Energy-Efficient Hybrid DRAM/NVM Main Memory

Ahmad Hassan, SAP Labs UK | Hans Vandierendonck, QUB | Dimitrios S. Nikolopoulos, QUB

Overview

DRAM-based main memory is a major consumer of energy

- Nowadays more data is kept in main memory for fast processing
- Large DRAM increases the Total Cost of Ownership (TCO)

DRAM Technology Limitations:

- Power – DRAM requires continuous leakage and refresh power
- Scaling – It is challenging to scale DRAM cells below 22nm

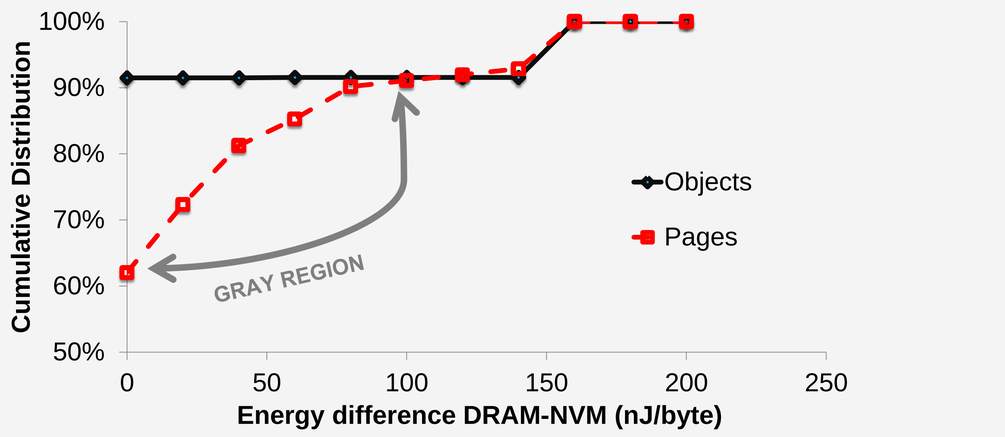
NVM Technology to Rescue:

- NVM consumes zero refresh power and offers high density
- NVM has higher latency and dynamic energy than DRAM

This work proposes energy-efficient hybrid DRAM/NVM system and data management policies at the granularity of application objects.

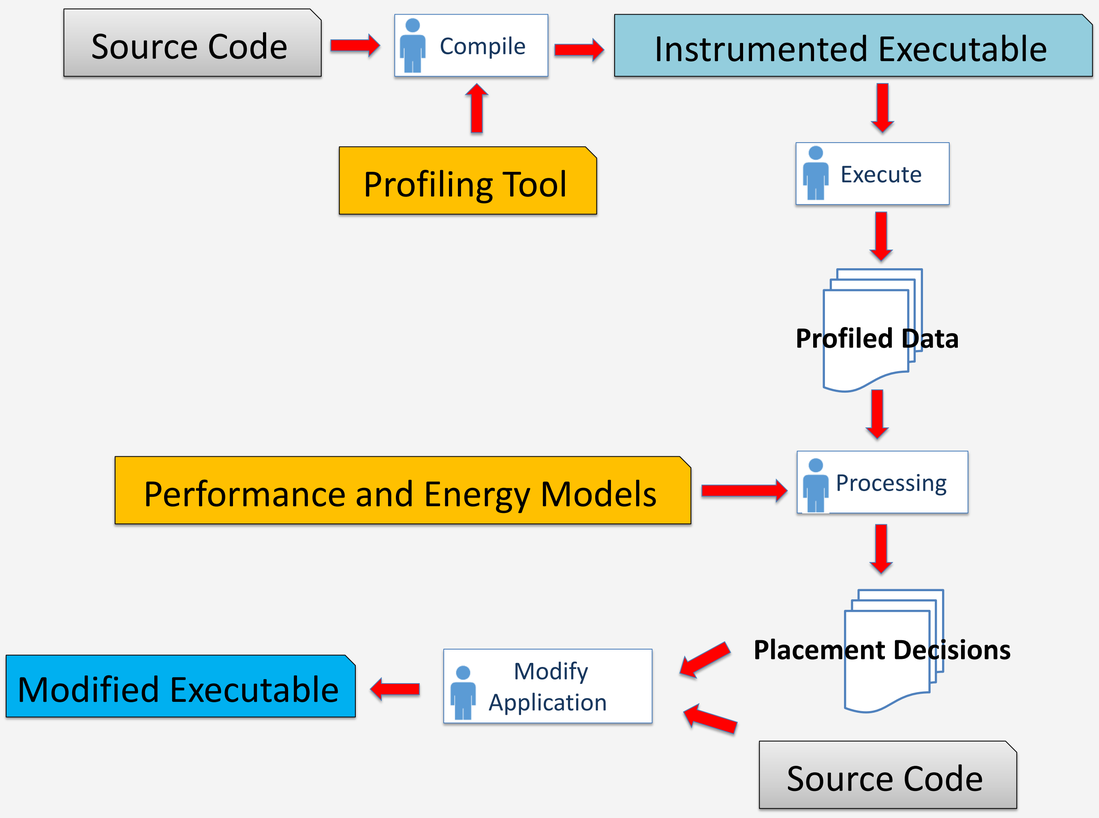
Why Application-level Management ?

For an energy-efficient hybrid memory system, application objects offer more precise granularity of data than OS-level pages.

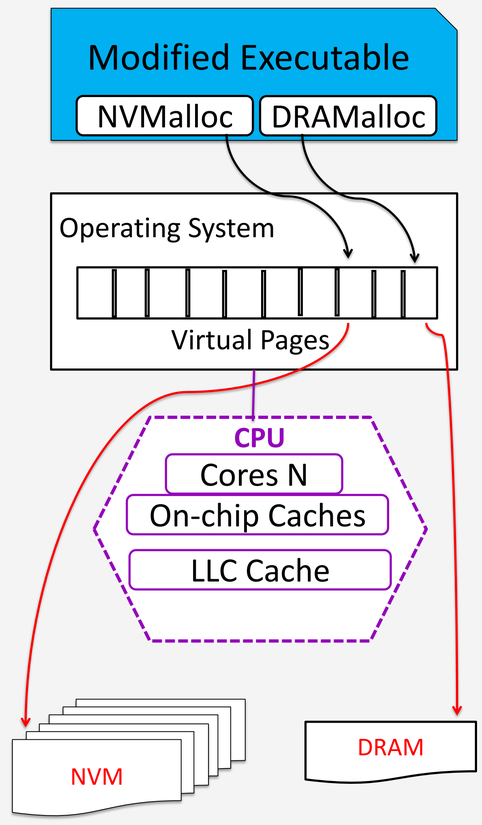


Histogram of DRAM energy minus NVM energy for objects and pages of the jpeg benchmark.

Object Placement Methodology



System Design



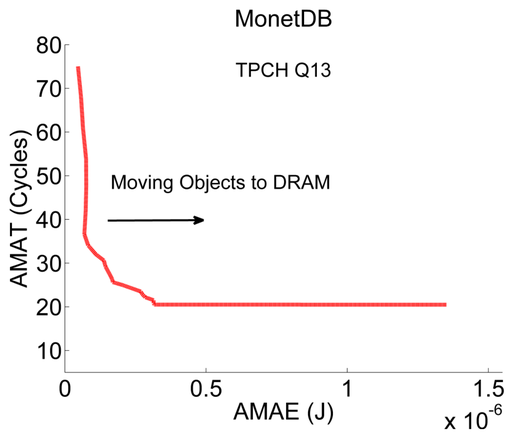
Analytical Models for Objects

Performance Model

$$AMAT_{DRAM} = \mu_r L_r + (1 - \mu_r)\, L_{LLC}$$

Energy Model

$$AMAE_{DRAM} = \mu_r E_r + \mu_w E_w + \phi S\, P_{DRAM} T$$



Hybrid Memory API

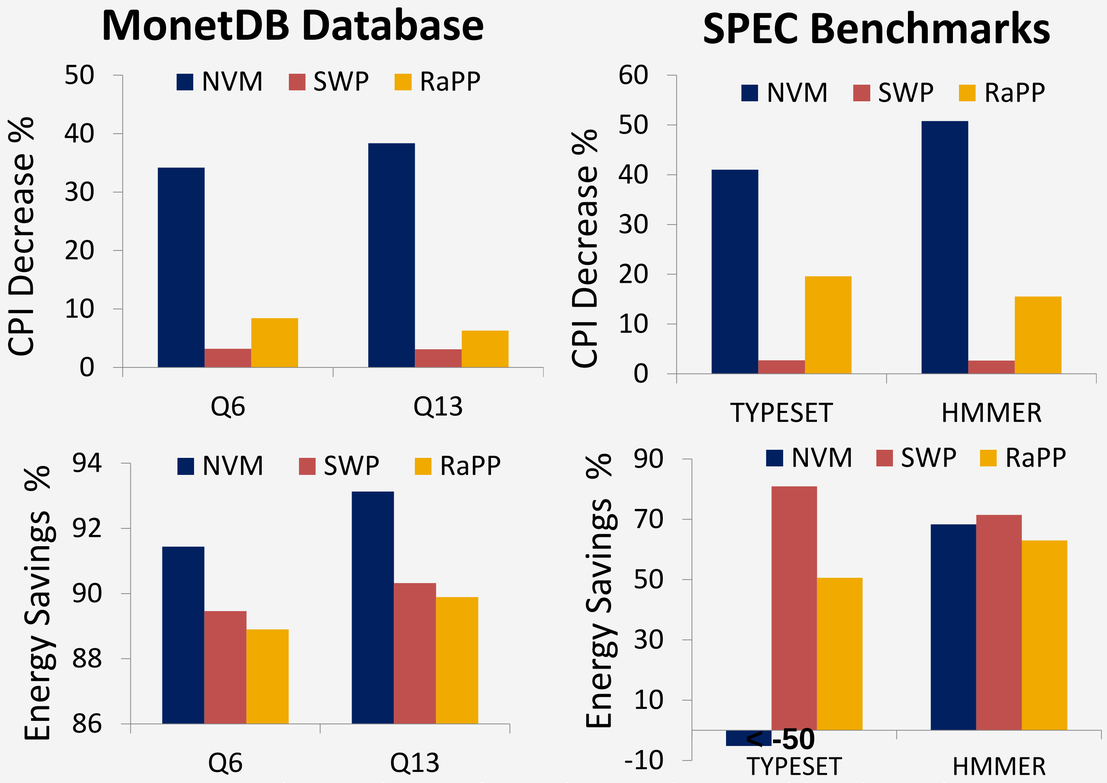
```
void * nvmalloc (size_t )
void * dramalloc (size_t )
void free (void *)
void * mmap (..., int MFLAG)
```

Set MFLAG to 1 for DRAM placement
Set MFLAG to 0 for NVM placement

Results

Hybrid memory simulation on GEM5

MonetDB Database **SPEC Benchmarks**



NVM: NVM only system | SWP: Software placement | RaPP: Rank based Page Placement

Conclusion & Ongoing Work

- NVM, as a main memory technology, is inevitable for meeting main memory capacity demands of emerging applications.
- For the workloads we studied, object placement saves more energy than state-of-the-art OS-level approaches on hybrid memory.
- Hybrid DRAM/NVM memory gives more energy savings than DRAM based system as the size of main memory grows.
- Ongoing work on scientific High Performance Computing (HPC) workloads to develop models for saving energy on hybrid memory.
- Ongoing work on dynamic data placement approaches to derive data placement decisions on hybrid memory.

Patents

- Analytical models and techniques for Software-Managed Energy-Efficient Hybrid DRAM/NVM Main Memory. *ID 83059367.*
- A novel technique to allocate memory on multiple types of main memory technologies from software application layer. ID *83093314.*

Ahmad Hassan https://uk.linkedin.com/in/ahmadhassan | Hans Vandierendonck http://www.cs.qub.ac.uk/~H.Vandierendonck | Dimitrios S. Nikolopoulos http://www.cs.qub.ac.uk/~D.Nikolopoulos