



Accelerating Dependent Cache Misses with an Enhanced Memory Controller

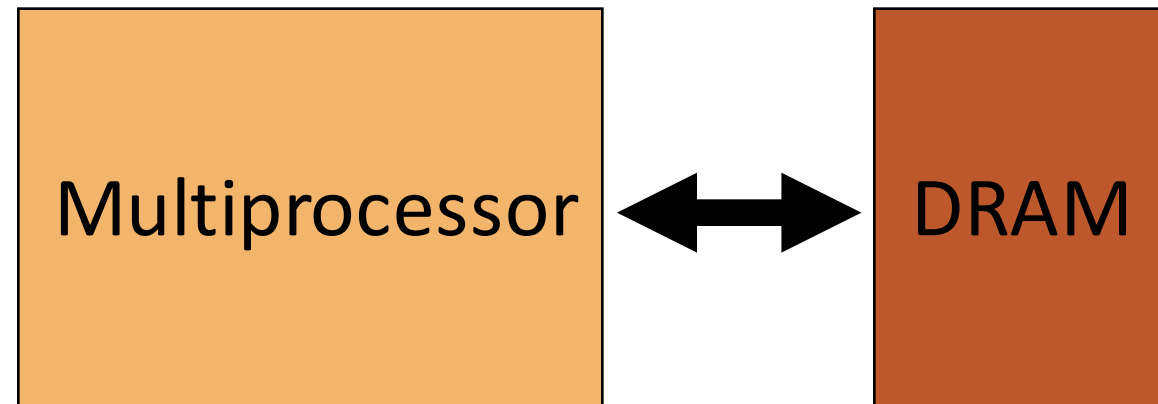
Milad Hashemi, Khubaib, Eiman Ebrahimi, Onur Mutlu, Yale N. Patt

Tuesday June 21: Session 7A, 3:30pm



Memory Access Latency

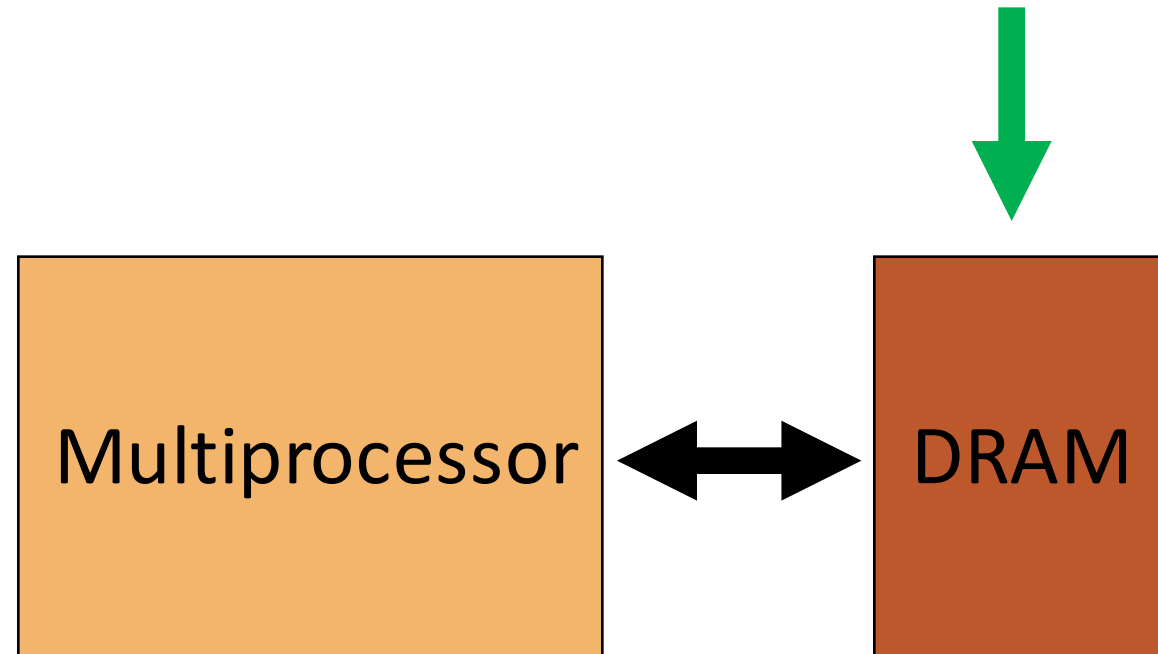
- The latency of accessing main memory is made up of two parts:





Memory Access Latency

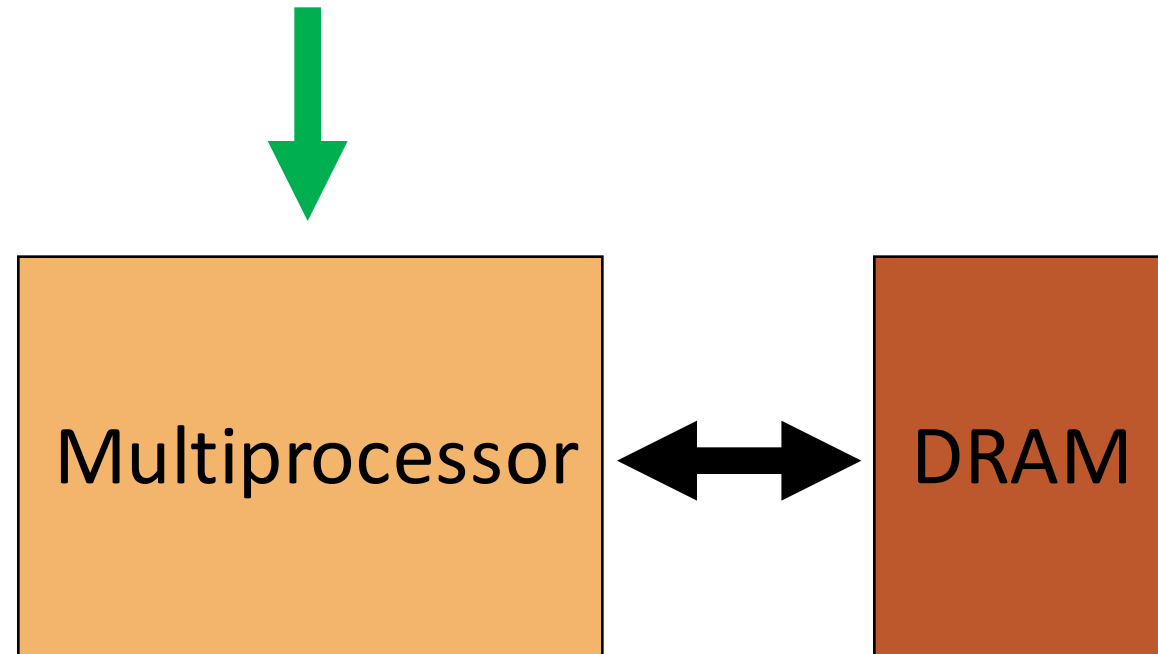
- The latency of accessing main memory is made up of two parts:
 - DRAM access latency





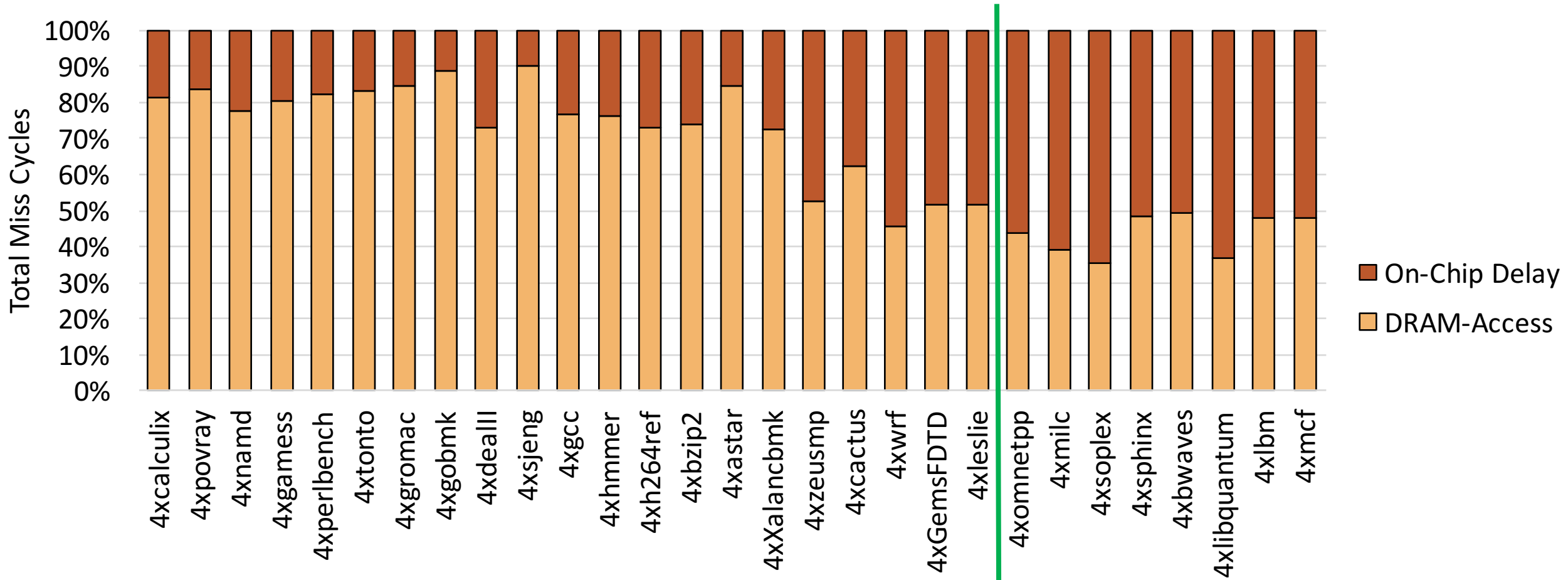
Memory Access Latency

- The latency of accessing main memory is made up of two parts:
 - DRAM access latency
 - On-chip latency





On-Chip Delay





Dependent Cache Misses

LD [R3] -> R5





Dependent Cache Misses

LD [R3] -> R5



ADD R4, R5 -> R9





Dependent Cache Misses

LD [R3] -> R5



ADD R4, R5 -> R9



ADD R9, R1 -> R6





Dependent Cache Misses

LD [R3] -> R5



ADD R4, R5 -> R9



ADD R9, R1 -> R6

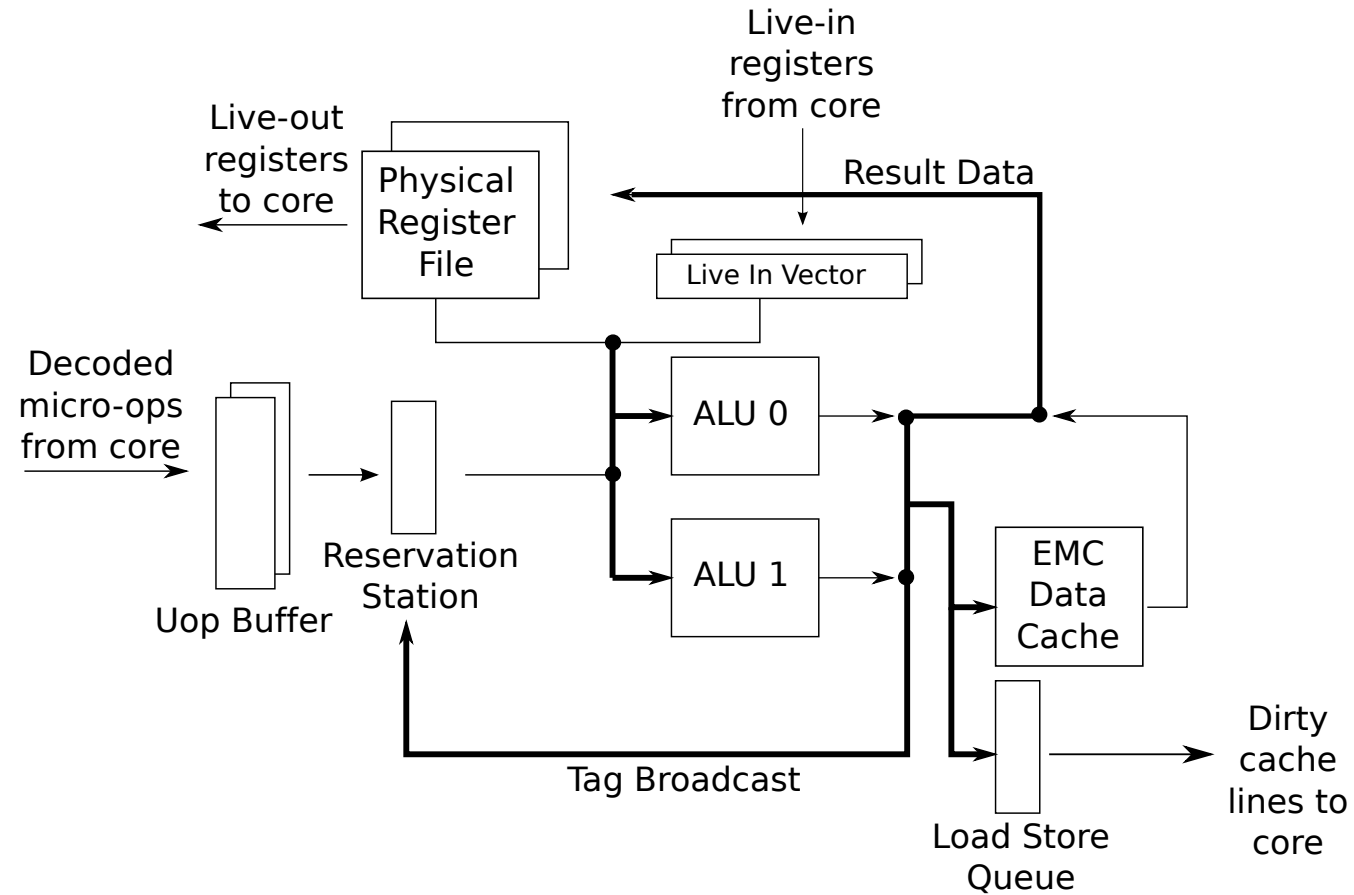


LD [R6] -> R8



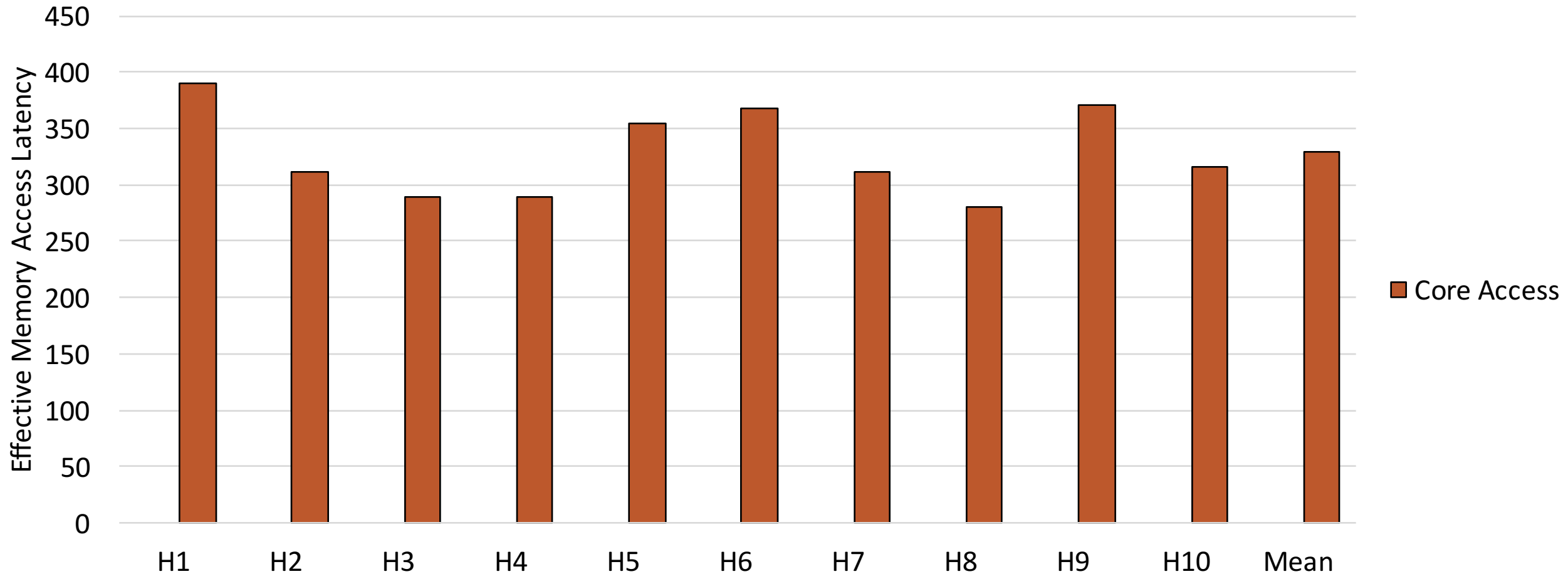


Compute Capable Memory Controller



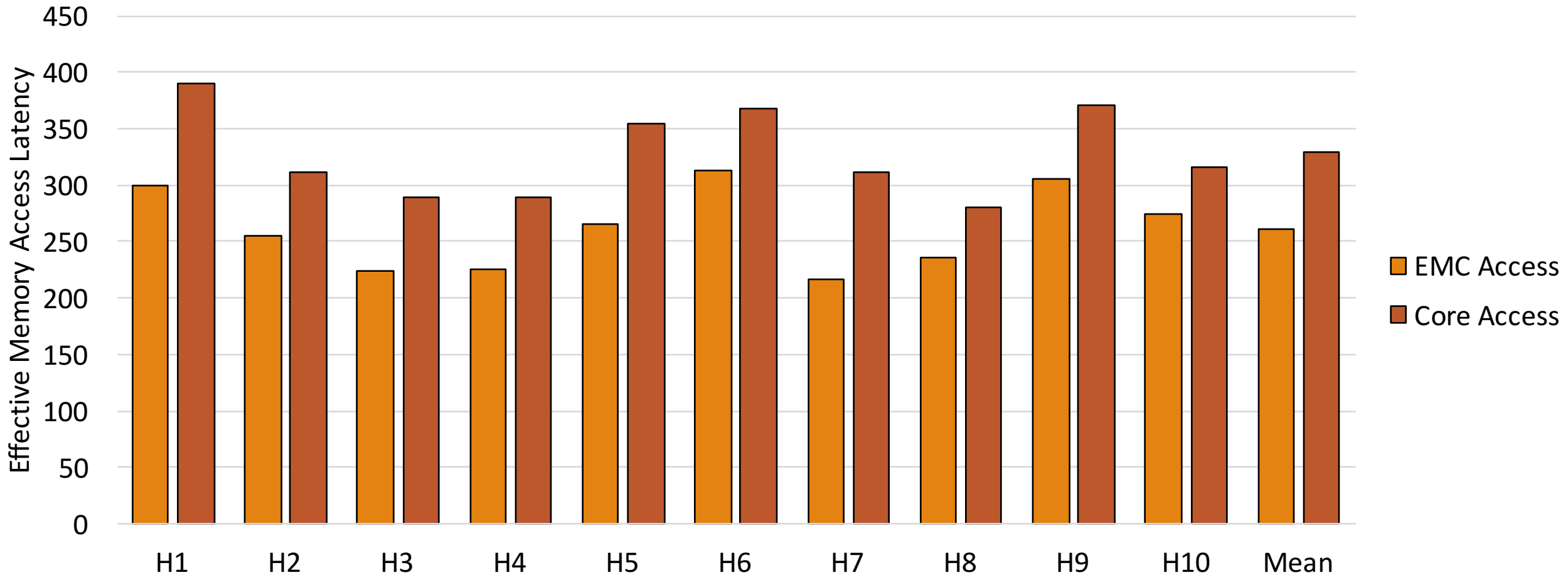


Effective Memory Access Latency Reduction





Effective Memory Access Latency Reduction





Accelerating Dependent Cache Misses with an Enhanced Memory Controller

Milad Hashemi, Khubaib, Eiman Ebrahimi, Onur Mutlu, Yale N. Patt

Tuesday June 21: Session 7A, 3:30pm