Announcements

• Put your name on HW#1 before turning in!
• Homework #2 will be posted. Articles to read.
• I’ll be in Arizona on Tuesday for ISPASS Program Committee meeting. I’ll (hopefully) be back Thursday.
<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Country</th>
<th>Arch</th>
<th>Proc</th>
<th>Cores</th>
<th>Max/Peak TFLOPS</th>
<th>Accel</th>
<th>Power kW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TaihuLight</td>
<td>China</td>
<td>Sunway</td>
<td>Sunway</td>
<td>10,649,600</td>
<td>93014/125435</td>
<td>?</td>
<td>15,000</td>
</tr>
<tr>
<td>2</td>
<td>Tianhe-2</td>
<td>China</td>
<td>x86</td>
<td>IVB</td>
<td>3,120,000</td>
<td>33862/54902</td>
<td>xeon-phi</td>
<td>18,000</td>
</tr>
<tr>
<td>3</td>
<td>Titan</td>
<td>USA/ORNL</td>
<td>x86</td>
<td>Opter</td>
<td>560,640</td>
<td>17590/27112</td>
<td>NVD K20</td>
<td>8,209</td>
</tr>
<tr>
<td>4</td>
<td>Sequoia</td>
<td>USA/LLNL</td>
<td>Power</td>
<td>BG/Q</td>
<td>1,572,646</td>
<td>17173/20132</td>
<td>?</td>
<td>7,890</td>
</tr>
<tr>
<td>5</td>
<td>Cori</td>
<td>USA/LBNL</td>
<td>x86</td>
<td>Cray</td>
<td>622,336</td>
<td>14014/27880</td>
<td>Xeon Phi</td>
<td>3,939</td>
</tr>
<tr>
<td>6</td>
<td>Oakforest</td>
<td>Japan</td>
<td>x86</td>
<td>??</td>
<td>556,104</td>
<td>13554/24913</td>
<td>Xeon Phi</td>
<td>2,719</td>
</tr>
<tr>
<td>7</td>
<td>K computer</td>
<td>Japan/RIKEN</td>
<td>SPARC</td>
<td>VIIIfx</td>
<td>705,024</td>
<td>10510/11280</td>
<td>?</td>
<td>12,660</td>
</tr>
<tr>
<td>8</td>
<td>Piz Daint</td>
<td>Switzerland</td>
<td>x86</td>
<td>Intel</td>
<td>206,720</td>
<td>9779/15998</td>
<td>NVD Tesla</td>
<td>1,312</td>
</tr>
<tr>
<td>9</td>
<td>Mira</td>
<td>USA/Argonne</td>
<td>Power</td>
<td>BG/Q</td>
<td>786,432</td>
<td>8586/10066</td>
<td>?</td>
<td>3,945</td>
</tr>
<tr>
<td>10</td>
<td>Trinity</td>
<td>USA/LANL</td>
<td>x86</td>
<td>??</td>
<td>301,056</td>
<td>8100/11078</td>
<td>?</td>
<td>4,233</td>
</tr>
<tr>
<td>11</td>
<td>UK Met</td>
<td>UK</td>
<td>x86</td>
<td>Cray</td>
<td>241,920</td>
<td>6765/8128</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>12</td>
<td>Marconi</td>
<td>Italy</td>
<td>x86</td>
<td>??</td>
<td>241,808</td>
<td>6223/10833</td>
<td>Xeon Phi</td>
<td>?</td>
</tr>
<tr>
<td>13</td>
<td>Pleiades</td>
<td>US/NASA</td>
<td>x86</td>
<td>SGI</td>
<td>241,108</td>
<td>5951/7107</td>
<td>?</td>
<td>4407</td>
</tr>
<tr>
<td>14</td>
<td>Hazel Hen</td>
<td>Germany</td>
<td>x86</td>
<td>??</td>
<td>185,088</td>
<td>5640/7403</td>
<td>?</td>
<td>3615</td>
</tr>
<tr>
<td>15</td>
<td>Shaheen II</td>
<td>Saudi Arabia</td>
<td>x86</td>
<td>SNB-EP</td>
<td>196,608</td>
<td>5537/7235</td>
<td>?</td>
<td>2,834</td>
</tr>
<tr>
<td>16</td>
<td>Pangea</td>
<td>France</td>
<td>x86</td>
<td>??</td>
<td>220,800</td>
<td>5283/6712</td>
<td>?</td>
<td>4150</td>
</tr>
<tr>
<td>17</td>
<td>Stampede</td>
<td>USA/TACC</td>
<td>x86</td>
<td>SNB-EP</td>
<td>462,462</td>
<td>5168/8520</td>
<td>XeonPhi</td>
<td>4,510</td>
</tr>
<tr>
<td>18</td>
<td>Theta</td>
<td>USA/Argonne</td>
<td>x86</td>
<td>??</td>
<td>207,360</td>
<td>5095/8626</td>
<td>XeonPhi</td>
<td>1087</td>
</tr>
<tr>
<td>19</td>
<td>Juqeen</td>
<td>DE/Julich</td>
<td>Power</td>
<td>BG/Q</td>
<td>458,752</td>
<td>5008/5872</td>
<td>?</td>
<td>2,301</td>
</tr>
<tr>
<td>20</td>
<td>Cheyenne</td>
<td>USA/NCAR</td>
<td>x86</td>
<td>??</td>
<td>144,900</td>
<td>4788/5332</td>
<td>?</td>
<td>1727</td>
</tr>
<tr>
<td>21</td>
<td>Vulcan</td>
<td>USA/LLNL</td>
<td>Power</td>
<td>BG/Q</td>
<td>393,216</td>
<td>4293/5033</td>
<td>?</td>
<td>1,972</td>
</tr>
<tr>
<td>22</td>
<td>Abel</td>
<td>USA/Geo</td>
<td>x86</td>
<td>??</td>
<td>145920</td>
<td>4042/5369</td>
<td>?</td>
<td>1800</td>
</tr>
</tbody>
</table>
How long does it take to run LINPACK? How much money does it cost to run LINPACK?

How much RAM? How much cooling?

Turnover since last time I taught the class?
What goes into a top supercomputer?

- Commodity or custom
  embedded vs high-speed?
- Memory
- Storage
  How much?
  Large hadron collider one petabyte of data every day
  Shared? If each node wants same data, do you need to
  replicate it, have a network filesystem, copy it around
with jobs, etc? Cluster filesystems?

• Reliability. How long can it stay up without crashing? Can you checkpoint/restart jobs?
  Sequoia MTBF 1 day.
  Blue Waters 2 nodes failure per day.
  Titan MTBF less than 1 day

• Power / Cooling
  Big river nearby?

• Accelerator cards / Heterogeneous Systems

• Network
  How fast? Latency? Interconnect? (torus, cube,
hypercube, etc)
Ethernet? Infiniband? Custom?

• Operating System
  Linux? Custom? If just doing FP, do you need overhead of an OS? Job submission software, Authentication

• Software – how to program?
  Too hard to program can doom you. A lot of interest in the Cell processor. Great performance if programmed well, but hard to do.

• Tools – software that can help you find performance problems
Introduction to Performance Analysis
What is Performance?

- Getting results as quickly as possible?
- Getting *correct* results as quickly as possible?
- What about Budget?
- What about Development Time?
- What about Hardware Usage?
- What about Power Consumption?
Motivation for HPC Optimization

HPC environments are expensive:

- Procurement costs: $40 million
- Operational costs: $5 million/year
- Electricity costs: 1 MW / year $1 million
- Air Conditioning costs: ??
Know Your Limitation

- CPU Constrained
- Memory Constrained (Memory Wall)
- I/O Constrained
- Thermal Constrained
- Energy Constrained
Performance Optimization Cycle

- Develop Code
  - Functionally Complete/Correct Code
  - Measure
  - Analyze
  - Modify / Tune
    - Functionally Complete/Correct/Optimized Code
  - Usage / Production
Wisdom from Knuth

“We should forget about small efficiencies, say about 97% of the time:

**premature optimization is the root of all evil.**

Yet we should not pass up our opportunities in that critical 3%. A good programmer will not be lulled into complacency by such reasoning, he will be wise to look carefully at the critical code; but only after that code has been identified” — Donald Knuth
Amdahl’s Law

- Original
- Speed up Blue 100x
- Speed up Red 2x
Speedup

• Speedup is the improvement in latency (time to run)

\[ S = \frac{t_{old}}{t_{new}} \]

So if originally took 10s, new took 5s, then speedup=2.
Scalability

- How a workload behaves as more processors are added
- Parallel efficiency: \( E_p = \frac{S_p}{p} = \frac{T_1}{pT_p} \)
- Linear scaling, ideal: \( S_p = p \)
- Super-linear scaling – possible but unusual
Strong vs Weak Scaling

- Strong Scaling – for fixed program size, how does adding more processors help
- Weak Scaling – how does adding processors help with the same per-processor workload
Strong Scaling

- Have a problem of a certain size, want it to get done faster.

- Ideally with problem size N, with 2 cores it runs twice as fast as with 1 core.

- Often processor bound; adding more processing helps, as communication doesn’t dominate.

- Hard to achieve for large number of nodes, as many
algorithms communication costs get larger the more nodes involved
Weak Scaling

- Have a problem, want to increase problem size without slowing down.

- Ideally with problem size N with 1 core, a problem of size 2\( \times \)n just as fast with 2 cores.

- Often memory or communication bound.
Scaling Example

LINPACK on Rasp-pi cluster. What kind of scaling is here?
Weak scaling. To get linear speedup need to increase problem size.
If it were strong scaling, the individual colored lines would increase rather than dropping off.