# HW 1(B): CACHE BASICS

DIHAN DAI

January 15, 2020

## Question 1 : Canvas 1(a)

$$
\begin{aligned}
&m = \log 256 = 8 \\
&b = \log B = \log 4 = 2 \\
&s = \log S = \log 4 = 2 \\
&t = m - b - s = 8 - 2 - 2 = 4
\end{aligned} \tag{1}
$$

## Question 2

The number of sets is $S = \frac{C}{BE}$ and the size of the set index is $s = \log S = \log \frac{C}{BE}$ bits. Thus the size of the tag is $t = (k - b - \log \frac{C}{BE})$ bits. The total capacity is

$$
C = SE(B + 0.125(t+1)) = SE(B + 0.125(k - b - \log \frac{C}{BE} + 1)) \text{ bytes} \tag{2}
$$

## Question 3 : Canvas 1(c)

$$
\begin{aligned}
&m = \log 256 = 8 \\
&C2 = C1 = 4 \times 4 \times 1 = 16 \\
&S = \frac{C2}{B \times E} = \frac{16}{4 \times 2} = 2 \qquad b = \log B = \log 4 = 2 \\
&s = \log S = \log 2 = 1 \\
&t = m - b - s = 8 - 2 - 1 = 5
\end{aligned} \tag{3}
$$

## Question 4

The total number of references in L1 cache is 10size

(a) In the first "it" loop, the caches are empty at the beginning. Thus block 1 of L2 cache is loaded from memory, and block 1 of L1 cache is loaded from the L2 cache. Since size is less than 8K, all the data are loaded in L1 cache after this loop and in the following "it" loop CPU can process data directly from L1 cache. Thus the total number of misses in L1 cache is $\lceil \frac{\text{size}}{16} \rceil$ and the total number of misses in L2 cache is $\lceil \frac{\text{size}}{32} \rceil$. The average memory access time is

$$AMAT = t_1 + \frac{\lceil \frac{\text{size}}{16} \rceil}{10\text{size}} \left( t_2 + \frac{\lceil \frac{\text{size}}{32} \rceil}{\lceil \frac{\text{size}}{16} \rceil} t_m \right)$$

(b) Since the size of the array is between 8K and 64K, the data are loaded in L2 cache after the first "it" loop, but for L1 cache, the data need to be loaded from L2 cache in every "it" loop. Thus the total number of misses in L1 cache is $10\lceil \frac{\text{size}}{16} \rceil$ and the total number of misses in L2 cache is $\lceil \frac{\text{size}}{32} \rceil$. The average memory access time is

$$AMAT = t_1 + \frac{\lceil \frac{\text{size}}{16} \rceil}{\text{size}} \left( t_2 + \frac{\lceil \frac{\text{size}}{32} \rceil}{10\lceil \frac{\text{size}}{16} \rceil} t_m \right)$$

(c) Since the size of the array is greater than 64K (the size of L2 cache), in every "it" loop, the data need to loaded in L2 cache from the memory and the data need to be loaded in L1 cache from the L2 cache. Thus the total number of misses in L1 cache is $10\lceil \frac{\text{size}}{16} \rceil$ and the total number of misses in L2 cache is $10\lceil \frac{\text{size}}{32} \rceil$. The average memory access time is

$$AMAT = t_1 + \frac{\lceil \frac{\text{size}}{16} \rceil}{\text{size}} \left( t_2 + \frac{\lceil \frac{\text{size}}{32} \rceil}{\lceil \frac{\text{size}}{16} \rceil} t_m \right)$$