Computer architecture

Homework week 5

Instructions

Submit by e-mail to the lecturer, as a PDF document with your name and student ID near the beginning. You can work in groups of 2; however, use different groups than week 4 and 5. Use the English language. Deadline: Oct 7th, 23:59.

Answers to bonus questions are counted only if you get at least half the points to each previous question.

Question 1 (3pt)

The DDR tRAS en tCCD delays are example minimum times between *successive requests* of a specific type to the DRAM bank. Using your expertise from last week's homework, name the relevant DDR timings that constrain the *rate* at which requests can be handled by a DDR channel.

Determine under which conditions the *maximum read bandwidth* and *maximum write bandwidth* can be obtained for a given set of timings given symbolically. Similarly, determine under which conditions the *minimum read bandwidth* and *minimum write bandwidths* can be obtained.

> **Note**
>
> The maximum bandwidth is the maximum rate at which requests can be handled.

> **Hint**
>
> The conditions should name the type of requests and diversity of addresses being used. For DDR, the maximum bandwidth is also called "sequential bandwidth".

Question 2 (2pt)

There are two sets of DDR timings in the MGSim configuration file provided with assignment series 1&2a (`minisim.ini`): one for MT41J128M8 and one for Kingston HyperX.

Determine quantitatively the maximum read and write bandwidths on both configurations.

Question 3 (5pt)

Consider a computer composed of a processor, a cache and a DDR channel connected to a MT41J128M8 controller/bank. Assume that the processor and cache operating frequency is set to 1GHz. Assume that the cache has a capacity of 27KiB. Assume that the processor can issue at most one load or store operation per cycle; and that the cache can handle at most one request from the processor and issue at most one request to the DDR channel per cycle.

Given that the processor issues one memory load per 10 cycles on average, and does not issue stores, estimate which cache hit rate yields maximum utilization of the DDR channel without stalling the processor.

For simplicity, you can consider that the pattern of cache misses can exploit the maximum read bandwidth.

Question 4 (bonus: 4pt)

Take the computer from the previous question. Consider that the processor and cache are duplicated, and their frequency reduced to 500MHz. They are connected to the same DDR channel, and need to take turns to issue requests to the memory.

Consider that the two processors issue the same patterns of loads as above, and that the cache hit rate is the same as above. Can this system achieve the maximum read bandwidth without stalling the processors? Why?