

COMP110: Principles of Computing

12: Machine Code



# Learning outcomes

- Explain the difference between interpretation, just-in-time compilation and ahead-of-time compilation
- Describe how common high-level code structures translate to machine code
- Write and understand simple assembly language programs





► CPUs execute machine code

- ► CPUs execute machine code
- Programs must be translated into machine code for execution

- ► CPUs execute machine code
- Programs must be translated into machine code for execution
- ► There are three main ways of doing this:

- CPUs execute machine code
- Programs must be translated into machine code for execution
- There are three main ways of doing this:
  - An interpreter is an application which reads the program source code and executes it directly

- CPUs execute machine code
- Programs must be translated into machine code for execution
- ► There are three main ways of doing this:
  - An interpreter is an application which reads the program source code and executes it directly
  - An ahead-of-time (AOT) compiler, often just called a compiler, is an application which converts the program source code into executable machine code

- ► CPUs execute machine code
- Programs must be translated into machine code for execution
- ► There are three main ways of doing this:
  - An interpreter is an application which reads the program source code and executes it directly
  - An ahead-of-time (AOT) compiler, often just called a compiler, is an application which converts the program source code into executable machine code
  - A just-in-time (JIT) compiler is halfway between the two — it compiles the program on-the-fly at runtime

#### Interpreted:

- ▶ Python
- ▶ Lua
- JavaScript (in old web browsers)
- Bespoke scripting languages

#### Interpreted:

- ► Python
- ▶ Lua
- JavaScript (in old web browsers)
- Bespoke scripting languages

#### Compiled:

- ▶ C
- ▶ C++
- ▶ Swift
- ► Rust

#### Interpreted:

- ► Python
- ▶ Lua
- JavaScript (in old web browsers)
- Bespoke scripting languages

#### Compiled:

- ▶ C
- ▶ C++
- Swift
- ► Rust

#### JIT compiled:

- ▶ Java
- ► C#
- JavaScript (in modern web browsers)
- ► Jython

#### Interpreted:

- ► Python
- ▶ Lua
- JavaScript (in old web browsers)
- Bespoke scripting languages

#### Compiled:

- ▶ C
- ▶ C++
- Swift
- ► Rust

#### JIT compiled:

- ▶ Java
- ► C#
- JavaScript

   (in modern web
   browsers)
- Jython

NB: technically any language could appear in any column here, but this is where they typically are

► Run-time efficiency: compiler > interpreter

- ► Run-time efficiency: compiler > interpreter
  - The compiler translates the program in advance, on the developer's machine

- Run-time efficiency: compiler > interpreter
  - ► The compiler translates the program **in advance**, on the developer's machine
  - The interpreter translates the program at runtime, on the user's machine — this takes extra time

► Portability: compiler < interpreter

- ▶ Portability: compiler < interpreter</p>
  - A compiled program can only run on the operating system and CPU architecture it was compiled for

- Portability: compiler < interpreter</li>
  - A compiled program can only run on the operating system and CPU architecture it was compiled for
  - An interpreted program can run on any machine, as long as a suitable interpreter is available

► Ease of development: compiler < interpreter

- ▶ Ease of development: compiler < interpreter</p>
  - Writing an AOT or JIT compiler (especially a good one) is hard, and required in-depth knowledge of the target machine

- ▶ Ease of development: compiler < interpreter</p>
  - Writing an AOT or JIT compiler (especially a good one) is hard, and required in-depth knowledge of the target machine
  - Writing an interpreter is easy in comparison



► Dynamic language features: compiler < interpreter

- ▶ Dynamic language features: compiler < interpreter
  - ► The interpreter is already on the end user's machine, so programs can use it e.g. to dynamically generate and execute new code

- ▶ Dynamic language features: compiler < interpreter</p>
  - ► The interpreter is already on the end user's machine, so programs can use it e.g. to dynamically generate and execute new code
  - The AOT compiler is not generally on the end user's machine, so this is more difficult



► JIT compilers have similar pros/cons to interpreters

- ▶ JIT compilers have similar pros/cons to interpreters
  - Runtime efficiency: JIT > interpreter (e.g. code inside a loop only needs to be translated once, then can be executed many times)

- JIT compilers have similar pros/cons to interpreters
  - Runtime efficiency: JIT > interpreter (e.g. code inside a loop only needs to be translated once, then can be executed many times)
  - ► Ease of development: JIT < interpreter

 Many modern interpreters and JIT compilers translate programs into bytecode

- Many modern interpreters and JIT compilers translate programs into bytecode
- Bytecode is essentially machine code for a virtual machine (VM)

- Many modern interpreters and JIT compilers translate programs into bytecode
- Bytecode is essentially machine code for a virtual machine (VM)
- Translation from source code to bytecode can be done ahead of time

- Many modern interpreters and JIT compilers translate programs into bytecode
- Bytecode is essentially machine code for a virtual machine (VM)
- Translation from source code to bytecode can be done ahead of time
- At runtime, translate the bytecode (by interpretation or JIT compilation) into machine code for the physical machine

- Many modern interpreters and JIT compilers translate programs into bytecode
- Bytecode is essentially machine code for a virtual machine (VM)
- Translation from source code to bytecode can be done ahead of time
- At runtime, translate the bytecode (by interpretation or JIT compilation) into machine code for the physical machine
- E.g. a Java JAR file, a .NET executable, a Python .pyc or .pyo file all contain bytecode for their respective VMs



#### **Assemblers**

#### **Assemblers**

 Assembly language is designed to translate directly into machine code

#### **Assemblers**

- Assembly language is designed to translate directly into machine code
- An ahead-of-time compile for assembly language is called an assembler

#### **Assemblers**

- Assembly language is designed to translate directly into machine code
- An ahead-of-time compile for assembly language is called an assembler
- Generally much simpler than an AOT compiler for a higher-level language





The MIPS architecture

 An example of a Reduced Instruction Set Computer (RISC) architecture

- An example of a Reduced Instruction Set Computer (RISC) architecture
  - Small number of simple instructions computational power comes from executing many instructions per second

- An example of a Reduced Instruction Set Computer (RISC) architecture
  - Small number of simple instructions computational power comes from executing many instructions per second
  - Compare with Complex Instruction Set Computer (CISC) architecture (e.g. Intel x86) — large number of complex instructions — fewer instructions per second, but shorter programs

- An example of a Reduced Instruction Set Computer (RISC) architecture
  - Small number of simple instructions computational power comes from executing many instructions per second
  - Compare with Complex Instruction Set Computer (CISC) architecture (e.g. Intel x86) — large number of complex instructions — fewer instructions per second, but shorter programs
- ► MIPS was popular in 1980s 2000s

- An example of a Reduced Instruction Set Computer (RISC) architecture
  - Small number of simple instructions computational power comes from executing many instructions per second
  - Compare with Complex Instruction Set Computer (CISC) architecture (e.g. Intel x86) — large number of complex instructions — fewer instructions per second, but shorter programs
- ► MIPS was popular in 1980s 2000s
  - ► Embedded systems

- An example of a Reduced Instruction Set Computer (RISC) architecture
  - Small number of simple instructions computational power comes from executing many instructions per second
  - Compare with Complex Instruction Set Computer (CISC) architecture (e.g. Intel x86) — large number of complex instructions — fewer instructions per second, but shorter programs
- ► MIPS was popular in 1980s 2000s
  - ► Embedded systems
  - Consoles (Nintendo 64, PlayStation 1 and 2)

- An example of a Reduced Instruction Set Computer (RISC) architecture
  - Small number of simple instructions computational power comes from executing many instructions per second
  - Compare with Complex Instruction Set Computer (CISC) architecture (e.g. Intel x86) — large number of complex instructions — fewer instructions per second, but shorter programs
- ► MIPS was popular in 1980s 2000s
  - Embedded systems
  - Consoles (Nintendo 64, PlayStation 1 and 2)
- Easier to understand than most CPU instruction sets in common use today

#### Online MIPS simulator

http://rivoire.cs.sonoma.edu/cs351/wemips/

Memory locations inside the CPU

- ► Memory locations inside the CPU
- Faster to access than main memory

- Memory locations inside the CPU
- Faster to access than main memory
- Registers in MIPS architecture include:

- Memory locations inside the CPU
- Faster to access than main memory
- Registers in MIPS architecture include:
  - ▶ \$zero: constant 0

- Memory locations inside the CPU
- Faster to access than main memory
- Registers in MIPS architecture include:
  - \$zero: constant 0
  - ▶ \$t0-\$t9: temporary storage

- Memory locations inside the CPU
- Faster to access than main memory
- Registers in MIPS architecture include:
  - \$zero: constant 0
  - ▶ \$t0-\$t9: temporary storage
  - ▶ \$s0-\$s7: saved temporary storage

- Memory locations inside the CPU
- Faster to access than main memory
- Registers in MIPS architecture include:
  - \$zero: constant 0
  - ▶ \$t0-\$t9: temporary storage
  - \$\$0-\$\$7: saved temporary storage
- Each register holds a single 32-bit value

```
add $d, $s, $t
```

```
add $d, $s, $t
```

▶ \$d, \$s and \$t are register names

```
add $d, $s, $t
```

- \$a, \$s and \$t are register names
- This adds the value of \$s to the value of \$t, and stores the result in \$d

```
add $d, $s, $t
```

- \$a, \$s and \$t are register names
- This adds the value of \$s to the value of \$t, and stores the result in \$d

```
sub $d, $s, $t
```

```
add $d, $s, $t
```

- \$a, \$s and \$t are register names
- This adds the value of \$s to the value of \$t, and stores the result in \$d

```
sub $d, $s, $t
```

 Subtracts the value of \$t from the value of \$s, and stores the result in \$d

```
addi $d, $s, C
```

```
addi $d, $s, C
```

ightharpoonup \$\ \square and \$\square are register names, c is an integer constant

```
addi $d, $s, C
```

- ▶ \$d and \$s are register names, c is an integer constant
- ► This adds the value of \$s to c, and stores the result in \$d

```
addi $d, $s, C
```

- ▶ \$d and \$s are register names, c is an integer constant
- ► This adds the value of \$s to c, and stores the result in \$d
- addi = "add immediate" as in c is specified immediately in the code, not looked up from a register

```
addi $d, $s, C
```

- \$d and \$s are register names, c is an integer constant
- ► This adds the value of \$s to c, and stores the result in \$d
- addi = "add immediate" as in c is specified immediately in the code, not looked up from a register
- ► There is no subi instruction to subtract c, add -c

### More fun with addi

#### More fun with addi

► Socrative FALCOMPED

#### More fun with addi

- ► Socrative FALCOMPED
- ► What does this code do?

```
addi $s0, $s1, 0
```

#### More fun with addi

- ► Socrative FALCOMPED
- What does this code do?

```
addi $s0, $s1, 0
```

▶ What does this code do?

```
addi $s0, $zero, 12
```

#### More fun with addi

- ► Socrative FALCOMPED
- ▶ What does this code do?

```
addi $s0, $s1, 0
```

▶ What does this code do?

```
addi $s0, $zero, 12
```

 MIPS does not have dedicated instructions for setting a register value to a constant or to the value of another register — it has to be done with addi





# Control flow in MIPS

▶ In assembly code, can set a **label** on any line:

```
MyLabel: add $s0, $s1, 1
```

▶ In assembly code, can set a label on any line:

```
MyLabel: add $s0, $s1, 1
```

 Some instructions use labels to refer to a location in the code

▶ In assembly code, can set a label on any line:

```
MyLabel: add $s0, $s1, 1
```

- Some instructions use labels to refer to a location in the code
- E.g. the j instruction simply jumps (backwards or forwards) to the specified line:

```
j MyLabel
```

► Branching is conditional jumping

► Branching is conditional jumping

```
beq $s, $t, Label
```

Branching is conditional jumping

```
beq $s, $t, Label
```

► This jumps to Label if and only if the value of \$s equals the value of \$t

Branching is conditional jumping

```
beq $s, $t, Label
```

► This jumps to Label if and only if the value of \$s equals the value of \$t

```
bne $s, $t, Label
```

Branching is conditional jumping

```
beq $s, $t, Label
```

► This jumps to Label if and only if the value of \$s equals the value of \$t

```
bne $s, $t, Label
```

► This jumps to Label if and only if the value of \$s does not equal the value of \$t

► Branching allows us to implement if statements

► Branching allows us to implement if statements

```
if s0 != 0:
    s1 += 1
else:
    s2 += 1
```

Branching allows us to implement if statements

```
if s0 != 0:
    s1 += 1
else:
    s2 += 1
```

```
beq $s0, $zero, Else
addi $s1, $s1, 1
j End
Else: addi $s2, $s2, 1
End:
```

► Branching allows us to implement while loops

► Branching allows us to implement while loops

```
i = 0
total = 0
limit = 10

while i != limit:
    total += i
    i += 1
# end while
```

► Branching allows us to implement while loops

```
i = 0
total = 0
limit = 10

while i != limit:
    total += i
    i += 1
# end while
```

```
addi $s0, $zero, 0
addi $s1, $zero, 0
addi $s2, $zero, 10

Loop: beq $s0, $s2, LoopEnd
add $s1, $s1, $s0
addi $s0, $s0, 1
j Loop
LoopEnd:
```

Function calls can be implemented using the jump instruction:

- Function calls can be implemented using the jump instruction:
  - ► To call the function: save the address of the instruction after the current one, then jump to the function

- Function calls can be implemented using the jump instruction:
  - To call the function: save the address of the instruction after the current one, then jump to the function
  - To return from the function: jump to the previously saved address

- Function calls can be implemented using the jump instruction:
  - To call the function: save the address of the instruction after the current one, then jump to the function
  - To return from the function: jump to the previously saved address
- MIPS has jal and jr instructions and \$ra register for this purpose

- Function calls can be implemented using the jump instruction:
  - To call the function: save the address of the instruction after the current one, then jump to the function
  - To return from the function: jump to the previously saved address
- MIPS has jal and jr instructions and \$ra register for this purpose
- Socrative FALCOMPED: why save the return address? Why not just hard-code it into the program?

- Function calls can be implemented using the jump instruction:
  - To call the function: save the address of the instruction after the current one, then jump to the function
  - To return from the function: jump to the previously saved address
- MIPS has jai and jr instructions and \$ra register for this purpose
- Socrative FALCOMPED: why save the return address? Why not just hard-code it into the program?
- Nested function calls require a stack of return addresses



# MIPS machine code



 Each line of MIPS assembly code can be translated into a machine code instruction

- ► Each line of MIPS assembly code can be translated into a **machine code instruction**
- ▶ 1 line of assembly = 1 instruction

- ► Each line of MIPS assembly code can be translated into a **machine code instruction**
- ▶ 1 line of assembly = 1 instruction
- ► Each instruction is a 32 bit value

- Each line of MIPS assembly code can be translated into a machine code instruction
- ▶ 1 line of assembly = 1 instruction
- ► Each instruction is a 32 bit value
- First 6 bits specify the opcode; how the remaining 26 bits are interpreted depends on which opcode it is

# Anatomy of an instruction

# Anatomy of an instruction

**R-type** instruction:

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

**R-type** instruction:

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

 opcode and function together specify the operation to execute

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000
  - ► E.g. sub has opcode 000000 and function 100010

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000
  - ► E.g. sub has opcode 000000 and function 100010
- Some instructions specify a shift amount

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000
  - ► E.g. sub has opcode 000000 and function 100010
- Some instructions specify a shift amount
  - ► For add sub etc these 5 bits are ignored

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000
  - ► E.g. sub has opcode 000000 and function 100010
- Some instructions specify a shift amount
  - ► For add sub etc these 5 bits are ignored
- Registers are identified by a 5-bit number

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000
  - ► E.g. sub has opcode 000000 and function 100010
- Some instructions specify a shift amount
  - ► For add sub etc these 5 bits are ignored
- Registers are identified by a 5-bit number
  - ► E.g. \$zero  $\rightarrow$  00000, \$s0  $\rightarrow$  01000, \$s1  $\rightarrow$  01001

opcode	\$s	\$t	\$d	shift	function
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- opcode and function together specify the operation to execute
  - ► E.g. add has opcode 000000 and function 100000
  - ► E.g. sub has opcode 000000 and function 100010
- Some instructions specify a shift amount
  - ► For add sub etc these 5 bits are ignored
- Registers are identified by a 5-bit number
  - ightharpoonup E.g. \$zero ightarrow 00000, \$s0 ightarrow 01000, \$s1 ightarrow 01001
  - ► There are 32 registers

# Example

```
add $s0, $s0, $s1
```



opcode s t d shift function 000000 01000 01001 01001 00000 100000

opcode	\$s	\$t	С
6 bits	5 bits	5 bits	16 bits

**I-type** instruction:

opcode	\$s	\$t	С
6 bits	5 bits	5 bits	16 bits

opcode specifies the operation to execute

opcode	\$s	\$t	С
6 bits	5 bits	5 bits	16 bits

- opcode specifies the operation to execute
  - ► E.g. addi has opcode 001000

opcode	\$s	\$t	С
6 bits	5 bits	5 bits	16 bits

- opcode specifies the operation to execute
  - ► E.g. addi has opcode 001000
- ▶ c is specified as a 16-bit number

# Example

```
addi $s0, $s1, 123
```

 $\downarrow$ 

```
opcode s t C 001000 01001 01000 000000001111011
```

opcode	address
6 bits	26 bits

**J-type** instruction:

opcode	address
6 bits	26 bits

▶ opcode specifies the operation to execute

opcode	address
6 bits	26 bits

- opcode specifies the operation to execute
  - ► E.g. J has opcode 000010

opcode	address
6 bits	26 bits

- opcode specifies the operation to execute
  - ► E.g. J has opcode 000010
- address is specified as a 26-bit number