Computer Architecture

Average Memory Access Time

- Average Memory Access Time (AMAT)

AMAT = Hit time + Miss rate × Miss penalty

- Time to access a cache for both hits and misses

- Example: Find the AMAT for a cache with
  - Cache access time (Hit time) of 1 cycle = 2 ns
  - Miss penalty of 20 clock cycles
  - Miss rate of 0.05 per access

- Solution:
- AMAT = 1 + 0.05 × 20 = 2 cycles = 4 ns
- Without the cache, AMAT will be equal to Miss penalty = 20 cycles

Improving Cache Performance

Average Memory Access Time (AMAT)

AMAT = Hit time + Miss rate * Miss penalty

Used as a framework for optimizations

Reduce the Hit time

- Small and simple caches

Reduce the Miss Rate

- Larger cache size, higher associativity, and larger block size

Reduce the Miss Penalty

- Multilevel caches

Two-Level Cache Performance – 1/2

- Average Memory Access Time:

$$AMAT = Hit\ Time_{L1} + Miss\ Rate_{L1} \times Miss\ Penalty_{L1}$$

- Miss Penalty for L1 cache in the presence of L2 cache

$$Miss\ Penalty_{L1} = Hit\ Time_{L2} + Miss\ Rate_{L2} \times Miss\ Penalty_{L2}$$

- Average Memory Access Time with a 2nd Level cache:

$$AMAT = Hit\ Time_{L1} + Miss\ Rate_{L1} \times (Hit\ Time_{L2} + Miss\ Rate_{L2} \times Miss\ Penalty_{L2})$$

- Memory Stall Cycles per Instruction =

$$Memory\ Access\ per\ Instruction \times (AMAT - Hit\ Time_{L1})$$

Two-Level Cache Performance – 2/2

- Average memory stall cycles per instruction = Memory Access per Instruction × Miss $Rate_{L1}$ × (Hit $Time_{L2}$ + Miss $Rate_{L2}$ × Miss $Penalty_{L2}$)
- Average memory stall cycles per instruction = Misses per $instruction_{L1}$ × Hit $Time_{L2}$ + Misses per $instruction_{L2}$ × Miss $Penalty_{L2}$
- Misses per $instruction_{L1}$ = MEM access per instruction × Miss $Rate_{L1}$
- Misses per $instruction_{L2}$ = MEM access per instruction × Miss $Rate_{L1}$ × Miss $Rate_{L2}$

Example on Two-Level Caches

- Problem:
  - Miss Rate_{L1} = 4%, Miss Rate_{L2} = 25%
  - Hit time of L1 cache is 1 cycle and of L2 cache is 10 cycles
  - Miss penalty from L2 cache to memory is 100 cycles
  - Memory access per instruction = 1.25 (25% data accesses)
  - Compute AMAT and memory stall cycles per instruction

- Solution:
- AMAT = 1 + 4% × (10 + 25% × 100) = 2.4 cycles
- Misses per instruction in L1 = 4% × 1.25 = 5%
- Misses per instruction in L2 = 4% × 25% × 1.25 = 1.25%
- Memory stall cycles per instruction = 5% × 10 + 1.25% × 100 = 1.75
- Can be also obtained as: (2.4 – 1) × 1.25 = 1.75 cycles