On the Future of High Performance Computing: How to Think for Peta and Exascale Computing

Jack Dongarra

University of Tennessee
Oak Ridge National Laboratory
University of Manchester
Overview

• Take a look at high performance computing
• What’s driving HPC
• Future Trends
Top500 List of Supercomputers

H. Meuer, H. Simon, E. Strohmaier, & JD

- Listing of the 500 most powerful Computers in the World
- Yardstick: Rmax from LINPACK MPP
  \[ Ax = b, \text{ dense problem} \]

- Updated twice a year
  SC'xy in the States in November
  Meeting in Germany in June

- All data available from www.top500.org
Example of typical parallel machine
Example of typical parallel machine
Example of typical parallel machine

Shared memory programming between processes on a board and a combination of shared memory and distributed memory programming between nodes and cabinets.
Example of typical parallel machine

Combination of shared memory and distributed memory programming
# November 2011: The TOP10

<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Computer</th>
<th>Country</th>
<th>Cores</th>
<th>Rmax [Pflops]</th>
<th>% of Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RIKEN Advanced Inst for Comp Sci</td>
<td>K computer Fujitsu SPARC64 VIIIfx + custom</td>
<td>Japan</td>
<td>705,024</td>
<td>10.5</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>Nat. SuperComputer Center in Tianjin</td>
<td>Tianhe-1A, NUDT Intel + Nvidia GPU + custom</td>
<td>China</td>
<td>186,368</td>
<td>2.57</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>DOE / OS Oak Ridge Nat Lab</td>
<td>Jaguar, Cray AMD + custom</td>
<td>USA</td>
<td>224,162</td>
<td>1.76</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>Nat. Supercomputer Center in Shenzhen</td>
<td>Nebulea, Dawning Intel + Nvidia GPU + IB</td>
<td>China</td>
<td>120,640</td>
<td>1.27</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>GSIC Center, Tokyo Institute of Technology</td>
<td>Tusbame 2.0, HP Intel + Nvidia GPU + IB</td>
<td>Japan</td>
<td>73,278</td>
<td>1.19</td>
<td>52</td>
</tr>
<tr>
<td>6</td>
<td>DOE / NNSA LANL &amp; SNL</td>
<td>Cielo, Cray AMD + custom</td>
<td>USA</td>
<td>142,272</td>
<td>1.11</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>NASA Ames Research Center/NAS</td>
<td>Plelades SGI Altix ICE 8200EX/8400EX + IB</td>
<td>USA</td>
<td>111,104</td>
<td>1.09</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>DOE / OS Lawrence Berkeley Nat Lab</td>
<td>Hopper, Cray AMD + custom</td>
<td>USA</td>
<td>153,408</td>
<td>1.054</td>
<td>82</td>
</tr>
<tr>
<td>9</td>
<td>Commissariat a l’Energie Atomique (CEA)</td>
<td>Tera-10, Bull Intel + IB</td>
<td>France</td>
<td>138,368</td>
<td>1.050</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>DOE / NNSA Los Alamos Nat Lab</td>
<td>Roadrunner, IBM AMD + Cell GPU + IB</td>
<td>USA</td>
<td>122,400</td>
<td>1.04</td>
<td>76</td>
</tr>
</tbody>
</table>

226 RPI IBM eServer Blue Gene Solution 32,768 .073 80
<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Computer</th>
<th>Country</th>
<th>Cores</th>
<th>Rmax [Pflops]</th>
<th>% of Peak</th>
<th>Power [MW]</th>
<th>MFlops/Watt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RIKEN Advanced Inst for Comp Sci</td>
<td>K computer Fujitsu SPARC64 VIIIfx + custom</td>
<td>Japan</td>
<td>705,024</td>
<td>10.5</td>
<td>93</td>
<td>12.7</td>
<td>830</td>
</tr>
<tr>
<td>2</td>
<td>Nat. SuperComputer Center in Tianjin</td>
<td>Tianhe-1A, NUDT Intel + Nvidia GPU + custom</td>
<td>China</td>
<td>186,368</td>
<td>2.57</td>
<td>55</td>
<td>4.04</td>
<td>636</td>
</tr>
<tr>
<td>3</td>
<td>DOE / OS Oak Ridge Nat Lab</td>
<td>Jaguar, Cray AMD + custom</td>
<td>USA</td>
<td>224,162</td>
<td>1.76</td>
<td>75</td>
<td>7.0</td>
<td>251</td>
</tr>
<tr>
<td>4</td>
<td>Nat. Supercomputer Center in Shenzhen</td>
<td>Nebulea, Dawning Intel + Nvidia GPU + IB</td>
<td>China</td>
<td>120,640</td>
<td>1.27</td>
<td>43</td>
<td>2.58</td>
<td>493</td>
</tr>
<tr>
<td>5</td>
<td>GSIC Center, Tokyo Institute of Technology</td>
<td>Tusbame 2.0, HP Intel + Nvidia GPU + IB</td>
<td>Japan</td>
<td>73,278</td>
<td>1.19</td>
<td>52</td>
<td>1.40</td>
<td>865</td>
</tr>
<tr>
<td>6</td>
<td>DOE / NNSA LANL &amp; SNL</td>
<td>Cielo, Cray AMD + custom</td>
<td>USA</td>
<td>142,272</td>
<td>1.11</td>
<td>81</td>
<td>3.98</td>
<td>279</td>
</tr>
<tr>
<td>7</td>
<td>NASA Ames Research Center/NAS</td>
<td>Plelades SGI Altix ICE 8200EX/8400EX + IB</td>
<td>USA</td>
<td>111,104</td>
<td>1.09</td>
<td>83</td>
<td>4.10</td>
<td>265</td>
</tr>
<tr>
<td>8</td>
<td>DOE / OS Lawrence Berkeley Nat Lab</td>
<td>Hopper, Cray AMD + custom</td>
<td>USA</td>
<td>153,408</td>
<td>1.054</td>
<td>82</td>
<td>2.91</td>
<td>362</td>
</tr>
<tr>
<td>9</td>
<td>Commissariat a l’Energie Atomique (CEA)</td>
<td>Tera-10, Bull Intel + IB</td>
<td>France</td>
<td>138,368</td>
<td>1.050</td>
<td>84</td>
<td>4.59</td>
<td>229</td>
</tr>
<tr>
<td>10</td>
<td>DOE / NNSA Los Alamos Nat Lab</td>
<td>Roadrunner, IBM AMD + Cell GPU + IB</td>
<td>USA</td>
<td>122,400</td>
<td>1.04</td>
<td>76</td>
<td>2.35</td>
<td>446</td>
</tr>
</tbody>
</table>

500 IT Service | IBM Cluster, Intel + GigE | USA | 7,236 | .051 | 53 |

**November 2011: The TOP10**
Japanese K Computer

K Computer > Sum(#2 : #8) ~ 2.5X #2

Linpack run with 705,024 cores at 10.51 Pflop/s (88,128 CPUs), 12.7 MW; 29.5 hours

Fujitsu to have a 100 Pflop/s system in 2014
Countries Share

Absolute Counts
US: 263
China: 75
Japan: 30
UK: 27
France: 23
Germany: 20
Performance of Countries

US

Teraflop/sec

Performance of Countries

Teraflop/sec

US
EU

Jun-00
Dec-00
Jun-01
Dec-01
Jun-02
Dec-02
Jun-03
Dec-03
Jun-04
Dec-04
Jun-05
Dec-05
Jun-06
Dec-06
Jun-07
Dec-07
Jun-08
Dec-08
Jun-09
Dec-09
Jun-10
Dec-10
Jun-11
Dec-11
Performance of Countries

- US
- EU
- Japan

Teraflop/sec

Performance of Countries

- US
- EU
- China
- Japan

**Pflop/s Club (Peak)**

- Japan (2)
- China (5)
- US (6)
- Germany (2)
- Russia (1)
- France (1)
## Russian Supercomputers

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Computer</th>
<th>Site</th>
<th>Manufacturer</th>
<th>Total Cores</th>
<th>Rmax</th>
<th>Rpeak</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Lomonosov</td>
<td>T-Platforms T-Blade2/1.1, Xeon X5570/X5670 2.93 GHz, Nvidia 2070 GPU, Infiniband QDR</td>
<td>Moscow State University</td>
<td>T-Platforms</td>
<td>33072</td>
<td>674105</td>
<td>1373060</td>
</tr>
<tr>
<td>108</td>
<td>MVS-100K</td>
<td>Cluster Platform 3000 BL460c/BL2x220, Xeon 54xx 3 Ghz, Infiniband</td>
<td>Joint Supercomputer Center</td>
<td>HP</td>
<td>11680</td>
<td>107448</td>
<td>140160</td>
</tr>
<tr>
<td>120</td>
<td></td>
<td>Cluster Platform 3000 BL2x220, E54xx 3.0 Ghz, Infiniband</td>
<td>Kurchatov Institute Moscow</td>
<td>HP</td>
<td>10304</td>
<td>101213</td>
<td>123648</td>
</tr>
<tr>
<td>122</td>
<td>SKIF Aurora</td>
<td>SKIF Aurora Platform - Intel Xeon X5680, Infiniband QDR</td>
<td>South Ural State University</td>
<td>RSC SKIF</td>
<td>8832</td>
<td>100400</td>
<td>117000</td>
</tr>
<tr>
<td>339</td>
<td></td>
<td>HP DL160 Cluster G6, Xeon E5645 6C 2.40 GHz, Gigabit Ethernet</td>
<td>Web Content Provider</td>
<td>HP</td>
<td>12024</td>
<td>59903</td>
<td>115430</td>
</tr>
</tbody>
</table>
Commodity plus Accelerator

Commodity
- Intel Xeon
  - 8 cores
  - 3 GHz
  - 8*4 ops/cycle
  - 96 Gflop/s (DP)

Accelerator (GPU)
- Nvidia C2070 “Fermi”
  - 448 “Cuda cores”
  - 1.15 GHz
  - 448 ops/cycle
  - 515 Gflop/s (DP)

Interconnect
- PCI-X 16 lane
  - 64 Gb/s
  - 1 GW/s

Device Memory
- 6 GB
39 Accelerator Based Systems

- Clearspeed CSX60022
- ATI GPU
- IBM PowerXCell 8i
- NVIDIA 2090
- NVIDIA 2070
- NVIDIA 2050

- 20 US
- 5 China
- 3 Japan
- 2 France
- 2 Germany
- 1 Australia
- 1 Italy
- 1 Poland
- 1 Spain
- 1 Switzerland
- 1 Russia
- 1 Taiwan
Accelerating Dense Linear Algebra with GPUs

LU Factorization in double precision (DP)  
[for solving a dense linear system]  
- GPU (MAGMA)  
- CPU (LAPACK)

Hessenberg factorization in DP  
[for the general eigenvalue problem]  
- GPU (MAGMA)  
- CPU (LAPACK)

Matrix Size  | GFlop/s  | Matrix Size  | GFlop/s
---|---|---|---
1024  | 1022 W*  | 1024  | 220 W*  
3072  | 1022 W*  | 3072  | 220 W*  
5184  | 1022 W*  | 5184  | 220 W*  
7040  | 1022 W*  | 7040  | 220 W*  
9088  | 1022 W*  | 9088  | 220 W*  

**GPU**  
Fermi C2050 [448 CUDA Cores @ 1.15 GHz]  
+ Intel Q9300 [4 cores @ 2.50 GHz]  
- DP peak 515 + 40 GFlop/s  
- System cost ~ $3,000  
- Power * ~ 220 W  
- ~1/10 price  
- ~1/5 power  
- ~4 X performance

**CPU**  
AMD ISTANBUL  
[8 sockets x 6 cores (48 cores) @2.8GHz]  
- DP peak 538 GFlop/s  
- System cost ~ $30,000  
- Power * ~ 1,022 W  
- ~4 X performance

* Computation consumed power rate (total system rate minus idle rate), measured with KILL A WATT PS, Model P430
Future Computer Systems

- Most likely be a hybrid design
  - Think standard multicore chips and accelerator (GPUs)
- Today accelerators are attached
- Next generation more integrated
- Intel’s MIC architecture “Knights Ferry” and “Knights Corner” to come.
  - 48 x86 cores
- AMD’s Fusion
  - Multicore with embedded graphics ATI
- Nvidia’s Project Denver plans to develop an integrated chip using ARM architecture in 2013.
Moore’s Law is Alive and Well

Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović
Slide from Kathy Yelick
But Clock Frequency Scaling Replaced by Scaling Cores / Chip

15 Years of exponential growth ~2x year has ended

Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović
Slide from Kathy Yelick
Performance Has Also Slowed, Along with Power

Power is the root cause of all this

A hardware issue just became a software problem

Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović
Slide from Kathy Yelick
Clock Scaling Hits Power Density Wall

Scaling clock speed (business as usual) will not work

Source: Patrick Gelsinger, Intel

Year


Power Density (W/cm²)

10000  1000  100  10

Hot Plate  Rocket  Nuclear Reactor  Nozzle  Sun’s Surface

Pentium®
**Power Cost of Frequency**

- **Power** $\propto$ **Voltage**$^2$ x **Frequency**  \((V^2F)\)
- **Frequency** $\propto$ **Voltage**
- **Power** $\propto$ **Frequency**$^3$

<table>
<thead>
<tr>
<th></th>
<th>Cores</th>
<th>V</th>
<th>Freq</th>
<th>Perf</th>
<th>Power</th>
<th>PE (Bops/watt)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Superscalar</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>“New” Superscalar</strong></td>
<td>1X</td>
<td>1.5X</td>
<td>1.5X</td>
<td>1.5X</td>
<td>3.3X</td>
<td><strong>0.45X</strong></td>
</tr>
</tbody>
</table>
## Power Cost of Frequency

- **Power** ∝ **Voltage**² x **Frequency** \( (V^2F) \)
- **Frequency** ∝ **Voltage**
- **Power** ∝ **Frequency**³

### Table

<table>
<thead>
<tr>
<th></th>
<th>Cores</th>
<th>V</th>
<th>Freq</th>
<th>Perf</th>
<th>Power</th>
<th>PE (Bops/watt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superscalar</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>&quot;New&quot; Superscalar</td>
<td>1X</td>
<td>1.5X</td>
<td>1.5X</td>
<td>1.5X</td>
<td>3.3X</td>
<td>0.45X</td>
</tr>
<tr>
<td>Multicore</td>
<td>2X</td>
<td>0.75X</td>
<td>0.75X</td>
<td>1.5X</td>
<td>0.8X</td>
<td>1.88X</td>
</tr>
</tbody>
</table>

(Bigger # is better)

50% more performance with 20% less power

Preferable to use multiple slower devices, than one superfast device
Broad Community Support and Development of the Exascale Initiative Since 2007

http://science.energy.gov/ascr/news-and-resources/program-documents/

- Town Hall Meetings April-June 2007
- Scientific Grand Challenges Workshops
  Nov, 2008 – Oct, 2009
  - Climate Science (11/08)
  - High Energy Physics (12/08)
  - Nuclear Physics (1/09)
  - Fusion Energy (3/09)
  - Nuclear Energy (5/09)
  - Biology (8/09)
  - Material Science and Chemistry (8/09)
  - National Security (10/09)
  - Cross-cutting technologies (2/10)
- Exascale Steering Committee
  - “Denver” vendor NDA visits (8/09)
  - SC09 vendor feedback meetings
  - Extreme Architecture and Technology Workshop (12/09)
- International Exascale Software Project
  - Santa Fe, NM (4/09); Paris, France (6/09); Tsukuba, Japan (10/09); Oxford (4/10); Maui (10/10); San Francisco (4/11); Cologne (10/11)
Energy Cost Challenge

At ~$1M per MW, energy costs are substantial

- 10 Pflop/s in 2011 uses ~10 MWs
- 1 Eflop/s in 2018 > 100 MWs

DOE Target: 1 Eflop/s in 2018 at 20 MWs
The High Cost of Data Movement

- Flop/s or percentage of peak flop/s become much less relevant

Approximate power costs (in picoJoules)

<table>
<thead>
<tr>
<th></th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP FMADD flop</td>
<td>100 pJ</td>
</tr>
<tr>
<td>DP DRAM read</td>
<td>4800 pJ</td>
</tr>
<tr>
<td>Local Interconnect</td>
<td>7500 pJ</td>
</tr>
<tr>
<td>Cross System</td>
<td>9000 pJ</td>
</tr>
</tbody>
</table>

Source: John Shalf, LBNL

- Algorithms & Software: minimize data movement; perform more work per unit data movement.
What’s Next?

- All Large Core
- Mixed Large and Small Core
- All Small Core
- Many Small Cores
- Many Floating-Point Cores

Different Classes of Chips
- Home
- Games / Graphics
- Business
- Scientific
# Potential System Architecture

<table>
<thead>
<tr>
<th>Systems</th>
<th>2011 K computer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System peak</strong></td>
<td>10.5 Pflop/s</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>12.7 MW</td>
</tr>
<tr>
<td>System memory</td>
<td>1.6 PB</td>
</tr>
<tr>
<td>Node performance</td>
<td>128 GF</td>
</tr>
<tr>
<td>Node memory BW</td>
<td>64 GB/s</td>
</tr>
<tr>
<td>Node concurrency</td>
<td>8</td>
</tr>
<tr>
<td>Total Node Interconnect BW</td>
<td>20 GB/s</td>
</tr>
<tr>
<td>System size (nodes)</td>
<td>88,124</td>
</tr>
<tr>
<td>Total concurrency</td>
<td>705,024</td>
</tr>
<tr>
<td>MTTI</td>
<td>days</td>
</tr>
</tbody>
</table>
Potential System Architecture with a cap of $200M and 20MW

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>System peak</td>
<td>10.5 Pflop/s</td>
<td>1 Eflop/s</td>
<td>O(100)</td>
</tr>
<tr>
<td>Power</td>
<td>12.7 MW</td>
<td>~20 MW</td>
<td></td>
</tr>
<tr>
<td>System memory</td>
<td>1.6 PB</td>
<td>32 - 64 PB</td>
<td>O(10)</td>
</tr>
<tr>
<td>Node performance</td>
<td>128 GF</td>
<td>1,2 or 15TF</td>
<td>O(10) – O(100)</td>
</tr>
<tr>
<td>Node memory BW</td>
<td>64 GB/s</td>
<td>2 - 4TB/s</td>
<td>O(100)</td>
</tr>
<tr>
<td>Node concurrency</td>
<td>8</td>
<td>O(1k) or 10k</td>
<td>O(100) – O(1000)</td>
</tr>
<tr>
<td>Total Node Interconnect BW</td>
<td>20 GB/s</td>
<td>200-400GB/s</td>
<td>O(10)</td>
</tr>
<tr>
<td>System size (nodes)</td>
<td>88,124</td>
<td>O(100,000) or O(1M)</td>
<td>O(10) – O(100)</td>
</tr>
<tr>
<td>Total concurrency</td>
<td>705,024</td>
<td>O(billion)</td>
<td>O(1,000)</td>
</tr>
<tr>
<td>MTTI</td>
<td>days</td>
<td>O(1 day)</td>
<td>- O(10)</td>
</tr>
</tbody>
</table>
Major Changes to Software & Algorithms

• **Must rethink the design of our algorithms and software**
  - Another disruptive technology
    - Similar to what happened with cluster computing and message passing
  - Rethink and rewrite the applications, algorithms, and software

• Data movement is expense
• Flop/s are cheap, so are provisioned in excess
Critical Issues at Peta & Exascale for Algorithm and Software Design

- Synchronization-reducing algorithms
  - Break Fork-Join model

- Communication-reducing algorithms
  - Use methods which have lower bound on communication

- Mixed precision methods
  - 2x speed of ops and 2x speed for data movement

- Autotuning
  - Today’s machines are too complicated, build “smarts” into software to adapt to the hardware

- Fault resilient algorithms
  - Implement algorithms that can recover from failures/bit flips

- Reproducibility of results
  - Today we can’t guarantee this. We understand the issues, but some of our “colleagues” have a hard time with this.
Fork-Join Parallelization of LU and QR.

Parallelize the update:

- Easy and done in any reasonable software.
- This is the $2/3n^3$ term in the FLOPs count.
- Can be done efficiently with LAPACK+multithreaded BLAS
1. Synchronization (in LAPACK LU)

- Fork-join, bulk synchronous processing

Allowing for delayed update, out of order, asynchronous, dataflow execution
Objectives

- High utilization of each core
- Scaling to large number of cores
- Synchronization reducing algorithms

Methodology

- Dynamic DAG scheduling (QUARK)
- Explicit parallelism
- Implicit communication
- Fine granularity / block data layout

Arbitrary DAG with dynamic scheduling
Pipelining: Cholesky Inversion
3 Steps: Factor, Invert L, Multiply L’s

48 cores
POTRF, TRTRI and LAUUM.
The matrix is 4000 x 4000, tile size is 200 x 200,

POTRF+TRTRI+LAUUM: 25 (7t-3)
Cholesky Factorization alone: 3t-2

Pipelined: 18 (3t+6)
Communication Avoiding Algorithms

- Goal: Algorithms that communicate as little as possible
- Jim Demmel and company have been working on algorithms that obtain a provable minimum communication. (M. Anderson yesterday)
- Direct methods (BLAS, LU, QR, SVD, other decompositions)
  - Communication lower bounds for all these problems
  - Algorithms that attain them (all dense linear algebra, some sparse)
- Iterative methods - Krylov subspace methods for $Ax=b$, $Ax=\lambda x$
  - Communication lower bounds, and algorithms that attain them (depending on sparsity structure)
- For QR Factorization they can show:

<table>
<thead>
<tr>
<th># flops</th>
<th>Lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta(mn^2)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># words</th>
<th>$\Theta(\frac{mn^2}{\sqrt{W}})$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th># messages</th>
<th>$\Theta(\frac{mn^2}{W^{3/2}})$</th>
</tr>
</thead>
</table>
Communication Reducing QR Factorization

Quad-socket, quad-core machine Intel Xeon EMT64 E7340 at 2.39 GHz. Theoretical peak is 153.2 Gflop/s with 16 cores.

Matrix size 51200 by 3200
Mixed Precision Methods

• Mixed precision, use the lowest precision required to achieve a given accuracy outcome
  - Improves runtime, reduce power consumption, lower data movement
  - Reformulate to find correction to solution, rather than solution \[ \Delta x \] rather than \[ x \].
Mixed Precision Solvers

MAGMA LU-based solvers on Fermi (C2050)

FERMI Tesla C2050: 448 CUDA cores @ 1.15GHz
SP/DP peak is 1030 / 515 GFlop/s

- Direct solvers
  - Factor and solve in working precision
- Mixed Precision Iterative Refinement
  - Factor in single (i.e. the bulk of the computation in fast arithmetic) and use it as preconditioner in simple double precision iteration, e.g.
  \[ x_{i+1} = x_i + (LU_{SP})^{-1} P (b - A x_i) \]

Similar results for Cholesky & QR
Conclusions

• For the last decade or more, the research investment strategy has been overwhelmingly biased in favor of hardware.

• This strategy needs to be rebalanced - barriers to progress are increasingly on the software side.

• High Performance Ecosystem out of balance
  ▪ Hardware, OS, Compilers, Software, Algorithms, Applications
    • No Moore’s Law for software, algorithms and applications