

# University of California, Berkeley – College of Engineering

Department of Electrical Engineering and Computer Sciences

Spring 2017

Instructors: Nicholas Weaver, Gerald Friedland

2017-05-09

# :( CS61C FINAL :)

After the exam, indicate on the line above where you fall in the emotion spectrum between “sad” & “smiley”...

<i>Last Name</i>	
<i>First Name</i>	
<i>Student ID Number</i>	
<i>CS61C Login</i>	cs61c-
<i>The name of your SECTION TA and time</i>	
<i>Name of the person to your LEFT</i>	
<i>Name of the person to your RIGHT</i>	
<i>All the work is my own. I had no prior knowledge of the exam contents nor will I share the contents with others in CS61C who have not taken it yet. (please sign)</i>	

## Instructions (Read Me!)

- This booklet contains 16 numbered pages including the cover page.
- Please turn off all cell phones, smartwatches, and other mobile devices. Remove all hats & headphones. Place your backpacks, laptops and jackets under your seat.
- You have 180 minutes to complete this exam. The exam is closed book; no computers, phones, or calculators are allowed. You may use three handwritten 8.5"x11" page (front and back) crib sheet in addition to the MIPS Green Sheet, which we will provide.
- There may be partial credit for incomplete answers; write as much of the solution as you can. We will deduct points if your solution is far more complicated than necessary. When we provide a blank, please fit your answer within the space provided.
- Scores for Q1-Q3 will be used for Midterm 1 clobbering and scores Q4-Q6 will be used for Midterm 2 clobbering

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
<b>Points Possible</b>	15	10	10	10	10	15	10	20	15	5	120

## Q1: Many, Many Matrices (15 points)

In this problem, you will be asked to implement operations on a data structure that represents a set of 2D matrices. As the structure is a collection of matrices, it can be conveniently defined as a triple pointer, as such: `int ***matrices`. For simplicity, we will only consider square matrices of integers.

Please fill in the blanks in the functions below to complete the desired implementation. **You may not need all of the lines provided.**

When completing the functions, you may use the following assumptions:

- All pointers are valid and non-NULL
- All calls to `malloc` are guaranteed to succeed and do not need to be checked
- All dimension and number counts are valid, meaning that we **never** pass in the dimension value of a matrix that is larger or smaller than the actual dimension
- For multiplication, you may assume that the dimensions align such that the matrix multiplication is valid
- For any matrix `m`, the value at the `i`th row and `j`th column is denoted by `m[i][j]`.

You MAY NOT assume that the size of an integer is 4 bytes, the size of a character is 1 byte etc.

Each function has a description of what it does. You may find it helpful to see the example of the product of a matrix and vector shown below.

$$\begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_{00}b_0 + a_{01}b_1 + a_{02}b_2 \\ a_{10}b_0 + a_{11}b_1 + a_{12}b_2 \\ a_{20}b_0 + a_{21}b_1 + a_{22}b_2 \end{bmatrix}$$

```
// Allocates all memory for a set of "numMatrices" matrices, each with
// dimensions "dim x dim". Does not initialize the values in each matrix.
```

```
int ***allocate(int numMatrices, int dim) {
    int ***matrices =
    for (int i = 0; i <          ;    ) {
        matrices[ ] =
        for (int j = 0;          ;    ) {
            matrices[ ][ ] =
        }
    }
    return matrices;
}
```

```

// Multiplies the kth matrix in the set with the vector provides and
// returns this new vector
int *multiply(int ***matrices, int *vector, int k, int dim) {
    int *product =

    int **matrix =
    for (int i = 0; i < dim; i++) {

    }
    return product;
}

// Free's all memory from the given set of matrices
void freeMatrices(int ***matrices, int numMatrices, int dim) {
    for (int i = 0;          ) {
        for (int j = 0;          ) {

        }

    }

}

```

## Q2: Memory State Warriors (10 points)

1) Consider the block of code below. You may assume all calls to malloc succeed.

```
#define TEAMS 30
char *steph = "curry";

int main(void) {
    char *kevin = "durant";
    char klay[9] = "thompson";
    int wins = 67;

    char **warriors = malloc(3 * sizeof(char *));
    warriors[0] = kevin;
    warriors[1] = steph;
    warriors[2] = (char *) klay;

    char ***nba = malloc(TEAMS * sizeof(char **));
    *nba = warriors;

    return wins;
}
```

For the code on the previous page, please list on what region of memory the following variables/quantities reside right before the function main returns. If there are multiple answers, write all possible regions.

Variable/Quantity	Region of Memory
nba	
*nba	
**nba	
**((*nba) + 1)	
klay	
steph	
TEAMS	

2) Consider the block of code below that, when given a value n, returns an array with the integer values from 1 to n. Are there any issues with the function? If yes, then **list all issues** with a brief explanation of each one on the lines below. Furthermore, for each issue you outline, describe how to fix the problem on following line. You may or may not need all lines.

```
int *array_of_n(int n) {
    int arr[n];
    for (int i = 0; i <= n; i++) {
        arr[i] = i + 1;
    }
    return arr;
}
```

1. Problem:

Solution:

2. Problem:

Solution:

3. Problem:

Solution:

### Q3: Back at it Again with Mipstery (10 points)

**Mipstery** is known to take in one positive integer as the only argument. Assume **Mipstery** resides at address **0xABCDEEFF**. [Hint: do not try to interpret Mipstery immediately (part d) -- parts a-c can be answered without first understanding the entire program!]

```
Mipstery: 1.   addiu    $t0, $0, 1
          2.   addu   $s0, $t0, $0
          3.   la     $t1, Mipstery
          4.   lw    $t2, 4($t1)
          5.   li    $t3, 0x7FFF
          6.   and   $t4, $t3, $t2
          7.   ori   $t4, $t4, 32768      # 32768 = 2^15
          8.   sll   $t2, $t2, 20
          9.   or    $t5, $t2, $t4
         10.   sw    $t5, 4($t1)
         11.   addiu   a0, a0, -1
         12.   bne   $a0, $0, Mipstery
         13.   move   $v0, $s0
         14.   j     done
          ... ..
```

- (a) Which instruction gets modified in the above code? Give the line number.
- (b) Notice that “la” and “li” (line 3 and line 5) are both pseudo-instructions. Assemble these instructions into TAL as efficiently as possible, recalling specifications for the pseudo-instructions from your project 2.
- ```
la    $t1, Mipstery          li    $t3, 0x7FFF
```
- (c) Convert the **bne** instruction (line 12 in the MAL code above) into machine code, assuming that “la” gets expanded into 1 instruction and “li” gets expanded into 3 instructions. Express your answer as a hexadecimal number.
- (d) (**Hard**) In one sentence: What does Mipstery do? Assume \$a0 is always positive. [Hint: first convert the modified instruction into hex!]

## Q4: When Memory Encounters Branches (10 points)

Consider the new instruction “memory branch equal”:

```
membeq $rs, $rt, label
```

It performs the following operation:

```
if (Mem[R[$rs]] == Mem[R[$rt]]):
    R[$at] ← 1
    PC ← PC + 4 + BranchAddr
Else:
    R[$at] ← 0
    PC ← PC + 4
```

Your task is to modify the single cycle MIPS CPU you have seen in lecture to support membeq while maintaining functionality of the rest of the MIPS ISA.

(a) First, assume you are free to modify the wires and gates, but cannot modify the major components (Instruction Memory, Regfile, ALU, Data Memory). Are you able to implement this instruction using our current CPU datapath? In one sentence, explain your answer.

(b) Now assume that you CAN modify ONE major component listed in part (a), and you may add to it one new input and one output, called newInput and newOutput, which you may define as you'd like. You may also modify wiring and gates in areas labelled (i), (ii), (iii) and (modified) in your diagrams packet. Select the correct implementation for the **blanks (i), (ii) and (iii)** from the choices in the packet. **If multiple solutions exist choose the one that uses the least amount of hardware. Also note that two of the control signals have been defined for membeq in part (c) below.**

Select the correct implementations for the given blanks:

| (i) | (ii) | (iii) |
|-----|------|-------|
|     |      |       |

(c) List the values of all control signals for the membeq instruction in the table below. If the value of a particular signal does not matter, you must put an 'X'. For the ALUCtr field, write the name of the operation the ALU should execute.

| Jump | membeq | RegDst | RegWr | ExtOp | ALUSrc | ALUCtr | MemWr | MemToReg | Branch |
|------|--------|--------|-------|-------|--------|--------|-------|----------|--------|
|      | 1      |        |       |       |        |        |       |          | 0      |

### Q5: Pipelining (10 points)

| RegFile Read | RegFile Setup | Register Clk-To-Q | Register Setup | Register Hold | Control Unit | ALU   | Adder |
|--------------|---------------|-------------------|----------------|---------------|--------------|-------|-------|
| 200ps        | 200ps         | 100ps             | 100ps          | 100ps         | 300ps        | 300ps | 100ps |

| Shifter | Mux   | Or Gate | Memory Read | Memory Write | Extender | Hazard Detection |
|---------|-------|---------|-------------|--------------|----------|------------------|
| 100ps   | 100ps | 100ps   | 400ps       | 400ps        | 200ps    | 100ps            |

See the simplified MIPS Pipeline attached in the packet and labelled “**QUESTION 5.**” This processor can execute the following instructions: ADD, ADDI, SUB, LW, SW, BEQ. **Note that branches change the PC in the MEM stage.** Register values must be read during ID. New values are available from the Register File after WB.

1. What is the highest frequency that this CPU can run at?
  - a. (Fill out to facilitate partial credit) In what stage is the critical path?
  - b. (Fill out to facilitate partial credit) What is the critical path delay of that path?

2. Fill in the bubbles next to lines of code that contain hazards for this program.

- LW \$s4, 0(\$s1)
- ADDI \$s4, \$s4, 8
- ADD \$s4, \$s4, \$s3
- ADDI \$s1, \$s1, -4
- BEQ \$s4, \$zero, DONE //Assume taken
- LW \$s4, 0(\$s1)
- SW \$s4, 4(\$s2)

DONE :

3. Now, assume we resolve hazards by stalling. How many cycles would the program from part 2 take to execute?

Number of Cycles:

## Q6: Cold Cache, Divine (15 points)

Given the following specifications of the following caches, fill in the blanks with the correct value or term. For each of these scenarios, assume you are working with a 24 bit physical address space on a byte-addressed machine. *Be as precise in your answer as possible* (don't leave variables in your answer).

|              |               |             |       |
|--------------|---------------|-------------|-------|
| Cache 1:     |               | Cache 2:    |       |
| Type:        | Direct Mapped | Type:       |       |
| Size:        |               | Size:       | 32KiB |
| # of Sets:   | 32            | # of Sets:  |       |
| Tag Bits:    | 16            | Tag Bits:   | 12    |
| Index Bits:  |               | Index Bits: |       |
| Offset Bits: |               | Block Size: | 64B   |

Consider the following C code:

```
#define ARRAY_SIZE 128*1024
double a[ARRAY_SIZE];
int i, j;
void normalize(int vect_size){
    for (i = 0; i < ARRAY_SIZE, i+=vect_size) {
        double squared_sum = 0.0;
        for (j = 0; j < vect_size; j += 1) { // InnerLoop1
            squared_sum += a[i+j] * a[i+j];
        }
        rooted_sum = sqrt(squared_sum);
        for (j = 0; j < vect_size; j += 1) { // InnerLoop2
            a[i+j] /= rooted_sum;
        }
    }
}
```

Once again, assume we are working with 24 bit physical addresses and the byte-addressed machine from above. Consider a 256 B direct mapped cache with 32B blocks, whose valid bits are all set to 0. Doubles are 8 bytes, and the doubles stored in the array `a` are double-word aligned (e.g. at addresses 0, 8, 16...).

- a) Suppose that we set `vect_size` to be 32 such that our vectors have 32 components (numbers) in them. We execute `normalize(32)`; Over the entire execution, what is the **hit rate** of:

Inner Loop1:

Inner Loop2:

- b) The cache is reinitialized (it is cold again). What is the smallest value of `vect_size` such that we have the **maximum** possible **hit** rate of Inner Loop1 over the entire execution of `normalize(size)`? What is the maximum possible hit rate you found?

`vect_size:`

maximum possible hit rate:

- c) Given that you are allowed to change a single parameter about the cache or the code, from the following options, what can you do to increase the hit rate of Inner Loop1? Circle all answers you believe are correct.
- i) Increase cache size to be the array size
  - ii) Make the cache fully associative
  - iii) Double the block size to 64B
  - iv) Change the line `squared_sum += a[i+j] * a[i+j];` to be `double val = a[i+j]; squared_sum += val * val;`

## Q7: Performance Programming (10 points)

1) For each snippet of code below, **fill in the circles for all outputs** that could be printed after running the code. You may assume that the function `printf` is atomic and that OpenMP parallelizes “for” loops using a blocking scheme as opposed to an adjacent scheme where threads process lower index values first. For both parts, assume there are **2 threads used to parallelize the loop**.

a.

```
#pragma omp parallel for
for (int i = 0; i < 4; i++) {
    printf(“%d ”, i);
}
```

Outputs (Fill in ALL possible outputs):

0 1 2 3

0 1 3 2

0 2 1 3

2 0 1 3

2 1 3 0

None of the outputs listed are possible

b.

```
int x = 0;
#pragma omp parallel for
for (int i = 1; i < 3; i++) {
    x += i;
    printf("%d ", x);
}
```

Outputs (Fill in ALL possible outputs):

- 1 2
- 3 3
- 2 2
- 1 1
- 1 3
- None of the outputs listed are possible

2) The goal of the piece of code below is to print "0 1 2". **Fill in one** of the following options regarding the code and its functionality when there are 3 threads used to parallelize the loop.

```
#pragma omp parallel for
for (int i = 0; i < 3; i++) {
    #pragma omp critical
    printf("%d\n", i);
}
```

- Always works, faster than serial
- Sometimes Works
- Always works, slower than serial
- Never Works

### **Q8: I have virtually no memory of this... (20 points)**

We have 2 machines, A and B, both with a **4 entry TLB**, **4 KiB pages**, **16 MiB physical memory**, and **4 GiB (32b address space) of virtual memory**. They both have the same TLB and Page Table state, shown below. The only difference between these 2 machines is that, to store additional data beyond the capacity of physical memory, machine **A uses a hard disk drive** and machine **B uses a SSD**. On both machines, we run the same program that has the following memory accesses made in the order given below.

Accesses:

**R 0xA0000008**

**W 0xA0004FFC**

**W 0xA0001234**

TLB:

| VPN     | PPN   | Valid Bit | Dirty Bit | LRU |
|---------|-------|-----------|-----------|-----|
| 0x80003 | 0x649 | 1         | 0         | 0   |
| 0xA0001 | 0x61F | 1         | 0         | 1   |
| 0xBFFEC | 0x613 | 1         | 1         | 2   |
| 0xBFFE8 | 0x612 | 1         | 1         | 3   |

Page Table: shown starting from index 0xA0000.

| PPN   | Valid Bit | Dirty Bit |
|-------|-----------|-----------|
| 0x609 | 1         | 1         |
| 0x61F | 1         | 0         |
| 0x620 | 0         | 0         |
| 0x625 | 1         | 1         |
| 0x629 | 0         | 0         |
| 0x62B | 1         | 0         |
| 0x62E | 0         | 0         |

1. Translate the virtual address 0xA0000008 to its corresponding physical address.
2. Let us now analyze how our accesses progress through memory.
  - a. Consider the possible access progressions below. Next to each access, write the letter corresponding to its access progression.
    - a. TLB Hit, neither Page Hit nor Page Fault
    - b. TLB Hit, Page Hit
    - c. TLB Hit, Page Fault
    - d. TLB Miss, neither Page Hit nor Page Fault
    - e. TLB Miss, Page Hit
    - f. TLB Miss, Page Fault

**R 0xA0000008**

**W 0xA0004FFC**

**W 0xA0001234**

- b. Given the following times to access each component, calculate the latency of **W 0xA0002FFC** on Machine A and Machine B. Remember to include in your calculations, the time to translate addresses, the time to get the corresponding data at that address, and the time to update any necessary caches and/or tables. Hardware accesses do **NOT** happen in parallel (e.g. accessing the TLB and physical memory at the same time is not possible, one must happen first and then the next). There is no data cache for physical memory.

| Accessing TLB (us) | Accessing Physical Memory (us) | Accessing HDD (us) | Accessing SSD (us) |
|--------------------|--------------------------------|--------------------|--------------------|
| 0.001              | 0.1                            | 3000               | 450                |

A:

B:

3. Both machines now starts running a new program that reads all of a 1 MiB file, which is stored entirely on disk. Assume that this program's page table has no virtual memory mappings for this file.

- a. The file is memory mapped through the virtual memory system, and the program simply touches every page in the file (e.g. reads 1 byte from each page). How many page faults would occur from reading this file?

- b. How long would it take to process all of the page faults from part (a) on both systems? Use the same access times as in the previous section 2.b. You may leave your answer as an unsimplified expression, and may use PF in place of your answer to part (a).

A:

B:

- c. Now assume the file is laid out contiguously on disk and the program instructs the operating system to load the entire 1 MiB file in a single request. If there is a sustained transfer rate of 20 MiB/us for both systems, how long does this take to process for system A? How long for system B?

## Q9: I/O (15 points)

An important advantage of interrupts over polling is the ability of the processor to perform other tasks while waiting for communication from an I/O device. Suppose that a **1 GHz processor** needs to read 1000 bytes of data from a particular I/O device. The I/O device supplies 1 byte of data every 0.02 ms. The time to process the data and store it in a buffer is negligible.

a) Assume a polling iteration takes 60 cycles. If the processor detects that a byte of data is ready through polling:

1. How many cycles does it take for the I/O device to supply 1 byte of data?
2. How many polling iterations does it take to read 1 byte of data? (round up to an integer)
3. How many cycles does it take to read the 1000 bytes of data?

b) If instead, the processor is interrupted when a byte is ready, and the processor spends the time between interrupts on another task, how many cycles of this other task can the processor complete while the I/O communication is taking place? The overhead for handling an interrupt is 2000 cycles.

c) The advantage of polling however arises when data rates become very large so that the interrupt overhead becomes substantial and at some point the system simply can't keep up. What is the data arrival time (in ms) at which point an interrupt-driven I/O scheme on this computer can't keep up with the data coming in? The overhead for handling an interrupt is 2000 cycles.

## Q10: Potpourri (5 points)

a) One byte of data is encoded in Hamming ECC with single-error correction. The result is the 12-bit number `0xdd3`. What was the original byte of data? [Remember to check that the data is correct!]

b) Fill in the bubble for all statements that are True: \_\_\_\_\_

RAID 5 has better fault tolerance than RAID 0

RAID 5 has better fault tolerance than RAID 1 (nope, same)

RAID 0 can store data more compactly than RAID 1 (no redundancy)

RAID 5 can store data more compactly than RAID 1 (more efficient)

RAID 0 reads faster than RAID 1 (either choice is fine, multi-small reads are better for RAID1 while single reads are better for RAID 0),

RAID 5 writes faster than RAID 1 (nope, Raid5 needs 2x read first)

c) Two machines are being considered for purchase. The SuperDuper machine has a Mean Time to Failure (MTTF) of 24 hours and a Mean Time to Repair (MTTR) of 6 hours, while the UnbelievablyGreat machine has a MTTF of 4 hours and a MTTR of 1 hour.

A. What's the availability of the SuperDuper Machine?

B. What's the availability of the UnbelievablyGreat Machine? \_\_\_\_\_

d) In which step of CALL is machine code first generated? \_\_\_\_\_

Compiler

Assembler

Linker

Loader