1. Problem Statement

How to minimize the application performance slowdown relative to ‘all-data-in-DRAM’?

2. Motivational Observations

1. Not all applications are slowed down, when executing over NVM.
2. Not all application data objects minimize the slowdown, when allocated in DRAM.

Insight: Need to identify the critical to performance data objects.

3. Contributions

- **CoMerge**: How to share heterogeneous memories across workloads, to maximize utilization and performance. *(MEMSYS '17)*

- **Mnemo**: How to size heterogeneous memories to achieve desired cost-performance tradeoff. *(Under Submission)*

4. CoMerge Solution

![Diagram showing CoMerge solution]

Equal Split  unused  CoMerge

- xsbench  clomp  stream

Higher utilization & Less slowdown

Higher Latency
Lower Bandwidth

All work is part of the projects:

- ECP SICM (Software Interface to Complex Memories)
- SSIO Unity (Unified Memory and Storage Space)