

Instructions: Answer **a total of 41** questions, 40 from Group A through Group F, and 1 from GROUP-G, choosing **a minimum number from each group as specified below:**

Group	Minimum number of questions to be answered from group	Total number of questions to be answered
A	5	40
B	3	
C	14	
D	3	
E	4	
F	4	
G	1 (no choice)	01

GROUPs A-F $40 \times 2 = 80$

Choose the unique correct answer

GROUP-A

For Q1-Q3: The MIPS Floating-Point Unit has latencies given in the table below. Latency is defined to be the number of **intervening** clock cycles needed to avoid a stall.

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

[Turn over

Now consider the following MIPS code:

```

-----
Loop:  L.D      F0, 0(R1)
        ADD.D   F4, F0, F2
        S.D     F4, 0(R1)
        DADDUI  R1, R1, #-8
        BNE     R1, R2, Loop
-----

```

In order to reduce the number of stalls, this code is scheduled. L.D is issued in cycle-1.
Answer the following with reference to the scheduled code:

1. DADDUI is issued in

- (a) cycle-2
- (b) cycle-3
- (c) cycle-4
- (d) cycle-5

2. ADD.D is issued in

- (a) cycle-2
- (b) cycle-3
- (c) cycle-4
- (d) cycle-5

3. S.D is issued in

- (a) cycle-5
- (b) cycle-6
- (c) cycle-7
- (d) cycle-8

For Q4-C06: Consider the MIPS Floating-Point Unit employing Dynamic Scheduling.. The following program is executed:

```

-----
Loop:  L.D      F0, 0(R1)
        MUL.D   F4, F0, F2
        S.D     F4, 0(R1)
        DADDIU  R1, R1, -8
        BNE     R1, R2, Loop
-----

```

This program multiplies all elements of an array by a scalar contained in F2.
The loop iterations are independent and can, in principle, be executed simultaneously.

The loop is UNROLLED DYNAMICALLY by the hardware

Fig Q4 shows the role of the registers.

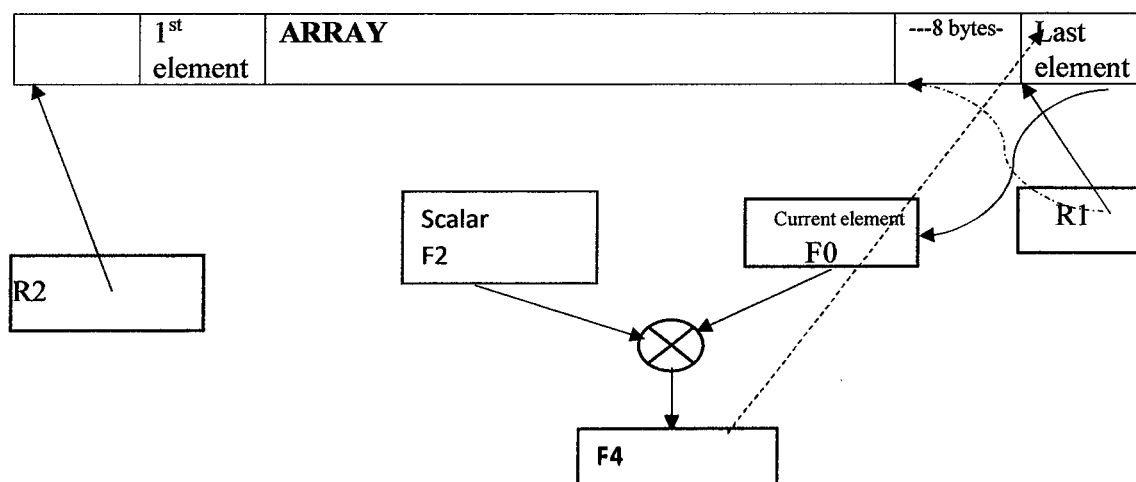


Figure Q4: Role of the registers

Let us consider the situation when the following two conditions are both true:

- (i) all instructions of the first TWO iterations of the loop have been issued; and
- (ii) Load1 and Load2 are executing but have not written result into registers.

[Load1 denotes “L.D F0, 0(R1)” for the first iteration.

Load2 denotes “L.D F0, 0(R1)” for the second iteration., etc.]

For a “Store” instruction, j refers to the memory address and k refers to the value to be written (stored).

4. For Mult1, Qj is

- (a) Regs[F0]
- (b) Load1
- (c) Store1
- (d) Regs[F4]

5. For Store1, Qk is

- (a) Regs[F4]
- (b) Regs[R1]
- (c) Mult1
- (d) Load1

6. For Store2, Vj is

- (a) Regs[F4]
- (b) Regs[R1]
- (c) Mult2
- (d) Regs[R1] – 8

[Turn over

7. WAR and WAW hazards can be resolved by

- (a) Pipeline scheduling
- (b) Loop Unrolling
- (c) Register Renaming
- (d) Branch Prediction

8. RAW hazards can be resolved by

- (a) forwarding or bypassing
- (b) stalling
- (c) both (a) and (b)
- (d) increasing the clock frequency

GROUP-B

Answer the following with respect to the CRAY X-MP:

9. Consider the issue of a 2-parcel instruction. In the cycle in which the first parcel moves from the NIP to the CIP, the second parcel is transferred

- (a) from the LIP to the NIP
- (b) from the instruction buffer to the LIP
- (c) from the instruction buffer to the NIP
- (d) none of the above

10. A vector operation is issued to an n-stage vector pipeline. The destination vector register of the vector operation becomes available for another operation after

- (a) $n + VL - 1$ cycles
- (b) $n + VL + 5$ cycles
- (c) $n + VL + 2$ cycles
- (d) none of the above

11. Each Vector Register of the CRAY X-MP has a length of

- (a) 16 words
- (b) 32 words
- (c) 64 words
- (d) 128 words

12. A one-parcel instruction is decoded in the

- (a) CIP
- (b) LIP
- (c) NIP
- (d) P (Program Address Register)

GROUP-C

13. At a network node, resources are allocated to packets instead of messages because

- (a) message structure depends on the protocol
- (b) messages are encoded
- (c) packets have a restricted maximum length
- (d) packets map to memory words

14. The packet header contains

- (a) crossbar switch settings
- (b) control state
- (c) buffer capacity
- (d) routing information

15. The basic unit of bandwidth and storage allocation is a

- (a) message
- (b) flit
- (c) packet
- (d) phit

16. The typical bit-length of a phit is

- (a) 8
- (b) 16
- (c) 64
- (d) 128

For Q17-Q21: Consider a Dropping Flow-Control mechanism with explicit negative acknowledgement. A 5-flit packet consists of a head flit H, three successive body flits B1, B2, and B3, and a tail flit T. The packet is routed through a 4-hop route consisting of channels c0, c1, c2, and c3. Each channel has two links, a forward link F, and a reverse link R. The head flit enters channel c0 (F-link) in cycle-0. In cycle-3, the packet is unable to proceed to channel c3 due to unavailability of the F-link of c3. As a result, the packet is DROPPED and a negative acknowledgement ("nack" for short) flit "N" is sent on the R-link of c2 in cycle-3. The N-flit travels backwards (towards the source) along the reverse links (R) from c2 to c0 in successive cycles and reaches c0 in cycle-5. In cycle-6, a retransmission of the packet is initiated.

17. In cycle-3, c2 (F-link) is occupied by

- (a) T
- (b) B3
- (c) B2
- (d) B1

[Turn over

18. Flit N occupies channel c0 (R-link) in

- (a) cycle-5
- (b) cycle-6
- (c) cycle-7
- (d) none of the above

19. Flit B2 of the retransmitted packet occupies c2 in

- (a) cycle-8
- (b) cycle-9
- (c) cycle-10
- (d) cycle-11

For Q20-Q23: Answer the following with reference to the MESI Protocol.

20. Cache-1 in the E (Exclusive) state observes a read request for a block in this cache from another cache Cache-2.

- (a) Cache-1 must send the block to cache-2 since cache-1 might have had modified the block
- (b) Cache-1 need not do anything since main memory has a valid copy of the block
- (c) Cache-1 moves to the I (invalid) state
- (d) none of the above

21. On receiving a PrWr request , a cache in the S (shared) state

- (a) generates a BusRdX transaction
- (b) flushes its data on the bus
- (c) moves to the E (exclusive) state
- (d) none of the above

22. When a cache in the M (modified) state observes a BusRd transaction, it

- (a) remains in the M state
- (b) moves to the shared state (S)
- (c) flushes its data on the bus
- (d) both (b) and (c)

23. When a cache in the I (invalid) state receives a PrRd request, it

- (a) move to the E-state if the shared signal (S) is asserted
- (b) moves to the S-state if the shared signal (S) is not asserted
- (c) moves to the M-state
- (d) none of the above

For Q24-Q26: Answer the following with reference to Directory-based Cache-coherence.

24. The directory entry for a block contains

- (a) the list of files that reference the block
- (b) the list of caches that currently contain a copy of the block
- (c) the list of threads that are using the block
- (d) none of the above

25. The node that currently holds the valid copy of a block and must supply the data when needed is called

- (a) Dirty node
- (b) Exclusive node
- (c) Owner node
- (d) Home node

26. A cache read-miss occurs at node-i. The home node for the block being a remote node, a network transaction is sent to the home node. At the home node, the dirty bit for the block is ON.

- (a) the home node supplies the data to node-i
- (b) the home node sends to node-i the identity of the node whose presence-bit is ON
- (c) node-i sends a request to the Owner node
- (d) both (b) and (c)

27. Simultaneous Multithreading (SMT)

- (a) allows multiple independent threads to execute simultaneously on the same core
- (b) allows two threads to use the same functional unit on a single core
- (c) cannot be employed in a multicore
- (d) none of the above

28. A Nehalem processor chip contains the following functional part:

- (a) Northbridge (Graphics and Memory controller hub)
- (b) Southbridge (I/O Controller hub)
- (c) both (a) and (b)
- (d) Un-core Interface Unit (UIU)

29. The Nehalem instruction pipeline contains

- (a) an in-order Execution Engine
- (b) an out-of-order Front-End Pipeline
- (c) an out-of-order Execution Engine
- (d) all of the above

30. In a multiprocessor system based on the Nehalem microarchitecture, each processor can access the memory region local to another processor through

- (a) Southbridge
- (b) Northbridge
- (c) QPI
- (d) PCI Express bus

GROUP-D

TABLE IV: Speedup of multi-threaded SPEC CPU2017 speed benchmarks

Speed 2017 FP	Number of Threads			
	2	4	6	12
603.bwaves_s	1.794	2.679	3.467	4.166
607.cactuBSSN_s	1.852	3.718	5.166	5.209
619.lbm_s	1.650	1.728	1.782	1.870
621.wrf_s	1.875	3.405	4.637	5.645
627.cam4_s	1.855	3.340	4.637	5.645
628.pop2_s	1.935	3.521	4.961	6.056
638.imagick_s	1.962	3.974	5.931	11.347
644.nab_s	1.252	2.452	3.646	5.036
649.fotonik3d_s	1.637	1.812	1.860	1.863
654.roms_s	1.863	2.980	3.708	4.051

31. Consider the above table of Benchmark results for different number of threads, obtained on an Intel Skylake machine (3.4 GHz, i7-6700 processor, 8MB last-level cache) running Ubuntu 14.04.

The benchmark 638.imagick_s has noticeable speed improvement with increasing number of threads because

- (a) it is an image processing application with a large amount of parallelism
- (b) it uses the cache efficiently
- (c) it uses special hardware on the bus
- (d) none of the above

32. A processor has the following instruction frequencies and corresponding clock-cycles

- (i) Integer ALU: 50% 1-cycle
- (ii) Load: 20 %, 5-cycles
- (iii) Store: 10%, 1-cycle
- (iv) Branch: 2 %, 2-cycles

The CPI is

- (a) 2
- (b) 9
- (c) 2.5
- (d) 5

For Q33-Q34:

A PARSEC (The Princeton Application Repository for Shared-Memory Computers) Benchmark is executed to compare the performance of two compilers: (i) GCC (GNU Compiler Collection; and (ii) ICC (Intel C++ Compiler). The gcc compiler has the versions gcc-O0 (No optimization), gcc-O2 (level-2 optimization), and gcc-O3 (level-3 optimization). Similarly, the icc compiler has the versions icc-O0, icc-O2, and icc-O3. The *Blackscholes* program for Financial Analysis is compiled using both the compilers.

The following table gives the execution times of Blackscholes using the binaries generated by the compilers.

Table: 2 Illustration - Execution Time of different threads for Blackscholes (native)

Number of Threads	gcc-O0	gcc-O2	gcc-O3	icc-O0	icc-O2	icc-O3
1	197.53s	133.8230s	137.765s	175.435s	91.646s	91.743s
2	108.977s	76.481s	78.355s	97.827s	54.865s	54.707s
4	65.09s	47.887s	48.926s	59.084s	36.260s	36.224s
8	51.158s	37.591s	41.431s	47.040s	31.277s	31.214s

33. Between gcc-O3 and icc-O3

- (a) gcc-O3 has a better performance
- (b) icc-O3 has a better performance
- (c) both have almost equal performance
- (d) none of the above

34. The speedup obtained by using icc-O3 for 2 threads relative to that from gcc-O0 using 1-thread is

- (a) 3.61
- (b) 2.02
- (c) 2.58
- (d) 1.81

GROUP-E

35. Between 2009 and 2019, most carbon emissions attributable to mobile devices shifted from

- (a) opex-related to capex-related
- (b) capex-related to opex-related

[capex : Capital Expenditure; opex: Operational Expenditure]

[Turn over

36. The Greenhouse Gas (GHG) Protocol is
- (a) an accounting standard for reporting carbon emissions
 - (b) a computer network protocol for fast data transmission
 - (c) a set of rules for chip manufacturing
 - (d) none of the above

37. Fabricating semiconductor circuits
- (a) contributes 31 % of global greenhouse gas emission
 - (b) generates considerable amount of waste
 - (c) requires Ultrapure water (UPW)
 - (d) all of the above

38. Superconducting qubit is used in
- (a) Vector Supercomputers
 - (b) Quantum computers
 - (c) Cloud Computers
 - (d) Microcore microprocessors

39. The Intel **Loihi** chip employs
- (a) Support Vector Machine
 - (b) Decision Tree Model
 - (c) Spiking Neural Network Model (SNN)
 - (d) Artificial Neural Network (ANN) model

GROUP-F

40. A system on chip (SoC) consumes comparatively less power due to
- (a) reduction in interconnect (bus) power
 - (b) on-chip memory
 - (c) both (a) and (b)
 - (d) none of the above

41. PAPI is
- (a) a mobile Operating System
 - (b) Benchmark
 - (c) a standard library that provides application-level interface to performance counters on multiple hardware and operating system platforms
 - (d) none of the above

For Q42-Q44: Answer the following with respect to GPU microarchitecture.

42. The instruction fetch loop includes
- (a) Scoreboard
 - (b) Operand Collector
 - (c) Decode
 - (d) ALU

43. The SIMD stack handles

- (a) nested control flow
- (b) skipped computation
- (c) both (a) and (b)
- (d) none of the above

44. The dynamic power dissipation in a CMOS device is proportional to

- (a) the leakage power
- (b) f^2 , where f is the clock frequency
- (c) V^2 , where V is the supply voltage
- (d) C^2 , where C is the node capacitance

45. Domain-specific accelerators employ the following technique for performance and efficiency gain:

- (a) Data specialization
- (b) Hardware-implemented interpreters
- (c) Associative memory
- (d) Quantum Entanglement

GROUP-G

46. Consider the following Cray X-MP code fragment:

```

-----
      A3      1
      S1      1.          ; Set S1 to Floating-point value 1.00
      A0      ADAT        ; Set A0 to starting address of vector in memory
      A1      10          ; Set A1 to 10
      VL      A1          ;Set the vector length to 10
i1     S7      RT          ;Read real-time-clock into S7 for noting time—1 cycle
i2     V1      ,A0,1       ;Read the vector into register V1, with stride =1
                          ; stride is the spacing of elements
i3     V2      S1 +FV1     ;Add 1.00 to each element of V1
i4     ,A0,1   V2          ; Write the result to memory
i5     S5      V2, A3      ; Set S5 to element A3 of V2
i6     S4      RT
-----

```

The execution times of the relevant instruction are given below:

<u>Id</u>	<u>Instruction</u>	<u>Execution-time(cycles)</u>
i1	S7 RT	1
i2	V1 ,A0,1	VL+17 (This includes 3 cycles for Setup. The first result is available after 17 cycles)
i3	V2 S1+FV1	VL + 11(This includes 3 cycles for Setup. The first result is available after 11 cycles)
i4	,A0,1 V2	VL+3 (including 3 cycles for Setup)
i5	S5 V2, A3	4

Here, Vector CHAINING is applicable. Thus, the output of i2 is chained to i3. The output of i3 is chained to i4.

NOTE:

(N1) If i2 starts execution (Setup) in cycle-x, it will produce its first result, i.e. V1[0], at the end of cycle $(x + 17 - 1) = x + 16$. In this calculation, the timing diagram has all the cycles in the range $x..(x+16)$ marked S (setup) or E (Execute). Any intervening H(hold) cycle is NOT counted. Thus, if there are h cycles marked H (hold) occurring after cycle-x, then the first output appears at the end of cycle $(x+16+h)$.

(N2) i2, i3, and i4, being Vector instruction, require SETUP (marked S in the timing diagram).

(N3) A chained instruction starts executing as soon as the first element of its Vector input is available. Thus, if i2 produces its first output (i.e., the first element of its output vector) at the end of cycle-y, the chained instruction i3 begins execution in cycle-(y+1).

(a) Draw a timing diagram for instructions i1 through i6, assuming i1..i4 to be issued in cycles 1..4. Instructions i2, i3, and i4 will go to the "hold" (H) state after setup if input data is not available. When the data is available, they start executing (E-state in the timing diagram).

15

(b) Give a brief explanation of (a) above.

5

-----END-----