

G22.2243: High Performance Computer Architecture

Assignment #5

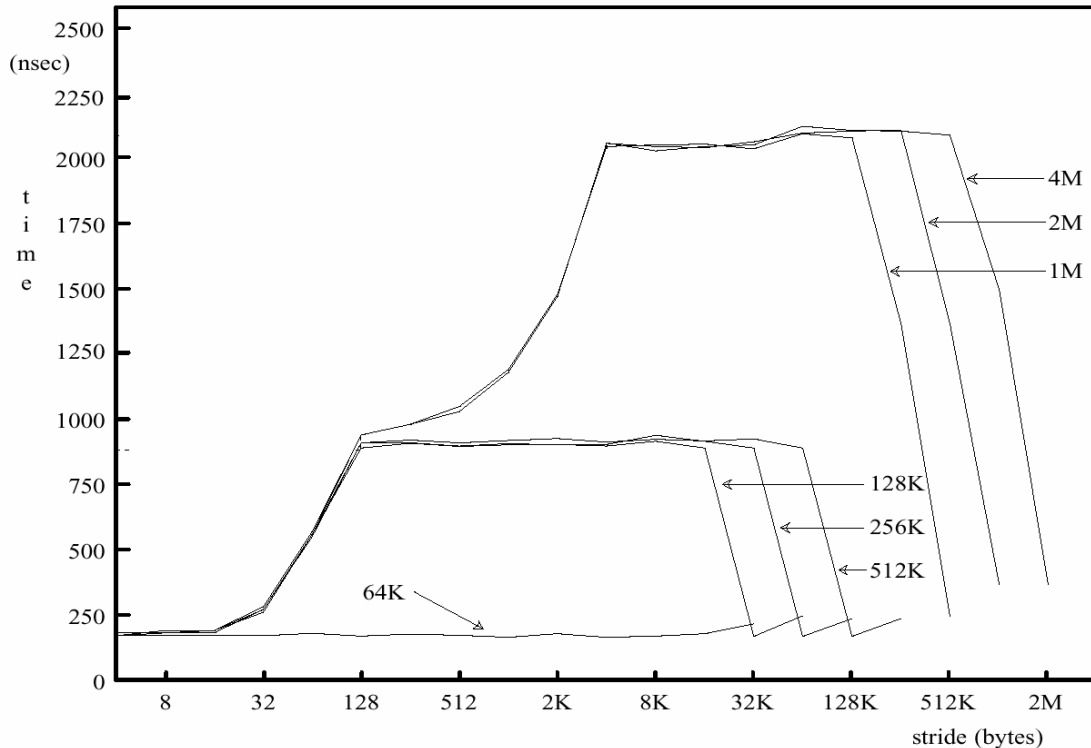
(Optional)

In this assignment, you will run a microbenchmark program to evaluate the behavior of a memory system. The question is derived from Exercise 5.2 of the textbook. (in particular, you should execute the program listing available there: I have posted this program to the mailing list so that you do not have to type it in).

Part 1:

Lets start off by looking at the results of the microbenchmark in an ideal setting.

The figure below shows the results of running the memory hierarchy microbenchmark program described above on a particular machine with a **single-level cache** and a **TLB**. The x-axis corresponds to the stride size (in bytes) and the y-axis to the average cost of a R+W operation (in nanoseconds). The curves correspond to different sizes of the array (in bytes). Note that the x-axis is shown using a log-scale.



To understand the graph, we shall start off by constructing a very simple analytical model of memory access costs for a machine with a single-level cache (with no TLB). Assume that our array size is N bytes, the stride is s , the cache size is C , with line size b (bytes) and associativity a .

- a. Complete the following table, assuming that the cache hit cost is T_{hit} , and the miss cost is T_{miss} . The rows in the table correspond to four different behaviors, corresponding to different values of the array size and stride (with respect to cache size, line size, and associativity). The "Frequency of Misses" column requires you to analyze how often would the microbenchmark result in a miss for the given

range of values; the “Time per Iteration” column should follow very simply once you have this number.

| <i>Behavior</i> | <i>Size of Array</i> | <i>Stride</i> | <i>Frequency of Misses</i> | <i>Time per Iteration</i> |
|-----------------|----------------------|-----------------------|----------------------------|---------------------------|
| 1 | $1 \leq N \leq C$ | $1 \leq s \leq N/2$ | | |
| 2.a | $C < N$ | $1 \leq s \leq b$ | | |
| 2.b | $C < N$ | $b \leq s \leq N/a$ | | |
| 2.c | $C < N$ | $N/a \leq s \leq N/2$ | | |

- b. With these expressions in hand, now identify the cache and TLB characteristics of the machine, specifically (i) the cache size, (ii) the line size, (iii) the cache associativity, (iv) the cache miss cost, (v) the page size, (vi) the TLB associativity, and (vii) the TLB miss cost.

The TLB appears to the microbenchmark as another cache with a line size corresponding to the page size, and with much larger miss latencies. Your answer should identify the 4 behaviors you analyzed as part of (a) in the graph (make a copy of the graph on the previous page and add labels to it).

Part 2:

Now run the microbenchmark program on at least two different machines you have access to. If you can, you should run the program shortly after a reboot: the curves you obtain are more likely to resemble the ideal curves depicted in Part 1.

For each of these machines, analyze the curves using the expressions derived above to identify the characteristics of the different levels of caches in each system. Justify your findings and compare them with what is publicly known for the machines you are running the benchmark on. You should be able to find information about the latter from the web. Note also that being off by a factor of 2 or so is expected: the microbenchmark results can be interpreted somewhat subjectively.

The solution you turn in should also discuss possible reasons for variations from the ideal scenario depicted in Part 1.

Submission Instructions

E-mail your solution in an electronic form (PDF preferred) to the instructor (mb@cs.nyu.edu) by the due date. There is no need to include the raw data of your measurements, just include the plots showing the curves formed by these measurements.