ENERGY-EFFICIENT VISUALIZATION PIPELINES
A CASE STUDY IN CLIMATE SIMULATION

Vignesh Adhinarayanan
Ph.D. (CS) Student
Synergy Lab, Virginia Tech
INTRODUCTION

“Supercomputers are constrained by power”

• Power budget for Los Alamos county = 66 MW
• Power budget for Trinity supercomputer alone = 15 MW

• Exceeding power budget $\Rightarrow$ Brownouts in Los Alamos
  – Installing and starting ASCI White believed to play a part in the rolling California brownouts in 2001
INTRODUCTION

“Supercomputers are constrained by energy”

• 1 MW power consumption → 1 million dollars per year
  – Operating cost of supercomputers is comparable to the acquisition cost
    • The gap is expected to narrow down in the future
THE ENERGY CHALLENGE

• Off-chip data movement cost nearly hundred times as much energy as on-chip data movement

TRADITIONAL “POST-PROCESSING” VISUALIZATION

HPC System

Nodes
(Simulation runs here)

Disks

Rendering Farm

Rendering Nodes
(Visualization takes place here)
Modern “Post-Processing” Visualization

HPC System

Nodes
(Simulation and visualization runs in HPC nodes)

Disks

Also write raw output only every few iterations (i.e., temporal sampling technique is used)

But you may miss out on important simulation events
**Post-processing vs In-situ Pipelines**

**Post-processing**
- **Simulation** → **Disk Write**
  - Large raw output
- **Disk Read** → **Visualization** → **Disk Write**
  - Large raw output
  - Small image

**In-situ**
- **Simulation** → **Disk Write**
- **Visualization** → **Disk Write**
  - Small image

**Traditional Post-Processing:** Post-processing without any sampling

**Modern Post-Processing:** Post-processing with temporal sampling (write output every few iterations – here every 24 iterations)

**In-situ:** Produce images on the fly and do so only every few iterations
**Goal**

“Study the performance, power, and energy trade-offs among traditional post-processing, modern post-processing, and in-situ visualization pipelines”

- Detailed sub-component level power measurements within a node to gain detailed insights
  - i.e., measure power consumption of CPU, memory, and disk
- Measurements at scale to understand problems unique to big supercomputers
APPLICATION

- **Modeling and Prediction Across Scale (MPAS) Ocean Simulation**
  - Solves an unstructured mesh problem
  - End goal: Identify eddies in the ocean

*Eddies near Southern Africa*
HARDWARE PLATFORM

• Compute nodes
  – 64 nodes
    • Each node contains 2x Intel Xeon E5-2670 and 64 GB of RAM
  – Nominal power consumption
    • 6000 W (idle) to 20000 W (workload such as MPAS)

• Storage nodes
  – Lustre file system
  – 5 nodes configured as 1 master + 2 MDS + 2 OSS
    • 1 RAID storage per MDS and OSS
  – Nominal power consumption
    • 2500W (idle) to 2800W (active)
**Experiments at Scale**

**Energy Comparison**

Real measurements

Partial measurement and estimation

In-situ consumes 19% lower energy than post-processing
# Single-Node Experiments

## Hardware Platform

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>2x Intel Xeon E5-2665</td>
</tr>
<tr>
<td>CPU frequency</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Last-level cache</td>
<td>20 MB</td>
</tr>
<tr>
<td>Memory</td>
<td>4x 16GB DDR3-1333</td>
</tr>
<tr>
<td>Memory size</td>
<td>64 GB</td>
</tr>
<tr>
<td>Hard disk</td>
<td>Seagate 7200rpm disk</td>
</tr>
<tr>
<td>Storage size</td>
<td>500GB</td>
</tr>
<tr>
<td>Disk bandwidth</td>
<td>6.0 Gbps</td>
</tr>
</tbody>
</table>

Hardware configuration
**DATA COLLECTION**

- **System under test**
  - Subsystem-level monitoring via RAPL

- **WattsUp power meter**
  - Full-system power provided by real power meter

- **Monitoring system**
  - Memory and processor power provided by validated power model

- **Power readings logged every one second**

- **Disk power consumption for micro-benchmarks estimated as Wattsup minus RAPL**
**Disk Power Model**

- Constant power from the *spinning* of disk
- Power consumption of read/write head dependent on *number of I/O operations*
- Power consumption of actual reads and writes dependent on *volume of data*
**Single-Node Experiments**

**Energy Comparison**

- Processor and memory consume a lot of energy while waiting for I/O.
- Worthwhile to minimize energy consumption while idling.

![Energy Comparison Chart]

![Sources of energy reduction]

- Energy saved from reduced off-chip data movement (17%)
- Energy saved from reduced system idling (83%)

---

**Graphs and Charts**

- Insitu vs. PostProcessing Energy Comparison
- Sources of energy reduction pie chart.
**Single-Node Experiments**

**Storage Requirements**

- ~97.5% lower storage requirement for the in-situ pipeline
  - Implies smaller storage cluster
  - Implies lower power consumption
RESOURCES FOR POST-PROCESSING

COMPUTE NODES

STORAGE NODES
RESOURCES FOR INSITU

COMPUTE NODES

STORAGE NODES
Redistributing Storage Power to Compute Nodes: Impact on Performance

Assuming reduced storage nodes results in 10% of total power redirected to compute nodes
- Performance improves by up to 6% for MPAS-O
**Findings**

- Most energy savings come from reducing system idling (i.e., from reducing the I/O wait time)
- Further savings possible if we can reduced size of the storage nodes
CONCLUSION

- In-situ visualization offers the following advantages:
  - Reduced energy consumption (by reducing system idling or I/O wait time)
  - Reduced power (by using fewer storage nodes)
  - Improved performance (by reducing I/O wait time and by making more power available for compute nodes)
EXPECTATIONS FOR A SUPERCOMPUTER

• Increased I/O wait time
  – Storage separated from compute by network
  – Longer execution time and corresponding increase in energy

• Additional energy consumption from data movement through the network
  – No data transfer via network cables in single-node

• Power/energy overhead for storage higher
  – Separate cluster for storage → additional CPUs, memory, cooling etc.
  – Storage sub-system shared with compute sub-system in single-node
Future Directions

• Enhancing HPC systems
  – Flash buffers and SSDs can reduce I/O wait time
    • Downside: Introducing more components can increase power consumption

• HPC system design changes
  – Bringing storage nodes and compute nodes together
    • Similar to Memory in Processor or Processor in Memory concepts in the computer architecture community

• Runtime system changes
  – Energy proportional computing and storage
    • Putting compute nodes to sleep states during I/O
    • Putting some storage nodes to deep sleep state when bandwidth and storage requirements are lower
In-situ consumes 7% lower execution time than modern post-processing
- Reduced I/O wait time
- The difference will be significant for an HPC system
  - Details later
**Single-Node Experiments**

**Power Comparison**

- In-situ consumes 3% more power than *modern* post-processing
  - Difficult trade-off choice

- Might not be the same for a supercomputer
  - Details later