

Assignment 3

Spiros-Daniel Mavroidakos - 260689391

Jastaj Virdee - 260689027

November 29, 2018

Contents

1	Question 1	1
1.1	1
1.2	1
1.3	1
1.4	1

1 Question 1

1.1

From the question, we know the following.

- We are reading L words from cache
- Cache Line is 4 words

L' is the number of words that can be stored in the cache given a maximum word separation of stride s . A stride can be a maximum size of $\left\lfloor \frac{L}{2} \right\rfloor$. We also know that the cache can have several cache lines. We will denote the amount of cache lines by k . Knowing this, we can come up with the following equation for L' .

$$L' = \left\lceil \frac{4k}{\left\lfloor \frac{L}{2} \right\rfloor} \right\rceil \quad (1)$$

Furthermore, t_0 shows the cache access latency as it does take some time to read from the cache. Given that $L \leq L'$, the stride size will either be 0 or a size that fits within the cache. Therefore, all the time delay for access comes from reading the cache.

1.2

Given that $L \leq L'$, the stride will be large enough that we cannot only read from the cache. The words will be too separated to read only from cache. However, at t_1 , we have peaked in the maximum read time. This implies that the stride is larger than the actual cache size. That is why it is a constant line. t_1 will therefore be showing the memory access latency since we are no longer reading from the cache. t_1 might also take into account the time to read the cache as we need to generate a cache miss to read from other storage.

1.3

Part 1 is when $L \leq L'$ so as stated in section 1.1, all of the data of the array will fit into the cache and will therefore be accessed at the time that it takes to read the cache. This is why it is the lowest of the 2 curves.

Part 2 is when $L \leq L'$ however, the stride has not become so big that we cannot use the cache in a meaningful way. Throughout part 2, the stride is continuously increasing which is why the curve is increasing as you must read from cache and then main memory. At the end of part 2 (the peak value where it becomes constant) is where part 3 begins and when the stride has become so large that the stride takes up all of the cache and you will always incur a cache miss and have to then go to main memory which is why it is a peak. Part 2 has a mix between some values are in cache and some are not while part 3 is where everything is not in the cache.

1.4