CS 351 Fall 2017 Exam 3 Mon. 12/11/2017

Name: _____

Rules and Hints

- You may use one handwritten 8.5 × 11″ cheat sheet (front and back). This is the only additional resource you may consult during this exam. *No calculators.*
- When applicable, you may write your answers in the form *[mathematical expression][units]*. There is no need to actually do the arithmetic.

Grade

	Your Score	**Max Score**
Problem 1: Short answer		15
Problem 2: Memory hierarchy performance		20
Problem 3: Cache tracing		30
Problem 4: RAID		25
Problem 5: Parallelism		10
Total		**100**

Problem 1: Short answer (15 points)

Part A (5 points)

You are iterating over a very large array. Does it matter, in terms of performance, whether you...

1. Iterate sequentially over each element
2. Iterate over the elements with even indices first, then do a second pass for the elements with odd indices?

Part B (5 points)

Which would you expect to have more bits of metadata: a fully associative cache, or a direct-mapped cache with the same amount of data? Assume that the block size is the same for both caches.

Part C (5 points)

Which would you expect to have better performance: making a single 2 MB disk access, or making two 1 MB accesses? Explain.

Problem 2: Memory hierarchy performance (20 points)

Consider this memory hierarchy:

- An L1 cache with a hit rate of 90% and an access time of 1 processor cycle
- An L2 cache with a hit rate of 75% and an access time is 6 processor cycles
- Main memory access time of 100 cycles

Part A: AMAT (10 points)

What is the average memory access time?

Part B: Memory access time per instruction (5 points)

You are executing a program whose dynamic instruction breakdown is 60% arithmetic operations, 25% loads and stores, and 15% branches. How much time does an average instruction in this program spend accessing memory?

Part C: Adding an L3 (5 points)

You are considering adding an L3 cache. Is it possible that this addition could actually worsen performance? Explain.

Problem 3: Cache tracing (30 points)

For all parts of this problem, assume that your addresses are 8 bits. Draw the final state of each of these caches, **and state the hit rate**, for the following sequence of read addresses:

```
0 0 0 0 1 1 0 1
1 1 0 0 1 0 0 1
0 0 0 0 1 1 0 0
1 0 0 0 1 1 0 0
0 0 1 0 1 0 0 1
1 0 0 0 0 1 1 1
1 0 0 0 0 1 1 0
```

Part A (10 points)

A direct-mapped cache with 8 bytes of data and 1-byte blocks

Part B (10 points)

A direct-mapped cache with 8 bytes of data and 2-byte blocks. Again, the sequence of references is:

```
0 0 0 0 1 1 0 1
1 1 0 0 1 0 0 1
0 0 0 0 1 1 0 0
1 0 0 0 1 1 0 0
0 0 1 0 1 0 0 1
1 0 0 0 0 1 1 1
1 0 0 0 0 1 1 0
```

Part C (10 points)

A 2-way set-associative cache with 8 bytes of data and 1-byte blocks. Again, the sequence of references is:

```
0 0 0 0 1 1 0 1
1 1 0 0 1 0 0 1
0 0 0 0 1 1 0 0
1 0 0 0 1 1 0 0
0 0 1 0 1 0 0 1
1 0 0 0 0 1 1 1
1 0 0 0 0 1 1 0
```

Problem 4: RAID (25 points)

Assume for all parts of this problem that you can afford to buy exactly 8 1-TB disks.

Part A (6 points)

How much (non-redundant) data can you fit on these disks if you configure them as...

- A RAID 0 array?
- A RAID 1 array?
- A RAID 5 array?

Part B (4 points)

If you choose a **RAID 0** configuration, how many disks can you access at a time if you're doing...

- A lot of small reads?
- A lot of small writes?
- A single large read?
- A single large write?

Part C (4 points)

If you choose a **RAID 1** configuration, how many disks can you access at a time if you're doing...

- A lot of small reads?
- A lot of small writes?
- A single large read?
- A single large write?

Part D (5 points)

Why is the latency of a small write on RAID 5 so high?

Part E (6 points)

If the latency of a small write is 10 ms when your 8 disks are configured as JBOD, what's the throughput (in small writes per second) for...

- A RAID 0 array?
- A RAID 1 array?
- A RAID 5 array?

Problem 5: Parallellsm (10 points)

You have a program that takes the following amounts of time when executed with the following number of threads on your 8-core laptop:

Threads	1	2	4	8	16
Execution time (ms.)	100	52	35	20	25

Part A: Speedup (5 points)

Plot the speedup (over 1 thread) vs. number of threads. Your axes should be labeled and with both a title and with numeric values.

Part B: Efficiency (5 points)

Plot the efficiency vs. number of threads. Your axes should be labeled and with both a title and with numeric values.