國 立 交 通 大 學 試 題 紙

一百零六學年度第二學期
博士班資格考

科目:計算機架構 A

日期： 107 年 7 月 30 日 第 1 頁 共 2 頁

請" ✓" 明 ✓不可看書 可看書

* 請將答案依題號順序寫入答案卷

答題時字跡需工整，否則不予計分。Write your answers legibly; otherwise you will get zero score.

1. (10%)
   (a) (2%) Describe *Amdahl's Law* and define the equation of the overall speedup using the original machine with the enhanced mode.
   (b) (4%) Describe the *weighted arithmetic mean* and *geometric mean* for summarizing the performance results of a benchmark suite, and explain in which circumstance each of them is more suitable to be applied.
   (c) (4%) Define the following two main measures of dependability and the equations with respect to *mean time to failure* (MTTF) and *mean time to repair* (MTTR): *module reliability* and *module availability*.

2. (10%) A 4-GHz processor was used to execute a benchmark program with the instruction mix and clock cycle counts shown in the following table.

Instruction type	Frequency	CPI
Integer ALU ops	30%	1
Floating-point ops	10%	25
Loads/Stores	40%	4
Branches	20%	2

   (a) (4%) Assume that the total number of instructions executed is 5×10^9, determine the *effective CPI*, and *execution time* of this program.
   (b) (3%) Assume that an enhancement proposal is to reduce the CPI of the floating-point operations to 9 with 20% lengthening of the clock cycle time. Calculate the *effective CPI* and the *speedup* of the enhancement.
   (c) (3%) Calculate the *fraction of the time* for executing load and store operations of this program in the original processor. Assume that we build an optimizing compiler to discard a half of the Load/Store operations from the original instruction mix, determine the *speedup* of the enhancement to the original design by applying *Amdahl's Law*.

3. (10%)
   (a) (3%) Describe the following two types of localities for memory hierarchy design and give one example for each type: *spatial locality* and *temporal locality*.
   (b) (3%) Describe the following three block placement approaches for cache and the block identification method corresponding to each approach: *direct mapped*, *set associative*, and *fully associative*.
   (c) (2%) Describe the following two write options for cache: *write through* and *write back*.
   (d) (2%) Describe the following two write miss options for cache: *write allocate* and *no-write allocate*.

◎請用深黑色鋼筆或原子筆出題

命題老師簽名：

國 立 交 通 大 學 試 題 紙

一百零六學年度第二學期
博士班資格考

科目:計算機架構 A

4. (10%) Typically, miss rate, miss penalty, hit time, and bandwidth are the main metrics to evaluate the performance of a cache memory. Describe each of the following cache optimizations and indicate its positive impact to the performance metrics:
   (a) *multilevel cache*
   (b) *nonblocking cache*
   (c) *virtual cache*
   (d) *merging write buffer*
   (e) *hardware prefetching*

5. (10%)
   (a) (2%) Describe the main idea of *vector architectures* for exploiting data-level parallelism.
   (b) (8%) Describe each of the following optimizations of vector architecture and the problem it solves:
      i. *vector-length register*
      ii. *vector mask register*
      iii. *stride*
      iv. *gatter-scatter*

◎請用深黑色鋼筆或原子筆出題

命題老師簽名：

國 立 交 通 大 學 試 題 紙

一百零六學年度第二學期
博士班資格考

科目:計算機架構 B

日期: 107 年 7 月 30 日 第 1 頁 共 4 頁

請"✓"明 ✓不可看書 可看書

* 請將答案依題號順序寫入答案卷

答題時字跡需工整，否則不予計分。Write your answers legibly; otherwise you will get zero score.

1. (15%) Pipelining is a well-known scheme to exploit instructional-level parallelism in executing machine programs.
   (a) Explain what pipelined execution actually means.
   (b) Compared with non-pipelined method, does pipelining make a single instruction run faster? In addition, does pipelining make a program run faster? Support your answers.
   (c) What typical pipeline stages do we have? (Use the stages that our reference textbook uses the most to answer this question. Make the stage names clear enough, or explain in detail, to show what each of these stages does. Be very careful not to leave out any necessary jobs in executing an instruction; that is, simply listing the stage names, especially the very simplified abbreviated stage names, will not be enough.)
   (d) To change a design from non-pipelined to pipelined, from the hardware or physical resource aspects, what costs must we pay? Be thorough, and support your answers well.
   (e) To use a pipeline to run a program well, from the software or system aspect, what costs must we pay? Again, be thorough, and support your answers well.

◎請用深黑色鋼筆或原子筆出題

命題老師簽名：

2. (8%) Given an assembly loop code, if the execution of an iteration in a particular pipeline takes 9 clock cycles, ignoring the delayed branch:

```
Loop: L.D    F0,0(R1)
      stall
      ADD.D  F4,F0,F2
      stall
      stall
      S.D    F4,0(R1)
      DADDUI R1,R1,#-8
      stall (assume DADDUI latency is 1)
      BNE    R1,R2,Loop
```

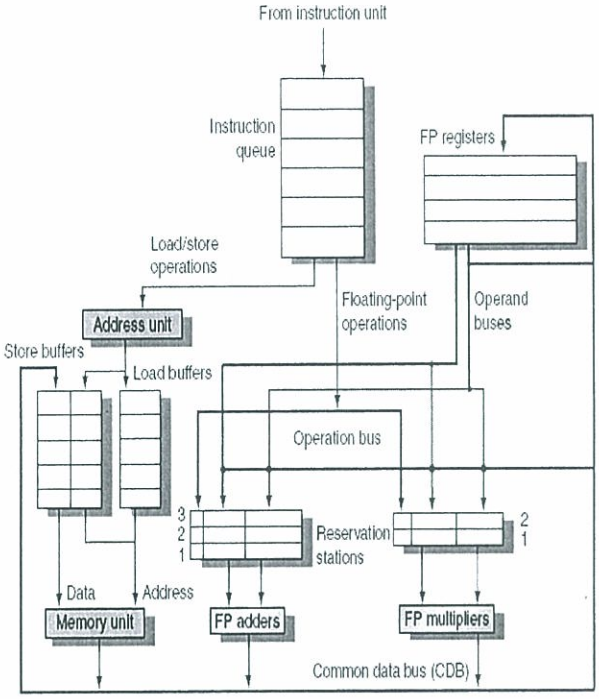
Instruction producing result	Instruction using result	Latency in clock cycles
FP ALU op	Another FP ALU op	3
FP ALU op	Store double	2
Load double	FP ALU op	1
Load double	Store double	0

(a) (3%) Reschedule the instructions to your best effort to speed the loop, showing how many cycles it takes.

(b) (5%) Unroll the loop the minimal number of times and schedule again to eliminate all stall cycles. You need to keep the loop body size to a minimum. Assume that the array length happens to be perfect for your unroll size. (Make any additional assumptions of yours only whenever necessary.)

3. (12%) Given the Tomasulo's algorithm diagram, and we know that it has two major features.

(a) Point out what these two major features are. Clearly explain what advantages each of these two features can bring us.
(b) Using the diagram, clearly explain how each of these two features is implemented.
(c) In some situations we want to put in a reorder buffer (ROB). Why do we do this?
(d) Continued from (c), explain how the system works so that it can now fulfill the requirement(s) making us put in the ROB (explicitly noting if any of the original Tomasulo's algorithm function(s) have been offloaded to other components).



◎請用深黑色鋼筆或原子筆出題

命題老師簽名：

國 立 交 通 大 學 試 題 紙

一百零六學年度第二學期
博士班資格考

科目:計算機架構 B

日期: 107 年 7 月 30 日

4. (15%) In multiprocessor architecture, there can be centralized shared-memory multiprocessors and distributed-memory multiprocessors. In accessing memory, on the other hand, there can be uniform memory access (UMA) and non-uniform memory access (NUMA) types.
   (a) Indicate which type multiprocessors exhibit what memory access features in general. Support your answers.
   (b) Indicate which type multiprocessors are suitable for what type programs. Support your answers.
   (c) Indicate the most noteworthy advantage(s) of each of the multiprocessor types.
   (d) Which of the two types of multiprocessors can contain more processors? Why?
   (e) Which of the two types of multiprocessors can be easier to program on? Why?

(Note: An incorrect or irrelevant statement in your answer can cost you all points in that sub-question.)

◎請用深黑色鋼筆或原子筆出題

命題老師簽名：