國立中正大學
111 學年度碩士班招生考試
試 題

[第 1 節]

科目名稱	計算機組織
系所組別	電機工程學系－ 計算機工程組 晶片系統組

－作答注意事項－

※作答前請先核對「試題」、「試卷」與「准考證」之系所組別、科目名稱是否相符。

1. 預備鈴響時即可入場，但至考試開始鈴響前，不得翻閱試題，並不得書寫、畫記、作答。
2. 考試開始鈴響時，即可開始作答；考試結束鈴響畢，應即停止作答。
3. 入場後於考試開始 40 分鐘內不得離場。
4. 全部答題均須在試卷（答案卷）作答區內完成。
5. 試卷作答限用藍色或黑色筆（含鉛筆）書寫。
6. 試題須隨試卷繳還。

國立中正大學 111 學年度碩士班招生考試試題

科目名稱：計算機組織　　本科目共 2 頁　第 1 頁

系所組別：電機工程學系-計算機工程組、晶片系統組

1. (20%) Draw a single-cycle datapath for MIPS R-type instructions SW and BNE (including instruction decoder and control signals to multiplexors and state elements).

2. (25%) Given a 32-bit machine with byte-addressable memories.

   (a) (10%) Draw a four-way set-associative cache that has a total of 1024 lines with one word per line.

   (b) (5%) Calculate the total amount of cache memory in bytes.

   (c) (10%) Find the binary values of (1)tag, (2)index, and (3)byte-offset for byte address 1200 in the cache of (a).

3. (25%) Assume a 5-stage pipelined processor similar to MIPS is going to execute the following code fragment.

```
Loop: LW    R2, 0(R1)
      ADDI  R1, R1, 4
      ADD   R3, R3, R2
      SUBI  R4, R4, 1
      BNEZ  R4, Loop      # if R4 != 0 go to Loop
```

   The initial value of R4 is 10. Assume that all possible forwarding paths are present in the pipeline. The cache is direct-mapped with a line size of 4 words. Initially the cache has all entries marked as invalid. The cache has a 1-cycle hit time and takes an additional 10 cycles on a cache miss. The branch prediction unit predicts each branch as the same as its last seen outcome. Each branch mis-prediction yields a 2-cycle penalty.
   Note: answers without clear reasoning are unacceptable!

   (a) (15%) What is the runtime in cycles to execute the above code?

   (b) (5%) If the order of the ADDI and ADD instructions is switched in the code shown above, will the runtime change? If so, by how much?

   (c) (5%) If the order of the ADD and SUBI instructions is switched in the code shown above, will the runtime change? If so, by how much?

4. (30%) Assume a fully-connected neural network containing 1 input layer, 3 hidden layers, and 1 output layer.

   - The input layer has 256 nodes, each of which represents an 8-bit pixel of a 16*16 grayscale image.

- Each of the 3 hidden layers contains 256 neurons, which computes a weighted sum of all nodes in its precedent layer, adds a bias value, and performs ReLU activation (i.e. the result remains the same if it is positive; otherwise it becomes 0). In other words, the j-th neuron of the i-th layer: $x_{i,j}$ compute $max(\sum_{k=0}^{255} x_{i-1,k} \times w_{i,j,k} + b_{i,j}, 0)$.
- The output layer has 10 nodes, each of which represents a digit (i.e. 0, 1, 2, … , 9). The j-th output node $x_{4,j}$ computes $\sum_{k=0}^{255} x_{3,k} \times w_{4,j,k} + b_{4,j}$ without ReLU.

(a) (5%) What is the amount of storage (in bytes) that are needed if the weights and biases are both represented as IEEE 754 single-precision floating-point numbers?

(b) (10%) Due to cost issues, only a 1KByte SRAM is allowed to store the weights on chip (assume biases are handled independently), and the design team decides to implement a direct-mapped cache mechanism to simplify the management of on-chip (i.e. 1Kbyte SRAM) and off-chip (i.e. containing all weights) storages. Assume one cache block stores 32-byte data. What are the on-chip storage requirements in addition to the 1Kbyte SRAM for weights?

(c) (5%) What is the miss rate of the weight cache in (b)?

(d) (10%) Describe how to improve the weight memory organization in (b) under the same cost constraint. Note: answers without clear reasoning are unacceptable!