COS 217: Introduction to Programming Systems

Assembly Language

Part 1
Context of this Lecture

“Under the hood”

- C Language
  - Assembly Language
    - Machine Language
Agenda

Language Levels

Architecture

Assembly Language: Performing Arithmetic

Assembly Language: Load/Store and Defining Global Data
High-Level Languages

Characteristics

• Portable (to varying degrees)
• Complex
  • One statement can do a lot of work – good ratio of functionality to code size
• Human readable
  • Structured: if(), for(), while(), etc.
  • Variable names can hide details of where data is stored (stack, heap, etc.)
• Type system allows compiler to check usage details without burdening reader

```c
int collatz(int n)
{
    int count = 0;
    while (n > 1) {
        count++;
        if (n & 1)
            n = 3 * n + 1;
        else
            n /= 2;
    }
    return count;
}
```
Machine Languages

Characteristics

- Not portable (hardware-specific)
- Simple
  - Each instruction does a simple task – poor ratio of functionality to code size
- Not human readable
  - Not structured
  - Requires lots of effort!
  - Requires tool support

```
0000 0000 0000 0000 0000 0000 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000
9222 9120 1121 A120 1121 A121 7211 0000
0000 0001 0002 0003 0004 0005 0006 0007
0008 0009 000A 000B 000C 000D 000E 000F
0000 0000 0000 FE10 FACE CAFE ACED CEDE
0000 0000 0000 0000 0000 0000 0000 0000
1234 5678 9ABC DEF0 0000 0000 F00D 0000
0000 0000 EEEE 1111 EEEE 1111 0000 0000
B1B2 F1F5 0000 0000 0000 0000 0000 0000
```
## Assembly Languages

### Characteristics

- **Not portable**
  - Each assembly language instruction maps to one machine instruction
- **Simple**
  - Each instruction does a simple task
- **Human readable**
  (In the same sense that Polish is human readable ... if you know Polish.)

```assembly
mov    w1, 0
loop:
    cmp   w0, 1
    ble   endloop
    add   w1, w1, #1
    ands  wzr, w0, #1
    beq   else
    add   w2, w0, w0
    add   w0, w0, w2
    add   w0, w0, 1
    b     endif
else:
    asr   w0, w0, 1
endif:
    b     loop
endloop:
```
Why Learn Assembly Language?

Knowing assembly language helps you:

- **Write faster code**
  - In assembly language
  - In a high-level language!
- **Write safer code**
  - Understanding mechanism of potential security problems helps you avoid them – even in high-level languages
- **Understand what’s happening “under the hood”**
  - Someone needs to develop future computer systems
  - Maybe that will be you!
- **Become more comfortable with levels of abstraction**
  - Become a better programmer at all language levels!
Why Learn ARM Assembly Lang?

Why learn ARMv8 (a.k.a. AARCH64 or A64) assembly language?

Pros

• ARM is the most widely used processor architecture in the world
  (in your phone, in your Mac, in your Chromebook, in Armlab,
  in internet-of-things devices)

• ARM has a modern and (relatively) elegant instruction set,
  compared to the expansive but ugly x86-64 instruction set

Cons

• x86-64 still has a huge presence in desktop/laptop/cloud (for now?)
## Lectures vs. Precepts

### Approach to studying assembly language:

<table>
<thead>
<tr>
<th>Lectures</th>
<th>Precepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study partial programs</td>
<td>Study complete programs</td>
</tr>
<tr>
<td>Begin with simple constructs; proceed to complex ones</td>
<td>Begin with small programs; proceed to large ones</td>
</tr>
<tr>
<td>Emphasis on reading code</td>
<td>Emphasis on writing code</td>
</tr>
</tbody>
</table>
Agenda

Language Levels

Architecture

Assembly Language: Performing Arithmetic

Assembly Language: Load/Store and Defining Global Data
John von Neumann (1903-1957)

In computing
- Stored program computers
  - Cellular automata, self-replication,
  - Game theory
  - Mergesort

Other interests
- Mathematics, statistics, game theory
- Nuclear physics

Princeton connection
- Princeton University & IAS, 1930-1957
  - [https://paw.princeton.edu/article/early-history-computing-princeton](https://paw.princeton.edu/article/early-history-computing-princeton)

Known for the “Von Neumann architecture”
- In which (machine-language) programs are just data in memory
- A.K.A. “Princeton architecture” – contrast to the now-mostly-obsolete “Harvard architecture”
Instructions (encoded within words) are fetched from RAM

Control unit interprets instructions:

• to shuffle data between registers and RAM

• to move data from registers to ALU (arithmetic+logic unit) where operations are performed
Von Neumann Architecture

Registers
Small amount of storage on the CPU
- Top of the “storage hierarchy”
- Very {small, expensive, fast}

ALU instructions operate on registers
ALU Arithmetic Example

Operation: src1 \rightarrow ALU \rightarrow dest
Operation: src2 \rightarrow ALU \rightarrow dest
Operation: ALU \rightarrow flags

RAM
Control Unit
Registers
CPU
Data bus
Von Neumann Architecture

RAM (Random Access Memory)
Conceptually: large array of bytes
(gigabytes+ in modern machines)

• Contains data
  (program variables, structs, arrays)
• and the program!

Instructions are fetched from RAM
Time to reminisce about old TOYs

Thinking back to COS 126, how did you feel about TOY?

A. Loved it!

B. Wasn't a fan.

C. I placed out, so I have no idea what you're talking about.
Time to reminisce about old TOYS

TOY REFERENCE CARD

INSTRUCTION FORMATS

| Format RR: | opcode | d | s | t | (0-6, A-B) |
| Format A:  | opcode | d | addr | (7-9, C-F) |

ARITHMETIC and LOGICAL
1: add
2: subtract
3: and
4: xor
5: shift left
6: shift right

TRANSFER between registers
7: load address
8: load
9: store
A: load indirect
B: store indirect

CONTROL
0: halt
C: branch zero
D: branch positive
E: jump register
F: jump and link

Word size. The TOY machine has two types of storage: main memory and registers. Each entity stores one word of information. On the TOY machine, a word is a sequence of 16 bits. Typically, we interpret these 16 bits as a hexadecimal integer in the range 0000 through FFFF. Using two’s complement notation, we can also interpret it as a decimal integer in the range -32,768 to +32,767. See Section 5.1 for a refresher on number representations and two’s complement integers.

Main memory. The TOY machine has 256 words of main memory. Each memory location is labeled with a unique memory address. By convention, we use the 256 hexadecimal integers in the range 000 through FFF. Think of a memory location as a mailbox, and a memory address as a postal address. Main memory is used to store instructions and data.

Registers. The TOY machine has 16 registers, indexed from 0 through F. Registers are much like main memory: each register stores one 16-bit word. However, registers provide a faster form of storage than main memory. Registers are used as scratch space during computation and play the role of variables in the TOY language. Register 0 is a special register whose output value is always 0.

Program counter. The program counter or PC is an extra register that keeps track of the next instruction to be executed. It stores 8 bits, corresponding to a hexadecimal integer in the range 00 through FF. This integer stores the memory address of the next instruction to execute.

https://introcs.cs.princeton.edu/java/62toy/
Registers and RAM

Typical pattern:
- **Load** data from RAM to registers
- **Manipulate** data in registers
- **Store** data from registers to RAM

On AARCH64, this pattern is enforced
- “Manipulation” instructions can only access registers
- This is known as a **load-store architecture**
  (as opposed to “register-memory” architectures)
- Characteristic of “RISC” (Reduced Instruction Set Computer) vs. “CISC” (Complex Instruction Set Computer) architectures, e.g. x86
Registers (ARM-64 architecture)

- x0
- w0
- x1
- w1
- ...
- x30 (LR)
- w30
- x29 (FP)
- w29
- xzr (all zeros)
- wzr
- sp (stack pointer)
- pc (program counter)

nzcv
pstate
General-Purpose 64-bit Registers

X0 ... X30

- Scratch space for instructions, parameter passing to/from functions, return address for function calls, etc.
- Some have special roles defined in hardware (e.g. X30) or defined by software convention (e.g. X29)
- Also available as 32-bit versions: W0 .. W30

XZR

- On read: all zeros
- On write: data thrown away
- Also available as 32-bit version: WZR
Special-purpose register...

- **SP (Stack Pointer):**
  Contains address of top (low memory address) of current function’s stack frame

Allows use of the STACK section of memory

(See *Assembly Language: Function Calls* lecture later)
PC Register

Special-purpose register...

- **PC (Program Counter)**
  - Stores the location of the next instruction
    - Address (in TEXT section) of machine-language instruction to be executed next
  - Value changed:
    - Automatically to implement sequential control flow (increment by 4 bytes)
    - By branch instructions to implement selection, repetition
Special-purpose register...

- Contains condition flags:
  - n (Negative), z (Zero), c (Carry), v (Overflow)
- Affected by compare (cmp) instruction
  - And many others, if requested
- Used by conditional branch instructions
  - beq, bne, blo, bhi, ble, bge, ...
- (See Assembly Language: Part 2 lecture)
Agenda

Language Levels
Architecture

**Assembly Language: Performing Arithmetic**

Assembly Language: Load/Store and Defining Global Data
ALU Arithmetic Example

operation → ALU → flags

src1 → ALU → dest

src2 → ALU → dest

RAM

Control Unit

ALU

Registers

CPU

Data bus
Many instructions have this format:

\[
\text{name}{,s} \ \text{dest}, \ \text{src1}, \ \text{src2} \\
\text{name}{,s} \ \text{dest}, \ \text{src1}, \ \text{immed}
\]

- **name:** name of the instruction (add, sub, mul, and, etc.)
- **s:** if present, specifies that condition flags should be Set

- dest and src1, src2 are \text{x} registers: 64-bit operation
- dest and src1, src2 are \text{w} registers: 32-bit operation
  - No mixing and matching between \text{x} and \text{w} registers
- src2 may be a constant (“immediate” value) instead of a register
64-bit Arithmetic

C code:

```c
static long length;
static long width;
static long perim;
...
perim =
  (length + width) * 2;
```

Assume that...

- there’s a good reason for having variables with file scope, process duration
- length held in x1
- width held in x2
- perim held in x3

We’ll see later how to make this happen

Assembly code:

```
add x3, x1, x2
lsl x3, x3, 1
```

Recall use of left shift by 1 bit to multiply by 2
More Arithmetic

```c
static long x;
static long y;
static long z;
...
z = x - y;
z = x * y;
z = x / y;
z = x & y;
z = x | y;
z = x ^ y;
z = x >> y;
```

Assume that...
- x held in x1
- y held in x2
- z held in x3

Assembly code:
```
sub x3, x1, x2
mul x3, x1, x2
sdiv x3, x1, x2
and x3, x1, x2
orr x3, x1, x2
eor x3, x1, x2
asr x3, x1, x2
```

Not xor!
More Arithmetic: Shortcuts

```
static long x;
static long y;
static long z;
...
z = x;
z = -x;
```

Assume that...
- x held in x1
- y held in x2
- z held in x3

Assembly code:

```
mov x3, x1
neg x3, x1
orr x3, xzr, x1
sub x3, xzr, x1
```

These are actually assembler shortcuts for instructions with XZR!
Signed vs Unsigned?

```
static long x;
static unsigned long y;
...
++x;
y--;
```

Assume that...
- x held in x1
- y held in x2

Assembly code:
```
add x1, x1, 1
sub x2, x2, 1
```

Mostly the same algorithms, same instructions!
- Can set different condition flags in PSTATE
- But some exceptions...
Assume that...
- x held in x1
- y held in x2

Assembly code:
- `sdiv x1, x1, 17`
- `udiv x2, x2, 42`
- `asr x1, x1, 1`
- `lsr x2, x2, 2`

"Arithmetic" right shift (shift in sign bit on left) vs. "logical" right shift (shift in zeros on left)
32-bit Arithmetic using “w” registers

C code:

```c
static int length;
static int width;
static int perim;
...
perim =
  (length + width) * 2;
```

Assume that...

- length held in w1
- width held in w2
- perim held in w3

Assembly code:

```asm
add w3, w1, w2
lsl w3, w3, 1
```
8- and 16-bit Arithmetic?

```c
static char x;
static short y;
...
++x;
--y;
```

No specialized arithmetic instructions

- Use “w” registers
- Specialized “load” and “store” instructions for transfer of shorter data types from / to memory – we’ll see these later
- Corresponds to C language semantics: all arithmetic is implicitly done on (at least) ints
Agenda

Language Levels
Architecture
Assembly Language: Performing Arithmetic
Assembly Language: Load/Store and Defining Global Data
Loads and Stores

Most basic way to load (from RAM) and store (to RAM):

\[
\begin{align*}
\text{ldr } & \text{dest, [src]} \\
\text{str } & \text{src, [dest]}
\end{align*}
\]

- dest and src are registers!
  - The addresses (src for \texttt{ldr}, dest for \texttt{str}) must be x-flavored
  - The other operands (dest for \texttt{ldr}, src for \texttt{str}) can be x-flavored or w-flavored
- Contents of registers in [brackets] must be memory addresses
  - Every memory access is through a “pointer”!

\[
\begin{align*}
\text{ldr } & \text{dest, [src]} \\
\text{str } & \text{src, [dest]}
\end{align*}
\]
Signed vs Unsigned, 8- and 16-bit

Special instructions for reading/writing **Bytes** (8 bit) and shorts ("**Half-words**": 16 bit)
- See appendix of these slides for information on ordering:
  - little-endian vs. big-endian

Special instructions for signed reads
- "**Sign-extend**” byte, half-word, or word to 32 or 64 bits
Loads and Stores

Most basic way to load (from RAM) and store (to RAM):

\[
\begin{align*}
\text{ldr} & \text{ dest, [src]} \\
\text{str} & \text{ src, [dest]}
\end{align*}
\]

• dest and src are registers!
  • The addresses (src for ldr, dest for str) must be x-flavored
  • The other operands (dest for ldr, src for str) can be x-flavored or w-flavored

• Contents of registers in [brackets] must be memory addresses
  • Every memory access is through a “pointer”!

• How to get correct memory address into register?
  • Depends on whether data is on stack (local variables), heap (dynamically-allocated memory), or global / static
  • For today, we’ll look only at the global / static case
Our First Full Program*

```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
    (length + width) * 2;
    return 0;
}
```

```assembly
.section .data
length: .word 1
width: .word 2
perim: .word 0

.section .text
.global main
main:
    adr    x0, length
    ldr    w1, [x0]
    adr    x0, width
    ldr    w2, [x0]
    add    w1, w1, w2
    lsl    w1, w1, 1
    adr    x0, perim
    str    w1, [x0]
    mov    w0, 0
    ret
```

* Sorry, I know by convention it should be “Hello, World!”. You’ll see that in precept.
Memory sections

```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
```

Sections (Stack/heap are different!)

- `.rodata`: read-only
- `.data`: read-write
- `.bss`: read-write (initialized to 0)
- `.text`: read-only, program code

```
.section .data
length: .word 1
width: .word 2
perim: .word 0

.section .text
.global main
main:
    adr  x0, length
    ldr  w1, [x0]
    adr  x0, width
    ldr  w2, [x0]
    add  w1, w1, w2
    lsl  w1, w1, 1
    adr  x0, perim
    str  w1, [x0]
    mov  w0, 0
    ret
```
Variable definitions

```
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
```

Declaring data

“Labels” for locations in memory

```
section .data
length: .word 1
width: .word 2
perim: .word 0

section .text
.global main
main:
adr  x0, length
ldr  w1, [x0]
adr  x0, width
ldr  w2, [x0]
add  w1, w1, w2
lsl  w1, w1, 1
adr  x0, perim
str  w1, [x0]
mov  w0, 0
ret
```

.word: 32-bit int and initial value

See appendix for variables in other sections, with other types.
main()

```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
```

Global visibility

**.global**: Declare “main” to be a globally-visible label

```
.section .data
length: .word 1
width: .word 2
perim: .word 0

.section .text
.global main
main:
    adr  x0, length
    ldr  w1, [x0]
    adr  x0, width
    ldr  w2, [x0]
    add  w1, w1, w2
    lsl  w1, w1, 1
    adr  x0, perim
    str  w1, [x0]
    mov  w0, 0
    ret
```
Make a “pointer”

```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim = (length + width) * 2;
    return 0;
}
```

Generating addresses
adr: put address of a label in a register

```asm
.section .data
length:.word 1
width:.word 2
perim:.word 0

.section .text
.global main
main:
    adr x0, length
    ldr w1, [x0]
    adr x0, width
    ldr w2, [x0]
    add w1, w1, w2
    lsl w1, w1, 1
    adr x0, perim
    str w1, [x0]
    mov w0, 0
    ret
```
Loads and Stores

```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
    (length + width) * 2;
    return 0;
}
```

Load and store

Use x0 as a “pointer” to load from and store to memory
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
     (length + width) * 2;
    return 0;
}

Return a value

ret: return to the caller, with
register 0* holding the return value

* w0 for int, x0 for long
```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim = 
        (length + width) * 2;
    return 0;
}
```

```assembly
section .data
length: .word 1
width: .word 2
perim: .word 0
section .text
global main
main:
    adr  x0, length
    ldr  w1, [x0]
    adr  x0, width
    ldr  w2, [x0]
    add  w1, w1, w2
    lsl  w1, w1, 1
    adr  x0, perim
    str  w1, [x0]
    mov  w0, 0
    ret
```
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}

.section .data
length: .word 1
width: .word 2
perim: .word 0

.section .text
.global main
main:
  adr  x0, length
  ldr  w1, [x0]
  adr  x0, width
  ldr  w2, [x0]
  add  w1, w1, w2
  lsl  w1, w1, 1
  adr  x0, perim
  str  w1, [x0]
  mov  w0, 0
  ret

Registers

<table>
<thead>
<tr>
<th>x0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>1</td>
</tr>
<tr>
<td>w2</td>
<td>2</td>
</tr>
</tbody>
</table>

Memory

| length | 1 |
| width | 2 |
| perim | 0 |
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
```c
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
        (length + width) * 2;
    return 0;
}
```

```asm
.section .data
length: .word 1
width: .word 2
perim: .word 0
    .section .text
    .global main
main:
    adr x0, length
    ldr w1, [x0]
    adr x0, width
    ldr w2, [x0]
    add w1, w1, w2
    lsl w1, w1, 1
    adr x0, perim
    str w1, [x0]
    mov w0, 0
    ret
```
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
    (length + width) * 2;
    return 0;
}
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
    (length + width) * 2;
    return 0;
}
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
  perim =
    (length + width) * 2;
  return 0;
}

Return value
Passed back in register w0
static int length = 1;
static int width = 2;
static int perim = 0;

int main()
{
    perim =
    (length + width) * 2;
    return 0;
}
Summary

Language levels

The basics of computer architecture
  • Enough to understand AARCH64 assembly language

The basics of AARCH64 assembly language
  • Instructions to perform arithmetic
  • Instructions to define global data and perform data transfer

To learn more
  • Study more assembly language examples
    • Chapters 2-5 of Pyeatt and Ughetta book
    • Study compiler-generated assembly language code (though it will be challenging!)
      • gcc217 –S somefile.c
Appendix 1

DEFINING DATA:
OTHER SECTIONS AND SIZES
### Defining Data: DATA Section 1

<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>char</td>
<td>'a'</td>
</tr>
<tr>
<td>short</td>
<td>12</td>
</tr>
<tr>
<td>int</td>
<td>345</td>
</tr>
<tr>
<td>long</td>
<td>6789</td>
</tr>
</tbody>
</table>

#### Notes:
- **.section** directive (to announce DATA section)
- **label definition** (marks a spot in RAM)
- **.byte** directive (1 byte)
- **.short** directive (2 bytes)
- **.word** directive (4 bytes)
- **.quad** directive (8 bytes)
Defining Data: DATA Section 2

char c = 'a';
short s = 12;
int i = 345;
long l = 6789;

Notes:
  Can place label on same line as next instruction
  .global directive can also apply to variables, not just functions
Defining Data: BSS Section

```
static char c;
static short s;
static int i;
static long l;
```

Notes:
- `.section` directive (to announce BSS section)
- `.skip` directive (to specify number of bytes)

```
.section ".bss"
   c:
      .skip 1
   s:
      .skip 2
   i:
      .skip 4
   l:
      .skip 8
```
Defining Data: RODATA Section

```
..."hello\n"...;
...
```

Notes:
- `.section "._rodata"` directive (to announce RODATA section)
- `.string "hello\n"` directive
Appendix 2

BYTE ORDER:
BIG-ENDIAN VS LITTLE-ENDIAN
Byte Order

AARCH64 is a **little endian** architecture
- **Least** significant byte of multi-byte entity is stored at lowest memory address
- “Little end goes first”

```
00000101
00000000
00000000
00000000
```

The int 5 at address 1000:
```
1000  00000101
1001  00000000
1002  00000000
1003  00000000
```

Some other systems use **big endian**
- **Most** significant byte of multi-byte entity is stored at lowest memory address
- “Big end goes first”

```
1000  00000000
1001  00000000
1002  00000000
1003  00000101
```

The int 5 at address 1000:
Byte Order Example 1

```c
#include <stdio.h>
int main(void)
{
    unsigned int i = 0x003377ff;
    unsigned char *p;
    int j;
    p = (unsigned char *)&i;
    for (j = 0; j < 4; j++)
        printf("Byte %d: %2x\n", j, p[j]);
}
```

Output on a little-endian machine
- Byte 0: ff
- Byte 1: 77
- Byte 2: 33
- Byte 3: 00

Output on a big-endian machine
- Byte 0: 00
- Byte 1: 33
- Byte 2: 77
- Byte 3: ff
Byte Order Example 2

Note:
Flawed code; uses “b” instructions to load from a four-byte memory area

AARCH64 is little endian, so what will be the value returned from w0?

What would be the value returned from w0 if AARCH64 were big endian?

```
.section "" .data
foo: .word 7

.section "" .text
.global "main"
main:
    adr   x0, foo
    ldrb  w0, [x0]
    ret
```

Note:
Flawed code; uses “b” instructions to load from a four-byte memory area

What would be the value returned from w0 if AARCH64 were big endian?