EECS 370 - Lecture 4 ARM

byte be like





Announcements

- Lab 2 pre-quiz due tonight on Canvas
- Lab 1 extended until tonight
- Lab 2 meets this Friday/Monday
 - Attendance will be taken
 - Assignment will be completed on paper by end of lab period
- Project 1a due next Thursday (week from today)
 - Can do ~80% of it now
 - Will have the rest after lecture today



Resources

- Many resources on 370 website
 - https://eecs370.github.io/#resources
 - ARMv8 references

			Assembly o	code	Semantics		Comments
d	ADD	Xd.	Xn.	Xm	X5 = X2 + X7		register-to-register
d & set flags	ADDS	Xd.	Xn.	Xm	X5 = X2 + X7 $X5 = X2 + X7$		flags NZVC
d immediate	ADDI	Xd,	Xn.	#uimm12	X5 = X2 + X7 X5 = X2 + #19		$0 \le 12$ bit unsigned ≤ 4095
d immediate & set flags	ADDIS	Xd,	Xn,	#uimm12	X5 = X2 + #15 X5 = X2 + #15		flags NZVC
btract	SUB	Xd.	Xn.	Xm	X5 = X2 + W1 X5 = X2 - X7	,	register-to-register
btract & set flags	SUBS	Xd.	Xn.	Xm	X5 = X2 - X7 X5 = X2 - X7		flags NZVC
btract immediate	SUBI	Xd,	Xn,	#uimm12	X5 = X2 - X7 X5 = X2 - #20		$0 \le 12$ bit unsigned ≤ 4095
btract immediate & set flags	SUBIS	Xd,	Xn.	Xm	X5 = X2 - #20 X5 = X2 - #20		flags NZVC
ottact miniculate ee set nags	30013	Au,	All,	AIII	A3 - A2 - #20		Hago NZ YC
ata Transfer Operations	As	sembly c	ode		Semantics	Comm	ents
ad register	LDUR	Xt.	[Xn, #simm	91	X2 = M[X6, #18]	double wo	rd load into Xt from Xn + #simm9
ad signed word	LDURSW	Xt,	[Xn, #simm		X2 = M[X6, #18]	word load	to lower 32b Xt from Xn + #simm9; sign extend upper 32b
ad half	LDURH	Xt,	[Xn, #simm	91	X2 = M[X6, #18]		and to lower 16b Xt from Xn + #simm9; zero extend upper 48b
ad byte	LDURB	Xt,	[Xn, #simm		X2 = M[X6, #18]		to least 8b Xt from Xn + #simm9 zero extend upper 56b
ore register	STUR	Xt.	[Xn, #simm		M[X5, #12] = X4		rd store from Xt to Xn + #simm9
ore word	STURW	Xt.	[Xn, #simm		M[X5, #12] = X4		from lower 32b of Xt to Xn + #simm9
ore half word	STURH	Xt.	[Xn, #simm		M[X5, #12] = X4	½ word lo	ad from lower 16b of Xt to Xn + #simm9
ore byte	STURB	Xt,	[Xn, #simm	91	M[X5, #12] = X4	byte load	from least 8b of Xt to Xn + #simm9
offset		#simm9	= -256 to +255			-256 ≤ 9	bits signed immediate ≤ +255
ove wide with zero	MOVZ	Xd,	#uimm16,	LSL N	X9 = 00N00	first (N =	Xd then place a 16b (#uimm) into the 0)/second (N = 16)/third (N = 32)/fourth (N = 48)
ove wide with keep	MOVK	Xd,	#uimm16,	LSL N	X9 = xxNxx		i Xd b (#uimm) into the first (N = 0)/second (N = 16)/ 32)/fourth (N = 48) 16b slot of Xd, without changing the other values (x's)
register aliases		X28 = S	P; X29 = FP;	X30 = LR; X31 = .	XZR		127, John C. Toy South Co. For Hand Co. Sp. St. Co. Sp. Sp. Sp. Sp. Sp. Sp. Sp. Sp. Sp. Sp
ogical Operations	Assemb	ly code			Semantic	es	Using C operations of & ^ << >>
d	AND	Xd.	Xn.	Xm	X5 = X2 & 2	C7	bit-wise AND
d immediate	ANDI	Xd,	Xn.	#uimm12	2 X5 = X2 & #	19	bit-wise AND with 0 ≤ 12 bit unsigned ≤ 4095
clusive or	ORR	Xd.	Xn.	Xm	$X5 = X2 \mid X$		bit-wise OR
clusive or immediate	ORRI	Xd.	Xn.	#uimm12			bit-wise OR with 0 ≤ 12 bit unsigned ≤ 4095
clusive or	EOR	Xd,	Xn.	Xm	$X5 = X2 ^X$		bit-wise EOR
clusive or immediate	EOR	Xd.	Xn,	#uimm12			bit-wise EOR with 0 ≤ 12 bit unsigned ≤ 4095
gical shift left	LSL	Xd,	Xn,	#uimm6	X1 = X2 < <		shift left by a constant ≤ 63
gical shift right	LSR	Xd,	Xn.	#uimm6	X5 = X3 > 3	> #20	shift right by a constant ≤ 63
nconditional branches	Accom	ıbly code	,	Semant	tics	Also know	vn as Jumps
anch	B	#simm26		PC + #1200	tics		ch PC + 26b offset; -2^25 ≤ #simm26
	-					≤ 2^25-1; 4b ins	struction
anch to register	BR	Xt		et in Xt	111000	Xt contains a ful	
anch with link	BL	#simm26	X30	= PC + 4; PC + #	F11000	16 million instru	ch to PC + 26b offset; ctions; ains return from subroutine address



Labels in LC2K

- The code on the right is awkward
 - Need to count lines to see what it's doing
- Labels make code easier to read/write
- Label definition: listed to the left of the opcode
 - Can be defined on any line (only once)
- Label use: replaces offset value in lw/sw/beq instructions (any number)
- For lw/sw, assembler will replace label use with the line number of definition
 - In this example, data is on line 2

```
lw 0 1 2
halt
.fill 12
```

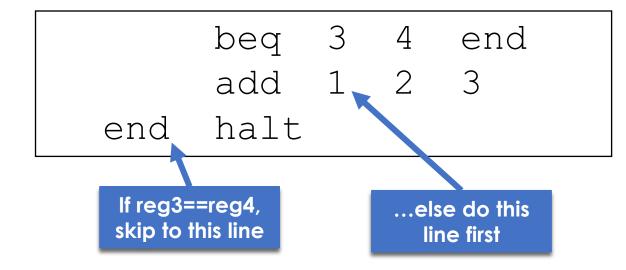




Labels in LC2K - beq

- Labels with beq indicate where we should branch
- Assembler's job is a little tricker
 - Doesn't just replace it with line number
 - Remember: target address is PC+1+offset

```
beq 3 4 1
add 1 2 3
halt
```





Exercise

```
// this is the assembly for:
while(x != y) {
  x *= 2;
}
```

 What are the values of the labels here?

```
loop beq 3 4 end
add 3 3 3
beq 0 0 loop
end halt
```

What are the labels replaced with?



Programming Assignment #1

- Write an assembler to convert input (assembly language program) to output (machine code version of program)
 - "1a"
- Write a behavioral simulator to run the machine code version of the program (printing the contents of the registers and memory after each instruction executes)
 - "1s"
- Write an efficient LC2K assembly language program to multiply two numbers
 - "1m"



Programming Assignment #1

- Where to start...
 - Write some test cases to check your code
 - Program 1: halt
 - Program 2: noop
 - halt
 - Program 3: add 1 1 1
 - halt
 - Program 4: nor 1 1 1
 - halt



Instruction Set Architecture (ISA) Design Lectures

- Lecture 2: ISA storage types, binary and addressing modes
- Lecture 3 : LC2K
- Lecture 4 : ARM
- Lecture 5 : Converting C to assembly basic blocks
- Lecture 6 : Converting C to assembly functions
- Lecture 7: Translation software; libraries, memory layout



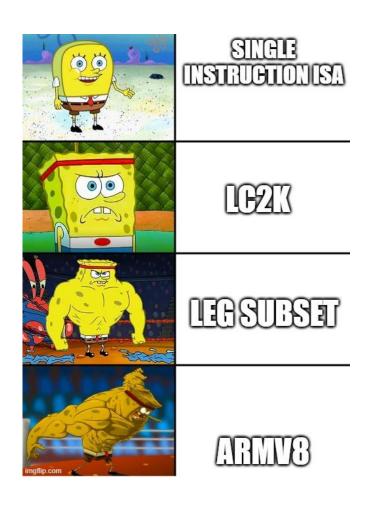
Agenda

- ARM overview and basic instructions
- Memory instructions
 - Handling multiple data widths
- Sample Problems



ARMv8 ISA

- LC2K is intended to be an extremely barebones ISA
 - "Bare minimum"
 - Easy to design hardware for, really annoying to program on (as you'll see in P1m)
 - Invented for our projects, not used anywhere in practice
- ARM (specifically v8) is a much more powerful ISA
 - Used heavily in practice (most smartphones, some laptops & supercomputers)
 - Subset (LEG) is focus of hw and lecture





ISA Types

Reduced Instruction Set Computing (RISC)

- Fewer, simpler instructions
- Encoding of instructions are usually the same size
- Simpler hardware
- Program is larger, more tedious to write by hand
- E.g. LC2K, RISC-V, ARM (kinda)
- More popular now

Complex Instruction Set Computing (CISC)

- More, complex instructions
- Encoding of instructions are different sizes
- More complex hardware
- Short, expressive programs, easier to write by hand
- E.g. x86
- Less popular now



ARM vs LC2K at a Glance

	LC2K	LEG
# registers	8	32
Register width	32 bits	64 bits
Memory size	2 ¹⁸ bytes	2 ⁶⁴ bytes
# instructions	8	40-ish
Addressability	Word	Byte

We'll discuss what this means in a bit



ARM Instruction Set—LEGv8 subset

- The main types of instructions fall into the familiar classes we saw with LC2K:
 - 1. Arithmetic
 - Add, subtract, (multiply not in LEGv8)
 - 2. Data transfer
 - Loads and stores—LDUR (load unscaled register), STUR, etc.
 - 3. Logical
 - AND, ORR, EOR, etc.
 - Logical shifts, LSL, LSR
 - 4. Conditional branch
 - CBZ, CBNZ, B.cond
 - 5. Unconditional branch (jumps)
 - B, BR, BL





LEGv8 Arithmetic Instructions

- Format: three operand fields
 - Dest. register usually the **first one** check instruction format
 - ADD X3, X4, X7 // X3 = X4 + X7
 - LC2K generally has the destination on the right!!!!

• C-code example: f = (g + h) - (i + j)

X1	→t0
VO	14

$$X2 \rightarrow t1$$

$$X5 \rightarrow i$$

ADD	t0, g, h	ADD	X1, X3,
ADD	t1, i, j	ADD	X2, X5,
SUB	f, t0, t1	SUB	X7, X1,



X4

X6

X2

ARMISA

LEGv8 R-instruction Encoding

- Register-to-register operations
- Consider ADD X3, X4, X7
 - R[Rd] = R[Rn] + R[Rm]
 - Rd = X3, Rn = X4, Rm = X7
- Rm = second register operand
- shamt = shift amount
 - not used in LEG for ADD/SUB and set to 0
- Rn = first register operand
- Rd = destination register
- ADD opcode is 10001011000, what are the other fields?

opcode	Rm	shamt	Rn	Rd	
11 bits	5 bits	6 bits	5 bits	5 bits	





I-instruction Encoding

- Format: second source operand can be a register or immediate—a constant in the instruction itself
- e.g., ADD X3, X4, #10 //although we write "ADD", this is "ADDI"
- Format: 12 bits for immediate constants 0-4095

opcode	immediate	Rn	Rd
10 bits	12 bits	5 bits	5 bits

- Don't need negative constants because we have SUBI
- C-code example: f = g + 10

ADDI X7, X5, #10

• C-code example: f = g - 10

SUBI X7, X5, #10





LEGv8 Logical Instructions

- Logical operations are bit-wise
- For example assume
- AND and OR correspond to C operators & and
- For immediate fields the 12 bit constant is padded with zeros to the left—zero extended

Category I	nstructionExample			Meaning	Comments
	and	AND	X1, X2, X3	X1 = X2 & X3	Three reg. operands; bit-by-bit AND
	inclusive or	ORR	X1, X2, X3	X1 = X2 X3	Three reg. operands; bit-by-bit OR
	exclusive or	EOR	X1, X2, X3	X1 = X2 ^ X3	Three reg. operands; bit-by-bit XOR
	and immediate	ANDI	X1, X2, 20	X1 = X2 & 20	Bit-by-bit AND reg. with constant
Logical	inclusive or immediate	ORRI	X1, X2, 20	$X1 = X2 \mid 20$	Bit-by-bit OR reg. with constant
	exclusive or immediate	EORI	X1, X2, 20	X1 = X2 ^ 20	Bit-by-bit XOR reg. with constant
	logical shift left	LSL	X1, X2, 10	X1 = X2 << 10	Shift left by constant
	logical shift right	LSR	X1, X2, 10	X1 = X2 >> 10	Shift right by constant





LEGv8 Shift Logical Instructions

• LSR X6, X23, #2

- C equivalent
 - X6 = X23 >> 2;
- LSL X6, X23, #2
 - What register gets modified?
 - What does it contain after executing the LSL instruction?
- In shift operations Rm is always 0—shamt is 6 bit unsigned

opcode	Rm	shamt	Rn	Rd
11 bits	5 bits	6 bits	5 bits	5 bits





Pseudo Instructions

 Instructions that use a shorthand "mnemonic" that expands to preexisting assembly instruction

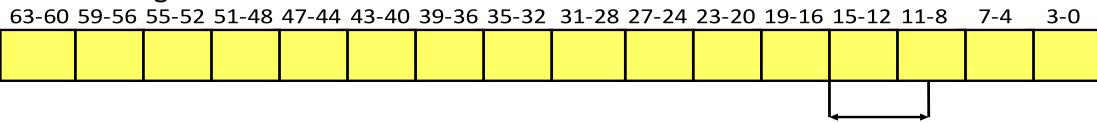
- Example:
 - MOV X12, X2 // the contents of X2 copied to X12—X2 unchanged
- This gets expanded to:
 - ORR X12, XZR, X2
- What alternatives could we use instead of ORR?





Class Problem #1

 Show the C and LEGv8 assembly for extracting the value in bits 15:10 from a 64bit integer variable



Assume the variable is in X1

$$x = x >> 10$$

 $x = x & 0x3F$



Agenda

- ARM overview and basic instructions
- Memory instructions
 - Handling multiple data widths
- Sample Problems



Memory Diagram:

Word vs Byte Addressing

- A word is a collection of bytes
 - Exact size depends on architecture
 - in LC2K and ARM, 4 bytes
 - **Double word** is 8 bytes
- LC2K is word addressable
 - Each address refers to a particular word in memory
 - Wanna move forward one int? Increment address by one
 - Wanna move forward one char? Uhhh...



- ARM (and most modern ISAs) is byte addressable
 - Each address refers to a particular byte in memory
 - Wanna move forward one int? Increment address by four
 - Wanna move forward one char? Increment address by one





LEGv8 Memory Instructions

- Like LC2K, employs base + displacement addressing mode
 - Base is a register
 - Displacement is 9-bit immediate ±256 bytes—sign extended to 64 bits
- Unlike LC2K (which always transfers 4 bytes), we have several options in LEGv8

	•	•		•
Category In	structionExample		Meaning	Comments
	load register	LDUR X1, [X2,40]	X1 = Memory[X2 + 40]	Doubleword from memory to
				register
	store register	STUR X1, [X2,40]	Memory[X2 + 40] = X1	Doubleword from register to
				memory
L	load signed word	LDURSW X1,[X2,40]	X1 = Memory[X2 + 40]	Word from memory to register
	store word	STURW X1, [X2,40]	Memory[X2 + 40] = X1	Word from register to memory
	load half	LDURH X1, [X2,40]	X1 = Memory[X2 + 40]	Halfword memory to register
	store half	STURH X1, [X2,40]	Memory[X2 + 40] = X1	Halfword register to memory
	load byte	LDURB X1, [X2,40]	X1 = Memory[X2 + 40]	Byte from memory to register
L	store byte	STURB X1, [X2,40]	Memory[X2 + 40] = X1	Byte from register to memory
	move wide with zero	MOVZ X1,20, LSL 0	$X1 = 20 \text{ or } 20 * 2^{16} \text{ or } 20$ * $2^{32} \text{ or } 20 * 2^{48}$	Loads 16-bit constant, rest zeros
	move wide with keep	MOVK X1,20, LSL 0	$X1 = 20 \text{ or } 20 * 2^{16} \text{ or } 20$ * $2^{32} \text{ or } 20 * 2^{48}$	Loads 16-bit constant, rest unchanged





D-Instruction fields

- Data transfer
- opcode and op2 define data transfer operation
- address is the ±256 bytes displacement
- Rn is the base register
- Rt is the destination (loads) or source (stores)
- More complicated modes are available in full ARMv8

opcode	address	op2	Rn	Rt
11 bits	9 bits	2 bits	5 bits	5 bits

Look over formatting on your own



Agenda

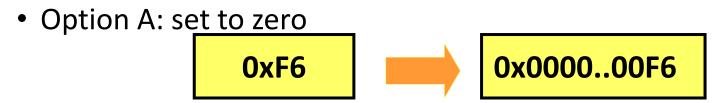
- ARM overview and basic instructions
- Memory instructions
 - Handling multiple data widths
- Sample Problems



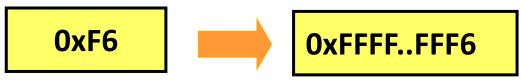
ARMISA

LEGv8 Memory Instructions

- Registers are 64 bits wide
- But sometimes we want to deal with non-64-bit entities
 - E.g. ints (32 bits), chars (8 bits)
- When we load smaller elements from memory, what do we set the other bits to?



• Option B: sign extend



We'll need different instructions for different options





Load Instruction Sizes

How much data is retrieved from memory at the given address?

Desired amount of data to transfer?	Operation	Unused bits in register?	Example
64-bits (double word or whole register)	LDUR (Load unscaled to register)	N/A	0xFEDC_BA98_7654_3210
16-bits (half-word) into lower bits of reg	LDURH	Set to zero	0x0000_0000_0000_ <mark>3210</mark>
8-bits (byte) into lower bits of reg	LDURB	Set to zero	0x0000_0000_0000_00 <mark>10</mark>
32-bits (word) into lower bits of reg	LDURSW (load signed word)	Sign extend (0 or 1 based on most significant bit of transferred word)	0x0000_0000_ 7 654_3210 or 0xFFFF_FFFF_ F 654_3210 (depends on bit 31)





Load Instruction in Action

```
struct {
 int arr[25];
 unsigned char c;
                                         LDURB X3, [X4, #100]
} my struct;
int func() {
 my struct.c++;
 // load value from mem into reg
 // then increment it
                                                                                 10
                                                                                       2600
  X3
                 10
                              Calculate address:
                               2500 + 100 = 2600
  X4
              2500
```



Load Instruction in Action – other example

```
int my big number = -534159618; // 0xE0295EFE in 2's complement
int inc number() {
  my big number++;
                                             LDURSW X3, [X4, #0]
  // load value from mem into reg
  // then increment it
};
                         Sign extend (0xE0295EFE) to
                         64 bits \rightarrow 0x FFFFFFFE0295EFE
                                                                     FE
                                                                            2604
 X3
                                                                     5E
                                                                            2605
      FFFF...5EFE
                         Calculate address:
                         2604 + 0 = 2604
                                                                     29
                                                                            2606
 X4
           2604
                                                                     E0
                                                                            2607
          Need to sign extend,
```



But wait...

int my_big_number = -534159618; // 0xE0295EFE in 2's complement

• If I want to store this number in memory... should it be stored like this?

FE	2604
5E	2605
29	2606
EO	2607

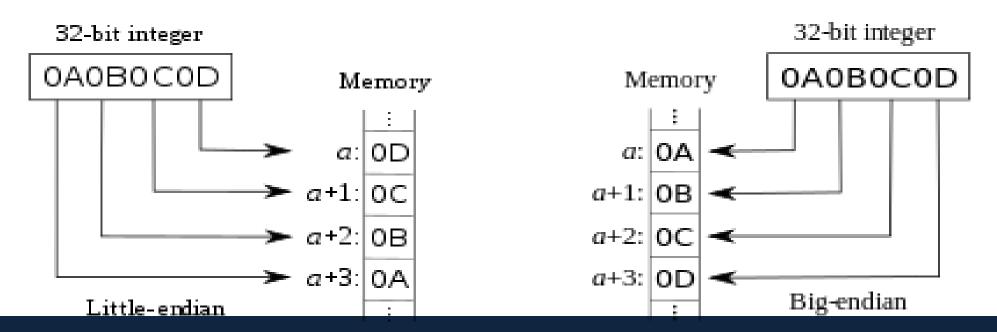
• ... or like this?

EO	2604
29	2605
58	2606
FE	2607



Big Endian vs. Little Endian

- Endian-ness: ordering of bytes within a word
 - Little Bigger address holds more significant bits
 - Big –Opposite, smaller address hold more significant bits
 - The Internet is big endian, x86 is little endian, LEG and ARMv8 can switch
 - But in general assume little endian. (Figures from Wikipedia)







Store Instructions

• Store instructions are simpler—there is no sign/zero extension to consider (do you see why?)

Desired amount of data to transfer?	Operation	Example
64-bits (double word or whole register)	STUR (Store unscaled register)	0xFEDC_BA98_7654_3210
16-bits (half-word) from lower bits of reg	STURH	0x0000_0000_0000_ <mark>3210</mark>
8-bits (byte) from lower bits of reg	STURB	0x0000_0000_0000_00 <mark>10</mark>
32-bits (word) from lower bits of reg	STURW	0xFFFF_FFFF_ F 654_3210



Agenda

- ARM overview and basic instructions
- Memory instructions
 - Handling multiple data widths
- Sample Problems





What is the final state of memory once you execute the following

instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STUR	X3, [X5, #100]
STURB	X4, [X5, #102]

	register file
Х3	
X4	

little e	ndiar I
0x02	100
0x03	101
0xFF	102
0x05	103
0xC2	104
0x06	105
0xFF	106
0xE5	107





What is the final state of memory once you execute the following

instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STUR	X3, [X5, #100]
STURB	X4, [X5, #102]

	register file
Х3	
X4	0xE5FF06C205FF0302

	O
little e	ndiar I
0x02	100
0x03	101
OxFF	102
0x05	103
0xC2	104
0x06	105
OxFF	106
0xE5	107





What is the final state of memory once you execute the following

instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STUR	X3, [X5, #100]
STURB	X4, [X5, #102]

register file	
Х3	0x000000000000FF
X4	0xE5FF06C205FF0302

	O
little e	ndiar I
0x02	100
0x03	101
0xFF	102
0x05	103
0xC2	104
0x06	105
OxFF	106
0xE5	107





What is the final state of memory once you execute the following

instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STUR	X3, [X5, #100]
STURB	X4, [X5, #102]

	register file
Х3	0x000000000000FF
X4	0xE5FF06C205FF0302

little e	o ndian
OxFF	100
0x00	101
0x00	102
0x00	103
0x00	104
0x00	105
0x00	106
0x00	107





What is the final state of memory once you execute the following

instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STUR	X3, [X5, #100]
STURB	X4, [X5, #102]

	register file
Х3	0x000000000000FF
X4	0xE5FF06C205FF0302

little e	ndiar I
0xFF	100
0x00	101
0x02	102
0x00	103
0x00	104
0x00	105
0x00	106
0x00	107

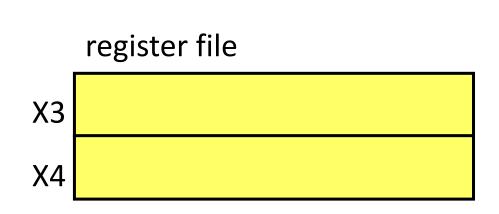




What is the final state of memory once you execute the following instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STURB	X3, [X5, #103]
LDURSW	X4, [X5, #100]

We shown the registers as blank. What do they actually contain before we run the snippet of code?



little er I	ndian I
0x02	100
0x03	101
OxFF	102
0x05	103
0xC2	104
0x06	105
0xFF	106
0xE5	107





What is the final state of memory once you execute the following instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STURB	X3, [X5, #103]
LDURSW	X4, [X5, #100]

	register file
X3	
X4	0xE5FF06C205FF0302
'	

little er	ndian I
0x02	100
0x03	101
OxFF	102
0x05	103
0xC2	104
0x06	105
0xFF	106
0xE5	107





What is the final state of memory once you execute the following instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STURB	X3, [X5, #103]
LDURSW	X4, [X5, #100]

	register file
Х3	0x000000000000FF
X4	0xE5FF06C205FF0302

little endian		
0x02	100	
0x03	101	
OxFF	102	
0x05	103	
0xC2	104	
0x06	105	
OxFF	106	
0xE5	107	





What is the final state of memory once you execute the following instruction sequence? (assume X5 has the value of 0)

X4, [X5, #100]
X3, [X5, #102]
X3, [X5, #103]
X4, [X5, #100]

	register file
Х3	0x000000000000FF
X4	0xE5FF06C205FF0302

little endian		
0x02	100	
0x03	101	
0xFF	102	
0xFF	103	
0xC2	104	
0x06	105	
0xFF	106	
0xE5	107	





What is the final state of memory once you execute the following instruction sequence? (assume X5 has the value of 0)

LDUR	X4, [X5, #100]
LDURB	X3, [X5, #102]
STURB	X3, [X5, #103]
LDURSW	X4, [X5, #100]

	register file
Х3	0x000000000000FF
X4	0xFFFFFFFFFF0302

Memory	
(each location is 1	byte)

little endian		
0x02	100	
0x03	101	
0xFF	102	
0xFF	103	
0xC2	104	
0x06	105	
0xFF	106	
0xE5	107	



Next Time

- More examples on doing stuff in ARM assembly
 - Like if/else, while loops, etc

