Beacon:
A Path to Energy-Efficient HPC

R. Glenn Brook
Director, Application Acceleration Center of Excellence
National Institute for Computational Sciences

[Email Address: glenn-brook@tennessee.edu]
The Beacon Project

- Funded by NSF to port and optimize scientific codes to the Intel® Xeon Phi™ coprocessor
- State-funded expansion focus on energy efficiency, big data applications, and industry
- The pre-production Intel® Xeon Phi™ coprocessors in the original Beacon cluster will be upgraded to commercial versions in 2013.

<table>
<thead>
<tr>
<th>Original Beacon Cluster by Appro</th>
<th>Fully Upgraded Beacon Cluster by Appro</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nodes</strong></td>
<td><strong>Nodes</strong></td>
</tr>
<tr>
<td></td>
<td>2 service, 16 compute</td>
</tr>
<tr>
<td><strong>CPU model</strong></td>
<td>Intel Xeon E5-2670</td>
</tr>
<tr>
<td><strong>CPUs per node</strong></td>
<td>2 8-core, 2.6 GHz</td>
</tr>
<tr>
<td><strong>RAM per node</strong></td>
<td>64 GB</td>
</tr>
<tr>
<td><strong>Intel® Xeon Phi™ coprocessors per node</strong></td>
<td>2 x pre-production</td>
</tr>
<tr>
<td><strong>Cores per Intel® Xeon Phi™ coprocessor</strong></td>
<td>50+</td>
</tr>
<tr>
<td><strong>RAM per Intel® Xeon Phi™ coprocessor</strong></td>
<td>8 GB GDDR5</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intel® Xeon Phi™ coprocessors per node</strong></td>
<td>4 x 5110P</td>
</tr>
<tr>
<td><strong>Cores per Intel® Xeon Phi™ coprocessor</strong></td>
<td>60</td>
</tr>
<tr>
<td><strong>RAM per Intel® Xeon Phi™ coprocessor</strong></td>
<td>8 GB GDDR5</td>
</tr>
</tbody>
</table>

Intel, Xeon, and Intel Xeon Phi are trademarks of Intel Corporation in the U.S. and/or other countries.
Sanity Check – Original Estimates

• Power:
  • 2x Intel® Xeon® E5-2670 @ 115W TDP
  • 4x Intel® Xeon Phi™ Coprocessor 5110Ps @ 225W TDP
  • 256 GB RAM: 16 DIMMs @ 6 W ea. = 96 W
  • Chassis: 120W (cooling + other)
  • Approximate total power: 1346 W (theoretical)

• Peak performance:
  • 4x Intel® Xeon Phi™ coprocessor 5110P: 4040 GFLOPS
  • 2x Intel® Xeon® E5-2670: 330 GFLOPS
  • Approximate total performance: 4370 GFLOPS
Sanity Check – Original Estimates

- Power: ~1346 W
- Performance: 4370 GFLOPS
- Assume 92% efficiency for power conversion:
  - $1342/0.92 = 1458.7$ W AC
- Assume 70% efficiency for HPL:
  - $4370 \times 0.7 = 3059$ GFLOPS
- Approximate MFLOPS/W:
  - $3059 / 1458.7 = 2097$ MFLOPS/W before optimizations
Green500 Hardware

- Appro Xtreme-X Supercomputer powered by ACE

<table>
<thead>
<tr>
<th>Beacon Green500 Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
</tr>
<tr>
<td>CPU model</td>
</tr>
<tr>
<td>CPUs per node</td>
</tr>
<tr>
<td>RAM per node</td>
</tr>
<tr>
<td>SSD per node</td>
</tr>
<tr>
<td>Intel® Xeon Phi™ Coprocessors 5110P per node</td>
</tr>
<tr>
<td>Cores per Intel® Xeon Phi™ coprocessor 5110P</td>
</tr>
<tr>
<td>RAM per Intel® Xeon Phi™ coprocessor 5110P</td>
</tr>
</tbody>
</table>
Assembled Team

• Intel:
  • Mikhail Smelyanskiy – software lead
  • Karthikeyan Vaidyanathan – MIC HPL
  • Ian Steiner – host HPL optimization
  • Many, many others: Joe Curley, Jim Jeffers, Pradeep Dubey, Susan Meredith, Rajesh Agny, Russ Fromkin, and many dedicated and passionate team members

• Technische Universität München:
  • Alexander Heinecke – MIC HPL

• Appro:
  • John Lee – hardware lead
  • David Parks – HPL and system software
  • Edgardo Evangelista, Danny Tran, others – system deployment and support

• NICS:
  • Glenn Brook – project lead
  • Ryan Braby, Troy Baer – system support
Design Approach

• Entire system designed to target energy-efficiency
• Custom HPL implementation – not MKL
  • Targets energy-efficiency
  • Optimized for 4 coprocessors per node
• Dynamic power management
• Minimized power consumed by unused components
  • For example, unused USB and Ethernet ports
• Summary: co-design of software and hardware to maximize energy efficiency
Measurement Methodology

• 9 Appro Greenblade subracks
  • 4x Appro GB824M servers
  • 4+1 highly efficient power supplies
• Each power supply implements the PMBus spec.
  • Crest factor: 6
  • Instantaneous AC input (RMS)
• Used iSCB interfaces to continuously poll every power supply in all of the subracks in the system
• Measured power of the InfiniBand switch separately
  • Peak just below 90W
  • Added to Rmax Power measured from power supplies
Power Sampling
Power Consumption

Beacon Green500 HPL Power Consumption

R_max: 112200 GFLOPS
Power: 44.894778 KW
MFLOPS/W: 2499.44
# Energy Efficient HPC System

## Workload Power Measurement Methodology

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect 1a: granularity of power measurements</td>
<td>1 instantaneous power sampling per second</td>
<td>1 instantaneous power sampling per second</td>
<td>continuously integrated total energy</td>
</tr>
<tr>
<td>Aspect 1b: timespan of power measurements</td>
<td>at least one power averaged measurement covering at least 20% of the run</td>
<td>a time series of equally spaced power averaged measurements</td>
<td>a time series of equally spaced integrated total energy values</td>
</tr>
<tr>
<td>Aspect 1c: reported analyzed measurements</td>
<td>core phase average power</td>
<td>core phase average power, whole application average power, idle power</td>
<td>core phase average power, whole application average power, idle power</td>
</tr>
<tr>
<td>Aspect 2: machine fraction</td>
<td>the greater of 1/64 of the machine or 1 kW</td>
<td>the greater of 1/4 of the machine or 10 kW</td>
<td>whole machine</td>
</tr>
<tr>
<td>Aspect 3: subsystems included</td>
<td>all participating subsystems, either measured or estimated</td>
<td>all participating subsystems, either measured or estimated</td>
<td>all participating subsystems must be measured</td>
</tr>
<tr>
<td>Aspect 4: power measurement point</td>
<td>upstream of power conversion OR power conversion loss modeled with manufacturer data</td>
<td>upstream of power conversion OR power conversion loss modeled with off-line measurements of single power supply</td>
<td>upstream of power conversion OR power conversion measured simultaneously during the same run</td>
</tr>
</tbody>
</table>
WORLD RECORD!
“Beacon” at NICS
Intel® Xeon® + Intel Xeon Phi™ Cluster
First to Deliver
2.499 GigaFLOPS / Watt
71.4% efficiency
#1 on current Green500

Other brands and names are the property of their respective owners.
Contact Information

R. Glenn Brook, Ph.D.
Director, Application Acceleration Center of Excellence
National Institute for Computational Sciences

glenn-brook@tennessee.edu