NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES
(FAST-NU), KARACHI CAMPUS
COMPUTER ARCHITECTURE-EE204
FINAL EXAMINATION, FALL 2016
16TH DECEMBER 2016

Max. Marks : 100 Time: 3 hours
Pages : Three

Note: Attempt **all** questions.

Q1(a) **i.** Why do benchmarks need periodic updates in the form of newer versions? (1)
ii. Name a few 'global compiler optimization techniques' that are used to enhance the performance of applications. (1.5)

Q1(b) The performance of an application is enhanced with a speedup of **1.75** when a graphics processor is added to the architecture. What is the percentage of graphics instructions in the application if the added hardware is **10** times better than its software counterpart? (4)

Q1(c) What are the three methods of encoding branch conditions in processors? Briefly explain each of them. (5)

Q2(a) **i.** The slowest stage of a pipelined processor takes **30 nSec** and the fastest stage takes **25 nSec**. What is the pipelined clock frequency of the processor if the pipeline latch delay is **1 nSec**? (1.5)
ii. What do you understand by the 'latency' and 'initiation interval' of an execution unit in a pipelined processor? (3)

Q2(b) Calculate the average CPI of a processor P2, if the frequency of occurrence and the CPI of each type of instruction is as given below: (5)

Type of instruction	CPI	Frequency of occurrence
A	1.7	22%
B	2.1	29%
C	2.7	17%
D	2.4	Remaining

Q2(c) What are RAW, WAR and WAW hazards? Suggest methods/techniques that are used in pipelined processors to alleviate the above problems. (5)

Q3(a) **i.** What are precise exceptions? (1)
ii. What are correlating branch predictors? (1.5)

iii. How does the Tomasulo's approach of dynamic scheduling perform internal forwarding to resolve RAW hazards? (1)

Q3(b) The following code sequence is executed in a processor using Tomasulo's approach of dynamic scheduling. Give the entries of Reservation Stations, the load and store buffers and register status assuming the following resources in the processor: (10)

- Two Reservation Stations for one Add/Sub Unit.
- Two Reservation Stations for one Mul/Div Unit
- One integer pipe for integer, load/store operation
- Four Load Buffers
- Three Store buffers
- 32 Floating Point registers.

			Instruction status		
			Issue	Execute	Write Result
1.	L.D	F0, 0(R1)	√	√	√
2.	L.D	F6, 0(R2)	√	√	
3.	MUL.D	F4, F0, F2	√		
4.	ADD.D	F8, F4, F6	√		
5.	S.D	F8, 0(R1)	√		
6.	L.D	F0, -8(R1)			
7.	L.D	F6, -8(R2)			

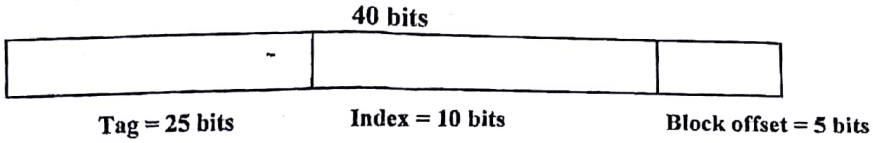
Q3 (c) What is the purpose of loop unrolling? Unroll and schedule the following code 4 times? Use register renaming and schedule the code to achieve no stalls in the final code: (10)

```
LOOP:   L.D     F0, 0(R1)
        ADD.D   F4, F0. F2
        S.D     F4, 0(R1)
        DADDI   R1, R1, 8
        BNE     R1, R2, LOOP
```

Q4(a) **i.** How does the block search take place in a set-associative cache? (1)
ii. Out of 1000 memory references, suppose there are 40 misses in the first-level cache and 20 misses in the second-level cache. Calculate the local and global miss rates of both levels of cache. (3)

Q4(b) Given the following fields in a direct-mapped cache, find the values of the following cache and main memory parameters: (10)

i. Size of the main memory and cache
ii. Block size and the number of blocks in cache
iii. The size of each field (in terms of number of bits) for a 4-way set-associative cache.



Q4(c) What are Compulsory, Conflict and Capacity misses of cache? What measures can be taken to reduce each of the above types of misses? (6)

Q5(a) **i.** Give the advantage of write-back policy over write-through policy of cache. (1.5)

ii. For the Hexadecimal address **4C2ABC95h**, give the content of Tag field and the index for a direct-mapped cache, if the cache is of **128KBytes** and the block size is of 32 bytes. (3)

Q5(b) Explain briefly the following cache optimization techniques, identifying the component of memory access time that is improved with the given technique: (10)

i. Multilevel cache
ii. Non-blocking cache
iii. Critical word first
iv. Hardware prefetching

Q5(c) **i.** What do you understand by the fast page mode operation of DRAMs? (2)
ii. Which write policy is used between main memory and secondary storage devices? (1)

Q6(a) **i.** List the four classes of computers based on Flynn's classification scheme.(1.5)
ii. What are the two software models used to exploit thread-level parallelism? (2)
iii. Differentiate between UMA and NUMA architecture. (1.5)

Q6(b) What are the two main challenges of parallel processing? Suggest methods and techniques that can be used to deal with these challenges. (5)

Q6 (c) **i.** How is the dependability of RAIDs improved? (1)
ii. How is RAID4 different from RAID5? (2)