Due on Thursday 14 February by 23:59′59″ EDT.

Getting Started To create your Lab 1 repository, visit the repository creation page at 6004.mit.edu/web/spring19/user/labs/lab1. Once your repository is created, you can clone it into your VM by running

```
git clone git@github.mit.edu:6004-spring19/labs-lab1-KERBEROS.git lab1
```

Turning in the Lab To turn in this lab, commit and push the changes you made to your git repository. After pushing, check the course website to verify that your submission passes all the tests. If you finish the lab in time but forget to push, you will incur the standard late submission penalties.

Grading the Lab To pass the lab, you must complete all sections and pass all tests. You must pass all labs to pass the course. However, you will only be graded on sections 3½ and 4.

Check-off Meeting After turning in this lab, you are required to go to the lab for a check-off meeting by Wednesday 20 February. See the course website for lab hours.

1 Introduction

**Most of this course** is dedicated to designing general-purpose, programmable machines — machines that can stream cat videos, solve differential equations, or display the front page of Reddit. But before we actually begin to design a programmable computer, we have to know how we want to program it. What things should it be able to do? Which operations will be designated by which inputs?

In 6.004, we’ll use an architecture\* called **RISC-V**. There are a few optional parts of RISC-V, but we’ll be focusing on the common denominator, called **RV32**. There is only one data type — a binary number called a **word** (1 word = 32 bits in RV32). You can only store up to 32 words at a time, in the so-called **registers**. If you want more, you have to load words from main memory and store them back in main memory when you’re done. Besides moving words around, you can do some basic arithmetic (like addition and subtraction), looping, and IF-THEN branching, but that’s it!

Let’s walk slowly through an example program. Three equivalent representations are shown on the next page; Python is on the left, RISC-V in human-readable form is in the middle, and RISC-V in binary is on the right.

\* An architecture, or more specifically an **instruction set architecture**, is a model of the way a computer is organized.
<table>
<thead>
<tr>
<th>Python</th>
<th>RISC-V (assembly)</th>
<th>RISC-V (binary)</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>n = 9</code></td>
<td><code>li a0, 9</code></td>
<td><code>0b: 0000000000100101010010111</code></td>
</tr>
<tr>
<td><code># check if n is even</code></td>
<td><code># check if n is even</code></td>
<td></td>
</tr>
<tr>
<td><code>if n % 2 == 0:</code></td>
<td><code>If:</code></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>andi a1, a0, 1</code></td>
<td><code>0c: 0100000010101010100010011</code></td>
</tr>
<tr>
<td></td>
<td><code>bnez a1, Else</code></td>
<td><code>08: 0000000001011000110001101</code></td>
</tr>
<tr>
<td></td>
<td><code>Then:</code></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>srai a0, a0, 1</code></td>
<td><code>0c: 0100000010101010100010011</code></td>
</tr>
<tr>
<td></td>
<td><code>j End</code></td>
<td><code>10: 0000001100000000000000011111</code></td>
</tr>
<tr>
<td>else:`</td>
<td><code>Else:</code></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>n = 3*n + 1</code></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>slli a2, a0, 1</code></td>
<td><code>14: 00000001011000101000001001</code></td>
</tr>
<tr>
<td></td>
<td><code>add a0, a2, a0</code></td>
<td><code>18: 00000011010001101000001011</code></td>
</tr>
<tr>
<td></td>
<td><code>addi a0, a0, 1</code></td>
<td><code>1c: 00000000101000001010000011</code></td>
</tr>
<tr>
<td></td>
<td><code>End:</code></td>
<td></td>
</tr>
<tr>
<td><code>...</code></td>
<td><code>...</code></td>
<td></td>
</tr>
<tr>
<td><code>...</code></td>
<td><code>...</code></td>
<td></td>
</tr>
</tbody>
</table>

A lot is going on here! Let's go through it line by line.

**li a0, 9**

The opcode **li** stands for *load immediate.* “Immediate” is just another term for “constant”, in this case 9, and **a0** is the name of the eleventh register, typically used to store the arguments of a function.† In English, this instruction means “load the constant 9 into register a0” – so we’ll be using a0 the way we use n in Python.

**If:**

This is a label. It doesn’t generate any instructions, but lets you name a specific place in the program.

**andi a1, a0, 1**

The opcode **andi** stands for *bitwise-AND with an immediate*. In English, this instruction means “take the bitwise-AND of [the contents of a0] and [the constant value 1], and place the result in register a1.” The immediate 1 is really being represented as a word (32 bits) inside the computer, so we are actually AND-ing with 0...01. The result that we place in a1 will be 1 if the number in a0 is odd, and 0 if the number in a0 is even.

**Check yourself:** Take a moment to figure out why this is the case!

---

* This is actually a pseudo-instruction; it is shorthand for **addi ao, zero, 9**. The register **zero** always contains zero.
† The thirty-two registers are numbered 0–31, and we could have referred to a0 by writing x10 instead.
To keep the words from getting tedious, from now on we'll use a right arrow ⇒ when we're describing the effect of an instruction, and the running prose will just be a commentary.

**bnez a1, Else**  
The opcode *bnez* stands for *branch if not equal to zero.* It can cause the computer to skip ahead (or back) to a different part of the program. Several instruction types do this; the term *branch* is used when the skip is conditional. The term *jump* is used otherwise.

⇒ CONTINUE to Then if the number in *a0* is even (*a1* = 0)

⇒ GO TO Else if the number in *a0* is odd (*a1* = 1)

Then:

This is just another label.

**srai a0, a0, 1**  
*Arithmetic right shift by immediate.* An arithmetic shift by 1 turns the number *abc...xyz* into *aab...wxy* by duplicating the most-significant bit (MSB) *a* and dropping the least-significant bit (LSB) *z*, whereas a *logical* shift would turn the same number into *eab...wxy.*† Since numbers are represented in binary, shifting the digits right by one has the effect of dividing by 2.

⇒ DIVIDE the number in *a0* by 2 and place the result in *a0*

**j End**   
⇒ GO TO **End**

Else:

**slli a2, a0, 1**  
*Logical left shift by immediate.*

⇒ MULTIPLY the number in *a0* by 2 and place the result in *a2*

**add a0, a2, a0**  
⇒ ADD the number in *a0* to the number in *a2* (≡ 2×*a0*) and place the result (≡ 3×*a0*) back in *a0*

**addi a0, a0, 1**  
⇒ ADD 1 to the number in *a0* and place the result in *a0*

End:  
We're done! The register *a0* now contains *a0* ÷ 2 if *a0* was even and *a0*×3 + 1 if *a0* was odd.

---

* This is also a pseudo-instruction; it is shorthand for *bne a1, zero, Else.* The register zero always contains zero.
† When the contents of a register represent a signed binary number, arithmetic right shifts preserve the sign of the number, whereas logical right shifts do not.
The Collatz conjecture states that for any $n_0$ the sequence $n_0, n_1, n_2, \ldots$ eventually reaches 1, where each $n_i$ after $i = 0$ is defined as

$$n_i = \begin{cases} \frac{n_{i-1}}{2} & \text{if } n_{i-1} \text{ is even} \\ 3n_{i-1} + 1 & \text{if } n_{i-1} \text{ is odd} \end{cases}$$

**Check yourself:** How could you modify this program so that it tests the Collatz conjecture?

In Python, the Collatz test might look something like this:

```python
n = 9
while n != 1:
    if n % 2 == 0:
        n = n / 2
    else:
        n = 3*n + 1
```

This is possible using only the instructions we've seen so far, but if you'd like to use others, take a look at the appendices.

### 2 The 9th of September, 1947

In this section we're going to introduce tools that help us run and debug our assembly.

The lab repository contains a copy of the code from page 2 (in a file called `collatz.S`). Also in the repo is `assemble-collatz.sh`, which will first assemble the program into binary and then generate a human-readable version.* To run it, open up the terminal and navigate to the lab repo; then enter the two following commands:

```
student@6.004-VM:~$ cd lab1
student@6.004-VM:~/lab1$ ./assemble-collatz.sh
student@6.004-VM:~/lab1$ cat collatz.dump
```

At the top of the output (just beneath “Contents of section .text.init”) is the hex representation of the RISC-V binary. Under that (beneath “Disassembly of section .text.init”) is an instruction-by-instruction listing of the program. The first five lines of the disassembly are reproduced at the top of the next page.

---

* Technically, `assemble-collatz` doesn’t generate an executable binary file, but something similar. One more step would turn it into something you could actually run on a RISC-V processor.
On the left (0:) is the address of each instruction. Beside that (00900513) is the hex representation of the instruction itself. On the right (li a0, 9) is the assembly. Notice that the addresses are numbered by fours (0, 4, 8, c, 10, 14, ...). The RISC-V address space is byte-indexed; and since each instruction is 32 bits, but a byte is 8 bits, each instruction spans four indices.

Fundamentally, all programs that your computer will ever run look like this. They may start an address other than 0, they might get moved around if they're included inside another program, and they may even be interrupted and later resumed. But your computer really does work by only doing many (many, many) small operations like these!

2½ Federal Regulations Require Me to Warn You that This Next Test... Is Looking Pretty Good

For the rest of 6.004, we'll be using test suites written by the staff, for two purposes. First, tests help us verify that our programs match the spec – in other words, that they actually do what we think they do. Second, the tests are a method for the staff to prove that you completed the lab (to complement the less exhaustive but more conceptual checkoffs).

The test framework will wrap each of your programs with extra code. If you take a look at a .dump file after assembling your programs, you'll see the extra instructions (with your work sandwiched in the middle). To generate the files necessary for testing your programs, simply enter make, like so:

```
student @ 6.004-VM:~/lab1 $ make
```

Anytime you change your code, be sure to re-run make.

Each lab release includes a RISC-V simulator; this is how you will run the tests. It also can help you debug your programs – plus give you a chance to see your code in action!

To get collatz to work with the simulator, you'll need to make some small tweaks. Fire up your editor of choice and delete the line containing li a0, 9 (but not the line containing the label collatz!) Then, add the instruction ret after the label End. Be sure to run make!
To start the simulator, enter python sim_gui.pyw, like so:

    student@6.004-VM:~/lab1 $ python sim_gui.pyw

On the left under **Program** select **collatz**, and under **Test Case** select **test 1**. Then, at the top, press **Load Program**.

**Note:** To select a different program or a different test, you must first press **Exit Program**. If you modify a program, you must re-run **make** and then re-load in the simulator.

If you press **Run**, you should see

```
$ Load program collatz with test 1
$ Run:
  10:Passed
Lab1.collatz1: PASSED
```

This is great news! Except it’s not very informative. If our program encountered an error or had a bug in it, all we would see is **FAILED**. We want something that gives us a better idea of what’s going on.

Hit **Exit Program** and then **Load Program** to reset the simulator. This time, press **Step** a few times. As you press **Step**, you should see **PC = 0x...** change and **Instrs executed = ...** begin to increment. “**pc**” stands for **program counter**; it is the address of the instruction that the simulator is about to execute. The first four dozen or so steps are setup for the test framework, but if you pressed **Step** four dozen times you would reach the code for **collatz**. That would be very tedious, so there’s an additional feature that lets you run the simulation until **PC** reaches particular addresses. These are called **break points**.

*You should **Exit** and **Load** again before adding break points.*

Open up **collatz.dump** and locate the first instruction from **collatz.S** (try searching for the symbol `<collatz>`). In the simulator, enter the address of this instruction under **Break Points** (at the bottom left) as a hexadecimal number with a **0x-** prefix, and then press **Apply**.

Now when you hit **Run**, the simulator will automatically step until **PC** equals the address you just entered; from this point you can step manually.

The default tab (**Console Output**) is not particularly useful since our programs don’t print anything yet. Instead, select **Registers** and note the current values of **a0**, **a1**, and **a2**.
With the simulator and source (or disassembly) open side-by-side, step through collatz. Before each step, predict how \(a_0\), \(a_1\), \(a_2\), and \(pc\) will change.

**Check yourself:** In test 1, does the simulation branch to <Else> (the case when \(a_0\) is initially odd) or continue on to <Then> (the case when \(a_0\) is initially even)? How about in test 2?

3 Μείζων δὲ τούτων ἐστίν …;

In this section we're going to **A** learn how to access main memory, and **B** get more practice by finishing another program. (After that, you'll write your first complete RISC-V program!)

At the top of the next page is a Python function designed to find the largest word in an array. An **array** is a collection of elements (such as bytes, words, double words, &c) that are evenly spaced. As a result, the location of the \(n\)-th element is always

\[
\text{array starting address} + (n \times \text{constant})
\]

For an array of words, the \(n\)-th word in the array is located at \((\text{start of array} + 4n)\), since RISC-V memory is indexed by byte and each word is four bytes long.

The function takes two arguments: \(p\), which is the start of the array,\(^*\) and \(n\), which is the number of elements in the array. To make our Python function a little more like assembly (so that it's easier to translate), we'll imagine that we have memory arranged just like it is in RISC-V (as a really, really long sequence of bytes).

\[
\text{Memory} = \# \text{ lots and lots of bytes}
\]

```python
def load_word(addr):
    return int(Memory[addr:addr+4])
```

* Traditionally called the **pointer** to the array.
Idiomatic Python

```python
def maximum(p, n):
    largest_so_far = 0
    for i in range(n):
        w = load_word(p + 4*i)
        if w > largest_so_far:
            largest_so_far = w
    return largest_so_far
```

Alternate Version (more similar to RISC-V)

```python
def maximum(p, n):
    largest_so_far = 0
    while n != 0:
        w = load_word(p)
        if w > largest_so_far:
            largest_so_far = w
        p = p + 4
        n = n - 1
    return largest_so_far
```

Our assembly version of this program will be subject to these constraints:

A  When our function is called, the argument `p` (the starting address of the array) will be stored in `a0` and the argument `n` (the size of the array) will be stored in `a1`.

B  Our function should store its result (the largest word in the array) in `a0` and end with a `ret` instruction.

C  Our function should only use the `a0-a7` registers.

One good recipe for translating Python code into RISC-V is the following four-step process. You'll be responsible for most of step 2!

1  Decide which registers will be used for which variables. It was stipulated that `p ←→ a0` and `n ←→ a1`. Let's (arbitrarily) decide that `largest_so_far ←→ a2` (which we'll copy to `a0` when we're done) and `w ←→ a3`.

2  Convert variable assignments and mathematical operations into register-register arithmetic and load/store instructions.

3  Convert if/else statements into branches.

4  Convert loops (for, while, &c) into branches and jumps.

There are several ways to implement loops; the next page demonstrates two of doing step 4. The ellipses (...) indicate an instruction that you will eventually fill in.
Condition at Bottom

maximum:
   li a2, 0  # load immediate 0 into a2 (largest_so_far)
loop:
   lw a3, 0(a0)  # load word at [[address stored in a0] + 0] into a3 (w)
   ble a3, a2, continue  # skip next instruction if a3 (w) ≤ a2 (largest_so_far)
   ...
continue:
   ...
   ...
   bnez a1, loop  # go to top of loop if a1 (n) ≠ 0
done:
   ...
   ...
   j loop  # go to top of loop

Condition at Top

maximum:
   li a2, 0  # load immediate 0 into a2 (largest_so_far)
loop:
   beqz a1, done  # go to <done> if a1 (n) = 0
   lw a3, 0(a0)  # load word at [[address stored in a0] + 0] into a3 (w)
   ble a3, a2, continue  # skip next instruction if a3 (w) ≤ a2 (largest_so_far)
   ...
continue:
   ...
   ...
   j loop  # go to top of loop
done:
   ...
   ...
The really new feature here is the `lw a3, 0(a0)` instruction. The opcode `lw` stands for load word. It gives us a way to move data from main memory (which is spacious but slow) to a register, so that we can do arithmetic with it. It works like this:

```
lw rd, imm(rs)
```

Suppose the content of the register `rs` is some number `n`. Then the word at address `(n + imm)` in memory will be loaded into register `rd`.

For example, if the value of `rs` was `0x00040100` and `imm` was 8, then the word located at address `0x00040108` will be put into `rd`.

To put a word back into memory, we can use the `sw` instruction, which works similarly.

```
sw rs2, imm(rs1)
```

Suppose the content of the register `rs1` is some number `n`. Then the word in register `rs2` will be copied into memory at address `(n + imm)`.

If you open up `maximum.S`, you'll notice that the staff have already written the load instruction for you. This gives you an opportunity to see it before you use it; in the last part of the lab, you'll have to write your own loads and stores.

**Complete maximum.S and get it to pass all of the tests in the simulator.**

---

**3½ A000217 30 points**

The triangular numbers count the number of objects in arrangements that look like triangles. The first few are 1, 3, 6, and 10.

If you open up `triangular.S`, you'll find a Python implementation of a function that computes the `n`-th triangular number given `n`. Your task is to translate this function into assembly. Your version will be subject to these constraints:

A. When the function is called, the argument `n` will be stored in `a0`.

B. The function should store its result in `a0` and end with a `ret` instruction.

C. The function should only use the `a0-a7` registers.

**Complete triangular.S and get it to pass all of the tests in the simulator.**
BUBBLE SORT – so named because smaller* elements float up to the top† of the array over time if you visualize it – is a relatively simple algorithm that works by swapping adjacent elements until everything is in order. The file bubblesort.S contains a template for bubble sort; inside is a blank function called sort. Write this function in RV32 assembly, subject to the following constraints:

A. The array to sort is stored in memory in a contiguous range of addresses, and is passed to sort using two arguments. When sort is called, the starting address of the array will be stored in a0 and the number of elements in the array will be stored in a1.

B. The sort function should modify the array directly, and need not return any value.

C. The sort function should only use the a0-a7 registers.

D. Your implementation should only use the subset of RV32 that we use in this course. This includes all its instructions except sub-word loads and stores (lb, lbu, lh, lhu, lb, sb, sh) and auipc. The simulator only supports this subset. Using an instruction outside this subset will cause an illegal instruction exception.

The file bubblesort.S also contains Python and C implementations of the Bubble sort algorithm to serve as the basis for your work.

As before, assemble your code by running make and execute the generated binary by running python sim_gui.pyw. Alternatively, you can run the tests by entering

```
student@6.004-VM:~/lab1 $ python sim.py bubblesort
```

To run tests one a time, enter

```
student@6.004-VM:~/lab1 $ python sim.py bubblesort k
```

where k is 1, 2, 3, 4, or 5. For more information on the command line interface to the simulator, see the appendices.

**Discussion Questions:** Be prepared to walk through collatz or maximum with a staff member during checkoff, and predict how the register contents will change after each instruction before actually stepping the simulator.

* Which apparently = lighter.
† It’s also called sinking sort for the opposite reason.
5 Appendix: RISC-V ISA Reference

For this lab, you should only use the subset of RV32I instructions presented in this appendix. Using unsupported instructions (sub-word loads and stores and AUIPC) will cause errors when you try to simulate the program (using sim.py or sim_gui.pyw). You are encouraged to use pseudo-instructions to simplify your code.

MIT 6.004 ISA Reference Card: Instructions

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Syntax</th>
<th>Description</th>
<th>Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUI</td>
<td>lui rd, immU</td>
<td>Load Upper Immediate</td>
<td>reg[rd]  &lt;= immU &lt;&lt; 12</td>
</tr>
</tbody>
</table>
| JAL         | jal rd, immJ       | Jump and Link           | reg[rd]  <= pc + 4  
|             |                    |                         | pc  <= pc + immJ    |
| JALR        | jalr rd, rs1, immJ | Jump and Link Register  | reg[rd]  <= pc + 4  
|             |                    |                         | pc  <= \{(reg[rs1] + immJ)[31:1], 1'b0\} |
| BEQ         | beq rs1, rs2, immB | Branch if =             | pc  <= (reg[rs1] == reg[rs2]) ? pc + immB  
|             |                    |                         | : pc + 4            |
| BNE         | bne rs1, rs2, immB | Branch if ≠             | pc  <= (reg[rs1] != reg[rs2]) ? pc + immB  
|             |                    |                         | : pc + 4            |
| BLT         | blt rs1, rs2, immB | Branch if < (Signed)    | pc  <= (reg[rs1] <= s reg[rs2]) ? pc + immB  
|             |                    |                         | : pc + 4            |
| BGE         | bge rs1, rs2, immB | Branch if ≥ (Signed)    | pc  <= (reg[rs1] >= s reg[rs2]) ? pc + immB  
|             |                    |                         | : pc + 4            |
| BLTU        | bltu rs1, rs2, immB| Branch if < (Unsigned)  | pc  <= (reg[rs1] < u reg[rs2]) ? pc + immB  
|             |                    |                         | : pc + 4            |
| BGEU        | bgeu rs1, rs2, immB| Branch if ≥ (Unsigned)  | pc  <= (reg[rs1] >= u reg[rs2]) ? pc + immB  
|             |                    |                         | : pc + 4            |
| LW          | lw rd, immI(rs1)   | Load Word               | reg[rd]  <= mem[reg[rs1] + immI]           |
| SW          | sw rs2, immS(rs1)  | Store Word              | mem[reg[rs1] + immS] <= reg[rs2]           |
| ADDI        | addi rd, rs1, imm1 | Add Immediate           | reg[rd] <= reg[rd] + imm1                  |
| SLTI        | slti rd, rs1, imm1 | Compare < Immediate (Signed) | reg[rd] <= (reg[rs1] < s imm1) ? 1 : 0 |
| SLTIU       | sltiu rd, rs1, imm1| Compare < Immediate (Unsigned) | reg[rd] <= (reg[rs1] < u imm1) ? 1 : 0 |
| XORI        | xorI rd, rs1, imm1 | Xor Immediate           | reg[rd] <= reg[rd] ^ imm1                   |
| ORI         | ori rd, rs1, imm1  | OrImmediate             | reg[rd] <= reg[rd] | imm1       |
| ANDI        | andi rd, rs1, imm1 | AndImmediate            | reg[rd] <= reg[rd] & imm1                  |
| SLLI        | slli rd, rs1, imm1 | Shift Left Logical Immediate | reg[rd] <= reg[rd] << imm1               |
| SRLI        | srlI rd, rs1, imm1 | Shift Right Logical Immediate | reg[rd] <= reg[rd] >> u imm1           |
| SRAI        | srai rd, rs1, imm1 | Shift Right Arithmetic Immediate | reg[rd] <= reg[rd] >> imm1                 |
| ADD          | add rd, rs1, rs2   | Add                      | reg[rd] <= reg[rd] + reg[rs2]               |
| SUB          | sub rd, rs1, rs2   | Subtract                 | reg[rd] <= reg[rd] - reg[rs2]               |
| SLL          | sll rd, rs1, rs2   | Shift Left Logical       | reg[rd] <= reg[rd] << reg[rs2]             |
| SRL          | srl rd, rs1, rs2   | Shift Right Logical      | reg[rd] <= reg[rd] >> u reg[rs2]            |
| SRA          | sra rd, rs1, rs2   | Shift Right Arithmetic   | reg[rd] <= reg[rd] >> reg[rs2]             |
| OR           | or rd, rs1, rs2    | Or                       | reg[rd] <= reg[rd] | reg[rs2]   |
| AND          | and rd, rs1, rs2   | And                      | reg[rd] <= reg[rd] & reg[rs2]              |
### MIT 6.004 ISA Reference Card: Pseudoinstructions

<table>
<thead>
<tr>
<th>Pseudoinstruction</th>
<th>Description</th>
<th>Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>li rd, constant</code></td>
<td>Load Immediate</td>
<td><code>reg[rd] &lt;= constant</code></td>
</tr>
<tr>
<td><code>mv rd, rs1</code></td>
<td>Move</td>
<td><code>reg[rd] &lt;= reg[rs1] + 0</code></td>
</tr>
<tr>
<td><code>not rd, rs1</code></td>
<td>Logical Not</td>
<td><code>reg[rd] &lt;= reg[rs1] ^ ~1</code></td>
</tr>
<tr>
<td><code>neg rd, rs1</code></td>
<td>Arithmetic Negation</td>
<td><code>reg[rd] &lt;= 0 - reg[rs1]</code></td>
</tr>
<tr>
<td><code>j label</code></td>
<td>Jump</td>
<td><code>pc &lt;= label</code></td>
</tr>
<tr>
<td><code>jal label</code></td>
<td>Jump and Link (with ra)</td>
<td><code>reg[ra] &lt;= pc + 4; pc &lt;= label</code></td>
</tr>
<tr>
<td><code>jr rs</code></td>
<td>Jump Register</td>
<td><code>pc &lt;= reg[rs1] &amp; ~1</code></td>
</tr>
<tr>
<td><code>jalr rs</code></td>
<td>Jump and Link Register (with ra)</td>
<td><code>reg[ra] &lt;= pc + 4; pc &lt;= reg[rs1] &amp; ~1</code></td>
</tr>
<tr>
<td><code>ret</code></td>
<td>Return from Subroutine</td>
<td><code>pc &lt;= reg[ra]</code></td>
</tr>
<tr>
<td><code>bgt rs1, rs2, label</code></td>
<td>Branch &gt; (Signed)</td>
<td><code>pc &lt;= (reg[rs1] &gt; s reg[rs2]) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>ble rs1, rs2, label</code></td>
<td>Branch ≤ (Signed)</td>
<td><code>pc &lt;= (reg[rs1] ≤ s reg[rs2]) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>bgtu rs1, rs2, label</code></td>
<td>Branch &gt; (Unsigned)</td>
<td><code>pc &lt;= (reg[rs1] &gt; u reg[rs2]) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>bleu rs1, rs2, label</code></td>
<td>Branch ≤ (Unsigned)</td>
<td><code>pc &lt;= (reg[rs1] ≤ u reg[rs2]) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>beqz rs1, label</code></td>
<td>Branch = 0</td>
<td><code>pc &lt;= (reg[rs1] == 0) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>bnez rs1, label</code></td>
<td>Branch ≠ 0</td>
<td><code>pc &lt;= (reg[rs1] != 0) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>bltz rs1, label</code></td>
<td>Branch &lt; 0 (Signed)</td>
<td><code>pc &lt;= (reg[rs1] &lt; s 0) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>bgez rs1, label</code></td>
<td>Branch ≥ 0 (Signed)</td>
<td><code>pc &lt;= (reg[rs1] ≥ s 0) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>bgtz rs1, label</code></td>
<td>Branch &gt; 0 (Signed)</td>
<td><code>pc &lt;= (reg[rs1] &gt; s 0) ? label : pc + 4</code></td>
</tr>
<tr>
<td><code>blez rs1, label</code></td>
<td>Branch ≤ 0 (Signed)</td>
<td><code>pc &lt;= (reg[rs1] ≤ s 0) ? label : pc + 4</code></td>
</tr>
</tbody>
</table>
6 Appendix: RISC-V Simulator Reference

This appendix details how to interact with the RISC-V simulator to debug your code. We provide two simulator versions, one with a command line interface (CLI) and one with a graphical user interface (GUI).

Before getting started, make sure that you have run ‘make’ successfully. For lab 1, check that the file `bubblesort.vmh` exists.

6.1 Command Line Interface

**Start the simulator** using the following command:

```
python -i sim.py <program_name> <test_number>
```

For lab 1, `<program_name>` is `bubblesort`, and `<test_number>` is the argument to the program, which should be an integer from 1 to 5. Omitting the `<test_number>` will run all test cases sequentially. The ‘-i’ option denotes you want to run in interactive mode. If you omit it, the simulator will simply run the program to completion.

**List all supported commands** by typing ‘help()’ after you start the simulator in interactive mode. The simulator will list all usable commands together with an example:

```
>>> help()
<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>run()</td>
<td>Run program until it either finishes or reaches a break point.</td>
<td></td>
</tr>
<tr>
<td>step(&lt;n&gt;)</td>
<td>Run &lt;n&gt; more instructions and stop.</td>
<td></td>
</tr>
<tr>
<td>setBps([&lt;pc0&gt;, ...])</td>
<td>Clear previous break points and set new ones.</td>
<td></td>
</tr>
<tr>
<td>addBps([&lt;pc0&gt;, ...])</td>
<td>Add break points.</td>
<td></td>
</tr>
<tr>
<td>clearBps([&lt;pc0&gt;, ...])</td>
<td>Clear break points.</td>
<td></td>
</tr>
<tr>
<td>showStats()</td>
<td>List statistics, e.g. instruction counts</td>
<td></td>
</tr>
<tr>
<td>showRegs()</td>
<td>List registers including pc register</td>
<td></td>
</tr>
<tr>
<td>showMem(&lt;start&gt;, &lt;end&gt;)</td>
<td>List memory layout from address &lt;start&gt; to &lt;end&gt;</td>
<td></td>
</tr>
<tr>
<td>showStack()</td>
<td>List memory layout of the stack specified by register x2(sp)</td>
<td></td>
</tr>
<tr>
<td>showInstr(&lt;pc&gt;)</td>
<td>Display the instruction at &lt;pc&gt;</td>
<td></td>
</tr>
<tr>
<td>showLablePC(&lt;label&gt;)</td>
<td>Display the PC of &lt;label&gt;</td>
<td></td>
</tr>
<tr>
<td>restart()</td>
<td>Restart the execution.</td>
<td></td>
</tr>
</tbody>
</table>

**Break points** let you interrupt program execution before a particular instruction executes. They are a useful debugging tool, as they let you inspect the state of the machine at specific points of interest. ?? discusses how to locate the address of a specific instruction to set a break point.

**Example:**

```
$ python -i sim.py bubblesort 1   # Run bubblesort.vmh with test case 1
>>> setBps([0x38])            # Set a break point at 0x38
>>> run()                    # Stop before executing the instruction at 0x38
>>> run()                    # Stop before executing the instruction at 0x38
>>> showRegs()               # Print out all registers including PC
>>> clearBps()               # Clear all break points
>>> run()                    # Run to completion
>>> restart()                # Restart the machine with the same program and the same test case(s)
>>> step()                   # Execute only one instruction
>>> exit()                   # Quit
```
6.2 Graphical User Interface

**Start the simulator** using the following command:

```python
python sim_gui.pyw
```

The GUI has similar functionality to that of the CLI.

**Example:** Select the program and the test case you want to run. Then click the ‘Load Program’ button. Click ‘Run’ to let the machine execute the program to completion. Click ‘Step’ to let the machine execute the next instruction only. If a break point is reached, the execution will stop.

You can optionally set and modify break points before clicking ‘Run’ or ‘Step’. Make sure you enter the PCs of the break points separated by spaces before clicking ‘Apply’. Section 6.3 explains how to locate the address of a specific instruction to set a break point.

Multiple tabs display the output of the program together with the machine state, including register contents and memory contents.

Click ‘Exit Program’ to quit the execution of the current program. After that, the machine is ready to load another program.

6.3 Finding the address of an instruction

After running ‘make’, a `.dump` file will be generated for each program. The `.dump` file shows the assembler mnemonics for the machine instructions from the program.

`bubblesort.dump` can be very useful to debug `bubblesort.S`. For instance, if you compile without any modification to `bubblesort.S`, and search `<sort>:` in the generated `bubblesort.dump`, you would see contents similar to the following:

```
00000038 <sort>:
38: 00008067 ret
```

This indicates that the address of both the label `<sort>` and the (pseudo)instruction `ret` is 0x38, and the instruction encoding is 0x00008067. Therefore, you can set a break point at 0x38 if you want your program to stop right before executing `ret`. 