Assembly Language

Readings: 2.1-2.7, 2.9-2.10, 2.14
Green reference card

Assembly language
Simple, regular instructions – building blocks of C, Java & other languages
Typically one-to-one mapping to machine language

Our goal
Understand the basics of assembly language
Help figure out what the processor needs to be able to do

Not our goal to teach complete assembly/machine language programming
Floating point
Procedure calls
Stacks & local variables

Aside: C/C++ Primer

```c
struct coord { int x, y; }; /* Declares a type */
struct coord start; /* Object with two slots, x and y */
start.x = 1; /* For objects "." accesses a slot */
struct coord *myLoc; /* "*" is a pointer to objects */
myLoc = &start; /* "+" returns thing's location */
myLoc->y = 2; /* "+->" is "++" plus "." */
```

```c
int scores[8]; /* 8 ints, from 0..7 */
scores[1]=5; /* Access locations in array */
int *index = scores; /* Points to scores[0] */
index++; /* Next scores location */
(*index)++; /* "++" works in arrays as well */
index = &(scores[3]); /* Points to scores[3] */
*index = 9;
```

0 1 2 3 4 5 6 7
ARM Assembly Language

The basic instructions have four components:
- Operator name
- Destination
- 1<sup>st</sup> operand
- 2<sup>nd</sup> operand

ADD <dst>, <src1>, <src2>  // <dst> = <src1> + <src2>
SUB <dst>, <src1>, <src2>  // <dst> = <src1> - <src2>

Simple format: easy to implement in hardware

More complex: \( A = B + C + D - E \)

Operands & Storage

For speed, CPU has 32 general-purpose registers for storing most operands
For capacity, computer has large memory (multi-GB)

Load/store operation moves information between registers and main memory
All other operations work on registers
Registers

32x 64-bit registers for operands

<table>
<thead>
<tr>
<th>Register</th>
<th>Function</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0-X7</td>
<td>Function arguments/Results</td>
<td></td>
</tr>
<tr>
<td>X8</td>
<td>Result, if a pointer</td>
<td></td>
</tr>
<tr>
<td>X9-X15</td>
<td>Volatile Temporaries</td>
<td>Not saved on call</td>
</tr>
<tr>
<td>X16-X17</td>
<td>Linker scratch registers</td>
<td>Don’t use them</td>
</tr>
<tr>
<td>X18</td>
<td>Platform register</td>
<td>Don’t use this</td>
</tr>
<tr>
<td>X19-X27</td>
<td>Temporaries (saved across calls)</td>
<td>Saved on call</td>
</tr>
<tr>
<td>X28</td>
<td>Stack Pointer</td>
<td></td>
</tr>
<tr>
<td>X29</td>
<td>Frame Pointer</td>
<td></td>
</tr>
<tr>
<td>X30</td>
<td>Return Address</td>
<td></td>
</tr>
<tr>
<td>X31</td>
<td>Always 0</td>
<td>No-op on write</td>
</tr>
</tbody>
</table>

Basic Operations

(Note: just subset of all instructions)

Mathematic: ADD, SUB, MUL, SDIV

Immediate (one input a constant)

ADD X0, X1, X2 // X0 = X1+X2
ADDI X0, X1, #100 // X0 = X1+100

Logical: AND, ORR, EOR

Immediate

AND X0, X1, X2 // X0 = X1∧X2
ANDI X0, X1, #7 // X0 = X1∧b0111

Shift: left & right logical (LSL, LSR)

LSL X0, X1, #4 // X0 = X1<<4

Example: Take bits 6-4 of X0 and make them bits 2-0 of X1, zeros otherwise:
Memory Organization

Viewed as a large, single-dimension array, with an address.
A memory address is an index into the array
"Byte addressing" means that the index points to a byte of memory.

```
0   8 bits of data
1   8 bits of data
2   8 bits of data
3   8 bits of data
4   8 bits of data
5   8 bits of data
6   8 bits of data
...```

Memory Organization (cont.)

Bytes are nice, but most data items use larger units.
Double-word = 64 bits = 8 bytes
Word = 32 bits = 4 bytes

```
0   64 bits of data
8   64 bits of data
16  64 bits of data
24  64 bits of data
```

Registers hold 64 bits of data

$2^{64}$ bytes with byte addresses from 0 to $2^{64}-1$
$2^{51}$ double-words with byte addresses 0, 8, 16, ... $2^{64}-8$
Double-words and words are aligned
i.e., what are the least 3 significant bits of a double-word address?
Addressing Objects: Endian and Alignment

Big Endian: address of most significant byte = doubleword address
Motorola 68k, MIPS, IBM 360/370, Xilinx Microblaze, Sparc

Little Endian: address of least significant byte = doubleword address
Intel x86, DEC Vax, Altera Nios II, Z80

ARM: can do either – this class assumes Little-Endian.

Data Storage

Characters: 8 bits (byte)
Integers: 64 bits (word)
Array: Sequence of locations
Pointer: Address (64 bits)

// G = ASCII 71
char a = 'G';
int x = 258;
char *b;
int *y;
b = new char[4];
y = new int[10];

(Note: real compilers place local variables (the “stack”) from beginning of memory, new’ed structures (the “heap”) from end. We ignore that here for simplicity)
**Loads & Stores**

Loads & Stores move data between memory and registers. All operations on registers, but too small to hold all data.

LDUR X0, [X1, #14]  // X0 = Memory[X1+14]

STUR X2, [X3, #20]  // Memory[X3+20] = X2

<table>
<thead>
<tr>
<th>General Purpose Registers</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0: 130</td>
</tr>
<tr>
<td>X1: 723</td>
</tr>
<tr>
<td>X2: 4</td>
</tr>
<tr>
<td>X3: 66</td>
</tr>
</tbody>
</table>

Note: LDURB & STURB load & store bytes.

**Addressing Example**

The address of the start of a character array is stored in X0. Write assembly to load the following characters:

X2 = Array[0]

X3 = Array[1]

X4 = Array[2]

X5 = Array[k]  // Assume the value of k is in X1
```c
/* Swap the kth and (k+1)th element of an array */
swap(int v[], int k) {
    int temp = v[k];
    v[k] = v[k+1];
    v[k+1] = temp;
}

// Assume v in X0, k in X1
```

**Array Example**

**Execution Cycle Example**

PC: Program Counter
IR: Instruction Register

Note: Word addresses
Instructions are 32b
Flags/Condition Codes

Flag register holds information about result of recent math operation
- Negative: was result a negative number?
- Zero: was result 0?
- Overflow: was result magnitude too big to fit into 64-bit register?
- Carry: was the carry-out true?

Operations that set the flag register contents:
- ADDS, ADDIS, ANDS, ANDIS, SUBS, SUBIS, some floating point.

Most commonly used are subtracts, so we have a synonym: CMP
  CMP X0, X1    same as SUBS X31, X0, X1
  CMP X0, #15  same as SUBIS X31, X0, #15

Control Flow

Unconditional Branch – GOTO different next instruction
- B START     // go to instruction labeled with "START" tag
- BR X30      // go to address in X30: PC = value of X30

Conditional Branches – GOTO different next instruction if condition is true
1 register: CBZ (==0), CBNZ (!= 0)
  CBZ X0, FOO  // if X0 == 0 GOTO FOO: PC = FOO

2 register: B.LT (<), B.LE(<=), B.GE (>=), B.GT(>), B.EQ(==), B.NE(!=)
  first compare (CMP X0, X1, CMPI X0, #12), then b.cond instruction
  CMP X0, X1  // compare X0 with X1 - same as SUBS X31, X0, X1
  B.EQ FOO   // if X0 == X1 GOTO FOO: PC = FOO

if (a == b)
  // X0 = a, X1 = b, X2 = c
  a = a + 3;  
  CMP X0, X1  // set flags
  B.NE ELSEIF // branch if a!=b
else
  b = b + 7;  
  ADDI X0, X0, 3  // a = a + 3
  B DONE      // avoid else
  c = a + b;  
  ELSEIF:     
    ADDI X1, X1, 7  // b = b + 7
  DONE:       
    ADD, X2, X0, X1  // c = a + b
**Loop Example**

Compute the sum of the values 0...N-1

```c
int sum = 0;
for (int I = 0; I != N; I++) {
    sum += I;
}
// X0 = N, X1 = sum, X2 = I
```

**String toUpper**

Convert a string to all upper case

```c
char *index = string;
while (*index != 0) { /* C strings end in 0 */
    if (*index >= 'a' && *index <= 'z')
        *index = *index + ('A' - 'a');
    index++;
}
// string is a pointer held at Memory[80].
// X0=index, 'A' = 65, 'a' = 97, 'z' = 122
```
### Machine Language vs. Assembly Language

<table>
<thead>
<tr>
<th>Assembly Language</th>
<th>Machine Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>mnemonic for easy reading</td>
<td>Completely numeric representation</td>
</tr>
<tr>
<td>labels instead of fixed addresses</td>
<td>format CPU actually uses</td>
</tr>
<tr>
<td>Easier for programmers</td>
<td></td>
</tr>
<tr>
<td>Almost 1-to-1 with machine language</td>
<td></td>
</tr>
</tbody>
</table>

**SWAP:**

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Machine Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSL X9, X1, #3</td>
<td>11010011011 00000000011 00001 01001</td>
</tr>
<tr>
<td>ADD X9, X0, X9 // Compute address of v[k]</td>
<td>10001011000 01001 0000000001 00001 01001</td>
</tr>
<tr>
<td>LDUR X10, [X9, #0] // get v[k]</td>
<td>11111000010 0000000000 00 01001 01010</td>
</tr>
<tr>
<td>LDUR X11, [X9, #8] // get v[k+1]</td>
<td>11111000010 0000010000 00 01001 01011</td>
</tr>
<tr>
<td>STUR X11, [X9, #0] // save new value to v[k]</td>
<td>11111000000 0000000000 00 01001 01011</td>
</tr>
<tr>
<td>STUR X10, [X9, #8] // save new value to v[k+1]</td>
<td>11111000000 0000010000 00 01001 01010</td>
</tr>
<tr>
<td>BR X30 // return from subroutine</td>
<td>11010100000 0000000000 0000000000 00000011110</td>
</tr>
</tbody>
</table>

### Labels

Labels specify the address of the corresponding instruction

Programmer doesn’t have to count line numbers

Insertion of instructions doesn’t require changing entire code

```assembly
// X0 = N, X1 = sum, X2 = I
ADD X1, X31, X31 // sum = 0
ADD X2, X31, X31 // I = 0

TOP:
CMP X2, X0 // Check I vs N
B.GE END // end when !(I<N)
ADD X1, X1, X2 // sum += I
ADDI X2, X2, #1 // I++
B TOP // next iteration

END:
```

Notes:

Branches are PC-relative

PC = PC + 4*(BranchOffset)

BranchOffset positive -> branch downward. Negative -> branch upward.
Labels Example

Compute the value of the labels in the code below.
Branches: PC = PC + 4*(BranchOffset)

// Program starts at address 100
LDUR X0, [X31, #100]
LOOP:
  LDURB X1, [X0, #0]
  CBZ X1, END
  CMPI X1, #97
  B.LT NEXT
  CMPI X1, #122
  B.GT NEXT
  SUBI X1, X1, #32
  STURB X1, [X0, #0]
NEXT:
  ADDI X0, X0, 1
  B LOOP
END:

Instruction Types

Can group instructions by # of operands

3-register
ADD X0, X1, X2
ADDI X0, X1, #100
AND X0, X1, X2
ANDI X0, X1, #7
LSL X0, X1, #4
LSR X0, X1, #2
LDUR X0, [X1, #14]

2-register
LDURB X0, [X1, #14]
STUR X0, [X1, #14]
STURB X0, [X1, #14]
B START
BR X30

1-register
CBZ X0, FOO
B.EQ DEST

0-register
**Instruction Formats**

All instructions encoded in 32 bits (operation + operands/immediates)

- **Branch (B-Type)**
  - \( \text{Instr}[31:21] = 0A0-0BF \)
  - BrAddr26

- **Conditional Branch (CB-Type)**
  - \( \text{Instr}[31:21] = 2A0-2A7, 5A0-5AF \)

- **Register (R-Type)**

- **Immediate (I-Type)**

- **Memory (D-Type)**
  - \( \text{Instr}[31:21] = 1C0-1C2, 7C0-7C2 \)

---

**B-Type**

Used for unconditional branches

\[ \text{BrAddr26} \]

0x05: B

\[ \text{PC} = \text{PC} + 4 \times -3 \]
### CB-Type

**Used for conditional branches**

<table>
<thead>
<tr>
<th>Opcode</th>
<th>CondAddr19</th>
<th>Rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x54:</td>
<td></td>
<td>B.cond</td>
</tr>
<tr>
<td>0xB4:</td>
<td></td>
<td>CBZ</td>
</tr>
<tr>
<td>0xB5:</td>
<td></td>
<td>CBNZ</td>
</tr>
</tbody>
</table>

**Condition Codes**

- 0x00: EQ (=)
- 0x01: NE (!=)
- 0x0A: GE (>=)
- 0x0B: LT (<)
- 0x0C: GT (>)
- 0x0D: LE (<=)

**Examples**

- `CBZ X12, -3` // if(X12==0) PC = PC + 4*-3
- `B.LT -5` // if (lessThan) PC = PC + 4*-5

### R-Type

**Used for 3 register ALU operations and shift**

<table>
<thead>
<tr>
<th>Opcode</th>
<th>Rm</th>
<th>SHAMT</th>
<th>Rn</th>
<th>Rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x450:</td>
<td>AND</td>
<td>Op2</td>
<td>Op1</td>
<td>Dest</td>
</tr>
<tr>
<td>0x458:</td>
<td>ADD</td>
<td>(0 for shift)</td>
<td>(0 for non-shift)</td>
<td></td>
</tr>
<tr>
<td>0x4D6:</td>
<td>SDIV, shamt=02</td>
<td>ADD</td>
<td>X3, X5, X6</td>
<td>// X3 = X5+X6</td>
</tr>
<tr>
<td>0x4D8:</td>
<td>MUL, shamt=1F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x550:</td>
<td>ORR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x558:</td>
<td>ADDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x650:</td>
<td>EOR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x658:</td>
<td>SUB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x69A:</td>
<td>LSR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x69B:</td>
<td>LSL</td>
<td>LSL</td>
<td>X10, X4, #6</td>
<td>// X10 = X4&lt;&lt;6</td>
</tr>
<tr>
<td>0x6B0:</td>
<td>BR, rest all 0's but Rd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x750:</td>
<td>ANDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x758:</td>
<td>SUBS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## I-Type

Used for 2 register & 1 constant ALU operations

<table>
<thead>
<tr>
<th>Opcode</th>
<th>ALU_Imm12</th>
<th>Rn</th>
<th>Rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x244: ADDI</td>
<td>Constant - Op2</td>
<td>Op1</td>
<td>Dest</td>
</tr>
<tr>
<td>0x248: ANDI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x164: ADDIS</td>
<td>ADDI X8, X3, #35 // X8 = X3 + 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x168: ORRI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x344: SUBI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x348: EORI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x2C4: SUBIS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x2C8: ANDIS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LDUR X6, [X15, #12] // X6 = Memory[X15+12]

## D-Type

Used for memory accesses

<table>
<thead>
<tr>
<th>Opcode</th>
<th>DAddr9</th>
<th>00</th>
<th>Rn</th>
<th>Rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x1C0: STURB</td>
<td>Address Constant</td>
<td>Address Reg</td>
<td>Target Reg</td>
<td></td>
</tr>
<tr>
<td>0x1C2: LDURB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x7C0: STUR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x7C2: LDUR</td>
<td>LDUR X6, [X15, #12] // X6 = Memory[X15+12]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LDUR X6, [X15, #12] // X6 = Memory[X15+12]
Conversion example

Compute the sum of the values 0…N-1

ADD X1, X31, X31
ADD X2, X31, X31
B TEST
TOP:
ADD X1, X1, X2
ADDI X2, X2, #1
TEST:
SUBS X31, X2, X0
B.LT TOP
END:

Assembly & Machine Language

Assembly

Machine Language