IT 209

Practice Problem

Date: 6th October, 2018

Problem Statement

A certain system with a 500 MHz clock uses a Harvard architecture (separate data and instruction caches) at the first level, and a unified second-level cache. The first-level data cache is a direct-mapped, write-through, writes-allocate cache with 8 Kbytes of data total and 8-Byte blocks, and has a perfect write buffer (never causes any stalls).

The first-level instruction cache is a direct-mapped cache with 4KBytes of data total and 8-byte blocks. The second-level cache is a two-way set associative, write-back, write-allocate cache with 2MBytes of data total and 32-Byte blocks. The first-level instruction cache has a miss-rate of 2%. The first-level data cache has a miss-rate of 15%. The unified second-level cache has a local miss rate of 10% (i.e. the miss rate for all accesses going to the second-level cache). Assume that 40% of all instructions are data memory accesses; 60% of those are loads, and 40% are stores. Assume that 50% of the blocks in the second-level cache are dirty at any time. Assume that there is no optimization for fast reads on an L1 or L2 cache miss.

All first-level cache hits cause no stalls. The second-level hit time is 10 cycles (That means that the L1 miss-penalty, assuming a hit in the L2 cache, is 10 cycles). Main memory access time is 100 cycles to the first bus width of data; after that, the memory system can deliver consecutive bus widths of data on each following cycle. Outstanding, non-consecutive memory requests can not overlap; an access to one memory location must complete before an access to another memory location can begin. There is a 128-bit bus from memory to the L2 cache, and a 64-bit bus from both L1 caches to the L2 cache. Assume that the TLB never causes any stalls.

Questions:

1. What fraction of all data memory references cause a main memory access (main memory is accessed before the memory request is satisfied)? **First show the equation, then the numeric result.**
2. How many bits are used to index each of the caches ie Instruction, Data and L2 ?
3. What is the average memory access time in cycles (including instructions and data memory references)? **First show the equation, then the numeric result.** *Hint: don’t forget* *to consider dirty lines in the L2 cache.*