1. Distribution and Toxicity Examples
2. Computational Infrastructures

February 28, 2006
Congo Red

- Red dye

- Not used for cotton or paper dyeing anymore because toxic

- Known to bind to beta amyloid oligomers and fibrils in vitro, but besides being toxic, can’t cross blood-brain barrier
“Vioxx is Here” No More

- Nonsteroidal anti-inflammatory drug for arthritis and pain
- Less gastrointestinal side effects than aspirin, ibuprofen, etc.
  - Reason: Only inhibits only COX-2 and not COX-1
- Widely marketed and prescribed
  - Sales of over $2 billion per year
- Voluntarily withdrawn in 2004 due to increased risk of heart attack and stroke
  - Allegations of scientific misconduct and much litigation ongoing. Estimated that Vioxx caused ~100 K heart attacks in 5 years.
  - Increased cardiac risk may be due to metabolites formed when compound becomes ionized
  - FDA advisory panel has recommended allowing resumption of sales saying benefit outweighs risk
Computational Infrastructures
Themes

- Moore’s Law
- Parallelism
- Clusters vs. supercomputers
- Special purpose vs. general purpose
Metrics

- Standard benchmarks are used to measure FLOPS (floating point operations per second)
  - Small differences can be misleading because people tune to benchmark, but fine for scale

- Power
- Cost per FLOP and power consumption per FLOP should also be considered

- Memory
- Networking
- Storage
- Flexibility (qualitative)
Parallelism

- Within processor
  - Cell processor

- Within computer
  - Multiprocessor, CPU with GPU

- Clusters/grids
  - Very widely adopted over past few years

- Global distributed computing
Processors

- **Standard desktop computer CPUs (Pentium 4, AMD Athlon, etc.)**
  - Few GFLOPS
  - Not increasing in speed as fast as before
  - SIMD helps (SSE, 3DNow, or AltiVec)

- **GPUs**
  - Couple hundred GFLOPS and speed growing faster than CPUs’
  - Small cache, stream programming
  - Floating point or less

- **Cell processors just out**
  - 300 GFLOPS (single precision)
  - One Power (PPE) plus 8 synergistic processing elements (SPEs)
Grid or Cluster Computing

- Very widely adopted over past few years
- Network together a number of computers
- Can operate on a single task through MPI but ideal for trivially parallelizable job

Sources of computers
- Dedicated computer’s cluster
- Grid of desktops (Novartis has all its employees’ computers on a grid)
- Utility computing (pay to use)
Global Distributed Computing

- Individuals around the world download software which runs calculations assigned by central servers

- Folding@Home
  - ~200,000 active clients for 200 TFLOPS
  - Achieved first unbiased M.D. simulation of a protein folding

- BOINC tries to make setting up (or running) a project easier
Folding@Home

~200,000 active clients worldwide

Client program screenshot

Earth’s city lights from space
Projects
(Outdated, from 2003)
Supercomputers

- Nice if you can get one
- Strength: Communication between processors
- IBM Blue Gene
  - Starting configuration priced over $1 M
  - BlueGene/L at Lawrence Livermore at top of Top 500 list right now
  - BlueGene/L - 280 TFLOPS
- Fujitsu BioServer
  - 1,920 low power (embedded) processors
  - Test system was used for Gromacs, CAChe, and MOPAC (typical computational chemistry programs)
**“Top 500” List**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Computer</th>
<th>Processors</th>
<th>Year</th>
<th>R_{max}</th>
<th>R_{peak}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DOE/NNSA/LLNL United States</td>
<td>BlueGene/L - eServer Blue Gene Solution IBM</td>
<td>131072</td>
<td>2005</td>
<td>280600</td>
<td>367000</td>
</tr>
<tr>
<td>2</td>
<td>IBM Thomas J. Watson Research Center United States</td>
<td>BGW - eServer Blue Gene Solution IBM</td>
<td>40960</td>
<td>2005</td>
<td>91290</td>
<td>114688</td>
</tr>
<tr>
<td>3</td>
<td>DOE/NNSA/LLNL United States</td>
<td>ASC Purple - eServer pSeries p5 575 1.9 GHz IBM</td>
<td>10240</td>
<td>2005</td>
<td>63390</td>
<td>77824</td>
</tr>
<tr>
<td>4</td>
<td>NASA/Ames Research Center/NAS United States</td>
<td>Columbia - SGI Altix 1.5 GHz, Voltaire Infiniband SGI</td>
<td>10160</td>
<td>2004</td>
<td>51870</td>
<td>60960</td>
</tr>
<tr>
<td>5</td>
<td>Sandia National Laboratories United States</td>
<td>Thunderbird - PowerEdge 1850, 3.6 GHz, Infiniband Dell</td>
<td>8000</td>
<td>2005</td>
<td>38270</td>
<td>64512</td>
</tr>
<tr>
<td>6</td>
<td>Sandia National Laboratories United States</td>
<td>Red Storm Cray XT3, 2.0 GHz Cray Inc.</td>
<td>10880</td>
<td>2005</td>
<td>36190</td>
<td>43520</td>
</tr>
<tr>
<td>7</td>
<td>The Earth Simulator Center Japan</td>
<td>Earth-Simulator NEC</td>
<td>5120</td>
<td>2002</td>
<td>35860</td>
<td>40960</td>
</tr>
<tr>
<td>8</td>
<td>Barcelona Supercomputer Center Spain</td>
<td>MareNostrum - JS20 Cluster, PPC 970, 2.2 GHz, Myrinet IBM</td>
<td>4800</td>
<td>2005</td>
<td>27910</td>
<td>42144</td>
</tr>
<tr>
<td>9</td>
<td>ASTRON/University Groningen Netherlands</td>
<td>Stella - eServer Blue Gene Solution IBM</td>
<td>12288</td>
<td>2005</td>
<td>27450</td>
<td>34406.4</td>
</tr>
<tr>
<td>10</td>
<td>Oak Ridge National Laboratory United States</td>
<td>Jaguar - Cray XT3, 2.4 GHz Cray Inc.</td>
<td>5200</td>
<td>2005</td>
<td>20527</td>
<td>24960</td>
</tr>
<tr>
<td>11</td>
<td>Lawrence Livermore National Laboratory United States</td>
<td>Thunder - Intel Itanium2 Tiger4 1.4GHz - Quadrics California Digital Corporation</td>
<td>4096</td>
<td>2004</td>
<td>19940</td>
<td>22938</td>
</tr>
<tr>
<td>12</td>
<td>Computational Biology Research Center, AIST Japan</td>
<td>Blue Protein - eServer Blue Gene Solution IBM</td>
<td>8192</td>
<td>2005</td>
<td>18200</td>
<td>22937.6</td>
</tr>
</tbody>
</table>

Nov. 2005 list, from www.top500.org
Cluster or Supercomputer

- Supercomputer needed for tightly coupled computation
  - But often this is not needed or can be circumvented
- Clusters are more flexible and can be upgraded more easily
- CPUs in clusters can be quite good today—supercomputers without advantage in processor speed
Special or General Purpose

- MDGRAPE an example of a specialized system for a problem domain
  - Also ASICS and certain supercomputers
- Specialized processors/computers may have immediate performance advantages
- General purpose gives more flexibility
- General will usually advance faster because many constituents for its development
## Cost and Power

<table>
<thead>
<tr>
<th></th>
<th>$/GFLOP</th>
<th>W/ GFLOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDGRAPE-3</td>
<td>15</td>
<td>0.2</td>
</tr>
<tr>
<td>Pentium 4</td>
<td>400</td>
<td>14</td>
</tr>
<tr>
<td>BlueGene/L</td>
<td>140</td>
<td>6</td>
</tr>
<tr>
<td>Earth Simulator</td>
<td>8000</td>
<td>128</td>
</tr>
</tbody>
</table>

Taiji 2004
Readings

- Building and managing production bioclusters (Dagdigian)

- Protein Explorer: A Petaflops Special-Purpose Computer System for Molecular Dynamics Simulations (Taiji, et. al.)

- Introduction to the Cell multiprocessor (Kahle, et. al.)