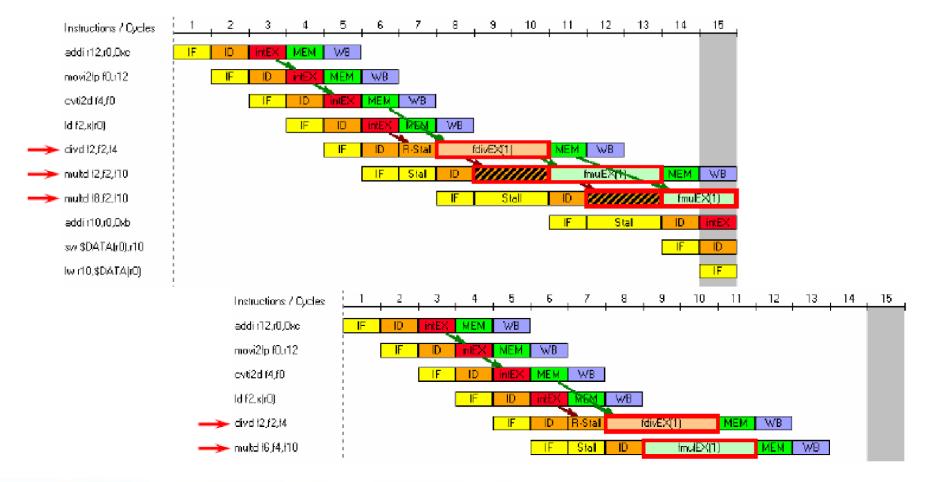
# Ingeniería de los Computadores

Unidad 2. Superescalares

# Unidad 2. Superescalares 2.1. Introducción y motivación

#### Motivación

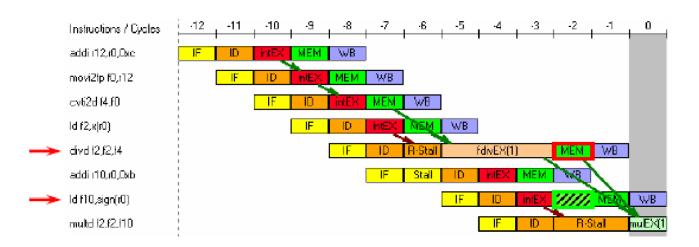
- Dependencias estructurales provocan pérdidas de ciclos
  - > Ejemplo de una unidad FP vs. Varias unidades FP



### Unidad 2. Superescalares 2.1. Introducción y motivación

#### Motivación

- Varias unidades funcionales permiten la ejecución fuera de orden
  - Validar riesgos WAR y WAW



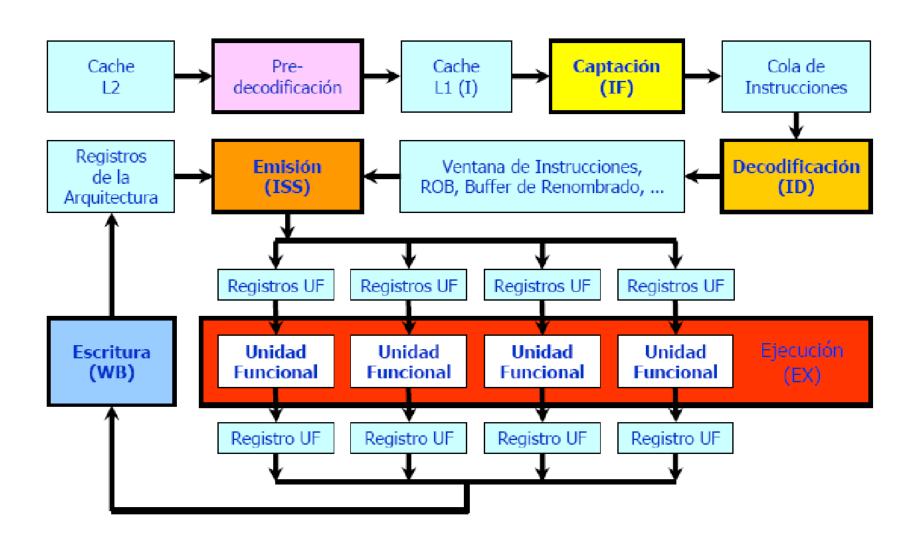
Motivación

Cauce

#### Etapas

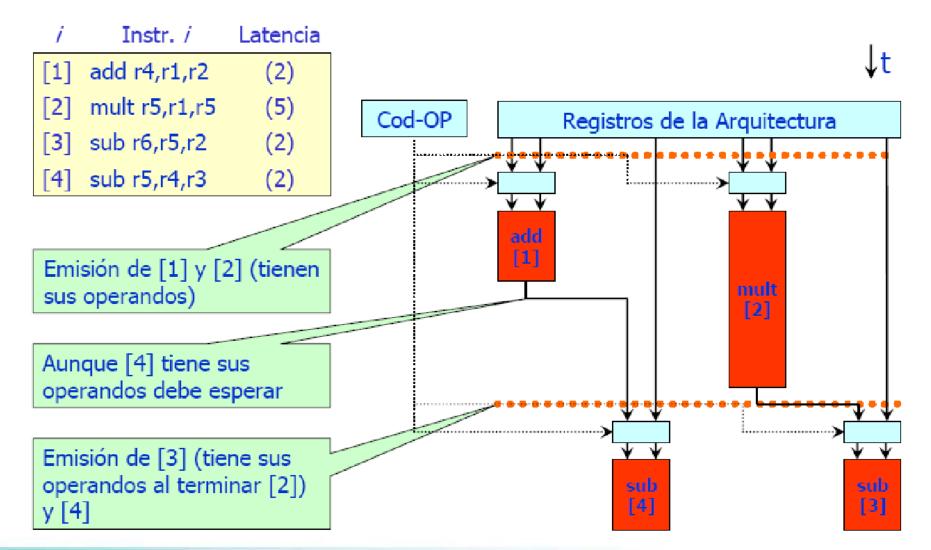
- Captación de instrucciones (IF)
- Decodificación de instrucciones (ID)
- Emisión de instrucciones (ISS)
- Ejecución de instrucciones (EX) Instrucción finalizada o "finish"
- Escritura (WB) Instrucción completada o "complete"
- Características del procesamiento superescalar
  - Diferentes tipos de órdenes: orden de captación, orden de emisión, orden de finalización
  - Capacidad para identificar ILP existente y organizar el uso de las distintas etapas para optimizar recursos

Motivación Cauce



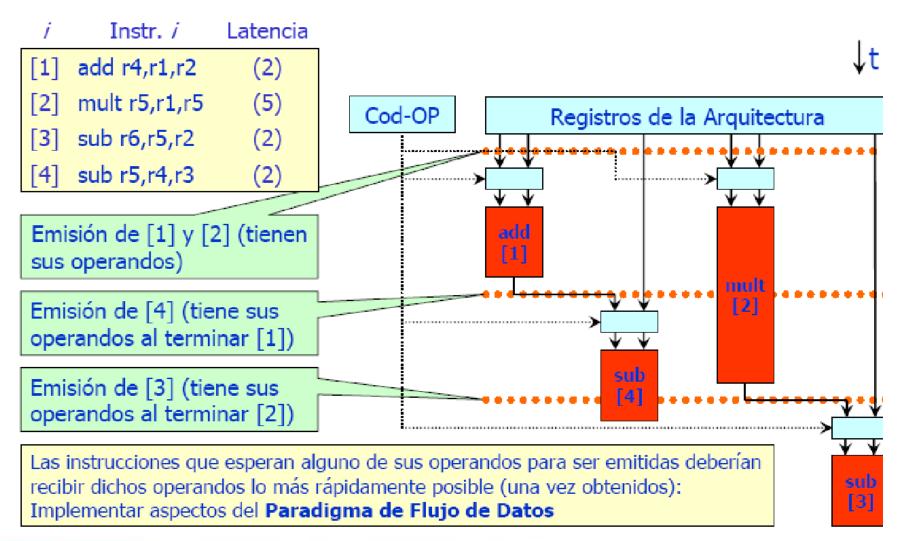
Motivación Cauce

#### Emisión ordenada



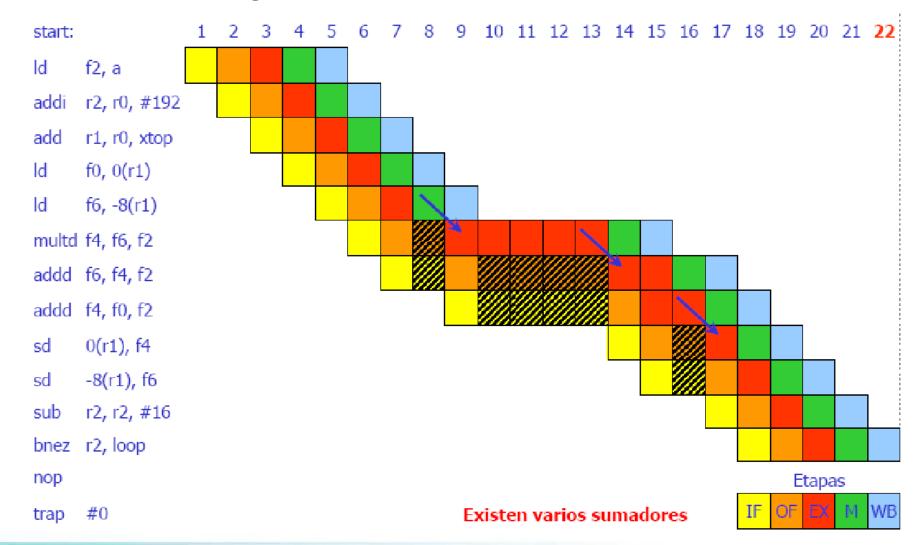
Motivación Cauce

#### Emisión desordenada



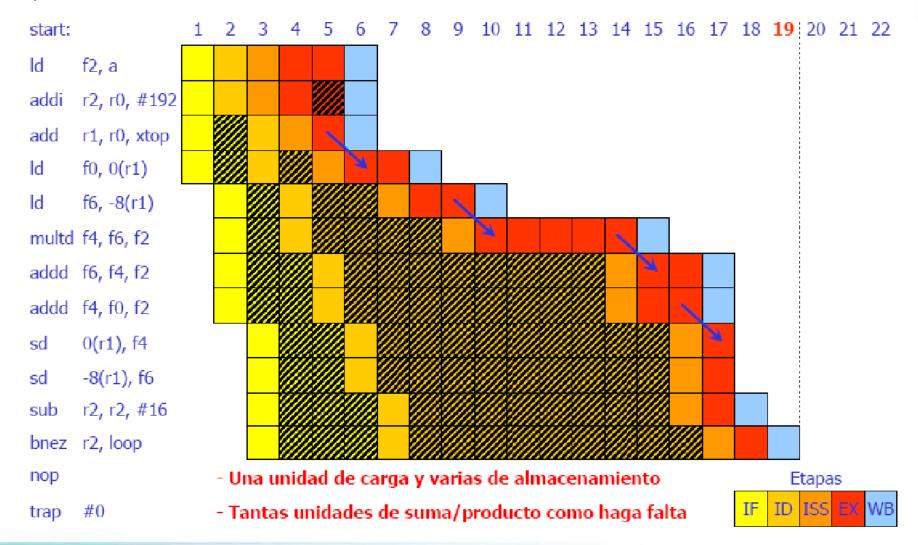
Motivación Cauce

#### Procesamiento segmentado



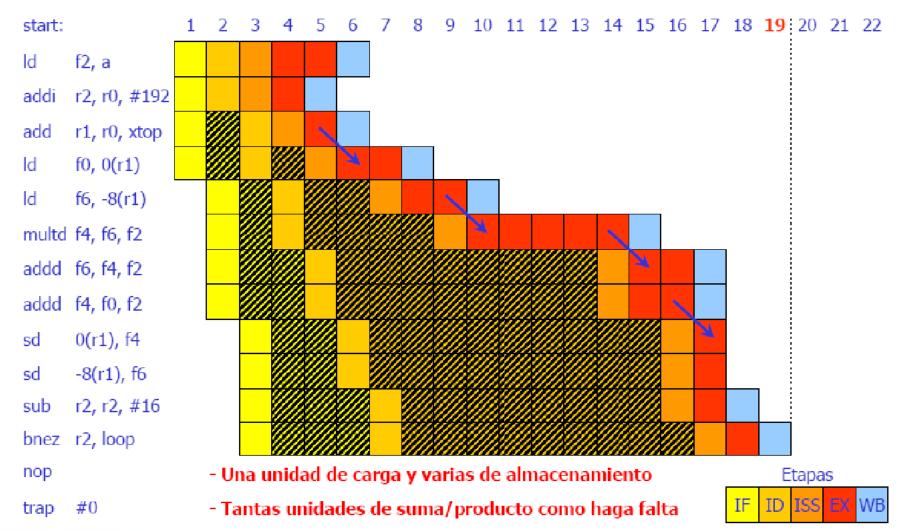
Motivación E Cauce

• Superescalar: emisión ordenada/finalización ordenada



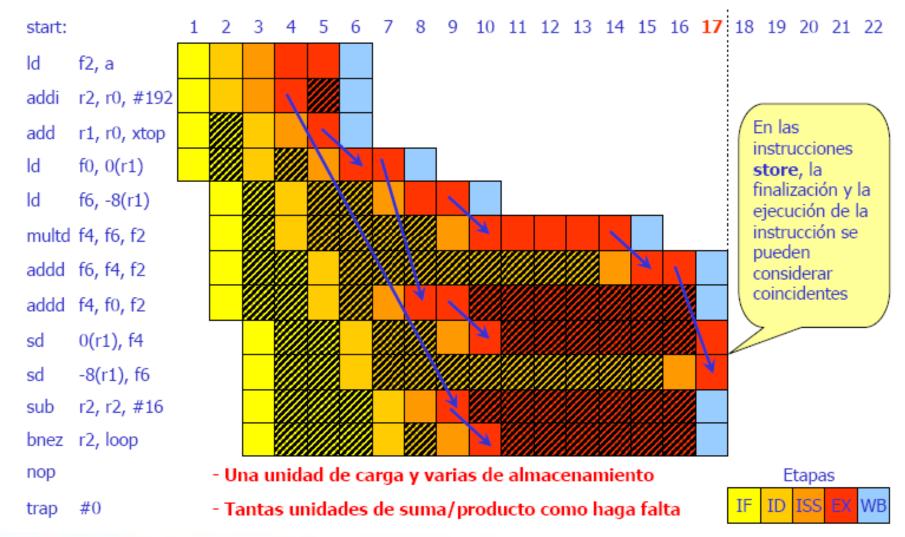
Motivación Cauce

Superescalar: emisión ordenada/finalización desordenada



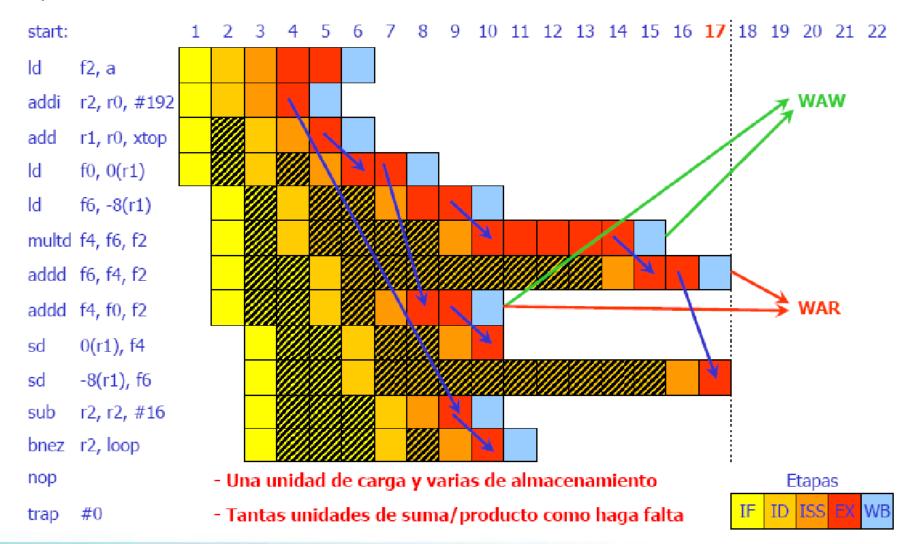
Motivación Cauce

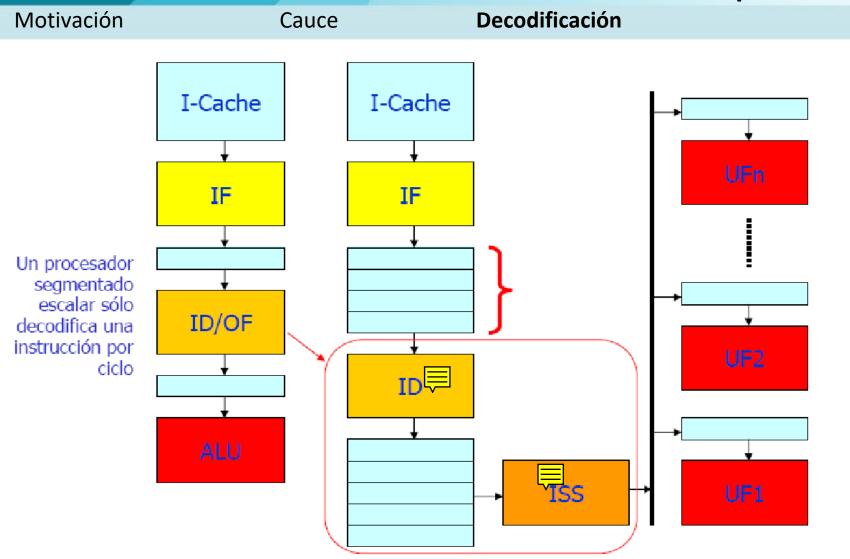
Superescalar: emisión desordenada/finalización ordenada



Motivación Cauce

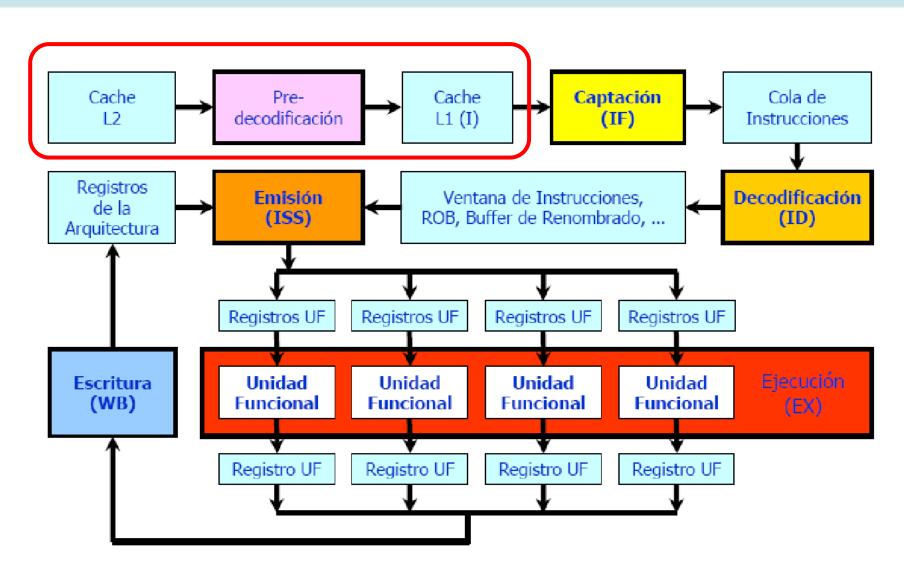
Superescalar: emisión desordenada/finalización desordenada





En un procesador superescalar se han de decodificar varias instrucciones por ciclo (y comprobar las dependencias con las instrucciones que se están ejecutando)

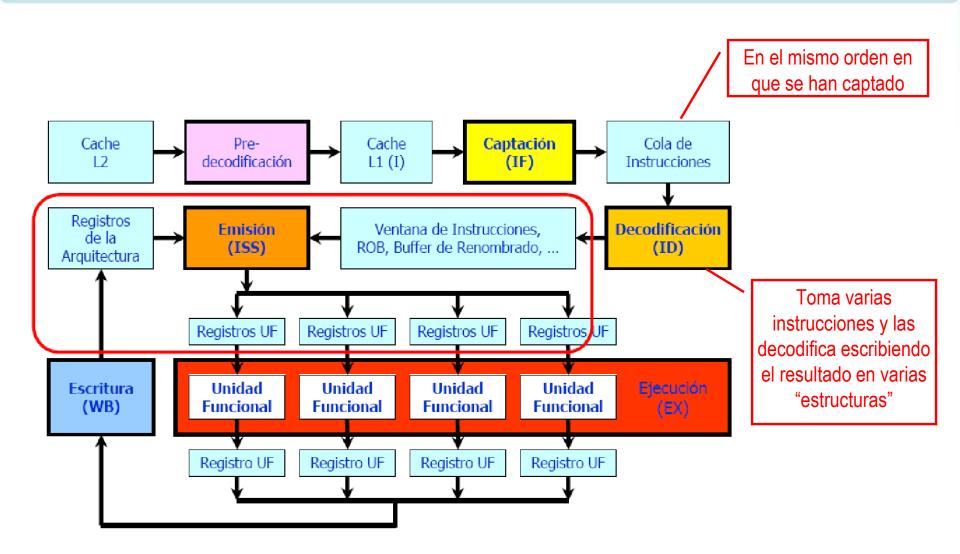
Motivación Cauce **Decodificación** 



Motivación Cauce **Decodificación** 

- Bits de predecodificación pueden indicar:
  - > Si es una instrucción de salto o no (se puede empezar su procesamiento antes)
  - El tipo de unidad funcional que va a utilizar (se puede emitir más rápidamente si hay cauces para enteros o coma flotante...)
  - > Si hace referencia a memoria o no

Motivación Cauce Decodificación **Emisión** 



Motivación Cauce Decodificación Emisión

#### Ventana de instrucciones:

- La ventana de instrucciones almacena las instrucciones pendientes (todas, si la ventana es centralizada o las de un tipo determinado, si es distribuida)
- Las instrucciones se cargan en la ventana una vez decodificadas y se utiliza un bit para indicar si un operando está disponible (se almacena el valor o se indica el registro desde donde se lee) o no (se almacena la unidad funcional desde donde llegará el operando)
- Una instrucción puede ser emitida cuando tiene todos sus operandos disponibles y la unidad funcional donde se procesará. Hay diversas posibilidades para el caso en el que varias instrucciones estén disponibles (características de los buses, etc.)

Motivación Cauce Decodificación **Emisión** 

Ejemplo de Ventana de Instrucciones

Dato no válido (indica desde dónde se recibirá el dato)

#	opcode	address	rb_entry	operand1	ok1	typ1	operand2	ok2	typ2	pred
2	MULTD	loop + 0x4	2	1	0	FPD	0	0	FPD	_
1	LD	loop	1	0	0	INT	0	1	IMM	_

Lugar donde se / almacenará el resultado Dato válido (igual a 0)

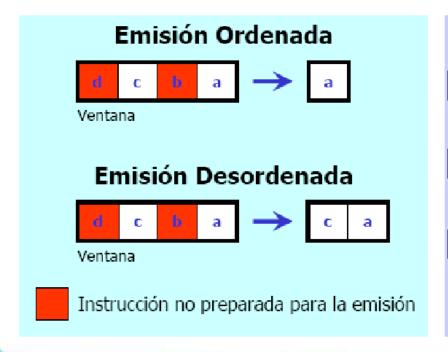
Motivación Cauce Decodificación **Emisión** 

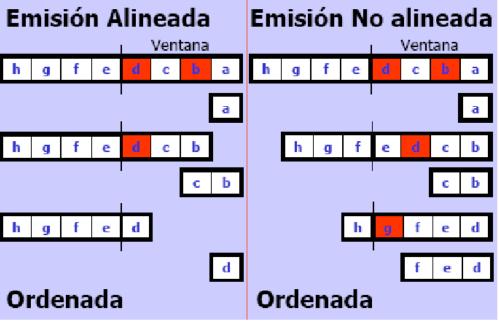
- Orden: Emisión Ordenada o Desordenada
- Alineamiento: Emisión Alineada o No alineada

#### **Ejemplos:**

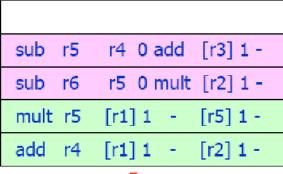
**Alineada:** SuperSparc (92), PowerPC (93, 95, 96), PA8000 (96), Alpha (92, 94, 95), R10000 (96)

No Alineada: MC88110 (93), PA7100LC (93), R8000 (94), UltraSparc (95)





Decodificación Motivación Cauce **Emisión** 





Ventana de Registros

add	/sub: t: 1	2
mul	t: 1	

[2] mult r5,r1,r5 (5)

(2)

[1] add r4,r1,r2

[3] sub r6,r5,r2 (2)

[4] sub r5,r4,r3 (2)

J	L
٨	,
	▼

Se han emitido [1] y [2]

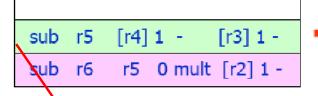
sub	r5	r4 0 add	[r3] 1 -
sub	r6	r5 0 mult	[r2] 1 -

Instrucc.	ISS	EXE
add	(1)	(2)-(3)
mult	(1)	(2)-(6)
sub	(7)	(8)-(9)
sub	(7)	(8)-(9)



Ha terminado [1]

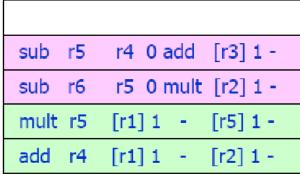
Ha terminado [2]: pueden emitirse [3] y [4]



sub	r5	[r4] :	- ا	[r3] 1 -
sub	r6	[r5] :	L -	[r2] 1 -

[4] puede emitirse pero debe esperar a la [3]

Motivación Cauce Decodificación Emisión





Ventana de

Registros

add/sub: 2 mult: 1

[1] add r4,r1,r2 (2)

[2] mult r5,r1,r5 (5)

[3] sub r6,r5,r2 (2)

[4] sub r5,r4,r3 (2)



Se han emitido [1] y [2]

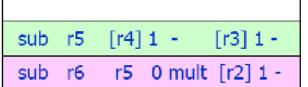
sub	r5	r4 0 add	[r3] 1 -
sub	r6	r5 0 mult	[r2] 1 -



Instrucc.	ISS	EXE
add	(1)	(2)-(3)
mult	(1)	(2)-(6)
sub	(7)	(8)-(9)
sub	(4)	(5)-(6)



Se ha emitido [4] y ha terminado [2]: puede emitirse [3]





sub	r6	[r5]	1 -	[r2] 1 -

Motivación Cauce Decodificación Emisión

sub r5 r4 0 add [r3] 1 sub r6 r5 0 mult [r2] 1 mult r5 [r1] 1 - [r5] 1 add r4 [r1] 1 - [r2] 1 -



add/sub: 1

mult: 1

Ventana de Registros

1	add	r4,r1,r2	(2)
		, ,	

	L	
١	7	
	₹	

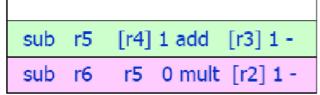
Se han emitido [1] y [2]

sub	r5	r4 0 add	[r3] 1 -
sub	r6	r5 0 mult	[r2] 1 -

Instrucc.	ISS	EXE
add	(1)	(2)-(3)
mult	(1)	(2)-(6)
sub	(7)	(8)-(9)
sub	(10)	(11)-(12)

Ha terminado [1]

Ha terminado [2] y se ha emitido [3]. Cuando la unidad quede libre se emitirá [4]





sub r5 [r4] 1 - [r3] 1 -

Decodificación **Emisión** Motivación Cauce

sub	r5	r4 0 add	[r3] 1 -
sub	r6	r5 0 mult	[r2] 1 -
mult	r5	[r1] 1 -	[r5] 1 -
add	r4	[r1] 1 -	[r2] 1 -
		_	



add/sub: 1

mult: 1

Se ganan

3 ciclos

[1] add r4,r1,r2 (2)

[2] mult r5,r1,r5 (5)

[3] sub r6,r5,r2 (2)

[4] sub r5,r4,r3 (2)

mult [r2] 1 -	Ventana de
	Registros
- [r2] 1 -	



Se han emitido [1] y [2]



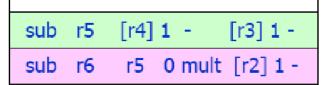
Instrucc.	155	EXE
add	(1)	(2)-(3)
mult	(1)	(2)-(6)
sub	(7)	(8)-(9)
sub	(4)	(5)-(6)

sub	r5	r4 0 add	[r3] 1 -
sub	r6	r5 0 mult	[r2] 1 -



Ha terminado [1]

Se ha emitido [4] y ha terminado [2]: puede emitirse [3]



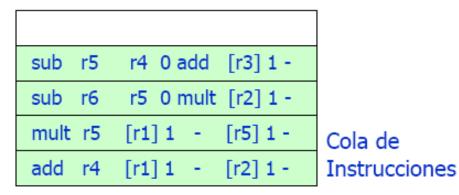


sub r6 [r5] 1 -[r2] 1 -

Motivación Cauce Decodificación **Emisión** 

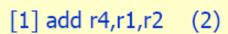


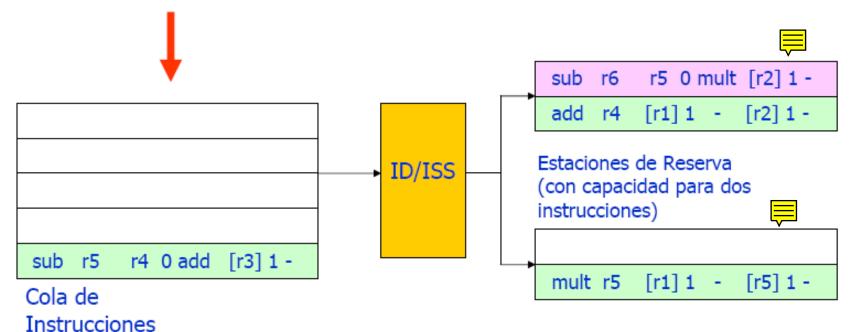
#### Estaciones de reserva



add/sub: 1

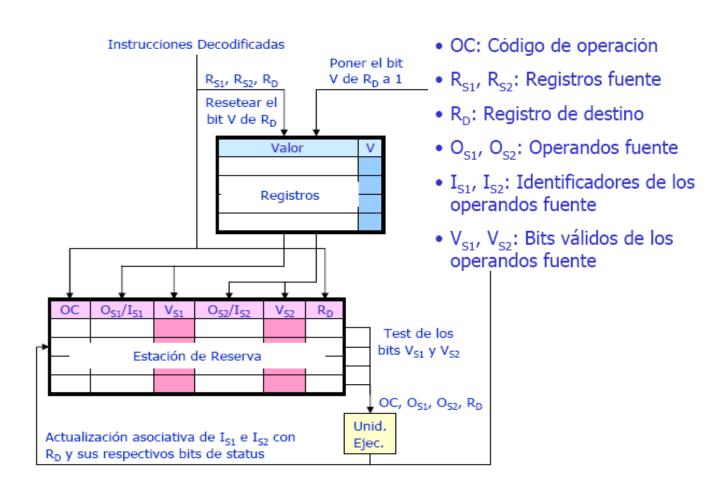
mult: 1





Motivación Cauce Decodificación Emisión

Estaciones de reserva



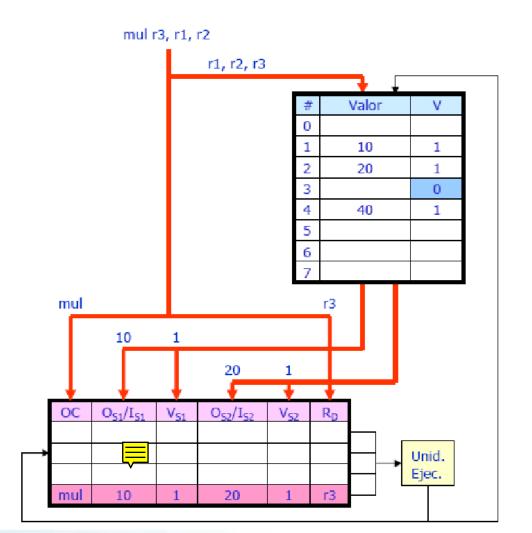
Motivación Cauce Decodificación Emisión

Estaciones de reserva. Ejemplo de uso (1)

Ciclo i: mul r3, r1, r2
Ciclo i+1: add r5, r2, r3
add r6, r3, r4

#### Ciclo i:

- Se emite la instrucción de multiplicación, ya decodificada, a la estación de reserva
- Se anula el valor de r3 en el banco de registros
- Se copian los valores de r1 y r2 (disponibles) en la estación de reserva

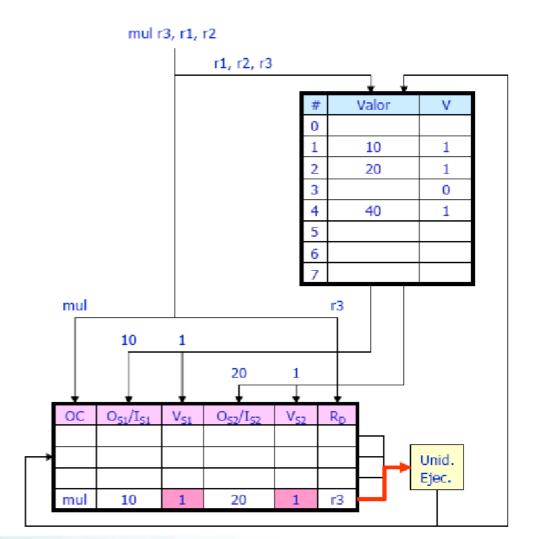


Motivación Cauce Decodificación **Emisión** 

Estaciones de reserva. Ejemplo de uso (2)

#### Ciclo i + 1:

- La operación de multiplicación tiene sus operadores preparados (V<sub>S1</sub> = 1 y V<sub>S2</sub> = 1)
- Así que puede enviarse a la unidad de ejecución

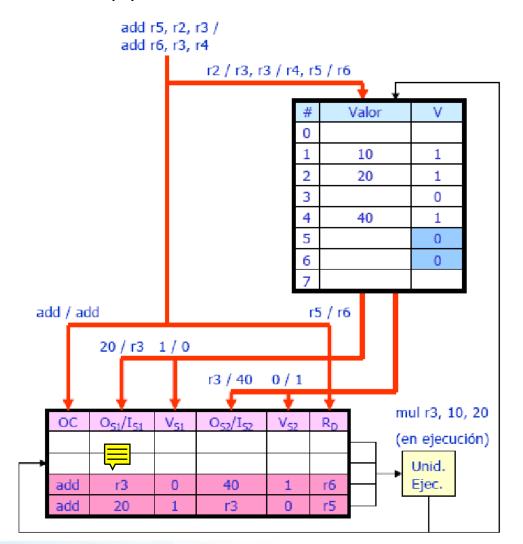


Motivación Cauce Decodificación Emisión

Estaciones de reserva. Ejemplo de uso (3)

#### Ciclo i + 1 (*cont.*):

- Se emiten las dos instrucciones de suma a la estación de reserva
- Se anulan los valores de r5 y r6 en el banco de registros
- Se copian los valores de los operandos disponibles y los identificadores de los operandos no preparados



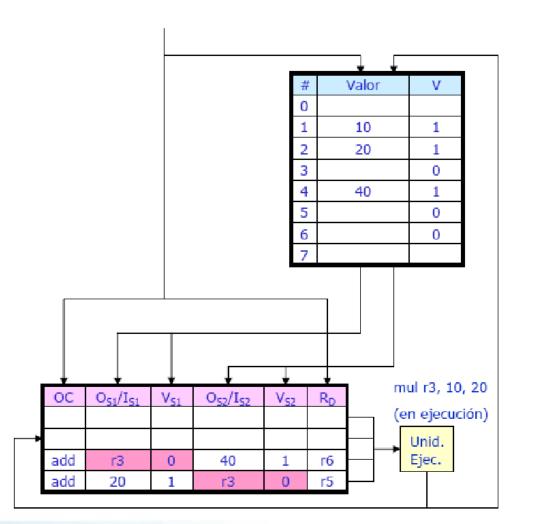
Motivación Cauce Decodificación Emisión

Estaciones de reserva. Ejemplo de uso (4)

Ciclo i: mul r3, r1, r2
Ciclo i+1: add r5, r2, r3
add r6, r3, r4

#### Ciclos i + 2 ... i + 5:

- La multiplicación sigue ejecutándose
- No se puede ejecutar ninguna suma hasta que esté disponible el resultado de la multiplicación (r3)



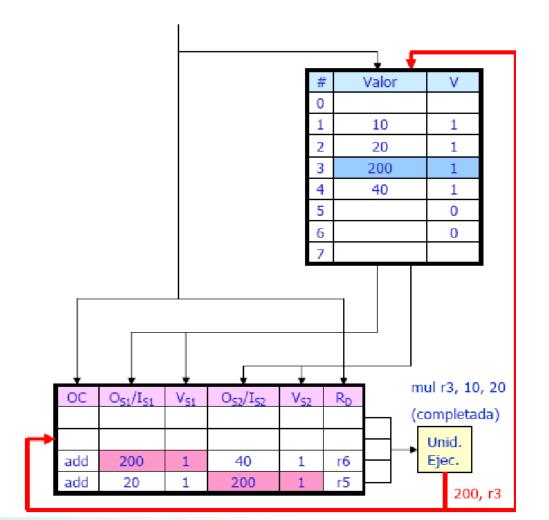
Motivación Cauce Decodificación Emisión

Estaciones de reserva. Ejemplo de uso (5)

Ciclo i: mul r3, r1, r2
Ciclo i+1: add r5, r2, r3
add r6, r3, r4

#### Ciclo i + 6:

- Se escribe el resultado de la multiplicación en el banco de registros y en las entradas de la estación de reserva
- Se actualizan los bits de disponibilidad de r3 en el banco de registros y en la estación de reserva



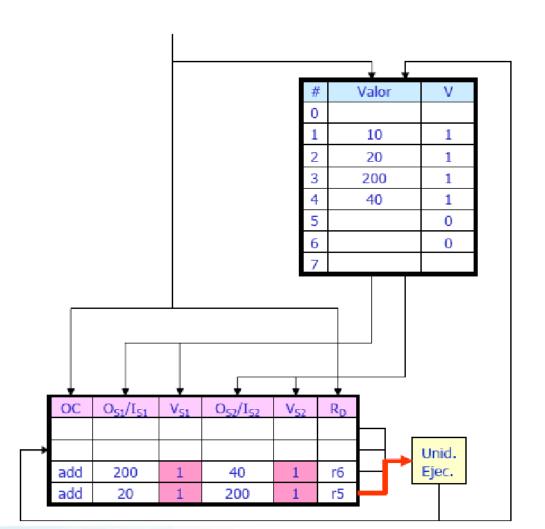
Motivación Cauce Decodificación **Emisión** 

Estaciones de reserva. Ejemplo de uso (6)

Ciclo i: mul r3, r1, r2
Ciclo i+1: add r5, r2, r3
add r6, r3, r4

#### Ciclo i + 6 (cont.):

- Las sumas tienen sus operadores preparados (VS1 = 1 y VS2 = 1)
- Así que pueden enviarse a la unidad de ejecución



Motivación Cauce Decodificación **Emisión** 

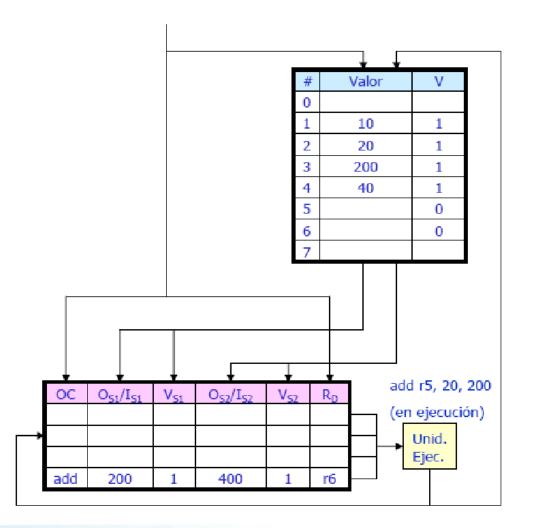
Estaciones de reserva. Ejemplo de uso (7)

Ciclo i: mul r3, r1, r2 Ciclo i+1: add r5, r2, r3

add r6, r3, r4

#### Ciclo i + 6 (*cont.*):

 Como sólo hay una unidad de ejecución, se envía la instrucción más antigua de la estación de reserva, la primera suma



Motivación Cauce Decodificación Emisión

Ejercicio

Emisión ordenada/desordenada

Recursos: 1 lw/sw, 1 add/sub, 1 mult/div

Recursos: 1 lw/sw, 2 add/sub, 2 mult/div

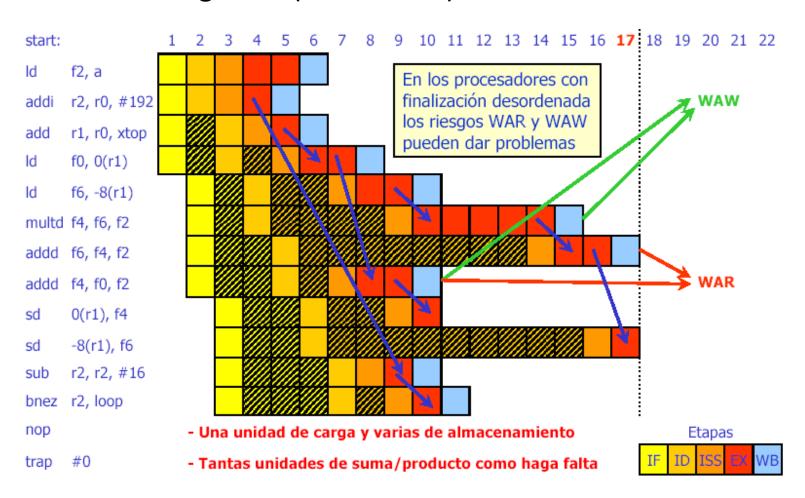
Latencia: lw/sw (5 ciclos), add/sub (2 ciclos), mult/div (5 ciclos)

Lw r1,0(r2)
Add r2,r1,r3
Mult r3,r1,r2
Sub r4,r1,r2
Add r4,r1,r2
Div r4,r5,r6

#### Unidad 2. Superescalares 2.3. Gestión de riesgos (dependencia datos)

#### Renombrado

Renombrado de registros (motivación)



### Unidad 2. Superescalares 2.3. Gestión de riesgos (dependencia datos)

#### Renombrado

Técnica para evitar el efecto de las dependencias WAR, o Antidependencias (en la emisión desordenada) y WAW, o Dependencias de Salida (en la ejecución desordenada).

R4 := R3 + 1

R3 := R5 + 1

R7 := R3 \* R4



Cada escritura se asigna a un registro físico distinto R3b := R3a - R5a

R4a := R3b + 1

R3c := R5a + 1

R7a := R3c \* R4a

Sólo RAW

R.M. Tomasulo (67)

Implementación Estática: Durante la Compilación

Implementación Dinámica: Durante la Ejecución (circuitería adicional y registros extra)

#### Características de los Buffers de Renombrado

- **Tipos de Buffers** (separados o mezclados con los registros de la arquitectura)
- Número de Buffers de Renombrado
- Mecanismos para acceder a los Buffers (asociativos o Indexados)

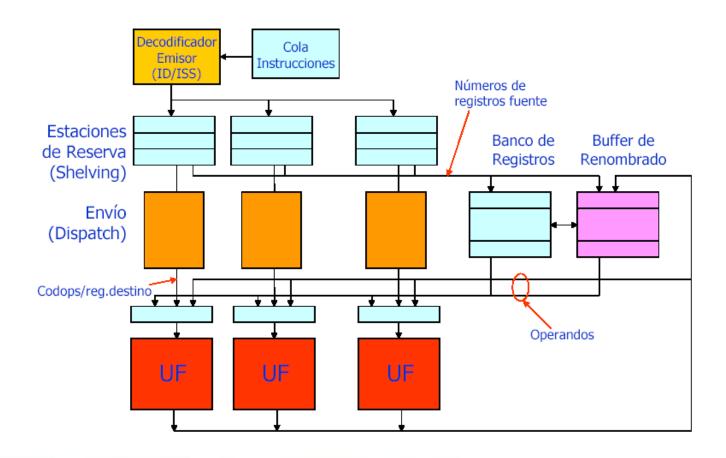
#### Velocidad del Renombrado

Máximo número de nombres asignados por ciclo que admite el procesador

### Unidad 2. Superescalares 2.3. Gestión de riesgos (dependencia datos)

#### Renombrado

- Renombrado de registros
  - Estaciones de reserva + buffer de renombrado

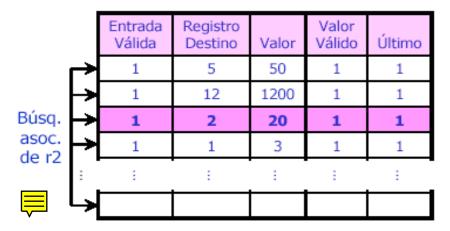


#### Renombrado

Tipos de buffers de renombrado

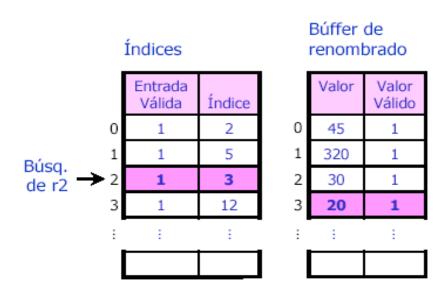
### Buffer de Renombrado con Acceso Asociativo

#### Búffer de Renombrado



- Permite varias escrituras pendientes a un mismo registro
- Se utiliza el bit último para marcar cual ha sido la más reciente

### Buffer de Renombrado con Acceso Indexado



- Sólo permite una escritura pendiente a un mismo registro
- Se mantiene la escritura más reciente

#### Renombrado

## Ejemplo de Renombrado (I)

	Entrada Válida	Registro Destino	Valor	Valor Válido	Último
0	1	4	40	1	1
1	1	0	0	1	1
2	1	1	10	1	0
3	1	1	15	1	1
4	0				
5	0				
6	0				
7	0				
8	0				
9	0				
10	0				
11	0				
12	0				
13	0				
14	0				

### Situación Inicial:

- Existen dos renombrados de r1 en las entradas 2 y 3 del buffer
- La última está en la entrada 3

#### Renombrado

## Ejemplo de Renombrado (II)

	Entrada Válida	Registro Destino	Valor	Valor Válido	Último			
0	1	4	40	1	1			
1	1	0	0	1	1			
2	1	1	10	1	0			
3	1	1	15	1	1			
2 3 4 5 6 7 8	1	2		0	1			
5	0							
6	0							
7	0							
8	0							
9	0							
LO	0							
11	0							
12	0							
13	0							
14	0							
r0: 0, "válido" r1: 15, "válido" 4								

mul r2, r0, r1 add r3, r1, r2 sub r2, r0, r1

### Ciclo i:

- Se emite la multiplicación
- Se accede a los operandos de la multiplicación, que tienen valores válidos en el buffer de renombrado
- Se renombra r2 (el destino de mul)

#### Renombrado

## Ejemplo de Renombrado (III)

	Entrada Válida	Registro Destino	Valor	Valor Válido	Último
0	1	4	40	1	1
1	1	0	0	1	1
2	1	1	10	1	0
2 3 4 5 6 7	1	1	15	1	1
4	1	2		0	1
5	1	3		0	1
6	0				
7	0				
8	0				
9	0				
10	0				
11	0				
12	0				
13	0				
14	0				
r1	. 15, "vá	lido"	r2: "n	J o válid	o" 5

mul r2, r0, r1 add r3, r1, r2 sub r2, r0, r1

### Ciclo i + 1:

- Se emite la suma
- Se accede a sus operandos, pero r2 no estará preparado hasta que termine la multiplicación
- Se renombra r3 (destino de add)

#### Renombrado

## Ejemplo de Renombrado (IV)

	Entrada Válida	Registro Destino	Valor	Valor Válido	Último
0	1	4	40	1	1
1	1	0	0	1	1
2	1	1	10	1	0
1 2 3 4 5	1	1	15	1	1
4	1	2		0	0
	1	3		0	1
6	1	2		0	1
7 8	0				
	0				
9	0				
10	0				
11	0				
12	0				
13	0				
14	0				
rC		ido"	r1: 15	, "válid	o" 6

mul r2, r0, r1 add r3, r1, r2 sub r2, r0, r1

### Ciclo i + 2:

- Se emite la resta
- Se accede a sus operandos
- Se vuelve a renombrar r2 (destino de sub)

#### Renombrado

## Ejemplo de Renombrado (V)

	Entrada Válida	Registro Destino	Valor	Valor Válido	Último
0	1	4	40	1	1
1	1	0	0	1	1
2	1	1	10	1	0
3	1	1	15	1	1
4	1	2	0	1	0
5	1	3		0	1
6	1	2		0	1
7	0				
8	0				
9	0				
10	0				
11	0				
12	0	·			
13	0				
14	0				

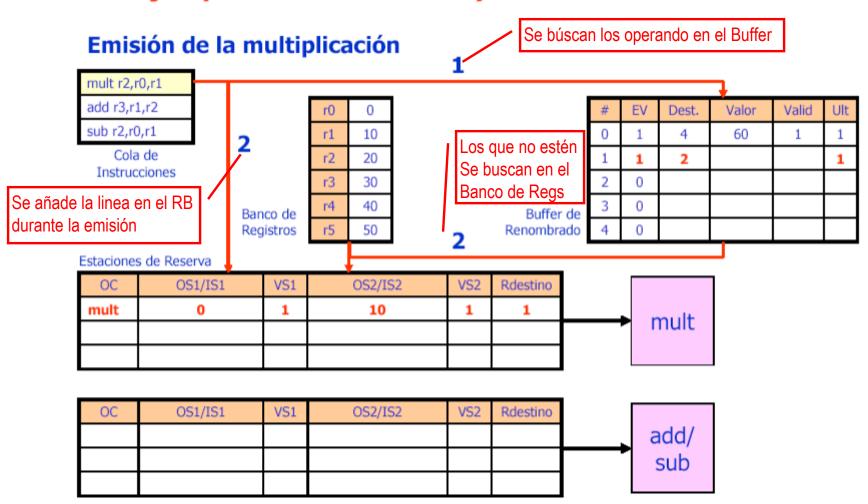


### Ciclo i + 5:

- Termina la multiplicación
- Se actualiza el resultado en el buffer de renombrado
- Ya se puede ejecutar la suma con el valor de r2 de la entrada 4
- Cuando termine la resta escribirá otro valor para r2 en la entrada 6
- Sólo se escribirá en el banco de registros el último valor de r2

#### Renombrado

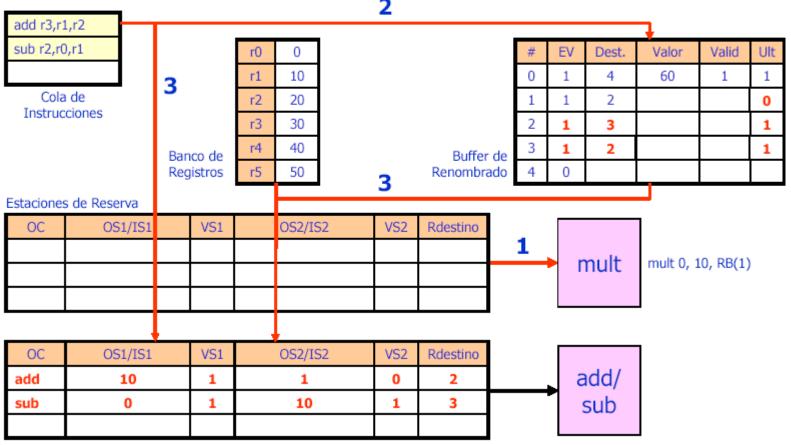
### Ejemplo de Renombrado y Estaciones de Reserva



Renombrado

### Ejemplo de Renombrado y Estaciones de Reserva (II)





#### Renombrado

### Ejemplo de Renombrado y Estaciones de Reserva (III)

### Envío de la resta



		-
	r1	10
	r2	20
	r3	30
Banco de	r4	40
Registros	r5	50

#	EV	Dest.	Valor	Valid	Ult
0	1	4	60	1	1
1	1	2			0
2	1	3			1
3	1	2			1
4	0				

#### Estaciones de Reserva

OC	OS1/IS1	VS1	OS2/IS2	VS2	Rdestino		
						mult	mult 0, 10, RB(1)
						muit	muit 0, 10, Kb(1)

Buffer de

Renombrado

OC	OS1/IS1	VS1	OS2/IS2	VS2	Rdestino			
add	10	1	1	0	2	1	add/	cub 0 10 PP/3\
							sub	sub 0, 10, RB(3)

#### Renombrado

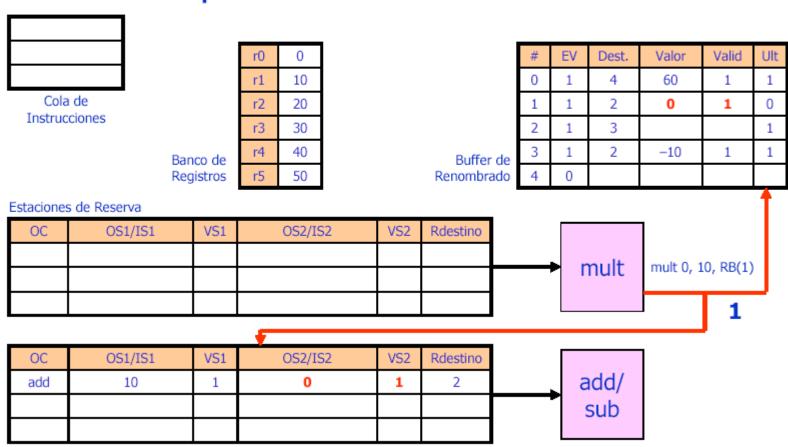
## Ejemplo de Renombrado y Estaciones de Reserva (IV)

#### Termina la resta 0 ΕV Dest. Valor Valid Ult 10 60 Cola de 20 0 Instrucciones r3 30 3 40 -10r4 Banco de Buffer de 50 Renombrado Registros Estaciones de Reserva OC OS1/IS1 VS1 OS2/IS2 VS2 Rdestino mult mult 0, 10, RB(1) OS1/IS1 OC VS1 OS2/IS2 VS<sub>2</sub> Rdestino add/ add 0 10 1 sub 0, 10, RB(3) sub

#### Renombrado

### Ejemplo de Renombrado y Estaciones de Reserva (V)

### Termina la multiplicación

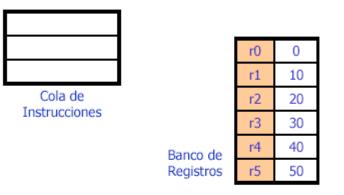


#### Renombrado

## Ejemplo de Renombrado y Estaciones de Reserva (VI)

Renombrado

#### Envío de la suma



	#	EV	Dest.	Valor	Valid	Ult
	0	1	4	60	1	1
	1	1	2	0	1	0
	2	1	3			1
Buffer de	3	1	2	-10	1	1
nombrado	4	0				

#### Estaciones de Reserva

OC	OS1/IS1	VS1	OS2/IS2	VS2	Rdestino	
						mult
						 muit

OC	OS1/IS1	VS1	OS2/IS2	VS2	Rdestino			]
						1	add/	add 10, 0, BB(2)
							sub	add 10, 0, RB(2)

#### Renombrado

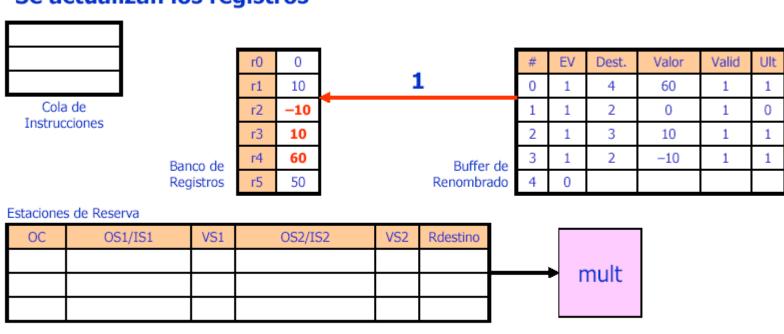
### Ejemplo de Renombrado y Estaciones de Reserva (VII)

#### Termina la suma Dest. Valor Valid 10 60 Cola de 20 r2 2 0 Instrucciones 30 r3 3 10 40 3 r4 -10Buffer de Banco de 50 Renombrado Registros Estaciones de Reserva OS1/IS1 VS1 OS2/IS2 VS<sub>2</sub> Rdestino OC mult OC OS1/IS1 VS1 OS2/IS2 VS2 Rdestino add/ add 10, 0, RB(2) sub

#### Renombrado

### Ejemplo de Renombrado y Estaciones de Reserva (VII)

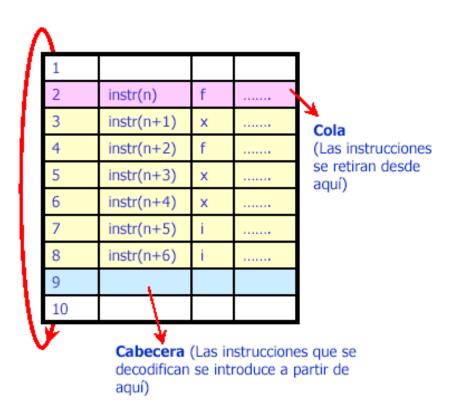
### Se actualizan los registros



OC	OS1/IS1	VS1	OS2/IS2	VS2	Rdestino	
						add/
						sub

Renombrado

Reorden



La gestión de interrupciones y la ejecución especulativa se realizan fácilmente mediante el ROB

- El puntero de cabecera apunta a la siguiente posición libre y el puntero de cola a la siguiente instrucción a retirar.
- Las instrucciones se introducen en el ROB en orden de programa estricto y pueden estar marcadas como emitidas (issued, i), en ejecución (x), o finalizada su ejecución (f)
- Las instrucciones sólo se pueden retirar (se produce la finalización con la escritura en los registros de la arquitectura) si han finalizado, y todas las que les preceden también.
- La consistencia se mantiene porque sólo las instrucciones que se retiran del ROB se completan (escriben en los registros de la arquitectura) y se retiran en el orden estricto de programa.

# Unidad 2. Superescalares

## 2.3. Gestión de riesgos (dependencia datos)

Renombrado

Reorden

## Ejemplo de Uso del Buffer de Reorden (I)

**I1:** mult r1, r2, r3

**I2:** st r1, 0x1ca

**I3:** add r1, r4, r3

**I4:** xor r1, r1, r3

Dependencias:

**RAW:** (I1,I2), (I3,I4)

**WAR:** (I2,I3), (I2,I4)

**WAW:** (I1,I3), (I1,I4), (I3,I4)

I1: Se puede empezar a ejecutar inmediatamente (se suponen disponibles r2 y r3)

12: Se envía a la unidad de almacenamiento hasta que esté disponible r1

I3: Se puede empezar a ejecutar inmediatamente (se suponen disponibles r4 y r3)

I4: Se envía a la estación de reserva de la ALU para esperar a r1

Estación de Reserva (Unidad de Almacenamiento)

codop	dirección	op1	ok1
st	0x1ca	3	0

Estación de Reserva (ALU)

codop	dest	op1	ok1	op2	ok2
xor	6	5	0	[r3]	1



Renombrado

Reorden

## Ejemplo de Uso del Buffer de Reorden (II)

#### Ciclo 7

#	codop	Nº Inst.	Reg. Dest.	Unidad	Resultado	ok	marca	'ready'
3	mult	7	r1	int_mult	-	0	Х	12
4	st	8	1	store	-	0	_	12
5	add	9	r1	int_add	1	0	X	9
6	xor	10	r1	int_alu	-	0	i	-

#### Ciclo 9 No se puede retirar add aunque haya finalizado su ejecución

#	codop	Nº Inst.	Reg.Dest.	Unidad	Resultado	ok	marca	'ready'
3	mult	7	r1	int_mult	-	0	х	12
4	st	8	-	store	1	0	_	12
5	add	9	r1	int_add	17	1	f	9
6	xor	10	r1	int_alu	1	0	X	10

### **Ciclo 10** Termina **xor**, pero todavía no se puede retirar

#	codop	Nº Inst.	Reg.Dest.	Unidad	Resultado	ok	marca	'ready'
3	mult	7	r1	int_mult	-	0	х	12
4	st	8	-	store	1	0	i	12
5	add	9	r1	int_add	17	1	f	9
6	xor	10	r1	int_alu	21	1	f	10

Renombrado

Reorden

## Ejemplo de Uso del Buffer de Reorden (III)

#### Ciclo 12

#	codop	Nº Inst.	Reg. Dest.	Unidad	Resultado	ok	marca	'ready'
3	mult	7	r1	int_mult	33	1	f	12
4	st	8	-	store	-	1	f	12
5	add	9	r1	int_add	17	1	f	9
6	xor	10	r1	int_alu	21	1	f	10

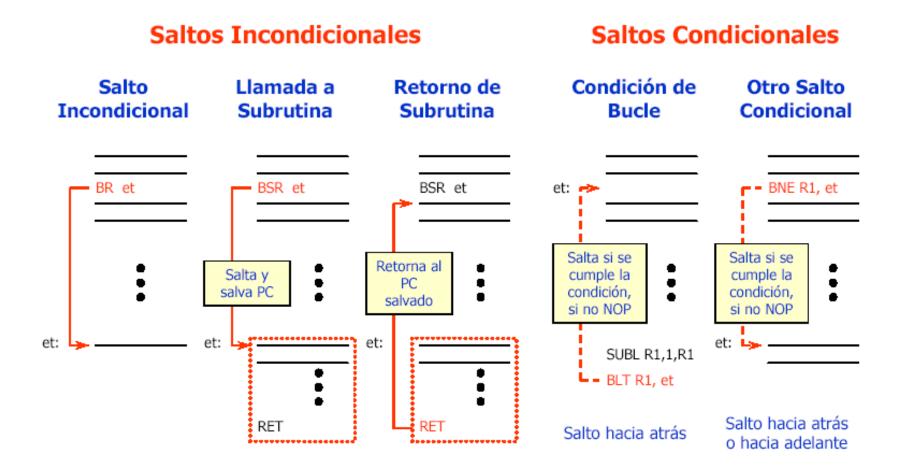
#### Ciclo 13

#	codop	Nº Inst.	Reg. Dest.	Unidad	Resultado	ok	marca	'ready'
5	add	9	r1	int_add	17	1	f	9
6	xor	10	r1	int_alu	21	1	f	10

- Se ha supuesto que se pueden retirar dos instrucciones por ciclo.
- Tras finalizar las instrucciones **mult** y **st** en el ciclo 12, se retirarán en el ciclo 13.
- Después, en el ciclo 14 se retirarán las instrucciones add y xor.

Riesgos

### Clasificación de los Saltos



### Riesgos

El efecto de los saltos en los procesadores superescalares es más **pernicioso** ya que, al emitirse varias instrucciones por ciclo, *prácticamente en cada ciclo* puede haber una instrucción de salto.

El **salto retardado no tiene mucho interés** porque la unidad de emisión decide las instrucciones que pasan a ejecutarse teniendo en cuenta las dependencias.

#### Detección de la Instrucción de Salto

Cuanto antes se detecte que una instrucción es de salto menor será la posible penalización. Los saltos se detectan usualmente en la fase de decodificación.

#### Gestión de los Saltos Condicionales no Resueltos

Si en el momento en que la instrucción de salto evalúa la condición de salto ésta no se haya disponible se dice que *el salto o la condición no se ha resuelto*. Para resolver este problema se suele utilizar el *procesamiento especulativo del salto*.

### Acceso a las Instrucciones destino del Salto

Hay que determinar la forma de acceder a la secuencia a la que se produce el salto

### La implementación física de las Unidades de Salto

### Riesgos

- Detección
  - En etapa de decodificación



- Detección temprana ('early branch detection')
  - Detección paralela a decodificación



Detección anticipada ('look-ahead branch detection')



Detección integrada en captación



Riesgos **Gestión** 

Uso de los ciclos que siguen a la inst. de salto condicional	Salto Retardado	Se utilizan los ciclos que siguen a la captación de una instrucción de salto para insertar instrucciones que deben ejecutarse independientemente del resultado del salto (Primeras arquitecturas RISC y posteriores)
Gestión de Saltos	Bloqueo del Procesamiento del Salto	Se bloquea la instrucción de salto hasta que la condición esté disponible (68020, 68030, 80386)
Condicionales no Resueltos  (Una condición de salto no se puede comprobar si no se ha	Procesamiento Especulativo de los Saltos	La ejecución prosigue por el camino más probable (se especula sobre las instrucciones que se ejecutarán). Si se ha errado en la predicción hay que recuperar el camino correcto.  (Típica en los procesadores superescalares actuales)
terminado de evaluar)	Múltiples Caminos	Se ejecutan los dos caminos posibles después de un salto hasta que la condición de salto se evalúa. En ese momento se cancela el camino incorrecto.  (Máquinas VLIW experimentales: Trace/500 , URPR-2)
Evitar saltos condicionales	Ejecución Vigilada ( <i>Guarded Exec.</i> )	Se evitan los saltos condicionales incluyendo en la arquitectura instrucciones con operaciones condicionales  (IBM VLIW, Cydra-5, Pentium, HP PA, Dec Alpha)

Riesgos

Gestión

## Esquemas de Predicción de Salto

### Predicción Fija

Se toma siempre la misma decisión: el salto siempre se realiza, 'taken', o no, 'not taken'

### Predicción Verdadera

La decisión de si se realiza o no se realiza el salto se toma mediante:

- Predicción Estática:
  - Según los atributos de la instrucción de salto (el código de operación, el desplazamiento, la decisión del compilador)
- Predicción Dinámica:

Según el resultado de ejecuciones pasadas de la instrucción (historia de la instrucción de salto)

Riesgos **Gestión** 



Riesgos

Gestión

# • Predicción fija

Aproximación 'Siempre No Tomado'	
- Toda condición de salto no resuelta se predice que no da lugar a un salto	SuperSparc (1992) (TP: 1; NTP: 0)
<ul> <li>Se continúa la ejecución por donde iba aunque se puede adelantar algo el procesamiento de la secuencia de salto (calculo de la dirección de salto, BTA)</li> </ul>	Alpha21064
- Cuando se evalúa la condición se comprueba si la predicción era buena.	Power I (1990) (TP: 3; NTP: 0)
<ul> <li>Si la predicción era buena el procesamiento continúa y se borra la BTA, y si era mala se abandona el procesamiento de la secuencia predicha (no se considera su efecto) y se captan instrucciones a partir de la BTA</li> </ul>	Power 2 (1993) (TP: 1; NTP: 0)
* Es más fácil de implementar que la aproximación de 'Siempre Tomado'	
Aproximación 'Siempre Tomado'	
- Toda condición de salto no resuelta se predice que da lugar a un salto	
<ul> <li>En previsión de error de predicción se salva el estado de procesamiento actual (PC) y se empieza la ejecución a partir de la dirección de salto.</li> </ul>	MC68040 (1999)
- Cuando se evalúa la condición de salto se comprueba si la predicción era buena	(TP: 1; NTP: 2)
<ul> <li>Si la predicción es correcta se continúa, y si es errónea se recupera el estado almacenado y no se considera el procesamiento de la secuencia errónea</li> </ul>	
* Necesita una implementación más compleja que la aproximación anterior aunque suele proporcionar mejores prestaciones	

Riesgos **Gestión** 

Predicción estática

### Predicción basada en la DIRECCIÓN del salto



La mayoría de saltos hacia atrás corresponden a bucles

La mayoría de saltos hacia delante corresponden a IF-THEN-ELSE

Mal comportamiento en programas con pocos bucles y muchos IF-THEN-ELSE

### **EJEMPLOS**

- Alpha 21064 (1992) (Opción seleccionable)
- PowerPC 601/603 (1993)

Riesgos **Gestión** 

### Predicción estática

### Predicción basada en el Código de Operación

Para ciertos códigos de operación (ciertos saltos condicionales específicos) se predice que el salto se toma, y para otros que el salto no se toma MC88110 (93) PowerPC 603(93)

### Ejemplo: Predicción Estática en el MC88110

Formato	Instrucció	in	Predicción
Formato	Condición Especificada	Bit 21 de la Instr.	Prediccion
	<b>≠</b> 0	1	Tomado
	= 0	0	No Tomado
	> 0	1	Tomado
bcnd ( <i>Branch</i>	< 0	0	No Tomado
Conditional)	≥ 0	1	Tomado
Conditionary	≤ 0	0	No Tomado
	bb1 (Branch on Bit Set)		Tomado
	bb0 (Branch on Bit Clear)		No Tomado

Riesgos **Gestión** 

Predicción estática

Predicción dirigida por el COMPILADOR

Basada en el tipo de CONSTRUCCIÓN

El compilador predice si el salto será tomado o no dependiendo del tipo de construcción de control Basada en PROFILING

El compilador predice en función del comportamiento de esa instrucción en ejecuciones previas del programa Especificado por el PROGRAMADOR

El programador indica al compilador si el salto debe ser tomado o no (mediante directivas específicas)

- Se añade un Bit de Predicción al opcode de la instrucción
- · El compilador activa o desactiva este bit para indicar su predicción

#### **EJEMPLOS**

- AT&T 9210 Hobbit (1993)
- PowerPC 601/603 (1993)
- PA 8000 (1996)

Riesgos **Gestión** 

### Predicción dinámica

- La predicción para cada instrucción de salto puede cambiar cada vez que se va a ejecutar ésta según la historia previa de saltos tomados/no-tomados para dicha instrucción.
- El presupuesto básico de la predicción dinámica es que es más probable que el resultado de una instrucción de salto sea similar al que se tuvo en la última (o en las n últimas ejecuciones)
- Presenta mejores prestaciones de predicción, aunque su implementación es más costosa

### Predicción Dinámica Implícita

No hay bits de historia propiamente dichos sino que se almacena la dirección de la instrucción que se ejecutó después de la instrucción de salto en cuestión

### Predicción Dinámica Explícita

Para cada instrucción de salto existen unos bits específicos que codifican la información de historia de dicha instrucción de salto

Riesgos **Gestión** 

### Predicción dinámica explícita

#### Predicción con 1 bit de historia

La designación del estado, Tomado (1) o No Tomado (0), indica lo que se predice, y las flechas indican las transiciones de estado según lo que se produce al ejecutarse la instrucción (T o NT)

#### Predicción con 2 bits de historia

Existen cuatro posibles estados. Dos para predecir Tomado y otros dos para No Tomado

La primera vez que se ejecuta un salto se inicializa el estado con predicción estática, por ejemplo 11

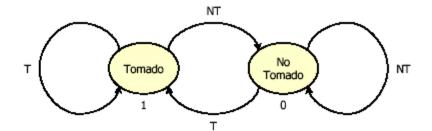
Las flechas indican las transiciones de estado según lo que se produce al ejecutarse la instrucción (T o NT)

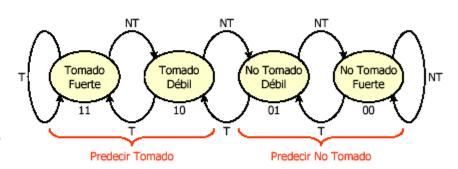
#### Predicción con 3 bits de historia

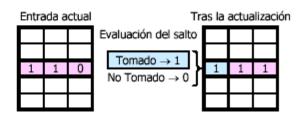
Cada entrada guarda las últimas ejecuciones del salto

Se predice según el bit mayoritario (por ejemplo, si hay mayoría de unos en una entrada se predice salto)

La actualización se realiza en modo FIFO, los bits se desplazan, introduciéndose un 0 o un 1 según el resultado final de la instrucción de salto







Riesgos **Gestión** 

Predicción dinámica explícita

# Implementación de los Bits de Historia

• En la Cache de Instrucciones	Alpha 21064 (92) (2k x 1 bit) Alpha 21064A (94) (4k x 2 bits) UltraSparc (92) (2k x 2 bits)
• En una Tabla de Historia de Salto (BHT)	PA8000(96) (253 x 3bits) PowerPC 620(95) (2K x 2bits) R10000 (96) (512 x 2bits)
<ul> <li>En una Cache para las Instrucciones a las que se produce el Salto (BTIC) o,</li> <li>En una Cache para las Direcciones a las que se produce el Salto (BTAC)</li> </ul>	Pentium (94) (256 x 2 bits) (BTAC) MC 68069 (93) (256 x 2bits) (BTAC)

Riesgos

## Predicción dinámica explícita

### 1) Branch Target Buffer (BTB): bits acoplados

Gestión

#### La BTB almacena

- La dirección destino de los últimos saltos tomados
- · Los bits de predicción de ese salto

#### Actualización de la BTB

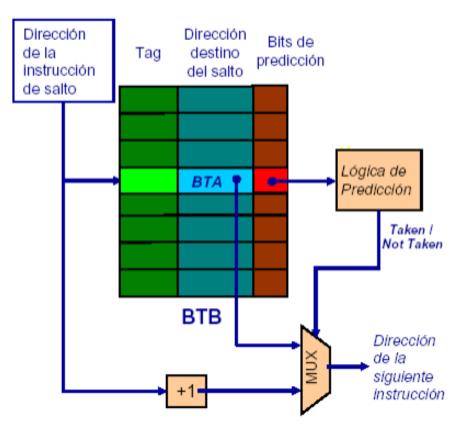
Los campos de la BTB se actualizan después de ejecutar el salto, cuando se conoce:

- Si el salto fue tomado o no ⇒ Actualizar bits de predicción
- La dirección destino del salto ⇒ Actualizar BTA

### Predicción Implícita (sin bits de predicción)

Aplicable con un sólo bit de predicción

- · Sí la instrucción de salto está en la BTB
  - ⇒ El salto se predice como tomado
- Sí la instrucción de salto no está en la BTB
  - ⇒ El salto se predice como no tomado



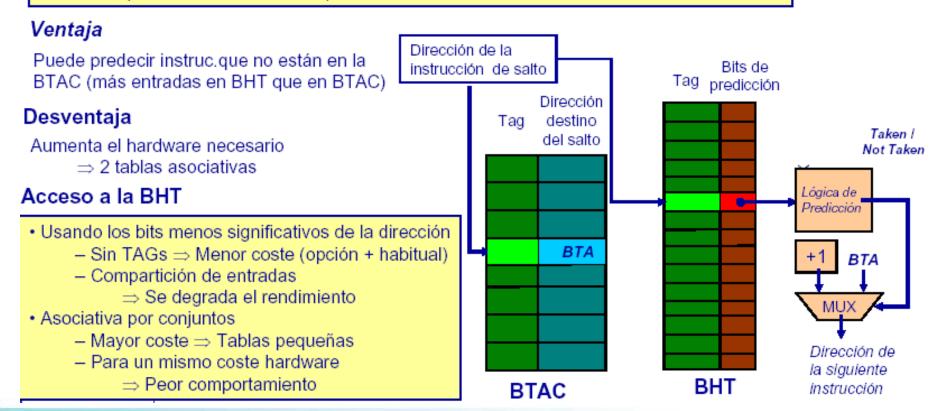
DESVENTAJA: Sólo se pueden predecir aquellas instrucciones de salto que están en la BTB

Riesgos **Gestión** 

Predicción dinámica explícita
 2) Tabla de historia de saltos (BHT): bits desacoplados

#### Existen dos tablas distintas:

- La BTAC, que almacena la dirección destino de los últimos saltos tomados
- La BHT, que almacena los bits de predicción de todos las instrucciones de salto condicional.



Riesgos **Gestión** 

- Predicción dinámica explícita
  - 3) Bits de predicción en la l-cache

#### Funcionamiento

Cuando se capta la instrucción de la cache Si se trata de una instrucción de salto condicional

- Se accede en paralelo a los bits de predicción
- Si el salto se predice como tomado se accede a la instrucción destino del salto

#### Acceso a la instrucción destino del salto

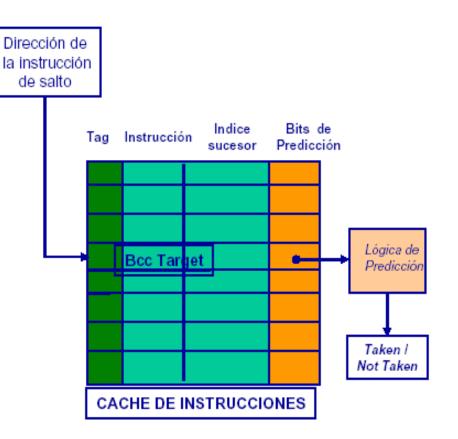
- BTB independiente
- Añadir índice sucesor a la l-cache

#### Alternativas de diseño

- Bits de predicción por cada instrucción de la cache
- Bits de predicción por cada línea de cache

### Ventajas

- Puede predecir instrucciones que no están en la BTB
- No añade una cantidad extra de hardware excesiva



Riesgos

Gestión

## Extensión del Procesamiento Especulativo

- Tras la predicción, el procesador continúa ejecutando instrucciones especulativamente hasta que se resuelve la condición.
- El intervalo de tiempo entre el comienzo de la ejecución especulativa y la resolución de la condición puede variar considerablemente y ser bastante largo.
- En los procesadores superescalares, que pueden emitir varias instrucciones por ciclo, pueden aparecer más instrucciones de salto condicional no resueltas durante la ejecución especulativa.
- Si el número de instrucciones que se ejecutan especulativamente es muy elevado y la predicción es incorrecta, la penalización es mayor.

Así, cuanto mejor es el esquema de predicción mayor puede ser el número de instrucciones ejecutadas especulativamente.

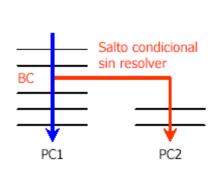


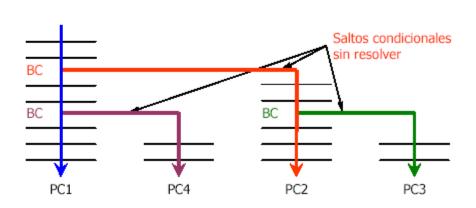
- Nivel de Especulación: Número de Instrucciones de Salto Condicional sucesivas que pueden ejecutarse especulativamente (si se permiten varias, hay que guardar varios estados de ejecución). Ejemplos: Alpha21064, PowerPC 603 (1); Power 2 (2); PowerPC 620 (4); Alpha 21164 (6)
- Grado de Especulación: Hasta qué etapa se ejecutan las instrucciones que siguen en un camino especulativo después de un salto. Ejemplos: Power 1 (Captación); PowerPC 601 (Captación, Decodificación, Envío); PowerPC 603 (Todas menos la finalización)

Riesgos **Gestión** 

Ramificación multicamino

- Se siguen las dos secuencias de instrucciones que aparecen a partir de una instrucción de salto (la correspondiente al salto efectuado y al salto no efectuado). Una vez resuelto el salto la secuencia incorrecta se abandona
- Para implementar esta técnica hacen falta varios contadores de programa
- Se necesitan gran cantidad de recursos hardware y el proceso para descartar las instrucciones que se han ejecutado incorrectamente es complejo





Riesgos

Gestión

### Ejecución condicional

### Instrucciones de Ejecución Condicional (Guarded Execution)

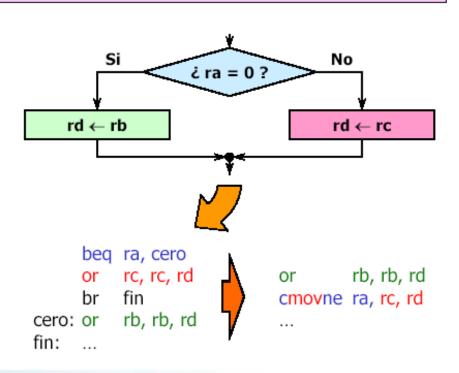
- Se pretende reducir el número de instrucciones de salto incluyendo en el repertorio máquina instrucciones con operaciones condicionales ('conditional operate instructions' o 'guarded instructions')
- Estas instrucciones tienen dos partes: la condición (denominada guardia) y la parte de operación

### Ejemplo: cmovxx de Alpha

#### cmovxx ra.rq, rb.rq, rc.wq

- xx es una condición
- ra.rq, rb.rq enteros de 64 bits en registros ra y rb
- rc.wq entero de 64 bits en rc para escritura
- El registro ra se comprueba en relación a la condición xx y si se verifica la condición rb se copia en rc.

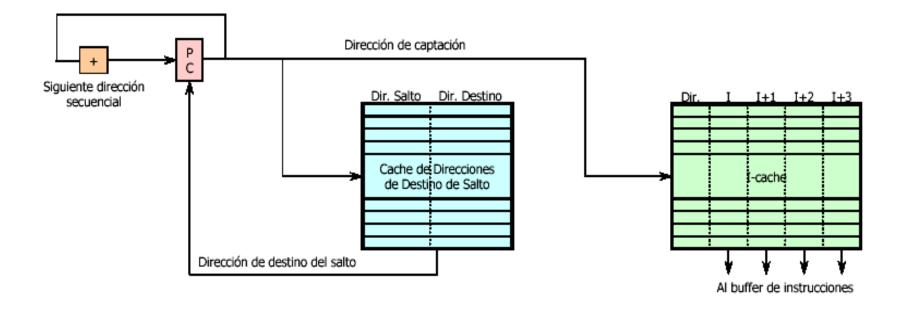
Sparc V9, HP PA, y Pentium ofrecen también estas instrucciones.



Riesgos Gestión **Estructuras** 

### Esquema de cache de direcciones de destino de salto (BTAC)

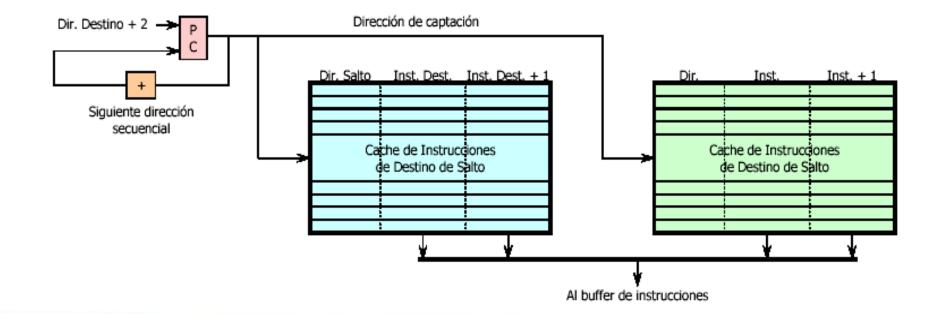
- Se añade una cache que contiene las direcciones de las instrucciones destino de los saltos, junto con las direcciones de las instrucciones de salto.
- Se leen las direcciones al mismo tiempo que se captan las instrucciones de salto.



Riesgos Gestión **Estructuras** 

### Esquema de cache de instrucciones de destino de salto (BTIC)

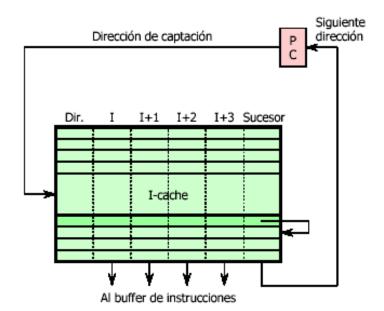
- Se añade una cache que contiene las instrucciones siguientes a la dirección de destino de los saltos, junto con las direcciones de las instrucciones de salto.
- Sólo tiene sentido si la cache de instrucciones tiene una latencia muy alta.
- Mientras se procesan estas instrucciones se calcula la dirección de las siguientes.



Riesgos Gestión **Estructuras** 

### Esquema de índice sucesor en la cache de instrucciones

 La cache de instrucciones contiene un índice sucesor que apunta a la siguiente línea de la cache de instrucciones que hay que captar (la siguiente, o la que se predice que se debe captar si hay una instrucción de salto condicional en esa línea).



### Ejemplos y evolución de los esquemas de Acceso

Calcular/captar	BTIC	BTAC	Índice sucesor
i486 (1989)	$\rightarrow$	Pentium (1993)	
MC68040 (1990)	$\longrightarrow$	MC 68060 (1993)	
	Am 29000 (1988)	$\rightarrow$	Am 29000 superscalar (1995)
Sparc CYC 600 (1992) SuperSparc (1992)		$\longrightarrow$	UltraSparc (1995)
R4000 (1992) R10000 (19996)		$\longrightarrow$	R8000 (1994)
PowerPC 601 (1993) PowerPC 603 (1993)	$\longrightarrow$	PowerPC 604 (1995) PowerPC 620 (1996)	

# Unidad 2. Superescalares Problemas

### Ejecución fuera de orden - Problemas

Considérese el siguiente fragmento de código:

```
(1) subd f2, f2, f1
(2) addd f4, f2, f3
(3) subd f5, f2, f3
(4) multd f6, f2, f5
(5) subd f1, f2, f5
(6) subd f7, f4, f6
```

Se va a ejecutar en un procesador superescalar capaz de captar, decodificar y emitir mediante ventana centralizada 3 instrucciones por ciclo, pudiendo retirar hasta 2 instrucciones por ciclo. Cuenta con 2 UF de suma/resta (2 ciclos) y 1 UF de multiplicación (4 ciclos). Si se dispone de un ROB que permite la ejecución desordenada. Estimar los ciclos necesarios para ejecutar por completo el código anterior en los siguientes casos:

- a) Emisión ordenada (no alineada)
- b) Emisión desordenada (alineada)
- c) Emisión desordenada (no alineada)

# Unidad 2. Superescalares Problemas

### Ejecución fuera de orden - Problemas

- 2. Se dispone de un procesador superescalar con la siguiente configuración:
  - una estación de reserva RS1 para las sumas y restas
  - una estación de reserva RS2 para las multiplicaciones y divisiones
  - un buffer de reorden ROB
  - dos unidades de ejecución de sumas/restas con una latencia de 2 ciclos
  - una unidad de ejecución de multiplicaciones con una latencia de 5 ciclos
  - una unidad de ejecución de divisiones con una latencia de 40 ciclos

El procesador es capaz de emitir dos instrucciones por ciclo, las estaciones de reserva pueden realizar envíos no alineados y desordenados a las unidades de ejecución, y se pueden retirar hasta dos instrucciones por ciclo.

Si en la cola de instrucciones se encuentran las siguientes instrucciones:

```
(1) addd f3, f1, f2; f3 = f1 + f2

(2) addd f2, f3, f2; f2 = f3 + f2

(3) multd f4, f3, f2; f4 = f3 * f2

(4) divd f5, f2, f1; f5 = f2 / f1

(5) subd f2, f3, f1; f2 = f3 - f1
```

y los registros f1 y f2 tienen inicialmente los valores 10 y 5, ¿qué valores y en qué ciclos se escribirá en los registros de la arquitectura?

# Unidad 2. Superescalares Problemas Saltos

#### **Saltos - Problemas**

1. En un programa, una instrucción de salto condicional a una dirección anterior tiene el siguiente comportamiento tras 18 ejecuciones ('S': salta, 'N': no salta)

S, S, N, N, N, S, S, N, S, N, S, N, S, S, S, S, N

Indicar la penalización efectiva (porcentaje de casos en que falla la predicción) que se produce si se utiliza:

- a) Predicción fija (siempre Not Taken)
- b) Predicción estática (si salto atrás Taken, si hacia adelante Not Taken)
- c) Predicción dinámica con 2 bits (inicialmente en 11)
- d) Predicción dinámica con 3 bits (inicialmente en 111)

# Unidad 2. Superescalares Problemas Saltos

#### **Saltos - Problemas**

2. Se desea ejecutar el siguiente programa en un procesador superescalar:

```
(1) lw r1, N
(2) add r2, r0, r0
(3) bucle: lw r3, X(r2)
(4) sgt r4, r3, r0
(5) bnz r4, mayor
(6) sub r3, r0, r3
(7) mayor: sw X(r2), r3
(8) add r2, r2, #4
(9) sub r1, r1, #1
(10) bnz r1, bucle
```

Si el procesador usa <u>predictores dinámicos de dos bits</u> que se inicializan con un predictor estático que predice como tomados los saltos hacia atrás y como no tomados los saltos hacia delante, y que la penalización en caso de errar la predicción es de 5 ciclos, ¿qué penalización se obtendrá si X (0) = 1 y X (i+1) = 1 – X (i)?