SMAUG
System-Level Modeling and Optimized Use of Disruptive Memory Technologies

Memory Hierarchy with DMTs? Complex!

In-memory and near-memory computing can save bandwidth and offer parallelism
Multi-core and many-core CPUs with shared and distributed caches
Novel memory technologies have different properties
RDMA turns this into Giga
On-chip/on-board networks lead to non-uniform I/O and memory architectures

As memory hierarchies are becoming increasingly complex, system software needs a hardware/software model to manage resources efficiently.

Approach / SMAUG Project

Phase 1: Models
- Set up a test bed for disruptive memory technologies
  - Hardware: Cluster that supports NVRAM, RDMA, PIM, HBM, NMC
  - Software: Linux and MxCluster (extended MxKernel for bare-metal tests)
- Detailed measurements
  - Bottlenecks, synchronization effects, access patterns, ...
- Creation of models for HW components, systems, and applications
- Model validation based on system simulator
  - Early experiments with model-based management strategies

Phase 2: Management
- Systematic analysis of resource management strategies
- Integration into MxCluster operating system and evaluation

Milestones in Phase 1

Work structured in three full iterations:
3x (test bed → modeling → measurements → validation)

NVRAM (latest gen. Optane DCPMMs)
RDMA ( Mellanox Infiniband)
PIM (UPMEM DIMMs)
NMC & HBM ( Xilinx Alveo US0280)

<table>
<thead>
<tr>
<th>Work Package</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP1 platform model</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td>WP2 measurements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP3 testbed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP4 SW model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP5 SW analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP6 validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP7 optimized use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP8 collaboration &amp; dissemination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Role within the Priority Program

Huge potential for collaboration!

State of the Art / Models

Analyses/models of disruptive memory technology

NVM
[Izrailevitz ‘19, Wang ‘20, Pohl ‘17, George ‘20, Scargall ‘20, Köppen ‘19]

RDMA
[MacArthur ‘12, Kalia ‘16, Nelson ‘19, Wei ‘21, Shen ‘20, Qiu ‘18]

NMC/PIM
[Singh ‘19, Hsieh ‘16, Mutlu ‘20, Khan ‘20, Corda ‘21, Lee ‘20]

HBM
[Das ‘20, Laffighe ‘17]

Whole-system models for OS decision making

Barrelfish SKB
[Schöpbach ‘12, Baumann ‘09]

MCTOP
[Chatzopoulos ‘17]

hwloc
[Broquedis ‘10]

• Almost no support for disruptive memory technology, yet
• Focus on NUMA (latencies, bandwidth)
• No application model, no prediction support

Precise system-level models are crucial for the optimized use of disruptive memory technology!

Requirements / Decisions to Make

Mapping decision
- Data to memory region
- Control flows to CPU or other processing elements (NMC, PIM)

Optimization goals
- Performance, energy, WCET, memory, aging, ...

Input data
- HW component properties (performance, energy consumption, ...)
- Application profiles (R/W patterns, atomicity requirements, ...)
- System (bottlenecks, e.g. interconnect and PCIe, parallelism, concurrency, ...)

References


Marcel Köppen and Olaf Spinczyk  {marcel.koeppen, olaf}@uos.de
Embedded Software Systems Group  Computer Science Institute  Osnabrück University  ess.csuos.de