CSE 560 – Practice Problem Set 6 Solution

1. You are building a system around a processor with in-order execution that runs at 2.66 GHz and has a CPI of 0.7 excluding memory accesses. The only instructions that read or write data from memory are loads (20% of all instructions) and stores (5% of all instructions).

   The memory system for this computer is composed of a split L1 cache that imposes no penalty on hits. Both the I-cache and the D-cache are direct mapped and hold 32 KB each. The I-cache has a 2% miss rate and 32-byte blocks, and the D-cache is write through with a 5% miss rate and 16-byte blocks. There is a write buffer on the D-cache that eliminates stalls for 95% of all writes (i.e., there is no stall for a hit, even though the cache is write through).

   The 512 KB write-back, unified L2 cache has 64-byte blocks and an access time of 15 ns. Of all memory references sent to the L2 cache in this system, 80% are satisfied without going to main memory.

   The main memory has an access latency of 60 ns, after which any number of bus words may be transferred at the rate of one per cycle on the 128-bit-wide 133 MHz main memory bus.

   Assume that writes are similar to reads.

   Note, you may want to define a few symbols such as $\mathrm{t}_{\mathrm{HIT,I\$}}$, $\mathrm{t}_{\mathrm{MISS,I\$}}$, $\%\mathrm{miss}_{\mathrm{I\$}}$, $\mathrm{t}_{\mathrm{AVG,I\$}}$, etc. to handle each cache and memory in your formulation.

   (a) What is the average memory access time for instruction accesses?

   So as to not get lost in the notation, let's first define a few symbols.

Symbol	Meaning	Value (if given in problem statement)
$t_{HIT,I\$}$	Time (penalty) for hit in I$	0 clocks
$t_{MISS,I\$}$	Time for miss in I$	
$t_{HIT,D\$}$	Time (penalty) for hit in D$	0 clocks
$t_{MISS,D\$}$	Time for miss in D$	
$t_{HIT,L2}$	Time for hit in L2	15 ns = 40 clocks
$t_{MISS,L2}$	Time for miss in L2	
$t_{HIT,M}$	Time for hit in main memory	
$\%miss_{I\$}$	Miss rate for I$	0.02
$\%miss_{D\$}$	Miss rate for D$	0.05
$\%miss_{L2}$	Miss rate for L2	0.20
$t_{AVG,I\$}$	Average time for I$	
$t_{AVG,D\$}$	Average time for D$	
$t_{AVG,L2}$	Average time for L2	

This problem is all about repeated use of the general form $t_{AVG} = t_{HIT} + \%miss \times t_{MISS}$.

The values needed to compute $t_{AVG,I\$}$ and that aren't in the table above are: $t_{MISS,I\$}$, $t_{MISS,L2}$, and $t_{HIT,M}$. Starting at the bottom of the memory hierarchy and moving up, we can compute the value of $t_{HIT,M}$ by adding the access time to the transfer time:

60 ns = 160 clocks
64 bytes X 8 bits/byte / 128 bits/transfer = 4 bus transfers/access
2.66 GHz processor freq. / 133 MHz bus freq. = 20 processor clocks per transfer
4 transfers X 20 clocks/transfer = 80 clocks to transfer data from main memory
$t_{HIT,M}$ = access time + transfer time = 160 + 80 = 240 clocks

and

$t_{MISS,L2} = t_{HIT,M}$ = 240 clocks

We can now apply the general expression for t_{AVG} on L2.

$t_{AVG,L2} = t_{HIT,L2} + \%miss_{L2} \times t_{MISS,L2}$
= 40 clocks + 0.20 X 240 clocks
= 88 clocks

Next, we observe that $t_{MISS,I\$} = t_{AVG,L2}$, and we are ready to apply the general expression again, this time for t_{AVG} on the I$.

$t_{AVG,I\$} = t_{HIT,I\$} + \%miss_{I\$} \times t_{MISS,I\$}$
= 0 + 0.02 X 88 clocks
= 1.76 clocks

(b) What is the average memory access time for data reads?

We have everything we need to directly apply the general expression for t_{AVG} on the D$.

$t_{AVG,D\$} = t_{HIT,D\$} + \%miss_{D\$} \times t_{MISS,D\$}$
$= t_{HIT,D\$} + \%miss_{D\$} \times t_{AVG,L2}$
= 0 + 0.05 X 88 clocks
= 4.4 clocks

(c) What is the average memory access time for data writes?

There is no reason given for writes to have a different access time than reads, given the statements made in the problem description.

(d) What is the overall CPI, including memory accesses?

The overall CPI is the base CPI + instruction cache CPI + data cache CPI due to loads + data cache CPI due to stores.

CPI = 0.7 + 1.76 + (0.20)(4.4) + (0.05)(4.4) = 3.56

Memory access took a 0.7 CPI system and turned it into a 3.56 CPI system. Clearly, some improvement in the memory system is warranted here, maybe an L3 cache?