

Notes for 10/19/09

Cache miss: when you have to get data from DRAM and not from cache

Three C's of cache misses: compulsory, capacity, conflict

Compulsory miss—something that you have never accessed before

Capacity—working set of program is bigger than cache

Conflict—your “spot” that you will be looked for in the cache is already full

Average memory access time = hit time + miss rate * miss penalty

Miss rate = 1 – hit rate

Cache size gets your miss rate down in terms of capacity but increases hit time

Associativity gets your miss rate down in terms of conflict but increases hit time

Bigger cache line size means you can exploit spatial locality

Smaller cache line size means that the hit time goes down

Valid bit—determines whether what is there is meaningful data

When you turn machine on, all valid bits are zero

Consider 1 kb cache size, 128 byte block, direct mapped

Bits 22 : 3 : 7 – tag : block index : block offset

Tag—rest of address to make sure that we have the right block

Consider 1 kb cache size, 128 byte block, 4-way set associative

Bits 24 : 1 : 7 – tag : set index : block offset