(iii) A updates the value of X.
Please state any assumption you have made for analyzing the above situation.

(d) Explain the role of Branch Target Buffer and Branch History Table in case of Dynamic Branch Prediction.

4 + 2 + 4 + 2 = 12

9. (a) Suppose that in an MIMD system, there are 20 processors. Each has its own cache. Suppose two processors each caches a single shared variable X. How many messages are sent across the system for maintaining cache coherency of X if (i) Snooping protocol is used? (ii) If a Centralized Directory Based Protocol is used? Explain your answer.

(b) State for each case mentioned below, which of the following design options are chosen for a RISC based architecture and why?
(i) Fixed vs. variable length instruction format
(ii) Simple vs. complex addressing mode.

(c) Consider the following program: (assume Opcode <srec>,<dest> format):

```
Add R3, R2
Sub R3,R4
Add R2,R1
Mov R1,[R4] ; writes to memory location
            ;pointed to by R4
Jnz R1, ThisPlace
: : :       : : :
ThisPlace: <some code>
```

Assuming a delay slot value of 3, rewrite the code to exploit the Delayed Branching mechanism. Explain briefly how performance is improved because of application of the above technique.

4 + 3 + 5 =12

ADVANCED COMPUTER ARCHITECTURE
(CSEN 5105)

Time Allotted: 3 hrs **Full Marks: 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and Any 5 (five) from Group B to E, taking at least one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1=10**

(i) Among the following types of Computer Architecture, which one puts more burden on the compiler?
(a) Superscalar (b) VLIW
(c) Data Flow (d) All of the above.

(ii) Among the following types of Computer Architecture, which one puts more burden on the hardware designer?
(a) Superscalar (b) VLIW
(c) RISC (d) All of the above.

(iii) What will be the time complexity of multiplying two n x n matrices on a hypercube architecture?
(a) $O(n^2)$ (b) O(n) (c) O(log n) (d) O(1).

(iv) For two instructions I and J WAR hazard occur, if
(a) $R(I) \cap D(J) \neq \varnothing$ (b) $R(I) \cap R(J) \neq \varnothing$
(c) $D(I) \cap R(J) \neq \varnothing$ (d) none of these.

(v) Match the correct combinations of multiple data processing on the right using various architecture combinations on the left:

(a) Vector Processors	(i) At the same time
(b) Array Processors	(ii) At different times
	(iii) Using same space
	(iv) Using different space.

(vi) VLIW philosophy is similar to which of the following architecture?
(a) Superscalar (b) RISC
(c) Data Flow (d) Array processor.

(vii) A 4-ary 3-cube hypercube architecture has
(i) 3 dimension with 4 nodes along each dimension
(ii) 4 dimension with 3 nodes along each dimension
(iii) both (i) and (ii)
(iv) none of these.

(viii) An initial collision vector is "1011010" for a pipeline. The corresponding reservation table has three rows with 3,2 and 3 cross marks respectively. The bounds of MAL for this pipeline is:
(a) 2,2 (b) 2,3 (c) 3,3 (d) 3,4.

(ix) In a five stage pipeline, a stage remains unused for a particular type of instruction. What is this problem known as?
(a) External fragmentation (b) Internal fragmentation
(c) Hazard (d) none of the above.

(x) The task of vectorizing a compiler is
(a) to find the length of vectors
(b) to convert sequential scalar instructions into vector instructions
(c) to process multi dimensional vectors
(d) to execute vector instructions.

Group - B

2. (a)

X	X					X	X
		X		X			
			X		X		

Estimate the Minimum Average Latency from the above Reservation Table.

(b) Prove that MAL is lower bounded by the maximum number of "X" marks occurring in any row of a reservation table.

(c) You have a requirement to implement a latency cycle (2,4). Create a set of compatibility classes to realize this.

(d) Can you create a modified Reservation Table from the one below to realize a Minimum Achievable Latency of 3?

		X		X		
	X		X		X	
X		X				X

3 + 2 + 3 + 4 = 12

Add R1, R2
Loop: Add R1,R2
Sub R1, #1
Bneq R1, #0, *Loop* (Go back to *Loop* if R1 is not equal to 0)
Store R2, B

Assume R1 is initially loaded with value 100 from B. Calculate the total time required to execute P.

3 + (3 + 1) + 5 = 12

7. (a) Draw a ILLIAC IV network (for 16 PEs). How many recalculating stages are required to route a message from PE5 to PE15.

(b) Show how the following two matrices can be multiplied using a mesh connected n/w? Make it 4 by 4.

$$\begin{pmatrix} 1 & 2 & 6 & 9 \\ 4 & 5 & 3 & 2 \\ 5 & 7 & 8 & 1 \\ 3 & 9 & 0 & 5 \end{pmatrix} \qquad \begin{pmatrix} 4 & 7 & 2 & 9 \\ 5 & 3 & 0 & 6 \\ 2 & 7 & 8 & 0 \\ 1 & 5 & 3 & 2 \end{pmatrix}$$

6 + 6 =12

Group - E

8. (a) Create a Data Flow Graph for the following sequence of instruction: (the instructions are in *OPcode Dest, Src1, Src2* format).

MUL R3 , R1, R2
ADD R5 , R3, R4
ADD R7 , R2, R6
ADD R10 , R8, R9
MUL R11 , R7, R10
ADD R5 , R5, R11

(b) What is the role played by the concept "token" in a Data Flow machine?

(c) Explain what problem will occur in the following situation of a multiprocessor machine consisting of two CPUs A and B each with its own cache:
(i) A reads from memory location X, then
(ii) B reads from X, and then

3. Consider a program P running on a machine M1.

Loop: Load R2,A
Add R3, #4
Beq R3,R4,*Loop* (go to *Loop* if R3 and R4 are equal)
Store R3,B

Assume the above Loop executes 10 times. The program runs on a machine with 4 stage pipeline viz. (IF, ID, EX, WB).

(a) Identify and classify any hazards occurring while executing P on M1.

(b) Calculate how much time is required to execute the program P if Operand Forwarding is not present.

(c) Rewrite P to take advantage of Delayed Branching (call it P').

(d) Assume M1 is fitted with Operand Forwarding now. Recalculate the execution time of the modified program P' as in (c).

3 + 3 + 3 + 3 = 12

Group - C

4. (a) Show in detail how the following two matrices are multiplied in $O(n^2)$ time on an SIMD machine where PEs are stored in a 2D Mesh:

$$\begin{matrix} 1 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 7 & 6 & 5 & 4 \\ 3 & 2 & 1 & 1 \end{matrix} \quad \textbf{and} \quad \begin{matrix} 2 & 3 & 2 & 3 \\ 4 & 5 & 4 & 5 \\ 6 & 7 & 6 & 7 \\ 3 & 2 & 3 & 2 \end{matrix}$$

(b) Draw the circuit diagram of a 8×8 Delta network using 2×2 switching elements.

(c) Consider the following statements i) through iii). Indicate which of the following are true and which are false. *Explain in one sentence in each case why you think so.*

(i) Compulsory miss can be reduced by decreasing the cache block size
(ii) Capacity miss can be reduced by decreasing the total size of the cache
(iii) Conflict miss can be reduced by decreasing the value of cache associativity

5 + 4 + 3 = 12

5. (a) Consider the sequence of SIMD instructions given below:
(i) Load A(0..3) to Ri's
(ii) Add 1 to each Ri
(iii) Multiply 2 to each Ri
(iv) Store Ri's to A(0..3).
Show how the above sequence can be executed by a Vector Processor as well as in an Array Processor with suitable diagrams.

(b) There is a Vector processor M1 available. It has two load units to load from memory to vector registers. The time to load a vector element is 50 cycles. Once a vector element is loaded, it takes 1 clock cycle to load the next vector element as the memory is organized in an interleaved manner. All vector units are chained. That means whenever a vector element has finished in one stage, it can move to the next stage without any further stalls. VADD takes 4 cycles, VMUL 16 cycles and VSHFT 1 cycle for each vector element. All the vector processing units are fully pipelined.
Vi 's below are all Vector registers.

If A and B below are two vectors each of length 32 elements, compute how many cycles are required to execute the following program P on M1?

VLD V1 <- A ; Vector Load
VLD V2 <- B
VADD V3 <- V1, V2
VMUL V4 <- V3, V1
VSHFT V5 <- V4

(3 + 3) + 6 = 12

Group - D

6. (a) Machine M1 with a single CPU core runs a program P in which 80% code is parallelizable. Another machine M2 has 40 parallel cores. What would be the speedup if P is executed on M2 following Amdahl's law?

(b) What would be the speedup if Gustafson's law is used instead in (a)? What is the basic conceptual difference in the above two cases?

(c) A machine M1 has a clock cycle time 10 ns. Each instruction using memory requires 100 clock cycles. Register only instructions require 1 clock cycle. A program P is given below:

Load R1, A
Load R2, B