

Coherence Protocol Optimization For Both Private and Shared Data

Nipun Katyal
nkatyal@andrew.cmu.edu

Chen Fu Hsu
chenfuh@andrew.cmu.edu

March 2024

URL :

1 Summary

We propose an optimization for cache coherence protocol for private and shared data inspired from POPS [1]. The optimization aims to delegate meta-data storage to only one of the caches (either L1 or L2), hence increasing effective cache capacity. We aim to utilize trace files to analyze the performance the proposed optimization and provide insightful metrics regarding performance and utilization.

2 Background

Chip multiprocessors tend to organize their caches in an increasing order of sizes, with $L_1 - L_i$ caches being private to a processor and L_{i+1} being shared by all the processors. This arrangement reduces the latency to access DRAM or lower level memory but requires significant work to keep the caches in the correct state when multiple processors access a common resource. Cache coherence is critical to maintain the correctness of the program and to allow the user to carefully orchestrate their programs to produce an expected result. Popular cache coherence protocols like MESIF and MOESI are based on snooping and require the processor to actively listen to the messages on the interconnect, on the other hand directory based protocols need to store additional information about the processors with whom they are sharing memory locations. We propose an optimization that is inspired from POPS [1] to find a common ground.

3 The Challenge

- Understanding the POPs protocol presented on the paper.

- Understanding the cache framework provided by Professor Railing because we are making modifications on top of it to implement our design.
- The complexity of implementing the POPs protocol since we need to add some hardware components, such as bloom filters and predictor tables, to the existing cache architecture.
- Performance and analysis on the optimized cache and base cache.

4 Resources

- The paper about using POPs protocol to optimize cache coherence.
- We'll start building our implementation on top of the simulator provided.
- We are using both GHC machines and PSC machines.
- We plan to scale our design from the provided simulator to larger simulator like gem5 if we have extra time.

5 Goals and Deliverables

5.1 Plan to Achieve

- Simulate the POPS inspired protocol on a simulator (preferably from 15-346) for multiprocessor NUMA caches (private L1 and shared L2).
- The simulator will provide the user with the functionality to set cache parameters and processor count.
- The simulator will provide the ability to profile the proposed system against MESIF and MOESI protocols and showcase relevant metrics like cache hits/misses, latency and interconnect traffic load.

5.2 Hope to Achieve

- Profile the proposed protocol against other coherence protocols such as directory based.
- The simulator will provide the ability to use a different interconnect.

5.3 Deliverables

- Analysis of MESIF/MOESI protocols against the proposed protocol.
- Ablation study of various tuneable parameters such as cache sizes, processor count.
- Performance of the protocol under trace-driven workloads.

6 Platform Choice

- The main developing environment will be on GHC machine; however, we might use PSC machine to do some in depth analysis.
- We will build our design on top of the cache framework that Professor Railing provided.
- We will use C++ as our programming language for the design.
- If we have extra time, we will build our design on top of gem5 simulator.

7 Schedule

Week 1	<ul style="list-style-type: none">• Study the paper of cache optimization using POPS and study the cache simulator provided by professors
Week 2	<ul style="list-style-type: none">• Start implement our design
Week 3	<ul style="list-style-type: none">• Finish implementation and verification on our design
Week 4	<ul style="list-style-type: none">• Perform analysis and gather data for both our optimized design and base design
Week 5	<ul style="list-style-type: none">• Work on the report and potentially scale our design from the simulator professor provided to gem5

References

- [1] Hemayet Hossain, Sandhya Dwarkadas, and Michael C. Huang. “POPS: Coherence Protocol Optimization for Both Private and Shared Data”. In: *2011 International Conference on Parallel Architectures and Compilation Techniques*. 2011, pp. 45–55. DOI: 10.1109/PACT.2011.11.