Computation Across the Curriculum

Charlie Peck, Earlham College
First Annual Tapia LittleFe Scientific Programming Competition

Brought to you by a cast of thousands...
Why Computer Science Matters

Charlie Peck, Mobeen Ludin, Ivan Babic, Sanele Mahlalela
MEAP @ IUPUI
July, 2012
Well, how did we get here?

- Computational power
- Storage volume
- Algorithmic and computational approaches
Digital World

- Simulations based on models
- Analysis of large data sets
- Visualization of large data sets
- Techniques that depend on high performance computing gear to be practical; automating vs enabling
- Third and fourth methods of scientific intellectual inquiry
- 20th vs 21st century science inquiry; atoms and icebergs the occupants of El Sidrón
Computational Methods

• Why use simulations?
  too small (atoms, molecules)
  too large (galaxies, the universe)
  too fast (photosynthesis, protein folding)
  too slow (geological processes, climate change)
  too complex (blood circulation, weather)
  too dangerous (toxic materials, nuclear stockpile stability)
# High Performance Computing Gear

<table>
<thead>
<tr>
<th></th>
<th>Bazaar</th>
<th>Cairo</th>
<th>BobSCEd</th>
<th>Al-salam</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year</strong></td>
<td>2000</td>
<td>2003</td>
<td>2006</td>
<td>2010</td>
</tr>
<tr>
<td><strong>GFLOPS</strong></td>
<td>18</td>
<td>128</td>
<td>666</td>
<td>~3000</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>32U</td>
<td>16U</td>
<td>8U</td>
<td>13U</td>
</tr>
</tbody>
</table>

Moore’s law, every 18 months the density of transistors in integrated circuits roughly doubles
## Moore's Law in Action

<table>
<thead>
<tr>
<th>Year</th>
<th>RAM Capacity</th>
<th>Storage Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>32 Bytes</td>
<td>$10^0$</td>
</tr>
<tr>
<td>1975</td>
<td>640 KB</td>
<td>$10^3$</td>
</tr>
<tr>
<td>2000</td>
<td>256 MB</td>
<td>$10^6$</td>
</tr>
<tr>
<td>2008</td>
<td>2 GB</td>
<td>$10^9$</td>
</tr>
<tr>
<td>2030</td>
<td>1 TB</td>
<td>$10^{12}$</td>
</tr>
</tbody>
</table>

Typical amount of RAM in a “desktop” computer
Data > Information > Knowledge

• Complete works of Shakespeare - ~ 5 MB
• Human genome - ~ 1 GB
• Complete works of Beethoven - ~ 20 GB
• Medical imaging - ~ 30 GB per scan
• Library of Congress - ~ 10 TB
• All US academic libraries - ~ 2 PB
• Large Hadron Collider - ~1.5 GB/second

Increasingly Readings > Data ...
Parallel and Distributed Computing

• Decomposing large problems into smaller ones, solving the smaller problems, and then reducing those answers to find “the answer”

• Domain decomposition

• Functional decomposition

• Shared memory systems

• Message passing systems
Natural Sciences

- **Modeling and Simulation**
  - Protein folding
  - Earthquakes
  - Phylogenetic reconstruction
  - Genome construction

- **Data Sets**
  - Sloan Digital Sky Survey
  - Protein Data Bank
  - Arctic aerial photographs
  - Geographical information systems (GIS)
Protein Folding

Result of an ensemble molecular dynamics simulation (Gromacs) of the villin headpiece
Earthquakes

Combined view with Camera Match

Source: San Diego Supercomputer Center
Climate Change

Source: National Center for Atmospheric Research (NCAR) and the University of Colorado’s National Snow and Ice Data Center (NSIDC)
Humanities

• Modeling and Simulation
  Game theory
  Topic modeling
  Text analysis

• Data Sets
  Library of Congress
  Project Gutenberg
  Newspaper morgues
Rome Reborn

Source: University of California Irvine

Looking out from the Roman Forum in a complete interactive 3D model of Rome
Arts

• Modeling and Simulation
  Animation and rendering
  Painting provenance
  Digital music

• Data Sets
  Photograph archives
  Scanned paintings, sculptures, buildings
  Digital recordings
It’s Everywhere

• Anthropology *
• Archeology
• Classics
• Economics
• Film
• History
• Language Acquisition
• Library Science
• Literature
• Painting *
• Politics
• Psychology
• Sculpture
Is that really a Van Gogh?

Source: Christian Science Monitor

Vase with 15 Sunflowers
Social Sciences

• **Modeling and Simulation**
  Teacher matching
  Social systems
  Derivatives analysis

• **Data Sets**
  Census
  Geographical information systems (GIS)
  Voting records
  Transaction records (commercial and civil)
What happened to the Anasazi?
Who were those people?

- Forensic analysis
- Archeology
- Genome sequencing
- Anthropology
Technology in the Wild
Science in the Field
Education, Outreach, and Training

- National Computational Science Institute/SC Education Program/Blue Waters UPEP
- Workshops for undergraduate faculty and students
- 1800 participants at 80 workshops over the past 7 years
- Specialize in first-touch to handoff to a regional provider
- 3D Internet - metaverse and web-based interactive applications
- Curriculum development
- Outreach to high schools
Why Do All This?

• President’s Information Technology Advisory Committee: “...computational science is one of the most important technical fields of the 21st century...”

• Rising Above The Gathering Storm: “...vastly improving K-12 and undergraduate science and mathematics education...”

• Bio2010: “...exposure during the early years of their undergraduate careers will help life science students use current computer methods and learn how to exploit emerging computer technologies as they arise...”
Future

- Software
  Making sense of all that stuff we are collecting
  [Readings] > Data > Information > Knowledge
  Visualization, interactive interfaces

- Grid
  Science portals, e.g. TeraGrid, Nanohub, Open Science Grid
  Humanities, Arts, and Social Science portals

- Hardware
  Specialized CPUs, e.g. FPGA and graphics chips
  Cores, cores, and more cores