EEL 5764 Computer Architecture
Fall 2018
Midterm 2

Name: _____ **UFID:** _____

Notes:

- Exam is 75 minutes long.
- Exam is closed book and closed notes, other than 1 US-letter-size crib sheet of handwritten notes
- Other than scientific calculators, all electronics must be turned off.
- Please write legibly. Otherwise your grades may be negatively affected.
- You must show your work for all problems.
- Use the back of the exam paper as necessary. Indicate clearly, which problems the answers on the back correspond to.

---

1. **(10x2 pts)** Decide whether the following statements are true or false. Add brief (1-2 sentence) explanations to justify your answers.

   a. Vector processors primarily use MIMD architectures.

False. Vector processors use SIMD.

   b. Distributed shared memory multiprocessors is a misnomer since each processor can only access its local memory.

False. Processors can access both local and remote memories.

   c. Linear speedups are needed to make multiprocessors cost effective.

True. Less than linear won't make them cost effective.

   d. Security assets in an SoC can be introduced by manufacturers as well as end users.

True.

   e. Deep learning evolved from neural networks.

True. (But not relevant to Fall 2019, since we didn't cover Deep Learning Architectures.)

f. You can get good vector performance without providing memory bandwidth.

False.

g. Cache coherence requires writes to the same location to be serialized.

True.

h. Post-silicon validation requires upfront planning to ensure relevant observability.

True.

i. Warehouse scale processors emphasize local computation over remote ones.

True.

j. Sequential consistency is a more relaxed consistency model than processor consistency.

False. Sequential consistency is the strictest.

2. **(15 pts total)** Consider the following code for a VMIPS processor, assume chaining is enabled, one load and store unit, 5 functional units, and setup time is ignored.

(1)	LV	V1,R1	(5) SV	R3,V3
(2)	LV	V2,R2	(6) SV	R4,V4
(3)	MULVS.D	V3,V2,V1	(7) SUBVV.D	V5,V4,V3
(4)	ADDVV.D	V4,V1,V3	(8) SV	R5,V5

a. (10 pts) How many convoys are required to execute the code?
b. (5 pts) Approximately how many cycles does it take to execute a 64-element vector?

(a) Convoys execute code segments that are free of structural hazard. Each load must be in one convoy. Chaining allows two vector instructions to start executing together. Inst 1 & 2need be in 2 convoys, 3-5 in 1, 6 in 1, 7, 8 in 1conv. Alternatively, 1 in 1conv, 2-4 in 1conv, 5 in 1conv, 2 & 7 in 1conv, 8 in 1conv. Total: 5 convoys.

(b) # of convoys * # of elements per vector = 5 * 64 = 320 cycles. Note they may have more/less convoys.

3. **(20 pts)** How many 2GHz multicores can a 1024-banks memory system support, assuming memory cycle time of 2ns and processors are capable of 2 LD and 2 ST operations per cycle.

For each ld/st each mem bank is busy for 2ns/0.5ns = 4cycles; Since there are 1024 banks, there can be 1024/4 = 256 simultaneous ld/st at a time; Maximum # of mem ref per processor is 4. No. of processors needed = 256/4 = 64 processors.

4. **(15 pts)** For a given video, the frame rate is 16 f/s, and each frame is composed of 1024 data elements, independent of other frames, and can be completely parallelized. A programmer writes code to process 32 seconds of video frames using a GPU, each processing subroutine processes one frame. The GPU has 16 SIMD processors, where each can execute threads of 32 data elements per instruction.
    a. What is the optimal grid size, in number of elements? (Recall: Grid size in GPU is analogous to vectorizable loops.)
    b. How many SIMD threads are required to process the video?
    c. How long will it take to process the video if each SIMD processor takes 1 second/thread?

(a) Grid: elements that can be completely parallelized. Each frame can be parallelized. So grid size is 1024.

(b) Number of frames = (32 sec)*(16 frames/sec) = 512 frames. Number of threads needed per each frame = (1024 elements/frame) / (32 elements/thread) = 32hreads/frame. Total number of threads = 32 threads/from * 512 frames = 16384 threads

(c) 16384 threads / 16 frames/second = 1024 seconds

5. **(10 bonus pts)** Consider the vector operation Y = a × X + Y where X and Y are both vectors of 64 elements. Show MIPS assembly code for the MIPS superscalar processor, and vector assembly code for VMIPS processor, and compare their performance in terms of the total number of instructions executed on these two processors.
    - If you do not remember the ISAs, use simple sentences like "load X[i]" for a scalar load, or "vector-scalar multiply X and a" for a × X, etc.
    - Assume the vector length of VMIPS is 64.

This was worked out in class. See class slides.

6. **(10 bonus pts)** Use the GCD test to determine whether dependencies exist in the following loop:

```
for (i = 0; i < 100; i++) {
   x[2*1+3] = x[2*i] * 5.0;
}
```

This was worked out in class. See class slides.