

# Appunti di Architettura degli Elaboratori

Alessandro Cheli - Prof. Marco Danelutto

A.A 2019-2020



# Indice

<b>I</b>	<b>1</b>
<b>1 Introduzione al Corso</b>	<b>3</b>
<b>2 Rappresentazioni Numeriche e Testuali</b>	<b>7</b>
2.1 Aritmetica Binaria . . . . .	7
2.2 Esadecimale . . . . .	7
2.3 Numeri in virgola mobile . . . . .	8
2.4 Codifica ASCII . . . . .	9
2.5 Porte Logiche e Algebra di Boole . . . . .	9
2.5.1 Algebra di Boole . . . . .	13
2.5.2 Teoremi dell'Algebra Booleana . . . . .	14
2.5.3 Mappe di Karnaugh . . . . .	15
2.5.4 Circuito con più output . . . . .	15
2.5.5 Operatori a più ingressi . . . . .	15
<b>3 Reti Logiche e Combinatorie</b>	<b>17</b>
3.1 Alcuni componenti standard . . . . .	17
3.2 Reti sequenziali e Automi . . . . .	25
3.2.1 Realizzazione di una rete sequenziale di un Automa di Mealy con un clock e registri . . . . .	27
<b>4 Reti Sequenziali, Verilog e RTL</b>	<b>29</b>
4.1 Scrivere e compilare System Verilog . . . . .	30
4.2 Esercizi . . . . .	32
4.2.1 Automa che riconosce "abba" . . . . .	32
4.2.2 Automa di Moore per riconoscere "abba" . . . . .	37
4.2.3 Mealy con Delay $\equiv$ Moore . . . . .	41
4.3 Componenti per il Calcolo . . . . .	44
<b>5 Memorie e Parallelismo</b>	<b>49</b>
5.1 Memorie . . . . .	49
5.1.1 Calcolatrice con registri di memoria . . . . .	56
5.2 Parallelismo . . . . .	57
5.2.1 Processore superscalare . . . . .	61
<b>II</b>	<b>65</b>
<b>6 Assembler ARM</b>	<b>67</b>
6.1 Introduzione all'Assembler ARM . . . . .	67
6.2 Istruzioni Assembler ARMv7 . . . . .	68



# Parte I



# Capitolo 1

## Introduzione al Corso

- Logica Booleana
- Aritmetica Binaria
- Reti Logiche
- Microarchitettura e Assembler ARM v7 e v8
- Gestione della memoria
- I/O

**Strumenti Software** A differenza degli A.A passati utilizzeremo Verilog e Assembler ARM. Utilizzeremo **iverilog** come compilatore Verilog e **gtkwave** come tool grafico. Un ambiente di sviluppo Verilog completo che vedremo è **Quartus**. Per la seconda parte del corso, Assembler ARM, useremo la **toolchain GNU**, in particolare:

### Se non hai una macchina ARM:

- cross-compiler per compilare
- QEMU per una macchina virtuale ARM
- gdb per debugging

### Se hai una macchina ARM come un Raspberry Pi:

- Toolchain GNU per compilare
- Cavo Ethernet
- Server SSH sulla macchina ARM per accesso remoto

**Storia degli Elaboratori** Nei corsi di Architettura degli Elaboratori negli anni 80 i processori studiati erano: il 6502 (8 bit, processore del computer Apple II, noto per essere stato costruito nel garage di Steve Jobs e Wozniak), Z80, processore a 16 bit del famoso computer ZX80 e l'Intel 8088. Tali processori raggiungevano al massimo una velocità di clock (detto molto a grandi linee, operazioni al secondo) dell'ordine di meno di una decina di MHz (Mega Hertz, milioni). I processori odierni raggiungono cicli di clock sull'ordine dei GHz (Giga Hertz, miliardi). Nel corso degli anni fino ad oggi, l'evoluzione dei processori ha seguito la **legge di Moore**. La "legge" spiega

Figura 1.1: Sinclair ZX80



che ogni 18 mesi la potenza dei processori in commercio raddoppia, perché la densità dei transistor contenuti all'interno aumenta. Negli ultimi decenni abbiamo miglioramenti architetturali come super pipeline e super scalari, ciò ha permesso di introdurre processori **multicore**, ovvero che contengono più "nuclei" interni (detti core) che elaborano le istruzioni dei processi in esecuzione in parallelo. Ad oggi il numero di core in uno smartphone raggiunge anche gli 8 core, mentre in processori per server sono stati raggiunti numeri di core anche intorno ai 64. I processori odierni utilizzano core a 64 bit, con architettura X86\_64 per Desktop o ARM per dispositivi mobili. Un componente fondamentale dell'evoluzione degli elaboratori è stato anche lo sviluppo dei processori grafici (GPU) con i quali ad oggi è possibile riprodurre grafica su schermo, ambienti tridimensionali molto complessi (videogiochi) o sfruttare la loro capacità di parallelizzazione per l'uso di reti neurali nell'intelligenza artificiale.

Osserveremo i calcolatori a diversi livelli di **astrazione**

I livelli di astrazione sono:

- Applicazioni utente
- Sistema Operativo
- Architettura (ASM, ad es. x86 o ARM)
- Microarchitettura
- Logica
- Circuiti digitali
- Device
- Fisica

Ogni livello si appoggia sul livello inferiore, ovvero è costruito sui componenti offerti dal livello inferiore. Dei principi fondamentali sono: **gerarchi, modularità e regolarità**

La modularità è fondamentale per avere moduli organizzati gerarchicamente, autonomi ed indipendenti.



**Set di Istruzioni** Distinguiamo due set di istruzioni dei processori, **CISC** e **RISC**. Gli acronimi sono rispettivamente **Complex Instruction Set Computer** e **Reduced Instruction Set Computer**, RISC contiene i processori ARM, che studieremo in dettaglio, mentre CISC comprende i processori più comuni nei desktop (X86 e X86\_64)



# Capitolo 2

## Rappresentazioni Numeriche e Testuali

### 2.1 Aritmetica Binaria

I calcolatori utilizzano valori discreti (differenze di potenziale) fra 0 e 1 per rappresentare valori numerici. Viene detta Aritmetica Binaria l'aritmetica con i numeri rappresentati in base 2.

Siamo abituati a ragionare in base 10, ad esempio il numero 413 in base 10 è

$$104 = 10^2 \cdot 1 + 10^1 \cdot 0 + 10^0 \cdot 4$$

Lo stesso numero rappresentato in base 2 (codice binario) è

$$104_{10} = 01101000_2 = 2^7 \cdot 0 + 2^6 \cdot 1 + 2^5 \cdot 1 + 2^4 \cdot 0 + 2^3 \cdot 1 + 2^2 \cdot 0 + 2^1 \cdot 0 + 2^0 \cdot 0$$

Un numero binario di 8 cifre è detto **byte**, un numero di 4 cifre è detto **nibble**. Una **parola** (**word**) è la quantità minima su cui viene rappresentato un intero in un calcolatore. Ad oggi le parole dei calcolatori sono 64 bit, alcuni calcolatori datati hanno parole da 32 bit.

La somma nell'aritmetica binaria è definita normalmente per i numeri positivi. Nei calcolatori i numeri hanno una dimensione finita (numero di bit) che indica il numero di cifre binarie con le quali è possibile rappresentare un numero. I positivi binari rappresentano numeri fino a  $2^N - 1$  dove  $N$  è il numero di cifre.

Per rappresentare i numeri negativi si utilizza il metodo **segno-magnitudo** dove il bit più a sinistra rappresenta il segno (0 se il numero è positivo e 1 se è negativo). Il problema del metodo segno-magnitudo è che non rispetta la somma aritmetica. Può rappresentare numeri da  $[-2^{N-1}, +2^{N-1}]$

Un metodo migliore per rappresentare i numeri negativi è il **complemento a due**. Nel complemento a due la cifra più a sinistra rappresenta sempre  $2^{N-1}$  ma **negativo**. Il resto delle cifre sono positive e vengono sommate alla prima cifra negativa. Questo metodo rispetta la somma aritmetica. Per moltiplicare un numero per  $-1$  si invertono le cifre binarie e si aggiunge 1 al numero. È possibile anche la sottrazione sommando un numero positivo ad uno negativo.

La somma fra due cifre può essere costruita con reti logiche. Il risultato della somma  $A + B = A \oplus B$  (operatore XOR) mentre il riporto della somma  $= A \wedge B$  (operatore AND)

### 2.2 Esadecimale

I numeri esadecimali sono numeri in base 16. Siccome non bastano le cifre decimali per rappresentare i numeri maggiori di 9 si usano le prime lettere dell'alfabeto. Una cifra esadecimale rappresenta un nibble (4 bit).

Figura 2.1: Gate XOR

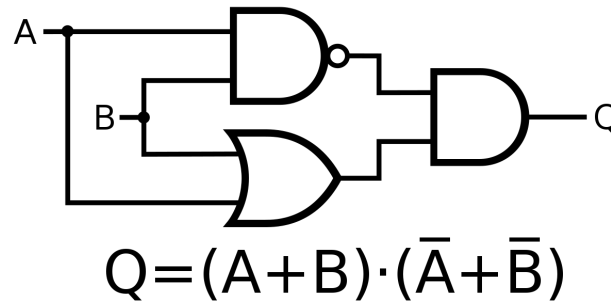
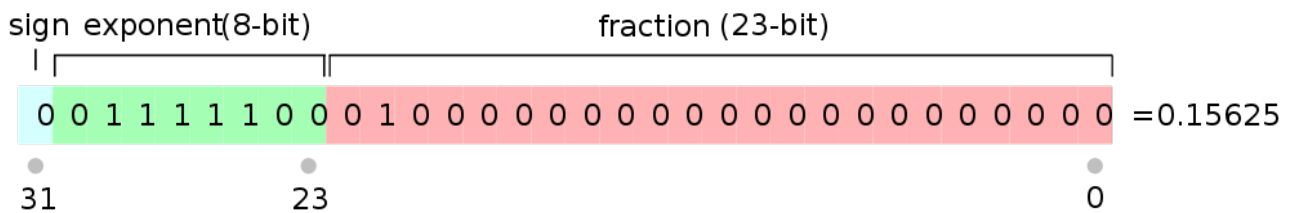


Figura 2.2: Standard IEEE 754 a 32 bit



## 2.3 Numeri in virgola mobile

I numeri in virgola mobile si rappresentano con lo standard IEEE 754 che definisce come si rappresentano i numeri in virgola mobile a singola precisione e doppia precisione (32 e 64 bit)

I bit del numero vengono divisi in 3 parti. Il primo bit denota il segno, la seconda parte rappresenta l'esponente e la terza parte si denota mantissa. L'esponente esprime dove la virgola verrà posizionata, come nella notazione scientifica di una calcolatrice l'esponente rappresenta  $10^n$  dove  $n$  è l'esponente. La mantissa è un numero di base moltiplicato per  $10^0$ , e viene successivamente moltiplicato per l'esponente. L'esponente può essere sia positivo che negativo.

Nello standard a 32 bit la sezione esponente ha 8 bit di lunghezza. Un numero ad 8 bit può rappresentare numeri da 0 a 255, per ottenere gli esponenti negativi nello standard dei numeri a virgola mobile il numero a 8 bit rappresenta invece numeri da -127 a +128

**Somma dei numeri a virgola mobile** Per sommare i numeri a virgola mobile il primo passo è allineare le mantisse, significa osservare gli esponenti e spostarli fino a che le cifre non sono sommabili in colonna. Il secondo passo consiste nel sommare e il terzo passo nel normalizzare la somma. Nei processori la somma floating point viene eseguita in dei moduli appositi che in input ricevono due o più numeri floating point ed eseguono in dei sotto-moduli i tre passaggi della somma in un tempo  $1/3t$  dove  $t$  è il tempo totale per eseguire una somma. I tre passaggi della

Figura 2.3: IEEE 754 a 64 bit

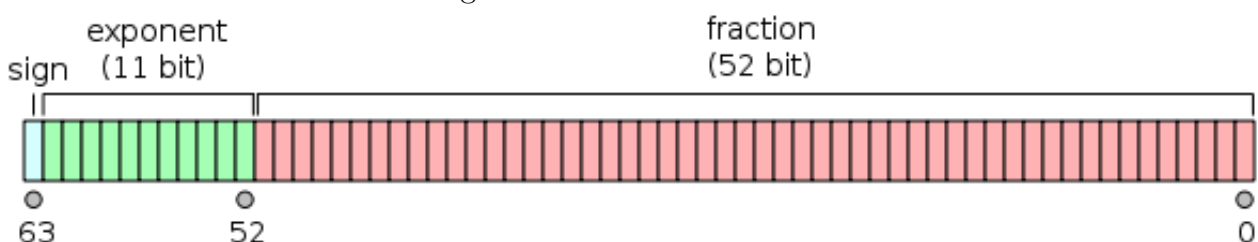


Figura 2.4: Tabella ASCII

# ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

somma possono essere sequenzializzati così che una volta che ogni sotto-modulo ha completato il passo, può ricevere subito l'input successivo (la somma di due numeri FP impiegherà  $t + 1/3$  invece che  $2t$ )

**Estensioni vettoriali** Alcuni processori permettono di eseguire operazioni contemporaneamente su un registro dividendolo in sottoregistri più piccoli.

## 2.4 Codifica ASCII

La codifica ASCII è una tabella di codifica di caratteri testuali con interi da 0 a 255 (8 bit). La codifica ASCII estesa è a 16 bit e comprende diversi caratteri non latini.

## 2.5 Porte Logiche e Algebra di Boole

I circuiti digitali vengono realizzati utilizzando componenti chiamati **porte logiche**. Sono realizzate con componenti fisici come transistor e resistenze, ma nella progettazione dei circuiti digitali le porte logiche vengono schematizzate con i simboli riportati nella Figura 3.1 per semplificare la progettazione **astruendo** il livello di complessità della circuiteria analogica. Solamente con la porta NAND si possono realizzare tutte le altre porte (NAND è funzionalmente completo), ma le porte in generale si costruiscono singolarmente con componenti appositi. Esse implementano la **logica booleana** che conseguentemente permette di realizzare operazioni di **aritmetica binaria** per costruire unità di calcolo in componenti elettronici e processori.

I componenti elettronici molto piccoli sono sensibili al **rumore**, per ovviare al problema i valori discreti (0 e 1) nei circuiti digitali non seguono un cambiamento istantaneo di differenza di potenziale (voltage), ma ammettono un margine per ridurre i problemi causati dal rumore.

I componenti (transistor) con cui si costruiscono porte logiche e circuiti sono realizzati con materiali semiconduttori, che possono essere di diversi tipi. Vedremo il tipo NMOS. Un transistor è composto da materiali come gallio e silicio.

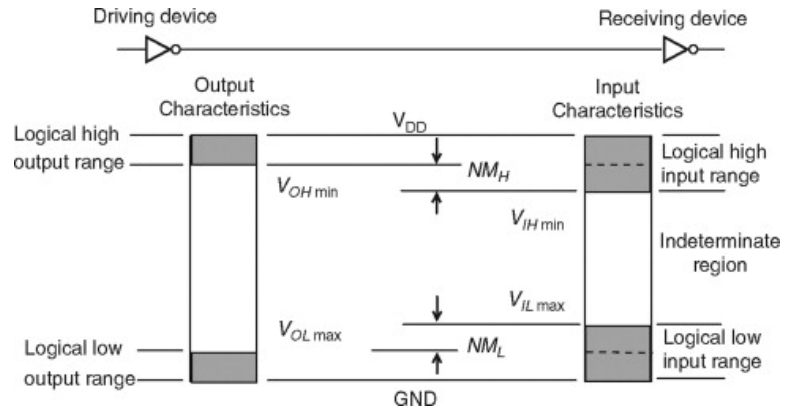


Figura 2.5: Margine di rumore nei circuiti digitali

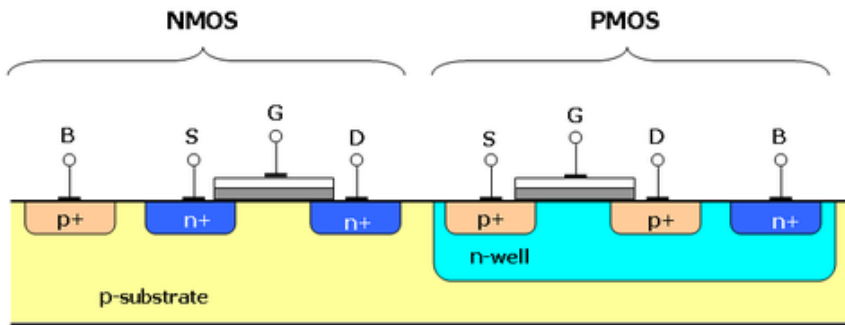


Figura 2.6: Transistor NMOS

Figura 2.7: Tabella delle porte logiche comuni

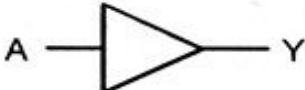
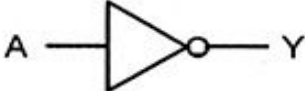






Logic function	Logic symbol	Truth table	Boolean expression															
Buffer		<table border="1" data-bbox="904 412 1067 546"><tr><th>A</th><th>Y</th></tr><tr><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td></tr></table>	A	Y	0	0	1	1	$Y = A$									
A	Y																	
0	0																	
1	1																	
Inverter (NOT gate)		<table border="1" data-bbox="904 591 1067 725"><tr><th>A</th><th>Y</th></tr><tr><td>0</td><td>1</td></tr><tr><td>1</td><td>0</td></tr></table>	A	Y	0	1	1	0	$Y = \bar{A}$									
A	Y																	
0	1																	
1	0																	
2-input AND gate		<table border="1" data-bbox="873 748 1099 949"><tr><th>A</th><th>B</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	A	B	Y	0	0	0	0	1	0	1	0	0	1	1	1	$Y = A \bullet B$
A	B	Y																
0	0	0																
0	1	0																
1	0	0																
1	1	1																
2-input NAND gate		<table border="1" data-bbox="873 972 1099 1173"><tr><th>A</th><th>B</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>0</td></tr></table>	A	B	Y	0	0	1	0	1	1	1	0	1	1	1	0	$Y = \overline{A \bullet B}$
A	B	Y																
0	0	1																
0	1	1																
1	0	1																
1	1	0																
2-input OR gate		<table border="1" data-bbox="873 1196 1099 1397"><tr><th>A</th><th>B</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	A	B	Y	0	0	0	0	1	1	1	0	1	1	1	1	$Y = A + B$
A	B	Y																
0	0	0																
0	1	1																
1	0	1																
1	1	1																
2-input NOR gate		<table border="1" data-bbox="873 1420 1099 1621"><tr><th>A</th><th>B</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td></tr></table>	A	B	Y	0	0	1	0	1	0	1	0	0	1	1	0	$Y = \overline{A + B}$
A	B	Y																
0	0	1																
0	1	0																
1	0	0																
1	1	0																
2-input EX-OR gate		<table border="1" data-bbox="873 1644 1099 1845"><tr><th>A</th><th>B</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>0</td></tr></table>	A	B	Y	0	0	0	0	1	1	1	0	1	1	1	0	$Y = A \oplus B$
A	B	Y																
0	0	0																
0	1	1																
1	0	1																
1	1	0																
2-input EX-NOR gate		<table border="1" data-bbox="873 1868 1099 2069"><tr><th>A</th><th>B</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	A	B	Y	0	0	1	0	1	0	1	0	0	1	1	1	$Y = \overline{A \oplus B}$
A	B	Y																
0	0	1																
0	1	0																
1	0	0																
1	1	1																

Figura 2.8: Porta NOT con transistor PMOS e NMOS

## Inverter (Not Gate)

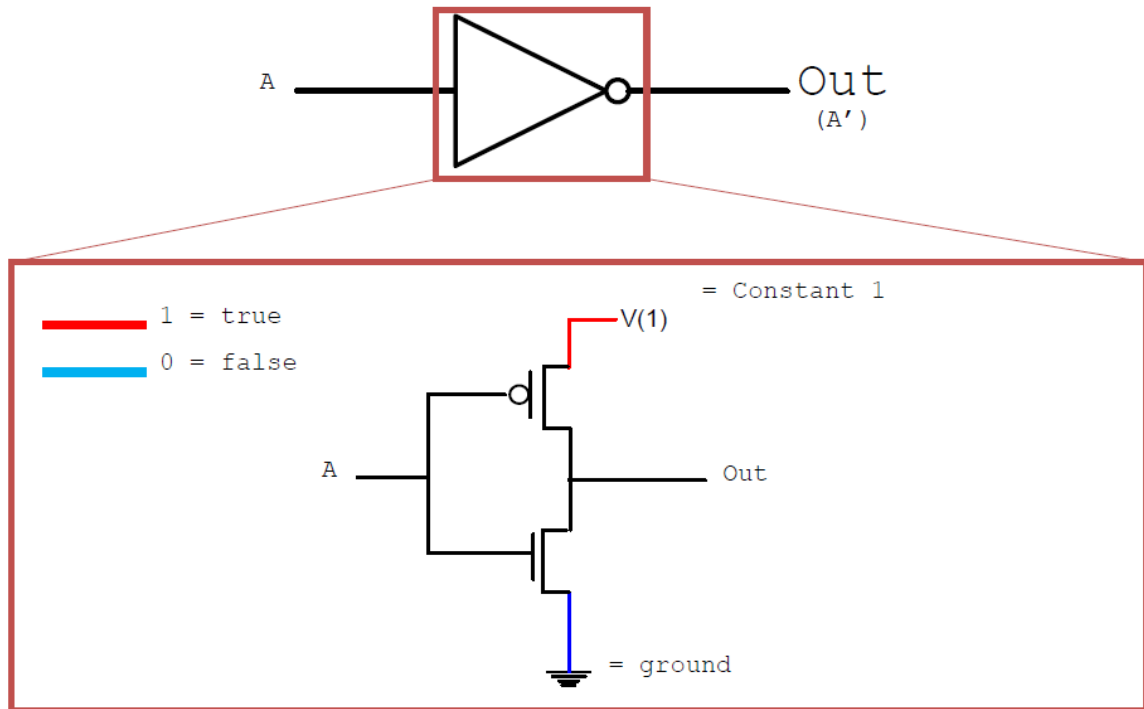
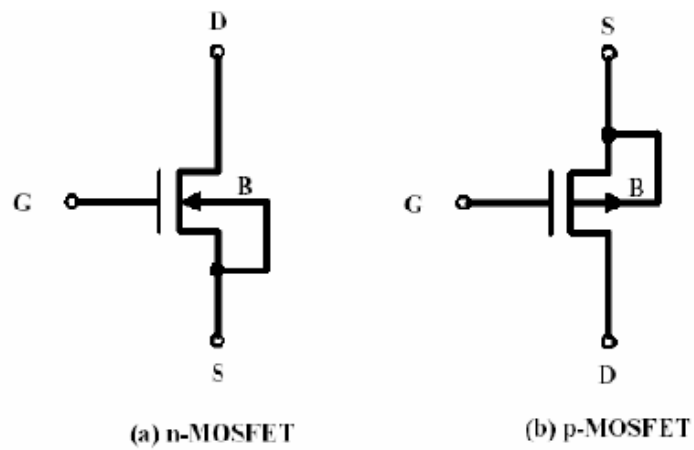


Figura 2.9: Transistor NMOS e PMOS





### 2.5.1 Algebra di Boole

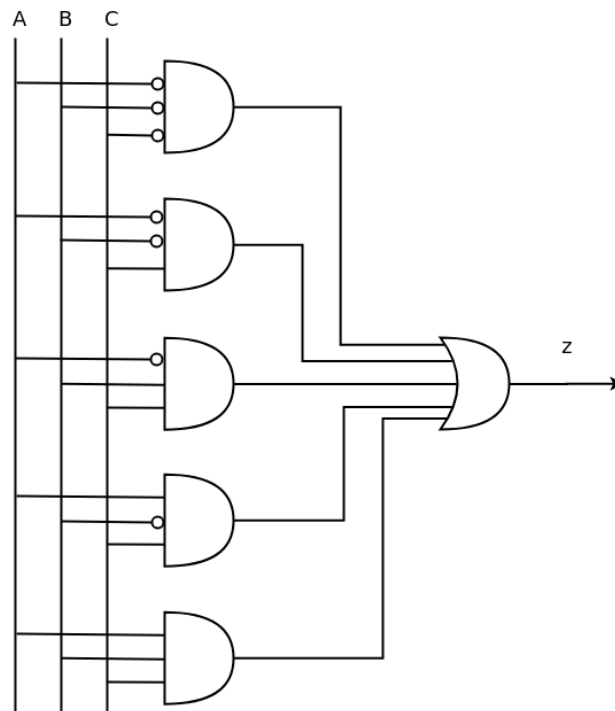


Figura 2.10: Conversione di  $z$  da formula booleana a circuito logico

Funzione	Notazione Usata	Notazione Logica
NOT(A)	$\bar{A}$	$\neg A$
AND(A,B)	$A \cdot B$	$A \wedge B$
OR(A,B)	$A + B$	$A \vee B$

Tabella 2.1: Notazione usata per l'Algebra di Boole

Prendiamo ad esempio un'espressione booleana in forma canonica in **somma di prodotti**:

$$z = \bar{A}\bar{B}\bar{C} + \bar{A}\bar{B}C + \bar{A}BC + A\bar{B}C + ABC$$

A	B	C	z
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Tabella 2.2: Tabella di verità di  $z$

$z$  si può anche esprimere come prodotto di somme:  $z = (A + \bar{B} + C)(\bar{A}BC)(\bar{A}\bar{B}C)$

### 2.5.2 Teoremi dell'Algebra Booleana

Breve ripasso dei teoremi della Logica Booleana.

#### Elemento Identità di prodotto e somma

$$A \cdot 1 = A \iff A \wedge T \equiv A$$

$$A + 0 = A \iff A \vee F \equiv A$$

#### Elemento assorbente

$$A \cdot 0 = 0 \iff A \wedge F \equiv F$$

$$A + 1 = 1 \iff A \vee T \equiv T$$

#### Idempotenza

$$A \cdot A = A \iff A \wedge A \equiv A$$

$$A + A = A \iff A \vee A \equiv A$$

#### Complemento

$$A \cdot \bar{A} = 0 \iff A \wedge \neg A \equiv F$$

$$A + \bar{A} = 1 \iff A \vee \neg A \equiv T$$

#### Commutatività

$$A + B = B + A$$

$$A \cdot B = B \cdot A$$

#### Associatività

$$(A \cdot B) \cdot C = A \cdot (B \cdot C)$$

$$(A + B) + C = A + (B + C)$$

#### Distributività

$$(A \cdot B) + C = (A + C) \cdot (B + C)$$

$$(A + B) \cdot C = AC + BC$$

#### DeMorgan

$$\overline{(A + B)} = \bar{A} \cdot \bar{B} \iff \neg(A \vee B) \equiv \neg A \wedge \neg B$$

$$\overline{(A \cdot B)} = \bar{A} + \bar{B} \iff \neg(A \wedge B) \equiv \neg A \vee \neg B$$

**Esempio** Semplifichiamo la formula booleana  $z = \bar{A}\bar{B}\bar{C} + \bar{A}\bar{B}C + \bar{A}BC + A\bar{B}C + ABC$

$$\begin{cases} \bar{A}\bar{B}\bar{C} + \bar{A}\bar{B}C \equiv \bar{A}\bar{B}(\bar{C} + C) \equiv \bar{A}\bar{B} \\ \bar{A}\bar{B}C + A\bar{B}C \equiv \bar{B}C(\bar{A} + A) \equiv \bar{B}C \end{cases} \quad (2.1)$$

$$\implies z = \bar{A}\bar{B} + \bar{B}C + AC \quad (2.2)$$

Le leggi della logica Booleana ci permettono di semplificare molto i componenti realizzati con porte logiche.

### 2.5.3 Mappe di Karnaugh

Una mappa di Karnaugh o K-Map è un metodo di semplificare un'espressione booleana. I valori sono trasferiti da una tabella di verità ad una mappa bidimensionale:

Approfondisci su Wikipedia

Prendiamo una formula a quattro variabili  $f(A, B, C, D)$ , una mappa di Karnaugh può essere:

AB\CD	00	01	11	10
00	1	1	1	1
01	1	1	0	0
11	0	1	0	0
10	0	0	1	1

Tabella 2.3: Tabella di verità della mappa di Karnaugh di  $f$

Facciamo ad esempio la mappa di Karnaugh di  $z = f(A, B, C)$  vista nella sezione precedente:

A\BC	00	01	11	10
0	1	1	1	0
1	0	1	1	0

Tabella 2.4: Tabella di verità della mappa di Karnaugh di  $z$

Possiamo riconoscere un'implicante nella seconda e terza colonna che corrisponde esattamente a  $C$ . Osserviamo un'altra implicante nella prima riga, prima e seconda colonna che corrisponde esattamente a  $\overline{A}\overline{B}$ . Possiamo poi sommare le implicanti per ottenere una formula equivalente a quella di partenza, ciò implica che  $z = \overline{A}\overline{B} + C$

### 2.5.4 Circuito con più output

Se abbiamo una tabella di verità di un circuito con  $n$  input e  $2^n$  righe, con più output  $z_1, \dots, z_k$ , gli output si suddividono in  $k$  tabelle con un solo output ( $z_k$ ) e  $2^n$  righe di input.

### 2.5.5 Operatori a più ingressi

Gli operatori a più ingressi AND, OR, etc..., se presentano più di due ingressi si rappresentano con una rete logica che sfrutta la proprietà associativa degli operatori logici. Ciò comporta un limite massimo di ingressi perché viene introdotto un ritardo di stabilizzazione determinato e piccolo. Ad esempio, un AND a 4 ingressi sarà rappresentato come  $z = x_1 \cdot x_2 \cdot x_3 \cdot x_4 = (x_1 \cdot x_2) \cdot (x_3 \cdot x_4)$ . Perciò gli operatori associativi a più ingressi si rappresentano come un albero k-ario di porte logiche. Il numero di livelli di porte sarà  $\log_k(n)$  dove  $k$  è l'arietà delle singole porte ed è  $n$  il numero di ingressi nel circuito.



# Capitolo 3

## Reti Logiche e Combinatorie

### 3.1 Alcuni componenti standard

Le tabelle di verità che vediamo si convertono in reti logiche. Una tabella di verità con  $n$  ingressi ha  $2^n$  righe e corrisponde ad un componente logico di un circuito con  $n$  input. Vediamo alcuni oggetti utili. Ricordiamo che i passaggi per definire un componente in forma di rete logica sono:

1. Definizione della funzione con tabella di verità
2. Conversione della funzione normalizzata in somma di prodotti
3. Conversione a circuito con porte AND/OR/NOT

**Multiplexer (k commutatore)** Un multiplexer o commutatore è un circuito, ad esempio, con due 2 ingressi, con un ingresso aggiuntivo chiamato di *controllo* che permette di alternare quale sarà l'input che verrà copiato in output.

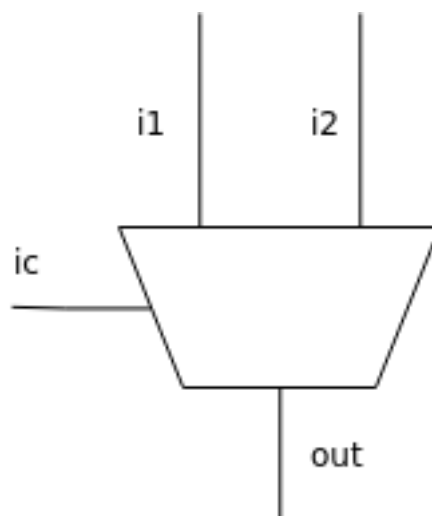


Figura 3.1: Multiplexer 2 vie 1 bit

ictrl	$in_1$	$in_2$	out
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Tabella 3.1: Tabella di verità di un multiplexer a due vie da un bit

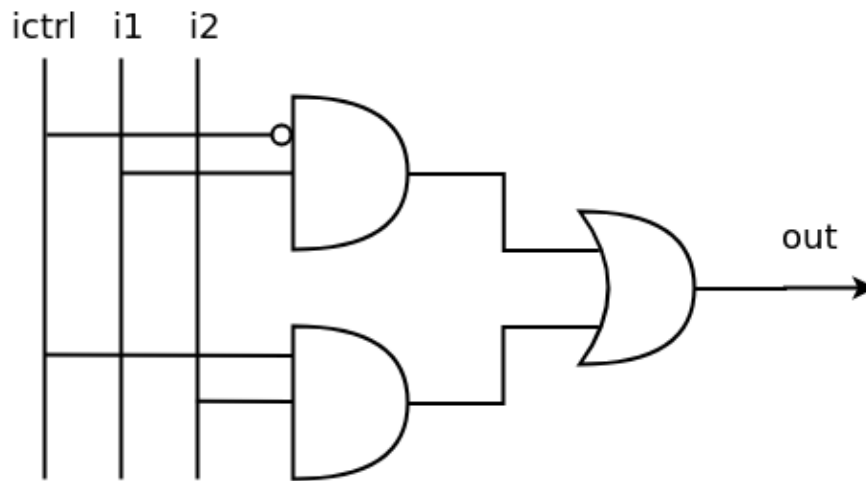


Figura 3.2: Circuito logico di un Multiplexer 2 vie 1 bit

La funzione è definita come  $out = \bar{ic} \cdot in_1 \cdot \bar{in}_2 + \bar{ic} \cdot in_1 \cdot in_2 + ic \cdot \bar{in}_1 \cdot in_2 + ic \cdot in_1 \cdot in_2$  e si può ridurre in  $out = \bar{ic} \cdot in_1 + ic \cdot in_2$ . Le tabelle di verità semplificate ci permettono di dedurre direttamente la formula ridotta. Ad esempio, la tabella seguente corrisponde a quella antecedente.

ictrl	in1	in2	out
0	0	-	0
0	1	-	1
1	-	0	0
1	-	1	1

Tabella 3.2: Tabella di verità semplificata di un multiplexer a due vie da un bit

**Multiplexer a 4 Input** Un multiplexer a 4 input ha bisogno di due bit di controllo. Avendo 6 ingressi, con porte da 8 ingressi massimo  $\Rightarrow \lceil \log_8(6) \rceil$  livelli di porte.

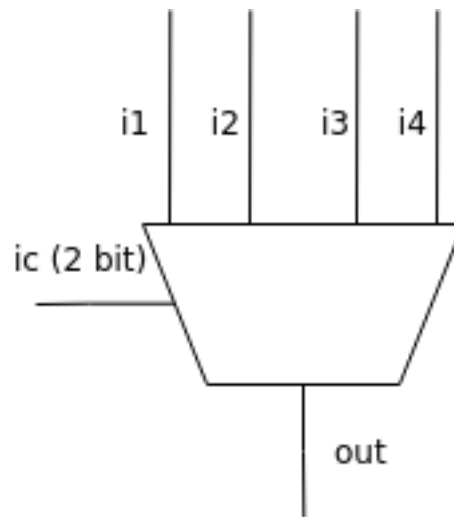


Figura 3.3: Multiplexer 4 vie 1 bit

$ic_1$	$ic_2$	$in_1$	$in_2$	$in_3$	$in_4$	out
0	0	1	-	-	-	1
0	1	-	1	-	-	1
1	0	-	-	1	-	1
1	1	-	-	-	1	1

Tabella 3.3: Tabella di verità semplificata di un multiplexer a quattro vie da un bit

La formula di verità corrispondente sarà

$$\text{out} = (\bar{ic}_1 \cdot \bar{ic}_2 \cdot in_1) + (\bar{ic}_1 \cdot ic_2 \cdot in_2) + (ic_1 \cdot \bar{ic}_2 \cdot in_3) + (ic_1 \cdot ic_2 \cdot in_4)$$

I livelli delle porte AND saranno  $\log_2(n + 1)$

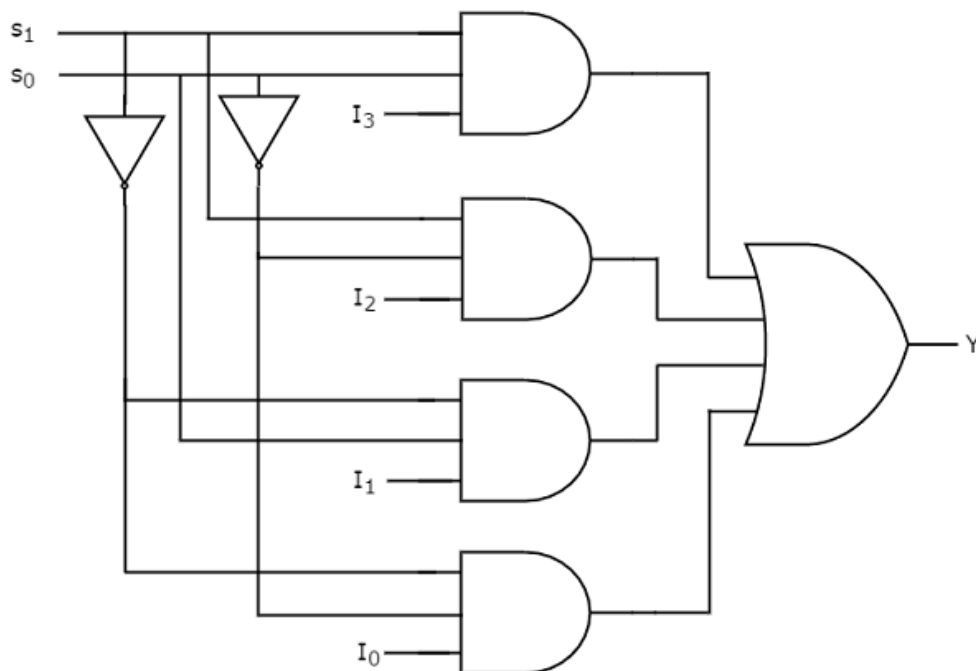


Figura 3.4: Circuito logico di un Multiplexer 4 vie 1 bit

**Commutatori (multiplexer) composti** Un commutatore multiplo da 1 bit con un numero di vie  $y$ , con  $y = 2^x \wedge x \in \mathbb{N}$  e  $y > 2$  si può costruire a partire da un albero con  $\log_2 y$  livelli di multiplexer da 2 vie a 1 bit. Dove ogni bit di controllo del multiplexer complessivo controlla un livello singolo dell'albero.

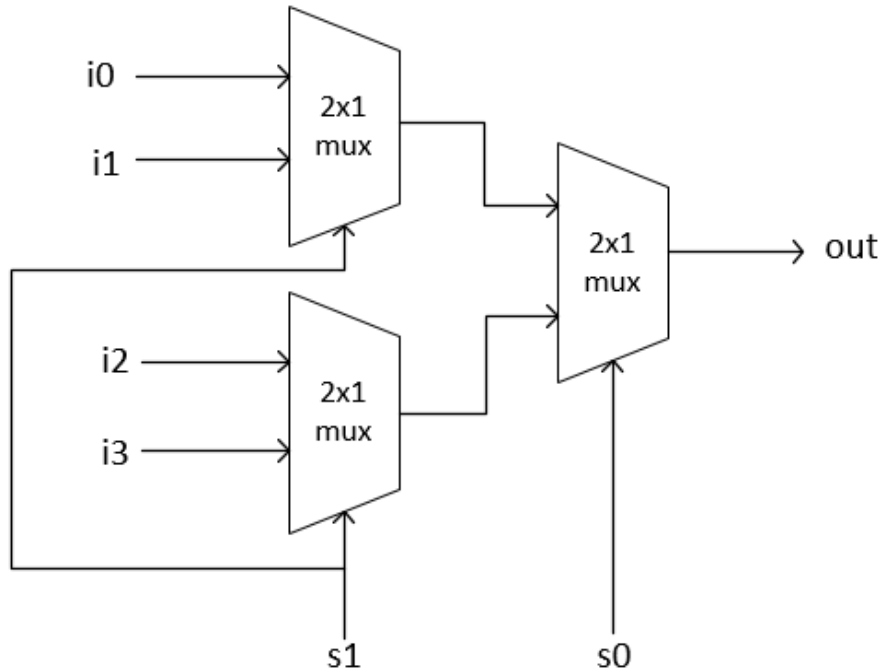


Figura 3.5: Multiplexer 4 vie 1 bit composto da due Multiplexer 2x1

**Multiplexer a 2 vie da  $k$  bit** Un commutatore a 2 vie da  $k$  bit si costruisce con  $k$  commutatori a 2 vie ad 1 bit. Dove ogni commutatore 2x1 accetta in input le corrispettive vie di ogni bit, gli output saranno i  $k$  bit corrispondenti ad ogni Multiplexer.

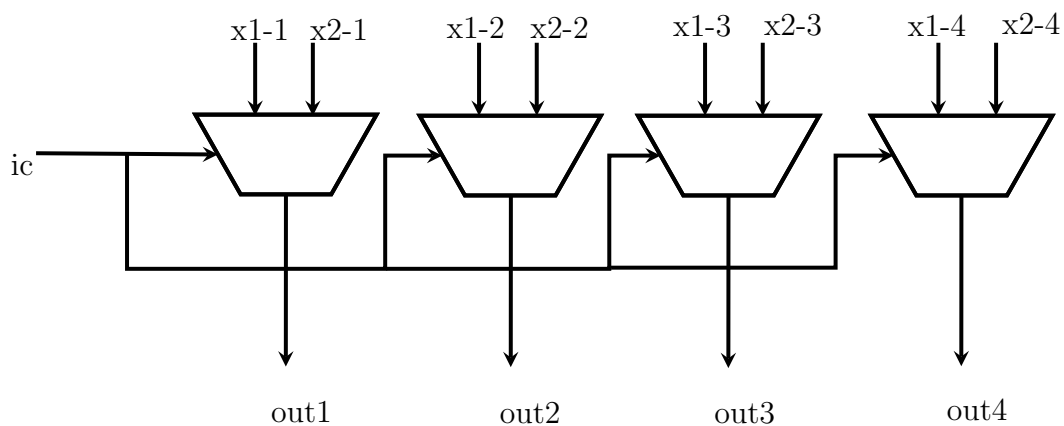


Figura 3.6: Multiplexer 2 vie da 4 bit

**Sommatore di numeri** Un sommatore è un componente che somma due numeri in input e restituisce in output un risultato ed un riporto. Un sommatore di due numeri che non accetta un riporto in input è detto **Half Adder** mentre un sommatore che accetta un riporto è detto Full Adder.



$x_1$	$x_2$	$r_0$	$z_1$	$r_1$
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	1	0
1	0	0	1	0
1	1	0	0	1
1	0	1	0	1
1	1	1	1	1

Tabella 3.4: Tabella di verità di un Full Adder.

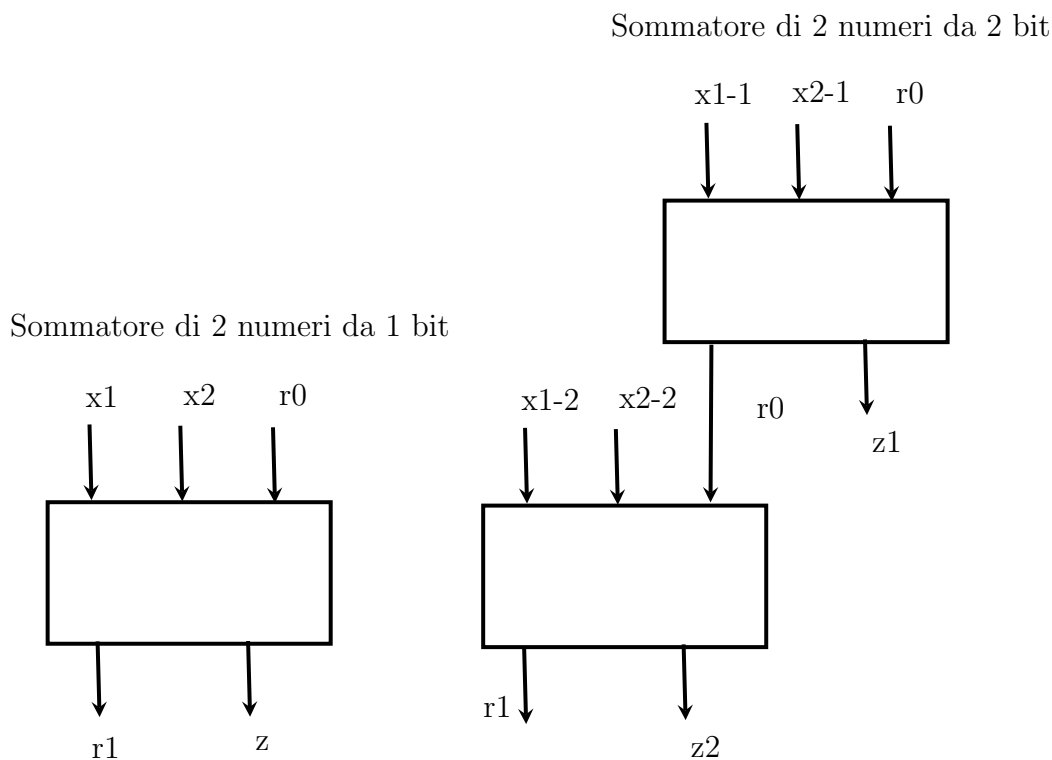


Figura 3.7: Sommatori di due numeri

Analogamente ai commutatori si possono costruire sommatore di 2 numeri a più bit a partire da sommatore di 2 numeri da 1 bit.

**Esercizio** Realizzare la tabella di verità e circuito di un demultiplexer a 2 vie da 1 bit.

**Encoder** Un encoder è un circuito che converte  $2^n$  linee in input in un codice di  $n$  bit.

a	b	c	d	z	t
0	0	0	1	0	0
0	0	1	0	0	1
0	1	0	0	1	0
1	0	0	0	1	1

Tabella 3.5: Tabella di verità di encoder da 2 bit.

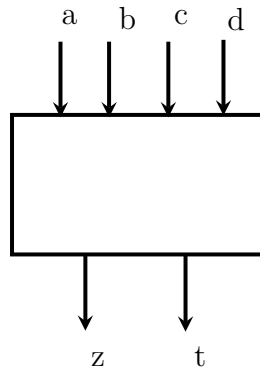


Figura 3.8: Encoder di un numero da 2 bit.

Gli output dell'encoder a 2 bit saranno

$$z = \bar{a}bcd + a\bar{b}cd = \bar{c}d(\bar{a}b + a\bar{b})$$

$$t = \bar{a}bcd + a\bar{b}cd$$

**Decoder** Un decoder esegue l'operazione opposta.

a	b	u	v	z	t
0	0	0	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
1	1	1	0	0	0

Tabella 3.6: Tabella di verità di un decoder da 2 bit.

Gli output saranno  $u = ab, v = a\bar{b}, z = \bar{a}b, t = \bar{a}\bar{b}$

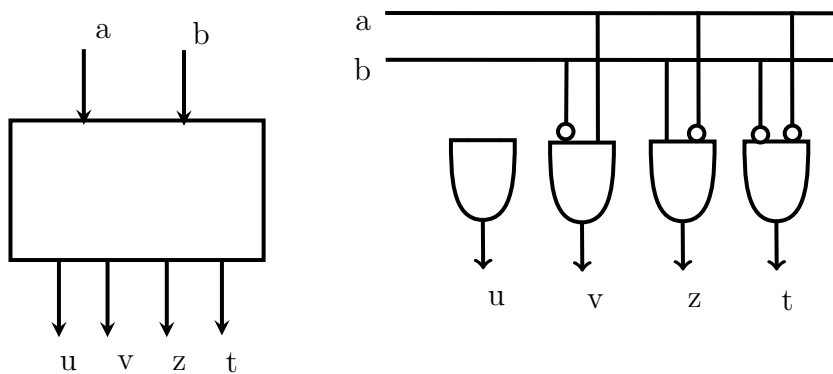


Figura 3.9: Decoder di un numero da 2 bit.

**Confrontatore** Un confrontatore (comparator) confronta due configurazioni di bit e controlla che siano uguali. Un confrontatore da 1 bit è esattamente un gate NOT XOR.

$$z = ab + \bar{a}\bar{b}$$

a	b	t
0	0	1
0	1	0
1	0	0
1	1	1

Tabella 3.7: Tabella di verità di un confrontatore da 1 bit.

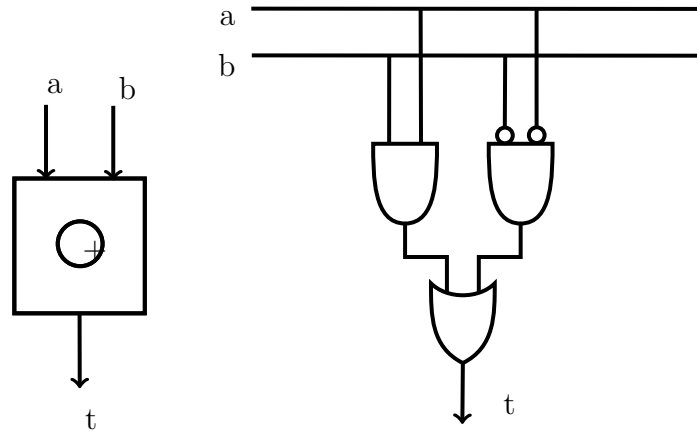


Figura 3.10: Confrontatore da 1 bit

Per realizzare un confrontatore da più bit si mettono in parallelo più confrontatori da un solo bit e si uniscono gli output con un albero di AND.

Un confrontatore a 2 bit impiegherà tempo  $2\Delta t + \Delta t$

Per confrontare valori da  $n$  bit si impiegherà però un tempo di  $\lceil \log_k n \rceil (\Delta t \cdot \text{ogni livello AND})$

**Confronto maggiore** In maniera simile ad un confrontatore si può realizzare un componente che controlla se un numero di  $n$  bit è maggiore di un altro.

$x_0$	$x_1$	$y_0$	$y_1$	$z$
0	0	-	-	0
0	1	0	0	1
1	-	0	-	1
1	1	-	0	1

Tabella 3.8: Tabella di verità di un confronto maggiore da 2 bit.

$y_1 y_0 \backslash x_1 x_0$	00	01	11	10
00	0	1	1	1
01	0	0	1	1
11	0	0	0	0
10	0	0	1	0

Tabella 3.9: Mappa di Karnaugh di un confrontatore dell'operatore maggiore da 2 bit.

La formula booleana ottenuta è  $z = x_0 \overline{y_1} \overline{y_0} + x_1 \overline{y_1} + x_1 x_0 \overline{y_0}$

**ALU** Una ALU, ovvero Arithmetic Logic Unit è un componente che in genere, in input accetta due parole  $x, y$  entrambe da  $n$  bit, e può fare due o più operazioni. L'operazione viene selezionata attraverso uno o più bit di controllo. In output ci sarà un uscita da  $n$  bit. Ad esempio, prendiamo una ALU a 4 bit con somma/sottrazione, che accetterà un solo bit di controllo.

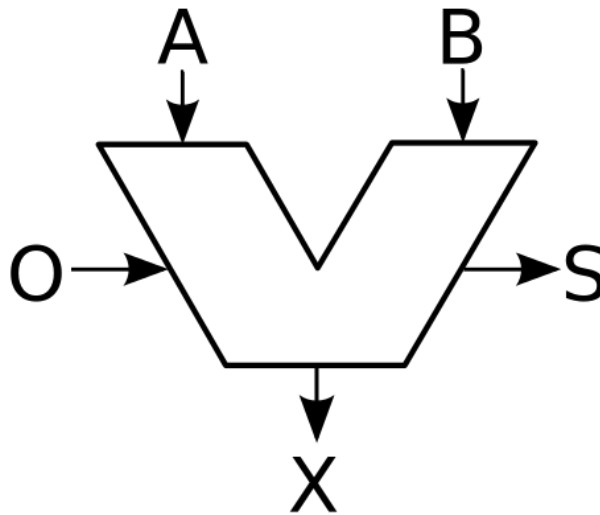


Figura 3.11: Simbolo di una ALU

I livelli di porte AND saranno  $\lceil \log_k 8 \rceil$  che verranno sommati ai livelli di porte OR che saranno  $\log_k(128)$ . Le ALU si distinguono fra intere e a floating point. Per semplicità vedremo le ALU intere con somma e moltiplicazione.

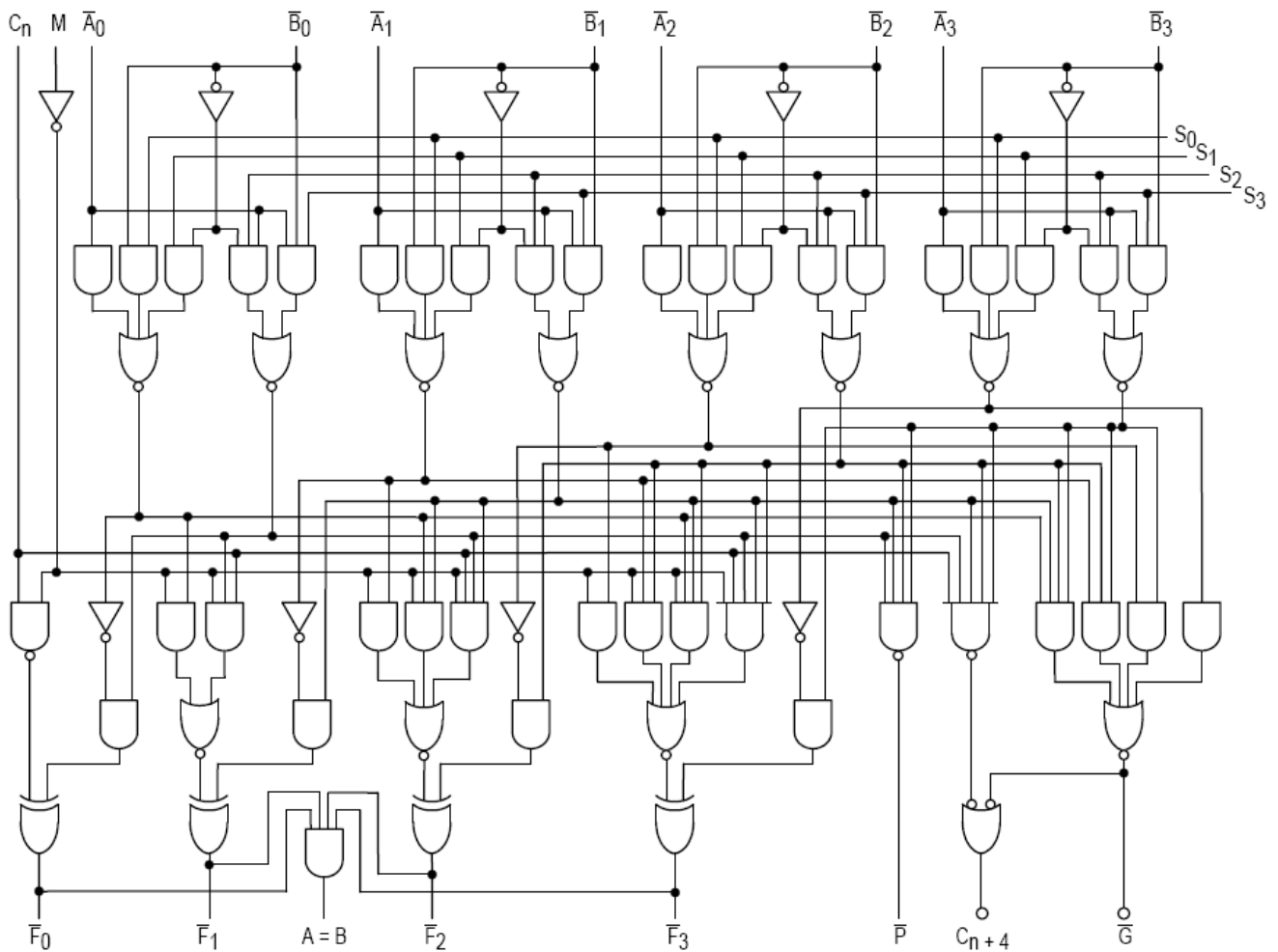


Figura 3.12: Schematica del circuito integrato della ALU 74181, la prima ALU completa su un singolo chip.

**Multiply and Add** Un componente Multiply and Add (MUL&ADD) è un componente ALU che realizza l'operazione matematica  $\sum_i x_i y_i$ , ovvero prende gli ingressi  $x_i, y_i$ , li moltiplica e li accumula sommando.

**Ritardi di propagazione** Sappiamo che il cambiamento fra HIGH e LOW (0 e 1) nei circuiti digitali non è istantaneo. A volte, il ritardo di propagazione dei transistor può causare dei glitch (fluttuazioni) che potrebbero causare una lettura incorretta dell'output di un circuito. Per ovviare a questo problema si introduce un componente chiamato **clock**, che scandisce un ciclo preciso per permettere ai valori di propagarsi nei circuiti e leggere l'output preciso alla fine della propagazione  $\Delta t$ .

## 3.2 Reti sequenziali e Automi

Abbiamo visto nel corso di programmazione 1 gli automi a stati finiti (Finite State Machines o FSM). Per ora, abbiamo visto circuiti logici che implementano funzioni senza stato. Per implementare lo stato possiamo utilizzare gli automi a stati finiti. In un automa di Moore gli output sono determinati solo dallo stato corrente, mentre negli automi di Mealy gli output sono determinati sia dallo stato corrente che dagli input correnti.

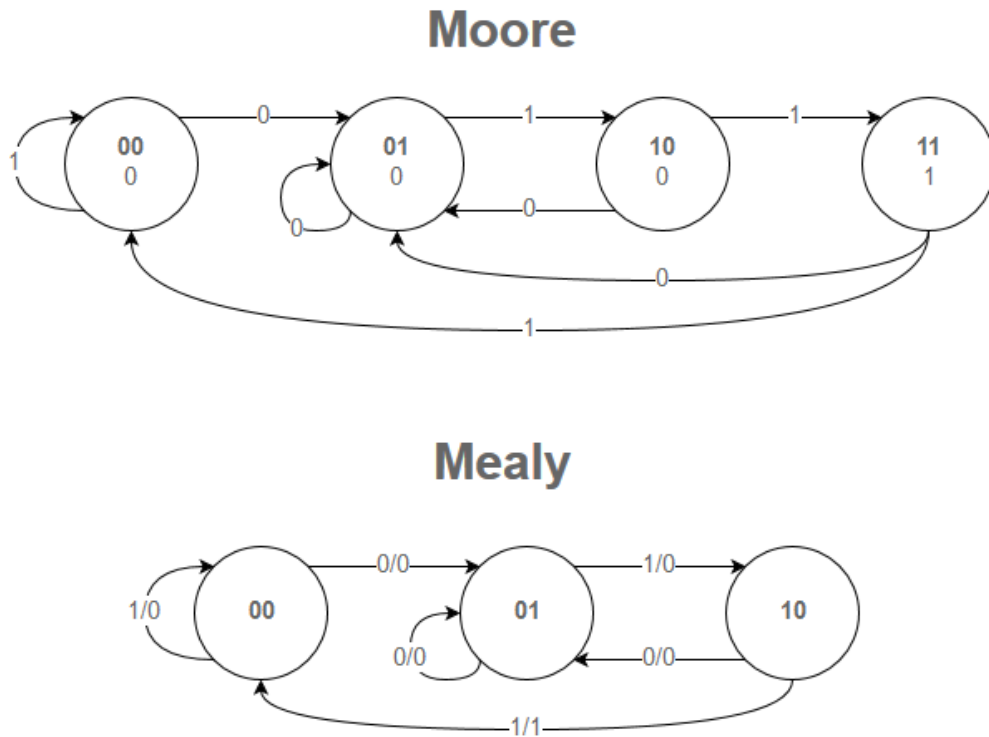


Figura 3.13: Automi di Moore e Automi di Mealy

Gli automi di Moore hanno sempre un numero di stati **maggiore o uguale** agli automi di Mealy. Per portare un Automa di Mealy ad una rete sequenziale abbiamo bisogno di tre blocchi di una rete combinatoria. Uno che dallo stato precedente  $s$  e dall'input  $x$  dia in output l'uscita  $z$ , e un altro che ricevendo  $x, s$  in input restituisca  $s^1$  (lo stato del prossimo nodo del grafo) e un **SR Latch** (o **flip-flop**). I latch possono essere di tipo  $SR, D, T$  e  $JK$ . Un **flip-flop** è un latch che accetta in input un segnale di clock e i due bit  $S, R$  (Set e Reset)

**SR Latch** Un componente SR Latch è un componente che implementa un bit di memoria. Impostando l'input  $S = 1$  (bit di set) e  $R = 0$  (bit di reset) l'output sarà  $Q = 1, \bar{Q} = 0$  mentre impostando gli input  $S = 0, R = 1$  l'output è  $Q = 0, \bar{Q} = 1$ . Il fulcro della funzionalità di un SR Latch è che se entrambi gli input  $S = 0, R = 0$  allora l'output è in uno stato detto **latch**, ovvero che "ricorda" lo stato precedente.

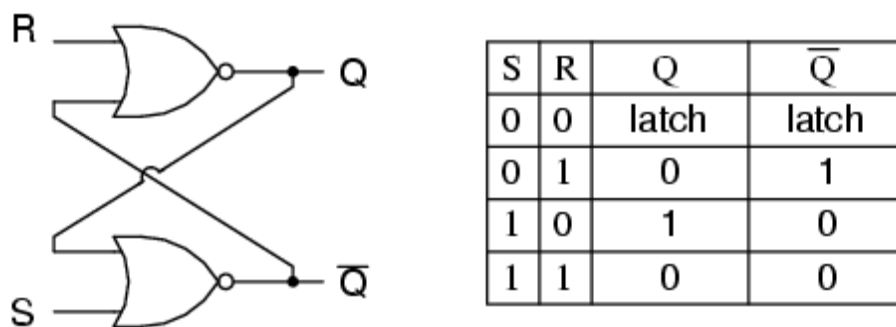


Figura 3.14: Tabella di Verità e circuito di un SR Latch

**Clock e D-latch** Introduciamo un componente fondamentale, il **clock**, che introduce una dimensione temporale nelle reti combinatorie scandendo un **ciclo**. Un ciclo di clock ha lunghezza  $\tau$ . Possiamo così estendere un SR Latch rendendolo un **D-Latch**, che riceve in input soltanto il ciclo di clock e un bit  $D$ , ovvero il bit che va "ricordato". Per evitare **glitch** possiamo scrivere sul D-Latch soltanto quando il segnale di clock è nello stato HIGH.

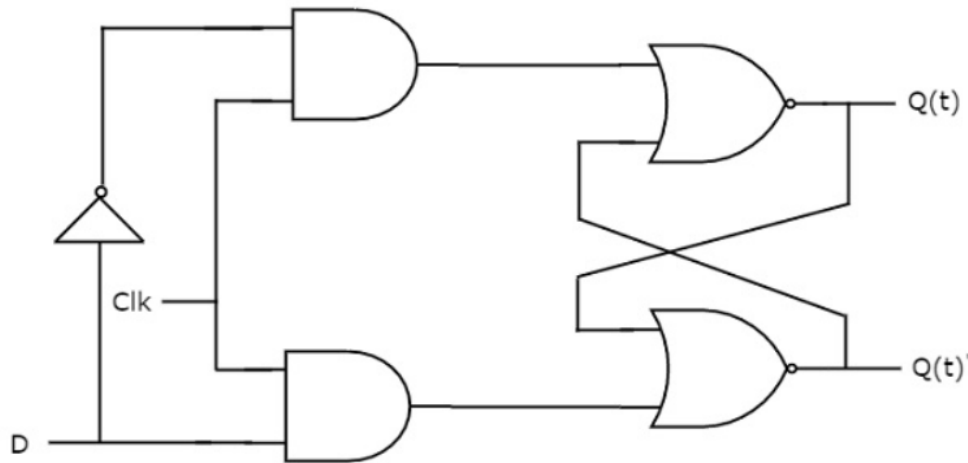


Figura 3.15: Un componente D-latch

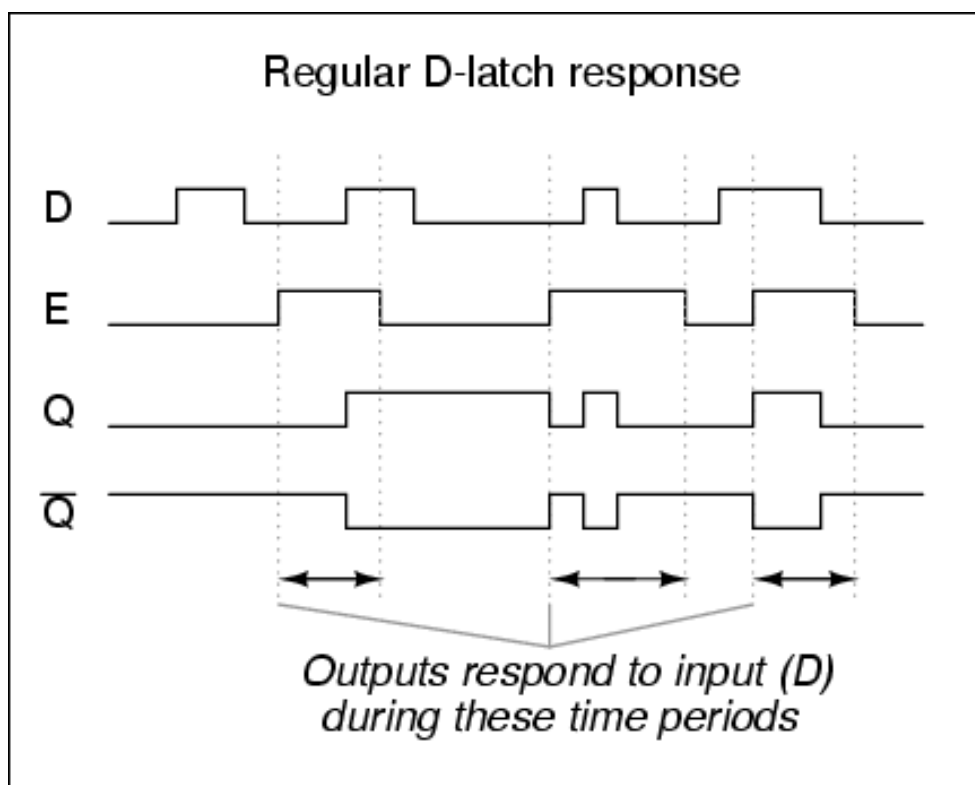


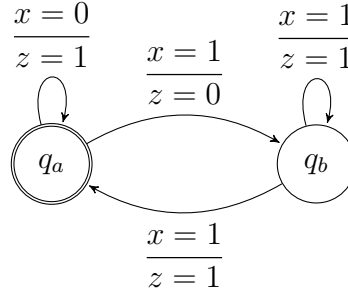
Figura 3.16: Diagramma temporale di un D-latch

### 3.2.1 Realizzazione di una rete sequenziale di un Automa di Mealy con un clock e registri

Consideriamo, ad esempio, un automa che calcola la parità del numero di 1 in una sequenza di bit. Ricevendo in input una sequenza di bit 1,0,0,0,1,0,1 ad esempio restituirebbe in output

0, 1, 1, 1, 1, 0, 0 (l'output diventa high)

Figura 3.17: Automa di parità di una sequenza di bit



Per trasformare l'automa di Mealy in una rete sequenziale avremo una rete combinatoria che calcola lo stato successivo, un registro che mantiene lo stato e una rete combinatoria che calcola l'uscita. Siccome l'automa di Mealy usa l'input  $x$  per decidere le transizioni, l'input  $x$  dovrà essere passato ad entrambe le reti combinatorie dei nodi del grafo.

Chiamiamo la prima rete combinatoria  $\sigma$  e la seconda  $\omega$

x	s	s'
0	0	0
1	0	1
0	1	1
1	1	0

Tabella 3.10: Tabella di verità della rete combinatoria  $\sigma$

x	s	z
0	0	1
1	0	0
0	1	0
1	1	1

Tabella 3.11: Tabella di verità della rete combinatoria  $\omega$

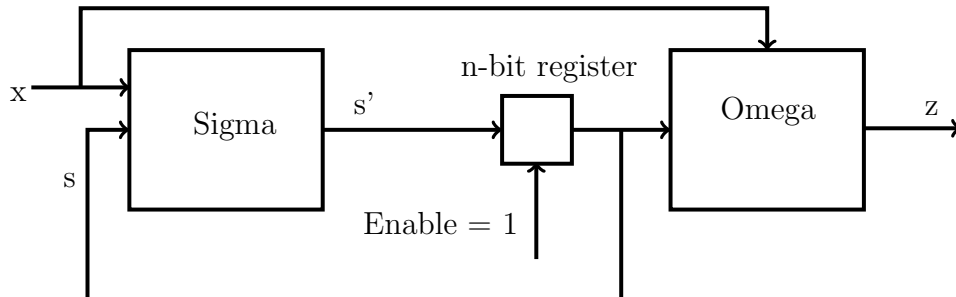


Figura 3.18: Rete sequenziale dell'automa di Mealy in figura precedente



# Capitolo 4

## Reti Sequenziali, Verilog e RTL

Gli RTL, o Register Transfer Language, permettono di descrivere cosa succede a livello di circuito fra registri. Vengono utilizzati per descrivere l'hardware. Vedremo il linguaggio **Verilog**. Gli RTL permettono di descrivere e comporre dei moduli. Il libro di testo propone il dialetto **System Verilog** che mette a disposizione due metodi per descrivere i moduli. Un metodo è il metodo *constructive*, noi vedremo il metodo *behavioral* dove ad esempio un Multiplexer da 2 vie 1 bit è descritto da:

```
1      z = (ic == 0 ? x : y)
```

Verilog è un linguaggio compilato. Un file System Verilog compilato produce una traccia di esecuzione e un eseguibile che simula il comportamento dei moduli. Viene detta **simulazione**.

Un programma Verilog può anche essere dato in input a un programma detto **synthetizer**, che produce una **netlist**, ovvero una lista dei componenti e dei collegamenti per realizzare il modulo fisicamente. Un altro modo per realizzare la sintesi è utilizzare un **FPGA**, o Field-programmable gate array. Un FPGA è un circuito integrato composto da una matrice di celle, e una singola cella può:

1. Eseguire una funzione booleana di 3-5 ingressi con 1 uscita
2. Implementare un bit di memoria
3. Routing

Un FPGA moderno comprende, oltre a delle celle, delle righe che contengono diversi componenti come delle ALU.

Listing 4.1: mux.sv

```

1 module mux (output logic z, input logic x, y, ic );
2     assign z = (ic == 0 ? x : y);
3 endmodule

```

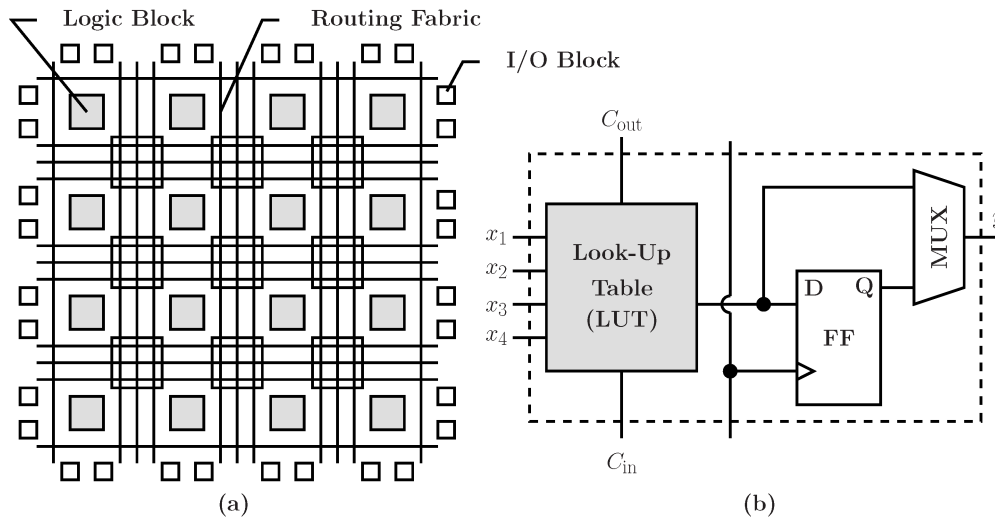


Figura 4.1: Schema FPGA

## 4.1 Scrivere e compilare System Verilog

Creiamo un multiplexer da due bit con System Verilog, compiliamo e visualizziamo con **GTK-Wave**

Per compilare, eseguiamo da terminale

```
iverilog -g2005-sv nome_sorgente.sv -o nome_eseguibile
```

Quindi, per compilare entrambi i file e caricarli in GTKWave:

```
iverilog -g2005-sv test_mux.sv mux.sv -o test_mux
```

```
# Eseguiamo la simulazione
```

```
./test_mux
```

```
# Viene creato il file provamux.vcd, carichiamolo in GTKWave
```

```
gtkwave provamux.vcd &
```

Listing 4.2: test\_mux.sv

```

1 module test_mux();
2
3 reg my_x, my_y, my_ic;
4 wire my_z;
5
6 mux mymux(my_z, my_x, my_y, my_ic);
7
8     initial
9         begin
10             // Dump log to a file
11             $dumpfile("provamux.vcd");
12             $dumpvars;
13             my_x = 0;
14             my_y = 0;
15             my_ic = 1;
16
17             #10
18                 my_x = 1;
19
20             $finish;
21         end
22 endmodule // test_mux

```

Listing 4.3: Multiplexer da 4 vie ad 1 bit

```

1 module mux4(output logic z,
2     input logic [3:0] x,
3     input logic [0:1] ic);
4
5     logic k12k3;
6     logic k22k3;
7
8     mux k1(k12k3,x[0],x[1],ic[0]);
9     mux k2(k22k3,x[2],x[3],ic[0]);
10    mux k3(z,k12k3,k22k3,ic[1]);
11 endmodule

```

Listing 4.4: Multiplexer di variabili booleane

```

1 module mux(output logic z, input logic x, y, ic);
2
3     assign
4         z = ((~ic) & x) | (ic & y);
5
6 endmodule

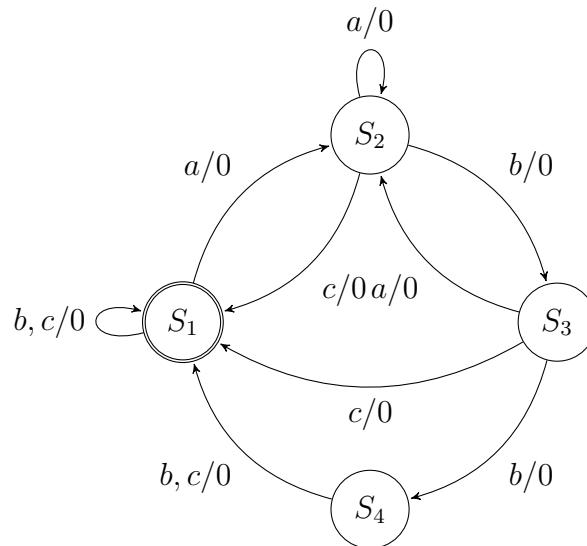
```

## 4.2 Esercizi

### 4.2.1 Automa che riconosce "abba"

Realizziamo un automa di Mealy che riconosce le stringhe "abba" da un insieme  $\{a, b, c\}$ . La rete sequenziale dell'automa si realizzerà con i componenti visti in figura 3.18. Consideriamo la rappresentazione binaria dell'alfabeto con  $a = 00, b = 01, c = 11$ . Per gli stati usiamo la codifica  $S_1 = 00, S_2 = 01, S_3 = 11, S_4 = 10$

Figura 4.2: Automa di Mealy che riconosce "abba"



	$s_1$	$s_2$	$x_1$	$x_2$	$z$
Stato $S_1$	0	0	-	-	0
Stato $S_2$	0	1	-	-	0
Stato $S_3$	1	1	-	-	0
Stato $S_4$	1	0	0	0	1
	1	0	0	1	0
	1	0	1	1	0

Tabella 4.1: Tabella di verità dell'output della rete sequenziale dell'automa "abba" ( $\omega$ )

	$s_1$	$s_2$	$x_1$	$x_2$	$s'_1$	$s'_2$
Stato $S_1$	0	0	0	0	0	1
	0	0	0	1	0	0
	0	0	1	1	0	0
Stato $S_2$	0	1	0	0	0	1
	0	1	0	1	1	1
	0	1	1	1	0	0
Stato $S_3$	1	1	0	0	0	1
	1	1	0	1	1	0
	1	1	1	1	0	0
Stato $S_4$	1	0	0	0	0	0
	1	0	0	1	0	0
	1	0	1	1	0	0

Tabella 4.2: Tabella di verità del cambio di stato dell'automa per riconoscere "abba" ( $\sigma$ )

Avremo che la formula booleana sarà per il primo e secondo bit di stato:

$$s'_1 = \overline{s_1}s_2\overline{x_1}x_2 + s_1s_2\overline{x_1}x_2 = s_2\overline{x_1}x_2$$

$$s'_2 = \overline{s_1}\overline{s_2} + \overline{s_1}s_2\overline{x_1} + s_2\overline{x_2}$$

Nella componente **sigma** in Verilog avremo l'assegnamento ad un nuovo stato **news**:

```

1 assign
2     news[1] = s[2] & ~x[1] & x[2];
3
4 assign
5     news[2] = ~s[1] & ~x[1] & ~x[2]
6             | s[2] & ~x[1] & ~x[2]
7             | ~s[1] & s[2] & ~x[1];

```

La formula per l'uscita sarà  $z = s_1\overline{s_2}\overline{x_1}x_2$ . La componente **omega** effettuerà l'assegnamento

```

1 assign
2     z = s[1] & ~s[2] & ~x[1] & ~x[2];

```

Ci rimane da definire il registro da 2 bit per definire tutta la rete sequenziale. Abbiamo supposto che la rete sequenziale funzioni ricevendo un segnale di clock ad intervalli regolari. La rete di output  $\omega$  è definita utilizzando soltanto AND ed impiegherà  $\Delta t$  per eseguire l'operazione. La funzione di cambio di stato  $\sigma$  è definita invece con 1 AND e 1 OR ed impiegherà  $2\Delta t$  per l'operazione. Il ciclo di clock dev'essere almeno  $\tau = \delta + \max\{t_\sigma, t_\omega\}$  dove  $\delta$  è la durata del segnale HIGH nel clock.

Abbiamo 3 moduli. Uno per il registro, uno per il modulo  $\omega$  e uno per il modulo  $\sigma$ . Scriviamolo in Verilog.

**Note sulla sintassi e semantica Verilog** La sintassi  $[a:b]$  denota un array indicizzato dal numero  $a$  al numero  $b$ . Utilizziamo gli array per rappresentare valori a più bit. In questo caso, il registro è generalizzato per un parametro  $N$  e può ad esempio essere inizializzato con **registro #4 nome(...)** dove 4 è il parametro  $N$  e corrisponde al numero di bit del registro.

La negazione con **!** nega un singolo valore, mentre la notazione **~** viene detta negazione *bit wise* e nega tutti i valori di una sequenza di bit.

Listing 4.5: Registro a N bit

```

1 module registro(output [1:N]z, input clock, input enable, input [1:N]inval);
2
3     parameter N = 2;
4
5     reg [1:N] stato;
6
7     initial
8         begin
9             stato = 0;
10        end
11
12    always @(posedge clock)
13        begin
14            if(enable)
15                stato <= inval;
16        end
17
18    assign
19        z = stato;
20
21 endmodule

```

Normalmente, in un blocco **assign** assegno un valore ad una variabile booleana istantaneamente. Posso invece aggiungere un delay inserendo **#x** di fronte all'identificatore della variabile, dove **x** è un numero. Ad esempio:

```

1 assign
2     #1 z = (ic == 0 ? x : y)

```

Per sincronizzare il modulo della rete sequenziale dell'automa di Mealy mettiamo un registro di fra l'input **x** e le reti sequenziali **sigma** e **omega**.

Listing 4.6: Modulo di  $\sigma$  o funzione di cambio di stato

```

1 module sigma(output [1:2]news, input [1:2]s, input [1:2]x);
2
3     assign
4         news[1] = s[2] & ~x[1] & x[2];
5
6     assign
7         news[2] = ~s[1] & ~x[1] & ~x[2]
8                 | s[2] & ~x[1] & ~x[2]
9                 | ~s[1] & s[2] & ~x[1];
10
11 endmodule

```

Listing 4.7: Modulo di  $\omega$ 

```
1 module omega(output z, input [1:2]s, input [1:2]x);
2
3   assign
4     z = s[1] & ~s[2] & ~x[1] & ~x[2];
5
6 endmodule
```

Listing 4.8: Modulo della rete sequenziale

```
1 module m1(output z, input [1:2]x, input clock);
2
3   wire [1:2]outreg;
4   wire [1:2]inreg;
5
6   registro s(outreg, clock, 1'b1, inreg);
7   sigma sm(inreg, outreg, x);
8   omega om(z, outreg, x);
9
10 endmodule
```

Listing 4.9: Modulo della rete sequenziale (con Sincronizzazione)

```
1 module m1(output z, input [1:2]x, input clock);
2
3   wire [1:2] outreg;
4   wire [1:2] inreg;
5   wire [1:2] xsy;
6
7   registro #(2) r(xsy, clock, 1'b1, x);
8
9   registro #(2) s(outreg, clock, 1'b1, inreg);
10  sigma sm(inreg, outreg, xsy);
11  omega om(z, outreg, xsy);
12
13
14 endmodule
```

Listing 4.10: Modulo di test della rete sequenziale

```
1 module testm1();
2
3     reg [1:2] x;
4     reg      clock;
5     wire z;
6
7     m1 m(z, x, clock);
8
9     initial
10         clock = 0;
11
12     always
13         begin
14             #2 clock = 1;
15             #1 clock = 0;
16         end
17
18     initial
19         begin
20             $dumpfile("test-mealy.vcd");
21             $dumpvars;
22
23             x = 2'b00;
24             #4 x = 2'b01;
25             #5 x = 2'b01;
26             #1 x = 2'b00;
27
28
29             #10 $finish;
30         end
31 endmodule // testm1
```



Listing 4.11: Makefile

```

SOURCES = test-m1.v m1.v sigma.v omega.v reg.v
CC = iverilog
VIEW = gtkwave
EXECS = a.out Mealy.out

mealy: $(SOURCES)
    $(CC) $(SOURCES) -o Mealy.out

clean:
    rm -f $(EXECS)

run:    mealy
        ./Mealy.out
        $(VIEW) test-mealy.vcd

```

Figura 4.3: Diagramma dei tempi della Rete Sequenziale dell'automa di Mealy per riconoscere "abba" **senza sincronizzazione**

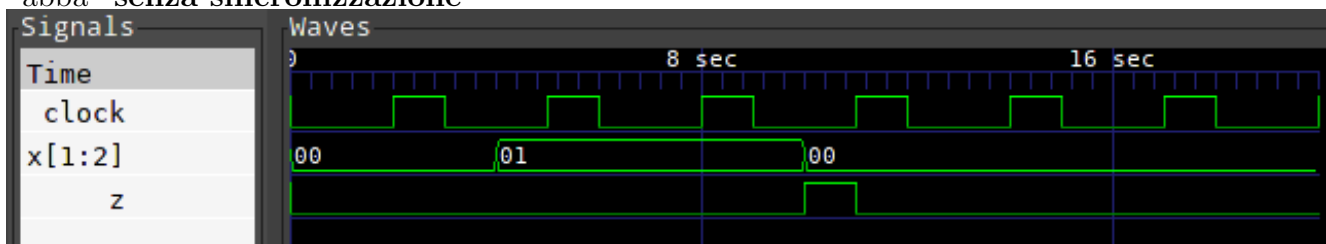
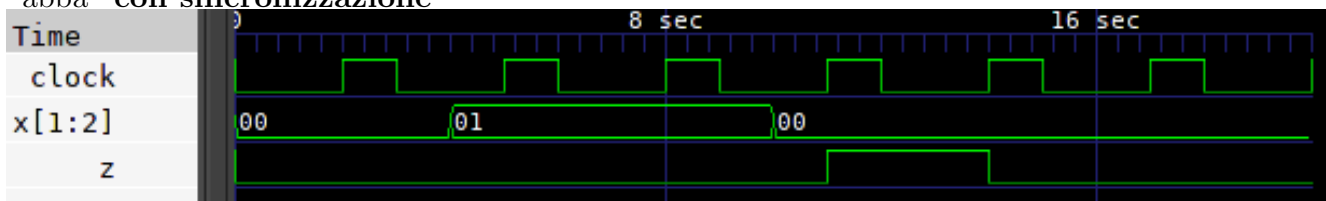


Figura 4.4: Diagramma dei tempi della Rete Sequenziale dell'automa di Mealy per riconoscere "abba" **con sincronizzazione**



## 4.2.2 Automa di Moore per riconoscere "abba"

Vediamo adesso un automa di Moore per riconoscere la stringa "abba" nell'alfabeto  $\{a, b, c\}$ . La differenza sta nel fatto che i singoli nodi non hanno accesso all'input  $x$ . Teniamo nota che la differenza fra un automa di Mealy di un automa di Moore, consiste che nell'automa di Moore la funzione  $\omega$  non accetta riceve l'input

Figura 4.5: Automa di Moore che riconosce "abba"

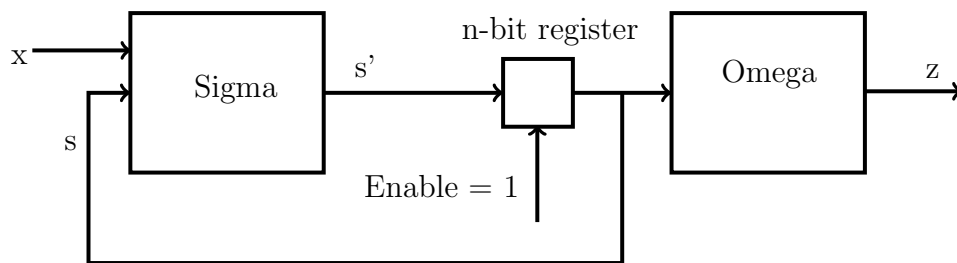
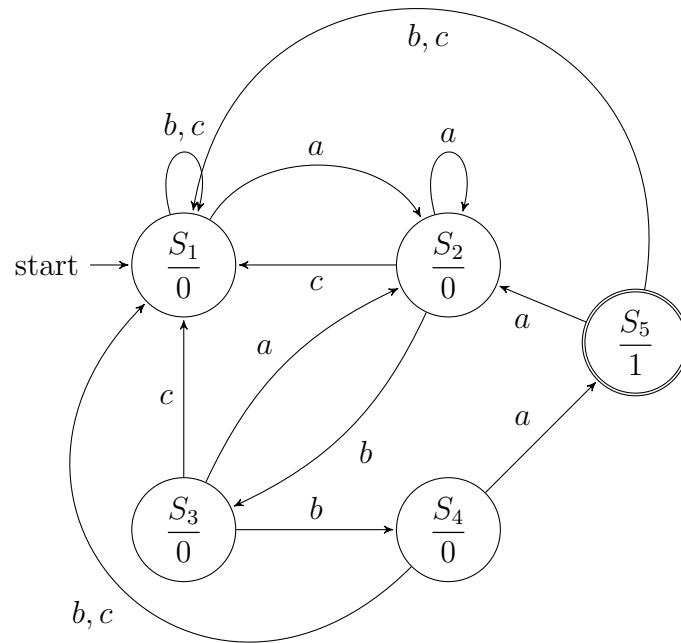


Figura 4.6: Rete Sequenziale dell'automa di Moore per riconoscere "abba"

Listing 4.12: Modulo di  $\sigma$  o funzione di cambio di stato (Moore)

```

1 module sigma(output [1:3]t, input [1:3]s, input [1:2]x);
2   // s[1] & s[2] & s[3] & x[1] & x[2]
3   assign
4     t[1] = ~s[1] & s[2] & s[3] & ~x[1] & ~x[2];
5   assign
6     t[2] = ~s[1] & ~s[2] & s[3] & ~x[1] & x[2] |
7           ~s[1] & s[2] & ~x[1] & x[2];
8   assign
9     t[3] = ~s[1] & ~s[2] & ~x[1] & ~x[2] |
10          ~s[1] & s[2] & ~s[3] & ~x[1] |
11          s[1] & ~s[2] & ~s[3] & ~x[1] & x[2];
12 endmodule

```

Listing 4.13: Modulo di  $\omega$  (Moore)

```

1 module omega(output z, input [1:3]s, input [1:2]x);
2   // z = 1 solo se ci troviamo nello stato 5 (100)
3   assign
4     z = s[1] & ~s[2] & ~s[3];
5
6
7 endmodule

```

Listing 4.14: Modulo della rete sequenziale (Moore)

```

1 module m1(output z, input [1:2]x , input clock);
2
3   wire [1:3]sinp;
4   wire [1:3]sout;
5
6   sigma sm(sinp, sout, x);
7   omega om(z, sout, x);
8   registro #(3) rs(sout, clock, 1'b1, sinp);
9
10
11 endmodule

```

Listing 4.15: Modulo di test della rete sequenziale (Moore)

```
1 module testm1();
2
3     reg [1:2] x;
4     reg      clock;
5     wire z;
6
7     m1 m(z, x, clock);
8
9     initial
10         clock = 0;
11
12     always
13         begin
14             #2 clock = 1;
15             #1 clock = 0;
16         end
17
18     initial
19         begin
20             $dumpfile("test-moore.vcd");
21             $dumpvars;
22
23             x = 2'b00;
24             #4 x = 2'b01;
25             #5 x = 2'b01;
26             #1 x = 2'b00;
27
28
29             #10 $finish;
30         end
31 endmodule // testm1
```

### 4.2.3 Mealy con Delay $\equiv$ Moore

Vedremo come in termini di esecuzione, introducendo dei delay nell'automa di Mealy otteniamo un comportamento analogo all'automa di Moore.

Listing 4.16: Modulo di  $\sigma$  o funzione di cambio di stato (Mealy con Delay)

```
1 module sigma(output [1:2]news, input [1:2]s, input [1:2]x);
2
3 assign
4     #1 news[1] = s[2] & ~x[1] & x[2];
5
6 assign
7     #2 news[2] = ~s[1] & ~x[1] & ~x[2]
8                 | s[2] & ~x[1] & ~x[2]
9                 | ~s[1] & s[2] & ~x[1];
10
11 endmodule
```

Listing 4.17: Modulo di  $\omega$  (Mealy con Delay)

```

1 module omega(output z, input [1:2]s, input [1:2]x);
2
3   assign
4     #1 z = s[1] & ~s[2] & ~x[1] & ~x[2];
5
6 endmodule

```

Listing 4.18: Registro a N bit con delay

```

1 module registro(output [1:N]stato, input clock, input enable, input [1:N]inval);
2
3   parameter N = 2;
4
5   reg [1:N] st;
6
7   initial
8     begin
9       st = 0;
10    end
11
12   always @(posedge clock)
13     begin
14       if(enable)
15         st <= inval;
16     end
17
18   assign
19     #1 stato = st;
20
21 endmodule

```

Listing 4.19: Modulo della rete sequenziale (Mealy con Delay)

```

1 module m1(output z, input [1:2]x, input clock);
2
3   wire [1:2]outreg;
4   wire [1:2]inreg;
5
6   registro s(outreg, clock, 1'b1, inreg);
7   sigma sm(inreg, outreg, x);
8   omega om(z, outreg, x);
9
10 endmodule

```

Listing 4.20: Modulo di test della rete sequenziale (Mealy con Delay)

```
1 module testm1();
2
3     reg [1:2] x;
4     reg      clock;
5     wire z;
6
7     m1 m(z, x, clock);
8
9     initial
10         clock = 0;
11
12     // ciclo di clock tau = 4 (leggermente piu' grande del necessario (3))
13     always
14     begin
15         // funz
16         #3 clock = 1;
17         // non funz
18         //#1 clock = 1;
19
20         #1 clock = 0;
21     end
22
23     initial
24     begin
25         $dumpfile("test-mealy-delay.vcd");
26         $dumpvars;
27
28         x = 2'b00;
29         #4 x = 2'b01;
30         // #4 fa aba e non da' mai 1
31         // #8 fa abba e mi da' l'uno
32         // se metto il clock a 2 (1+1) non funziona piu'
33         #8 x = 2'b01;
34         // col clock a 1+1 non funz
35         //#5 x = 2'b01;
36
37         #1 x = 2'b00;
38
39
40         #10 $finish;
41     end
42 endmodule // testm1
```

### 4.3 Componenti per il Calcolo

Nel capitolo precedente abbiamo visto gli **adder**, che accettano due bit e restituiscono bit sommato e riporto, e abbiamo visto i **full adder**, che accettano in input anche un riporto per poter essere composti a cascata. Tutti i tempi di stabilizzazione dei full adder si sommano l'uno con l'altro. Ciò è chiaramente un problema perché se dobbiamo sommare due numeri da 32 bit dobbiamo aspettare la stabilizzazione di 32 sommatore. Oppure, sommando due numeri da 4 bit avremo  $4 \cdot 2\Delta t = 8\Delta t$

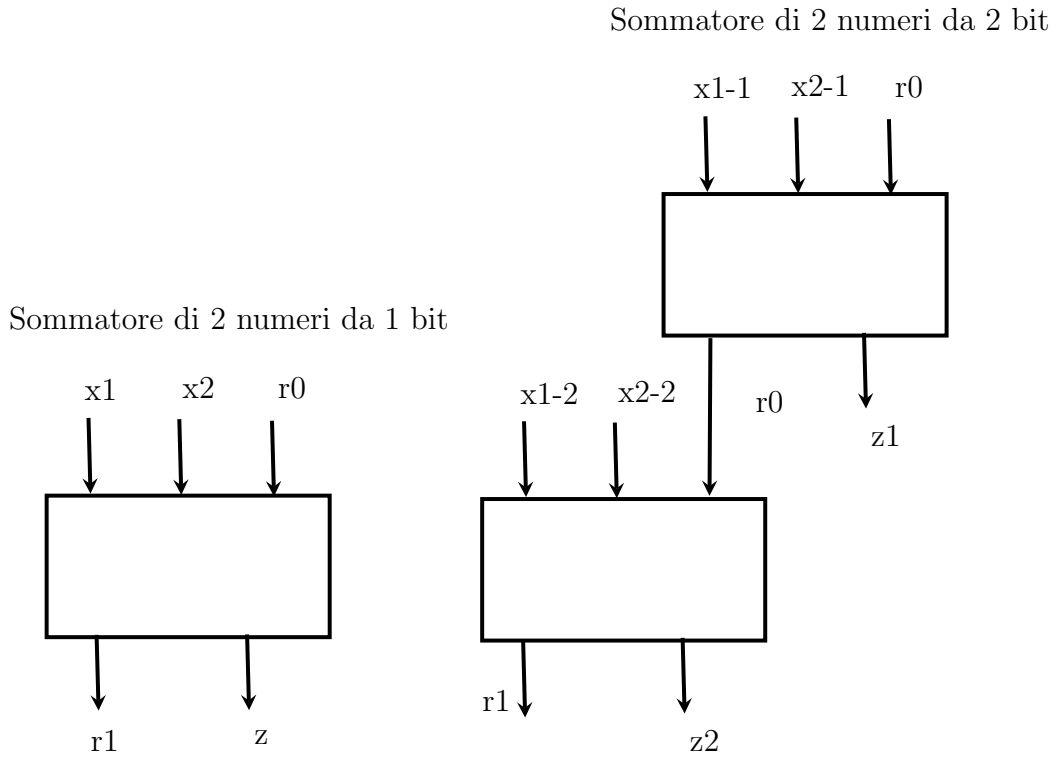


Figura 4.7: Sommatore di due numeri

Per ovviare al problema possiamo realizzare  $n + 1$  tabelle di verità (ognuna per ogni bit di input, più il riporto che indica se è avvenuto un overflow). Le tabelle di verità avranno  $2^{2n}$  righe. Ad esempio, per sommare due numeri da 4 bit avremo 1 livello AND e 3 livelli OR perché usiamo porte che accettano al massimo 8 entrate. Ciò impiegherà (per numeri da 4 bit)  $4\Delta t$  a differenza della somma con Full Adder a cascata che impiega  $8\Delta t$ .

Se volessi sommare numeri da 8 bit, la situazione varia leggermente. Servono 2 livelli di AND, e per esprimere la tabella di verità servono al minimo  $2^{15}$  righe. I livelli di OR saranno quindi  $\lceil \frac{\log_2 2^{15}}{\log_2 2^3} \rceil = \lceil \frac{15}{3} \rceil$ .

**ALU con sottrattore** Possiamo realizzare la sottrazione in questo modo (funziona anche per i numeri negativi se utilizziamo il complemento a due):

1. Vogliamo sottrarre un numero  $B = 4 \equiv 00000100$  ad un altro numero  $A = 12 \equiv 00001100$ :  
 $A - B = 8 \equiv 00001000$
2. Invertiamo B e aggiungiamo 1:  $(\bar{B} + 1) \equiv 11111100$
3. Otteniamo che  $A - B \equiv A + (\bar{B} + 1) \equiv 000001000$  (non va considerato l'overflow)



Un possibile (ma errato) circuito che realizza sottrazione e addizione può essere il seguente:

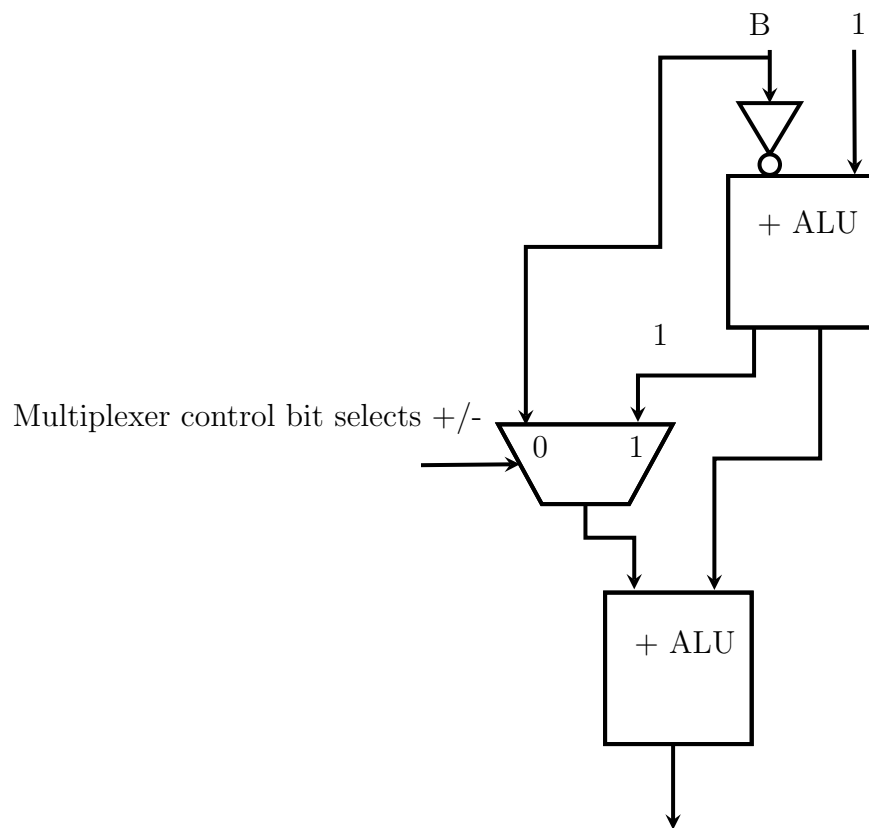


Figura 4.8: ALU per addizione e sottrazione

Si può semplificare molto il circuito utilizzando un ALU per la somma che accetta un riporto iniziale come i full adder a cascata.

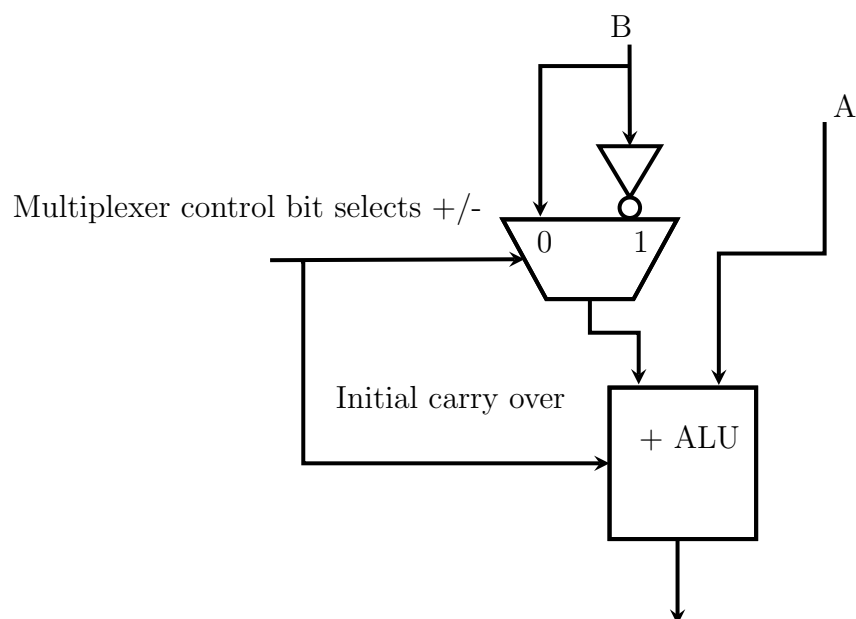


Figura 4.9: ALU per addizione e sottrazione semplificata

**Uscite aggiuntive delle ALU o flag** Le ALU, oltre all'uscita del risultato possono avere delle uscite aggiuntive, dette **flag**. I flag più comuni sono il flag **Zero**, il flag **Negative** ed il flag **Carry**.



Listing 4.21: Half Adder in Verilog

```

1 module ha(output c, output z, input x, input y);
2
3     assign
4         c = x&y;
5
6     assign
7         z = ~x&y | x&~y;
8
9 endmodule //fa

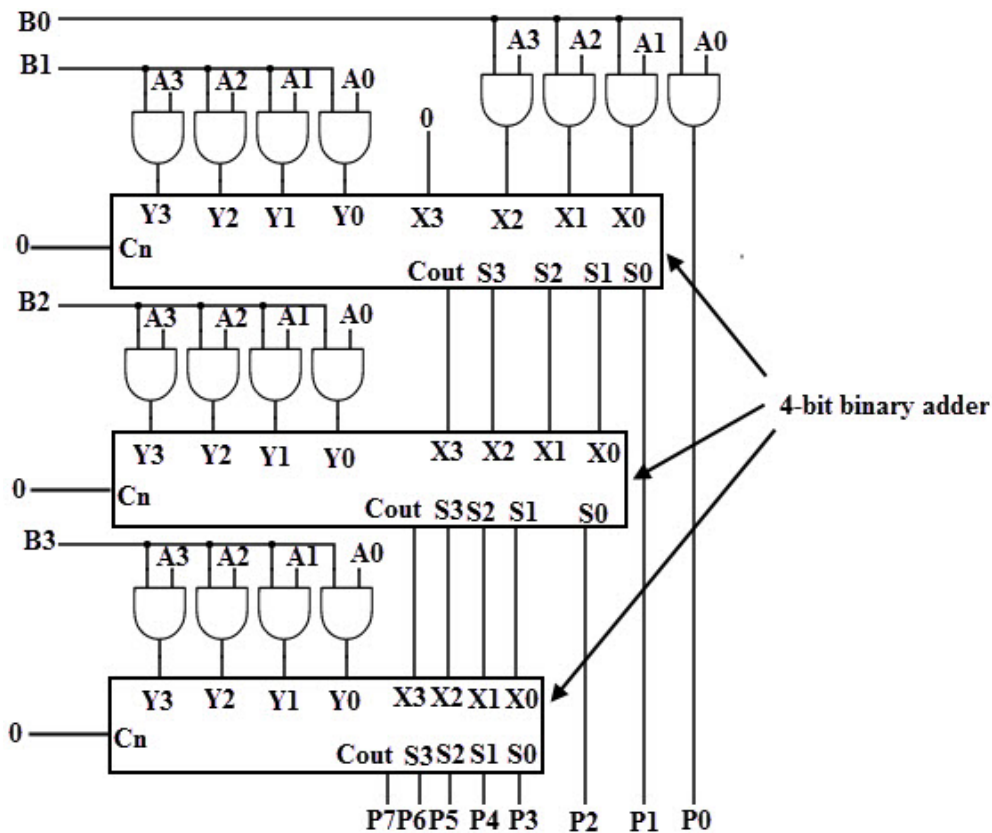
```

Listing 4.22: Full Addder in Verilog

```

1 module fa(output c, output z, input x, input y, input r);
2
3     assign
4         c = ~x&y&r | x&~y&r | x&y&~r | x&y&r;
5
6     assign
7         z = ~x&~y&r | ~x&y&~r | x&~y&~r | x&y&r;
8
9 endmodule //fa

```

Figura 4.10: Circuito di un moltiplicatore interno ad una ALU di due numeri da 4 bit  $A$  e  $B$

Listing 4.23: mul4bit.v

```

1 module mul(output [7:0]z, // 8 bit number result
2   input [3:0] x, // 2 4-bit numbers input
3   input [3:0] y);
4
5   wire [11:0] c; // carry wire
6   wire [6:1] w; // adder result wire
7
8   assign
9     z[0] = x[0]&y[0]; // first bit of output (least significative)
10    // is just AND between the first bits of the two inputs
11
12    // first row of adders
13    //carryout  output  addend1      addend2      carryin
14    ha h1(c[0],    z[1],    x[1]&y[0],    x[0]&y[1]); // second bit of output
15    fa f2(c[1],    w[1],    x[2]&y[0],    x[1]&y[1],    c[0]);
16    fa f3(c[2],    w[2],    x[3]&y[0],    x[2]&y[1],    c[1]);
17    ha h4(c[3],    w[3],    x[3]&y[1],    c[2]);
18
19    // second row of adders
20    ha h5(c[4],    z[2],    w[1],        x[0]&y[2]); // third bit of output
21    fa f6(c[5],    w[4],    w[2],        x[1]&y[2],    c[0]);
22    fa f7(c[6],    w[5],    w[3],        x[2]&y[2],    c[1]);
23    fa f8(c[7],    w[6],    c[3],        x[3]&y[2],    c[6]);
24
25    // third row of adders
26    ha h9(c[8],    z[3],    w[4],        x[0]&y[3]); // third bit of output
27    fa f10(c[9],   z[4],    w[5],        x[1]&y[3],    c[8]);
28    fa f11(c[10],  z[5],    w[6],        x[2]&y[3],    c[9]);
29    fa f12(z[7],   z[6],    c[7],        x[3]&y[3],    c[10]);
30    // on last bit, carry out is the last bit of output
31
32 endmodule //mul4bit

```

# Capitolo 5

## Memorie e Parallelismo

### 5.1 Memorie

Un componente fondamentale di un processore è la **memoria**. Si implementa utilizzando dei registri da  $n$  bit (dette parole). Vogliamo realizzare una memoria il cui accesso sia simile ad un array nel linguaggio C. Ovvero vogliamo avere un vettore di registri a cui possiamo accedere con un offset intero. Data una memoria  $A$  si accederà alla parola  $k$  – *esima* tramite  $A[k]$ . Per implementare tale selezione si può utilizzare un multiplexer che riceve in input le uscite dei singoli registri. I singoli registri ricevono tutti un segnale di clock in input, e permettono la scrittura attraverso un demultiplexer (decoder) che dato un indirizzo invia un 1 all'enable del registro selezionato. Si può mettere in AND il segnale enable del decoder insieme ad un input enable generale.

L'implementazione reale delle memorie (in particolare memorie flash), avviene ponendo delle linee dette "di parola" (orizzontali) e linee dette "bit line" (verticali) in una matrice. All'incrocio delle linee vi è un condensatore. Il condensatore mantiene una carica per un breve periodo di tempo. Ho modo di controllare se nel condensatore alla posizione  $xy$  era "ricordato" un valore 1, mandando della corrente lungo la "linea di parola"  $x$ . Se il condensatore era carico esso si scarica lungo la linea "bit line"  $y$ . Inviando della corrente sulla linea di parola  $x$  leggeremo quindi dalle bit line (in alcuni casi indicate come source lines in output) la parola  $x$ . Prima di un condensatore, si inserisce un transistor nella giunzione fra WL e BL. Utilizzando un condensatore, nelle RAM dinamiche (Dynamic Random Access Memory) possono insorgere problemi dovuti ai tempi di carica e scarica del condensatore. Una volta letto un valore su una Bit Line, il condensatore si scarica e "perde" il valore precedentemente "ricordato".

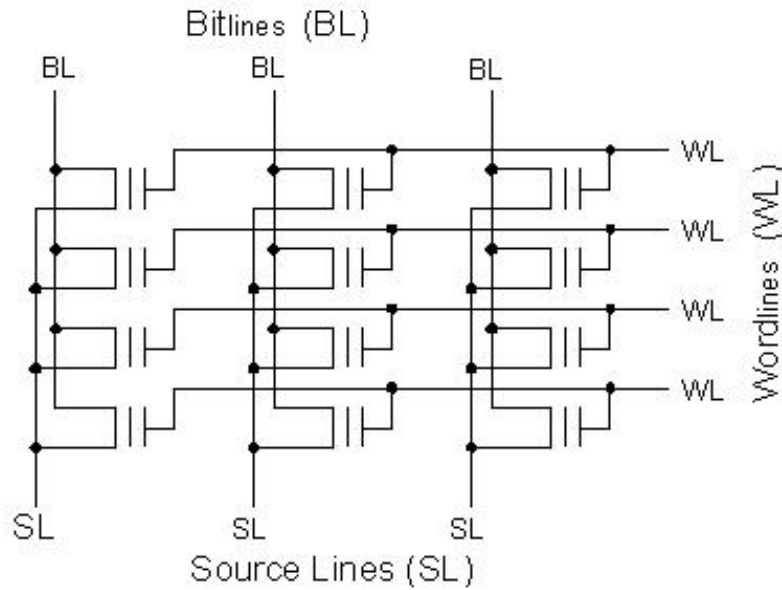


Figura 5.1: Memoria Flash

Per selezionare la Word Line utilizziamo sempre un demultiplexer (decoder) che riceve un indirizzo. Nelle ram DDR comuni sono presenti spesso 8 o 9 chip. Ad esempio, su uno stick di ram DDR3 da 1GB (1 Giga Byte) sono presenti 8 memorie da 1Gb (1 Giga Bit), e generalmente anche una nona memoria che ricorda la "parità" per il controllo degli errori. Le RAM dinamiche perdono il loro contenuto se non sono alimentate. Visto il comportamento dei condensatori (la loro carica si perde lentamente) le DRAM hanno bisogno di un **refresh** periodico. Le DRAM sono molto economiche (1 transistor per 1 bit) ma possono risultare più lente rispetto alle alternative SRAM. Le memorie SRAM (Static RAM) sono meno economiche ma più rapide rispetto alle DRAM. Utilizzano flip-flop invece dei condensatori e per ogni bit utilizzano 6 transistor. Le memorie all'interno del processore (calcolatore) sono memorie statiche (realizzate con flip-flop) molto rapide. Un meccanismo sia a livello hardware che a livello di S.O. permette la suddivisione dei dati processati fra vari livelli di cache, dalla memoria statica del processore alla più capiente ma lenta DRAM.

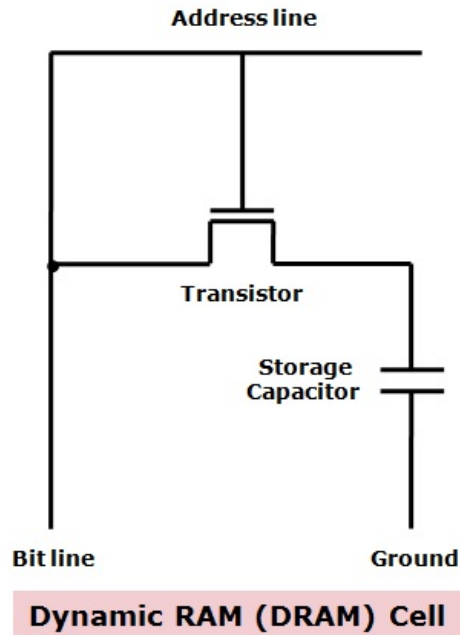


Figura 5.2: Schema di una cella di una memoria DRAM

**Definizione 5.1.1. RAM, ROM, PROM e EEPROM:** L'acronimo RAM (Random Access Memory) indica in generale delle memorie sulle quali possiamo leggere e scrivere. Le memorie ROM sono memorie a sola lettura (Read Only Memory). Le memorie ROM non hanno condensatori nelle giunzioni della matrice, ma a momento di produzione vengono "incisi" dei collegamenti o isolamenti nelle giunzioni fra Word Line e Bit Line. Le memorie **PROM** utilizzano dei fusibili nelle giunzioni fra WL e BL per permettere agli acquirenti di scrivere nella memoria una volta sola. Di fabbrica tutti i bit di una PROM sono impostati a 1. Bruciando i fusibili nelle giunzioni si interrompono i collegamenti realizzando una vera e propria scrittura permanente.

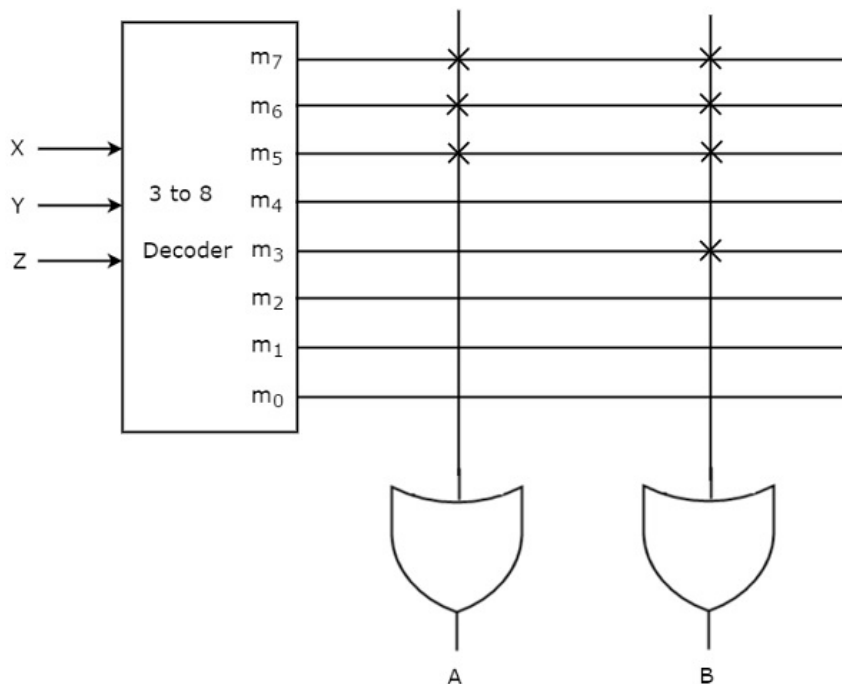


Figura 5.3: Esempio di memoria PROM

Un altro tipo di memorie sono le **EEPROM** (Electrically Erasable ROM). Utilizzano dei transistor particolari che possono essere "reimpostati" permettendo di cancellare e riscrivere la memoria con una corrente maggiore.

Nella seconda parte del corso vedremo dettagliatamente Assembly ARM. Le istruzioni Assembly eseguono delle operazioni sui registri interni alla CPU. Un esempio di istruzione banale è la somma di numeri fra registri. `ADD R1 R2 R3`. Nell'Assembly ARM la lettura avviene da destra a sinistra, ed il risultato viene memorizzato nel registro più a sinistra, in questo caso R1.

**Memorie Multi Porta e Tempi di Accesso** Le memorie multiporta permettono la lettura parallela di due o più registri. Invece di un singolo multiplexer per leggere un registro, alle uscite dei registri è collegato un altro multiplexer che ricevendo un indirizzo diverso dal primo multiplexer, permette di leggere un altro valore dalla memoria, abilitando alla lettura parallela di due registri contemporaneamente. È fondamentale avere una memoria con almeno due porte per eseguire istruzioni su due registri alla volta. Indichiamo i tempi di accesso alla memoria con  $T_a$ . Ad oggi siamo in un range di tempi inferiori ad un nanosecondo per i tempi di accesso ai flip-flop (poche decine di picosecondi), sull'ordine dei nanosecondi per le SRAM (cache) e sull'ordine dei microsecondi per le DRAM.

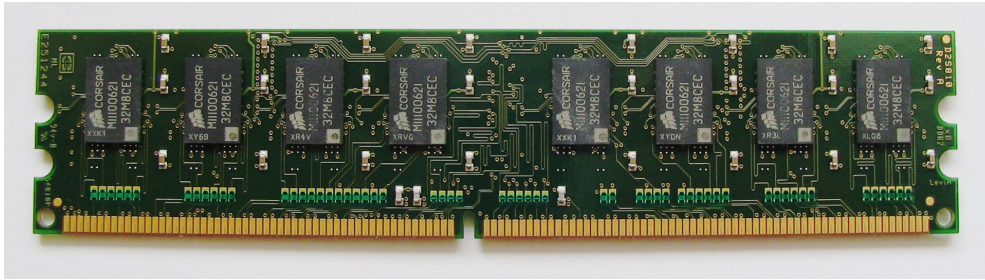


Figura 5.4: Modulo RAM Corsair DDR2-533

**Blocco memoria** Introduciamo un blocco per la memoria da utilizzare nei diagrammi dei circuiti. La memoria riceve in input un dato  $x$  da scrivere nell'indirizzo indicato dall'input di controllo **Address**. Se il segnale **enable** è HIGH allora avviene la scrittura di  $x$  all'interno dell'indirizzo puntato. Altrimenti, se **enable** è LOW avviene la lettura. Assumiamo che l'uscita sia stabile per il ciclo di clock

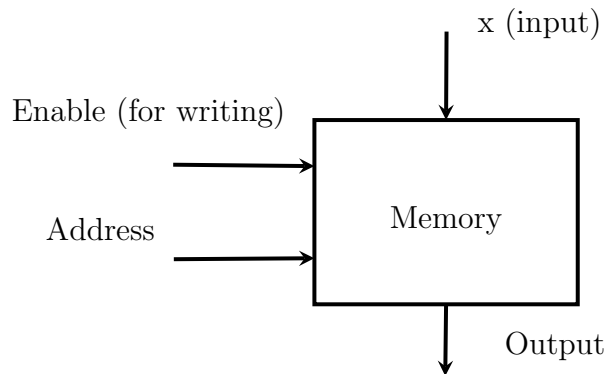


Figura 5.5: Blocco Memoria

**LUT (Lookup Table)** Una rete combinatoria che riceve  $n$  bit in input può essere implementata con una memoria ROM. Data una tabella di verità di una rete combinatoria, si può portare



la tabella di verità in forma vettoriale per creare una LUT. Una LUT o **Lookup Table** è in generale un array che rimpiazza una computazione, possibilmente complessa, con un'operazione di accesso, nel caso delle reti combinatorie con una memoria ROM. I bit in input della tabella di verità vengono portati a bit rappresentanti l'indirizzo di accesso della memoria, e i risultati della tabella di verità saranno memorizzati in un vettore di  $2^n$  elementi (se la tabella di verità ha un solo output). Per quanto alcune reti combinatorie possono risultare più complesse a livello di circuito se implementate con una LUT, un vantaggio non banale delle Lookup Tables è che possono essere riprogrammate se viene utilizzata una memoria EEPROM. Ciò svolge un ruolo fondamentale negli FPGA. Le LUT possono essere utilizzate in parallelo.

**Esempio 5.1.1. Rete Sequenziale di un contatore** Se volessimo realizzare un contatore, con le sole operazioni di incremento e decremento con un ASF, dovremmo avere un numero di stati pari alla capacità massima del contatore. Per realizzare un contatore sono necessari un registro da  $n$  bit, (prendiamo ad esempio 8 bit) ed una ALU che fa due sole operazioni (+1 e -1), quindi con un solo bit di controllo. Scegliamo se aumentare o decrementare il contatore impostando il bit di controllo della ALU, e scegliamo se scrivere nel prossimo ciclo di clock impostando ad HIGH il bit Enable (Event) del registro. Essendo una rete sequenziale, la funzione  $\sigma$  è implementata dalla ALU, e la funzione  $\omega$  è semplicemente l'identità. È una rete di Moore poiché la funzione identità non dipende dagli ingressi (Inc/Dec ed Event).

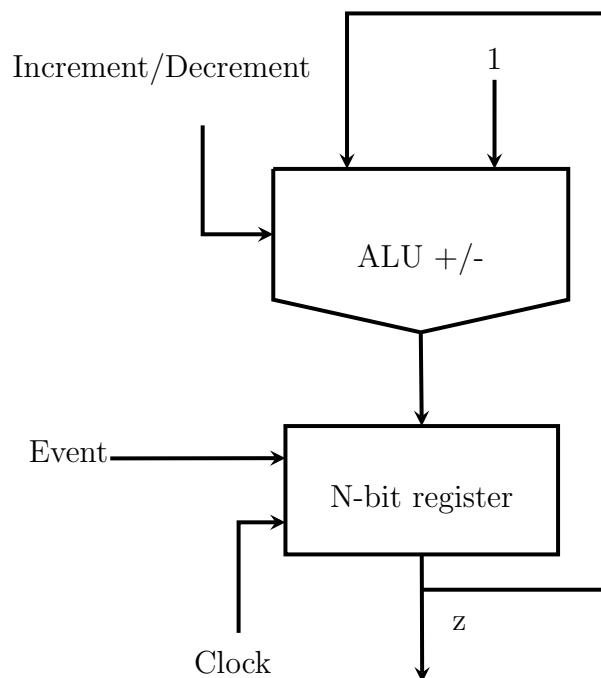


Figura 5.6: Rete sequenziale di un contatore

**Esempio 5.1.2. Semplice Calcolatrice** Vogliamo realizzare una calcolatrice molto semplice da 32 bit per eseguire addizione, sottrazione, moltiplicazione e divisione. La calcolatrice riceverà in input due numeri  $A, B$  e restituirà in output il risultato, che potrà essere re-inserito nei due registri contenenti  $A$  e  $B$ . Essa è una rete sequenziale che possiamo inserire in un componente che avrà in input i due numeri  $A, B$ , la selezione dei due multiplexer per gli input, segnali di enable per la scrittura nel registro  $A$  e nel registro  $B$  e due bit di controllo per selezionare l'operazione. In output avrà il numero in uscita e 4 bit di flag (Overflow, Carry, Negative e Zero). Definiamo questa parte del calcolatore "parte operativa". Abbiamo bisogno anche di una "parte di controllo" che dati due input si occuperà di controllare le entrate di controllo della parte operativa della calcolatrice.

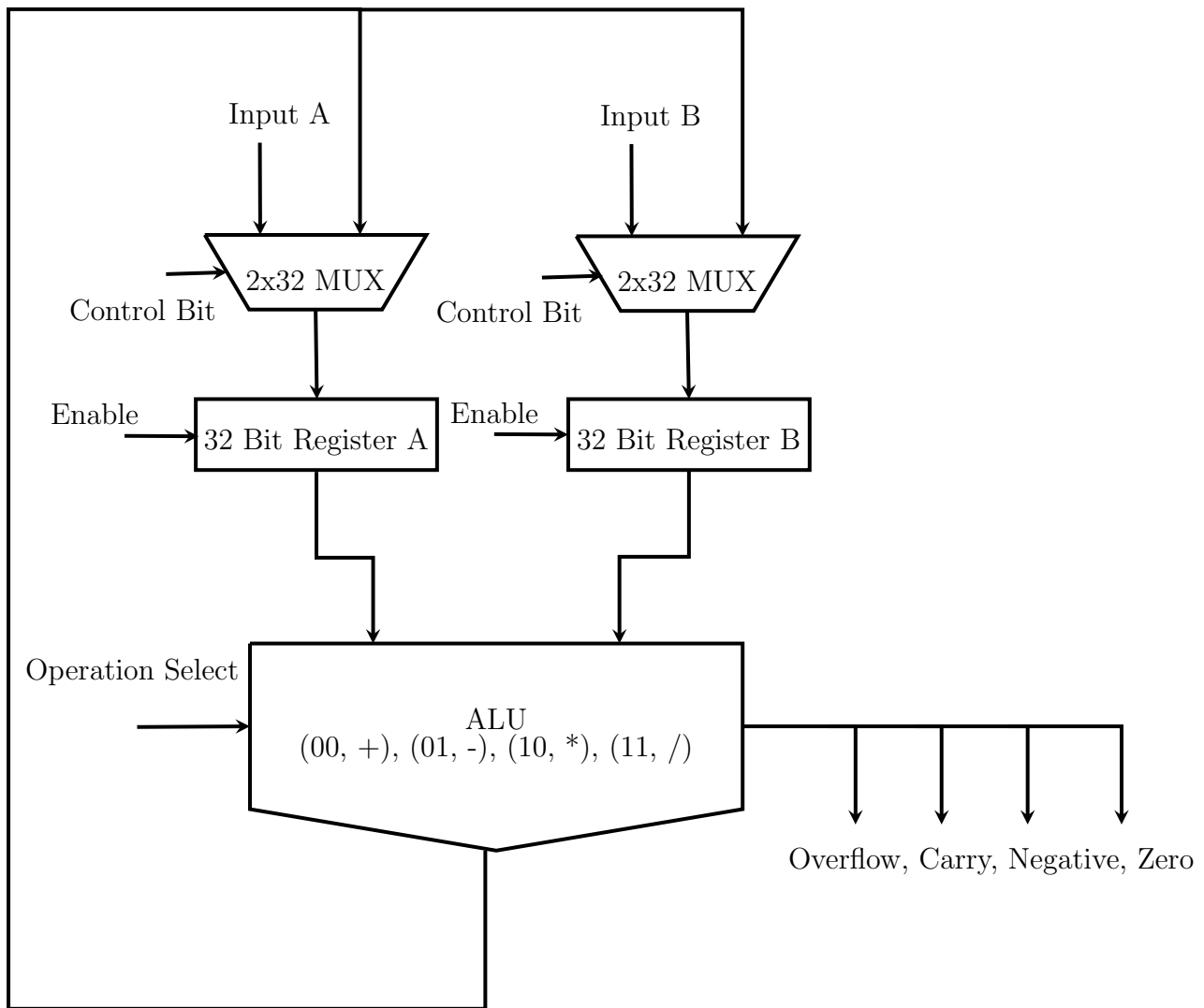


Figura 5.7: Semplice Calcolatore (Omesso clock per semplicità del disegno, è presente)

**Definizione 5.1.2. if-then-else:** Supponiamo di voler realizzare un importante costrutto per la programmazione (*if-then-else*) utilizzando solo componenti standard. Abbiamo due reti combinatorie  $f, g$ , una parola memorizzata in un registro  $A$  e un registro  $B$  in cui sarà memorizzato il risultato. Vogliamo costruire una rete per valutare l'espressione:

```

if(expr) then
    g(A) -> B
else
    f(A) -> B

```

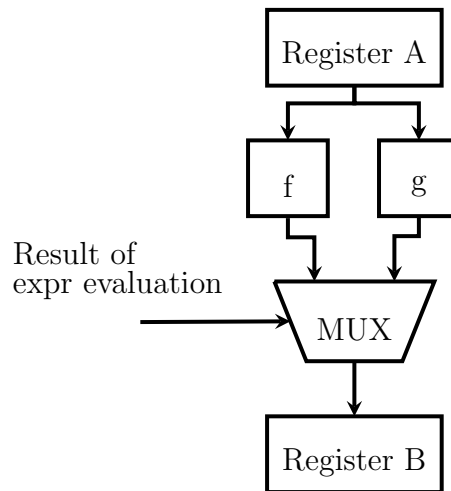


Figura 5.8: Rete che implementa if-then-else

**Definizione 5.1.3. while:** Se volessimo invece implementare un ciclo while della forma:

```
while(expr) do
  f(A) -> A
```

Possiamo realizzarlo con la seguente rete:

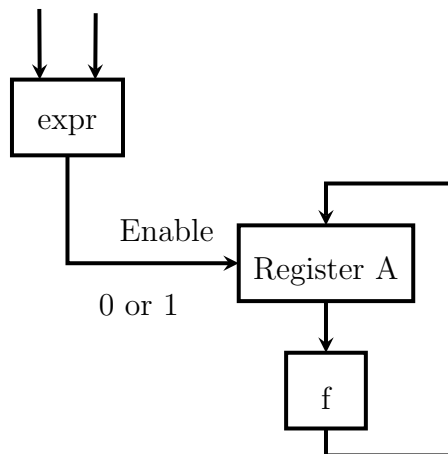


Figura 5.9: Rete che implementa while

In maniera simile è possibile realizzare anche un costrutto switch-case. Icarus Verilog, per trasformare programmi Verilog di moduli behavioural in netlist di componenti di reti logiche utilizza meccanismi simili.

### Esercizio 5.1.3. Confronto maggiore

Vediamo ora diversi metodi per realizzare una rete combinatoria che controlla se un numero è maggiore di un altro, utile per realizzare operazioni condizionali.

Listing 5.1: Esercizio di confronto maggiore in Verilog (approccio algoritmico)

```

1 module maggiore(output reg gt, input [N-1:0] x, input [N-1:0] y)
2   parameter N = 2;
3   always @(x or y)
4     begin
5       if(x[1] > y[1])
6         gt <= 1;
7       else
8         begin
9           if(y[1] > x[1])
10            gt <= 0;
11          else
12            begin
13              if(x[0] > y[0])
14                gt <= 1;
15              else
16                gt <= 0;
17            end
18          end
19        end
20    endmodule //maggiore

```

### 5.1.1 Calcolatrice con registri di memoria

Realizziamo adesso una calcolatrice con sottrazione e addizione che permette di leggere due numeri in input e fare operazioni utilizzando anche numeri precedentemente memorizzati in degli altri registri. Abbiamo bisogno di risorse di calcolo (una ALU) e registri o memorie. Abbiamo bisogno di una memoria multiporta particolare per realizzare correttamente il circuito. Questa memoria accetta 3 indirizzi, i primi due sono per la lettura (ha due parole in output) e il terzo indirizzo è per la scrittura. È una rete sequenziale, per questo motivo abbiamo bisogno di un meccanismo per evitare i problemi di temporizzazione. Essendo una rete sequenziale il tempo di esecuzione sarà  $\tau = \delta + \max\{t_\sigma, t_\omega\}$ . Dovremmo prendere un tempo del ciclo di clock  $\tau = \delta + \max\left\{t_{\text{MUX}} + t_{\text{ALU}}, t_{\text{ALU}} + t_{\text{MUX}}\right\}$  dove  $t_\sigma = t_{\text{MUX}} + t_{\text{ALU}}$  e  $t_\omega = t_{\text{ALU}} + t_{\text{MUX}}$

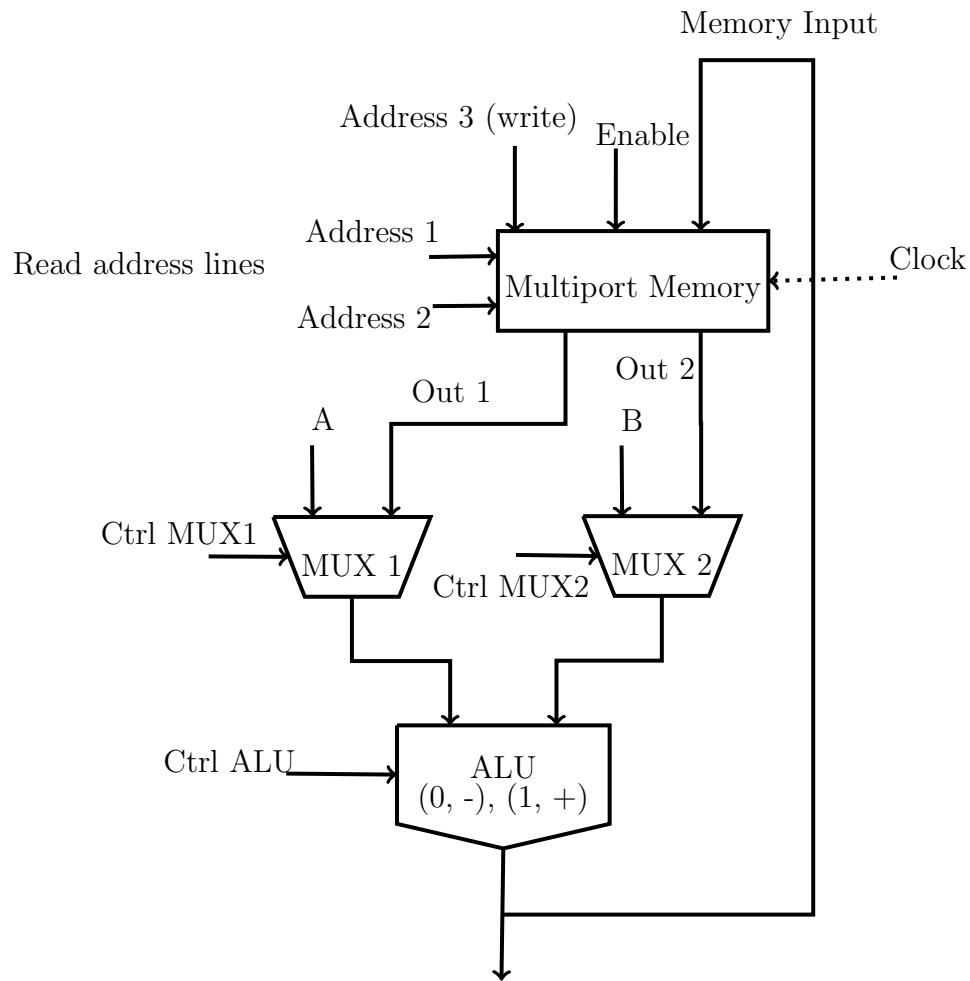


Figura 5.10: Calcolatrice con memoria.

## 5.2 Parallelismo

**Definizione 5.2.1.** Supponiamo di avere una serie di libri  $L_1, L_2, L_3, L_4$  da tradurre. Vogliamo ridurre il tempo totale di traduzione di tutti i libri. Se a tradurre un libro impiego un tempo da  $t_0$  a  $t_1$ , suddividendo la traduzione a metà fra due persone, impiegheremo  $t_{1/2}$  a tradurre il libro. L'obiettivo del parallelismo è ridurre la **latenza**. Se ho un insieme di libri, la latenza  $L$  di un libro è il tempo necessario per tradurlo (il tempo a cui ho terminato la traduzione, meno il tempo in cui ho iniziato la traduzione). Se ho  $m$  task da completare, il **tempo di completamento**  $T_c = m \cdot L$  è il tempo totale necessario a completare tutti i task  $m$ . Con il parallelismo vogliamo diminuire il tempo di completamento. Con **tempo di servizio** intendiamo l'intervallo medio di tempo fra la "consegna" di due risultati successivi. Definiamo anche **throughput** come il numero di calcoli eseguiti per unità di tempo.

Data una lista di istruzioni  $i_0, i_1, i_2, i_3, \dots$  il nostro processore esegue (in pseudocodice):

```
while(true) {
  leggi un istruzione (ASM)
  interpreta
  esegui
}
```

**Definizione 5.2.2. Tempo ideale:** Dato  $n$  = numero di worker (o grado di parallelismo) e  $T_{\text{seq}}$  = il tempo sequenziale, il **tempo ideale**  $T_{\text{id}}(n) = \frac{T_{\text{seq}}}{n}$ .

**Definizione 5.2.3. Speedup:** Definiamo anche lo **speedup** come  $\frac{T_{\text{seq}}}{T(n)}$ , ovvero il miglior tempo sequenziale fratto il tempo parallelo con grado di parallelismo  $n$ .

**Definizione 5.2.4. Scalabilità:**  $\text{Scalab}(n) = \frac{T(1)}{T(n)} \leq n$

**Definizione 5.2.5. Efficienza:**  $\text{Efficienza}(n) = \frac{T_{\text{id}}}{T(n)} = \frac{T_{\text{seq}}/n}{T(n)} = \frac{T_{\text{seq}}}{n \cdot T(n)} \leq 1$

**Definizione 5.2.6. Parallelismo temporale o Pipelining:** Il processore deve sempre eseguire dei task (istruzioni) sequenzialmente per eseguire un programma, e si può diminuire la latenza dei singoli task per ridurre il tempo di esecuzione del programma.

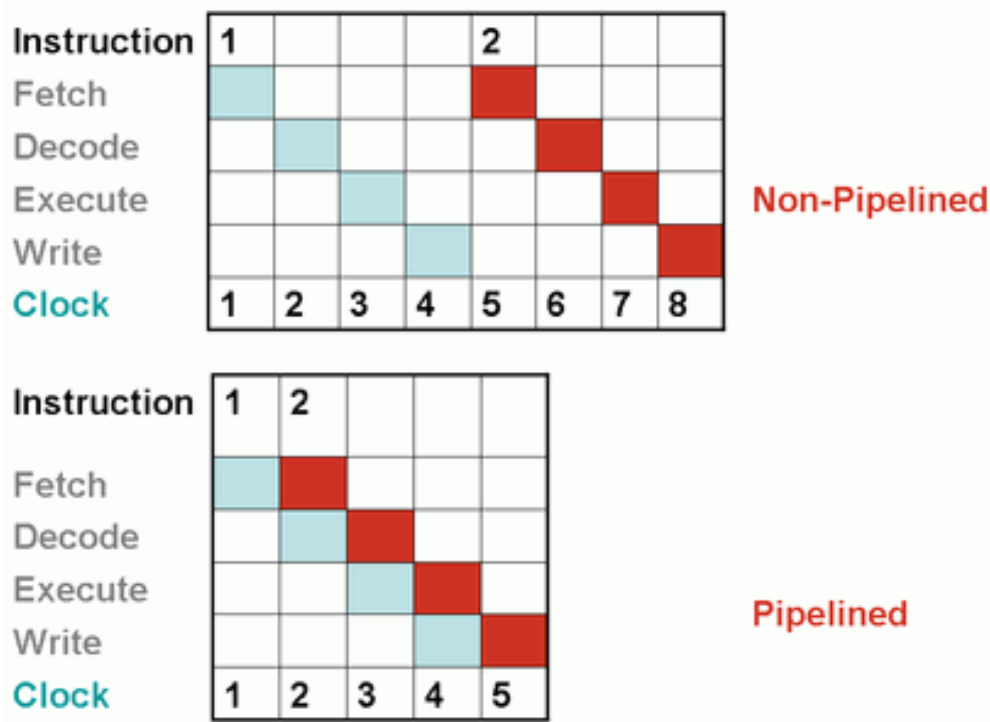


Figura 5.11: Diagramma temporale del pipelining di una CPU.

Avremo che  $T_c$  = tempo iniziale di "riempimento" della pipeline + tempo ( $\propto (m \cdot \text{tempo delle singole istruzioni})$ )

In alternativa, il nostro processore può assegnare un "ciclo" ad ogni istruzione

```
// Per istruzione 0
while(true) {
  leggi, decodifica, esegui
}
```

```
// Per istruzione 1
while(true) {
```

```
leggi, decodifica, esegui
}
```

```
// Per istruzione 2...
while(true) {
leggi, decodifica, esegui
}
```

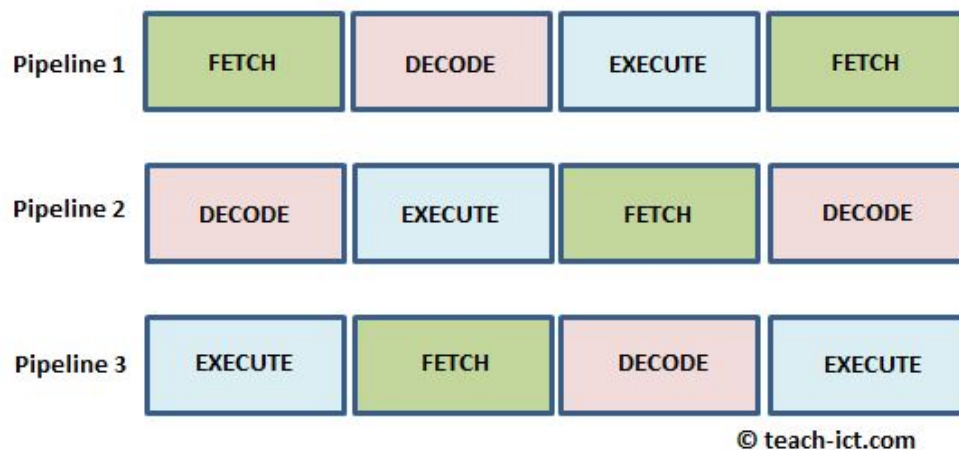


Figura 5.12: Pipelining con più worker (parallelismo temporale).

Usando il **parallelismo temporale**, ad ogni periodo  $L$  otterremo 3 risultati. Ciò riduce il tempo di servizio a  $\frac{1}{3}L$ . Le istruzioni dei task sono comunque eseguite sequenzialmente su worker separati.

**Definizione 5.2.7. Parallelismo spaziale:** Nel **parallelismo spaziale**, suddividiamo le istruzioni di ogni task su worker diversi. Significa che i worker possono eseguire in parallelo le istruzioni di un singolo task. Il parallelismo spaziale serve a diminuire la latenza, il parallelismo temporale serve invece ad aumentare il throughput.

**Pipeline (stream parallel)** Avendo 3 worker ed eseguendo 3 istruzioni a cascata che impiegano rispettivamente  $2t, 3t, 1t$ , se eseguiamo dei task successivamente, le istruzioni dei task successivi avverranno con un delay dovuto alla differenza del tempo di esecuzione. Dopo un numero di istruzioni le istruzioni si sincronizzeranno. Se i worker impiegano rispettivamente dei tempi  $t_1, t_2, t_3$ , il tempo di completamento di  $m$  istruzioni sarà  $T_c = \sum t_i + (m + 1) \cdot \max\{t_1, t_2, t_3\}$ . Ciò crea una situazione non ottimale

$$\begin{aligned}
 n &= \text{numero di stadi della pipeline} \\
 T_s &= \max\{T_{Si}\} \\
 \max(\text{speedup}(n)) &= \# \text{ di stadi della pipeline} \\
 T_c(n) &= \sum_i L_i + n - 1T_s
 \end{aligned}$$



Figura 5.13: Processore senza pipeline (sottoscalare)



Figura 5.14: Processore con pipeline a cinque stadi (scalare)

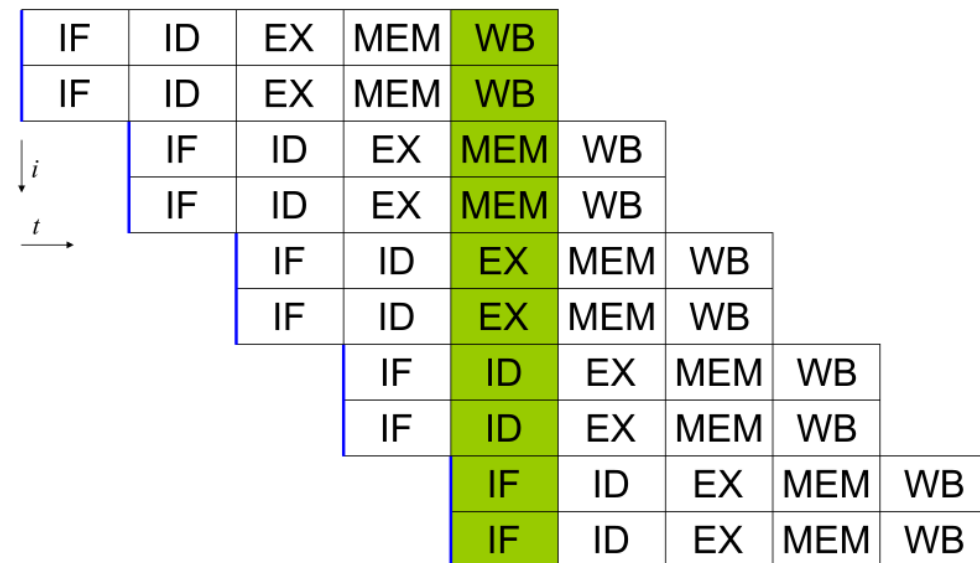


Figura 5.15: Processore con pipeline superscalare

**Map e Farm** Farm è di tipo *stream parallel* mentre map è di tipo *data parallel*. Nel caso del **Map** (e anche del Farm), se le istruzioni impiegano un tempo  $t$  che viene precisamente suddiviso in  $t/2$  allora il problema non sussiste. Se invece vengono suddivise in tempi diversi si creano spazi temporali in cui non è possibile sfruttare tutta la potenza di calcolo. Nella map posso suddividere un dato in arrivo e fare del calcolo in parallelo su di esso (*data parallelism*). Nella farm e nella pipeline i dati in arrivo possono essere passati a diversi worker (verranno schedulati) che eseguono computazioni in parallelo (*stream parallelism*). I dati non possono essere suddivisi nello *stream parallelism* come nel *data parallelism*.



### Tempi della Farm

$nw$  = grado di parallelismo

$T_w$  = il tempo che 1 worker impiega per calcolare 1 risultato

$$T_s = \max \left\{ T_{\text{sched}}, \frac{T_w}{nw}, T_{\text{coll}} \right\}$$

### Tempi della Map

$$T_c = T_{\text{split}} + \frac{m}{nw} T_f + T_{\text{merge}}$$

$$T_s = \max \left\{ T_{\text{split}}, T_{\text{merge}}, \frac{m}{nw} t_f \right\}$$

#### 5.2.1 Processore superscalare

Scriviamo un loop di esecuzione del processore in maniera più accurata.

```
while(true) {
I = fetch(PC) // PC = program counter
decode(I)
exec(I)
|--> update(PC)
}
```

**Composizione nel parallelismo** Possiamo comporre i diversi metodi di parallelismo. Ad esempio si può realizzare una pipeline che contiene una farm. Prendiamo ad esempio

$$\text{pipe}(S_1, \text{farm}(S_2), S_3)$$

Un possibile schema dei worker della pipeline è

$$\rightarrow f \rightarrow \begin{cases} (\text{farm}) \\ g \\ g \end{cases} \rightarrow h$$

Calcoliamo i tempi

$$T_s = \max \{ T_{S_1}, T_{S_2}, T_{S_3} \}$$

$$\begin{aligned} T_s &= \max \left\{ T_f, \max \left\{ T_{\text{sched}}, \frac{T_g}{nw}, T_{\text{coll}} \right\}, T_h \right\} \\ &= \max \left\{ T_f, T_{\text{sched}}, \frac{T_g}{nw}, T_{\text{coll}}, T_h \right\} \end{aligned}$$

#### Definizione 5.2.8. Bottleneck o Colli di Bottiglia

Supponiamo di avere due computazioni composte  $f \circ g$ . Se  $f$  impiega  $5t$  e  $g$  impiega  $2t$ , eseguire in successione  $\rightarrow f \rightarrow g \rightarrow$  può causare dei bottleneck (rallentamenti). Se  $f$  è data parallel si può suddividere la sua computazione in parti uguali (ad esempio due) e poi passare il risultato a  $g$ . In alternativa si può dividere  $f$  in 5 parti uguali e ad ogni parte calcolare la composizione sequenziale (in una farm) in modo che  $T_s = \max \left\{ T_{\text{sched}}, \frac{T_w}{nw}, T_{\text{coll}} \right\} \approx \frac{T_w}{nw} \approx \frac{7t}{5} \approx 1$

#### Esempio 5.2.1. Pipeline in Verilog

Vediamo adesso una pipeline per parallelizzare l'operazione di moltiplicazione per 2 composta due volte.

Listing 5.2: Modulo moltiplicazione per 2

```
1 module per2(output [N-1:0]out, input [N-1:0]in);
2
3     parameter N = 8;
4
5     assign
6         #2 out = in*2;
7
8 endmodule // per2
```

Listing 5.3: Modulo moltiplicazione per 4

```
1 module per4(output [N-1:0]out, input [N-1:0]in, input clock);
2
3     parameter N = 8;
4
5     wire [N-1:0] pass;
6     reg [N-1:0] ingresso;
7
8     per2 p1(pass, ingresso);
9     per2 p2(out, pass);
10
11     always @(posedge clock)
12         ingresso = in;
13
14 endmodule // per2
```

Listing 5.4: Programma di test

```

1 module test();
2
3     parameter N = 8;
4
5     reg [N-1:0] in;
6     reg         clock;
7
8     wire [N-1:0] out;
9     integer     i;
10
11     per4 p4(out,in,clock);
12
13     always
14     begin
15         #3 clock = 1;
16         #1 clock = 0;
17     end
18
19     initial
20     begin
21         clock = 0;
22         in = 0;
23
24         $dumpfile("test.vcd");
25         $dumpvars;
26
27         #3 in=0;
28
29         for(i=1; i<9; i=i+1)
30             #4 in = i;
31
32         #10 $finish;
33
34     end
35 endmodule // test

```

Listing 5.5: Modulo moltiplicazione per 2 (pipelined)

```

1 module per2(output [N-1:0]out, input [N-1:0]in);
2
3     parameter N = 8;
4
5     assign
6         #2 out = in*2;
7
8 endmodule // per2

```

Listing 5.6: Modulo moltiplicazione per 4 (pipelined)

```

1 module per4(output [N-1:0]out, input [N-1:0]in, input clock);
2
3     parameter N = 8;
4
5     wire [N-1:0] pass;
6     stadio s1(pass, in, clock);
7     stadio s2(out, pass, clock);
8
9 endmodule // per2

```

Listing 5.7: Programma di test (pipelined)

```

1 module test();
2
3     parameter N = 8;
4
5     reg [N-1:0] in;
6     reg        clock;
7
8     wire [N-1:0] out;
9     integer     i;
10
11     per4 p4(out,in,clock);
12
13     always
14     begin
15         #1 clock = 1;
16         #1 clock = 0;
17     end
18
19     initial
20     begin
21         clock = 0;
22         in = 0;
23
24         $dumpfile("test.vcd");
25         $dumpvars;
26
27         #1 in = 0;
28
29         for(i=1; i<9; i=i+1)
30             #2 in = i;
31
32         #10 $finish;
33
34     end
35 endmodule // test

```

## Parte II



# Capitolo 6

## Assembler ARM

### 6.1 Introduzione all'Assembler ARM

Abbiamo visto un ciclo che viene eseguito dal processore.

```
while(true) {  
    preleva un istruzione (ASM)  
    interpreta  
    esegui  
}
```

**Definizione 6.1.1.** Prelevare un istruzione significa estrarre dalla memoria una parola (configurazione di 0 e 1) contenente un'istruzione **Assembly**. L'Assembly, spesso abbreviato **asm** è un linguaggio di programmazione a basso livello con una forte corrispondenza fra il linguaggio e le istruzioni in linguaggio macchina di una particolare architettura. Vedremo il linguaggio Assembler per ARM, architettura acronimo di Advanced RISC Machines, dove RISC sta per Reduced Instruction Set Computer. RISC utilizza molte meno istruzioni di CISC, più comune nelle architetture X86 e X86\_64. I principi di RISC sono:

1. La regolarità  $\implies$  semplicità
2. Supporto al caso più frequente
3. "Piccolo è bello"
4. Un buon risultato è frutto di un buon compromesso

Il linguaggio asm viene compilato a linguaggio macchina. asm è un linguaggio mnemonico più vicino possibile alle istruzioni macchina. Una volta compilato e caricato in memoria, le istruzioni di un programma Assembler risiedono in dei registri.

**Definizione 6.1.2.** Nell'architettura **Von Neumann** è presente un **Program Counter** (abbreviato con PC), un contatore (registro). Nel caso dell'architettura ARMv7 (32 bit) il PC è un registro generale in cui è abilitata la scrittura. In ARMv8 (64 bit) non è più possibile utilizzare il PC come un registro generale. Il contenuto del PC indica l'indirizzo di memoria dove risiede l'indirizzo dell'istruzione corrente eseguita dal ciclo del processore, e dev'essere aggiornato per contenere l'indirizzo della prossima istruzione alla fine del ciclo.

**Definizione 6.1.3. Tipi di istruzioni Assembler:** Vedremo istruzioni **aritmetico-logiche**, istruzioni di **load/store**, istruzioni di **controllo di flusso** ed alcune istruzioni **speciali**.

## 6.2 Istruzioni Assembler ARMv7

*Nota.* Il linguaggio Assembler ARM è radicalmente diverso dai linguaggi Assembler della famiglia X86! Vi sono diversi tipi di istruzioni ARM asm, vediamo i tipi di operandi (parametri) che le istruzioni primitive ricevono.

**Definizione 6.2.1. Sintassi delle istruzioni ARMv7:**

`<operazione>{cond}{flags} Rd,Rn, OP2`

- `<operazione>` è l'istruzione da eseguire
- `{cond}` Un codice di condizione di due lettere (opzionale).
- `{flags}` Flag opzionali aggiuntivi
- `Rd` Il registro di destinazione dell'operazione
- `Rn` Il primo registro sorgente
- `OP2` Un secondo operando flessibile (opzionale)

**Definizione 6.2.2. Registri ARMv7:** Nelle CPU ARMv7 sono presenti soltanto 16 registri da 32 bit, della forma `Rn` dove  $0 \leq n < 16$ .

- `R0` : Parametri temporanei o valori di ritorno. (Standard per i valori di ritorno della libreria standard C)
- `R1 - R3` : Argomenti, valori temporanei.
- `R4 - R12` : Valori temporanei.
- `R13` : Stack Pointer (SP).
- `R14` : Link register.
- `R15` : Program counter.

**Definizione 6.2.3. Operandi Immediati:** Operandi che non risiedono in dei registri ma sono costanti presenti all'interno della configurazione della parola, ad esempio `#1` è costante per il numero 1. `#0x1234` corrisponde in esadecimale al numero 4661

**Definizione 6.2.4. Operandi di Memoria:** Gli operandi di memoria sono della forma:

`[Rx, offset]`

Tale operando va a leggere il contenuto della memoria all'indirizzo puntato dal registro `Rx`, a cui va aggiunto l'`offset` costante

Ad esempio, un'istruzione per caricare una parola dalla memoria in un registro è

`LDR R0, [R2, #0]`

Tale istruzione caricherà nel registro `R0` il contenuto dell'indirizzo puntato dal contenuto di `R2` a cui va aggiunto 0. L'`offset` è utile per accedere a strutture dati (memorizzate sequenzialmente) come `struct` e `array`, presenti nel linguaggio C.



*Nota.* La memoria di ARM, come microarchitettura, è indirizzata a byte (gli indirizzi corrispondono a singoli byte). Ad esempio, quando carico un valore dalla memoria in un registro con

```
MOV R2, #0x400
LDR R0, [R2, #0]
```

All'interno del registro R0 saranno caricati i 4 byte a partire dall'indirizzo 1024 (400 in esadecimale).

**Definizione 6.2.5. Big Endian e Little Endian:** Se vogliamo caricare un numero da 4 byte (32 bit) in un registro dalla memoria, può sorgere il dubbio, in quale ordine vengono letti i byte? Nel metodo **Big Endian** carica partendo dal byte più significativo (quello più a sinistra), fino a quello meno significativo. Nel metodo **Little Endian**, il primo byte caricato è quello meno significativo (quello più a destra del registro/numero), mentre l'ultimo caricato è il byte più significativo. ARMv7 supporta entrambi i metodi di **endianness**.

#### Definizione 6.2.6. Istruzioni Logiche

Realizza l'AND bit per bit dei registri R1 (maschera) e R2 (sorgente), memorizzando in RD

```
AND RD, R1, R2
```

Realizza l'OR bit per bit dei registri R1 e R2, memorizzando in RD

```
ORR RD, R1, R2
```

Realizza l'Exclusive OR bit per bit dei registri R1 e R2, memorizzando in RD

```
EOR RD, R1, R2
```

Realizza l'Exclusive OR bit per bit dei registri R1 e R2, memorizzando in RD

```
EOR RD, R1, R2
```

Imposta a 0 i bit di R2 laddove nella posizione corrispondente della maschera R1 vi sia un 1, memorizzando il risultato in RD

```
BIC RD, R1, R2
```

**Definizione 6.2.7. Bit di flag in ARMv7:** In ARMv7 sono presenti 4 bit di flag (che possono essere alterati dopo l'esecuzione di un istruzione). Il flag **Z** indica se il risultato di un operazione è 0. Il flag **N** indica se il risultato dell'operazione è negativo. Il flag **C** indica se è presente un carry (riporto) sull'ultima cifra. Il flag **V** (overflow) indica se l'operazione ha dato overflow. I flag sono memorizzati in un registro detto "parola di stato", che non è accessibile come gli altri registri general purpose. Vi è presente anche un bit che indica l'endianness delle operazioni di memoria.

#### Definizione 6.2.8. Istruzioni Aritmetiche

Realizza la somma dei registri R1 e R2, memorizzando il risultato in RD

```
ADD RD, R1, R2
```

Realizza la somma dei registri R1 e R2 **con riporto**, memorizzando il risultato in RD

**ADC** RD, R1, R2

Realizza la sottrazione dei registri R1 - R2, memorizzando il risultato in RD

**SUB** RD, R1, R2

Realizza la sottrazione dei registri R1 - R2 **con riporto**, memorizzando il risultato in RD

**SBC** RD, R1, R2

Compare accetta solamente due operandi, realizza la sottrazione dei registri R1 - R2 ed **imposta i bit di flag**. Viene utilizzato per realizzare esecuzione condizionale nei programmi, laddove le istruzioni successive utilizzino un codice condizionale.

**CMP** R1, R2