

### Recap

- The lea Instruction
- Logical and Arithmetic Operations

### Recap: lea

The **lea** instruction <u>copies</u> an "effective address" from one place to another.

lea src,dst

Unlike **mov**, which copies data <u>at</u> the address src to the destination, **lea** copies the value of src *itself* to the destination.

The syntax for the destinations is the same as **mov**. The difference is how it handles the **src**.

### Recap: Unary Instructions

The following instructions operate on a single operand (<u>register</u> or <u>memory</u>):

Instruction	Effect	Description
inc D	D ← D + 1	Increment
dec D	D ← D - 1	Decrement
neg D	D ← -D	Negate
not D	D ← ~D	Complement

Examples: incq 16(%rax)

dec %rdx

not %rcx

### Recap: Binary Instructions

The following instructions operate on two operands (<u>both can be register or memory</u>, <u>source can also be immediate</u>). <u>Both cannot be memory locations!</u> Read it as, e.g., "Subtract S from D":

Instruction	Effect	Description
add S, D	$D \leftarrow D + S$	Add
sub S, D	D ← D - S	Subtract
imul S, D	D ← D * S	Multiply
xor S, D	D ← D ^ S	Exclusive-or
or S, D	D ← D   S	Or
and S, D	D ← D & S	And

```
Examples: addq %rcx,(%rax)
xorq $16,(%rax, %rdx, 8)
subq %rdx,8(%rax)
```

### Recap: Large Multiplication

- Multiplying 64-bit numbers can produce a 128-bit result. How does x86-64 support this with only 64-bit registers?
- If you specify two operands to **imul**, it multiplies them together and truncates until it fits in a 64-bit register.

imul S, D D 
$$\leftarrow$$
 D \* S

• If you specify one operand, it multiplies that by **%rax**, and splits the product across **2** registers. It puts the high-order 64 bits in **%rdx** and the low-order 64 bits in **%rax**.

Instruction	Effect	Description
imulq S	$R[%rdx]:R[%rax] \leftarrow S \times R[%rax]$	Signed full multiply
mulq S	$R[%rdx]:R[%rax] \leftarrow S \times R[%rax]$	Unsigned full multiply

### Recap: Division and Remainder

Instruction	Effect	Description
idivq S	R[%rdx] ← R[%rdx]:R[%rax] mod S; R[%rax] ← R[%rdx]:R[%rax] <b>÷</b> S	Signed divide
divq S	R[%rdx] ← R[%rdx]:R[%rax] mod S; R[%rax] ← R[%rdx]:R[%rax] ÷ S	Unsigned divide

- Terminology: dividend / divisor = quotient + remainder
- x86-64 supports dividing up to a 128-bit value by a 64-bit value.
- The high-order 64 bits of the dividend are in **%rdx**, and the low-order 64 bits are in **%rax**. The divisor is the operand to the instruction.
- The quotient is stored in %rax, and the remainder in %rdx.

### Recap: Division and Remainder

Instruction	Effect	Description
idivq S	R[%rdx] ← R[%rdx]:R[%rax] mod S; R[%rax] ← R[%rdx]:R[%rax] <b>÷</b> S	Signed divide
divq S	R[%rdx] ← R[%rdx]:R[%rax] mod S; R[%rax] ← R[%rdx]:R[%rax] <b>÷</b> S	Unsigned divide
cqto	R[%rdx]:R[%rax] ← SignExtend(R[%rax])	Convert to oct word

- Terminology: dividend / divisor = quotient + remainder
- The high-order 64 bits of the dividend are in **%rdx**, and the low-order 64 bits are in **%rax**. The divisor is the operand to the instruction.
- Most division uses only 64-bit dividends. The **cqto** instruction sign-extends the 64-bit value in **%rax** into **%rdx** to fill both registers with the dividend, as the division instruction expects.

### Recap: Shift Instructions

The following instructions have two operands: the shift amount **k** and the destination to shift, **D**. **k** can be either an immediate value, or the byte register **%c1** (and only that register!)

Instruction	Effect	Description
sal k, D	D ← D << k	Left shift
shl k, D	D ← D << k	Left shift (same as sal)
sar k, D	$D \leftarrow D >>_A k$	Arithmetic right shift
shr k, D	D ← D >> <sub>L</sub> k	Logical right shift

### Recap: Shift Amount

Instruction	Effect	Description
sal k, D	D ← D << k	Left shift
shl k, D	D ← D << k	Left shift (same as sal)
sar k, D	$D \leftarrow D >>_A k$	Arithmetic right shift
shr k, D	D ← D >> <sub>L</sub> k	Logical right shift

- When using **%c1**, the width of what you are shifting determines what portion of **%c1** is used.
- For w bits of data, it looks at the low-order log2(w) bits of %cl to know how much to shift.
  - If %c1 = 0xff (0b11111111), then: **shlb** shifts by 7 because it considers only the low-order log2(8) = 3 bits, which represent 7. **shlw** shifts by 15 because it considers only the low-order log2(16) = 4 bits, which represent 15.

### Recap: A Note About Operand Forms

- Many instructions share the same address operand forms that mov uses.
  - Eg. 7(%rax, %rcx, 2).
- These forms work the same way for other instructions, e.g. sub:
  - sub 8(%rax,%rdx),%rcx → Go to 8 + %rax + %rdx, subtract what's there from %rcx

- The exception is **lea**:
  - It interprets this form as just the calculation, not the dereferencing
  - lea 8(%rax,%rdx),%rcx → Calculate 8 + %rax + %rdx, put it in %rcx

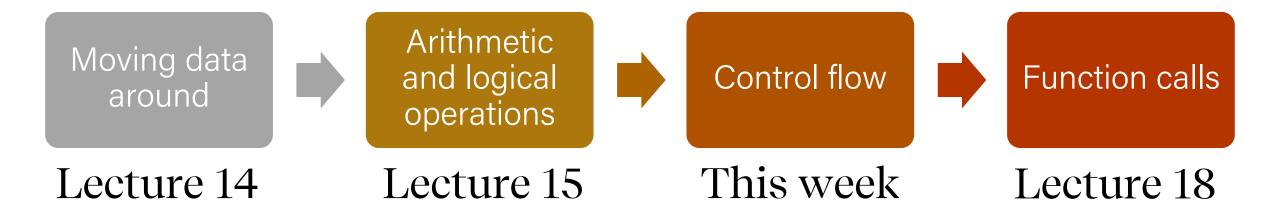
### Plan for Today

- Assembly Execution and %rip
- Control Flow Mechanics

**Disclaimer:** Slides for this lecture were borrowed from

—Nick Troccoli's Stanford CS107 class

### Learning Assembly



### Lecture Plan

- Assembly Execution and %rip
- Control Flow Mechanics

### Learning Goals

- Learn about how assembly stores comparison and operation results in condition codes
- Understand how assembly implements loops and control flow

### **Executing Instructions**

# What does it mean for a program to execute?

### **Executing Instructions**

#### So far:

- Program values can be stored in memory or registers.
- Assembly instructions read/write values back and forth between registers (on the CPU) and memory.
- Assembly instructions are also stored in memory.

### Today:

Who controls the instructions?
 How do we know what to do now or next?

#### Answer:

• The program counter (PC), %rip.

4004fd	fa
4004fc	eb
4004fb	01
4004fa	fc
4004f9	45
4004f8	83
4004f7	00
4004f6	00
4004f5	00
4004f4	00
4004f3	fc
4004f2	45
4004f1	c7
4004f0	e5
4004ef	89
4004ee	48
4004ed	55

### Register Responsibilities

Some registers take on special responsibilities during program execution.

- **%rax** stores the return value
- %rdi stores the first parameter to a function
- **%rsi** stores the second parameter to a function
- %rdx stores the third parameter to a function
- **%rip** stores the address of the next instruction to execute
- **%rsp** stores the address of the current top of the stack

See the x86-64 Guide and Reference Sheet on the Resources webpage for more!

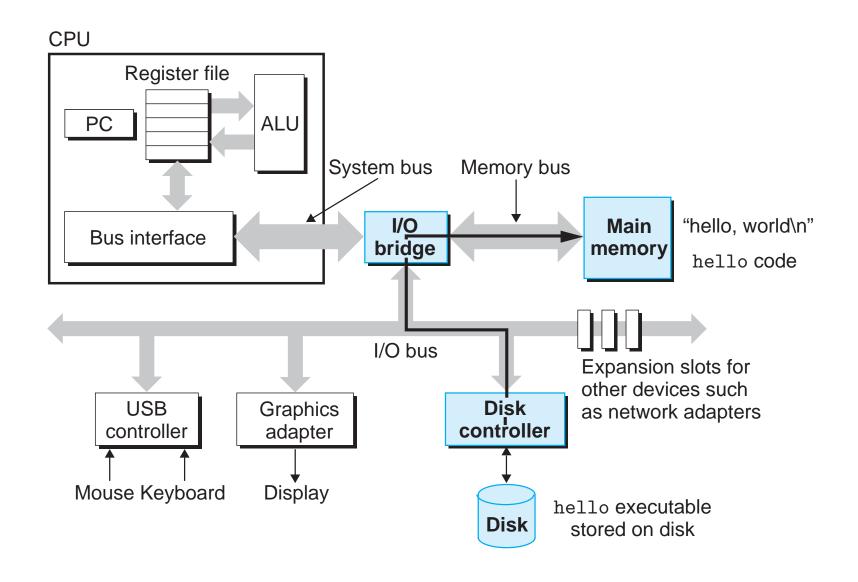
### Register Responsibilities

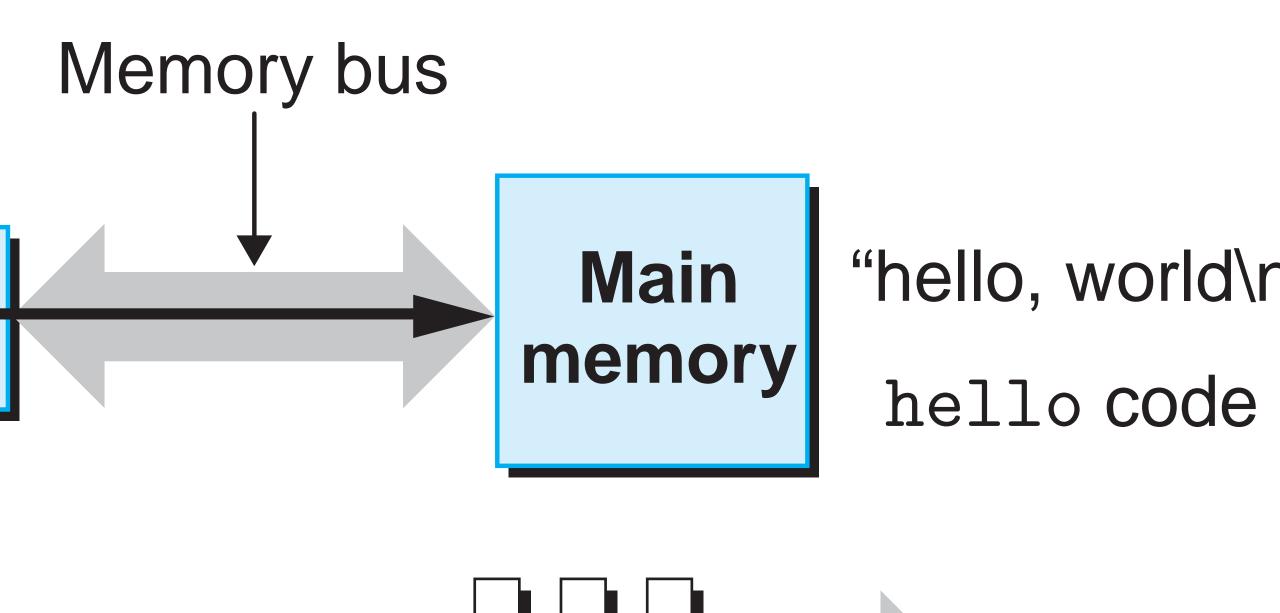
Some registers take on special responsibilities during program execution.

- %rax stores the return value
- %rdi stores the first parameter to a function
- **%rsi** stores the second parameter to a function
- %rdx stores the third parameter to a function
- **%rip** stores the address of the next instruction to execute
- **%rsp** stores the address of the current top of the stack

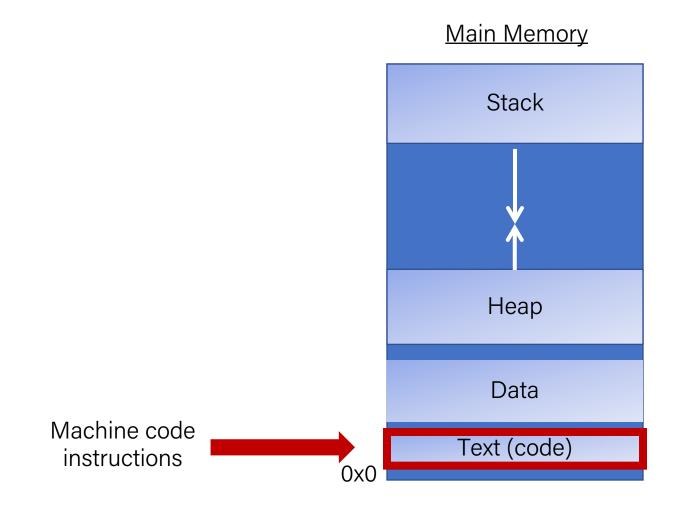
See the x86-64 Guide and Reference Sheet on the Resources webpage for more!

### Instructions Are Just Bytes!





### Instructions Are Just Bytes!



4004ed: 55

#### 00000000004004ed <loop>:

 4004ee:
 48
 89
 e5
 mov
 %rsp,%rbp

 4004f1:
 c7
 45
 fc
 00
 00
 00
 movl
 \$0x0,-0x4(%rbp)

 4004f8:
 83
 45
 fc
 01
 addl
 \$0x1,-0x4(%rbp)

 4004fc:
 eb
 fa
 jmp
 4004f8 <loop+0xb>

push

%rbp

4004fd	fa
4004fc	eb
4004fb	01
4004fa	fc
4004f9	45
4004f8	83
4004f7	00
4004f6	00
4004f5	00
4004f4	00
4004f3	fc
4004f2	45
4004f1	<b>c7</b>
4004f0	e5
4004ef	89
4004ee	48
4004ed	55

#### Main Memory

Stack Heap Data Text (code)

#### 00000000004004ed <loop>:

4004ed: 55

4004ee: 48 89 e5

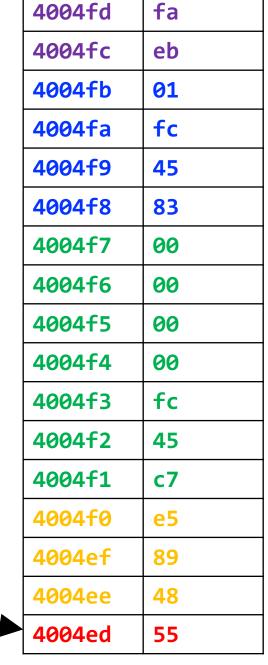
4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The **program counter** (PC), known as %rip in x86-64, stores the address in memory of the **next instruction** to be executed.

push	%rbp
mov	%rsp,%rbp
movl	\$0x0,-0x4(%rbp)
addl	\$0x1,-0x4(%rbp)
imp	4004f8 <loop+0xb></loop+0xb>



#### 00000000004004ed <loop>:

4004ee: 48 89 e5

4004ed: 55

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The **program counter** (PC), known as %rip in x86-64, stores the address in memory of the **next instruction** to be executed.

push	%rbp
mov	%rsp,%rbp
movl	\$0x0,-0x4(%rbp)
addl •	\$0x1,-0x4(%rbp)
jmp	4004f8 <loop+0xb></loop+0xb>



#### 00000000004004ed <loop>:

4004ed: 55

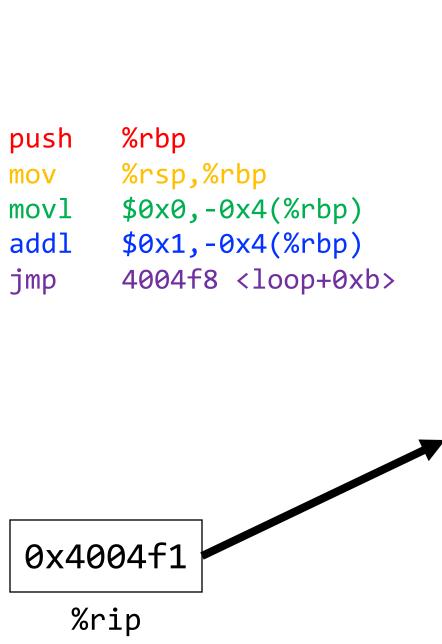
4004ee: 48 89 e5

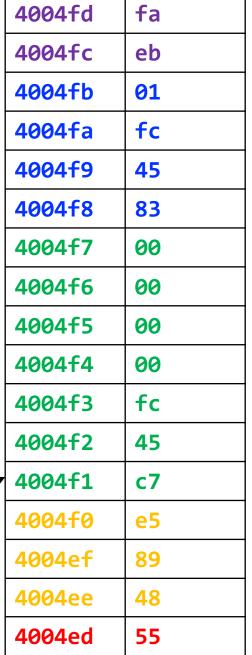
4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The **program counter** (PC), known as %rip in x86-64, stores the address in memory of the *next instruction* to be executed.





#### 00000000004004ed <loop>:

4004ed: 55

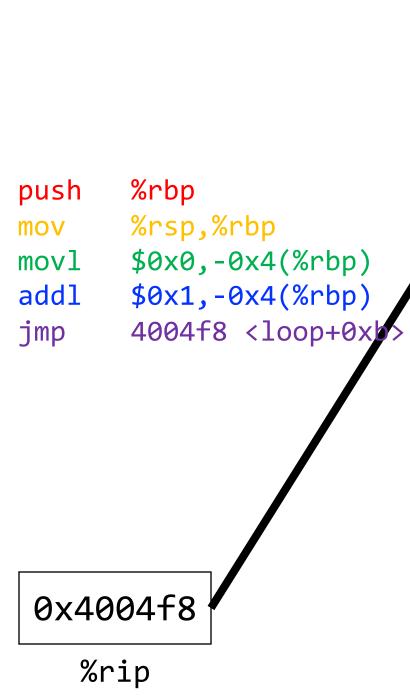
4004ee: 48 89 e5

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The **program counter** (PC), known as %rip in x86-64, stores the address in memory of the *next instruction* to be executed.



4004fd fa 4004fc eb 4004fb **01** fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 fc 4004f3 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

00000000004004ed <loop>:

4004ed: 55

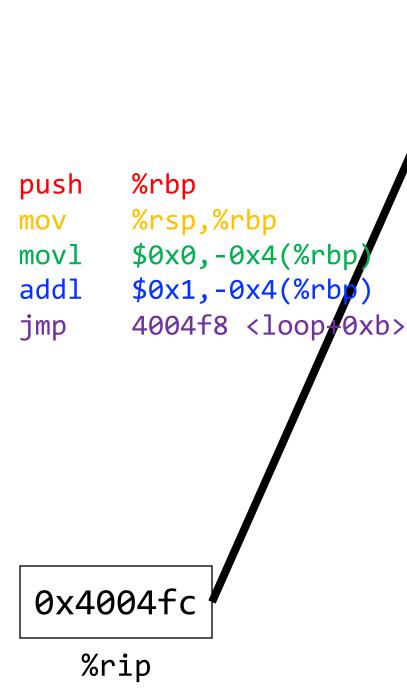
4004ee: 48 89 e5

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The **program counter** (PC), known as %rip in x86-64, stores the address in memory of the **next instruction** to be executed.



4004fd fa 4004fc eb 4004fb 01 fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 4004f3 fc 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

#### 00000000004004ed <loop>:

4004ed: 55 push 4004ee: 48 89 e5 mov

4004ee: 48 89 e5 mov %rsp,%rbp 4004f1: c7 45 fc 00 00 00 00 movl \$0x0,-0x4(%rbp)

4004f8: 83 45 fc 01

4004fc: eb fa

addl \$0x1,-0x4(%rbp) jmp 4004f8 <loop+0xb>

%rbp

Special hardware sets the program counter to the next instruction:

%rip += size of bytes of current
instruction

0x4004fc

%rip

4004fd	fa
4004fc	eb
4004fb	01
4004fa	fc
4004f9	45
4004f8	83
4004f7	00
4004f6	00
4004f5	00
4004f4	00
4004f3	fc
4004f2	45
4004f1	с7
4004f0	e5
4004ef	89
4004ee	48
4004ed	55

## Going In Circles

- How can we use this representation of execution to represent e.g. a loop?
- **Key Idea:** we can "interfere" with **%rip** and set it back to an earlier instruction!

4004ed: 55

#### 00000000004004ed <loop>:

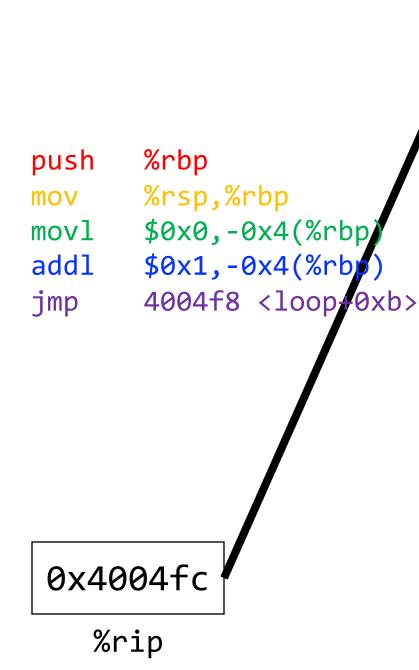
4004ee: 48 89 e5

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The jmp instruction is an unconditional jump that sets the program counter to the jump target (the operand).



4004fd fa 4004fc eb 4004fb 01 fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 4004f3 fc 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

4004ed: 55

#### 00000000004004ed <loop>:

100100: 10 00 05

4004ee: 48 89 e5

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The jmp instruction is an unconditional jump that sets the program counter to the jump target (the operand).



4004fd fa 4004fc eb 4004fb **01** fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 fc 4004f3 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

00000000004004ed <loop>:

4004ed: 55

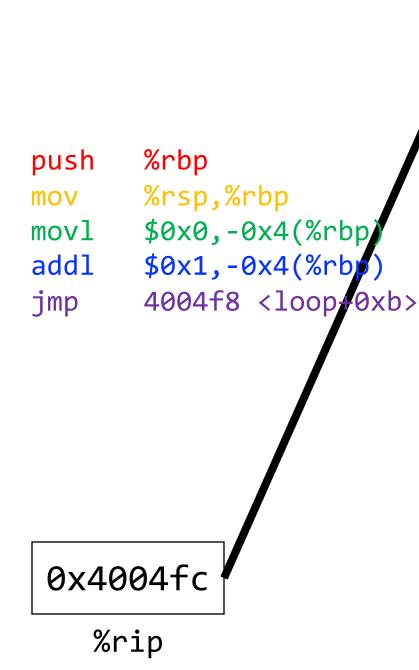
4004ee: 48 89 e5

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The jmp instruction is an unconditional jump that sets the program counter to the jump target (the operand).



4004fd fa 4004fc eb 4004fb 01 fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 4004f3 fc 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

#### 00000000004004ed <loop>:

4004ed: 55

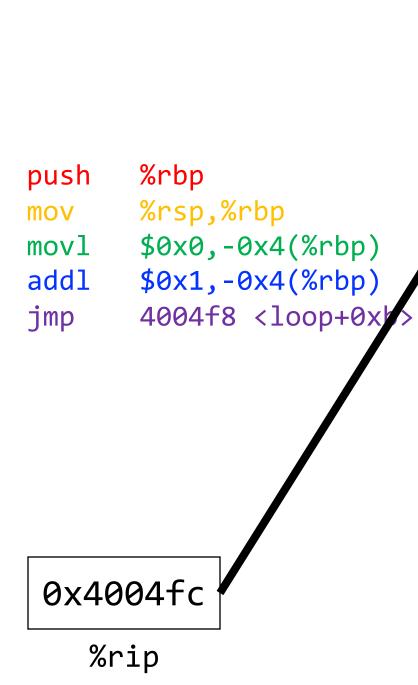
4004ee: 48 89 e5

4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

The jmp instruction is an unconditional jump that sets the program counter to the jump target (the operand).



4004fd fa 4004fc eb 4004fb **01** fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 fc 4004f3 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

4004ed: 55

```
00000000004004ed <loop>:
```

4004ee: 48 89 e5

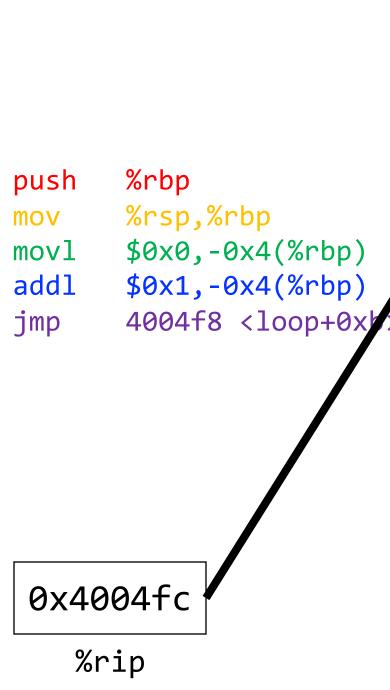
4004f1: c7 45 fc 00 00 00 00

4004f8: 83 45 fc 01

4004fc: eb fa

This assembly represents an infinite loop in C!

while (true) {...}



4004fd fa 4004fc eb 4004fb **01** fc 4004fa 4004f9 45 4004f8 83 4004f7 00 4004f6 00 4004f5 00 4004f4 00 fc 4004f3 45 4004f2 4004f1 **c7** 4004f0 **e5** 89 4004ef 4004ee 48 55 4004ed

### jmp

The **jmp** instruction jumps to another instruction in the assembly code ("Unconditional Jump").

```
jmp Label (Direct Jump)
jmp *Operand (Indirect Jump)
```

The destination can be hardcoded into the instruction (direct jump):

```
jmp 404f8 <loop+0xb> # jump to instruction at 0x404f8
```

The destination can also be one of the usual operand forms (indirect jump):

```
jmp *%rax  # jump to instruction at address in %rax
```

# "Interfering" with %rip

# 1. How do we repeat instructions in a loop?

jmp [target]

 A 1-step unconditional jump (always jump when we execute this instruction)

What if we want a conditional jump?

#### Lecture Plan

- Assembly Execution and %rip
- Control Flow Mechanics
  - Condition Codes
  - Assembly Instructions

- In C, we have control flow statements like **if**, **else**, **while**, **for**, etc. to write programs that are more expressive than just one instruction following another.
- This is *conditional execution of statements*: executing statements if one condition is true, executing other statements if one condition is false, etc.
- How is this represented in assembly?

```
if (x > y) { | In Assembly:
1. Calculate the condition result
2. Page described to the condition of the cond
                                                                                                                                                                                                                      // a
        } else {
```

- 2. Based on the result, go to a or b

- In assembly, it takes more than one instruction to do these two steps.
- Most often: 1 instruction to calculate the condition, 1 to conditionally jump

#### Common Pattern:

equal"

not equal"

less than"

# Conditional Jumps

There are also variants of **jmp** that jump only if certain conditions are true ("Conditional Jump"). The jump location for these must be hardcoded into the instruction.

Instruction	Synonym	Set Condition
je Label	jz	Equal / zero
jne <i>Label</i>	jnz	Not equal / not zero
js Label		Negative
jns <i>Label</i>		Nonnegative
jg Label	jnle	Greater (signed >)
jge <i>Label</i>	jnl	Greater or equal (signed >=)
jl Label	jnge	Less (signed <)
jle <i>Label</i>	jng	Less or equal (signed <=)
ja <i>Label</i>	jnbe	Above (unsigned >)
jae <i>Label</i>	jnb	Above or equal (unsigned >=)
jb Label	jnae	Below (unsigned <)
jbe <i>Label</i>	jna	Below or equal (unsigned <=)

Read cmp **S1,S2** as "compare S2 to S1":

```
// Jump if %edi > 2
                                 // Jump if %edi == 4
                                 cmp $4, %edi
cmp $2, %edi
                                 je [target]
jg [target]
                                 // Jump if %edi <= 1
// Jump if %edi != 3
                                 cmp $1, %edi
cmp $3, %edi
                                 jle [target]
jne [target]
```

Read cmp **S1,S2** as "compare S2 to S1":

```
// Jump if %edi > 2
                                    // Jump if %edi == 4
                                    cmp $4, %edi
cmp $2, %edi
                                    je [target]
jg [target]
                                             if %edi <= 1
// Jump if %ed Wait a minute - how does the
                                             %edi
                jump instruction know anything
cmp $3, %edi
                                             getl
                about the compared values in
jne [target]
                the earlier instruction?
```

- The CPU has special registers called *condition codes* that are like "global variables". They *automatically* keep track of information about the most recent arithmetic or logical operation.
  - **cmp** compares via calculation (subtraction) and info is stored in the condition codes
  - conditional jump instructions look at these condition codes to know whether to jump
- What exactly are the condition codes? How do they store this information?

## **Condition Codes**

Alongside normal registers, the CPU also has <u>single-bit</u> condition code registers. They store the results of the most recent arithmetic or logical operation.

#### Most common condition codes:

- **CF**: Carry flag. The most recent operation generated a carry out of the most significant bit. Used to detect overflow for unsigned operations.
- **ZF**: Zero flag. The most recent operation yielded zero.
- SF: Sign flag. The most recent operation yielded a negative value.
- **OF**: Overflow flag. The most recent operation caused a two's-complement overflow-either negative or positive.

## **Condition Codes**

Alongside normal registers, the CPU also has <u>single-bit</u> condition code registers. They store the results of the most recent arithmetic or logical operation.

Example: if we calculate t = a + b, condition codes are set according to:

- CF: Carry flag (Unsigned Overflow). (unsigned) t < (unsigned) a
- **ZF**: Zero flag (Zero). (t == 0)
- SF: Sign flag (Negative).  $(t < \theta)$
- OF: Overflow flag (Signed Overflow). (a<0 == b<0) && (t<0 != a<0)

# Setting Condition Codes

The **cmp** instruction is like the subtraction instruction, but it does not store the result anywhere. It just sets condition codes. (**Note** the operand order!)

CMP S1, S2

S2 - S1

Instruction	Description
cmpb	Compare byte
стрш	Compare word
cmpl	Compare double word
cmpq	Compare quad word

Read **cmp S1,S2** as "compare S2 to S1". It calculates S2 – S1 and updates the condition codes with the result.

```
// Jump if %edi > 2
                               // Jump if %edi == 4
// calculates %edi - 2
                               // calculates %edi - 4
cmp $2, %edi
                               cmp $4, %edi
jg [target]
                               je [target]
// Jump if %edi != 3
                               // Jump if %edi <= 1
                               // calculates %edi - 1
// calculates %edi - 3
                               cmp $1, %edi
cmp $3, %edi
jne [target]
                               jle [target]
```

# Conditional Jumps

Conditional jumps can look at subsets of the condition codes in order to check their condition of interest.

Instruction	Synonym	Set Condition
je <i>Label</i>	jz	Equal / zero (ZF = 1)
jne <i>Label</i>	jnz	Not equal / not zero (ZF = 0)
js Label		Negative (SF = 1)
jns <i>Label</i>		Nonnegative (SF = 0)
jg Label	jnle	Greater (signed >) (ZF = 0 and SF = OF)
jge <i>Label</i>	jnl	Greater or equal (signed >=) (SF = OF)
jl Label	jnge	Less (signed <) (SF != OF)
jle <i>Label</i>	jng	Less or equal (signed <=) (ZF = 1 or SF! = OF)
ja <i>Label</i>	jnbe	Above (unsigned $>$ ) (CF = 0 and ZF = 0)
jae <i>Label</i>	jnb	Above or equal (unsigned $\geq$ =) (CF = 0)
jb Label	jnae	Below (unsigned <) (CF = 1)
jbe <i>Label</i>	jna	Below or equal (unsigned $\leq$ =) (CF = 1 or ZF = 1)

# Setting Condition Codes

The **test** instruction is like **cmp**, but for AND. It does not store the & result anywhere. It just sets condition codes.

TEST S1, S2

S2 & S1

Instruction	Description
testb	Test byte
testw	Test word
testl	Test double word
testq	Test quad word

**Cool trick:** if we pass the same value for both operands, we can check the sign of that value using the **Sign Flag** and **Zero Flag** condition codes!

## Condition Codes

- Previously-discussed arithmetic and logical instructions update these flags. **lea** does not (it was intended only for address computations).
- Logical operations (xor, etc.) set carry and overflow flags to zero.
- Shift operations set the carry flag to the last bit shifted out and set the overflow flag to zero.
- For more complicated reasons, **inc** and **dec** set the overflow and zero flags, but leave the carry flag unchanged.

# Recap

- Assembly Execution and %rip
- Control Flow Mechanics

**Next time:** Conditional branches