※ 注意：請於試卷上依序作答，並應註明作答之部份及其題號。

PART I: COMPUTER ARCHITECTURE

1. (15 pts) Based on their applications, computers can be divided into three categories: desktop, server, and embedded. Let us focus on embedded applications here, and please answer the following questions:
   (a) [2 pts] How would you define embedded applications?
   (b) [2 pts] List three examples and explain why they are embedded applications.
   (c) [2 pts] Which of the following techniques can be used to reduce the footprint (i.e. the memory usage) of an embedded application?
   A. code compression
   B. loop unrolling
   C. dynamic loaded library
   D. software pipelining
   (d) [3 pts] For the following program, which techniques could be used to improve the performance?

```
for( int i = 0; i<128 ; ++i ){
  a[i] = b[ i ] + c[ i ]
}
```

   A. SIMD (single-instruction multiple-data) operations
   B. loop unrolling
   C. memory access prefetch
   D. larger memory capacity
   (e) [3 pts] What are the digital signal processors? How would an embedded application take advantage of a digital signal processor?
   (f) [3 pts] Multi-core embedded systems have become popular. Are they mostly homogeneous or heterogeneous? Why?

2. (15 pts) Memory cycle times have been much longer than processor cycle times, which is also known as the processor-memory speed gap. Please answer the following questions:
   (a) [3 pts] Which of the following techniques can be used to overcome the processor-memory speed gap?
   A. cache memory
   B. replacing DRAM with SRAM
   C. flash memory
   D. moving data from ROM to RAM
   (b) [3 pts] When you increase the processor-memory bus from 32-bit to 64-bit, which of the following are true?
   A. latency for loading a 32-bit word from main memory remains unchanged
   B. more memory modules are needed
   C. the bus bandwidth is doubled
   D. the memory address space increases
   (c) [3 pts] For a virtual memory system, which of the following might happen during a memory access?
   A. Hit in TLB, Hit in Page Table, Miss in Cache
   B. Hit in TLB, Miss in Page Table, Miss in Cache
   C. Miss in TLB, Miss in Page Table, Hit in Cache
   D. Miss in TLB, Hit in Page Table, Hit in Cache
   (d) [3 pts] Suppose we have a quad-core computer, where processor cores may communicate via the shared memory. What are the potential advantages and disadvantages if an on-chip shared cache is used?

見背面

(e) [3 pts] Suppose we have an 8-core shared-memory computer built with two quad-core processor chips, and every two cores share one cache. For one processor core to load the word from location X after another core has written a word to location X, what would be the latency for the load instruction? Please list all possibilities.

3. (10 pts) Network bandwidth and storage capacity are growing even faster than the Moore's Law. Today, network-attached storage (NAS) is used popularly by people to share files between multiple computers via the network.
   (a) [3 pts] For a computer to acquire a file from a NAS server in a local area network (LAN) environment, which of the following might affect the time needed to acquire the file?
   A. the network bandwidth
   B. the physical distance between the computer and the NAS server
   C. the speed of the processor on the NAS server
   D. the speed of the disk on the NAS server
   (b) [2 pts] Calculate the minimum time to backup 100GB of data to the NAS server over the 802.11g wireless network. The peak network bandwidth is 54Mbit per second.
   (c) [2 pts] Would it be better if you copy the data onto a USB hard disk first, move the USB disk to connect to the NAS server, and then copy data to the NAS server? The peak bandwidth of the USB 2.0 interface is 480Mbps.
   (d) [3 pts] Compared to hard disks, what are the advantages and disadvantages of flash memory?

4. (10 pts) Consider two different processors, M1 and M2, of the same RISC instruction set. There are five classes of instructions (A, B, C, D and E). When the cache hit rate is 100%, the CPI for each instruction class for M1 and M2 are listed in the following table:

Class	M1	M2
A	8	4
B	6	6
C	10	2
D	8	2
E	8	8

Suppose M1 has a clock rate of 200 MHz. An engineer analyzed the workload and found out the instruction mix - A: 20%, B: 10%, C: 10%, D: 40%, E: 20%.
   (a) [3 pts] Based on this information, what would be the minimum clock rate required by M2 to execute the workload as fast as M1, assuming 0% cache miss rate?

Another engineer further analyzed the workload from the above question and found out the following: 30% of the instructions access memory for data, D-cache miss rate was 10%, I-cache miss rate was 5%, and memory access time is 50 nanoseconds.
   (b) [3 pts] Considering this new information, what is the average CPI for M1?
   (c) [4 pts] What would be the minimum clock rate required by M2 to execute the workload as fast as M1?

接次頁

PART II: Operating Systems

一. 是非題：請依題號在選擇題作答區作答，選答案 A 表示 True， 選答案 B 表示 False。每題正確答案得 2 分，錯誤答案倒扣 1 分，不答不計分。

1. Monitors can deadlock.
2. Increasing the page size will likely decrease working set size (in bytes).
3. Increasing the page size will likely decrease the chance that a replaced (kick-out) page requires a disk write.
4. Ignoring the overhead of context switching, I/O utilization on a round-robin system will decrease as the time-slice length is increased.
5. Again assume a zero-overhead context switch, some programs take longer to finish as the time-slice length decreases on a round-robin system.
6. The average response time of jobs decreases when round robin is used on jobs of equal length.
7. It makes sense to spin on a lock on a uniprocessor.
8. A faster hard disk reduces thrashing.
9. A TLB can contain two entries that have the same physical address.
10. Two independent processes P1 and P2 that run correctly with paging may deadlock when paging is disabled.
11. Strict2PL (phase locking) produces serializable but not recoverable schedules.
12. DMA uses cycle stealing to improve total system performance as well as the CPU utilization.
13. In Unix Semantics, writes to an open file by a user are not visible immediately to other users that have the same file open simultaneously.
14. In two phase commit protocol, the atomicity of transaction is ensured.
15. The happen-before relation ensures the global ordering across processes.

二. 簡答題：每題 3 分

1. Your friend, who did not study the operating system course well, decides to sell the "No Race Condition" scheduler, as a way to eliminate their concurrency bugs. His idea is that since race conditions are caused by concurrent access to shared state that he will simply disable time-sharing and run programs, first come first served, until they terminate. Write a **short, self-contained C program** that is harmless under timesharing but will cause your friend's system to freeze.

```
main() {
    // your very short code below



}
```

見背面

2. How does a deadlock recovery algorithm (i.e., aborting some deadlock process to break a deadlock cycle) ensure that a process starvation will not occur? Describe the avoiding-*starvation mechanism in 1-2 sentences.*

3. If we use majority protocol to lock a data item replicated in n different sites, where each site has a lock manager to control the locks for data stored at that site. Each lock manager determines whether the lock can be granted immediately in a distributed manner. How many messages needed to handle lock and unlock respectively?

4. Describe why there might be false cycles when detecting deadlock in distributed systems.

5. Describe how we generate a unique timestamp in distributed systems.

三. 申論題：5分

If you were to design a book recommendation system in a peer-to-peer way, such that the book title, abstract, and rating are stored locally in a database in your own disk space. You can issue a robot request to collect the ratings of your friends in a social network for certain books. The system will also solicit some high-rating books to you once in a while. You can borrow the book from your friends or buy from some on-line bookstores and contribute your ratings to the system. The problem is all the peers in this system might not be always on-line. Therefore, the recommendation is very dynamic. In this kind of distributed system, would you design the different transactions in a stateful or stateless way? Why? List different transactions and brief your reasons.

試題隨卷繳回