國立陽明交通大學 111 學年度碩士班考試入學試題

科目：計算機系統(1103)　　考試日期：111 年 2 月 9 日 第 3 節

系所班別：資訊聯招　　第 1 頁，共 9 頁

【不可使用計算機】＊作答前請先核對試題、答案卷(試卷)與准考證之所組別與考科是否相符！！

請使用答案卡作答

一、複選題(80%)，共二十題，每題全對得 4 分，答對一個選項得 1 分(答對兩個選項得 2 分，以此類推)，答錯一個選項倒扣 1 分(答錯兩個選項倒扣 2 分，最多扣至本科目計算機系統 0 分為止)，整題未作答不給分。(舉例每題有 a、b、c、d 四個選項，某題答案為 a、b，而作答時選 a、b、c，則三個選項正確一個錯誤，該題得 3-1=2 分)

1. Consider the following C code. Which outputs are possible?
   (a) ABCED.
   (b) AEBCD.
   (c) ABDCE.
   (d) AEBDC.

```
int main() {
  printf("A");
  if (fork() == 0) {
    usleep(((rand() % 5) + 1) * 10000); /* wait for a random time */
    printf("B");
    if (fork() == 0) {
      usleep(((rand() % 5) + 1) * 10000); /* wait for a random time */
      printf("C");
    }
    else {
      wait(NULL);
      printf("D");
    }
  }
  else {
    usleep(((rand() % 5) + 1) * 10000); /* wait for a random time */
    printf("E");
    wait(NULL);
  }
  return 0;
}
```

2. Consider a system with 15 resources and 4 threads: T_0, T_1, T_2, and T_3. The second column in the following table shows the maximum number of resources need by each thread and the third column shows the number of resources each thread is currently holding. Let $\hat{x}$ be the maximum value of x (shown below) that can still keep the system in a safe state. Which statements are correct?

	Maximum Needs	Current Needs
T_0	5	3
T_1	4	1
T_2	9	5
T_3	x	4

   (a) $\hat{x} = 13$.
   (b) With $\hat{x}$, $\langle T_0, T_1, T_2, T_3 \rangle$ is a safe sequence.
   (c) Even for some value of x that makes the system in an unsafe state, it may not lead to a deadlock.
   (d) The way we use here to see whether a system is safe only applies to one type of resource. The banker's algorithm extends it to multiple types of resources.

3. Which statements about semaphore are correct?
   (a) Two processes sharing two common binary semaphores may result in a deadlock.
   (b) If we use binary semaphore to implement a mutex lock, the semaphore should be

initialized to 0.

(c) If we invoke a `x.signal()` operation on a condition variable x but no process is suspended on x, then the state of x remains unchanged.

(d) If we invoke a `Signal(S)` operation on a semaphore S but no process is suspended on S, then the state of S remains unchanged.

4. A process could be in new, running, waiting, ready, or terminated state. Which statements about process states are correct?
   (a) If a process invokes a blocking `send()` for message passing, the process will enter the waiting state.
   (b) When a process must wait on a spin-lock, the process will enter the waiting state.
   (c) A process kept staying in the waiting state will not incur context-switching overhead.
   (d) A `wakeup()` operation changes the state of a process from the waiting state to the running state.

5. Which statements about CPU scheduling are correct?
   (a) Compared with first-come first-serve (FIFO), nonpreemptive shortest-job-first scheduling results in fewer context switches.
   (b) Shortest-remaining-time-first scheduling can lead to starvation.
   (c) A voluntary context switch occurs to a process when its time slice has expired.
   (d) Processor affinity, which tries keeping a thread running on the same processor, is a load-balancing technique for multiple-processor scheduling in a symmetric multiprocessing (SMP) system.

6. In addition to a reference bit, a modify bit can be added to form the bit pair (reference, modify) associated with a page to improve the second-chance algorithm used in page replacement, where the reference bit is 1 if the page has been referenced and the modify bit is 1 if the page has been updated in the past round. Divide the pages into four classes: (0,0), (0,1), (1,0), (1,1).
   (a) Both reference bit and modify bit, if being 1, are cleared to 0 on a circular round of the page replacement algorithm, as a way to give the second chance.
   (b) Pages in class (0,0) should be replaced first, while the pages in class (1,1) should be replaced last.
   (c) Pages in class (0,1) should be replaced before pages in class (1,0).
   (d) Pages in class (0,1), if picked to replace and dirty, should be written out before replacement.

7. Consider memory-mapped files where disk blocks of a file are mapped to pages in memory, so that file I/O can be treated as routing memory access.
   (a) To load a file into the memory, the mmap() system call loads initially the entire file into the memory pages.
   (b) By default, a memory-mapped file is mapped into the user process address space, instead of the kernel space.
   (c) After invoking mmap(), read() and write() system calls are automatically translated by the kernel into memory access instructions.
   (d) If the standard I/O is memory-mapped by mmap(), the mapping will be to kernel address space instead of user address space, and thus read() and write() would trigger memory copy between kernel space and user space.

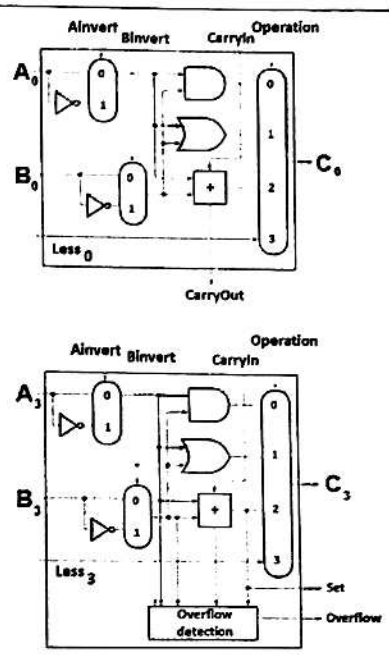
8. A file can be stored in a file system with different allocation methods: contiguous, extent-based, linked, or indexed. Suppose directories are already in memory. An extent has an average of 16 blocks. Assume that a file has 512 blocks, numbered from 0 to 511, and the target of a random access read is block 239.
   (a) With contiguous allocation, only one block need to be read in order to access block 239.
   (b) With extent-based allocation, 15 blocks need to be read in order to access block 239.
   (c) With linked allocation, 240 blocks need to be read in order to access block 239.
   (d) With indexed allocation, 2 blocks need to be read in order to access block 239.

9. Consider UNIX UFS with 4 KB blocks. In an inode, suppose there are 2048 entries in the "direct block" pointing to at most 2048 data blocks. There is also an entry for each of "single indirect", "double indirect", and "triple indirect", pointing to an indirect node with 2048 entries for data blocks or recursively indirect nodes.
   (a) With only direct block, the maximum file size is around 8 MB.
   (b) With direct block and single indirect, the maximum file size is around 16 MB.
   (c) With direct block, single indirect, and double indirect, the maximum file size is around 16 GB.
   (d) With direct block, single indirect, double indirect, and triple indirect, the maximum file size is around 32 TB.

10 Consider various mechanisms used in Linux, which statements are valid?
   (a) A dynamically loaded kernel module will not be unloaded automatically by the kernel unless explicitly requested.
   (b) The kernel maintains a table of functions for loading programs of different binary formats, and try them one by one when loading an executable binary.
   (c) Slab allocation is for contiguous memory allocation needed for user processes.
   (d) For a library of functions statically linked to multiple programs, it is loaded into memory multiple times as multiple copies, but only once and one copy if dynamically linked.

11. For instruction set architecture (ISA), which of the following statements are true?
   (a) RISC CPUs usually use fixed and simple instruction formats, which may offer better execution efficiency for pipelining than CISC CPUs.
   (b) The choice of Big/Little endian determines the memory ordering and is nothing to do with the ISA design.
   (c) The purpose of Open ISA is to allow "open source" of CPU RTL codes.
   (d) An ISA keeping the source (rs1 and rs2) and destination (rd) registers at the same position in all formats can simplify decoding (such as RISC-V), whereas MIPS has overlapped fields for rs2 and rd, which needs multiplexors.

12. Given the normal 1-bit ALU (top) and the MSB 1-bit ALU (bottom), you can implement the C-language statement by giving tuple (Ainvert, Binvert, CarryIn, Operation). Each A,B,C represent 4 bits wide. For condition determination, C is 0001 if TRUE, and is 0000 if FALSE.



E.g. Give (0, 0, 0, 2) for C=A+B

(a) By giving appropriate tuples, we can make possible statements: C=A-B; C=B-A, but not possible for C=A-B-1 and C=B-A-1, which need an additional substractor.

(b) For C = A<B, let us give (0, 1, 1, 3) and also connect set to $Less_0$. All other $Less_{1\sim3}$ are set 0.

(c) For C = A>B, let us give (0, 1, 1, 3) and also connect ($\overline{set}$) to $Less_0$. All other $Less_{1\sim3}$ are set 0.

(d) For D = A==B, let us give (0, 1, 1, 2) and also connect all $C_{0\sim3}$ to a NOR gate to obtain the final D_0. All $Less_{0\sim3}$ are set 0 and connect other $C_{1\sim3}$ to $D_{1\sim3}$.

13. Considering both DMA(direct memory access) and I/O, which of the following statements are true?

(a) If a machine cannot handle interrupt, the system will not be able to perform I/O via DMA.

(b) By using memory-mapped I/O, CPU employs DMA to release the requirements of keeping busy in I/O movement.

(c) DMA is an I/O processor, which can accept I/O requests from CPU and acts as a slave device on a bus.

(d) The primary problem that DMA will create with a cache memory system is because DMA uses physical addresses whereas CPU caches may use virtual addresses.

The following text is excerpted from an article on the web about CPU performance.

"The total clock speed or CPU performance of a device indicates how many processing cycles per second can be executed by a CPU, considering all of its processing units (more commonly referred to as cores). The CPU speed is measured in gigahertz (GHz), with each gigahertz being equivalent to one thousand million cycles per second. Given the same clock rate, more CPU parallelism can contribute better performance. However, power dissipation and circuit temperature usually constrain modern processors clock frequency and thus they can be dynamically changed so that circuit temperature stays below a certain limit.

When considering the total CPU speed, let us look for more GHz and a higher core count. Although this is not the single determining factor for an electronic device's performance, high processor speed is crucial if you are looking to perform demanding tasks such as gaming or video editing. As for the core count, a CPU chip with multiple processing units has the advantage of being able to run various tasks at the same time.

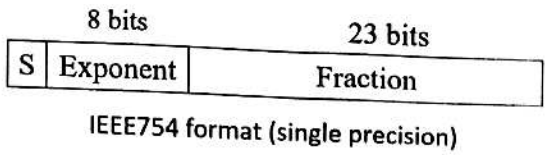
Nowadays, a CPU chip can comprise cores with varying speeds, thanks to ARM's big.LITTLE technology; the advantage is that tasks are assigned to different cores depending on how demanding they are, optimizing system usage, and preventing overheating. For instance, sending e-mails doesn't require much processing power. Instead of assigning this task to a sophisticated core, the system will delegate these tasks to LITTLE cores that produce less heat than the BIG cores with an aggressive execution such as multi-issue pipelines, hardware multithreading or out-of-order execution. Besides, even a single-core may be able to operate on different clock speeds: namely "base" and "turbo."

The current victor when it comes to total clock speed for computers is AMD's Epyc 7742, flaunting an overall clock rate of 64 x 2.25GHz (64 cores with 2.25 GHz each). The fierce competition among manufacturers means that higher CPU performance is a never-ending race."

14. For CPU clock rate, which of the following statements are true?
    (a) Executing programs on a 1.5GHz CPU could be faster than on another CPU of 2GHz even within the same brand.
    (b) Powerful branch prediction is essential for CPU with a high clock rate so as to minimize the penalty due to branch hazard.
    (c) Given that the big.Little consisting of two cores, one big core can run at 4GHz and the Little core runs at 2GHz. The average clock speed will be 3GHz.
    (d) Unbalance of pipeline stages affects cycle time (thus clock rate). We can push more clock rate by adding more stages and thus improve the CPU throughput without increasing the instruction latency.

15. Considering CPUs with full parallelism, which of the following statements are true?
    (a) To fully utilize AMD 64-core CPUs, the software has to exploit full parallelism by both multiprocessing and multithreading.
    (b) In big.LITTLE technology, the Little cores will keep pipeline as deep as possible to achieve higher clock rate, more focused on ILP (instruction level parallelism). In contrast, big cores usually embed TLP (task level parallelism) such as to maintain a low clock rate.
    (c) With two clock rates in a core, the *base* clock time is determined by the longest stage of the pipeline and the *turbo* mode is limited by the maximum clock generator at run time.
    (d) CPU parallelism can be obtained by multi-issue, SIMD, MIMD, and vector techniques. The multi-issue, SIMD, and vector need compiler optimization, whereas MIMD needs supports from OS.

16. The single-precision (32-bit) IEEE 754 floating-point format is as follows. Which of following statements are correct?
    (a) To represent 3.125 in the IEEE754 format, the exponent part is 00000001b.
    (b) To represent 3.125 in the IEEE754 format, the fraction part is 480000h.
    (c) $1.101_2 \times 2^{-130}$ cannot be represented using the IEEE754 format.
    (d) To represent $1.101_2 \times 2^{-130}$ in the IEEE754 format, the exponent part is 00000000b.



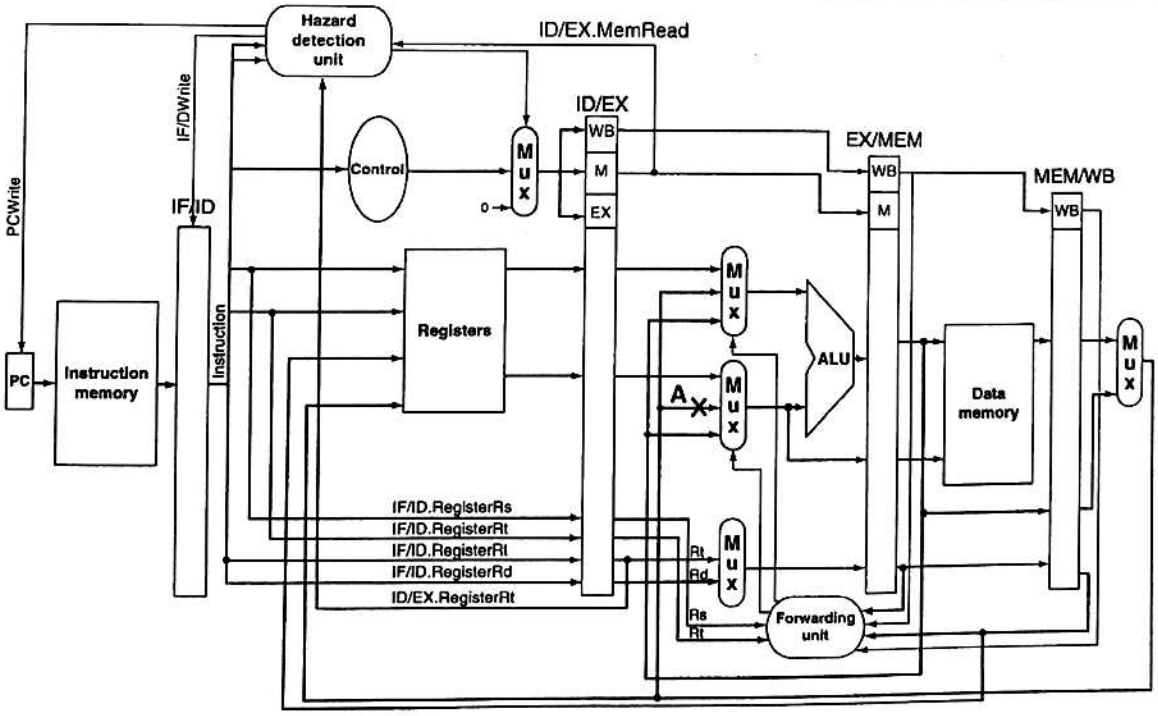
IEEE754 format (single precision)

Single Precision		Meaning
Exp.	Significant	
0	0	0
0	Non-zero	+/- de-normalized number
1-254	Anything	+/- normalized floating-point number
255	0	+/- infinity
255	Non-zero	NaN (Not a number)

17. Consider a five-stage piplelined MIPS CPU. Please select the signals that are required to insert a stall cycle when data hazard is detected at ID stage.
    (a) Set IF.Flush = 1 (to clear IF/ID)
    (b) Set PCWrite = 0 (to disable the write to Program Counter (PC))
    (c) Set IF/IDwrite = 0 (to disable the write to IF/ID)
    (d) Set ID.Flush = 1 (to set 9 control signals used in EX, MEM and WB stages to 0)

18. Given the pipelined CPU in the figure below, which code sequence(s) will fail to run correctly if the data-path A is cut?

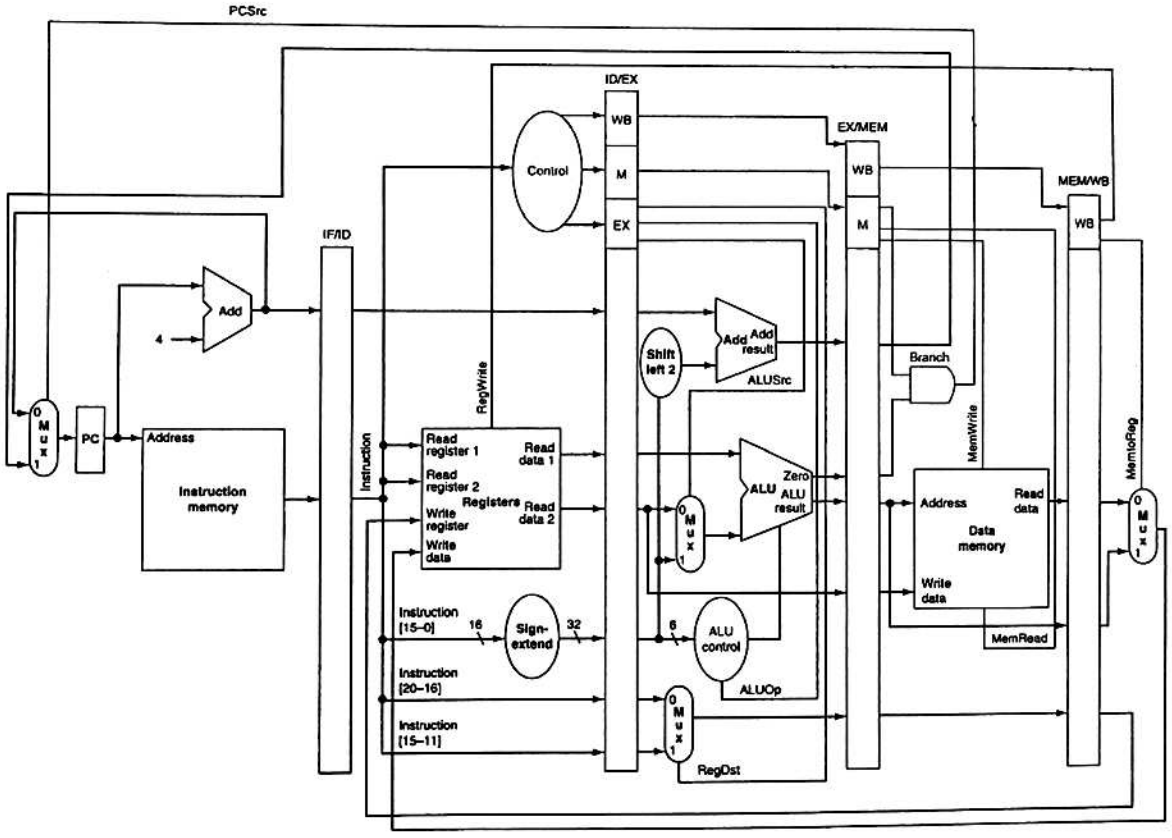
(a)	(b)	(c)	(d)
add $s1, $t0, $t1 . add $s3, $s2, $s1 . add $s2, $s3, $s1 .	add $s3, $t0, $t1 . add $s1, $s2, $s1 . add $s2, $s3, $s1 .	add $s1, $t0, $t1 . add $s1, $s2, $s1 . add $s2, $s3, $s1 .	add $s1 $t0, $t1 . add $s2, $s2, $s1 . add $s2, $s3, $s1 .



19. For a MIPS 5-stage pipelined CPU with data forwarding and stall mechanisms been implemented, which of following statements are true for the code sequence below? You may refer to the following code sequence and the CPU pipeline figure.
    (a) It takes 12 clock cycles to finish the execution of the code sequence.
    (b) There are total 2 stall cycles when running the code sequence.

(c) The instruction in EXE stage is *addi* when the code runs to the 5th cycle.
(d) The ALUsrc signal is set to 0 when the code runs to the 5th cycle.

```
add  $s2, $s2, $s3
lw   $s1, 0($s2)
addi $t1, $s2, 8
sw   $t2, 0($s1)
lw   $t3, 0($s1)
add  $s3, $t3, $s2
sub  $s3, $s3, $s2
```



20. For the code sequence written in MIPS assembly, assume that the $s2, $t1 are initialized to 0x0, and 0x3, respectively. Which of following statements are correct for the execution of this code sequence?
(a) Assume 1-bit predictor is used and the predictor is initialized as *taken*. For the inner ***bne*** (marked as *branch1*), the rate of prediction accuracy is 2/3.
(b) Assume 2-bit predictor is used and the predictor is initialized as *taken*. For the inner ***bne*** (marked as *branch1*), the rate of prediction accuracy is 5/6.
(c) Assume 1-bit predictor is used and the predictor is initialized as *taken*. For the outer ***bne*** (marked as *branch2*), the rate of prediction accuracy is 1/3.
(d) Assume the predictors are initialized as *taken*. For the outer ***bne*** (marked as *branch2*), 2-bit predictor has higher accuracy rate than 1-bit predictor.

```
Loop:  addi $t2, $zero, 0x6.
Loop2: addi $s2, $s2, 1.
       addi $t2, $t2, -1.
       bne $t2, $zero, Loop2   //branch1.
       addi $t1, $t1, -1.
       bne $t1, $zero, Loop    //branch2.
Done:
```

二、題組(20%)，共四個題組，每一題組內所有小題完全答對得 5 分，答錯任一小題或未作答得 0 分。題組 A 包含題號 21~24，題組 B 包含題號 25~28，題組 C 包含題號 29~32，題組 D 包含題號 33~35。

A. Consider using semaphores for the first readers-writers problem, which allows either one writer or at most *n* readers in the critical session at the same time (no reader be kept waiting unless a writer has already obtained permission to use the shared object). We use an integer `read_count` to count the number of readers that are currently reading the data. We use two binary semaphores: `rw_mutex` and `mutex`. The former is common to both reader and writer processes. The `mutex` semaphore is used to ensure mutual exclusion when the variable `read_count` is updated. The reader's code is shown below.

```
sem_wait(&①);
read_count++;
if (read_count == ②)
   sem_wait(&③);
sem_post(&④);
(read the data)
sem_wait(&⑤);
read_count--;
if (read_count == ⑥)
   sem_post(&⑦);
sem_post(&⑧);
```

21. What is the initial value of `mutex`? (a) 0 (b) 1 (c) *n* (d) none above.
22. What should be in ①? (a) `mutex` (b) `rw_mutex` (c) `read_count` (d) *n*
23. What should be in ②? (a) 0 (b) 1 (c) *n* (d) none above.
24. What should be in ⑦? (a) `mutex` (b) `rw_mutex` (c) `read_count` (d) *n*

B. Compare Oracle SPARC and ARM in supporting the page table. The former adopts 2-level caching and a hashed table with the feature of span (the number of pages an entry represent), while the latter uses also 2-level caching and a hybrid 1/2-level hierarchical page table (1-level for big pages and 2-level for small pages). Suppose 100 ns for memory access, 10 ns for 2nd cache access with cache hit ratio of 90%, and 1 ns for 1st cache access with cache hit ratio of 60%, and the remaining 10% for table lookup in the memory. Note that the 1st cache is looked up first before the 2nd cache, and then the table in the memory.

25. What is the minimum access time for a desired word in a frame? (a) 11 ns (b) 101 ns (c) 111 ns (d) 201 ns
26. In ARM, what is the maximum access time for a desired word in a frame? (a) 201 ns (b) 211 ns (c) 311 ns (d) 411 ns

27. In SPARC, if the maximum length of a linked list entry is 4, what is the maximum access time for a desired word in a frame? (a) 301 ns (b) 501 ns (c) 611 ns (d) 811 ns
28. In ARM, for a large process using big pages, what is the average access time for a desired word in a frame? (a) 33 ns (b) 60 ns (c) 115 ns (d) 215 ns

C. You have designed a brand new CPU: NYCU (New Younker Compute Unit). We know:

$$real\ \ CPI = ideal\ \ CPI + data\ \ access\ \ prob * avg\ \ stall\ \ cycles$$

$$avg\ \ stall = miss\ \ rate * miss\ \ penalty$$

Assume that NYCU can run at 2GHz and its lw/sw share 25% of total N instructions. Let us ignore the instruction cache issue and ideal CPI is 1. The access time of main memory is 100 ns.

29. We build a direct-mapped L1 data cache to match the NYCU1 pipeline. Assume that the miss rate is 4% for each cache access. What is the real CPI of NYCU1?
    (a) 2 (b) 3 (c) 4 (d) 5
30. Now we build a 4-way set-associative L1 data cache for new NYCU2 pipeline. Which of the following effects is NOT true for the new L1 compared to NYCU1?
    (a) the CPU clock rate could be reduced
    (b) better access time since address tags can be compared in parallel
    (c) an extra multiplexor to select blocks
    (d) conflict cache misses can be reduced
31. Now we design a NYCU3 by providing a 128 KB L2 to NYCU1, where L2 has a hit time of 8 ns and a miss rate 20%. What is the real CPI of NYCU3?
    (a) 0.28 (b) 0.56 (c) 1.28 (d) 1.56
32. We have to consider TLB in cache accessing for the NYCU memory hierarchy. Which of the following statement is NOT true for TLB?
    (a) TLB= translation-lookaside buffer (b) TLB miss will result into a page fault
    (c) small TLB uses fully-associative (d) TLB is accessed before L1 (physical addr)

D. Given the MIPS code sequence below, if we assume it starts at location 8800 4000h in memory. Which of following statements are true?

8800 4000h

```
        add     $t0, $zero, $zero
loop:   beq     $s2, $zero, finish
        add     $t0, $t0, $s1
        sub     $s2, $s2, 1
        j       loop
finish: sll     $t0, $t0, 2
        add     $v0, $t0, $zero
```

33. Assume the register number of $*s2* is 18 and the OP code of ***beq*** is 4. The MIPS machine codes of the ***beq*** in this code sequence is (a) 0x12400003 (b) 0x12400004 (c) 0x1240000C (d) 0x12400010
34. The MIPS machine code of the ***j*** (OP code is 2) in this code sequence is (a) 0x 08001001 (b) 0x22001001 (c) 0x0a001001 (d) 0x02001001
35. Assume both $s1 and $s2 initially contain integers 5 and 6, respectively. What is the value that $v0 will contain after the execution of the whole code sequence? (a) 6 (b) 30 (c) 60 (d) 120