

L1 Memory Systems

Joyen Benitto

State of the Main Memory System

- DRAM and memory controller , as we design them today are and will be unable to satisfy all requirements.
- Memory and Storage merging (some emerging technologies allow you to do this) (*We need to rethink the main memory system*)

Trends affecting Main Memory

1. Need more main memory capacity, bandwidth and QoS increasing
2. Main memory energy/power is a key system design concern
3. DRAM technology scaling is ending

1. Need for main memory capacity, bandwidth and QoS

- Data-intensive application pushing bounds of memory (very hungry for data)
- Multi-core: increasing number of cores and agents
- Consolidation: cloud computing, GPUs, Mobile and heterogeneity

This is a 2009 study that shows how the number of cores is increasing and then the DRAM DIMM capacity is doubling every 3 years but yes the number of cores eventually saturated so did DIMM capacity at of 2025, but whats happening is folks using larger cache sizes, but yes in-case of GPUs the core count seems to be increasing rapidly.

One thing you could do to succeed in research is analyze trends and figure out a way to fix those dropping curves

The above is a 2017 study conducted by safari group ETHZ. The capacity had a significant growth over the years about 128x, this is majorly the goal of a memory design engineer to increase bandwidth and at the same time increase capacity by keeping the latency low.

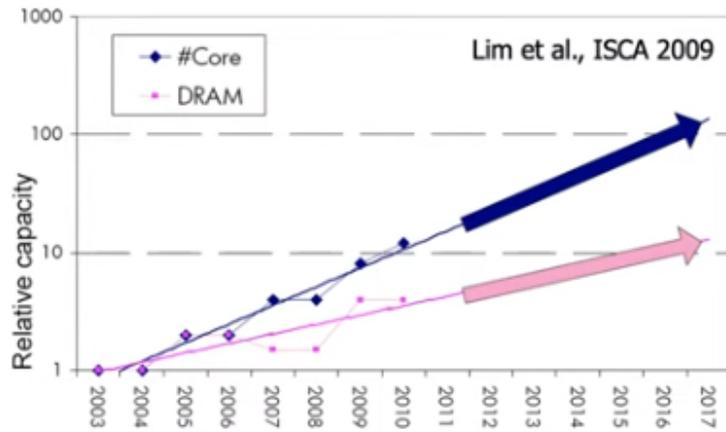


Figure 1: isca hp and umich study dram vs core

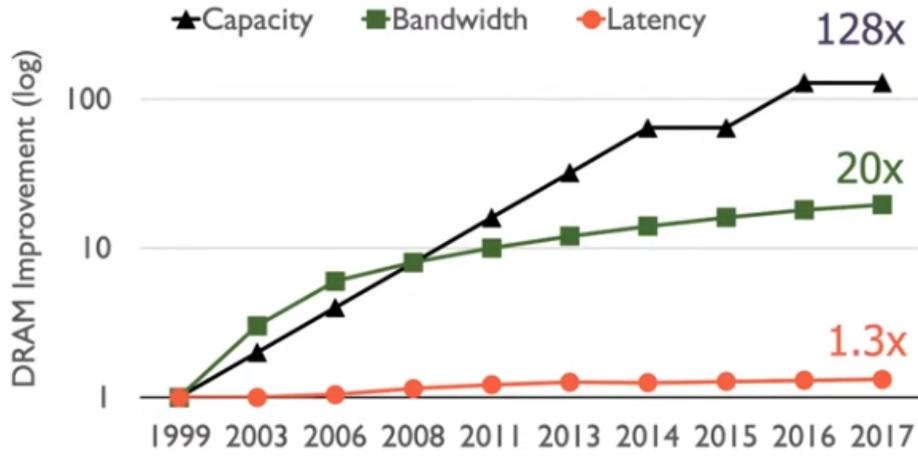


Figure 2: Latency vs badwidth vs capacity, 2017

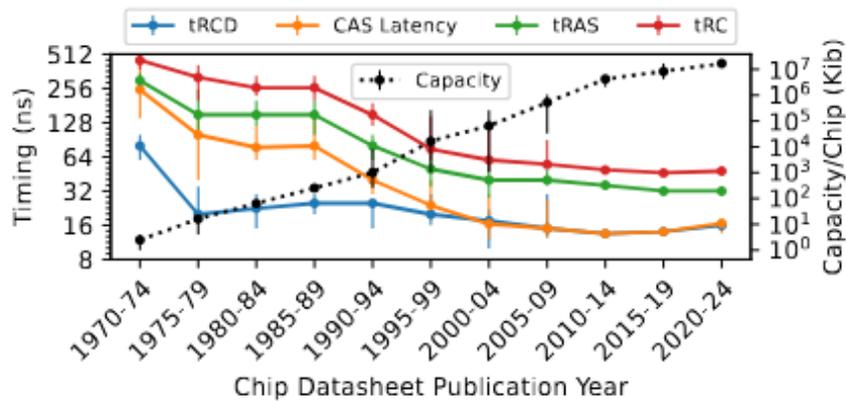
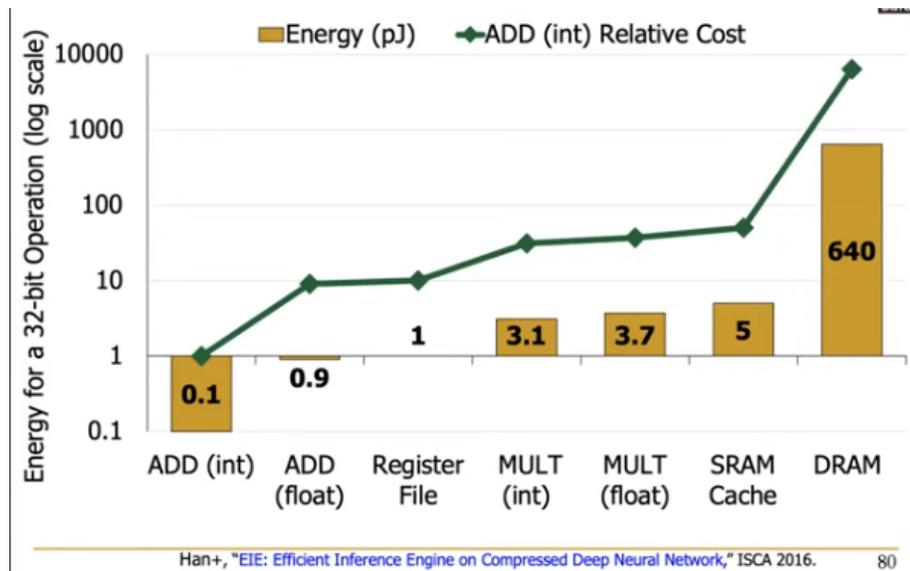


Figure 1: Semi-log plot showing the evolution of key DRAM access timings (left) and per-chip storage capacity (right) across each 5-year period of time.

Figure 3: <https://arxiv.org/pdf/2204.10378>

2.Data Movement vs Computation Energy



The above study shows how much it energy we consume in a DRAM operation (about 6400x) compared to an ADD but this is only the DRAM with storage it can be 10x more

3. DRAM Technology Scaling is ending

Scaling has provided many benefits but the higher capacity(density), lower cost, lower energy.

- Difficulties in scaling create robustness problems
- https://people.inf.ethz.ch/omutlu/pub/memory-scaling__imw13.pdf