Istanbul Technical University

Department of Computer Engineering

BLG 322E - COMPUTER ARCHITECTURE

Assignment 4

Due Date: Wednesday, May 8, 2024, 23:59.

- Please **write your full name** (first name and last name) **and Student ID** at the top of your solution.
- Please show ALL your work. Answers with no supporting explanations or work will not receive any partial credit. Your homework is not just a final report of your results; we want to see your steps. Upload all the papers you worked on to get to the solution.
- **Submissions:** Submit your solution as a **PDF file** to Ninova before the deadline.
- **No late submissions** will be accepted. Do not send your solutions by e-mail. We will only accept files that have been uploaded to the official Ninova e-learning system before the deadline. Do not risk leaving your submission to the last few minutes.
- **Consequences of plagiarism/cheating:** Assignments have to be done individually. Any cheating will be subject to disciplinary action.

If you have any questions, please e-mail **M. Alpaslan Tavukçu (**tavukcu22@itu.edu.tr**).**

QUESTION 1 (60 Points)

A computer system includes **256 KB** (B: Byte) **(physical address is 18 bit)** main memory and a cache memory that can hold **1 KB** data. Data transfers between main and cache memories are performed using blocks of **16 bytes**. The cache control unit uses set associative mapping technique where each set contains **two frames** (2-way set associative). In necessary cases FIFO is used as a replacement technique.

a) In what fields is the physical address divided by the cache control unit? Give the lengths of the fields. **(5 Points)**

b) What is the size of the tag memory (how many rows, the length and contents of each row)? **(5 Points)**

c) The CPU runs the piece of code given below. This program iterates over dynamically allocated byte matrix **A** (array of arrays) with the shape of 8x10. **Each iteration of the program involves accessing data at the byte level, indicating that the program processes data byte by byte (data is accessed in bytes):**

```
for i = 0 to rows-1:
   for j = 0 to cols-1:
      Read A[i][j]
```

The starting address and sizes of the arrays:

ID	Starting Address	Size
A[0] (A[0][0])	$00210	10 bytes
A[1] (A[1][0])	$00410	10 bytes
A[2] (A[2][0])	$00820	10 bytes
A[3] (A[3][0])	$01020	10 bytes
A[4] (A[4][0])	$02010	10 bytes
A[5] (A[5][0])	$04010	10 bytes
A[6] (A[6][0])	$08020	10 bytes
A[7] (A[7][0])	$10020	10 bytes

Assume that the cache memory is empty. Which arrays are placed in which frames of the cache memory over the iterations? What is the total number of the replacements? **(20 Points)**

d) What is the total number of replacements when column-wise iteration is used (like code given below)? Which arrays are placed in which frames of the cache memory over the iterations? What is the total number of the replacements? Compare with row-wise, discuss briefly. **(20 Points)**

```
for j = 0 to cols-1:
   for i = 0 to rows-1:
      Read A[i][j]
```

e) To increase the hit ratio, what can be done? (Hint: Consider changing the starting addresses of the arrays.) Please explain briefly. **(10 Points)**

QUESTION 2 (40 Points)

A computer system includes **64 KB** (B: Byte) **(physical address is 16 bit)** main memory and a cache memory that can hold **1 KB** data. Data transfers between main and cache memories are performed using blocks of **8 bytes**. The cache control unit uses set associative mapping technique where each set contains **two frames** (*2-way set associative*). In necessary cases **FIFO** is used as a replacement technique. In read operations **Read Through**, in write operations **Write Through** (WT) with **Write Allocate** (WA) methods are used. Cache memory is used only for data, not for instructions. The cache access time is $\mathbf{t_c}$ ns, the main memory access time is $\mathbf{t_m}$ ns and the block transfer time between cache and main memories is $\mathbf{t_B}$ ns. Assume that the cache memory is empty at the beginning. The CPU performs the following arithmetic operation **C = A + B**, where each variable is **one** byte.

a) Assign proper main memory addresses to three variables in this operation A, B, C so that they can be accessed as **fast** as possible, give exemplary addresses. What is the total time spent, in terms of $\mathbf{t_c}$, $\mathbf{t_m}$ and $\mathbf{t_B}$, to access these variables in this case? **(20 Points)**

b) Assign proper main memory addresses to three variables in this operation A, B, C so that they can be accessed as **slow** as possible, give exemplary addresses. What is the total time spent, in terms of $\mathbf{t_c}$, $\mathbf{t_m}$ and $\mathbf{t_B}$, to access these variables in this case? **(20 Points)**