ELECTRICAL & COMPUTER ENGINEERING
TEXAS A&M UNIVERSITY

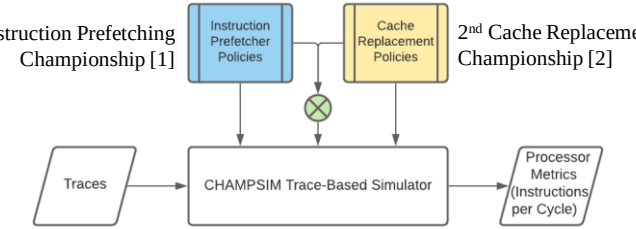
Combining Instruction Prefetchers and Cache Replacement Policies in the L1I Cache

By: Alex Christian[1] and David Chapa[2]

Purpose and Background

- Processor performance is hindered by long DRAM latencies, so caches are critical to processor performance.
  - Cache misses cause accesses to lower-level caches and memory, stalling the processor and lowering **Instructions Per Cycle (IPC)**
- **Instruction Prefetching –** Predicting future instruction accesses and prefilling them in the cache to avoid miss latency
- **Cache Replacement Policies –** Decides what data will be kept in the cache and what will be evicted when new data must be added to the cache.
- Our research attempts to explore the possible benefits of combinations of instruction prefetchers and cache replacement policies in the L1I cache

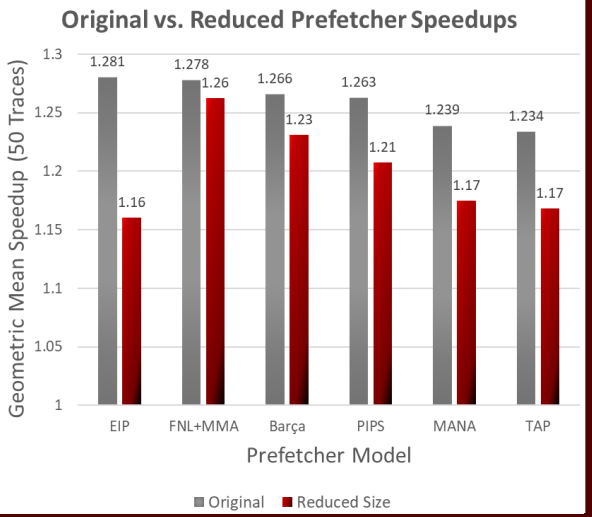
1st Instruction Prefetching Championship [1]



2nd Cache Replacement Championship [2]

Instruction Prefetchers

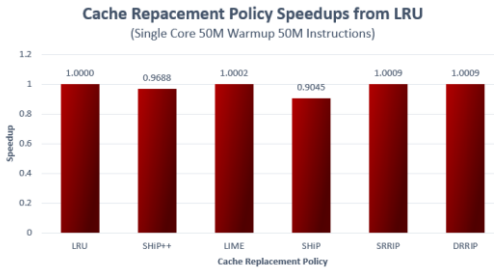
- Instruction prefetching reduces cache misses in a processor's instruction cache
  - Predicts/fetches cache lines that would have missed & stalled processor
  - Increases IPC
- Simulated prefetchers to find their speedups over a processor with no prefetching
- Reduced each prefetcher's overhead to ~32KB and simulated again



Instruction Prefetcher Results: Top three prefetchers were FNL+MMA, Barça, and PIPS

Cache Replacement Policies

- Cache replacement policies decide what data should be evicted from a cache to make room for new data that must be added to the cache.
- Replacement policies are typically used to manage data caches.
  - Used to maintain the L1I cache in this context.



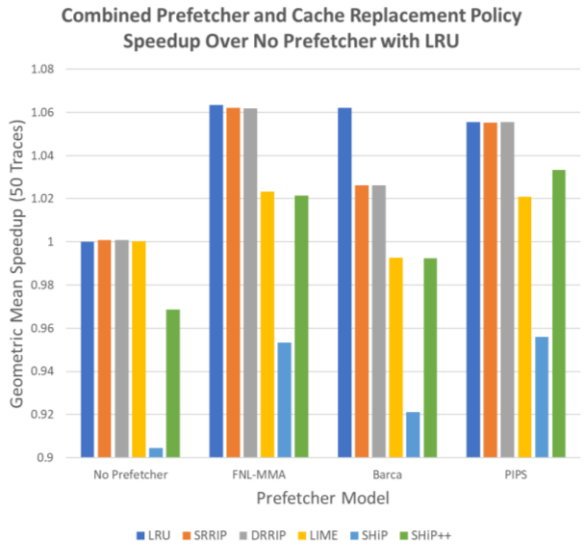
Cache Replacement Results:

- Replacement policies do not show much of an improvement in the L1I.
- Minor speedup from SRRIP, DRRIP, and LIME when compared to the baseline of LRU.
- SHiP++ did not have a speedup despite being the champion of The 2nd Cache Replacement Championship [2].

Combination Testing

- Ported cache replacement policies designed for the LLC to work on the L1I
- Ran simulations with FNL+MMA, Barça, and PIPS combined with each cache replacement policy

Combination Policy Results: LRU combinations performed the best. Naively combining instruction prefetchers and cache replacement policies does not provide any benefit without making modifications. Next, we will explore integrating replacement and prefetching into a combined scheme.



Modifications

```
if(type == PREFETCH){
    if(!hit){
        if(conf < conf0){
            rrpv[set][way] = 3;
        }
        else if(conf < conf1){
            rrpv[set][way] = 2;
        }
        else if(conf < conf2){
            rrpv[set][way] = 1;
        }
        else if(conf >= conf2){
            rrpv[set][way] = 0;
        }
    }
    else{
        if((conf >= conf2) && (rrpv[set][way] > 0)){
            rrpv[set][way]--;
        }
    }
}
else{
    if (hit)
        rrpv[set][way] = 0;
    else
        rrpv[set][way] = maxRRPV-1;
}
```

- **Modifying both Barça and SRRIP**
  - Barça generates confidence values for every prefetch.
  - SRRIP was changed to set the RRPV of each prefetch based on the confidence value.
  - Speedup of 1.000245972.
- Modifications show potential and proof-of-concept.

Incoming Instruction	Confidence Value	Hit/Miss	Way0	Way1	Way2	Way3
data0	90%	Miss	Data: RRPV: 3	Data: RRPV: 3	Data: RRPV: 3	Data: RRPV: 3
data1	30%	Miss	Data: data0 RRPV: 0	Data: RRPV: 3	Data: RRPV: 3	Data: RRPV: 3
data2	67%	Miss	Data: data0 RRPV: 0	Data: data1 RRPV: 2	Data: RRPV: 3	Data: RRPV: 3
data3	45%	Miss	Data: data0 RRPV: 0	Data: data1 RRPV: 2	Data: data2 RRPV: 1	Data: RRPV: 3
data2	67%	Hit	Data: data0 RRPV: 0	Data: data1 RRPV: 2	Data: data2 RRPV: 1	Data:data3 RRPV: 2
data4	10%	Miss	Data: data0 RRPV: 0	Data: data1 RRPV: 2	Data: data2 RRPV: 0	Data:data3 RRPV: 2
data0	90%	Hit	Data: data0 RRPV: 1	Data: data4 RRPV: 2	Data: data2 RRPV: 1	Data:data3 RRPV: 3
data5	80%	Miss	Data: data0 RRPV: 0	Data: data4 RRPV: 2	Data: data2 RRPV: 1	Data:data3 RRPV: 3

Conclusion

- LRU outperformed other replacement policies designed for the LLC
  - LRU has smallest overhead and highest speedup → most suitable for L1I
- Combination policies in the L1I cache with LLC replacement policies did not show significant improvement over baseline LRU with no prefetcher
- Instruction Prefetcher & Cache Replacement Policy communication has potential
  - Our modified Barça-SRRIP results showed a very small speedup over the baseline without communication
  - With algorithm improvement & fine tuning we believe communication between instruction prefetchers and cache replacement policies can improve speedup and help performance with minimal added overhead

References

[1] "The 1st Instruction Prefetching Championship," Ncsu.edu, 2020. https://research.ece.ncsu.edu/ipc/.

[2] "THE 2ND CACHE REPLACEMENT CHAMPIONSHIP – Co-located with ISCA June 2017," Tamu.edu, 2017. https://crc2.ece.tamu.edu/.

Acknowledgements

We would like to thank Dr. Paul V. Gratz for his help throughout our project. We also want to thank all those who contributed to ChampSim and those who maintain the CESG Cluster since both were vital tools in our research.