Level3 Measurement of "Shoubu System B"

2018/11/14
ExaScaler Inc.
Sunao Torii
PEZY Computing K.K.
Ryo Sakamato
# Executive Summary

Comparison between now and then

<table>
<thead>
<tr>
<th>Shoubu SystemB</th>
<th>Nov. 2018</th>
<th>June 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rmax (GFLOPS)</td>
<td>1063.3TFLOPS</td>
<td>857.6TFLOPS</td>
</tr>
<tr>
<td>Nmax (Matrix Size)</td>
<td>1,904,640</td>
<td>1,388,800</td>
</tr>
<tr>
<td>Num. of Nodes</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>Node CPU Memory</td>
<td>64GB</td>
<td>32GB</td>
</tr>
<tr>
<td>GFLOPS/W</td>
<td>17.6</td>
<td>18.4</td>
</tr>
<tr>
<td>Measurement Quality</td>
<td>Level3</td>
<td>Level1</td>
</tr>
<tr>
<td>Measurement Nodes</td>
<td>Full (60)</td>
<td>Partial (20)</td>
</tr>
<tr>
<td>Measurement NW comp.</td>
<td>Full (2)</td>
<td>Full (2)</td>
</tr>
<tr>
<td>Including internal cooling device</td>
<td>Yes (Coolant Circulation Pump and Valve)</td>
<td>No</td>
</tr>
<tr>
<td>Including storage device</td>
<td>No (Not Use in HPL)</td>
<td></td>
</tr>
<tr>
<td>Efficiency Enhancement</td>
<td>Liquid Temp. Optimizing</td>
<td>N.A</td>
</tr>
</tbody>
</table>
Our Motivation

• Maintain #1 place of Green500 List
  Rmax performance within TOP500 List is required.
  
  Boosting Rmax performance
  Our target is Rmax over 1.0 Peta FLOPS.

• Level 3 quality power measurement
  Some people say, “Are ExaScaler machines really energy efficient?” and I promised, we will try Level2 or 3 next time.

  Whole system power consumption is measured.
  Coolant Temperature optimization is added.
Agenda

• Executive Summary of latest TOP500 measurement
• Our motivation
• How to Performance boost
• Level1 & Level3 measurement
• Cooling power optimization
• Estimated PUE of Shoubu System B
• Collaboration progress
• Conclusion
How to performance boost

• Increasing the number of Nodes
  • Previous: 50nodes
  • Current: 60nodes

• Expanding Nmax(Matrix size)
  • Previous: 1,388,800 -> 69,440x69,440 / Each Accelerator (PEZY-SC2)
  • Current: 1,904,640 -> 79,360x79,360 / Each Accelerator (PEZY-SC2)

To realize them, Node CPU memory is enlarged from 32GB to 64GB
Shoubu System B block diagram

- **Storage**
- **FE**
- **Interconnect**
- **Liquid Immersion Tank**
- **EPU(PSU)**

- **Brick 4 x compute node**
- **Control Panel**
- **HVDC**
- **Pump**
- **HX** (Heat Exchanger (passive))
- **Flow Ctrl. Valve**

- **Supplying Cooling Water**
- **Shared with Other systems**

- **1 Φ 1W_AC200V**
- **1 Φ 3W_AC200V**
- **3 Φ 3W_AC200V**
- **1 Φ 2W_AC100V**
- **3 Φ 3W_AC200V**

- **MCCB**

2018/11/14  Green500 BOF@SC2018
Level 1 Measurement

[Diagram of a level 1 measurement setup:
- Green500 BOF@SC2018
- 2018/11/14
- Diagram includes elements such as
  - Liquid Immersion Tank
  - Compute Node
  - Interconnect
  - Storage
  - EPU (PSU)
  - HVDC
  - MCCB
  - Flow Ctrl. Valve
  - Heat Exchanger (passive)
  - Measurement Point
  - Sub System
- Diagram arrows indicate flow and connections]
Level3 Measurement

- Liquid Immersion Tank
- Interconnect
- Brick (4 x compute node)
- EPU (PSU)
- Storage
- FE
- Measurement Point
- Flow Ctrl. Valve
- HX: Heat Exchanger (passive)
- Supplying Cooling Water
- Shared with Other systems
- Sub System
- MCCB
- HVDC
- Control Panel
- Pump
- Measurement Point

1 φ 1W_AC200V
1 φ 3W_AC200V
3 φ 3W_AC200V
1 φ 2W_AC100V
3 φ 3W_AC200V
1 φ 2W_AC200V
1 φ 1W_AC200V

2018/11/14 Green500 BOF@SC2018
Level3 Measurement Pictures

Compute Node @ Input point of HVDC

Sub-systems@ In MCCB

Clump sensors are used at all measurement points
1st Measurement result

15.99 GF/W@Level3, 17.90 GF/W@Level1

Compute Node (Brick): 55.6 kW
Sub System: 6.4 kW
Breakdown of subsystem power

Cooling pump is using power almost all in subsystems.

Total Subsystem: 6.4kW
Cooling Pump(included FlowCtl): 6.2kW
FE: 0.2kW
Cooling power optimization

- Reduction of Pump Power
- Add Inverter (VVVF) control

Flow Ctrl. Valve
Heat Exchanger (passive)
Measurement Point

Sub System
MCCB
Pickup Handles

Control Panel
VVVF
Pump
HX
DC48V
DC380V
HVDC
EPU (PSU)
Liquid Immersion Tank
4 x compute node
Brick
Interconnect
X15
4 x compute node

Supplying Cooling Water
Shared with Other systems

2018/11/14
Green500 BOF@SC2018
Cooling power optimization

- Coolant circulation speed is controlled to keep around 35°C during Core phase.
  Freq: 50Hz to 17Hz

![Coolant Temperature Graph](image)

Total Subsystem: 0.7kW
Cooling Pump(+inv, flow ctl): 0.5kW
FE: 0.2kW
Final Measurement result

17.6 GFlops/W @ Level 3
PUE (Power Usage Effectiveness)

\[
\text{PUE} = \frac{\text{cooling + server power}}{\text{server power}}
\]

• System Internal PUE : \((0.5+59.9)/59.9 = 1.01\)

• Estimated Total PUE including facilities is around 1.2
  • Chilling water is generated by ACCS facilities
    • RIKEN ACCS COP is around 4-5 in October.
      \textit{COP: Coefficient Of Performance}
    • Combine Turbo-chiller, Free-Cooling Tower, and well water
    • Shared with HOKUSAI, AHU and ShoubuB

<table>
<thead>
<tr>
<th></th>
<th>Air Cooling</th>
<th>Liquid Immersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>On board Fan Power</td>
<td>Accumulated to server power</td>
<td>Not Required</td>
</tr>
<tr>
<td>Leakage Power</td>
<td>High (Hi-temperature)</td>
<td>Low (Low-temperature)</td>
</tr>
<tr>
<td>Absolute Total Power</td>
<td>High</td>
<td>\textbf{Low}</td>
</tr>
<tr>
<td>PUE at same calculation</td>
<td>Lower (Advantageous)</td>
<td>Higher (Dis-advantageous)</td>
</tr>
</tbody>
</table>
Collaboration: Heterogeneous Many Core Project

Large-Scale Computational Science on Heterogeneous Many Core Computers

Visit Booth: 821

We have been done:

1. MD Simulation of Water
   - Measured effective performance 3.8PFLOPs(10.6%) on Gyoukou

2. Simulation of Rings of Saturn
   - Measured effective performance 47.9PFLOPs(40%) on SunwayTaihuLight
   - 10.6PFLOPs(23.5%) on Gyoukou

   - Measured effective performance 4.78PFLOPs(21.5%) on Gyoukou

4. Real Time Simulation of Cerebellum of Macaca
   - Realize Cerebellum of Macaca using 8 billion neuron and synapse and simulate learning process of eye movement.
   - Measured effective performance 1.85PFLOPs on Gyoukou

5. Accelerate Homology Analysis for Bacteria analysis in Flora
   - About 10-100 times speed up

6. Double double precision operation in BLAS (Rgemm)
   - 76% of peak performance
Level3 Measurement is not so difficult for such under 100KW scale machines.

- Understanding the regulation is not so easy.
- Great help to find out hiding overhead.

We never stop developing MOST power efficient supercomputer!

PEZY-SC3 coming soon!