

5.1 Appendix B, page B-21 3rd edition

5.4

- a) hit time_{L1} = 1ns
 miss rate_{L1} = 0.02
 hit time_{L2} = 15ns
 miss rate_{L2} = 1 - 0.8 = 0.2

miss penalty_{L2} = memory latency + time to transfer_{one L2 block}.

transfer rate of bus = 0.8bytes / ns

L2 block transfer = 80ns

miss penalty_{L2} = 20 + 80 = 100ns

Since 50% are dirty miss penalty_{L2} = 100 + 0.5 x 100 = 150ns

average memory access time = hit time_{L1} x (hit time_{L2} + miss rate_{L2} x
 miss penalty_{L2})

avg. mem. acc. time = 1.9ns

- b) miss rate_{L1} (for reads) = 0.05
 avg. mem. acc. time = 1 + 0.05 x (15 + 0.2 x 150) = 3.25ns

- c) miss penalty for d.w. = 150ns
 write time into L2 buffer = hit time_{L2} + 0.2 x miss penalty_{L2}
 = 1 x 15 + 0.2 x 150
 = 45ns

avg. mem. acc. time = hit time_{L1} + 0.05 x write time to L2 buffer
 = 1 + 0.05 x 45
 = 3.25 ns

- d) overall CPI
 CPI_{overall} = CPI_{base} + Clock (InstrRead * AMAT_{inst} + DataRead * AMAT_{dataread} +
 DataWriteFreq * AMAT_{datawrite})
 = 0.7 + 1.1 GHz (1.0 * 1.9 ns + 0.20 * 3.25ns + 0.05 * 3.25 ns)
 = 3.68

- e) CPU_{time} = InstructionCount * CPI_{overall} * ClockCycleTime
 = 1.5 * 10⁶ * 3.68 * (2.1 GHz)⁻¹
 = 2.63 ms

- f) Improving memory access time would improve the system.

5.7

Combining the two will naturally decrease miss rate; however the miss penalty might increase.