

6 Review: “Tiered-Latency DRAM: A Low Latency and Low Cost DRAM Architecture” (6)

Summary

The paper discusses the issue of DRAM speed as a major drag on overall performance of a computer system. It elaborates that the main source of DRAM latency is the bitline capacitance which is proportional to the bitline length. This issue is commonly addressed by DRAM producers as a tradeoff between speed and cost. The paper offers a solution taking advantage of both the speed of short bitlines and the cost of long bitlines.

- The paper explains the fundamental tradeoff in conventional DRAMs. A shorter bitline means a lower bitline capacitance which is easier to drive to a desired voltage. But a shorter bitline means that more bitlines are needed to provide the same amount of memory. As each bitline needs a sense-amplifier this increases the cost of the DRAM considerably. For a longer bitline the exact opposite is the case.
- The paper proposes a solution that utilizes a gate in the middle of the bitline to split it into two sections. A close and fast section and a far and slow section. Therefore allowing both the benefits of short and long bitlines in conventional designs at small cost.
- The paper establishes the issue of managing the now partitioned memory. The easiest but also worst performing option is to use it like normal DRAM that has a faster and a slower memory partition. But performance gain can be increased in two further ways.
 - The fast memory can be handled as a cache only visible to the DRAM for the slow memory allowing a better on average performance but decreasing the amount of memory on the DRAM. The paper proposes several caching strategies to optimally use the memory and maximise performance.
 - The fast memory can be made visible to OS allowing the OS to implement a tailored solution by deciding what data to store in the fast and what data to store in the slow memory.
- The paper evaluates the performance of the TL-DRAM by performing extensive simulations of

the system.

Strengths

The paper proposes a novel approach to DRAM implementation. The proposed system achieves very high performance on minimal cost, or at least area gain, on the DRAM chip. The paper very thoroughly simulates the system and provides a comprehensive analysis of the proposed implementation. The paper addresses an important issue in DRAM design.

Weaknesses

The paper's cost calculations are insufficient and biased. The paper assumes that the cost of the DRAM chip is linear in its size and then compares these metrics to the price of conventional DDR3-RAM. This is not a fair comparison as the cost of DDR3-RAM is lowered by its high production volumes. The TL-DRAM does not achieve the performance of fast DRAM chips and I do not see a valid space for this design on the market as it cannot realistically beat the cheap implementations in cost and does not outperform the fast implementations. The paper does not provide a physical implementation of the proposed system and therefore all results have to be taken with a grain of salt. To optimally use the TL-DRAM, i.e. showing it to the OS, the OS needs to be redesigned to be able to handle it which incurs small interoperability if it is not widely adopted.

My POV

The issue addressed in the paper is an important one and the solution proposed is both innovative and effective but I cannot see this system being used in a commercial system in the near future. I find the cost estimation very biased and therefore disagree with the applicability claims of the paper. If the paper would have provided a physical implementation then the TL-DRAM could have been tested in a real context and the results would have more weight. Further the cost and viability could be assessed more fairly.

Takeaways

- DRAM-latency is an important issue that needs to be addressed.
- The proposed TL-DRAM is a smart implementation using preexisting structures with minimal modifications to achieve remarkable results.

- The real-world applicability of this paper is flawed at best as it proposes a theoretical concept and assumes very optimistic cost estimations.