ECP2046 Computer Organization and Architecture
SOLUTION FOR TUTORIAL 1

1. Briefly explain the drawbacks of mechanical computer as compared to electronic computer.

   (a) Computing speed is limited by the inertia of its moving parts
   (b) Transmission of digital information by mechanical means is quite unreliable.

2. To what extent does each of the following items play the role of processor and/or memory when used in numerical computations:

   (a) Abacus
   (b) Slide rule
   (c) Electronic Pocket Calculator

   An abacus is not a memory. It only has data processing capability. A slide rule provides result like log A when given A, so it has some limited processing capability of the table lookup variety. The electronic pocket calculator on the other hand, has both an extensive memory and a powerful processor capable of automatic execution of complex, keyboard entered instructions.

3. Discuss the advantages and disadvantages of storing programs and data in the same memory (the stored program concept). Under what circumstances is it desirable to store programs and data in separate memories?

   The advantages:

   Instructions can be modified during program execution, making it possible to write programs that modify themselves. Only one set of addressing circuits is needed for accessing both instructions and data. The implementation of memory management software is simplified by the fact that no a priori distinction need be made between instructions and data. More efficient utilization of the available memory space results since either program or data can be assigned to any available free space.

   The disadvantages:

   Self-modifying programs are difficult to design, debug, and use. In fact, they are of questionable value. It is more difficult to debug programs when they are stored with data. Care must be taken to prevent inadvertent modification of instructions by treating them as data. Conversely, it is necessary to ensure that no attempt is made to execute data as instructions. The criteria for determining instruction and data word lengths are different. If separate program and data memories are used *then the word size of each memory can be optimized separately. This is not possible if a single memory is used for instructions and data. Separate memories allow instruction and data to be fetched in parallel, which may lead to an increase in processing speed.*

Separate program and data memories can be used to avoid the disadvantages of a single memory noted above. Many microprocessors place instructions and data in separate caches. An I-cache for instructions and a D-cache for data. It is also common in embedded microprocessor-based system, where the same software is used continuously, to put programs in ROMs and data in RAMs. This prevents the inadvertent alteration of instructions by treating them as data. It is also protects the programs in the event of a power failure since ROMs are invariably nonvolatile whereas most RAMs are volatile. Separate memories allow instructions and data to be accessed simultaneously. This means that the program and data access methods can be optimized independently, leading to improved system performance. For example, each memory can be assigned a different word size and a different address interleaving scheme. In applications such as real time digital signal processing where data can change independently of the CPU, separate program and data memories are also desirable.

4. Computers with separate program and data memories implemented in RAM and ROMs, respectively, are sometimes called *Harvard Class* machines after the Harvard Mark 1 computer. Computers with a single (RAM) memory for program and data storage are then called *Princeton Class* after the IAS computer. Most currently installed computers belong to one of these classes. Which one? Explain why the class you selected is the most widely used.

   Most of the computers in the world today are fairly simple 4 to 16 bits machines which are embedded as special purpose controllers in toys, appliances, automobiles and so on. They execute a fixed set of programs and frequently are not user programmable. Consequently their programs are stored in nonwritable ROM memories, while their working data is stored in a writable RAM. This separation of programs and data into different types of memories is characterized of Harvard class computers. Hence, the majority of computers can call Harvard class. Most user programmable machines such as personal computers and workstations are of course Princeton class.

5. A vector of 10 non-negative numbers is stored in consecutive location beginning in location 100 in the memory of IAS computer. Using provided instruction set, write a program that computes the address of the largest number in this array. If several locations contain the largest number, specify the smallest address.

   Figure 7 shows a program that solves this problem. A(100:109) denotes the vector to be examined. Beginning at the lowest addresses, A(I) is compared to A(J). If A(J) is greater than A(I), I is replaced by J. The final value of I is the required answer.

6. Early computer literature describes the IAS and other first generation computers as "parallel", unlike some of their predecessors. In what sense was the IAS a parallel computer? What forms of parallelism do modern computers have that are lacking in the IAS?

   Earlier computers like ENIAC had bit serial architectures, that is a data processing instruction operated on 1 bit position of its operands in one CPU clock cycle, so an n-bit operation like addition required n clock cycles. The IAS on the other hand operated on all bits of a 40 bits word in parallel, hence its designation as a parallel machine. The IAS lacks instruction level parallelism such as multiple E-units, pipelining, and multiple instruction issue, except perhaps for the fact that it fetches two instructions at once, a primitive form of instruction queuing. It also has no support for the processor level parallelism found in multiprocessors.

7. The IAS had no call or return instructions designed for transferring control between programs.

   Describe how call and return can be programmed using the IAS's original instruction set.

   The task of programming call or return instructions in the IAS is complicated by the absence of instructions that manipulate the contents of the program counter PC directly. The first instruction of the called subroutine must be placed at a known location, say X(0:19), so CALL can be implemented by a branch instruction of the form go to M(X,0:19). Similarly, the instruction to be executed after RETURN should be stored at a known address, say locations Y(0:19), so the RETURN can be implemented by go to M(Y, 0:19).

8. Briefly explain the shortcomings of IAS computer.

   In general writing and debugging a program whose instruction change themselves is difficult and error prone. Further before every execution of the program, the original version must be reloaded into M. Later computers employ special instruction types and registers for index control, which eliminates the need for address modify instruction.

   The small amount of storage space in the CPU results in a great deal of unproductive data transfer traffic between the CPU and main memory M. It also adds to program length. Later computers have more CPU register and a special memory called a cache that acts as a buffer between the CPU registers and M.

   No facilities were provided for structuring programs. For example, the IAS has no procedure call and return instructions to link different programs..

The instruction set is biased toward numerical computation. Programs for non-numerical tasks such as text processing were difficult to write and executed slowly.

9. Define the terms software compatibility and hardware compatibility. What role have they played in the evolution of computers ?

Two computers are said to be software compatible if they can execute the same machine language programs (object code compatibility) or high level language (source code compability). The requirement of running under a common operating system is often implied. A new computer is often designed to be backward software compatible with an earlier computer in the same family. This means that the new computer can directly execute software written for the older machine, but the converse is not usually true. Sometimes a new computer maintains a special program (an emulator) to translate from the older machine's object code to that of the new machine.

Hardware compatible refers to the situation where two computers share major buses, especially the system and IO buses, as well as common interface standards. Even within the same computer family, hardware compatibility tends to be quite limited.

10. What are the usual definitions of the terms CISC and RISC? Identify two key architectural features that distinguish recent RISC and CISC machines.

CISC and RISC refer to complex instruction set computer and reduced instruction set computer, respectively. Some key features of RISCs not found in CISCs are:

- *a small, streamlined instruction set with few formats and addressing modes.*
- *Single cycle execution of most instructions.*
- *Load/Store architecture*
- *Pipelined and superscalar operations have also been implemented successfully in CISC.*

11. In the IBM 360 Models 65 and 75, addresses are staggered in two separate main memory units (e.g. all even-numbered words in one unit and all odd-numbered words in another). What might be the purpose of this technique?

The purpose is to increase performance. When an address is presented to a memory module, there is some time delay before the read or write operation can be performed. While this is happening, an address can be presented to the other module. For a series of requests for successive words, the maximum rate is doubled.

Location	Instruction/data	Comment
0	1	Constant.
1	100	Address I (result).
2	101	Address J.
3	108	COUNT.
4L	AC := M(100)	Transfer A(I) to AC.
4R	AC := AC − M(101)	Compute A(I) − A(J).
5L	If AC ≥ 0 then go to M(7,0:19)	
5R	AC := M(2)	Replace I by J.
6L	M(1) := AC	
6R	M(4,8:13) := AC(28:39)	Modify address in 4L.
7L	AC := M(3)	Decrement COUNT.
7R	AC := AC − M(2)	
8L	If AC ≥ 0 then go to M(9,0:19)	Test COUNT.
8R	go to M(8,20:39)	Halt.
9L	AC := M(2)	Increment J.
9R	AC := AC + M(0)	
10L	M(2) := AC	
10R	M(4,28:39) := AC(28:39)	Modify address in 4R.
11L	go to M(4,0:19)	

Figure 7 IAS program for problem 1.18.