Proposal of Topic
Memory Power Tracing on Hybrid Memory Cube

By Jie Li

Objective. The objective of this research is to investigate the impacts of the memory accesses on the power consumptions in an advanced memory devices named hybrid memory cube(HMC).

Background and Motivation. The modern microprocessor architecture mainly utilizes multi-level data caches as a primary optimization to reduce the memory access latency and power consumptions as well as increase the perceived bandwidth from an application. This mechanism works well with significant memory reuse or linear memory access patterns. However, the irregular or non-deterministic pattern of memory access will result in the high miss rate in both cache and Transaction Look-aside Buffer(TLB), which subsequently leads to the high memory access latency and power consumptions. In response to these data-intensive applications, Leidel et al. [1] developed the GoblinCore-64(GC64) micro architecture with using a large degree of hardware-managed concurrency coupled to a high bandwidth memory subsystem. Wang et al. [2] proposed the Concurrent Dynamic Memory Coalescing within the GC64 architecture explicitly designed to exploit memory performance from irregular memory access patterns. In the aforementioned works, the performance is measured by the relative efficiency, of which is calculated by dividing the number of coalesced requests by the total number of input requests. However, the memory power consumptions savings benefited by the memory system optimization are not measured. In this research, we will investigate the memory power consumptions savings to evaluate the efficacy of the proposed design and algorithms.

Challenges and Proposed Approach. HMC-Sim[3] will be used as our testing tool, this simulation framework is developed specifically for the HMC specification to track the memory access latency and power consumptions of HMC based on the provided memory traces, it also provides the visualization tools that automatically plots the figures and generates the videos. The evaluations have confirmed that HMC-Sim can provide insightful guidance in designing and developing highly efficient systems, algorithms and applications, considering the next-generation three-dimensional stacked memory devices. We will use HMC-Sim to test the memory traces with/without the memory coalescing, just as the same test environment as [2], the same set of applications that represent dense (linear) memory request, random memory requests and real application workloads are used to test the power consumptions. In addition to the relative efficiency, the power savings measurement provides a more comprehensive evaluation to the proposed design.

Deliverables and Timeline. A major deliverable of this research is a report in which presents the memory power consumptions savings using *address partitional algorithm*(APA) and *work partitioned algorithm*(WPA) for the target applications. The timeline of the proposed research is as follows. **Week 1** (10.29-11.2): Task 1, read related papers. **Week 2**(11.5-11.9): Task 2, play with HMC-Sim. **Week 3**(11.12-11.16) & **Week 4**(11.19-11.23): Task 3, run the HMC-Sim and test different cases.**Week** 5(11.26-11.30) & **Week 6**(12.3-12.7): Task 4, summarize research results and write final report.

References

[1] John D. Leidel, Xi Wang, and Yong Chen. GoblinCore-64: Architectural Specification. Technical report, Texas Tech University, September 2015.

[2] Xi Wang, John D. Leidel, and Yong Chen. Concurrent Dynamic Memory Coalescing on GoblinCore-64 Architecture. In *Proceedings of the Second International Symposium on Memory Systems(MEMSYS), 2016*, pages 177-187.

[3]John D Leidel and Yong Chen. HMC-Sim: A Simulation Framework for Hybrid Memory Cube Devices. Parallel Processing Letters, 24(04):1442002, 2014.