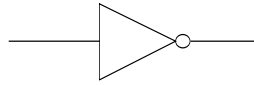


CIRCUITI DIGITALI  
MOS

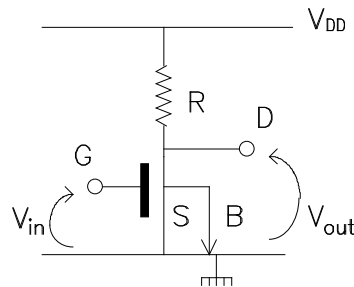
## PORTE LOGICHE

### Inverter

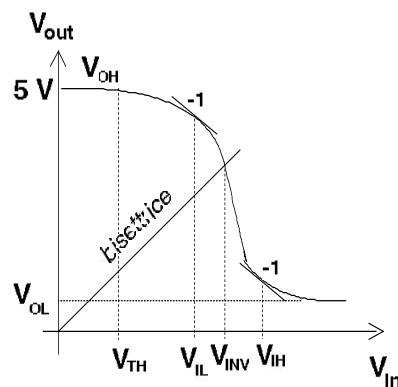
Lo schema logico dell' inverter è il seguente



mentre il corrispondente schema circuitale è



La caratteristica ingresso-uscita dell' inverter è :



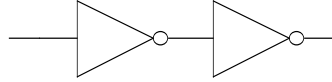
L'inverter è definito da cinque tensioni notevoli ovvero da:

- $V_{OH}$  che rappresenta il valore del livello alto della tensione di uscita corrispondente ad una tensione di ingresso sufficientemente bassa ;
- $V_{OL}$  cioè il livello basso della tensione di uscita che si ottiene quando  $V_{in}$  è sufficientemente elevata;
- $V_{IL}$  o tensione di ingresso a livello basso e  $V_{IH}$  o tensione di ingresso a livello alto che rappresentano i valori di tensione che hanno tangente  $= -1$  sulla caratteristica ingresso-uscita dell'inverter;
- $V_{INV}$  o tensione di inversione che corrisponde al punto di intersezione tra la caratteristica di ingresso-uscita dell' inverter e la bisettrice del primo quadrante.

Se  $V_{in} < V_{TH}$  nel transistor a canale n di tipo enhancement non scorre, con buona approssimazione, corrente e quindi la  $V_{out}$  è a livello alto. Essendo  $V_{TH} > 0$  il canale si crea richiamando elettroni per mezzo di una polarizzazione di gate sufficientemente potente. Supponendo il bulk cortocircuitato a massa se  $G=S$  allora  $V_{in} = V_{GS} = 0$  per cui il dispositivo lavora sotto soglia e quindi  $V_{out} = V_{DD}$  in quanto, poiché il transistor non conduce, la caduta ai capi della resistenza  $R$  è nulla.

Facendo crescere la tensione di ingresso  $V_{in}$  il transistorore rimane interdettato fino a che detta  $V_{in}$  non raggiunge il valore di soglia  $V_{TH}$ : a questo punto comincia a scorrere corrente nel dispositivo. Il passaggio di una corrente  $I_{DS}$  determina una caduta di tensione ai capi della resistenza  $R$  con conseguente abbassamento della tensione di uscita  $V_{out}$ .

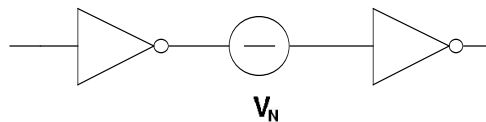
Si consideri ora lo schema di una rete più complessa del tipo



in cui le uscite del primo stadio sono gli ingressi di quello successivo.

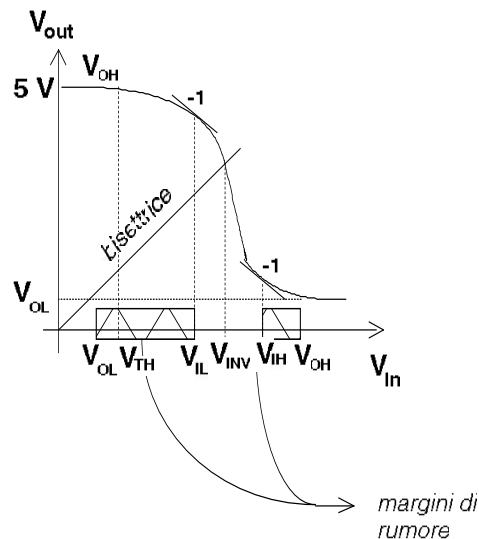
La tensione di uscita del circuito riportato è alta se l'uscita del primo inverter è pari a  $V_{OL}$  viceversa è bassa se l'uscita del primo inverter è pari a  $V_{OH}$ . Poichè le definizioni di  $V_{OH}$  e di  $V_{OL}$  dipendono l'una dall'altra è necessario aggiungere la condizione che impone  $V_{OL} < V_{OH}$ .

La linea che interconnette l'uscita del primo gate all'ingresso del secondo è soggetta a disturbi esterni che sono simulabili attraverso un generatore di tensione opportunamente dimensionato:



Si definiscono margini di rumore gli intervalli in cui la tensione di ingresso all'inverter  $V_{in}$  può variare senza pregiudicare il funzionamento della porta logica in particolare si parla di:

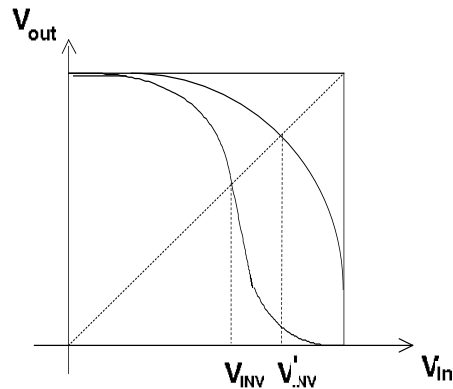
- margine di rumore a livello basso se  $V_{OL} < V_{in} < V_{IL}$ ;
- margine di rumore a livello alto se  $V_{IH} < V_{in} < V_{OH}$ .



Si possono definire due tensioni  $V_{NL}$  e  $V_{NH}$  cioè

$$\begin{cases} V_{NL} = V_{IL} - V_{OL} \\ V_{NH} = V_{OH} - V_{IH} \end{cases}$$

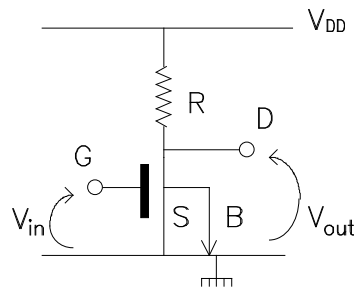
ed in particolare è utile avere margini di rumore equivalenti ( $V_{NL} = V_{NH}$ ). Per non avere margini di rumore troppo sbilanciati è importante non portare  $V_{inv}$  a valori troppo elevati oppure troppo bassi e questo dipende dalla forma della caratteristica ingresso-uscita cioè:



Analizzando gli andamenti appena riportati si vede che la caratteristica di uscita meno panciuta è quella più adatta perché  $V_{inv}$  cade a metà del range di  $V_{out}$  ( $V'_{INV}$  è invece troppo sbilanciata verso l'alto).

#### ESEMPIO

Lo scopo è quello di caratterizzare il circuito:



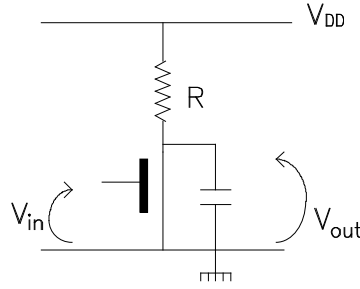
Per trovare le espressioni di  $V_{INV}$ ,  $V_{IH}$ ,  $V_{IL}$ ,  $V_{OL}$  e  $V_{OH}$  è necessario conoscere le equazioni del circuito e cioè

$$- V_{TH} = V_{FB} + 2\Phi_F + \gamma\sqrt{2\Phi_F - V_{BS}};$$

$$I_D = \mu C_{OX} \frac{W}{L} \left( V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS} \text{ se il dispositivo funziona in zona lineare;}$$

$$- I_D = \mu C_{OX} \frac{W}{2L} (V_{GS} - V_{TH})^2 \text{ se il dispositivo funziona in zona di saturazione.}$$

Nei circuiti integrati di tipo MOS il carico è sempre di tipo capacitivo ovvero



Si passa ora al dimensionamento statico dell'inverter che porta alla formulazione delle seguenti equazioni:

$$\begin{cases} V_{DS} + V_R = V_{DD} & (1) \\ I_R = I_{DS} & (2) \end{cases}$$

Si è già visto che se la tensione di ingresso  $V_{in} = 0$  si ha che  $V_{OH} = V_{DD}$  infatti il gate coincide con il source. Si vuole ora ricavare l'equazione della tensione di uscita a livello basso  $V_{OL}$  e per fare ciò è necessario stabilire se il dispositivo lavora in zona lineare od in zona di saturazione.

Per capire qual'è la zona di funzionamento del dispositivo bisogna valutare se  $V_{DS} > V_{DS_{SAT}} = V_{GS} - V_{TH} = (5-1)V = 4V$  essendo  $V_{TH} = 1V$  e  $V_{GS} = 5V$ , poiché si suppone che l'ingresso dell'inverter sia l'uscita  $V_{OH}$  di un inverter posto allo stadio precedente.

Poiché  $V_{DS} = V_{OL} \cong 200 \text{ mV}$ , infatti  $V_{OL} < V_{TH}$ , il dispositivo funziona in zona lineare allora si sceglie la relazione  $I_D = \mu C_{OX} \frac{W}{L} \left( V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS}$

Per la (1) si ha che  $V_{DS} = V_{OL} = V_{DD} - V_R$  essendo inoltre  $I_R = I_{DS}$  si ha che

$$\frac{V_{DD} - V_{OL}}{R} = \mu C_{OX} \frac{W}{L} (V_{DD} - V_{TH}) V_{OL}$$

cioè

$$V_{OL} = \frac{V_{DD}}{1 + \mu C_{OX} \frac{W}{L} R (V_{DD} - V_{TH})}$$

Per ricavare  $V_{IL}$  bisogna nuovamente andare a valutare la zona di funzionamento:

$V_{DS} = 4V$  e  $V_{DS_{SAT}} = V_{GS} - V_{TH} \cong 0.5V \Rightarrow V_{DS} > V_{DS_{SAT}}$  e quindi il dispositivo lavora in zona di saturazione per cui  $I_D = \mu C_{OX} \frac{W}{2L} (V_{GS} - V_{TH})^2$ .

Poiché  $I_D = I_R = \frac{V_{DD} - V_{out}}{R}$  si ha che  $\mu C_{OX} \frac{W}{2L} (V_{GS} - V_{TH})^2 = \frac{V_{DD} - V_{out}}{R}$  e derivando quest'ultima relazione rispetto a  $V_{in} = V_{GS}$  si ottiene  $2\mu C_{OX} \frac{W}{2L} (V_{in} - V_{TH}) = -\frac{\partial V_{out}}{\partial V_{in}} \frac{1}{R}$ .

Sostituendo le relazioni  $V_{in} = V_{IL}$  e  $\frac{\partial V_{out}}{\partial V_{in}} = -1$  si ottiene che  $\mu C_{OX} \frac{W}{L} (V_{IL} - V_{TH}) = \frac{1}{R}$  e quindi

$$V_{IL} = V_{TH} + \frac{1}{\mu C_{OX} \frac{W}{L} R}$$

Per trovare una espressione per  $V_{IH}$  si effettuano gli stessi passaggi di quelli fatti per  $V_{IL}$  con unica differenza che si lavora in zona lineare e quindi  $I_D = \mu C_{OX} \frac{W}{L} \left( V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS}$ .

Per calcolare la tensione di inversione  $V_{inv}$  deve valere la condizione  $V_{in} = V_{out} \Rightarrow V_G = V_D$  da cui si ricava che il dispositivo lavora in zona di saturazione essendo verificata la condizione di saturazione per cui  $V_{DS} > V_{DS_{SAT}} = V_{GS} - V_{TH}$  dove  $V_{TH} = 1$ .

Dal circuito in esame si ottiene che  $\mu C_{OX} \frac{W}{2L} (V_{GS} - V_{TH})^2 = \frac{V_{DD} - V_{out}}{R}$ , poiché siamo nel caso in cui  $V_{GS} = V_{out} = V_{INV}$ , la relazione può essere riscritta come  $\mu C_{OX} \frac{W}{2L} (V_{INV} - V_{TH})^2 = \frac{V_{DD} - V_{INV}}{R}$  che è un'equazione di secondo grado da cui si ricava che

$$V_{INV} = \frac{R \beta V_{TH} - 1 + \sqrt{1 + 2 R \beta (V_{DD} - V_{TH})}}{\beta R}$$

essendo  $\beta = \mu C_{OX} \frac{W}{L}$ .

#### ESEMPIO

Se il circuito in esame è caratterizzato dai seguenti dati

$$\begin{cases} \mu C_{OX} = 20 \frac{\mu A}{V^2} \\ \frac{W}{L} = 2 \\ R = 100 \text{ k}\Omega \\ V_{DD} = 5 \text{ V} \\ V_{TH} = 1 \text{ V} \end{cases}$$

si ottengono le tensioni notevoli:

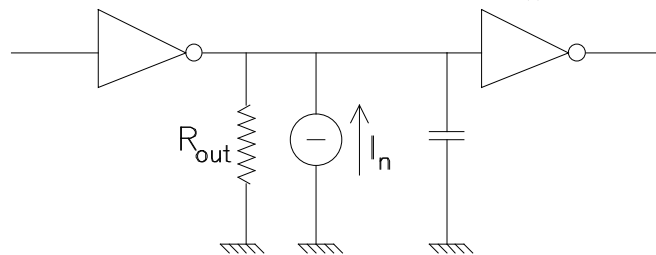
$$V_{OH} = 5 \text{ V}, V_{OL} = 0.3 \text{ V}, V_{INV} = 2.2 \text{ V}, V_{IH} = 3 \text{ V}, V_{IL} = 1.25 \text{ V}.$$

Si può notare che essendo  $V_{INV} = 2.2 \text{ V}$  non c'è simmetria (quest'ultima si avrebbe se  $V_{INV} = 2.5 \text{ V}$ ) ed in particolare il margine di rumore a livello alto è più elevato.

La tensione dovuta al rumore è legata all'ingresso capacitivo dell'ultimo inverter ed alla resistenza di uscita del primo gate, tale tensione nello schema riportato a lato, è modellata attraverso un generatore di corrente  $I_n$ .

Se la tensione di ingresso  $V_{in}$  è bassa allora  $R_{out} = R = 100 \text{ k}\Omega$ .

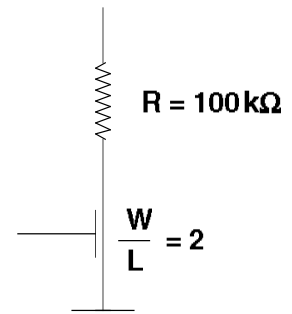
Se la tensione di ingresso  $V_{in}$  è alta la tensione di uscita del primo inverter è  $V_{OL}$  ed in particolare vale l'equazione  $V_{DD} \frac{R_m}{R_m + R} = V_{out}$  in cui  $R_m$  rappresenta la resistenza del transistore  $\Rightarrow$



$\frac{R_m}{R_m + R} = \frac{V_{OL}}{V_{DD}} = \frac{0.3 \text{ V}}{5 \text{ V}} = 0.06$  da cui si vede che se l'uscita del primo inverter è a livello basso si ha una resistenza di uscita pari a  $R_m \cong 6 \text{ k}\Omega$ .

La tensione in ingresso al secondo inverter dovuta al rumore è direttamente proporzionale alla resistenza di uscita dello stadio di ingresso e quindi è più consistente quando in uscita dal primo gate si ha  $V_{OH}$ : per questo motivo il margine di rumore a livello alto è più elevato.

Nello schema circuitale a lato si è fissato sia il valore della resistenza  $R$  a  $100 \text{ k}\Omega$  sia il rapporto  $\frac{W}{L}$ , relativo alla larghezza ed alla lunghezza del transistor, a 2.

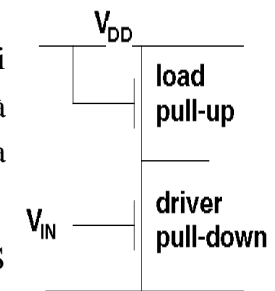


Se la tecnologia utilizzata per realizzare il circuito prevede come limite inferiore alla lunghezza dei transistori una  $L_{min} = 2 \mu\text{m}$  scegliendo come  $L$  del dispositivo detto valore minimo si ha, per il vincolo  $\frac{W}{L} = 2$ , che  $W = 4 \mu\text{m}$ .

Per realizzare sul layout la resistenza si sfrutta la relazione  $R = \frac{\mathcal{R}}{\square} \frac{W}{L}$  in cui la quantità  $\frac{\mathcal{R}}{\square}$  rappresenta gli "ohm al quadro" ed è specifica per ciascun tipo di pista, cioè:

- polisilicio  $40 \div 100 \frac{\mathcal{R}}{\square}$ ;
- diffusione  $20 \div 50 \frac{\mathcal{R}}{\square}$ ;
- metal  $0.0 \dots \frac{\mathcal{R}}{\square}$ .

In particolare una  $R = 100 \text{ k}\Omega$  si realizza disponendo in fila 1000 quadratini di polisilicio di  $2 \mu\text{m} \times 2 \mu\text{m}$  (supponendo che il polisilicio abbia una resistività di  $100 \frac{\mathcal{R}}{\square}$ ) e quindi sul layout detta resistenza è rappresentata da una striscia di poly di  $2 \mu\text{m} \times 2 \text{ mm}$  !



La realizzazione degli elementi passivi attraverso la tecnologia CMOS provoca una notevole occupazione di area sul chip (pari circa al 99% della superficie) per questo si preferisce sostituire la resistenza con un transistor ed un possibile schema circuitale è riportato a lato. Il transistor di pull-up ha il gate cortocircuitato a  $V_{DD}$  per permettere che il canale si formi sempre. Per caratterizzare il circuito è necessario calcolare le tensioni notevoli  $V_{OL}$ ,  $V_{OH}$ ,  $V_{IH}$ ,  $V_{IL}$  e  $V_{INV}$ .

### Calcolo di $V_{OH}$ .

Il circuito è esprimibile dal seguente sistema:

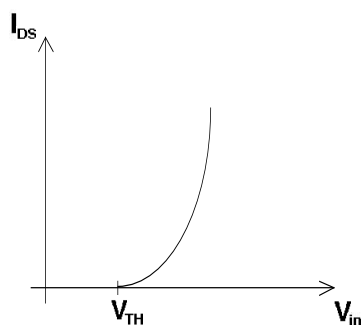
$$\begin{cases} V_{DS_{pd}} + V_{DS_{pu}} = V_{DD} \\ I_{DS_{pd}} = I_{DS_{pu}} \end{cases}$$

Bisogna stabilire la zona di funzionamento del transistor di pull-up ovvero ci si chiede se

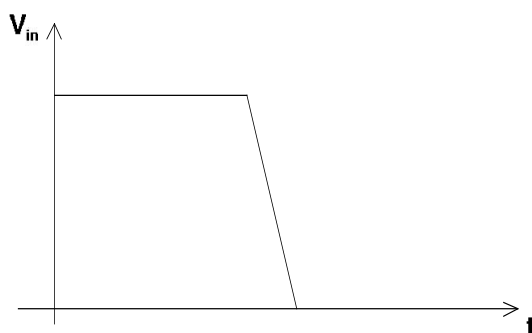
$$V_{DS_{pu}} > V_{DS_{SAT}} = V_{GS_{pu}} - V_{TH_{pu}} \cong V_{GS_{pu}} - 1 \text{ V}.$$

Poiché la disuguaglianza è verificata il transistor di pull-up lavora, se non è interdetto, in zona di saturazione ed allora la corrente che passa attraverso il dispositivo è rappresentata dalla relazione  $I_D = \mu C_{ox} \frac{W}{2L} (V_{GS} - V_{TH})^2$ . Quando la tensione di uscita è alta la corrente deve essere nulla e questo avviene, come si può facilmente rilevare dall'espressione di  $I_{DS}$ , quando  $V_{GS} = V_{TH} = V_{DS}$  per cui  $V_{OH} = V_{DD} - V_{DS} = V_{DD} - V_{TH_{pu}}$ .

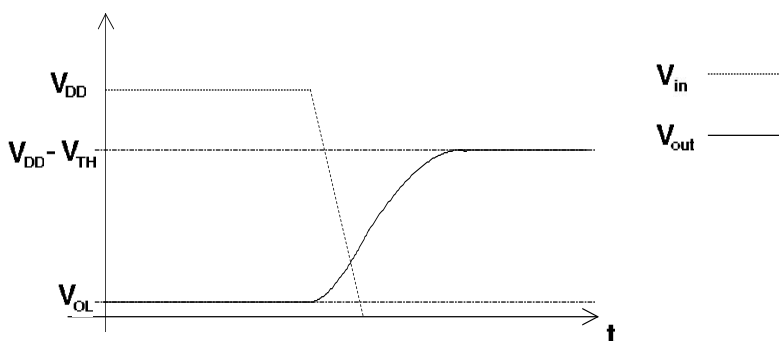
La corrente  $I_{DS}$  decresce se la tensione di ingresso  $V_{in}$  diminuisce infatti l'andamento di  $I_{DS} (V_{in})$  è il seguente:



Supponendo di avere una tensione di ingresso  $V_{in}$  con il seguente andamento nel tempo



il corrispondente andamento della tensione di uscita dell'inverter è



L'andamento del transitorio di  $V_{out}$  è dovuto al carico capacitivo si può inoltre sottolineare che non tutta l'escursione dell'alimentazione  $[0 \div V_{DD}]$  è disponibile in uscita infatti  $V_{out} \in [0 \div V_{DD} - V_{TH_{pu}}]$  e  $V_{INV}$  cade proprio al centro di detto range; questo circuito funziona peggio di quello con la resistenza  $R$  essendo i margini di rumore più stretti.



### Calcolo di $V_{IL}$

Per calcolare il valore della tensione  $V_{IL}$  è necessario uguagliare le correnti del transistore di pull-up e di pull-down, per far ciò bisogna determinare la zona di funzionamento dei due dispositivi.

Si è visto che il *load* (transistore di pull-up) può essere in interdizione, nel caso non scorra corrente  $I_{DS}$ , oppure in saturazione. Nel caso specifico, cioè quando è applicata una  $V_{in} = V_{IL}$ , si trova nella seconda zona di funzionamento mentre per il *driver* (transistore di pull-down) è necessario andare a valutare la disuguaglianza:

$$V_{DS_{pd}} > V_{DS_{SAT}} = V_{GS_{pd}} - V_{TH_{pd}}$$

Il dispositivo di pull-down lavora in zona di saturazione infatti la disuguaglianza sopra riportata è verificata in quanto  $V_{DS}$  è elevata, essendo l'uscita dell'inverter a livello alto, mentre  $V_{GS_{pd}} - V_{TH_{pd}}$  è una tensione bassa in quanto  $V_{GS_{pd}} = V_{IL}$ .

Definiti i valori  $\beta_d$  e  $\beta_l$  come  $\beta_d = \mu C_{OX} \frac{W_{driver}}{L_{driver}}$  e  $\beta_l = \mu C_{OX} \frac{W_{load}}{L_{load}}$  si ha che:

- $I_{DS_{driver}} = \frac{\beta_d}{2} (V_{in} - V_{TH_d})^2$  essendo  $V_{GS_{driver}} = V_{in}$ ;
- $I_{DS_{load}} = \frac{\beta_l}{2} (V_{DD} - V_{out} - V_{TH_l})^2$  essendo  $V_{GS_{load}} = V_{DS_{load}} = V_{DD} - V_{out}$ .

Uguagliando  $I_{DS_{load}}$  e  $I_{DS_{driver}}$  si ha che

$$\frac{\beta_d}{2} (V_{in} - V_{TH_d})^2 = \frac{\beta_l}{2} (V_{DD} - V_{out} - V_{TH_l})^2$$

definito  $k_r = \frac{\beta_d}{\beta_l}$  si ha

$$k_r (V_{in} - V_{TH_d})^2 = (V_{DD} - V_{out} - V_{TH_l})^2 \quad (*)$$

La tensione  $V_{IL}$  è per definizione tale che  $\left. \frac{\partial V_{out}}{\partial V_{in}} \right|_{V_{in}=V_{IL}} = -1$  perciò derivando rispetto alla tensione di

ingresso l'equazione (\*) si ottiene

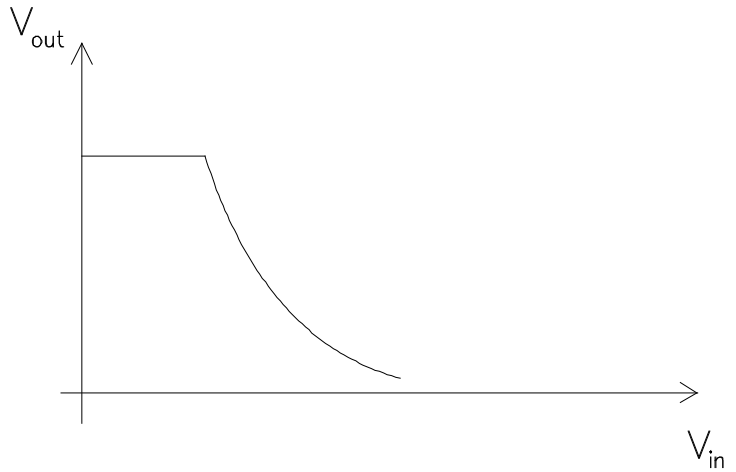
$$k_r (V_{in} - V_{TH_d}) = - (V_{DD} - V_{out} - V_{TH_l}) \frac{\partial V_{out}}{\partial V_{in}}$$

ma essendo  $\left. \frac{\partial V_{out}}{\partial V_{in}} \right|_{V_{in}=V_{IL}} = -1$  si conclude che

$$k_r (V_{IL} - V_{TH_d}) = (V_{DD} - V_{out} - V_{TH_l})$$

Se si ricava  $V_{IL}$  dall'equazione  $k_r (V_{IL} - V_{TH_d}) = (V_{DD} - V_{out} - V_{TH_l})$  e detto valore si sostituisce nell'equazione (\*) si ottiene un assurdo a meno che  $k_r = 1$ .

Se  $k_r \neq 1$  la caratteristica ingresso-uscita dell'inverter ha un andamento simile a quello riportato a fianco cioè è presente uno spigolo vivo perché la pendenza passa bruscamente da 0 a  $-\sqrt{k_r}$  e  $V_{IL} = V_{TO} = V_{TH_d}$ .



Si ricorda che l'equazione che determina la tensione di soglia è  $V_{TH} = V_{FB} + 2 \Phi_F + \gamma \sqrt{2 \Phi_F - V_{BS}}$  e che il substrato driver è collegato a massa, trattandosi di un dispositivo a canale n, per cui  $V_{BS_d} = 0$  e di conseguenza si verifica che  $V_{TH_d} = V_{TO}$ .

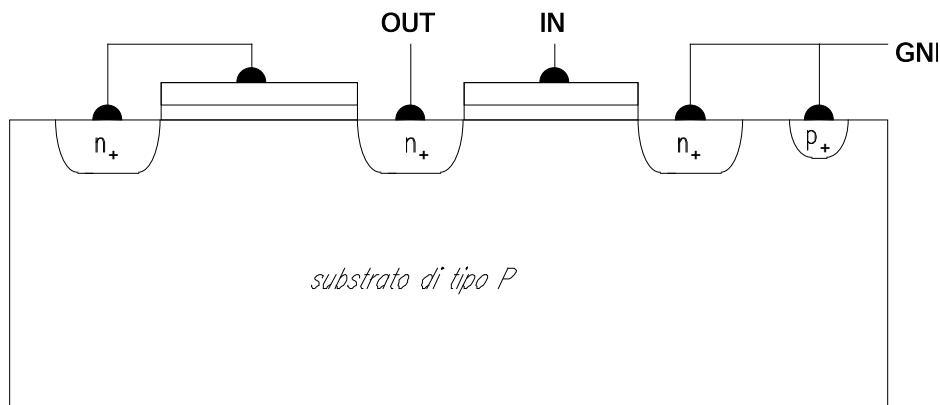
Per ricavare la  $V_{INV}$  si uguaglia la tensione di ingresso  $V_{in}$  alla tensione di uscita  $V_{out}$ .

La tensione  $V_{TO}$  è un parametro introdotto da SPICE e corrisponde alla quantità  $V_{TO} = V_{FB} + \gamma \sqrt{2 \Phi_F} + 2 \Phi_F$ .

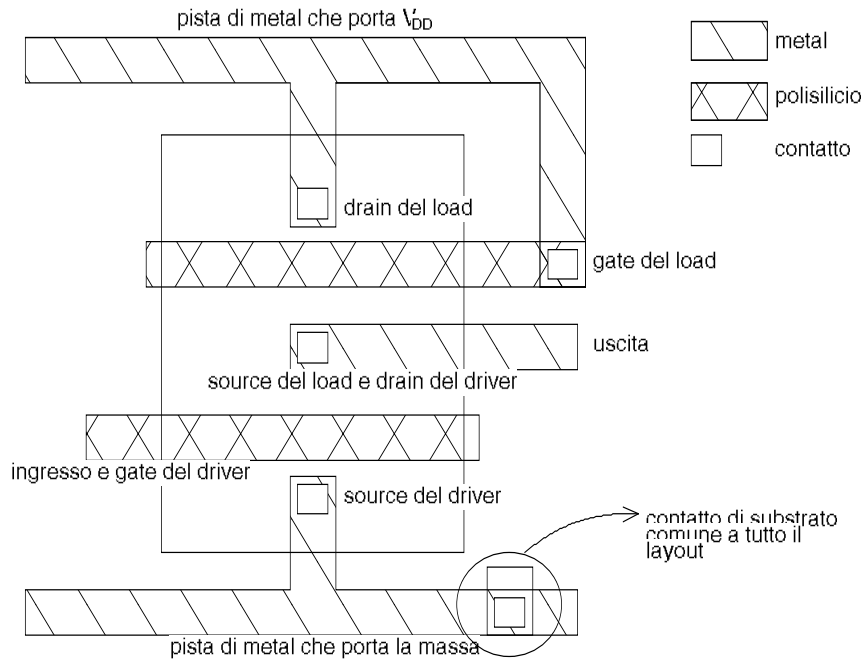
Se  $V_{out} = V_{OH}$  si ha che  $V_{BS_{load}} = -V_{OH}$ , ma  $V_{OH}$  dipende  $V_{TH}$  infatti  $V_{OH} = V_{DD} - V_{TH}$ , per cui l'espressione della tensione di soglia diventa  $V_{TH} = V_{FB} + 2 \Phi_F + \gamma \sqrt{2 \Phi_F + V_{DD} - V_{TH}}$ .

Si nota che se la tensione di soglia cresce la tensione di uscita  $V_{OH}$  si abbassa e la tensione  $V_{BS}$  varia con conseguente modifica di  $V_{TH}$ : da queste considerazioni si conclude che  $V_{TH}$  va calcolata in modo ricorsivo.

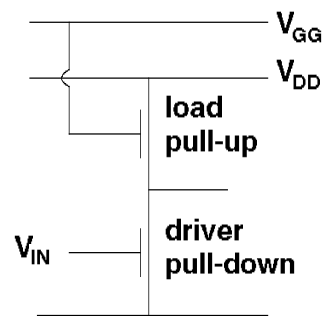
Una sezione verticale del wafer di silicio su cui è realizzato lo schema circuitale analizzato è:



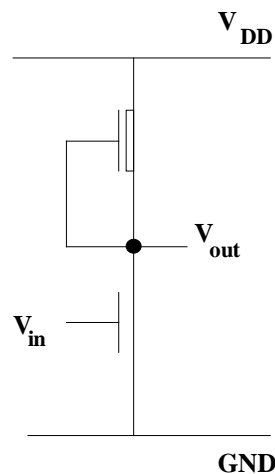
Il corrispondente layout è:



Per concludere l'analisi degli inverter realizzati con *load* di tipo enhancement si riporta a lato una struttura a doppia alimentazione, in cui il gate del carico è collegato ad una tensione  $V_{GG}$ . Se in questo schema la tensione  $V_{GG}$  è fissata ad un valore pari a  $V_{DD} + V_{TH_{pull\ up}}$  si riesce ad ottenere un range di variazione della tensione di uscita pari a  $[0..V_{DD}]$  ovvero si ristabilisce il funzionamento del circuito con carico resistivo.



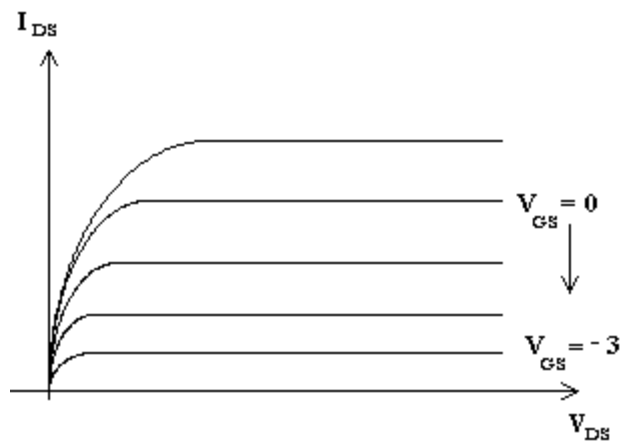
Una soluzione alternativa per realizzare l'inverter è quella di utilizzare come *load* un transistor di tipo depletion, cioè un dispositivo con canale già presente e con tensione di soglia  $V_{TH} < 0$ , il circuito corrispondente è il seguente:



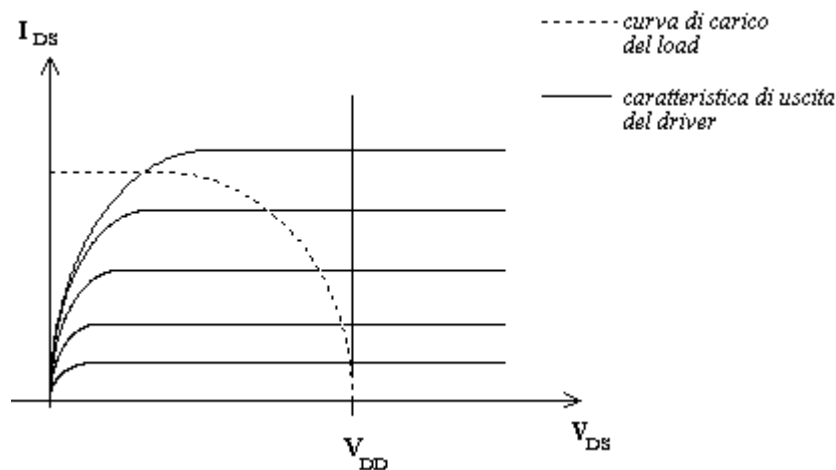
Essendo  $V_{GS_{depletion}} = 0$ , poiché gate e source sono cortocircuitati, e supponendo che  $V_{TH_{depletion}} = -3$  V si ha che  $V_{GS_{depletion}} - V_{TH_{depletion}} = 3$  V e quindi il transistor conduce sempre.

Con questo nuovo schema  $V_{OH} = V_{DD}$  cioè il range di uscita dell'inverter  $\in [0 \div V_{DD}]$  ovvero si è ripristinato, senza dover ricorrere ad una doppia alimentazione, il funzionamento dell'inverter con resistenza di carico R.

La caratteristica  $I_{DS}(V_{DS})$  del transistor depletion (*load*) per  $V_{BS} = 0$  è la seguente:

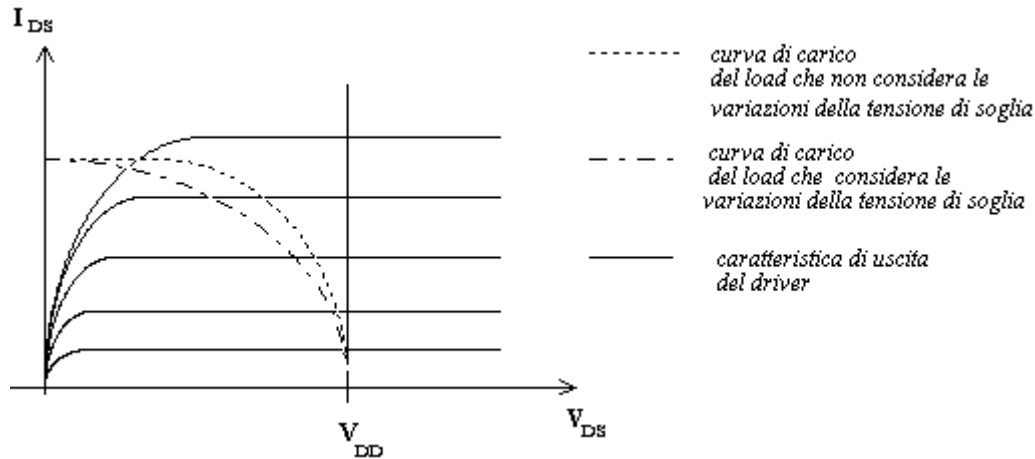


Per valutare il punto di lavoro del circuito in esame si riporta, opportunamente ribaltata, la curva  $I_{DS}(V_{DS})$  per  $V_{BS} = 0$  e  $V_{GS} = 0$  sulla caratteristica  $I_{DS}(V_{DS})$  del *driver*:

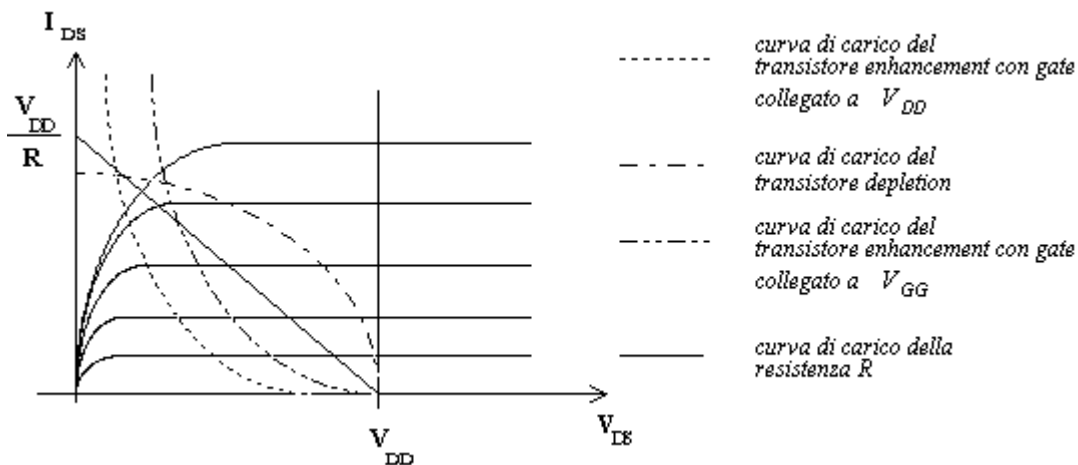


L'andamento della curva di carico in figura non è corretto perché è valutato per  $V_{BS} = \text{cost}$  mentre nel caso dell'inverter detta tensione varia essendo  $V_{BS_{load}} = -V_{OH}$ ; il cambiamento di  $V_{BS}$  provoca inoltre l'aggiornamento di  $V_{TH}$  essendo  $V_{TH} = V_{FB} + 2 \Phi_F + \gamma \sqrt{2 \Phi_F - V_{BS}}$ .

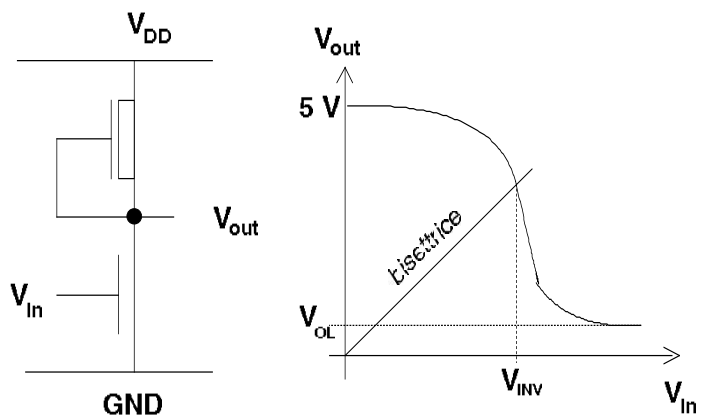
Il grafico con curva di carico più realistica è il seguente:



Può essere interessante confrontare gli andamenti delle curve di carico al variare del *load* ovvero a seconda che questo sia una resistenza, un transistor enhancement oppure un transistor depletion:



Dato lo schema dell'inverter a lato si ha che il *driver* è in saturazione quando  $V_{in} = V_{INV}$ , essendo  $V_{INV} = m V_{DD}$  con  $m = 0.5$ , in quanto è verificata la disuguaglianza  $V_{DS} > V_{GS} - V_{TH}$ .



Per analizzare la zona di funzionamento del *load* si suppone che il transistor depletion sia modellabile come un dispositivo di tipo enhancement ma con tensione di soglia negativa. Poiché  $V_{DS} = V_{DD} - V_{out} = V_{DD} - 0.5 V_{DD} = 2.5 V$  essendo  $V_{in} = V_{out} = V_{INV} = 0.5 V_{DD}$ , supponendo inoltre che  $V_{TH_{depletion}} = -0.8 V_{DD} = -4 V$ , la disuguaglianza

$V_{DS_{depletion}} > V_{DS_{SAT}} = V_{GS_{depletion}} - V_{TH_{depletion}}$  non è verificata, in quanto  $2.5 V > 4 V$ , perciò il dispositivo depletion lavora in zona lineare.

Dall'uguaglianza delle correnti del *driver* e del *load* si ha che

$$\frac{\beta_d}{2} (mV_{DD} - V_{TH_d})^2 = \beta_l \left( 0 - (-0.8 V_{DD}) + \frac{(1-m)V_{DD}}{2} \right) (1-m)V_{DD}$$

essendo

- $V_{GS_d} = m V_{DD}$ ;
- $V_{GS_l} = 0$ ;
- $V_{TH_d} = 0.2 V_{DD}$ ;
- $V_{TH_l} = 0.8 V_{DD}$ ;
- $V_{DS_l} = (1-m)V_{DD}$ ;
- $\beta_d = \mu C_{OX} \left( \frac{W}{L} \right)_d$ ;
- $\beta_l = \mu C_{OX} \left( \frac{W}{L} \right)_l$ .

Ricordando che  $k_r = \frac{\beta_d}{\beta_l} = \left( \frac{W}{L} \right)_d / \left( \frac{W}{L} \right)_l$  risulta che

$$\frac{k_r}{2} (m - 0.2)^2 V_{DD}^2 = \left( 0.8 + \frac{(1-m)}{2} \right) (1-m)V_{DD}^2$$

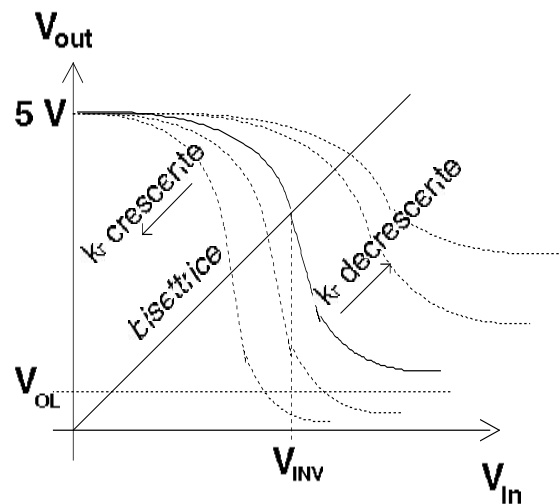
per cui, utilizzando quest'ultima relazione, si può esprimere  $k_r$  in funzione di  $m$  cioè

$$k_r = \frac{(1.6 - 1 + m)(1 - m)}{(m - 0.2)^2}$$

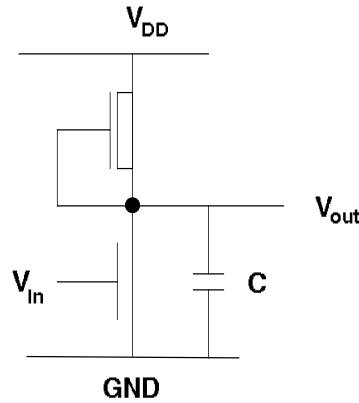
Per ottenere un funzionamento ragionevole del circuito (per avere cioè una  $V_{OL} < V_{TH}$ ) è necessario che  $k_r > 1$ , si vuole infatti che il partitore caratterizzato dalle resistenze dei due dispositivi sia sbilanciato verso massa ovvero che  $\left( \frac{W}{L} \right)_d > \left( \frac{W}{L} \right)_l$ .

L'aumento o la diminuzione del valore  $k_r$  provoca il cambiamento della caratteristica di uscita dell'inverter ed in particolare:

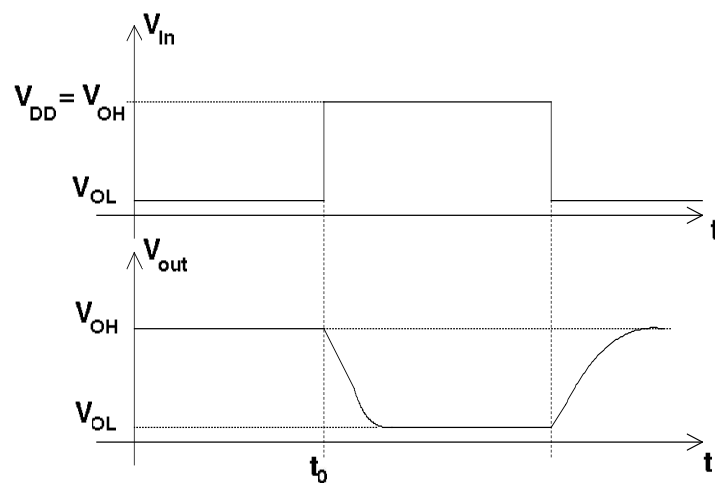
- se  $k_r$  cresce la caratteristica di uscita si abbassa ed il margine di rumore a livello basso diminuisce e  $V_{INV}$  si abbassa;
- se  $k_r$  decresce la caratteristica di uscita si alza.



Si aggiunge all'inverter un carico capacitivo:



L'andamento di  $V_{out}$  in funzione di  $V_{in}$  in questo nuovo schema è:



Quando la tensione di ingresso è alta la tensione di uscita è bassa ma cosa accade quando  $V_{in}$  passa da 0 a  $V_{DD}$  ?

In  $t_0^+$  il driver funziona in zona di saturazione, essendo verificata la disuguaglianza  $V_{DS} > V_{GS} - V_{TH}$  poichè  $5V > 5V - 1V$ . Se  $V_{GS} = V_{DS}$  il driver è in saturazione per cui in esso scorre una corrente pari a  $\frac{\beta_d}{2}(V_{GS} - V_{TH})^2$ . A causa della presenza del condensatore non vale più la caratteristica ingresso-

uscita fino ad ora analizzata perchè questa è valida se è verificata l'uguaglianza tra le correnti del *load* e del *driver*. Se  $k_r \gg 1$  si ha che  $I_{driver} \gg I_{load}$  ed  $I_C \approx I_{driver}$  per cui fino a che vale la relazione

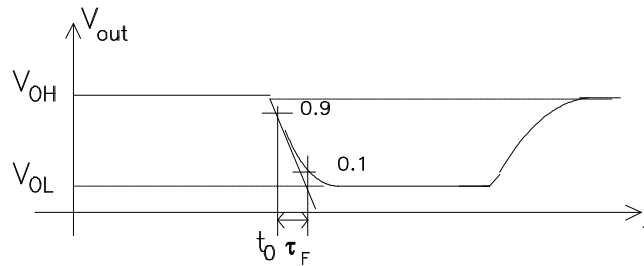
$I_{driver} = \frac{\beta_d}{2}(V_{GS} - V_{TH})^2$ , cioè fino a che il driver lavora in saturazione, il condensatore si scarica con

andamento lineare, essendo  $I_{driver}$  costante; quando  $V_C$  scende al di sotto di una certa soglia il driver non lavora più saturazione, bensì in zona lineare, quindi il condensatore termina di scaricarsi con un andamento circa esponenziale, essendo un transistor che lavora in zona lineare assimilabile ad una resistenza. L'andamento diventa un esponenziale perfetto quando  $V_{DS}$  è sufficientemente bassa cioè

quando il termine  $\frac{V_{DS}}{2}$  nell'espressione di  $I_d$  è trascurabile, essendo  $I_d = \beta_d \left( V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS}$ .

Per calcolare il tempo di transizione  $\tau_F$  (*fall time* o tempo di caduta), che valuta quanto impiega il

condensatore per passare da una tensione pari a 0.9 del suo range ad una pari a 0.1 di detto range, si traccia la sottotangente all  
a curva di scarica che in pratica è il prolungamento del tratto lineare:

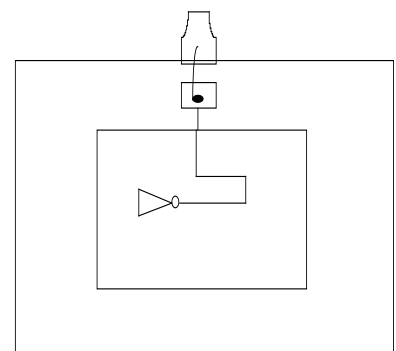


Il tempo di salita  $\tau_R$  (*rise time*) del circuito è esprimibile come  $\tau_R = k_r \tau_F$ ; per stabilire l'andamento della carica del condensatore è necessario conoscere la zona di funzionamento del *load*. Si vede che quest'ultimo lavora in saturazione, ma l'andamento della carica non è una rampa perchè, contrariamente al caso del *driver*, la corrente di saturazione del *load* non è costante a causa del *body effect*, cioè della variazione di  $V_{BS}$  in funzione della tensione di uscita infatti:

$$I_l = \frac{\beta_l}{2} (V_{GS} - V_{TH})^2 \neq \text{costante in quanto } V_{TH} = V_{FB} + 2 \Phi_F + \gamma \sqrt{2 \Phi_F - V_{BS}} \text{ è variabile.}$$

Dalla relazione  $\tau_R = k_r \tau_F$  si capisce perchè  $k_r$ , benchè debba essere maggiore di 1, non possa essere troppo elevato: un coefficiente  $k_r$  alto determinerebbe un tempo di salita  $\tau_R$  eccessivamente consistente. Poichè la velocità è un requisito fondamentale per un circuito solitamente si hanno  $k_r$  intorno a 5 oppure 6; nei circuiti in cui non è richiesta una elevata velocità  $k_r$  può aumentare a vantaggio dei consumi, che diminuiscono.

Ci si pone ora il problema di come connettere un inverter a dimensione minima con un *pad* del chip, cioè ci si chiede come si collega l'inverter all'ambiente esterno; se il collegamento avvenisse attraverso un filo il tempo di salita  $\tau_F$  e quello di discesa  $\tau_R$  aumenterebbero considerevolmente, essendo la capacità del circuito esterno molto più elevata di quella dell'inverter.



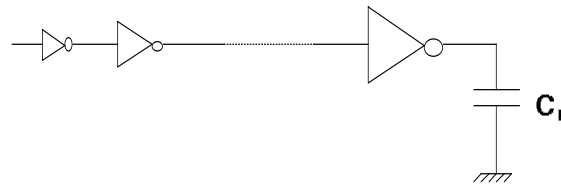
Si suppone di avere un inverter con  $L = 1.2 \mu\text{m}$  e  $W = 1.8 \mu\text{m}$ , ovvero a dimensioni minime rispetto alla tecnologia con cui è stato realizzato, la sua capacità di ingresso è  $C_g = C_{OX} W L$  dove  $W L$  è l'area del gate:  $C_g$  è parecchi ordini di grandezza inferiore alla capacità della struttura costituita da pad, pin, piste ed ingressi ad altri chip.

Per caratterizzare un processo, costituito per esempio da due inverter in cascata, è necessario conoscere sia le dimensioni minime sia il ritardo della porta, valutabile da  $\tau_F$  e  $\tau_R$  attraverso la relazione  $\tau = \frac{\tau_F + \tau_R}{2}$ . Normalmente i tempi tipici di ritardo per un processo sono circa di 0.5 nsec

ma nei processi più moderni raggiungono valori pari a 0.1, 0.2 nsec.



Essendo inverosimile collegare la porta logica con un filo si utilizza un sistema di inverter scalati cioè una serie di inverter in cascata che presentano, l'uno rispetto all'altro, dimensioni crescenti ma coefficiente  $k_r$  fisso:



Per costruire un insieme di inverter scalati con fattore di scala  $f = 2$  si procede in questo modo:

1. si realizza l'inverter a dimensioni minime, in cui il driver ha  $L$  e  $W$  dimensionati con i valori più bassi consentiti dalla tecnologia;
2. l'inverter successivo è realizzato con una  $W$  raddoppiata per il *driver*, ed una  $L$  dimezzata per il *load*, in questo modo il coefficiente  $k_r$  è mantenuto costante;
3. se l'insieme di inverter scalati non è completo si ritorna al punto 2.

### ESEMPIO

Si vuole realizzare un insieme di inverter scalati con  $k_r = 4$  e  $f = 2$  per cui si può effettuare la seguente scelta:

	driver	load
1° inverter	$W=1$	$W=1$
	$L=1$	$L=4$
2° inverter	$W=2$	$W=1$
	$L=1$	$L=2$
3° inverter	$W=4$	$W=1$
	$L=1$	$L=1$

Se il primo inverter ha una capacità  $C_g = C_{OX} W L$  allora il secondo inverter ha una capacità pari a  $f C_g$ , essendo la  $W$  del driver della seconda porta logica  $f$  volte più grande di quello della prima; per quanto riguarda i tempi di ritardo si ha che se il ritardo del primo inverter è pari a  $\tau$  quello del secondo è pari a  $f \tau$  (essendo ritardo =  $\frac{\text{capacità}}{\text{corrente}}$ ) in particolare dopo  $N$  stadi il ritardo totale è pari a

$$T = N f \tau.$$

Definita  $y = \frac{C_L}{C_g}$  si ha che  $C_L = f \cdot (\text{capacità } N\text{-esimo inverter})$  inoltre, essendo la capacità dell' $N$ -esimo inverter pari a  $f^{N-1} C_g$ , risulta che  $C_L = f^N C_g$  da cui si ha che

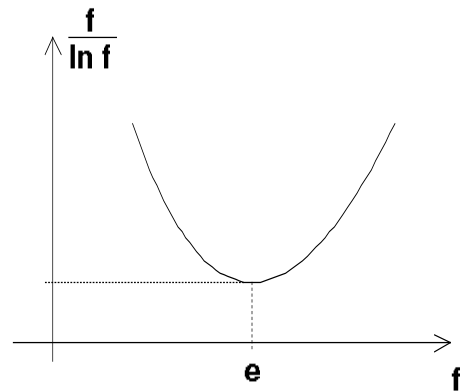
$$N = \log_f \frac{C_L}{C_g} = \frac{\ln \frac{C_L}{C_g}}{\ln f} = \frac{\ln y}{\ln f}$$

Sostituendo l'espressione di N nell'equazione relativa al ritardo totale si ottiene

$$T = \ln(y)\tau \frac{f}{\ln f}$$

in cui il valore  $\ln(y)\tau$  dipende dalle caratteristiche del processo.

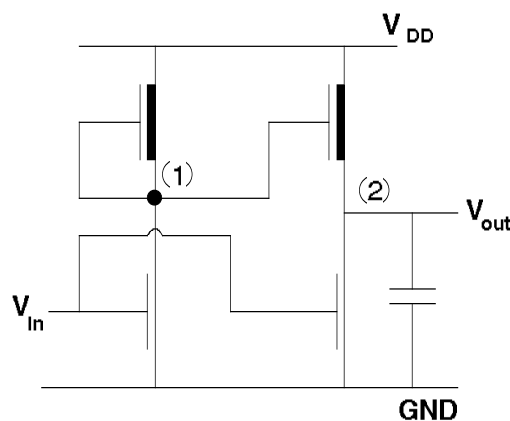
Per minimizzare il ritardo totale T è necessario minimizzare la quantità  $\frac{f}{\ln f}$  rispetto alla variabile f: valutando l'andamento



riportato a lato si vede che T è minimo se  $f = e = 2.718$ .

Quando il valore di f cresce il ritardo per singolo stadio cresce ma il numero totale di stadi N diminuisce mentre se f decresce il ritardo del singolo stadio diminuisce ed il numero totale degli stadi aumenta: per  $f = e$  il numero di stadi ed il ritardo sul singolo stadio sono bilanciati in modo tale da minimizzare la quantità  $\frac{f}{\ln f}$  e conseguentemente T.

Si consideri ora il circuito sottostante:



Lo schema sopra riportato è caratterizzato da un primo stadio costituito da un inverter ed un secondo stadio collegato in modo inusuale.

Applicando una tensione di ingresso  $V_{in} = 0$  i due *drivers* sono interdetti ed i nodi ① e ② sono ad un potenziale pari a  $V_{DD}$ ; se si impone, invece, una tensione  $V_{in} = V_{DD}$  i due *drivers* conducono e, se le dimensioni dei due dispositivi enhancement sono uguali, i nodi ① e ② sono allo stesso potenziale e quindi la struttura si comporta come un inverter perciò  $V_{out} = V_{OL}$ .

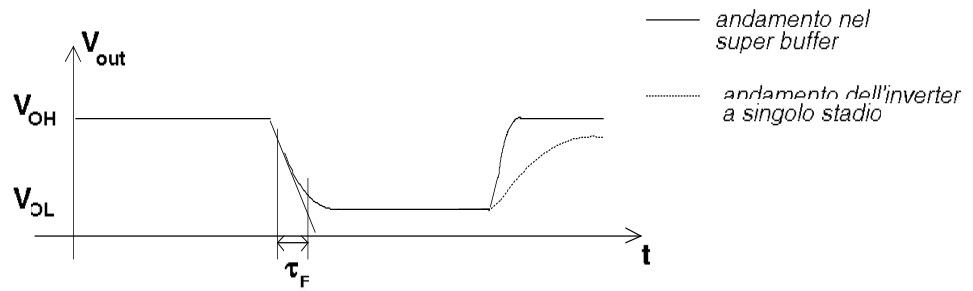
Il nuovo schema ha però un funzionamento dinamico diverso dagli inverter fino ad ora analizzati.

Il *load* del secondo stadio, che nello schema dell'inverter a singolo stadio con carico realizzato con un dispositivo depletion aveva  $V_{GS} = 0$ , ha ora  $V_{GS} = V_{DD} - V_{OL} \approx V_{DD}$  per cui la corrente di detto pull-up, che lavora in saturazione, è esprimibile come:

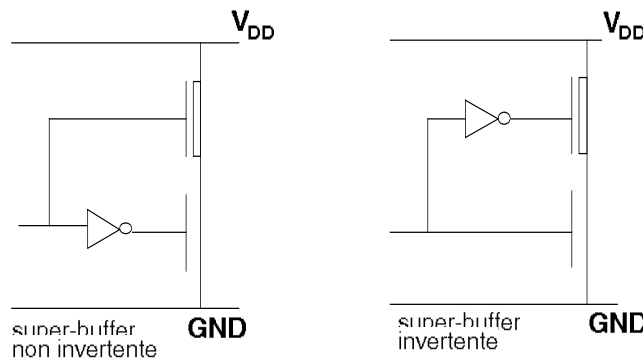
$$I_l = \frac{\beta_l}{2} (V_{GS} - V_{TH})^2 = \frac{\beta_l}{2} [V_{DD} - (-0.8)V_{DD}]^2 = \frac{\beta_l}{2} 3.24 V_{DD}^2$$

mentre nello schema dell'inverter a singolo stadio  $I_1 = \frac{\beta_1}{2} 0.64 V_{DD}^2$ .

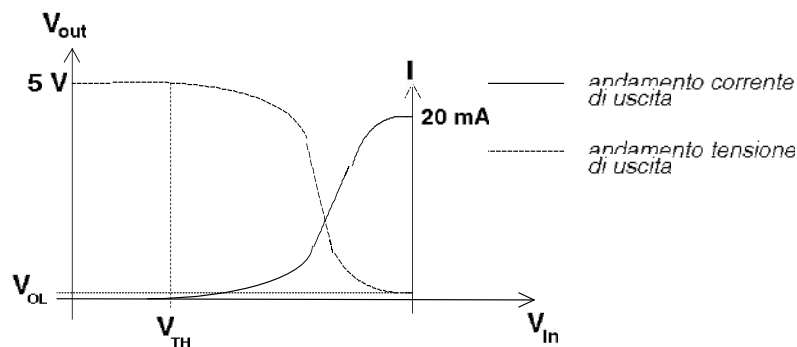
Poichè la corrente del load di quest'ultimo esempio è più elevata rispetto agli altri casi analizzati, si ha che il condensatore in uscita si carica più



velocemente con conseguente diminuzione del rise time  $\tau_R$ . In particolare se prima  $\tau_R = 4$  oppure 5 volte  $\tau_F$  ora  $\tau_R = \tau_F$ , il circuito è quindi notevolmente più veloce e per queste sue proprietà è definito **super-buffer**. Esistono strutture per realizzare super-buffers invertenti od, in alternativa, super-buffers non invertenti che, utilizzando una notazione mista, hanno la seguente configurazione:



Si affronta adesso il problema di come, data una certa logica interna, pilotare i pad del chip. Fino ad ora non si è analizzato l'andamento della corrente assorbita dal componente, riportando sulla caratteristica di ingresso-uscita anche detta corrente si ha:



Confrontando gli andamenti della tensione e della corrente di uscita della porta logica si vede che quando  $V_{out} = V_{OL}$  la corrente che scorre nel componente è piuttosto elevata.

Un buffer di uscita deve avere tempi di commutazione ragionevoli, dell'ordine dei ns, cioè durante il transitorio deve scorrere una corrente elevata:

se si vuole un  $\tau = 5$  ns sapendo che  $Q = I \tau = C V$ ,  $C = 100$  pF,  $V = 5$  V si ottiene che

$$I = \frac{C V}{\tau} = \frac{100 \cdot 10^{-12} \cdot 5 \text{ [F] [V]}}{5 \cdot 10^{-9} \text{ [sec]}} = 100 \text{ mA}$$

Utilizzando in uscita lo schema circuitale a lato con  $C = 100 \text{ pF}$ , per avere una corrente di scarica pari a  $100 \text{ mA}$  ed un  $\tau_F = 5 \text{ ns}$ , il tempo di salita si ricava dalla nota formula  $\tau_R = k_r \tau_F$  e, fissando  $k_r = 5$ , risulta  $\tau_R = 25 \text{ ns}$ .

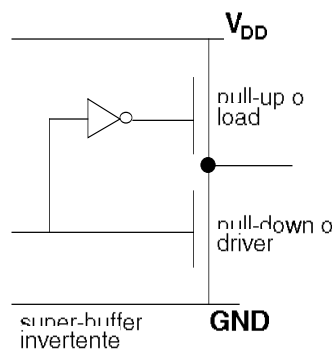
Poichè il *rise time* è 5 volte più lungo del *fall time*, si ricava che la corrente che scorre nel condensatore durante il processo di carica è pari ad  $\frac{1}{5}$  della corrente dello stesso durante il transitorio che lo scarica, cioè detta corrente è di  $20 \text{ mA}$ .

La corrente che, in condizioni statiche, passa nel ramo di uscita (si ricorda che il condensatore è, in condizioni statiche, aperto), quando la tensione  $V_{out}$  è a livello basso, risulta pari a  $20 \text{ mA}$  per cui la potenza che dissipa la porta logica analizzata è pari a:

$$P = V I = 20 \cdot 10^{-3} \cdot 5 \text{ [A] [V]} = 100 \text{ mW}$$

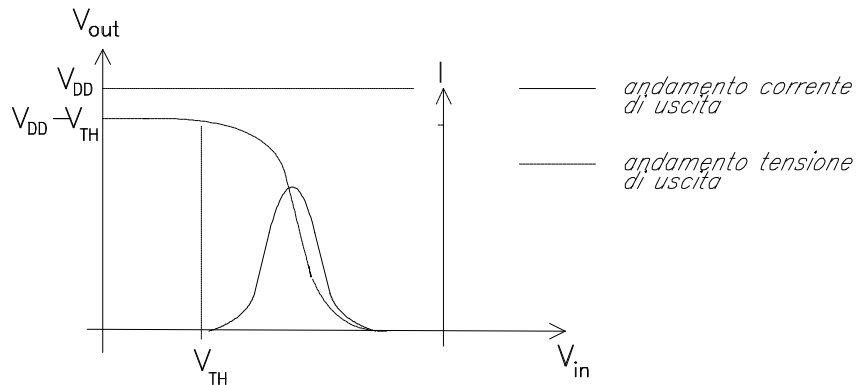
Considerando che tipicamente un chip ha una potenza di  $1 \text{ W}$  la soluzione appena proposta non può essere utilizzata a causa dell'elevata dissipazione.

Uno schema alternativo, in cui il load è realizzato con un transistore di tipo enhancement, è il seguente:

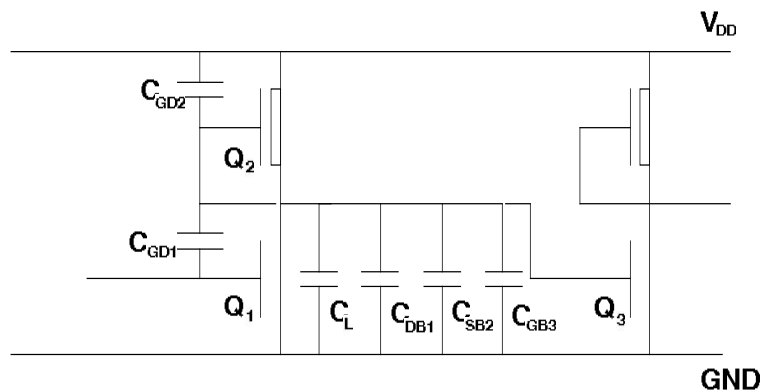


Poichè il pull-up è interdetto quando l'uscita è a livello basso, quest'ultimo buffer consuma solo sul pilotaggio dell'inverter mentre lo stadio di uscita, costituito dai due transistori enhancement, staticamente non dissipa potenza. Questo buffer ha  $V_{OL} = 0$ , perché  $I = 0$  ed il transistore di pull-down lavora in zona lineare, inoltre  $V_{OH} = V_{DD} - V_{TH}$ , in quest'ultima situazione si ha che il driver è interdetto, il load è in saturazione ed infine la corrente  $I = \frac{\beta}{2} (V_{GS} - V_{TH})^2 = 0$  perché sia nel pull-down che nel carico, che si suppone sempre di tipo capacitivo, non scorre corrente.

La rappresentazione grafica della tensione di uscita e della corrente in funzione della tensione di ingresso è la seguente:



Se si considera la struttura costituita da due inverter in cascata e si vogliono evidenziare le capacità che costituiscono questo circuito si ricava questo schema:



dove

- $C_{GD1}$  e  $C_{GD2}$  rappresentano le capacità gate-drain dei transistori 1 e 2;
- $C_L$  è la capacità relativa al collegamento metallico della linea;
- $C_{DB1}$  e  $C_{GB3}$  sono capacità di giunzione;
- $C_{SB2}$  è la capacità tra source e substrato del dispositivo 2.

### PROBLEMA

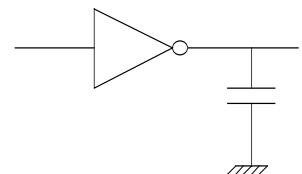
Una tipico problema da risolvere è il seguente:

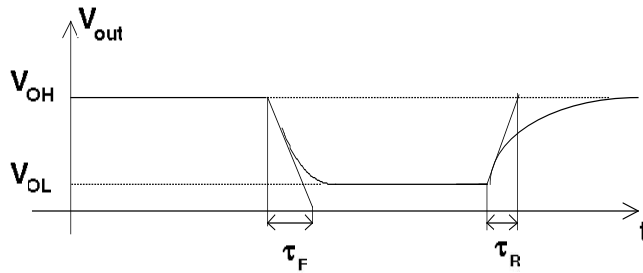
dato lo schema circuitale a lato, di cui si conoscono le grandezze  $L_{min} = 2 \mu m$ ,  $C = 1 pF$ ,  $V_{DD} = 5 V$ ,  $\mu_n C_{OX} = 20 \frac{\mu A}{V^2}$ ,  $V_{THN} = -V_{THP} = 1 V$  e

$\frac{\mu_p}{\mu_n} \cong \frac{1}{3}$ , si vogliono ricavare i valori di  $W_n$  e  $W_p$  che garantiscono che

$$\tau_R = \tau_F = 5 \text{ nsec.}$$

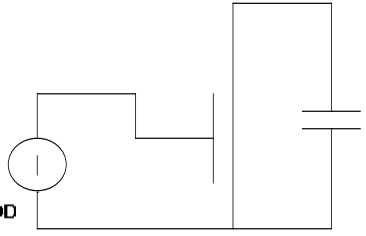
Per risolvere il quesito si suppone, per semplificare la trattazione, che sia  $\tau_F$  che  $\tau_R$  siano valutabili dalle sottotangenti all'andamento della tensione di uscita in funzione del tempo:





Bisogna per prima cosa stabilire se il transistoro che caratterizza il circuito lavora in zona lineare oppure in saturazione. Poiché si è nella situazione rappresentata nello schema a lato si vede che si lavora in saturazione per cui

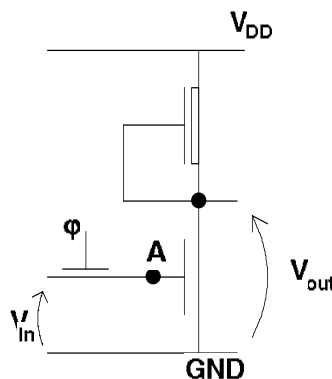
l'espressione della corrente è  $I = \frac{\beta}{2} (V_{GS} - V_{TH})^2$ .

Riprendendo la relazione  $Q = CV = I \tau$ , si ha che  $V_{DD}$  

$\frac{\beta}{2} (V_{DD} - V_{TH})^2 = \frac{C V_{DD}}{\tau}$ , essendo inoltre  $\beta = \mu_n C_{OX} \frac{W_n}{L_n}$  si ha che

$$\begin{cases} W_n = \frac{2 C V_{DD}}{\tau} \frac{1}{(V_{DD} - V_{TH})^2} \frac{L}{\mu_n C_{OX}} = \frac{2 \cdot 10^{-12} \cdot 5}{5 \cdot 10^{-9}} \frac{1}{16} \frac{2 \cdot 10^{-6}}{20 \cdot 10^{-6}} = 12.5 \mu m \\ W_p = 3 W_n = 37.5 \mu m \end{cases}$$

Si considera ora un esempio di circuito con carico costituito da un transistoro di tipo depletion e munito, in ingresso, di un pass-transistor regolato da una tensione  $\phi$ :



Si suppone che  $\phi$  sia a livello alto ( $\phi = V_{DD}$ ) mentre la tensione di ingresso  $V_{in}$  sia bassa, si analizza allora il comportamento del circuito al nodo A. Se inizialmente il nodo A è a zero rimane a questo potenziale mentre se è ad una tensione, per esempio, di 2 V lo stadio di ingresso assume una configurazione come quella a lato, in cui si vede che il nodo D è quello a tensione più elevata, ed il condensatore si scarica. Il pass-transistor lavora infatti in zona lineare in quanto la relazione  $V_{DS} > V_{GS} - V_{TH}$  non è verificata essendo  $V_{DS} = 2$  V,  $V_{GS} = 5$  V e  $V_{TH} = 1$ . La capacità evidenziata in

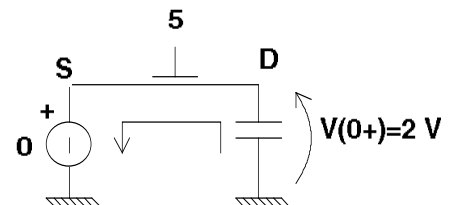
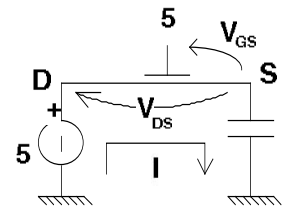


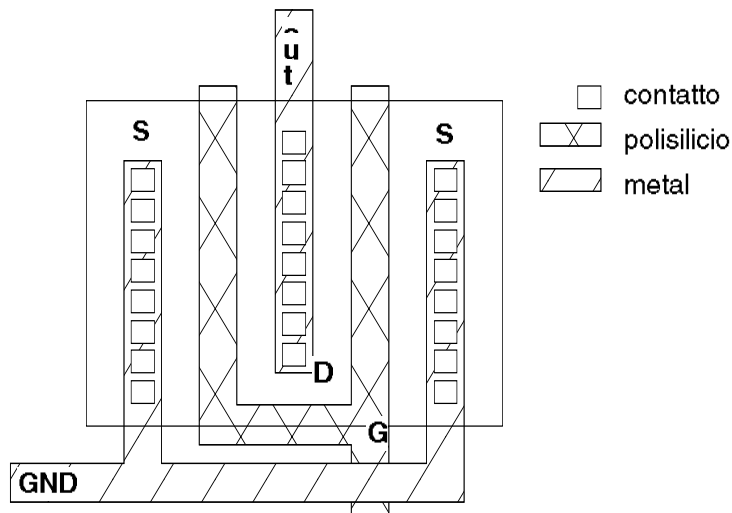
figura è la capacità che vede il nodo A e corrisponde alla capacità di gate del driver; poiché detta C tende a scaricarsi il dispositivo di pull-down ha una tensione di gate nulla per cui l'uscita della porta logica risulta a livello alto.

Si considera ora il comportamento del circuito al nodo A se la tensione di ingresso  $V_{in}$  è fatta crescere, in questo caso il pass-transistor ha gli elettrodi di drain e di source invertiti rispetto al caso precedente ed è caratterizzato dalle tensioni  $V_G = 5\text{ V}$ ,  $V_D = 5\text{ V}$  e  $V_S = 0\text{ V}$ .



Poiché  $V_{DS} = V_{GS}$  la relazione  $V_{DS} > V_{GS} - V_{TH}$  è soddisfatta perciò il pass-transistor lavora in zona di saturazione e quindi è attraversato da una corrente  $I_{pass-transistor} = \frac{\beta}{2}(V_{GS} - V_{TH})^2$ . Al termine del transitorio di  $V_{in}$  il nodo A è ad una tensione pari a  $V_{DD} - V_{TH_{pass-transistor}}$  questo perché nel pass-transistor la corrente scorre fino a che la tensione  $V_S$ , che cresce in quanto si carica il condensatore che rappresenta la capacità del gate del driver, non raggiunge un valore tale che  $V_{GS} = V_{TH_{pass-transistor}}$  per cui il pass-transistor entra in interdizione. In questa ultima situazione la tensione di uscita della porta logica è pari a  $V_{OL}$ .

A questo punto è utile effettuare una piccola digressione mettendo a confronto gli elettrodi di drain e di source: non è sempre vero che essi sono tecnologicamente uguali. Se infatti le larghezze  $W$  dei transistori superano i  $50\ \mu\text{m}$  è necessario, per mantenere le dimensioni ridotte, realizzare il dispositivo con una struttura a pettine. Supponendo che il transistor abbia una  $W = 60\ \mu\text{m}$  esso può, per esempio, essere realizzato come in figura sotto, dove si vede chiaramente che l'area occupata dalla giunzione di drain è pari alla metà di quella occupata dal source:



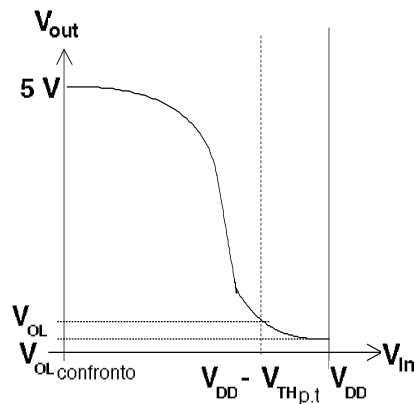
Il fatto che l'area di drain sia dimezzata determina una riduzione della capacità di carico.

Si è visto che, se il nodo A dello schema circuitale in analisi è ad un potenziale pari a  $V_{DD} - V_{TH_{pass transistor}}$ , l'uscita del circuito è a livello basso (essendo  $V_{DD} - V_{TH_{pass transistor}} > V_{TH_{driver}}$ ).

Confrontando  $V_{OL}$  di questo schema circuitale fornito di pass-transistor con la tensione di uscita a livello basso, definita  $V_{OL_{confronto}}$ , del circuito costituito esclusivamente dall'inverter risulta che

$V_{OL_{confronto}} < V_{OL}$ , infatti al gate del driver del circuito in esame è applicata una tensione pari a  $V_{DD} - V_{TH_{pass-transistor}}$  e non pari a  $V_{DD}$ .

Visualizzando la situazione appena descritta sulla caratteristica ingresso-uscita dell'inverter si ha:



Se si vuole fare in modo che  $V_{OL} = V_{OL_{confronto}}$  è necessario utilizzare nella struttura un inverter che abbia una caratteristica ingresso-uscita che decresce più velocemente di quella fino ad ora considerata, ovvero una caratteristica che garantisca una  $V_{out} = V_{OL_{confronto}}$  quando  $V_{in} = V_{DD} - V_{TH_{pass-transistor}}$ . Per realizzare un inverter con la caratteristica desiderata è necessario avere una resistenza di carico più elevata e per far ciò si può modificare il *load* dell'inverter aumentando  $W_{load}$ , lasciando invariato il *driver*, oppure, in alternativa, si può cambiare la tensione di inversione  $V_{INV}$  e portarla dal valore  $\frac{V_{DD}}{2}$  a  $\frac{V_{DD} - V_{TH_{pass-transistor}}}{2}$ . Se si sceglie di variare la caratteristica ingresso-uscita diminuendo la tensione di inversione si determina un aumento del coefficiente  $k_r$  infatti, data l'espressione di  $k_r$  in funzione di  $m$  cioè  $k_r = \frac{(0.6 + m)(1 - m)}{(m - 0.2)^2}$  ed essendo

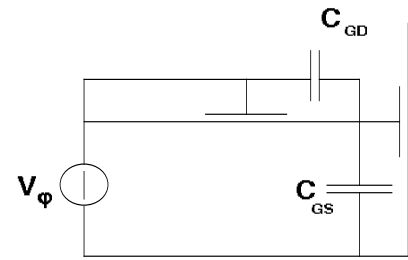
$V_{INV} = m V_{DD}$ , si ha che:

- se  $V_{INV} = \frac{V_{DD}}{2}$ , allora  $m = 0.5$ , per cui  $k_r \cong 5$ ;
- se  $V_{INV} = \frac{V_{DD} - V_{TH_{pass-transistor}}}{2}$ , per esempio  $m = 0.35$ , per cui  $k_r \cong 10$ .

Se la tensione di controllo va a 0V il pass-transistor si interdice, cioè non lascia più passare corrente, ed il nodo A diventa un elemento di memoria in quanto mantiene la sua tensione più o meno costante. Si possono avere piccole variazioni nei valori di tensione del nodo A a causa della possibile presenza di maglie CE attraverso cui possono passare impulsi di corrente. Se ho una configurazione caratterizzata da due pass-transistor in ingresso ed uno di essi, a causa dell'applicazione di una tensione di controllo nulla, si apre si manifesta un fenomeno di *charge injection* o di distribuzione di carica del canale del dispositivo che si è aperto; se la capacità del *driver* è sufficientemente grande l'effetto dovuto alla distribuzione della carica è poco evidente. Il lento processo di scarica del nodo A può essere dovuto a:

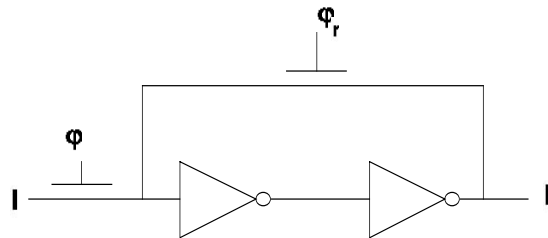


1. *clock-feed through* che è un effetto dovuto alle capacità  $C_{GS}$  del *driver* e  $C_{GD}$  del pass-transistor, dette capacità possono infatti dare origine a maglie CE che, quando la tensione di controllo si annulla, lasciano passare impulsi di corrente;
2. *charge injection*;
3. *correnti di diffusione*.



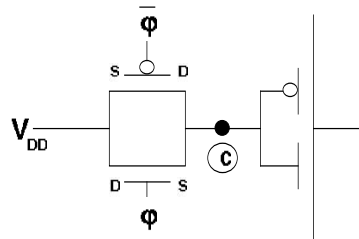
Il tempo di scarica del nodo A è dell'ordine dei msec ed è molto più lungo delle costanti tipiche del circuito (che sono dell'ordine dei  $\mu$ sec). La struttura appena analizzata è l'elemento base di tutti i dispositivi di memorizzazione dinamici.

E' importante che l'informazione rimanga memorizzata il più a lungo possibile per far ciò devo avere una struttura del tipo:



Per mantenere più a lungo l'informazione memorizzata sul nodo A è necessario riscrivere il dato con frequenza opportuna cioè è necessario effettuare un refresh dell'informazione memorizzata.

Utilizzando la tecnologia CMOS la struttura prima realizzata con pass-transistor risulta avere lo schema circuitale riportato di seguito in cui l'interruttore è ottenuto usando un doppio transistore:

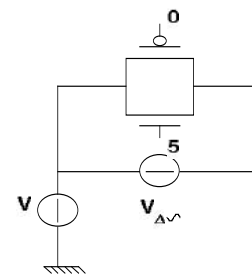


Questo schema presenta il vantaggio che il nodo C ha una tensione a livello alto pari a  $V_{DD}$ .

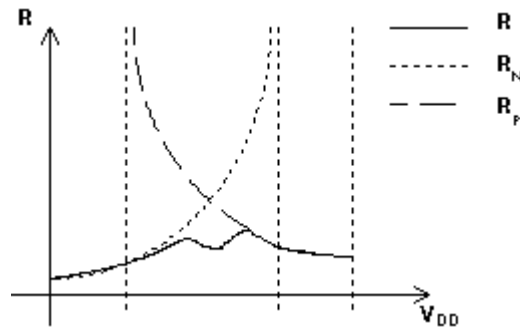
Analizzando in dettaglio il funzionamento del transfer gate si misura, con lo schema a lato in cui  $V_{\Delta\sim}$  è un piccolo segnale, la resistenza R, che si ottiene misurando le resistenze del transistore a canale n e del transistore a canale p, infine si fa il parallelo:

se  $V = 0$  si ha che  $R_N = \frac{1}{\beta_N (V_{GS_N} - V_{TH_N})}$  e se  $V = 5$  si ha che

$$R_P = \frac{1}{\beta_P (V_{GS_P} - V_{TH_P})}$$



Si riportano sotto gli andamenti delle resistenze  $R$ ,  $R_P$  ed  $R_N$ :

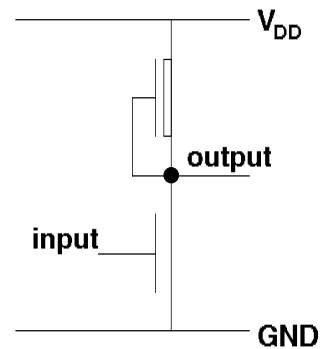


Le realizzazioni CMOS sono dal punto di vista circuitale più complicate perché necessitano la presenza del segnale di clock.

A questo punto è opportuno domandarsi quale dei due transistori che costituiscono la struttura dell'inverter sia quello che effettua l'operazione logica; è facile vedere che il dispositivo che realizza la funzione logica è il *driver* infatti, come risulta dallo schema circuitale riportato a lato, solo il *driver* è collegato al segnale di ingresso. Il meccanismo che effettua l'operazione logica è basato sull'attivazione di un collegamento tra l'uscita ed una delle due alimentazioni.

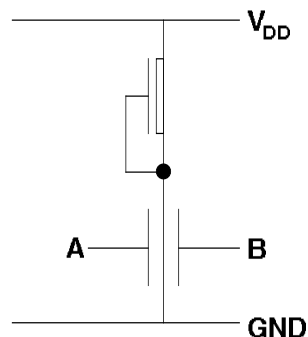
Prendendo in considerazione l'alimentazione di massa si può evidenziare

come l'applicazione del segnale di ingresso determina il collegamento a GND in un inverter:



input	collegamento a GND	output
0	0	1
1	1	0

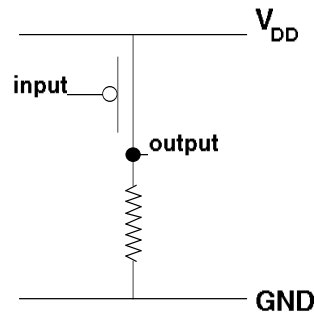
Si vuole ora analizzare la dinamica del collegamento a massa di una porta logica NOR il cui schema circuitale è il seguente:



La tabella seguente riporta gli andamenti dell'uscita e del collegamento verso massa del gate NOR:

A	B	coll. a GND	output
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

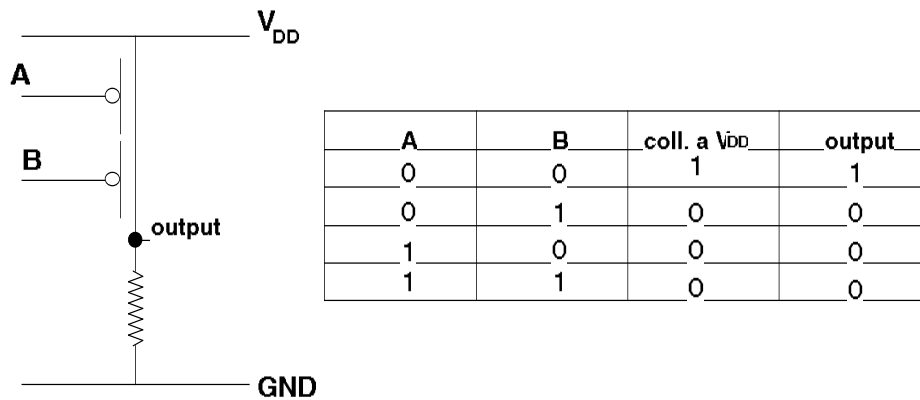
Un possibile schema alternativo per la realizzazione di un inverter è il seguente:



La tabella che segue riporta la dinamica del collegamento verso la tensione di alimentazione  $V_{DD}$  e dell'uscita in funzione del segnale di ingresso:

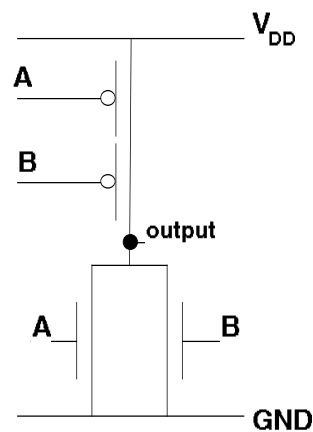
input	collegamento a $V_{DD}$	output
0	1	1
1	0	0

Per costruire, con la struttura appena analizzata, un gate NOR si utilizza lo schema:

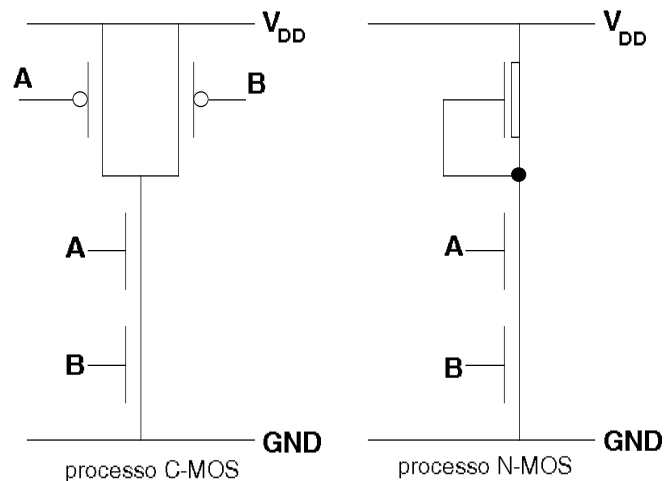


Dalla tabella sopra riportata si vede che il collegamento a  $V_{DD}$  è attivo solo se sia l'ingresso A che l'ingresso B sono a livello basso.

Un ulteriore schema che permette la realizzazione di una porta logica NOR si ottiene sostituendo la resistenza del circuito precedente con una struttura costituita da due transistori, cioè:



La porta logica NAND ha una struttura duale sia per la realizzazione con tecnologia C-MOS che per quella con tecnologia N-MOS, ed in particolare si hanno i seguenti schemi:



Si focalizza ora l'attenzione sui problemi che sorgono nel dimensionamento statico dei circuiti realizzati con tecnologia N-MOS (gli schemi C-MOS non danno alcuna complicazione dal punto di vista statico).

Con porte logiche NOR si effettua il dimensionamento nel *worst case*, che si ha quando solo uno tra gli ingressi A e B è attivo; se A e B sono attivi contemporaneamente, sia  $\tau_F$  che  $V_{OL}$  sono più bassi e questo è un vantaggio. Nel caso di un gate NOR non vi è sostanziale differenza tra struttura a 2 e struttura a 100 ingressi, anzi le prestazioni tendono a migliorare al crescere del numero degli ingressi.

Si hanno più difficoltà con porte logiche NAND di tipo N-MOS in quanto lo schema circuitale ha i *drivers* in serie, per cui, se si vuole dimensionare il circuito con N ingressi, si possono mantenere le dimensioni del *load* costanti mentre la larghezza del *driver* deve essere moltiplicata per N, cioè deve essere portata ad un valore pari a  $W \times N$  con conseguente aumento nell'occupazione di area. Quando il numero degli ingressi diventa maggiore di 6 i problemi si complicano ulteriormente in quanto la corrente che deve scorrere nella struttura raggiunge valori abbastanza elevati. Il circuito

che realizza l'operazione logica di NAND presenta complicazioni anche dal punto di vista elettrico in quanto si manifesta l'*effetto body*, infatti il *driver* superiore può condurre meno di quello inferiore cioè può avere una tensione di soglia più alta.

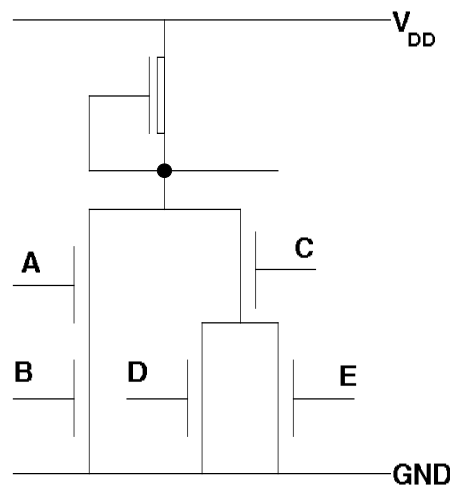
Per realizzare un circuito NAND a N ingressi si può, in alternativa, mantenere le dimensioni del *driver* costanti ed aumentare la lunghezza L del *load*; questa soluzione non elimina il *body effect* ed inoltre introduce un aumento nei ritardi.

Un inconveniente delle porte logiche C-MOS è rappresentato dalla logica doppia, infatti la funzione logica è realizzata sia da transistori di tipo p che da transistori di tipo n, e questo fatto rappresenta uno spreco.

Le porte logiche di tipo C-MOS fino ad ora analizzate presentano inoltre qualche problema dal punto di vista dinamico in quanto i tempi di discesa crescono al crescere degli ingressi; al fine di mantenere contenuti detti tempi di ritardo si impone che il numero degli ingressi sia minore di 10.

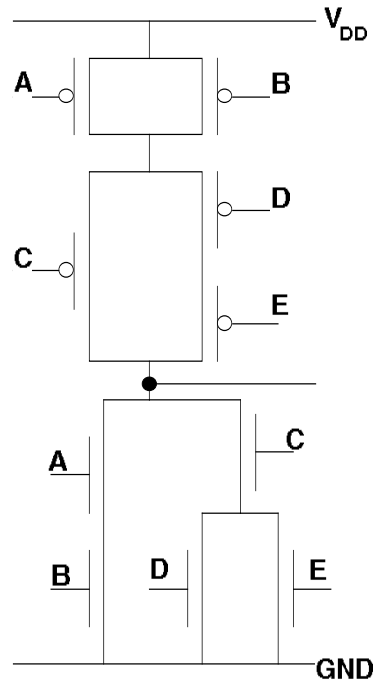
### ESERCIZIO

Si vuole realizzare un circuito che implementi la funzione logica  $z = \overline{A B + C(D + E)}$ . Scegliendo di utilizzare un processo N-MOS la struttura che si ottiene è la seguente:



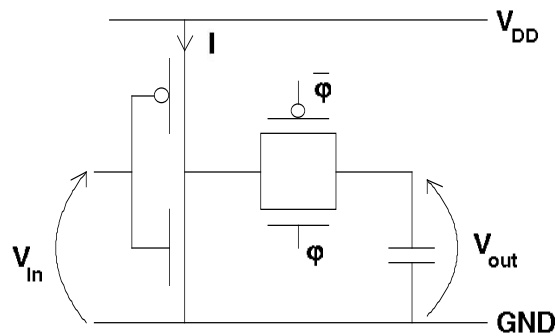
In un circuito N-MOS l'OR è ottenuto dal parallelo dei transistori mentre l'AND dalla serie.

La funzione logica  $z = \overline{A B + C(D + E)}$  è invece realizzata dal seguente circuito C-MOS:



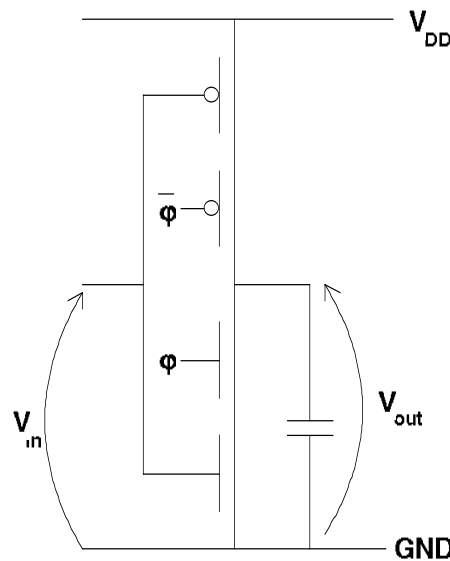
Nello schema appena riportato si vede che la serie dei dispositivi realizza l'OR mentre il parallelo permette di ottenere l'AND.

Si consideri ora la struttura costituita da inverter e *transfer gate* cioè:



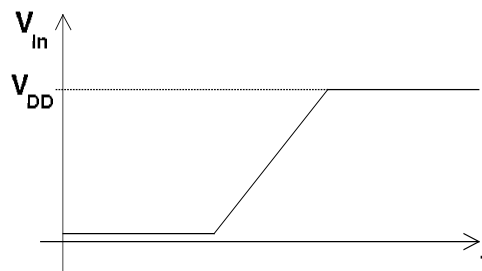
Data la struttura sopra, se la tensione di controllo  $\phi$  del *transfer gate* è a livello alto, cioè  $\phi = 1$ , si ha che  $V_{out} = \overline{V_{in}}$ , mentre se  $\phi = 0$  si ha che  $V_{out_n} = V_{out_{n-1}}$  ovvero la struttura rappresenta una cella di memoria.

Uno schema circuitale che effettua la stessa funzione logica di quello appena analizzato è il *clocked C-MOS* o *C<sup>2</sup>MOS* che ha questa configurazione:



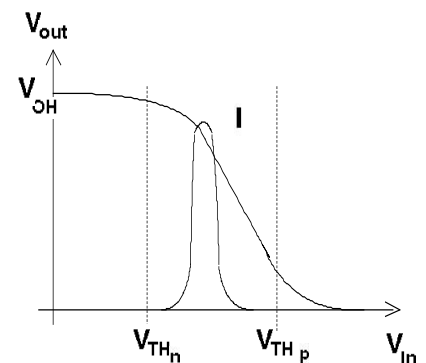
Valutando il comportamento del clocked C-MOS si vede che se  $\phi = 1$  allora  $V_{out} = \overline{V_{in}}$ , mentre se  $\phi = 0$  si ha  $V_{out_n} = V_{out_{n-1}}$ . Il processo di carica è dovuto o alla serie dei due transistori di tipo p o a quella dei due transistori di tipo n.

Si vuole studiare l'andamento della corrente nell'inverter durante le transizioni. Supponiamo che la tensione in ingresso all'inverter abbia il seguente andamento nel tempo:



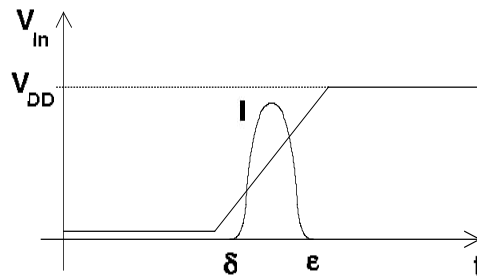
Riportando sulla caratteristica dell'inverter la tensione di soglia sia del dispositivo a canale n che del dispositivo a canale p si vede che:

- se  $V_{in} < V_{TH_n}$  conduce solo il transistor di tipo p perché quello a canale n è interdetto;
- se  $V_{TH_n} \leq V_{in} \leq V_{TH_p}$  entrambi i dispositivi sono in conduzione;
- se  $V_{in} > V_{TH_p}$  conduce solo il dispositivo di tipo n perché quello a canale p è interdetto.



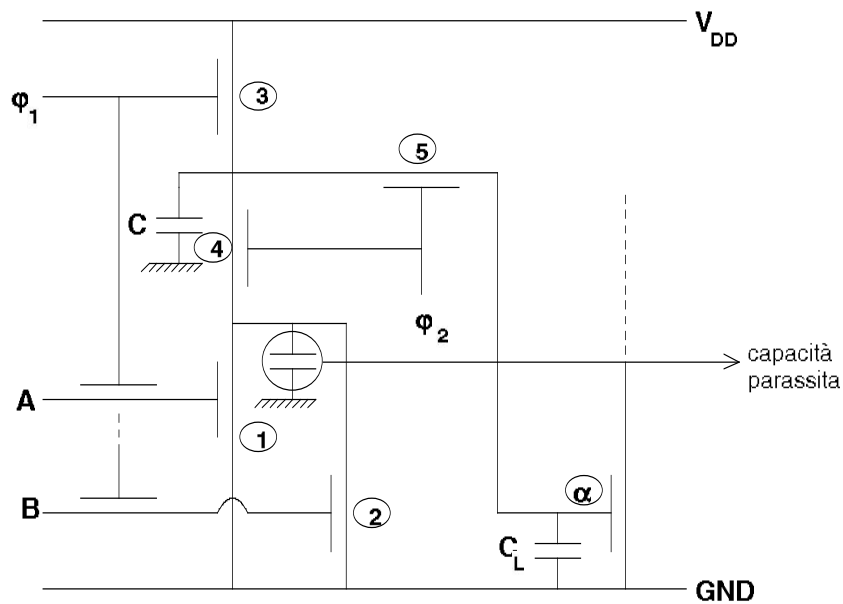
Risulta chiaro che nell'inverter scorre corrente quando entrambi i transistori conducono cioè quando  $V_{TH_n} \leq V_{in} \leq V_{TH_p}$ .

Essendo  $t = \delta$  l'istante in cui la tensione di ingresso  $V_{in} = V_{TH_n}$  e  $t = \epsilon$  l'istante in cui  $V_{in} = V_{TH_p}$  si possono mettere a confronto gli andamenti nel tempo della tensione di ingresso e della corrente  $I$  che scorre nell'inverter:



Da quanto appena riportato risulta che l'inverter consuma una potenza proporzionale all'impulso di corrente moltiplicato per la durata del tempo di salita della tensione di ingresso. Poiché il tempo di salita non può mai essere nullo, l'inverter consuma una potenza proporzionale alla frequenza di funzionamento cioè alla frequenza delle transizioni della tensione di ingresso. Per evitare questo spreco di potenza bisogna fare in modo che la tensione di controllo  $\phi$  sia nulla quando la tensione di ingresso  $V_{in}$  sta variando; in questo modo infatti non esiste più il collegamento diretto con massa ed il condensatore si carica con  $V_{in}$ .

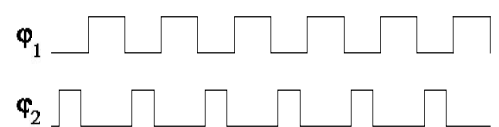
Si considera ora un circuito N-MOS che non consuma staticamente e che ha il seguente schema circuitale:



La struttura appena visualizzata rappresenta parte di un NOR N-MOS dinamico, il circuito completo è costituito da un'ulteriore porta con configurazione uguale a quella riportata.

L'andamento nel tempo dei segnali di controllo  $\phi_1$  e  $\phi_2$  è tale da garantire che  $\phi_1 \text{ AND } \phi_2 = 0$  cioè i due segnali di controllo non si sovrappongono (si tratta di *non overlap phases*) come si vede dalle forme d'onda riportate a lato. Il funzionamento del circuito è il seguente:

quando la tensione di controllo  $\phi_1$  è a livello alto i valori A e B





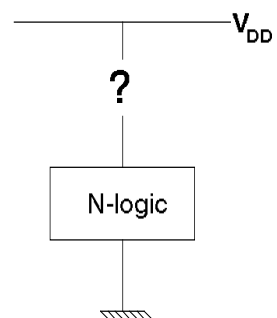
sono copiati sui gate dei transistori ① e ②, inoltre contemporaneamente il condensatore C si carica ad una tensione definita  $V^*$  e di valore pari a  $V_{DD} - V_{TH_3}$ , il dispositivo ③ infatti è in saturazione. In particolare  $V^*$  è pari a 4V, se si trascura il body effect, altrimenti, considerando il body effect, vale 3.6 V. Essendo  $\varphi_1$  a livello alto,  $\varphi_2$  è a livello basso per cui sul condensatore  $C_L$  rimane memorizzato il dato del ciclo precedente. Quando la tensione  $\varphi_1$  si abbassa e  $\varphi_2$  si alza il condensatore C non è più collegato a  $V_{DD}$  ed i segnali A e B sono staccati dai gate dei dispositivi ① e ②.

In questo caso il comportamento del circuito dipende dalla storia precedente ed in particolare:

- se  $C_L$  è carico e se A era, al ciclo precedente, a '1' i condensatori C e  $C_L$  si scaricano (in particolare C si scarica attraverso il transistor ① mentre  $C_L$  attraverso i transistori ⑤, ④ e ①) per cui il nodo  $\alpha$  va a '0';
- se  $C_L$  è scarico allora il nodo  $\alpha$  è a '0';
- se  $A = B = 0$  e  $C_L$  è carico i transistori ① e ② non conducono per cui, nonostante i dispositivi ④ e ⑤ siano "on", C e  $C_L$  rimangono carichi ed il nodo  $\alpha$  è a '1';
- se  $C_L$  è scarico ed  $A = B = 0$  la tensione ai capi del condensatore  $C_L$ , definita  $V_{C_L}$ , alla fine della transizione vale  $\frac{C}{C+C_L} V^*$  infatti i dispositivi ① e ② sono aperti, ma i dispositivi ④ e ⑤ sono chiusi perciò si realizza un partitore capacitivo di tensione (il valore della tensione  $V_{C_L}$  è alto se  $C \gg C_L$ ).

La tecnologia N-MOS è più semplice e meno costosa della C-MOS ma presenta il problema di come collegare la tensione di alimentazione  $V_{DD}$  al circuito logico, le alternative per realizzare questa connessione sono:

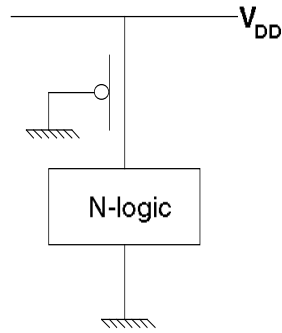
1. utilizzazione di una resistenza R;
2. uso di dispositivo enhancement;
3. uso di dispositivo depletion;
4. uso di logica di tipo p.



La soluzione presentata al punto 4 è valida, in quanto funziona bene, poiché gode di alcune proprietà quali quella di non necessitare di un dimensionamento statico nullo, ha però come inconveniente quello di richiedere un numero di transistori elevato.

Per diminuire il numero di transistori è necessario fare ricorso ad una logica C-MOS.

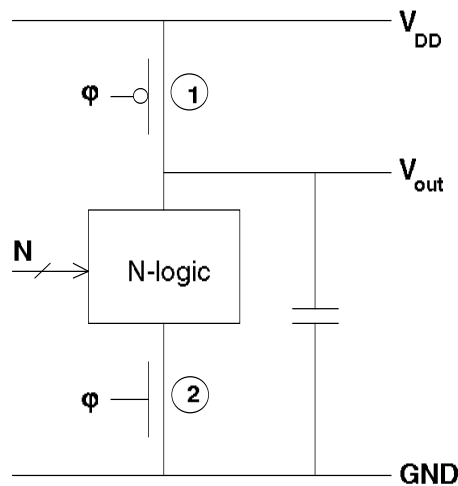
Una logica alternativa è la logica **pseudo N-MOS** la cui struttura può essere così schematizzata:



La soluzione sopra riportata non è consigliata perché se da un lato riduce il numero di transistori da  $2N$  a  $N+1$  (dove  $N$  è il numero di ingressi) si tratta di una logica che consuma troppo, infatti il transistor di carico conduce sempre essendo il gate di detto transistor cortocircuitato a massa.

## LOGICHE DINAMICHE C-MOS

Il blocco fondamentale delle logiche dinamiche C-MOS ha la seguente struttura:

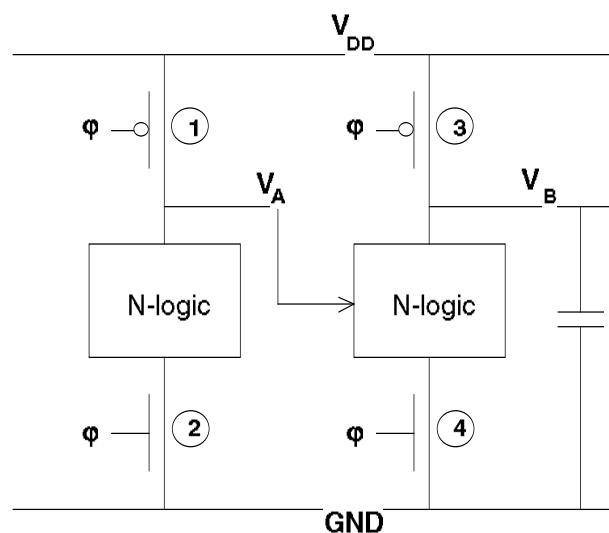


Il funzionamento del circuito sopra riportato si articola in due fasi:

- nella prima la tensione  $\phi$  è a livello basso perciò  $V_{out} = V_{DD}$  ovvero il condensatore di uscita è precaricato, infatti è collegato a  $V_{DD}$  attraverso il dispositivo ①, che conduce;
- nella seconda fase  $\phi$  è a livello alto per cui il dispositivo ① è interdetto mentre quello contrassegnato da ② conduce, il condensatore di uscita si scarica se è collegato a massa mentre rimane carico se è collegato a  $V_{DD}$ .

Durante la fase due, definita fase di valutazione, il collegamento a GND o a  $V_{DD}$  è imposto dalla logica, che a sua volta dipende dagli N-ingressi; il livello della tensione di uscita alla fine della fase di valutazione è il risultato da prendere in considerazione, cioè è il valore di  $V_{out}$  esatto.

Si prende ora in considerazione l'insieme di due stadi con struttura uguale a quella appena analizzata:



La struttura costituita dai due stadi in cascata ha un funzionamento articolato ancora in due fasi: la precarica, che fa assumere alle tensioni  $V_A$  e  $V_B$  un valore pari a  $V_{DD}$ , e la fase valutazione.

All'inizio del processo di valutazione  $V_A$  e  $V_B$  sono a livello alto e durante questa fase possono presentarsi due casi:

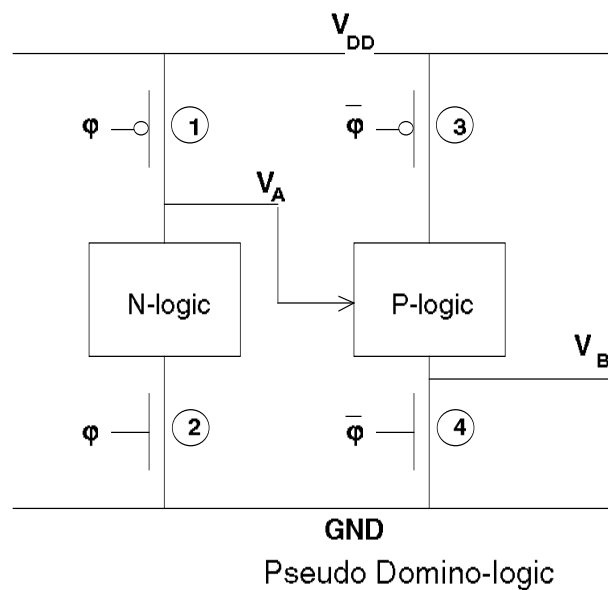
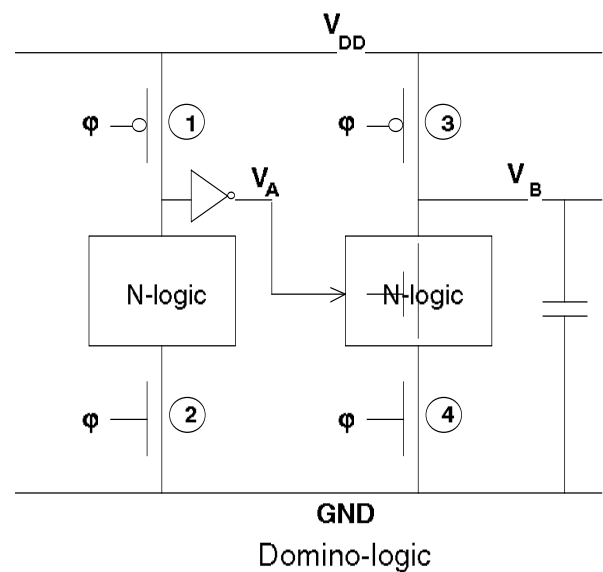
- se  $V_A$  si mantiene alto allora  $V_B$  va a livello basso perché il condensatore si scarica attraverso il transistor pilotato da  $V_A$  ed il dispositivo ④ (si ricorda che durante la fase di valutazione  $\phi$  è a '1');
- se  $V_A$  si abbassa il dispositivo pilotato da questa tensione conduce fino a che  $V_A$  non va sotto il valore di soglia, per questo la tensione  $V_B$  decresce fino a che il percorso che collega il condensatore a massa non si apre, da questo istante in poi  $V_B$  non decresce più, fermandosi ad un valore di tensione che non è corretto perché non classificabile come livello alto e neppure come basso.

Il problema del circuito appena analizzato è legato al fatto che, durante la fase di valutazione, nonostante  $V_A$  vada a livello basso, il transistor che costituisce l'N-LOGIC inizialmente conduce, in quanto è un dispositivo a canale n a cui è applicata una tensione di gate per un certo tempo più elevata della tensione di soglia.

Per risolvere il problema si può inserire un inverter in modo tale che durante la fase di precarica la tensione  $V_A$  vada a livello basso, in questo modo durante la valutazione se  $V_A$  si alza  $V_B$  va a livello basso, mentre se  $V_A$  resta a '0' allora  $V_B$  rimane alto.

Bisogna notare che l'aggiunta dell'inverter porta ad avere un OR invece che NOR inoltre una grandezza critica risulta essere la frequenza di funzionamento poiché se non è sufficientemente bassa si hanno dei problemi.

Un tipo di struttura alternativa è la seguente:



Il problema del primo circuito visto è quello che la tensione  $V_A$ , anche se si sta abbassando, per un certo periodo durante la fase di valutazione, è sufficientemente alta da permettere la conduzione del transistor che pilota: questo comportamento è dovuto al fatto che il dispositivo pilotato da  $V_A$  è di tipo n. Sostituendo l'N-logic del secondo stadio con una P-logic si risolve il problema perché i transistori a canale p non conducono se il segnale ad essi applicato è a livello alto.

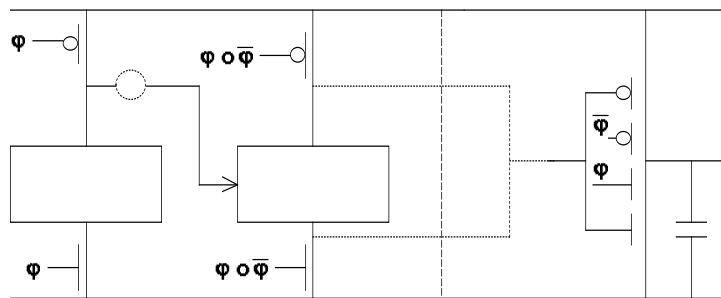
Il nuovo schema circuitale funziona in questo modo:

- durante la fase di precarica ( $\phi = 0$ ) la tensione  $V_A$  va a  $V_{DD}$  e la tensione  $V_B$  va a zero.
- durante la fase di valutazione ( $\phi = 1$ ) i dispositivi ① e ④ sono in interdizione per cui la tensione  $V_A$  viene condizionalmente scaricata, se la funzione logica lo permette, verso massa tramite il transistor di valutazione.

Con questa struttura si ha il solito inverter ovvero se  $V_A$  va a '0' allora  $V_B$  va ad un livello di tensione alto mentre se  $V_A$  rimane a '1' conseguentemente  $V_B$  rimane a '0'.

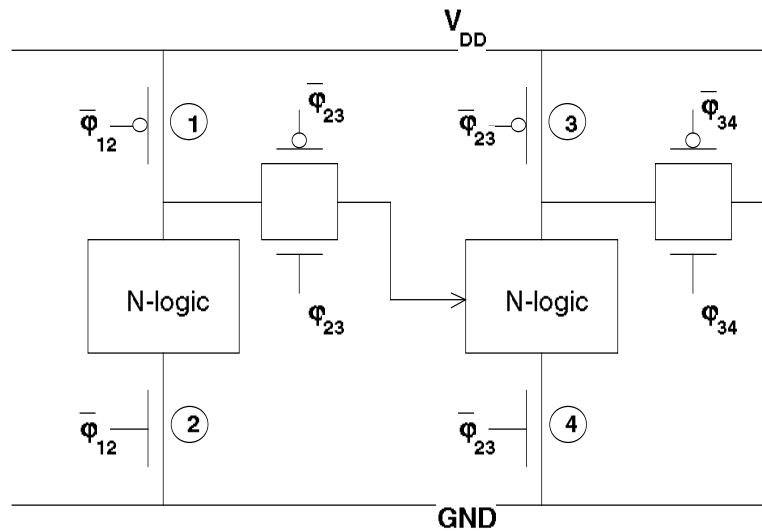
Lo schema a logica pseudo domino è una struttura abbastanza bella e con essa si fanno le PLA, che sono logiche programmabili a due livelli. Esempio di logiche a due livelli sono la NOR-NOR oppure la NAND-NAND, le prime si usano con processi N-MOS a carico depletion mentre le seconde si usano con i processi C-MOS dinamici.

Si può aggiungere in cascata sia alla logica pseudo domino che a quella domino un inverter  $C^2MOS$  ovvero si può ricorrere ad una struttura del tipo:



Questo nuovo circuito durante la fase di precarica mantiene in uscita il valore del ciclo precedente perché i transistori intermedi dello stadio di uscita sono interdetti (essendo  $\phi = 0$ ). Durante la fase di valutazione i due transistori intermedi dell'inverter  $C^2MOS$  si attivano e propagano il dato attuale in uscita: in conclusione l'inverter  $C^2MOS$ , inserito come ultimo stadio, serve a mantenere stabile il dato in uscita durante la fase di precarica del ciclo successivo. La frequenza massima di funzionamento di questo schema dipende dai tempi di ritardo; se quindi si vuole realizzare una funzione logica a 10 livelli sufficientemente veloce non si può inglobare tutto in uno stesso ciclo di precarica-valutazione, ma bisogna ricorrere ad una organizzazione a *pipeline*, utilizzando i flip-flop D-latch (che servono a mantenere il dato).

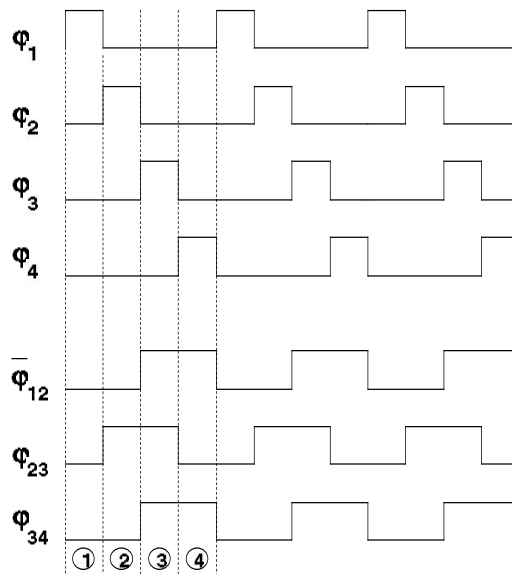
Come ultima alternativa si presenta la seguente struttura:



Logica dinamica C-MOS a quattro fasi

Nel circuito appena riportato la trasmissione dei segnali da uno stadio all'altro è ritardata tramite i *transfer gates*. Dal punto di vista strutturale la logica dinamica C-MOS a quattro fasi è semplice ma è delicata la temporizzazione dei clock: il segnale  $\phi_{12}$  deve essere a livello alto nella prima e nella seconda sotto-fase mentre  $\phi_{34}$  deve essere a '1' nella terza e nella quarta sotto-fase.

Si riportano gli andamenti temporali dei segnali di controllo:



Questo circuito attraverso i transfer gate garantisce che la tensione all'ingresso del secondo stadio, durante la fase di valutazione, rimanga fissa, in quanto, attraverso l'interdizione del transfer gate che collega i due stadi, rimane memorizzato il dato di uscita del primo stadio. In altre parole si superano i problemi che potrebbe causare la variazione di  $V_A$  al secondo stadio bloccando strutturalmente, attraverso i transfer gate, la tensione di ingresso allo stadio di uscita.

Descrivendo dettagliatamente il funzionamento dei quattro sotto-cicli si ha che:

- durante il sotto-ciclo ① lo stadio di ingresso è in precarica ed il transfer gate che collega i due stadi è interdetto;
- durante il sotto-ciclo ② il transfer gate che collega i due stadi conduce e fa passare un '1' anche allo stadio 2 che entra in fase di precarica;
- durante il sotto-ciclo ③ inizia la fase di valutazione del primo stadio, perché la tensione  $\varphi_{12}$  si abbassa, mentre il secondo stadio rimane in precarica;
- durante il sotto-ciclo ④ la fase di valutazione dello stadio di ingresso è conclusa ed inizia la fase di valutazione dello stadio di uscita.