一. **CHOICE**: (60 points) (note: only **one** is correct)

1. Which of the following I/O mechanisms requires the least hardware support? ( ) A

   A. Polling B. Interrupt
   C. DMA D. None of the above

2. What is the primary characteristic of a CISC architecture? Circle the best answer. ( ) C

   A. Large number of registers
   B. Small number of registers
   C. Variable length instructions
   D. Fixed length instructions

3. Computer A has an overall CPI of 1.5 and can be run at a clock rate of 750MHz. Computer B has a CPI of 2.5 and can be run at a clock rate of 1GHz. A program has exactly 100,000 instructions when compiled for computer A. How many instructions would the program need to have when compiled for Computer B, in order for the two computers to have exactly the same execution time for this program? ( ) B

   A. 45000 B. 80000 C. 100000 D. 120000

4. A key factor in determining the cost of an integrated circuit is volume. Which of the following is **not** reason why a chip made in high volume should cost less? ( ) B

   A. With high volumes, the manufacturing process can be tuned to a particular design, increasing the yield.
   B. It is less work to design a high-volume part than a low-volume part.
   C. The masks used to make the chip are expensive, so the cost per chip is lower for higher volumes.
   D. Engineering development costs are high and largely independent of volumes, thus, the development cost per die is lower with high-volume parts.

5. A group of students were debating the efficiency of the five-stage pipeline when one student pointed out that not all instructions are active in every stage of the pipeline. After deciding to ignore the effect of hazards, they made the following four statements. Which one is **correct?** ( ) D

   A. Allowing jumps, branches, and ALU instructions to take fewer stages than the five required by the load instruction will increase pipeline performance under all circumstances.
   B. Trying to allow some instructions to take fewer cycles does not help, since the throughput is determined by the clock cycle; the number of pipe stages per instruction affects throughput, not latency.
   C. You cannot make ALU instructions take fewer cycles because of the write-back of the result, but branches and jumps can take fewer cycles, so there is some opportunity for improvement.
   D. Instead of trying to make instructions take fewer cycles, we should explore making the pipeline longer, so that instructions take more cycles, but the cycles are shorter. This could improve performance.

6. How the cache conflict misses will be affected by the following modifications? Assume the baseline cache is set associative. ( ) B

(a). Double the associativity while keep the capacity and line size constant

(b). Double the number of sets while keep the capacity and line size constant

A. Increase; Decrease B. Decrease; Increase

C. Increase; Increase D. Decrease; Decrease

7. Here is a series of address references given as word addresses: 2, 3, 11, 16, 21, 13, 64, 48, 19, 11, 3, 22, 4, 27, 6, 11. Determine how many misses will happen in each condition.

(a) A direct-mapped cache with 16 one-word blocks that is initially empty. ( ) D

A. 11 B. 12 C. 13 D. 15

(b) A direct-mapped cache with four-word blocks and a total size of 16 words. ( ) B

A. 11 B. 12 C. 13 D. 15

(c) A two-way set-associative cache with four-word blocks and a total size of 16 words. Use LRU replacement. ( ) C

A. 11 B. 12 C. 13 D. 15

10. If a data cache does not contain a dirty bit, then it must be using a _____ policy. ( ) A

A. Write through B. Write back

C. Read back D. None of the above

11. Which of the following situation will not happen?( ) B

A. TLB miss, Cache hit, Page hit

B. TLB hit, Cache hit, Page miss

C. TLB miss, Cache hit, Page hit

D. TLB hit, Cache hit, Page hit

12. Answer the following questions assuming the following memory/cache organization:

- 4 GiB Byte-Addressed Memory
- 256 KiB Direct-Mapped Cache
- 8 Byte Words
- 4 Word Blocks

(a) Give the Tag : Index : Offset breakdown for the above cache: ( ) C

A. 14:13:3 B. 16:13:3

C. 14:13:5 D. 16:13:5

(b) Suppose we switch to a word-addressed memory.

Give the Tag : Index : Offset breakdown: ( ) D

A. 14:13:3 B. 16:13:3

C. 14:13:5 D. 16:13:5

(c) Calculate AMAT for a machine with the following specs: L1 hits take 3 cycles, L1 local miss rate is 25%. L2 hits take 10 cycles, L2 local hit rate is 60%. L3 hits take 100 cycles, L3 global miss rate is 9%. Main memory accesses take 1000 cycles and all data is available in memory. ( ) C

A.105.5 cycles B. 1000 cycles

C. 24.5 cycles D, 25.5 cycles

ACBBD BDBCA BCDC

二. **TRUE (T) OR FALSE (F)** (10 points)

1. In the MIPS processor we studied, all instructions were 32-bits wide.（ ）T

2. MIPS can best be described as a CISC architecture because the instruction set has more than 20 opcodes. （ ） F

3. Adding a lower level cache reduces miss penalty. （ ）T

4. Increasing set associativity increases hit time. （ ）T

5. An instruction takes less time to execute on a pipelined processor than on a nonpipelined processor (all other aspects of the processors being the same). （ ）F

6. A denormalized binary floating point number is any non-zero floating point number that is not in the form $1.a \times 2^b$, where *a* and *b* are integers represented in binary. （ ）T

7. The MIPS **slt** instruction only works correctly if both operand registers contain non-negative values. （ ）F

8. The CPI of superscalar processors can be less than one. （ ）T
s

9. With a 4-entry TLB, 2 physical pages, and 4 virtual pages, the TLB will never be full. （ ）T

10. in a function that makes lots of function calls, it is more efficient to save local variables in temporary registers than in saved registers. （ ）F

三. **PIPELINE** (10 points)

In this exercise, we examine how data dependences affect execution in the basic 5-stage pipeline. Problems in this exercise refer to the following sequence of instructions:

```
lw $5, -16($5)
sw $5, -16($5)
add $5, $5, $5
```

The basic pipeline is following the Figure.1.

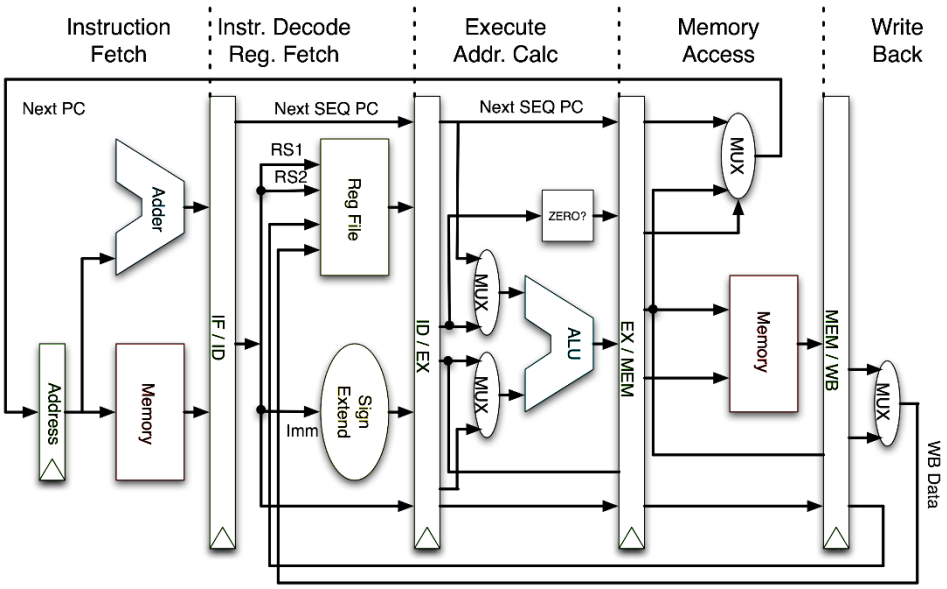


Figure.1 Basic Pipeline

a) Find all data dependences in this instruction sequence, write like ($2) I1 to I4.

WAR	RAW	WAW
($5) I1 to I3 ($5) I2 to I3	($5) I1 to I2 ($5) I1 to I3	($5) I1 to I3

b) Assume there is no forwarding in this pipelined processor. Add NOP instructions to eliminate them.

```
lw $5, -16($5)
nop
nop
sw $5, -16($5)
add $5, $5, $5
```

c) Assume there is full forwarding. Add NOP instructions to eliminate them.

```
lw $5, -16($5)
nop
sw $5, -16($5)
add $5, $5, $5
```

The remaining problems in this exercise assume the following clock cycle times:

Without forwording	With full forwarding
180ps	240ps

d) What is the total execution time of this instruction sequence without forwarding and with full forwarding? What is the speedup achieved by adding full forwarding to a pipeline that had no forwarding?

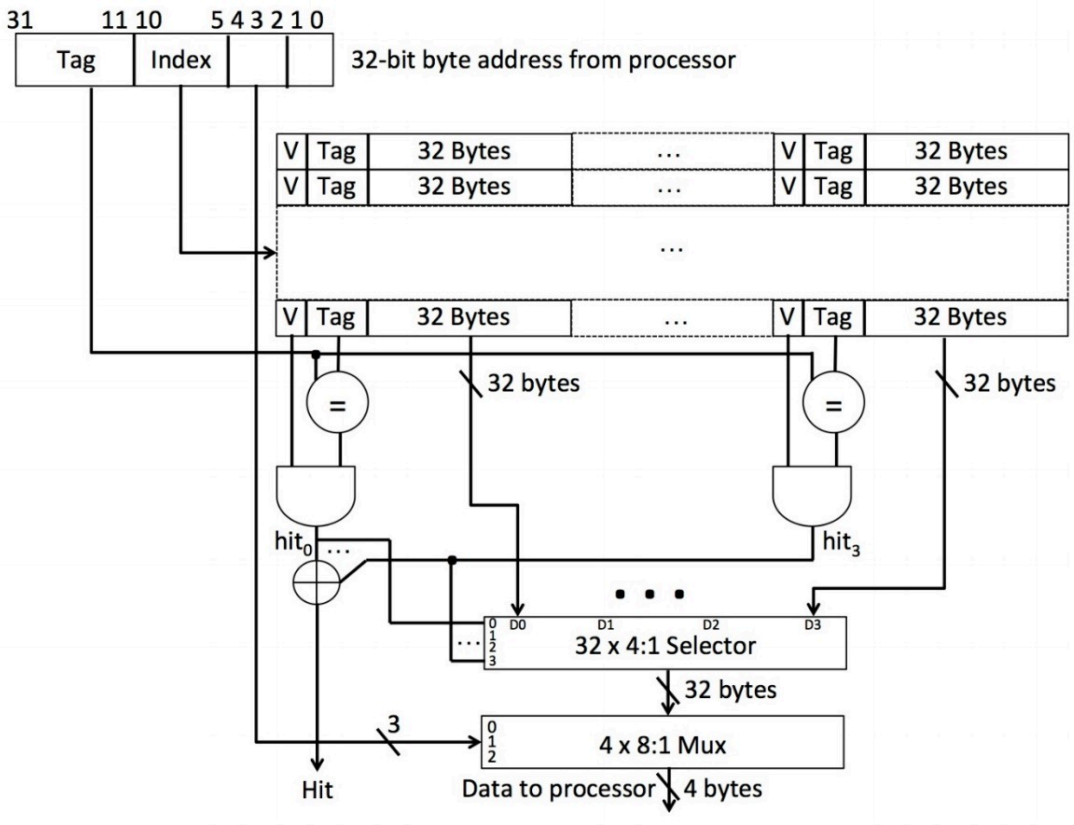
no forwarding: 180ps * 9 =1620ps

forwarding used: 240 * 8 = 1920ps

speedup = 1620/1920 = 0.84

四. **CACHE** (12 points)

You are given the sketch of a cache design below:



Note: a 4:1 Selector has four control inputs labeled 0, 1, 2, 3 and four data inputs D0, D1, D2, D3. It connects Di to the output when the hit control signal is true. Assume that **at most one** control input is true at any time. The selector function is undefined when none of the control inputs are true.

Answer the following questions about the cache above.

1. What is the block size of the cache in bytes?
   32Bytes
2. What is the number of blocks in this cache?
   256
3. What is the total data capacity of the cache in bytes?
   8192Bytes
4. What is the associativity of the cache?
   4-way
5. Is this a write-through or write-back cache?
   Write-through
6. What is the total number of valid bits in the cache?
   256
7. What is the total number of tag bits in the cache?
   256*21=5376

五. **VIRTUAL MEMORY** (8 points)

In this problem, we are running two different processes on our computer. Our system has the following properties:

- 1 MiB of Physical Address Space – 20-bit physical addresses (PA)
- 4 GiB Virtual Address Space – 32-bit virtual addresses (VA)
- 32 KiB page size – 15-bit page offset
- 4-entry fully-associative TLB, LRU replacement
- Page tables (PTs) use write-back policy with permission bits for read (rd), write (wr), and execute (ex).

a) Numbers – Fill in the blanks with the appropriate numbers:

_____#tag bits for TLB 32-15=17

_____ bit-width of PT address register 20

_____of valid entries in a PT $2^{20-15}=32$

_____bit-width of an entry in a PT (there are 5 extra bits) 5+5=10

fully-associative TLB means no index bits, PT sits in physical mem (PT address is PA), PT valid entries limited by physical mem capacity, PT only holds PPN (5 bits), one PT per process.

b) Below is an excerpt from one of the processes (Process 1), which uses a large square matrix of 32-bit integers. What is the largest gap between successive memory accesses (in the virtual address space) in **bytes** taken in the `for` loop?

```
 #define MAT_SIZE = 2048
for(int i=0; i<MAT_SIZE; i++)
    mat[i*(MAT_SIZE)] = i;
```

stride length is MAT_SIZE+1 integers, which is 4*(MAT_SIZE+1) bytes
(2048+1)*4=8196

c) Assume that the matrix `mat` is stored contiguously in memory and that `mat[0]` is at the beginning of a page. Assuming that Process 1 is the only process running, calculate the following hit rates (HRs) for the first execution of the `for` loop:

_____ PT Hit Rate 0%

_____ TLB Hit Rate 75%

We access memory in a strictly increasing manner: (i+1)*(MAT_SIZE+1) > i*(MAT_SIZE+1). Because of this, we never revisit a page once we have left it. TLB hits do NOT count as PT hits, so the PT hit rate is 0%.