Comments

Antoniu Pop

Question 1:

- Good overall understanding of the key characteristics of caches (temporal/spatial locality) and of cache misses.

- For part (d) a majority of students did not account for the factthat, in case of a cache miss, it is necessary to first determine that the data is not in cache before going to main memory, which incurs a cost (t_h).

- A significant number of students did not know how to define and calculate a speedup.

- Very few misunderstandings on questions, only a couple of students did not relate the parameters in part (b) to the types of cache misses from part (a) as asked.

Question 2:

- The vast majority did very well on textbook questions about HDD operation and RAID.

- Most students were comfortable with the HDD-only scenario (question a.2) that was rather similar to the examples presented in the lectures.

- A rather common mistake in part (b) was to consider that latency is paid for every 4kB in an access. It is only paid once, when initiating a transfer.

- In a few cases, students did not know what a microsecond is.

- Also for part (b), no student fully and correctly analysed the HDDvs. SSD scenarios, so I have accepted additional answers for that question. Specifically, were accepted papers where the latency was added to the transfer time (similarly to seek/search time in the case of HDDs) for the entire transfer rather than just on the total size less the 4kB for which latency was provided.

  E.g., for question (b.1) 8kB read from SSD would take 12microseconds for the first 4kB and then 4/500,000 = 8 microseconds for the next 4kB for a total of 20 microseconds, but I accepted 12us + 8/500,000 = 28 microseconds as well.

  When students considered that the latency was paid for each 4kB portion of the transfer, of course, this was not accepted.

Javier Navaridas-Palma

I'm generally happy with students' performance in Q3 and Q4, overall good marks were achieved, specially in Q3.

From Q3 it seems that most students seem to understand dependencies and the concept of pipelining. A common inaccuracy was to define instructions in a pipeline to be run in parallel as they are actually interleaved. Some students drew a dependency between the CMP and the DIV, while the dependency is with the SUB. This is a somehow understandable error. In the discussion about the suitability of the code for superscalar, most students only discussed about 4-way superscalar and not many noticed that there were 8 instructions with only 4 levels of dependency so 2-way could do perfectly.

With regards to Q4, most students seem to have grasped the concepts of superscalar, multithreading and multicore, although there were some minor inaccuracies here and there. However, there weren't many students that were able to explain cache trashing and even less false sharing, which is a bit disappointing. Furthermore, I was a rather surprised with the low performance of part c as I wasn't expecting it to be too challenging. The most common mistakes here where to believe that GHz are the only thing that determines processor speed and to multiply threads and superscalar lanes to compute the number of instructions per cycle of a core.