Revised: 3/21/2020

# Assembly Language I

Michael C. Hackett
Assistant Professor, Computer Science

Community College of Philadelphia

### Lecture Topics

- Assembly Language
- Registers
- Assembly Statements
- Arithmetic Instructions
  - Addition
  - Subtraction
  - Set If Less Than

- Logical Instructions
  - AND, OR, XOR, NOT
- Shift Instructions
- Immediate Instructions
- Printing Integers
- An Example in MARS

• If you're in this course, you've previously used a high-level programming language like Python, Java, etc.

- A high-level language is either compiled or interpreted.
  - The source code is converted to either:
  - Executable instructions the processor can understand. (Compiled)
  - Executable instructions that another intermediate piece of software can understand and execute. (Interpreted)

Languages like Python and Perl are interpreted.

• C and C++ source code is compiled.

- Some languages, like Java, are a mix of both.
  - Java source code is *compiled* to bytecode (Java's machine code).
  - The compiled bytecode is *interpreted* by the Java Virtual Machine.

- When programming with a low-level language, there is less abstraction.
  - It's not as easily read/written like with a high-level language.
  - You're basically writing in the language of the processor.
- *Machine Language/Machine Code*: Binary instructions the processor can understand.
- **Assembly Language**: Text instructions and symbols that map to the binary machine language instructions.

 Assembly programs are written in plain-text editors like Notepad or WordPad.

We'll be using the MIPS instruction set for assembly programming.

- We'll be using a simulator in this course, MARS, that will run our assembly programs.
  - MIPS Assembler and Runtime Simulator

### Registers

- Assembly instructions make use of a limited number of **registers**: small, but fast storage available to the processor.
- In the MIPS architecture, there are 32 general registers
  - Each register can store one 32-bit (4 bytes) word
  - A word is the term used by an instruction set that reflects both:
    - The size of a register
    - The size of an instruction
  - The word size in MIPS is 32-bits (4 bytes)
- When performing any instruction, the operation typically involves using the data currently in one or more of these registers.

### Registers

- The registers are referred to by their number, preceded by a dollar sign.
  - \$0 through \$31

- Registers more commonly are referenced by their *name*.
  - Many registers have predefined purposes.
  - The names make their purposes easier to remember.

# Registers

| Register Number | Name      | Used For                  |
|-----------------|-----------|---------------------------|
| \$0             | \$zero    | Must always contain 0     |
| \$1             | \$at      | Assembler Temporary Value |
| \$2-\$3         | \$v0-\$v1 | Function Return Values    |
| \$4-\$7         | \$a0-\$a3 | Function Arguments        |
| \$8-\$15        | \$t0-\$t7 | Temporary Values          |
| \$16-\$23       | \$s0-\$s7 | Saved Temporary Values    |
| \$24-\$25       | \$t8-\$t9 | Temporary Values          |
| \$26-\$27       | \$k0-\$k1 | Reserved for OS Kernel    |
| \$28            | \$gp      | Global Pointer            |
| \$29            | \$sp      | Stack Pointer             |
| \$30            | \$fp      | Frame Pointer             |
| \$31            | \$ra      | Function Return Address   |

Assembly statements have the following format:

```
[label:] mnemonic operand, operand, operand [#comment]
[label:] mnemonic operand, operand [#comment]
[label:] mnemonic operand
```

- The fields in brackets are optional
  - The brackets themselves, [ and ], are not literally in the instruction
- Each instruction has 1, 2 or 3 operands

```
[label:] mnemonic operand, operand, operand [#comment]
[label:] mnemonic operand, operand [#comment]
[label:] mnemonic operand
```

- The label is optional and used to identify a statement or group of statements
  - Will be demonstrated in the next lecture

```
[label:] mnemonic operand, operand, operand [#comment]
[label:] mnemonic operand, operand [#comment]
[label:] mnemonic operand [#comment]
```

- The mnemonic is the operation to be performed, such as **add** (for addition) and **sub** (for subtraction)
  - Mnemonics are easier to remember than their equivalent binary operation codes

```
[label:] mnemonic operand, operand, operand [#comment]
[label:] mnemonic operand, operand [#comment]
[label:] mnemonic operand [#comment]
```

- The instruction's operands are separated by commas.
- Each operand is usually a general register
  - Later lectures will show the use of memory references as operands

```
[label:] mnemonic operand, operand, operand [#comment]
[label:] mnemonic operand, operand [#comment]
[label:] mnemonic operand
```

- The end of a statement may include an optional comment.
  - Comments always begin with the # character
- Comments can exist on their own lines.
- There are no multi-line comments.
  - Unlike Java, for example, which uses /\* and \*/ for multi-line comments

• The first instruction we will see is the **add** mnemonic, which adds the values of two registers and stores the result in a third register.

- \$rd = Destination Register
- \$rs = Source Register 1
- \$rt = Source Register 2

- This is an example of a Register Format Instruction (or R-format)
  - More details on this in a later lecture

• The following instruction:

will add the values of registers \$t1 and \$t2 together and store the result in register \$t0

- One way to think of this instruction is: \$t0 = \$t1 + \$t2
  - However, this is not valid assembly code

$$$t1 = $t2 + $t0$$

$$$t1 = $t2 + $t0$$

• Convert the following to a valid assembly instruction:

$$$t1 = $t2 + $t0 + $t3$$

• Hint: It's not add \$t1, \$t2, \$t0, \$t3

$$$t1 = $t2 + $t0 + $t3$$

- Let's assume the following values are in these registers
  - \$t0 = 5
  - \$t1 = 0
  - \$t2 = 3
  - \$t3 = 7

• Convert the following to a valid assembly instruction:

$$$t1 = $t2 + $t0 + $t3$$

• We first need to sum \$t2 (3) and \$t0 (5)

• This assigns 8 to \$t1

Convert the following to a valid assembly instruction:

$$$t1 = $t2 + $t0 + $t3$$

• Next, we can add \$t3 (7) to \$t1 (8), and assign the result to \$t1

• This assigns 15 to \$t1

• Convert the following to a valid assembly instruction:

$$$t1 = $t2 + $t0 + $t3$$

add \$t1, \$t2, \$t0 add \$t1, \$t1, \$t3

• Similar in format to the **add** mnemonic, the **sub** mnemonic finds the difference between the values of two registers and stores the result in a third register.

sub \$rd, \$rs, \$rt

- \$rd = Destination Register
- \$rs = Source Register 1
- \$rt = Source Register 2

• The following instruction:

will subtract the value of registers \$t2 from \$t1 and store the result in register \$t0

- One way to think of this instruction is: \$t0 = \$t1 \$t2
  - However, this is not valid assembly code

$$$t1 = $t2 - $t0$$

$$$t1 = $t2 - $t0$$

$$$t1 = $t2 + $t0 - $t3$$

$$$t1 = $t2 + $t0 - $t3$$

```
add $t1, $t2, $t0
sub $t1, $t1, $t3
```

$$$t1 = $t2 + ($t0 - $t3)$$

$$$t1 = $t2 + ($t0 - $t3)$$

```
sub $t1, $t0, $t3
add $t1, $t1, $t2
```

#### Set If Less Than

• The **slt** mnemonic compares the numeric values of two registers and stores the either a 1 or 0 in a third register.

- \$rd = Destination Register
- \$rs = Source Register 1
- \$rt = Source Register 2

- The slt mnemonic stores a 1 in \$rd if \$rs < \$rt</li>
- The slt mnemonic stores a 0 in \$rd if \$rs >= \$rt

#### Set If Less Than

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 5
  - \$t2 = 13
- The slt mnemonic stores a 1 in \$t0 if \$t1 < \$t2
- The slt mnemonic stores a 0 in \$t0 if \$t1 >= \$t2

#### Set If Less Than

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 13
  - \$t2 = 11
- The slt mnemonic stores a 1 in \$t0 if \$t1 < \$t2
- The slt mnemonic stores a 0 in \$t0 if \$t1 >= \$t2

### Logical Instructions

Logical instructions return a true or false value.

• The operands of a logical instruction must be either true or false.

- For the rest of the course:
  - 1 means true
  - 0 means false

### Logical Instructions

• An **AND** logical operation results in true (1) only when both operands are themselves true (1)

| $\boldsymbol{x}$ | y | x AND y |
|------------------|---|---------|
| 0                | 0 | 0       |
| 0                | 1 | 0       |
| 1                | 0 | 0       |
| 1                | 1 | 1       |

- May be expressed as:
  - *x AND y*
  - $x \wedge y$
  - *x* · *y*

• The **and** mnemonic performs an and operation on two registers, and assigns the result in a third.

and \$rd, \$rs, \$rt

- \$rd = Destination Register
- \$rs = Source Register 1
- \$rt = Source Register 2

For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1
  - \$t2 = 0

This instruction will assign 0 to \$t0

| x | y | x AND y |
|---|---|---------|
| 0 | 0 | 0       |
| 0 | 1 | 0       |
| 1 | 0 | 0       |
| 1 | 1 | 1       |

For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1
  - \$t2 = 1

This instruction will assign 1 to \$t0

| $\boldsymbol{x}$ | y | x AND y |
|------------------|---|---------|
| 0                | 0 | 0       |
| 0                | 1 | 0       |
| 1                | 0 | 0       |
| 1                | 1 | 1       |

• Logical instructions perform bitwise operations.

- Let's assume the following values are in these registers
  - \$t0 = 0000000
  - \$t1 = 1001001
  - \$t2 = 1100111

• We perform the following instruction:

and \$t0, \$t1, \$t2

and \$t0, \$t1, \$t2

 Since registers are 32-bits in length, each logical instruction will perform 32 bitwise operations

• An **OR** logical operation results in true (1) only when at least one (or both) of the operands are themselves true (1)

| x | y | x OR y |
|---|---|--------|
| 0 | 0 | 0      |
| 0 | 1 | 1      |
| 1 | 0 | 1      |
| 1 | 1 | 1      |

- May be expressed as:
  - *x OR y*
  - $x \vee y$
  - x + y

• The **or** mnemonic performs an or operation on two registers, and assigns the result in a third.

or \$rd, \$rs, \$rt

- \$rd = Destination Register
- \$rs = Source Register 1
- \$rt = Source Register 2

For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1
  - \$t2 = 0

This instruction will assign 1 to \$t0

| x | y | x OR y |
|---|---|--------|
| 0 | 0 | 0      |
| 0 | 1 | 1      |
| 1 | 0 | 1      |
| 1 | 1 | 1      |

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1
  - \$t2 = 1

| • | This | instruction | will | assign 1 | L to \$t0 |
|---|------|-------------|------|----------|-----------|
|---|------|-------------|------|----------|-----------|

| $\boldsymbol{x}$ | y | x OR y |
|------------------|---|--------|
| 0                | 0 | 0      |
| 0                | 1 | 1      |
| 1                | 0 | 1      |
| 1                | 1 | 1      |

or \$t0, \$t1, \$t2

\$t0 = 0000000 \$t1 = 1001001 \$t2 = 1100111 1001001 OR <u>1100111</u> **1101111** 

\$t0 = **1101111** \$t1 = 1001001 \$t2 = 1100111

• An **XOR** (exclusive or, "x or"; "zor") logical operation results in true (1) only when one of the operands are true (1)

| x | у | x XOR y |
|---|---|---------|
| 0 | 0 | 0       |
| 0 | 1 | 1       |
| 1 | 0 | 1       |
| 1 | 1 | 0       |

- May be expressed as:
  - *x XOR y*
  - $x \oplus y$

• The **xor** mnemonic performs an xor operation on two registers, and assigns the result in a third.

- \$rd = Destination Register
- \$rs = Source Register 1
- \$rt = Source Register 2

For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1
  - \$t2 = 0

This instruction will assign 1 to \$t0

| x | y | x XOR y |
|---|---|---------|
| 0 | 0 | 0       |
| 0 | 1 | 1       |
| 1 | 0 | 1       |
| 1 | 1 | 0       |

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1
  - \$t2 = 1

This instruction will assign 0 to \$t0

| $\boldsymbol{x}$ | y | x XOR y |
|------------------|---|---------|
| 0                | 0 | 0       |
| 0                | 1 | 1       |
| 1                | 0 | 1       |
| 1                | 1 | 0       |

xor \$t0, \$t1, \$t2

\$t0 = 0000000 \$t1 = 1001001 \$t2 = 1100111 1001001 XOR <u>1100111</u> **0101110** 

\$t0 = **0101110** \$t1 = 1001001 \$t2 = 1100111

• A **NOT** logical operation negates/inverts a single operand

| x | NOT x |
|---|-------|
| 0 | 1     |
| 1 | 0     |

- May be expressed as:
  - NOT x
  - $\overline{x}$
  - $\bullet$   $\neg \chi$
  - x'
  - $\sim \chi$

• The **not** mnemonic performs an not operation on one register, and assigns the result in a second.

not \$rd, \$rs

- \$rd = Destination Register
- \$rs = Source Register

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 1

| x | NOT x |
|---|-------|
| 0 | 1     |
| 1 | 0     |

This instruction will assign 0 to \$t0

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0
  - \$t1 = 0

| х | NOT x |
|---|-------|
| 0 | 1     |
| 1 | 0     |

• This instruction will assign 1 to \$t0

not \$t0, \$t1

\$t0 = 0000000 \$t1 = 1001001 NOT <u>1001001</u> **0110110** 

\$t0 = **0110110** \$t1 = 1001001

Shifting moves the bits of a word to the left or right

- Shift instructions have a destination register, source register, and the shift amount
  - The shift instruction does not alter the data in the source register
- The first type of shifting process we will see is *logical shift*

• The **s11** mnemonic performs a logical left shift operation on one register, and assigns the result in a second.

sll \$rd, \$rs, shamt

- \$rd = Destination Register
- \$rs = Source Register
- shamt = Shift Amount

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0000000
  - \$t1 = 1010101

- This operation will shift the bits in \$11 to the left by two bits
  - Inserts 0's to the right

sll \$t0, \$t1, 2

\$t0 = 0000000 \$t1 = 1010101 1010101 1010100

\$t0 = **1010100** \$t1 = 1010101

• The **sr1** mnemonic performs a logical right shift operation on one register, and assigns the result in a second.

srl \$rd, \$rs, shamt

- \$rd = Destination Register
- \$rs = Source Register
- shamt = Shift Amount

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0000000
  - \$t1 = 0010101

- This operation will shift the bits in \$11 to the right by two bits
  - Inserts 0's to the left

srl \$t0, \$t1, 2

\$t0 = 0000000 \$t1 = 0010101 0010101

0000101

\$t0 = **0000101** \$t1 = 0010101

- Right shifting can be either
  - Logical Shift (The previous example)
  - Arithmetic Shift
- Arithmetic Shift ensures a negative number (begins with 1) remains negative.
  - It fills the left bits with 1's instead of 0's for negative numbers
- The **sra** mnemonic is used for arithmetic right shift.
  - There is no arithmetic left shift

• For the following instruction:

- Let's assume the following values are in these registers
  - \$t0 = 0000000
  - \$t1 = 1110101
- This operation will shift the bits in \$11 to the right by two bits
  - Inserts 1's to the left since the left-most bit is a 1
  - If the number in \$11 started with 0, 0's would be inserted

sra \$t0, \$t1, 2

\$t0 = 0000000 \$t1 = 1110101 1110101

1111101

\$t0 = **1111101** 

\$t1 = 1110101

 A useful application of left shifting is to perform multiplication by powers of 2

- $00001100_2 = 12_{10}$ 
  - Left shifting by **2** yields **00110000**<sub>2</sub> =  $48_{10}$
  - $12_{10} \times 2_{10}^2$
- $00001111_2 = 15_{10}$ 
  - Left shifting by **3** yields **01111000**<sub>2</sub> =  $120_{10}$
  - $15_{10} \times 2_{10}^3$

 A useful application of right shifting is to perform division by powers of 2

- $01001100_2 = 76_{10}$ 
  - Right shifting by **2** yields **00010011**<sub>2</sub> =  $19_{10}$
  - $76_{10} \div 2_{10}^2$
- $11110000_2 = 240_{10}$ 
  - Right shifting by **3** yields **00011110**<sub>2</sub> =  $30_{10}$
  - $240_{10} \div 2_{10}^3$

- The R-Format instructions seen so far lack the ability to:
  - Assign a constant value (other than \$zero) to a register
  - Perform arithmetic instructions with a constant value
  - Perform logical instructions with a constant value
  - Transfer data to registers from main memory and vice versa
- Immediate Format Instructions (I-Format) allow the use of immediate/constants
  - More details on this in a later lecture
- The maximum size of an immediate constant is 16-bits (half-word)
  - -32,768 to 32,767

• The first immediate instruction we will see is the **addi** mnemonic, which adds the value of a register with a constant and stores the result in a destination register.

addi \$rd, \$rs, constant

- \$rd = Destination Register
- \$rs = Source Register
- constant = A literal/constant value

• The following instruction:

will add the values of register \$11 and the number 5 together and store the result in register \$10

- One way to think of this instruction is: \$t0 = \$t1 + 5
  - However, this is not valid assembly code

• The **subi** mnemonic subtracts the value of a register with a constant and stores the result in a destination register.

subi \$rd, \$rs, constant

- \$rd = Destination Register
- \$rs = Source Register
- constant = A literal/constant value

• The following instruction:

will subtract 5 from the values of register \$t1 and store the result in register \$t0

- One way to think of this instruction is: \$t0 = \$t1 5
  - However, this is not valid assembly code

- There are immediate instruction mnemonics for the logical instructions.
  - No need for an immediate instruction for NOT operations

```
andi $rd, $rs, constant ori $rd, $rs, constant xori $rd, $rs, constant
```

- \$rd = Destination Register
- \$rs = Source Register
- constant = A literal/constant value

- Loading describes the process of placing data into a register
  - Either as a constant value or data loaded from main memory
- The **li** (load immediate) mnemonic loads a constant to a register
  - 1i is a pseudo-operation; it is not included as part of the MIPS instruction set
  - The assembler converts an 1i instruction to appropriate instructions

li \$rd, constant

- \$rd = Destination Register
- constant = A literal/constant value

li \$t0, 17

- This instruction would load the value 17 to register \$t0
  - Essentially, \$t0 = 17

- Moving describes the process of copying data from one register to another
- The **move** mnemonic copies data from one register to another
  - move is a pseudo-operation; it is not included as part of the MIPS instruction set
  - The assembler converts a **move** instruction to appropriate instructions

move \$rd, \$rs

- \$rd = Destination Register
- \$rs = Source Register

move \$t1, \$t0

• This instruction would copy the value in \$t0 to register \$t1

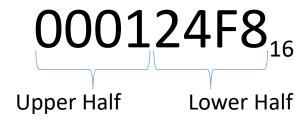
| li   | \$t0, | <b>17</b> |  |
|------|-------|-----------|--|
| move | \$t1, | \$t0      |  |

- As previously stated, the constants can only be 16 bits in size
  - In two's complement this gives us a range of -32,768 through 32,767
  - Or 0 through 65,535 for unsigned numbers
- When using the **li** pseudo-operation, it assigns the constant to the *lower-order half* of the register's 32 bits.
- Another instruction is used to assign a constant to the *higher-order* half of the register's 32 bits.

• To demonstrate, we'll use hexadecimal constants as this will easily show the higher and lower halves of the register.

• Let's say we want to load the number 75,000 (a number that requires at least 17 bits) in register \$t0

•  $75,000_{10} = 0000000000000000010010011111000_2 = 000124F8_{16}$ 



• To load a constant to the higher-order half of the register, the lui (load upper immediate) pseudo-operation is used.

lui \$rd, constant

- \$rd = Destination Register
- constant = A literal/constant value

 The li pseudo-instruction assigns data to the lower half and clears/zeros the upper half

• The **lui** instruction assigns data to the upper half and clears/zeros the lower half

- This demonstrates using hexadecimal literals instead of the decimal literals used in previous slides
  - A 0x prefix indicates a hexadecimal literal/constant

```
$t0 = 0x00000000 lui $t0, 0x0001 $t0 = 0x00010000
$t0 = 0x00000000 li $t0, 0x24f8 $t0 = 0x0000024f8
```

Both examples below will yield the wrong value assigned to \$t0

| \$t0 = 0x00000000 | lui | \$t0, | 0x0001 | \$t0 = 0x <b>00010000</b> |
|-------------------|-----|-------|--------|---------------------------|
| \$t0 = 0x00010000 | li  | \$t0, | 0x24f8 | \$t0 = 0x <b>000024F8</b> |

Start by loading the upper half

\$t0 = 0x00000000 **lui** \$t0, 0x0001 \$t0 = 0x00010000

• Then, perform an or operation with the lower half

```
$t0 = 0x00000000 lui $t0, 0x0001 $t0 = 0x000100000 $t0 = 0x000100000 ori $t0, 0x24f8 $t0 = 0x000124f8
```

### Printing Integers

• MIPS has special *system calls* to provide services like reading input and printing output.

• We'll focus only on printing integers and show examples of printing other data types and reading keyboard input in later lectures.

### Printing Integers

• To print an integer, the number **must** be placed in register \$a0

- The system call code for printing an integer is 1
  - 1 must be placed in register \$v0
- When the above two steps are complete, the syscall instruction is used to perform the function indicated by the number in \$v0

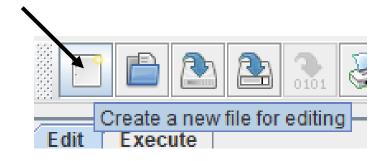
### Printing Integers

```
move $a0, $t0 #Copies the value in $t0 to $a0

li $v0, 1 #Sets the syscall code for printing an integer

syscall #Prints the integer in register $a0
```

- Open MARS
  - See MARS Download document in Canvas
- Click the New button to create a new assembly source code file



• Enter the following:

```
mips1.asm*

1  #Prints the number 15

2  
3  li $t1, 15   #Loads the constant 15 to $t1

4  
5  move $a0, $t1   #Copies $t1 (15) to $a0

6  
7  li $v0, 1   #Sets the syscall code for printing an integer

8  
9  syscall   #Prints the integer ($v0 = 1) in $a0 (15)

10
```

- Save the .asm file
  - The default name "mips1.asm" is fine

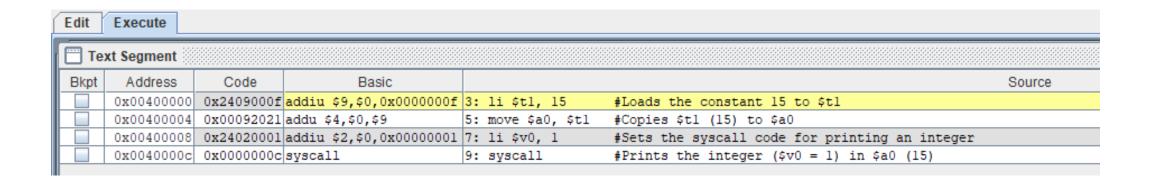




Click the Assemble button



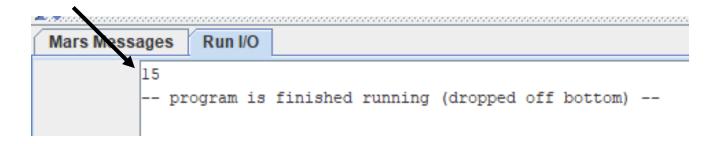
You should see something similar to what is shown below



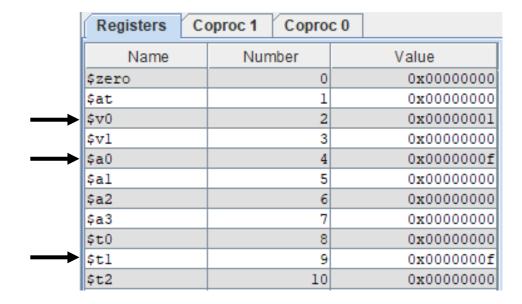
Click the Run Current Program button



• 15 should be displayed in the Run I/O pane at the bottom of MARS



- Note that you can see the data in the registers at the conclusion of the program
  - Values are displayed in hexadecimal



• To re-run a program, click the Reset MARS button



• Then click the Run Current Program button again

