal and staphylococcal bacteria strains
- Vancomycin resistance exists (Michigan)
- Can’t just synthesize variants and test
- Need structure-based approach to predict
- Solution with *SnB (Shake-and-Bake)*
  - Pat Loll
  - George Sheldrick
Integrated 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.

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

Dynamic Resource Allocation
- Develop automated procedures for dynamic computational resource allocation.
**ACDC Data Grid Overview**

(Grid-Available Data Repositories)

**Joplin:** Compute Cluster
- 300 Dual Processor
- 2.4 GHz Intel Xeon
- RedHat Linux 7.3
- 38.7 TB Scratch Space

**Nash:** Compute Cluster
- 75 Dual Processor
- 1 GHz Pentium III
- RedHat Linux 7.3
- 1.8 TB Scratch Space

**Mama:** Compute Cluster
- 9 Dual Processor
- 1 GHz Pentium III
- RedHat Linux 7.3
- 315 GB Scratch Space

**Young:** Compute Cluster
- 16 Dual Sun Blades
- 47 Sun Ultra5
- Solaris 8
- 770 GB Scratch Space

**Crosby:** Compute Cluster
- SGI Origin 3800
- 64 - 400 MHz IP35
- IRIX 6.5.14m
- 360 GB Scratch Space

**CSE Multi-Store**
- 40 TB

**Storage Area Network**
- 75 TB

**Network Attached Storage**
- 1.2 TB

**ACDC: Grid Portal**
- 4 Processor Dell 6650
- 1.6 GHz Intel Xeon
- RedHat Linux 9.0
- 66 GB Scratch Space

**100 GB Storage**

**100 GB Storage**

**136 GB Storage**

**Note:** Network connections are 100 Mbps unless otherwise noted.
ACDC-Grid Collaborations

- High-Performance Networking Infrastructure
- WNY Grid Initiative
- Grid3+ Collaboration
- iVDGL Member
  - Only External Member
- Open Science Grid Member
  - Organizational Committee
  - Blueprint Committee
  - Security Working Group
  - Data Working Group
- Grid-Based Visualization
  - SGI Collaboration
- Grid-Lite: Campus Grid
  - HP Labs Collaboration
- Innovative Laboratory Prototype
  - Dell Collaboration
Grid-Enabling Application Templates (GATs)

- Structural Biology
  - *SnB* and *BnP* for Molecular Structure Determination/Phasing
- Groundwater Modeling
  - *Ostrich*: Optimization and Parameter Estimation Tool
  - *POMGL*: Princeton Ocean Model Great Lakes for Hydrodynamic Circulation
  - *Split*: Modeling Groundwater Flow with Analytic Element Method
- Earthquake Engineering
  - *EADR*: Evolutionary Aseismic Design and Retrofit; Passive Energy Dissipation System for Designing Earthquake Resilient Structures
- Computational Chemistry
  - *Q-Chem*: Quantum Chemistry Package
- Geographic Information Systems & BioHazards
ACDC-Grid

Browser view of “miller” group files published by user
Advanced Computational Data Center Grid Jobs

**Grid Job Submission:**
This section contains forms for the selection of a grid-enabled application, modification of an application template, grid job definition review and grid job submission.

**Grid Job Status:**
This section contains grid user based specific grid job completion status, grid job current state (COMPLETE, RUNNING, QUEUED, BLOCKED, FAILED, ETC.), detailed information on all running or queued grid jobs and grid-enabled application specific intermediate and post processing grid job graphics, plots and tables.
Instructions and Description for Running a Job on ACDC-Grid

Advanced Computational Data Center Grid Job Submission Instructions

The grid-enabling application templates used on the ACDC-Grid are created from the application developers' grid user profiles that contain the users' standard information such as name, organization, address, etc., and more specific information such as group id and access level for each grid-enabled application. This information is stored in a database for each of the grid-enabled applications and can be accessed through a user profile through the ACDC-Grid Web Portal.

Additionally, each grid-enabled scientific application profile contains information about specific execution parameters, required data files, optional data files, computational requirements, etc., and statistics on historical ACDC-Grid jobs for predictive runtime estimates. MySQL provides the speed and reliability required for this task and is currently being used as the ACDC-Grid Web Portal database provider.

The grid-enabled versions of many well-defined scientific and engineering applications have very similar general requirements and core functionality that are required for execution in the ACDC-Grid environment. We have identified that sequentially defining milestones for the grid user to complete intuitively guides them through the application workflow.

**Software Application:** Grid user chooses a grid-enabled software application.

**Template:** Grid user selects the required and/or optional data files from the ACDC Data Grid. User-defined computational requirements are input or a template defines computational requirement runtime estimate is selected.

**Job Definition:** Grid user defines application specific runtime parameters or accepts default template parameter definitions.

**Review:** Grid user accepts the template complete job definition workflow or corrects any part of job definition.

**Execution Scenario:** The grid user has the ability to input an execution scenario or select a ACDC-Grid determined template defined execution scenario.

**Grid Job Status:** The grid user can view specific grid job completion status, grid job current state (COMPLETE, RUNNING, QUEUED, BLOCKED, FAILED, etc.), detailed information on all running or queued grid jobs, and grid-enabled application specific intermediate and post-processing grid job graphics, plots and tables.

Each item of the job definition workflow is then stored in the ACDC-Grid Web Portal database so the grid user may use/modify any previously created workflow in creating new job definitions. The job definitions can also be accessed via batch script files for executing hundreds of similar workflows in an automated fashion. For example, a grid user would first define/save a relatively generic job workflow template for the grid-enabled application and then use the batch script capabilities to change the job definition workflow data files or application parameters and execute a series of new grid jobs.
Default Parameters Based on Template
Initial Data Sets

Add Dataset  
Delete Dataset

Select dataset to delete: 

Datasets

Name (8 chars max): 

Dataset Type: 

File Name (*.hkl): 

Browse

File Type: 

Wavelength: 

Max. Resolution: 

Anomalous Dispersion: 

Heavy Element Type: 

Nat. Element Replaced: 

No. Expected Sites: 

F Prime (F): 

F Double Prime (F''): 

Continue  
Reset Sequence  
Reset Current Stage  
Cancel

Default Parameters (cont’d)
Generating Reflections (Drear)

Reflections and Invariants

Drear Table

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Job Type</th>
<th>Native Data</th>
<th>Derivative Data</th>
<th>Norm Method</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>iledhkl</td>
<td>BASIC</td>
<td>iledhkl</td>
<td>NULL</td>
<td>Wilson (Anisotropic)</td>
<td></td>
</tr>
</tbody>
</table>

Normalization Data

- Data resolution cutoffs (in Angstroms)?
  - Low: 0.99
  - High: 0.94
- Use Bayesian estimates for weak reflections?
  - No
- Min |F| / |sig(F)| for local scaling:
  - 0.0
- SIR and SAS cutoffs:
  - Tmax: 6.0
  - ZMax: 3.0
  - XMin: 3.0
  - YMin: 1.0

Generate Invariants

- Data resolution cutoffs ?
  - Low: 0.99
  - High: 0.94
- Minimum allowed |IE| / |sig(IE)|:
  - 3.0
  - Maximum |IE| : 5.0
- Minimum allowed invariants / reflection ratio:
  - 6.0

Initial values for adjustable parameters

- Minimum |IE| / |sig(IE)| = ZMin:
  - 3.0
- Number of reflections to use:
  - 0
- Number of invariants to save:
  - 0
Reflections and Invariants

Drear Table

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Job Type</th>
<th>Native Data</th>
<th>Derivative Data</th>
<th>Norm Method</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>iledhkl</td>
<td>BASIC</td>
<td>iledhkl</td>
<td>NULL</td>
<td>Wilson (Anisotropic)</td>
<td></td>
</tr>
</tbody>
</table>

Normalization Data

Data resolution cutoffs (in Angstroms)?
Low: 399.0, High: 0.94

Use Bayesian estimates for weak reflections?
No

Min |E| / |E| for local scaling:
3.0

SIR and SAS cutoffs:

<table>
<thead>
<tr>
<th>Tmax</th>
<th>ZMax</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Generate Invariants

Data resolution cutoffs?
Low: 399.0, High: 0.94

Minimum allowed |E| / |E|:
3.0

Minimum allowed |E| / |E| reflection ratio:
5.0

Initial values for adjustable parameters:
Minimum |E| / |E| = ZMin:
3.0

Number of reflections to use:
340

Number of invariants to save:
5400

Invariant Generation
SnB Setup

Grid Parameters
- Preferred resource name: Grid Scheduler
- Number of processors: 8
- Wallclock time requested: 720 mins
- Job Prefix for results: job00
- Queue: grid

SnB Run Parameters
- Invariants
  - Number of triplet invariants to use: 8400
- Trials To Process
  - Random seed (prime): 11909
  - Number of Trials: 1000
  - Starting Trial: 1
- Input Phase File: none
- Input Atom File: none
- Keep complete (every trial) peak file?: Yes

- Cycles Information
  - Number of Shake-and-Bake cycles: 20
  - Keep complete (every cycle) trace file?: No
  - Terminate trials failing the R-Ratio test?: No
  - R-Ratio cutoff: 0.20

SnB Setup
**Phase Refinement Method**

**Number of passes through phase set:** 3

**Phase shift:** 30.0

**Number of shifts:** 2

**Real-Space Constraints**

**Number of peaks to select:** 84

**Minimum interpeak distance:** 3

**Minimum distance between symmetry-related peaks:** 0.0

**Number of special position peaks to keep:** 0

**Fourier grid size:** 0.31

**Perform extra cycles with more peaks?** No

**Number of extra cycles:** 3

**Number of peaks:** 84

**Twice Baking**

**Trials for E-Fourier filtering (fourier refinement)?** None

**Number of cycles:** 3

**Number of peaks:** 84

**Minimum |E|:** 0.75

**Automatic solution identification criteria**

**Rmin Improvement (%):** 95.0

**Rcryst Improvement (%):** 25.0

SnB Setup (cont’d)
### SnB Job Review

**Grid Job ID:** 447

**Selected resource:** clearwater.ccr.buffalo.edu

**Number of processors:** 5

**Wallclock time requested:** 720

**Number of triplet invariant to use:** 8400

**Start Phases From:** Random Atoms

**Number of trials:** 11909

**Starting Trial:** 1

**Input Phase File:** Unused

**Input Atom File:** Unused

**Keep complete (every trial) peak file?** Yes

**Number of Shake-and-bake cycles:** 20

**Keep complete (every cycle) trace file?** No

**Terminate trials failing the R-Ratio test?** No

**R-Ratio cutoff:** Unused

**Phase Refinement Method:** Parameter Shift(Fast)

**Number of passes through phase set:** 3

**Phase shift:** 90.0

**Number of shifts:** 2

**Number of peaks to select:** 84

**Minimum interpeak distance:** 3

**Minimum distance between symmetry-related peaks:** 3.0

**Number of special position peaks to keep:** 0

**Fourier grid size:** 0.31

**Perform extra cycles with more peaks?** No

**Number of extra cycles:** Unused

**Number of peaks:** Unused

**Trials for E-Fourier filtering (fourier refinement)?** None

**Number of cycles:** Unused

**Number of peaks:** Unused

**Minimum |E|:** Unused
Graphical Representation of Intermediate Job Status
Histogram of Completed Trial Structures
Grid Job 447 Walltime Summary
Walltime Consumed: 2 (0.3%)

Walltime Left: 99.72%

Walltime Consumed: 0.28%

Walltime Summary Chart
### Grid Job Status


#### Job Filter Criteria

- **Show GATS**
  - SnB
  - SnB DREAM

<table>
<thead>
<tr>
<th>Job Id</th>
<th>Job Name</th>
<th>Resource</th>
<th>Num Procs</th>
<th>Status</th>
<th>Percent Complete</th>
<th>Last Update</th>
<th>Cancel Job</th>
<th>Drilldown</th>
</tr>
</thead>
<tbody>
<tr>
<td>447</td>
<td>iiledhkl</td>
<td>clearwater.ccr.buffalo.edu</td>
<td>5</td>
<td>RUNNING</td>
<td>28.5</td>
<td>15-Mar-2005 10:22:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>trllys</td>
<td>clearwater.ccr.buffalo.edu</td>
<td>10</td>
<td>RUNNING</td>
<td>1</td>
<td>15-Mar-2005 10:22:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>444</td>
<td>64chkl</td>
<td>nash.ccr.buffalo.edu</td>
<td>3</td>
<td>COMPLETE</td>
<td>100</td>
<td>14-Mar-2005 22:00:01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>443</td>
<td>trllys</td>
<td>clearwater.ccr.buffalo.edu</td>
<td>10</td>
<td>COMPLETE</td>
<td>100</td>
<td>10-Mar-2005 22:48:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>442</td>
<td>pr43hkl</td>
<td>nash.ccr.buffalo.edu</td>
<td>5</td>
<td>COMPLETE</td>
<td>100</td>
<td>10-Mar-2005 17:26:01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>vancohkl</td>
<td>clearwater.ccr.buffalo.edu</td>
<td>10</td>
<td>COMPLETE</td>
<td>100</td>
<td>10-Mar-2005 18:08:01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>434</td>
<td>15chkl</td>
<td>clearwater.ccr.buffalo.edu</td>
<td>5</td>
<td>COMPLETE</td>
<td>100</td>
<td>10-Mar-2005 14:42:01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>433</td>
<td>15chkl</td>
<td>clearwater.ccr.buffalo.edu</td>
<td>8</td>
<td>COMPLETE</td>
<td>100</td>
<td>10-Mar-2005 14:38:01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
User starts up – default image of structure.
Molecule scaled, rotated, and labeled.
Remove Carbon Atoms (and Links)
User Adds Bond Between Atoms
HS Summer Workshops in Computational Science

- Chemistry, Bioinformatics, Visualization
- 10-14 HS Students Participate Each Summer for 2 weeks
- Project-Based Program
Pilot HS Program in Computational Science

- Year long extracurricular activity at Mount St. Mary’s, City Honors, and Orchard Park HS
- Produce next generation scientists and engineers
- Students learn Perl, SQL, Bioinformatics
- $50,000 startup funding from Verizon, PC’s from HP
**Business First**

**HMOs cut Medicare premiums**

Patio home development proposed for Town of Aurora

**Grammy: Designs on Buffalo**

Ani DiFranco, art director up for award

**UB brings bioinformatics to a younger generation**

By DMITRY D. SNYDER

For most of Daisy Brown's childhood, science classes have been intrusive, but somewhat interesting. This all changed last year when the 11-year-old learned about genetics, DNA, and how genes are copied and divided.

That's the impact of the miniaturized version of a test for DNA activity that University at Buffalo biology professor Donald J. Grandinetti has developed. For the first time, the test can be used to create a DNA barcode.

Grandinetti has used the test to create a DNA barcode and is now using it to analyze DNA samples. He has also created a DNA barcode for a variety of other applications, including the study of cancer genetics.

**An early look at bioinformatics**

By DMITRY D. SNYDER

For most of Daisy Brown's childhood, science classes have been intrusive, but somewhat interesting. This all changed last year when the 11-year-old learned about genetics, DNA, and how genes are copied and divided.

That's the impact of the miniaturized version of a test for DNA activity that University at Buffalo biology professor Donald J. Grandinetti has developed. For the first time, the test can be used to create a DNA barcode.

Grandinetti has used the test to create a DNA barcode and is now using it to analyze DNA samples. He has also created a DNA barcode for a variety of other applications, including the study of cancer genetics.
Acknowledgments

- Mark Green
- Amin Ghadersohi
- Naimesh Shah
- Steve Gallo
- Jason Rappleye
- Jon Bednasz
- Sam Guercio
- Martins Innus
- Cynthia Cornelius

- George DeTitta
- Herb Hauptman
- Charles Weeks
- Steve Potter
- Bruce Holm
- Janet Penksa

- NSF, NIH, NYS, NIMA, NTA, Oishei, Wendt, DOE
www.ccr.buffalo.edu