

**ECE 366—Fall 2001, Instructor: Prof. Shantanu Dutt**

**Homework 4 : Due Fri Nov. 30**

1. The  $t_{av} = hit\_time^{cache} + miss\_rate^{cache} miss\_penalty^{cache}$  metric discussed in class is also called the AMAT (average memory access time) metric. Find the AMAT for a computer with a 2ns clock, and the following cache parameters: a miss penalty of 20 ccs, a miss rate of 0.05, and a hit time of 1cc. Assume that read and write penalties are the same and ignore other write stalls. **30**
  
2. Suppose we can improve the cache miss rate in the above computer to 0.03 by doubling the cache size. This causes the cache hit time to become 1.2 cc's (a larger memory unit is also slower to access). Does the AMAT improve compared to the computer in the above problem? Show your work and explain. **30**
  
3. Consider a computer memory hierarchy with a cache, main memory (MM) and secondary storage (SS). The cache hit rate is 0.97, cache hit time = 2cc, block transfer time from MM to cache is 20cc (this is the same as the hit time for MM), MM hit rate of 0.99999, and page transfer time from SS to MM of 100,000 cc's (this is the same as the hit time for SS). Note the SS hit rate is always 1. The AMAT  $t_{av}$  is given by  $t_{av} = hit\_time^{cache} + miss\_rate^{cache} miss\_penalty^{cache}$ . Determine the AMAT for this computer. **50**
  
4. Suppose a computer's address size is  $k$ -bits (using byte addressing), the cache size is  $S$  bytes, the cache block size is  $B$  bytes, and the cache is  $A$ -way set associative. Assume that  $B$  is a power of two, so  $B = 2^b$ . Figure out what the following quantities are in terms of  $S, B, A, b$  and  $k$ :
  - (a) The number of sets in the cache. **10**
  - (b) The number of index bits in the address. **10**
  - (c) The total number of bits needed to to implement the cache—this includes data bits, tag bits and valid bits. **20**
  
5. (a) Following is a string of address references given as word addresses: 1, 4, 8, 5, 20, 17, 19, 56, 9, 11, 33, 43, 5, 1, 9, 17. Assuming a DM cache with 16 1-word blocks that is initially empty, label each reference in the list as a hit or miss and show the final contents of the cache. In case of a miss, classify which type of miss (compulsory, capacity, conflict) it is. A miss occurring because a block is being accessed for the very first time is a *compulsory* miss. If a miss is not compulsory, then to determine if it is a *conflict* miss, determine if this miss would have occurred in a FA cache with the same block size and for the same string of block accesses till the current miss; if not then this miss is a conflict miss. In other words, if block  $i$  is being currently accessed, and a miss occurs, go back to the previous miss, and assuming block  $i$  to be present in cache at that point, determine if it would have been the least recently used block at that point; if so, then block  $i$  would have been replaced at least at this previous miss in a FA cache, and the current miss when accessing block  $i$  is not a conflict miss. Otherwise, it

is a conflict miss. Any miss that is neither a conflict nor a compulsory miss is a *capacity* miss. **30**

(b) Do the same for the above reference string for an FA cache with 4-word blocks and total size of 16 words. Assume LRU replacement. **40**