102台大電機

單選題

1. Pipelining in CPU design is aimed to provide the optimal
(A) throughput (B) latency (C) space parallelism (D) caching of instruction execution.

Answer: (A)

2. Given a non-pipelined CPU operating at 50 MHz, we are optimizing the CPU with a 4-stage pipeline design. In an ideal case, what is the operating clock frequency for the 4-stage pipeline CPU? (A) 50 MHz (B) 100 MHz (C) 150 MHz (D) 200 MHz (E) 400 MHz

Answer: (D)

3. Direct Memory Access (DMA) is (A) worse (B) better (C) the same
in performance for small transfers than interrupt driven I/O.

Answer: (A)

4. Suppose a 32-bit CPU with physical address bus $A_{31}A_{30}$..... A_1A_0 and assume that the data cache has the following structure:

Cache structure is set associative with 2 lots per set
Cache size is 128 KBytes
Cache block size is 16 Bytes and a word is 32-bit (4 Bytes)
Cache is indexed with physical address

The address bits (A) $A_{16}A_{15}...A_4$ (B) $A_{15}A_{14}...A_4$ (C) $A_{14}A_{13}...A_4$ (D) $A_{13}A_{12}...A_4$ (E) $A_{15}A_{14}...A_3$ are used as index address for the cache.

Answer: (B)

Remark: Number of sets = 128KB / (16B × 2) = 4K

Tag	Index	Offset
16	12	4
$A_{31} \sim A_{16}$	$A_{15} \sim A_4$	$A_3 \sim A_0$

5. Suppose a 32-bit CPU with physical address bus $A_{31}A_{30}$..... A_1A_0 and assume that the data cache has the following structure:

   Cache structure is direct-mapped

   Cache size is 128 KBytes

   Cache block size is 16 Bytes and a word is 32-bit (4 Bytes)

   Cache is indexed with physical address

   The address bits (A) $A_{16}A_{15}...A_4$ (B) $A_{15}A_{14}...A_4$ (C) $A_{14}A_{13}...A_4$ (D) $A_{13}A_{12}...A_4$ (E)$A_{13}A_{12}...A_3$ are used as index address for the cache.

Answer: (A)

Remark: Number of sets = 128KB / (16B) = 8K

Tag	Index	Offset
15	13	4
$A_{31} \sim A_{17}$	$A_{16} \sim A_4$	$A_3 \sim A_0$

6. A program is compiled into 10 billion instructions and is to be executed by a 5-stage pipelined CPU with 1 GHz clocks. Each instruction results in an average of 2.2 stall cycles. What is the execution time for this application? (A) 42 (B) 32 (C) 22 (D) 4.4 (E) 6.4 seconds.

Answer: (B)

Remark: Execution time = $[(5 - 1) + 10G + 2.2 \times 10G] / 1G = 32$

7. What instruction in ARM processors does not affect the conditional code?
   (A) ADDS r0, r1, r2 (B) ADD r0, r1, r2 (C) CMP r1, r2 (D) TST r1, r2

Answer: (B)

8. An application spends 80% of its time doing multiply instructions. If the multiplier is sped up by 4 times, the application will run (A) 5 (B) 4 (C) 3 (D) 2.5 (E) 2 times faster.

Answer: (D)

9. An application spends 80% of its time doing multiply instructions. If the multiplier is sped up by infinite times, the application will run (A) 5 (B) 4 (C) 3 (D) 2.5 (E) 2 times faster.

Answer: (A)

複選題

10. (A) instruction pre-fetch (B) instruction buffer (C) branching (D) cache misses (B) resource constraints will keep the pipeline from being full.

Answer: (A), (B)

11. Features that are typically found in RISC architectures include:
    (A) large number of registers
    (B) uniform instruction format
    (C) load-store operations
    (D) hardwired control unit
    (E) arithmetic instructions directly operating on memory data

Answer: (A), (B), (C), (D)

12. If we run the following program on a 32-bit machine, what outputs might be generated?

```
#include <stdio.h>
int main ( )
{
    int A[3] = {1, 2, 3};
    int *ptr;
    ptr A;
    printf(" %p : %d \n", ptr, *ptr) ;
    ptr++;
    printf(" %p : %d \n", ptr, *ptr) ;
    return 0;
}
```

    (A) 0xbfe5a870:1
        0xbfe5a874:2
    (B) 0xbfela5e0:l
        0xbfela5e4:2
    (C) 0xbfe5a870:2
        0xbfe5a874:3
    (D) 0xbfela5e0 :2
        0xbfela5e4:3
    (E) 0xbfe5a870:1
        0xbfe5a874:3

Answer: (A), (B)

102 台聯大電機

1. Assume that the original design of a processor runs at 100 MHz clock rate and provides A, B, C instruction classes. An engineer plans to redesign the processor to be able to run at X MHz (X > 100) for improvement. It is found that the CPI and usage of the instruction class for a benchmark program will be affected by the value of X in the redesigned version. The CPIs and usages of each instruction class for the original and the redesigned processor and are provided in the following table:

Instruction Class	Original Processor		Redesigned Processor	
	CPI	Usage	CPI	Usage
A	6	0.2	6 - X/100	0.6
B	9	0.3	9 - X/50	0.2
C	15	0.5	15 – X/20	0.2

   (1) What is the average CPI of the original processor?

   (2) If the redesigned processor needs two times of the instruction count as the original design for the same benchmark, then what is the minimum value of X that the performance can be improved?

   (3) Alternatively, we can add cache memory to improve the performance. Assume that, in average, each instruction needs 2.5 clocks for memory access in the original design and the cache access time is 5 times faster than the memory access. If we can design an ideal cache in the original processor (i.e. without any cache miss), then what is the minimum value of X for the redesigned processor (without cache) to achieve better performance than the original processor equipped with the ideal cache?

Answer

(1) The average CPI of the original processor = $6 \times 0.2 + 9 \times 0.3 + 15 \times 0.5 = 11.4$

(2) $\frac{1\times 11.4}{100} > \frac{2\times\left[\left(6-\frac{X}{100}\right)\times 0.6+\left(9-\frac{X}{50}\right)\times 0.2+\left(15-\frac{X}{20}\right)\times 0.2\right]}{X}$ → X > 109.09

(3) The average CPI after adding cache for original processor = 11.4 – 2.5 + 0.5 = 9.4

$\frac{1\times 9.4}{100} > \frac{2\times\left[\left(6-\frac{X}{100}\right)\times 0.6+\left(9-\frac{X}{50}\right)\times 0.2+\left(15-\frac{X}{20}\right)\times 0.2\right]}{X}$ → X > 125.37

2. For each of the three MIPS instructions below:

   (a) addi $15, $0, -8

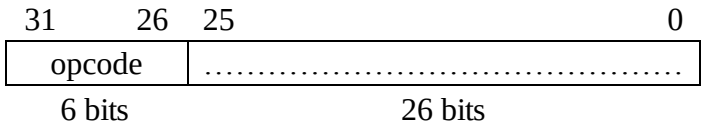
   (b) sltu $16, $15, $0

   (c) jalr $16, $15

   [Hint] The special version of the jump-and-link instruction, jalr rs, rd, jumps to the address in register rs and puts the return address in register rd.

   (1) Specify the machine code (bit 0 ~ bit 25) of the above three instructions, respectively. You should show how many bits are in each field and leave unknown fields blank. Furthermore, you do not specify the opcode for above three instructions, respectively,

and leave it blank.



The rs, rt, rd fields should be translated into binary number, respectively. For example, $15 → 01111, $0 → 00000. The bit codes should be written here and put in correct position.

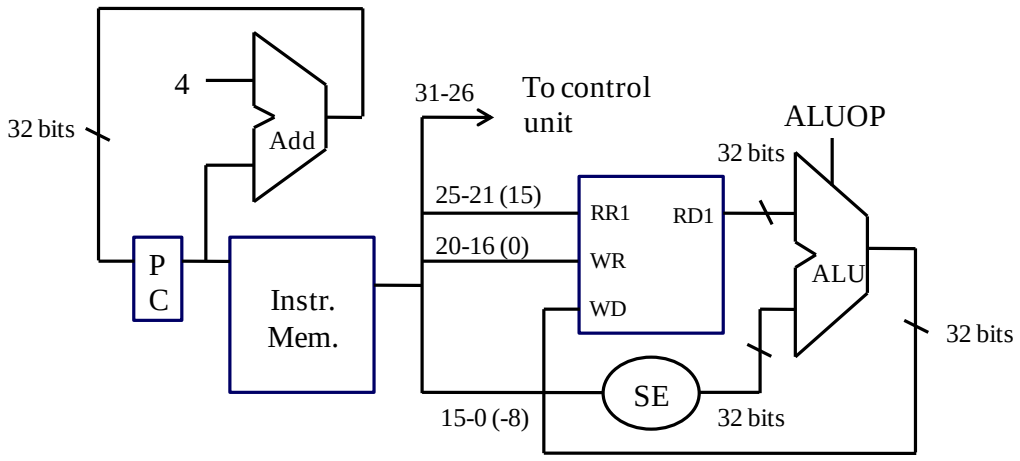
(2) Draw the datapath and controls for a single cycle implementation of the above three instructions, respectively; only include parts of the datapath that are used in the instruction; specify the bit width of any lines you draw in the datapath, and write any known values on the lines.

Answer

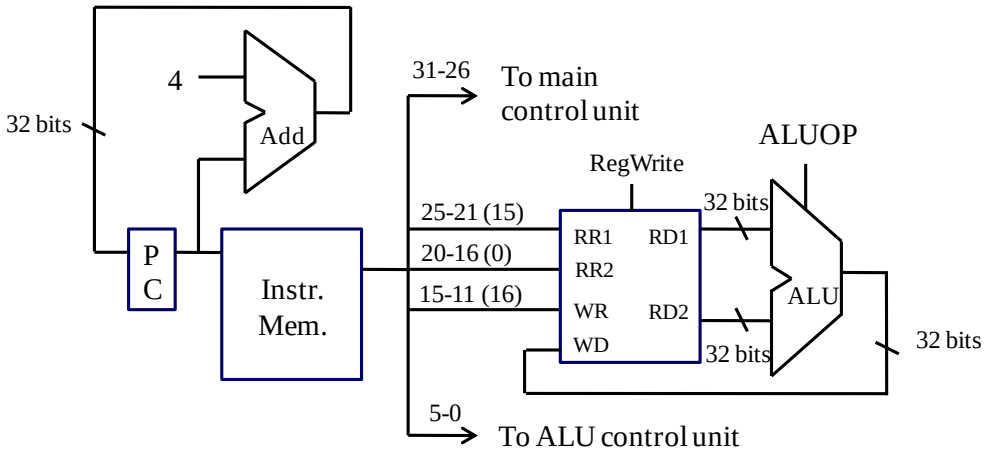
(1)

Instruction	Opcode	rs	rt	rd	shamt	funct
addi $15, $0, -8		00000	01111	1111 1111 1111 1000		
sltu $16, $15, $0		01111	00000	10000	00000	
jalr $16, $15		10000	00000	01111	00000	

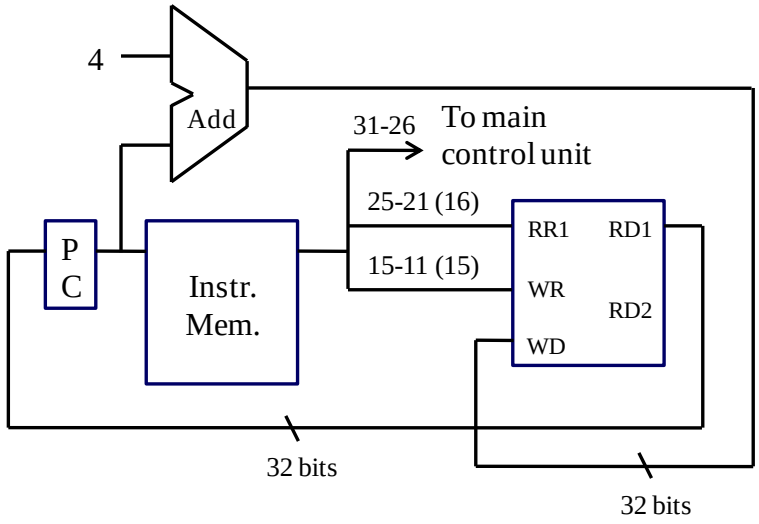
(2) addi instruction datapath:



sltu instruction datapath:



jalr instruction datapath:



3. Suppose there exists a 12-bit IEEE 754 floating point format, with 1 sign bit, 6 exponent bits, and 5 mantissa bits.
   (1) How would -∞ be represented in this 12-bit format? And what is the smallest positive normalized number. Give the value in decimal of the second number, and show both either as 12 bits or as 3 hexadecimal digits.
   (2) Give the nearest representation n of 5.612 in this format.
   (3) What is the actual value of n? Hence, work out its relative error r, to 3 significant digits. You may use the fact that a / 5.612 ≈ a × 0.1782.
   (4) Calculate n^2 using binary floating point multiplication. Show rounding, normalization and where you might check for overflow. Give the result as a 12-bit IEEE 754 number.

Answer

(1)

	Decimal value	Binary format	Hexadecimal format
Smallest positive normalized number →	- ∞	1 111111 00000	FE0
	1.0×2^{-30}	0 000001 00000	020

(2) $5.612_{10} = 101.10011\ldots. \times 2^0 = 1.0110011\ldots. \times 2^2$
→ 0 100001 01101

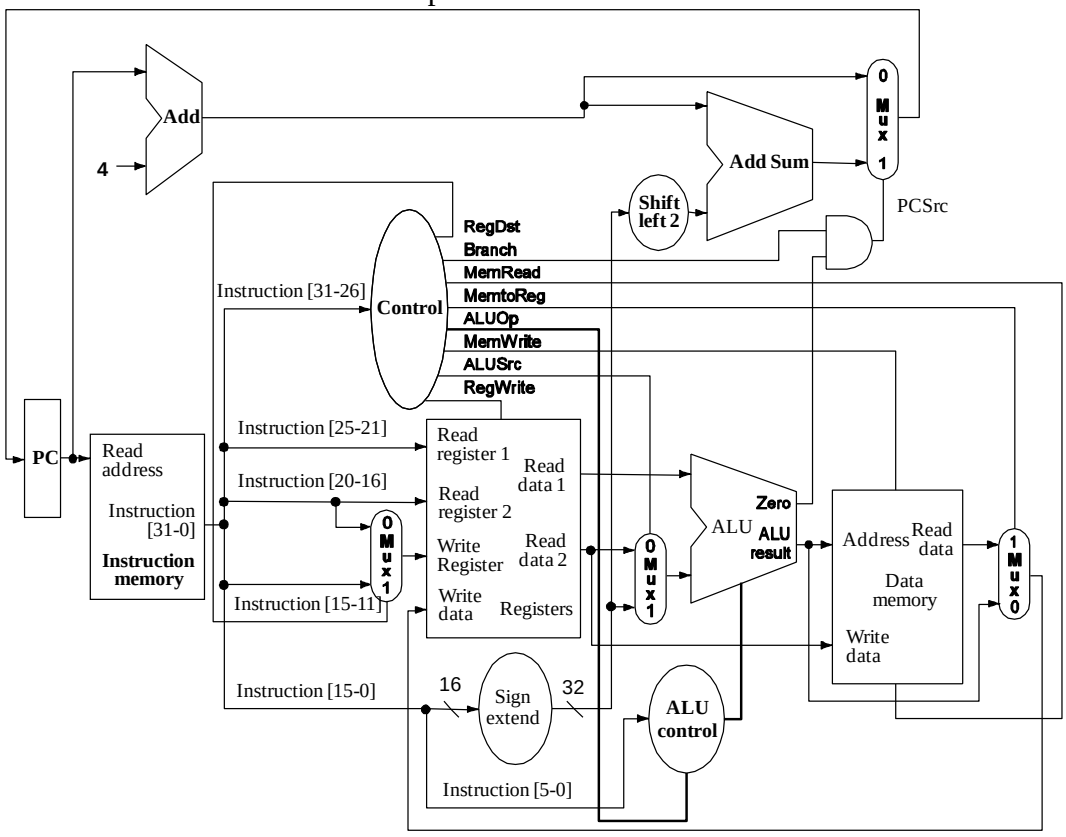
(3) The actual value of $n = 101.101_2 = 5.625$
$r = (5.625 - 5.612) / 5.612 \approx 0.013 \times 0.1782 \approx 2.32 \times 10^{-3}$

(4)

$$
\begin{array}{rl}
 & 1.0\ 1\ 1\ 0\ 1 \times 2^2 \\
\times & 1.0\ 1\ 1\ 0\ 1 \times 2^2 \\
\hline
 & 1.0\ 1\ 1\ 0\ 1 \\
 & \quad 1\ 0\ 1\ 1\ 0\ 1 \\
 & \quad\quad 1\ 0\ 1\ 1\ 0\ 1 \\
+ & \quad\quad\quad 1\ 0\ 1\ 1\ 0\ 1 \\
\hline
Round & 1.1\ 1\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ 1 \times 2^4 \\
\hline
Result & 1.1\ 1\ 1\ 1\ 1 \times 2^4
\end{array}
$$

No normalization necessary; 6-bit excess-31 exponent can accommodate $4 = 100011_2$, so no overflow. Result: $0\ 100011\ 11111_2 = 11111.1_2 = 31.5_{10}$

4. The simple datapath with the control unit for the MIPS processor is shown below. The input to the ALU control unit is the 6-bit opcode field from the instruction.



Consider the following instructions:

36: lw $1, 50($7)
40: add $4, $5, $6
44: sw $4, 50($7)
48: beq $1, $4, 8

(1) With the aids of the control signals, different instructions will utilize different hardware blocks of the datapath to perform the adequate function. What are the values of the 1-bit control signals generated by the control unit for each instruction? Please complete the following table. Note that the input to the ALU control unit is the 6-bit opcode field from the instruction. Use "1" for true, "0" for false, and "x" for don't care.

Instruction	RegDst	ALUSrc	Memto-Reg	Reg-Write	Mem-Read	Mem-Write	Branch
add							
lw							
sw							
beq							

(2) What is the new PC address after the beq instruction is executed?

(3) Assume that the latencies for logic blocks in the datapath are given below:

I-Men	Add	Mux	ALU	Regs	D-Men	Sign-extend	Shift-left-2
500ps	150ps	100ps	180ps	220ps	1000ps	90ps	20ps

Assuming zero latency for the control unit, what is the clock cycle time if the processor must support add, beq, lw, and sw instructions?

(4) Suppose that the latency for ALU control block in the datapath is 55ps. To avoid lengthening the critical path, how much time can the control unit take to generate the MemWrite signal?

(5) For the speed-up, the processor is pipelined into 5 stages: IF, ID, EX, MEM, and WB. Please indicate what data or control dependencies affect execution of the given instructions.

(6) There are several ways to reduce the branch delay. What is the technique of the "delayed branch"? Try to re-schedule the instructions by "from-before" delayed branch scheduling, if there is a 1-cycle branch delay.

Answer

(1)

Instruction	RegDst	ALUSrc	Memto-Reg	Reg-Write	Mem-Read	Mem-Write	Branch
add	1	0	0	1	0	0	0
lw	0	1	1	1	1	0	0
sw	×	1	×	0	0	1	0
beq	×	0	×	0	0	0	1

(2) If the instruction beq is taken, the new PC address = $52 + 8 \times 4 = 84$
If the instruction beq is not taken, the new PC address = 52

(3) The clock cycle time is 2220 ps.

Instruction	Latency
add	500 + 220 + 100 + 180 + 100 + 220 = 1300 (ps)
beq	500 + 220 + 100 + 180 + 100 =1100 (ps)
lw	500 + 220 + 180 + 1000 + 100 + 220 = 2220 (ps)
sw	500 + 220 + 180 + 1000 =1900 (ps)

(4) 2220 – 500 – 1000 = 720 (ps)

(5) The RAW data dependency between instruction pairs (add, sw) and (add, beq) will affect the correctness of the given instructions execution.

(6) Delayed branch is a type of branch where the instruction immediately following the branch is

always executed, independent of whether the branch condition is true or false.
The following shows the code after "from before" scheduling.

```
lw    $1, 50($7)
add   $4, $5, $6
beq   $1, $4, 8
sw    $4, 50($7)
```

5. In the original processor shown in problem 4, some actions will be taken when an exception occurs. Please pick up the right things from the following table and give correct order about those actions. (ex: (a) → (b) → (c))

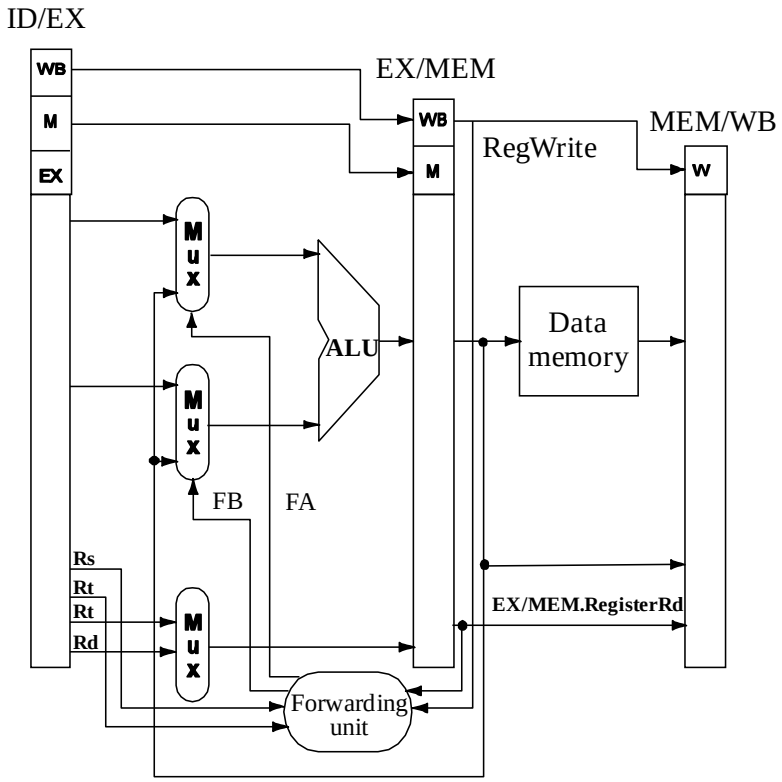
(a) stopping the execution of the program	(b) automatically execute a jal instruction
(c) save all register values into the stack	(d) execute the predefined actions for exceptions
(e) flushing all instructions	(f) transfer the control to OS at some specified address
(g) automatically stall one cycle	(h) save the address of the offending instruction in EPC

Answer: (h) → (f) → (d) → (a)

6. In this problem, we assume that the processor in problem 4 has been pipelined into 5 stages: IF, ID, EX, MEM, and WB with necessary pipeline registers.
   (1) For the datapath shown in problem 4, please modify the datapath to implement the forwarding capability that can eliminate the stalls due the data dependency in the code sequence given in problem 4. In order to simplify the answers, only the datapath of EX and MEM stages are required to be redrawn on your answer sheet.
   (2) Please explain the meanings of the added signals in (1) with the forwarding capability. The triggering conditions of the added control signals are also required.
   (3) With the forwarding capability in (1), please give the values of the control signals (RegDst, ALUSrc, MemtoReg, RegWrite, Branch) at the fifth cycle while executing the given code sequence in problem 4.

Answer

(1)



(2)

Input	Bits	Usage
ID/EX.RegisterRs	5	operand reg number, compare to see if match
ID/EX.RegisterRt	5	operand reg number, compare to see if match
EX/MEM.RegisterRd	5	destination reg number, compare to see if match
EX/MEM.RegWrite	1	TRUE if writes to the destination reg
Output	**Bits**	**Usage**
FA	2	forwarding signal
FB	2	forwarding signal

(3)

Control signal	RegDst	ALUSrc	MemtoReg	RegWrite	Branch
Value	X	1	1	1	0

7. State whether the following techniques are associated primarily with a software- or hardware-based approach to exploiting ILP (Instruction-Level Parallelism). If it is associated with a software-based approach, please give an 'S' as the answer. If it is associated with a hardware-based approach, please give an 'H' as the answer. In some cases, the answer may be both.
   (1) Branch prediction
   (2) Register renaming
   (3) Speculation
   (4) Superscalar
   (5) VLIW

Answer

(1)	(2)	(3)	(4)	(5)
B	B	B	H	S

8. Memory Hierarchy:
   (1) What is the advantage of separating instruction cache and data cache when comparing to the unified cache?
   (2) Please state why it is "virtual" in the so-called virtual memory?
   (3) A computer system has 1G bytes main memory and 512K bytes cache with 2-way set associativity and 4-byte block size. The partial content of the cache is listed as follows at a specific time instance. Please identify the possible two lowest addresses of the data that is shaded (i.e. D2).
   (Note: 1... 1 denotes consecutive 1 s and 0.. .0 denotes consecutive 0s)

Index	Tag	Data(11)	Data(10)	Data(01)	Data(00)
0	100…0	81	82	83	84
	011…1	91	92	93	94
1	000…0	A1	A2	A3	A4
	011…1	B1	B2	B3	B4
2	111…1	C1	C2	C3	C4
	100…0	D1	D2	D3	D4
3	100…0	E1	E2	E3	E4
	000…0	F1	F2	F3	F4

Answer

(1) Split cache compares with unified cache has the advantage of doubling the cache bandwidth.

(2) In virtual memory system, all physical memory is controlled by the operating system. Some memory can be stored in physical RAM chips while other memory is stored on a hard drive. When a program needs memory, it requests it from the operating system. Computer programmers no longer need to worry about where the memory is physically stored or whether the user's computer will have enough memory. The word “virtual” is used to represent the fact that some memory is actually stored on a hard drive.

(3)

Address format	Tag	Index	offset
Field length	12 bits	16 bits	2 bits
Binary value	100000000000	0000000000000010	10

102 交大資聯

單選題

1. Which one of the following statements is TRUE?
   (a) From the point of view of a network service provider, latency is a suitable metric to measure the performance of the network service server.
   (b) Mini-computer has disappeared from the computer market since main-frame computers have been well designed to have much more computing power than mini-computers.
   (c) The CPU operating clock rate has not been continuously and significantly improved these years because Moor's law has been invalidate since several years ago.
   (d) CPU and RF/analog designs are generally regarded as two main streams/representatives that can represent the semiconductor technology level of a nation.
   (e) With the fixed number of defects per area, a small die-area IC design has better manufacture yield than the big die-area one.

Answer: (e)

Remark(a): should be throughput
Remark(b): should be microcomputer
Remark(c): because power wall and memory wall

Remark(e): Manufacture Yield = $\frac{1}{(1+(\text{Defects per area} \times \text{Die area}/2))^2}$

2. Which one of the following statements is ~~in~~correct?
   (a) With the same ISA and circuit design, the CPU with higher clock rate does not necessarily run through a program in shorter runtime than the CPU with lower clock rate.
   (b) Million Instructions Per Second (MIPS) is not a unfair metric to compare the performances of two computers because MIPS metric only counts the number of executed instructions but does not consider the CPU idle time for stalling CPU requested by cache miss.
   (c) Programming languages and compilers heavily influence the number of executed machine instructions per operation.
   (d) Algorithms determine the number of executed operations.
   (e) Computer *A* runs through program *B* with 50 seconds, where 20 and 30 seconds are for integer and floating computations respectively. Computer A is promoted by replacing a new CPU running faster by two times and three times in integer and floating point operations respectively. New computer *A* can run through program *B* faster than the old one by 2.5 times.

Answer: (e)

Remark(a): Same ISA means same IC and same circuit design means same CPI.

3. Given the values of registers $t0 and $t1 in the table below, which of following statements is correct.

$t0 = 0xAAAAAAAA
$t1 = 0x22222222

(a) For the initial values of registers in the table above, the value of$t2 will become 0x55555552 after the execution of code sequence below.

```
sll  $t2, $t0, 3
sr   $t2, $t2, $t1
```

(b) For the initial values of registers in the table above, the value of$t2 will become 0x37777777 after the execution of code sequence below.

```
srl  $t2, $t0, 3
or   $t2, $t2, $t1
```

(c) For the initial values of registers in the table above, the value of $t2 will become 0xF7777777 after the execution of code sequence below.

```
srl  $t2, $t0, 3
or   $t2, $t2, $t1
```

(d) For the initial values of registers in the table above, the value of$t2 will become 0 after the execution of instruction below.

```
slt  $t2, $t0, $t1
```

(e) For the initial values of registers in the table above, the value of$t2 will become 1 after the execution of instruction below.

```
sltu $t2, $t0, $t1
```

Answer: (b)

4. Suppose the program counter (PC) is at address 0x80000000, which of following statements is incorrect.
   (a) It is possible to use one branch instruction to get to address 0x7fffff00
   (b) It is possible to use one jump instruction to get to address 0x7fffff00
   (c) It is possible to use one branch instruction to get to address 0x80000040
   (d) It is possible to use one branch instruction to get to address 0x80010000
   (e) It is possible to use one jump instruction to get to address 0x8fff0000

Answer: (b)

Remark: Actually statement (a) is also incorrect.

5. Consider the following sequence of actual outcomes for a branch, where *T* means the branch is taken, *N* means not taken, which of following statements is incorrect?

T-T-N-T-N-N-N-T-N

(a) It will be mispredicted *five* times if we always predict the branch outcomes as taken.
(b) It will be mispredicted *four* times if we always predict the branch outcomes as NOT taken.
(c) It will be mispredicted *four* times if a 1-bit predictor is used and this predictor is initialized as taken.
(d) It will be mispredicted *five* times if 1-bit predictor is used and this predictor is initialized as taken.
(e) It will be mispredicted *four* times if a 2-bit predictor is used and this predictor is initialized as taken

Answer: (c)

Remark(c): 5 times

複選題

6. Each bit's carry out of a carry lookahead adder (CLA) is a function of each bit's *generate*, *propagate*, and *carry in* signals. Consider a two-level 9-bit CLA that is composed of several 3-bit CLA units with c_0 as the carry in and c_1 to c_9 as the carry outs. Each CLA unit realizes 3-bit group generate and group propagate functions based on 3-bit inputs, A signal is regarded validate if its value is correct; or, it is regarded as invalidate. Signals g_i and p_i are the *generate* and *propagate* signals at bit *i* while G_{i-j} and P_{i-j} are the *group generate* and *group propagate* signals associated with bits *i* to *j*. Assume all input signals have been available, which ones of the following statements are correct ?

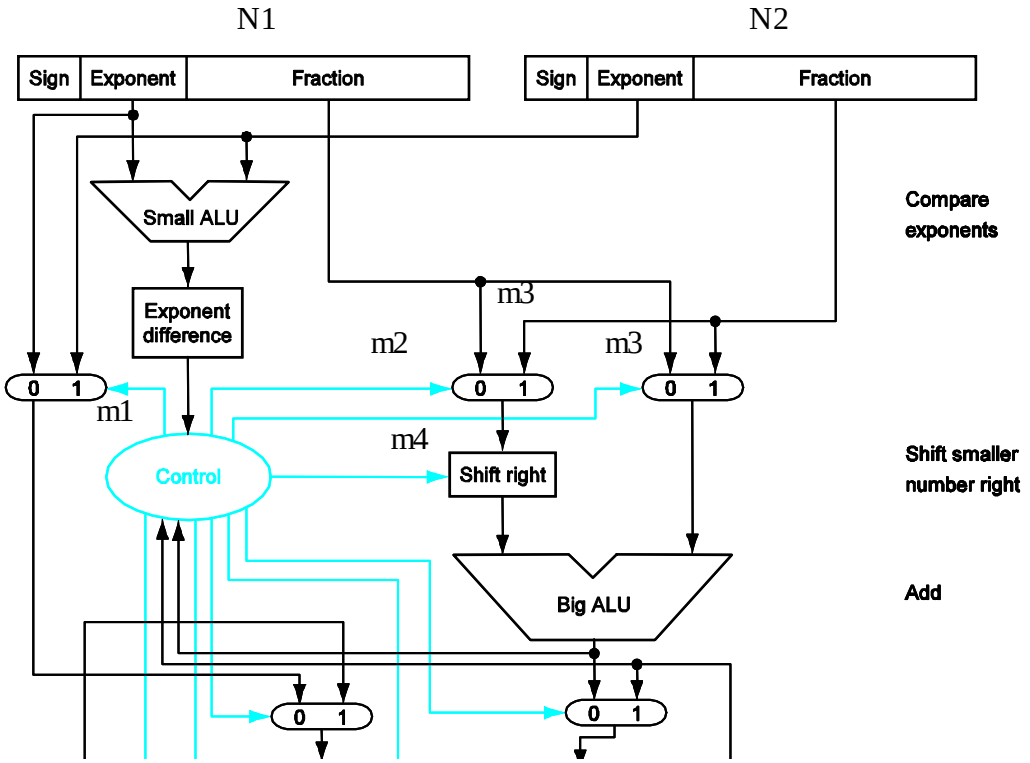
(a) A number of 4 CLA units in total are required to assemble the two-level 9-bit CLA.
(b) Carry outs c_1, c_2 and c_3 become validate earlier than other carry out signals.
(c) $G_{0-2} = g_2 + p_2g_1 + p_2p_1g_0$ and $P_{0-2} = p_2p_1p_0$.
(d) Carry outs c_4 and c_5 become validate earlier than c_6.
(e) Assume the delays of each *NOT*, isolated *AND*, *XOR*, *AND-OR* structure gates are 1, 2, 2, and 3 time units respectively, and the *sum* is realized by 2 *XOR* gates while *the propagate* is realized by a *XOR* gate. Then the longest delay of the 2-level 9-bit CLA is the 12 time units.

Answer: (b), (c)

Remark(e): 15 time units

$a_ib_i \rightarrow p_ig_i$	$p_ig_i \rightarrow P_iG_i$	$P_iG_i \rightarrow C_i$	$C_i \rightarrow c_i$	$c_i \rightarrow s_i$
2	3	3	3	4

7. Which ones of the following statements are correct?
   (a) The biased exponents in IEEE standard 754 demand signed number exponent comparison.
   (b) The number with all bits in exponent and fraction fields being 1 is referred to as infinity in IEEE standard 754.
   (c) The fraction and exponent parts of the smallest positive single precision denormalized number in IEEE standard 754 are both 0.
   (d) The below figure shows partial design for the first three steps of a floating-point adder. Assume $N1 = 3.63 \times 10^{10}$ and $N2 = 2.28 \times 10^{7}$. ($m1$, $m2$, $m3$, $m4$) = (0, 1, 0, 3).
   (e) Floating-point unit (FPU) was called co-processor before because CPU and FPU were manufactured into two independently chips. Currently modern CPU and FPU have been designed and integrated together in a single chip.



Answer: (d), (e)

8. Which ones of the following statements are correct?
   (a) Consider the code sequence {k = 10; while (k >= 1) {$Dat[k] = Dat[k] + 2$; k-- ;}}; spatial locality occurs on accessing variable k.
   (b) Consider the above code sequence; spatial locality occurs on accessing array Dat.
   (c) Assume a compiler does not have any optimization technique; Consider two sequential code sequences: Code 1: for (i = 0; i < $maxX$; i++) *for* (j = 0; j < maxY; j++) Dat[i][j]++; Code 2: for (j= 0; j < $maxY$; j++) for (i = 0; i < $maxX$; i++) Dat[i][j]++; The performances

of two sequential codes are the same since modern cache design has very high cache hit rate (almost near 100%) and thus memory access time can be reduced significantly.

(d) Consider 32-bit word-addressing and a 32-word (only data) direct-mapped cache with 4-word blocks. The address sequence for data accessing is (0, 40, 80, 340, 56, 68, 172, 348, 48, 80). A conflict miss occurs for the last access (address 80).

(e) Continue statement (d). If the cache size is increased by 4 words (36 words in total), a cache hit occurs for the last access (address 80). As compared to the accessing result of 36-word cache, the miss for the last access to the 32-word cache (in statement (d)) can be regarded as a capacity miss.

Answer: (b), (d)

Remark(d):

Word address	Block address	Tag	Index	Hit/Miss	3C
0	0	0	0	Miss	Compulsory
40	10	1	2	Miss	Compulsory
80	20	2	4	Miss	Compulsory
340	85	10	5	Miss	Compulsory
56	14	1	6	Miss	Compulsory
68	17	2	1	Miss	Compulsory
172	43	5	3	Miss	Compulsory
348	87	10	7	Miss	Compulsory
48	12	1	4	Miss	Compulsory
80	20	2	4	Miss	Conflict

Remark(e):

Word address	Block address	Tag	Index	Hit/Miss	3C
0	0	0	0	Miss	Compulsory
40	10	1	1	Miss	Compulsory
80	20	2	2	Miss	Compulsory
340	85	9	4	Miss	Compulsory
56	14	1	5	Miss	Compulsory
68	17	1	8	Miss	Compulsory
172	43	4	7	Miss	Compulsory
348	87	9	6	Miss	Compulsory
48	12	1	3	Miss	Compulsory
80	20	2	2	Hit	

9. A system has a 16-bit virtual address size with 256B pages and a 16-bit byte addressing physical memory with physically index 2-way set associative data cache (LRU replacement policy) of512B cache size and 16B blocks. The first six accesses to the page table are: 0x1332, 0xad58, 0x0162, 0x2ea9, 0x813d, 0x9f5a. The system uses a single level page table. Assume the TLB and data cache are initially empty and their contents have been updated accordingly by the first six accesses (page table is shown below; VPN: virtual page number, PPN: physical page number). The CPU includes a fully associative TLB with 2 entries and an LRU replacement policy. Data reads from the following virtual addresses are performed (in the order listed): 0x813f, 0xl36f, 0x2e5f, 0x9f50, 0x1370. Which ones of following statements are correct?

VPN	PPN	VPN	PPN
0x01	0x27	0x81	0x8a
0x13	0x45	0x9f	0xcd
0x2e	0xe8	0xad	0x7e

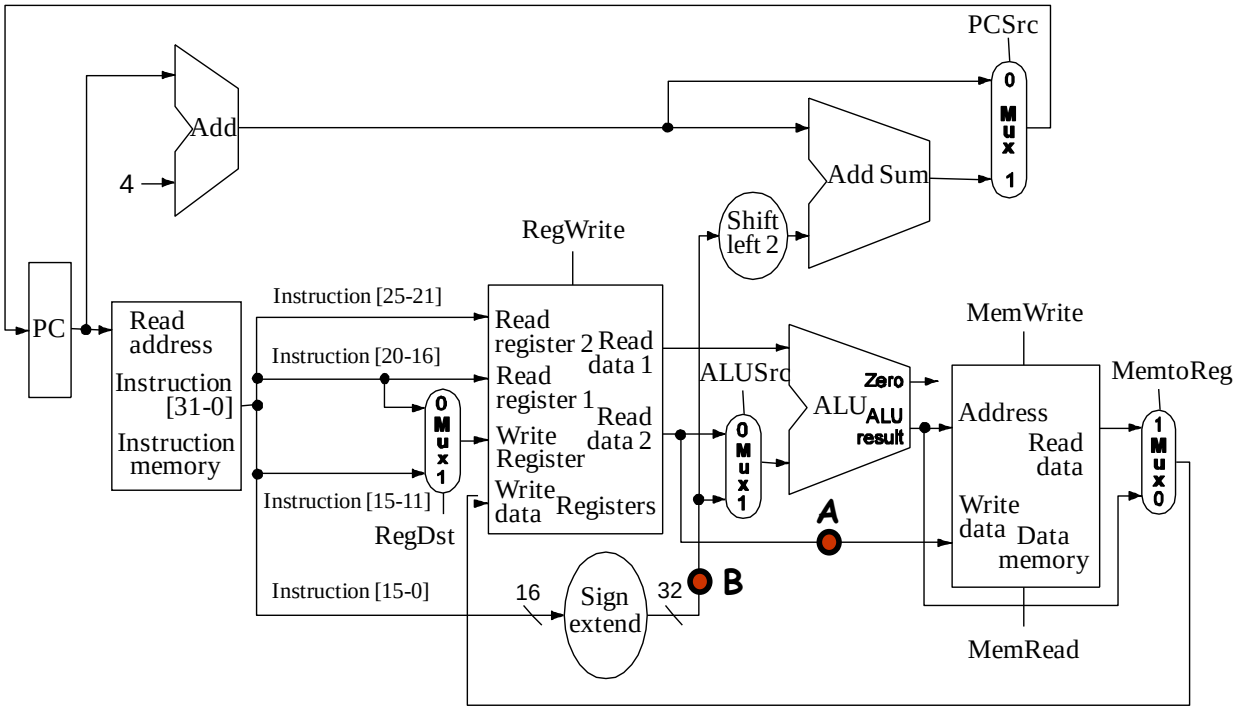
(a) The required numbers of bits for virtual page number and page offset in virtual address are all 8 bits.
(b) The required number of bits for index and tag in physical address are 4 and 8 bits respectively.
(c) Data access to 0x813fhas access hit in TLB and cache.
(d) Data access to 0X9f50 has access miss in TLB and access hit in cache.
(e) Data access to 0X1370 has access miss in TLB and cache.

Answer: (a) , (b) , (c) , (d) , (e)

Remark:

	Virtual address		Physical address			TLB	Cache
	VPN	PO	tag	index	offset		
(a), (b)	8bits	8bits	8bits	4bits	4bits	H/M	H/M
	13	32	45	3	2	M	M
	ad	58	7e	5	8	M	M
	01	62	27	6	2	M	M
	2e	a9	e8	a	9	M	M
	81	3d	8a	3	d	M	M
	9f	5a	cd	5	a	M	M
(c)	81	3f	8a	3	f	H	H
	13	6f	45	6	f	M	M
	2e	5f	e8	5	f	M	M
(d)	9f	50	cd	5	0	M	H
(e)	13	70	45	7	0	M	M

10. For the Single-Cycle MIPS CPU shown below, which of following statements are correct?



(a) If the path labeled A has been cut, the instructions add, slt, and sw still can run correctly.

(b) If the path labeled B has been cut, the instructions lw, sw and beq may fail.

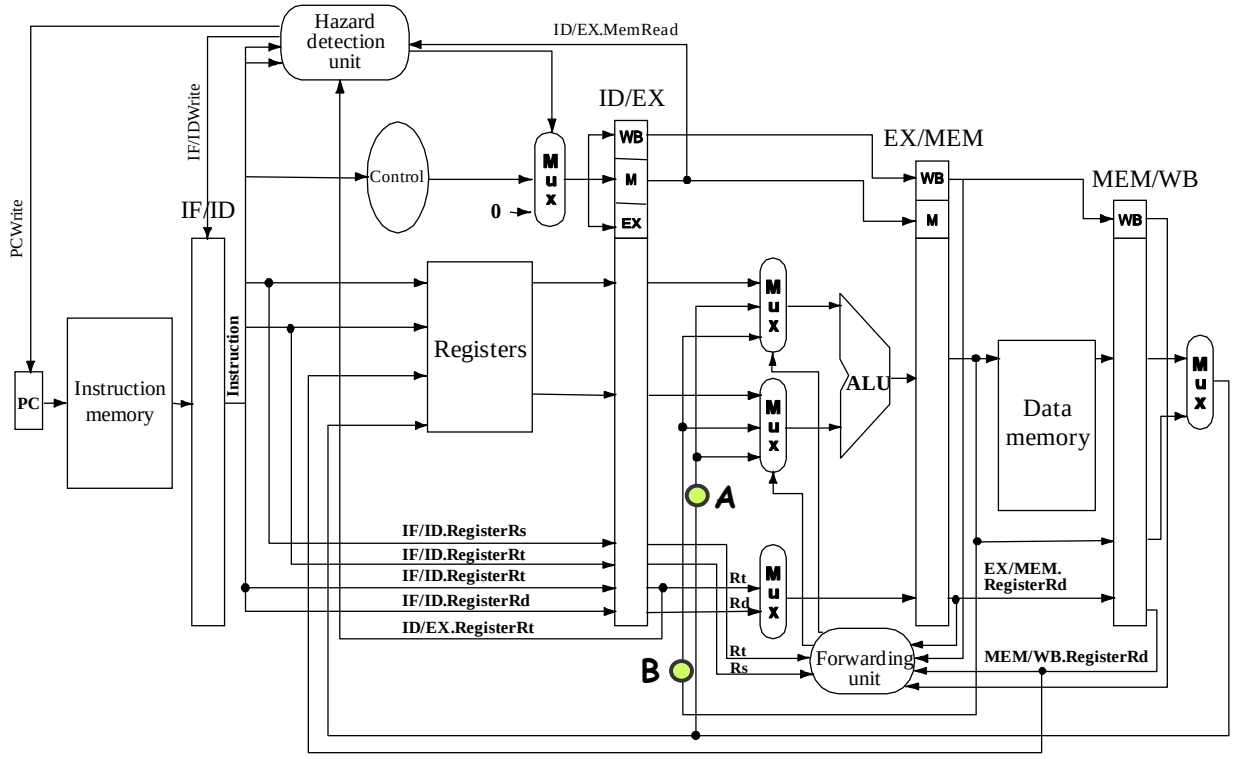
(c) If the control signal *ALUsrc* is stuck on 0, the instructions lw, sw, and beq may fail to run correctly.

(d) If the control signal *RegDst* is stuck on 0, the instructions lw and sw still can run correctly.

(e) If the control signal *MemToReg* is stuck on 1, the instructions add, sub, lw, and slt may fail to run correctly.

Answer: (b), (d)

11. Given the pipelined MIPS CPU below where assume both forwarding and stall mechanisms have been designed, which of following statements are correct?



(a) Assume the path labeled A has been cut, the following code snippet will fail.

```
add  $s2, $s0, $s1
add  $s3, $s0, $s2
```

(b) Assume the path labeled A has been cut, the following code snippet will fail.

```
add  $t1, $t1, $s2
add  $t1, $t0, $t2
add  $s2, $s1, $t1
```

(c) Assume the path labeled B has been cut, the following code snippet will fail.

```
add  $t1, $t1, $s2
add  $t1, $t0, $t2
add  $s2, $s1, $t1
```

(d) Assume the path labeled A has been cut, the following code snippet will fail.

```
lw   $s0, 4($t1)
add  $s2, $s0, $s1
```

(e) Assume the path labeled B has been cut, the following code snippet will fail.

```
lw   $s0, 4($t1)
add  $s2, $s0, $s1
```

Answer: (c), (d)

12. Consider the code sequence below, where we predict *beq* as taken, but actually it is NOT taken, which of following statements are correct?

```
lw   $s0, 0($t0)
lw   $s1, 0($t0)
beq  $s0, $s1, L1
add  $s2, $s0, $s1
```

(a) For the pipelined MIPS CPU which has data forwarding mechanism and the branch outcome is determined in MEM stage, it takes 8 cycles to complete the execution of the above code sequence.

(b) For the pipelined MIPS CPU which has data forwarding mechanism and the branch outcome is determined in MEM stage, it takes 12 cycles to complete the execution of the above code sequence.

(c) For the pipelined MIPS CPU which has data forwarding mechanism and the branch outcome is determined in ID stage, it takes 7 cycles to complete the execution of the code sequence.

(d) For the pipelined MIPS CPU which has data forwarding mechanism and the branch outcome is determined in ID stage, it takes 11 cycles to complete the execution of the above code sequence.

(e) For the pipelined MIPS CPU which has data forwarding mechanism and the branch outcome is determined in ID stage, it takes 10 cycles to complete the execution of the above code sequence.

***Answer*:** (d)

Remark(a): (5 – 1) + 4 + 1 + 1 = 10
Remark(d): (5 – 1) + 4 + 1 + 2 = 11

102清大資工

1. Given the table of latencies of individual stages of a MIPS instruction, please answer the following questions:

Instruction fetch	Register read	ALU operation	Data Access	Register write
200ps	100ps	250ps	400ps	100ps

(a) What is the clock cycle time in a nonpipelined and pipelined processor?
(b) What is the total latency of a lw instruction in a nonpipelined and pipelined processor?
(c) If the time for an ALU operation can be shortened/increased by 25%, how much speedup/slowdown will it be from pipelining? Explain the reason.
(d) If you can split one stage into two, each with half the original latency, to improve the pipelining performance. Which stage would you split and what is the new clock cycle time in a pipelined processor?

Answer

(a) The clock cycle time in a nonpipelined processor is 1050ps
The clock cycle time in a pipelined processor is 400ps
(b) The total latency of a lw instruction in a nonpipelined processor is 1050ps
The total latency of a lw instruction in a pipelined processor is 2000ps
(c) Shortening/Increasing the ALU operation will not affect the speedup/slowdown obtained from pipelining. It would not affect the clock cycle.
(d) Data Access stage
New clock cycle time = 250ps

2. Please answer the following questions about interfacing I/O devices to the processor and memory:
(a) Describe the following terminologies:
- Memory-mapped I/O.
- I/O instruction.
- Device polling.
- Interrupt-driven communication.

(b) Which communication pattern is most appropriate for a "Video Game Controller"? Explain.
(c) Prioritize the following interrupts caused by different devices:
- Mouse Controller.
- Reboot.
- Overheat.

Answer

(a)

Memory-mapped I/O: An I/O scheme in which portions of address space are assigned to I/O devices and reads and writes to those addresses are interpreted as commands to the

I/O device.

I/O instruction: A dedicated instruction that is used to give a command to an I/O device and that specifies both the device number and the command word (or the location of the command word in memory).

Device polling: The process of periodically checking the status of an I/O device to determine the need to service the device.

Interrupt-driven communication: An I/O scheme that employs interrupts to indicate to the processor that an I/O device needs attention.

(b) Device polling

(c) Mouse Controller → 3
Reboot → 2
Overheat → 1

3. In general, a fully associative cache is better than a direct-mapped cache in term of miss rate. However, it is not always the case for a cache, especially for a small cache. Please design an example to demonstrate that a direct-mapped cache outperforms a fully associative cache in term of miss rate under the replacement policy. Least Recently Used (LRU).

Answer: Suppose the cache has 2 blocks and CPU has the follow references of block address

1, 2, 3, 4, 5, 1, 2, 5, 1, 2, 3, 4, 5

	Direct-mapped			Fully associative	
Block address	Tag	Index	Hit/Miss	Tag	Hit/Miss
1	0	1	Miss	1	Miss
2	1	0	Miss	2	Miss
3	1	1	Miss	3	Miss
4	2	0	Miss	4	Miss
1	0	1	Miss	1	Miss
2	1	0	Miss	2	Miss
5	2	1	Miss	5	Miss
1	0	1	Miss	1	Miss
2	1	0	**Hit**	2	Miss
3	1	1	Miss	3	Miss
4	2	0	Miss	4	Miss
5	2	1	Miss	5	Miss

This example shows the miss rate for fully associative cache is higher than that of the direct-mapped cache.

4. Consider adding a new index addressing mode to the machine A such that the following code sequence

```
ADD   R1, R1, R2          // R1 = R1 + R2
LW    Rd, 0(R1)           // load
```

can be combined as

```
LW    Rd, 0(R1 + R2)
```

(a) Assume that the load and store instructions are 10% and 20% of all the benchmarks considered, respectively, and this new addressing mode can be used for 5% of both of them. Determine the ratio of instruction count on the machine A before adding the new addressing mode and after adding the new addressing mode.

(b) Assume that adding this new addressing mode also increases the clock cycle by 3%, which machine (either machine A *before* adding the new addressing mode or machine A *after* adding the new addressing mode) will be faster and by how much?

Answer

(a) $\frac{\text{IC}_{\text{old}}}{\text{IC}_{\text{new}}} = \frac{\text{IC}_{\text{old}}}{\text{IC}_{\text{old}} \times (1 - 0.05 \times 0.3)} = 1.01523$

(b) $\text{speedup} = \frac{(1 - 0.5 \times 0.3) \times 1.03 \times 1}{1} = 1.01455$

Machine A before adding the new addressing mode is faster by 1.01455 times.

5. The following Boolean equation describes a 3-output logic function.

O1 =A'BC'D'+AB'C+FGH+E'

O2 = BCD + CDE 'H + FGH'

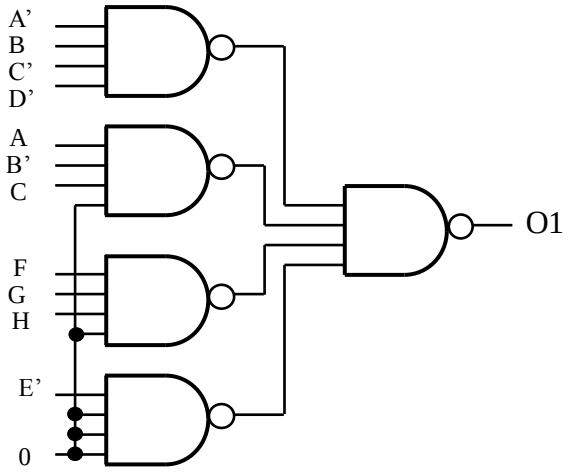
O3 = AB 'CD + ABEF + F'G'

(a) Draw a circuit diagram for *O1* using only 4-input NAND gates.

(b) If the 3-output function is implemented with a read-only memory, what size of ROM is needed?

Answer

(a)



(b) Since there are total 8 Boolean variables included in the 3 functions, the length of the ROM address is 8 bits. The ROM size would be $2^8 \times 3$ bits = 768 bits

6. Assume the floating point is stored in a 32 bits wide format. The leftmost bit is the sign bit, the exponent of 16 in excess-64 format is 7 bits wide, and the mantissa is 24 bits long and a hidden 1 is assumed.
   (a) What is the smallest positive number representable in this format (write the answer in hex and decimal)?
   (b) What is the largest positive number representable (write the answer in hex and decimal)?
   (c) Assume A = 2.34375×10^{-2} and B = 2.9015625×10^{2}. Write down the bit pattern using this format for *A* and *B*.
   (d) Using the format presented in (c), calculate the *A* – *B*.

Answer

	Binary format	Hexadecimal	Decimal value
(a)	0 0000001 00000000000000000000000	01000000	$+1.0 \times 16^{-63}$
(b)	0 1111110 11111111111111111111111	7EFFFFFF	$\approx +2.0 \times 16^{62}$

(c) A = $2.34375 \times 10^{-2} = 0.0234375_{10} = 0.0000011_2 = 0.011_2 \times 16^{-1}$
The format is 0 1111111 01100000000000000000000
B = $2.9015625 \times 10^{2} = 290.15625_{10} = 100100010.00101_2 = 0.00010010001000101_2 \times 16^{3}$
The format is 0 1000011 00010010001000100000000

(d) $0.011_2 \times 16^{-1} = 0.00000000000000000011_2 \times 16^{3}$
$0.00000000000000000011_2 \times 16^{3} - 0.00010010001000101_2 \times 16^{3}$
$= - 0.00010010001000100001_2 \times 16^{3} = - 0000100100010.0010001_2$
$= - 290.1328125_{10}$

7. Assume garage collection comprises 30% of the cycles of a program. There are two methods used to speed up the program. The first one would be automatically handle garbage collection in hardware. This causes an increase in cycle time by a factor of 1.4. The second one provides new hardware instructions that could be used during garbage collection. This would use only one-third of instructions needed for garbage collections but increase the cycle time by 1.2. Compute the program execution time for both methods and show which one is better.

Answer

Automatic garbage collection by hardware: The execution time of the new machine is $(1 - 0.3) \times 1.4 = 0.98$ times that of the original.

Special garbage collection instructions: The execution time of the new machine is $(1 - 2 \times 0.3/3) \times 1.2 = 0.96$ times that of the original.

Therefore, the second option is the best choice.

102成大電機

Choose the correct answers for the following multiple choice problems. Each question may have more than one answer, 10 points each, no partial point, no penalty.

1. Which of the following is (are) true for a 128KB cache with a line size of 64 bytes? Assume that the cacheable memory is 4 GB. Address bits are numbered from A0 to A31
   (a) In a direct-mapped implementation, the index field uses address bit A6 to A16 for cache line selection.
   (b) In a direct-mapped implementation, the tag length is 13 bits.
   (c) In a direct-mapped implementation, the tag length is 15 bits; the field determining the line size is 6 bits in length.
   (d) In a direct-mapped implementation, the total tag size is 30720 bits.

Answer: (a), (c), (d)

Remark: Number of blocks = 128KB / 64B = 2K

Address format:

Tag	Index	Offset
15	11	6
A31 ~ A17	A16 ~ A6	A5 ~ A0

Total tag bit = 2K × 15 = 2048 × 15 = 30720

2. Which of the following is (are) true for programmed I/O?
   (a) The transfers of I/O data are performed by a DMA device in the I/O unit.
   (b) The programmed I/O operations are initiated by an interrupt and some special commands.
   (c) The processor executes some load/store instructions and others to transfer the I/O data.
   (d) I/O devices move the data between themselves.

Answer: (c)

3. Which of the following is (are) true for hazards in a pipelined processor?
   (a) RAW hazard comes from an instruction cache miss.
   (b) Considering two instructions i and j, with i occurring before j, j tries to read a source before i writes it, so j incorrectly gets the old value. This is a RAW hazard.
   (c) Considering two instructions i and j, with i occurring before j, j tries to read a source before i writes it, so j incorrectly gets the old value. This is a WAR hazard.
   (d) Considering two instructions i and j, with i occurring before j, j tries to write a destination before it is read by i, so i incorrectly gets the new value. This is a RAW hazard.

Answer: (b)

4. Which of the following is (are) true for virtual memory system?
   (a) Virtual memory function can be enabled through software control.
   (b) Virtual memory technique treats part of the main memory as a fully-set associative write-back cache for program execution.
   (c) A translation lookaside buffer can be seen as the cache of a page table.
   (d) A page table is shared among the programs in execution.

Answer: (b), (c)

5. Which of the following is (are) true as a processor is booted from power-on?
   (a) The program counter value is fetched by the processor from a specified memory location through a load instruction.
   (b) The first instruction is fetched using the program counter value specified by operating system.
   (c) The BIOS provides the address of the first instruction being fetched.
   (d) The first instruction is fetched by the processor using the address specified in the program counter.

Answer: (d)

Remark: When the PC powers on, the processor's registers are set to some predefined values. One of the registers is the *instruction pointer* (or program counter) register and its value after a power on is well defined: it is a 32-bit value of 0xfffffff0. The instruction pointer register points to code to be executed by the processor. The computer's hardware translates this address so that it points to a BIOS memory block.

BIOS is a chip on the motherboard that has a relatively small amount of read-only memory (ROM). This memory contains various low-level routines that are specific to the hardware supplied with the motherboard. So, the processor will first jump to the address 0xfffffff0, which really resides in the BIOS's memory. Usually this address contains a jump instruction to the BIOS's POST routines. POST stands for *Power On Self Test*. This is a set of routines including the memory check, system bus check and other low-level stuff so that the CPU can initialize the computer properly. The important step on this stage is determining the boot device. All modern BIOS's allow the boot device to be set manually, so you can boot from a floppy, CD-ROM, harddisk etc.

Remark(c): The BIOS contains the first instruction being fetched.

6. Which of the following is (are) true about instruction set architecture (ISA)?
   (a) ISA is an abstraction which is the interface between the hardware and the low-level software (assembly instructions).
   (b) ISA enables the different implementations of a processor using the same ISA to run identical software.
   (c) ISA specifies how a processor pipeline should be implemented.
   (d) MIPS and ARM are both RISC-type ISA and use the same instructions set architecture.

Answer: (b)

Remark: ISA is an abstraction which is the interface between the hardware and the low-level software (Machine language).

7. Which of the following is (are) true about page fault and TLB exception?
   (a) A page fault is signaled by operating system so that the OS can fetch the missing page.
   (b) A TLB exception is handled by the DMA controller in a hard drive.
   (c) Both a page fault and a TLB miss are exceptions and signaled by the processor hardware.
   (d) When a requested page is not found in the main memory, this causes a page fault.

Answer: (c), (d)

Remark: A page fault is a trap to the software raised by the hardware. The hardware that detects a page fault is the memory management unit in a processor. The exception handling software that handles the page fault is generally part of the operating system.

8. Which of the following is (are) true about cache operations?
   (a) When a cache write miss occurs, the written data are directly updated in the next level of memory. This is the write-around policy.
   (b) When a cache write hit occurs, the written data are also updated in the next level of memory. This is the write-through policy.
   (c) There is no cache coherency problem for using the write-through cache since the data are written into the next level of memory.
   (d) Cache is pronounced as [kæʧ].

Answer: (a), (b)

Remark: Cache is pronounced as [kæʃ]

9. Which of the following is (are) true about the label specified in a branch instruction, for instance, beq r1, r2, label?
   (a) The label is a pseudo-instruction and the compiler compiles it into a memory location.
   (b) The label is transformed into an offset specifying the difference between the branch instruction (or the instruction after the branch) and the branch target.
   (c) The label is transformed into an absolute address in memory for the branch target.
   (d) If the label is an external reference, the linker can figure out its value and finalize the binary format for the branch instruction.

Answer: (b), (d)

10. Which of the following is (are) true?
    (a) Pipelining improves the instruction throughput, i.e., IPC, rather than individual instruction execution time.
    (b) Pipelining improves the instruction throughput, i.e., IPC, other than individual instruction execution time.
    (c) Pipelining improves the instruction throughput, i.e., CPI, rather than individual instruction execution time.
    (d) Pipelining improves the instruction throughput, i.e., CPI, other than individual instruction execution time.

Answer: (a)

102成大電通

1. Read the following paragraph and choose the correct answers from the following multiple choice questions. Each question may have more than one answer. No partial point, no penalty. "Just as CPU programmers were forced to explicitly manage CPU memories in the days before virtual memory, for almost a decade, GPU programmers directly and explicitly managed the GPU memory hierarchy. The recent release of NVIDIA's Fermi architecture, however, has brought GPUs to an inflection point: it implements a unified address space that eliminates the need for explicit memory movement to and from GPU memory structures. Yet, without demand paging, something taken for granted in the CPU space, programmers must still explicitly reason about available memory. The drawbacks of exposing physical memory size to programmers are well known. Which of the followings are true?"
   (a) Without virtual memory technology, CPU programmers must explicitly manage the physical memory.
   (b) There is no need for the programmers of the Fermi architecture to worry about the available memory to use.
   (c) With virtual memory, a CPU programmer implements demand paging policy in his/her program.
   (d) GPU stands for General Processing Unit.

Answer: (a)

2. Answer the following questions about virtual memory.
   (a) What event triggers a page fault?
   (b) What event triggers a TLB miss?
   (c) Does a process have its own page table or all the active processes share a page table?

Answer

(a) The even when an accessed page in not present in main memory.
(b) The even when there is no matching entry in the TLB for a page.
(c) Each active process has its own page table.

3. For CPUs, the problem of exception support was solved at a relatively early stage. This support was a key enabler to their success, and instrumental in this success was the definition of precise exception handling. In a pipelined processor, multiple exceptions may occur at the same clock cycle. Assume this is a five-stage pipeline. Write down which exception (if any) may occur at the IF, ID, EXE, MEM, WB stage.

Answer

Stage	Exception
IF	(1) Instruction TLB miss (2) Hardware malfunction
ID	(1) Undefined Instruction (2) Hardware malfunction
EXE	(1) Arithmetic Overflow (2) Hardware malfunction
MEM	(1) Memory protection error (2) Data TLB miss (3) Hardware malfunction
WB	(1) Hardware malfunction

102 成大資工

1. Please describe the steps involved in developing and executing assembly language programs?

Answer

Step1: Write an assembly language program.
Step2: The assembler translates assembly language statements to their binary equivalents (object code).
Step3: Separately assembled modules are combined into one single load module, by the linker.
Step4: The program loader copies the program into the computer's main memory, and at execution time, program execution begins.

2. Please write the MIPS code to compute the mathematic formula $a \times b + 3c - 10$.

Answer:

Assume that the values of variables a, b, and c are contained in registers $s0, $s1, and $s2, respectively, and the computation result will be included in register $t0 and is smaller than 32 bit binary number can represent.

```
mul   $t0, $s0, $s1
sll   $t1, $s2, 1
add   $t1, $t1, $s2
add   $t0, $t0, $t1
addi  $t0, $t0, -10
```

3. (a) If processor A has a higher clock rate than processor B, and processor A also has a higher MIPS rating than processor B, explain whether processor A will always execute faster than processor B.
   (b) Computer A has an overall CPI of 1.5 and can be run at a clock rate of 700MHz. Computer B has a CPI of 2.0 and can be run at a clock rate of 650MHz. We have a particular program to run and this program has exactly 120,000 instructions when compiled for computer A. How many instructions would the program need to have when compiled for Computer B if we want the two computers to have exactly the same execution time for this program?

Answer

(a) We cannot differentiate which machine is faster from the measure of MIPS before the capabilities of the ISA of these two machines are given.
(b) Suppose IC_B represents instruction count for Computer B
Execution time for Computer A= Execution time for Computer B → $(120000 \times 1.5) / 700M = (IC_B \times 2) / 650M$ → $IC_B = 83571.429$

4. (a) What are the two characteristics of program memory accesses that caches exploit?
   (b) What are three types of cache misses?

Answer

(a) **Temporal locality:** if an item is referenced, it will tend to be referenced again soon.
Spatial locality: if an item is referenced, items whose addresses are close by will tend to be referenced soon.

(b) **Compulsory misses:** a cache miss caused by the first access to a block that has never been in the cache.
Capacity misses: a cache miss that occurs because the cache even with fully associativity, can not contain all the block needed to satisfy the request.
Conflict misses: a cache miss that occurs in a set-associative or direct-mapped cache when multiple blocks compete for the same set.

5. Answer TRUE or FALSE for the following questions.
   (a) A virtual cache access time is always faster than that of a physical cache.
   (b) Both DRAM and SRAM must be refreshed periodically using a dummy read/write operation,
   (c) High associativity in a cache reduces compulsory misses.
   (d) A write-through cache typically requires less bus bandwidth than a write-back cache.
   (e) Memory interleaving is a technique for reducing memory access time through increased bandwidth utilization of the data bus.

Answer

(a)	(b)	(c)	(d)	(e)
True	False	False	False	True

Remark:

(a) True (for a single level cache) since the virtual to physical address translation is avoided.
(b) False. Only DRAM needs to be refreshed periodically.
(c) False. It reduces conflict misses.
(d) False. It is the other way round

102中央資工

複選題

1. Which of the following statement(s) should be true?
   (a) There are more addressing modes in CISC than in RISC.
   (b) Microprogramming is usually used in CISC since this kind of control can be executed faster.
   (c) Horizontal microinstructions are executed faster than Vertical microinstructions.
   (d) Using ROM in the control designs can usually save the space or cost, when compared with PLA.
   (e) None of the above is true.

Answer: (a), (c)

2. About the performance analysis, which of the following statement(s) should be true?
   (a) Assume that a C program is compiled into 1000 machine instructions, which are related to the size of the executable file. Then, the average execution time is usually equal to multiplying 1000 (instructions) by its average CPI and the clock cycle time.
   (b) In evaluating the performance by using the benchmark tests, we usually use the geometric mean to calculate the average value among various test results.
   (c) MIPS is not a reliable metric since it provides the wrong results when we compare the performances of a compiled program running on two machines with the same instruction set architecture.
   (d) It is usually a preferred approach to consider only one of the three factors: clock rate, CPI or the instruction count, and then try to improve it to decrease the execution time. That is called a divide-and-conquer methodology.
   (e) None of the above is true.

Answer: (e)

3. About the memory hierarchy, which of the following statement(s) should be true?
   (a) A page fault happens if a page that is not in the physical memory is accessed.
   (b) TLB is a data cache of main memory.
   (c) Adding a secondary cache can help to reduce the miss rate.
   (d) We can always increase the size of cache to improve the cache performance.
   (e) None of the above is true.

Answer: (a)

4. What are the advantages of register-register ALU instruction type compared with register-memory ALU instruction type and memory-memory ALU instruction type?
   (a) Data can be accessed without loading first.
   (b) Fixed-length
   (c) Simple code-generation
   (d) Instructions take similar Clock Cycle per Instruction (CPI) to execute.
   (e) Large variation in instruction size

Answer: (b), (c), (d)

Remark: advantage for register-register ISA (from "Computer architecture: quantity approach" book)
1. Simple, fixed length instruction encoding
2. Simple code generation model
3. Instructions take similar number of clocks to execute

5. What are the properties critical to program correctness that are normally preserved by control dependency?
   (a) Temporal locality
   (b) Spatial locality
   (c) Exception behavior
   (d) Dataflow
   (e) None of the above

Answer: (c), (d)

Remark: (from "Computer architecture: quantity approach" book)

Control dependence is not the critical property that must be preserved; instead, the two properties critical to program correctness and normally preserved by maintaining both data and control dependence are:
1. The exception behavior – any change in the ordering of instruction execution must not change how exceptions are raised in the program (or cause any new exceptions)
2. The data flow – the actual flow of data values among instructions that produce results and those that consume them.

6. What of the following statements are true?
   (a) If some combination of instructions cannot be accommodated because of resource conflicts, the machine is said to have a structural hazard.
   (b) The potential overlap among instructions is called Instruction-Level Parallelism.
   (c) If there exists a loop-carried dependence, it is not possible to transform the code and unroll the loop to make iterations be executed in parallel.
   (d) Hazards usually have smaller impact on longer pipelines.
   (e) For a cache with write through strategy, read misses cannot result in writes.

Answer: (a), (b), (c), (e)

Remark(c):

Data accesses in later iterations are dependent on data values produced in earlier iterations; such a dependence is called a loop-carried dependence.
Loops with no loop carried dependence can parallelized (iterations executed in parallel)
Loops with loop carried dependence cannot be parallelized (must be executed in the original order)

單選題

7. Assume that a computer uses 32-bit addressing and has a 1MB cache with the block size equal to 32 bytes. The size of the Tag field in the address for a direct mapped cache is L. The size of the Tag field in the address for an 8 way set associative cache is M. The size of the Tag field in the address for a fully associative cache is N. L + M + N = K and
   (a) $K < 50$
   (b) $50 \leq K < 55$
   (c) $55 \leq K < 60$
   (d) $60 \leq K < 65$
   (e) $65 \leq K$

***Answer*:** (b)

Remark: L = 12; M = 15; N = 27

8. A cache with 64 blocks and a block size of 32 bytes. The byte address 2473537 maps to the block number K ($0 \leq K < 64$). K mod 5 equals to?
   (a) 0
   (b) 1
   (c) 2
   (d) 3
   (e) 4

***Answer*:** (a)

Remark: K = 50

9. The following table shows the information about a machine executing instructions. There are 5 classes of instructions, each of which takes 3-5 steps.

Instruction Class	Fetch	Register Read	ALU	Memory access	Register Write
Load Word (LW)	2ns	Ins	2ns	3ns	Ins
Store Word (SW)	2ns	Ins	2ns	3ns	
R-Format (R)	2ns	Ins	2ns		Ins
Branch (B)	2ns	Ins	2ns		
Jump (J)	2ns	Ins	2ns		

For a certain program, 25% of the instructions are LW, 10% are SW, 45% are R, 15% are B, 5% are J. Consider the following three cases: (i) If there is no pipelining, the average time required for executing one instruction is L (ns). (ii) If a 5-stage pipeline is adopted, the average time required for executing one instruction in the ideal case is M (ns). (iii) Assume that the destinations of Branch (B) and Jump (J) will not be known until the end of the ALU stage. If every B or J instruction incurs two pipeline bubbles, the average time required for executing one instruction is N (ns). L + M + N = K. Then

(a) $K < 13$
(b) $13 \leq K < 14$
(c) $14 \leq K < 15$
(d) $15 \leq K < 16$
(e) $16 \leq K$.

Answer: (e)

Remark: L = 9 ns; M = 3 ns; N = 4.2 ns

10. Suppose that there is a machine with 32-bit addresses and a two-level page table. Note that the first 10 bits of an address is an index into the first level page table and the next 10 bits are an index into a second level page table. The page tables are stored in memory and each entry in the page tables is 4 bytes. Suppose the space allocated to a specific process is 64 Mbytes. How many pages are occupied by this process?

(a) 2^{26} pages
(b) 2^{14} pages
(c) 2^{12} pages
(d) 2^{10} pages
(e) None of the above

Answer: (b)

Remark: Page offset = 32 – 10 – 10 = 12 → page size = 4KB → number of pages = $2^{26} / 2^{12} = 2^{14}$

11. (Continued from the previous question.) How much space is occupied in memory by the page tables for the specific process?
    (a) 68 kbytes
    (b) 64 kbytes
    (c) 16 kbytes
    (d) 4 kbytes
    (e) None of the above

Answer: (e)

Remark: The size of a small page table = $2^{10} \times 4B = 4KB$
Since only one small page table in each level will reside in memory, the space occupied for the page table is $2 \times 4KB = 8KB$

12. Assume that an un-pipelined machine has 9ns clock cycles. The machine uses four cycles for ALU operations, four cycles for branches, and five cycles for memory operations. The relative frequencies of these operations are 30%, 20%, and 50%, respectively. Suppose that pipelining the machines adds Ins of overhead to the clock cycle time. Assume that the ideal CPI is one. Ignore any other impact. What is the average clock cycle per instruction (CPI) of the un-pipelined machine?
    (a) 5.5
    (b) 4
    (c) 3.5
    (d) 4.5
    (e) 5

Answer: (d)

13. (Continued from the previous question.) What is the speedup in the instruction execution rate gained from pipelining the machine?
    (a) 4.05
    (b) 4.75
    (c) 5.15
    (d) 5.85
    (e) 6.28

Answer: (a)

102中山資工

NOTE: If some questions are unclear or not well defined to you, you can make your own assumptions and state them clearly in the answer sheet.

1. You are the lead designer of a new processor. The processor design and compiler are complete, and now you must decide whether to produce the current design as it stands or spend additional time to improve it. Please consider and answer the following problems.
   (1) You discuss this problem with your hardware engineering team and arrive at the following options:
      (a) *Leave the design as it stands.* Call this base machine M_{base}. It has a clock rate of 500 MHz.
      (b) *Optimize the hardware.* The hardware team claims that it can improve the processor design to give it a clock rate of 600 MHz. Call this machine M_{opt}.

   The following measurements have been made using a simulator for M_{base} and M_{opt}.

Instruction class	Frequency	CPI	
		Mbase	*Mopt*
A	40%	2	2
B	25%	3	2
C	25%	3	3
D	10%	5	4

   What are the CPI and MIPS (million instructions per second) rate for M_{base} and M_{opt} Copy Table 1-1 to your answer sheet and fill in the blanks with your answers.

   (2) The compiler team proposes to improve the compiler for the machine to further enhance performance. Call this combination of the improved compiler and the base machine M_{comp}. The instruction improvements from this enhanced compiler have been estimated as follows.

Instruction class	Percentage of instruction executed vs. base machine
A	90%
B	90%
C	85%
D	95%

   For example, if the base machine executed 500 class A instructions, M_{comp} would execute $0.9 \times 500 = 450$ class A instructions for the same program. What is the CPI for M_{comp}? How much faster is M_{comp} than M_{base}? Copy Table 1-2 to your answer sheet and fill in the blanks with your answers.

Table 1-1:

Machine	CPI	MIPS rate
M_{base}		
M_{opt}		

Table 1-2:

Machine	CPI	Execution time *Mbase 1* Execution time *Mcomp*
M_{comp}		

Answer

(1)

Table 1-1:

Machine	CPI	MIPS rate
M_{base}	2.8	178.57
M_{opt}	2.35	255.32

CPI for $M_{base} = 0.4 \times 2 + 0.25 \times 3 + 0.25 \times 3 + 0.1 \times 5 = 2.8$

CPI for $M_{opt} = 0.4 \times 2 + 0.25 \times 2 + 0.25 \times 3 + 0.1 \times 2 = 2.35$

MIPS for $M_{base} = 500 / 2.8 = 178.57$

MIPS for $M_{opt} = 600 / 2.35 = 255.32$

(2)

Table 1-2:

Machine	CPI	Execution time $_{Mbase}$ /Execution time M_{comp}
M_{comp}	2.81	1.12

CPI for $M_{comp} = (0.4 \times 0.9 \times 2 + 0.25 \times 0.9 \times 3 + 0.25 \times 0.85 \times 3 + 0.1 \times 0.95 \times 5) / (0.4 \times 0.9 + 0.25 \times 0.9 + 0.25 \times 0.85 + 0.1 \times 0.95) = 2.5075 / 0.8925 = 2.81$

$$\frac{\text{Execution Time for } M_{base}}{\text{Execution Time for } M_{comp}} = \frac{(IC \times 2.8) / 500M}{(IC \times 0.8925 \times 2.81) / 500M} = 1.12$$

2. Here is a loop in **C:**

```
Loop:   g = g + A[i];
        i = i + j;
        if (i != h) goto Loop;
```

Assume that A is an array of 100 elements and the variables g, h, i, and j are in registers $s1, $s2, $s3, and $s4, respectively, In addition, assume that the base of the array A is in $s5. Other registers that can be used are $t0 and $t1. A possible MIPS assembly code corresponding to this C loop is showed as follows:

```
Loop:   add   $t1, $s3, OPA
        add   $t1, $t1, $t1
        add   $t1, $t1, OPB
        lw    $t0, 0(OPC)
        add   OPD, $s1, $t0
        add   $s3, $s3, $s4
        bne   $s3, OPE, Loop
```

Please determine proper values for the operands (OPA, OPB, OPC, OPD, OPE). Copy the Table 2 to your answer sheet and fill in the operand values.

Table 2:

Operand	OPA	OPB	OPC	OPD	OPE
Value					

Answer

Table 2:

Operand	OPA	OPB	OPC	OPD	OPE
Value	$s3	$s5	$t1	$s1	$s2

3. Figure 1 shows a 16-bit adder consisting of four 4-bit ALUs using carry lookahead. Assume that $gi = ai \cdot bi$ (generate) and p; $= ai + bi$ (propagate). Determine the value of P0, P1, P2, P3, G0, G1, G2, G3, and C4 (CarryOut) when adding two 16-bit numbers 0001101000110011 and 1110010111101011 with Carry In bit equal to 0. Please copy Table 3 to your answer sheet and fill in the blanks with your answers.

Table 3:

Signal	P0	P1	P2	P3	G0	G1	G2	G3	C4
Value									

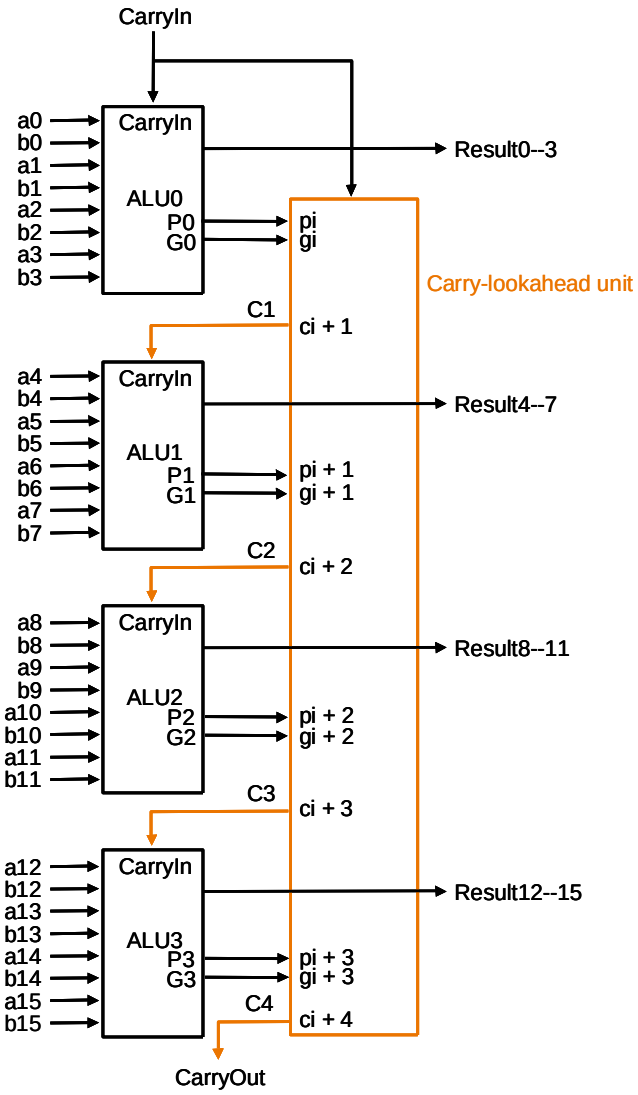


Figure 1

Answer

Table 3:

Signal	P0	P1	P2	P3	G0	G1	G2	G3	C4
Value	0	1	1	1	0	1	0	0	1

4. Compare the single-cycle implementation, in which all instructions take 1 clock cycle, with the five-stage (IF, ID, EX, MEM, WB) pipelined implementation using the following eight instructions: load word (lw), store word (sw), subtract (sub), and (and), or (or), set-less-than (slt) and branch-on-equal (beq). The operation times for the major functional units are 2 ns for memory access, 2 ns for ALU operation, and 1 ns for register file read or write.
   (1) Please complete the Table 4-1 by filling in the time for each component to calculate the total time of each instruction executed in the single-cycle implementation. Assume that the multiplexors, control unit, PC accesses, and sign extension unit have no delay. Please copy Table 4-1 to your answer sheet and fill in the blanks with your answers.
   (2) What is the clock cycle time for the single-cycle implementation and the pipelined implementation? Please copy Table 4-2 to your answer sheet and fill in the blanks with your answers.
   (3) Consider a program consisting of 100 lw instructions and in which each instruction is dependent upon the instruction before it. What would the actual CPI be if the program were run on single-cycle implementation and the pipelined implementation with a forwarding unit and a hazard detection unit? Please copy Table 4-2 to your answer sheet and fill in the blanks with your answers.

   Table 4-1:

Instruction class	Instruction fetch	Register read	ALU operation	Data access	Register write	Total time
Load word (lw)						
Store word (sw)						
R-format (add, sub, and, or, slt)	2ns	1 ns	2ns		1 ns	6 ns
Branch (beq)						

Table 4-1:

Design	Clock cycle time	Actual CPI
Single-cycle implementation		
Pipelined implementation		

Answer

(1)

Table 4-1:

Instruction class	Instruction fetch	Register read	ALU operation	Data access	Register write	Total time
Load word (lw)	2ns	1 ns	2ns	2ns	1 ns	8 ns
Store word (sw)	2ns	1 ns	2ns	2ns		7 ns
R-format (add, sub, and, or, slt)	2ns	1 ns	2ns		1 ns	6 ns
Branch (beq)	2ns	1 ns	2ns			5 ns

Table 4-2:

	(2)	(3)
Design	Clock cycle time	Actual CPI
Single-cycle implementation	8 ns	1
Pipelined implementation	2 ns	2

Remark: Actual CPI for pipeline = [(5 – 1) + 100 + 99] / 100 ≈ 2

5. Increasing associativity requires more comparators, as well as more tag bits per cache block. Assuming a cache of 4K blocks, a four-word block size, and a 32-bit address, find the total number of sets and the total number of tag bits for caches that are direct mapped, two-way and four-way set associative, and fully associative. Please copy Table 5 to your answer sheet and fill in

Table 5:

Cache	The total number of sets	The total number of tag bits
Direct mapped		
Two-way set associative		
Four-way set associative		
Fully associative		

Answer

Table 5:

Cache	The total number of sets	The total number of tag bits
Direct mapped	4K	(32 – 12 – 4) × 4K = 64K bits
Two-way set associative	4K / 2 = 2K	(32 – 11 – 4) × 4K = 68K bits
Four-way set associative	4K / 4 = 2K	(32 – 10 – 4) × 4K = 72K bits
Fully associative	1	(32 – 4) × 4K = 112K bits

6. Here is a series of address references given as word addresses: 1, 4, 8, 5, 20, 17, 19, 56, 11, 4, 43, 5, 6, 9, 17. Assuming a direct-mapped cache with four-word blocks and a total size of 16 words that is initially empty, label each reference in the list as a hit or a miss and show the final contents of the cache. Please copy Table 6-1 and Table 6-2 to your answer sheet and fill in your answers.

Table 6-1:

Ref.	1	4	8	5	20	17	19	56	9	11	4	43	5	6	9	17
Hit/Miss																

Table 6-2:

Block#	0	1	2	3
Starting Address				

Answer

Table 6-1:

Ref.	1	4	8	5	20	17	19	56	9	11	4	43	5	6	9	17
Hit/Miss	M	M	M	H	M	M	H	M	M	H	M	M	H	H	M	H

Table 6-2:

Block#	0	1	2	3
Starting Address	16	4	8	

Remark:

Word address	1	4	8	5	20	17	19	56	9	11	4	43	5	6	9	17
Block address	0	1	2	1	5	4	4	14	2	2	1	10	1	1	2	4
Tag	0	0	0	0	1	1	1	3	0	0	0	2	0	0	0	1
Index	0	1	2	1	1	0	0	2	2	2	1	2	1	1	2	0
Hit/Miss	M	M	M	H	M	M	H	M	M	H	M	M	H	H	M	H

Block#	0	1	2	3
Final content	16, 17 , 18, 19	4, 5 , 6 , 7	8, 9 ,10 ,11	

7. Suppose we have a processor with a base CPI of 1.0, assuming all references hit in the primary cache, and a clock rate of 1 GHz. Assume a main memory access time of 100 ns, including all the miss handling.
   (1) Suppose the miss rate per instruction at the primary cache is 5%. What is the effective CPI for the machine with one level of caching?
   (2) If we add a secondary cache that has a 10-ns access time for either a hit or a miss and is large enough to reduce the miss rate to main memory to 2%. What is the effective CPI for the machine with the two-level cache?

Answer

(1) Effective CPI for one level cache = $1 + 0.05 \times 100 = 6$

(2) Effective CPI for two level cache = $1 + 0.05 \times 10 + 0.02 \times 100 = 3.5$

8. Suppose you want to achieve a speedup of 50 with 100 processors. What fraction of the original computation can be sequential?

Answer

$\text{Speedup} = 50 = \frac{1}{\frac{f}{100} + (1-f)} \rightarrow f = 0.9898 \rightarrow 98.98\%$

102中興電機

1. To compare the performance of two different computers: M1 and M2. Table 1 shows the measurements on these computers.

Table 1

Program	Time on M1	Time on M2
Program 1	4 second	3 second
Program 2	10 second	20 second

Table 2 shows the additional measurements on these computers.

Table 2

Program	Instructions executed on M1	Instructions executed on M2
Program 1	5×10^9	6×10^9

(a) If the clock rates of computers M1 and M2 are 4GHz and 6GHz, respectively, find the clock cycles per instruction (CPI) for program 1 on both computers ?

(b) Assuming the CPI for program 2 on each computer is the same as the CPI for program 1, find the instruction counts for program 2 running on each computer using the execution time in Table 1 ?

Answer:

(a) $CPI_{M1} = (4 \times 4G) / (5 \times 10^9) = 3.2$; $CPI_{M2} = (3 \times 6G) / (6 \times 10^9) = 3$

(b) $IC_{M1} = (10 \times 4G) / 3.2 = 12.5 \times 10^9$; $IC_{M2} = (20 \times 6G) / 3 = 40 \times 10^9$

2. For a given instruction set architecture, please list three sources to increase in the CPU performance?

Answer:

(1) smart compiler
(2) advance computer organization
(3) advance VLSI technology

3. Convert the following MIPS codes to C codes. The parameter variable *n* corresponds to the argument register $a0. The compiled program starts with the label of the procedure and then saves two registers on the stack, the return address and $a0. The MIPS assembly codes are listed as follows:

```
test:
        addi  $sp, $sp, -8
        sw    $ra, 4($sp)
        sw    $a0, 0($sp)
        slti  $t0, $a0, 1
        beq   $t0, $zero, L1
```

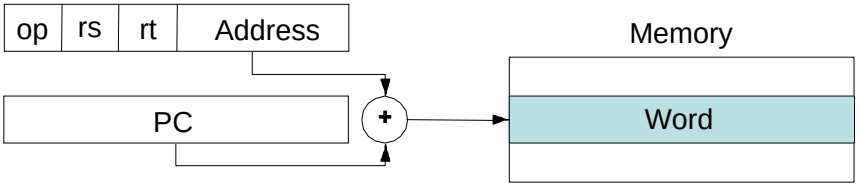
```
        addi  $v0, $zero, 1
        addi  $sp, $sp, 8
        jr    $ra
L1:     addi  $a0, $a0, -l
        jal   test
        lw    $a0, 0($sp)
        lw    $ra, 4($sp)
        addi  $sp, $sp, 8
        mul   $v0, $a0, $v0
        jr    $ra
```

Answer:

```
int test (int n)
{
    if (n < 1) return (1);
        else return (n*fact(n – 1));
 }
```

4. Please illustrate and describe the PC-relative addressing mode in MIPS machines.

Answer:



The Target address for a PC-relative instruction address is the offset parameter added to the address of the next instruction. This offset is usually signed to allow reference to code both before and after the instruction.

5. A program consisting of a sequence of ten instructions without branch or jump instructions is to be executed in an 8-stage pipelined RISC computer with a clock period of 0.5 ns. Determine (a) the latency time for the pipeline, (b) the maximum throughput for the pipeline, and (c) the time required for executing the program.

Answer:

(a) The latency time for the pipeline = 8×0.5 ns = 4ns
(b) The maximum throughput for the pipeline = 1 /0.5 ns = 2G instruction/second
(c) The time required for executing the program = $[(8 – 1) + 10] \times 0.5$ ns = 8.5 ns

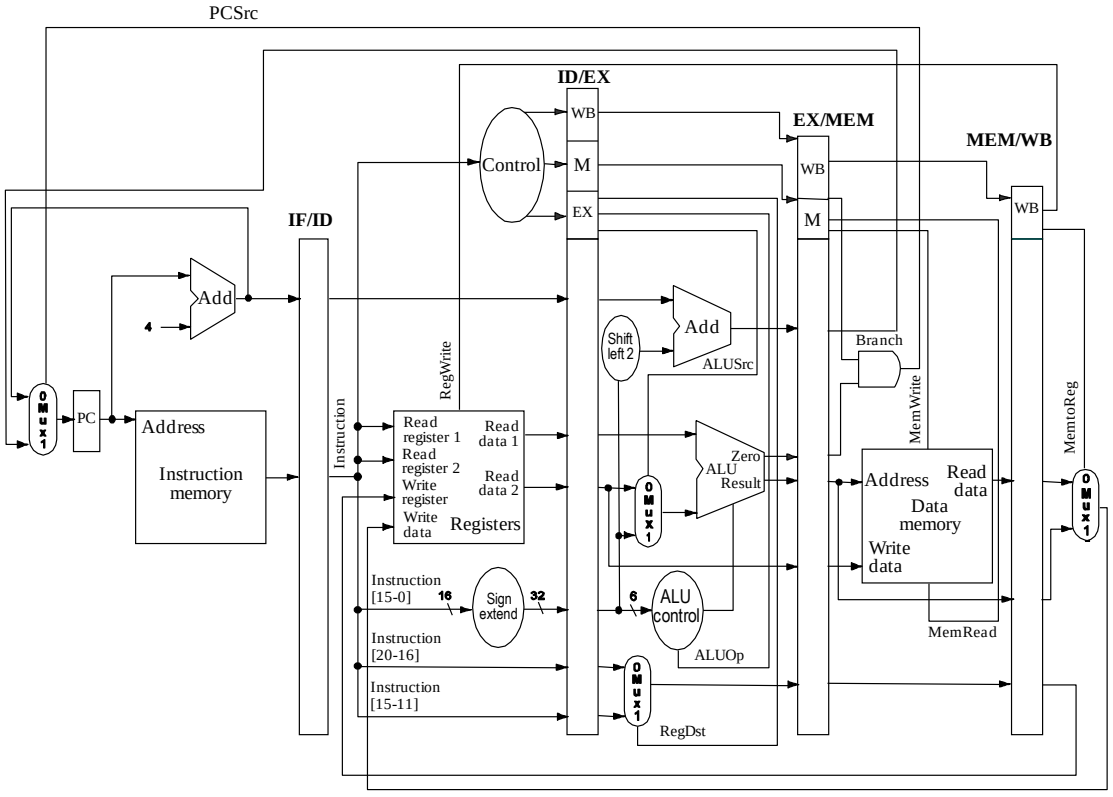
6. Consider executing the following code on the pipelined datapath.

```
add   $2, $3, $1
sub   $4, $3, $5
add   $5, $3, $7
sub   $3, $4, $5
add   $7, $6, $1
add   $8, S2, $6
```

(a) At the end of the fourth cycle of execution, which registers are being read and which register will be written?

(b) At the end of the sixth cycle of execution, which registers are being read and which register will be written?

(c) How many cycles does it take for the code to execute?



Answer:

(a) At the end of the fourth cycle of execution, registers \$3 and \$7 are being read.

(b) At the end of the sixth cycle of execution, register \$4 is being written and registers \$6 and \$1 are being read.

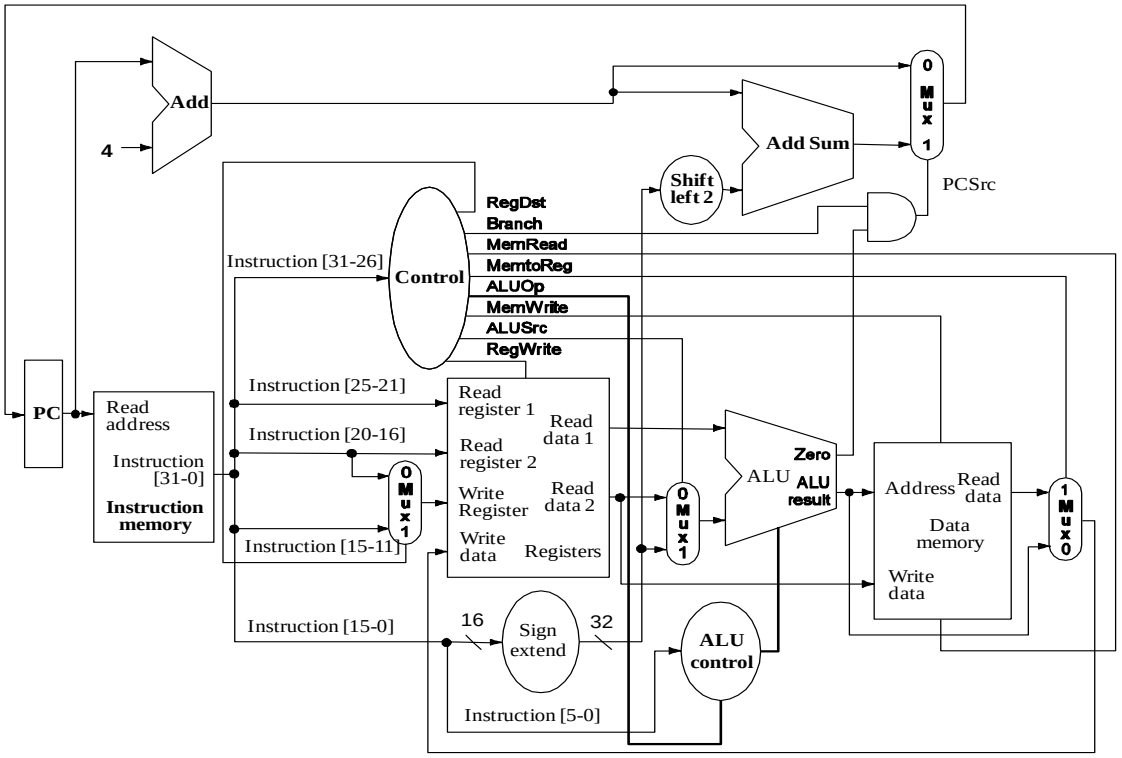
(c) Execute this program needs (5 – 1) + 6 = 10 clock cycles

Remark(c): there are data hazards in the given code but no forwarding scheme is used in the datapath. Hence, the code cannot execute correctly.

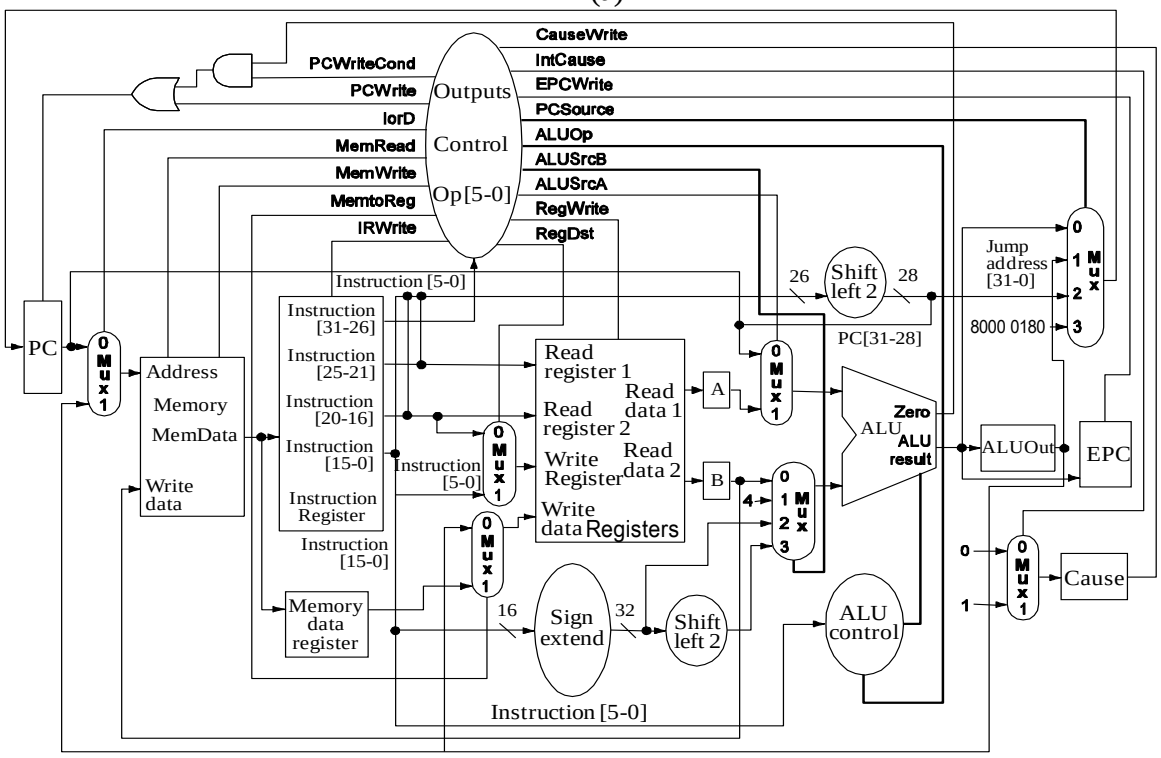
7. Considering the following datapath shown below, draw paths (use color pens if possible) to show the flow of both data and the program counter for the following instructions.

(a) add $t0, $t0, $t1 # single cycle datapath

(b) lw $a0, 0($t1) # multicycle datapath



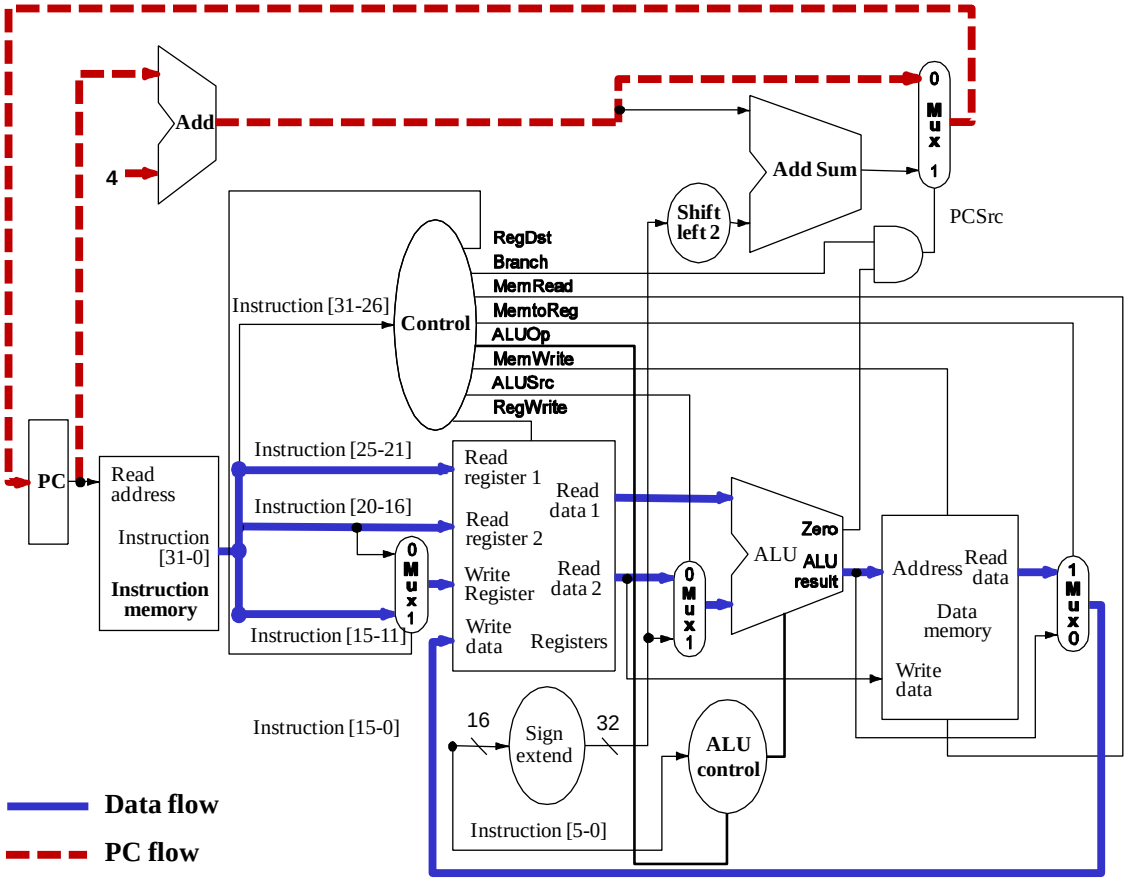
(a)



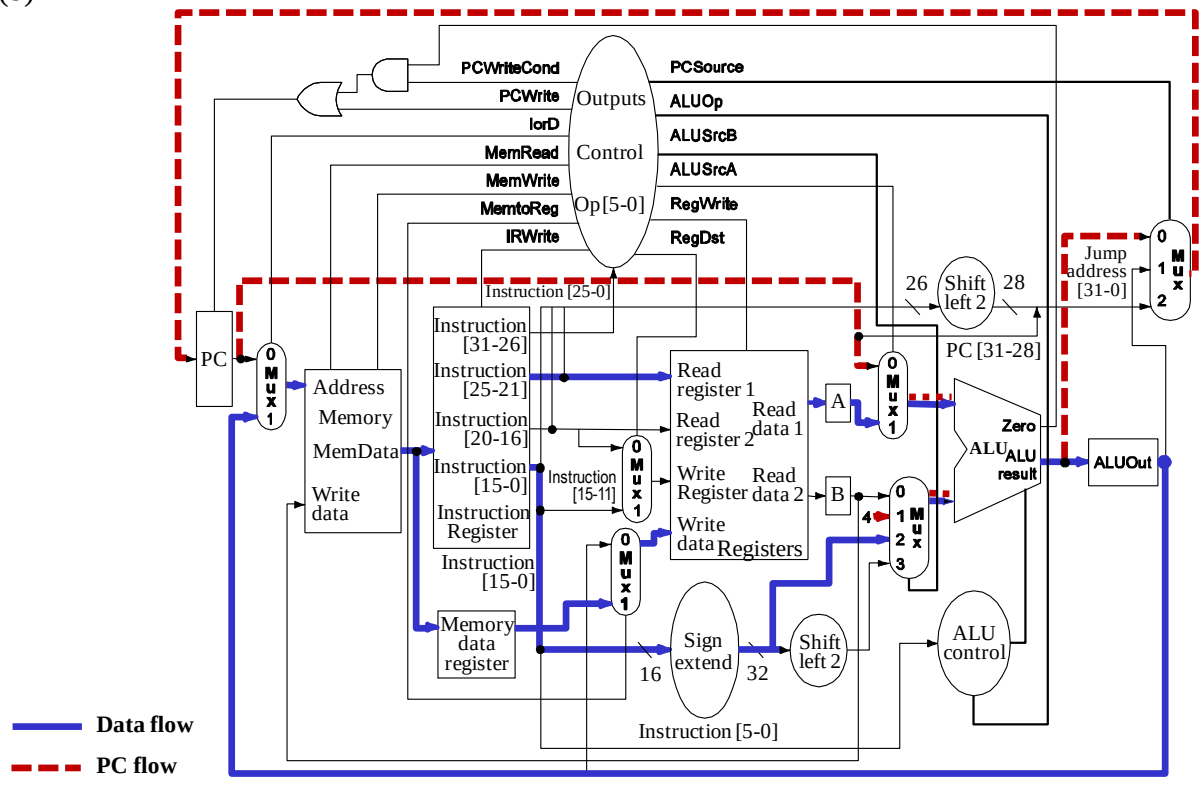
(b)

Answer:

(a)


Add
4
Add Sum
0 M u x 1
PCSrc
Shift left 2
RegDst
Branch
MemRead
MemtoReg
ALUOp
MemWrite
ALUSrc
RegWrite
Control
Instruction [31-26]
Instruction [25-21]
Instruction [20-16]
Instruction [15-11]
Instruction [15-0]
Instruction [5-0]
PC
Read address
Instruction [31-0]
Instruction memory
Read register 1
Read register 2
Write Register
Write data
Read data 1
Read data 2
Registers
ALU
Zero
ALU result
Address
Read data
Data memory
Write data
16
Sign extend
32
ALU control
1 M u x 0
Data flow
PC flow

(b)


PCWriteCond
PCWrite
IorD
MemRead
MemWrite
MemtoReg
IRWrite
Outputs Control Op[5-0]
PCSource
ALUOp
ALUSrcB
ALUSrcA
RegWrite
RegDst
Instruction [25-0]
26
Shift left 2
28
PC [31-28]
Jump address [31-0]
0 1 M u x 2
PC
0 M u x 1
Address
Memory
MemData
Write data
Instruction [31-26]
Instruction [25-21]
Instruction [20-16]
Instruction [15-0]
Instruction Register
Instruction [15-11]
Instruction [15-0]
Memory data register
Read register 1
Read register 2
Write Register
Write data
Read data 1
Read data 2
Registers
A
B
4
0 1 M u x 3
Zero
ALU
ALU result
ALUOut
Sign extend
16
32
Shift left 2
ALU control
Instruction [5-0]
Data flow
PC flow

102中興資工

1. (Single choice) Which statement is wrong?
   (a) Pipeline improves throughput but cannot improve instruction latency.
   (b) Although a left shit instruction can replace an integer multiply by a power of 2, a right shift cannot be performed as an integer division by a power of 2.
   (c) Forwarding resolves data hazard by adding extra hardware to retrieve the missing item early from the internal resources.
   (d) Capacity misses is caused by the first access to a block that has never been in the cache.

Answer: (d)

2. A compiler designer is trying to decide between two code sequences for a particular computer. The hardware designers have supplied the following facts:

	CPI for each instruction class		
	A	B	C
CPI	1	2	3

For a particular high-level language statement, the compiler writer is considering two code sequences that require the following instruction counts:

Code sequence	Instruction counts for each instruction class		
	A	B	C
1	2	1	2
2	4	1	1

   (a) Which code sequence will be faster? Why?
   (b) What is the CPI for each sequence?

Answer:

(a) Total clock cycle for code sequence 1 = $1 \times 2 + 2 \times 1 + 3 \times 2 = 10$
Total clock cycle for code sequence 2 = $1 \times 4 + 2 \times 1 + 3 \times 1 = 9$
Code sequence 1 is faster

(b) CPI for code sequence 1 = 10 / 5 = 2
CPI for code sequence 2 = 9 / 6 = 1.5

3. Convert 16-bit binary version of 2_{10} and -2_{10} to 32-bit binary number.
   $1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1100_2$

Answer:

2_{10} = 0000 0000 0000 0000 0000 0000 0000 0010_2
-2_{10} = 1111 1111 1111 1111 1111 1111 1111 1110_2

Remark: 題目給的二進位數為4

4. In ARM, suppose register r0 has the binary number

$1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111_2$

and that register r1 has the binary number

$0000\ 0000\ 0000\ 0000\ 0000\ 0000\ 0000\ 0001_2$

and the following instruction is executed.

```
CMP   r0, r1
```

Which conditional branch is taken? Please explain your answer.

```
BLO   L1 : unsigned branch (branch on lower)
BLT   L2 : signed branch (branch on less than)
```

Answer:

Instruction BLO is not taken to L1

Instruction BLT is taken to L2

5. In ARM, the branch instruction is encoded using 24-bit address. You might think that the 24-bit address would extend the program limit to 2^{24} or 16MB, which would be fine for many programs but constrain some large ones. How to solve this problem? Hint: specify a register that would always be added to the branch address.

Answer:

We can specify the program counter register (PC) that would always be added to the branch address, so that a branch instruction would calculate the following:

Branch target address = PC + branch address

This sum allows the program to be as large as 2^{32}, solving the branch address size problem.

6. There are situations in pipelining when the next instruction cannot execute in the following clock cycle. These events are called *hazards*. In the following situation, please point out what kind of hazards is occurred.
Suppose you found a sock (短襪) at the folding station for which no match existed. One possible strategy is to run down to your room and search through your clothes bureau (衣櫃) to see if you can find the match. Obviously, while you are doing the search, loads that have completed drying and are ready to fold and those that have finished washing and are ready to dry must wait.

Answer: Data hazard is occurred.

7. Assume a five stage pipelined MIPS processor (i.e., instruction fetch, register read, ALU operation, data access, register write) and the following two sequences of instructions.

	Instruction sequence

a	lw $1, 40($6)
	add $6, $2, $2
	sw $6, 50($1)

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

(a) Assume there is no forwarding in this pipelined processor. Add *nop* instructions in above two sequences of instructions to eliminate the hazards.

(b) Assume there is full forwarding. Add *nop* instructions in above two sequences of instructions to eliminate the hazards.

Answer:

(a)

	Instruction sequence	

a.	lw $1, 40($6)
	add $6, $2, $2
	nop
	nop
	sw $6, 50($1)
		Delay I3 to avoid RAW hazard on $1 and $6 from I1 and I2

b.	lw $5, -16($5)
	nop
	nop
	sw $5, -16($5)
	add $5, $5, $5
		Delay I2 to avoid RAW hazard on $5 from I1.

(b)

	Instruction sequence	

a.	lw $1, 40($6)
	add $6, $2, $2
	sw $6, 50($1)
		No RAW hazard on $1 from I1 (forwarded)

b.	lw $5, -16($5)
	nop
	sw $5, -16($5)
	add $5, $5, $5
		Delay I2 to avoid RAW hazard on $5 from I1

102 台科大電子

1. What is the power wall? How does it affect CPU development?

Answer:

Power Wall is the increasing power consumption and resulting heat generation of the processing unit. The power wall can be mitigated by "shrinking" the processor by using smaller traces for the same logic. This poses manufacturing, system design, and deployment problems.

The power wall has forced a dramatic change in the design of microprocessor. Rather than continuing to decrease the response time of a single program running on the single processor, as currently all desktop and server companies are shipping microprocessors with multiprocessors per chip, where benefit is often more on throughput than on response time.

2. Please explain the following terms with examples:
   (a) Structure hazard
   (b) Data hazard
   (c) Speculation

Answer:

(a) Structural hazards: hardware cannot support the instructions executing in the same clock cycle (limited resources)

(b) Data hazards: attempt to use item before it is ready (Data dependency)

Example:

```
1       lw   $5,   50($2)
2       add  $2,   $5,   $4
3       add  $4,   $2,   $5
4       beq  $8,   $9,   L1
5       sub  $16, $17, $18
6       sw   $5,   100($2)
7   L1:
```

Type	Example
Structure hazard	假設上列指令是在只有單一記憶體的 datapath 中執行，則在第 4 個時脈週期，指令 1 讀取記憶體資料同時指令 4 也在從同一個記憶體中擷取指令，也就是兩個指令同時對一記憶體進行存取。在這樣情形下就會發生 structural hazard
Data hazard	指令 1 在第 5 個時脈週期時才會將計算結果寫回暫存器$5，但指令 2 和 3 分別在時脈週期 3 跟 4 時便需要使用暫存器$5 的內容。此時指令 2 和 3 擷取到暫存器$5 的內容並非最新的，因此在這樣情形下就會發生 data hazard

(c) Speculation: an approach whereby the compiler or processor guesses the outcome of an instruction to remove it as a dependence in executing other instructions.

For example, the compiler can use speculation to move the *lw* instruction across *sw*, if we guess the memory addresses for the two instructions are different.

```
sw   $1, 100($2)
lw   $3, 100($4)
```

3. Suppose we have a 32-bit virtual address, 4KB pages, and 4 bytes per page table entry. What is the total page table size?

Answer

Number of entries in page table = 4GB / 4KB = 1M

The page table size = 4B × 1M = 4MB

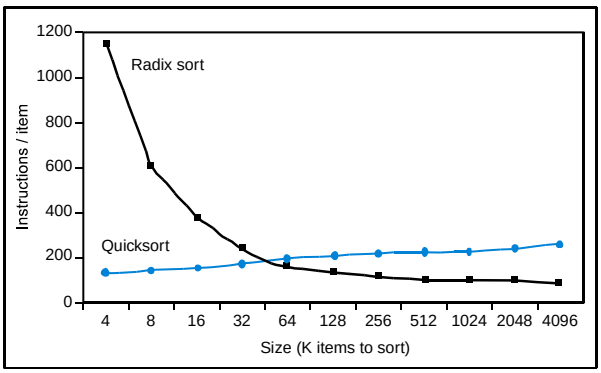
4. Please use CPU execution time formula to explain why superscalar and super pipeline could improve performance.

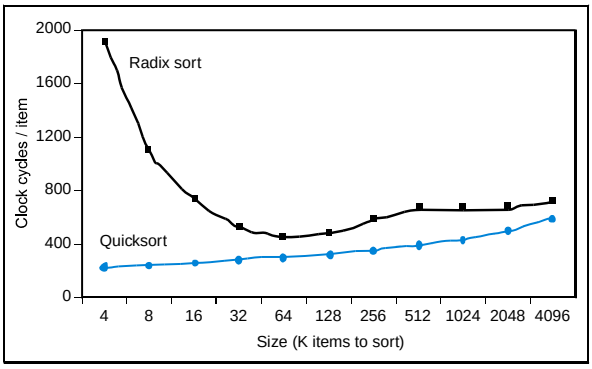
Answer: CPU execution time = Instruction count × CPI × clock cycle time

Superscalar enables the processor to execute more than one instruction per clock cycle. This will increase IPC (or reduce the average CPI) and thus can improve performance.

Superpipeline increases the depth of the pipeline (more pipeline stage) to overlap more instructions. Performance is potentially greater since the clock cycle time can be shorter.

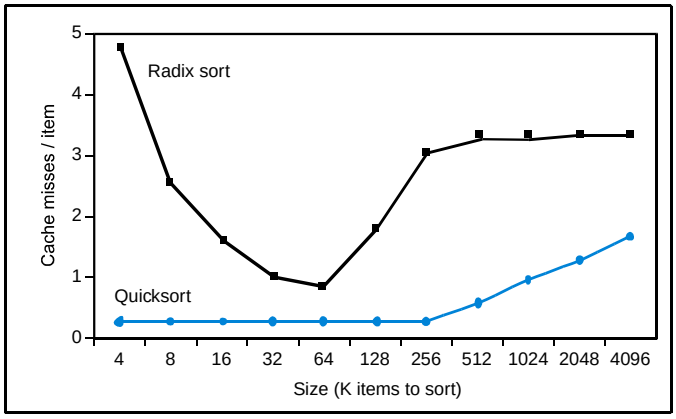
5. Please explain why the *clock cycle/item* in the right-hand side is not consistent with *instruction/item* in the left-hand side?





Answer: Quicksort consistently has many fewer misses per item to be sorted.

Remark:


Cache misses / item
5
4
3
2
1
0
Radix sort
Quicksort
4
8
16
32
64
128
256
512
1024
2048
4096
Size (K items to sort)

102 台科大資工

1. A compiler designer is trying to decide between two code sequences for a particular computer. The hardware designers have supplied the following facts:

	CPI for instruction class		
	A	B	C
CPI	3	4	2

For a particular high-level-language statement, the compiler writer is considering two code sequences that require the following instruction counts:

Code sequence	Instruction counts for instruction class		
	A	B	C
1	2	2	6
2	2	5	1

(a) Which code sequence executes the most instructions?

(b) Which will be faster?

(c) What is the CPI for each sequence?

Suppose we measure the code for the same program from two different compilers and obtain the following data:

Code form	Instruction counts (in billions) for instruction class		
	A	B	C
Complier 1	1	2	7
Compiler 2	2	1	5

Assume that the computer's clock rate is 5 GHz.

(d) Which code sequence will execute faster according to execution time?

(e) Which code sequence will execute faster according to MIPS?

Answer:

(a) Instruction count for code sequence 1 = 2 + 2 + 6 = 10

Instruction count for code sequence 2 = 2 + 5 + 1 = 8

Hence, code sequence 1 executes the most instructions

(b) Clock cycles for code sequence 1 = $2 \times 3 + 2 \times 4 + 6 \times 2 = 26$

Clock cycles for code sequence 2 = $2 \times 3 + 5 \times 4 + 1 \times 2 = 28$

Hence, code sequence 1 is faster than code sequence 1

(c) CPI for code sequence 1 = 26 / 10 = 2.6

CPI for code sequence 2 = 28 / 8 = 3.5

(d) Execution Time for compiler 1 = $\frac{(1\times 3+2\times 4+7\times 2)\times 10^9}{5\times 10^9} = 5$ s

Execution Time for compiler 2 $= \frac{(2\times 3+1\times 4+5\times 2)\times 10^9}{5\times 10^9} = 4$ s

According to execution time, code sequence from compiler 2 is faster.

(e) MIPS for compiler 1 $= \frac{(1+2+7)\times 10^9}{5\times 10^6} = 2000$

MIPS for compiler 2 $= \frac{(2+1+5)\times 10^9}{4\times 10^6} = 2000$

According to MIPS, code sequence from compiler 1 is as fast as from compiler 2.

2. Consider three different cache configurations below:
Cache 1: direct-mapped with four-word blocks.
Cache 2: two-way set associative with two-word blocks and LRU replacement.
Cache 3: fully associative with four-word blocks and LRU replacement.
Assuming that each cache has total data size of 16 32-bit words and all of them are initially empty. 20-bit word address is used. Consider the following sequence of address references given as word addresses: 20, 23, 33, 34, 28, 35, 45, 47, 56, 57 and 20. For caches 1, 2, and 3, please label each reference in the list as a hit or a miss.

Answer:

Cache1: number of blocks = 4

Word addr.	Block addr.	Tag	Index	Hit/Miss
20	5	1	1	Miss
23	5	1	1	Hit
33	8	2	0	Miss
34	8	2	0	Hit
28	7	1	3	Miss
35	8	2	0	Hit
45	11	2	3	Miss
47	11	2	3	Hit
56	14	3	2	Miss
57	14	3	2	Hit
20	5	1	1	Hit

Cache2: number of sets = 4

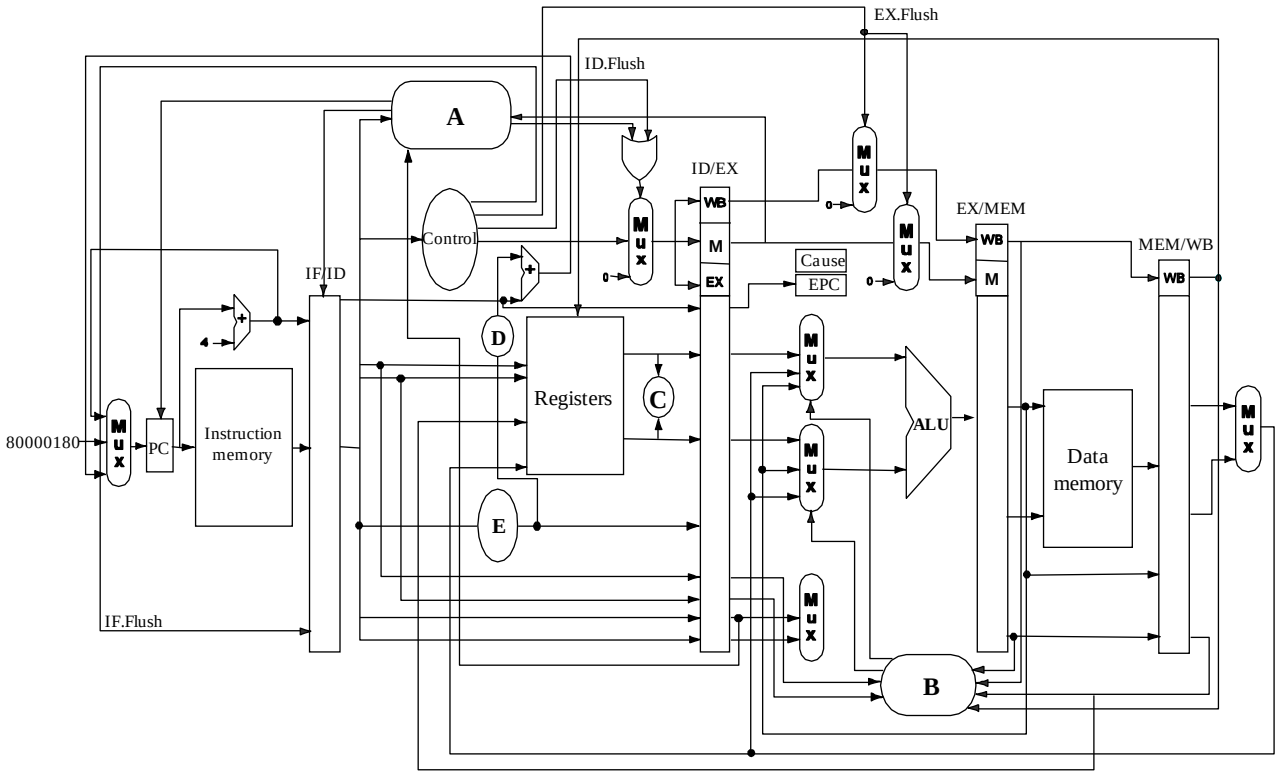
Word addr.	Block addr.	Tag	Index	Hit/Miss
20	10	2	2	Miss
23	11	2	3	Miss
33	16	4	0	Miss
34	17	4	1	Miss
28	14	3	2	Miss

35	17	4	1	Hit
45	22	5	2	Miss
47	23	5	3	Miss
56	28	7	0	Miss
57	28	7	0	Hit
20	10	2	2	Miss

Cache3: number of blocks = 4

Word addr.	Block addr.	Tag	Hit/Miss
20	5	5	Miss
23	5	5	Hit
33	8	8	Miss
34	8	8	Hit
28	7	7	Miss
35	8	8	Hit
45	11	11	Miss
47	11	11	Hit
56	14	14	Miss
57	14	14	Hit
20	5	5	Miss

3. The following figure shows datapath of a pipelining processor.



(a) Please make the right binding between each of units "A", "B", "C", "D", "E" and one of following names.

(1) AND array (2) Hazard detection unit (3) Forwarding unit
(4) XOR array (5) Shift left 2 unit (6) Sign extend unit
(7) MUX

(b) For the following instruction sequence and assuming that branch (B) is taken. Please describe the instruction pair that causes data hazard, the corresponding depend register, and the unit required to solve the data hazard.

```
(S)          sub   $5, $3, $4     # Reg5 = Reg3 - Reg4
(A)          add   $1, $4, $5     # Reg1 = Reg4 + Reg5
(D)          and   $3, $2, $5     # Reg3 = (Reg2 OR Reg5)
(L)          lw    $7, 0($1)      # Load from MEM (compute base on $1) to $7
(O)          ori   $9, $3, $7     # Reg9 = (Reg3 OR Reg7)
(B)          beq   $7, $9, loop   # Branch to loop if Reg7 = Reg9
(R)          sra   $3, $9, 2      # Reg3 = Reg9 >> 2
(T)          slt   $9, $3, S7     # Reg9 = 1 if Reg3 < Reg7
(I)  loop:   addi  $9, $7, 40     # Reg9 = Reg7 + 40
```

(c) Continue with (b). How many cycles does it take to execute the above instruction sequence?

Answer:

(a)

A	B	C	D	E
(2) Hazard detection unit	(3) Forwarding unit	(4) XOR array	(5) Shift left 2 unit	(6) Sign extend unit

(b)

Instruction pair that cause data hazard	Depend register	Function units required
(S & A)	$5	B
(S & D)	$5	B
(A & L)	$1	B
(D & O)	$3	B
(L & O)	$7	A, B
(L & B)	$7	A
(O & B)	$9	A
(O & R)	$9	B
(R & T)	$3	B

Remark: Since the datapath cannot forward any data to ID stage, forwarding unit (B) is not needed for (O & B)

(c) Instructions between (L & O) and (O & B) required one clock delay for data hazard, respectively.

Branch is taken imply the instruction (R) is fetched into the pipeline and will be flushed and also cause one clock delay. Hence, the total clock cycles required to execute the code is (5 – 1) + 7 + 3 = 14 clocks

102 台師大資工

1. Consider a program written in high level programming language is executed separately on two processors with different instruction set architectures. Processor A executes 10 billion instructions with clock rate 4GHz and CPI 1.0. Processor B executes 8 billion instructions with the same clock rate but CPI 1.1.
   (a) Find the MIPS rates of those processors.
   (b) Find the total execution time for each of the processors.
   (c) Compare their performance.

Answer

(a) $MIPS_A = \frac{4G}{1.0\times10^6} = 4000$, $MIPS_B = \frac{4G}{1.1\times10^6} = 3640$

(b) Execution Time for processor A = $\frac{10\times10^9\times1}{4\times10^9} = 2.5$

Execution for processor B = $\frac{8\times10^9\times1.1}{4\times10^9} = 2.2$

(c) Processor B is 2.5/2.2 = 1.136 times faster than Processor A

2. Assume a decimal value -1234 is stored in a 32-bit general purpose register. What is the value (in binary or hexadecimal format) stored in the register if it is represented as
   (a) a signed integer, and
   (b) a single-precision floating-point number in the IEEE 754 standard.

Answer: 1234_{10} = 0000 0000 0000 0000 0000 0100 1101 0010_2

(a) -1234_{10} = 1111 1111 1111 1111 1111 1011 0010 1110_2 = $FFFFFB2C_{16}$

(b) -0000 0000 0000 0000 0000 0100 1101 0010_2 = $-1.0011010010 \times 2^{10}$ →

	S	E	F
binary	1	01110101	00110100100000000000000
hexadecimal	BA9A4000		

3. Consider a computer datapath divides the execution of an instruction into five stages, i.e. instruction fetch, instruction decode, execution, memory access, and write back. The processing times of these stages are 400ps (10^{-12} seconds), 200ps, 120ps, 350ps, and 200ps, respectively,
   (a) What is the highest operating frequency if it is a single cycle (not pipelined) datapath?
   (b) A pipeline register adds 20ps delay. What is the highest operating frequency if it is a five-stage pipeline datapath?
   (c) Assume an infinite code sequence is executed, and no hazard is found between instructions. Compare the performances of the single cycle datapath and the pipelined datapath.

Answer

(a) 1 / (400 ps + 200 ps + 120 ps + 350 ps + 200 ps) = 787.4MHz

(b) 1 / 420 ps = 2.38 GHz

(c) Pipeline datapath is (400 + 200 + 120 + 350 + 200) / 420 = 3.02 times faster than single cycle datapath

4. Consider the code sequence executing in a five-stage (IF, DE, EX, MEM, and WB) pipelined datapath.

```
        lw    $1, 0($6)      # $1 ← MEM[$6]
        lw    $2, 1($6)      # $2 ← MEM [$6 + 1]
        add   $3, $1, $2     # $3 ← $1 + $2
        sub   $4, $3, $5     # $4 ← $3 + $5
        beq   $3, $0, L1     # jump to L1 if $3 == $0
        ….
    L1:
```

(a) Show all the possible hazards between instructions.

(b) Describe how the hazards can be resolved in recent computer architectures.

Answer

(a)

Data hazard	
Register	Instruction pair
$1	(1, 3)
$2	(2, 3)
$3	(3, 4), (3, 5)

(b) Data hazard can be resolved by forwarding technique in recent computer architectures. However, forwarding technique cannot save the day is when an instruction tries to read a register following a load instruction that writes the same register. Hence, the pipeline need a stall between the load and its use.

5. Consider a 128 bytes fully associative mapped cache with 4-word blocks (a word has 32 bits). Its replacement policy is LRU. Suppose the following memory word addresses (in decimal) are accessed in sequence:
40, 41, 42, 400, 43, 400, 60, 61, 62, 63, 64, 800, 40, 41, 42, 800, 43, 44, 60, 61, 62, 120, 121, 122, 123, 168, 169, 41, 42, and 400.

(a) Show whether it is a hit or a miss in the cache for each of the memory accesses.

(b) What is the miss rate?

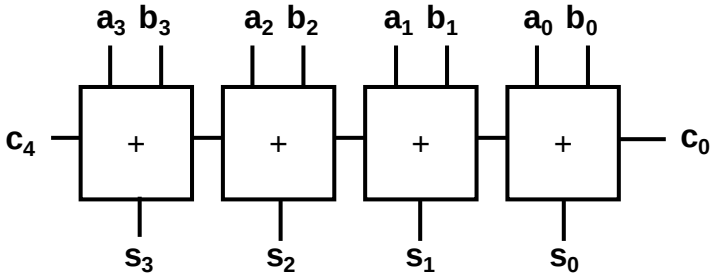
Answer: There $\frac{128}{4\times4}$ = 8 blocks in the cache.

		(a)
Word address	Block address	Hit/Miss
40	10	Miss
41	10	Hit
42	10	Hit
400	100	Miss
43	10	Hit
400	100	Hit
60	15	Miss
61	15	Hit
62	15	Hit
63	15	Hit
64	16	Miss
800	200	Miss
40	10	Hit
41	10	Hit
42	10	Hit
800	200	Hit
43	10	Hit
44	11	Miss
60	15	Hit
61	15	Hit
62	15	Hit
120	30	Miss
121	30	Hit
122	30	Hit
123	30	Hit
168	42	Miss
169	42	Hit
41	10	Miss
42	10	Hit
400	100	Miss
(b)	Miss rate	10/30 = 1/3

102海洋電機

1. Draw a schematic diagram for a 4-bit adder, using 1-bit full-adder as a basic unit

Answer



2. Explain sequential logic and combinational logic. Give one example for each class.

Answer

Sequential logic: sequential logic is a type of logic circuit whose output depends not only on the present value of its input signals but on the past history of its inputs. That is, sequential logic has state (memory).

A common example of a circuit employing sequential logic is the flip-flop.

Combinational logic: The elements that operate on data values are all combinational, which means that their outputs dependent only on the current inputs. Given the same input, a combinational element always produces the same output.

The 4-bit ripple carry adder is an example of a combinational; element.

3. Mark each of the following statements as either True or False.
   (1) Pipelining always increases throughput and reduces individual instruction execution time.
   (2) A write-through cache will not have the same miss rate as a write-back cache.
   (3) The Physical Address Space must be smaller than the Virtual Address Space.
   (4) A fully-associative cache must require more logic circuits than a direct-mapped cache of the same size.
   (5) Set-associative caches not necessarily always have higher hit rates than direct-mapped caches of the same size.

Answer

(1)	(2)	(3)	(4)	(5)
False	False	False	True	True

Remark (5): if the size of cache is large, set-associative and direct-mapped cache both may have the same hit rate.

4. Describe briefly the following allocation algorithms:
   (1) First fit
   (2) Best fit

Answer

(1) First Fit - A resource allocation scheme (usually for memory). First Fit fits data into memory by scanning from the beginning of available memory to the end, until the first free space which is at least big enough to accept the data is found. This space is then allocated to the data.
(2) A resource allocation scheme (usually for memory). Best Fit tries to determine the best place to put the new data. One example might be to try and minimize the wasted space at the end of the block being allocated - i.e. use the smallest space which is big enough.

Remark: This topic belongs to OS

5. Explain the four conditions that must hold simultaneously in a system for causing a deadlock situation.

Answer

(1) Mutual Exclusion Condition: At least one resource must be held in a non-shareable mode, that is, only one process at a time claims exclusive control of the resource. If another process requests that resource, the requesting process must be delayed until the resource has been released.
(2) Hold and Wait Condition: There must exist a process that is holding a resource already allocated to it while waiting for additional resource that are currently being held by other processes.
(3) No-Preemptive Condition: Resources cannot be removed from the processes are used to completion or released voluntarily by the process holding it.
(4) Circular Wait Condition: The processes in the system form a circular list or chain where each process in the list is waiting for a resource held by the next process in the list.

Remark: This topic belongs to OS

6. Consider a 4K × 8 RAM chip (i.e. 4096 bytes) .
   (a) How many address lines and data lines are there in this chip ?
   (b) How many such chips do you need to construct a 64K × 32 memory ?
   (c) How many address lines and data lines are there in the 64K × 32 memory?
   (d) What kind of decoder do you need to connect this 64K × 32 memory? Draw the schematic diagram for this 64K × 32 memory construction

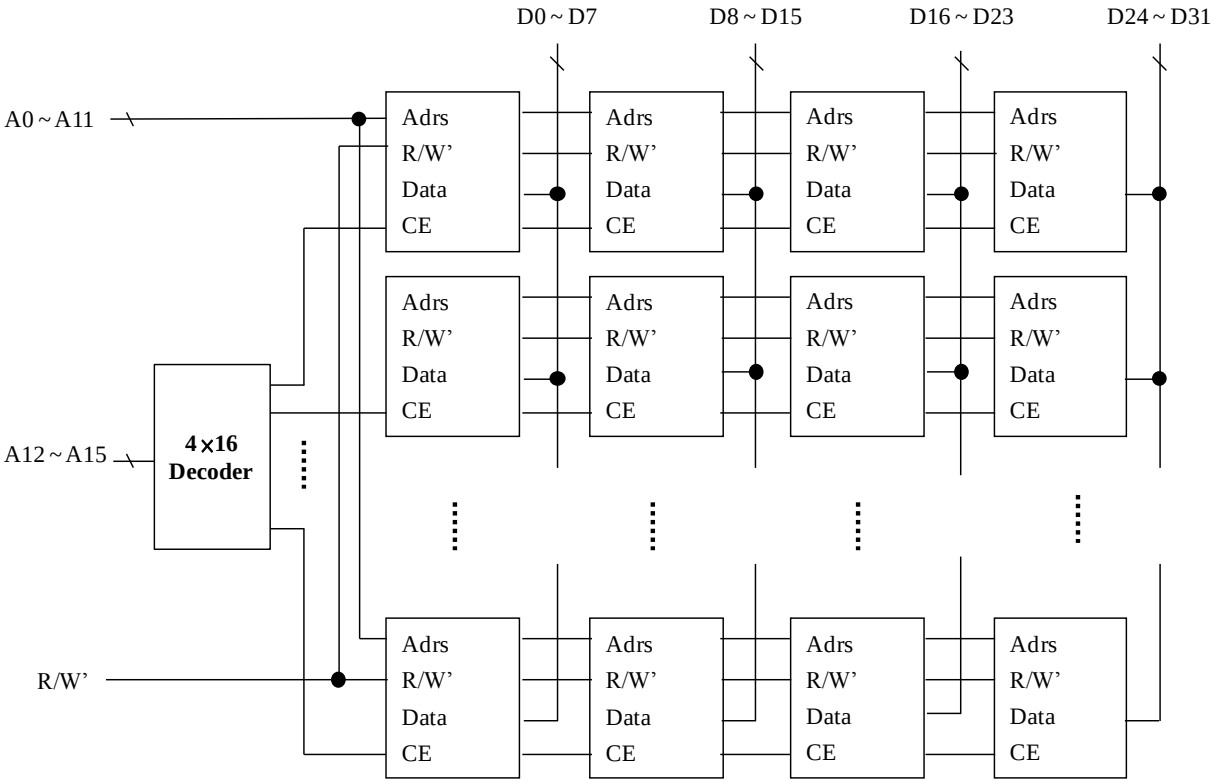
Answer

(a) address lines: $\log_2 4K = 12$, data line: 8
(b) $(64K/4K) \times (32/8) = 64$

(c) address lines: $\log_2 64K = 16$, data line: 32

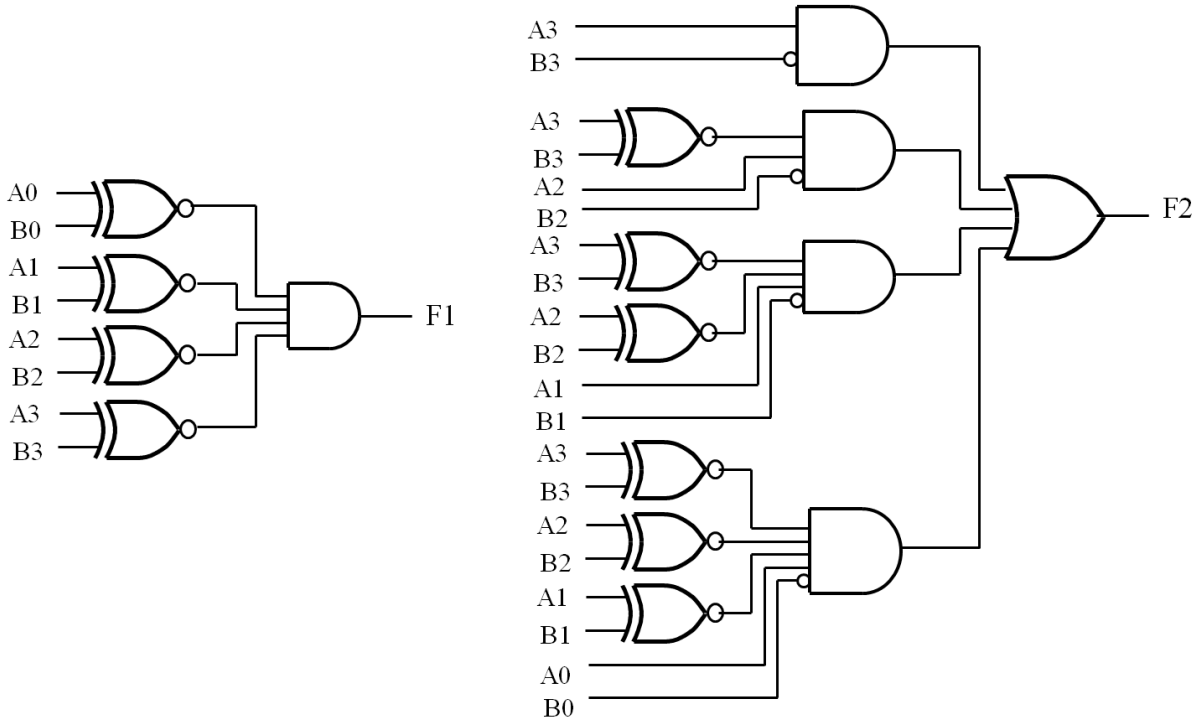
(d) 4×16 decoder is needed.

The following figure shows the $64K \times 32$ memory schematic



102海大資工

1. 問答與計算：
   (a) 說明南橋(South bridge)電路及北橋(North bridge)電路之功能。
   (b) 說明NOR flash memory與Nand flash memory的差異。
   (c) 說明Cache中，Write miss如何發生，如何處理。
   (d) 資料有500筆，若CPU的Cache hit rat為98%，檢查Cache的TLB要2ns，對Cache讀寫要10ns，查主記憶體的TLB要4ns，而實際位址的計算讀取要55ns，若讀取此500筆資料，共需多少時間？
   (e) 管線(Pipeline)的指令切割成6個步驟，其執行時間分別為7ns、8ns、4ns、4ns、6ns、5ns，此管線知增速(Speedup)為何？
   (f) 下圖F1及F2的功能為何？



Answer

(a) The *south bridge* handles all of a computer's I/O functions, such as USB, the system BIOS, and the interrupt controller.
The *north bridge* typically handles communications among the CPU, in some cases RAM, and PCI Express (or AGP) video cards, and the south bridge.

(b)

	NOR flash	Nand flash
名稱由來	Bit cell like a NOR gate	Bit cell like a NAND gate
存取方式	Random read/write access	Block-at-a-time access
價格	More expensive	Cheaper
用途	Instruction Mem. in embedded systems	USB keys, media storage, ...

(c) A cache write miss refers to a failed attempt to write a piece of data in the cache, which results in a main memory access with much longer latency. There are two ways to handle write miss.

Write allocate: the block is fetched from memory and then the appropriate portion of the block is overwritten in the cache.

No write allocate: Update the portion of the block in memory but not put it in the cache.

(d) $500 \times 0.98 \times (10 + 2) + 500 \times 0.02 \times (4 + 55 + 2 + 10) = 6590$ns

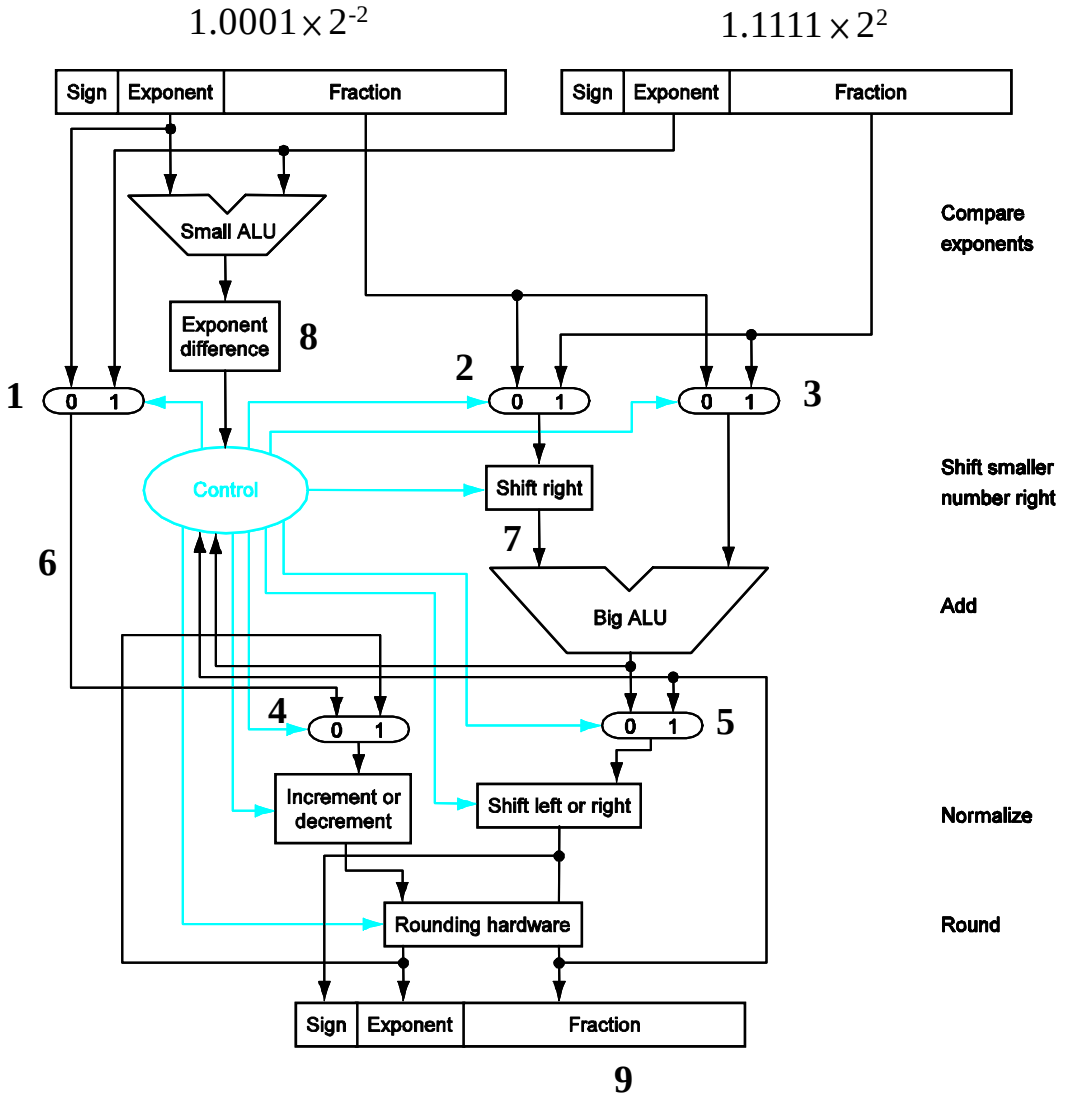
(e) Speedup = $(7 + 8 + 4 + 4 + 6 + 5) / 8 = 4.25$

(f) **Function of F1:** if (A = B) then F1 = 1 else F1 = 0;
Function of F2: if (A > B) then F2 = 1 else F2 = 0

2. 單精準度浮點數(IEEE 754)加法電路，參考下圖，執行$1.1101\times2^{-2} + 1.1111\times2^{2}$

(a) 下圖中，編號6, 7, 8, 9所指處其值為何？

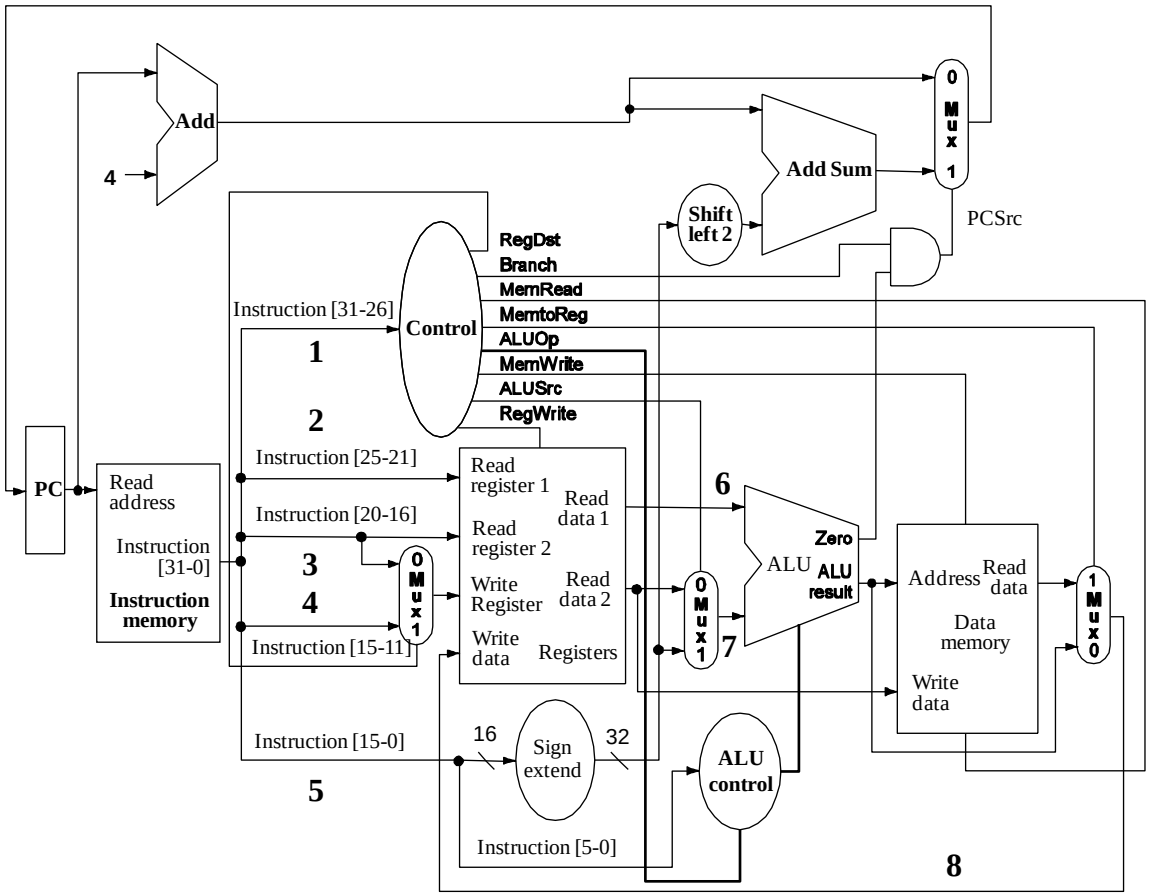
(b) 說明標示1, 2, 3, 4, 5的多工器功能為何？



Answer

(a)	No. 6	No. 7	No. 8	No. 9	
	2	0.00010001	4	000000001	
(b)	No. 1	No. 2	No. 3	No. 4	No. 5
	1	0	1	0	0

3. Datapath問題



Reg. number	Reg. content	Reg. number	Reg. content
00000	800h	10000	0
00001	200h	10001	0
00010	11111111h	10010	0
00011	50h	10011	0
00100	0	10100	0
00101	0	10101	0
00110	0	10110	0
00111	0	10111	0
01000	100h	11000	0

Instr. Mem. Addr.	Instr. Mem. content
100h	lw $t9, 10($t1)
104h	lw $t10, 14($t1)
108h	add $t3, $t9, $t10
10ch	sw $t3, 4($t1)
110h	beq $t3, $t4 100
114h	
118h	

Data Mem. Addr.	Data Mem. content
200h	0
204h	0
208h	0
20ch	0
210h	55555555h
214h	aaaaaaaah
218h	0

01001	0	11001	0				
01010	0	11010	0				
01011	0	11011	0				
01100	0	11100	0				
01101	0	11101	0				
01110	0	11110	0				
01111	0	11111	0				

注意：

(1) PC內容為100h

(2) 執行指令lw $t9, 10($t1)，其中$t9以01001表示，$t1以00001表示，10為16進制

(3) sw指令說明同(2)

問題：

(a) 執行第一個指令時(lw)，上圖中箭號所指編號1, 2, 3, 4, 5, 6, 7, 8處之值為何？(5, 6, 7, 8請以16進制表示)

(b) 執行第三個指令時(add)，上圖中箭號所指編號1, 2, 3, 4, 5, 6, 7, 8處之值為何？

(c) 所有指令執行完畢後，PC內容為何？請以16進制表示。

Answer

(a)	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8
	100011_2	00001_2	01001_2	00000_2	10h	200h	10h	55555555h
(b)	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8
	000000_2	01001_2	01010_2	00011_2	1820h	0	0	0
(c)	$PC = 114_{16} + (4 \times 100)_{10} = 114_{16} + 190_{16} = 2A4_{16}$							