The HPC Workforce

Barbara Chapman
Stony Brook University
University of Houston
Brookhaven National Lab

Rice Oil and Gas Conference
March 3, 2016
Highly Trained HPC Workforce

• Variety of skills and expertise needed to
  – Develop, deploy and maintain large-scale facilities
  – Create, utilize and enhance large-scale applications
  – Gather, store and analyze large-scale data sets

• Importance of HPC growing in science and technology, as well as in commerce
National Strategic Computing Initiative

• Whole-government initiative to create systems that apply exaflops of compute power to exabytes of data
  – Keep US at forefront of HPC capabilities
  – Improve HPC application developer productivity.
  – Make HPC readily available
  – Establish hardware technology for future HPC systems

• Increase capacity and capability of national HPC ecosystem; path beyond “Moore’s Law”

• Provide workforce training

Success of initiative depends critically on HPC workforce
DOE’s Office of Science

Delivers scientific discoveries and tools to transform our understanding of nature and advance the energy, economic, and national security of the U.S.

- 47% of the U.S. Federal support of basic research in the physical sciences;
- ~22,000 Ph.D. scientists, grad students, engineers, and support staff at >300 institutions, including all 17 DOE labs;
- U.S. and world leadership in high-performance computing and computational sciences;
  » Major contributions to HPC applications, libraries, system software, expertise in all aspects of use of high-end computers

Reliance on DOE to fund/drive/develop much of HPC technology
Need for Skilled Computing Workforce in DOE Laboratories

• Scientific computation underlies much of Office of Science R&D
  – ASCR facilities, experts in their utilization are essential to all areas of scientific activity in the DOE national laboratories.
• Maintaining a sufficient workforce is this area is critical
  – To enable research outcomes
  – Amortize significant investment in facilities
• Breadth of expertise required to effectively deploy tools is increasing.
Workforce Problems at DOE

- Anecdotal evidence of hiring and retention problems
  - DOE Fellowship program discontinued
- Committee set up to evaluate situation
  - Subcommittee of ASCAC (Advanced Scientific Computing Advisory Committee)
  - Asked to assess workforce needs: demand vs. availability
  - Suitability of existing graduate/postdoc-level workforce training to meet DOE Office of Science mission needs?
- Committee decided to take a broader perspective
  - Also considered bigger picture; problems of recruitment and retention in labs

Report can be downloaded from ASCAC website (Reports tab)
Recruitment/Retention Challenges

Labs invited to provide information on areas where they experience recruitment / retention difficulties

• Computing / Computational Sciences figured prominently in responses, especially in large labs
  – Low number of qualified applicants for open positions in these areas, many of which are foreign nationals

• Broad range of needed expertise was identified:
  – Algorithms; Applied Mathematics; Data Analysis, Management and Visualization; Cybersecurity; Software Engineering and High Performance Software Environments; and High Performance Computer Systems.
  – Termed “Computing Sciences” in following

• Strong demand for M.Sc, Ph.D.-level Computing Sciences positions
  – Especially at LBNL and ORNL where open positions are ca. 25% of total staff in Computing Sciences

Labs spend significant effort on recruitment in this area
<table>
<thead>
<tr>
<th>Competencies</th>
<th>National Laboratories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced computing architectures</td>
<td>LLNL, ORNL</td>
</tr>
<tr>
<td>Applied mathematics (including advanced modeling and methods)</td>
<td>ANL, LBNL, LLNL, ORNL</td>
</tr>
<tr>
<td>Computational sciences/simulation; scientific software</td>
<td>ANL, BNL, INL, LBNL, LLNL, ORNL, SNL</td>
</tr>
<tr>
<td>Cyber security</td>
<td>INL, LLNL, ORNL, SNL</td>
</tr>
<tr>
<td>Data acquisition software</td>
<td>FNAL, ORNL</td>
</tr>
<tr>
<td>Data informatics (data mining, machine learning, big data, statistical techniques)</td>
<td>ANL, LBNL, LLNL, ORNL, PNNL</td>
</tr>
<tr>
<td>Dynamic mesh algorithms</td>
<td>LLNL</td>
</tr>
<tr>
<td>HPC /extreme-scale/exascale computing</td>
<td>ANL, INL, LANL, LBNL, LLNL, ORNL, PNNL, SNL</td>
</tr>
<tr>
<td>Performance analysis of HPC applications</td>
<td>LBNL, ORNL</td>
</tr>
<tr>
<td>Software quality assurance</td>
<td>LLNL, ORNL</td>
</tr>
<tr>
<td>Solvers</td>
<td>LBNL, LLNL</td>
</tr>
<tr>
<td>Storage systems</td>
<td>LBNL, ORNL</td>
</tr>
<tr>
<td>Uncertainty quantification</td>
<td>LLNL, ORNL</td>
</tr>
<tr>
<td>Visualization and scientific data analysis</td>
<td>LLNL, ORNL</td>
</tr>
</tbody>
</table>
# Recruitment/Retention at Labs

<table>
<thead>
<tr>
<th>Lab</th>
<th># Open Posns</th>
<th>Ave. Time To Fill (Days)</th>
<th>Total Technical Staff</th>
<th>% Foreign Nationals</th>
<th>Declined Job Offers</th>
<th>Attrition Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LANL</td>
<td>148*</td>
<td>263*</td>
<td>1903*</td>
<td>5.4*</td>
<td>21/173*</td>
<td>4.9*</td>
</tr>
<tr>
<td>LBNL1</td>
<td>56</td>
<td>112</td>
<td>206</td>
<td>38.4</td>
<td>2/39</td>
<td>8.0</td>
</tr>
<tr>
<td>LLNL2</td>
<td>146</td>
<td>311</td>
<td>2094*</td>
<td>7.4*</td>
<td>7/36</td>
<td>4.8*</td>
</tr>
<tr>
<td>ORNL3</td>
<td>87</td>
<td>110</td>
<td>379</td>
<td>38</td>
<td>11/73</td>
<td>7.6</td>
</tr>
<tr>
<td>PNNL4</td>
<td>44</td>
<td>107</td>
<td>1113*</td>
<td>16*</td>
<td>16/50</td>
<td>8.9**</td>
</tr>
</tbody>
</table>

* Data for all scientific and engineering disciplines, M.S. and Ph.D. level
** Data for all scientific and engineering disciplines, all degree levels
1 LBNL data for “all scientists and engineers on the Computer Science curve”
2 LLNL data based on best attempt to identify positions in the Computing Sciences; time-to-fill may be skewed by indefinite postings; attrition rate corrected for voluntary separation program
3 ORNL data for “lab-wide computing/computational science” positions; attrition rate corrected to account for voluntary separation program (37% of terminations)
4 PNNL attrition rate is uncorrected for voluntary separation program; historical rate is 4-5%; total number of job offers is estimated.
Workforce Recruitment/Retention Challenges

• Labs take much longer than industry to fill positions in Computing Sciences
  – 4 times as long when security clearance does not permit foreigner
  – On average over 100 days to fill an Office of Science position
  – More than 200 days for NNSA lab positions
  – Industry needs 39 days on average for Computer and Mathematical occupations (all degrees)
  – 48-50 days on average for M.Sc., Ph.D. positions in STEM

• Acceptance and retention rates mostly favorable
• Attrition rates compare favorably with industry
  – 10% in industry, 5% NNSA labs, ca. 8% open science labs
  – However loss of expertise can be catastrophic
### Number of PhDs in CS and CE

<table>
<thead>
<tr>
<th>PhD Specialty</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Intelligence</td>
<td>181</td>
<td>193</td>
<td>203</td>
<td>171</td>
<td>748</td>
</tr>
<tr>
<td>Databases/Info Retrieval</td>
<td>99</td>
<td>106</td>
<td>122</td>
<td>125</td>
<td>452</td>
</tr>
<tr>
<td>Graphics/Vis</td>
<td>87</td>
<td>111</td>
<td>99</td>
<td>99</td>
<td>396</td>
</tr>
<tr>
<td>HW/Architecture</td>
<td>78</td>
<td>70</td>
<td>92</td>
<td>91</td>
<td>329</td>
</tr>
<tr>
<td>High Performance Computing</td>
<td>29</td>
<td>37</td>
<td>49</td>
<td>60</td>
<td>175</td>
</tr>
<tr>
<td>Networks</td>
<td>150</td>
<td>140</td>
<td>147</td>
<td>152</td>
<td>589</td>
</tr>
<tr>
<td>Operating Systems</td>
<td>59</td>
<td>55</td>
<td>66</td>
<td>55</td>
<td>235</td>
</tr>
<tr>
<td>Scientific/ Numerical Computing</td>
<td>33</td>
<td>27</td>
<td>32</td>
<td>29</td>
<td>121</td>
</tr>
<tr>
<td>Software Engineering</td>
<td>126</td>
<td>147</td>
<td>149</td>
<td>140</td>
<td>562</td>
</tr>
</tbody>
</table>

Taulbee Survey 2014
Multidisciplinary Education

Computational Science disciplines taught in:
• Computer Science
• Computer Engineering
• Information Science

Also in interdisciplinary studies:
• Computational Science and Engineering
• 2005 PITAC report described difficulties establishing these

From PITAC Report, 2005
Multidisciplinary Education

• Interdisciplinary Computational Science and Engineering (CS&E) studies in some institutions
  – Domain sciences, applied mathematics, numerical analysis, computer science
  – Problem-solving methodologies, science and engineering tools
  – Degree program, an area of specialization or a certificate
  – Overall number of graduates remains low

• Interdisciplinary Data Science education emerging, not likely to satisfy demand

• NSF taskforce (2011): Universities not teaching essential skills for applying CS&E in the field, not preparing students to harness powerful new supercomputing infrastructure

• Degree programs complemented by variety of training courses via NSF’s XSEDE, university centers, labs
UH: HPC User Training Road Map

- Linux
- Basic Programming
- General HPC concepts
- Visualization
- Libraries
- Profiling & Tuning
- Debugging

- Parallel Programming
- Parallel Programming Interfaces
- Parallel Visualization
- Parallel Libraries
- Parallel Profiling & Tuning
- Parallel Debugging

Contacts: Jerry Ebalunode, jebaluno@central.uh.edu
Martin Huarte-Espinosa, mhuartee@central.uh.edu
Lab Training and Outreach

• Labs all put significant effort into attracting, training and retaining workers in Computing Sciences
• Engage graduate students and postdocs through programs such as summer internships, university subcontracts
  – Exposure to lab environment is key
  – Postdoctoral fellowships such as ANL’s Wilkinson and ORNL’s Householder
  – Programs where student is co-advised by lab staff
• Many staff recruits come from these programs
DOE’s Computational Science Graduate Fellowship Program

- DOE’s Computational Science Graduate Fellowships (CSGF) train graduate students to meet national workforce needs in computational sciences
- Program established 1991; terminated 2011; now reinstated
- Effective elements
  - Interdisciplinary program of study
  - Research practicum at DOE laboratories
  - Annual review that enables networking
  - Careful selection process
Big Picture

• Huge demand for graduates with computing expertise
  – Taulbee survey reports very low unemployment in US for computing graduates (under 1%)
  – Large majority of graduates enter industry (70%)
  – Today, most remain in US (8.2% take jobs outside US)
• Computing Sciences disciplines in very high demand
• US Bureau of Labor statistics project very high job growth in computing related jobs
• Bill Gates: adding 100Ks of jobs p.a. but only 15K graduates
• Retirement of current workforce is expected to grow workforce gap over coming decade

Huge growth in computing-related jobs in entire economy. The workforce gap will grow.
What About HPC Positions?

• 70% of HPC graduates enter industry
  – Just 7% of graduates in Scientific and Numerical Computing took Government jobs

• Computing Science skills not only useful for extreme-scale computing
  – Strong demand from Silicon Valley and financial industry
  – Big Data skills in high demand in industry

• DOE Labs, HPC centers cannot compete with industry compensation
  – Awareness of lab careers among graduates low
  – Regulations decrease attractiveness of jobs
  – Lack of well-defined career path in academic setting

Council on Competitiveness: HPC is a proven game-changing technology
An Incomplete Talent Pool

Lack of diversity in US graduates in CS and CE is a major contributing factor in national shortage:

- US citizens among graduates are mostly white male
- African American / Hispanic graduates very low (ca. 1% each)
- Many females among CS graduates are Asian

### Participation Rate in Natural Sciences and Engineering Bachelor’s Degrees in 1998

<table>
<thead>
<tr>
<th>Race/ethnicity and sex</th>
<th>Total 24-year-old Population</th>
<th>Total bachelor’s degrees</th>
<th>Total NS&amp;E degrees</th>
<th>Bachelor’s degrees per 100 24-year-olds</th>
<th>NS&amp;E % of bachelor’s</th>
<th>NS&amp;E degrees per 100 24-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>3,403,039</td>
<td>1,199,579</td>
<td>205,355</td>
<td>35.3</td>
<td>17.1</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,714,571</td>
<td>525,714</td>
<td>128,481</td>
<td>30.7</td>
<td>24.4</td>
<td>7.5</td>
</tr>
<tr>
<td>Female</td>
<td>1,688,468</td>
<td>673,865</td>
<td>76,874</td>
<td>39.9</td>
<td>11.4</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2,251,292</td>
<td>878,018</td>
<td>142,500</td>
<td>39.0</td>
<td>16.2</td>
<td>6.3</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>149,413</td>
<td>69,988</td>
<td>22,003</td>
<td>46.8</td>
<td>31.4</td>
<td>14.7</td>
</tr>
<tr>
<td>Underrepresented minority</td>
<td>1,002,334</td>
<td>181,709</td>
<td>25,820</td>
<td>18.1</td>
<td>14.2</td>
<td>2.6</td>
</tr>
<tr>
<td>Black</td>
<td>473,402</td>
<td>95,878</td>
<td>12,731</td>
<td>20.3</td>
<td>13.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>497,620</td>
<td>78,125</td>
<td>12,006</td>
<td>15.7</td>
<td>15.4</td>
<td>2.4</td>
</tr>
<tr>
<td>American Indian/Alaskan Native</td>
<td>31,312</td>
<td>7,706</td>
<td>1,083</td>
<td>24.6</td>
<td>14.1</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Demographic Trends in Industry

Apple

- White: 54
- Asian: 23
- Hispanic: 7
- Undeclared: 6
- Mixed Race: 8
- Black: 2

Twitter

- White: 58
- Asian: 34
- Other: 2
- Mixed Race: 3
- Hispanic: 2
- Black: 1
Demographic Trends in Industry

Google

Facebook

- White
- Asian
- Other
- Mixed Race
- Hispanic
- Black

- White
- Asian
- Other
- Mixed Race
- Hispanic
- Black
US Citizens / Permanent Residents as Percentage of PhDs in CS, CE Areas

<table>
<thead>
<tr>
<th>PhD Specialty</th>
<th>Citizens, Permanent Residents</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Intelligence</td>
<td>439</td>
<td>58.7%</td>
</tr>
<tr>
<td>Databases/Info Retrieval</td>
<td>203</td>
<td>44.9%</td>
</tr>
<tr>
<td>Graphics/Vis</td>
<td>228</td>
<td>57.6%</td>
</tr>
<tr>
<td>HW/Architecture</td>
<td>147</td>
<td>44.7%</td>
</tr>
<tr>
<td>High Performance Computing</td>
<td>78</td>
<td>44.6%</td>
</tr>
<tr>
<td>Networks</td>
<td>205</td>
<td>34.8%</td>
</tr>
<tr>
<td>Operating Systems</td>
<td>108</td>
<td>46.0%</td>
</tr>
<tr>
<td>Scientific/Numerical Computing</td>
<td>78</td>
<td>64.5%</td>
</tr>
<tr>
<td>Software Engineering</td>
<td>328</td>
<td>58.4%</td>
</tr>
</tbody>
</table>

Figures accumulated over past 4 years; Taulbee Report, 2014
## LBNL Demographics in STEM

### Types of Jobs at Berkeley Labs

<table>
<thead>
<tr>
<th>Types of Jobs at Berkeley Labs</th>
<th>TTL</th>
<th>Women</th>
<th>%</th>
<th>URM</th>
<th>%</th>
<th>OPC</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientists and Engineers</td>
<td>640</td>
<td>100</td>
<td>15.6%</td>
<td>29</td>
<td>4.5%</td>
<td>131</td>
<td>20.5%</td>
</tr>
<tr>
<td>(Conducting Research)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postdoctoral Scientists</td>
<td>486</td>
<td>133</td>
<td>27.4%</td>
<td>26</td>
<td>5.3%</td>
<td>209</td>
<td>43.0%</td>
</tr>
<tr>
<td>Engineers</td>
<td>483</td>
<td>102</td>
<td>21.1%</td>
<td>51</td>
<td>10.6%</td>
<td>118</td>
<td>24.4%</td>
</tr>
<tr>
<td>(Information, Mechanical, and Electrical)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Support</td>
<td>907</td>
<td>390</td>
<td>43.0%</td>
<td>145</td>
<td>16.0%</td>
<td>207</td>
<td>22.8%</td>
</tr>
<tr>
<td>(Non S&amp;Es in programmatic divisions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ops Support</td>
<td>677</td>
<td>324</td>
<td>47.9%</td>
<td>161</td>
<td>23.8%</td>
<td>117</td>
<td>17.3%</td>
</tr>
<tr>
<td>(Non S&amp;Es in Operational Divisions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>3193</td>
<td>1049</td>
<td>32.9%</td>
<td>412</td>
<td>12.9%</td>
<td>782</td>
<td>24.5%</td>
</tr>
</tbody>
</table>

Date of Data: October 1, 2013
Career, term, and postdoctoral employees only
URM=Underrepresented Minorities (African American/Black, Hispanic/Latino, and American Indian/Ala
OPC=Other People of Color (Asian/Asian American, Middle Eastern/Southwest Asian/North African anc

Representative of demographics in labs
Data indicate retention problem for female postdocs
Workforce Demographics

- % of foreign nationals in graduate population growing steadily
  - 60% of graduates in Computing Sciences are now foreign nationals
- Lack of diversity in US graduates in CS and CE is a major contributing factor in national shortage
  - Current US citizens among graduates are mostly white male
  - African American / Hispanic graduates very low (ca. 1% each)
  - Percentage of females among graduates is declining
- Data from DOE labs reflect national demographics
  - Also indicate retention problem for female postdocs
- Lack of STEM diversity acknowledged, not effectively addressed

The growing demand for graduates in Computing Sciences far exceeds the supply. A continued underrepresentation of minorities and females is expected.
Role of Labs: Workforce Retention

• Labs face many challenges in maintaining workforce
• Need to re-examine career paths to offer competitive choices, provide a more attractive workplace
  – Current funding model makes it hard for young staff to establish themselves
  – Consider how to give employees opportunities to grow professionally e.g. through new opportunities
  – Provide resources to address work/life balance
  – Adapt to shorter-term commitment that is becoming common
  – Facilitate mid-career entry to labs
  – Engage in education
Status of HPC Workforce

• HPC growing in national significance
  – Strong job growth expected
  – Demand for Computing Science expertise far exceeds supply of trained graduates

• Multidisciplinary training available: degrees, certificate classes

• More need for practical, “real-world” training

• Need to do more to broaden appeal of HPC if workforce is to grow

• In government, this includes addressing retention problems
Questions? Comments?

SCENIC VIEW OF NEXT GOAL
BACK-UP SLIDES
Summary of Findings

• The multidisciplinary national labs face workforce recruiting and retention challenges in Computing Sciences
• Insufficient educational opportunities are available at academic institutions in areas of the Computing Sciences most relevant to the DOE mission.
• There is a growing demand for graduates in Computing Sciences that far exceeds the supply. A larger workforce gap and continued underrepresentation of minorities and females are expected.
• The exemplary DOE CSGF program, deemed highly effective in multiple reviews, is uniquely structured and positioned to provide the future workforce with the interdisciplinary knowledge, motivation, and experiences necessary for contributing to the DOE mission.
• The DOE laboratories have individually developed measures to help recruitment and retention, yet more can be done at the national level to amplify and extend the effectiveness of local programs.
Summary of Recommendations

• Leverage and expand the successful DOE CSGF program scope to include research in HPC-enabling sciences such as computational mathematics, computer science, and data analytics

• Establish graduate level curricular competencies specifically to fulfill DOE’s Computing Sciences workforce needs.

• Develop a recruiting and retention program that increases DOE’s visibility on college campuses and that increases attractiveness of lab positions

• Build a diverse workforce that spans demographics and universal accessibility for a broader awareness of career opportunities within DOE.

• Expand support for local laboratory programs and encourage greater inter-laboratory sharing of information about locally successful programs as well as workforce related data.