The Center for Computational Research

Russ Miller
Director, Center for Computational Research
UB Distinguished Professor, Computer Science & Engineering
Senior Research Scientist, Hauptman-Woodward Medical Inst

University at Buffalo
The State University of New York
Gordon E. Moore

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

A computation that took 1 year to run on a PC in 1985 would only take 5 mins to run on a PC today!

A computation that runs in 2 hours on a PC today would have taken 24 years to run on a PC in 1985!
Supercomputers

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

Cray1 - 1976

“Seymour Cray is the Thomas Edison of the supercomputing industry”
- Larry L. Smarr

Seymour Cray
1925-1996

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

- Industry Standard Hardware and Software
  - PC-Based Components (Intel or AMD)
  - Ethernet or Myrinet
  - Linux, PBS, MPI
  - “Commodity Off-The-Shelf” (COTS)

- Operates as a Single System

- Rivals Performance of Traditional Supercomputer at a Fraction of the Price
## Fastest Computers

<table>
<thead>
<tr>
<th>Year</th>
<th>Mach</th>
<th>Procs</th>
<th>GFlops</th>
<th>Year</th>
<th>Mach</th>
<th>Procs</th>
<th>GFlops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>Cray 1</td>
<td>1</td>
<td>0.1</td>
<td>1993</td>
<td>Cray T3D</td>
<td>1024</td>
<td>152</td>
</tr>
<tr>
<td>1982</td>
<td>Cray X-MP</td>
<td>4</td>
<td>0.9</td>
<td>1994</td>
<td>Fujitsu VPP</td>
<td>140</td>
<td>236</td>
</tr>
<tr>
<td>1986</td>
<td>Cray 2</td>
<td>4</td>
<td>2</td>
<td>1996</td>
<td>Hitachi SR2</td>
<td>2048</td>
<td>368</td>
</tr>
<tr>
<td>1989</td>
<td>Cray Y-MP</td>
<td>8</td>
<td>2.7</td>
<td>1997</td>
<td>Intel ASCI-R</td>
<td>9152</td>
<td>1830</td>
</tr>
<tr>
<td>1989</td>
<td>TMC CM-2</td>
<td>8192</td>
<td>28</td>
<td>1999</td>
<td>SGI ASCI-BM</td>
<td>6144</td>
<td>3072</td>
</tr>
<tr>
<td>1992</td>
<td>TMC CM-5</td>
<td>1024</td>
<td>131</td>
<td>2000</td>
<td>IBM ASCI-W</td>
<td>8192</td>
<td>12,288</td>
</tr>
<tr>
<td>2002</td>
<td>NEC E.S.</td>
<td>5120</td>
<td>40,960</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A 1-year calc in 1980 = 5.4 sec today
A 1990 HPC = a laptop today
Growth of Peak Performance
Earth Simulator

- 40TFlops Peak
- Homogeneous, Centralized,

640 NEC SX/6 Nodes (5120 CPUs)

Footprint = 4 tennis courts

$6M/year in power
Center for Computational Research
1999-2003 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
    - $13.5M as lead
    - $97.5M in support
  - $41.8M Vendor Donations

- Deliverables
  - 350+ Publications
  - Software, Media, Algorithms, Consulting, Training, CPU Cycles…
Major CCR Resources

- 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 SN
  - Restricted Use (Skolnick)
- 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
**Advanced CCR Data Center (ACDC)**

**Computational 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

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

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

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

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

- **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

- **FDDI**: 100 Mbps
- **1.54 Mbps (T1) - RPCI**
- **44.7 Mbps (T3) - BCOEB**
- **1.54 Mbps (T1) - HWI**
- **BCOEB**
- **100 Mbps**
- **1000 Mbps**
- **155 Mbps (OC-3) I2**
- **155 Mbps (OC-3) I1**
- **NYSERNet 350 Main St**
- **Commercial**
- **622 Mbps (OC-12)**
- **Abilene**
- **Medical/Dental**

**CCI**

University at Buffalo  The State University of New York  Center for Computational Research
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.
Shake-and-Bake Method: Dual-Space Refinement

Shake-and-Bake
“Top Algorithms of the 20th Century”

DeTitta, Hauptman, Miller, Weeks

Parameter Shift
FFT
FFT-1

Trial Structures
Structure Factors

Trial Phases

Phase Refinement

Density Modification (Peak Picking) (LDE)

Reciprocal Space “Shake”

Real Space “Bake”

Solutions

“Shake” and “Bake” Method:
Dual-Space Refinement

Trial Structures

Solutions
Phasing and Structure Size

Number of Atoms in Structure

- Conventional Direct Methods
- Se-Met
- Multiple Isomorphous Replacement
- Se-Met with Shake-and-Bake
- Shake-and-Bake
- Vancomycin
- 190kDa

0 100 1,000 10,000 100,000
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

Core Services
- Metacomputing Directory Service
- Globus Security Interface
- GRAM
- GASS

Applications
- Shake-and-Bake
- Apache
- MySQL
- Oracle

High-level Services and Tools
- Globus Toolkit
- MPI
- MPI-IO
- C, C++, Fortran, PHP
- globusrun

NWS

ACDC-Grid Data Resources

Local Services
- Condor
- Stork
- MPI
- LSF
- PBS
- Maui Scheduler
- RedHat Linux
- WINNT
- TCP
- UDP
- Irix
- Solaris

Adapted from Ian Foster and Carl Kesselman

University at Buffalo  The State University of New York  Center for Computational Research
ACDC-Grid Portal

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 consists 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"?

Advanced Center for Computational Research

Data Center
Data Grid Capabilities

Browser view of “miller” group files published by user “rappleye”
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.

The charts are based on:
- total CPU hours
- usage data for:
  - running jobs
  - all resources

Usage data for:
- running jobs
- queued jobs
- individual or all resources
- grouped by:
  - group
  - user
  - queue

Charts are updated every 15 minutes.
CondorView integrated into ACDC-Grid Portal
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
Risk Mitigation

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

- Ability of proteins to perform biological function is attributed to their 3-D structure.
- Protein folding problem refers to the challenge of predicting 3-D structure from amino-acid sequence.
- Solving the protein folding problem will impact drug design.
CCR 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
CCR Visualization Projects

- **Peace Bridge Visualization Project**
  - CCR, Niagara College of Canada, Bergmann, IBC, eMedia Inc, Parsons Engineering

- **StreetScenes®**
  - Virtual Reality (VR) software solution for 3D visualization of surface traffic

- **Buffalo Niagara Medical Campus**
  - CCR, IBC Digital Inc.

- **Emergency response / GIS – Earthquake**
  - CCR

- **Williamsville Toll Barrier**
  - CCR, TVGA

- **Accident Reconstruction**
  - CCR, TVGA

- **Biomedical Imaging**
  - CCR
Initial Photo Match incorporating real and computer-generated components
Key Receptor Sites
Multiple Viewpoints
Fully Interactive
Aerial Photography
Accident Reconstruction
The Accident
Accident Animation
(Driver’s View)
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
3D Medical Visualization App

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

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

- IBC Digital
  - Gov. Pataki Visit
  - Peace Bridge (Early & Current)
  - Buffalo-Niagara Medical Campus
  - Compute Cycles for Animation

- Bergmann Associates
  - Peace Bridge (Current)
  - NYS Thruway Toll Plaza

- Azar & More
  - Reenactment of 1901 Pan Am Exhibition
  - PHSCologram & Courses
  - Avid Digital Editing

- Niagara College
  - Start up
  - Peace Bridge (Current)

- Hauptman-Woodward Medical Research Institute
  - Computing
  - Collaboratory

- The Children’s Hospital of Buffalo
  - Medical Visualization

- Veridian
  - Battlespace Management
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

Grammy: Designs on Buffalo

UB brings bioinformatics to a younger generation

An early look at bioinformatics

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

Media Coverage
miller@buffalo.edu
www.ccr.buffalo.edu