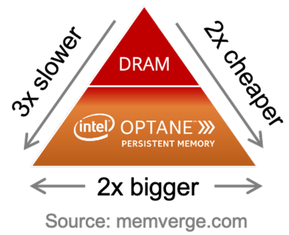
Machine Learning Augmented Hybrid Memory Management

Thaleia Dimitra Doudali thdoudali@gatech.edu - **Advisor: Ada Gavrilovska** ada@cc.gatech.edu

1. Problem Space

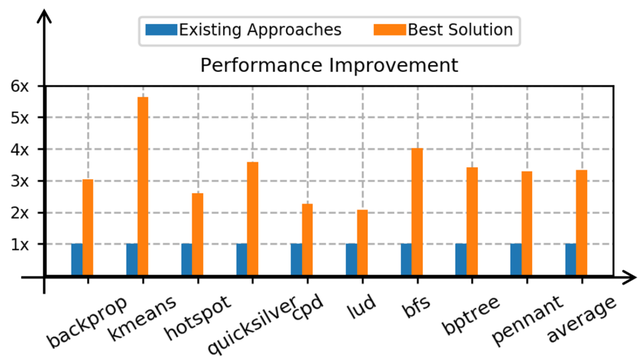


Source: memverge.com

1. Hybrid Memories: New platforms accelerate analytics and simulations of ever exploding data sizes, by incorporating new memory technologies.

2. Complexity Increases: The difference in access speeds and characteristics is more distinct, compared to traditional DRAM-only systems.

3. Traditional Approaches Break: Assumptions and heuristics, now break, resulting in significant loss in performance and efficiency.



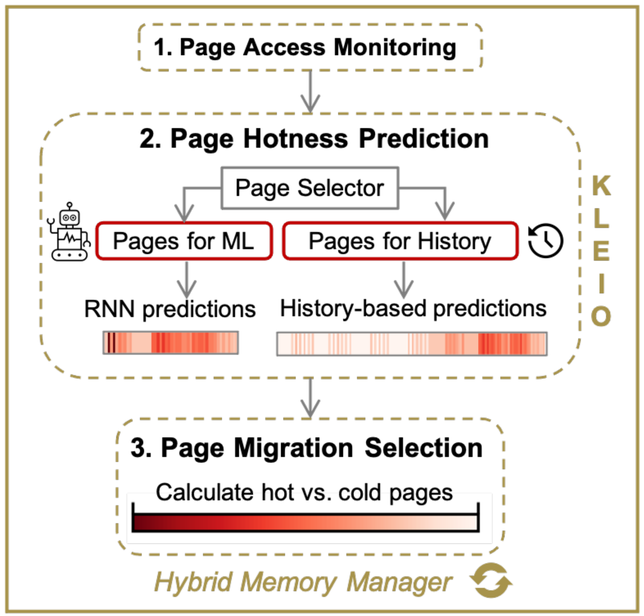
Funding: This thesis was partially supported by NSF award SPX-1822972, the DOE ECP project on Simple Interfaces for Complex Memories (SICM) and the DOE SSIO Unity project.

2. Thesis Goal

Build system-level mechanisms and insights to augment hybrid memory management with Machine Learning (ML).

Thesis Questions / Challenges

- Where to use ML?
- Which ML method?
- What to predict?
- Which data needs ML?
- How to configure the ML?
- How to reduce the ML overheads?
- How to maximize performance?
- How to maximize resource efficiency?



3. Thesis Contributions

Recurrent Neural Networks (RNNs).
1 RNN per page, for a small page subset.

Kleio @ HPDC '19
ML-based Page Scheduling
(best paper award finalist)

Group pages with identical patterns.
1 RNN per page group.

Coeus
Page Grouping

Synchronize data movement periods with data reuse times.

Cori @ IPDPS '21
Frequency Tuner

Use as little DRAM as needed by the actively hot memory footprint.

Mnemo @ HPBDC '19
Memory Cost Efficiency

Share DRAM by merging data objects across co-located workloads.

CoMerge @ MEMSYS '17
Hybrid Memory Sharing