

OS Interaction with Cache Memories

Dr. Sohum Sohoni

School of Electrical and Computer Engineering

Oklahoma State University



Outline

- Technically 'light' talk- broad audience
- Background terminology and one example of current work
- Wild predictions about the future
 - Many thanks to the National Science Foundation CNS #0720741
 - Any views presented here are my own, and not reflective of the NSF's views and policies



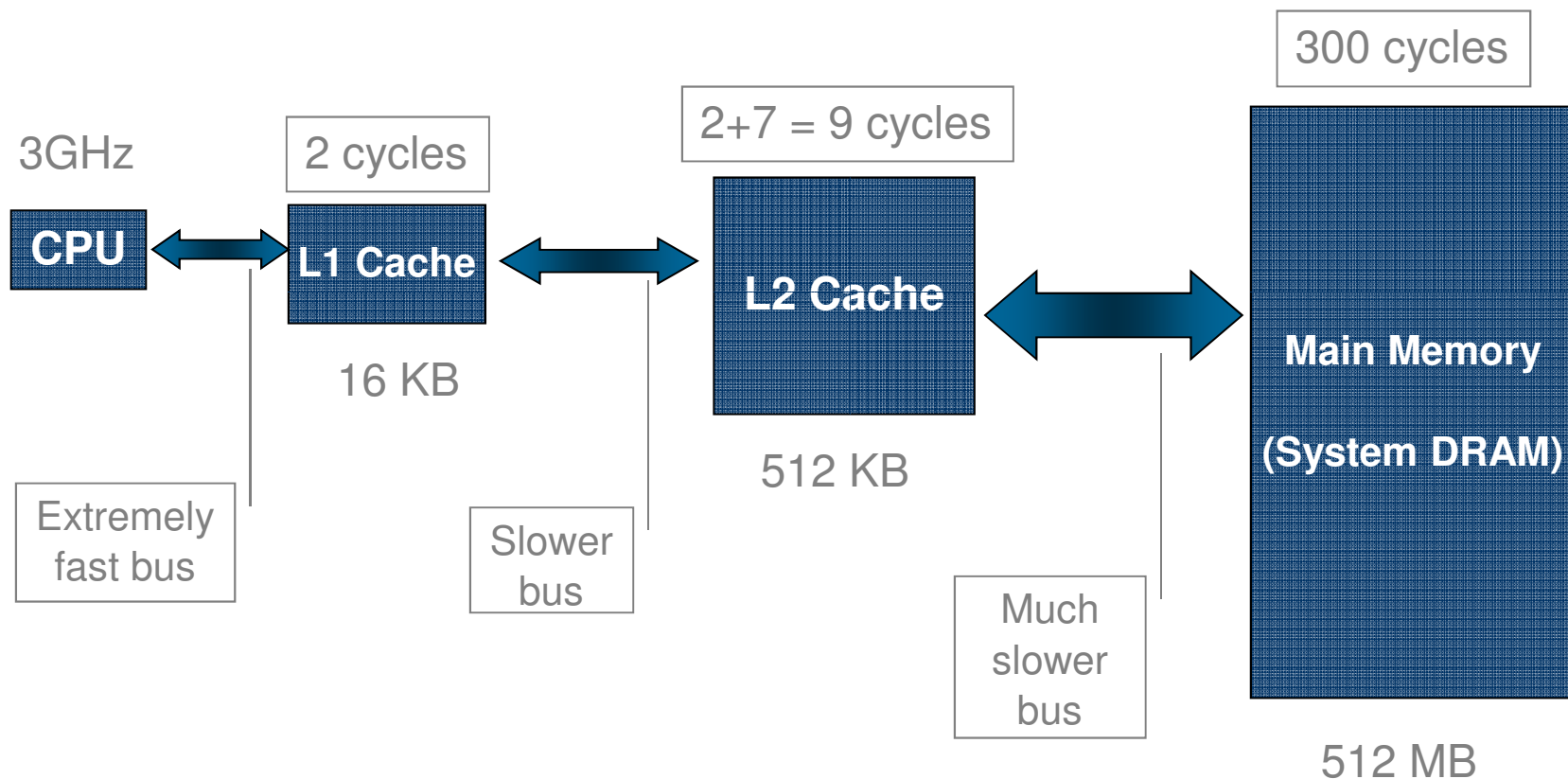
Some Terminology

- CPU Caches
 - Miss rates
 - Locality
 - Prefetching
- Context Switches
 - ‘state’
 - Working set or memory footprint
 - Process queue



Goal of Memory Hierarchy

- Low latency, high bandwidth, high capacity, low cost



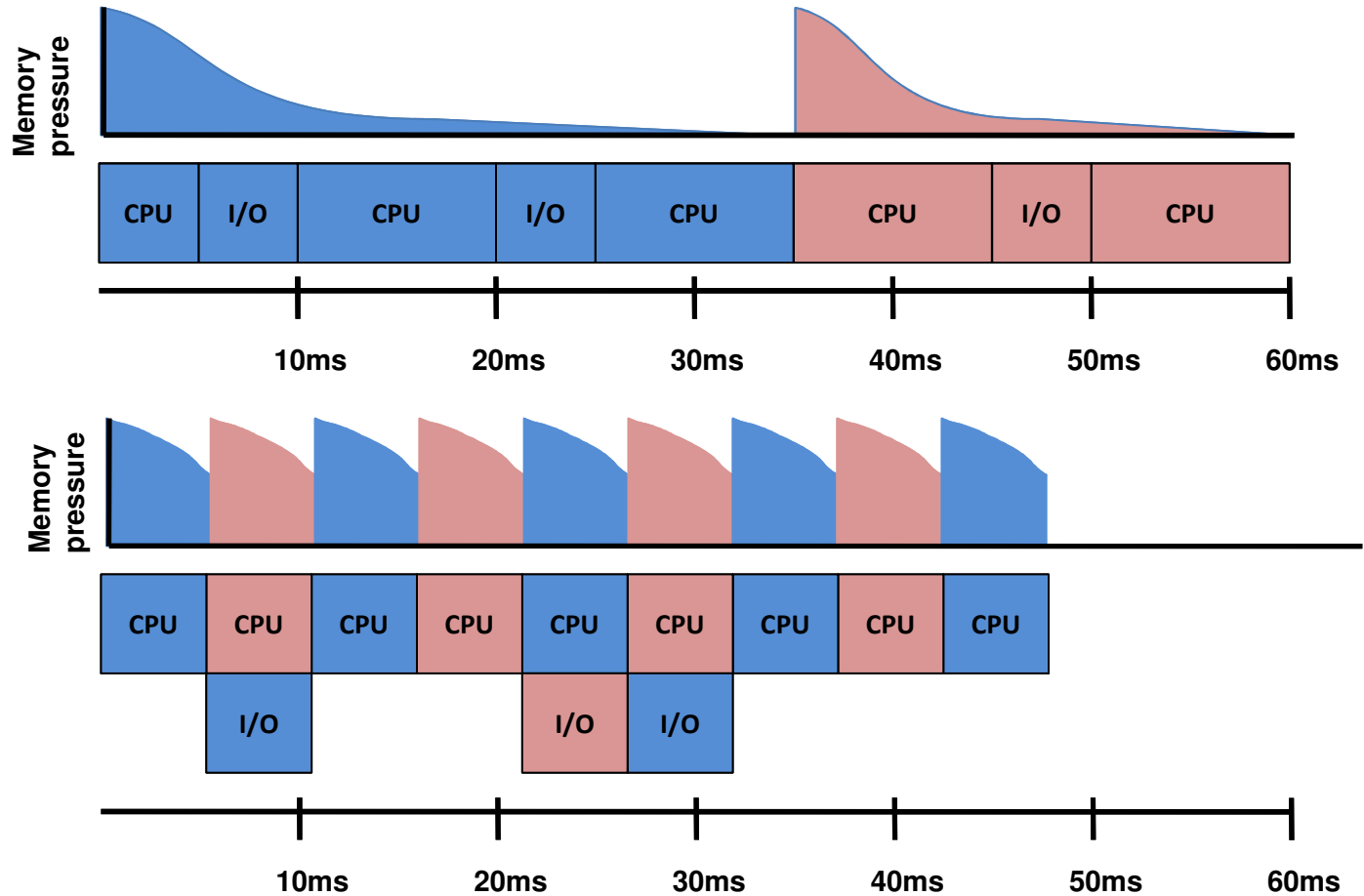
What Happens in a Context Switch

- Current process 'state' is saved
- Scheduler is invoked
- Next process is 'brought in'
- TLB's are flushed
- L1 cache may be flushed
- New process executes for its time slice
- Interrupt, state saved, scheduler ...

- Effective locality gets wiped out

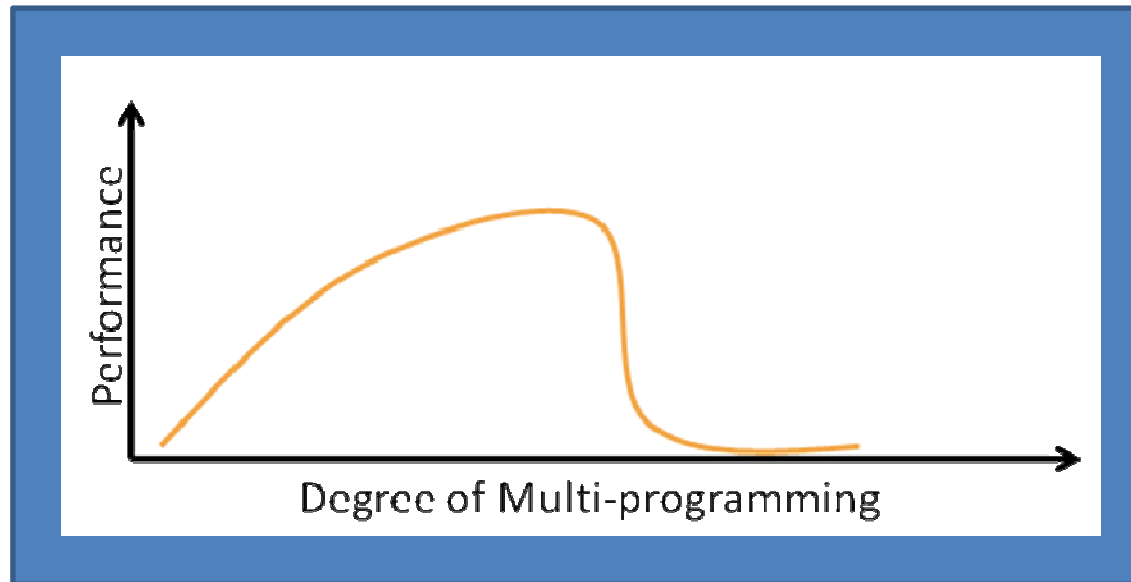


Effect of MultiProgramming



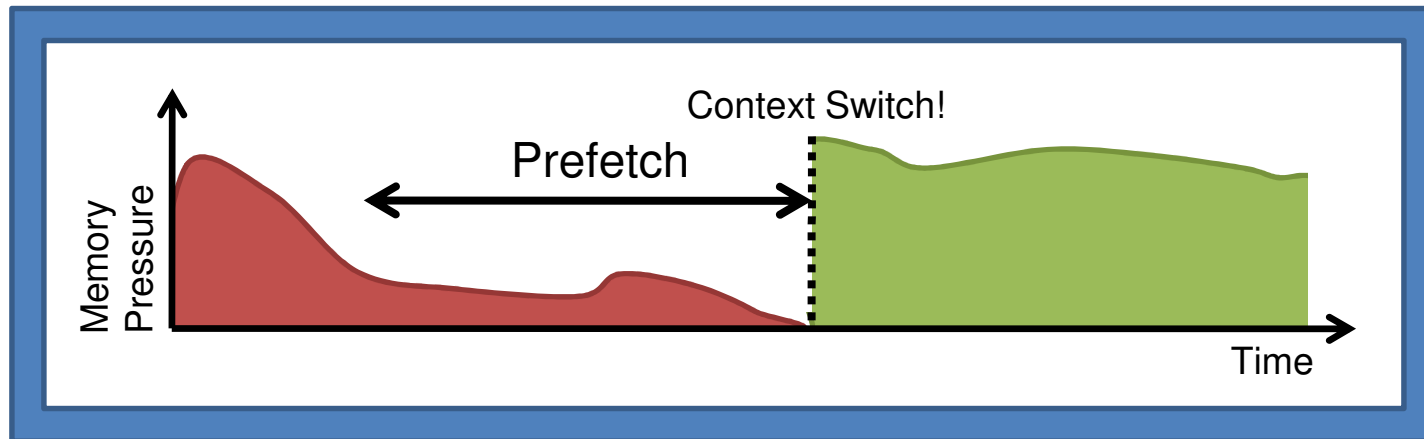
Problem with MultiProgramming

- Increasing multi-programming increases cache miss rates
- Loss of locality of reference
- Diminishing returns from multi-programming
- Eventual thrashing

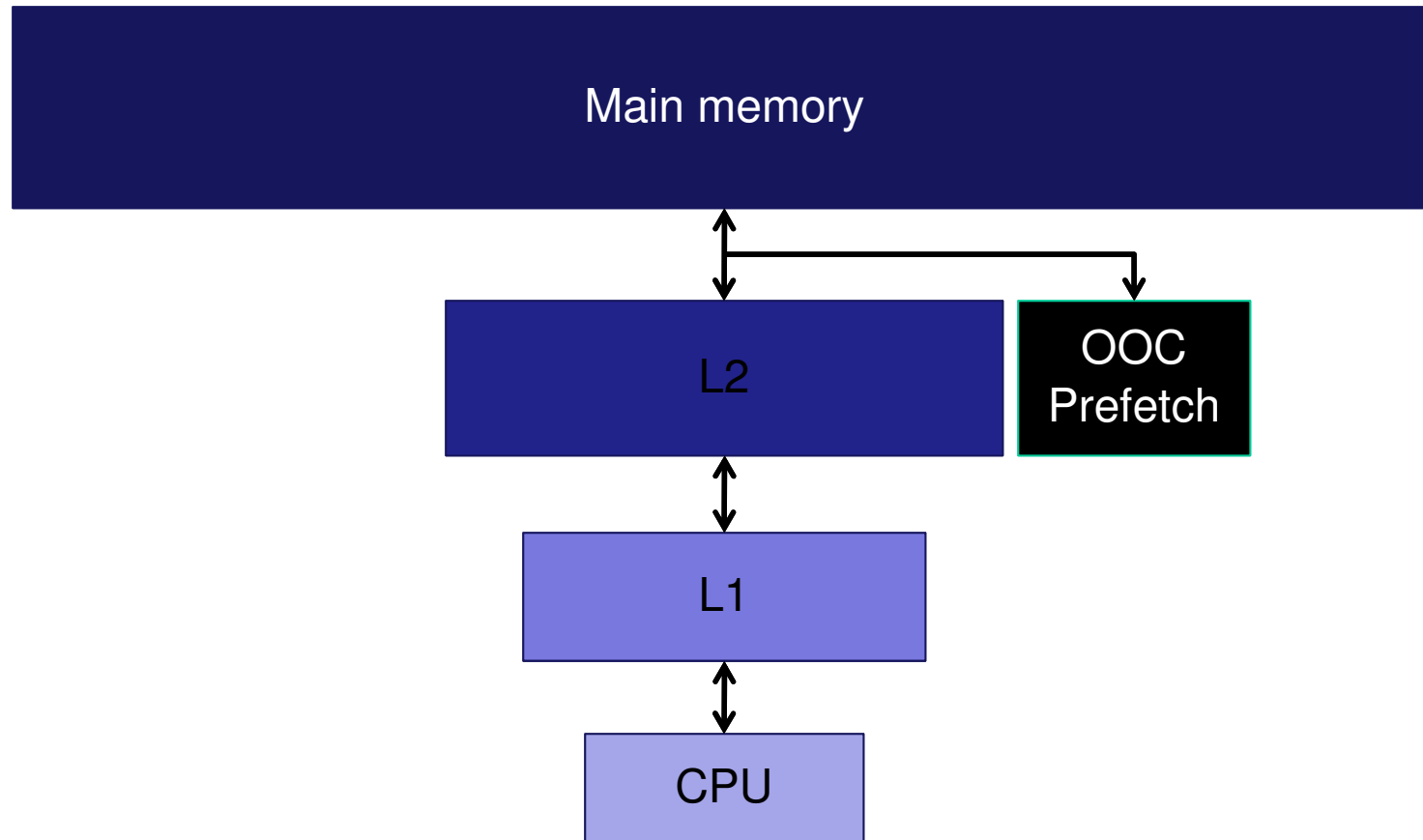


Out of Context Prefetching

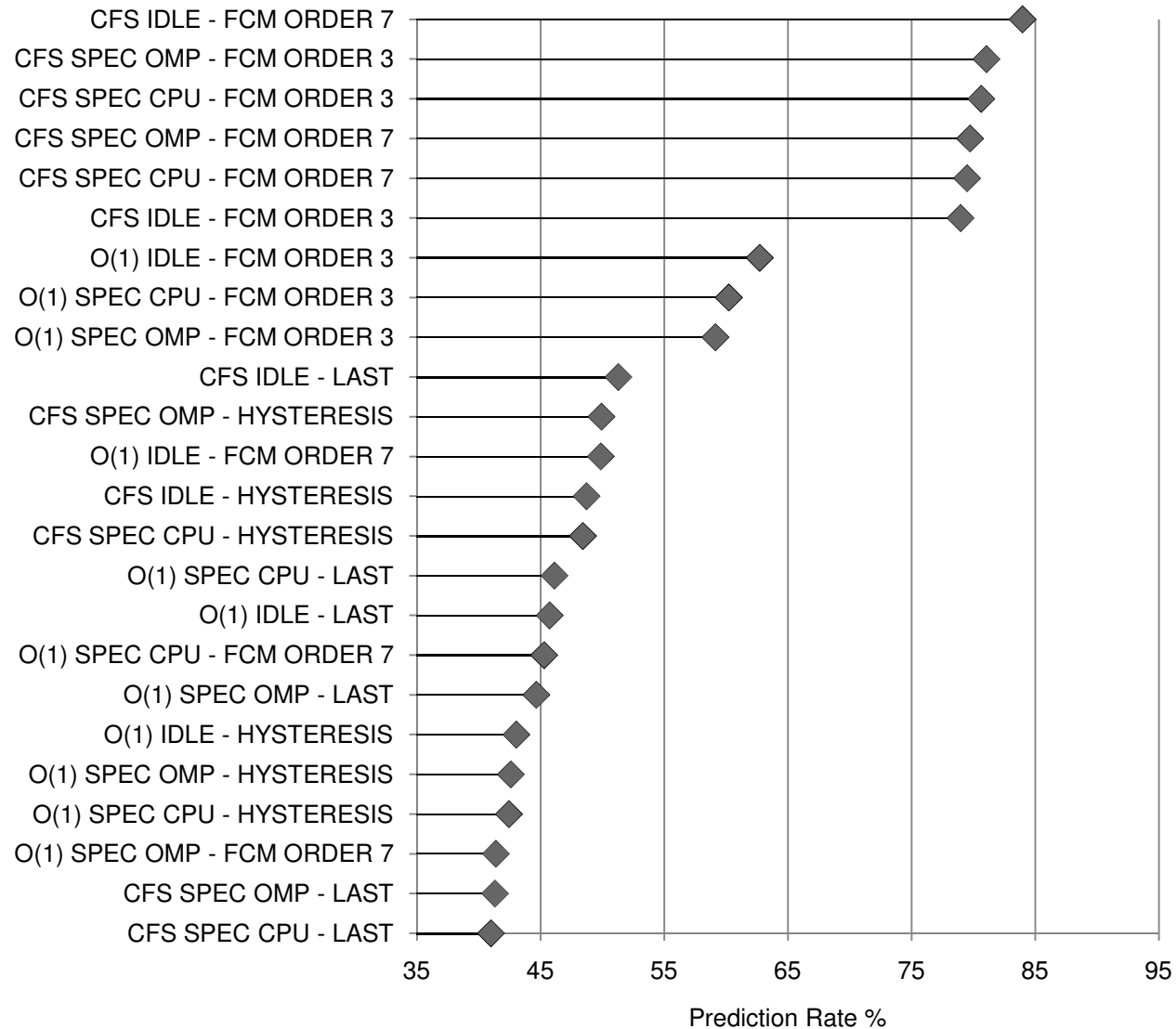
- Reduce the negative effect of multi-programming on CPU cache performance
 - Predict future context switches
 - Prefetch the working set of the next process



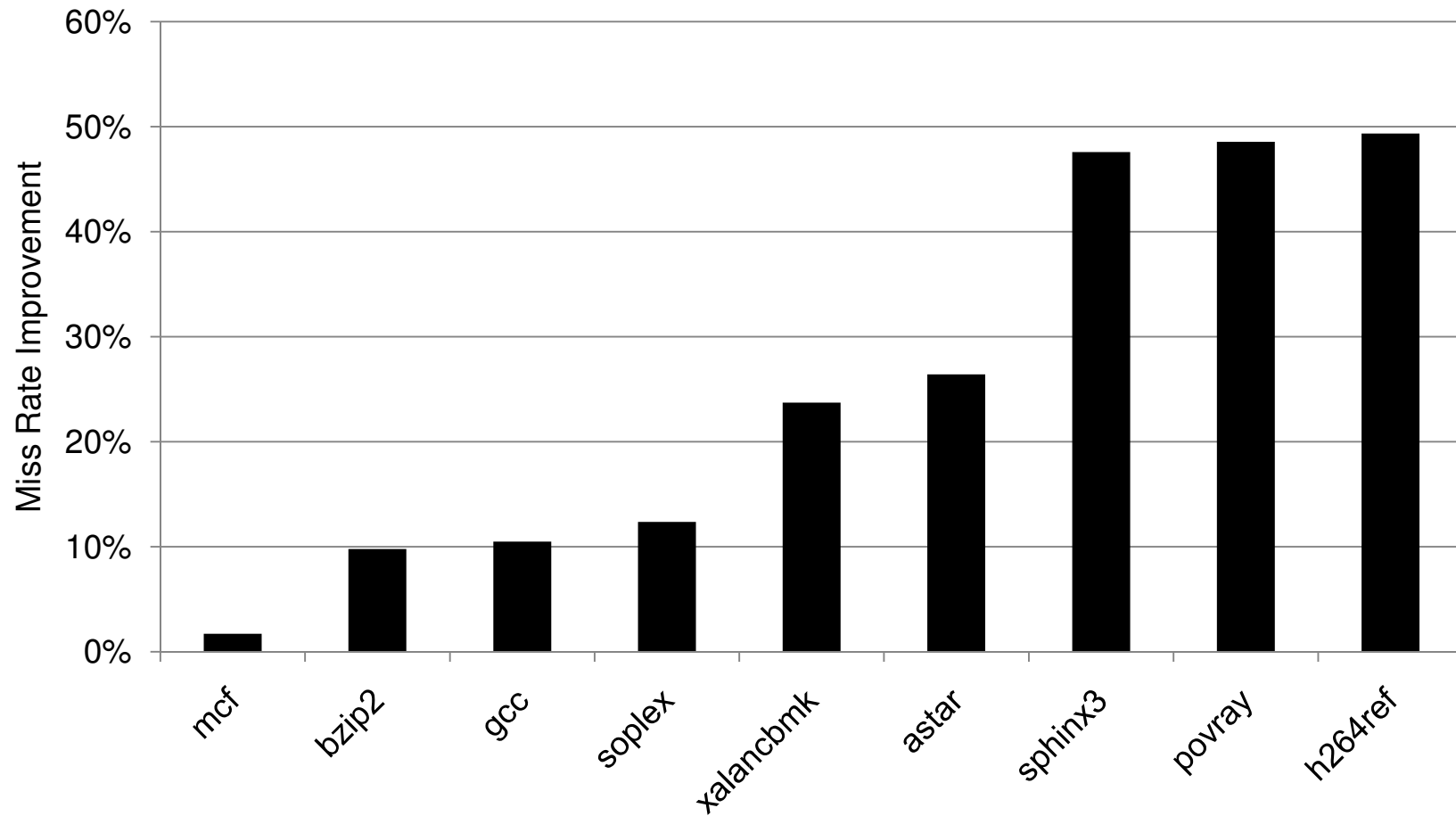
OOC Prefetching Picture



Context Switch Prediction



Miss Rate Improvement



**THANK
YOU!**

Q AND A

