

DIGITAL DESIGN AND COMPUTER ARCHITECTURE (252-0028-00L), SPRING 2021  
OPTIONAL HW 3: MICROARCHITECTURE, ISA, AND PERFORMANCE EVALUATION

Instructor: Prof. Onur Mutlu

TAs: Juan Gomez-Luna, Jisung Park, Hasan Hassan, Mohammed Alser, Lois Orosa, Minesh Patel,  
Jawad Haj-Yahya, Haiyu Mao, Behzad Salami, Jeremie Kim, Giray Yaglikci, Can Firtina,  
Geraldo De Oliveira Junior, Rahul Bera, Konstantinos Kanellopoulos, Nika Mansouri, Gagandeep Singh

Released: Thursday, April 15, 2021

## 1 Big versus Little Endian Addressing

Consider the 32-bit hexadecimal number 0xcafe2b3a.

1. What is the binary representation of this number in *little endian* format? Please clearly mark the bytes and number them from low (0) to high (3).

2. What is the binary representation of this number in *big endian* format? Please clearly mark the bytes and number them from low (0) to high (3).

## 2 The MIPS ISA

### 2.1 Warmup: Computing a Fibonacci Number

The Fibonacci number  $F_n$  is recursively defined as

$$F(n) = F(n-1) + F(n-2),$$

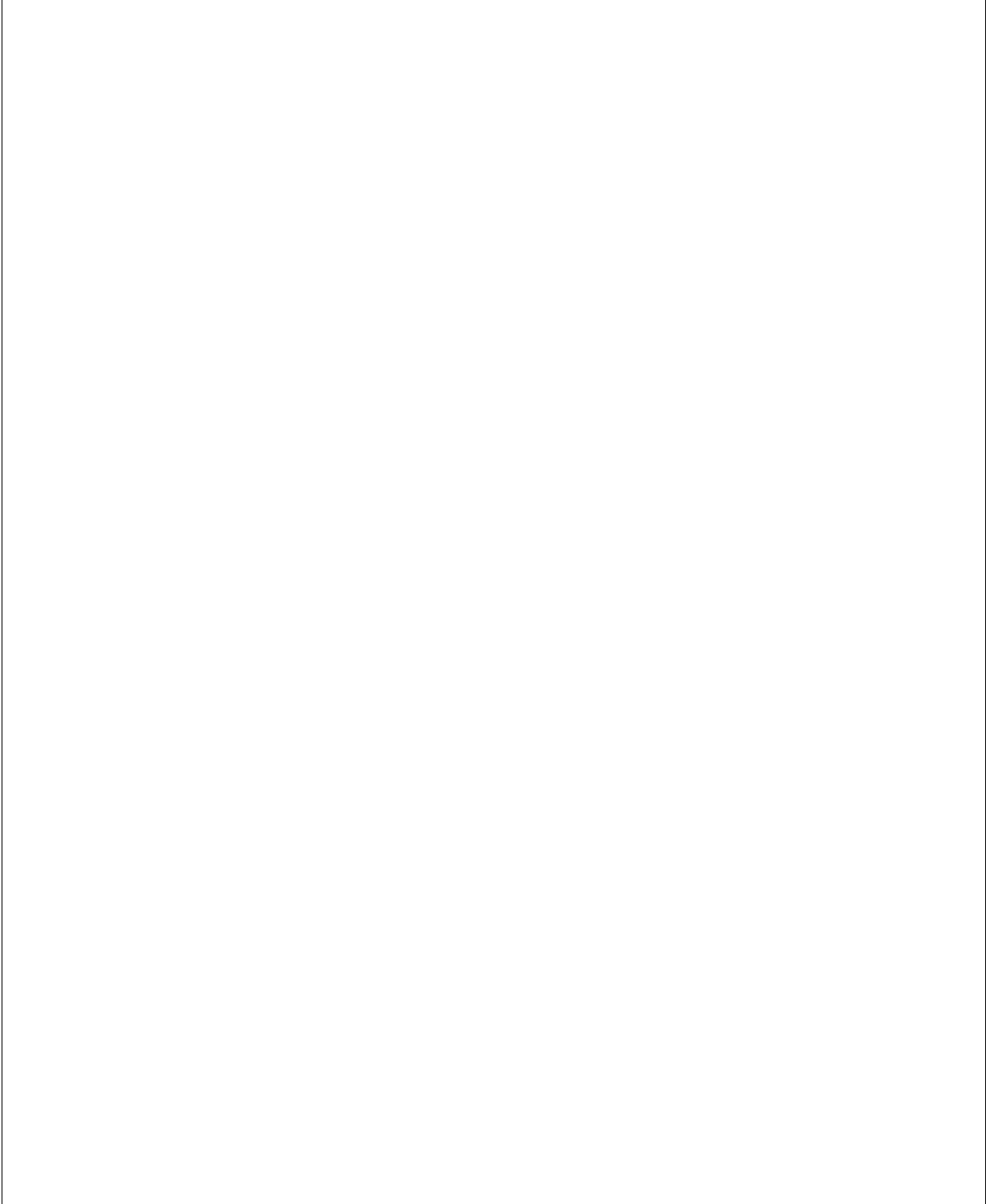
where  $F(1) = 1$  and  $F(2) = 1$ . So,  $F(3) = F(2) + F(1) = 1 + 1 = 2$ , and so on. Write the MIPS assembly for the `fib(n)` function, which computes the Fibonacci number  $F(n)$ :

```
int fib(int n)
{
    int a = 0;
    int b = 1;
    int c = a + b;
    while (n > 1) {
        c = a + b;
        a = b;
        b = c;
        n--;
    }
    return c;
}
```

Remember to follow MIPS calling convention and its register usage (just for your reference, you may not need to use all of these registers):

- The argument `n` is passed in register `$4`.
- The result (i.e., `c`) should be returned in `$2`.
- `$8` to `$15` are caller-saved temporary registers.
- `$16` to `$23` are callee-saved temporary registers.
- `$29` is the stack pointer register.
- `$31` stores the return address.

Note: A summary of the MIPS ISA is provided at the end of this handout.



## 2.2 MIPS Assembly for REP MOVSB

MIPS is a simple ISA. Complex ISAs—such as Intel’s x86—often use one instruction to perform the function of many instructions in a simple ISA. Here you will implement the MIPS equivalent for a single Intel x86 instruction, REP MOVSB, which is specified as follows.

The REP MOVSB instruction uses three fixed x86 registers: ECX (count), ESI (source), and EDI (destination). The “repeat” (REP) prefix on the instruction indicates that it will repeat ECX times. Each iteration, it moves one byte from memory at address ESI to memory at address EDI, and then increments both pointers by one. Thus, the instruction copies ECX bytes from address ESI to address EDI.

- (a) Write the corresponding assembly code in MIPS ISA that accomplishes the same function as this instruction. You can use any general purpose register. Indicate which MIPS registers you have chosen to correspond to the x86 registers used by REP MOVSB. Try to minimize code size as much as possible.

- (b) What is the size of the MIPS assembly code you wrote in (a), in bytes? How does it compare to REP MOVSB in x86 (note: REP MOVSB occupies 2 bytes)?

- (c) Assume the contents of the x86 register file are as follows before the execution of the REP MOVSB:

```
EAX: 0xccccaaaa
EBP: 0x00002222
ECX: 0xFEE1DEAD
EDX: 0xfeed4444
ESI: 0xdecaffff
EDI: 0xdeaddeed
EBP: 0xe0000000
ESP: 0xe0000000
```

Now, consider the MIPS assembly code you wrote in (a). How many total instructions will be executed by your code to accomplish the same function as the single REP MOVSB in x86 accomplishes for the given register state?

(d) Assume the contents of the x86 register file are as follows before the execution of the REP MOVSB:

```
EAX: 0xccccaaaa  
EBP: 0x00002222  
ECX: 0x00000000  
EDX: 0xfed4444  
ESI: 0xdecaffff  
EDI: 0xdeaddeed  
EBP: 0xe0000000  
ESP: 0xe0000000
```

Now, answer the same question in (c) for the above register values.

### 3 Dataflow (I)

Draw the data flow graph for the `fib(n)` function from Question 2.1. You may use the following data flow nodes in your graph:

- `+` (addition)
- `>` (left operand is greater than right operand)
- `Copy` (copy the value on the input to both outputs)
- `BR` (branch, with the semantics discussed in class, label the True and False outputs)

You can use constant inputs (e.g., 1) that feed into the nodes. Clearly label all the nodes, program inputs, and program outputs. Try to use the fewest number of data flow nodes possible.

## 4 Dataflow (II)

- We define the *switch node* in Figure 1 to have 2 inputs (**I**, **Ctrl**) and 1 output (**O**). The *Ctrl* input always enters perpendicularly to the switch node. If the *Ctrl* input has a *True* token (i.e., a token with a value of 1), the **O** wire propagates the value on the **I** wire. Else, the 2 input tokens (**I**, **Ctrl**) are consumed, and no token is generated at the output (**O**).
- We define the *inverter node* in Figure 2 to have 1 input (**I**) and 1 output (**O**). The node negates the input token (i.e.,  $O = !I$ ).
- We define the *TF node* in Figure 3 to have 3 inputs ( $I_F$ ,  $I_T$ , **Ctrl**) and 1 output (**O**). When **Ctrl** is set to True, **O** takes  $I_T$ . When **Ctrl** is set to False, **O** takes  $I_F$ .
- The  $\geq$  node outputs True only when the left input is greater than or equal to the right input.
- The  $+1$  node outputs the input plus one.
- The  $+$  node outputs the sum of the two inputs.
- A node generates an output token when tokens exist at *every* input, and *all* input tokens are consumed.
- Where a single wire splits into multiple wires, the token travelling on the wire is replicated to all wires.

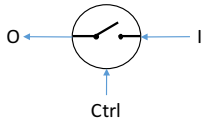


Figure 1: Switch Node

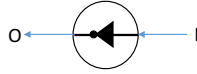


Figure 2: Inverter Node

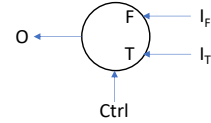
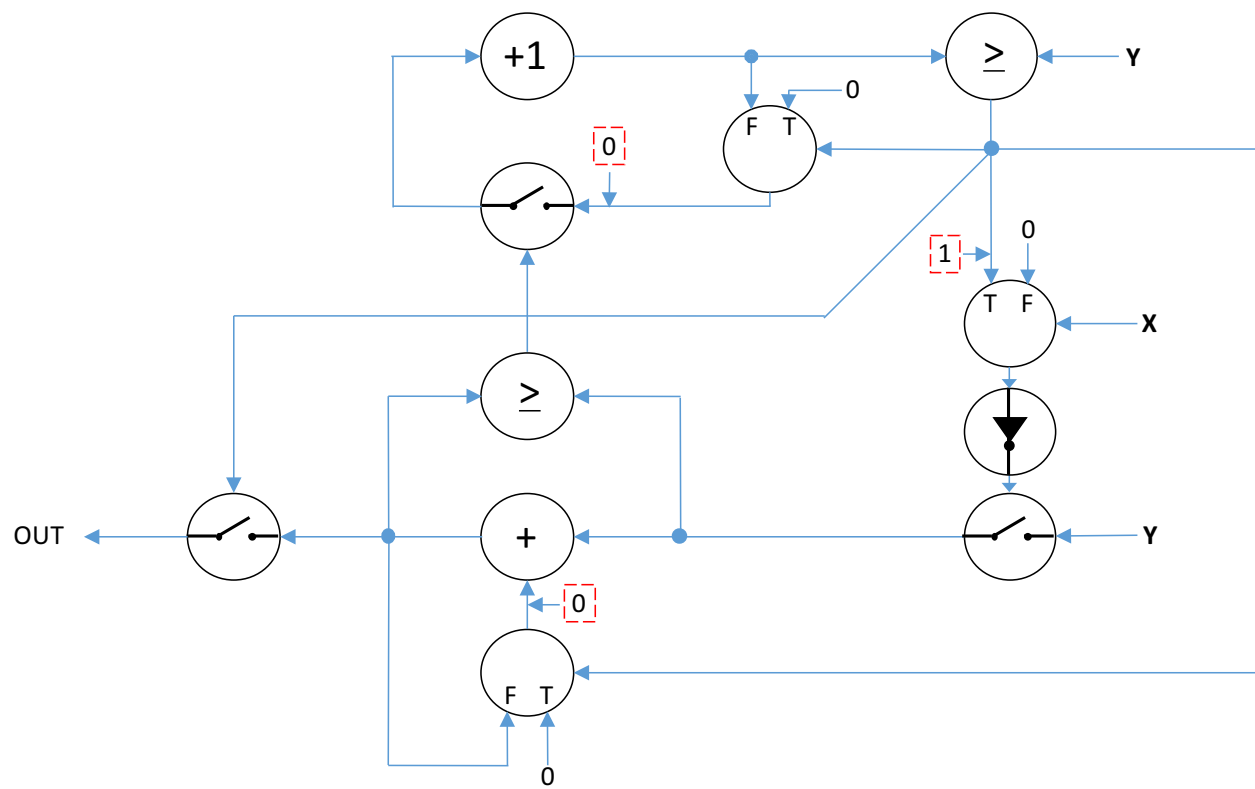


Figure 3: TF Node

Consider the dataflow graph on the following page. Numbers in dashed boxes represent tokens (with the value indicated by the number) in the initial state. The **X** and **Y** inputs automatically produce tokens as soon as the previous token on the wire is consumed. The order of these tokens follows the pattern (*note, the following are all single digit values spaced appropriately for the reader to easily notice the pattern*):

**X:** 0 01 011 0111 01111  
**Y:** 1 22 333 4444 55555

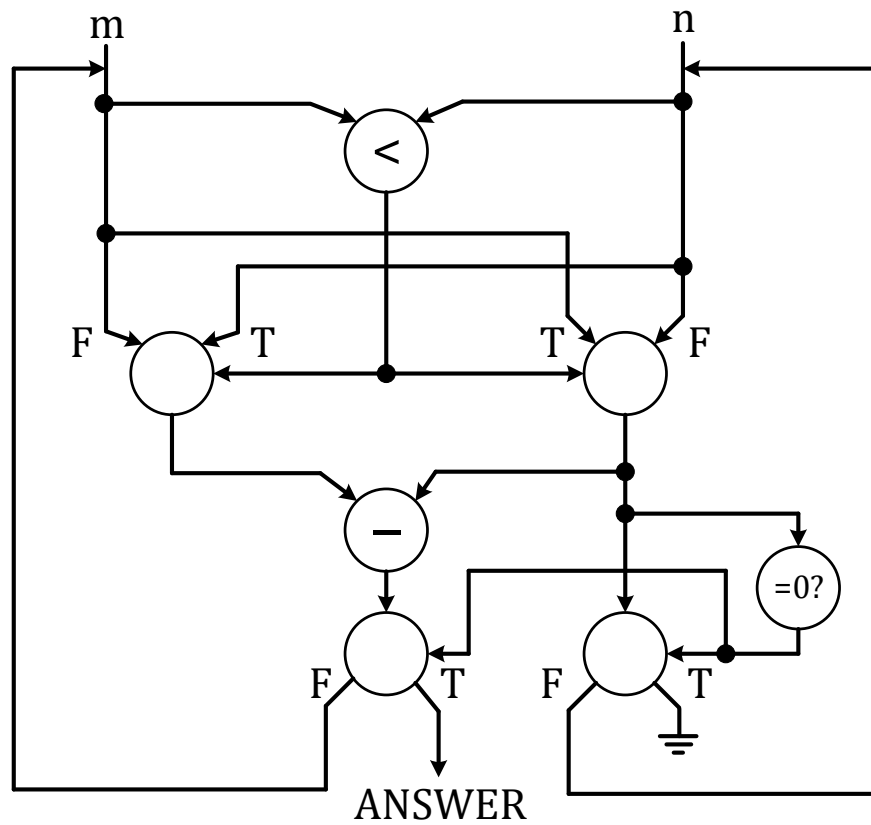
Consider the dataflow graph on the following page. Please clearly describe the sequence of tokens generated at the output (OUT).



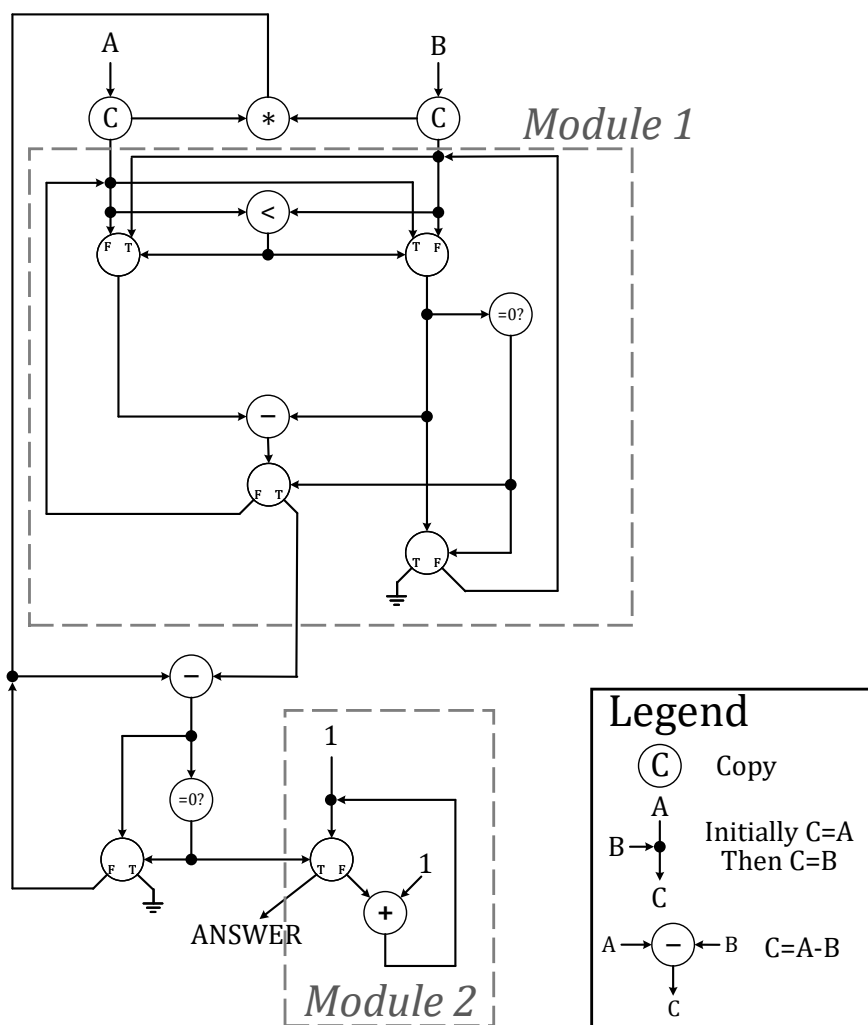


## 5 Dataflow (III)

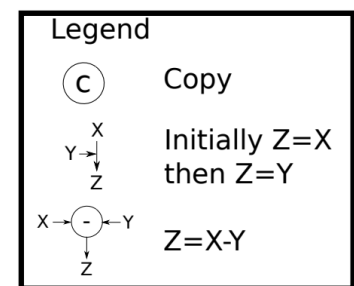
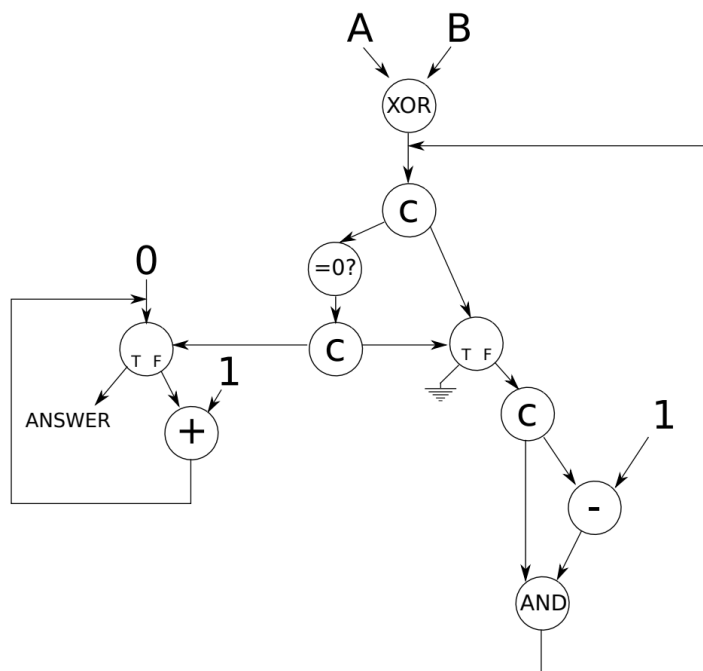
- (a) What does the following dataflow program do? Specify clearly in less than 10 words (one could specify this function in three words).



- (b) What does the following dataflow graph do (10 words or less)? (Hint: Identify what Module 1 and Module 2 perform.)



- (c) What does the following dataflow graph do (15 words or less)?  
 (Note that the inputs,  $A$  and  $B$ , are non-negative integers.)



## 6 Microarchitecture vs. ISA (I)

- a) Briefly explain the difference between the *microarchitecture* level and the *ISA* level in the transformation hierarchy. What information does the compiler need to know about the microarchitecture of the machine in order to compile a given program correctly?

- b) Classify the following attributes of a machine as either a property of its microarchitecture or ISA:

Microarchitecture?	ISA?	Attribute
		The machine does not have a subtract instruction
		The ALU of the machine does not have a subtract unit
		The machine does not have condition codes
		A 5-bit immediate can be specified in an ADD instruction
		It takes n cycles to execute an ADD instruction
		There are 8 general purpose registers
		A 2-to-1 mux feeds one of the inputs to ALU
		The register file has one input port and two output ports

## 7 Microarchitecture vs. ISA (II)

A new CPU has two comprehensive user manuals available for purchase as shown in Table 2.

Manual Title	Cost	Description
the_isa.pdf	CHF 1 million	describes the ISA in detail
the_microarchitecture.pdf	CHF 10 million	describes the microarchitecture in detail

Table 1: Manual Costs

Unfortunately, the manuals are extremely expensive, and you can only afford one of the two. If both manuals might be useful, you would prefer the cheaper one.

For each of the following questions that you would like to answer, decide which manual is more likely to help. *Note: we will subtract 1 point for each **incorrect** answer.*

1. The latency of a branch predictor misprediction.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

2. The size of a physical memory page.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

3. The memory-mapped locations of exception vectors.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

4. The function of each bit in a programmable branch-predictor configuration register.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

5. The bit-width of the interface between the CPU and the L1 cache.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

6. The number of pipeline stages in the CPU.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

7. The order in which loads and stores are executed by a multi-core CPU.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

8. The memory addressing modes available for arithmetic operations.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

9. The program counter width.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

10. The number of cache sets at each level of the cache hierarchy.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

## 8 ISA vs. Microarchitecture (III)

A new CPU has two comprehensive user manuals available for purchase as shown in Table 2.

Manual Title	Cost	Description
the_isa.pdf	CHF 1 million	describes the ISA in detail
the_microarchitecture.pdf	CHF 10 million	describes the microarchitecture in detail

Table 2: Manual Costs

Unfortunately, the manuals are extremely expensive, and you can only afford one of the two. If both manuals might be useful, you would prefer the cheaper one.

For each of the following questions that you would like to answer, decide which manual is more likely to help. *Note: we will subtract 1 point for each **incorrect** answer, and award 0 points for unanswered questions.*

1. The integer multiplication algorithm used by the ALU.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

2. The program counter width.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

3. Branch misprediction penalty.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

4. The ability to flush the TLB from the OS.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

5. The size of the Reorder Buffer in an Out-of-Order CPU.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

6. The fetch width of a superscalar CPU.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

7. SIMD instruction support.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

8. The memory addresses of the memory-mapped devices of the CPU (e.g., keyboard).

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

9. The number of non-programmable registers in the CPU.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

10. The replacement policy of the L1 data cache.

1. the\_isa.pdf                      2. the\_microarchitecture.pdf

11. The memory controller's scheduling algorithm.

1. `the_isa.pdf`                      2. `the_microarchitecture.pdf`
12. The number of bits required for the destination register of a load instruction.
1. `the_isa.pdf`                      2. `the_microarchitecture.pdf`
13. Description of the support for division and multiplication between integers.
1. `the_isa.pdf`                      2. `the_microarchitecture.pdf`
14. The mechanism to enter in a system call in the OS.
1. `the_isa.pdf`                      2. `the_microarchitecture.pdf`
15. The size of the addressable memory.
1. `the_isa.pdf`                      2. `the_microarchitecture.pdf`

## 9 Performance Metrics

- If a given program runs on a processor with a higher frequency, does it imply that the processor always executes more instructions per second (compared to a processor with a lower frequency)? (Use less than 10 words.)

- If a processor executes more of a given program's instructions per second, does it imply that the processor always finishes the program faster (compared to a processor that executes fewer instructions per second)? (Use less than 10 words.)



## 10 Performance Evaluation (I)

Your job is to evaluate the potential performance of two processors, each implementing a different ISA. The evaluation is based on its performance on a particular benchmark. On the processor implementing ISA *A*, the best compiled code for this benchmark performs at the rate of 10 IPC. That processor has a 500 MHz clock. On the processor implementing ISA *B*, the best compiled code for this benchmark performs at the rate of 2 IPC. That processor has a 600 MHz clock.

- What is the performance in Millions of Instructions per Second (MIPS) of the processor implementing ISA *A*?

- What is the performance in MIPS of the processor implementing ISA *B*?

- Which is the higher performance processor:    *A*    *B*    Don't know  
Briefly explain your answer.

## 11 Performance Evaluation (II)

You are the leading engineer of a new processor. Both the design of the processor and the compiler for it are already done. Now, you need to decide if you will send the processor to manufacturing at its current stage or if you will delay the production to introduce last-minute improvements to the design. To make the decision, you meet with your team to brainstorm about how to improve the design. Together, after profiling the target applications for the processor, you come up with two options:

- **Keep the current project.** For version A of the processor, the clock frequency is 600 MHz, and the following measurements are obtained:

Instruction Class	CPI	Frequency of Occurrence
A	2	40%
B	3	25%
C	3	25%
D	7	10%

- **Include optimizations to the design.** For version B of the processor, the clock frequency is 700 MHz. The ISA for processor B includes three new types of instructions. Those three new types of instructions increase the total number of executed instructions for processor B by 50%, in comparison to processor A. The following measurements are obtained:

Instruction Class	CPI	Frequency of Occurrence
A	2	15%
B	2	15%
C	4	10%
D	6	10%
E	1	10%
F	2	20%
G	2	20%

- (a) What is the CPI of each version? Show your work.

$CPI_A$ :

$CPI_B$ :

- (b) What are the MIPS (Million Instructions Per Second) of each version? Show your work.

$MIPS_A$ :

$MIPS_B$ :

- (c) Considering your team is aiming to release to the market the processor that gives better performance when executing the target application, which processor version will you choose as the final design? Show your work.

## 12 Performance Evaluation (III)

A multi-cycle processor  $P1$  executes *load instructions* in **10 cycles**, *store instructions* in **8 cycles**, *arithmetic instructions* in **4 cycles**, and *branch instructions* in **4 cycles**. Consider an application  $A$  where 20% of all instructions are load instructions, 20% of all instructions are store instructions, 50% of all instructions are arithmetic instructions, and 10% of all instructions are branch instructions.

- (a) What is the CPI of application  $A$  when executing on processor  $P1$ ? Show your work.

- (b) A new design of the processor doubles the clock frequency of  $P1$ . However, the latencies of the load, store, arithmetic, and branch instructions increase by 2, 2, 2, and 1 cycles, respectively. We call this new processor  $P2$ . The compiler used to generate instructions for  $P2$  is the same as for  $P1$ . Thus, it produces the same number of instructions for program  $A$ . What is the CPI of application  $A$  when executing on processor  $P2$ ? Show your work.

- (c) Which processor is faster ( $P1$  or  $P2$ )? By how much? Show your work.

- (d) There is some extra area available in the chip of processor  $P1$ , where extra hardware can fit. You can decide to include in your processor a faster branch execution unit or a faster memory device. The faster branch execution unit reduces the latency of branch instructions by a factor of 4. The memory device reduces the latency of the memory operations by a factor of 2. Which design do you choose? Show your work.

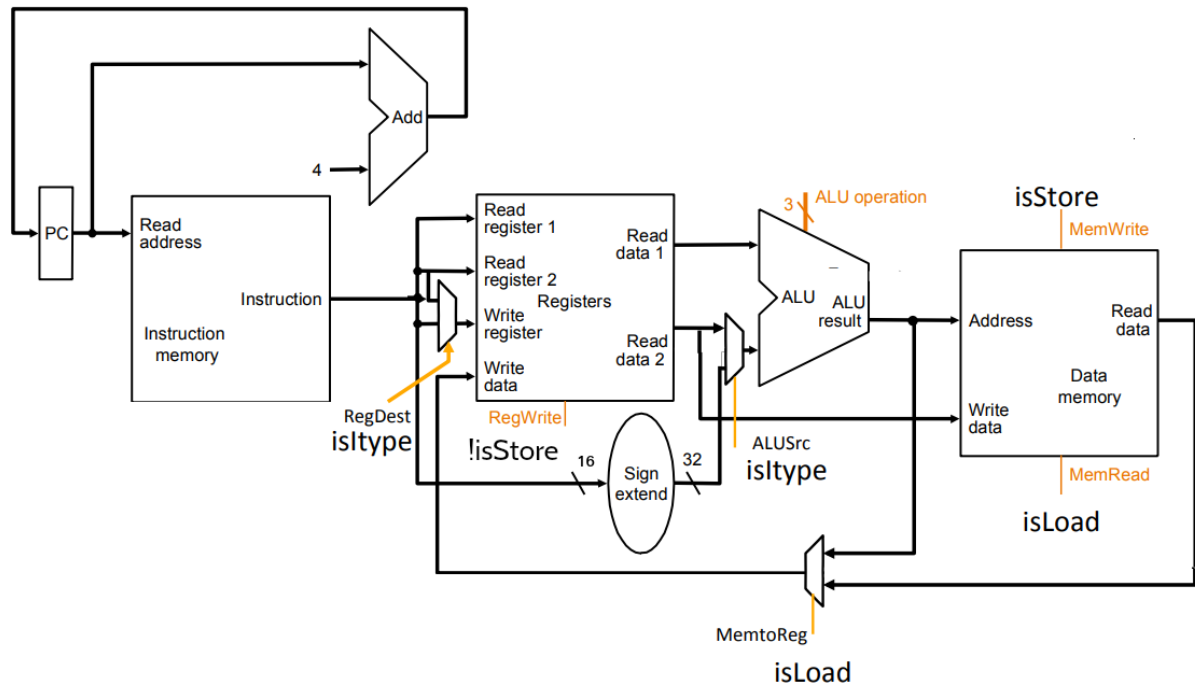
## 13 Single-Cycle Processor Datapath

In this problem, you will modify the single-cycle datapath we built up in Lecture 11 to support the JAL instruction. The datapath that we will start with is provided below. Your job is to implement the necessary data and control signals to support the JAL instruction, which we define to have the following semantics:

$$\text{JAL : } R31 \leftarrow PC + 4$$

$$PC \leftarrow PC_{31..28} \parallel \text{Immediate} \parallel 0^2$$

Add to the datapath on the next page the necessary data and control signals to implement the JAL instruction. Draw and label all components and wires very clearly (give control signals meaningful names; if selecting a subset of bits from many, specify exactly which bits are selected; and so on).



## 14 REP MOVSB

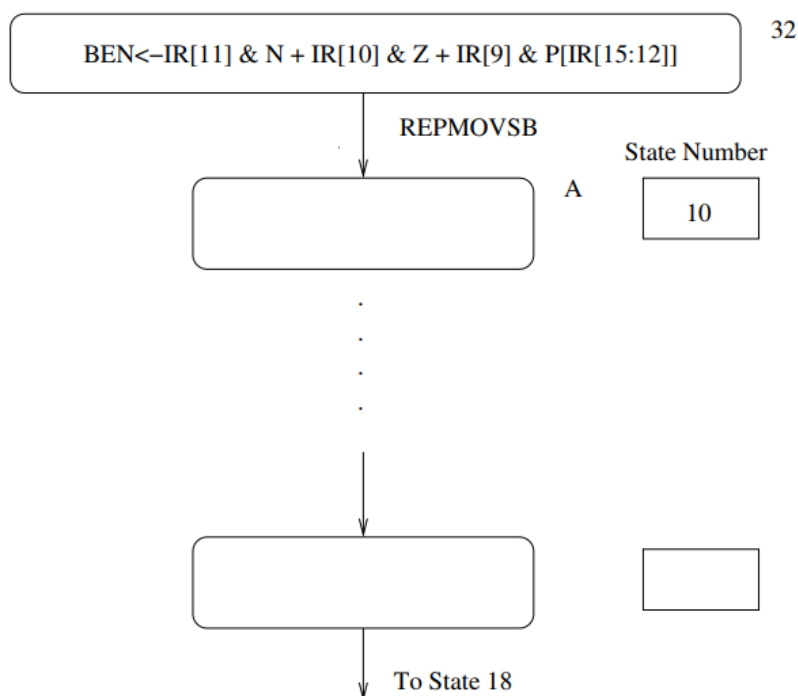
Let's say you are the lead architect of the next flagship processor at Advanced Number Devices (AND). You have decided that you want to use the LC-3b ISA for your next product, but your customers want a smaller semantic gap and marketing is on your case about it. So, you have decided to implement your favorite x86 instruction, REP MOVSB, in LC-3b.

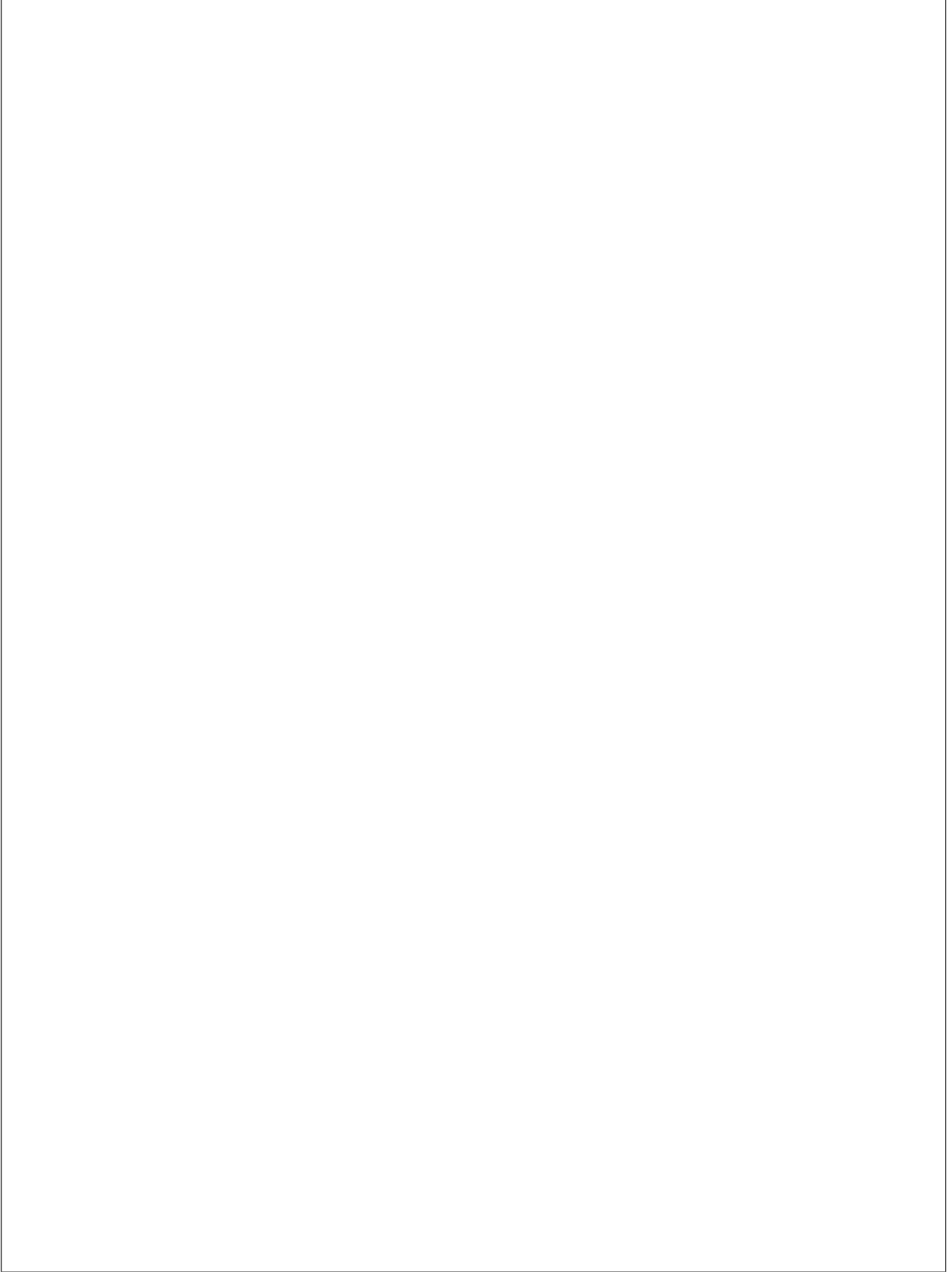
Specifically, you want to implement the following definition for REP MOVSB (in LC-3b parlance): REP-MOVSB SR1, SR2, DR which is encoded in LC-3b machine code as:

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1010				DR			SR1			0	0	0	SR2		

REPMOVSB uses three registers: SR1 (count), SR2 (source), and DR (destination). It moves a byte from memory at address SR2 to memory at address DR, and then increments both pointers by one. This is repeated SR1 times. Thus, the instruction copies SR1 bytes from address SR2 to address DR. Assume that the value in SR1 is greater than or equal to zero.

- (a) Complete the state diagram shown below, using the notation of the LC-3b state diagram. Describe inside each bubble what happens in each state and assign each state an appropriate state number. Add additional states not present in the original LC-3b design as you see fit.







- (b) Add to the LC-3b datapath any additional structures and any additional control signals needed to implement REPMOVS<sub>B</sub>. Clearly label your additional control signals with descriptive names. Describe what value each control signal would take to control the datapath in a particular way.

- (c) Describe any changes you need to make to the LC-3b microsequencer. Add any additional logic and control signals you need. Clearly describe the purpose and function of each signal and the values it would take to control the microsequencer in a particular way.

## MIPS Instruction Summary

Opcode	Example Assembly	Semantics
add	add \$1, \$2, \$3	$\$1 = \$2 + \$3$
sub	sub \$1, \$2, \$3	$\$1 = \$2 - \$3$
add immediate	addi \$1, \$2, 100	$\$1 = \$2 + 100$
add unsigned	addu \$1, \$2, \$3	$\$1 = \$2 + \$3$
subtract unsigned	subu \$1, \$2, \$3	$\$1 = \$2 - \$3$
add immediate unsigned	addiu \$1, \$2, 100	$\$1 = \$2 + 100$
multiply	mult \$2, \$3	hi, lo = $\$2 * \$3$
multiply unsigned	multu \$2, \$3	hi, lo = $\$2 * \$3$
divide	div \$2, \$3	lo = $\$2 / \$3$ , hi = $\$2 \bmod \$3$
divide unsigned	divu \$2, \$3	lo = $\$2 / \$3$ , hi = $\$2 \bmod \$3$
move from hi	mfhi \$1	$\$1 = \text{hi}$
move from low	mflo \$1	$\$1 = \text{lo}$
and	and \$1, \$2, \$3	$\$1 = \$2 \& \$3$
or	or \$1, \$2, \$3	$\$1 = \$2   \$3$
and immediate	andi \$1, \$2, 100	$\$1 = \$2 \& 100$
or immediate	ori \$1, \$2, 100	$\$1 = \$2   100$
shift left logical	sll \$1, \$2, 10	$\$1 = \$2 \ll 10$
shift right logical	srl \$1, \$2, 10	$\$1 = \$2 \gg 10$
load word	lw \$1, 100(\$2)	$\$1 = \text{memory}[\$2 + 100]$
store word	sw \$1, 100(\$2)	$\text{memory}[\$2 + 100] = \$1$
load upper immediate	lui \$1, 100	$\$1 = 100 \ll 16$
branch on equal	beq \$1, \$2, label	if ( $\$1 == \$2$ ) goto label
branch on not equal	bne \$1, \$2, label	if ( $\$1 \neq \$2$ ) goto label
set on less than	slt \$1, \$2, \$3	if ( $\$2 < \$3$ ) $\$1 = 1$ else $\$1 = 0$
set on less than immediate	slti \$1, \$2, 100	if ( $\$2 < 100$ ) $\$1 = 1$ else $\$1 = 0$
set on less than unsigned	sltu \$1, \$2, \$3	if ( $\$2 < \$3$ ) $\$1 = 1$ else $\$1 = 0$
set on less than immediate	sltui \$1, \$2, 100	if ( $\$2 < 100$ ) $\$1 = 1$ else $\$1 = 0$
jump	j label	goto label
jump register	jr \$31	goto \$31
jump and link	jal label	$\$31 = \text{PC} + 4$ ; goto label