Problem Set I

Problems selected for assignment shall be solved and submitted

Assignment: {3, 11, 12, 13, 14} working in group is allowed but sharing answers will disqualify both involved parties and result in ZERO score (Never share your answer sheet; no mercy!)

1. You are to write an IAS program to compute the results of the following equation

$$Y = \sum_{X=1}^{N} X$$

Assume that the result of the computation does not arithmetic overflow and that X, Y, and N are positive integers with N *Note*: The IAS did not have assembly language only machine language.

(a) Use the equation Sum(Y)= $N(N+1)/2$ when writing the IAS program.

(b) Do it the "hard way," without using the equation from part (a).

2. Given the memory contents of the IAS computer shown below, show the assembly language code for the

Address	Contents
08A	010FA210FB
08B	010FA0F08D
08C	020FA210FB

program, starting at address 08A. Explain what this program does.

3. Using the IAS instruction set, write a program that reads two values from address 000H &001H and write back their sum at 002H.

4. Consider two different machines, with two different instruction sets, both of which have a clock rate of 200 MHz. The following measurements are recorded on the two machines running a given set of benchmark programs:

Instruction Type	Instruction Count (millions)	Cycles Per Instruction
Machine A		
Arithmetic and logic	8	1
Load and store	4	3
Branch	2	4
Others	4	3
Machine A		
Arithmetic and logic	10	1
Load and store	8	2
Branch	2	4
Others	4	3

(c) Determine the effective CPI, MIPS rate, and execution time for each machine.

(d) Comment on the results.

5. The following table shows the execution times, in seconds, for five different benchmark programs on three machines.

Benchmark	Processor		
	R	M	Z
E	417	244	134
F	83	70	70
H	66	153	135
I	39,449	35,527	66,000
K	772	368	369

(a) Compute the speed metric for each processor for each benchmark, normalized to machine R. That is, the ratio values for R are all 1.0. Other ratios are calculated using Equation (2.5) with R treated as the reference system. Then compute the arithmetic mean value for each system using Equation (2.3).

(b) Repeat part (a) using M as the reference machine.

(c) Which machine is the slowest based on each of the preceding two calculations?
(d) Repeat the calculations of parts (a) and (b) using the geometric mean, defined in Equation (2.6). Which machine is the slowest based on the two calculations?

6. To clarify the results of the preceding problem, we look at a simpler example.

Benchmark	Processor		
	X	Y	Z
1	20	10	40
2	40	80	20

(a) Compute the arithmetic mean value for each system using X as the reference machine and then using Y as the reference machine. Argue that intuitively the three machines have roughly equivalent performance and that the arithmetic mean gives misleading results.
(b) Compute the geometric mean value for each system using X as the reference machine and then using Y as the reference machine. Argue that the results are more realistic than with the arithmetic mean.

7. Consider a hypothetical 32-bit microprocessor having 32-bit instructions composed oftwo fields: the first byte contains the opcode and the remainder the immediate operand or an operand address.
(a) What is the maximum directly addressable memory capacity (in bytes)?
(b) Discuss the impact on the system speed if the microprocessor bus has
  (i) a 32-bit local address bus and a 16-bit local data bus, or
  (ii) a 16-bit local address bus and a 16-bit local data bus.
(c) How many bits are needed for the program counter and the instruction register?

8. Consider a hypothetical microprocessor generating a 16-bit address (for example, assume that the program counter and the address registers are 16 bits wide) and having a 16-bit data bus.
(a) What is the maximum memory address space that the processor can access directly if it is connected to a "16-bit memory"?
(b) What is the maximum memory address space that the processor can access directly if it is connected to an "8-bit memory"?
(c) What architectural features will allow this microprocessor to access a separate "I/O space"?
(d) If an input and an output instruction can specify an 8-bit I/O port number, how many 8-bit I/O ports can the microprocessor support? How many 16-bit I/O ports? Explain.

9. Consider a computer system that contains an I/O module controlling a simple keyboard/printer teletype. The following registers are contained in the processor and connected directly to the system bus:

INPR: Input Register, 8 bits
OUTR: Output Register, 8 bits
FGI: Input Flag, 1 bit
FGO: Output Flag, 1 bit
IEN: Interrupt Enable, 1 bit

Keystroke input from the teletype and printer output to the teletype are controlled by the I/O module. The teletype is able to encode an alphanumeric symbol to an 8-bit word and decode an 8-bit word into an alphanumeric symbol.

(a) Describe how the processor, using the first four registers listed in this problem, can achieve I/O with the teletype.
(b) Describe how the function can be performed more efficiently by also employing IEN.

10. A microprocessor has an increment memory direct instruction, which adds 1 to the value in a memory location. The instruction has five stages: fetch opcode (four bus clock cycles), fetch operand address (three cycles), fetch operand (three cycles), add 1 to operand (three cycles), and store operand (three cycles).
(a) By what amount (in percent) will the duration of the instruction increase if we have to insert two bus wait states in each memory read and memory write operation?
(b) Repeat assuming that the increment operation takes 13 cycles instead of 3 cycles.

11. Use the online INTERACTIVE Block Replacement Simulator (click the hyperlink) to answer the following problems

i. Consider the input sequence of length 120 given below and the table below.

220 66 131 35 94 172 126 217 73 176 250 84 114 187 201 116 4 102 84 22 44 87 114 82 144 28 211 131 25 192 12 134 176 157 197 211 223 67 199 203 30 154 51 123 140 172 218 249 27 91 5 51 202 59 196 240 238 71 100 217 49 231 226 12 118 233 204 222 220 31 220 66 173 5 6 94 62 126 124 250 21 81 74 116 233 9 167 62 20 4 161 35 152 102 79 73 86 84 182 22 92 44 66 159 187 240 167 100 169 201 174 114 232 82 187 87 175 131 156 301

	2- way			4 – way			8 - way		
	LRU	FIFO	RAND	LRU	FIFO	RAND	LRU	FIFO	RAND
8 blocks									
16 blocks									
32 blocks									
64 blocks									

(a) Analyze the effectiveness of different block replacement techniques by listing down the miss rate in each case.
(b) What other block replacement technique can be used and is proved to be the ideal? Explain.

ii. In a N-way set-associative cache, blocks are mapped to different sets when N changes. Also, for a particular sequence, the number of compulsory and conflict misses change with the cache type. Consider the following sequence 4 0 9 7 8 11 7 5 2 1 12 6 8.

(a) List the compulsory and conflict misses for different replacement techniques for the caches below.

MISSES	**LRU**		**FIFO**		**RANDOM**	
	Comp	***Conflict***	***Comp***	***Conflict***	***Comp***	***Conflict***
4blocks 2sets						
8blocks 2sets						
16blocks 2sets						

(b) Try to define compulsory, capacity and conflict misses. Explain the difference between them.
(c) What is the best way to reduce conflict misses? Can it be used?
(d) List which set in the given cache will the following blocks be mapped

BLOCK	**CACHE**	**#SET**
0	***8block,2sets***	
9		
11		
4		
2	***8blocks,4sets***	
9		
10		
4		
7	***16blocks,2sets***	
1		
12		
3		

12. Use the online INTERACTIVE Multitask Cache Simulator (click the hyperlink) to answer the following problem

> Note:
>
> 1) The trick to use the same set of memory references is to click on the "GENERATE MEMORY REFERENCES" only for the first time while setup. Do not click on this button until your simulation-based comparison is done.
> 2) Always choose to run the complete simulation

i. When multiple tasks are scheduled in a system, the number of cycles that one task is allowed to run before the CPU switches to another task is very important.

Assume the following setup for the animation.

Specification	**Value**
Replacement Policy	LRU
Scheduling Mechanism	Round Robin
Use Random Access Sequence	**NO**
Number of Tasks	4
Number of memory references for Task A,B,C,D	25 each

Consider the following sequences for the 4 tasks A, B, C and D

# 25	9	24	39	63
# 24	9	28	10	14
# 23	9	8	10	63
# 22	57	46	46	63
# 21	29	35	26	14
# 20	15	28	63	8
# 19	57	29	2	63
# 18	57	56	13	53
# 17	48	56	24	53
# 16	60	22	32	53
# 15	57	46	2	29
# 14	48	52	6	62
# 13	49	29	17	29
# 12	49	22	10	8
# 11	60	0	13	8

# 10	24	0	62	8
# 9	60	46	32	8
# 8	15	52	10	5
# 7	49	8	10	56
# 6	40	29	28	29
# 5	49	29	2	26
# 4	49	52	34	14
# 3	49	52	40	49
# 2	9	39	10	62
# 1	27	21	39	44
Task	**A**	**B**	**C**	**D**

(a) Find the reload transient for the following caches with different CPU slice times

Cache Size, # Sets	CPU Time Slice	Reload Transients
16, 2	1	
	3	
	4	
	5	
	8	
16,8	1	
	3	
	4	
	6	
	8	
32,2	1	
	3	
	4	
	6	
	8	

(b) What is the CPU time slice number (in cycles) that reduces the number of reload transients in each of the following caches?

(c) In a few sentences, explain Reload Transient and Task Footprint. How are they related?

(d) Explain why it is important to find the ideal time for a CPU slice for a particular cache.

---

ii. Use the following settings as simulation input

Specification	**Value**
Replacement Policy	LRU
Use Random Access Sequence	**NO**
Number of Tasks	5
Number of memory references for Task A,B,C,D,E	5 each
CPU Time slice	1
Priority A,B,C,D,E	3, 1, 0, 2, 4

Memory references of each task

# 5	52	93	113	28	9
# 4	29	93	113	28	9
# 3	19	93	104	125	9
# 2	93	19	104	89	31
# 1	19	76	113	125	9
Task	**A**	**B**	**C**	**D**	**E**

(a) Does the task scheduling technique affect the hit rate of the overall system? Explain.

(b) Find the hit rate in each of the following caches after 12 simulation cycles.

Cache Hit Rate % Cache Size, # Sets	FIFO	Round Robin	Priority Based 3-1-0-2-4 priority
16, 4			
16,8			
32,2			

(c) When does the round robin technique behave like the FIFO technique?
(d) When can the priority-based technique behave like the round robin technique?

13. Study the Influence of the Mapping on the miss rate for several cache sizes

- Download the simulator (setup.exe) from the course page and install on a Windows system.
- Open the simulator and configure a system with the following architectural characteristics:
  - Processors in SMP = 1.
  - Cache coherence protocol = MESI. (Read Chapter 17.3 CACHE COHERENCE AND THE MESI PROTOCOL)
  - Scheme for bus arbitration = Random.
  - Word wide (bits) = 32.
  - Words by block = 64 (block size = 256 bytes).
  - Blocks in main memory = 4096 (main memory size = 1024 KB).
  - Replacement policy = LRU.
- Configure the mapping using the following configurations: Direct, two-way set associative, four-way set associative, eight-way set associative, and fully-associative (remember: Number_of_ways = Number_of_blocks_in_cache / Number_of_cache_sets). For each of the configurations of mapping, configure the number of blocks in cache in order to get the following cache sizes: 4 KB (16 blocks in cache), 8 KB, 16 KB, and 32 KB (128 blocks in cache). For each configuration obtain the miss rate using the memory trace: Ear. You can find the trace inside the same directory above.

(a) Does the miss rate increase or decrease as the associativity increases? Why? What does it happen with the conflict misses when you enlarge the associativity grade?
(b) Does the influence of the associativity grade increase or decrease as the cache size increases? Why?
(c) In conclusion, does the increase of associativity improve the system performance? If the answer is yes, in general, which is the step with more benefits: from direct to 2-way, from 2-way to 4-way, from 4-way to 8-way, or from 8-way to fully-associative?

14. Study the Influence of the Replacement Policy of cache memory on the miss rate

- Download the simulator (setup.exe) from the course page and install on a Windows system.
- Open the simulator and configure a system with the following architectural characteristics:
  - Processors in SMP = 1.
  - Cache coherence protocol = MESI.
  - Scheme for bus arbitration = Random.
  - Word wide (bits) = 16.
  - Words by block = 16 (block size = 32 bytes).
  - Blocks in main memory = 8192 (main memory size = 256 KB).
  - Blocks in cache = 128 (cache size = 4 KB).
  - Mapping = 8-way set-associative (cache sets = 16).
- Configure the replacement policy using the following configurations: Random, LRU,LFU, and FIFO. For each of the configurations, obtain the miss rate using the trace files (extension ".prg"): Hydro, Nasa7, Cexp, Mdljd, Ear, Comp, Wave, Swm and UComp. You can find the trace files inside the *installation_directory**Traces**PRG.*

(a) In general, which is the replacement policy with the best miss rate? And which does it have the worst? Do the benefits of LFU and FIFO policies happen for all the benchmarks or do they depend on the different locality grades?
(b) For a direct-mapped cache, would you expect the results for the different replacement policies to be different? Why or why not?
(c) In conclusion, does the use of a concrete replacement policy improve the system performance? If the answer is yes, in general, which is the step with more benefits: from Random to LRU, from Random to LFU, or from Random to FIFO? Why (consider the cost/performance aspect)?