



UNIVERSITÀ  
DI TRENTO

Dipartimento di  
Ingegneria e Scienza dell'Informazione

Dipartimento di Ingegneria e Scienza dell'Informazione

Corso di Laurea in  
Ingegneria Informatica, delle Comunicazioni ed Elettronica

# FONDAMENTI DI ELETTRONICA DIGITALE

Docente

Roberto Passerone

Studente

Cristiano Berardo 234428

Anno accademico 2023/2024



# Indice

<b>1 Il Silicio</b>	<b>6</b>
1.1 Materiali . . . . .	6
1.1.1 Struttura cristallina del silicio . . . . .	7
1.1.2 Densità di elettroni liberi . . . . .	7
1.2 Corrente di drift . . . . .	9
1.2.1 Velocità di saturazione . . . . .	10
1.2.2 Calcolo della corrente di drift . . . . .	11
1.3 Impurità . . . . .	12
1.3.1 Atomi pentavalenti . . . . .	12
1.3.2 Atomi trivalenti . . . . .	12
1.3.3 Conduzione . . . . .	13
1.3.4 Resistività in materiali estrinseci . . . . .	14
1.4 Corrente di diffusione . . . . .	15
1.5 Corrente totale . . . . .	15
1.6 Take away . . . . .	16
1.7 Processo costruttivo . . . . .	16
<b>2 Il Diodo</b>	<b>17</b>
2.1 Come è fatto . . . . .	17
2.2 Calcolo del potenziale di built-in $\phi_j$ . . . . .	18
2.3 Potenziale esterno . . . . .	19
2.3.1 Potenziale positivo . . . . .	19
2.3.2 Potenziale negativo . . . . .	20
2.4 Modello matematico del Diodo . . . . .	21
2.5 Modello di diodo ideale per grandi tensioni . . . . .	21
2.6 Diodo nei circuiti . . . . .	21
2.6.1 Rettificatore a semi onda . . . . .	22
2.6.2 Rettificatore ad onda intera, il ponte di diodi . . . . .	22
2.6.3 Raddrizzatore AC/DC con filtro capacitivo . . . . .	23
2.7 Modello di diodo ideale per piccole tensioni . . . . .	24
2.8 I circuiti logici realizzabili con i diodi . . . . .	24
2.8.1 Circuito AND . . . . .	25
2.8.2 Circuito OR . . . . .	25
<b>3 Dispositivi MOSFET</b>	<b>26</b>
3.1 Cos'è il condensatore MOS . . . . .	26
3.2 Come è fatto . . . . .	26
3.2.1 Regione di accumulazione . . . . .	26
3.2.2 Regione di svuotamento . . . . .	27
3.2.3 Regione di inversione . . . . .	27
3.3 Il transistore nMOS . . . . .	28
3.3.1 Simbolo circuitale . . . . .	28
3.3.2 Funzionamento qualitativo . . . . .	28
3.3.3 Calcoliamo il valore della corrente di drain $I_D$ . . . . .	30
3.4 Regione lineare, ohmica o triodo . . . . .	32
3.4.1 Se $V_{DS}$ è piccola . . . . .	32
3.5 Saturazione . . . . .	33
3.5.1 E quanto satura? . . . . .	34

3.6	Modulazione di lunghezza di canale . . . . .	35
3.7	Modello matematico del transistore nMOS . . . . .	35
3.8	Abbassare la tensione di soglia nel canale . . . . .	36
3.9	Transistore nMOS a svuotamento . . . . .	36
3.9.1	Effetto body . . . . .	36
3.10	Transistori pMOS . . . . .	37
3.11	Modello matematico del pMOS . . . . .	37
3.12	Riassunto . . . . .	38
3.13	Modello di analisi . . . . .	39
3.13.1	Calcolo con la retta di calcolo . . . . .	39
3.13.2	Come determinare l'uscita $V_{DS}$ in funzione dell'ingresso $V_{GS}$ . . . . .	41
3.14	Saturazione di velocità . . . . .	43
3.14.1	Conduzione sotto soglia . . . . .	44
<b>4</b>	<b>Invertitore Logico</b>	<b>45</b>
4.1	Funzione di trasferimento e caratteristica . . . . .	45
4.1.1	Schema di base dell'invertitore . . . . .	45
4.1.2	Interruttore con carico resistivo . . . . .	46
4.2	Caratteristica reale . . . . .	46
4.3	Margini di rumore . . . . .	47
4.4	Obiettivi di progetto . . . . .	48
4.5	Circuito invertitore nMOS . . . . .	48
4.5.1	Scelta di $V_L$ . . . . .	49
4.5.2	Dimensionamento . . . . .	49
4.5.3	Calcoliamo i margini di rumore: $V_{IL}$ e $V_{OH}$ . . . . .	50
4.6	Inverter con carico saturato . . . . .	52
4.6.1	Dimensionamento . . . . .	52
4.6.2	Calcolo di $V_H$ . . . . .	53
4.6.3	Margini di rumore: . . . . .	53
4.7	Carico a depletion mode . . . . .	55
4.7.1	Dimensionamento . . . . .	55
4.7.2	Margini di rumore: . . . . .	56
4.8	Inverter pseudo nMOS . . . . .	57
4.8.1	Dimensionamento . . . . .	57
4.8.2	Margini di rumore . . . . .	58
4.9	Confronto di inverter . . . . .	59
4.9.1	A carico resistivo . . . . .	59
4.9.2	A carico saturato . . . . .	59
4.9.3	A carico deplation/svuotamento . . . . .	60
4.9.4	A carico pseudo nMOS . . . . .	60
4.10	Impatto della saturazione di velocità . . . . .	61
4.11	Impatto della saturazione di velocità . . . . .	61
<b>5</b>	<b>Porte logiche</b>	<b>62</b>
5.1	Costruzione di porte logiche . . . . .	63
5.2	Porta logica: NOR . . . . .	64
5.3	Porta logica: NAND . . . . .	65
5.4	Dimensionamento dei transistori in fase di conduzione . . . . .	67
5.4.1	Per $M_A$ . . . . .	67
5.4.2	Per $M_B$ . . . . .	67
5.5	AND . . . . .	68
5.6	Porte logiche complesse . . . . .	69
5.7	Tempistiche . . . . .	70
5.7.1	Comportamento dinamico: i ritardi . . . . .	70
5.7.2	Capacità nel transistor MOS . . . . .	71
5.7.3	Transitorio dell'invertitore a carico resistivo . . . . .	71
5.7.4	Tempo di salita . . . . .	72
5.7.5	Tempo di propagazione . . . . .	72
5.7.6	Transitorio in discesa . . . . .	73
5.7.7	Transitorio complesso . . . . .	74

5.7.8	Transitorio, invertitore pseudo nMOS . . . . .	76
5.7.9	Confronto di inverter . . . . .	76
5.8	Consumo statico . . . . .	77
<b>6</b>	<b>Logica CMOS</b>	<b>78</b>
6.1	Inverter CMOS . . . . .	79
6.2	Vantaggi dei CMOS . . . . .	80
6.2.1	Ingresso zero . . . . .	81
6.2.2	Accensione dell'nMOS . . . . .	82
6.2.3	Aumentiamo ancora l'ingresso . . . . .	82
6.2.4	Zona triodo nMOS . . . . .	83
6.2.5	Zona cut-off pMOS . . . . .	83
6.3	Rapporto tra n e p MOS . . . . .	84
6.4	Margini di rumore . . . . .	85
6.4.1	$V_I = V_{IH}$ . . . . .	85
6.4.2	$V_I = V_{IL}$ . . . . .	85
6.5	Comportamento dinamico . . . . .	85
6.5.1	Transitorio in discesa . . . . .	85
6.5.2	Tempo di discesa: . . . . .	85
6.5.3	Tempo di propagazione da alto a basso dell'uscita: . . . . .	86
6.5.4	Transitorio in salita . . . . .	86
6.6	Osservazioni . . . . .	86
6.6.1	Ritardi proporzionali a K . . . . .	86
6.7	Dipendenza da $V_{DD}$ . . . . .	87
6.7.1	Teniamo conto di $V_{DD}$ . . . . .	87
6.8	Saturazione di velocità . . . . .	88
6.8.1	Calcolo di resistenza media . . . . .	88
6.9	Consumo di potenza statico . . . . .	89
6.10	Consumo di potenza dinamico . . . . .	89
6.11	Corrente di corto circuito . . . . .	92
6.12	Compromessi . . . . .	92
6.13	Costruzione di porte logiche . . . . .	93
6.14	Dimensionamento dei transistori . . . . .	95
6.15	Take away . . . . .	96
<b>7</b>	<b>Logica a pass transistor</b>	<b>97</b>
7.1	Livelli dei segnali . . . . .	97
7.2	Transmission gate . . . . .	99
7.3	Funzione che usa la logica transmission gate . . . . .	100
7.4	Considerazioni . . . . .	101
<b>8</b>	<b>Logica dinamica</b>	<b>102</b>
8.1	Due fasi . . . . .	102
8.1.1	Fase 1: precarica . . . . .	102
8.1.2	Valutazione . . . . .	103
8.2	Condizioni di funzionamento . . . . .	104
8.2.1	Proprietà della logica dinamica (pro's) . . . . .	104
8.2.2	Proprietà della logica dinamica (con's) . . . . .	104
8.3	Cascata di porte dinamiche . . . . .	104
8.3.1	Logica domino . . . . .	105
8.4	Altre logiche: take away . . . . .	105
<b>9</b>	<b>Latches, Flip-Flop e memorie</b>	<b>106</b>
9.1	Circuiti sequenziali . . . . .	106
9.2	Circuiti bistabili . . . . .	106
9.2.1	Realizzazione con due inverter . . . . .	107
9.2.2	Buffer invertente . . . . .	107
9.3	Elemento di memoria base . . . . .	108
9.4	Sostituiamo ciascuno degli inverter con una NOR . . . . .	109
9.5	Flip flop SR . . . . .	109

9.5.1	Funzionamento . . . . .	110
9.6	Flip flop SR con porte NAND . . . . .	110
9.7	Flip flop SR con clock . . . . .	111
9.8	Realizzazione ottimizzata . . . . .	111
9.8.1	Schema circuitale . . . . .	112
9.8.2	Funzionamento . . . . .	112
9.8.3	Dimensionamento . . . . .	112
9.9	Ritardo di commutazione . . . . .	113
9.10	Latch basato su multiplexer . . . . .	113
9.10.1	Caratteristiche: . . . . .	114
9.11	Flip flop di tipo Master Slave . . . . .	115
9.11.1	Circuito di memoria utilizzando flip flop edge trigger . . . . .	116
9.11.2	Tempo di set-up . . . . .	116
9.11.3	Tempo di propagazione . . . . .	116
9.12	Alternative realizzatible . . . . .	117
9.12.1	Rimuoviamo il transmission gate sul feedback . . . . .	117
9.12.2	Sostituire i transmission gate con pass transistor . . . . .	117
9.12.3	Usare logica dinamica . . . . .	118
<b>10</b>	<b>Memorie RAM</b> . . . . .	<b>119</b>
10.1	Cella statica - la cache . . . . .	120
10.1.1	Lettura della cella . . . . .	120
10.1.2	Scrittura della cella . . . . .	121
10.2	Cella RAM dinamica . . . . .	121
10.2.1	Lettura e Scrittura . . . . .	121
<b>11</b>	<b>Sense amplifier</b> . . . . .	<b>123</b>
11.1	Precarica e lettura . . . . .	123
11.2	Consumi . . . . .	124
11.3	Row Decoders . . . . .	124
11.4	Column Decoders . . . . .	125
<b>12</b>	<b>Memorie a stato solido</b> . . . . .	<b>126</b>
12.1	Memorie ROM . . . . .	126
12.2	Memorie Flash - riscrivibili e non volatili . . . . .	127
12.2.1	Il Floating gate transistor . . . . .	127
12.2.2	Programmazione . . . . .	127
12.3	Memorie EEPROM . . . . .	128
12.4	SSD basate su memoria Flash NAND . . . . .	129
12.4.1	Organizzazione delle celle . . . . .	129
12.4.2	NOR vs. NAND layout . . . . .	130
12.5	Organizzazione interna memorie NAND - pagine e blocchi . . . . .	131
12.5.1	Cancellazione e scrittura . . . . .	131
12.6	Piani . . . . .	132
12.7	Dies . . . . .	132
12.8	Distribuzioni dei livelli logici per una memoria NAND . . . . .	133
12.8.1	Cella multi bit a 8 livelli . . . . .	133
12.9	Programmazione . . . . .	133
12.10	Lettura . . . . .	134
12.11	Conclusioni . . . . .	135

# Capitolo 1

## Il Silicio

### 1.1 Materiali

Per approcciare alla materia servirebbe studiare la meccanica quantistica

I materiali possono essere di tre tipologie;

- Isolanti;
- Conduttori;
- Semiconduttori.

Tutti si distinguono per la loro resistività, vedremo in modo particolare i **semiconduttori**, ottenuti dalla colonna IV, oppure dalla combinazione della colonna III e V della tavola periodica.

TABLE 2.1 Electrical Classification of Solid Materials	
MATERIALS	RESISTIVITY $\rho$ ( $\Omega \cdot \text{cm}$ )
Insulators	$10^5 < \rho$
Semiconductors	$10^{-3} < \rho < 10^5$
Conductors	$\rho < 10^{-3}$

III A	IV A	V A	VIA
<sup>5</sup> Boro	<sup>6</sup> Carbonio	<sup>7</sup> Azoto	<sup>8</sup> Ossigeno
13 Alluminio	14 Silicio	15 Fosforo	16 Zolfo
<sup>30</sup> Zinco	<sup>31</sup> Gallo	<sup>32</sup> Germanio	<sup>33</sup> Arsenico
48 Cadmio	49 Indio	50 Stagno	51 Antimonia
<sup>60</sup> Mercurio	<sup>81</sup> Titano	<sup>82</sup> Piombo	<sup>83</sup> Bismuto
			<sup>84</sup> Polonio

<span style="background-color: #00FFFF; color: black;">■</span>	Conduttori	$\rho < 10^{-5} \Omega \cdot \text{m}$	(rame: $3 \cdot 10^{-8} \Omega \cdot \text{m}$ )
<span style="background-color: #FF8C00; color: black;">■</span>	Semiconduttori	$10^{-5} < \rho < 10^5 \Omega \cdot \text{m}$	(silicio: $2300 \Omega \cdot \text{m}$ )
<span style="background-color: #FFFFE0; color: black;">■</span>	Isolanti	$\rho > 10^5 \Omega \cdot \text{m}$	(diamante: $10^{14} \Omega \cdot \text{m}$ )

Figura 1.1: Eletrical Classification of Solid Materials

**Esempio:** Prendiamo un filo lungo 1 cm, area  $10\mu\text{m} \cdot 10\mu\text{m}$ .  $L/A = 1/0.001 \cdot 0.001 = 10^6$

Prendendo come resistività  $\rho = 10^5$  e poi  $\rho = 10^{-3}$  ed utilizzando la relazione  $\rho = R \frac{A}{L}$ , otteniamo dei valori di resistenza di diversi ordini di grandezza differenti:  $R = 100 G\Omega \quad 1 K\Omega$ .

### 1.1.1 Struttura cristallina del silicio

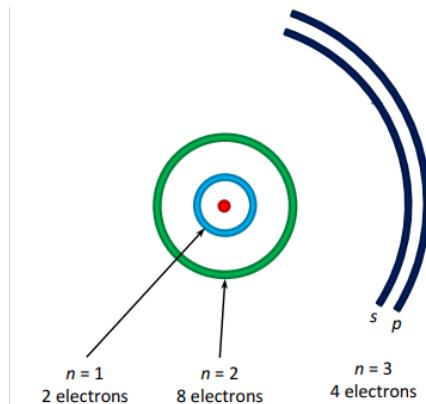


Figura 1.2: Atomo di silicio

Alla temperatura dello zero assoluto,  $0^{\circ} K = -273.15^{\circ} C$ , tutti gli elettroni sono fissi in un legame covalente e dunque il materiale risulta isolante. Questo dovuto al fatto che nessuna carica, elettrone, si può muovere.

A temperature più alte, fornendo quindi dell'energia, alcuni legami si rompono e l'elettrone, sull'orbitale più esterno, è libero per la conduzione.

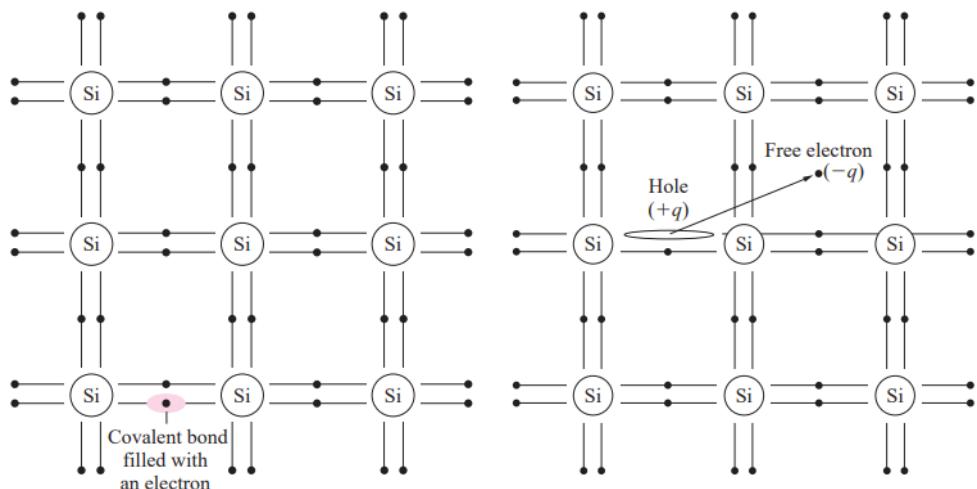


Figura 1.3: Struttura cristallina Silicio

A questo punto si vede anche in figura che un elettrone che si stacca da un legame covalente lascia dietro di sé un buco chiamato **hole** o **lacuna** che dovrà essere riempito.

### 1.1.2 Densità di elettroni liberi

Il numero di elettroni liberi in dipende da due fattori:

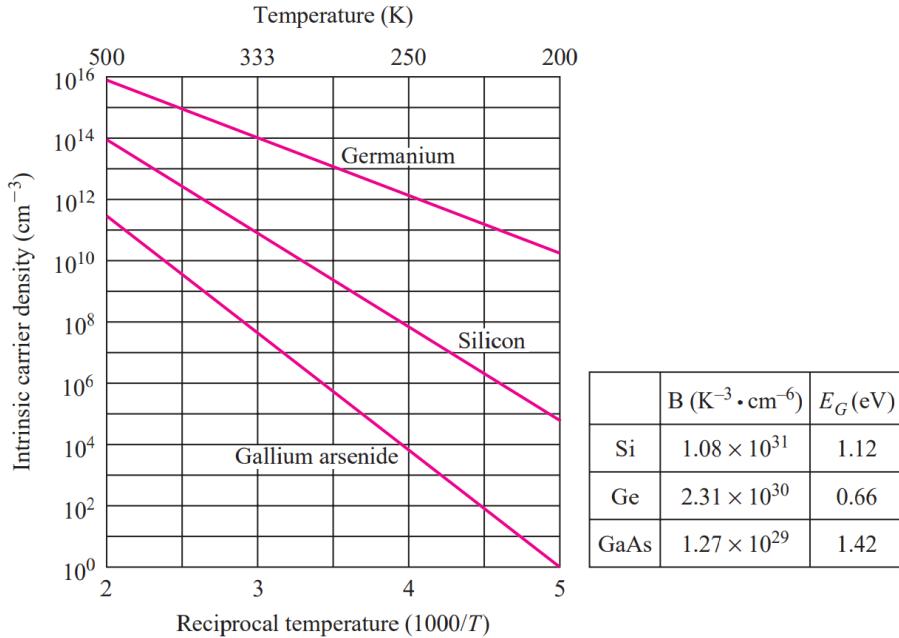
- Temperatura;
- Materiale;

Utilizzando la seguente formula si possono trovare la densità di cariche:

$$n_i^2 = BT^3 \exp -\frac{E_G}{kT} \quad (1.1)$$

dove:

- $n_i$  = concentrazione di elettroni per  $cm^3$
- $E_G$  = semiconductor bandgapenergy [eV] (elettronvolt, energia minima per rompere un legame)
- $k$  = costante di Boltzmann equivalente a  $8.62 \times 10^{-5}$  [eV/K]
- $T$  = temperatura, [K]
- $B$  = parametro del materiale,  $1.08 \times 10^{31}$  [ $K^{-3} cm^{-6}$ ] per il Si



**Calcoliamo le cariche nel silicio:** Sappiamo che la densità del silicio è:  $5 \times 10^{22}/cm^3$ , applicando la formula a temperatura ambiente otteniamo:

$$n_i = 6.73 \times 10^9 \approx 10^{10}/cm^3$$

**Lacune:** Un elettrone libero lascia dietro di sé un buco, il quale dovrà essere riempito. Un elettrone da un altro legame può riempire il buco, che si sposta dalla parte opposta all'elettrone.

La lacuna la possiamo considerare come una carica **positiva**.

All'equilibrio, la densità di elettroni liberi e di lacune è identica.

$$n = n_1 = p \quad pn = n_i^2$$

dove

- $n$  = concentrazione di elettroni liberi al  $cm^3$ ;
- $n_1$  = numero di elettroni al  $cm^3$ , densità di e al  $cm^3$ ;
- $p$  = concentrazione di lacune libere al  $cm^3$ ;

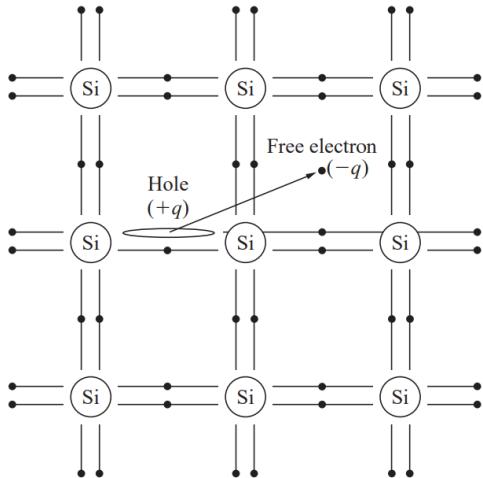


Figura 1.4: Spostamento di lacuna

## 1.2 Corrente di drift

Una elettrone libero, che rappresenta dunque una carica, può generare una corrente elettrica.

La corrente rappresenta la quantità **netta** di cariche per unità di tempo che attraversano la selezione del materiale. Il valore della corrente dipende dalla **velocità**. All'equilibrio termico, le cariche si muovono caoticamente e casualmente in tutte le direzioni, cambiando direzione ad ogni urto con un atomo del cristallo.

Questo fa sì che:

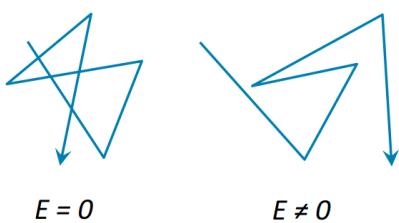
- La velocità media sia **nulla**;
- La corrente totale risulti essere **zero**



Applicando un **campo elettrico**  $E$ , le cariche tendono a muoversi nella sua direzione (drift).

- Legge di Coulomb:  $\vec{F} = q\vec{E}$ ;
- Seconda legge di Newton:  $\vec{F} = m\vec{a}$
- Quindi:  $\vec{a} = \frac{q}{m}\vec{E}$

L'effetto della struttura dei materiali, essendoci gli atomi, fa sì che le collisioni avvengano e modifichino la traiettoria dell'elettrone. Questo scontro fa perdere energia agli elettroni, effetto Jule, e l'effetto totale è quello di godere di un moto costante verso una direzione.



La velocità risulta essere proporzionale al campo elettrico:

$$v_n = -\mu_n \vec{E} \quad v_p = \mu_p \vec{E} \quad (1.2)$$

dove

- $v_n$  = Velocità elettroni
- $\mu_n$  = mobilità degli elettroni,  $1350 \text{ cm}^2/\text{V} \cdot \text{s}$  in Si intrinseco (puro).
- $v_p$  = Velocità lacune
- $\mu_p$  = mobilità degli elettroni,  $500 \text{ cm}^2/\text{V} \cdot \text{s}$  in Si intrinseco (puro).

Le **lacune** hanno **minore mobilità** perché costrette a muoversi nella struttura, mentre gli elettroni godono di maggiore libertà.

### 1.2.1 Velocità di saturazione

Come visto prima, la velocità delle cariche risulta essere proporzionale, in media, al campo elettrico. Questo è vero fino ad un certo valore perché anche le cariche non possono superare la velocità della luce,  $300\,000 \text{ [Km/s]}$ .

Dunque esiste un limite di saturazione il quale limita la risposta in frequenza e dunque anche la velocità di una porta logica (clock limitato).

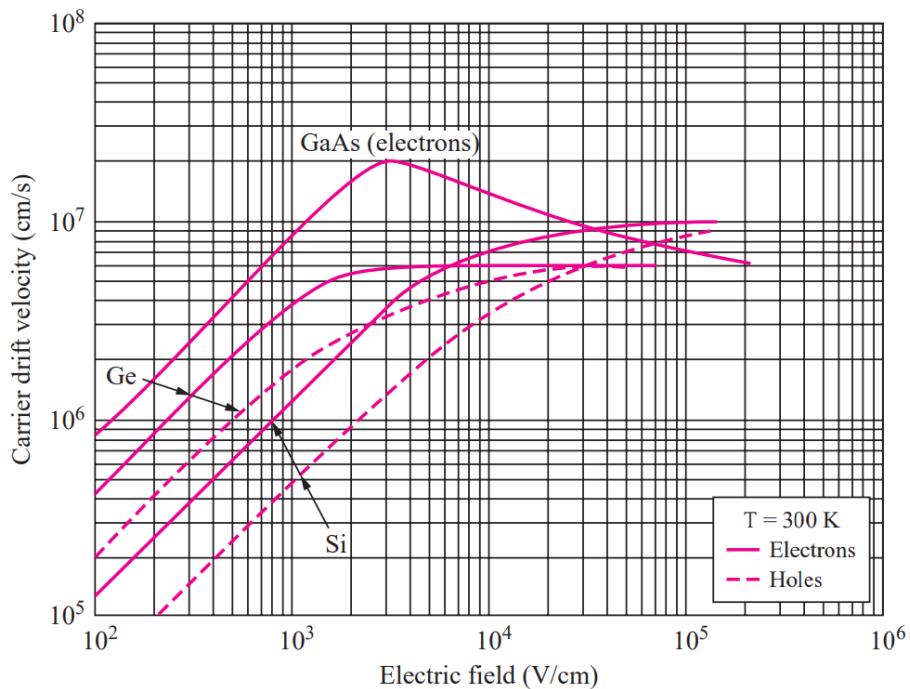


Figura 1.5: Velocità di saturazione

## 1.2.2 Calcolo della corrente di drift

Valutiamo la carica che attraversa una superficie nell'unità di tempo.

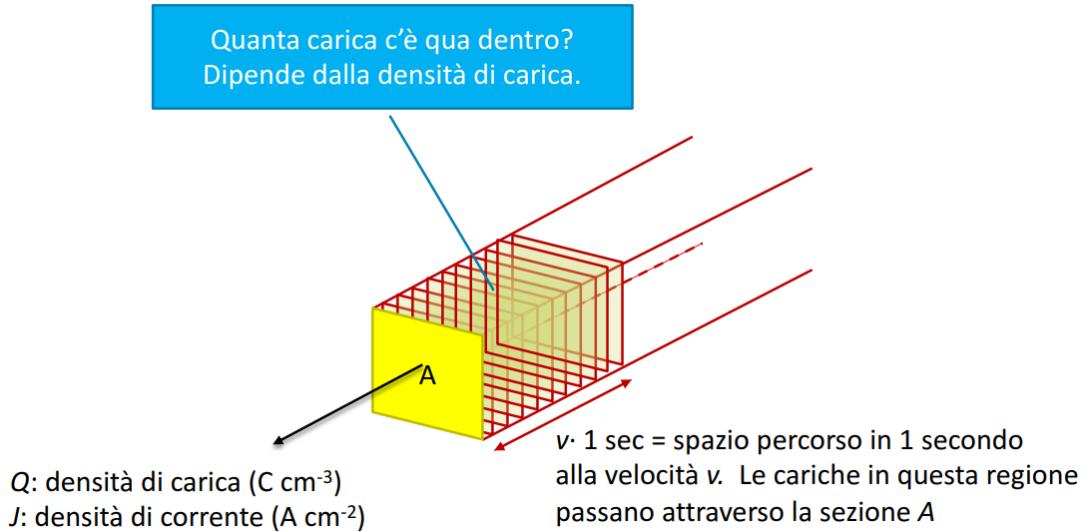


Figura 1.6: Corrente di Drift

Ricordiamo le seguenti formule:

$$I = JA = \sigma \frac{A}{L} \vec{E} L = GV \quad q = 1.6 \cdot 10^{-19}$$

$$j_n^{drift} = Q_n v_n = (-qn)(-\mu_n \vec{E}) = qn\mu_n \vec{E} \quad [A/cm^2] \quad (1.3)$$

$$j_p^{drift} = Q_p v_p = (-qp)(-\mu_p \vec{E}) = qp\mu_n \vec{E} \quad [A/cm^2] \quad (1.4)$$

$$j_T^{drift} = j_n^{drift} + j_p^{drift} = q(n\mu_n + p\mu_p)\vec{E} = \sigma \vec{E} \quad (1.5)$$

$\sigma : q(n\mu_n + p\mu_p)$  rappresenta la conduttività.

### A temperatura ambiente

$$n = n_i = p = 10^{10}$$

$$\sigma = (1.6 * 10^{-19}) 10^{10} (1350 + 500) = 2.96 * 10^{-6}$$

$$j_T^{drift} = j_n^{drift} + j_p^{drift} = q(n\mu_n + p\mu_p)\vec{E} = \sigma \vec{E}$$

$$\rho = \frac{1}{\sigma} = 3.38 * 10^5 \omega cm$$

Dunque il silicio a temperatura ambiente rientra nella famiglia degli **isolanti**.

### Conduttore in rame lungo 1m, di diametro 1mm e tensione di 1V

$$R = \rho \frac{L}{A} = 1.6810^{-8} \frac{1}{\pi 0.0005^2} = 0.0214 \omega$$

$$I = \frac{V}{R} = \frac{1}{0.0214} = 46.7 A$$

$$J = \frac{I}{A} = \frac{46.7}{\pi 0.0005^2} = 5910^6 Am^{-2}$$

$$v = \frac{J}{nq} = \frac{5910^6}{8.4610^{28} 1.610^{-19}} = 0.0044 ms^{-1} = 4.4 mms^{-1}$$

Questa risulta essere la velocità netta di drift, gli elettroni si muovono a velocità molto più elevata, ma caoticamente. Dunque gli elettroni impiegano delle ore per fare quale metro in un filo di rame, come è possibile dunque che la luce di una lampadina di accenda subito?

Questo è dovuto alla **reazione a catena** che da al via al tutto, e in generale è il **campo elettromagnetico** che si sposta rapidamente e che fa muovere gli elettroni.

## 1.3 Impurità

L'aggiunta di impurità ci permette di controllare la resistività.

### 1.3.1 Atomi pentavalenti

Contribuiscono con un elettrone in più, alcuni di questi atomi sono nella colonna V della tavola periodica:

Fosforo, Arsenico, Antimonio

Questi atomi, superati i  $0^\circ K$ , si ionizzano più velocemente ed essendo che hanno un elettrone in più, rispetto al Si, questo sarà libero di viaggiare libero nel reticolo cristallino. Essendo che P, As e Sb possono donare cariche, vengono chiamati **donatori**.

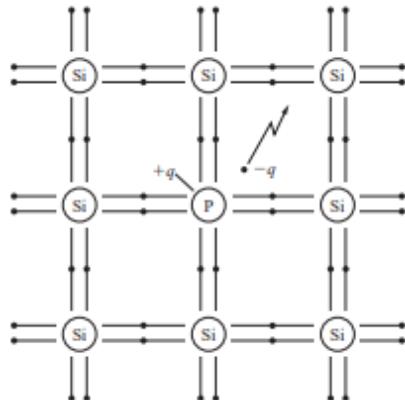
### 1.3.2 Atomi trivalenti

Contribuiscono con un **elettrone** in meno e dunque contribuiscono ad una **lacuna**, alcuni di questi atomi sono nella colonna III della tavola periodica:

Boro

Questi atomi, a differenza di quelli sopracitati, vengono chiamati **accettori** in quanto avendo sull'orbitale più esterni un elettrone in meno rispetto a Si, formano una lacuna la quale dovrà essere riempita da un elettrone del Si.

La conduzione è affidata prevalentemente a elettroni o lacune.

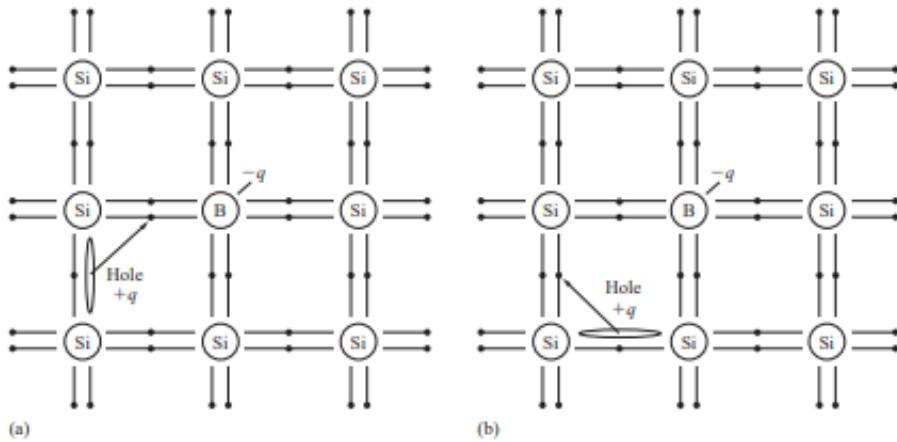


Poca energia/calore è sufficiente per liberare le cariche, dunque queste impurità modificano la concentrazione di portatori:

- $n = p$ : Stiamo parlando del silicio senza impurità chiamato **intrinseco**
- $n \neq p$ : Parliamo di silicio con aggiunta di impurità chiamato **estrinseco**

$n > p$ , il silicio è di **tipo n**, ovvero i portatori di maggioranza sono gli **elettroni**, donatori di ioni positivi. Questo silicio ha una migliore mobilità.

$n < p$ , il silicio è di **tipo p**, ovvero i portatori di maggioranza sono le **lacune**, donatori di ioni negativi. Questo silicio ha una peggiore mobilità.



Le concentrazioni di impurità sono normalmente dell'ordine di  $10^{14}$  fino a  $10^{21}$  atomi per  $cm^3$ , per confronto, la concentrazione intrinseca di portatori è dell'ordine di  $10^{10}$  cariche per  $cm^3$ .

La concentrazione dei portatori **maggioritari** è quindi praticamente uguale alla concentrazione di impurità ed è praticamente costante. Sono le **impurità** che portano il più alto contenuto di elettroni liberi, l'equilibrio termico ne porta molte meno. Infatti aggiungere  $10^{14}$  impurità significa aggiungere la stessa quantità di cariche o lacune.

### 1.3.3 Conduzione

Occorre calcolare la concentrazione di portatori. La carica totale del conduttore deve essere globalmente nulla.

- $N_D$  = concentrazione donatori (ioni positivi)
- $N_A$  = concentrazione accettori (ioni negativi)

$$q(N_D + p - N_A - n) = 0$$

Vale ancora la relazione (legge dell'azione di massa):  $pn = n_i^2$

Quindi per i semiconduttori di tipo n:

$$n^2 - (N_D - N_A)n - n_i^2 \quad n = \frac{(N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2}}{2}$$

In pratica se  $(N_D - N_A) \gg 2n_i$  allora:

$$n = N_D - N_A \quad p = \frac{n_i^2}{n}$$

Per il silicio di tipo p:

$$p = N_A - N_D \quad n = \frac{n_i^2}{p}$$

**Pertanto:** La concentrazione dei portatori **maggioritari** è praticamente costante e **indipendente dalla temperatura**.

La concentrazione dei portatori **minoritari** è invece proporzionale a  $n_i^2$  e fortemente **dipende dalla temperatura**

Per esempio, drogando con Boro a  $10^{16}/cm^3$  e con Fosforo a  $2 \cdot 10^{15}/cm^3$  a temperatura ambiente:

$$N_A = 10^{16} \quad N_D = 2 \cdot 10^{15}$$

$$p = (N_A - N_D) = 8.00 \cdot 10^{15}$$

$$n = \frac{n_i^2}{p} = 1.25 \cdot 10^4$$

**A 400 Kelvin (126,85°C):**

$$n_i^2 = 5.4 \cdot 10^{24} \quad p = (N_A - N_D) = 8.00 \cdot 10^{15} \quad n = \frac{n_i^2}{p} = 6.75 \cdot 10^8$$

### 1.3.4 Resistività in materiali estrinseci

Il drogaggio riduce la mobilità dei portatori, infatti un silicio drogato con  $N_D = 2 \cdot 10^{15}/cm^3$

$$n = 2 \cdot 10^{15}/cm^3 \quad p = 10^{20}/2 \cdot 10 \cdot 10^{15} = 5 \cdot 10^4/cm^3$$

Questo silicio di tipo p ha una mobilità inferiore:

$$\mu_n = 1320 \quad \mu_p = 460$$

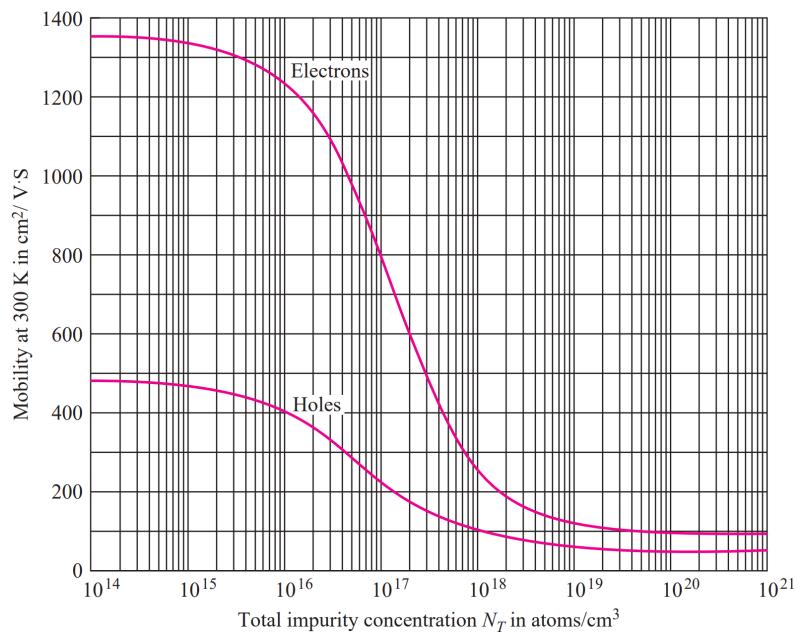
La conduttività e resistività sono:

$$\sigma = (1.6 \cdot 10^{-19})[(2 \cdot 10^{15})1320 + (5 \cdot 10^4)460] = 0.422$$

$$\rho = 1/\sigma = 2.37 \text{ } [\Omega cm]$$

Il silicio ora è un semiconduttore, da notare che una piccola frazione di impurità ( $2 \cdot 10^{15}$  contro  $5 \cdot 10^{22}$ ) cambia la resistività di 5 ordini di grandezza, prima era:  $3.38 \cdot 10^5$ .

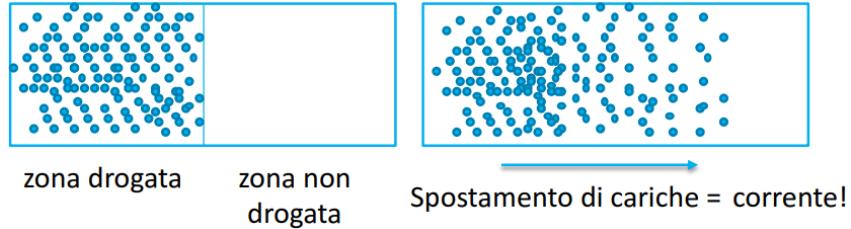
Dunque possiamo controllare la resistività del silicio.



## 1.4 Corrente di diffusione

Se il drogaggio non è uniforme le concentrazioni dei portatori variano lungo il cristallo. Si generano quindi correnti di diffusione proporzionali al gradiente di concentrazione.

I portatori vanno da regioni a maggiore concentrazione a regioni a minore concentrazione, proprio come succede con lo scambio termico.



Si è notato che questa corrente di diffusione segue le seguenti ugualianze:

$$J_p^{diff} = (+q)D_p \left( -\frac{\partial p}{\partial x} = -qD_p \frac{\partial p}{\partial x} \right)$$

$$J_n^{diff} = (-q)D_n \left( -\frac{\partial n}{\partial x} = -qD_n \frac{\partial n}{\partial x} \right)$$

Dove  $D_p$  e  $D_n$  sono le diffusività di lacune ed elettroni che dipendono dalla mobilità e temperatura secondo la relazione di Einstein:

$$\frac{D_n}{\mu_n} = \frac{kT}{q} = \frac{D_p}{\mu_p} \quad \text{dove } V_t = kT/q = 0.0258V \text{ a } 300K$$

## 1.5 Corrente totale

Per ottenere la corrente totale, applicando un campo elettrico, dobbiamo sommare la corrente di drift e di diffusione:

$$j_n^T = q\mu_n n \vec{E} + qD_n \frac{\partial n}{\partial x} \quad (1.6)$$

$$j_p^T = q\mu_p p \vec{E} + qD_p \frac{\partial p}{\partial x} \quad (1.7)$$

Oppure applicando la relazione di Einstein:

$$j_n^T = q\mu_n n \left( \vec{E} + V_T \frac{1}{n} \frac{\partial n}{\partial x} \right) \quad (1.8)$$

$$j_p^T = q\mu_p p \left( \vec{E} - V_T \frac{1}{p} \frac{\partial p}{\partial x} \right) \quad (1.9)$$

Per completare il sistema dobbiamo considerare la dipendenza del campo dalla carica:

$$\nabla \cdot \varepsilon \vec{E} = Q$$

Dove  $\varepsilon$  è la permittività e  $Q$  la densità di carica nello spazio.

## 1.6 Take away

Tramite il drogaggio possiamo controllare la concentrazione delle cariche nel silicio, cariche costituite da **elettroni e/o lacune**.

La corrente si può dividere in due componenti:

- Corrente di **drift** dovuta ad un campo elettrico applicato;
- Corrente di **diffusione** dovuta a gradienti di concentrazione.

I dispositivi elettronici sono realizzati tramite aree di tipo p ed n collegate.

## 1.7 Processo costruttivo

I circuiti sono/erano realizzati secondo un processo planare. Un substrato di silicio drogato di tipo p o n funge da supporto, altre aree drogate diversamente di tipo p o n vengono create sulla superficie.

Queste aree sono collegate tra loro tramite piste in alluminio, disposte su diversi strati separati da isolante.

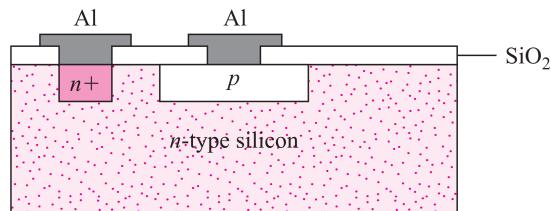


Figura 1.7: Transistor vista frontale

# Capitolo 2

## Il Diodo

### 2.1 Come è fatto

Il diodo è costruito da una giunzione di silicio di tipo p e n.

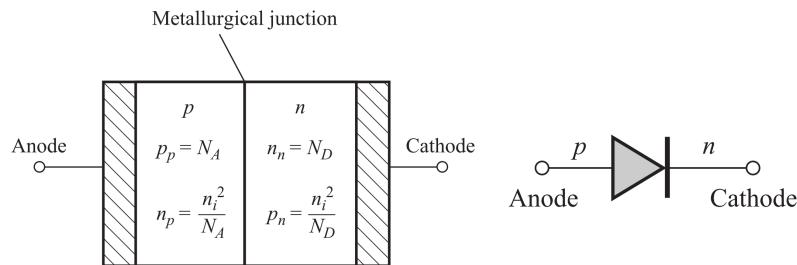


Figura 2.1: Diodo

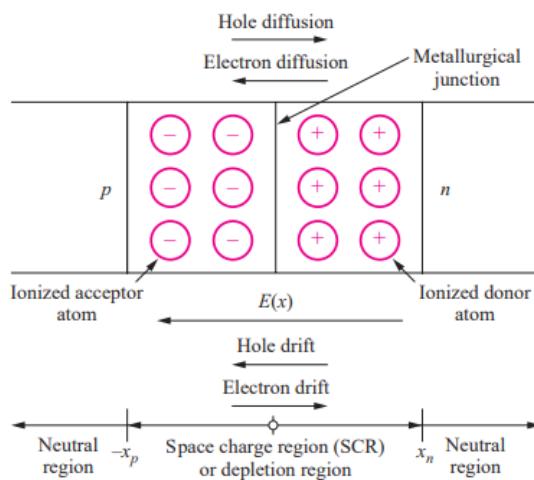
Supponiamo

- $N_A = p_p = 10^{17}/cm^3$        $N_D = n_n = 10^{16}/cm^3$
- $n_p = 10^3/cm^3$        $p_n = 10^4/cm^3$

C'è un forte gradiente di concentrazione di carica, infatti le lacune tendono a diffondere dalla zona p alla zona n, mentre gli elettroni tendono a diffondere dalla zona n alla zona p.

#### Carica spaziale:

La corrente totale deve però essere nulla, infatti una corrente di drift deve bilanciare quella di diffusione. La corrente di drift è dovuta alla creazione di una regione di carica spaziale, che genera un campo elettrico.



Da dove viene il campo elettrico? A sinistra vi erano le lacune che si sono spostate nella parte n ed hanno lasciato degli atomi di Boro scoperti, questo fa sì che si lasciano degli atomi ionizzati negativamente a sinistra. Gli elettroni che da destra sono andati a sinistra sono ioni positivi di fosforo.

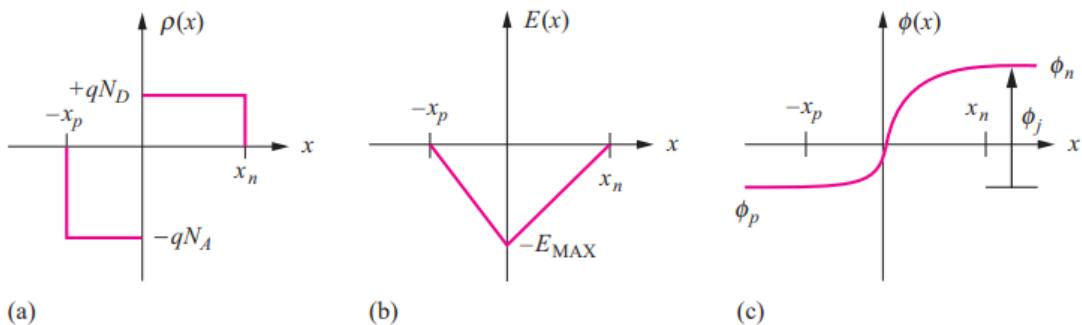
Man mano che le cariche si diffondono, a ridosso della giunzione si forma una zona di ioni che sono cariche fisse e prendono in nome di **carica spaziale**, queste cariche fisse formano un dipolo con il positivo a destra e il negativo a sinistra formando così un campo elettrico dovuto a queste cariche.

Il CE fa insorgere una corrente di trascinamento che va in verso opposto a quelle della diffusione (il CE va dalle cariche positive a quelle negative, la corrente di diffusione il contrario).

Dunque all'inizio si forma la corrente di diffusione, e inizia a formarsi la carica spaziale a ridosso della giunzione e quest'ultima ostacola la diffusione ed ad un certo punto si arriva ad un equilibrio.

Questo equilibrio non significa che non c'è più movimento ma le cariche che diffondono e driftano si annullano a vicenda e dunque la corrente totale diventa nulla.

$$qN_Ax_p = qN_Dx_n \quad \nabla \cdot E = \frac{\rho_c}{\epsilon_s} \quad \phi_j = - \int E(x) dx$$



Per capire quando questo processo va in equilibrio possiamo calcolare il campo elettrico. Possiamo calcolarlo perché la quantità di questi ioni, concentrazione, la sappiamo a priori dato che li abbiamo inseriti noi consapevolmente, sotto l'ipotesi che si siano tutti ionizzati.

## 2.2 Calcolo del potenziale di built-in $\phi_j$

Il potenziale di **built-in** è il fenomeno che si oppone alla diffusione delle cariche tra una zona e l'altra, viene chiamato di built-in perché è già all'interno dopo la giunzione dei due materiali con drogaggi diversi.

Sarà necessario **superare** questo potenziale per far scorrere una corrente all'interno del diodo facendolo lavorare come un corto circuito.

Utilizzando l'equazione 1.9 e ponendola uguale a zero, dato che la corrente si deve bilanciare, avremo che:

$$\begin{aligned} E - V_T \frac{1}{p} \frac{\partial p}{\partial x} &= 0 \quad \rightarrow \quad E = - \frac{\partial V}{\partial x} \\ - \frac{\partial V}{\partial x} &= V_T \frac{1}{p} \frac{\partial p}{\partial x} \quad \rightarrow \quad dV = - V_T \frac{dp}{p} \end{aligned}$$

integrando da ambo le parti otteniamo che:

$$V_2 - V_1 = \phi_j = V_T \cdot \ln \left( \frac{N_A N_D}{n_i^2} \right) \quad (2.1)$$

dove:

- $N_A$  e  $N_D$  li conosciamo dato che abbiamo deciso noi le impurità
- $n_i$  dipende dalla temperatura
- $V_T$  è una costante

$$\phi_j = 0.026 \ln(10^{17} 10^{16} / 10^{20}) = 0.026 \ln(10^{13}) \approx 0.7V$$

Questo potenziale di built-in  $\phi_j$  risulta essere circa  $0.75V$  per il silicio e  $0.3V$  per il germanio, tensione che contrasta la diffusione delle cariche dei portatori.

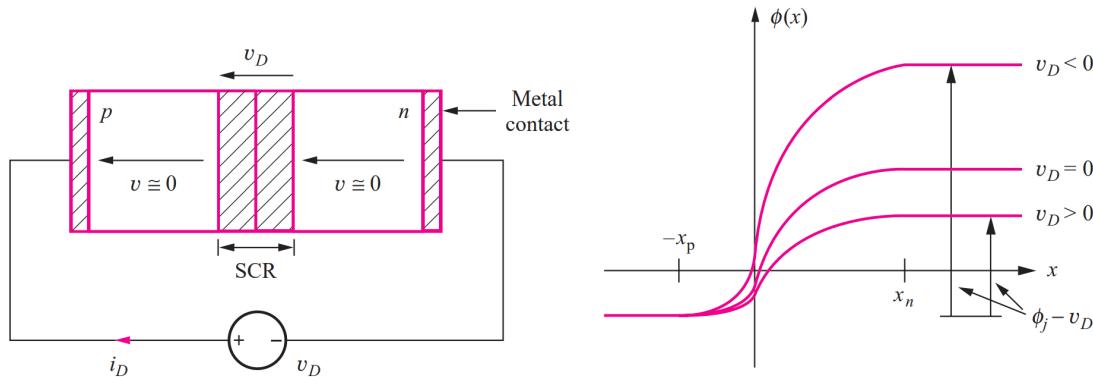
Questo significa che una lacuna a sinistra deve avere abbastanza energia per superare la barriera di potenziale.

## 2.3 Potenziale esterno

Applicando un potenziale esterno si altera l'equilibrio.

- Un **potenziale positivo** riduce la barriera, e una corrente può scorrere nel diodo
- Un **potenziale negativo** incrementa la barriera. La corrente scorre, ma è molto piccola perché fatta solo dai portatori minoritari

Per la legge della sovrapposizione degli effetti il potenziale esterno si somma a quello interno.



La barriera si può molto facilmente abbassare o azzerare, basta applicare una differenza di potenziale tra anodo e catodo di  $0.7V$  e a questo punto tutte le cariche possono spostarsi liberamente.

### 2.3.1 Potenziale positivo

L'equilibrio è perturbato, la regione di carica spaziale si restringe, inoltre passerà molta corrente in quanto la differenza di densità di carica è di diversi ordini di grandezza superiore.

La corrente di drift non è in grado di bilanciare quella di diffusione dunque:

- le lacune si diffondono dalla zona p alla zona n;
- gli elettroni si diffondono dalla zona n alla zona p;
- le correnti fluiscano nello stesso verso.

### 2.3.2 Potenziale negativo

In questo caso mettiamo una batteria in polarità inversa: il + sul catodo e il - sull'anodo. Così facendo non solo le cariche sono contrastate dalla barriera di spostamento ma vi è anche la barriera aggiuntiva della batteria collegata e dunque le cariche non diffondono per niente e non passa corrente.

La barriera di potenziale si amplia, la regione di carica spaziale si allarga. La corrente di diffusione viene ostacolata, mentre i portatori minoritari possono scivolare per la barriera:

- le lacune si diffondono dalla zona n alla zona p;
- gli elettroni si diffondono dalla zona p alla zona n.

Scorre quindi una debole corrente, dovuta ai portatori minoritari generati termicamente, in prima approssimazione indipendente dal potenziale, talmente piccola che può essere considerata nulla.

Il diodo a differenza di un resistore permette di dare un solo verso alla corrente, **raddrizzarla**: infatti si para di **polarizzazione**.

- **Polarizzazione diretta**: tensione esterna applicata in modo da ridurre la barriera energetica permettendo il flusso di corrente;
- **Polarizzazione inversa**: tensione esterna applicata in modo da aumentare la barriera energetica impedendo il flusso di corrente.

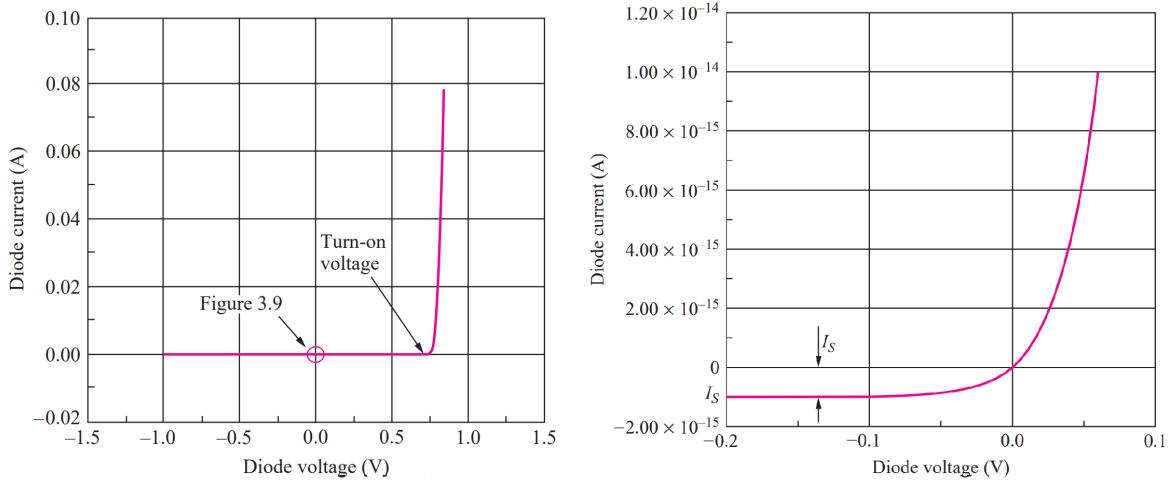


Figura 2.2: Grafico andamento tensione-corrente

## 2.4 Modello matematico del Diodo

$$I_D = I_S \left[ \exp \left( \frac{q v_D}{n k T} \right) - 1 \right] = I_S \left[ \exp \left( \frac{v_D}{n V_T} \right) - 1 \right] \quad (2.2)$$

Dove

- $I_S$  = corrente inversa di saturazione
- $v_D$  = tensione applicata ai capi del diodo
- $q$  = carica elettronica  $1.6 \cdot 10^{-19}$  [C]
- $k$  = costante di Boltzmann  $1.38 \cdot 10^{-23}$  [J/K]
- $n$  = fattore di non idealità (tipicamente 1 per Si)
- $V_T = kT/q$  = tensione termica (0.025V, Si, temp. amb)

## 2.5 Modello di diodo ideale per grandi tensioni

Quando si lavora con tensioni molto maggiori della tensione minima per far oltrepassare la barriera alle cariche (0.7V) allora possiamo applicare una prima approssimazione generale del diodo:

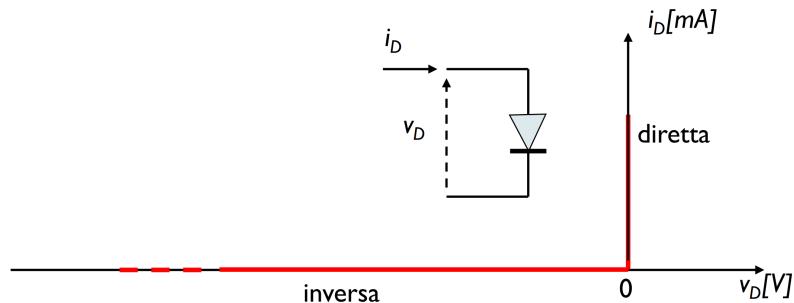


Figura 2.3: Diodo ideale

Questa grossolana approssimazione è applicabile se il nostro circuito è sottoposto a decine di volt.

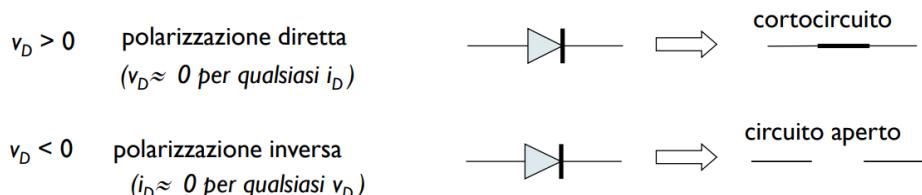


Figura 2.4: Comportamento in una rete

## 2.6 Diodo nei circuiti

I diodi sono dei componenti molto utilizzati soprattutto negli alimentatori di apparecchiature elettroniche in quanto esse, funzionando in regime di corrente continua, hanno bisogno di rendere positiva la sinusoide di rete per poter poi estrarre, grazie ad un condensatore, la componente continua.

Vediamo alcuni esempi di circuiti.

### 2.6.1 Rettificatore a semi onda

Il diodo è molto utilizzato per raddrizzare la sinusode di rete.

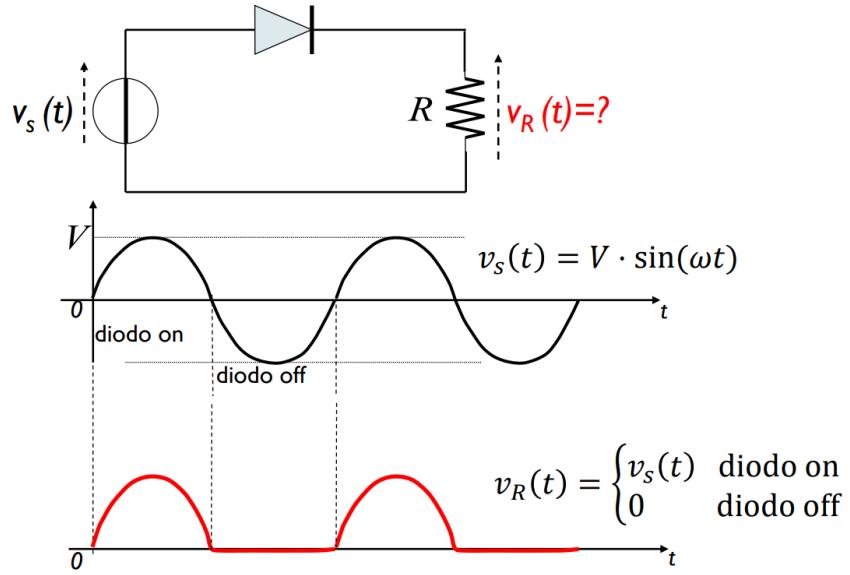


Figura 2.5: Rettificatore a semi onda

### 2.6.2 Rettificatore ad onda intera, il ponte di diodi

Ovviamente possiamo raddrizzare tutta la sinusode di rete in una sinusode con soli semi onde positive, grazie al ponte di diodi:

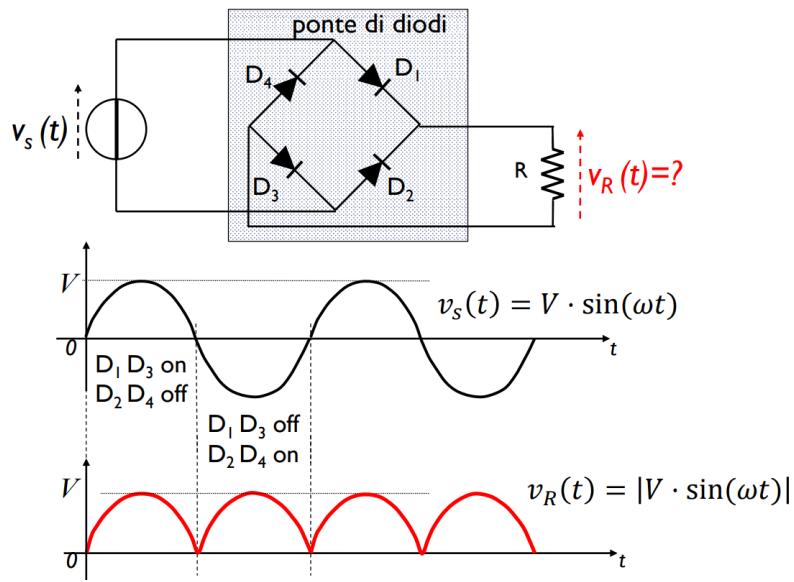


Figura 2.6: Ponte di diodi

### 2.6.3 Raddrizzatore AC/DC con filtro capacitivo

Normalmente i dispositivi elettronici funzionano in corrente continua. La rete però fornisce una tensione sinusoidale a 50Hz. Grazie al seguente circuito possiamo far tendere la nostra sinusoide di rete ad una tensione costante.

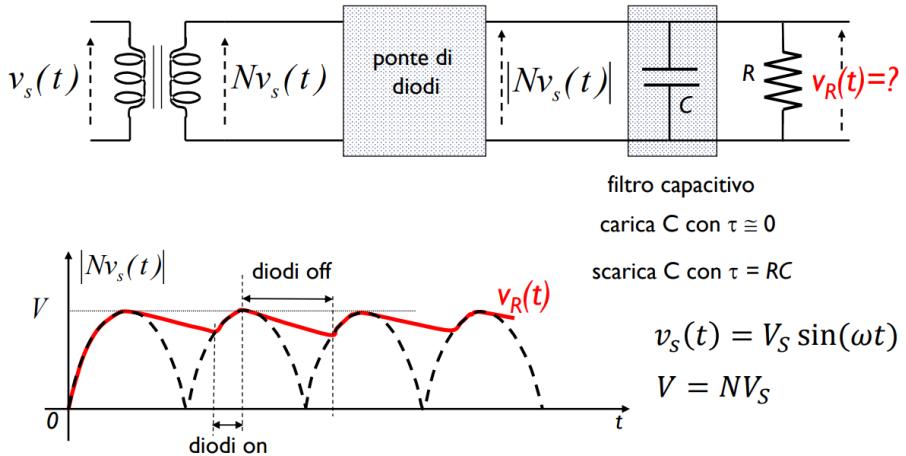


Figura 2.7: Raddrizzatore AC/DC con filtro capacitivo

La fluttuazione della tensione di uscita  $V_R(t)$  può essere controllata modificando la costante di tempo di scarica  $\tau = RC$  ( $\gg T$ )

Più il condensatore è grosso più riesce a mantenere costante la tensione durante la scarica, dunque perché non utilizzarne uno con molta capacità?

Ci sono alcuni motivi per evitare l'utilizzo di un condensatore grande:

- **Corrente di carica elevata:** Un condensatore di grandi dimensioni richiede una corrente di carica elevata durante l'avvio del circuito. Questo può causare picchi di corrente eccessivi che possono danneggiare i diodi raddrizzatori e altri componenti del circuito.
- **Tempo di carica lungo:** Un condensatore di grandi dimensioni impiega più tempo per caricarsi. Questo può portare a un ritardo significativo nella tensione di uscita all'accensione del circuito.
- **Aumento del ripple:** Un condensatore di grandi dimensioni può aumentare il ripple della tensione di uscita. Il ripple è la componente alternata residua presente nella tensione continua dopo la raddrizzatura. Un ripple eccessivo può essere dannoso per i circuiti elettronici sensibili.
- **Dimensioni e costo:** Un condensatore di grandi dimensioni è più grande e costoso di un condensatore di dimensioni più ridotte.
- **Non adatto a carichi variabili:** Un condensatore di grandi dimensioni non è adatto a carichi variabili. Se il carico varia, la tensione di uscita varierà di conseguenza.

In generale, è consigliabile utilizzare un condensatore di dimensioni sufficienti per ridurre il ripple a un livello accettabile, ma non troppo grande da causare i problemi sopra menzionati. Il valore ottimale del condensatore dipende da diversi fattori, tra cui la frequenza della tensione di rete, il tipo di raddrizzatore, il valore del carico e il livello di ripple desiderato.

## 2.7 Modello di diodo ideale per piccole tensioni

A differenza di prima ora stiamo prendendo in esame un circuito che lavora su basse tensioni (es. 5V). Per questo utilizzo non si può approssimare come prima in quanto la barriera di potenziale iniziale diventa molto rilevante e potrebbe creare problemi ai carichi che si attaccano al circuito.

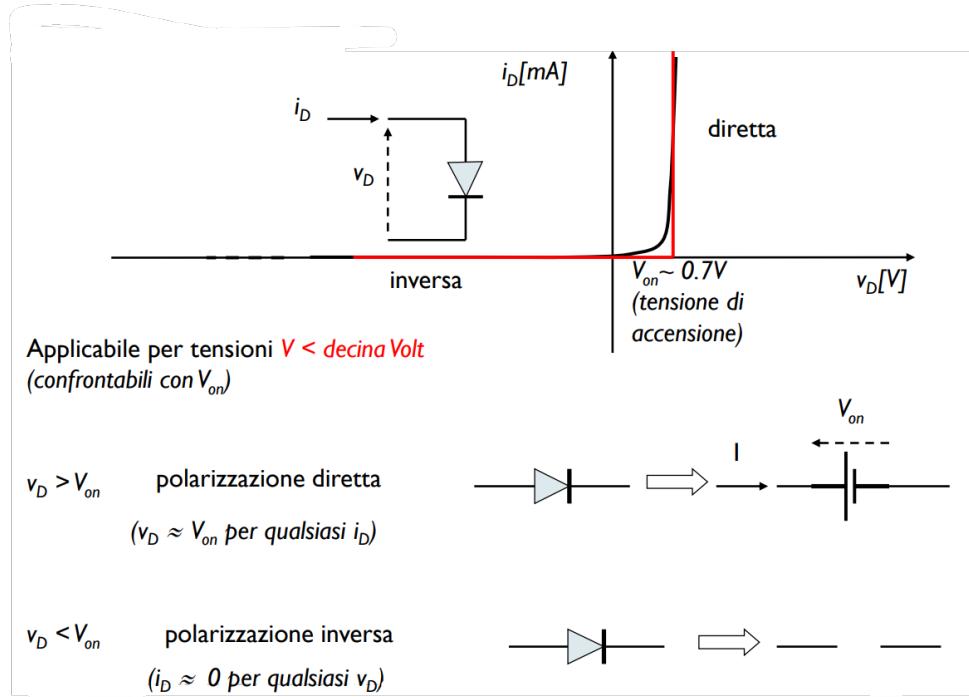


Figura 2.8: Modello di diodo per piccole tensioni

## 2.8 I circuiti logici realizzabili con i diodi

I diodi possono essere utilizzati per costruire porte logiche quali ad esempio la AND e OR.

Di seguito verrà mostrata una loro possibile implementazione.

### 2.8.1 Circuito AND

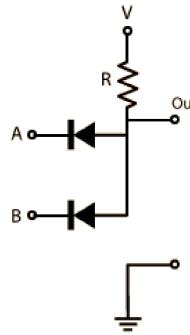


Figura 2.9: Diode AND gate

Vediamo i tre casi:

- $V_A = V_B = 5V$ 
  - entrambi i diodi sono interdetti
  - non scorre corrente
  - l'uscita va a 5V
- $V_A = 0V, V_B = 5V$ 
  - Il diodo B conduce, il diodo A è interdetto
  - l'uscita scende a  $V_{on}$  (0 per grandi tensioni)
- $V_A = V_B = 0V$ 
  - entrambi i diodi conducono
  - l'uscita scende a  $V_{on}$  (0 per grandi tensioni)

### 2.8.2 Circuito OR

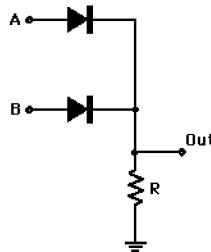


Figura 2.10: Diode OR gate

Vediamo i tre casi:

- $V_A = V_B = 5V$ 
  - entrambi i diodi conducono
  - l'uscita va a 5V -  $V_{on}$  (5 per grandi tensioni)
- $V_A = 5V, V_B = 0V$ 
  - Il diodo A conduce, il diodo B è interdetto
  - l'uscita va a 5V -  $V_{on}$  (5 per grandi tensioni)
- $V_A = V_B = 0V$ 
  - entrambi i diodi sono interdetti
  - l'uscita scende a 0V (0 per grandi tensioni)

# Capitolo 3

## Dispositivi MOSFET

### 3.1 Cos'è il condensatore MOS

Partiamo dal nome:

- MOS: Metal–Oxide–Semiconductor
- FET: field–effect transistor

ovvero transistore metallo–ossido–semiconduttore ad effetto di campo, questo dispositivo è alla base dei circuiti a MOSFET.

### 3.2 Come è fatto

Costituito da un layer conduttore detto **gate**, separato da un substrato in silicio drogato (di tipo p o n) e da uno strato isolante di biossido di silicio ( $SiO_2$ ), utilizzato per indurre una carica superficiale all'interfaccia tra il silicio e il biossido di silicio.

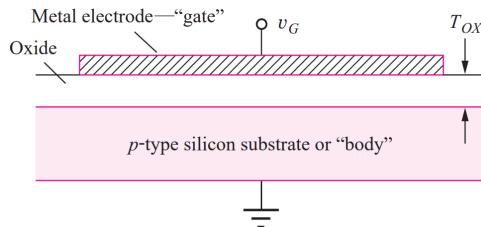


Figura 3.1: Condensatore MOSFET

Ricorda proprio un condensatore a piatti, tranne che al posto di esserci un metallo da entrambe le facce troviamo un semiconduttore e un metallo, dunque alteriamo le normali caratteristiche di un condensatore.

#### 3.2.1 Regione di accumulazione

La polarizzazione negativa avviene quando sul gate troviamo cariche negative, mentre sotto al gate troviamo cariche positive composta da lacune.

Dunque si dice che la superficie del semiconduttore è in accumulo. Sostanzialmente abbiamo accumulato le cariche presenti nel substrato drogato di silicio sulla sua superficie.

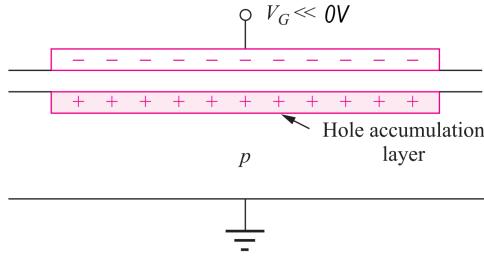


Figura 3.2: Regione di accumulazione

### 3.2.2 Regione di svuotamento

Il potenziale del gate viene incrementato ( $> 0V$ ). Dunque il campo elettrico è diretto verso il basso e allontana le lacune dalla superficie. Allontana le lacune ma rimangono gli ioni negativi degli atomi accettori che abbiamo utilizzato per produrre il substrato di tipo p, i quali controbilanceranno la carica positiva che si è formata sul gate.

Dunque ecco perché il nome zona di svuotamento, perché abbiamo allontanato, svuotato, la zona dalle lacune.

Da notare che gli ioni nel substrato sono **fissi** perché fanno parte del reticolo cristallino del silicio.

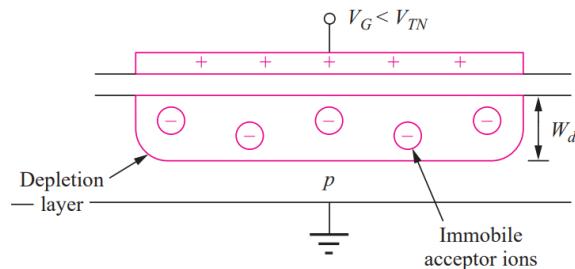


Figura 3.3: Regione di svuotamento

### 3.2.3 Regione di inversione

Se aumentiamo ancora la tensione sul gate, gli elettroni sono attratti sulla superficie del substrato, dunque la carica positiva nel metallo ora viene bilanciata dagli ioni fissi e da elettroni nel substrato.

Quando il numero di elettroni supera il numero di lacune, ci troviamo in condizione della cosiddetta **inversione**. La tensione alla quale si forma lo strato di inversione è chiamata **tensione di soglia**, indicata con  $V_T$  (o  $V_{TN}$  per gli elettroni).

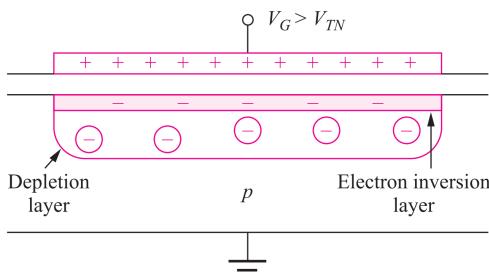


Figura 3.4: Regione di inversione

Dunque è come se avessimo fatto uno strato di tipo n sopra le lacune del substrato, abbiamo creato uno **strato conduttore**, praticamente fatto a comando variando la tensione sul gate, che mette in contatto il lato destro e sinistro del condensatore.

Ci fa comodo questo strato, infatti ai lati possiamo mettere due terminali ed otteniamo appunto il **transistore**.

### 3.3 Il transistore nMOS

Ai lati del gate, mettiamo due regioni di tipo n+ (tanti elettroni) chiamati: **source** e **drain**.

Questi due terminali sono messi in modo tale che possano mettersi in contatto con lo strato di inversione discusso in precedenza, formando di fatto una sorta di U rovesciata.

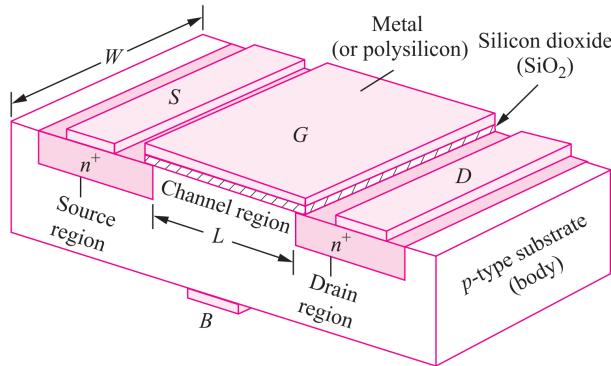


Figura 3.5: Transistore nMOS

Il transistore ha due parametri molto importanti che influenzano le caratteristiche elettriche del circuito:

- Distanza tra i terminali, lunghezza L del canale;
- Larghezza del transistore

Con più la lunghezza è piccola più il costo per stampare il circuito diventerà costoso  $\approx 10nm$ . Questa tipologia di transistori veniva più utilizzata una decina di anni fa, ora si è passati ad un nuovo transistore chiamato **FinFET** tridimensionale e non più planare, consentendo ulteriormente la miniaturizzazione.

#### 3.3.1 Simbolo circuitale

Le zone di source e drain sono terminali fortemente drogati con atomi donatori, e possono anche essi donare elettroni per formare lo strato di inversione. Le stesse regioni formano anche due diodi con il substrato (giunzione p-n), i quali sono normalmente **polarizzati inversamente**, in prima approssimazione è come se non ci fossero. Creano solo una carica spaziale intorno alle due connessioni.

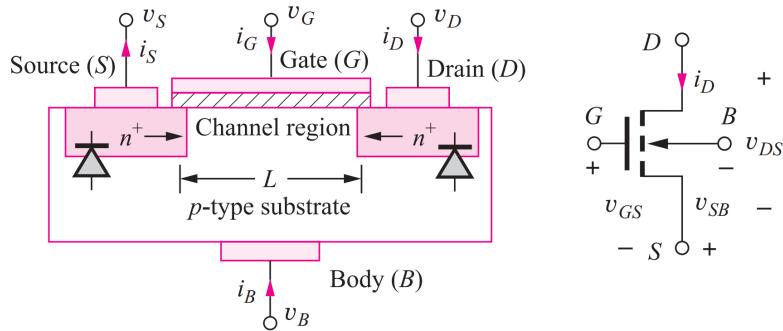


Figura 3.6: Transistore nMOS

#### 3.3.2 Funzionamento qualitativo

Assumiamo ora di mettere  $V_B = V_S = 0$  e  $V_D = \text{piccola}$ , e studiamo le tre situazioni:

- $V_{GS}^1 = 0$  o comunque molto minore di  $V_T^2$
- $V_{GS} < V_T$
- $V_{GS} > V_T$  circa un corto circuito

<sup>1</sup> $V_{GS}$  indica la tensione tra il gate e source che è la tensione che conta per formare la zona di inversione

<sup>2</sup>Tensione di soglia

$$V_{GS} \ll V_T$$

Innanzitutto troviamo una zona di carica spaziale intorno ai due terminali, e in questa condizione non è formata nessuna zona di inversione ma in ogni caso non può passare corrente perché non abbiamo cariche mobili all'interno del substrato che possono condurre.

Anche se al drain mettessimo una tensione positiva non scorrerebbe corrente.

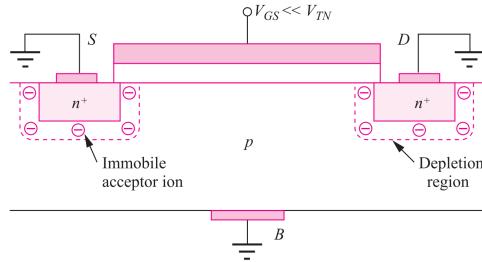


Figura 3.7: Nessuna inversione

$$V_{GS} < V_T$$

Aumentando la tensione sul gate inizia a formarsi la regione di carica spaziale sotto il gate che si unisce a quella dei diodi (giunzione pn), ma continua a non passare nessuna corrente. Questo perché la regione di svuotamento non fornisce cariche per la conduzione ma è costituita solo da ioni fissi.

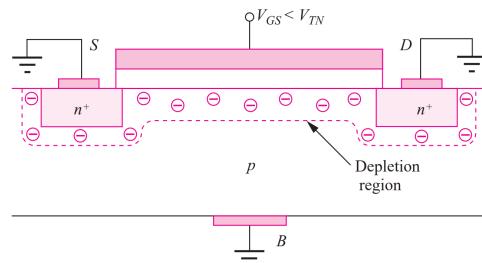


Figura 3.8: Giunzione pn

$$V_{GS} > V_T$$

Quando la tensione sul gate raggiunge e supera la tensione di sogli  $V_T$  ecco che si forma un canale di tipo n tra le sue regioni n<sup>+</sup>, un canale di elettroni mobili.

Ora se sul drain avessimo una tensione superiore del source ecco che potrebbe passare una corrente. Infatti la corrente dipende sia da  $V_{GS}$ , il quale determina il canale (quanti elettroni liberi) e da  $V_D$ .

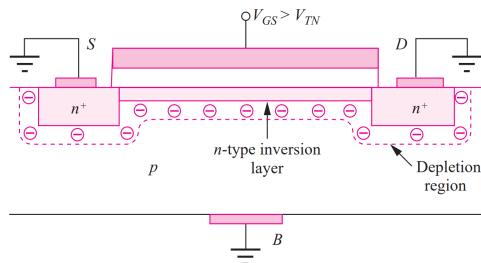


Figura 3.9: Canale di elettroni

**Drain e Source** Come si fa a distinguere il drain dal source? Teoricamente il transistore è un dispositivo a quattro terminali simmetrico dunque possiamo deciderli noi quale è l'uno e quale è l'altro, nei circuiti integrati funziona proprio così. Però in commercio i transistori (i componenti discreti) hanno solitamente tre terminali, questo significa che il costruttore ha deciso per noi chi è chi collegando un terminale con il substrato il quale è il **source**.

### 3.3.3 Calcoliamo il valore della corrente di drain $I_D$

Vogliamo calcolare la corrente che scorre tra drain e source in funzione della tensione sul gate e della tensione sul drain, supponendo di mettere il source a massa.

Per prima cosa possiamo calcolare la corrente del gate  $I_G$ , non essendo collegato a nulla le cariche non possono scorrere anche se il gate si trova alla stessa tensione del generatore  $V_{GS}$ , quindi  $I_G = 0A$ . Si comporta come un condensatore carico.

Solo all'inizio scorre un po' di corrente fino a che il gate si carica.

Supponiamo che  $V_{DS}$  si positiva ma piccola e che  $V_{GS} > V_T$ .

La tensione  $v(x)$  dentro il canale non è costante ma varierà lungo tutto il canale, in prima approssimazione possiamo affermare che vada linearmente da  $V_S = 0$  a  $V_D = V_{DS}$ . Inoltre la corrente di drain è sostanzialmente una corrente di drift.

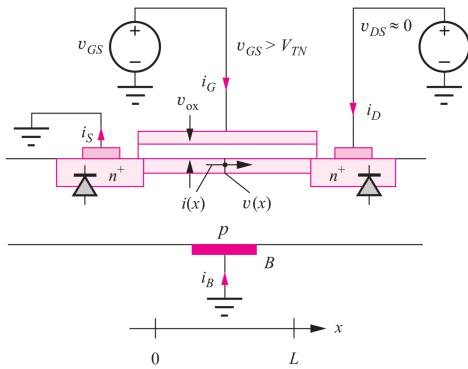


Figura 3.10: Corrente di drain

**Carica nel canale** Quanto vale la corrente di drift? Essa dipende dal campo elettrico, dalla mobilità, dalla velocità e dalla carica.

La carica presente in ogni punto del canale dipende dalla tensione sul dielettrico, come in un condensatore.

**Carica per unità di lunghezza**

$$Q' = -WC''_{ox}(v_{ox} - V_T) \quad (3.1)$$

dove:

- $C''_{ox} = \frac{\epsilon_{ox}}{T_{ox}}$ ,  $\epsilon_{ox}$  permittività dell'ossido e  $T_{ox}$  è il suo spessore
- $v_{ox}(x) = V_{GS} - v_x$ ,  $v_x$  è la tensione nel canale.

$v_{ox}$  deve essere maggiore di  $V_T$  perché ci sia lo strato di inversione su tutto il canale e dunque cariche libere;  $WC''_{ox}$  capacità per unità di lunghezza;  $v_{ox} - V_T$  tensione dell'ossido sottratta alla tensione soglia ovvero elimina il contributo degli ioni negativi fissi i quali non concorrono a creare una corrente.

La tensione sull'ossido varia lungo il canale, se in un condensatore formato da piastre metalliche ogni punto è equipotenziale, in questo caso nel semiconduttore questo non accade, quindi pure la quantità di carica non sarà costante, la differenza di potenziale sarà più alta tra source e gate a sinistra che tra drain e gate a destra Fig. 3.10.

## Corrente

La corrente come al solito è data da:

$$I(x) = Q'(x)s_x(x) \quad s_x(x) \text{ velocità delle cariche dovuta al campo elettrico} \quad (3.2)$$

$$I(x) = -WC_{ox}''(v_{ox} - V_T)(-\mu_n E_x) \quad (3.3)$$

Ricordando il campo elettrico:

$$\vec{E}_x = -\frac{dv(x)}{dx}$$

Sostituendo tutto otteniamo:

$$I(x) = -\mu_n WC_{ox}''(V_{GS} - v_{ox} - V_T) \frac{dv(x)}{dx} \quad (3.4)$$

Integrando lungo x:

- per  $x = 0$ ,  $v(0) = V_S = 0$
- per  $x = L$ ,  $v(L) = V_{DS}$
- $i(x) = I_D$ , la corrente non varia perché non ci sono altri percorsi alternativi

$$\int_0^L i(x)dx = \int_0^L -\mu_n WC_{ox}''(V_{GS} - v_{ox} - V_T)dv(x) \quad (3.5)$$

Risolvendo l'integrale:

$$\begin{aligned} I_D \cdot L &= \mu_n WC_{ox}''(V_{GS} - V_T - \frac{V_{DS}}{2})V_{DS} \\ I_D &= \mu_n \frac{W}{L} C_{ox}''(V_{GS} - V_T - \frac{V_{DS}}{2})V_{DS} \end{aligned} \quad (3.6)$$

Da notare che la corrente dipende dal rapporto di forma del transistore:  $\frac{W}{L}$ , in particolare quando è largo rispetto alla lunghezza. W grande corrente grande, W piccolo corrente piccola.

La corrente dipende dalla  $V_{DS}$  in quanto è la "forza motrice" delle cariche e da  $V_{GS}$  che mette a disposizione gli elettroni per il passaggio di carica. Notiamo qui che la corrente dipende in modo **quadratico** con  $V_{DS}$ .

Valida solo se la tensione sull'ossido è superiore alla soglia  $V_T$  in ogni punto:

- $v_{ox} \geq V_T$
- $V_{GS} - v_{ox} \geq V_T$ ,  $v_{ox}$  è più grande al drain

**In particolare:**

- $V_{GS} - V_{DS} \geq V_T$
- $V_{DS} \leq V_{GS} - V_T$
- $V_{GS} \geq V_{DS} + V_T$

Analizzando l'ultima disequazione: Perché ci sia il canale, sicuramente ci deve essere una tensione  $V_T$ , in più aggiungiamo una tensione sul drain, ne risulta che la tensione gate-source deve essere maggiore di una  $V_T$  rispetto alla tensione del drain.

### 3.4 Regione lineare, ohmica o triodo

Questa regione, che varia quadraticamente, viene chiamata lineare o triodo, la formula della corrente può essere riscritta in modo più compatto utilizzando la transconduttanza ( $K'_N = \mu_n C''_{ox}$ ) che dipende dalla tecnologia (parametro tecnologico), la mobilità dei portatori dipende dal silicio e la capacità che dipende dal fattore geometrico, quale lo spessore, dell'ossido, dunque dipende da chi ve lo produce.

Talvolta nella transconduttanza viene anche inserito il rapporto tra larghezza e lunghezza:  $K_N = \mu_n C''_{ox} \frac{W}{L}$ .

$$I_D = K_n \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS}$$

$$I_D = K_n \left( (V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2} \right) \quad (3.7)$$

**Esempio:** Calcoliamo il valore di  $K'_n$  per un transistore con

- $\mu_n = 500 \text{ cm}^2/\text{Vs}$  mobilità <sup>3</sup>
- $T_{os} = 25 \text{ nm}$  spessore dell'ossido

$$K'_n = \mu_n \frac{\varepsilon_{ox}}{T_{ox}} = 500 \cdot 3.9 \cdot 8.854 \cdot 10^{-14} / 25 = 69.1 \mu\text{A}/\text{V}^2$$

#### 3.4.1 Se $V_{DS}$ è piccola

Si chiama zona lineare perché se  $V_{DS}$  è piccola la formula non compare il termine  $V_{DS}^2$  e dunque il transistore si comporta come una resistenza <sup>4</sup> (da cui anche zona ohmica) il cui valore può essere controllato da  $V_{GS}$ , quindi il modello diventa lineare:

$$I_D = K_n(V_{GS} - V_T)V_{DS} \quad (3.8)$$

E per calcolare la resistenza equivalente basta fare il reciproco della conduttanza G:

$$R_{on} = \frac{1}{K_n(V_{GS} - V_T)} \quad (3.9)$$

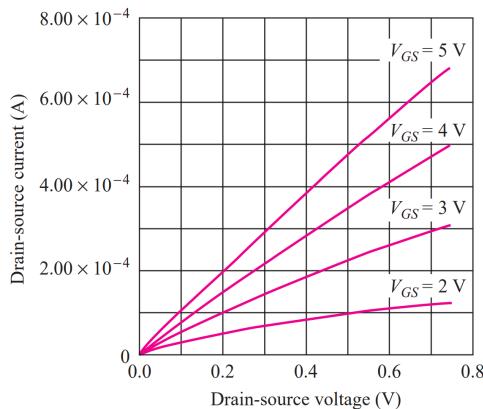


Figura 3.11: Regione lineare

Più si aumenta la  $V_{GS}$  e più la resistenza risulterà bassa perché maggiori saranno i portatori sul canale.

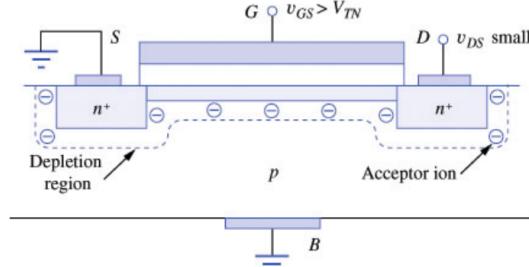
<sup>3</sup>Notiamo che la mobilità ha un valore molto più basso rispetto ai  $1350 \text{ cm}^2/\text{Vs}$ , questo è dovuto al fatto che le cariche si muovono sulla superficie che è a contatto con l'ossido.

<sup>4</sup> $I = G \cdot V$ , dove G è la conduttanza  $G = 1/R$

### 3.5 Saturazione

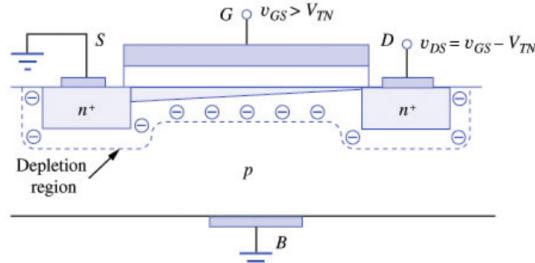
Il modello visto fino ad ora è valido fin tanto che:

- $V_{GS} \geq V_{DS} + V_T$
- La corrente  $I_D$  aumenta in modo quadratico con  $V_{DS}$ .



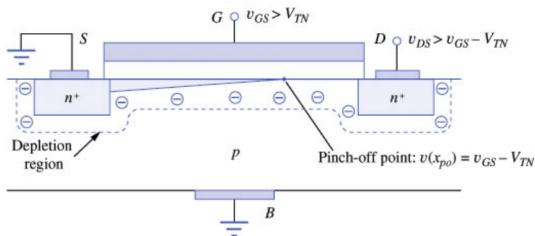
Con l'aumento di  $V_{DS}$  il canale si assottiglia verso il drain fino a scomparire per

- $V_{DS} = V_{GS} - V_T$



Con ulteriori aumenti di  $V_{DS}$  lo stato di inversione diventa nullo al drain e la corrente, non aumentando più, satura. Quindi in prima approssimazione possiamo affermare che la corrente diventa costante, ma vedremo che non è del tutto così, leggermente aumenterà in quanto alla relazione:  $E_x = -\frac{dV(x)}{dx}$ , essendo che la lunghezza del canale diminuisce, aumenterà la tensione.

- $V_{DS} \geq V_{GS} - V_T$



### 3.5.1 E quanto satura?

La corrente satura negli ultimi due casi, in particolare inizia a saturare dal valore per cui:  $V_{DS} = V_{GS} - V_T$ . E la corrente di saturazione sarà data da:

$$I_D = K'_n \frac{W}{L} \left( V_{GS} - V_T - \frac{V_{GS} - V_T}{2} \right) (V_{GS} - V_T)$$

$$I_D = \frac{K'_n W}{2} \frac{W}{L} \left( V_{GS} - V_T \right)^2 \quad (3.10)$$

Quindi vediamo subito che rispetto alla corrente di drain, Formula 3.7, la corrente è **indipendente** da  $V_{DS}$ .

La formula sopra riportata è valida per  $V_{DS} \geq V_{DSAT} = V_{GS} - V_T$ . La tensione  $V_{DSAT}$  è chiamata tensione di **saturazione** o di **pinch-off**.

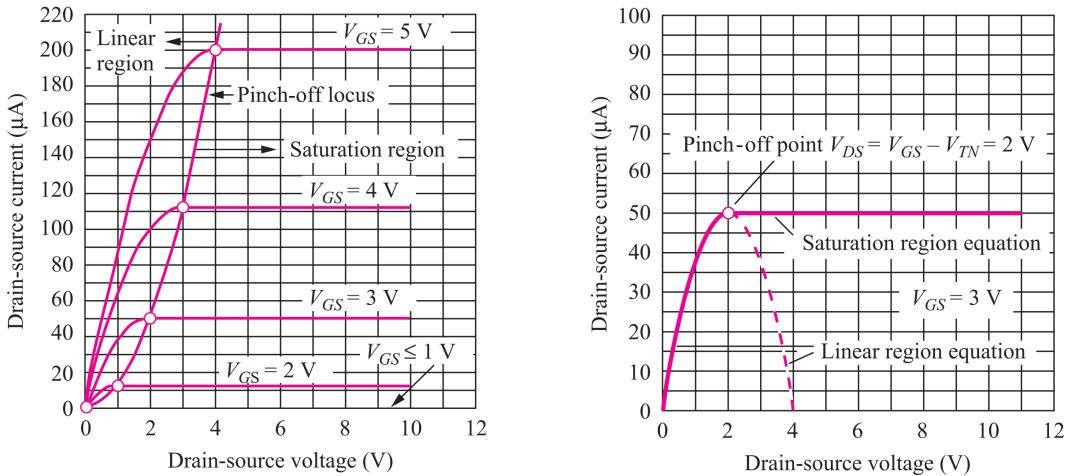


Figura 3.12: Relazione corrente in funzione della tensione al drain

Osservando questo grafico troviamo dunque tre casi:

- $I_{DS} = 0\text{ A}$ 
  - Questo è il caso più semplice perché significa che  $V_{GS} \leq V_T$
- $V_{GS} \geq V_{DS} + V_T$ 
  - parte del grafico che si forma una parabola rivolta verso il basso
  - ci troviamo nella zona lineare
- $V_{DS} \geq V_{GS} - V_T$ 
  - parte del grafico che si forma una retta
  - ci troviamo nella zona di saturazione

**Esempio:** Supponiamo

- $V_T = 1\text{ V}$
- $V_{GS} = 5\text{ V}$
- $V_{DS} = 10\text{ V}$
- $K'_n = 40\mu\text{A}/\text{V}^2$
- $L = 0.35\mu\text{m}$ ,  $W = 8.75\mu\text{m}$

Per trovare  $I_D$  intanto dobbiamo capire in che regione ci troviamo utilizzando le disequazioni viste in precedenza. In questo caso vediamo subito che siamo nella regione di saturazione e dunque possiamo applicare la relativa Formula 3.10 e sostituendo tutti i dati otterremmo  $8\text{ mA}$ .

Da notare il rapporto di forma  $W : L = 25 : 1$

### 3.6 Modulazione di lunghezza di canale

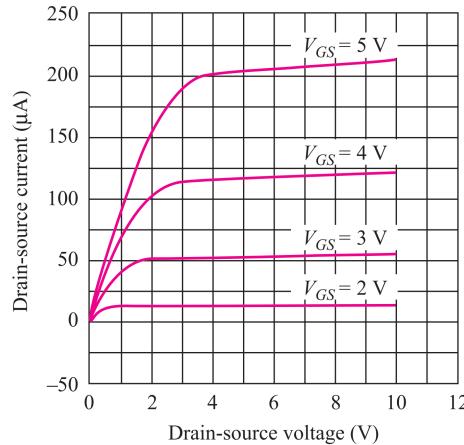
Abbiamo detto che all'aumentare di  $V_{DS}$  la lunghezza del canale L effettiva diventa più piccola con la conseguenza che la corrente  $I_D$  aumenta.

Normalmente nei circuiti ad alte prestazioni si vuole tenere una corrente elevata per andare più velocemente ma ovviamente ciò comporta un aumento di calore e una diminuzione della carica della batteria.

Per tener traccia di questo aumento nella formula vista in precedenza si aggiunge una costante moltiplicativa  $\lambda$  chiamata parametro di modulazione di lunghezza del canale.

Normalmente  $0 \leq \lambda \leq 0.2 V^{-1}$

$$I_D = \frac{K'_n}{2} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \quad (3.11)$$



### 3.7 Modello matematico del transistore nMOS

#### Transconduttanza

$$K_n = K'_n \frac{W}{L} \quad K'_n = \mu_n C''_{ox}$$

#### Zona triodo

$$I_D = K_n \left( V_{GS} - V_{TN} - \frac{V_{DS}}{2} \right) V_{DS} \quad \text{per} \quad 0 \leq V_{TN} + V_{DS} \leq V_{GS}$$

#### Regione di saturazione

$$I_D = \frac{K_n}{2} \left( V_{GS} - V_{TN} \right)^2 \quad \text{per} \quad V_{DS} \geq V_{GS} - V_{TN}$$

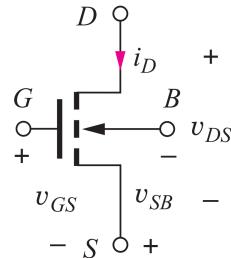


Figura 3.13: Transistore nMOS

### 3.8 Abbassare la tensione di soglia nel canale

Arrivati a questo punto però ci potrebbe essere comodo abbassare questa tensione di soglia, il quale abbassamento farebbe in modo che il transistore possa condurre prima.

Vi sono due metodi: **Transistore nMOS a svuotamento** e **Effetto body**.

### 3.9 Transistore nMOS a svuotamento

Si potrebbe pensare di inserire direttamente il canale n direttamente nel substrato.

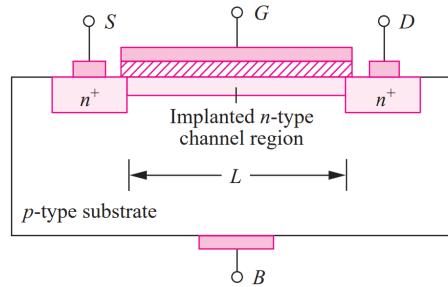


Figura 3.14: Transistore con canale di tipo n già inserito

Il canale dunque esiste già e il transistore conduce anche per  $V_{GS} = 0$ , la tensione di soglia  $V_T$  è negativa, per il resto non cambia nulla.

Questo transistore prende il nome anche di transistore a **depletion mode**.

#### 3.9.1 Effetto body

Altrimenti si potrebbe applicare una tensione al bulk  $V_B$  diversa da zero per modificare la tensione di soglia

$$V_T = V_{T0} + \gamma(\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F}) \quad (3.12)$$

dove

- $V_{T0}$  valore di  $V_T$  per  $V_{SB} = 0$  ,  $-5 \leq V_{T0} \leq 5$
- $\gamma$  parametro effetto body,  $0 \leq \gamma \leq 3$
- $2\phi_F$  potenziale di superficie,  $0.3 \leq 2\phi_F \leq 1$

### 3.10 Transistori pMOS

In modo simmetrico, si può realizzare un transistor su substrato di tipo n, con source e drain p+. Funziona tutto in maniera duale: tensioni e correnti invertire, mobilità delle lacune inferiore.

Per ottenere delle tensioni negative, non utilizzate nei dispositivi elettronici, basta mettere il source alla tensione più alta presente nel circuito, prima era a massa, e di conseguenza tutte le altre tensioni saranno inferiori o al più uguali a  $V_S$ .

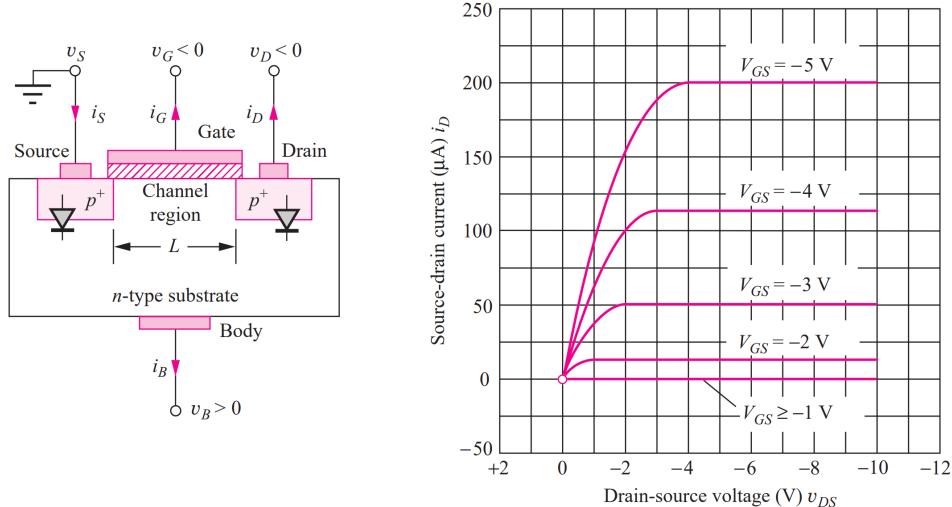


Figura 3.15: Transistore pMOS

### 3.11 Modello matematico del pMOS

#### Transconduttanza

$$K_p = K'_p \frac{W}{L} \quad K'_p = \mu_p C_{ox}''$$

#### Zona triodo

$$I_D = K_p \left( V_{GS} - V_{TP} - \frac{V_{DS}}{2} \right) V_{DS} \quad \text{per} \quad 0 \leq |V_{DS}| \leq |V_{GS} - V_{TP}|$$

#### Regione di saturazione

$$I_D = \frac{K_p}{2} \left( V_{GS} - V_{TP} \right)^2 \quad \text{per} \quad |V_{DS}| \geq |V_{GS} - V_{TP}|$$

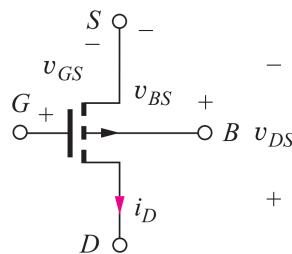


Figura 3.16: Transistore pMOS

### 3.12 Riassunto

Nel transistore MOSFET troviamo quattro terminali:

Gate	Source	Drain	Body
------	--------	-------	------

e tre zone di funzionamento:

**Cut-off:** se  $V_{GS}$  è sotto la tensione di soglia  $V_T$ , ciò non fa scorrere corrente tra  $I_D$  tra drain e source

**Triodo:** se  $V_{GS}$  è sopra la tensione di soglia  $V_T$  lungo tutto il canale, la corrente  $I_D$  tra drain e source dipende da  $V_{GS}$  e da  $V_{DS}$  secondo la legge quadratica :

$$I_D = K_n \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS} \quad \text{per} \quad V_T \leq V_{GS} \geq V_{DS} + V_T$$

**Saturazione** se  $V_{GS}$  è sopra la tensione di soglia  $V_T$  per parte del canale, la corrente  $I_D$  tra drain e source non dipende da  $V_{DS}$  e dipende da  $V_{GS}$  secondo una legge quadratica:

$$I_D = \frac{K_n}{2} \left( V_{GS} - V_T \right)^2 \quad \text{per} \quad V_{GS} \leq V_{DS} + V_T$$

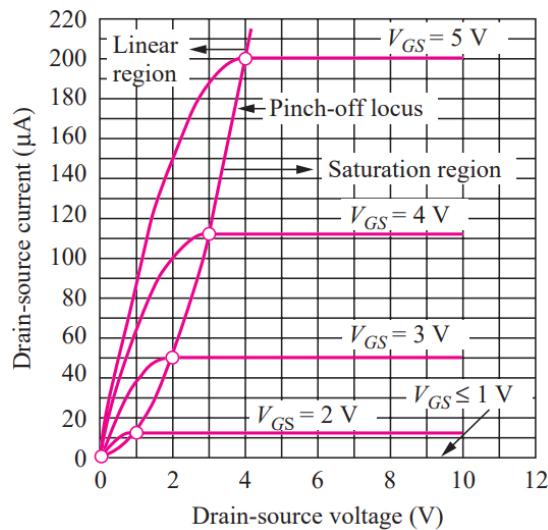


Figura 3.17: Caratteristica I-V

### 3.13 Modello di analisi

Occorre stabilire la regione di funzionamento del transistore in modo da applicare la formula giusta. Se difficile da dire a priori, si fanno delle ipotesi e poi si verifica che i risultati siano consistenti con le ipotesi (come già fatto per i diodi).

Spesso si faranno delle semplificazioni, considerando solo il punto di inizio e di fine di un transitorio, per evitare di dover integrare lungo tutta la caratteristica.

#### Circuiti digitali e amplificatori

Per circuiti digitali si lavora in cut-off (interruttore aperto) o in regione lineare (interruttore chiuso)

#### Amplificatori

Per amplificatori si lavora in regione di saturazione (generatore di corrente costante, controllata dalla tensione di ingresso)

Bisogna fare attenzione che la  $V_{DS}$  dipende dal resto del circuito, infatti non è imposta direttamente da un generatore di tensione ma dipende dalla corrente di drain.

#### 3.13.1 Calcolo con la retta di calcolo

Dato il seguente circuito, formato da un singolo transistore con source a massa e gate e drain collegato ad un generatore di tensione.

Assumendo  $V_T = 1V$  e  $K_n = 220\mu A/V^2$ , trovare la corrente e la tensione tra drain e source.

$V_T$  è possibile trovarla facilmente. Basta mettersi in un punto di pinch-off e fare la differenza tra  $V_{GS}$  e  $V_{DS}$ .

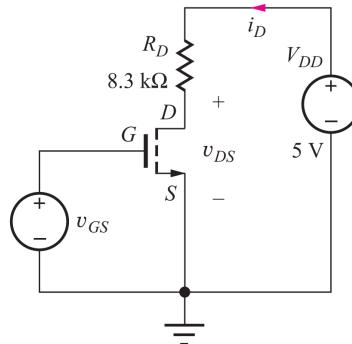


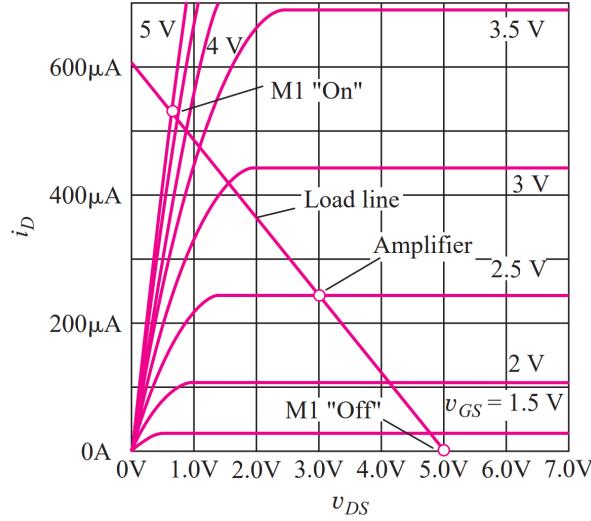
Figura 3.18: Circuito

Per prima cosa dobbiamo trovare la retta di carico, ed utilizziamo l'equazione alla maglia del circuito a destra.

$$V_{DD} = I_D R + V_{DS}$$

Trovando che se la corrente è zero allora  $V_{DS} = V_{GS}$ , e per trovare la massima corrente applicheremo la legge di Ohm tra la tensione  $V_{DD}$  e la resistenza R.

Fatto ciò possiamo tracciare la retta di carico:



Ora vediamo a vari valori di  $V_{GS}$  come si comporta.

$$V_{GS} \leq V_T$$

In questo caso vediamo che la corrente di drain è zero pure la corrente nella resistenza è zero

**Troviamo M1** Supponiamo ora di essere nella regione lineare e che  $V_{GS} = V_{DD}$ , e che la tensione sia tutta sulla resistenza, dunque che  $V_{DS}$  sia piccola.

Utilizzando la Formula 3.6 e l'equazione della retta di calcolo vista prima, mettendole a sistema possiamo trovare  $V_{DS}$  e  $I_D$ .

$$V_{DD} = K_n \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS} R + V_{DS}$$

$$\frac{K_n R}{2} V_{DS}^2 - [1 + K_n R(V_{GS} - V_T)] V_{DS} + V_{DD} = 0$$

$$V_{DS} = \frac{1 + K_n R(V_{GS} - V_T) \pm \sqrt{\left(1 + K_n R(V_{GS} - V_T)\right)^2 - 2K_n R V_{DD}}}{K_n R}$$

$$V_{DS} = 0.65V \quad I_D = 525\mu A$$

### Supponiamo di essere in saturazione

Con  $V_{GS} = 2.5V$  notiamo immediatamente dal grafico che ci troviamo in saturazione, proviamo dunque a calcolare come prima tensione e corrente, utilizzando sempre l'equazione della retta di calcolo e anche la Formula 3.10.

$$V_{DD} = \frac{K_n}{2} (V_{GS} - V_T)^2 R + V_{DS}$$

$$V_{DS} = V_{DD} - \frac{K_n}{2} (V_{GS} - V_T)^2 R \quad (3.13)$$

$$V_{DS} = 2.95V \quad I_D = 247.5\mu A$$

In questa zona il transistore ha una corrente costante e dunque si comporta come un generatore di corrente, e il valore della corrente lo controlliamo con la  $V_{GS}$ . Per piccole variazioni di  $V_{GS}$ , questo comporta il muoversi sulla retta di carico, causano grosse variazioni di  $V_{DS}$ , funziona come un amplificatore.

### 3.13.2 Come determinare l'uscita $V_{DS}$ in funzione dell'ingresso $V_{GS}$

Vorremmo capire che relazione ci sia tra  $V_{DS}$  e  $V_{GS}$ , sicuramente dovremo vedere tutte le zone di funzionamento.

#### Zona di Cut-off

Partendo da  $V_{GS}$  piccola, fintantoché questa è inferiore a  $V_T$ , corrente nel drain non ne scorre, non c'è caduta di potenziale sulla resistenza e l'uscita è 5V.

Questa zona infatti risulterà essere piatta.

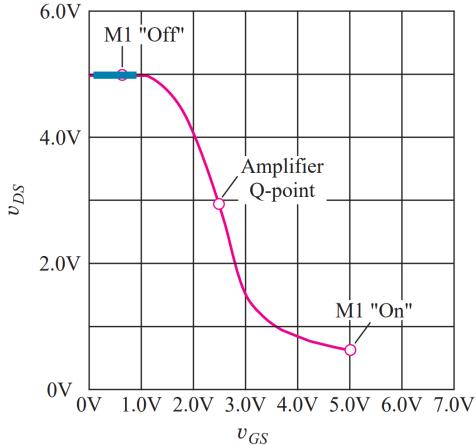


Figura 3.19:  $V_{GS} < V_T$

#### Zona di amplificazione/saturazione

Per ricavare la curva in questo caso basta assegnare dei valori a  $V_{GS}$ . Prendendo le equazioni della corrente e della retta di carico come fatto prima per ottenere la Formula 3.13 e sostituendoci  $V_{GS} = V_{DS} + V_T$  per trovare la zona di confine.

Questo vale fino a che  $V_{GS} \leq V_{DS} + V_T$

$$V_{DS} = \frac{-2 \pm \sqrt{4 + 8K_n R V_{DD}}}{2K_n R}$$

Una soluzione verrà negativa che ovviamente va scartata perché non c'è nessuna tensione negativa in questo circuito.

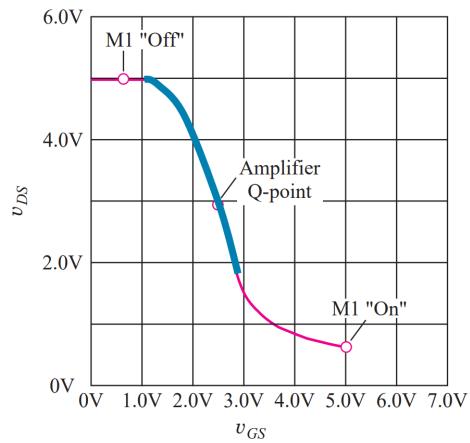


Figura 3.20:  $V_{GS} \leq V_{DS} + V_T$

## Zona di triodo

Successivamente per  $V_{GS} > V_{DS} + V_T$  il transistore funziona in zona trodo.

$$V_{DS} = \frac{1 + K_n R(V_{GS} - V_T) \pm \sqrt{(1 + K_n R(V_{GS} - V_T))^2 - 2K_n R V_{DD}}}{K_n R}$$

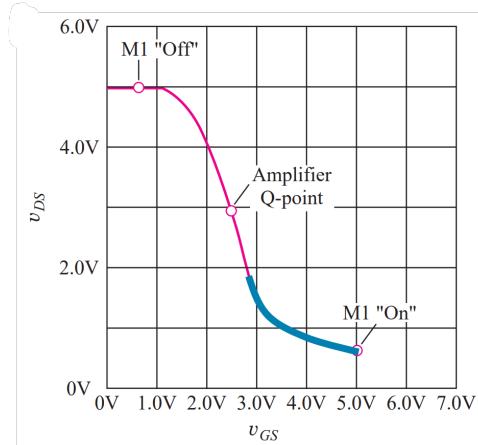


Figura 3.21:  $V_{GS} > V_{DS} + V_T$

Si può vedere che se l'ingresso è basso l'uscita è alta; se l'ingresso è alto l'uscita è bassa, anche perché se fosse l'ingresso ad un altro transistore sarebbe sotto soglia (1V). Si comporta come invertitore logico.

Nella zona intermedia si comporta in modo differente: un piccolo cambiamento di  $V_{GS}$ , l'ingresso, comporta un grande cambiamento per l'uscita  $V_{DS}$ . La derivata in valore assoluto è maggiore di 1.

### 3.14 Saturazione di velocità

Come accennato in precedenza, all'aumentare del campo elettrico i portatori raggiungeranno un limite di velocità, oltre ( $\vec{E} > 10^5 \text{ V/cm}$ ) non aumentano più.

In maniera approssimata possiamo dire che  $\vec{E} = \frac{V_{DS}}{L}$ , dunque aumentiamo il campo elettrico aumenta la corrente (perché aumenta la velocità dei portatori). Se la velocità satura, aumentando il campo elettrico ma la velocità non aumenta e nemmeno la corrente la quale diventa costante.

Una saturazione di velocità porta inevitabilmente ad una saturazione di corrente, non dovuta al fatto che non vi è più inversione ma è dovuta al fatto che i portatori non possono andare più velocemente. Si potrebbe arrivare prima in saturazione di velocità rispetto alla saturazione di pinch-off, questo dipende dalla lunghezza del canale.

Normalmente con le tecnologie moderne, siamo sempre in saturazione di velocità, e soprattutto vogliamo correnti elevate.

Dunque ad un certo punto, pur aumentando  $V_{DS}$  e quindi il campo elettrico  $\vec{E}$ , la velocità non aumenta più.

Supponiamo che la saturazione si raggiunga per  $V_{DS} = V_{SAT} = \vec{E} \cdot L$ , la corrente non aumenta più in funzione a  $V_{DS}$ .

**Nota:** Non confondete la **regione di saturazione** con la **saturazione di velocità**. Si può avere saturazione di velocità in zona triodo, o non avere saturazione di velocità in regione di saturazione!

Se siamo in regione di saturazione con **saturazione di velocità**, la corrente cresce linearmente con  $V_{GS}$ , invece che quadraticamente!

**Esempio:** si consideri  $L = 1\mu\text{m}, 100\text{nm}, 22\text{nm}, 5\text{nm}$ . Troviamo a che valore di  $V_{DS}$  causa la saturazione di velocità:

$$V_{SAT} = 10^{-4} \cdot 10^5 = 10\text{V}$$

$$V_{SAT} = 10^{-5} \cdot 10^5 = 1\text{V}$$

$$V_{SAT} = 22 \cdot 10^{-6} \cdot 10^5 = 0.22\text{V}$$

$$V_{SAT} = 5 \cdot 10^{-7} \cdot 10^5 = 0.05\text{V}$$

Se prendiamo in considerazione il secondo caso, è normale usare 1V nei circuiti, ma talvolta potremmo avere tensioni più alte come 1.8V e a quel punto saremo in saturazione di velocità.

Con 22nm siamo praticamente sempre in saturazione di velocità perché la tensione di alimentazione non è ancora arrivata ad essere 0.22V, ma non potrà mai essere sotto a 0.6 V perché il transistore ha una tensione di soglia e se si scende sotto ad essa non si può proprio accendere. Per ovviare a questo ci sono altri metodi.

#### Saturazione di velocità

La zona di saturazione è un limite che sotto al quale possiamo essere o in zona **triodo** o in nella regione di **pinch-off**. L'espressione singola che riassume il tutto e la seguente:

$$I_D = K_n \left( V_{GS} - V_T - \frac{V_{MIN}}{2} \right) V_{MIN} (1 - \lambda V_{DS}) \quad (3.14)$$

dove  $V_{MIN}$  :

$$V_{MIN} = \min((V_{GS} - V_T), V_{DS}, V_{SAT})$$

quindi il minimo tra, in ordine, saturazione, triodo e saturazione di velocità.

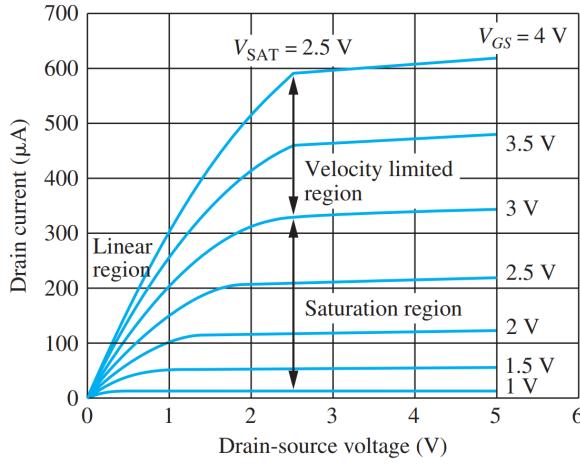


Figura 3.22: Regione di saturazione

### 3.14.1 Conduzione sotto soglia

Abbiamo detto che per  $V_{GS} < V_T$  la corrente di drain è nulla. Ma anche questa è una approssimazione, infatti ci sono comunque dei portatori e ci sono le correnti inverse dei diodi.

Se  $V_{DS} > 0$  scorre una debole corrente, più la tensione di soglia è bassa, più è alta la corrente sotto soglia. Ciò può diventare significativa quando ci sono miliardi di dispositivi!

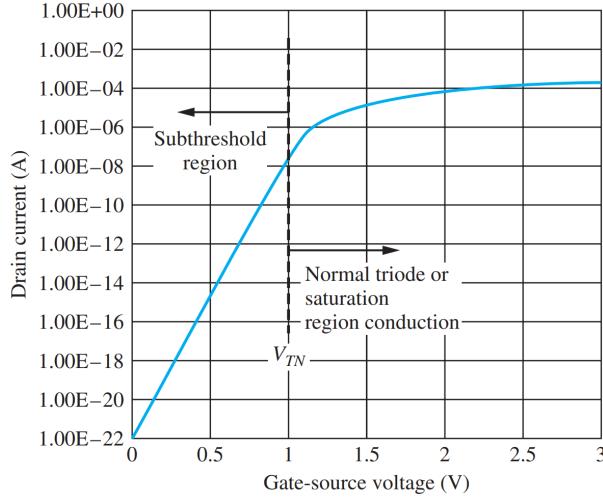


Figura 3.23: Corrente sotto soglia, con  $V_{DS}$  fissata

Inoltre tanto più bassa è la soglia del dispositivo, tanto più alta è la corrente sotto soglia. Sarebbe bello fare dei transistori con tensione di soglia molto piccola perché diminuirebbe l'assorbimento di potenza, ma ciò non è possibile perché conduce.

Alcuni circuiti basano il loro funzionamento proprio sul fatto che vi è questa corrente sotto soglia, ma ciò comporta che questi circuiti vadano bene se vogliamo costruire dispositivi non preformanti: il sensore di temperatura.

Ad alte prestazioni questa corrente sotto soglia comporta che il dispositivo consuma anche quando sta fermo, idealmente una porta consuma quando si deve scaricare e caricare, non quando sta ferma. Questo consumo non possiamo evitarlo il quale si attesta sull'ordine dei  $mA$ .

# Capitolo 4

## Invertitore Logico

Come primo esempio di porta logica vediamo una delle porte più semplici: l'**invertitore**.

Quando la tensione in ingresso è piccola fornisce una tensione alta in uscita, e viceversa.

### 4.1 Funzione di trasferimento e caratteristica

- $V_+$  e  $V_-$  sono le tensioni di alimentazione,  $V_+$  alle volte è chiamata  $V_{DD}$  (drain) e altre  $v_{CC}$  (collettore) per i transistori bipolari.
- $V_H$  e  $V_L$  sono i livelli alti e bassi della tensione di uscita

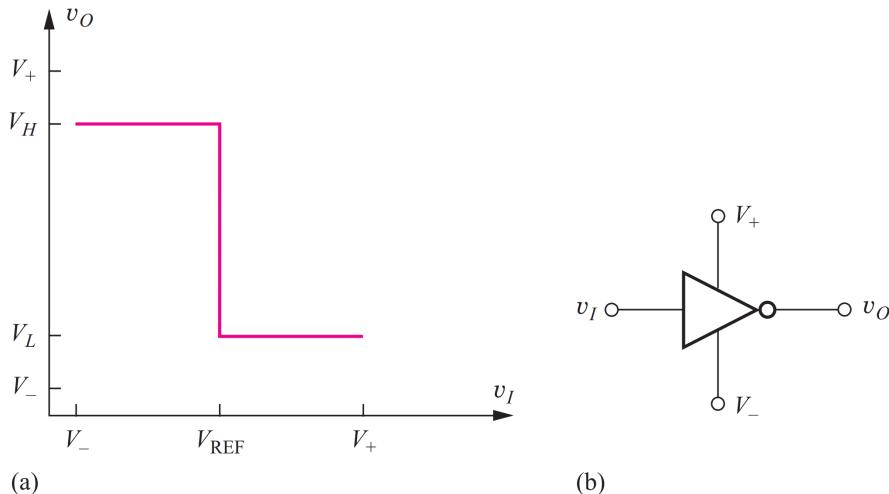


Figura 4.1: Funzione ingresso/uscita e simbolo circuituale

#### 4.1.1 Schema di base dell'invertitore

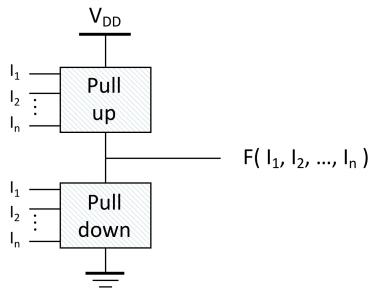
Per ottenere il risultato che ci aspettiamo da un invertitore dobbiamo utilizzare due circuiti in modo tale da pilotare l'uscita verso il valore logico alto o basso, chiamati **pull-up** e **pull-down**.

La rete di pull down porta l'uscita al valore basso;

La rete di pull up porta l'uscita al valore alto.

Dunque le tensioni in ingresso NON vengono usate per portare l'uscita al valore desiderato per solamente per attivare i due circuiti, è il circuito di pull-up/down che pilota l'uscita.

In questo modo la tensione in uscita tende verso il valore del circuito che "tira più forte".



	Pull-up OFF	Pull-up ON
Pull-down OFF	Z (float) <sup>a</sup>	1
Pull-down ON	0	Dipende <sup>b</sup>

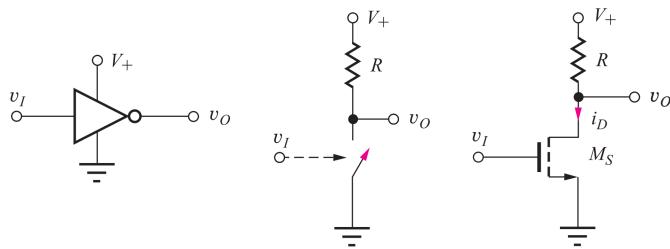
<sup>a</sup>alta impedenza, l'uscita può mantenere il livello, scaricarsi o caricarsi anche rispetto alle condizioni circostanti dato il fenomeno del campo magnetico generato attorno al filo

<sup>b</sup>Il circuito è come un partitore di tensione, chi tira di più vince

#### 4.1.2 Interruttore con carico resistivo

Il modo di realizzare un invertitore è quello di creare dei cammini verso  $V_{DD}$  e verso massa. L'inverter più semplice è quello denominato a **carico resistivo** costituito da una resistenza che collega sempre l'uscita a  $V_{DD}$  con un impedenza diversa da zero la quale tira verso l'alto; e un circuito di pull-down realizzato con un transistore MOS, interruttore, che quando  $v_i$  è bassa, l'interruttore è aperto (non tira verso massa) e l'uscita è alta, fase di **pull-up**.

Quando  $v_i$  è alta, l'interruttore è chiuso, e l'uscita è connessa a massa dal transistore, se quest'ultimo tira più forte di R, a massa, fase di **pull down**.



## 4.2 Caratteristica reale

Ovviamente prima abbiamo visto un grafico dove si mostrava il comportamento ideale di questo circuito, in realtà l'andamento non sarà lineare. Agli estremi troveremo le zone che interessano ai circuiti elettronici digitali, i quali rappresentano 0 o 1, mentre della zona centrale è detta di **amplificazione**.

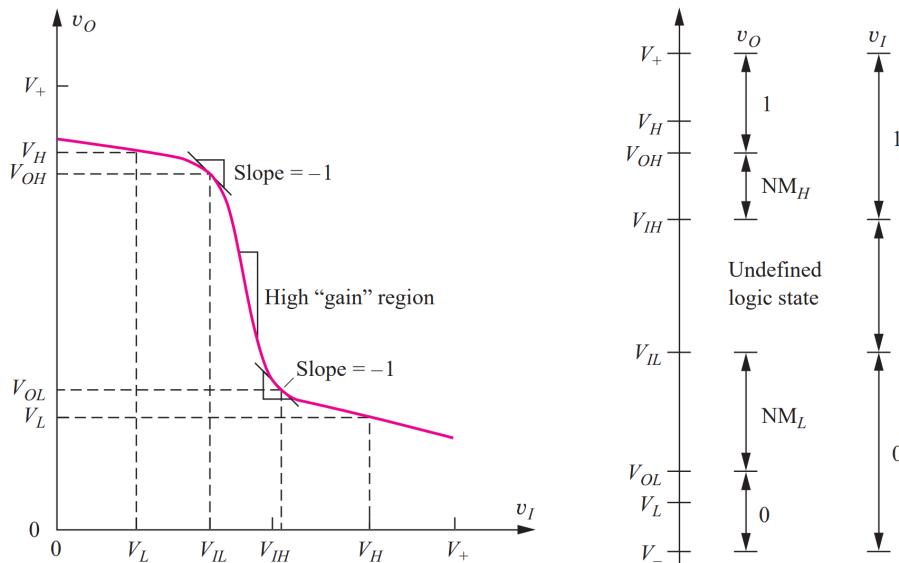


Figura 4.2: Caratteristica reale invertitore resistivo

### 4.3 Margini di rumore

Come posso definire il cambiamento dello stato? Una prima grossolana approssimazione potrebbe essere quella di dividere a metà, con una soglia, la tensione di alimentazione. Tutto quello che sta al di sopra lo possiamo considerare altro e quello che sta sotto basso.

Questo metodo però non porterebbe a risultati ottimali in quanto se un segnale è a ridosso della soglia, solo un po' di rumore porterebbe a far variare l'uscita.

Dunque una soluzione è quello di stare lontano dalla soglia e avere un **margine di errore**. Come possiamo definirlo questo margine di errore?

Osservando quando nel grafico della caratteristica si incontrano i punti in cui la tangente alla curva tende al valore  $-1$ : le tensioni per le quali la tangente vale  $-1$  discriminano il livello alto e basso.

Questo perché possiamo discriminare le parti del circuito che si comportano come amplificazione, infatti nella zona centrale la derivata ha un numero in valore assoluto più grande di  $1$ , quindi presenta un amplificazione.

Le zone più esterne invece, dove la tangente ha un valore  $-1 < x < 0$ , l'inverter attenuerà, in particolar modo se vi è del rumore ne riduce di molto la componente.

**Notazione:**

- $V_{IL}$  : tensione massima per la quale l'ingresso è considerato basso;
- $V_{IH}$  : tensione minima per la quale l'ingresso è considerato alto;
- $V_{OL}$ : valore estremo di uscita per  $v_I = V_{IH}$ ;
- $V_{OH}$ : valore estremo di uscita per  $v_I = V_{IL}$ .
- $V_{OL}$ : valore estremo di uscita per  $v_I = V_{IH}$ ;
  
- $NM_H = V_{OH} - V_{IH}$
- $NM_L = V_{IL} - V_{OL}$

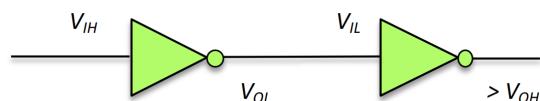
La zona di mezzo, amplificazione, per questa tipologia di scopo e circuiti non si dovrà considerare.

Utilizzando più porte logiche, se l'ingresso abbiamo  $V_{IH}$ , in uscita ci troviamo  $V_{OL}$ , la quale a sua volta deve essere riconoscere come tensione bassa.

Dunque  $V_{OL}$  deve essere inferiore alla massima tensione che l'inverte a valle riconosce come bassa che è  $V_{IL}$ :  $V_{OL} \leq V_{IL}$ , se verificata il circuito funziona.

Nel caso opposto avremo che:  $V_{OH} \geq V_{IH}$ .

Questa differenza è definita **Margine di rumore**.



## 4.4 Obiettivi di progetto

Ora dobbiamo progettare l'inverter in modo tale da renderlo migliore:

- Minimizzare la regione indefinita, la caratteristica deve essere la più verticale possibile;
- Massimizzare i margini di rumore, dunque oltre al rumore nel circuito potremmo utilizzare componenti con tolleranze più grandi (processo meno preciso = costo minore);
- Dispositivo unidirezionale, l'uscita non deve influenzare l'ingresso;
- I livello di I/O devono essere compatibili, ovvero margini di rumori positivi, se fossero negativi non potremmo costruire circuiti in casata;
- Massimizzare gli indici le prestazioni:
  - Area più piccola possibile (costo minore);
  - Minimizzare i consumi di potenza;
  - Operare alla massima velocità;

Normalmente gli indici di prestazione sono in conflitti tra di loro.

## 4.5 Circuito invertitore nMOS

Il circuito per realizzare un invertitore è uguale all'amplificatore utilizzato in zona non lineare.

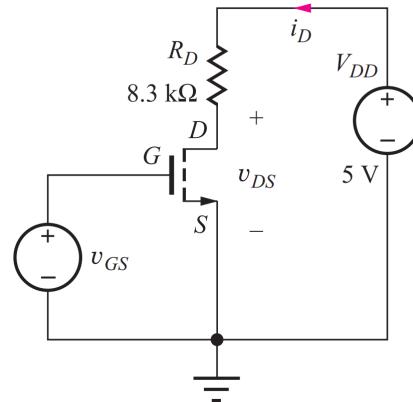


Figura 4.3: Circuito

L'uscita la prendiamo attaccando un filo sul drain:  $V_{DS} = V_O$ . A questo punto, al contrario dell'analisi, ora dobbiamo progettarlo con determinate prestazioni.

Messo per iscritto che utilizzeremo un inverter a carico resistivo e che la topologia sia quella in figura, dobbiamo scegliere: la dimensione di  $R$  e determinare il valore del rapporto  $W/L$ .

La  $L$  normalmente viene fissata con la tecnologia minima di produzione in quel momento (180 nm, 20 nm, ecc) si sceglie  $L$  più piccola in modo da portare più corrente possibile e minore dunque sarà la  $W$  e in generale il transistore diventa più piccolo.

### Obiettivo:

- Rendere  $V_L$  abbastanza bassa (sicuramente più bassa della  $V_{GS}$ ) e  $V_H = V_{DD}$ ;
- Minimizzare il consumo di potenza;
- Massimizzare le prestazioni

### 4.5.1 Scelta di $V_L$

**La tensione di uscita:** deve pilotare la porta logica successiva, ovvero deve mettere in interdizione il gate del transistor successivo per metterlo in cut-off:  $V_L < V_{TN}$ , solitamente un valore appropriato si aggira tra 25% e il 50% di  $V_{TN}$ .

Nei transistori normalmente la  $V_{TN} = 0.6V$ , dunque possiamo scegliere una  $V_L = 0.2V$ , ovvero un terzo.

**Consumi di potenza:** Se il transistore è interdetto il generatore non eroga corrente (a meno quella per caricare il gate del transistor a valle), anche se la tensione rimane alta.

Quando  $v_I$  è alta, scorre corrente e si consuma la batteria. Supponendo  $V_{DD} = 2.5V$ , vicino agli standard di oggi, scegliendo un consumo di potenza che ci aggrada,  $P = 0.2 \text{ mW}$ , dettato dai parametri come:

- energia disponibile nella batteria;
- numero di porte logiche;
- la media di quante queste hanno l'uscita a zero;
- quanta potenza il sistema riesce a dissipare.

Partendo dal consumo totale ad esempio 0.2 W perché così so che la batteria mi dura due giorni, calcolando 2000 porte logiche di cui la metà a valore logico altro (potenza = 0), ottengo che ogni porta logica dovrà consumare 0.2 mW.

### 4.5.2 Dimensionamento

Come detto precedentemente supponiamo:

$$V_{TN} = 0.6V \quad V_{DD} = 2.5V \quad K'_n = 100\mu A/V^2$$

Per soddisfare i consumi di potenza:

$$V_{DD} \cdot I_D = 0.2mW \quad \rightarrow \quad I_D = 0.2/2.5 = 80\mu A$$

Questa è la corrente che deve scorrere quando l'uscita è bassa (la porta consuma).

**Calcolo R:**

$$I_D \cdot R = 2.5 - V_L = 2.5 - 0.2 \quad \rightarrow \quad R = (2.5 - 0.2)/0.00008 = 28.75K\Omega$$

Si poteva anche ottenere grazie alla retta di carico, supponendo di avere un transistore discreto.

**Calcolo il rapporto di forma W/L:**

$$V_i = V_{GS} = V_h = 2.5V, \quad V_{DS} = 0.2V$$

Trovandoci in zona triodo ( $V_{GS}$  molto alta,  $V_{DS}$  piccola, la tensione di soglia è 0.6 V, quindi  $0.2 + 0.6 = 0.8V$  la  $V_{GS}$  è molto più alta e dunque siamo in zona triodo) possiamo utilizzare la formula della zona triodo per trovare il rapporto W/L:

$$\frac{W}{L} = \frac{80\mu A}{100\frac{\mu A}{V^2}(2.5 - 0.6 - \frac{0.2}{2})0.2} = 2.22 = \frac{2.22}{1}$$

Ora abbiamo tutti le dimensioni dei componenti.

Non abbiamo considerato in questo caso le tempistiche. Queste ultime dipendono da quanto veloce la resistenza riesca a caricare il/i condensatore/i (gate) a valle.

Più piccola sarà R più passa corrente e più verrà caricato in fretta il gate, per calcolare il tempo dobbiamo considerare il circuito come se fosse un RC, e con  $\tau = RC$  ci possiamo ricavare quanto tempo impiega a caricarsi e scaricarsi.

Scegliendo una R piccola però significa che, si andremo più veloci, ma stiamo consumando anche più corrente. Vi è sempre lo scontro prestazioni e consumi.

### 4.5.3 Calcoliamo i margini di rumore: $V_{IL}$ e $V_{OH}$

Quando  $v_I = V_{GS}$  è piccola, ma maggiore della tensione di soglia, e  $v_o = V_{DS}$  è grande ci troviamo sicuramente in **saturazione**.

Utilizzando la corrente di drain e il valore dell'uscita  $v_o = V_{DD} - I_D R$ , otteniamo:

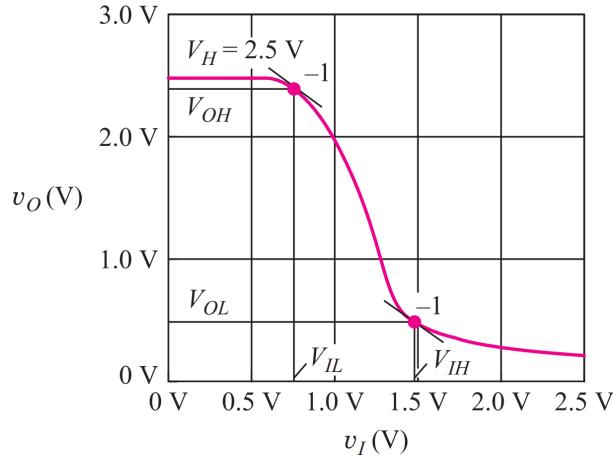
$$V_o = V_{DD} - \frac{K_n}{2}(V_{GS} - V_{TN})^2 R$$

Cerchiamo dunque quando la tangente vale  $-1$ : la si ottiene facendo il rapporto tra la derivata in uscita e la derivata in ingresso

$$\frac{dV_o}{dV_{GS}} = -K_n(V_{GS} - V_{TN})R = -1$$

$$V_{IL} = V_{GS} = V_{TN} + \frac{1}{K_n R}$$

$$V_{OH} = V_{DD} - \frac{1}{2K_n R}$$



Per il valore logico alto, quando  $v_I = V_{GS}$  è grande  $v_o = V_{DS}$  è piccola siamo in zona triodo. Come prima dobbiamo sempre derivare e uguagliare a  $-1$ , in questo caso la derivata risulterà più complicata, infatti risulta più semplice assumere che:  $\frac{dV_o}{dV_{GS}} = \frac{dV_{GS}}{dV_o}^{-1}$

Il risultato diverrà:

$$V_O = V_{DD} - K_n \left( V_{GS} - V_{TN} - \frac{V_O}{2} \right) V_O R$$

$$K_n R V_{GS} - K_n R V_{TN} - \frac{K_n R}{2} V_O - \frac{V_{DD}}{V_O} = -1$$

$$K_n R V_{GS} = -1 + K_n R V_{TN} - \frac{K_n R}{2} V_O + \frac{V_{DD}}{V_O}$$

$$V_{GS} = -\frac{1}{K_n R} + V_{TN} + \frac{1}{2} V_O + \frac{1}{K_n R} \frac{V_{DD}}{V_O}$$

$$\frac{dV_{GS}}{dV_O} = \frac{1}{2} - \frac{1}{K_n R} V_{DD} \frac{1}{V_O^2} = -1$$

$$\frac{3}{2} V_O^2 = \frac{V_{DD}}{K_n R} \quad \rightarrow \quad V_O = \left( \frac{2}{3} \frac{V_{DD}}{K_n R} \right)^{\frac{1}{2}}$$

$$V_{GS} = V_T - \frac{1}{K_n R} + \frac{1}{2} \left( \frac{2}{3} \frac{V_{DD}}{K_n R} \right)^{\frac{1}{2}} + \frac{V_{DD}}{K_n R} \left( \frac{3}{2} \frac{K_n R}{V_{DD}} \right)^{\frac{1}{2}} = V_T - \frac{1}{K_n R} + \frac{1}{2} \left( \frac{2}{3} \frac{V_{DD}}{K_n R} \right)^{\frac{1}{2}} + \left( \frac{3}{2} \frac{K_n R}{V_{DD}} \right)^{\frac{1}{2}}$$

Pertanto:

$$V_{IH} = V_{TN} - \frac{1}{K_n R} + 1.63 \left( \frac{V_{DD}}{K_n R} \right)^{\frac{1}{2}}$$

$$V_{OL} = \sqrt{\frac{2V_{DD}}{3K_n R}}$$

**Esempio di calcolo** Per il nostro invertitore abbiamo deciso i seguenti valori:  $V_{TN} = 0.6V$ ,  $K_n = 100 \cdot 2.22 = 222 \mu A/V^2$ ,  $R = 28.8 K\Omega$ ,  $K - nR = 6.39$ , sostituendo i valori otteniamo i seguenti valori:

- $V_{IL} = 0.756V$  il massimo valore che viene riconosciuto come basso dall'inverter;
- $V_{IH} = 1.46V$  ingresso minimo che viene riconosciuto come alto;
- $V_{OH} = 2.42V$  uscita in corrispondenza di  $V_{IL}$ , vicina a 2.5V;
- $V_{OL} = 0.51V$  uscita in corrispondenza di  $V_{IH}$ , vicina a 0.5V comunque sotto la soglia anche se di poco.

Calcolo dei margini di rumore:

$$NM_H = V_{OH} - V_{IH} = 0.96V \text{ è un buon valore: si ha quasi un volt di tolleranza}$$

$$NM_L = V_{IL} - V_{OL} = 0.25V \text{ non un buon valore: si ha poco margine}$$

### Resistore

Dobbiamo preoccuparci anche di come costruire il resistore sul circuito integrato. La resistenza sappiamo che dipende dalla resistività:

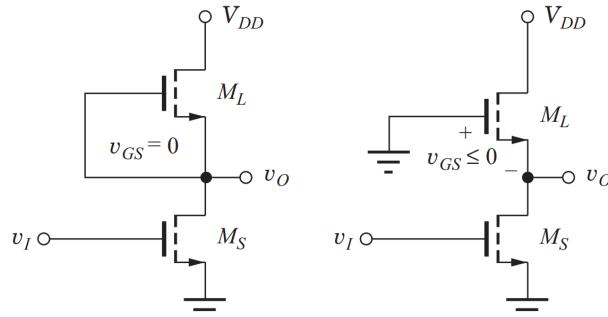
$$R = \rho \frac{L}{TW} = 28.8 K\Omega$$

Vogliamo T e W piccole e la lunghezza L lunga per avere una resistenza elevata. Supponiamo che  $T = 1 \mu m$ ,  $\rho = 0.001 \Omega cm$ , e con una tecnologia produttiva di un micron mettiamo  $W = 1 \mu m$ . Facendo così però la lunghezza dovrà essere:  $2880 \mu m$ , una cosa gigantesca se si pensa che un circuito integrato abbastanza grande è  $2 \times 2 cm$ .

Ma soprattutto è enorme rispetto al solo transistor ( $W/L = 22.2$ , con  $L = 1 \mu m$ ) il quale avrà un area di  $2.22 \mu m^2$  a confronto con i  $2880 \mu m^2$  della resistenza.

Dunque le resistenze nei circuiti integrati sono di difficile realizzazione anche per il fatto che il silicio non è un materiale adatto all'isolamento elettrico.

Per ovviare a questo problema cerchiamo di realizzare il carico con un altro transistor in particolare un altro nMOS con gate a potenziale fisso.

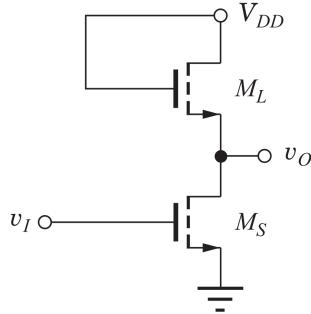


I due circuiti sopra in figura non funzionano quando:

- Nel primo caso,  $V_{GS} = 0$ , quindi  $M_L$  è sempre interdetto
- Nel secondo caso,  $V_{GS} < 0$ , quindi  $M_L$  è sempre interdetto

Una topologia che potrebbe funzionare potrebbe essere la seguente

## 4.6 Inverter con carico saturato



Questo transistore,  $M_L$ , si potrà trovare o in zona di cut-off o in zona di saturazione, quasi sempre sarà in quest'ultima. Questo tipo di carico viene chiamato saturato perché appunto il transistore sarà in saturazione.

Il problema di questo circuito è che il source di  $M_L$  è collegato all'uscita e quest'ultima quando è bassa rimane tale, ma quando è alta vogliamo un uscita alta ma il source rischia di non essere così alto, inoltre se vi è una differenza di tensione tra source e substrato che comporta ad avere un effetto body che **modifica la tensione di soglia** ( $V_{SBL} = V_O$ ), e in questo caso non potremmo trascurarlo quando l'uscita dovrà andare a valore alto.

Utilizziamo gli stessi parametri a carico resistivo:

- $I_D = 80\mu A$
- $V_{DD} = 2.5V$
- $V_L = 0.2V$  tensione che deve assumere l'uscita quando si vuole un output un valore logico basso
- $I_D$ , formula per la zona di saturazione
- $V_{GSL} = V_{DD} - V_O$
- $\gamma = 0.5$
- $2\phi_f = 0.6$

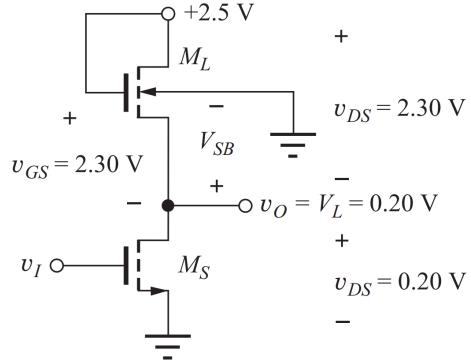
### 4.6.1 Dimensionamento

$$V_{T_{NL}} = V_{T0} + \gamma(\sqrt{V_{SB} + 2\phi_f} - \sqrt{2\phi_f}) = 0.6 + 0.5(\sqrt{0.2 + 0.6} - \sqrt{0.6}) = 0.66V$$

Imponendo la corrente di drain troviamo il rapporto di forma:

$$\frac{W}{L} = \frac{2I_D}{K'_n(V_{GS} - V_{T_{NL}})^2} = \frac{2 \cdot 80\mu A}{100 \frac{\mu A}{V^2} (2.3 - 0.66)^2} = \frac{1}{1.68}$$

Otteniamo dunque che la dimensione del carico è addirittura più piccola,  $1.68 \mu m^2$ , che è simile a quella del driver, ma è più grande della tecnologia utilizzata ( $1 \mu m$ ) perché non vogliamo che il circuito resistivo porti troppa corrente (ne porta circa un quarto rispetto al pull-down), infatti vogliamo che il pull-down vinca quando deve tirare giù, e dunque il pull-up deve essere più debole (sempre per il partitore di tensione).



#### 4.6.2 Calcolo di $V_H$

Ovviamente anche in questo caso vi sono degli svantaggi. Infatti quando l'ingresso è basso,  $V_I = V_L$ , il pull-down si interdice e l'uscita dovrebbe andare a livello logico ad 1.

Il problema è che il condensatore messo a valle mentre si carica, la  $V_{GSL} = V_{DD} - V_S$  (inizialmente a 0.2 V), la tensione sul source aumenta fino al punto da portare la  $V_{GSL}$  alla tensione di soglia, spegnendo di fatto il transistore e portando il condensatore ad un massimo di  $V_O = V_{DD} - V_{GS} = 2.5 - 0.6 = 1.9V$ .

In realtà non è così semplice perché man mano che  $V_O$  sale e più l'effetto body ha una sua rilevanza, dunque l'uscita arriva ad un punto dell'uscita tale per cui la  $V_{GS} = V_{TN}$  ma quest'ultima cambia e dunque l'uscita arriverà a  $V_{DD} - V_{TNL}$  e la tensione di soglia dobbiamo calcolarla tramite l'effetto body:

$$V_O = V_{DD} - V_{TNL} = V_{DD} - \left[ V_{TO} + \gamma(\sqrt{V_O + 2\phi_F} - \sqrt{2\phi_F}) \right]$$

$$V_O = 2.5 - \left[ 0.6 + 0.5(\sqrt{V_O + 0.6} - \sqrt{0.6}) \right]$$

$$V_O = 1.55 \text{ o } 3.27V, \text{ quest'ultima non possibile, la tensione massima è } 2.5 \text{ V}$$

Dunque l'uscita al posto di arrivare a  $2.5V$ , arriverà a solamente  $1.55V$ , questo è un problema nell'utilizzo del carico saturato. Il transistore messo a valle dunque si verrà arrivare una tensione più bassa, dunque dovremmo ridimensionare pure quello.

Inoltre vengono peggiorati pure i margini di rumore.

#### Dimensionamento $M_S$

Dopo lo svolgimento di tutti i conti visti in precedenza con una tensione in ingresso di  $1.55V$  al posto dei  $2.5V$  troviamo che il transistore sotto deve avere un rapporto  $\frac{w}{L} = \frac{4.71}{1}$ .

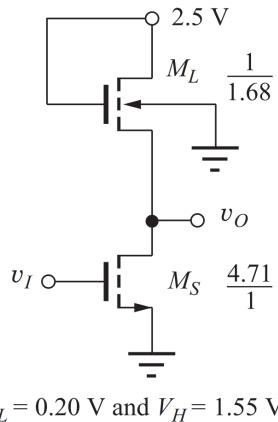
Diminuendo l'ingresso, il pull-down deve sempre tirare  $80\mu A$ , e per riuscirci avendo una tensione più bassa basta aumentare il transistor. Dobbiamo infatti farlo più del doppio più grosso.

#### 4.6.3 Margini di rumore:

$$NM_H = 1.55 - 1.12 = 0.43V, \text{ si è dimezzato}$$

$$NM_L = 0.6 - 0.38 = 0.22V$$

In compenso abbiamo un circuito molto piccolo.



$$V_L = 0.20 \text{ V} \text{ and } V_H = 1.55 \text{ V}$$

**Complicanze:** il problema di questo inverter è che il pull-up si spegne perché ha una tensione di soglia troppo alta e il suo gate dopo un certo punto scende sotto questa soglia. Si potrebbe mettere allora un transistore che abbia una tensione di soglia più piccola, addirittura negativa: **carico depletion mode** o **transistore a svuotamento**.

## 4.7 Carico a depletion mode

In questo caso l'ingresso  $V_{GSL}$  lo collego all'uscita  $V_O$ , per questo il pull-up sarà sempre in conduzione, qualunque sia l'uscita, come la resistenza.

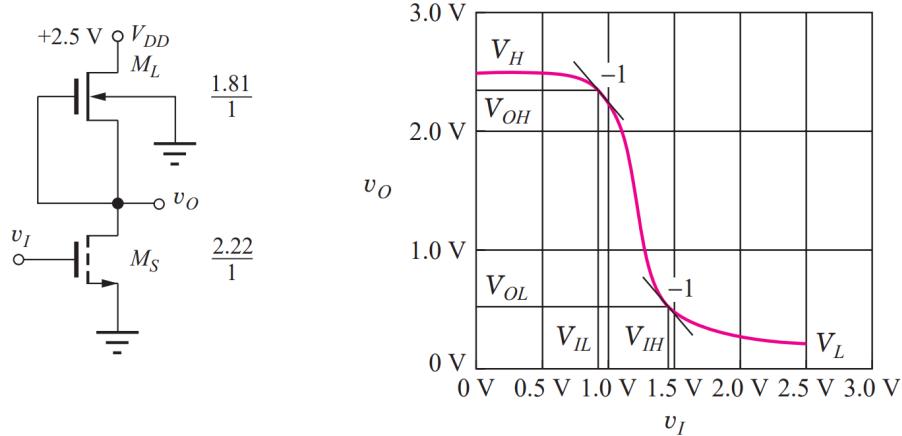


Figura 4.4: Pull-up con transistore nMOS a svuotamento

Vi è sempre un leggero effetto body ma è inferiore a prima.

Essendo a svuotamento, richiede un passo in più per la produzione, quindi potrebbe costare leggermente di più.

### 4.7.1 Dimensionamento

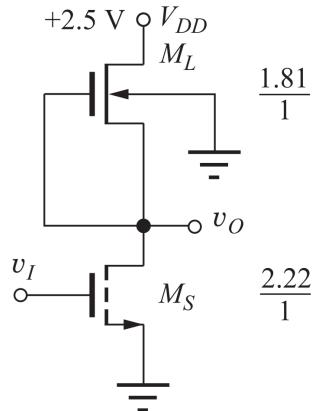
Usando i soliti valori e supponendo una tensione di soglia di  $V_{TO} = -1 \text{ V}$ , anche qui è presente l'effetto body: supponiamo l'uscita bassa vediamo dove finisce la tensione di soglia.

$$V_{TNL} = V_{TO} + \gamma(\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F}) = -1 + 0.5(\sqrt{0.2 + 0.6} - \sqrt{0.6}) = -0.94 \text{ V}$$

Ci troviamo in saturazione:  $V_{DS} = 2.3 > V_{GS} - V_{TN} = 0 - (-0.94) = 0.94$  e dunque possiamo dimensionare il pull-up utilizzando la formula della saturazione

$$\frac{W}{L} = \frac{I_D}{\frac{K_n}{2}(V_{TN}^2)} = \frac{80}{50 \cdot 0.94^2} = \frac{1.81}{1}$$

L'uscita in questo caso ci va a  $V_{DD}$  dunque il pull-down lo dimensioniamo in egual modo se ci fosse una resistenza.



#### 4.7.2 Margini di rumore:

$$NM_H = 2.35 - 1.45 = 0.90 \text{ V}, \text{ si è dimezzato}$$

$$NM_L = 0.93 - 0.50 = 0.43 \text{ V}$$

Questo è un inverter con caratteristiche migliori a quello con carico resistivo, su molti aspetti: dimensione e margini di rumore.

Se collegassimo il gate del pull-up a  $V_{DD}$ ? Tutto funziona ma il pull-up sarà molto forte il che vuol dire che il pull-down deve essere più forte (dimensione più grande per portare più corrente) oppure dobbiamo ridurre la potenza del pull-up facendolo molto lungo per compensare il fatto che porta tutta quella corrente.

## 4.8 Inverter pseudo nMOS

Un'ultima soluzione potrebbe essere quella di realizzare il pull-up tramite un pMOS. In questo caso il source è il terminale a tensione più alta il quale sarà attaccato a  $V_{DD}$ , il suo substrato lo attacchiamo a  $V_{DD}$ .

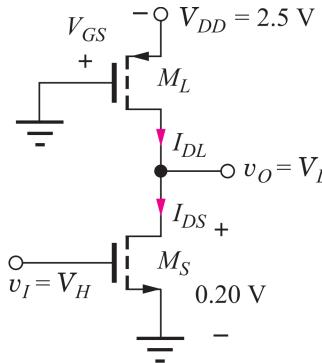


Figura 4.5: Inverter con carico un transistore pMOS

### 4.8.1 Dimensionamento

#### pMOS per uscita bassa:

$V_{GS} = -V_{DD} = -2.5V$ ;  $V_{DS} = 0.2 - 2.5 = -2.3V$ ;  $V_L = 0.2V$ ;  $V_{TP} = -0.6$ . Il transistore è sempre attivo. Quando l'uscita è bassa il pMOS sarà in zona di saturazione perché  $V_{GS}$  non supera  $V_{DS}$  per più di una tensione di soglia  $V_{TP} = -0.6V$ .

Calcoliamo dunque il rapporto di forma del transistore.

$$I_D = \frac{K'_p}{2} \frac{W}{L} (V_{GS} - V_{TP})^2$$

$$80\mu A = \frac{1}{2} 40 \frac{\mu A}{V^2} \frac{W}{L} (-2.5 + 0.6)^2$$

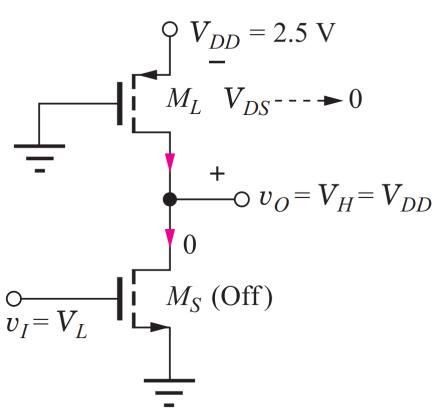
$$\frac{W}{L} = \frac{1.11}{1}$$

Di nuovo viene un transistore molto piccolo, o comunque simile al pull-down.

Questo oggetto qui ci permette di far andare l'uscita fino a  $V_{DD}$ , quando l'uscita è alta il circuito non consuma dato che la  $V_{DS} = V_{DD}$ .

#### Uscita alta:

Rifacendo i conti come prima viene esattamente un rapporto  $\frac{W}{L} = \frac{2.22}{1}$  come con il carico resistivo.



#### 4.8.2 Margini di rumore

Basta eguagliare la corrente dell'nMOS con quella del pMOS, differenziare  $V_O$  rispetto a  $V_I$  e quindi trovare il punto in cui vale  $-1$ , Molto tedioso

$$NM_H = 2.33 - 1.58 = 0.75 V$$

$$NM_L = 0.95 - 0.49 = 0.46 V$$

I margini di rumore sono molto buoni.

Tra il carico a svuotamento e il carico a pMOS le due cose si equivalgono, magari l'altro aveva un margine del rumore  $NM_H$  migliore. Anche per il pMOS vi sono dei passi in più da fare in fase di produzione dato il substrato di tipo n.

## 4.9 Confronto di inverter

### 4.9.1 A carico resistivo

$$I_D = \frac{V_R}{R} = \frac{V_{DD} - V_{DS}}{R} = -\frac{1}{R}V_{DS} + \frac{V_{DD}}{R}$$

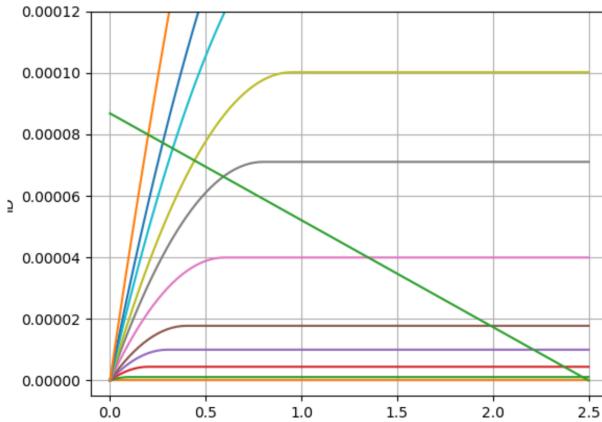


Figura 4.6: Grafico di confronto inverter resistivo

### 4.9.2 A carico saturato

Bisogna fare attenzione in questo caso all'effetto body non trascurabile.

In saturazione:

$$I_D = \frac{K_n}{2}(V_{GS} - V_{TNL})^2$$

$$I_D = \frac{K_n}{2}(V_{DD} - V_{DS} - V_{TNL})^2$$

Questo rappresenta l'arco di parabola in verde sul grafico.

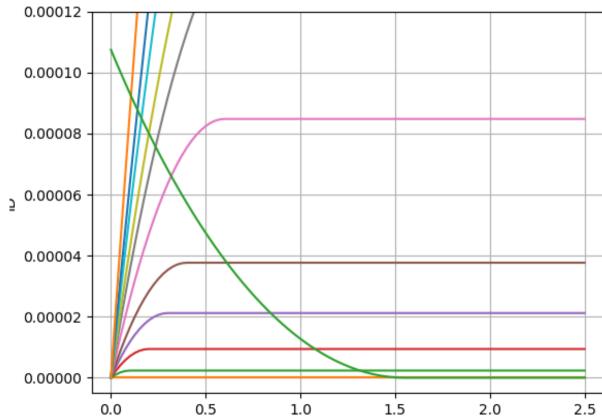


Figura 4.7: Grafico di confronto inverter a carico saturato

### 4.9.3 A carico deplation/svuotamento

Vi sono due zone a seconda delle tensioni:

$$I_D = \frac{K_n}{2}(-V_{TNL})^2 \quad \text{Per } V_{GSL} \leq V_{DSL} + V_{TNL}$$

$$I_D = K_n \left( -V_{TNL} - \frac{V_{DSL}}{2} \right) V_{DSL} \quad \text{Per } V_{GSL} \geq V_{DSL} + V_{TNL}$$

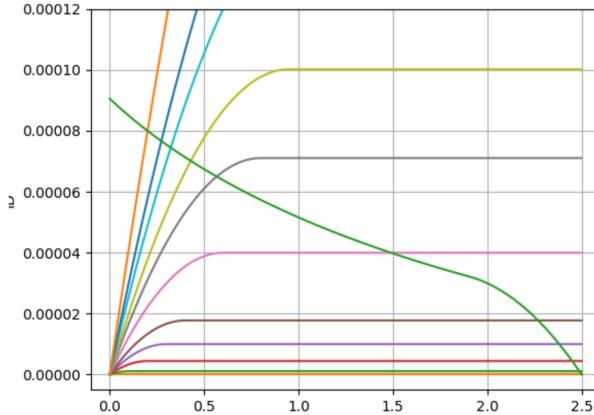


Figura 4.8: Grafico di confronto inverter a svuotamento

### 4.9.4 A carico pseudo nMOS

Per quanto riguarda questo caso come carico ci ritroviamo la caratteristica di un transistore pMOS.

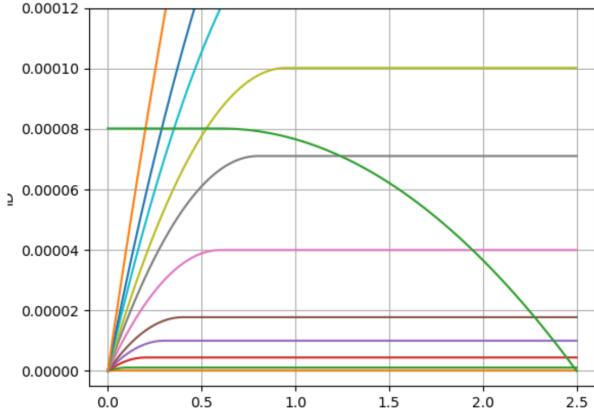


Figura 4.9: Grafico di confronto inverter pseudo nMOS

Questi grafici fanno vedere quanta corrente possono far scorrere i vari carichi. Quindi a parità di tutto il resto il pMOS conduce meglio la corrente soprattutto nel transitorio e potremmo caricare la capacità un po' più velocemente. Al netto anche dei margini di rumore leggermente inferiori al carico a svuotamento.

## 4.10 Impatto della saturazione di velocità

Se i componenti sono più piccoli del micron entra in gioco anche la saturazione di velocità delle cariche, in particolare se il campo elettrico è al di sopra di  $10^5 V$ .

Vediamo alcuni casi:

- Se il campo elettrico è troppo alto, la velocità dei portatori satura, supponiamo  $V_{DS} = V_{SAT} = 1.2 V$  sia la soglia di saturazione di velocità
- Per il pull down non c'è differenza

La saturazione ha effetto sul pull-up infatti la formula per definire la corrente tenendo presente la saturazione di velocità è la seguente:

$$I_D = K_n \left( V_{GS} - V_{TN} - \frac{V_{MIN}}{2} \right) V_{MIN} (1 - \lambda V_{DS})$$

la formula è uguale alla zona triodo tranne che sostituiamo  $V_{DS}$  con  $V_{MIN}$ :

$$V_{MIN} = \min((V_{GS} - V_{TN}), V_{DS}, V_{SAT})$$

Rispettivamente il minimo tra: saturazione, triodo e saturazione di velocità.

## 4.11 Impatto della saturazione di velocità

**Se usiamo un carico saturato:**

$$V_{GS} - V_{TN} = 2.3 - 0.66 = 1.64 V$$

$$V_{DS} = 2.3 V$$

Occorre allora usare  $V_{SAT} = 1.2 V$ , in quanto interviene prima della saturazione normale del transistore

$$\frac{W}{L} = \frac{I_D}{K'_n (V_{GS} - V_{TN} - \frac{V_{SAT}}{2}) V_{SAT}} = \frac{1}{1.56}$$

Confrontato con  $1/68$ , ottenuto precedentemente, il canale deve essere leggermente accorciato per compensare la perdita di corrente dovuta alla saturazione di velocità.

**Se usiamo un carico svuotamento:**

$$V_{GS} - V_{TN} = 0 + 0.94 = 0.94 V$$

$$V_{DS} = 2.3 V$$

$V_{GS} - V_{TN} < V_{SAT}$  non dobbiamo cambiare nulla.

**Se usiamo uno pseudo nMOS:**

$$V_{GS} - V_{TN} = 2.5 - 0.6 = 1.9 V$$

$$V_{DS} = 2.3 V$$

Occorre usare  $V_{SAT} = 1.2 V$  ottenendo  $\frac{W}{L} = \frac{1.28}{1}$ . Confrontare con  $1.11/1$  ottenuto precedentemente, la larghezza W deve essere leggermente aumentata per compensare la perdita di corrente dovuta alla saturazione di velocità.

Con tecnologie moderne siamo sempre in saturazione di velocità, infatti inizialmente si parte sulla carta e poi si passa su Spice per confrontare i vari modelli con tecnologie di dimensioni diverse.

# Capitolo 5

## Porte logiche

Fino ad ora ci siamo focalizzati sull'invertitore con logia nMOS o anche chiamata logica a rapporto perché il livello delle uscite dipende dal dimensionamento del **pull-up** e del **pull-down** e a seconda li dimensioniamo abbiamo dei livelli di uscita differenti. In particolare è il livello basso che viene particolarmente influenzato, quello alto arriva a  $V_{DD}$  a meno che non si usi un carico saturato che vada in cut-off e dunque non porta più l'uscita a  $V_{DD}$ .

Con questa logica il pull-up viene sempre acceso, e per portare all'uscita un valore basso si deve accendere il transistore sottostante per poter portare giù l'uscita, dopo un corretto dimensionamento.

Nulla vieta di fare il contrario: mettere un pull-down sempre attivo e un pull-up, pMOS, che viene controllato dall'ingresso.

Quando l'ingresso è alto non conduce, e il pull-down porta giù l'uscita.

Quando l'ingresso è basso, invece, è il pull-up che tira su l'uscita.

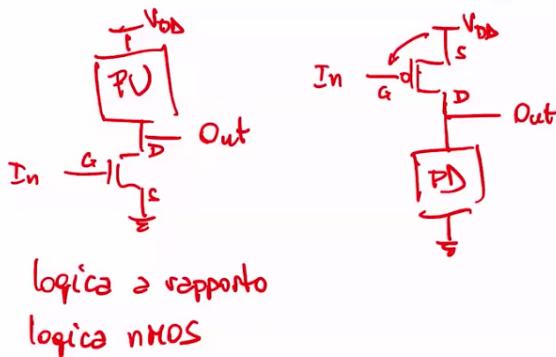


Figura 5.1: Logia a rapporto

**Come mai non si fa mai una logia a rapporto con un pMOS come pull-up?** Ricordiamo che i pMOS e gli nMOS, data la stessa tecnologia, i transistori nMOS conducono meglio dei pMOS, questo perché la mobilità degli elettroni è più alta di quella delle lacune.

A parità di corrente che si vuole far passare l'nMOS è più piccolo, più corrente vogliamo più W deve essere grande.

Ma anche guardando la transconduttanza  $K'_N = \mu_n C''_{ox} \frac{W}{L}$ , vediamo che troviamo un coefficiente moltiplicativo  $\mu_n$ , che riguarda la mobilità dei portatori, se più alta è la mobilità possiamo fare W più piccola, questo implica che si possono fare transistori più piccoli.

Dunque normalmente, per questa ragione, si cerca di utilizzare un nMOS.

Per fare altre porte logiche possiamo constatare che il pull-up rimarrà sempre uguale, quello che varierà sarà il pull-down a seconda delle porte logiche da costruire.

## 5.1 Costruzione di porte logiche

Per costruire le porte logiche fondamentali possiamo usare delle combinazione di transistori. In particolar modo si useranno transistori collegati in serie o in parallelo.

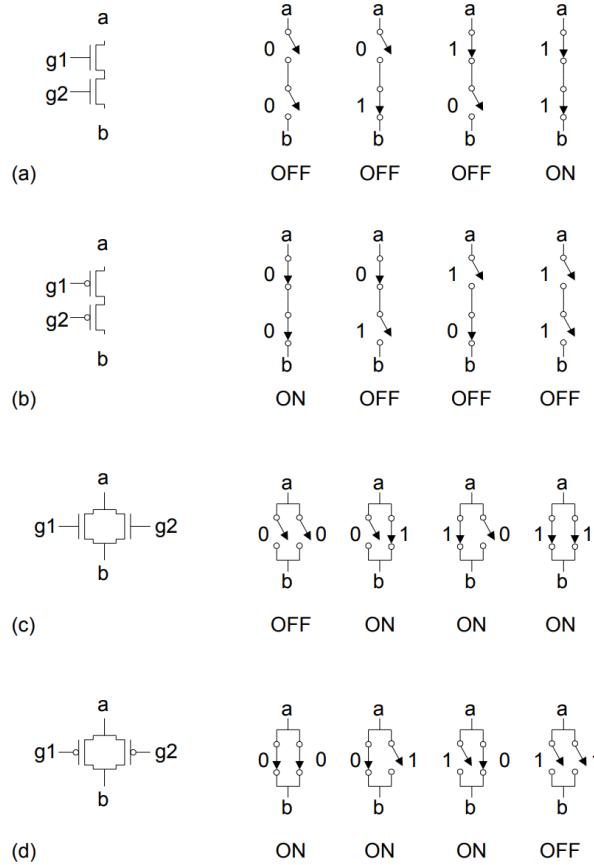


Figura 5.2: Vari modi per costruire le porte AND, NAND, OR e NOR

Nel caso delle Figura 5.2 nei casi a) nMOS e b) pMOS, i transistori in serie conducono se e solo se conducono entrambi.

Nel caso delle Figura 5.2 nei casi c) nMOS e d) pMOS, i transistori in parallelo conducono quando almeno un transistore è in conduzione.

## 5.2 Porta logica: NOR

Vogliamo che l'uscita sia alta se e solo se entrambi gli ingressi sono bassi, altrimenti, vogliamo che l'uscita sia a zero quando almeno uno dei due ingressi è al livello alto. In questa maniera dovremo realizzare una costruzione parallela: il pull-down si deve attivare quando almeno un ingresso sia al livello alto.

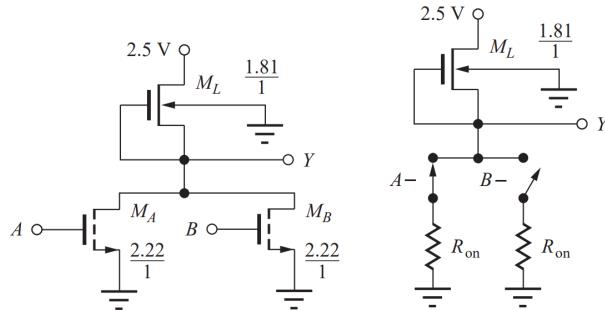


Figura 5.3: Porta logica NOR

**NOTA:** la resistenza interna dei due transistori  $R_{ON}$  deve essere inferiore alla resistenza del pull-up, per vincere il partitore e portare l'uscita a livello basso, definita a  $0.2\text{ V}$  o inferiore .

Affinché l'uscita sia al livello logico basso, dobbiamo dimensionare i due transistori di pull-down in modo tale che da soli siano in grado di tirare giù l'uscita. Il dimensionamento dunque è uguale a quello fatto per l'inverter:  $\frac{W}{L} = \frac{2.22}{1}$ .

**A quanto arriva l'uscita nel caso in cui  $A = B = 2.5\text{ V}$ ?**

Entrambi i transistori conducono e si trovano in zona triodo, mentre il pull-up si trova in zona di saturazione ( $V_{GS} = 0\text{ V}$  e  $V_{DS}$  è alta sicuramente  $V_{DS} \leq (2.5 - 0.2)\text{ V}$ ).

Quindi per trovare quanto vale l'uscita ci basta uguagliare le due equazioni:

$$2K'_n 2.22 \left( V_{GS} - V_{TN} - \frac{V_{DS}}{2} \right) V_{DS} = \frac{K'_n}{2} 1.81 (V_{GSL} - V_{TNL})^2$$

dove:

- $V_{TN}$ , supponiamo che non ci sia effetto body e la uguagliamo a  $0.6\text{ V}$ .
- $V_{GSL}$  questa risulta essere zero, per costuzione del circuito
- $V_{TNL}$  dovremmo considerare l'effetto body,  $V_{TNL} = V_{TO} + \gamma \sqrt{V_{DS} + 2\phi_F} - \sqrt{2\phi_F}$
- L'equazione a sinistra viene moltiplicata per un fattore 2 perché la corrente, essendo i sue transistori uguali, si splitterà in due.

**Ipotesi:** mettiamo una  $V_{TNL}^2 = 0.94$ , cerchiamo di indovinarla. Sicuramente ci sarà un errore ma non importa in quanto di per sé, già il modello che usiamo contiene degli errori ed ulteriori semplificazioni possono aiutare al posto di risolvere l'equazione con due radici non comode da gestire.

$$8.88 \left( 1.9 - \frac{V_{DS}}{2} \right) V_{DS} = 1.81 \cdot 0.94$$

$$4.44V_{DS}^2 - 16.872V_{DS} + 1.7 = 0$$

$$V_{DS} = 0.103\text{ V}$$

Quest'ultima era  $0.2\text{ V}$ , con due pull-down scende di quasi la metà.

### 5.3 Porta logica: NAND

Essendo la NAND non attiva se e solo se entrambi gli ingressi sono alti, il pull-down deve portare l'uscita a zero solo quando entrambi gli ingressi sono alti e per questo utilizzeremo due transistori in serie. Se uno dei due ingressi è interdetto, la corrente non va a massa e il pull-up porta l'uscita al livello alto.

Questi due resistori devono portare una corrente di  $80 \mu A$ , come definito dalle specifiche, però sono in serie e la tensione, essendo i due transistori uguali, si ripartisce circa a metà. Per far in modo che portino la stessa corrente di prima dovremo fare dei transistori più grossi (in prima approssimazione dovranno essere grossi il doppio:  $W/L = 4.44$ ).

**Come mai diventa più grosso?** Come per i condensatori in parallelo, o i resistori in serie, le lunghezze dei due transistori si sommano, formando quindi una lunghezza doppia, e per ottenere sempre lo stesso rapporto, che ci garantisce che possa passare tanta corrente quanta le specifiche dicono, dobbiamo aumentare W. Aumentiamo del doppio L dobbiamo aumentare del doppio anche W per mantenere il medesimo rapporto  $W/L$ .

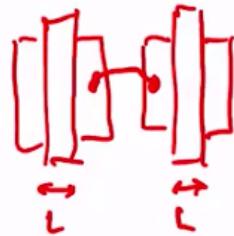


Figura 5.4: Transistori in serie

**Se fossero in parallelo:** i transistori in questo caso è come se fosse W doppia, quindi L la facciamo molto piccola, il quale implica che due transistori in parallelo conducano molto bene.



Figura 5.5: Transistori in parallelo

Ritornando a noi, dimensionandoli in questa maniera siamo sicuri che riescano a trasportare  $80 \mu A$ .  
Tutto questo in prima approssimazione perché:

Non siamo sicuri che la tensione si ripartisca esattamente a metà, infatti il source di B non è collegato a massa, bensì è collegato al drain di A, questo provoca il fenomeno dell'effetto body per il transistore B. Questo effetto gli fa alzare la tensione di soglia.

Questo comporta che:

- $V_{GSMA} = V_{DD}$  perché il source è collegato a massa e l'effetto body non si presenta;
- $V_{GSMB} \neq V_{DD}$  perché il source è collegato al drain di  $M_A$  e per questo si presenta l'effetto body il quale innalza la tensione di soglia.

In particolare  $V_{GSMB}$  è in una situazione meno favorevole di  $M_A$  in quanto conduce di meno perché ha una  $V_{GSMA}$  più bassa e ha una tensione di soglia più alta, questo comporta che  $M_B$  dovrà essere più grosso di  $M_A$  per compensare la differenza di conduzione che vi è tra i due.

## 5.4 Dimensionamento dei transistori in fase di conduzione

Supponiamo che  $V_0$  si divida esattamente in due, dunque su ogni transistore, se l'uscita va a  $0.2\text{ V}$ , cadrà esattamente  $0.1\text{ V}$ :  $V_{DS} = 0.1\text{ V}$ , si entrambi. Questo per semplificare, sempre dato dal fatto che già il modello non è preciso, qualche semplificazione possiamo farcela per fare i conti in maniera più agile.

### 5.4.1 Per $M_A$

Le cose sono sempre come prima, è cambiata solo la  $V_{DS}$  si è dimezzata:

$$K'_n \left( \frac{W}{L} \right)_A \left( V_{GSA} - V_{TNA} - \frac{V_{DSA}}{2} \right) V_{DSA} = I_D = 80 \mu A$$

dove

- $V_{GSA} = 2.5\text{ V}$ ;
- $V_{TNA} = 0.6\text{ V}$ ;
- $V_{DSA} = 0.1\text{ V}$

$$\left( \frac{W}{L} \right)_A = \frac{80}{100(2.5 - 0.6 - \frac{0.1}{2})0.1} = 4.32$$

Vediamo che non viene esattamente 4.44 bensì un po' meno, questo è anche dovuto al fatto che abbiamo fatto i conti in maniera un pelo più oculata.

### 5.4.2 Per $M_B$

La tensione di soglia è alterata, quindi dovremmo usare la formula che tiene conto dell'effetto body:

$$V_{TNB} = V_{TN} + \gamma(\sqrt{V_{DSA} + 2\phi_F} - \sqrt{2\phi_F}) = 0.63\text{ V}$$

Per determinare il valore di  $V_{DSA}$  entra in gioco la semplificazione detta pocanzi:  $0.1\text{ V}$ . Anche questo transistore si trova in zona triodo, pertanto:

$$K'_n \left( \frac{W}{L} \right)_B \left( V_{GSB} - V_{TNB} - \frac{V_{DSB}}{2} \right) V_{DSB} = I_D = 80 \mu A$$

dove

- $V_{GSB} = 2.4\text{ V}$ ;
- $V_{TNB} = 0.63\text{ V}$ ;
- $V_{DSB} = 0.1\text{ V}$

$$\left( \frac{W}{L} \right)_A = \frac{80}{100(2.4 - 0.6 - \frac{0.1}{2})0.1} = 4.65$$

Dunque il transistore più in alto risulterà essere più grosso.

## 5.5 AND

Il pull-down deve essere attivo quando almeno uno degli ingressi si trova a zero, quindi risulterà essere una NAND con una negazione in cascata.

Infatti la AND, con questa logica, non è possibile farla sul pull-down perché è fatto con nMOS e se gli ingressi sono a zero, si mettono in interdizione e il circuito ritorna il valore alto:  $V_{DD}$ .

Dunque una logia fatta nel modo in cui l'abbiamo vista è **sempre invertente**, ecco anche perché è possibile utilizzare un inverter dopo la porta logica NAND.

La AND costruita in questo modo sarà costituita da 5 transistori e sarà anche più lenta dovuto al fatto che vi è anche un inverter. Dunque conviene sempre usare porte invertenti piuttosto che usare le porte non invertenti.

Infatti un circuito a due livelli, studiato nel corso Reti Logiche, costituito da un livello di AND seguito in seguito da porte OR, non verrà mai fatto con porte non invertenti, bensì con porte invertenti perché sono più veloci, portando al medesimo risultato.

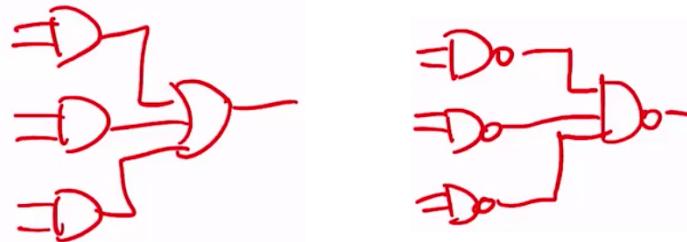


Figura 5.6: Circuito a due livelli

Nelle NOR i circuiti vanno ancora più veloce perché i transistori sono inseriti in parallelo, oppure più piccole comunque.

Infatti una volta, quando si utilizzava questa logica a rapporto, si facevano tutti i circuiti con le NOR, per intenderci si sintetizzano gli zeri nelle mappe di Karnaugh.

## 5.6 Porte logiche complesse

Non c'è bisogno di fare circuiti solo con NAND e NOR, ma basta mettere nel pull-down la combinazione adatta per ottenere un uscita desiderata, possiamo dunque costruire funzioni complesse direttamente nel pull-down.

**Esempio:**  $Y = (A+BC+BD)' = (A+B(C+D))'$

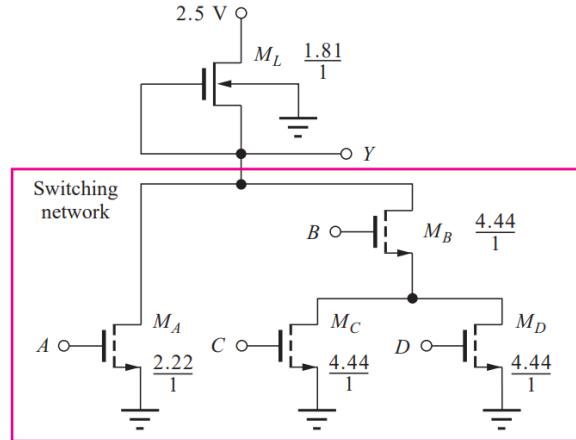


Figura 5.7: Esempio di funzione complessa

**NOTA:** nella parte di destra si sarebbe potuto benissimo mettere C e D sopra, al posto di B, e mettere quest'ultimo al posto dei transistori C e D. Il problema sorge dal fatto che vi è l'effetto body, dove nel caso della figura si ha solo su B, nel caso in cui volessimo invertire l'effetto si presenterebbe su due transistori, C e D, al posto che solamente su B.

Il dimensionamento è sempre fatto nei casi peggiori:

- se solo A attivo, deve potere portare l'uscita bassa (NOT);
- se solo B e C attivi, assieme, devono portare l'uscita bassa (NAND);
- se solo B e D attivi, assieme, devono portare l'uscita bassa (NAND).

## 5.7 Tempistiche

Andiamo a vedere ora le tempistiche di questi circuiti, in particolare che metriche ci interessano dal punto di vista delle tempistiche? Dal momento in cui cambia l'ingresso, quanto impiega l'uscita a cambiare?

Questo è essenziale per determinare la velocità di un circuito scandito dal clock: nel tempo di salita/discesa del clock la logica deve aver calcolato il risultato, ovvero deve aver raggiunto un uscita stabile (vedi tempo di Set-Up e tempo di Hold).

### 5.7.1 Comportamento dinamico: i ritardi

#### Ritardo di propagazione

Tempo che intercorre tra la transizione dell'ingresso e quella dell'uscita. Di norma i ritardi sono diversi a seconda della transizione.

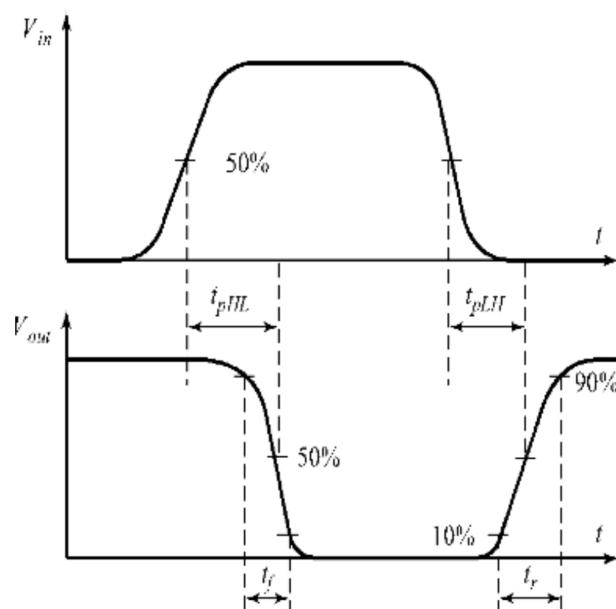
Per determinarlo si osserva quando la transizione passa per il 50% della tensione totale. Questo tempo potrebbe anche venire influenzato dal secondo tipo di ritardo: il tempo di rise e fall.

Questa metrica ci dice quanto **veloce** va il circuito.

#### Ritardo di salita e discesa (rise e fall)

Tempo che impiega l'uscita a passare dal 10% al 90% (o viceversa).

Non prediamo in considerazione la derivata in quanto essendo molto bassa all'inizio comporterebbe gradi errori di tempistiche.



Questo ritardo è noto come **slew rate**, questa metrica ci dice quanto velocemente cambiano entrambi i fronti.

### 5.7.2 Capacità nel transistor MOS

I ritardi, l'uscita che non segue immediatamente l'ingresso, sono insiti nel circuito del transistore, in quanto il suo gate si comporta proprio un **condensatore** a piatti piani paralleli con differenti capacità:

- tra gate e canale;
- tra gate e source/drain (overlap);
- capacità dei diodi (non lineare);
- capacità non lineari, che dipendono anche dalla zona di funzionamento.

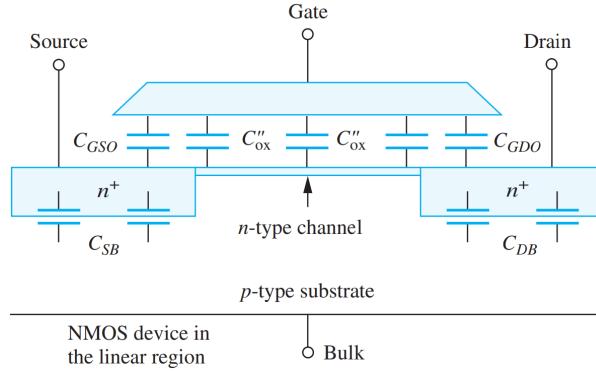


Figura 5.8: Capacità lineari e non in un transistore

La tensione sul condensatore non può cambiare istantaneamente, perché servirebbe una potenza infinita e dunque un picco di corrente infinita, tutto questo non è possibile e dunque la tensione cambia in maniera graduale.

### 5.7.3 Transitorio dell'invertitore a carico resistivo

Calcoliamo, nel circuito più semplice visto fino ora, quanto tempo impiega il gate a valle a caricarsi. Poniamo inizialmente  $V_I = V_{DD} = 2.5V$  e l'uscita  $V_L = 0.2V$ , successivamente, a  $t = 0s$  vi è una transizione sull'ingresso infatti ora  $V_I = V_L = 0.2V$ . Figura 5.9 a).

Per  $t < 0$  l'uscita valeva  $0.2V$  quindi il condensatore a valle aveva una differenza di potenziale ai suoi capi di  $0.2V$ .

Per  $t \leq 0$  il gate del transistore va a  $0.2V$  e l'uscita viene portata a  $2.5V$  dal pull-up e il condensatore inizia a caricarsi. Figura 5.9 b).

Si caricherà secondo lo stesso principio di caricamento di un circuito RC: con la carica esponenziale.

$$V_O = V_2 - (V_2 - V_1)e^{-\frac{t}{RC}}$$

dove:

- $V_1 = 0.2V$
- $V_2 = 2.5V$
- R la sappiamo dato che il circuito lo abbiamo progettato
- C lo sappiamo dato che è la somma di tutte le capacità dei gate a valle

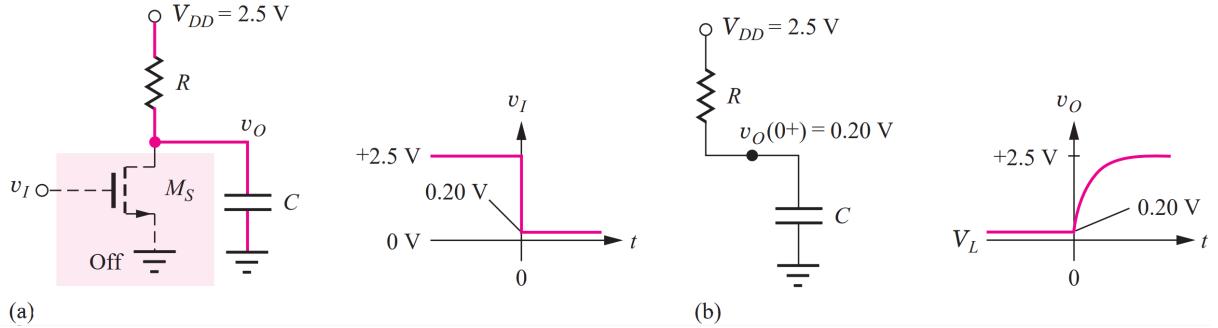


Figura 5.9: transizione in salita

#### 5.7.4 Tempo di salita

Dobbiamo vedere quando l'uscita passa tra il 10% e il 90% del range di valori (non è  $2.5\text{ V}$ , bensì  $(2.5 - 0.2)\text{ V}$ ):

- 10% :  $V_1 + 0.1(V_2 - V_1) = 0.2 + 0.1(2.5 - 0.2) = 0.43\text{ V}$
- 90% :  $V_1 + 0.9(V_2 - V_1) = 0.2 + 0.9(2.5 - 0.2) = 2.27\text{ V}$

$$0.43 = 2.5 - 2.3e^{-\frac{t}{RC}} \quad t_1 = -RC \ln(0.9)$$

$$2.27 = 2.5 - 2.3e^{-\frac{t}{RC}} \quad t_1 = -RC \ln(0.1)$$

$$t_r = t_2 - t_1 = 2.22RC$$

Quindi in questo caso il tempo di salita è un multiplo della costante di tempo  $\tau$ , e la costante è appunto 2.22.

#### 5.7.5 Tempo di propagazione

Questo caso è più semplice, mi basta trovare quanto tempo impiega per passare a metà, 50%, della sua escursione massima.

$$50\% : V_1 + 0.5(V_2 - V_1) = 0.2 + 0.5(2.5 - 0.2) = 1.35\text{ V}$$

$$1.35 = 2.5 - 2.3e^{-\frac{t}{RC}} \quad t_1 = -RC \ln(0.5)$$

Quando è che passa per il valore cercato? A tempo  $t = 0\text{ s}$  perché abbiamo presupposto che cambi istantaneamente: all'istante 0 passerà per tutti i valori.

$$t_{PLH} = t_1 = 0.69RC$$

Dunque l'inverter andrà tanto più in fretta, tanto più  $C$  è piccola (transistori a valle più piccoli e pochi così che la capacità non cresca troppo). Il tempo di calcolo è fortemente dipendente anche dalla topologia del circuito e non solamente dalla porta logica in se.

Un altro parametro che possiamo cambiare è  $R$ , tanto più piccola è  $R$  tanta più corrente può passare (al limite è un filo) e tanto più in fretta carica il/i condensatore/i a valle.

La  $R$  però la abbiamo ricavata dalle specifiche di consumo totale diviso il numero di porte (con stima della metà) accese e da lì abbiamo ricavato quanta doveva essere la corrente che fluiva su ogni resistenza e attraverso la legge di Ohm ci siamo ricavati il valore della resistenza. Quindi va da se che se diminuiamo  $R$  passa più corrente e questa porterà ad un consumo maggiore a fronte di un miglioramento di velocità.

## Esempio

1. Sia  $R = 28.8 K\Omega$  e  $C = 0.2 pF$  i tempi saranno rispettivamente:

$$t_r = 2.2RC = 12.7 ns$$

$$t_{PLH} = 0.69RC = 3.97 ns$$

2. Sia  $R = 28.8 K\Omega$  e  $C = 0.5 pF$  i tempi saranno rispettivamente:

$$t_r = 2.2RC = 31.7 ns$$

$$t_{PLH} = 0.69RC = 9.94 ns$$

Facendo un rapido calcolo notiamo che sono tempistiche davvero enormi se confrontati con i tempi di clock odierni: supponiamo un clock di  $4Ghz$ , il completamento di un'operazione deve essere effettuato entro

$$\frac{1}{4 \cdot 10^9} = 2.5 \cdot 10^{-10} s = 0.25 \cdot 10^{-9} s$$

stiamo parlando di più di un ordine di grandezza di differenza.

### 5.7.6 Transitorio in discesa

In questo caso è più complicato perché ad agire è il transistore e non il resistore ma questo rimane ugualmente.

Inizialmente l'ingresso è a  $0.2 V$  e l'uscita si trova a  $V_{DD}$ , al tempo  $t > 0$  l'ingresso si trova a  $V_{DD}$  e l'uscita deve scendere fino a  $0.2 V$ . Al tempo  $t > 0$  il transistore si trova in zona di saturazione perché la  $V_{DS}$  è alta il quale inizia a scaricare la capacità, dunque il valore del drain inizia ad abbassarsi perché la capacità viene scaricata dal transistore.

In questa fase la scarica è costante, quasi lineare; non passa nessuna corrente per R in quanto vi è la medesima differenza di potenziale.

La tensione ai capi di C scende fintantoché scende ad una tensione di soglia sotto il gate,  $V_O < G_{GS} - V_T$ , e dunque entra in zona triodo.

In questa fase la scarica non è più costante e inizia a scendere anche quella del transistor e il condensatore si scarica sempre meno in fretta, in più vi è anche la resistenza R da tenere in considerazione. Infatti man mano che scende la tensione sul drain vi è corrente che passa per la resistenza la quale aumenta più il condensatore viene portato a massa.

Dunque la maggior parte della corrente passa attraverso la resistenza e non dal condensatore il quale fatica a scaricare le cariche depositate sulla capacità, infatti quando arriviamo a  $0.2V$  (specificata di progetto) ai capi di C, tutta la corrente che passa per il transistore è passata prima da R. Infatti più in basso di  $0.2V$  C non arriva.

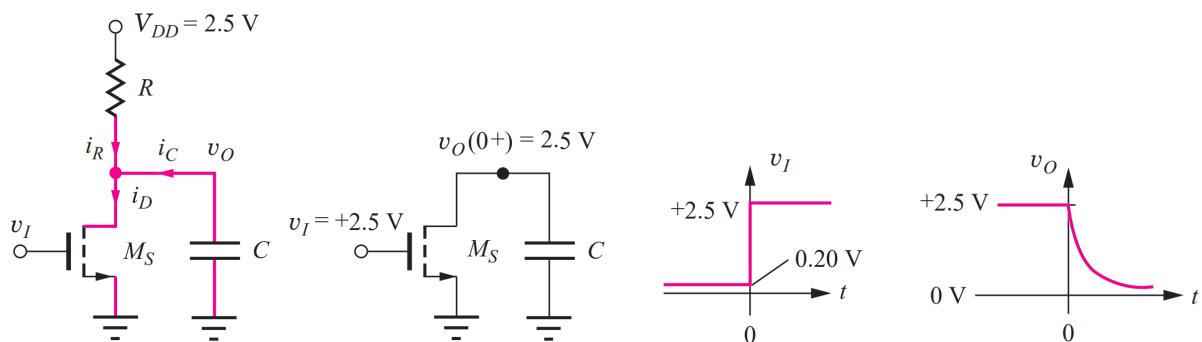


Figura 5.10: Transizione in discesa

### 5.7.7 Transitorio complesso

Il transitorio passa per diverse regioni di funzionamento:

- Comincia in zona di saturazione, perché la tensione di uscita inizialmente a 2.5V non può cambiare istantaneamente;
- Si segue quindi la caratteristica fino ad arrivare sulla retta di carico

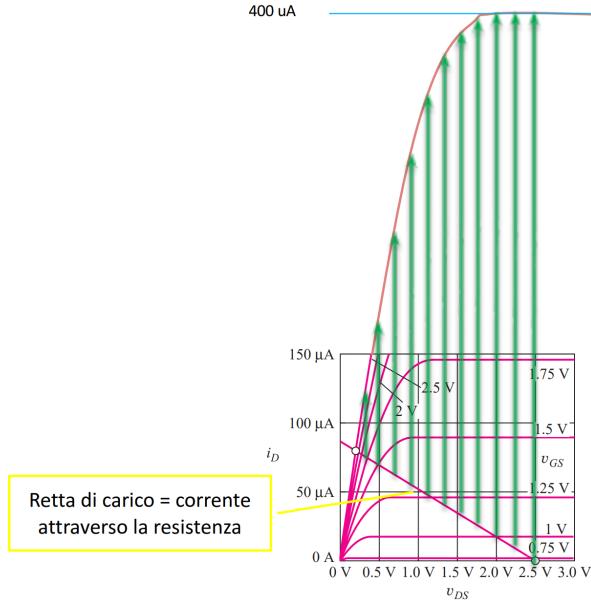


Figura 5.11: Transitorio complesso

La freccia in verde raffigura la corrente che serve per scaricare la capacità C, si vede che scende e anche che non è particolarmente lineare perché aumenta la corrente sulla resistenza.

#### Transitorio in discesa

I conti sono decisamente più complessi, infatti una prima semplificazione possiamo considerare che la corrente che passa attraverso la resistenza è piccola, ce la dimentichiamo, tanto il modello non è preciso tanto vale semplificare al massimo.

Per approssimare il valore della resistenza senza fare innumerevoli conti possiamo determinare la resistenza da utilizzare partendo da  $R_{on}$ , resistenza che vediamo quando  $V_{DS}$  tende a zero in Figura 5.11, corretta per un fattore correttivo.

$$I_D = K - n(V_{GS} - V_T)V_{DS} \quad \text{Se } V_{DS} \text{ è bassa}$$

$$R_{on} = \frac{1}{K_n(V_{GS} - V_T)}$$

Per determinare R si utilizza un fattore correttivo del valore di 1.7:

$$R = 1.7R_{on}$$

$$t_f = 2.2(1.7R_{on})C = 3.7R_{on}C$$

$$t_{PHL} = 0.69(1.7R_{on})C = 1.2R_{on}C$$

Prendiamo queste due ultime espressioni per calcolare il tempo di salita e discesa HL.

## Esempio

- Sia  $R = 28.8 \text{ K}\Omega$  e  $C = 0.5 \text{ pF}$  i tempi saranno rispettivamente:

$$\bullet \quad K_m = 2.22(100 \mu\text{A}/\text{V}^2) = 222 \mu\text{A}/\text{V}^2$$

$$t_r = 2.2RC = 31.7 \text{ ns}, \text{ calcolo precedente}$$

$$t_{PLH} = 0.69RC = 9.94 \text{ ns}, \text{ calcolo precedente}$$

Calcoliamo ora il tempo di discesa:

$$R_{on} = \frac{1}{(222(2.5 - 0.6))} = 2.37 \text{ K}\Omega$$

Da notare che è molto più piccola di R, perché il pull-down deve essere molto piccola, altrimenti il partitore non tira giù l'uscita

$$t_r = 3.7R_{on}C = 4.39 \text{ ns}$$

$$t_{PHL} = 1.2R_{on}C = 1.42 \text{ ns}$$

La media:

$$t_p = (t_{PHL} + t_{PLH})/2 = 5.68 \text{ ns}$$

**Osservazione:** il comportamento del circuito è molto asimmetrico ed è dovuto al fatto che il pull down deve avere una resistenza in conduzione molto minore del carico per raggiungere un valore di  $V_L$  sufficientemente basso.

La simulazione con Spice conferma i risultati ottenuti:

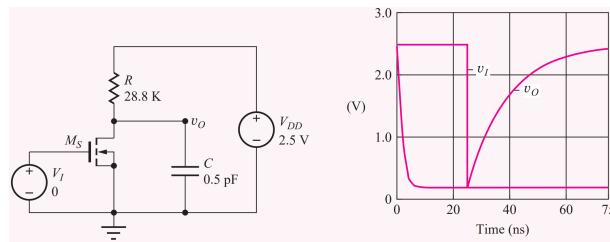


Figura 5.12: Simulazione Spice

Il gradino raffigura l'ingresso, le curve rappresentano l'uscita.

### 5.7.8 Transitorio, invertitore pseudo nMOS

Situazione simile alla precedente, approssimiamo anche la resistenza del pMOS di carico

$$R_p = \frac{1}{K_p(V_{DD} - V_{TP})} \quad K - p = 1.11 \cdot 40 \frac{\mu A}{V^2}$$

$$R_p = \frac{1}{44.4(2.5 - 0.6)} = 11.9 K\Omega$$

$$t_r = 3.7 \cdot 11.9 \cdot 0.5 = 22 ns$$

$$t_{PLH} = 1.2 \cdot 11.9 \cdot 0.5 = 7.14 ns$$

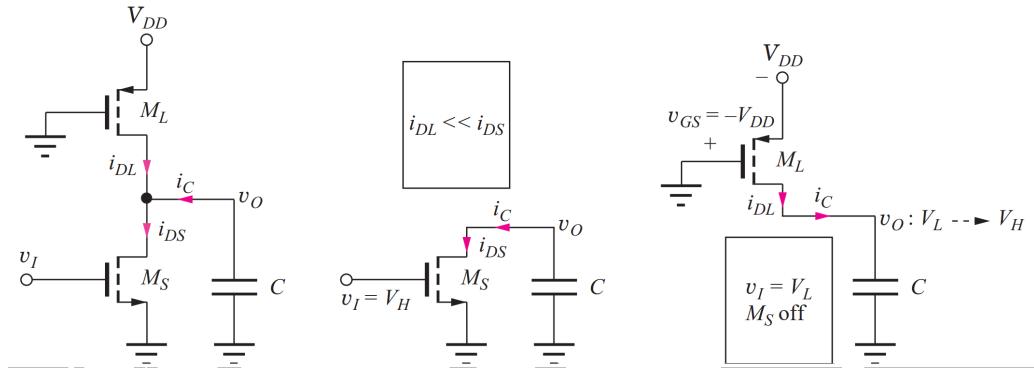


Figura 5.13: Transistore nMOS

### 5.7.9 Confronto di inverter

I diversi inverter nMOS hanno lo stesso comportamento in discesa, c'è sempre un nMOS che scarica il condensatore.

In salita, dipende dal pull up, il pMOS è quello con la maggiore capacità di corrente, ed ha quindi il ritardo minore.

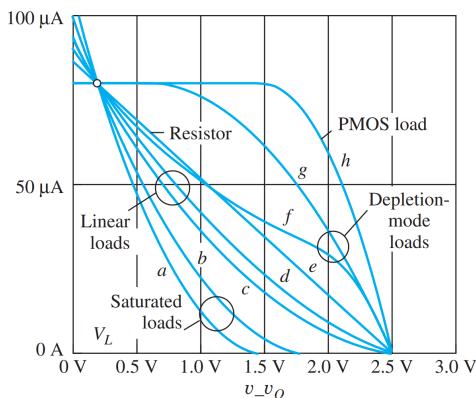


TABLE 6.9  
NMOS Inverter Delays\* (ns)

	$\tau_{PHL}$	$\tau_{PLH}$	$t_f$	$t_r$
Resistor	0.35	2.0	0.72	6.3
Pseudo NMOS	0.31	1.3	0.84	3.0
Depletion load	0.33	2.0	0.73	4.6
Linear load	0.32	1.8	0.83	9.6
Saturated load	0.33	1.6	0.65	14.2

Figura 5.14: Confronto dei vari inverter

In tutti i casi trattati abbiamo considerato il cambiamento del ingresso istantaneo, ma quest'ultimo non avverrà istantaneamente e anche questo causa un ritardo.

## 5.8 Consumo statico

La porta logica consuma ogni qualvolta che passa corrente. Quando l'uscita è alta il pull-down è interdetto e il pull-up è attivo ma non consuma corrente (il gate a valle è carico, le cariche non devono passare), e infatti ai capi del resistore la differenza di potenziale risulta 0.

Consuma quando:

- Deve passare dal valore zero al valore uno, passa corrente perché si deve caricare il condensatore
- Deve passare dal valore uno al valore zero, passa corrente perché deve passare corrente per scaricare il condensatore, in realtà stiamo dissipando la corrente caricata in precedenza.

Nel caso della logica a rapporto però, la porta consuma anche quando l'uscita è bassa e sta ferma, non consuma quando sta alta e ferma, perché il pull-down conduce, il pull-up pure e vi è una corrente che passa da  $V_{DD}$  fino a massa.

Questo consumo è definito **consumo statico** della porta logica.

**Di quanta potenza si parla?**

$$P = V_{DD}I_{DD} = 2.5 \cdot 80 = 200 \mu W = 0.2 mW \text{ per non fare nulla!}$$

Il consumo medio si suppone essere la metà,  $0.1 mW$ , dato la casualità dei due stati.

**Esempio:** con 10 000 porte logiche si ha un consumo statico di  $1 W$ , oltre al consumo dinamico per far cambiare lo stato alla porta.

Si potrebbe diminuire la corrente statica aumentando la resistenza, ma andremo a peggiorare le prestazioni: aumenterà la costante di tempo per caricare il gate della porta logica.

Questo tipo di logica rimane un grande problema per sistemi complessi, non possiamo fare grossi circuiti perché costruiremo delle grandi stufe elettriche.

Queste porte non sono più usate, le abbiamo studiate per il concetto di base le quali ci introducono il **metodo di analisi** del pull-down, infatti ora analizzeremo il pull-up.

# Capitolo 6

## Logica CMOS

La logica tuttora utilizzata è la cosiddetta logica di tipo CMOS: complementary metal-oxide semiconductor.

Il problema della logica a rapporto è che quando l'uscita è a zero, il pull-down conduce, ma anche il pull-up conduce e c'è uno **spreco** di corrente.

Un modo per risolvere la situazione è fare in modo che quando l'uscita deve andare a zero, il pull-down conduca ma il pull-up no, si deve staccare. Questa è la logica CMOS.

Infatti essa utilizza in maniera simmetrica i transistori nMOS e pMOS, simile all'inverter pseudo pMOS.

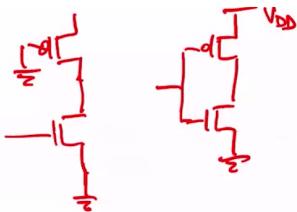


Figura 6.1: Pseudo nMOS a sinistra, logica CMOS a destra

Nella logica CMOS anche il pMOS funziona da interruttore invece che da carico.

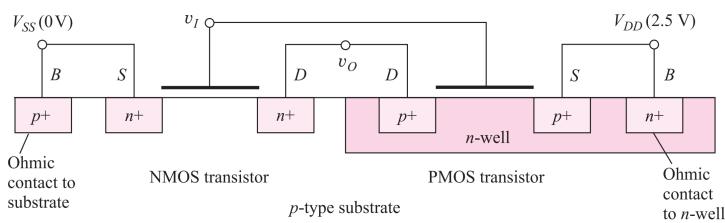


Figura 6.2: Logica CMOS

Si vede che con questa logica quando l'ingresso è alto, il pMOS si stacca e l'nMOS non ha nessuno con cui competere; quando l'ingresso è basso, invece, è l'nMOS che si stacca. Logica appunto **complementare**.

**Perché allora la logica CMOS non viene usata anche per costruire l'inverter?** Perché si complicherebbe pure il pull-up, creando dunque un circuito più grande. Soprattutto se si pensa a funzioni logiche complicate.

Ad esempio una NAND a tre ingressi nella logica a CMOS avremo 6 transistori; mentre nella logica a rapporto avremo solamente 4 transistori, dato che il pull-up non cambia.

## 6.1 Inverter CMOS

Entrambi i transistori controllati da tensione di ingresso, attivi in fasi differenti e funzionamento simmetrico. I parametri non sono completamente simmetrici.

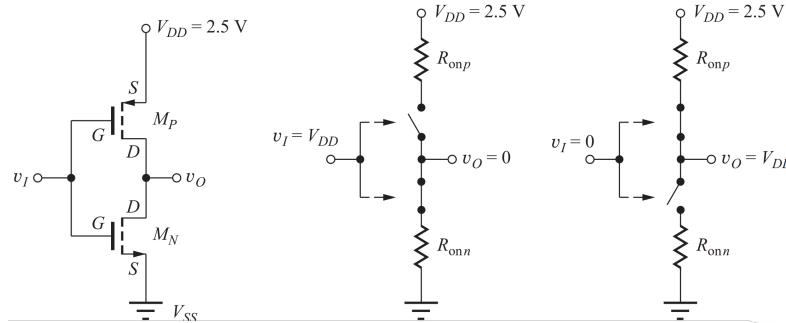


Figura 6.3: Inverter CMOS

In questa maniera, prima avevamo dimensionato il circuito tale per cui l'uscita andava a  $0.2 \text{ V}$ , ora invece l'uscita può andare proprio a  $0 \text{ V}$ , e non ci sarà nessuna competizione (partitore di tensione).

L'uscita non dipende dalle dimensioni, rapporto  $\frac{W}{L}$ , dei transistori, bensì si possono fare tutti della **dimensione minima**. Il rapporto varia solo la transconduttanza, contiene sempre il rapporto, e dunque influenza la resistenza, quest'ultima influisce sulle tempistiche e ora il dimensionamento si può fare sulle tempistiche al posto che sui livelli, che in CMOS sono o  $0 \text{ V}$ , oppure  $V_{DD}$ .

Inoltre si migliorano pure i margini di rumore, dato che utilizziamo tutta l'escursione e non solo una parte.

**NOTA:** Per capire quale è il pMOS e l'nMOS si guarda la freccia della corrente. Nella Figura 6.3 la freccia che entra nel canale indica il pMOS, quella che esce indica il nMOS. Inoltre il tratteggiato indica un canale ad arricchimento, si deve creare lo strato; il non tratteggiato indica un transistore a svuotamento, il canale lo ha già e al limite lo si toglie abbassando la  $V_{GS}$ . Tutto ciò in elettronica digitale, essendo dispositivi ad arricchimento, per semplicità di disegno la linea del canale la si fa solida.

**Per  $V_I = V_{DD}$**

- $V_{GS} = V_{DD}$  per nMOS, che conduce e scarica la capacità C
- $V_{GS} = 0$  per pMOS, che quindi si trova in regione di cut-off e non conduce corrente
- La tensione di uscita finale  $VL = 0$

**Nota:** inizialmente il transistore nMOS è in saturazione, quando l'uscita è ancora alta e man mano che C si scarica, si passa alla zona triodo.

**Per  $V_I = 0$**

- $V_{GS} = -2.5$  per pMOS, che conduce e carica la capacità C
- $V_{GS} = 0$  per nMOS, che quindi si trova in regione di cut-off e non conduce corrente
- La tensione di uscita finale  $VL = V_{DD}$

**Nota:** inizialmente il transistore pMOS è in saturazione, quando l'uscita è ancora bassa e man mano che C si carica, si passa alla zona triodo.

## 6.2 Vantaggi dei CMOS

- Lo 0 e 1 logico corrispondono alle tensioni di alimentazione, dunque ci aspettiamo che margini di rumore saranno grandi;
- I livelli non dipendono dalle dimensioni dei transistori, si possono fare entrambi delle dimensioni minime a vantaggio dell'area;
- Sia lo 0 che 1 logico sono pilotati direttamente, bassa impedenza di uscita ( $\approx K\Omega$ , o meno), alta immunità ai disturbi
- Alta impedenza di ingresso, ogni porta può pilotare moltissime altre porte
- Consumo statico di potenza nulla, regime non c'è un flusso di corrente tra  $V_{DD}$  e massa, solo consumo dinamico (in prima approssimazione)

### Caratteristica statica in tensione

La caratteristica è derivabile incrociando le caratteristiche I-V dei due transistori, imponendo che le correnti di drain siano uguali si individuano precise zone di funzionamento per i due transistori. Dimensioniamo i transistori in modo che  $K_P = K_N$  per avere una simmetria.

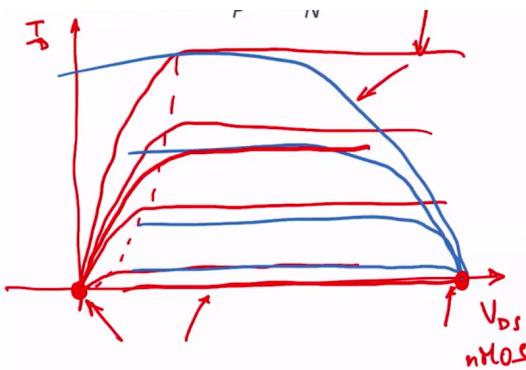
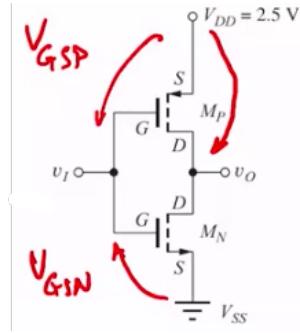


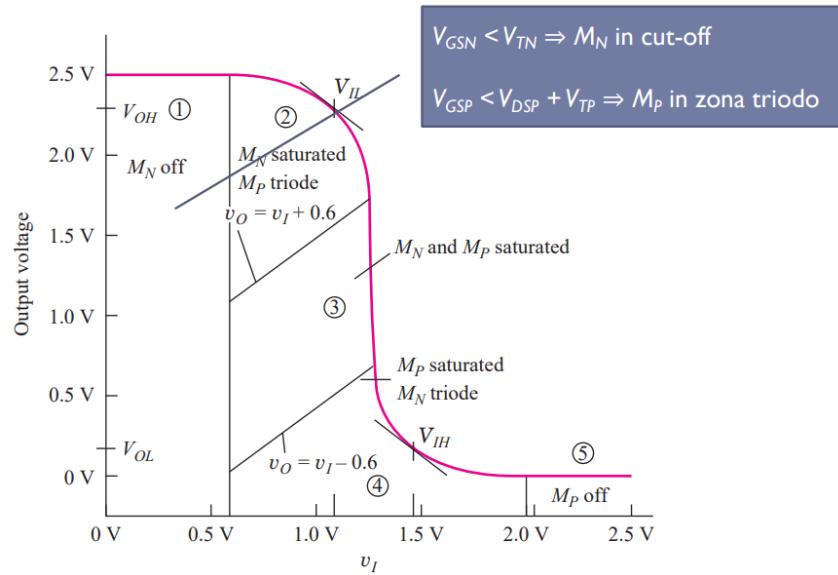
Figura 6.4: Caratteristica statica in tensione CMOS

Partendo dal nMOS alto e pMOS basso, man mano che aumenta l'ingresso la tensione sul nMOS scende mentre aumenta quella del pMOS.

### 6.2.1 Ingresso zero



L'nMOS è in cut-off, dato che la sua  $V_{GS} = 0V$ . Mentre il pMOS conduce e l'uscita conduce alla tensione di  $V_{DD}$ .

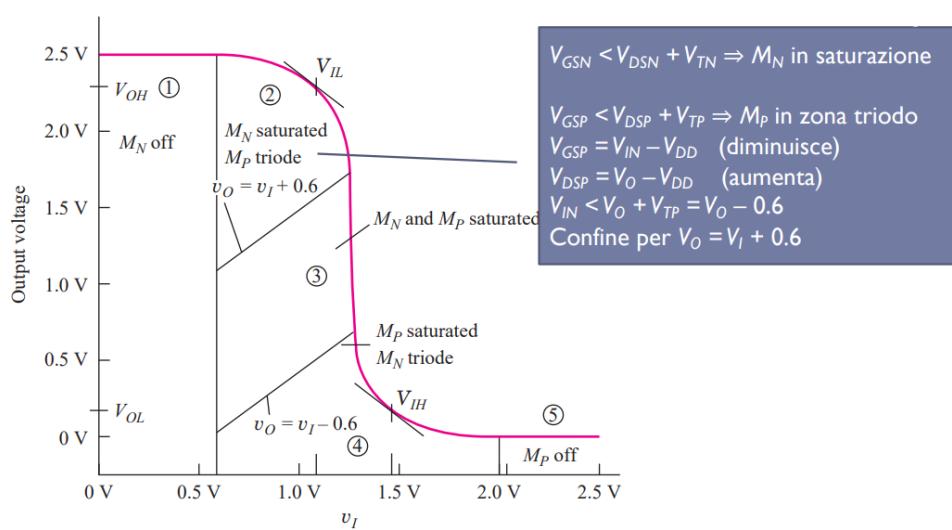


## 6.2.2 Accensione dell'nMOS

Aumentando l'ingresso la  $V_{GSP}$  diminuirà, in valore assoluto avvicinandosi alla sua soglia ( $-0.6 V$ ) ancora molto lontana per il momento. Quando infatti arriviamo ad un ingresso di  $0.6 V$  l'nMOS si attiva, in questa fase di transizione entrambi i transistori sono in conduzione, il pMOS continua ad essere in zona triodo ( $-2.5 + 0.6 V = -1.9 V$ ) molto al di sotto della soglia di  $-0.6 V$ .

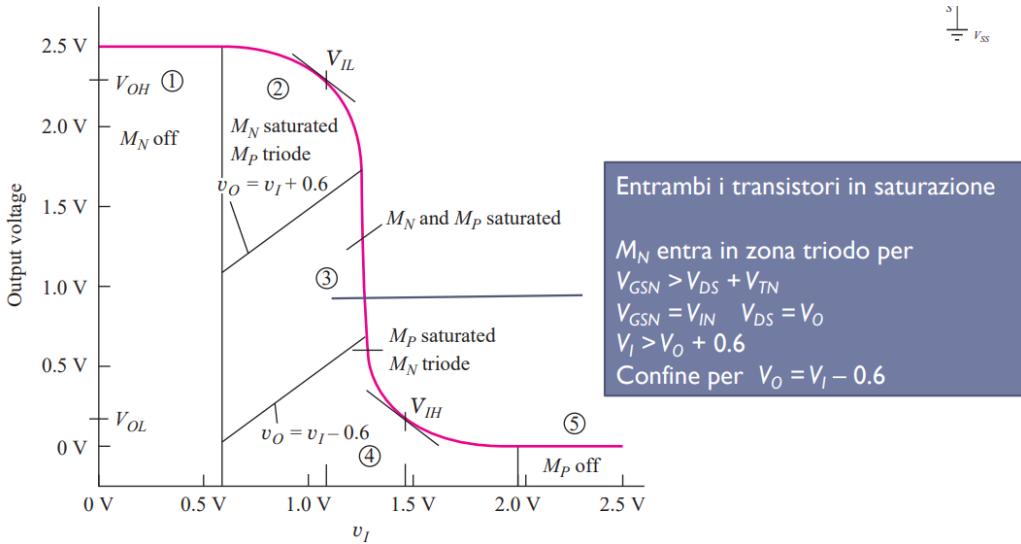
Mentre l'nMOS, accendendosi, si ritrova l'uscita alta ma il suo ingresso è appena sopra la soglia e non siamo dunque una soglia sopra l'uscita, per questo motivo ci troviamo in saturazione.

Aumentando sempre più l'ingresso il pMOS condurrà di meno, viceversa per l'nMOS e l'uscita scende.



## 6.2.3 Aumentiamo ancora l'ingresso

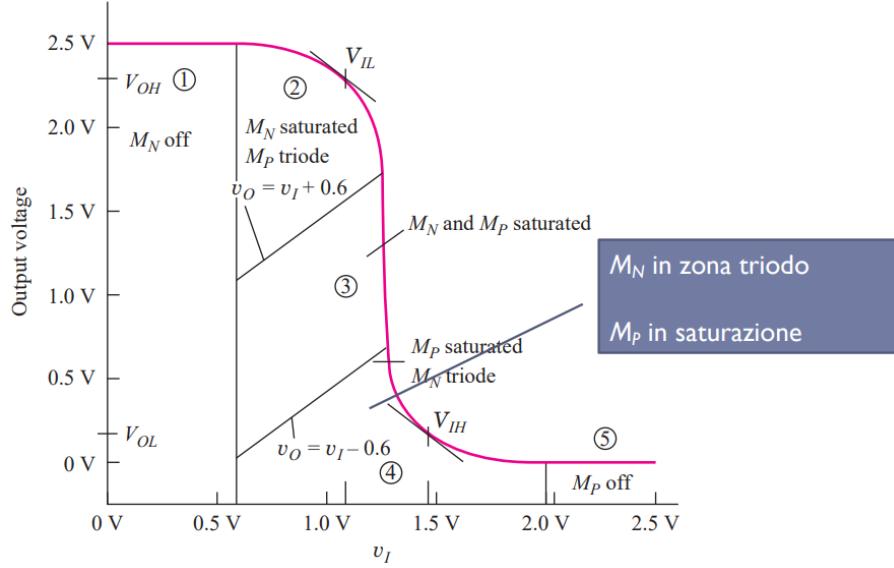
Siamo per entrambi i transistori in zona di saturazione e stanno tirando in modo simile, siamo alla commutazione. Aumentando ancora l'ingresso possiamo far entrare in zona triodo il transistore nMOS.



#### 6.2.4 Zona triodo nMOS

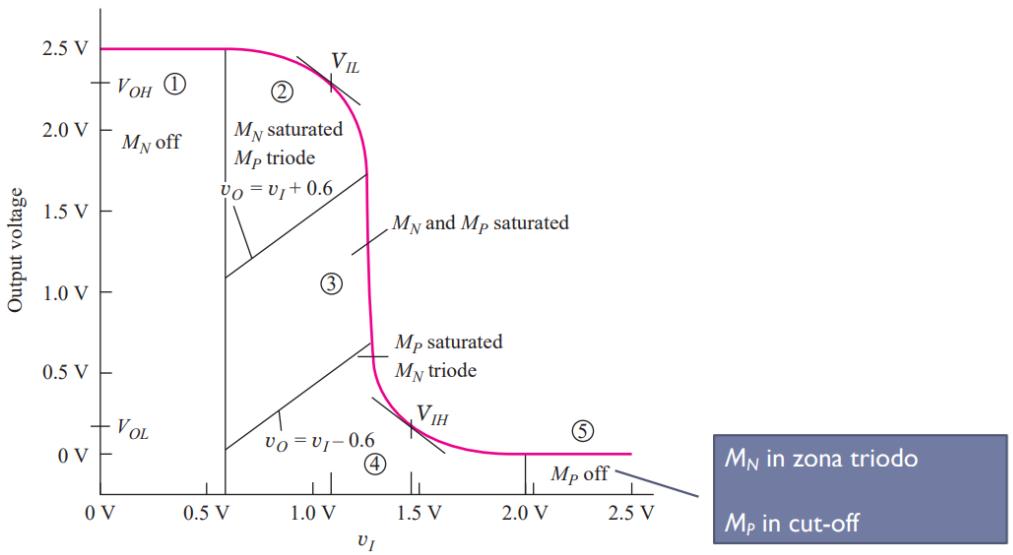
Aumentando ancora l'ingresso a questo punto la  $V_{GSN}$  supererà di una tensione di soglia l'uscita e passerà in zona triodo, con il pMOS in zona di saturazione.

Con l'nMOS che tira di più, la corrente è sempre la stessa però tira l'uscita a massa.



#### 6.2.5 Zona cut-off pMOS

Il pMOS si spegnerà quando la sua  $|V_{GSP}| < |0.6| \text{ V}$ , e a questo punto la tensione in uscita arriva effettivamente a zero.



Se consideriamo dei transistori senza modulazione di lunghezza di canale, la zona 3 risulta effettivamente molto verticale, questo perché le caratteristiche di pMOS e nMOS si sovrappongono. Mentre se si considera la modulazione di lunghezza di canale questa zona sarà meno verticale.

### 6.3 Rapporto tra n e p MOS

Nel caso precedente le transconduttanza erano uguali, ma quest'ultima non è detto che siano uguali. Intanto le mobilità possono essere diverse e il rapporto  $\frac{W}{L}$  anch'esso potrebbe variare.

**Comportamento simmetrico se  $K_n = K_p$**

Per ottenere questo valore basta uguagliare le transconduttanza:

$$K'_n \frac{W_n}{L_n} = k'_p \frac{W_p}{L_p} \quad \rightarrow \quad \frac{W_p}{W_n} = \frac{K'_n L_p}{K'_p L_n}$$

Assumendo  $L_p = L_n$

$$K'_p = 40, K'_n = 100 \quad \rightarrow \quad W_p = 2.5W_n$$

Normalmente si fa una  $W_p$  più grande per compensare il fatto che le lacune hanno una mobilità inferiore, e dunque la corrente del p è inferiore a n.

**Asimmetria**

$$K_R = \frac{K_n}{K_p}$$

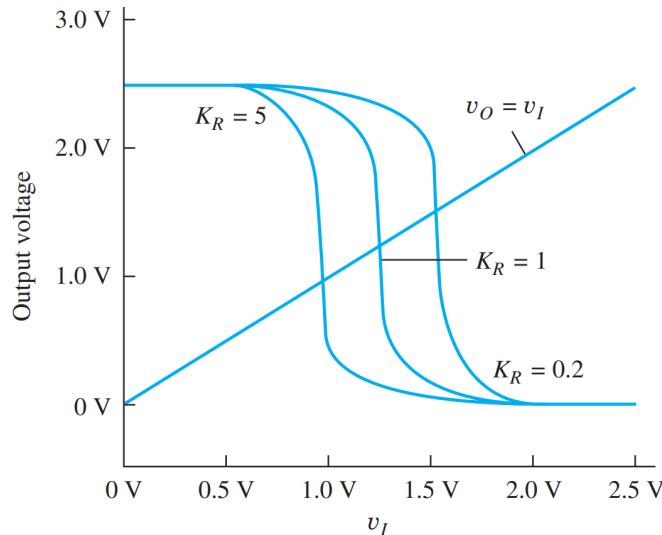


Figura 6.5: Asimmetria

Spesso si cerca di farla simmetrica, ma non è detto infatti spostare la caratteristica si spostano i valori  $V_{OH}$  e  $V_{OL}$  e questo cambia i valori di margine di rumore. Se simmetrico i due margini sono uguali, mentre se è asimmetrico si diminuisce e contemporaneamente aumenta il margine di 1 o 0.

## 6.4 Margini di rumore

### 6.4.1 $V_I = V_{IH}$

Ipotizzando che l'ingresso si alto il pMOS è in saturazione, mentre l'nMOS è in zona triodo, le correnti di drain devono essere uguali:

$$K_n \left( V_I - V_{TN} - \frac{V_O}{2} \right) V_O = \frac{K_p}{2} (VI - V_{DD} - V_{TP})^2$$

Risolvendo per  $V_O$ , e derivando rispetto a  $V_I$  e uguagliando a  $-1$  otterremo:

$$V_{IH} = \frac{5V_{DD} + 3V_{TN} + 5V_{TP}}{8} \quad V_{OL} = \frac{V_{DD} - V_{TN} + V_{TP}}{8}$$

### 6.4.2 $V_I = V_{IL}$

Stessi passaggi visti prima ma in questo caso si applica quando l'ingresso è basso:

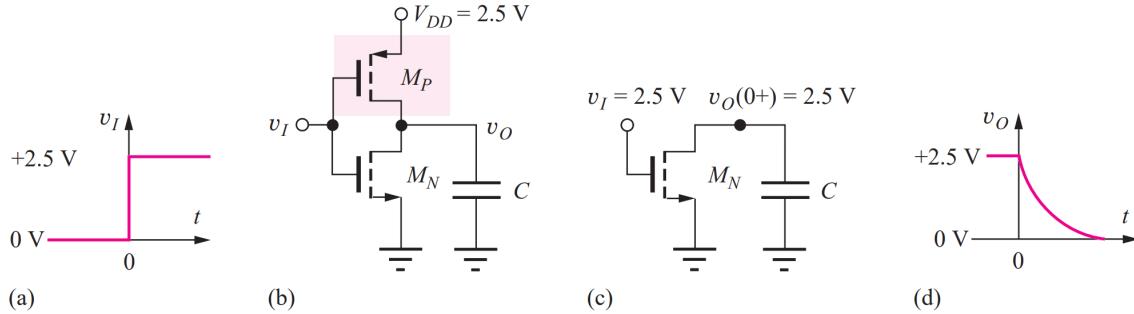
$$V_{IL} = \frac{3V_{DD} + 5V_{TN} + 3V_{TP}}{8} \quad V_{OL} = \frac{7V_{DD} - V_{TN} + V_{TP}}{8}$$

Nel nostro caso i margini di rumore sono entrambi di  $0.93\text{ V}$ , molto buoni se confrontati con l'inverter con logica a rapporto.

## 6.5 Comportamento dinamico

Supponiamo che l'ingresso passi dal valore basso al valore alto, l'nMOS si attiva e il pMOS si interdice e la capacità verrà scaricata.

Per calcolarla useremo la resistenza efficace dell'nMOS e ipotizzando una scarica esponenziale.



### 6.5.1 Transitorio in discesa

Funzionamento simile all'invertitore nMOS a carico resistivo, calcoliamo la resistenza equivalente dell'nMOS supponendo che questo sia in zona triodo:

$$R_{on} = \frac{1}{K_n(V_{GS} - V_T)}$$

Dobbiamo aggiustarlo per un fattore, dato che all'inizio la corrente è costante:

$$R = 1.7R_{on}$$

### 6.5.2 Tempo di discesa:

$$t_f = 2.2(1.7R_{on}C) = 3.7R_{on}C$$

### 6.5.3 Tempo di propagazione da alto a basso dell'uscita:

$$t_{PHL} = 0.69(1.7R_{on}C) = 1.2R_{on}C = \frac{0.63C}{K_n}$$

### 6.5.4 Transitorio in salita

Comportamento speculare al precedente, ma al posto di calcolare la  $R_{on}$  dell'nMOS la calcoliamo del pMOS.

Tutto identico a prima, ora abbiamo  $K_p$  al posto di  $K_n$ .

Nel caso in cui fossero simmetrici, i due risultati ovviamente saranno **simmetrici**.

## 6.6 Osservazioni

### 6.6.1 Ritardi proporzionali a K

Il ritardo dipende dalla capacità e dalla transconduttanza  $K$ , possiamo modificare i tempi di propagazione cambiando il valore di W, la L normalmente la si fa sempre più piccola possibile, più W è grande più  $K_n$  è grande è minore sarà il ritardo questo perché il transistore porta più corrente e carica/scarica più velocemente la capacità.

Questo dimensionamento, con logica CMOS, lo possiamo fare molto agilmente in quanto non influisce sul livello dell'uscita (mentre in logica a rapporto influiva). Il dimensionamento in questo caso lo utilizziamo per migliorare le tempistiche.

**NOTA I:** La capacità aumenta all'aumentare di W, dato che W è il piatto del condensatore.

**NOTA II:** Aumentando W si andrà più in fretta ma si consumerà molta più corrente. Bisogna prestare attenzione anche ad un altro aspetto, vedi NOTA I. Mettendo in cascata delle porte logiche, è vero che singolarmente il transistor va più veloce, ma facendo una cascata di porte logiche con W grande, la capacità del gate a valle aumenta e di conseguenza diminuiamo la velocità del transistore a monte in quanto si trova una capacità molto elevata da pilotare.

**Catena di porte logiche:** se si vuole fare una catena di porte logiche per pilotare carichi ad alta corrente, non si devono fare tutte uguali, bensì si parte da transistori piccoli fino ad arrivare a transistori grandi, questo per ottimizzare il ritardo complessivo.

### Il segnale in ingresso NON è uno step

Questo comporta al fatto che:

- Ci saranno dei tempi di salita e di discesa non nulli
- Questo effetto aumenta i ritardi
- Mettendo vari inverter in cascata, e facendo simulazioni accurate, vediamo che il ritardo è più o meno il doppio di quanto previsto dalle formule precedenti

## 6.7 Dipendenza da $V_{DD}$

Apparentemente, dalle formule viste precedentemente, i ritardi non dipendono dalla tensione di alimentazione. Infatti le formule sono molto approssimate, possiamo usarle per avere una stima dell'ordine di grandezza di cui si sta parlando.

Ovviamente il ritardo dipende dalla tensione di alimentazione. Infatti vi è una dipendenza da  $V_{DD}$  dovuta ad effetti del secondo ordine.

Per ottenere una formula che tenga conto pure dell'alimentazione dovremmo integrare la corrente di scarica e considerare la saturazione di velocità.

È importante tenerne conto in quanto anche i consumi dipendono dalla tensione di alimentazione, più questa è bassa più i consumi sono bassi però rallentiamo. Infatti i processori moderni, quando sono in idle, abbassano la tensione di alimentazione e di conseguenza anche il clock (altrimenti si rischia di fare degli errori dato che si va più piano).

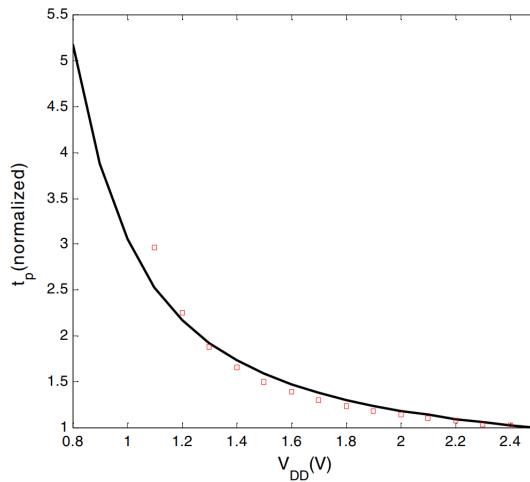


Figura 6.6: Simulazione di un ritardo di un transistore normalizzato

Man mano che diminuiamo la tensione di alimentazione, aumenta notevolmente il ritardo: a  $1.5V$  rispetto a  $2.5V$  aumentiamo il ritardo del 50%.

### 6.7.1 Teniamo conto di $V_{DD}$

Possiamo tentare, in modo approssimato, di ricavare una formula che tiene conto di  $V_{DD}$ .

Corrente di scarica in saturazione è costante (non siamo in realtà sempre in saturazione!):

$$I_D = \frac{1}{2} K_n (V_{GS} - V_{TN})^2$$

Dove  $V_{GS} = V_{DD}$

Il tempo di propagazione lo abbiamo definito come il tempo che passa dal 50% e per calcolarlo sulla  $V_{DS}$  notiamo che siamo ancora abbastanza in saturazione. Quindi la corrente è abbastanza costante durante tutto il primo 50%.

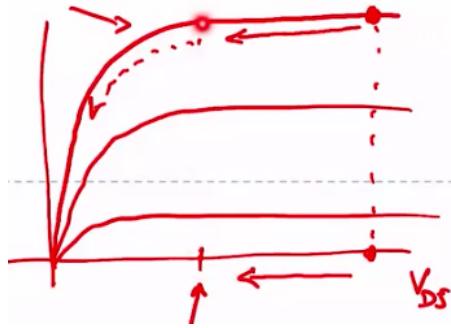


Figura 6.7: 50% di  $V_{DS}$

Quanto impieghiamo per scaricare l'uscita da  $V_{DS}$  a  $V_{DS}/2$ ?

$$I_D \cdot t_{PHL} = \frac{Q}{2} = \frac{1}{2}C \cdot V_{DD}$$

$$t_{PHL} = \frac{C \cdot V_{DD}}{K_n(V_{GS} - V_{TN})^2} = \frac{C \cdot V_{DD}}{K_n(V_{DD} - V_{TN})^2}$$

**Esempi:** dato  $C = 10pF$ ,  $K_n = 500 \mu A/V^2$

1. @3.3 → 9 ns
2. @2.5 → 13.8 ns
3. @1.8 → 25 ns

Se avessimo usato la formula che non teneva in considerazione il cambiamento della tensione in ingresso avremmo avuto:

$$t_{PHL} = \frac{0.63C}{K_n} = 12.6 \text{ ns}$$

Altrimenti si fanno delle simulazioni con i vari Spice.

## 6.8 Saturazione di velocità

Un altro effetto di cui occorrerebbe tenere conto, soprattutto con la miniaturizzazione attuale, è che vi è una saturazione di velocità ovvero che si entra in saturazione molto prima. Dunque questa saturazione riduce la corrente di carica e scarica, i ritardi peggioreranno.

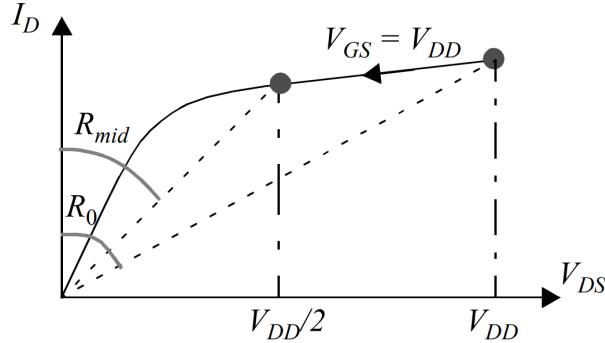
### 6.8.1 Calcolo di resistenza media

Consideriamo il tratto tra  $V_{DD}$  e  $V_{DD}/2$  per quanto riguarda il tempo di ritardo, non è possibile usare la  $R_{on}$ , utilizzata per quando  $V_{DD}$  è nulla.

$$R_{eq} = \frac{1}{2} \left( \frac{V_{DD}}{I_{DSAT}(1 + \lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1 + \lambda V_{DD}/2)} \right) = \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left( 1 - \frac{5}{6} \lambda V_{DD} \right)$$

$$I_{DSAT} = K'_n \frac{W}{L} \left( V_{DD} - V_T - \frac{V_{SAT}}{2} \right) V_{SAT}$$

Minore è il valore di  $V_{SAT}$ , minore è la corrente  $I_{DSAT}$ , e maggiore sarà la resistenza equivalente.



L'inclinazione delle dure rette che collegano l'origine ai punti sopra indicati, è una resistenza.

## 6.9 Consumo di potenza statico

Le porte CMOS non presentano un consumo statico, dovuta all'uscita, di potenza come nel caso dell'inverter nMOS, questo significa che se la porta non sta facendo transizioni, non c'è un cammino da  $V_{DD}$  a massa e dunque nessun consumo.

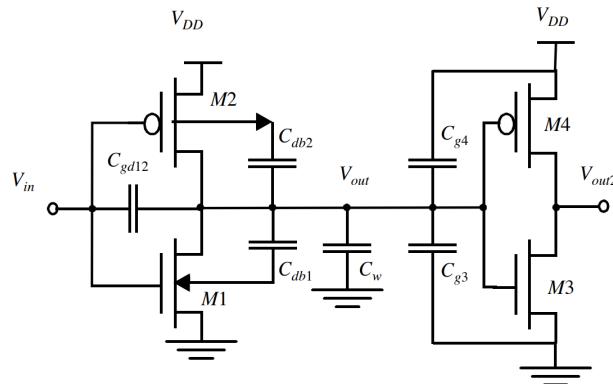
Però, come nel caso di una logica a rapporto, non vi è solo questo fattore di consumo statico bensì:

- stiamo lavorando con componenti reali e il consumo vi è sempre solamente dall'uso dei fili;
- nei diodi tra drain/source e substrato vi è una corrente di polarizzazione inversa, moltiplicata per milioni di dispositivi diventa una corrente non irrisiona;
- Correnti dovute ad effetti del secondo ordine, quali la conduzione sotto soglia;
- Correnti tra l'n-well dei pMOS e il substrato di tipo p.

Consumo comunque molto ridotto rispetto alla tecnologia nMOS, ecco perché molto usata questa logica CMOS.

## 6.10 Consumo di potenza dinamico

Unico consumo significativo per i dispositivi CMOS, dovuto all'energia necessaria per caricare e scaricare la capacità presente all'uscita. Ha luogo solo durante le transizioni delle porte logiche.



Capacità dovuta a vari fattori

- Capacità tra il drain e il gate;
- Capacità di diffusione della giunzione p-n del drain;
- Capacità di gate delle porte a valle;
- Capacità dei collegamenti.

Assumiamo che inizialmente la capacità sia scarica e al tempo  $t = 0$  l'interruttore si chiude e il pMOS carica tramite la sua resistenza interna il condensatore:

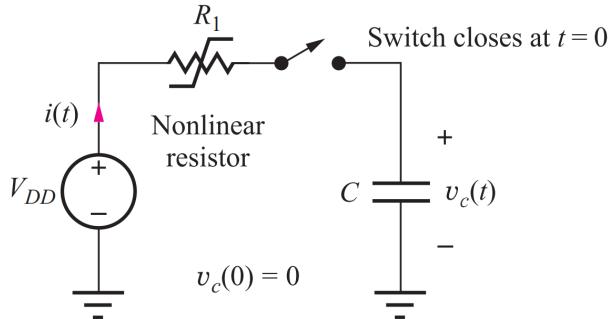
$$E_D = \int_0^\infty P(t)dt$$

La potenza è:

$$P(t) = V_{DD}I(t)$$

essendo  $V_{DD}$  costante possiamo riscrivere la formula precedente:

$$E_D = V_{DD} \int_0^\infty I(t)dt$$



La corrente è quella che scorre nel condensatore:

$$I(t) = C \frac{dV}{dt}$$

Sostituendo a  $E_D$ :

$$E_D = V_{DD} \int_0^\infty C \frac{dV}{dt} dt = CV_{DD} \int_{V_C(0)}^{V_C(\infty)} dV_C$$

Con le condizioni a contorno messe precedentemente:  $V_C(0) = 0 V$  e  $V_C(\infty) = V_{DD}$

$$E_D = CV_{DD}^2$$

L'energia immagazzinata in C è

$$\varepsilon_s = \frac{1}{2}CV_{DD}^2$$

mentre l'energia dissipata risulta essere:

$$\varepsilon_l = CV_{DD}^2 - \frac{1}{2}V_{DD}^2 = \frac{1}{2}V_{DD}^2$$

Dunque **solo metà dell'energia** donata dal generatore ha effettivamente caricato il condensatore, l'altra metà è stata dissipata dal transistore e dai collegamenti.

**Ora facciamo il contrario, scarichiamo la capacità** L'energia  $\varepsilon_s$  immagazzinata nel condensatore viene dissipata nella transizione contraria, pertanto l'energia totale dissipata in una transizione completa allora è:

$$\varepsilon_{TD} = \varepsilon_l + \varepsilon_s = \frac{1}{2}CV_{DD}^2 + \frac{1}{2}CV_{DD}^2 = CV_{DD}^2$$

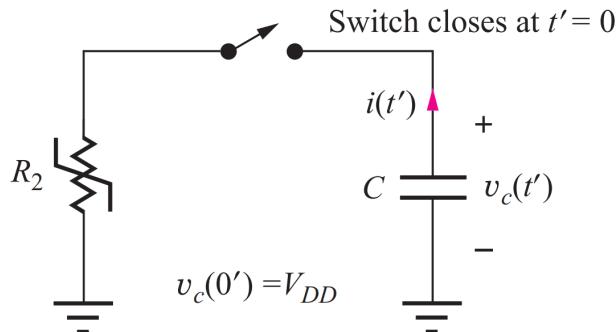
La potenza dissipata equivale all'energia nell'unità di tempo, se la porta commuta con frequenza  $f$  (transizioni complete nell'unità di tempo):

$$P_D = CV_{DD}^2 f \alpha$$

Dove:

$\alpha$  = fattore di attività [0 (segnale costante), 1 (segnale che commuta ad ogni ciclo)]

Più è elevata la capacità e più si consuma, da notare che la tensione compare come termine al quadrato, questo significa che diminuendo la tensione di alimentazione consumeremo molto meno. Diminuendola di metà consumeremo quattro volte in meno, ecco le continue diminuzioni di tensione di alimentazione nei circuiti moderni, ma non troppo dato che vi è il trade-off con le prestazioni (diminuzione di tensione di alimentazione: consumi minori, ma prestazioni peggiori).



La potenza dunque dipende da:

- frequenza di commutazione
- capacità di carico (architettura)
- quadrato della tensione di alimentazione

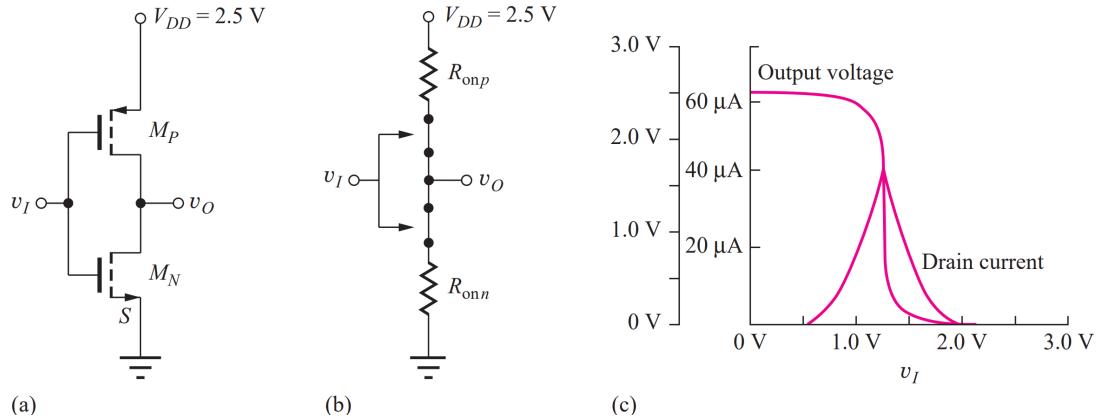
**NOTA:** ad ogni glitch del circuito avremo una transizione completa.

**Circuiti adiabatici:** si vede come il ciclo di carica è efficiente solo per il 50%, ci sono dei circuiti chiamati circuiti adiabatici che provvedono a recuperare il calore speso e trasformarlo in lavoro. Circuiti molto difficili da costruire.

## 6.11 Corrente di corto circuito

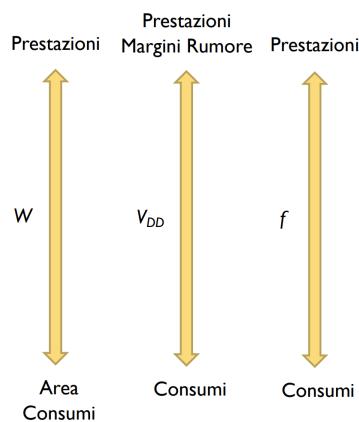
Il segnale di ingresso non passa da 0 a  $V_{DD}$  istantaneamente, infatti i due transistori in questa fase sono in conduzione entrambi. In questo periodo vi è un passaggio di corrente tra  $V_{DD}$  e massa che non contribuisce al caricamento della capacità.

Questo valore di corrente che viene buttata può arrivare fino a valori del 20-30% di quella necessaria per scaricare la capacità. Anche questo inciderà sul consumo di potenza effettivo. Tutto questo dipende dal dimensionamento dei transistori e da quanto vanno veloci i transistori a monte.



## 6.12 Compromessi

- Area: in CMOS possiamo fare transistori piccoli, ciò comporta ad una scarsa conduzione di corrente;
- Consumi: ridurre la tensione di alimentazione, e in concomitanza anche la frequenza (altrimenti si produrrebbero errori), porta a consumi minori, scarse prestazioni. Riduzione di C;
- Margini di rumore: transistori bilanciati (area non minima), tensione di alimentazione alta;
- Prestazioni: transistori grandi, tensione di alimentazione alta



## 6.13 Costruzione di porte logiche

Come abbiamo visto per l'inverter, il pull-up e il pull-down dovranno essere sostanzialmente simmetrici, appunto complementari. Non dovranno mai condurre contemporaneamente, se una ha una connessione in serie, l'altra è in parallelo, e viceversa.

### Esempio della NAND

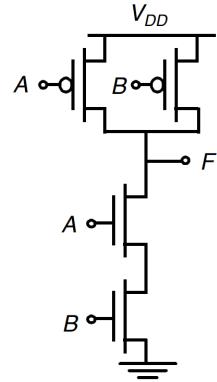


Figura 6.8: NAND in logica CMOS

### Esempio della NOR

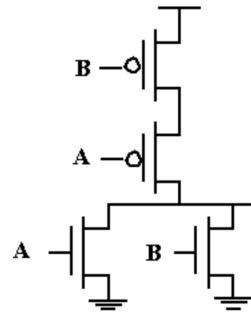
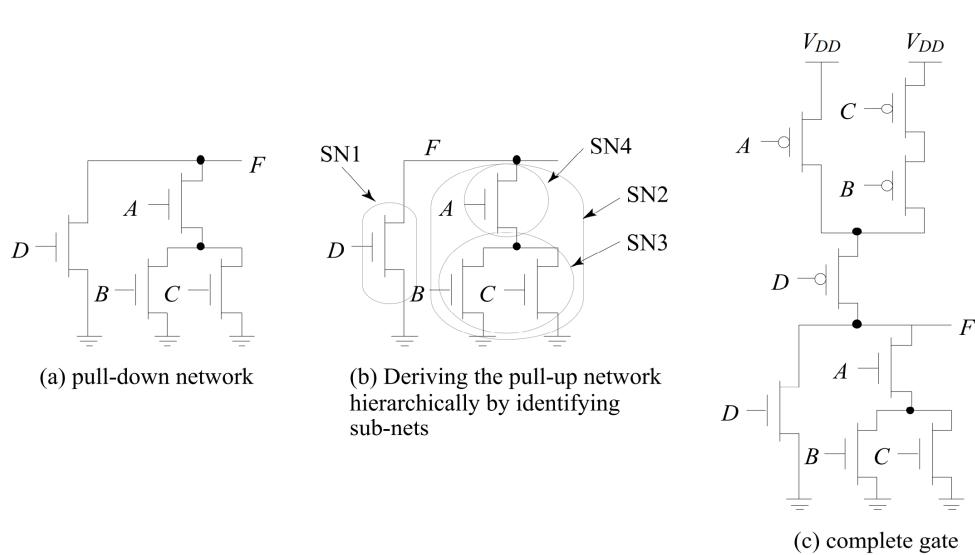


Figura 6.9: NOR in logica CMOS

Realizzare la funzione  $F = (D + A \cdot (B + C))'$



### NOTA:

- Ogni variabile che appare nella formula dà luogo a due transistori;
- Se una variabile appare più volte, darà luogo ad altrettante coppie di transistori;
- Costa di più in quanto è più grosso di una logica a nMOS, questo però si recupera in termini di potenza.

## 6.14 Dimensionamento dei transistori

In questo caso non è necessario per i livelli di uscita, infatti l'output range sempre tra 0V e  $V_{DD}$ , ma è utile per determinare i ritardi, infatti verranno dimensionati per i valori di ritardo che vogliamo ottenere.

Per fare ciò possiamo usare le formule già viste in precedenza, infatti il ritardo è inversamente proporzionale al rapporto  $W/L$ .

$$R_{on} = \frac{1}{K'_n \frac{W}{L} (V_{GS} - V_{TN})} \quad (6.1)$$

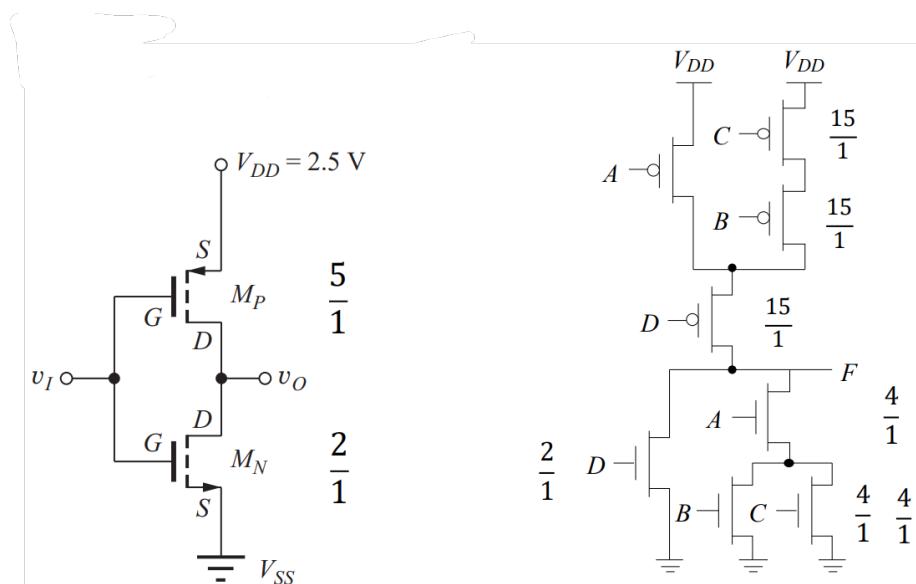
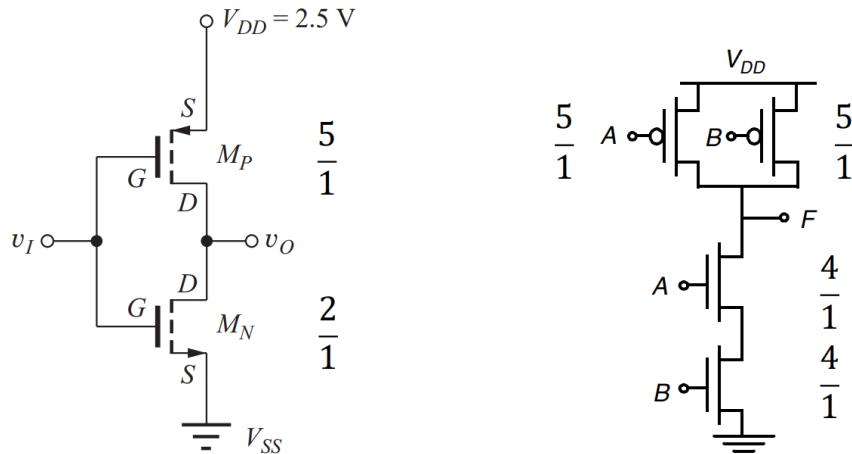
oppure usando un modello che utilizza la velocità di saturazione:

$$R_{eq} = \frac{1}{2} \left( \frac{V_{DD}}{I_{DSAT}(1 + \lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1 + \lambda V_{DD}/2)} \right)$$

dove

$$I_{DSAT} = K'_n \frac{K}{L} \left( V_{DD} - V_T - \frac{V_{SAT}}{2} \right) V_{SAT}$$

Si prenda a riferimento un inverter, e si considerano i casi peggiori, considerando  $K'_n = 2.5K'_p$  mantenendo dei ritardi simmetrici:



Per l'ultimo pMOS si assuma che  $R_{on}$  sia la resistenza di un transistor 1/1. Per un transistor di dimensione W/L la resistenza equivalente sarà:

$$R_{eq} = \frac{R_{on}}{\frac{W}{L}}$$

Per la serie vogliamo che:

$$\begin{aligned} \frac{R_{on}}{\frac{W}{L} D} + \frac{R_{on}}{\frac{W}{L} A} &= \frac{R_{on}}{\frac{5}{1}} \rightarrow \frac{R_{on}}{\frac{15}{1} D} + \frac{R_{on}}{\frac{W}{L} A} = \frac{R_{on}}{\frac{5}{1}} \\ 1 + \frac{15}{\frac{W}{L} A} &= 3 \rightarrow \frac{W}{L A} = 7.5 \end{aligned}$$

## 6.15 Take away

- Margini di rumore elevati: le uscite sfruttano l'intero range di valori, da 0 a  $V_{DD}$ , inoltre i livelli logici non dipendono dal rapporto di dimensione dei transistori, e sono ripristinati ad ogni porta logica;
- Alta impedenza di ingresso, bassa impedenza di uscita, semplifica l'accoppiamento tra stadi successivi;
- Consumi di potenza statici ridotti nulli se non si considera la corrente di leakage;
- Tempi di salita e discesa equivalenti e facilmente controllabili, occorre dimensionare opportunamente i transistori per compensare la differenza di mobilità tra transistori n e p;
- Alta occupazione di area, richiede una coppia di transistori per ogni ingresso.

# Capitolo 7

## Logica a pass transistor

Ora vogliamo trovare una logica che abbia i vantaggi del CMOS e la semplicità dell'nMOS. Una possibile soluzione è la logica a pass transistor.

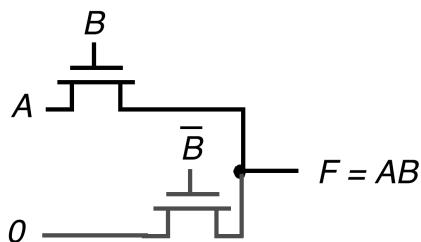
Fino ad ora i segnali in ingresso viene usato per attivare il transistor ed è quest'ultimo che sposta l'uscita, dunque il segnale in ingresso non agisce direttamente sull'uscita ma ha azione indiretta.

In questa logica il segnale di ingresso viene usato anche per pilotare l'uscita, dunque l'ingresso non è più solo collegato al gate del transistore bensì è anche collegato al source e drain.

Quando il gate è a 1, il segnale in ingresso al source del transistore passa in uscita. Si può effettuare una notevole riduzione nel numero dei transistor.

### Porta AND

- Quando  $B = 0$ , è abilitato solo il pass inferiore e l'uscita è 0
- Quando  $B = 1$ , in uscita si trova A

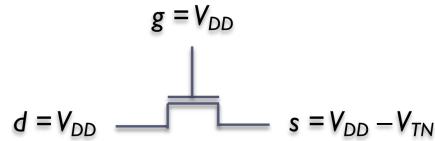


In logica CMOS sono 6 i transistor da utilizzare al posto di 4 come in questo caso (2 in figura e 2 per  $B'$ ), vi è un notevole risparmio!

### 7.1 Livelli dei segnali

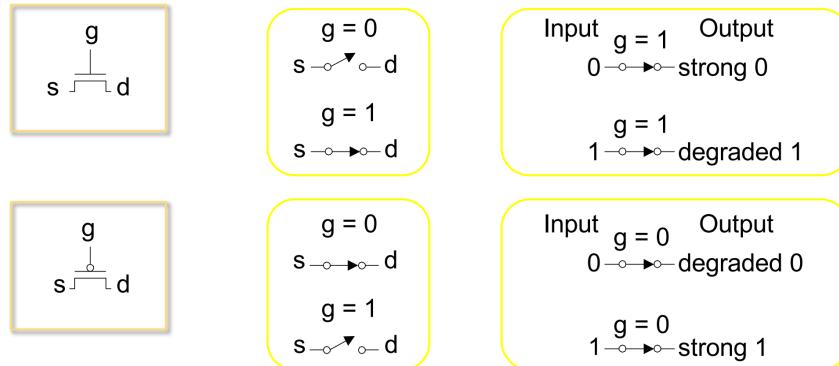
Come visto per l'inverter con carico a saturazione, il transistore nMOS non è in grado di caricare l'uscita fino a  $V_{DD}$ .

- L'uscita arriva al più a  $V_{DD} - V_{TN}$  perché poi il canale svanisce;
- La situazione, come visto, è peggiorata dall'effetto body, visto che  $V_{SB}$  non è nulla
- Nel caso dell'inverter, si arrivava a 1.55 V (2.5 - 0.6 - effetto body) invece di 2.5 V



In pinch-off da entrambe le parti!

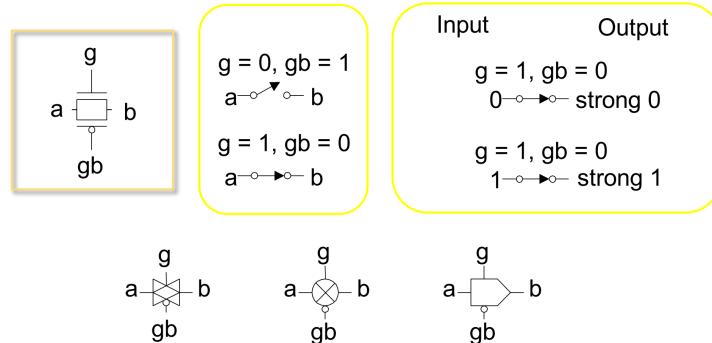
Dunque conduce bene solo uno dei due valori. Possiamo usare transistori nMOS o pMOS, degradano il segnale a causa della tensione di soglia in maniera complementare.



## 7.2 Transmission gate

Un'idea è quella di metterli in parallelo, i transmission gate sono coppie complementari, fanno passare bene entrambi i livello.

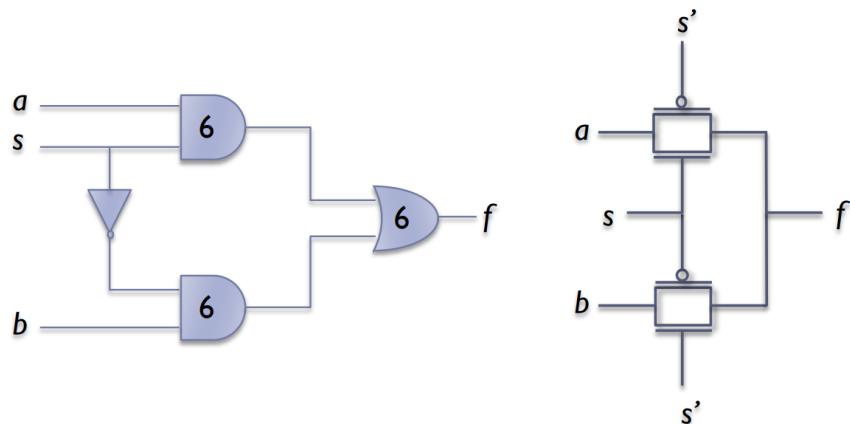
Vi è un problema: **non restoring** il rumore di ingresso viene passato sull'uscita. In logica CMOS questo non avviene, il segnale viene ripulito.



Questo meccanismo spreca un transistore in più, e notare che serve anche il complemento.

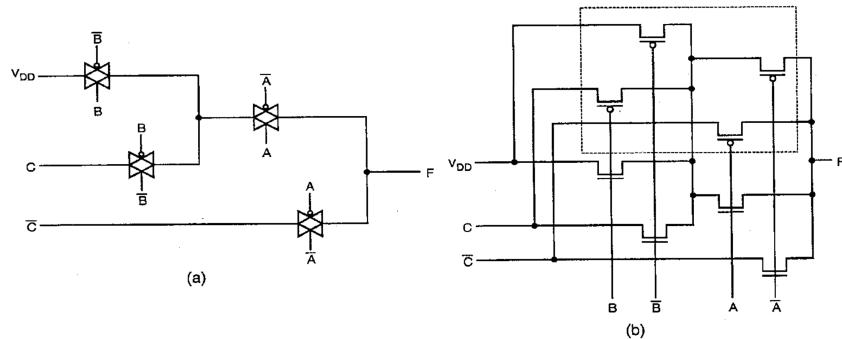
Si realizzi la funzione  $f = as + bs'$

In CMOS occorrono 20 transistori (14 soluzione con NAND). Con transmission gate ne bastano 6.



### 7.3 Funzione che usa la logica transmission gate

Data  $F = AB + A'C' + AB'C$



Funzione di 3 variabili realizzata tramite transmission gate (a), 14 transistori inclusi gli invertitori contro 34 per realizzazione CMOS.

Layout ottimizzato raggruppando i transistori p ed n assieme (b), richiede un solo n-well.

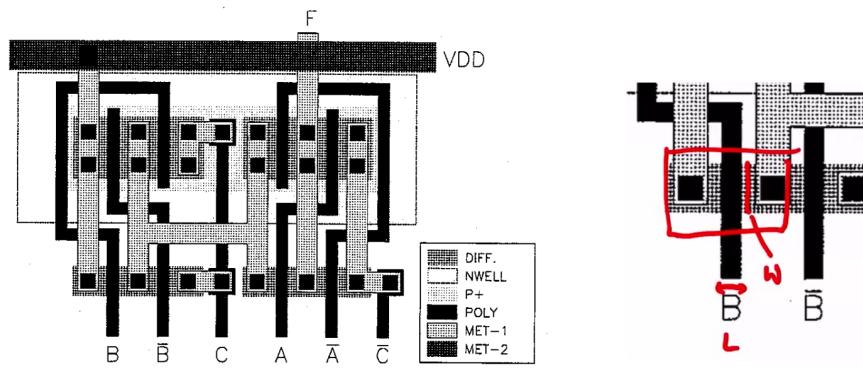


Figura 7.1: Il circuito integrato della funzione (1999)

Per gli attuali circuiti è normale avere 6/8 livelli di metallizzazione, semplifica molto il passaggio delle piste.

## 7.4 Considerazioni

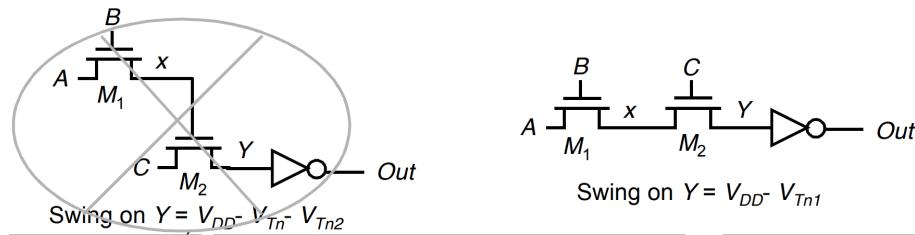
La logica a pass transistor o a transmission gate è non **restoring**: tutto il rumore lo ritroviamo in uscita, per eliminare il rumore basta mettere un inverter in uscita.

La logica CMOS ripulisce invece il segnale a causa dei margini di rumore, dunque la soluzione è quella di introdurre dei buffer CMOS regolarmente lungo il cammino logico in modo da ripristinare i livelli.

**Errore da non fare 1:** Evita di lasciare l'uscita flottante, deve esserci sempre uno ed un solo percorso (altrimenti non sappiamo dove vada l'uscita) da ingresso ad uscita.

**Errore da non fare 2:** La logica a pass transistor non porta l'uscita alla tensione di alimentazione. Anche in questo caso si utilizza un buffer o un inverter in serie al segnale. Si perdono però un po' i vantaggi della riduzione di area.

Poiché il segnale è degradato, attenzione a non pilotare il gate di un altro pass, o si perde una ulteriore tensione di soglia! Quindi si perde una tensione di soglia ogni volta che si passa da GS.



Nell'immagine più a destra invece non si perdono due tensioni di soglia, bensì solamente una.

# Capitolo 8

## Logica dinamica

Questa logica cerca di mettere assieme tutti i vantaggi visti del CMOS, dal punto di vista dei consumi, e della logica a rapporto per quanto concerne le dimensioni.

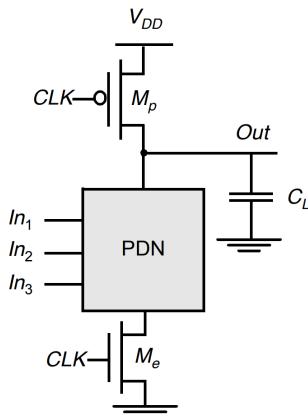
La logica dinamica usa la memoria dei valori dei segnali insita nei nodi di uscita (dato che sono dei condensatori) ad alta impedenza.

- Una volta caricati, se lasciati liberi i nodi non si scaricano subito;
- Usa un solo transistore per ogni ingresso, rispetto a due della logica CMOS (più un paio di servizio)

### 8.1 Due fasi

Il circuito del pull down è identico al CMOS, la quale calcola la funzione.

Inoltre vi sono due transistori aggiuntivi,  $M_p$  e  $M_e$  controllati da un segnale CLK (non necessariamente uguale a quello di sistema).



Operazioni in due fasi  
Prelatura (Clk = 0)  
Valutazione (Clk = 1)

#### 8.1.1 Fase 1: precarica

Nella fase di precarica il segnale CLK = 0, in questo caso :

- il transistore pMOS  $M_p$  caricherà la capacità  $C_L$  a  $V_{DD}$ ;
- il transistore nMOS  $M_e$  stacca il PND da massa in modo che non possa influire sulla fase di carica.

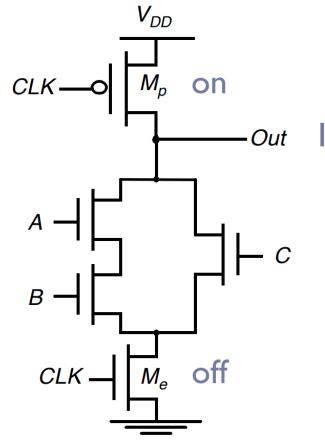
### 8.1.2 Fase 2: valutazione

Nella fase di valutazione il segnale  $CLK = 1$ , in questo caso :

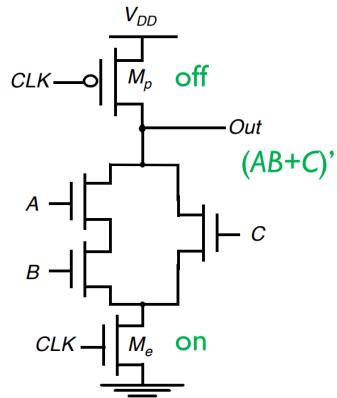
- il transistore pMOS  $M_p$  si disattiva, mentre  $M_e$  si attiva;
- se la rete PND conduce (a causa dei valori di ingresso), allora l'uscita viene portata a livello logico basso: 0 V;
- se la rete PDN non conduce, l'uscita rimane a 1 grazie alla carica immagazzinata nel condensatore  $C_L$ .

Dunque ora invece che avere una funzione complessa, numerosi transistori, nel pull-up abbiamo semplificato tutto mettendo un semplice transistore di precarica.

**Esempio:** Precarica:  $CLK = 0$ ; L'uscita Out va al valore 1 qualunque sia il valore degli ingressi A, B e C.



Valutazione:  $CLK = 1$ ; Si abilita  $M_e$  e si disabilita  $M_p$ ; C'è un collegamento tra Out e massa se e solo se  $A \cdot B + C$  ha valore 1, altrimenti l'uscita rimane 1.



In ogni caso la funzione che ci interessa è quella realizzata con  $CLK = 1$ :  $CLK' + CLK(AB + C)'$ .

## 8.2 Condizioni di funzionamento

Abbiamo visto che ci offre vantaggi dal punto di vista delle dimensioni e da notare anche che non vi è mai un cammino tra  $V_{DD}$  e massa, rispetto ad una logica a rapporto, come nel CMOS. Ci sono anche dei lati negativi...

Una volta che l'uscita della porta dinamica è stata scaricata, non può essere più caricata fino al successivo ciclo di precarica. Quindi i glitch nel pull-down possono avere effetti devastanti.

**Conseguenza:** gli ingressi alla porta dinamica possono fare al più una transizione durante la fase di valutazione.

L'uscita si può trovare in uno stato di alta impedenza (nessun transistore che la pilota, più soggetta a rumore; di fatto staccata dal circuito)

Sia durante che dopo la valutazione (PDN off)

Lo stato è immagazzinato nella capacità  $C_L$

### 8.2.1 Proprietà della logica dinamica (pro's)

- La funzione logica è realizzata solo dalla rete di pull-down, il numero di transistori è pari a  $N + 2$  (contro i  $2N$  per la logica statica CMOS);
- L'output usa l'intero range di valori:  $V_{OL} = Gnd$  e  $V_{OH} = V_{DD}$ ;
- Logica non a rapporto, dimensionamento dei dispositivi non influenza sui livelli logici;
- Compattezza: capacità di carico ridotta per il minore numero di ingressi e transistori in uscita;
- Nessuna corrente statica, tutta la corrente fornita da PDN è usata per scaricare  $C_L$ .

### 8.2.2 Proprietà della logica dinamica (con's)

- Sommando tutto, il consumo di potenza è più elevato che per la logica statica CMOS, ma minore della logica a rapporto

No consumo statico;

L'uscita commuta con più facilità (se fosse sempre a 0, commuta lo stesso);

Il clock ha un carico ulteriore.

- La rete di pull-down comincia a lavorare appena il segnale di ingresso eccede  $V_{TN}$

Quindi  $V_{IH}$  e  $V_{IL}$  sono uguali a  $V_{TN}$

I margini di rumore sono più ridotti ( $NM_L$ )

- Richiede un ciclo di precarica/valutazione

Controllo più complesso

Il periodo del ciclo non può superare un certo valore, altrimenti lo stato viene perso a causa delle correnti di leakage

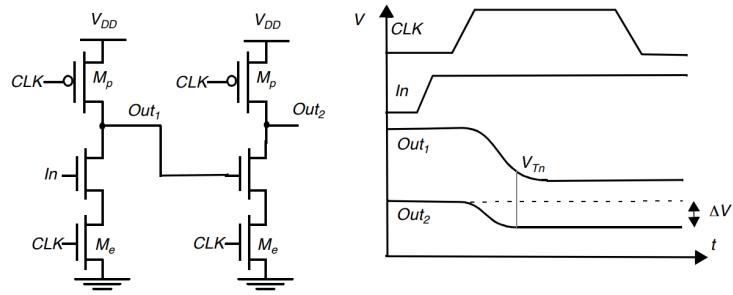
## 8.3 Cascata di porte dinamiche

Mentre in CMOS è molto semplice usare porte in cascata, nella logica dinamica invece è difficile realizzare cascate di porte.

**Esempio:** Durante la precarica le uscite dei due inverter vanno entrambe a  $V_{DD}$ .

Supponiamo l'ingresso In vada a 1: all'inizio della fase di valutazione l'uscita Out1 comincia a scaricarsi, mentre Out2 dovrebbe rimanere a 1.

Però c'è un tempo durante il quale l'ingresso del secondo inverter è ancora a 1 e anche Out2 comincia a scaricarsi fino a quando Out1 non scende sotto  $V_{TN}$ . Non si può recuperare il livello perso, e si rischiano malfunzionamenti.



### 8.3.1 Logica domino

Per operare correttamente, gli ingressi dovrebbero solo fare transizioni da 0 a 1 durante la valutazione, in questo modo non si rischia di attivare per errore la PDN.

La logica domino inserisce un inverter dopo ogni porta, e grazie a ciò si potranno fare solo transizioni da 0 a 1: le uscite quindi sono inizialmente a 0 durante la precarica, ed eventualmente vanno a 1 se si attiva la PDN.

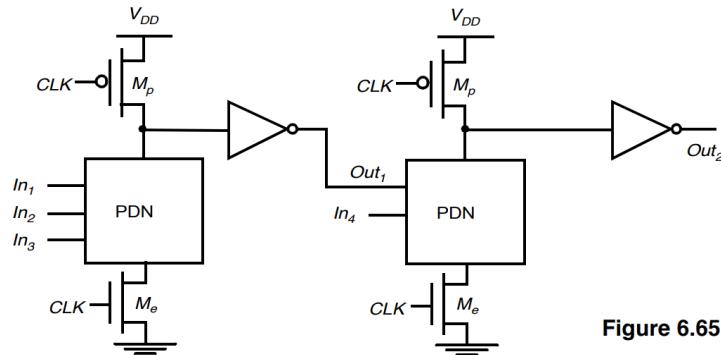


Figure 6.65

Facendo così però, la PDN da invertente, introducendo l'invertente, la logica domino permette di realizzare solo funzioni non invertenti. Questo è un grande svantaggio in quanto non si possono realizzare tutte le porte logiche, AND e OR, che sono le due porte non invertenti, non costituiscono un insieme completo di operatori, serve anche la NOT.

## 8.4 Altre logiche: take away

Abbiamo visto varie soluzioni per la costruzione di porte logiche:

- Logiche a rapporto;
- Pass transistor;
- Logiche dinamiche.

con vari compromessi:

range dei livelli;

i consumi;

margini di rumore;

velocità;

rumore.

Ovviamente nulla viete di usare un mix di modelli per la costruzione: usare CMOS dove servono consumi ridotti a scapito delle dimensioni; logica dinamica quando non abbiamo problemi di interconnessioni e sappiamo che gli ingressi non variano così spesso; parti con pass transistor e transistor gate come ad esempio i multiplexer i quali vengono molto semplici soprattutto quando servono multiplexer molto grossi.

Tipicamente la logica a rapporto non è molto utilizzata visto il grande consumo statico, usato nei casi in cui vogliamo un pull-up semplificato.

# Capitolo 9

## Latches, Flip-Flop e memorie

Fino ad ora abbiamo visto inverter grazie al quale possiamo ottenere tutte le logiche che abbiamo bisogno cambiano il pull-down/up, dunque sappiamo anche implicitamente creare i circuiti combinatori con varie possibili soluzioni.

A questo punto ci resta da vedere come creare gli elementi di memoria, utilizzate nei **circuiti sequenziali**.

### 9.1 Circuiti sequenziali

I circuiti sequenziali, come visto in Reti Logiche, sono formati da una rete combinatoria, che sappiamo costruire, e da una parte di memoria che mantiene i valori della rete combinatoria.

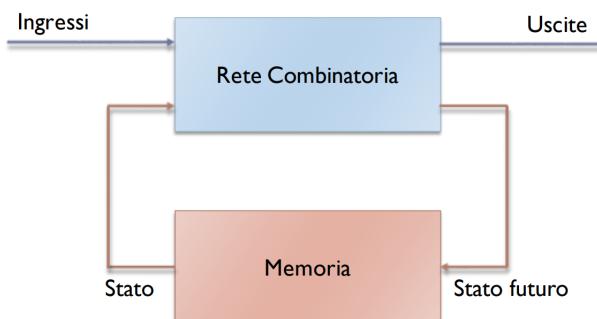


Figura 9.1: Circuito sequenziale

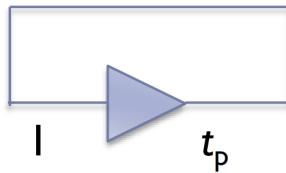
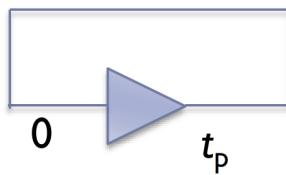
Il problema dunque è quello di mantenere un valore logico. Una primissima soluzione sarebbe quella di sfruttare la memoria insita in un componente elettronico data dal fatto che vi è un dato tempo di propagazione del segnale (ritardo).

Un buffer mantiene un valore al suo interno per un tempo pari al ritardo di propagazione del segnale  $t_p$ , questo ritardo ovviamente non basta vogliamo memorizzare il valore per molto più tempo. Una soluzione potrebbe essere, sfruttando la logica dinamica, quella di usare i condensatori, i quali però si scarica e si dovrebbe quindi "rinfrescare" il valore di tanto in tanto.

Alcune memorie RAM le utilizzano.

### 9.2 Circuiti bistabili

Un modo per non perdere il valore all'ingresso è quello di creare un anello tra uscita ed ingresso, così facendo otteniamo il cosiddetto **loop di feedback**.



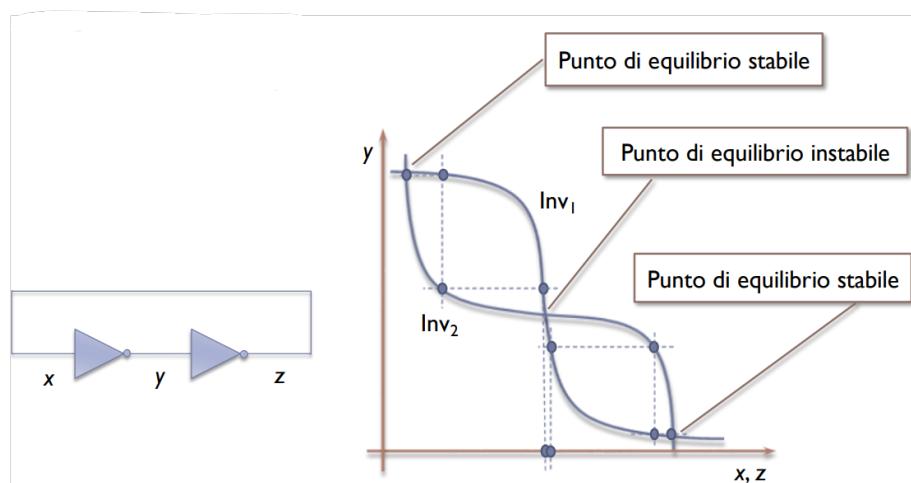
Supponiamo in ingresso ci sia uno 0, se il valore all'ingresso permane almeno per il ritardo di propagazione, questo sarà mantenuto per almeno un altro ritardo, e così via.

Il circuito è **stabile**, è stabile sia per uno 0 che per un 1 e si definisce: **bistabile**.

Per realizzare questo circuito, buffer, basta unire due inverter in cascata.

### 9.2.1 Realizzazione con due inverter

Il circuito di memoria è normalmente realizzato come due invertitori in serie



Partendo da metà, il circuito si porta ad uno dei due punti stabili. Questo si verificherà sempre in quanto nel circuito vi è sempre del rumore. Il problema è che non sappiamo dove andrà, e men che meno quanto tempo impiega per decidere quale valore preservare!

Questo punto è chiamato anche definito come metastabile, ed è utilizzato per creare numeri casuali.

### 9.2.2 Buffer invertente

Proviamo a fare questo ragionamento anche per un buffer invertente.

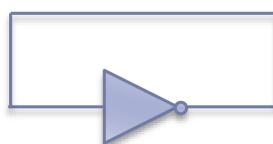


Figura 9.2: Oscillatore ad anello

Supponiamo ci sia uno 0 all'ingresso, dopo un tempo  $t_p$  troviamo un 1 in uscita, questo ritorna all'ingresso e dopo un tempo  $t_p$  diventa uno 0 in uscita, che ritorna all'ingresso, e riparte da capo.

Il circuito non è stabile, abbiamo fatto un oscillatore e il periodo di oscillazione è pari a 2 volte  $t_p$  utile per generare dei clock. Mettendo una resistenza e condensatore si può regolare la frequenza di oscillazione. In laboratorio questo circuito si può fare utilizzando due inverter per allungare un po' il ritardo, altrimenti capiteremo nel punto metastabile.

Attenzione che creando circuiti complessi può capitare che si creino dei feedback, i quali possono creare un circuito oscillante!

Anche gli oscillatori a cristalli di quarzo vengono realizzati con un circuito simile a questo.

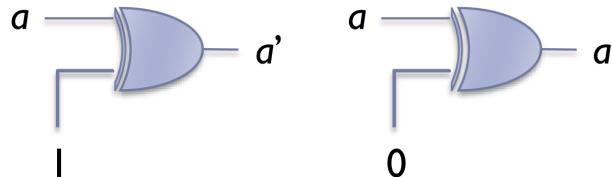
### 9.3 Elemento di memoria base

Abbiamo capito che si può usare un circuito bistabile (due inverter in cascata) per memorizzare un bit. Vogliamo anche poter modificare tale valore a piacere in modo da controllare l'uscita.

Le porte invertenti hanno un comportamento da inverter.

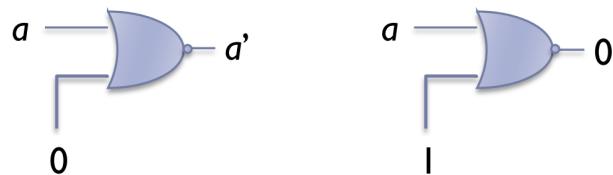
#### EXOR

Possiamo scegliere se invertire oppure no, cambiando il valore sull'ingresso . Pericoloso: se non invertiamo il circuito diventa instabile.



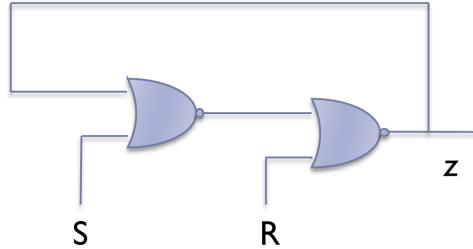
#### NOR o NAND

O invertono o impongono un valore, possiamo dunque decidere cosa fargli fare. Queste porte sembrano soddisfare entrambi i nostri requisiti.



## 9.4 Sostituiamo ciascuno degli inverter con una NOR

Un ingresso prende il posto dell'ingresso dell'inverter, l'altro lo utilizziamo per controllare l'uscita della porta.



### Uso dei comandi

- Per  $R = 1$  l'uscita si porta a 0
- Per  $S = 1$  ed  $R = 0$  l'uscita si porta a 1
- Quando  $S = R = 0$  il circuito si comporta come memoria

## 9.5 Flip flop SR

a	b	NOR
0	0	1
0	1	0
1	0	0
1	1	1

Tabella 9.1: Tabella di verità NOR

### NOR incrociati

Lo stesso circuito si può disegnare in un altro modo:

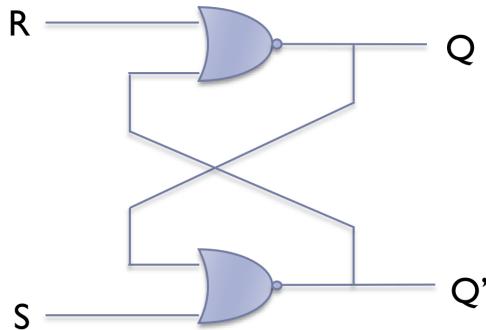


Figura 9.3: Flip Flop SR

a	b	Q	Q'	
0	0	$Q_{-1}$	$Q'_{-1}$	mem
0	1	0	1	reset
1	0	1	0	set
1	1	0	0	nd

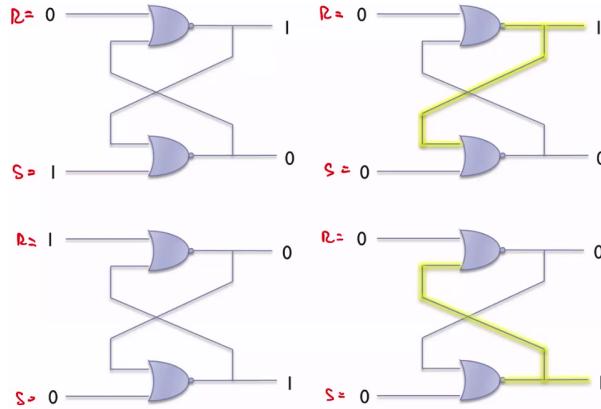
Figura 9.4: Tabella di verità Flip Flop SR

Si evidenzia la struttura simmetrica del circuito.

$$Q = (R + (S + Q)')' = R'(S + Q)$$

$$Q' = (S + (R + Q)')' = S'(R + Q)$$

### 9.5.1 Funzionamento



#### Osservazioni

Le uscite non sono complementari anche se le abbiamo indicate con  $Q$  e  $Q'$ . Però per  $S = R = 1$  le uscite valgono entrambe 0, ma questa è in realtà una situazione da evitare! Si creerebbero delle **possibili oscillazioni**.

Supponiamo  $S = R = 1$ . Supponiamo che vadano entrambi a 0 contemporaneamente.

Tutti i fili sono a 0, le uscite diventano 1 e quindi di nuovo 0 e così via. Il circuito oscilla (0 e 1 si rincorrono, si arriva a regime se i ritardi non sono uguali).

Oscilla sono in questa situazione con gli ingressi a 1, mettendogli poi uno zero il circuito non sa più dove andare!

## 9.6 Flip flop SR con porte NAND

Utilizziamo due NAND incrociati.

$$Q = (S(RQ)')' = S' + RQ$$

$$Q' = (R(SQ)')' = R' + SQ$$

Funziona in modo duale, in questo caso gli ingressi sono attivi bassi.

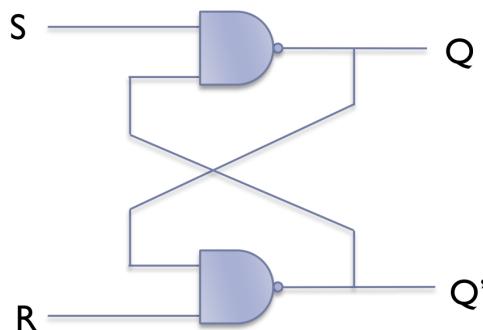


Figura 9.5: Flip flop SR con porte NAND

S	R	$Q'$	$Q'$	
0	0	1	1	nd
0	1	1	0	set
1	0	0	1	reset
1	1	$Q_{-1}$	$Q'_{-1}$	mem

Figura 9.6: Tabella di verità

A questo punto abbiamo realizzato un circuito che riesce ad immagazzinare un bit. A questo punto però vogliamo controllare con un clock quando è il momento di leggere oppure tenere in memoria un dato.

## 9.7 Flip flop SR con clock

Un ingresso di abilitazione controlla quando il flip flop può cambiare di stato, quindi mascheriamo i valori di ingresso con il segnale C!

Il Flip flop è operativo solo quando  $C = 1$ , altrimenti è in memoria. Facendo così abbiamo un inversione la quale rende SR attivo alto, funziona in logica positiva.

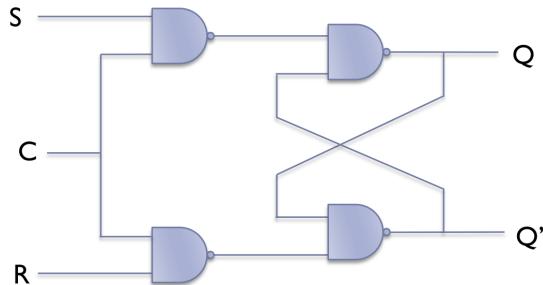


Figura 9.7: Flip flop SR con porte NAND

C	S	R	Q	$Q'$	
0	-	-	$Q_{-1}$	$Q'_{-1}$	mem
1	0	0	$Q_{-1}$	$Q'_{-1}$	mem
1	0	1	0	1	reset
1	1	0	1	0	set
1	1	1	-	-	nd

Figura 9.8: Tabella di verità di Flip flop

Nel caso dell'ultima riga abbiamo realizzato proprio quello che prima abbiamo detto che non si doveva fare, infatti quando S e R sono ad uno e il clock passa da 1 a 0 mette il circuito nello stato di instabilità! In questo caso è più semplice di prima portare S ed R a 1, prima era molto più difficile dato che erano due fili separati, ora i due fili sono controllati dal medesimo clock.

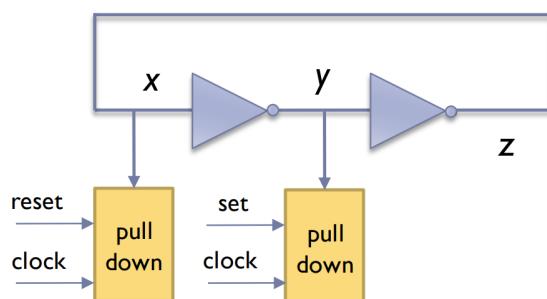
Per evitare questo disguido si potrebbe aggiungere un inverter: con un solo filo, S, si controlla una porta NAND e con  $S'$  si controlla l'altra. Questa soluzione prende il nome di **Latch-D** !

## 9.8 Realizzazione ottimizzata

La realizzazione con NOR o NAND il circuito viene un po' grosso come circuito (17/18 transistori a seconda che sia un flip-flop o un latch).

Una soluzione sarebbe quella di realizzare il circuito con una logica a rapporto.

Collegiamo i due inverter incrociati costruendo un circuito bistabile e proviamo ad imporre un valore ai nodi usando un pull down che sia più forte del pull up dell'inverter.



- Se tiriamo giù la x, mettiamo l'uscita z a 0 (reset);
- Se tiriamo giù la y, mettiamo l'uscita z a 1 (set).

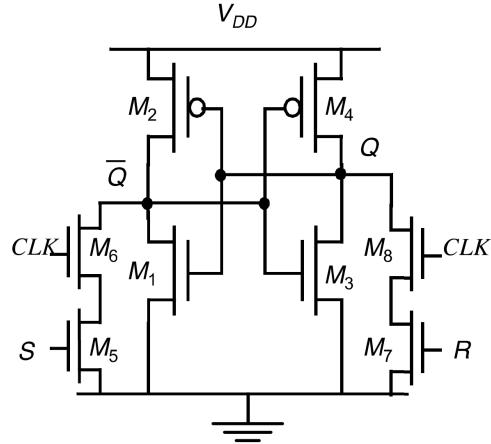
Pull down attivo solo quando il clock è attivo.

Il problema della logica a rapporto: il consumo in condizioni statiche, in questo caso però il vantaggio è che una volta importo il valore il pull-down lo posso staccare perché è il feedback che mantiene il valore.

### 9.8.1 Schema circuitale

Combinazione di

- Inverter incrociati (M1, M2 – M3, M4)
- Doppio pull down (M5, M6 – M7, M8)
- Pull down attivi solo per CLK attivo



### 9.8.2 Funzionamento

Affinché il circuito funzioni, occorre che i pull-down siano più forti del pull-up. Non deve portare l'uscita a zero, infatti se i due inverter sono simmetrici (soglia di commutazione a  $V_{DD}/2$ ), l'importante che il pull-down deve portare il segnale appena sotto  $V_{DD}/2$ .

### 9.8.3 Dimensionamento

Supponiamo che  $M_7$  o  $M_8$  sia disattivato oppure consideriamoli un unico transistor. L'uscita bassa dell'inverter  $M_4 - M_8M_7$  deve essere minore di  $V_{DD}/2$ .

Eguagliamo le correnti di  $M_4$  e di  $M_8M_7$  per  $V_Q = V_{DD}/2$ , ed entrambi i transistori sicuramente in zona triodo.

$$\text{Per } M87: V_{GS} = V_{DD} = 2.5 \text{ V} \quad V_{DS} = V_{DD}/2 = 1.25 \text{ V}$$

$$\text{Per } M4: V_{GS} = -V_{DD} = -2.5 \text{ V} \quad V_{DS} = -V_{DD}/2 = -1.25 \text{ V}$$

$$I_{87} = K'_n \left( \frac{W}{L} \right)_{87} ((2.5 - 0.6)1.25 - 0.78125) = 1.6K'_n \left( \frac{W}{L} \right)_{87}$$

$$I_4 = K'_n \left( \frac{W}{L} \right)_4 ((-2.5 + 0.6)(-1.25) - 0.78125) = 1.6K'_n \left( \frac{W}{L} \right)_4$$

$$K'_n \left( \frac{W}{L} \right)_{87} = K'_p \left( \frac{W}{L} \right)_4$$

Assumendo che  $(W/L)_4 = 6$  e  $K'_n = 3K'_p$

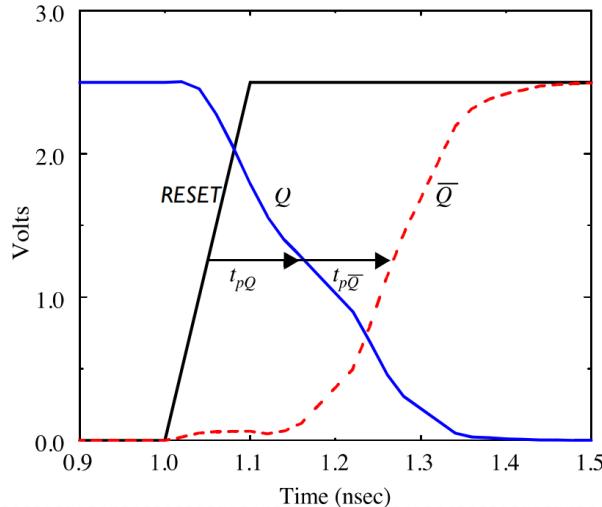
$$(W/L)_{87} = 2$$

$$(W/L)_7 = (W/L)_8 = 4$$

## 9.9 Ritardo di commutazione

Il ritardo tra l'attivazione del reset e la commutazione di Q' è dato dalla somma di 2 componenti.

- Il ritardo per portare Q a  $V_{DD}/2$  dovuto all'inverter pseudo nMOS fatto da M4 e M8 / M7
- Il ritardo dovuto all'inverter CMOS fatto dai transistori M2 ed M1



## 9.10 Latch basato su multiplexer

Un altro modo di realizzare il flip flop è quello di usare un multiplexer in retroazione, in VHDL scriveremo:

```
q <= d when clk = 1 else q;
```

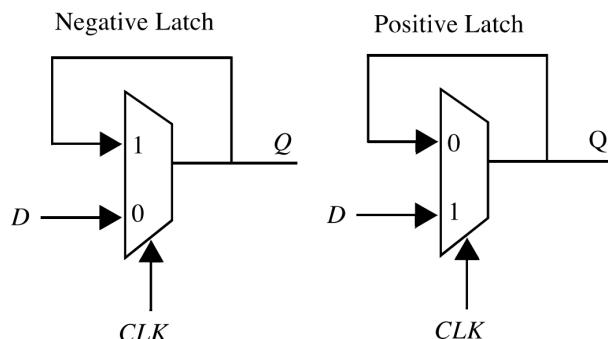
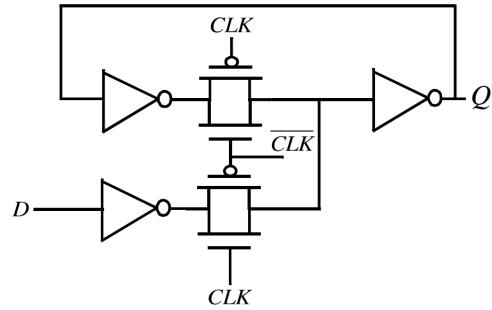


Figura 9.9: Latch basato su multiplexer

Questo dispositivo può essere realizzato con logica a transmission gate:

Quando CLK = 0 si instaura il feedback

Quando CLK = 1 si apre l'anello e si può configurare un nuovo valore



Non possiamo semplicemente prendere il filo e riportarlo al suo ingresso, bensì dobbiamo inserire la funzione identità utilizzando due inverter i quali provvedono a mantenere un segnale stabile dall'inizio alla fine del canale.

Un altro inverter è inserito prima del transistore di sotto per disaccoppiare lo switch da chi lo pilotava e abbassare il rumore.

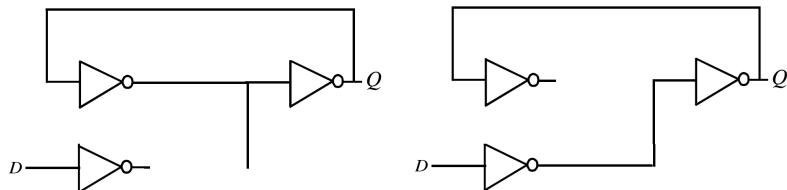
Di fatti abbiamo creato un circuito che mantiene il valore, aprendo o chiudendo l'interruttore (transistor superiore) posso decidere se settare o mantenere un bit.

### 9.10.1 Caratteristiche:

- Il funzionamento non dipende più dal rapporto tra le dimensioni dei transistori (il pull-down non deve concorrere con il pull-up);
- Dato il punto precedente possiamo creare transistori della dimensione più piccola, da tenere in mente che però porteranno meno corrente (logica lenta);
- Il numero di transistori utilizzato è più alto rispetto alla soluzione precedente;
- Il carico sul segnale CLK è più alto.

Il numero totale di transistori aumenta a 12 transistori contro 8 del circuito precedente.

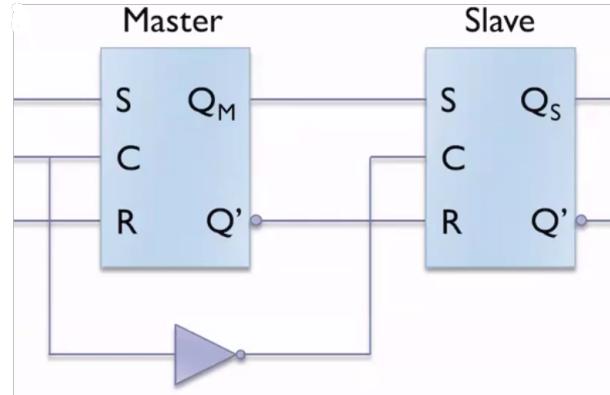
Funzionamento del circuito quando il clk=0, a destra, e quando clk=1, a sinistra.



## 9.11 Flip flop di tipo Master Slave

Nei circuiti sequenziali è molto rischioso usare dei latch sensibili al clock infatti quando il  $\text{clk}=1$ , il latch è trasparente e nulla vieta al segnale di ingresso di variare provocando errori nella logica combinatoria.

Per risolvere questo problema utilizziamo la soluzione vista a Reti Logiche, due flip flop con clock in cascata.



Il primo, detto master, aperto quando il clock è alto; Il secondo, detto slave, aperto quando il clock è basso. In questo modo non avremo mai un cammino che va dall'ingresso all'uscita.

Ma vi è un cammino indiretto tra ingresso e uscita:

Quando  $C = 1$  il primo flip flop è aperto e si può impostare il valore di  $Q_M$ ;

Quando  $C$  esegue la transizione da 1 a 0 il valore di  $Q_M$  viene trasferito nel secondo flip flop e sull'uscita  $Q_S$ ;

Mentre  $C = 0$  tutto sta fermo perché il primo non è sensibile, e il secondo ha gli ingressi fissi.

Vediamo che questo circuito è sensibile al fronte di discesa: falling edge.

Una variante del circuito è il **Flip flop edge triggered**, al posto di metter un flip flop RS come ingresso, utilizziamo un flip flop di tipo D.

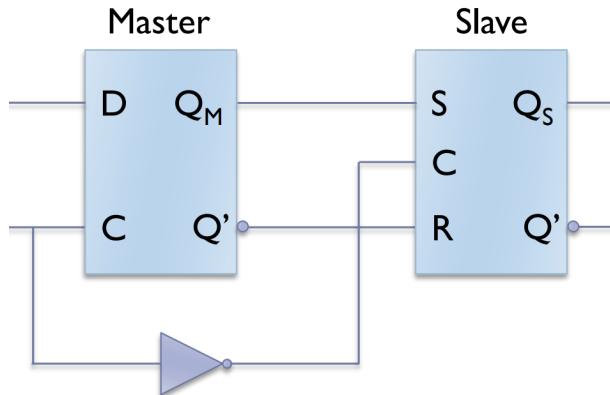
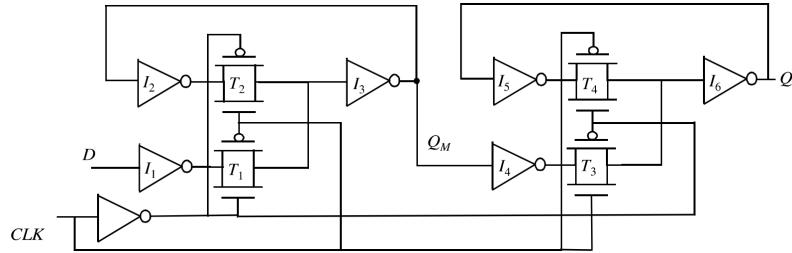


Figura 9.10: Flip flop edge triggered

Il valore di  $Q$  è sempre uguale a quello di  $D$  nel momento in cui il clock va da 1 a 0. Un solo ingresso  $D$ , serve per pilotare  $S$  ed  $R$ . Il resto funziona in modo analogo a quello precedente.

### 9.11.1 Circuito di memoria utilizzando flip flop edge trigger



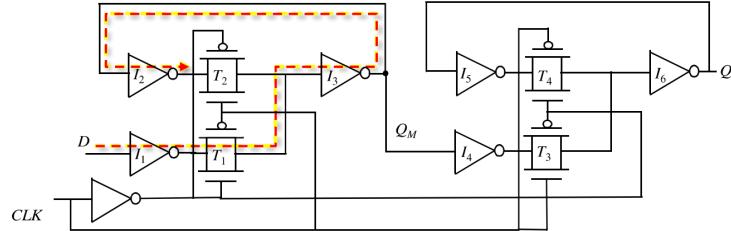
Questo circuito è da caratterizzare in termini di tempistiche, sono due molto importanti:

- **Tempo di set-up:** tempo prima del fronte del clock in cui il segnale D deve essere stabile;
- **Tempo di propagazione:** tempo perché il segnale si propaghi fino all'uscita Q a partire dal fronte del clock;

In particolare quando vogliamo settare un valore, set-up, dobbiamo aspettare un tempo per il quale il valore sia passato per  $I_3$  fino a  $I_2$  altrimenti avremo devi valori che si rincorrono e continuando a cambiare dato i due inverter. Dunque si dovrà inserire il valore da memorizzare un po' prima del fronte attivo del clock in modo tale che il valore si è già propagato almeno fino all'ingresso del transmission gate  $T_2$ .

### 9.11.2 Tempo di set-up

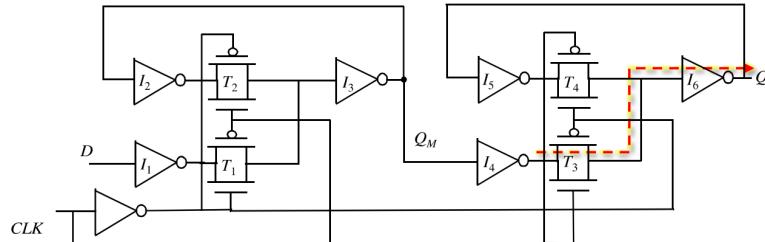
Il segnale all'ingresso di  $T_2$  deve essere stabile quando arriva il fronte del clock, il segnale D deve passare per  $I_1$ ,  $T_1$ ,  $I_3$  e  $I_2$ . Il tempo di set-up è quindi la somma dei ritardi per 3 inverter e un transmission gate.



### 9.11.3 Tempo di propagazione

Il segnale all'uscita di  $I_4$  è già stabile al fronte, perché equivale a quello di  $I_2$  che deve essere stabile per soddisfare il tempo di set-up.

Il ritardo è allora dovuto solo a  $T_3$  e  $I_6$ . Tempo di propagazione uguale alla somma dei ritardi per un transmission gate e un inverter



Contando tutti i transistori il circuito possiede 22 transistori, numero abbastanza elevato, infatti si può fare di meglio.

## 9.12 Alternative realizzatible

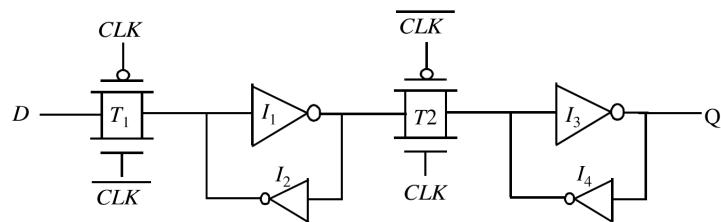
Nei circuiti si vuole sempre minimizzare le dimensioni, questo lo si può fare rimuovendo componenti non strettamente necessari al circuito.

### 9.12.1 Rimuoviamo il transmission gate sul feedback

Rimuovendo il transmission gate sul feedback avremo che:

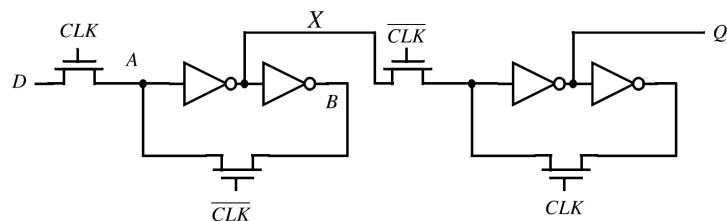
- Meno transistori caricano il clock;
- Ora però chi pilota D assieme al transmission gate T1 deve pilotare più forte di  $I_2$ ;
- Occorre dimensionare correttamente i transistori;
- Anche  $I_4$  deve essere debole, oppure potrebbe influenzare lo stato del primo latch!

Ricadiamo in una logica a rapporto! Per questo ci deve essere una comunicazione tra chi produce la cella e chi la deve utilizzare.

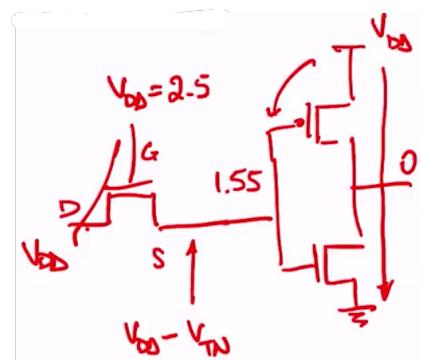


### 9.12.2 Sostituire i transmission gate con pass transistor

In questo caso il problema è che verrà degradato il livello dei segnali, aumentano i consumi.

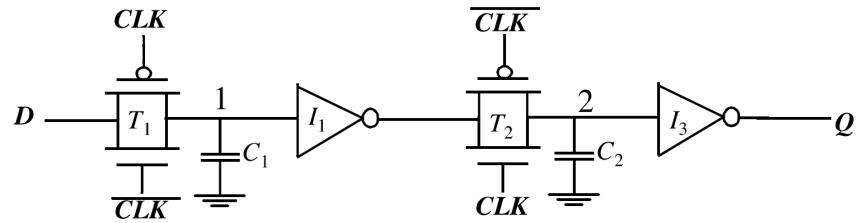


Uno svantaggio di utilizzare la logica a pass transistor con un inverter è che vi è un consumo statico di corrente. Questo è dovuto al fatto che l'inverter viene pilotato con una tensione che si trova in mezzo, e non agli estremi della tensione di lavoro.



### 9.12.3 Usare logica dinamica

Frequenza di commutazione deve essere sufficientemente alta altrimenti la capacità si scarica. Questo può essere utilizzato anche come circuito di campionamento.



Da notare che  $C_1$  e  $C_2$  non ci sono, bensì è la raffigurazione della capacità dell'ingresso dei due transistori dell'inverter.

# Capitolo 10

## Memorie RAM

La memoria non è molto diversa da un flip-flop, la differenza è che solitamente si vogliono memorizzare miliardi di bit ( $10^9$  GB) e la vogliamo molto veloce.

La memoria è un insieme di latch strutturati su una matrice quadrata.

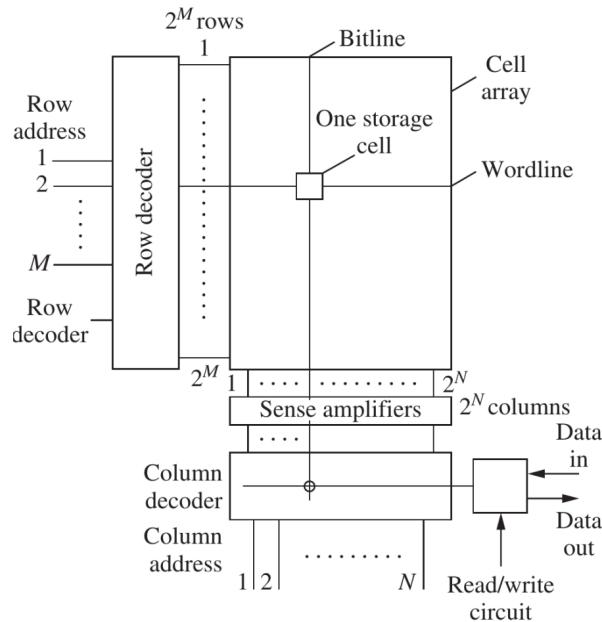


Figura 10.1: RAM

Un decoder seleziona la riga (**wordline**) della cella desiderata usando metà dell'indirizzo, vengono lette tutte le celle della riga selezionata, i sense amplifier amplificano e talvolta ripristinano i valori nelle celle lette. Infine un multiplexer seleziona la colonna (**bitline**) della cella desiderata usando l'altra metà dell'indirizzo.

La lettura di un intera riga non fa perdere molto tempo in quanto avviene in parallelo, potrebbe essere però uno spreco di corrente. Questo spreco normalmente non vi è in quanto se si accede ad una posizione di memoria solitamente si vogliono pure leggere i dati contenuti negli indirizzi successivi i quali probabilmente si troveranno proprio sulla stessa riga. In questo modo posso direttamente leggere la colonna senza leggere ancora la riga, con il solo vantaggio di andare più veloci.

**Tecniche di programmazione** In C/C++, e probabilmente anche in molti altri linguaggi, il compilatore ottimizza il codice scrivendo in celle adiacenti il contenuto delle strutture dati come ad esempio le liste.

Un esempio di codice non ottimo è utilizzare una struct Punto per memorizzare le coordinate x,y. Infatti in memoria verranno memorizzati tante coordinate in questo modo: XYXYXYXY. Questo non è ottimo, soprattutto per calcoli paralleli fatti dalle GPU, se si vuole leggere solo il valore di una variabile, infatti la cache richiederà alla DDR celle successive Y anche se queste non verranno utilizzate per il calcolo.

## 10.1 Cella statica - la cache

Le celle di memoria sono sostanzialmente dei latch, la cella statica non viene utilizzata per la costruzione delle DDR, perché troppo grossa, bensì viene utilizzata per la memoria cache in quanto è più veloce per via del suo funzionamento.

Il suo funzionamento altro non è che il solito loop tra i due inverter incrociati con controllo di accesso.

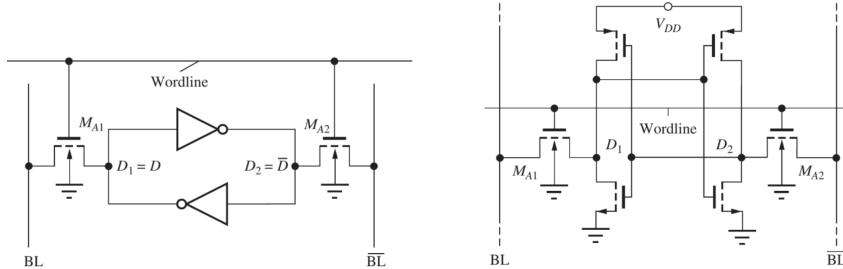


Figura 10.2: Schema circuitale cache

L'unica cosa che è un po' diversa nell'accesso a una cella di memoria è proprio il modo in cui cambiamo il valore che è contenuto nella cella. In questo caso utilizziamo un metodo definito **bilanciato** o **differenziale**: accesso, lettura, di Q e Q' tramite pass transistor  $M_{A1}$  e  $M_{A2}$ . La wordline, generata dal controllore di riga, attiva l'accesso controllando il gate dei transistori.

Le linee che consentono di accedere alla lettura e scrittura di una cella viene chiamata **bitline**. Ogni colonna ha una coppia di bitline, questo è conveniente per essere veloci a determinare il valore della cella osservando la differenza di valori tra BL e BL' utilizzando il **sense amplifier**.

### 10.1.1 Lettura della cella

Le bitline in lettura vengono precaricate a metà della tensione di funzionamento  $V_{DD}/2 = 1.5V$ , mentre la wordline è ancora disattivata. Si potrebbe lasciarle flottanti con il valore precedente, ma si rischia di andare a scrivere invece di leggere.

A questo punto quando la wordline si attiva, le bitline sono flottanti e se guardo la **differenza** tra i due transistori vedo subito da che parte sta andando il cambiamento dei valori, e il sense amplifier lo nota e trasmette subito il valore al chiamante.

**Esempio:** Supponiamo  $D_1 = 0 V$  e  $D_2 = 3V$

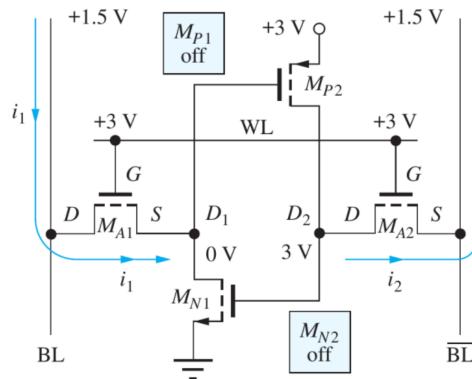


Figura 10.3: Cella 6-T

$M_{A1}$  in zona triodo,  $M_{A2}$  in saturazione.  $D_1$  tende a salire,  $D_2$  tende a scendere. I transistori sono dimensionati in modo tale da evitare che  $D_1$  salga troppo, e si inverta il valore memorizzato!

BL scende, e BL' sale, a questo punto il sense amplifier porta velocemente BL a 0 e BL' a 3 V e termina la lettura.

### 10.1.2 Scrittura della cella

Si pilotano le bitline con il valore desiderato.

Supponiamo di voler scrivere 0 quando prima c'era memorizzato un 1, nel caso in cui volessimo scrivere lo stesso valore contenuto già nella cella non succederebbe nulla.  $M_{A1}$  è in saturazione e scarica il nodo  $D_1$ ,  $M_{A2}$  è anche in saturazione, e fa salire  $D_2$  verso  $V_{DD} - V_{TN}$ , quando  $D_2$  supera  $D_1$ , il feedback positivo fa cambiare il valore nella cella.

**NOTA:** Le celle che pilotano le bitline devono essere dimensionate in modo da vincere contro gli inverter della cella, bisogna portare il valore oltre il punto di equilibrio instabile.

Questa cella funziona molto bene, ha consumi molto bassi dato che se non cambia nulla il consumo statico è zero, i tempi di accesso sono molto rapidi date le relative alte correnti che scorrono e sono dotate di un sense amplifier che aumenta notevolmente la loro già ottima velocità.

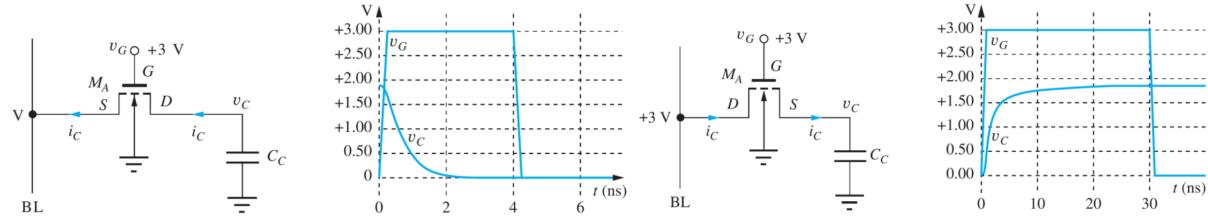
Il problema principale è che sono grosse dati i 6 transistori.

## 10.2 Cella RAM dinamica

A questo punto potremmo pensare di utilizzare la logica dinamica e salvare il valore in un condensatore, in questo caso vero, creato all'interno del silicio.

Questo circuito è molto semplice, tramite la bitline si accede a un transistor il quale carica o scarica la capacità posta sull'altro terminale. Questo condensatore, essendo che deve essere piccolo per avere determinate velocità, nel giro di pochi millisecondi si scarica e dunque il controllore della RAM deve "rinfrescare" i valori per mantenerli invariati, ovvero li legge e li riscrive.

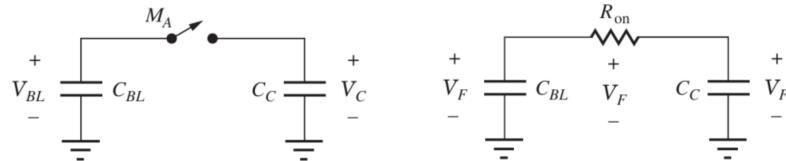
Durante questa operazione di **refresh** la RAM non è accessibile. Ecco perché non può essere usata come cache, quest'ultima deve essere sempre disponibile, deve rispondere ad un ciclo di clock.



Anche in questo caso abbiamo lo stesso problema dato che viene utilizzato un pass-transistor: non arriviamo al valore di tensione della bitline bensì sarà minore, come si può vedere dal grafico Spice di destra. Per far arrivare a 3V il condensatore basta mettere ad esempio  $V_G = 5V$ , la tensione sulla bitline lasciarla invariata e sul condensatore avremo il valore desiderato. Ecco il perché delle varie tensioni che vi sono sulla scheda madre: 3V, 5V, 12V.

### 10.2.1 Lettura e Scrittura

In questo caso, a differenza delle celle statiche, non c'è un inverter che pilota la bitline, ma c'è appunto il condensatore che da solo deve far in modo di variare abbastanza la bitline per leggere il valore correttamente. La bitline invece è realmente raffigurabile come un condensatore molto grosso.



Come prima si carica a metà, della tensione immagazzinata nel condensatore, il valore della bitline. Per leggere il valore chiudiamo l'interruttore, attivando il transistore di accesso. A questo punto abbiamo un circuito costituito dai due condensatori e da una resistenza equivalente  $R_{on}$  data dal condensatore stesso.

A fronte ci ciò, si avvierà un trasferimento di carica da una parte all'altra a seconda se il condensatore sia carico o meno, fino a quando le tensioni si equivalgono. Quindi calcolando la carica iniziale e finale, possiamo definire il valore.

La carica iniziale è data da:

$$Q_I = C_C V_C + C_{BL} V_{BL}$$

$$Q_F = (C_C + C_{BL}) V_F$$

$$Q_I = Q_F$$

$$V_F = \frac{C_C V_C + C_{BL} V_{BL}}{C_C + C_{BL}}$$

Se  $C_{BL} \gg C_C$ , allora  $V_F \approx V_{BL}$  ed il calore  $C_C$  va praticamente perso. Quindi la tensione finale cambia di pochissimo e per rilevarla occorre un sense amplifier molto sensibile per rilevare la variazione di  $V_{BL}$

### Esempio:

$$C_{BL} = 49C_C$$

$$V_0 = 0V \text{ e } V_1 = 2V$$

precarichiamo la  $V_{BL} = 1V$ , il valore finale risulterà:

$$V_F = \frac{C_C V_C + C_{BL} V_{BL}}{C_C + C_{BL}} = \frac{C_C V_C + 49C_C V_{BL}}{C_C + 49C_C} = \frac{V_C + 49 \cdot 1}{50} = \frac{1}{50V_C} + \frac{49}{50}$$

Dunque:

$$\text{se } V_C = 0V, \text{ allora } V_F = 0.98V$$

$$\text{se } V_C = 2V, \text{ allora } V_F = 1.02V$$

Lo scarto è di soli  $20mV$ , il sense amplifier porta la bitline al valore finale, e riscrive anche il valore nella cella, perché è andato perso, vi è un refresh dell'intera riga implicito.

Una volta letta la riga, non c'è bisogno di rileggerla, bensì si salverà il dato in un vero flip-flop con anello di feedback.

# Capitolo 11

## Sense amplifier

Questo dispositivo funziona esattamente come un latch posto in una situazione di equilibrio instabile e vogliamo che la cella lo faccia cadere da una parte o dall'altra a seconda del suo contenuto.

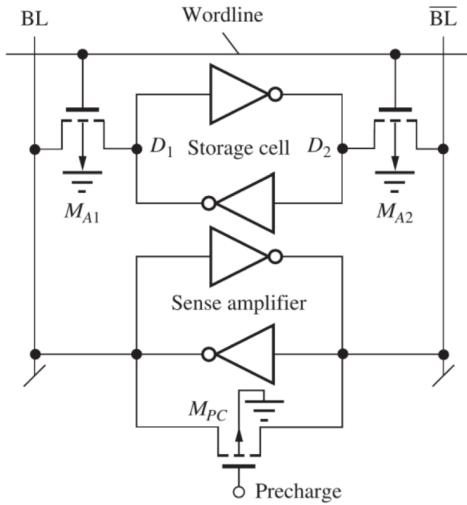


Figura 11.1: Sense amplifier

### 11.1 Precarica e lettura

Per raggiungere questo punto si imposta il valore, sia a destra che a sinistra, alla stessa tensione che si imposta sulla  $BL$ , ovvero metà della tensione di lavoro. Dopo aver caricato il valore si stacca il precharge, si attiva la wordline, la cella fa variare le bitlines e il sense amplifier porta effettivamente al valore finale dato che è nel punto di equilibrio instabile, il quale amplifica questa variazione dato che i suoi transistori possiamo farli grossi, ce lo possiamo permettere perché sono posti solo sulle colonne.

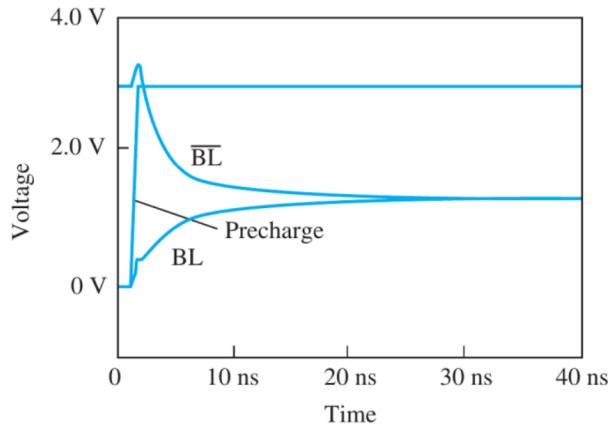


Figura 11.2: Precarica delle bitlines

## 11.2 Consumi

Consumi significativi in fase di precarica, i due inverter nel sense amplifier, i gate sono posti a  $V_{DD}/2$ , dunque entrambi i transistori conducono nella fase di precarica con un conseguente passaggio di corrente e consumo statico.

## 11.3 Row Decoders

I decoder CMOS generalmente sono troppo grossi, infatti MOS a rapporto, sia con porte NOR (uscita attiva alta), sia con porte NAND (uscita attiva bassa). Attenzione però ai consumi!

Dal punto di vista dei consumi è meglio utilizzare delle porte NAND, e per ottenere l'uno in uscita basta metterci un inverter. Il problema che i transistori sono in serie il che comporta un circuito più lento.

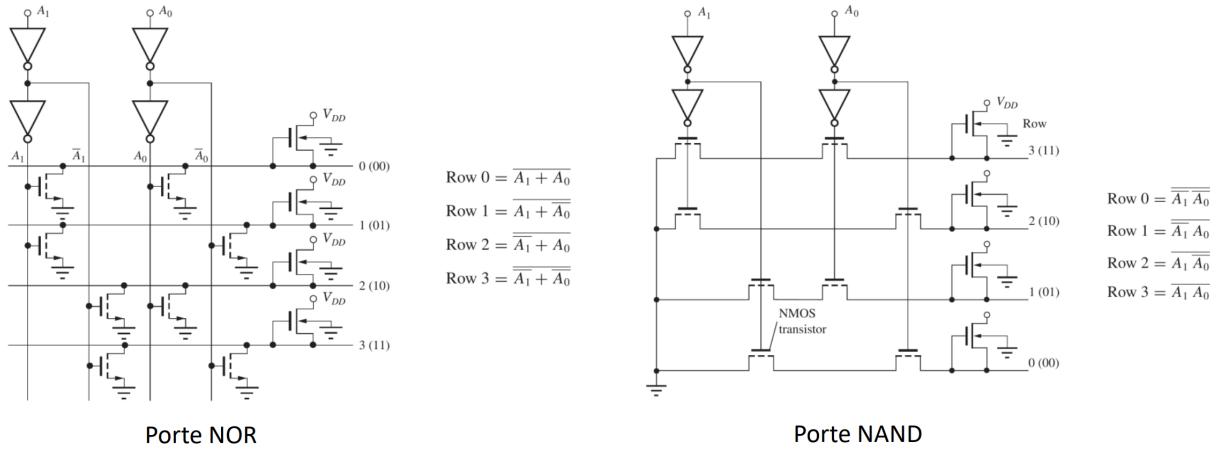
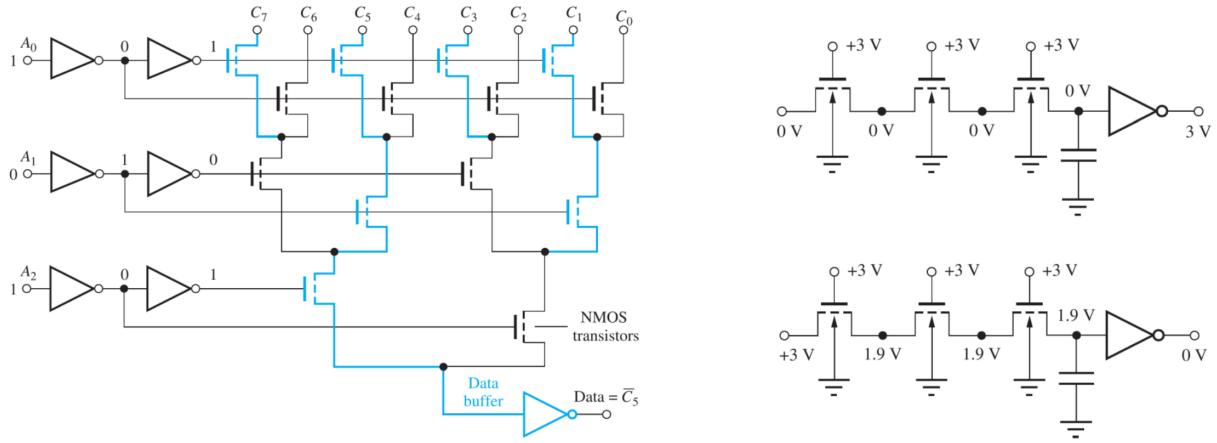


Figura 11.3: Row Decoders

Questa figura rappresenta una decodifica in binario da 2 bit a 4 bit, vogliamo che solo una bit dei 4 valga 1.

## 11.4 Column Decoders

Questa decodifica di colonna non è altro che un multiplexer. La sua realizzazione può essere realizzata tramite CMOS ma diventa troppo onerosa per vie delle risorse. Altre soluzioni possono essere quella di usare transmission gate o pass transistor.



Multiplexer 8:1 (3 bit) a pass transistor

Figura 11.4: Column Decoders

# Capitolo 12

## Memorie a stato solido

Abbiamo già visto come è fatta una memoria RAM, alle bitline si attaccano due inverter messi in feedback, con dei pass transistor accediamo al contenuto il quale verrà letto grazie al sense amplifier.

### 12.1 Memorie ROM

Tutto questo per una cella ROM, memoria non scrivibile, tutto questo processo è più semplice. Per far sì che in una cella venga memorizzato un bit (0,1) non è necessario che si inserisca un Flip-Flop o un condensatore come nella RAM dinamica, l'importante che la bitline vada a 0 oppure 1 quando si seleziona la cella corrispondente.

Dunque è un po' come la DRAM, attacchiamo al transistor da una parte la bitline e dall'altra a massa al posto del condensatore. Il transistore porta a zero il contenuto quindi la bitline leggerà zero, per leggere invece un uno dobbiamo proprio togliere il transistore!

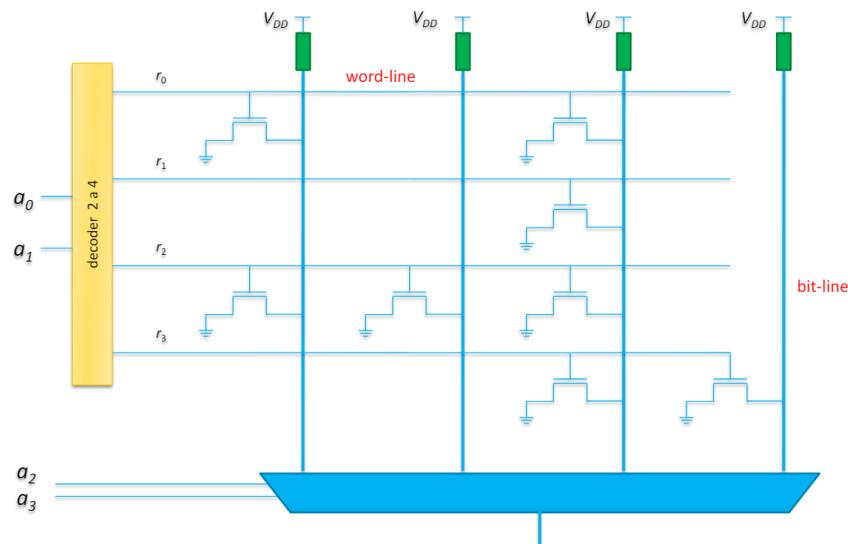


Figura 12.1: 16-bit ROM

**NOTA:** la bitline viene caricata prima con il valore 1. Inoltre questa è una memoria costruita così infatti non è programmabile ma è di sola lettura.

## 12.2 Memorie Flash - riscrivibili e non volatili

La memoria Flash è non volatile e riscrivibile. A differenza della memoria DRAM, prima di potere procedere alla scrittura devono essere cancellati dei blocchi di dati, con conseguente riduzione delle prestazioni di scrittura rispetto a quelle di lettura. La memoria Flash supporta solo un determinato numero di operazioni di scrittura, che varia a seconda della tecnologia utilizzata.

La memoria Flash è disponibile come NAND o NOR. I prodotti SSD utilizzano la memoria Flash NAND perché garantisce una durata maggiore, è meno costosa, le celle sono più dense e le operazioni di cancellazione/scrittura sono più veloci rispetto alla memoria Flash NOR.

La memoria Flash NOR viene utilizzata per memorizzare il codice binario di programmi in quanto è caratterizzata da prestazioni molto elevate in lettura.

### 12.2.1 Il Floating gate transistor

Per rendere la cella di memoria riscrivibile si utilizza un particolare transistor chiamato **Floating gate transistor**.

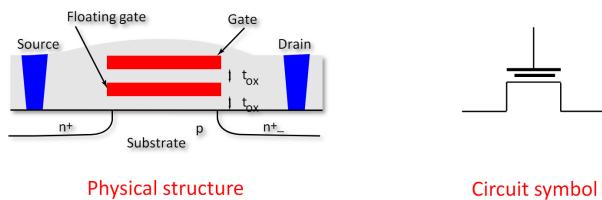
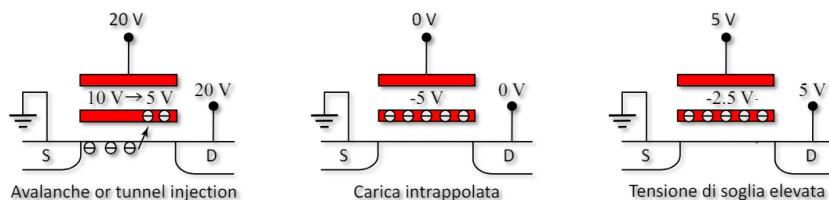


Figura 12.2: Floating gate transistor

La particolarità è che ha due gate, il più in alto è quello standard, quello più vicino al gate ed il canale è **isolato e flottante**, il quale viene interposto tra il gate ed il canale.

A parte l'incremento dello spessore dell'ossido, il floating gate altera anche la tensione di soglia e dunque possiamo cambiarla a piacimento, vediamo come.

### 12.2.2 Programmazione



Vogliamo mettere, spostare, delle cariche su questo flottante, inizialmente il metallo è neutro, ha tanti protoni quanti elettroni. Per far in modo di depositarvi sopra delle cariche si deve aumentare di molto la tensione di alimentazione posta sul gate.

Il campo elettrico attira gli elettroni nel canale verso l'isolante, se questa energia è abbastanza elevata questi elettronni passano attraverso l'ossido di silicio per **effetto valanga** (alta energia), si deve dare tanta tensione al drain, o per **effetto tunnel**, fenomeno della meccanica quantistica che si verifica quando lo strato di ossido è particolarmente sottile e applicando una grande tensione al gate le cariche si spostano, arrivando nel gate di mezzo.

Caricato il gate di mezzo con degli elettroni, le quali caricate bilanciano già le cariche sul gate, ora il gate per formare il canale per mettere in comunicazione il drain e il source, la tensione di gate deve essere ancora più alta di quella di un transistore normale, senza il floating gate (cariche in più).

Dunque mettendo delle cariche nel floating gate l'effetto è stato quello di **alzare la tensione di soglia** del transistore. Questa carica potrebbe permanere anche per diversi anni!

Il transistor non può più essere portato in conduzione (con le tensioni normali), ed è come se non ci fosse, circuito aperto.

## 12.3 Memorie EPROM

Le memorie EPROM, Erasable Programmable Read Only Memory, funzionano proprio così, sono delle memorie con dei floating gate.

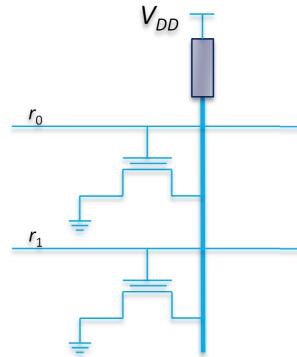


Figura 12.3: Memorie EPROM

Se abbiamo messo le cariche negative sul floating gate, abbiamo programmato un uno, perché la tensione applicata al gate (2.5 V, tensione normale) non è abbastanza per superare la tensione di soglia (si è sempre in **cut-off**). Se non le avessimo messe, allora c'è uno zero, perché attivando il gate con 2.5 V il canale si forma e la bitline va a massa e legge zero, si comporta come un transistore normale.

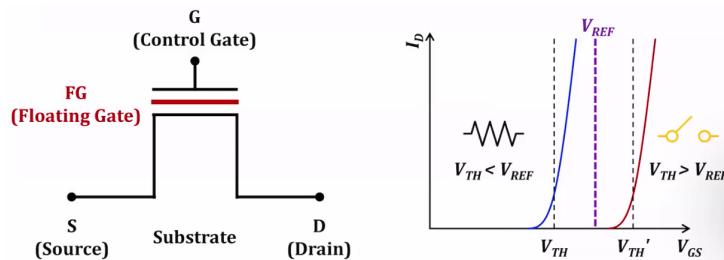
Come detto prima, essendo il gate flottante, queste cariche non vanno più via. Una volta per eliminare la carica nei circuiti caricati con **effetto valanga** si lasciava piccola finestrella la quale serve perché una volta irraggiato di raggi UV per numerosi minuti, la carica scendeva attraverso il substrato e il circuito ritornava programmabile. Questo processo però degrada notevolmente il substrato il quale dopo un po' di cicli non regge più e il circuito risulta da buttare.

Nei circuiti caricati con **effetto tunnel**, il processo di rimozione è molto semplice, basta fare il contrario, quindi applicare una tensione molto negativa al gate per far scivolare via le cariche dal floating gate.

Comunemente oggi si utilizzano le **EEPROM** o **E<sup>2</sup>PROM**: Eletrically Erasable Programmable Read Only Memory, tuttora non sono nemmeno più di sola lettura ma anche di scrittura.

**NOTA:** per vedere se il transistore è stato programmato oppure no, al gate basta mettere una tensione che sta tra le due soglie e:

- Se non è programmato il transistore si accende;
- Se è programmato non si accende.



## 12.4 SSD basate su memoria Flash NAND

Nelle moderne architetture Flash, si può decidere quanta carica mettere nel floating gate, infatti sono multi livello, ovvero possiamo programmare diversi modelli di logica all'interno del floating gate.

In una cella normale la cella o è carica oppure si possono rappresentare multi livelli, quindi più bit.

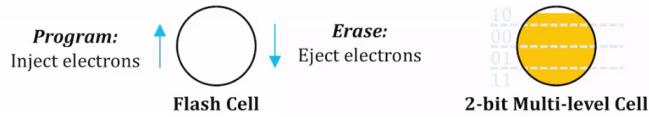
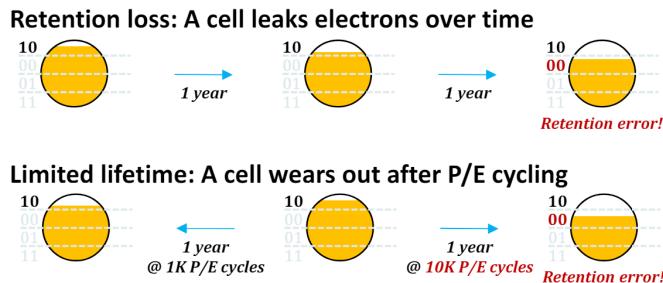


Figura 12.4: Multi-level cell

Il problema è che la carica dopo del tempo si perde fino ad arrivare ad avere un errore questo può capitare spesso nelle celle multi livello. Per ovviare a questo le Flash normalmente hanno dei bit aggiuntivi per della ridondanza e correzione.

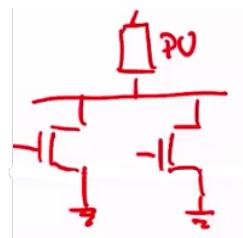
Oltre a questo man mano che si fanno letture e scritture abbiamo elettroni che passano attraverso il substrato grazie all'effetto tunnel i quali danneggiano quest'ultimo tendendo sempre di più ad essere meno isolante. Nel caso in cui si rilevano errori frequenti, quella cella verrà marchiata come non usabile. Per ovviare a questo problema, il controllore dell'SSD, molto sofisticato, cerca di scrivere un po' ovunque per mantenere una usura quantomeno omogenea.



Le celle inoltre diventano più lente in quanto si fa sempre più fatica a discriminare il livello che vi è salvato al suo interno.

### 12.4.1 Organizzazione delle celle

Prima abbiamo visto che l'organizzazione delle celle, ad esempio la cella di memoria EPROM, era organizzata come una NOR.



Questa configurazione è possibile, il problema è che occupano molto spazio, vedremo in seguito il perché.

Quello che invece possiamo fare è quello di organizzare le celle in serie, creando appunto una **NAND string**.

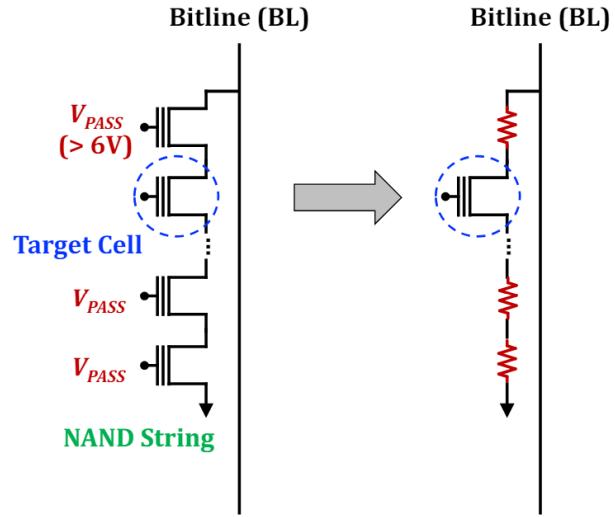


Figura 12.5: NAND string

Se nella NOR per leggere una cella agli altri ingressi mettiamo zero, nella NAND invece tutti gli ingressi dobbiamo metterli ad uno! Perché si comportino come dei cortocircuiti, dobbiamo mettere una tensione  $V_{PASS}$ , maggiore della più alta soglia possibile che si può programmare, sulle celle che non vogliamo leggere, mentre una tensione normale per la cella che vogliamo leggere.

#### 12.4.2 NOR vs. NAND layout

Anche il layout è in favore della NAND, in quanto ha una densità maggiore, questo è dovuto al fatto che non vi è un contatto, la parte N collega un transistor al successivo senza doversi attaccare alla bitline.

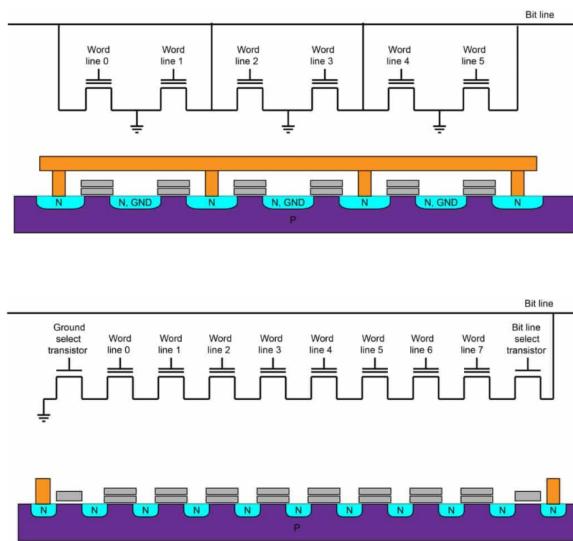


Figura 12.6: NOR vs. NAND layout

Tabella 12.1: Confronto tra celle NAND e NOR

Caratteristica	Celle NAND	Celle NOR
Costo	Più economico	Più costoso
Densità	Più densa	Meno densa
Prestazioni di scrittura	Più veloci	Più lente
Durata	Più lunga	Più corta
Accesso casuale	Supportato	Non supportato
Esecuzione del codice	Non supportato	Supportato
Velocità di lettura	Più lenta	Più veloce

**NOTA:** un possibile uso delle memorie Flash NOR è quello di installarci il BIOS dato che si vuole avere una lettura molto veloce anche a scapito della dimensione.

## 12.5 Organizzazione interna memorie NAND - pagine e blocchi

Nelle moderne architetture le memorie contengono un elevato numero ( $> 100\,000$ ) di celle che operano in modo concorrente, sempre organizzate con uno schema a matrice.

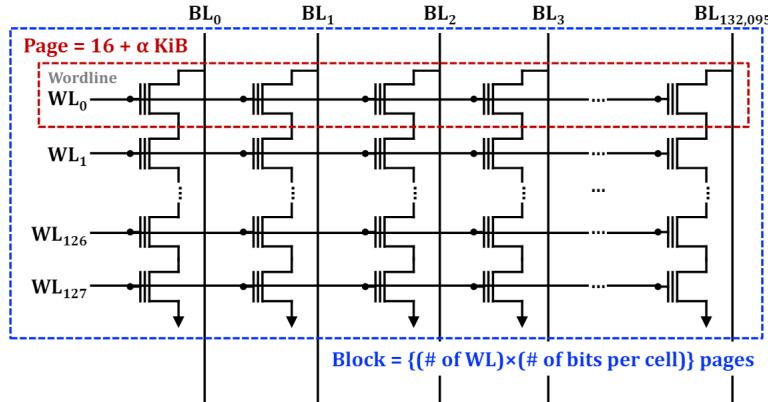


Figura 12.7: NAND - 128 wordline

Possiamo vedere che un blocco contiene molte pagine, in questo caso da  $16 + \alpha KiB$ , dove  $\alpha$  è un numero di bit usato per la ridondanza e correzione di errori.

### 12.5.1 Cancellazione e scrittura

Il processo di cancellazione, decremento della tensione  $V_{TH}$ , e scrittura, incremento della tensione  $V_{TH}$ , nella SSD avviene per blocchi, perché la cancellazione è molto lenta essendo che dobbiamo regolare con precisione la quantità di carica che finisce nei transistori. Togliendogli troppa carica si crea un condensatore che starà sempre acceso il quale porterà sempre allo stesso valore la bitline e quella bitline non potrà essere più utilizzata.

Questo processo infatti solitamente è iterativo in cui si aggiunge carica a poco a poco e si verifica dove si è arrivati. Ecco perché della scrittura a blocchi, ricordiamo anche che pure nei vecchi dischi magnetici si scriveva a blocchi, infatti il file system pressoché è rimasto invariato.

Solitamente dunque quando si deve cambiare dei valori in un blocco, il blocco vecchio lo si elimina piano piano e si scrive il nuovo valore in un'altra parte del disco, così da mantenere anche un usura uniforme. Per cambiare il valore di una cella **NON È POSSIBILE** farlo, proprietà "cancella prima di scrivere".

## 12.6 Piani

Un gran numero di blocchi ( $> 1000$ ), uno sopra l'altro, condividono la bitline e anche altri circuiti per la scrittura e lettura, in un **piano**.

Un possibile elemento condiviso potrebbe essere il **decoder** il quale seleziona il blocco opportuno. Per far sì che il blocco venga selezionato si devono aggiungere due pass transistor, i quali verranno entrambi attivati, per far attaccare la stringa alla bitline.

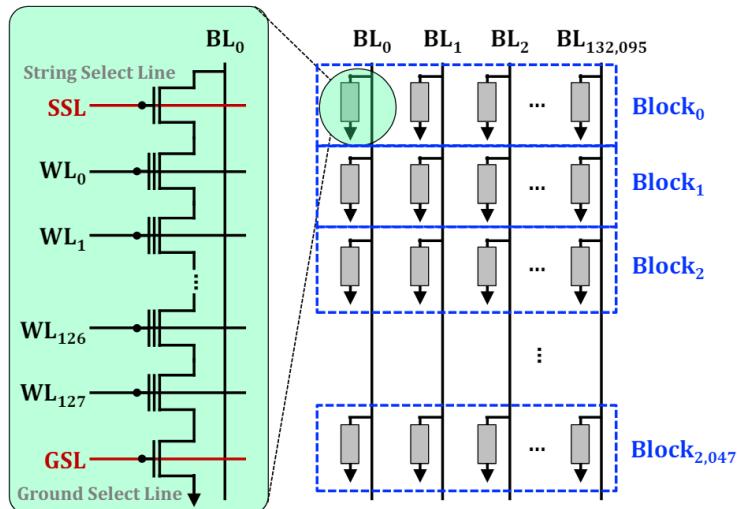


Figura 12.8: Piano con 2048 righe

## 12.7 Dies

Un Die o un chip contengono molteplici piani (2 - 4).

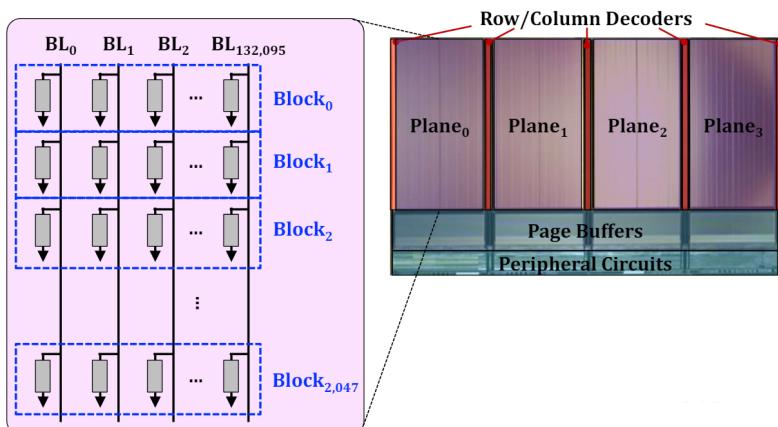


Figura 12.9: 21-nm 2D NAND Flash Die

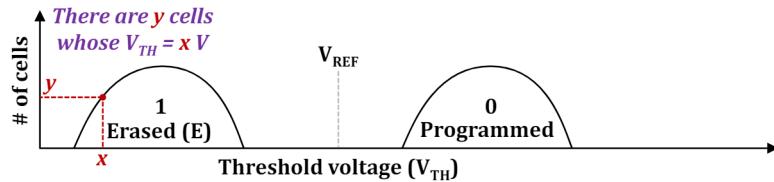
I piani condividono i decoders dunque vi sono dei limiti interni per il parallelismo, una sola operazione allo stesso istante di tempo, a dire il vero si possono fare delle operazioni in parallelo su piani diversi, però sulla stessa riga, utile per della ridondanza.

## 12.8 Distribuzioni dei livelli logici per una memoria NAND

Ogni transistore non ha un valore specifico per quando il dato è programmato o cancellato, dunque i livelli compressivi di tutti i transistori seguono l'andamento di una gaussiana.

Tra le sue gaussiane vi deve essere abbastanza spazio in mezzo per discriminare, un po' come il margine di rumore, se si incontrassero le due gaussiane non riusciremmo più a distinguere se la cella è programmata o meno.

Tutto questo discorso vale ancora di più se si pensa alle celle multi-bit.



### 12.8.1 Cella multi bit a 8 livelli

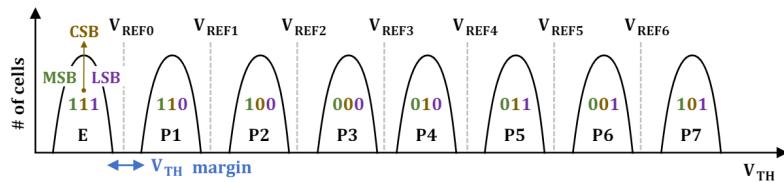
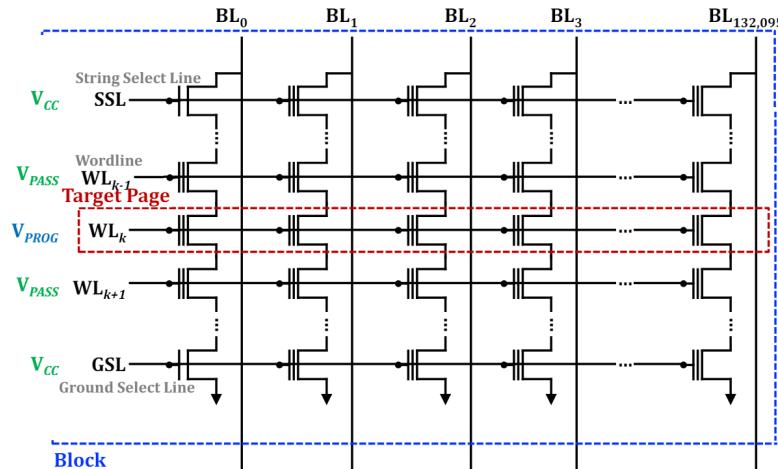


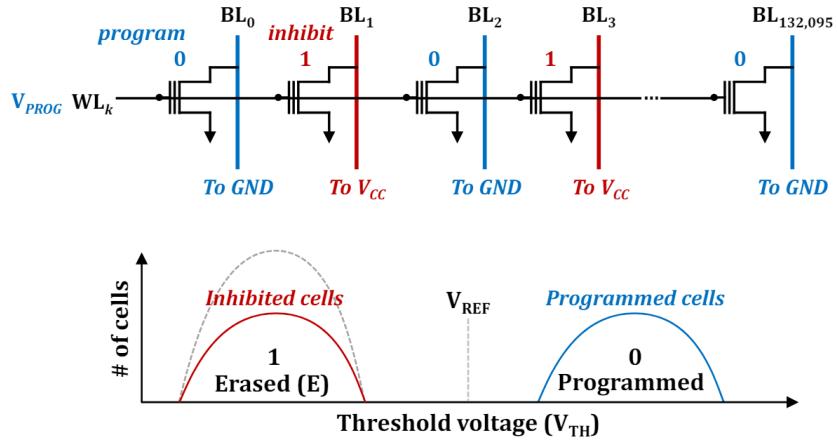
Figura 12.10: Triple Level Cell

## 12.9 Programmazione

Abbiamo detto che la programmazione avviene per pagine intere, quindi innanzitutto si deve mettere il livello di programmazione,  $V_{PROG}$ , 20 V, alla bitline. Tutti gli altri floating gate si deve mettere un livello tale da superare anche la soglia più alta,  $V_{PASS}$ ,  $V_{PASS} < V_{PROG}$ . I transitori di accesso basta attivarli,  $V_{DD} - V_{CC}$ .



Dunque i transistor che non devono essere scritti si comportano come una resistenza, mentre la pagina verrà tutta programmata. Noi non vogliamo però programmarla tutta con lo stesso valore, bensì con valori diversi: alcuni vogliamo programmarli altri no.

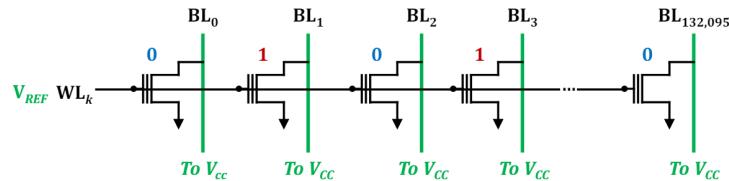


Per programmare basta agire sulla bitline, mettendola a GND tra il gate e il substrato c'è la tensione massima, se non si vuole programmare, basta portare alla stessa tensione del gate pure la bitline facendo si che la differenza di potenziale sia zero e dunque non programmiamo nulla.

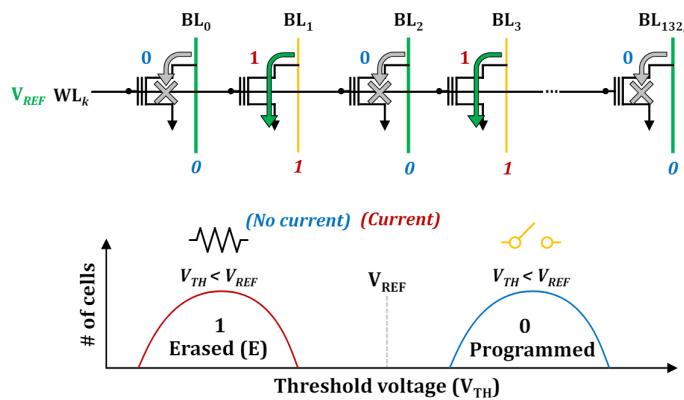
Agendo sulle bitline possiamo decidere che valore vogliamo dentro ogni singola cella di tutta la pagina selezionata con la wordline.

## 12.10 Lettura

Ora non mettiamo più la tensione di programmazione,  $V_{PROG}$ , bensì la tensione che sta a metà,  $V_{REF}$ . Quindi sopra la soglia normale  $V_T$  ma sotto  $V_{PROG}$ . Le bitline possiamo metterle alla tensione di alimentazione  $V_{CC}$  e andare a vedere su quale stringa sta passando corrente, altrimenti si possono precaricare e si può vedere quelle che si scaricano.



Qualunque metodo utilizziamo, riusciamo a leggere un intera pagina. Se passa corrente significa che il valore è uno, viceversa zero.



## 12.11 Conclusioni

Un SSD è un componente molto più complesso di quello che possa sembrare, monta un processore il quale deve tenere conto di varie problematiche quali:

- l'usura deve essere uniforme,
- deve gestire dove scrivere i dati,
- quando eliminare un blocco,
- spostare qualche blocco,
- supervisione della salute dell'SSD, S.M.A.R.T.,
- ecc.

Una ram statica (cache) a confronto è molto più semplice.