#### PennSim & LC4 Memory Introduction to Computer Systems, Fall 2022

**Instructor:** Travis McGaha

#### TAs:

Ali Krema Audrey Yang David LuoZhang Heyi Liu Katherine Wang Noam Elul Ria Sharma Andrew Rigas Craig Lee Eddy Yang Janavi Chadha Kyrie Dowling Patricia Agnes Sarah Luthra Anisha Bhatia Daniel Duan Ernest Ng Jason Hom Mohamed Abaker Patrick Kehinde Jr. Sofia Mouchtaris When poll is active, respond at **pollev.com/tqm** 

📧 Text TQM to 37607 once to join

## What is your favourite editor for code?



Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app

## Logistics

- HW03 Sequential Logic: This Friday 10/7 @ 11:59 pm
  - Written Homework, submitted to gradescope
  - NO EXTENSIONS OVER 72 HOURS
  - Should have everything you need
  - Practice in Recitations this week
- HW04 LC4 Programming: to be released this Friday
  - Programming assignment
  - May not have everything you need until Monday's lecture
- Check-in04 to be released tomorrow

## **In-Person Lecture Policies**

\* I ask that you wear a mask in lecture

- If you are using your electronics (outside of polls), please sit in the back
  - Having electronics out make it a lot easier to distracted by random notifications
  - Easy for people sitting nearby & behind you to get distracted by your distractions

### **Lecture Outline**

- ASM Files, Object Files, PennSim Demo
- LC4 Program Design (if/while/for)
- LC4 Memory

## More LC4 Syntax

- Integer Immediates (CONST, HICONST, ADD, SLL, etc.)
   can be either in hexadecimal or in decimal form
  - Hexadecimal constants **0xFF** or **xFF**
  - Decimal constants: #240, #-240
- Comments
  - Comments in LC4 are preceded by a ;

✤ Example:

```
CONST R0, 0x20
CONST R1, x10
CONST R2, #64
; this is a comment
DIV R3, R2, R1 ; this is a comment too
ADD R3, R3, R0
```

## LC4 Labels

- It can be cumbersome to calculate offsets for jumps.
- LC4 assembler allows us to put labels on memory. We can use labels for Jumps and Branch instructions to make our lives easier
  - A Label is just a "name" for a memory location. Like how we can refer to a memory location with an address.



## **Code in Memory**

- An instruction fits in 1 memory location (16 bits)
- These instructions are stored in memory and accessed sequentially
  - When we trace through the code, we are just accessing the next location in memory
     Index # Information



## **LC4 Memory Layout**

- The address space in LC4 is split into separate pieces for code and for data
- Separate regions for OS and User as well (more on OS in ~2 weeks)
- More to memory than this, but that will be discussed later

0x0000	User Code
0x1FFF	
0x2000	
	User Data
0x7FFF	
0x8000	
	OS Code
0x9FFF	
0xA000	
	OS Data
	+ Device Memory
0xFFFF	

## **LC4 ASM Directives**

- We can include directives to indicate where things in our ASM program should be loaded into memory
  - . CODE
    - Next instructions are in the CODE space
  - . DATA
    - Next values are in the DATA space
  - . ADDR
    - Set the current address to the specified value

Other directives exist, more on those later

## LC4 Example .asm file



## **LC4 ASM Files**

- LC4 Assembly Files are text files that contain a lot of conveniences for LC4 programmers
  - Instructions written as text (not as 16-bit patterns)
  - Comments
  - Initial values of some memory locations
  - Labels to refer to addresses and values
  - Directives
  - Pseudo Instructions
- ASM files are not directly run on the LC4 processor, this file needs to be processed into something machinereadable

### **Text Files**

- .asm files and source code files for most languages are text files:
  - They can only contain ascii (or sometimes Unicode) characters.
  - Are not directly executed by the computer
- Text files are different from .docx and .pdf files
  - You cannot write text files in Microsoft Word
- Text files are created and edited by a text editor
  - Vim, Emacs, Notepad, Notepad++, Nano, Sublime, Atom, etc.

### **ASM -> OBJ Process**

- ASM Files need to be processed by an assembler to become machine code
- Machine code can be executed directly by the computer hardware. (or a simulator in our case)
- Demo: multiply.obj vs multiply.asm



## PennSim

- A Java program written to:
  - Convert LC4 Assembly to machine code
  - Simulate the operations of LC4 ISA
  - Provide debugging tools for LC4 ISA

- PennSim Demo
  - (See the lecture recording)

## PennSim Commands Pt. 1

- PennSim has a command line at the top that you can type commands into
  - as <outfilename> <infilename>
    - <u>As</u>sembles a file from an assembly file into an object file .asm -> .obj
  - Id <filename>
    - Load an object file into memory
  - set <register> <value>
    - Loads a specific value into a register, works for special registers too (PC and PSR)
  - help <command>
    - Gives some helpful information on the specified command
  - reset
    - Resets the state of the simulator clears memory and registers

## PennSim Commands Pt. 2

#### clear

- Clears the output window
- break <set | clear> label
  - Set or clear a break point at the specific label

#### step

• Simulate the execution of the next instruction

#### continue

 Continue the simulation until a breakpoint or fatal error is encountered

#### script <filename>

• You can put a sequence of commands in a text file and then run them all at once using this command. Convenient and required for HW04.

## **Lecture Outline**

- ASM Files, Object Files, PennSim Demo
- LC4 Program Design (if/while/for)
- LC4 Memory

- Not all programming constructs have direct LC4 instructions
- How would we implement

```
if (R0 >= 3)
```

R1 = R0;



- Not all programming constructs have direct LC4 instructions
- How would we implement

```
if (R0 != R2) {
    R1 = R2;
} else {
    R1 = 0;
}
```

START	
	CMP R0, R2
	BRz ELSE
	ADD R1, R2, <b>#0</b>
	JMP AFTER
ELSE	CONST R1, #0
AFTER	
	;

- Not all programming constructs have direct LC4 instructions
- \* How would we implement
  while (R0 != 2) {
   // ...
  }



- Not all programming constructs have direct LC4 instructions
- \* How would we implement
  for (R0 = 0; R0 < R6; R0++) {
   // ...
  }
  CONST R0, #0
  START\_LOOP
  CMP R0, R6
  BRzp AFTER\_LOOP
   // ...
  ADD R0, R0, #1
  JMP START LOOP</pre>

AFTER LOOP

22

## **Note On Labels**

- When you are writing LC4 assembly, the labels you use must be unique.
  - If you use the same label more than once, the assembler will not know which location you are referring to with JMP <LABEL>
- To avoid name conflicts, it is common to number the labels or give more specific names.
- Instead of just using LOOP
  - LOOP\_1
  - LOOP\_2
  - SUM\_NUM\_LOOP
  - etc.

## **Assembly Programming Strategy**

- One approach
  - Start by writing a pseudo code program
    - Try to keep code "simple"
      - don't overuse variables, avoid recursion, etc
    - Comment while you do this
  - Translate each variable to a register
  - Translate each line/piece of code to Assembly
  - Test your assembly to make sure it works

## **Lecture Outline**

- ASM Files, Object Files, PennSim Demo
- LC4 Program Design (if/while/for)
- LC4 Memory

## **Data Memory**

- In LC4 we have 8 general purpose registers.
- Most programs need more than 8 variables
- User data is a portion of memory where we can store data that can't currently be in a register



### **Memory vs Registers**

- Registers
  - Quick access storage
  - Can easily directly modify values in registers
- Memory:
  - Memory can take longer to read/write
  - Read/write memory requires its own instructions
  - Need to read data from memory into a register before we can operate on it
  - After data is updated, it needs to be stored back in memory for memory to be updated.

## **Pointers**

- A <u>Pointer</u> is a variable (register in our case) that contains the address of a memory location.
  - We can use this pointer to read/write to that memory location
  - We can modify the address stored in the pointer through arithmetic to get a new address
- Dereferencing a pointer is when we take the address stored in the pointer and access that memory location for a read/write
- Pointers will be important for pretty much the rest of the course, especially for C programing.

### LDR and STR

- & LDR Rd, Rs, IMM6
  - Action: Rd = memory[Rs + SEXT(IMM6)]
  - Reads the data at address Rs + SEXT (IMM6) in memory and loads it in Rd
- str Rt, Rs, IMM6
  - Action: memory[Rs + SEXT(IMM6)] = Rt
  - <u>Stores</u> the value in Rt to the memory location at address
     Rs + SEXT (IMM6)

Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	R0 x20	
HICONST	R0 x40	•
LDR	R1, R0,	<b>#0</b>
ADD	R0, R0,	#1
LDR	R2, R0,	<b>#0</b>
LDR	R3, R0,	#1

Registers	Value
RO	
R1	
R2	
R3	

Set R0 to contain address 0x4020

x401E	<b>0xDEAD</b>
x401F	0xF00D
x4020	0xEF24
x4021	0x9823
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	R0 x20
HICONST	R0 x40
LDR	R1, R0, #0
ADD	R0, R0, <mark>#1</mark>
LDR	R2, R0, <mark>#0</mark>
LDR	R3, R0, <b>#1</b>

Registers	Value
RO	0x0020
R1	
R2	
R3	

Set R0 to contain address 0x4020

x401E	<b>0xDEAD</b>
x401F	0xF00D
x4020	0xEF24
x4021	0x9823
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	R0 x20		
HICONST	R0 x40		
LDR	R1, R0,	<b>#</b> 0	Load the value at R0 + 0 into R1
ADD	R0, R0,	#1	
LDR	R2, R0,	<b>#</b> 0	
LDR	R3, R0,	#1	

#### R0 now "points to" the value stored at 0x4020

Memory

**OxDEAD** 

**0xF00D** 

**0xEF24** 

**0x9823** 

**0x401E** 

**0x328F** 



Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	R0 x	20	
HICONST	R0 x	40	
LDR	R1,	R0,	<b>#</b> 0
ADD	R0,	R0,	#1
LDR	R2,	R0,	<b>#</b> 0
LDR	R3,	R0,	#1
•			

#### R0 now "points to" the value stored at 0x4020



x401E	<b>OxDEAD</b>
x401F	<b>0xF00D</b>
x4020	0xEF24
x4021	0x9823
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	<b>BU 420</b>	
CONSI		
HICONST	R0 x40	
LDR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
LDR	R2, R0,	<b>#</b> 0
LDR	R3, R0,	#1
•		

#### R0 is updated to "point to" the value stored at 0x4021

Registers	Value
RO	0x4021 ·
R1	OxFF24
KZ	
K3	

Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	R0 x20	
HICONST	R0 x40	
LDR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
LDR	R2, R0,	<b>#</b> 0
LDR	R3, R0,	#1
•		

Registers	Value		x401E	<b>0xDEAD</b>
RO	0x4021 -		x401F	<b>0xF00D</b>
R1	0xFF24		x4020	<b>0xEF24</b>
R2	0x0823		x4021	<b>0x9823</b>
	089025		x4022	0x401E
K3			x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

What is we wanted to read a value stored at address 0x4020 in memory?

CONST	R0	<b>x</b> 20	
HICONST	R0	<b>x</b> 40	
LDR	R1,	R0,	<b>#0</b>
ADD	R0,	R0,	#1
LDR	R2,	R0,	<b>#0</b>
LDR	R3,	R0,	#1
•			

Registers	Value		x401E	<b>0xDEAD</b>
RO	0x4021 -		x401F	<b>0xF00D</b>
R1	0xFF24		x4020	<b>0xEF24</b>
			x4021	0x9823
	0x3025		x4022	0x401E
K3	0X401E		x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

To store a value in memory, we follow similar steps

/	
CONST	R0 x20
HICONST	R0 x40
STR	R1, R0, <mark>#0</mark>
ADD	RO, RO, <mark>#1</mark>
ADD	R1, R1, <mark>#14</mark>
STR	R1, R0, <mark>#0</mark>
ADD	R1, R1, <mark>#9</mark>
STR	R1, R0, <mark>#1</mark>
•	

Registers	Value
RO	
R1	#24
R2	
R3	

Set R0 to contain address 0x4020

x401E	<b>0xDEAD</b>
x401F	0xF00D
x4020	0xEF24
x4021	0x9823
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

To store a value in memory, we follow similar steps

1	CONST	R0 x20	c
	HICONST	R0 x40	3
	STR	R1, R0, <mark>#0</mark>	
	ADD	R0, R0, <mark>#1</mark>	
	ADD	R1, R1, <b>#14</b>	
	STR	R1, R0, <mark>#0</mark>	
	ADD	R1, R1, <mark>#9</mark>	
	STR	R1, R0, <b>#1</b>	
	•		

Registers	Value
RO	0x0020
R1	#24
R2	
R3	

et R0 to contain address 0x4020

x401E	<b>0xDEAD</b>
x401F	0xF00D
x4020	0xEF24
x4021	0x9823
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

CONST	R0 x2	0
HICONST	R0 x4	0
STR	R1, R	0, <mark>#</mark> 0
ADD	R0, R	0, #1
ADD	R1, R	1, #14
STR	R1, R	0, #0
ADD	R1, R	1, #9
STR	R1, R	0, <mark>#1</mark>



Red arrow is the Program Counter (PC) which points to the next instruction to execute

CONST	R0 x20	
HICONST	R0 x40	
STR	R1, R0,	<b>#0</b>
ADD	R0, R0,	#1
ADD	R1, R1,	#14
STR	R1, R0,	<b>#0</b>
ADD	R1, R1,	<b>#9</b>
STR	R1, R0,	#1
-		



Red arrow is the Program Counter (PC) which points to the next instruction to execute

CONST	R0 x20	
HICONST	R0 x40	
STR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
ADD	R1, R1,	#14
STR	R1, R0,	<b>#</b> 0
ADD	R1, R1,	<b>#</b> 9
STR	R1, R0,	#1

Registers	Value
RO	0x4021
R1	#24
R2	
R3	



<b>0xDEAD</b>
0xF00D
0x0018
0x9823
0x401E
0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

/		-
CONST	R0 x20	
HICONST	R0 x40	
STR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
ADD	R1, R1,	#14
STR	R1, R0,	<b>#</b> 0
ADD	R1, R1,	<b>#9</b>
STR	R1, R0,	#1
<b>~</b>		

Registers	Value
RO	0x4021
R1	#38
R2	
R3	



x401E	<b>0xDEAD</b>
x401F	0xF00D
x4020	0x0018
x4021	0x9823
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

CONST	R0 x20	
HICONST	R0 x40	
STR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
ADD	R1, R1,	#14
STR	R1, R0,	<b>#</b> 0
ADD	R1, R1,	<b>#9</b>
STR	R1, R0,	#1
-		

Registers	Value
RO	0x4021
R1	#38
R2	
R3	



x401E	<b>0xDEAD</b>
x401F	0xF00D
x4020	0x0018
x4021	0x0026
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

CONST	R0 x20	
HICONST	R0 x40	
STR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
ADD	R1, R1,	#14
STR	R1, R0,	<b>#</b> 0
ADD	R1, R1,	<b>#9</b>
STR	R1, R0,	#1

Registers	Value
RO	0x4021
R1	#47
R2	
R3	



x401E	<b>0xDEAD</b>
x401F	<b>0xF00D</b>
x4020	0x0018
x4021	0x0026
x4022	0x401E
x4023	0x328F

Red arrow is the Program Counter (PC) which points to the next instruction to execute

CONST	R0 x20	
HICONST	R0 x40	
STR	R1, R0,	<b>#</b> 0
ADD	R0, R0,	#1
ADD	R1, R1,	#14
STR	R1, R0,	<b>#</b> 0
ADD	R1, R1,	<b>#9</b>
STR	R1, R0,	#1
•		

Registers	Value
RO	0x4021
R1	#47
R2	
R3	



<b>0xDEAD</b>
0xF00D
0x0018
0x0026
0x002F
0x328F

#### pollev.com/tqm

- What value is stored at address 0x4020 by the end of the program?
  - A. 110
  - **B. 210**
  - **C. 240**
- D. 320
- E. I'm not sure

Registers	Value
RO	
R1	
R2	

CONST	R0 x20
HICONST	R0 x40
LDR	R1, R0, <mark>#0</mark>
CONST	R2, <mark>#110</mark>
ADD	R1, R1, R2

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

pollev.com/tqm

- What value is stored at address 0x4020 by the end of the program?
  - A. 110
  - **B. 210**
- C. 240
- D. 320
- E. I'm not sure

Registers	Value
RO	
R1	
R2	

_	CONST	R0 x20
	HICONST	R0 x40
	LDR	R1, R0, <mark>#0</mark>
	CONST	R2, <b>#110</b>
	ADD	R1, R1, R2

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

pollev.com/tqm

- What value is stored at address 0x4020 by the end of the program?
  - A. 110
  - **B. 210**
  - **C. 240**
- D. 320
- E. I'm not sure

Registers	Value
RO	0x0020
R1	
R2	

CONST	R0 x20
HICONST	R0 x40
LDR	R1, R0, <mark>#0</mark>
CONST	R2, <b>#110</b>
ADD	R1, R1, R2

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

pollev.com/tqm

- What value is stored at address 0x4020 by the end of the program?
  - A. 110
  - **B. 210**
  - **C. 240**
  - D. 320
  - E. I'm not sure

Registers	Value
RO	0x4020
R1	
R2	

CONST	R0 x20
HICONST	R0 x40
- LDR	R1, R0, <mark>#0</mark>
CONST	R2, <b>#110</b>
ADD	R1, R1, R2

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

pollev.com/tqm

- What value is stored at address 0x4020 by the end of the program?
  - A. 110
  - **B. 210**
  - **C. 240**
  - **D. 320**
  - E. I'm not sure

Registers	Value
RO	0x4020
R1	210
R2	

CONST	R0 x20
HICONST	R0 x40
LDR	R1, R0, <mark>#0</mark>
CONST	R2, <mark>#110</mark>
ADD	R1, R1, R2

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

pollev.com/tqm

- What value is stored at address 0x4020 by the end of the program?
  - A. 110
  - **B. 210**
  - **C. 240**
  - **D. 320**
  - E. I'm not sure

Registers	Value
RO	0x4020
R1	210
R2	110

CONST	R0 x20
HICONST	R0 x40
LDR	R1, R0, <mark>#0</mark>
CONST	R2, <b>#110</b>
ADD	R1, R1, R2

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

pollev.com/tqm

What value is stored at address 0x4020 by the end of the program?

**B. 210** 

**C. 240** 

**D. 320** 

E. I'm not sure

Registers	Value
RO	0x4020
R1	320
R2	110

CONST	R0 x20
HICONST	R0 x40
LDR	R1, R0, <mark>#0</mark>
CONST	R2, <b>#110</b>
ADD	R1, R1, R2

#### Memory

x401F	240
x4020	210
x4021	107
x4022	380
x4023	333

Value was not written back to memory with STR! Memory is not changed

## **More LC4 Directives**

- These assembly directives provide information on how various data/constants should be assembled
  - .FALIGN
    - Pad current address to the next multiple of 16. Useful since a subroutine must start on an address that is a multiple of 16
  - .FILL IMM16
    - Set value at the current address to the specified 16 bit value
  - .BLKW UIMM16
    - Reserve UIMM16 words at the current address
  - .CONST IMM16
    - Associate IMM16 with the preceding label
  - .UCONST UIMM16
    - Associate UIMM16 with the preceding label

## **LC4 Pseudo Instructions**

- Pseudo instructions look like normal instructions, but the assembler translates them into real instructions.
  - Pseudo instructions act as convenient aliases to make code more readable for us
  - RET
    - <u>**Ret</u>urn from a subroutine**</u>
    - Actual implementation: **JMPR R7**
  - LEA Rd, <LABEL>
    - <u>Load</u> <u>effective</u> <u>a</u>ddress associated with a label into a register
    - Actual implementation: a CONST, HICONST Pair
  - LC Rd, <LABEL>
    - <u>L</u>oad a <u>c</u>onstant 16-bit value associate with a label via .CONST or .UCONST into a register
    - Actual implementation: a **CONST**, **HICONST** Pair

### Arrays

- Arrays are a common programming structure that consists of a list of other data
  - (we will work with arrays of integers)
- Arrays in LC4
  - Since integers are 16-bits and registers are 16-bits, we can only fit one integer into a register.
  - We cannot "assign" a register to an array the same way we can do so for integers
  - Instead, store the array into data memory and have a register with the address of the beginning of the array and a register with the length of the array

#### sum\_numbers.asm

- A sample program where we iterate over an array and get the sum of all integers in that array
  - Very useful to look at when working on HW04
- Demo

## Strings & strlen.asm

- Strings can be thought of as similar to arrays, but as a sequence of characters.
- Instead of having both the address to the start and the length, we only have the address to the start
  - The end of the string is marked by having a null-terminator character (which is equal to 0) to mark the end of the string.

- strlen.asm
  - Demo that shows how to calculate the length of a string in LC4