**Model Solutions**

**Class Test 2**

**High Performance Computer Architecture (CS60003)**

**Time 55 Minutes Marks 39**

1. Consider a computer having a 4-way set associative L2 cache of size **256 KB**. The cache line size is 8 words. The word size is 32 bits. The smallest addressable unit is a byte, and memory addresses are **64 bits** long. What percentage of the cache memory is used for tag bits? S**how your workout clearly, and put a box around your final answer.** **[5]**

**Solution:**

**The block size is 8 words (32 bytes).**

**Block offset is log232 = 5 bits. Award 1 Mark**

**Number of sets is 256 KB / (32 \* 4) = 2048 sets.**

**Index field = 11 bits. …. Award 1 Mark**

**Tag field = 64 – 11 – 5 = 48 bits . …. Award 1 Mark**

**Size of the cache line is 32 \* 8 = 256 bits.**

**Therefore, %cache memory used for tag bits = 48/(48 + 256) = 15.8%.**

**… Award 2 Marks**

1. A memory system uses a hierarchy of three caches: L1, L2,and L3. The L1, L2, and L3 caches have local hit rates of **90%**, **80%**, and **50%**; and the hit times of **1 ns**, **10ns**, and **20ns** respectively. The hit time of the main memory is **100 ns**. What is the average memory access time? S**how your work out clearly, and put a box around your final answer. [5]**

**Solution:**

**Access time = L1 hit time + L1 miss rate × L1 miss penalty**

**L1 miss penalty = L2 access time = L2 hit time + L2 miss rate × L2 miss penalty**

**L2 miss penalty = L3 access time = L3 hit time + L3 miss rate × L3 miss penalty = 20 ns + 0.5(100 ns) = 70 ns. … Award 1 Mark**

**L1 miss penalty = 10 ns + 0.2(70) ns = 24 ns. … Award 1 Mark**

**AMAT = 1 ns + 0.1(24 ns) = 3.4 ns. … Award 3 Marks**

1. A computer with a two level cache has a base CPI of 1.5 when all references hit the L1 cache. The computer has the following characteristics:

* Clock rate 250MHz
* Memory access time of 100ns
* Miss rate at the primary cache of 5%
* L2 cache access time of 10ns
* Global miss rate at secondary cache of 1%

What is the average CPI? S**how your work clearly, and put a box around your final answer.** **[5]**

**Solution:**

**Clock rate= 250 MHz implies clock cycle= 4 ns because 1/(250\*106) = 4 \* 10-9.**

**Therefore, memory access time= 25 cycles and L2 cache access time= 2.5 cycles. … Award 1 Mark**

**The effective CPI = 1.5 + miss rate L1 \* miss penalty L1 + Global miss rate L2\* miss penalty L2 = 1.5 + (0.05)\*2.5 + 0.01\*( 25 )**

**= 1.5 + 0.1 + 0.275 = 1.875 … Award 4 Marks**

1. The average memory access time (AMAT) for a microprocessor with only a single level of cache is **2.4** clock cycles. L1 hit time is 1 clock cycle and memory hit time is 100 clock cycles . The processor designers are trying to improve AMAT by 100% by adding a 2nd level of cache on-chip. The L2 cache hit time is 5 clock cycles. **To obtain the desired speedup, what should be L2 local hit rate?**

**Solution:**

**AMAT = Hit Time + Miss Rate x Miss Penalty**

**2.4 = 1 + Miss Rate x 100**

**L1 Miss Rate = 1.4%**

**Required speed up = 2. Therefore 2 = 2.4 / AMAT (new)**

**AMAT (new) = 1.2 clock cycles**

**Therefore, 1.2 = 1 + 0.014 x (5 + Miss RateL2\*100)**

**Solving for Miss RateL2 suggests that the highest possible miss rate is ~9.28%**

**Thus, L2 Local hit rate = (1 – Miss Rate) = ~90.72%.**

1. Consider the execution of a parallel loop on a bus-based multiprocessor with 2 processors. Processor P0 executes the even iterations and Processor P1 executes the odd ones. For each iteration, each processor reads elements A[i]and B[i], does the addition, and then writes element A[i], generating two memory reads and one memory write.

**Original Loop: for (i=0; i<N; i++) A[i] = A[i] + B[i];**

**After parallelization: Processor P0 executes even iterations: for (i=0; i<N; i=i+2) A[i] = A[i] + B[i];**

**Processor P1 executes odd iterations: for (i=1; i<N; i=i+2) A[i] = A[i] + B[i];**

Each processor has a private data cache with 16-byte blocks. Each cache block can fit 4 array elements, where each element is 4 bytes. Assume a cold start. The caches blocks are initially invalid. The elements of A and B array are assigned successive locations starting at a block boundary. The blocks containing A and B array elements do not conflict with each other in the cache.

1. Compute the number of bus transactions for N=100. **[2]**

**Answer:**

**Total number of bus transactions (all iterations) = 12 bus transactions (to process 4 elements) × 100/4 = 300 [2]**

b) Modify the code executed by P0 and P1 to minimize the cache misses and bus transactions. Find number of bus transactions for N=100. **[4+1]**

**Modified Code for P0 and P1**

**Processor P0 executes first N/2 iterations:**

**for (i=0; i<N/2; i++) A[i] = A[i] + B[i];**

**Processor P1 executes last N/2 iterations:**

**4 marks**

**for (i=N/2; i<N; i++) A[i] = A[i] + B[i];**

**Number of bus transactions for MSI (all iterations) =**

**4 bus transactions (to process 4 elements) × 100/4 = 100 [1]**

1. Assume that the caches implement MSI coherence protocol. Complete the table (next page) showing the execution of 2 iterations in P0 and in P1. **[12]**

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Read/Write Operation** | **Bus Transaction** | **P0 Cache**  **State** | **P1 Cache**  **State** |
| **Initial Cold Start** |  | **I** | **I** |
| **P0 reads A[0]** | **Bus Read** | **S** | **I** |
| **P1 reads A[1]** | **Bus Read** | **S** | **S** |
| **P0 reads B[0]** | **Bus Read** | **S** | **I** |
| **P1 reads B[1]** | **Bus Read** | **S** | **S** |
| **P0 writes A[0]** | **Bus ReadX** | **M** | **I** |
| **P1 writes A[1]** | **Bus ReadX, Flush** | **I** | **M** |
| **P0 reads A[2]** | **Bus Read** | **S** | **I** |
| **P1 reads A[3]** | **Bus Read** | **S** | **S** |
| **P0 reads B[2]** | **Bus Read** | **S** | **I** |
| **P1 reads B[3]** | **Bus Read** | **S** | **S** |
| **P0 writes A[2]** | **Bus ReadX** | **M** | **I** |
| **P1 writes A[3]** | **Bus ReadX, Flush** | **I** | **M** |