Cache II

3C's: Exclusive!

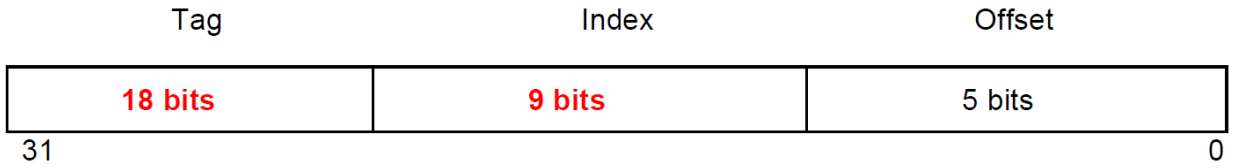
Sources of Cache Misses (3 C's)

- *Compulsory* (cold start, first reference):
  - 1st access to a block, "cold" fact of life, not a lot you can do about it.
    - If running billions of instructions, compulsory misses are insignificant
- *Capacity*:
  - Cache cannot contain all blocks accessed by the program
    - Misses that would not occur with infinite cache
- *Conflict* (collision):
  - Multiple memory locations mapped to same cache set
    - Misses that would not occur with ideal fully associative cache

3C's: Exclusive!

- Compulsory: Cannot be solved even if associativity and capacity are infinite
- Conflict: Can be eliminated by increasing associativity without changing cache capacity
- Capacity: Can be eliminated only by increasing cache capacity
- An example:
  - Cache capacity = 8KB, Associativity = 2, block size = 4B, LRU
  - Access the following addresses in order:
  - 0x00000000: compulsory
  - 0x00001000: compulsory
  - 0x00002000: compulsory
  - 0x00000000: conflict (increase associativity to 4 solves the problem)

3C's: Exclusive!



- Another example:
  - Consider a 32-bit physical memory space and a 32 KiB 2-way associative cache with LRU replacement. After the execution of the following code:

```
int ARRAY_SIZE = 64 * 1024;
int arr[ARRAY_SIZE]; // *arr is aligned to a cache block
/* loop 1 */ for (int i = 0; i < ARRAY_SIZE; i += 8) arr[i] = i;
/* loop 2 */ for (int i = ARRAY_SIZE - 8; i >= 0; i -= 8)
                arr[i+1] = arr[i];
```

  - 1. What is the hit rate of loop 1? What types of misses (of the 3 Cs), if any, occur as a result of loop 1? 0, Compulsory Misses
  - 2. What is the hit rate of loop 2? What types of misses (of the 3 Cs), if any, occur as a result of loop 2? 9/16, Capacity Misses

Multi-Level Cache

- Consider only L1 and L2 cache
- Local Miss Rate L2$ = L2$ Misses / L1$ Misses = L2$ Misses / Total_L2_accesses
- Global Miss Rate = L2$ Misses / Total Accesses

  =(L2$ Misses / L1$ Misses)x(L1$ Misses / Total Accesses)

  =Local Miss Rate L2$ x Local Miss Rate L1$
- AMAT = Time for a hit + Miss Rate x Miss Penalty

  = Time for a L1$ hit + (local) Miss Rate L1$ x

  (Time for a L2$ hit + (local) Miss Rate L2$ x L2$ Miss penalty)

Multi-Level Cache

- Suppose you have the following system that consists of an:
  - L1$ with a local hit rate of 80% and a hit time of 2 cycles
  - L2$ with a global miss rate of 8% and a hit time of 15 cycles
  - DRAM accesses take 50 cycles
- What is the AMAT of the L1 cache?_____
- Suppose we want to improve our AMAT, making sure that it is no greater than 6 cycles, by improving our L2$'s hit rate. What is the minimum possible local hit rate for L2$ that allows us to meet our AMAT requirement?_____

Multi-Level Cache

- Suppose you have the following system that consists of an:
  - L1$ with a local hit rate of 80% and a hit time of 2 cycles
  - L2$ with a global miss rate of 8% and a hit time of 15 cycles
  - DRAM accesses take 50 cycles
- What is the AMAT of the L1 cache? 2+0.2*(15+(0.08/0.2)*50)=9
- Suppose we want to improve our AMAT, making sure that it is no greater than 6 cycles, by improving our L2$'s hit rate. What is the minimum possible local hit rate for L2$ that allows us to meet our AMAT requirement? 90%