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
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 $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, Zeptometrix
  - Dell, HP, SGI, Stryker, Sun
  - AT&T, Sloan Foundation
  - InforMax, Q-Chem, 3M, Veridian
  - BioPharma Ireland, Confederation of Indian Industries
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
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

**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

**Fogerty: Condor Flock Master**
1 Dual Processor
250 MHz IP30
IRIX 6.5

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

**Expanding**
RedHat, IRIX, Solaris, WINNT, etc

**T1 Connection**

**Computer Science & Engineering**
25 Single Processor Sun Ultra5s

**School of Dental Medicine**
9 Single Processor Dell P4 Desktops

**Hauptman-Woodward Institute**
13 Various SGI IRIX Processors

Note: Network connections are 100 Mbps unless otherwise noted.
Network Connections

CCR Center for Computational Research
www.ccr.buffalo.edu

1000 Mbps

100 Mbps

1.54 Mbps (T1) - RPCI

1.54 Mbps (T1) - HWI

44.7 Mbps (T3) - BCOEB

155 Mbps (OC-3) I2

622 Mbps (OC-12)

100 Mbps

Medical/Dental

OC-3 - I1

NYSERNet 350 Main St

Commercial

Abilene

BCOEB

FDDI

Roswell Park Cancer Institute

University at Buffalo The State University of New York Center for Computational Research

NYSERNet 350 Main St

Commercial
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**
- 9 Dual Processor
- 1 GHz Pentium III
- RedHat Linux 7.3
- 315 GB Scratch Space

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

**Network Attached Storage**
- 1.2 TB

- Storage Area Network
- 75 TB

- CSE Multi-Store
- 2 TB

- 182 GB Storage

- 56 GB Storage

- 100 GB Storage

- 100 GB Storage

- 70 GB Storage

- 770 GB Scratch Space

- 66 GB Scratch Space

- 136 GB Storage

Note: Network connections are 100 Mbps unless otherwise noted.
Browser view of “miller” group files published by user
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