#### Grupo ARCOS

uc3m Universidad Carlos III de Madrid

# Tema 3 (III)

Fundamentos de la programación en ensamblador

Estructura de Computadores Grado en Ingeniería Informática



### Contenidos

- Fundamentos básicos de la programación en ensamblador
- Ensamblador del RISC-V 32, modelo de memoria y representación de datos
- Formato de las instrucciones y modos de direccionamiento
- Llamadas a procedimientos y uso de la pila

#### Información de una instrucción

#### Las instrucciones:

- Su tamaño se ajusta a una o varias palabras
- Están divididas en campos:
  - Operación a realizar
  - Operandos a utilizar
    - ☐ Puede haber operando implícitos

op.ALU	reg.	reg.	req.
0101	0011	1000	1010
add	r1	r2	r3

#### El formato de una instrucción:

- Forma de representación de una instrucción compuesta de campos de números binarios:
  - El tamaño de los campos limita el número de valores a codificar

### Información de una instrucción

- Se utiliza unos pocos formatos:
  - Cada instrucción pertenece a un formato
  - Según el código de operación se conoce el formato asociado
- Ejemplo: formatos básicos en RISC-V

31	25	5 24 20	0 19	15	14 12	2 11	7 6	0	
func	7	rs2	rs1		funct3	rd	opcode		R-type
i	mm[11:	0]	rs1		funct3	rd	opcode		I-type
imm[1	1:5]	rs2	rs1		funct3	imm[4:0]	opcode		S-type
		imm[31:12	2]			rd	opcode		U-type

# Instrucciones y pseudoinstrucciones del RISC-V<sub>32</sub>

 Una instrucción en ensamblador se corresponde con una instrucción máquina

```
▶ Ejemplo: addi t1, t1, 2
```

- Una pseudoinstrucción en ensamblador se corresponde con una o varias instrucciones de ensamblador
  - Ejemplo I:

```
La instrucción: mv reg2, reg1
```

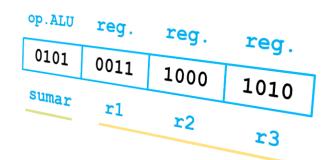
- ▶ Equivale a: add reg2, zero, reg1
- Ejemplo2:
  - ▶ La instrucción: li t1, 0x00800010
  - No cabe en 32 bits, pero puede usarse como pseudoinstrucción
  - Es equivalente a:

```
□ lui t1, 0x0080
```

□ ori t1, t1, 0x0010

# Campos de una instrucción

- En los campos se codifica:
  - Operación a realizar (código Op.)
    - Instrucción y formato de la misma



- Operandos a utilizar
  - Ubicación de los operandos
  - Ubicación del resultado
  - Ubicación de la siguiente instrucción (si op. salto)
    - □ Implícito: PC ← PC + '4' (apuntar a la siguiente instrucción)
    - □ Explícito: j 0x01004 (modifica el PC)

# Ubicaciones posibles para los operandos

I. En la propia instrucción

li  $t0 0 \times 123$ 

2. En registros (CPU)

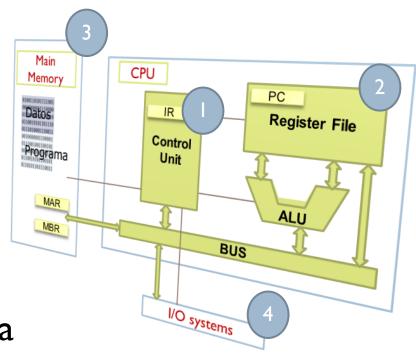
li t0 0x123

3. Memoria principal

lw t0 address(x0)

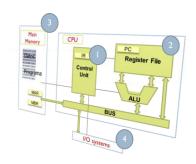
4. Unidades de Entrada/Salida

in t0 0xFEB



# Formas de indicar la ubicación de operandos: modos de direccionamiento

I. En la propia instrucción



2. En registros (CPU)

li t0 0x123

3. Memoria principal

lw t0 address(x0)

 num(registro): representa la dirección que se obtiene de sumar num con la dirección almacenada en el registro

4. Unidades de Entrada/Salida

in t0 0xFEB

#### Contenidos

- Fundamentos básicos de la programación en ensamblador
- Ensamblador del RISC-V<sub>32</sub>, modelo de memoria y representación de datos
- Formato de las instrucciones y modos de direccionamiento
- Llamadas a procedimientos y uso de la pila

## Modos de direccionamiento

▶ El modo de direccionamiento es un procedimiento que permite determinar la ubicación de un operando, un resultado o una instrucción



## Modos de direccionamiento en RISC-V

Inmediato	value
Directo	
A memoria	address
A registro	xr
Indirecto	
A memoria	
A registro	(xr)
Relativo a	
□ registro	offset(xr)
□ pila	offset(sp)
	beq label l

### Modos de direccionamiento

▶ El modo de direccionamiento es un procedimiento que permite determinar la ubicación de un operando, un resultado o una instrucción

Implícito
 Inmediato
 Directo - a registro

 a memoria

 Indirecto - a registro

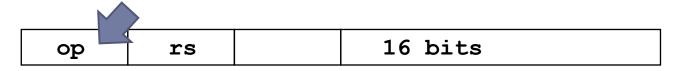
 a registro
 a memoria

 Relativo - a registro índice

 a registro base
 a PC
 a Pila

# Direccionamiento implicito

- El operando no está codificado en la instrucción, pero forma parte de esta
- ► Ejemplo: auipc a0 0x12345
  - a0 = PC + (0x12345 << 12).
  - ▶ a0 es un operando, PC es el otro (implícito)



- V/I (Ventajas/Inconvenientes)
  - ✓ Es rápido: no es necesario acceder a memoria.
  - Pero solo es posible en unos pocos casos.

#### Direccionamiento inmediato

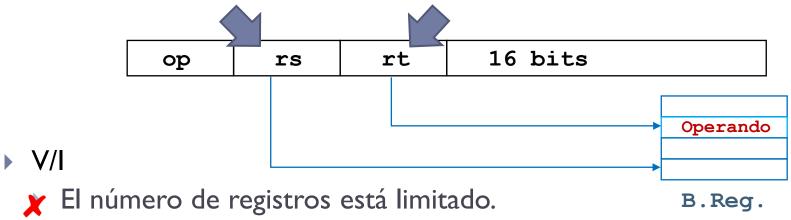
- El operando forma parte de la instrucción.
- Ejemplo: li a0 0x4f5 l
  - Carga en el registro a0 el valor inmediato 0x4f5 1.
  - El valor 0x00004f5 l está codificado en la propia instrucción.

op	rs	16 bits	

- V/I
  - ✓ Es rápido: no es necesario acceder a memoria.
  - X No siempre cabe el valor en una palabra:
    - No cabe en 32 bits, es equivalente a:
      - $\Box$  lui t1, 0x87654
      - $\square$  ori t1, t1,  $0 \times 321$

# Direccionamiento directo a registro (direccionamiento de registro)

- ▶ El operando se encuentra en el registro.
- Ejemplo: mv a0 a1
  - Copia en el registro a0 el valor que hay en el registro a1.
  - El identificador de a0 y a lestá codificado en la instrucción.



- ✓ Acceso a registros es rápido
- ✓ El número de registros es pequeño => pocos bits para su codificación, instrucciones más cortas

#### Direccionamiento directo a memoria

- El operando se encuentra en memoria, y la dirección está codificada en la instrucción.
- ► Ejemplo: LD RI #0xFFF0 (IEEE 694)



- V/I
  - X Acceso a memoria es más lento comparado con los registros
  - ✗ Direcciones largas => instrucciones más largas
  - ✓ Acceso a un gran espacio de direcciones (capacidad > B.R.)

## Modos de direccionamiento

▶ El modo de direccionamiento es un procedimiento que permite determinar la ubicación de un operando, un resultado o una instrucción



### Direccionamiento directo vs. indirecto

- En el direccionamiento directo se indica dónde está el operando:
  - En qué registro o en qué posición de memoria
- En el direccionamiento indirecto se indica dónde está la dirección del operando:
  - Hay que acceder a esa dirección en memoria
  - Se incorpora un nivel (o varios) de direccionamiento

# Direccionamiento indirecto de registro

Se indica en la instrucción el registro con la dirección del operando
 Ejemplo: | w a0 (a | )
 Carga en a0 el valor que hay en la dirección de memoria almacenada en a | .
 op rs rt 16 bits

- V/I
  - ✓ Amplio espacio de direcciones, instrucciones cortas
  - Pseudo-instrucción equivalente a lw a0 0(a l)

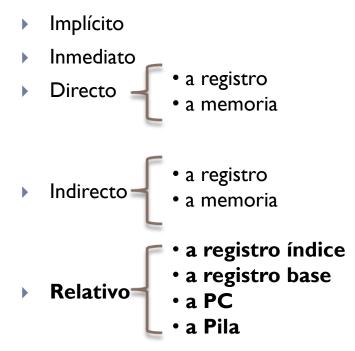
#### Direccionamiento indirecto a memoria

Se indica en la instrucción la dirección donde está la de la dirección del operando (no disponible en RISC-V)
 Ejemplo: LD RI [DIR] (IEEE 694)
 Carga en RI el valor que hay en la dirección de memoria que está almacenada en la dirección de memoria DIR.
 op rs rt 16 bits

- V/I
  - ✓ Amplio espacio de direcciones
  - ✓ El direccionamiento puede ser anidado, multinivel o en cascada
    - ► Ejemplo: LD R1 [[[.R1]]]
  - Puede requerir varios accesos memoria instrucciones más lentas de ejecutar

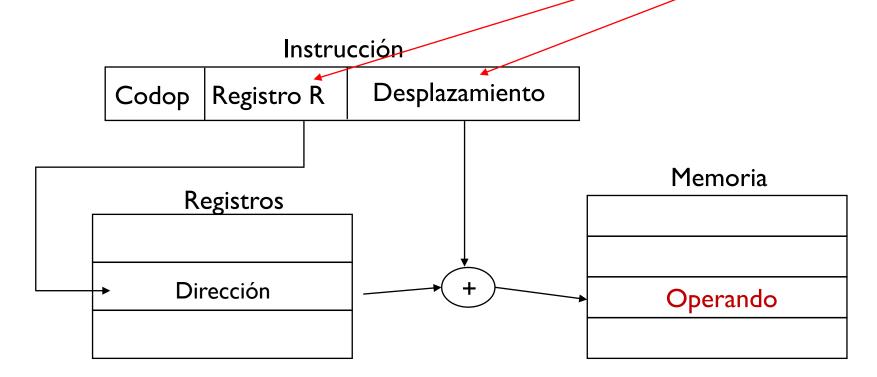
## Modos de direccionamiento

▶ El modo de direccionamiento es un procedimiento que permite determinar la ubicación de un operando, un resultado o una instrucción



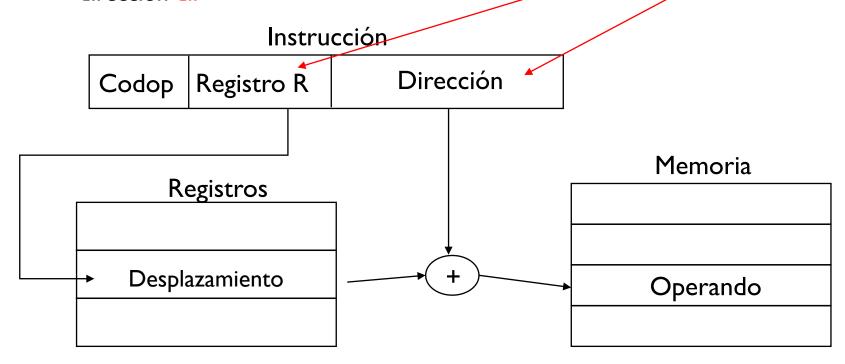
## Direccionamiento relativo a registro base

- Ejemplo: lw a0 12(t1)
  - Carga en a0 el contenido de la posición de memoria dada por tl + 12
  - Utiliza dos campos de la instrucción, t1 tiene la dirección base



## Direccionamiento relativo a registro indice

- Ejemplo: lw a0 dir(tl)
  - Carga en a0 el contenido de la posición de memoria dada por tl + dir
  - Utiliza dos campos: t1 representa el desplazamiento (índice) respecto a la dirección dir



## Utilidad: acceso a vectores

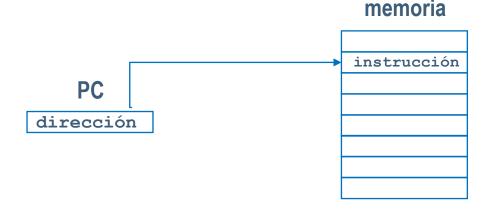
```
int v[5] ;
main ()
{
   v[3] = 5 ;

v[4] = 8 ;
}
```

```
.data
     .align 2#siguiente alineado a 4 byte
 v: .space 20 \# 5_{int} * 4_{bytes/int}
.text
main:
         la t0 v
         li t1 5
         sw t1 12(t0)
         li t0 16
         li t1 8
         sw t1 v(t0)
```

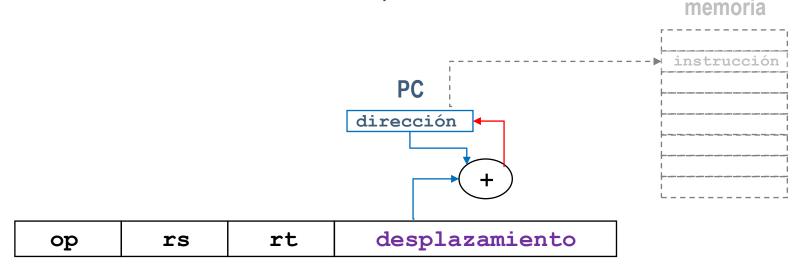
# Direccionamiento relativo al contador de programa

- El contador de programa PC:
  - Es un registro de 32 bits (4 bytes) en un computador de 32-bits
  - Almacena la dirección de la siguiente instrucción a ejecutar
    - Apunta a una palabra (4 bytes) con la instrucción a ejecutar
    - PC en un computador de 32-bits se actualiza por defecto como PC = PC + 4



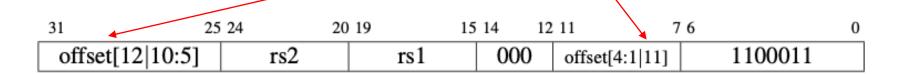
# Direccionamiento relativo al contador de programa

- ▶ Ejemplo: beq a0 x0 etiqueta
  - Se codifica etiqueta como el desplazamiento desde la dirección de memoria donde está esta instrucción, hasta la posición de memoria indicada en etiqueta.
    - ▶ Etiqueta se codifica como desplazamiento (dirección -> # instrucciones a saltar)
  - Si a0 es 0, entonces PC <= PC + "desplazamiento"</p>



#### Direccionamiento relativo a PC en el RISC-V

La instrucción beq t0, x1, ofsset se codifica en la instrucción:



- Etiqueta tiene que codificarse en el campo "ofsset"
- ¿Cómo se actualiza el PC si t0 == x1 y cuánto vale fin cuando se genera código máquina?

```
bucle: beq t0, x1, fin
add t8, t4, t4
addi t0, x0, -1
j bucle
fin: . . .
```

#### Direccionamiento relativo a PC en el RISC-V

#### Si se cumple la condición

- ▶ PC = PC + offest
- la el valor de offset puedes ser positivo o negativo

#### Por tanto en:

```
bucle: beq $t0, $1, fin
add $t8, $t4, $t4
addi $t0, $0, -1
j bucle
fin: . . .
```

- ▶ fin == 12
  - Cuando se ejecuta una instrucción, el PC apunta a la siguiente

```
li t0 8
li t1 4
li t2 1
li t4 0
while: bge t4 t1 fin
mul t2 t2 t0
addi t4 t4 1
j while
fin: mv t2 t4
```

- fin representa la dirección donde se encuentra la instrucción my
- while representa la dirección donde se encuentra la instrucción bge

Dinagaián

	li \$t0 8	
	li \$t1 4	
	li \$t2 1	
	li \$t4 0	
while:	bge \$t4 \$t1 fin	
	mul \$t2 \$t2 \$t0	
	addi \$t4 \$t4 1	
	j while	
fin:	move \$t2 \$t4	

Dirección	Contenido				
0x0000100	li	\$t0	8		
0x0000104	li	\$t1	4		
0x0000108	li	\$t2	1		
0x000010C	li	\$t4	0		
0x0000110	bge	\$t4	\$t1	fin	
0x0000114	mul	\$t2	\$t2	\$t0	
0x0000118	addi	\$t4	\$t4	1	
0x000011C	j	whil	Le		
0x0000120	move	\$t2	\$t4	1	

Cantanida

		Dirección		Contenido
	li \$t0 8 li \$t1 4	0x0000100	li :	\$t0 8
	li \$t2 1 li \$t4 0	0x0000104	li	\$t1 4
while:	bge \$t4 \$t1 <b>fin</b> mul \$t2 \$t2 \$t0	0x0000108	li	\$t2 1
	addi \$t4 \$t4 1	0x000010C	li	\$t4 0
fin:	j while move \$t2 \$t4	0x0000110	bge	\$t4 \$t1 fin
_	\$t4 \$t1 fin esenta un desplazamiento	0x0000114	mul	\$t2 \$t2 \$t0
respecto	o al PC actual $\Rightarrow$ 12 PC + 12	0x0000118	addi	\$t4 \$t4 1
En j while		0x000011C	j	while
while rep	presenta un desplazamiento	0x0000120	move	\$t2 \$t4

respecto al PC actual =>-16

PC = PC - 16

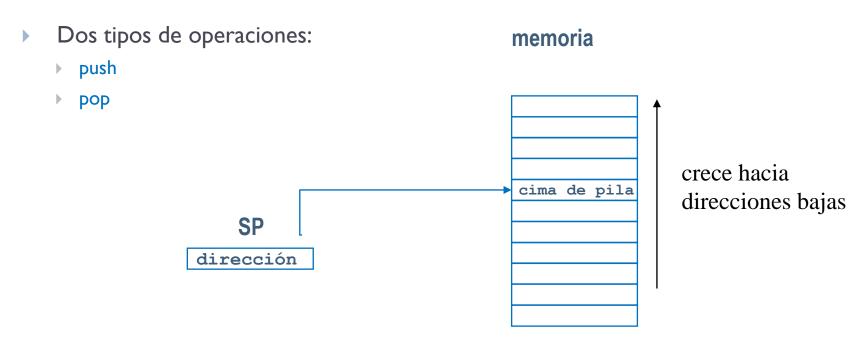
			Dirección	
	li	\$t0 8	0x0000100	1
	li	\$t1 4		
	li	\$t2 1	0x0000104	1
	li	\$t4 0		
while:	bge	\$t4 \$t1 <b>fin</b>	0x0000108	1
	mul	\$t2 \$t2 \$t0	0000100	
	addi	\$t4 \$t4 1	0x000010C	1
	j	while	0x0000110	
fin:	move	\$t2 \$t4	0110000110	b
En bge			0x0000114	m
fin repre	esenta un	n desplazamiento	0x0000118	
respecto	al PC a	ctual -> 12	0X0000110	а

fin representa un desplazamiento
respecto al PC actual => 12
PC = PC + 12
En j while
while representa un desplazamiento
respecto al PC actual =>-16
PC = PC - 16

Dirección	Contenido			
0x0000100	li :	\$t0	8	
0x0000104	li	\$t1	4	
0x0000108	li	\$t2	1	
0x000010C	li	\$t4	0	
0x0000110	bge	\$t4	\$t1	12
0×0000114	mul	\$t2	\$t2	\$t0
0×0000118	addi	\$t4	\$t4	1
0x000011C	j	-16	5	
0x0000120	move	\$t2	\$t4	1

# Direccionamiento relativo a pila

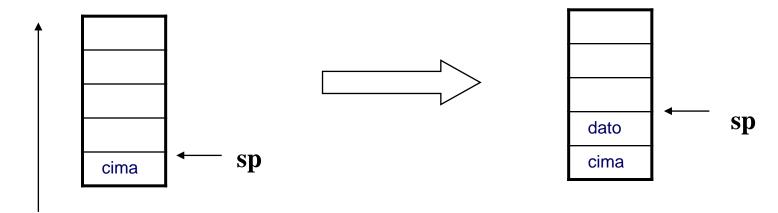
- ▶ El puntero de pila SP (Stack Pointer):
  - ► Es un registro de 32 bits (4 bytes) en el RISC-V<sub>32</sub>
  - Almacena la dirección de la cima de pila
    - Apunta a una palabra (4 bytes)



# Operación PUSH

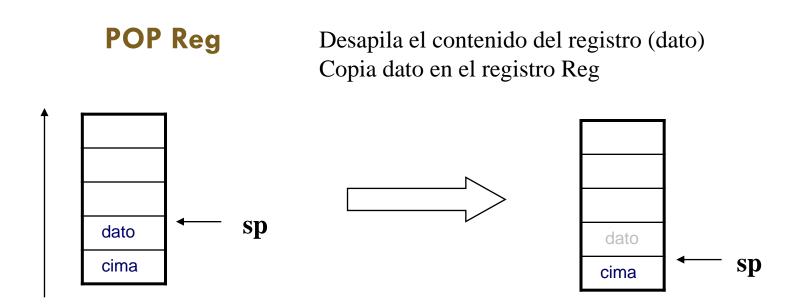
#### **PUSH Reg**

Apila el contenido del registro (dato)



crece hacia direcciones bajas

# Operación POP



crece hacia direcciones bajas

# Direccionamiento de pila en el RISC-V

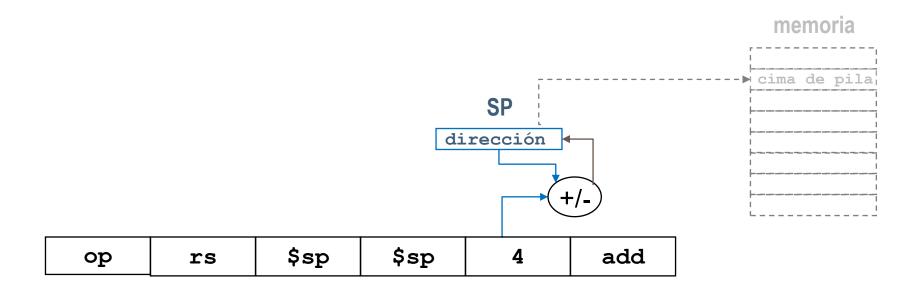
- RISC-V no dispone de instrucciones PUSH o POP.
- El registro puntero de pila (sp) es visible al programador.
  - Se va a asumir que el puntero de pila apunta al último elemento de la pila

#### PUSH t0

#### POP t0

## Operación PUSH en RISC-V

- ▶ Ejemplo: push a0
  - addi sp sp -4 # SP = SP 4
  - $\triangleright$  sw a0 0(sp) # memoria[SP] = a0



# Ejercicio

- Indique el tipo de direccionamiento usado en las siguientes instrucciones RISC-V:
  - I. li tl 4
  - 2. lw t0 4(a0)
  - 3. bne  $\times 0$  a0 etiqueta

1. li t | 4 -> directo a registro -> inmediato lw t0 4(a0) t0 -> directo a registro 4(a0) -> relativo a registro base bne x0 a0 etiqueta -> directo a registro a0 etiqueta -> relativo a contador de programa

# Ejemplos de tipos de direccionamiento

#### ▶ la t0 label

#### inmediato

- El segundo operando de la instrucción es una dirección
- PERO no se accede a esta dirección, la propia dirección es el operando

#### Iw t0 label

#### directo a memoria (!)

- El segundo operando de la instrucción es una dirección
- Hay que acceder a esta dirección para tener el valor con el que trabajar

#### bne t0 t1 label relativo a registro PC

- El tercer operando de la instrucción es desplazamiento respecto al PC
- label se codifica como un número en complemento a dos que representa el desplazamiento (como palabras) relativo al registro PC

#### Contenidos

- Fundamentos básicos de la programación en ensamblador
- Ensamblador del RISC-V<sub>32</sub>, modelo de memoria y representación de datos
- Formato de las instrucciones y modos de direccionamiento
- Llamadas a procedimientos y uso de la pila

#### Formato de instrucciones

- Una instrucción se divide en campos
- Una instrucción máquina es autocontenida e incluye:
  - Código de operación
  - Operandos
    - ▶ Tipos de representación de los operandos
  - Resultado
  - Dirección de la siguiente instrucción
- Ejemplo de campos en una instrucción del RISC-V:

op.	rd	fun.3	rs1	rs2	fun.7
-----	----	-------	-----	-----	-------

#### Formato de una instrucción

- El formato especifica, por cada campo de la instrucción:
  - El significado de cada campo
  - El número de bits de cada campo
  - Cómo se codifica cada campo
    - ▶ Binario, complemento a uno, a dos, etc.

#### Normalmente:

- Una arquitectura ofrece unos pocos formatos de instrucción.
  - Simplicidad en el diseño de la unidad de control.
- Campos del mismo tipo siempre igual longitud.
- Selección mediante código de operación.
  - Normalmente el primer campo.

# Longitud de formato

- La longitud del formato es el número de bits para codificar la instrucción
  - El tamaño habitual es una palabra (o múltiples palabras)
  - ▶ En RISC-V<sub>32</sub> el tamaño de todas las instrucciones es una palabra.

#### Dos tipos:

- Longitud única:
  - Todas las instrucciones tienen la misma longitud de formato.
  - Ejemplos:
    - ☐ MIPS32: 32 bits, PowerPC: 32 bits, ...
- ► Longitud variable:
  - Distintas instrucciones tienen distinta longitud de formato.

  - **Ejemplos:** 
    - □ IA32 (Procesadores Intel): Número variable de bytes.

# Código de operación

#### Tamaño fijo:

- n bits → 2<sup>n</sup> códigos de operación.
- ▶ m códigos de operación → log₂m bits.

#### Campos de extensión

- ▶ RISC-V (instrucciones aritméticas-lógicas)
- Op = 0; la instrucción está codificada en fun. X

31	25	24 20	19 15	5 14 12	11 7	6 0	)
	funct7	rs2	rs1	funct3	$^{\mathrm{rd}}$	opcode	R-type

#### Tamaño variable:

Instrucciones más frecuentes = Tamaños más cortos.

# Ejemplo: Formato de las instrucciones del RISC-V

R-type
I-type
S-type
U-type

# Ejemplo de formato en el RISC-V

#### RISC-V Instruction:

▶ add rd rs1 rs2

31 2	5 24 20	19 15	14 12	11 7	6	0
0000000	rs2	rs1	000	rd	0110011	7

# Ejercicio

Sea un computador de 16 bits de tamaño de palabra, que incluye un repertorio con 60 instrucciones máquina y con un banco de registros que incluye 8 registros.

#### Se pide:

Indicar el formato de la instrucción ADDx RI R2 R3, donde RI, R2 y R3 son registros.

palabra -> 16 bits
60 instrucciones
8 registros (en BR)
ADDx R1(reg.), R2(reg.), R3(reg.)

Palabra de 16 bits define el tamaño de la instrucción

I 6 bits

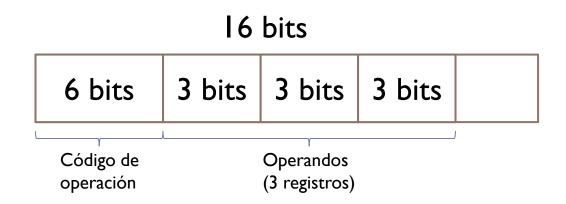
palabra -> 16 bits 60 instrucciones 8 registros (en BR) ADDx RI(reg.), R2(reg.), R3(reg.)

Para 60 instrucciones se necesitan 6 bits (mínimo)



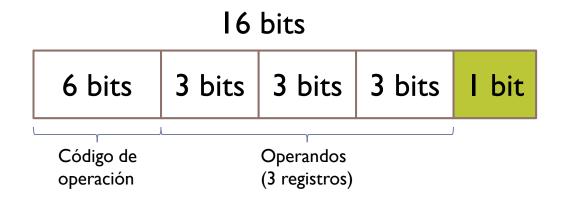
palabra -> 16 bits
60 instrucciones
8 registros (en BR)
ADDx R1(reg.), R2(reg.), R3(reg.)

Para 8 registros se necesitan 3 bits (mínimo)



palabra -> 16 bits
60 instrucciones
8 registros (en BR)
ADDx R1(reg.), R2(reg.), R3(reg.)

▶ Sobra I bit (16-6-3-3-3 = I), usado de relleno



# Juego de instrucciones

#### Queda definido por:

- Conjunto de instrucciones
- Formato de las instrucciones
- Registros
- Modos de direccionamiento
- Tipos de datos y formatos

# Juego de instrucciones

- Distintas formas para la clasificación de un juego de instrucciones:
  - Complejidad del juego de instrucciones
    - CISC vs RISC
  - Modo de ejecución
    - Pila
    - Registro
    - ▶ Registro-Memoria, Memoria-Registro, ...

#### CISC vs RISC

- Complex Instruction Set Computer
  - Muchas instrucciones
  - Instrucciones complejas
    - Más de una palabra
    - Unidad de control más compleja
    - Mayor tiempo de ejecución
  - Diseño irregular

- Alrededor del 20% de las instrucciones ocupa el 80% del tiempo total de ejecución de un programa
- ▶ El 80% de las instrucciones no se utilizan casi nunca
- 80% del silicio infrautilizado, complejo y costoso

- Reduced Instruction Set Computer
  - Instrucciones simples y ortogonales
    - Ocupan una palabra
    - Instrucciones sobre registros
    - Uso de los mismos modos de direccionamiento para todas las instrucciones (alto grado de ortogonalidad)
  - Diseño más compacto:
    - Unidad de control más sencilla y rápida
    - Espacio sobrante para más registros y memoria caché

# Modos de ejecución

- Los modos de ejecución indican el número de operandos y el tipo de operandos que pueden especificarse en una instrucción.
  - ▶ 0 direcciones → Pila.
    - □ PUSH 5; PUSH 7; ADD
  - ▶ I dirección → Registro acumulador.
    - □ ADD RI -> AC <- AC + RI
  - ▶ 2 direcciones → Registros, Registro-memoria, Memoria-memoria.
    - $\square$  ADD .R0, .RI (R0 <- R0 + RI)
  - ▶ 3 direcciones → Registros, Registro-memoria, Memoria-memoria.
    - □ ADD .R0, .R1, .R2

# Ejercicio

Sea un computador de 16 bits, que direcciona la memoria por bytes y que incluye un repertorio con 60 instrucciones máquina. El banco de registros incluye 8 registros. Indicar el formato de la instrucción ADDV RI, R2, M, donde RI y R2 son registros y M es una dirección de memoria.

# Ejercicio

- Sea un computador de 32 bits, que direcciona la memoria por bytes. El computador incluye 64 instrucciones máquina y 128 registros. Considere la instrucción SWAPM dir I, dir 2, que intercambia el contenido de las posiciones de memoria dir I y dir 2. Se pide:
  - Indicar el espacio de memoria direccionable en este computador.
  - Indicar el formato de la instrucción anterior.
  - Especifique un fragmento de programa en ensamblador del RISC-V 32 equivalente a la instrucción máquina anterior.
  - Si se fuerza a que la instrucción quepa en una palabra, qué rango de direcciones se podría contemplar considerando que las direcciones se representan en binario puro.

#### Grupo ARCOS

uc3m Universidad Carlos III de Madrid

### Tema 3 (III)

Fundamentos de la programación en ensamblador

Estructura de Computadores Grado en Ingeniería Informática

