Future of High Performance Computing

Thom H. Dunning, Jr.
National Center for Supercomputing Applications
Institute for Advanced Computing Applications and Technologies
Department of Chemistry

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Outline of Presentation

• Progress in High Performance Computing

• Directions in Computing Technology
  • Multi-core and many-core chips
  • Memory subsystem
  • Communications subsystem

• Era of Petascale Computing
  • Science @ Petascale
  • Petascale Computing Systems
  • Blue Waters Petascale Computing System

• Challenges of Petascale Computing
Progress in High Performance Computing
Progress in High Performance Computing

Relentless Increase in Performance

Top 500: #1
1 GF: late 1980s
1 TF: 1997
1 PF: 2008
Progress in High Performance Computing

Relentless Increase in Number of Cores

Number of Cores

\[10^2 \quad 10^3 \quad 10^4 \quad 10^5 \quad 10^6\]

\[ '92 \quad '94 \quad '96 \quad '98 \quad '00 \quad '02 \quad '04 \quad '06 \quad '08 \quad '10 \quad '12\]
Directions in Computing Technologies
“In the past, performance scaling in conventional single-core processors has been accomplished largely through increases in clock frequency (accounting for roughly 80 percent of the performance gains to date).”

Platform 2015
S. Y. Borkar et al., 2006
Intel Corporation
Directions in Computing Technologies

From Uni-core to Multi-core Processors

AMD
Uni-, Dual-, Quad-core, Processors

Intel
Multi-core Performance
Directions in Computing Technologies

Multi-core 2009: Intel’s Nehalem

Nehalem

• Modular
• Up to 8 cores
• 3 levels of cache
• Integrated memory controller
• Multiple QuickPath Interconnects
Directions in Computing Technologies

Switch to Multicore Chips

“For the next several years the only way to obtain significant increases in performance will be through increasing use of parallelism:

- 8× in 2009
- 16× in 2011
- 32× in 2013
- etc.
Directions in Computing Technologies

On to Many-core Chips

NVIDIA Tesla
(240 cores)

AMD Firestream
(800 cores)

IBM Cell
(1+8 cores)
Directions in Computing Technologies

New Technologies for HPC

Gflops

- NVIDIA (GPU)
- INTEL (CPU)

2002 2003 2004 2005 2006 2007

1.35 GHz G80

1.50 GHz G80

2.66 GHz Quad-core

3.4 GHz Dual-core

Courtesy of John Owens (UCSD) & Ian Buck (NVIDIA)
## NVIDIA: Selected Benchmarks

<table>
<thead>
<tr>
<th>Application</th>
<th>Description</th>
<th>Kernel X</th>
<th>App X</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.264</td>
<td>SPEC ’06 version, change in guess vector</td>
<td>20.2</td>
<td>1.5</td>
</tr>
<tr>
<td>LBM</td>
<td>SPEC ’06 version, change to single precision and print fewer reports</td>
<td>12.5</td>
<td>12.3</td>
</tr>
<tr>
<td>FEM</td>
<td>Finite element modeling, simulation of 3D graded materials</td>
<td>11.0</td>
<td>10.1</td>
</tr>
<tr>
<td>RPES</td>
<td>Rys polynomial equation solver, 2-electron repulsion integrals</td>
<td>210.0</td>
<td>79.4</td>
</tr>
<tr>
<td>PNS</td>
<td>Petri net simulation of a distributed system</td>
<td>24.0</td>
<td>23.7</td>
</tr>
<tr>
<td>LINPACK</td>
<td>Single-precision implementation of saxpy, used in Gaussian elimination routine</td>
<td>19.4</td>
<td>11.8</td>
</tr>
<tr>
<td>TRACF</td>
<td>Two Point Angular Correlation Function</td>
<td>60.2</td>
<td>21.6</td>
</tr>
<tr>
<td>FDTD</td>
<td>Finite-difference time domain analysis of 2D electromagnetic wave propagation</td>
<td>10.5</td>
<td>1.2</td>
</tr>
<tr>
<td>MRI-Q</td>
<td>Computing a matrix Q, a scanner’s configuration in MRI reconstruction</td>
<td>457.0</td>
<td>431.0</td>
</tr>
</tbody>
</table>

* For GeForce 8800 @ 346 GF (SP), W-m. Hwu et al., 2007
Directions in Computing Technologies

**Benchmarks: Direct SCF Calculations***

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Time/Iter (s)</th>
<th>Energy</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GPU</td>
<td>CPU**</td>
<td>GPU</td>
</tr>
<tr>
<td>Caffeine</td>
<td>0.16</td>
<td>4.1</td>
<td>-1605.91827</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>1.36</td>
<td>67.4</td>
<td>-3898.82189</td>
</tr>
<tr>
<td>Buckyball</td>
<td>7.32</td>
<td>279.4</td>
<td>-10709.0757</td>
</tr>
<tr>
<td>Taxol</td>
<td>4.91</td>
<td>269.2</td>
<td>-12560.6830</td>
</tr>
<tr>
<td>Valinomycin</td>
<td>8.44</td>
<td>691.2</td>
<td>-20351.9813</td>
</tr>
</tbody>
</table>

Differences due to use of 32-bit precision, will be eliminated in 64-bit version of INVIDIA chip

* GeForce 8800 @ 346 GF (SP), I. Ufimtsev and T. Martinez, *CiSE* 10, 26-34 (2008).
** Using GAMESS on AMD Opteron 175 CPU.
Directions in Computing Technologies

**NVIDIA: Tesla S1070**

- 4 Tesla T10s
- Frequency: 1.44 GHz
- 960 cores (240/T10)
- Performance
  - **SP:** 4.14 TF
  - **DP:** 0.34 TF
- 16 GB memory (4/T10)
- 408 GB/s memory bandwidth (104/T10)
- CUDA programming environment
Memory Subsystem

- Memory Wall
  - Limitation on computation speed caused by the growing disparity between processor speed and memory latency and bandwidth
  - From 1986 to 2000, processor speed increased at an annual rate of 55%, while memory speed improved by only 10% per year

- Memory Subsystem
  - Caches
    - Two to three levels: L1–L3
    - On chip (faster) and off chip (slower)
  - Main memory
    - DDR2: 3.2–8.5 GB/s
    - DDR3: 6.4–12.8 GB/s

- Issue
  - Memory latency and bandwidth limitations make it difficult to achieve major fraction of peak performance of chip
Communications Subsystem

- Communications Fabric
  - Infiniband
    - Standard for HPC systems
    - Used in TACC’s Ranger (Sun) system
  - SeaStar2+: Cray’s proprietary interconnect
  - IBM working on next generation (proprietary) interconnect

- Issue
  - Latency and bandwidth limitations make it difficult to scale science and engineering applications to large numbers of processors
Era of Petascale Computing
Science @ Petascale

Petascale computing will enable advances in a broad range of science and engineering disciplines:

- Molecular Science
- Weather & Climate Forecasting
- Astronomy
- Earth Science
- Health
Era of Petascale Computing

LANL Roadrunner Computer System

- **Computing resources**
  - 12,960 IBM PowerXCell 8i accelerators (116,640 cores)
  - 6,480 AMD dual-core Opterons (12,960 cores)
  - 1.46 PF peak
  - 1.1 Petaflop/s Linpack

- **Memory**
  - 52 TB (accelerators)
  - 104 TB total

- **Electrical power**
  - 3.9 MW (maximum)
  - ≥ 250 Megaflops/Watt

- **Floor space**
  - 296 racks, 6800 ft²
Era of Petascale Computing

ORNL Jaguar Computer System

- Computing resources
  - 37,544 AMD quad-core Opterons
  - 150,176 cores
  - 1.38 PF peak
  - 1.06 Petaflop/s Linpack

- Memory
  - 300 TB

- I/O Storage and Bandwidth
  - 10 PB
  - 240 GB/s

- Interconnect Bandwidth
  - 374 TB/s

- Floor space
  - 4400 ft²

Cray Jaguar (XT5) Petascale System
Era of Petascale Computing

NSF’s Strategy for High-end Computing

Track 1 System
- UIUC/NCSA (>1 PF sustained)

Track 2 Systems
- PSC (?)
- UT/ORNL (~1PF)
- TACC (500+TF)
- Track 2d

Track 3 Systems
- Leading University HPC Centers

Science and Engineering Capability (logarithmic scale)

FY’07 FY’08 FY’09 FY’10 FY’11
# Era of Petascale Computing

## NSF’s Track 2 Computing Systems

<table>
<thead>
<tr>
<th>System Attribute</th>
<th>TACC Ranger</th>
<th>UT-ORNL Kraken</th>
<th>PSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>Operational</td>
<td>In progress</td>
<td>In progress</td>
</tr>
<tr>
<td>Vendor</td>
<td>Sun</td>
<td>Cray</td>
<td>SGI</td>
</tr>
<tr>
<td>Processor</td>
<td>AMD</td>
<td>AMD</td>
<td>Intel</td>
</tr>
<tr>
<td>Peak Performance (TF)</td>
<td>579</td>
<td>~1,000</td>
<td></td>
</tr>
<tr>
<td>Number Cores/Chip</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Number Processor Cores</td>
<td>62,976</td>
<td>~100,000</td>
<td>~100,000</td>
</tr>
<tr>
<td>Amount Memory (TB)</td>
<td>123</td>
<td>~100</td>
<td>~100</td>
</tr>
<tr>
<td>Amount Disk Storage (PB)</td>
<td>1.73</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>External Bandwidth (Gbps)</td>
<td>10</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>
Blue Waters
Petascale Computing System
Blue Waters Petascale Computing System

Goals for Blue Waters

• **Maximize Core Performance**
  ...minimize number of cores needed for a given level of performance as well as lessen the impact of sections of code with limited scalability

• **Maximize Application Scalability**
  ...low latency, high-bandwidth communications fabric

• **Solve Memory-intensive Problems**
  ...large amount of memory
  ...low latency, high-bandwidth memory subsystem

• **Solve Data-intensive Problems**
  ...high-bandwidth I/O subsystem
  ...large quantity of on-line disk, large quantity of archival storage

• **Provide Reliable Operation**
  ...maximize system integration
  ...mainframe reliability, availability, serviceability (RAS) technologies
### Blue Waters Petascale Computing System

#### Blue Waters Computing System

<table>
<thead>
<tr>
<th>System Attribute</th>
<th>Abe</th>
<th>Blue Waters*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor</td>
<td>Dell</td>
<td>IBM</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel Xeon 5300</td>
<td>IBM Power7</td>
</tr>
<tr>
<td>Peak Performance (PF)</td>
<td>0.090</td>
<td></td>
</tr>
<tr>
<td>Sustained Performance (PF)</td>
<td>0.005</td>
<td>≥1</td>
</tr>
<tr>
<td>Number of Cores/Chip</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Number of Processor Cores</td>
<td>9,600</td>
<td>&gt;200,000</td>
</tr>
<tr>
<td>Amount of Memory (PB)</td>
<td>0.0144</td>
<td>&gt;0.8</td>
</tr>
<tr>
<td>Amount of Disk Storage (PB)</td>
<td>0.1</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Amount of Archival Storage (PB)</td>
<td>5</td>
<td>&gt;500</td>
</tr>
<tr>
<td>External Bandwidth (Gbps)</td>
<td>40</td>
<td>100–400</td>
</tr>
</tbody>
</table>

* Reference petascale computing system (no accelerators).
Blue Waters Petascale Computing System

Blue Waters Project

Blue Waters Project
Work Breakdown Structure

WBS 1.0
Blue Waters System
- WBS 1.1 IBM HW & SW
- WBS 1.2 Hardware (non-IBM)
- WBS 1.3 Software (non-IBM)
- WBS 1.4 Integ., Test & Trans

WBS 2.0
Applications/Research
- WBS 2.1 Petascale Apps
- WBS 2.2 Petascale Research

WBS 3.0
ETAO
- WBS 3.1 Undergrad. Ed.
- WBS 3.2 Graduate Ed.
- WBS 3.3 Training
- WBS 3.4 Broadening Part.

WBS 4.0
Industrial Engagement
- WBS 4.1 IPIPE
- WBS 4.2 ISV Forum
- WBS 4.3 Workforce Dev.

WBS 5.0
Facilities
- WBS 5.1 PCF
- WBS 5.2 Trans. From ACB

WBS 6.0
Project Management
- WBS 6.1 Tracking & Controls
- WBS 6.2 Finan., Contracts, HR
- WBS 6.3 Risk Management
- WBS 6.4 Config. Management
- WBS 6.5 Reporting & Reviews
- WBS 6.6 Communications
- WBS 6.7 Cybersecurity
- WBS 6.8 MIS Tools

Note: this is not an organizational chart
Goal: Facilitate the widespread and effective use of petascale computing to address frontier research questions in science, technology and engineering at research, educational and industrial organizations across the region and nation.

Charter Members

Argonne National Laboratory
Fermi National Accelerator Laboratory
Illinois Math and Science Academy
Illinois Wesleyan University
Indiana University*
Iowa State University
Illinois Mathematics and Science Academy
Krell Institute, Inc.
Los Alamos National Laboratory
Louisiana State University
Michigan State University*
Northwestern University*
Parkland Community College
Pennsylvania State University*
Purdue University*

The Ohio State University*
Shiloh Community Unit School District #1
Shodor Education Foundation, Inc.
SURΑ – 60 plus universities
University of Chicago*
University of Illinois at Chicago*
University of Illinois at Urbana-Champaign*
University of Iowa*
University of Michigan*
University of Minnesota*
University of North Carolina–Chapel Hill
University of Wisconsin–Madison*
Wayne City High School

* CIC universities
Blue Waters Petascale Computing System

Petascale Computing Facility

Partners
EYP MCF/
Gensler
IBM
Yahoo!

- Modern Data Center
  - 90,000+ ft² total
  - 20,000 ft² machine room

- Energy Efficiency
  - LEED certified (silver)
  - Efficient cooling system

www.ncsa.uiuc.edu/BlueWaters
Challenges in Petascale Computing
Challenges in Petascale Computing

Accuracy of Computational Models


Perturbation expansion not converging!
Challenges in Petascale Computing

Scalability of Algorithms

Challenges in Petascale Computing

More Challenges

- **Programming Models and Languages**
  
  *will MPI be adequate*
  
  - PGAS (partitioned global address space) programming model
  - Universal parallel C (UPC), Co-array Fortran (CAF)

- **New Computing Technologies**
  
  *new/revised algorithms will be needed*
  
  - Multicore and many-core chips
  - Heterogeneous multicore/many-core chips

- **Enhanced Reliability**
  
  *need to minimize impact of/ride through failure*
  
  - Systems level (e.g., virtualization)
  - Applications level
Questions?