**Investigating Hardware Caches for Terabyte-scale NVDIMMs**

Julian T. Angeles, Mark Hildebrand, Venkatesh Akella, Jason Lowe-Power

**Intro**
- Heterogenous NVRAM servers
  - DRAM acts as cache in 2LM to mask latency of NVRAM
- Performance and limitations of the DRAM cache at scale?

**Methods**
- Evaluate real DRAM cache
  - Custom microbenchmarks to identify characteristics
- Apply insights to real world workload
  - Graphs that fit and exceed capacity

**Results**

**Insights** - High Access Amplification, High miss-rate -> poor bandwidth

**Performance** - 50% bandwidth drop, 3-5x data movement

**Discussion**
- Software Managed Memory
  - Limited by core use and specialization
- New design that is implicit and high performance?