High Performance Computer Architecture (CS60003)

Tutorial - 1

Indian Institute of Technology Kharagpur

Question 1

Consider a memory system that uses a 32-bit address to address at the byte level, plus a cache that uses a 64-byte line size.

- Assume a direct mapped cache with a tag field in the address of 20 bits. Show the address format and determine the following parameters: number of addressable units, number of blocks in main memory, number of lines in cache, size of tag. Answer: 2^{32} bytes, 2^{26}, 64, 20 bits
- Assume an associative cache. Show the address format and determine the following parameters: number of addressable units, number of blocks in main memory, size of tag.Answer: 2^{32} bytes, 2^{26}, 26 bits
- Assume a four-way set-associative cache with a tag field in the address of 9 bits. Show the address format and determine the following parameters: number of addressable units, number of blocks in main memory, number of lines in set, number of sets in cache, number of lines in cache, size of tag. Answer: 2^{32} bytes, 2^{26}, 4, 2^{17}, 2^{19}, 9 bits

Question 2

Consider a single-level cache with an access time of 2.5 ns, a line size of 64 bytes, and a hit ratio of $H = 0.95$. Main memory uses a block transfer capability that has a first word (4 bytes) access time of 50 ns and an access time of 5 ns for each word thereafter.

- What is the access time when there is a cache miss? Assume that the cache waits until the line has been fetched from main memory and then re-executes for a hit. Answer: 130ns
- Suppose that increasing the line size to 128 bytes increases the H to 0.97. Does this reduce the average memory access time? Answer: Under the initial condition, the average access time is 8.875ns and under the revised scheme the average access time is 8.725ns

Question 3

Processor A has two 8 Kbyte, Level-1 caches – one for data and one for instruction. However, a design team is considering another option (i.e. Processor B) – a single, 16 Kbyte cache that holds both instructions and data. Additional specifications for the 16 Kbyte cache include:

- Each block will hold 32 bytes of data (not including tag, valid bit, etc.).
- The cache would be 2-way set associative.
- Physical addresses are 32 bits.
- Data is addressed to the word and words are 32 bits.

1. How many blocks would be in this cache? Answer: 512
2. How many bits of tag are stored with each block entry? Answer: 21

3. Each instruction fetch means a reference to the instruction cache and 35% of all instructions reference data memory. With the first implementation on Processor A, the average miss rate in the L1 instruction cache was 2%, the average miss rate in the L1 data cache was 10%, and in both cases, the miss penalty is 9 cycles. For Processor B, the average miss rate is 3% for the cache as a whole, and the miss penalty is again 9 cycles. Which design is better? Answer: Processor B with miss penalty of 0.27 against 0.495 of A