

Introduzione ai IC

Legge di Moore (1965) il n° di transistor su un chip raddoppiava ogni 18/24 mesi, osservazione dalla quale Moore previse che l'efficienza della tecnologia sarebbe raddoppiata ogni 18 mesi.

Evoluzione in complessità, n° di transistor, dimensione della matrice, frequenza, ma purtroppo aumento della potenza prodotta (e quindi necessità di una sempre più efficiente dissipazione del calore)

Livello di astrazione:

- Sistema;
- Modulo;
- Gate;
- Circuito;
- Dispositivo;

Stima dell'efficienza di un circuito digitale

1) Costi:

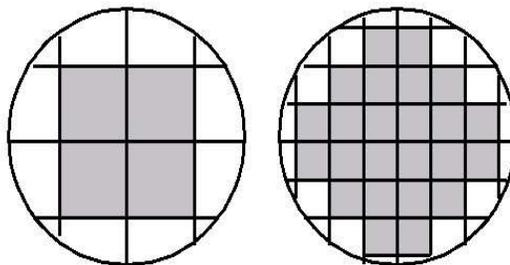
1. Costo fisso: tempo e sforzo nel progetto, creazione delle maschere, R & D (ricerca e sviluppo), spese dell'azienda, ecc...;
2. Costo variabile: costi dovuti alla lavorazione del Si (silicio), alla fase di packaging, ai test ecc... tali costi sono proporzionali al volume e all'area dei chip;

$$C_{ic} = C_v + \frac{C_f}{Vol} \quad C_v = \frac{C_{die} + C_{die\ test} + C_{pack}}{Y} \quad Y = \frac{n^\circ \text{ good chip per wafer}}{\text{total chips per wafer}} \cdot 100$$

$$C_{die} = \frac{C_{wafer}}{\text{dies per wafer} \cdot Y} \quad \text{dies per wafer} = \frac{\pi \cdot r_{wafer}^2}{A_{die}} - \frac{2 \pi \cdot r_{wafer}}{\sqrt{2} A_{die}} \quad \text{considerando}$$

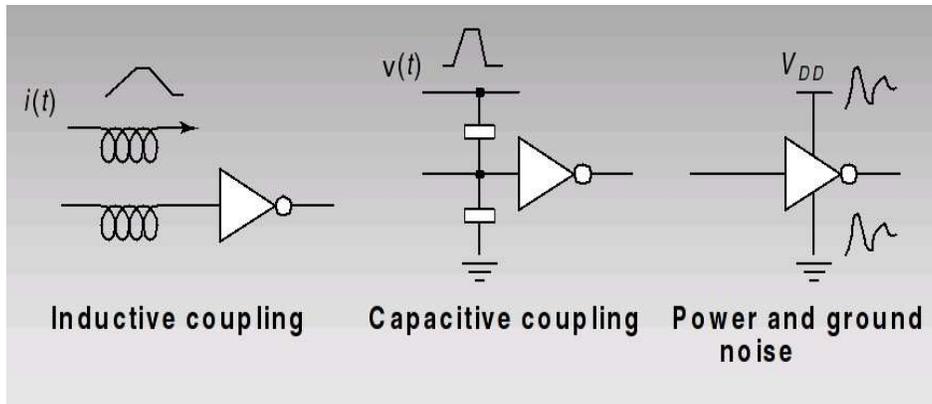
anche i difetti... $Y_{die} = 1 + \left(\frac{\text{difetti per unità di area} \cdot A_{die}}{\alpha} \right)^{-\alpha}$ con $\alpha \simeq 3$

PS: $C_{die} = f(A_{die})^4$



2) Affidabilità

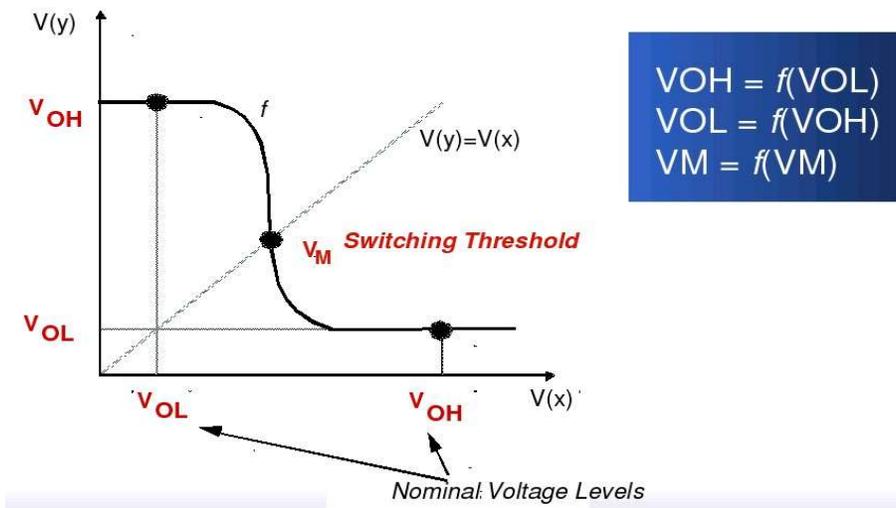
rumore: variazioni di tensione e di corrente indesiderate ai nodi logici



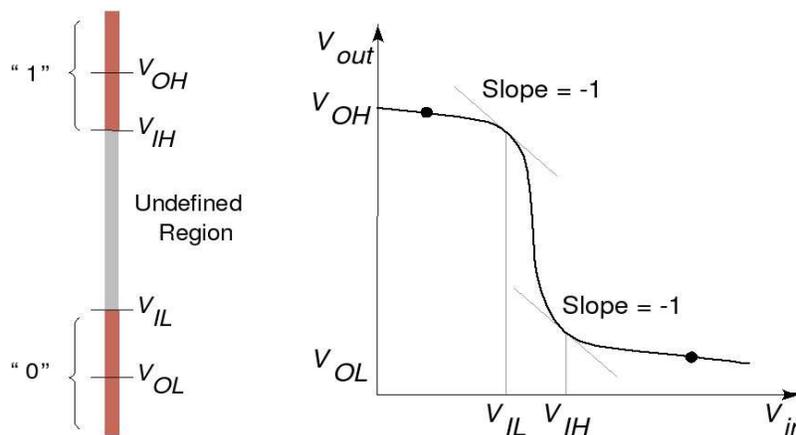
3) Performance:

1. VTC --> caratteristica del trasferimento di tensione

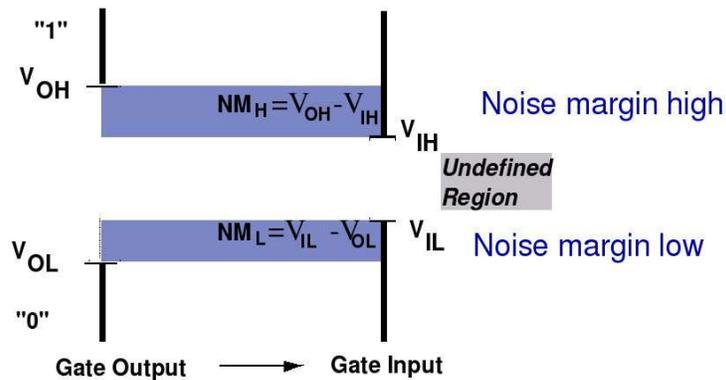
$V(x)$ $V(y)$ (esprime la V_0 in funzione di V_i)



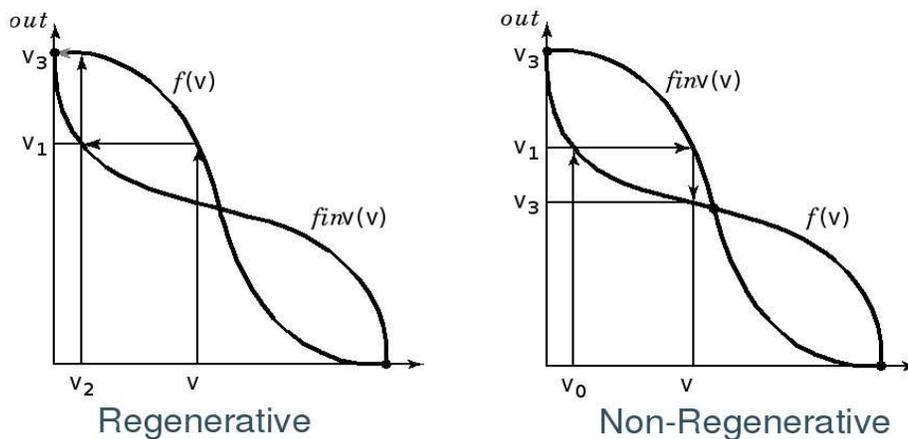
2. mapping da analogico a digitale



3. margini di rumore: più largo è il margine di rumore, migliore è il gate. I margini di rumore esprimono la capacità di un circuito di sopprffare una fonte di rumore --> in valore assoluto, sono però ingannevoli: la loro reale influenza dipende dal circuito. L'immunità al disturbo esprime l'abilità del sistema nel processare e trasmettere informazioni correttamente in presenza di rumore.



4. proprietà di rigenerazione: un gate con tali proprietà assicura che un segnale disturbato converga gradualmente a un livello di tensione normale ("1" o "0") dopo essere passato attraverso un numero finito di stati logici.



nel primo grafico V_0 converge verso uno dei valori normali, mentre nel secondo grafico V_0 casuale rimane nella regione indefinita.

Condizione necessaria per la rigenerazione: VTC deve avere una regione transitoria con un guadagno > 1 (in valore assoluto) circondata da 2 zone con un guadagno < 1 .

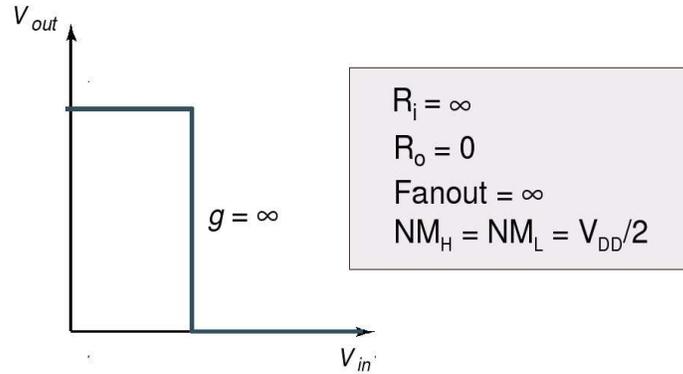
5. Fan-in e fan-out:

PS: i gate dovrebbero essere unidirezionali: un cambio in un output non dovrebbe riflettersi su un input stabile.

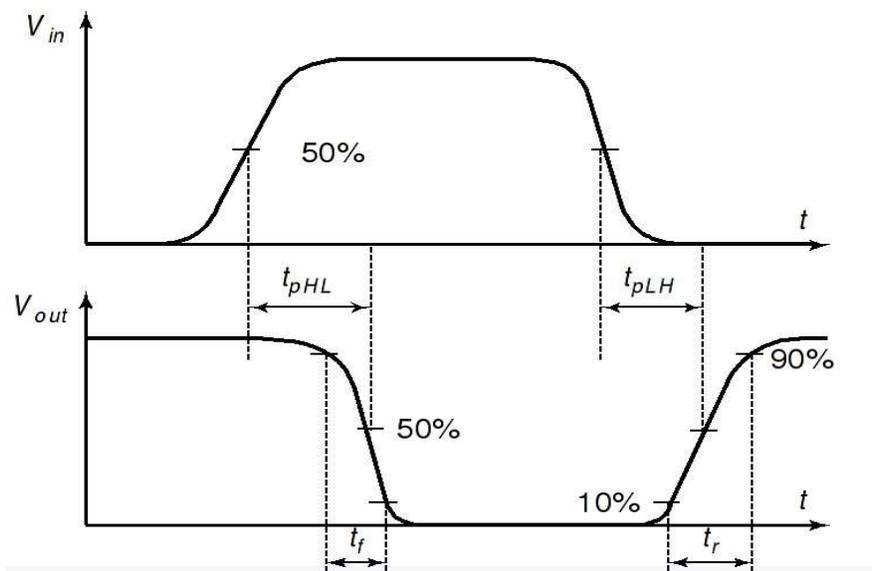
Fan-out: n° di gate di carico connessi all'output (più lenti all'aumentare del n° di output);

Fan-in: n° di input in un gate (più grossi e lenti all'aumentare del n° di input);

6. Gate digitale ideale: soglia di switch in mezzo alla zona definita



7. Delay



t_f = fall time

t_r = rise time dal 10 al 90% della tensione di regime.

$t_p = \frac{t_{pHL} + t_{pLH}}{2} \rightarrow$ tempo di propagazione

4) Dissipazione di energia e potenza

Consumo di potenza: quanta energia è consumata per ogni operazione e quanto calore il circuito dissipa $p(t) = V(t) \cdot i(t) = V_{supply} \cdot i(t)$

1. Potenza media dissipata: per dimensionare i dissipatori (e da eei dipende la durata delle batterie) $P_{ave} = \frac{1}{T} \int_t^{t+T} p(t) dt = \frac{V_{supply}}{T} \int_t^{t+T} j_{supply}(t) dt$

2. Potenza di picco: per dimensionare i conduttori

$P_{peack} = V_{supply} \cdot j_{peack}$ il ritardo di propagazione e il consumo di potenza sono strettamentecollegati. Il ritardo di propagazione è determinato dalla velocità con la quale un certo quantitativo di energia può essere immagazzinata nei condensatori

del gate (trasferimento più veloce = gate più veloce)

PDP: Power Delay Product --> energia consumata dal gate per la commutazione

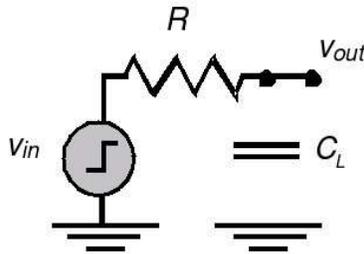
$$PDP = P_{ave} \cdot t_p ;$$

ADP: Energy Delay Product --> $EDP = (PDP)^2$ (anche questo misura la qualità di un gate, più precisa in quanto combina il gate ideale e il gate reale;

Ci sono 2 tipi di dissipazioni:

- 1) Dissipazione dinamica: al momento della commutazione, dovuta alla scarica dei condensatori e alle correnti in circolo, è funzione della frequenza;
- 2) Dissipazione statica: quando il gate non stà commutando, dovuta ai percorsi di conduzione o alle correnti di dispersione;

Esempio: analizzando un circuito RC del I° ordine si può osservare che:



$$E_n = \int_0^{\infty} i_{in}(t) V_{in}(t) dt = C V_{DD} V_{DD} = 1 = \text{tensione alta per convenzione} \quad \text{mentre}$$

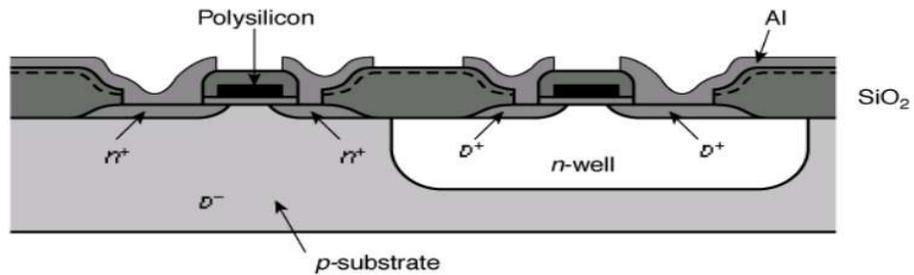
$$E_n = \text{energia necessaria per caricare il condensatore} = \int_0^{\infty} P_c(t) dt = \frac{C V_{DD}^2}{2} \rightarrow \text{uguale}$$

quindi all'energia dissipata in calore sul resistore.

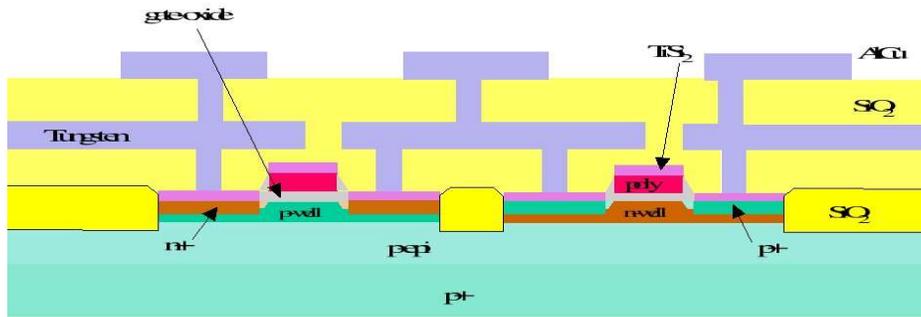
NB: i parametri di progetto da valutare sono quindi:

- dissipazione di potenza ed energia;
- costo;
- affidabilità e robustezza --> capacità di resistere alle sollecitazioni e al rumore (margini di rumore, immunità al rumore, impedenza);
- velocità;
- smaltimento;

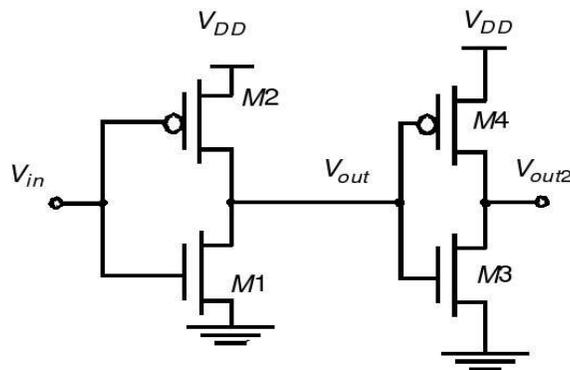
Processo di Fabbricazione



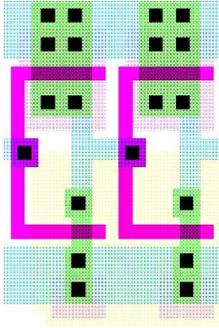
aree p: drogate con ioni accettori --> le lacune sono i portatori in maggioranza;
 aree n: drogate con ioni donatori --> gli e⁻ sono i portatori in maggioranza;
 Attualmente il processo di produzione usato si dice “a doppia finestra isolato in trincea”:



Uno schema circuitale:



quando un mos è ON, l'altro è OFF, quindi non c'è un cammino diretto tra V_{DD} e massa, ne consegue che $P_{dis} \simeq 0$ in condizioni statiche



Un CMOS visto nella fase di progettazione “dall’alto” (layout), tecnologia planare (viene disegnata solo la superficie)
 i quadratini neri sono i contatti, le piste viola sono i gate, la parte azzurra è il V_{DD} ed infine la parte bianca è il V_{SS}

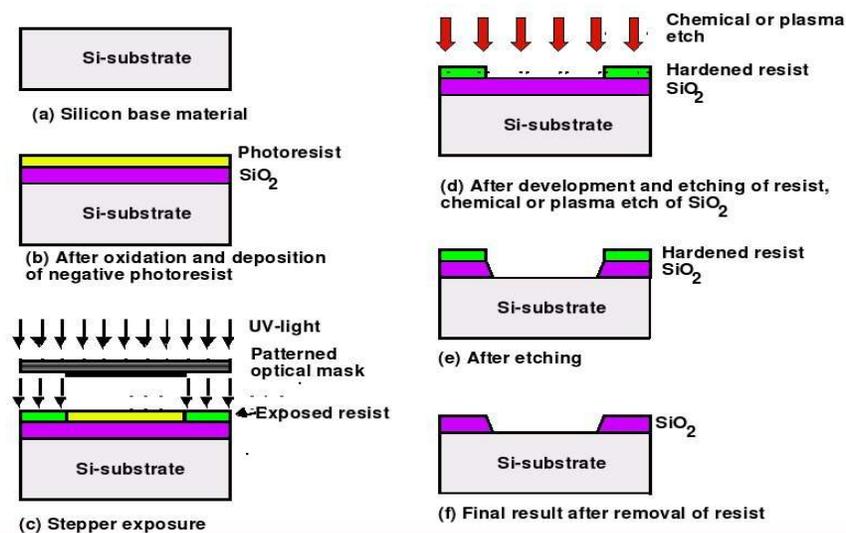
Fabbricazione di un CMOS

1) Wafer di silicio: derivati da singoli cristalli di silicio puro, tagliati con diametro dai 10 ai 30 cm e con spessore 1 mm;

2) Fotolitografia:

PS: per il patterning (modellazione) del SiO_2

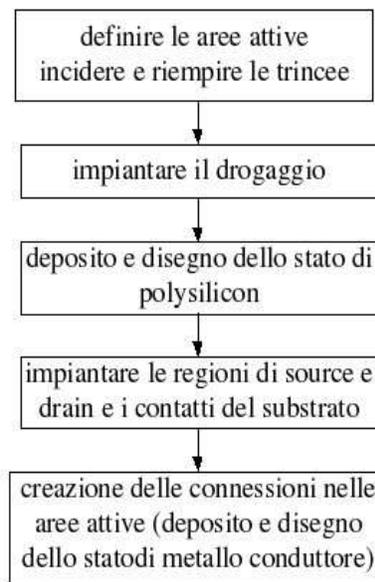
1. oxidation layering: applicazione di uno strato di SiO_2 ;
2. photoresist coating: strato di photoresist (materiale sensibile alla luce) che diventa solubile se esposto alla luce;
3. stepper exposure: applicazione della maschera di vetro con il disegno (opaca e trasparente) --> esposizione ai raggi ω : le parti trasparenti rendono il photoresist sottostante solubile;
4. photoresist development and bake: sviluppo del photoresist;
5. acid etching: bagno nell'acido --> solo le zone non esposte alla luce, avranno il photoresist bucato e cotto;
6. spin, rinse e dry: pulizia del wafer (deve essere perfetto);
7. altri processi;
8. photoresist removal (or ashing): rimozione del photoresist rimanente;



3) Altri importanti processi:

1. diffusione e impiantazione ionica: il primo processo diffonde il drogaggio sia verticalmente che orizzontalmente all'interno della superficie esposta --> la concentrazione maggiore c'è in superficie, mentre decresce gradualmente all'interno. Con il secondo processo, il drogante viene inserito nel silicio sotto forma ionica, in base alla velocità (= profondità) e alla corrente (=dosaggio) del raggio --> permette una maggior precisione e un maggior controllo nel processo produttivo (molto usato attualmente);
2. deposizione: applicazione di uno stato (sacrificale) di SiN_4 (per la formazione dell'ossido) tramite CVD (deposizione chimica a vapore);
3. incisione: per formare i percorsi e i buchi per i contatti;
4. planarizzazione: CMP (planarizzazione chimico-matematica) per rendere perfettamente piano il wafer;

4) Schema generale:



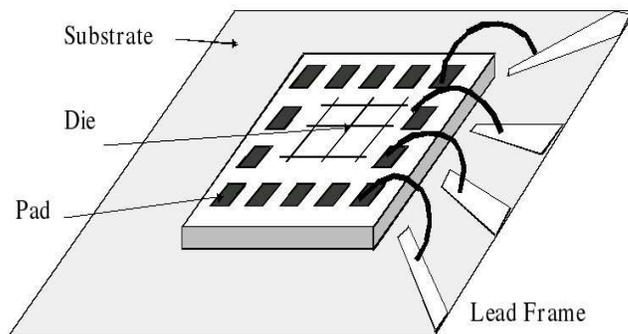
Packaging

Per il packaging servono:

- a) buoni dielettrici (scarsi effetti capacitivi parassiti);
- b) buone caratteristiche meccaniche (affidabilità robustezza [protezione del chip da agenti esterni]);
- c) buone caratteristiche termiche (capacità di dissipare calore);
- d) materiali economici (plastici) --> anche se spesso il packaging è più costoso del chip stesso;

Tecniche di interconnessione (Livello 1)

Wire bonding:

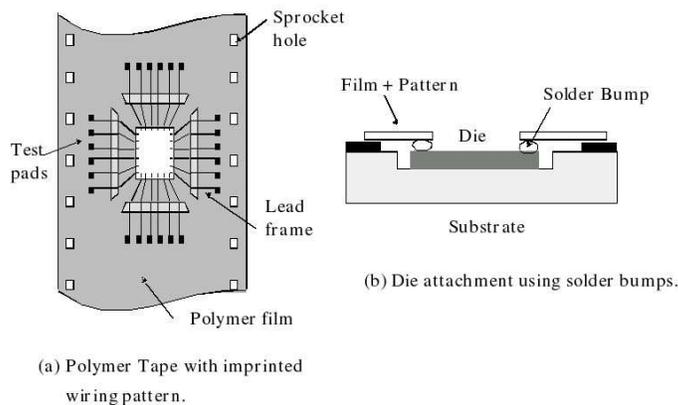


la parte anteriore del die è attaccata al substrato con della colla (con buona conduttività termica) successivamente, le piazzole dei chip sono singolarmente connesse alla struttura dei piedini con fili d'alluminio o d'oro.

Svantaggi:

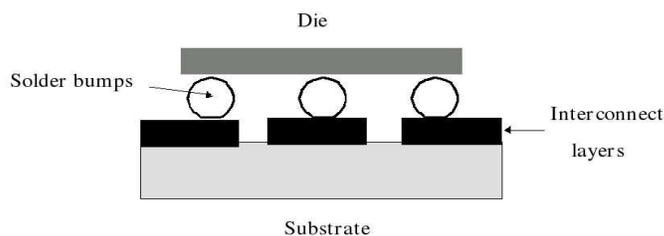
- 1) i fili devono essere connessi uno dopo l'altro;
- 2) difficoltà tecnologiche all'aumentare del n° di connessioni (per evitare cortocircuiti tra i fili);
- 3) difficile stimare il calore esatto degli effetti parassiti;

tape-automated bonding:



il die è attaccato ad una struttura di piedini stampata su una pellicola di polimeri. La connessione tra i fili della pellicola e le piazzole delle die è fatta usando palline di metallo saldate (in genere stagno) e uno strato di metallo. Vantaggio: tutte le connessioni vengono eseguite simultaneamente.

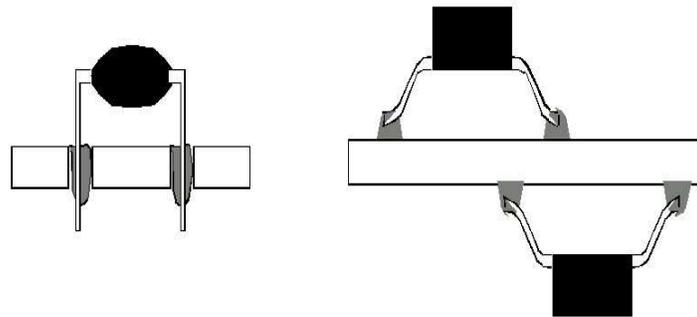
flip-chip bonding: (brevetto IBM, 1960)



un'operazione salda tutti i piedini).

si appoggia il die direttamente sullo stato di interconnessione tramite palline di metallo saldate (stagno). Vantaggi: migliore performance elettrica, accorcio i collegamenti, non serve fare singolarmente ogni connessione di I/O (con

Tecniche di interconnessione (Livello 2) --> dimensionare piastra-supporto

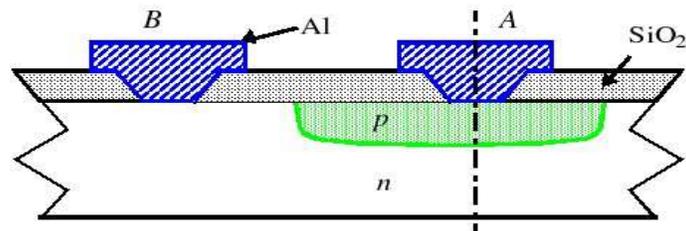


(a) Through-Hole Mounting

(b) Surface Mount

I dispositivi

IL DIODO



usati per:

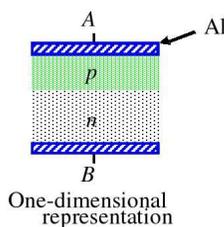
- 1) le loro capacità parassite in funzione della tensione;
- 2) proteggere i dispositivi di input contro le cariche statiche;

La regione di svuotamento: la giunzione PN è la più semplice dei dispositivi a semiconduttore:

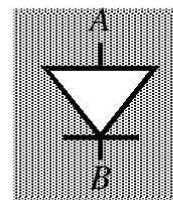
Giunzione brusca: - n donatori (fosforo, arsenio);
- P accettori (boro) + lacune;

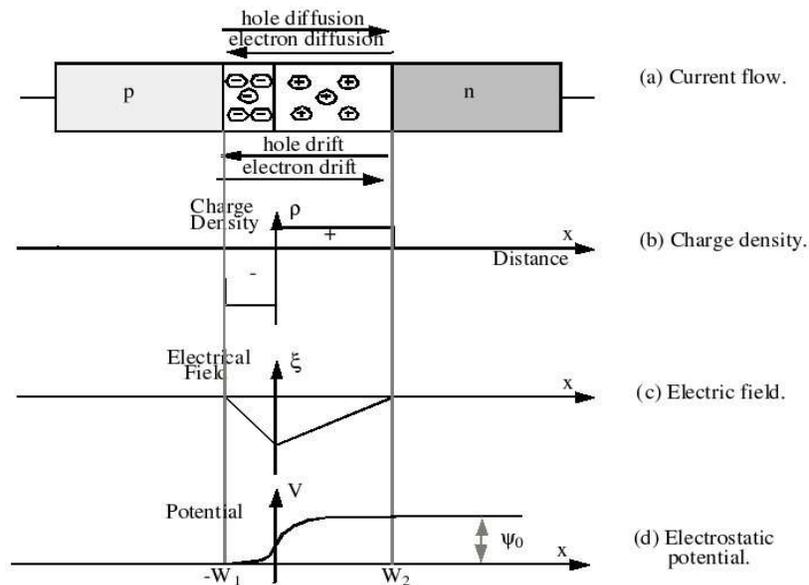
Semiconduttore --> caratteristiche intermedie tra quelle di un conduttore e quelle di un isolante;

Modello semplificato unidimensionale



Simbolo





La giunzione Pn causa una grande differenza di concentrazione di cariche al confine (molti e^- nella zona n e pochi nella zona P, viceversa per quanto riguarda le lacune). Tale gradiente provoca la diffusione $n \rightarrow P$ degli e^- , e $P \rightarrow n$ delle lacune. lo spostamento di tali cariche crea quindi una regione ionizzata negativamente dalla parte P e positivamente dalla parte n, vicino al confine: tale regione è detta di svuotamento. Qui si genera un campo elettrico $n \rightarrow P$ che tende a portare le lacune da P a n e gli elettroni da n a P, andando perciò a contrastare la diffusione. All'equilibrio le 2 correnti (quella dovuta alla diffusione e quella dovuta al campo) sono uguali ed opposte.

In tali condizioni di equilibrio, esiste un potenziale ϕ_0 (detto potenziale di built-in)

che vale: $\phi_0 = \phi_T \cdot \log_e \frac{N_A \cdot N_D}{n_i^2}$ dove:

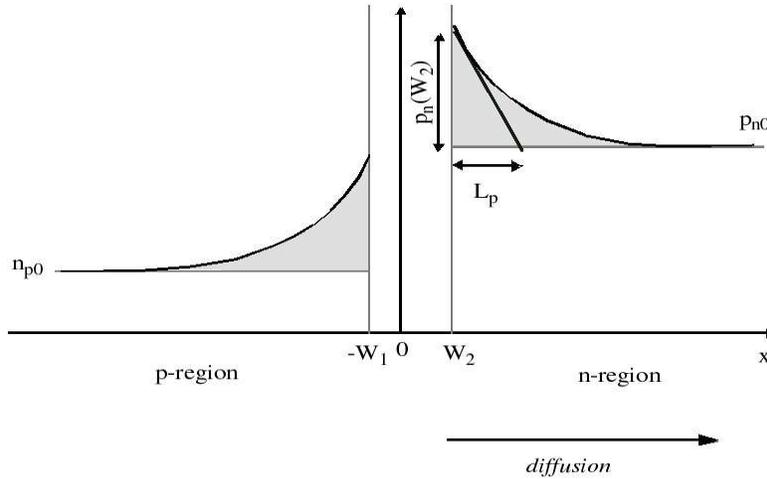
$$\phi_T = \text{potenziale termico} = \frac{K \cdot T}{q} = 26 \text{ mV a } 300^\circ \text{ K}$$

N_A = n° di lacune; N_D = n° di e^- ;

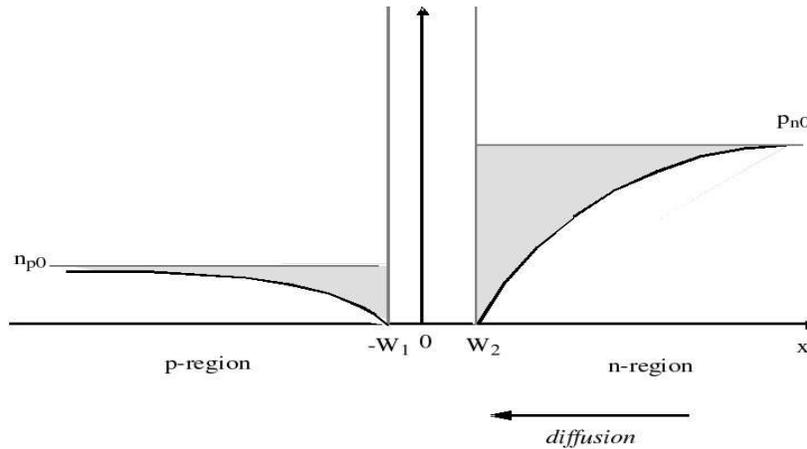
n_i^2 = concentrazione di portatori, in un semiconduttore puro $\rightarrow 1,5 \times 10^{10} \text{ cm}^{-3}$ per il Si

Comportamento statico

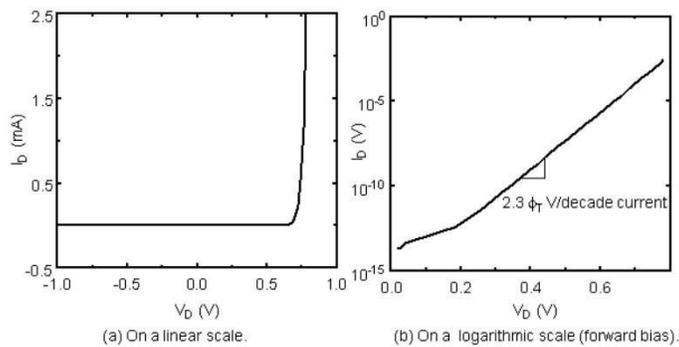
1) Corrente nel diodo: quando il diodo è in diretta (V_D diretto, ossia $V_p > V_n$) la componente maggiore è data dalla corrente di diffusione. (il grafico indica la concentrazione dei portatori nei pressi della regione di svuotamento in diretta).



Quando invece è in inversa (V inversa, ossia $V_n > V_p$), la componente maggiore è data dalla corrente di drift (dovuta al campo elettrico) [il grafico indica la concentrazione dei portatori nei pressi della regione di svuotamento inversa].



Da notare, la dipendenza esponenziale della corrente in funzione della tensione applicata.

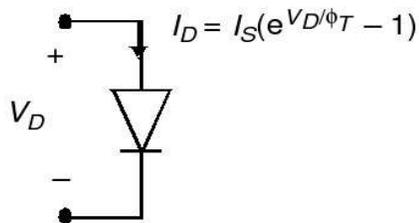


$$I_D = I_S \cdot \left(e^{\frac{V_D}{\phi_T}} - 1 \right) \text{ dove:}$$

I_S è la corrente di saturazione, proporzionale all'area del diodo in funzione dei livelli di drogaggio e della larghezza della regione neutra $\approx 10^{-17} \frac{A}{\mu \cdot m^2}$;

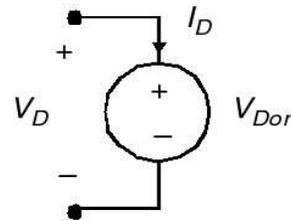
ϕ_T è la tensione termica;

2) Modelli per l'analisi



(a) Ideal diode model

svantaggi: è fortemente non lineare



(b) First-order diode model

si approssima $V_{Don} \approx 0,7V$

Comportamento dinamico

Determina la velocità massima alla quale il dispositivo può operare --> il comportamento dinamico del diodo dipende essenzialmente da quanto velocemente le cariche si possono muovere al suo interno.

In ogni IC digitale odierno, il diodo funziona sempre in inversa (funziona in diretta solamente in alcune particolari situazioni, peraltro da evitare).

PS: in inversa, il diodo ha un comportamento capacitivo (la giunzione tende a diventare un isolante).

1) Capacità della regione di svuotamento

In diretta, la barriera di potenziale è ridotta --> equivale a un restringimento della zona di svuotamento.

In inversa, la barriera di potenziale aumenta --> equivale ad un allargamento della zona di svuotamento;

NB: il numero totale di donatori ed accettori deve essere numericamente uguale (il diodo deve rimanere neutro).

$$Q_j = A_D \cdot \sqrt{2 \cdot \epsilon \cdot q \cdot \frac{N_A \cdot N_D}{N_A + N_D} \cdot (\phi_0 - V_D)}$$

che è la carica nella regione di svuotamento,

dove: $\epsilon = \epsilon_{Si}$ che è la permittività del Si = $1,053 \times 10^{-10}$ F/m;

$$W_j = \sqrt{\frac{2 \cdot \epsilon}{q} \cdot \frac{N_A + N_D}{N_A \cdot N_D} \cdot (\phi_0 - V_D)} = W_2 - W_1 \quad \text{che è la larghezza;}$$

$$E_j = \sqrt{\frac{2 \cdot q}{\epsilon} \cdot \frac{N_A \cdot N_D}{N_A + N_D} \cdot (\phi_0 - V_D)} = W_2 - W_1 \quad \text{che risulta essere il campo elettrico max;}$$

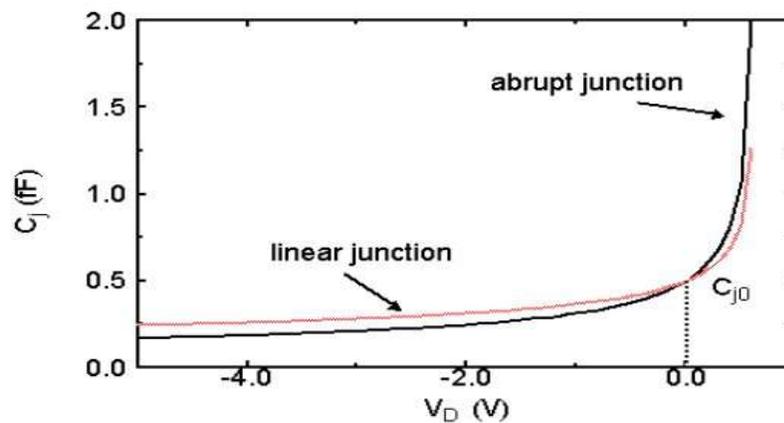
$$C_j = \frac{d\phi_j}{dV_D} = A_D \sqrt{\frac{\epsilon \cdot q}{2} \cdot \frac{N_A \cdot N_D}{N_A + N_D} \cdot (\phi_0 - V_D)^{-1}} \quad \text{che è la capacità;}$$

Con un'approssimazione si ottiene: $C_j = \frac{C_{j_0}}{\left(1 - \frac{V_D}{\phi_0}\right)^m}$, C_{j_0} è in assenza di

polarizzazione (cioè è in funzione solo delle caratteristiche fisiche del diodo), dove:

a) $m = 1/2$ quando il diodo si comporta come una giunzione brusca (ossia c'è una transizione istantanea tra il materiale P e il materiale n);

b) $m = 1/3$ quando il diodo ha una funzione lineare (come nel caso degli odierni dispositivi, in cui il passaggio è graduale, grazie al drogaggio per compensazione)



2) Capacità della regione di svuotamento a regime

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{\phi_j \cdot V_{high} - \phi_j \cdot V_{low}}{V_{high} - V_{low}} = K_{eq} \cdot C_{j_0}$$

3) Capacità di diffusione: vicino alla regione di svuotamento, c'è una capacità maggiore in quanto sono presenti un maggior numero di cariche;

Effetti secondari

Nel diodo reale, le zone "neutre" hanno una certa caduta di potenziale (resistenza), e non tutte le tensioni appaiono direttamente sulla giunzione...;

C'è inoltre la possibilità di un breakdown e valanga. Se si raggiunge una certa tensione di breakdown, la corrente in inversa aumenta vertiginosamente, questa va a sua volta ad aumentare l'ampiezza del campo elettrico nella giunzione, e così facendo la velocità dei portatori. Questi ad un certo valore di E ($E_{critico}$), ottengono infine abbastanza energia per

creare, nelle collisioni con gli atomi di silicio, altri portatori, in un processo (appunto) a valanga. Mantenendo troppo tempo il diodo in queste condizioni, si incorre a danni permanenti alla sua struttura (a causa dell'aumento della temperatura).

PS: *limite termico di funzionamento del silicio*: quando il numero di questi portatori intrinseci (creati dalla temperatura) è uguale al numero di portatori dati dal drogaggio, non riesco più a distinguere la zona P dalla zona n, dato che non c'è più la barriera di potenziale.

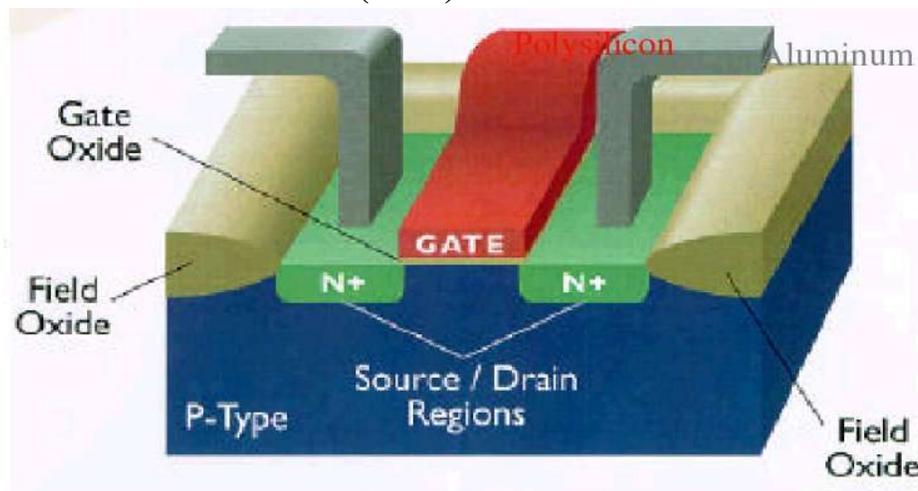
In altre parole, non posso applicare una tensione qualsiasi al diodo, infatti nel caso particolare del Si, si ha:

- breakdown a valanga con $V > 5V$;
- effetto tunnel (in altre condizioni): in determinate condizioni, si ha una probabilità non nulla che una particella passi attraverso una barriera di potenziale anche se la sua energia è minore;

La temperature infine influenza il funzionamento del diodo in altri 2 modi:

- 1) ϕ_T dipende linearmente dalla temperatura, e se aumento ϕ_T diminuisce la corrente;
- 2) Anche I_S dipende dalla temperatura (raddoppia circa per ogni aumento di $8^\circ C$);

Il Transistor MOS(FET) [metal-oxide-semiconductor(field-effect-transistor)]



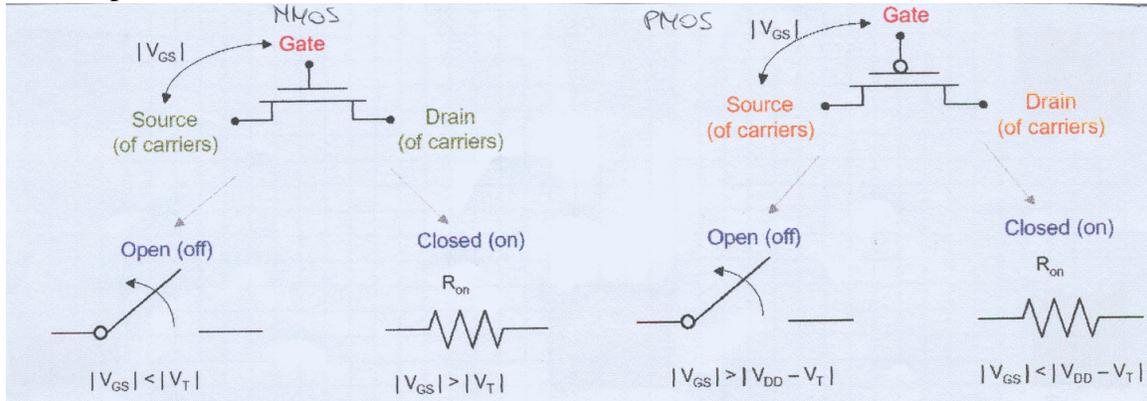
Elemento fondamentale dei IC: è un dispositivo a 4 parametri: Gate, Drain, Source e il suo stesso Corpo (per modulare alcune sue caratteristiche);

Transistor:

- BJT (bipolari): definito dalle condizioni di polarizzazione delle giunzioni;
- MOS: funzioni sempre polarizzate in inversa;

a) NMOS --> S/D = n, substrate = P, portatori = e^- , in un canale n (a differenza dei diodi, nei quali entrambe lacune ed e^- erano portatori);

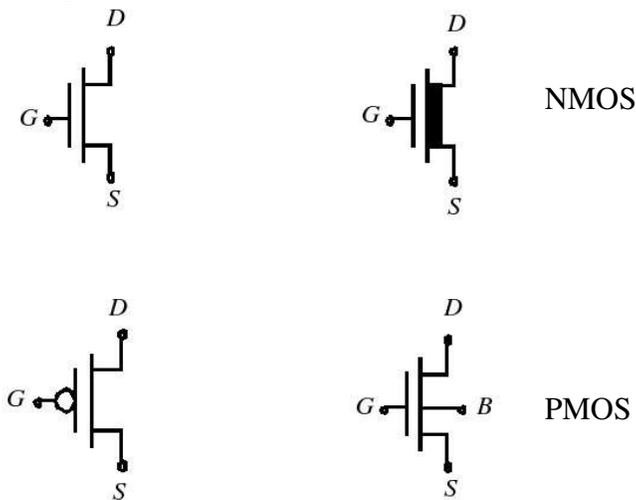
- b) PMOS --> S/D = P, sub = n, portatori = lacune, in un canale P;
 c) CMOS (complementari MOS) --> PMOS e NMOS in serie;
 Il MOS può essere considerato un "interruttore":



La conduttività del canale è governata dal gate: più alto è ΔV_{GS} minore è la resistenza del canale e quindi maggiore è la corrente che passa. Se $V_{GS} < V_T$, il canale è chiuso, dove V_T è il potenziale di soglia.

PS: la lunghezza effettiva del segnale è un pò minore della lunghezza nel layout (inevitabile a causa della tolleranza nei processi tecnologici) --> effetto sul comportamento dinamico: l'iniziale lentezza dei mos era causata da questi effetti parassiti (diffusione laterale).

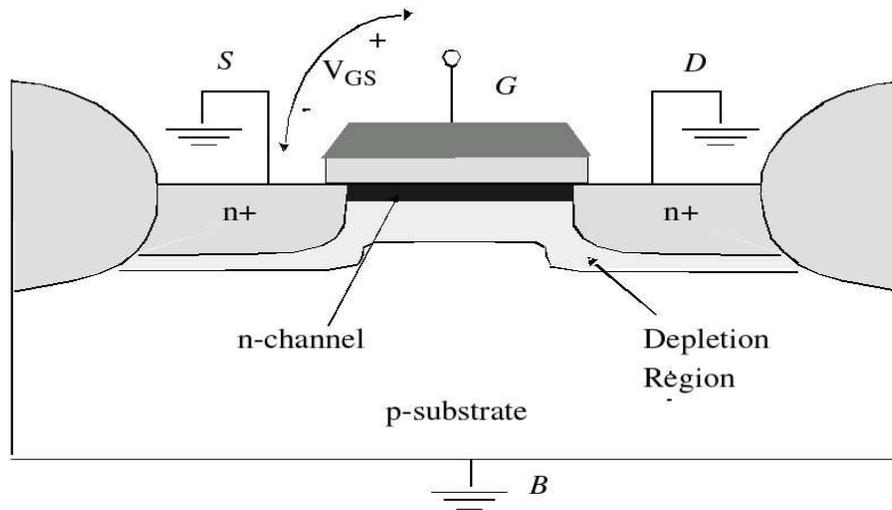
Simbologia dei mos



Comportamento Statico

Concetto di tensione di soglia ed effetto body

Tensione di soglia: tensione che mi consente un passaggio significativo di corrente nel canale (dipende da fattori tecnologici e fisici del dispositivo)



$V_{GS}=0$, tra D e S non passa corrente (resistenza ∞). Se ora applico una tensione, gate e substrato diventano una specie di condensatore, con l'ossido del gate come dielettrico. Il potenziale positivo del gate carica l'accumulo di cariche positive sulla parte del gate e negative sulla parte del substrato. In particolare, il secondo inizialmente si manifesta respingendo le lacune: in questo modo si forma una piccola regione di svuotamento, per la quale si possono usare le formule di W_d e Q_d definite per il diodo,

essendo una situazione analoga: $W_q = \sqrt{\frac{2 \cdot \epsilon \cdot \phi}{q \cdot N_A}} \quad \sqrt{2 \cdot \phi \cdot \epsilon \cdot q \cdot N_A}$

Con N_A drogaggio del substrato del substrato (concentrazione degli accettori);

ϕ Tensione lungo lo stato svuotato;

Al crescere di V_{GS} , inizia un fenomeno chiamato **INVERSIONE FORTE**, che si concretizza quando il potenziale raggiunge il doppio del potenziale di Fermi:

$$\phi_F = \phi_T \log_e \left(\frac{N_A}{n_i} \right)$$

Tale fenomeno porta all'instaurazione di uno stato n che (da ora) aumenta all'aumentare di V_{GS} (viene fornito di e^- dal gate):

$$Q_{B_0} = \sqrt{|2 \cdot q \cdot N_A \cdot \epsilon - 2 \cdot \phi_F|}$$

Il valore di tensione al quale avviene tale inversione è detto: **TENSIONE DI SOGLIA**:

$V_T = V_{T_0} + y \cdot \sqrt{|-2 \cdot \phi_F + V_{SB}|} - \sqrt{|-2 \cdot \phi_F|}$ --> formula semplificata in cui V_T è solo in funzione dei parametri costruttivi.

V_T con $V_{SB} = 0$ (B: bulk [substrato])

$$y = \frac{\sqrt{2 \cdot q \cdot \epsilon \cdot N_A}}{C_{OX}} \quad \text{che è il coefficiente di effetto body (dovuto alla tecnologia)}$$

PS: $V_T > 0$ per gli nmos e $V_T < 0$ per i pmos

Senza semplificare risulta: $V_T = \phi_{ms} - 2 \cdot \phi_F - \frac{Q_B}{C_{OX}} - \frac{Q_{SS}}{C_{OX}} - \frac{Q_I}{C_{OX}}$ dove:

$\frac{Q_B}{C_{OX}}$ è la carica nella zona svuotata;

$\frac{Q_{SS}}{C_{OX}}$ è la carica superficiale;

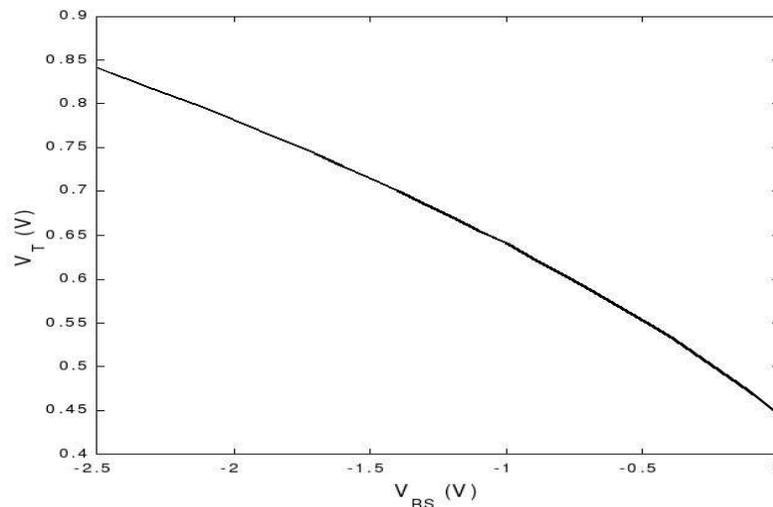
$\frac{Q_I}{C_{OX}}$ carica dovuta all'impiantazione logica --> tale processo permette di ottenere una

V_T molto più precisa --> i mos naturali avrebbero una soglia molto vicina a ϕ (a causa di ϕ_{ms}) e quindi sarebbero difficili da controllare.

$$V_{T_0} = \phi_{ms} - 2 \cdot \phi_F - \frac{Q_{B_0}}{C_{OX}} - \frac{Q_{SS}}{C_{OX}} - \frac{Q_I}{C_{OX}}$$

EFFETTO BODY

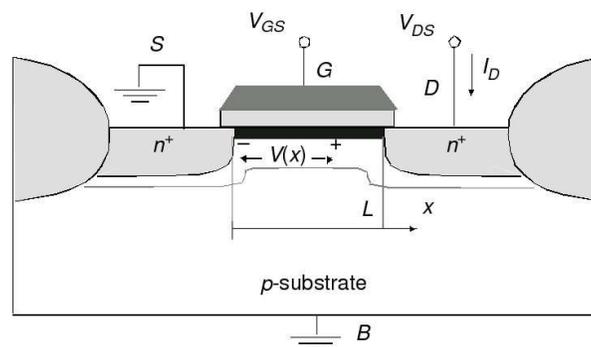
esprime gli effetti sulla V_T al variare di V_{SB} (tensione tra S e il substrato)



Aumentando V_{SB} , si verifica un allargamento della zona svuotata dal lato del source e del drain --> ho più cariche da equilibrare, quindi devo aumentare V_T (quindi, polarizzando in modi diversi il substrato si possono ottenere V_T diversi)

PS: $V_{SB} < 0,6$ perchè altrimenti il mos va in diretta

Transistor in modalità lineare



Dato $V_{SB} > V_T$ e un piccolo V_{DS} . Tale tensione provoca un passaggio di corrente da D a S ($I_D = f(V_{SB}, V_{DS})$).

Data una tensione generica $V(x)$ nel canale, si ha che:

$$Q_j(x) = -C_{OX} \cdot (V_{GS} - V(x) - V_T) \quad \text{dove:}$$

$$C_{OX} = \frac{\epsilon_{OX}}{T_{OX}} \quad \epsilon_{OX} \rightarrow 3,5 \times 10^{-11} \text{ F/m} \quad \text{ed } T_{OX} \rightarrow \text{spessore dell'ossido}$$

$$I_D = -V_n(x) \cdot Q_j(x) \cdot W \quad \text{qui la velocità dei portatori è } V_n(x) = -\mu_n \cdot E(x) = \mu_n \cdot \frac{dv}{dx}$$

qui μ_n è la mobilità dei portatori mentre $E(x)$ è il campo elettrico. Sostituendo ed integrando, si ottiene inoltre un parametro K_n detto: PARAMETRO DEL PROCESSO DI TRANSCONDUITTANZA

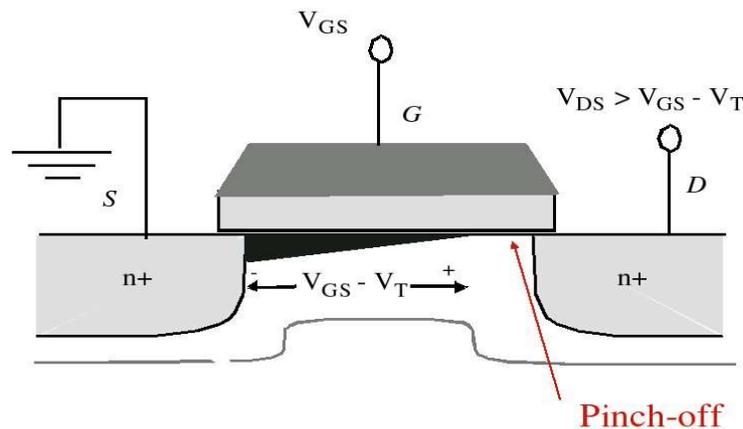
$$K_n = \mu_n \cdot C_{OX} = \mu_n \cdot \frac{\epsilon_{OX}}{T_{OX}}$$

che moltiplicato al rapporto W/L dà il guadagno K_n^1 del dispositivo.

$$I_D = K_n \cdot \frac{W}{L} \cdot \left[(V_{GS} - V_T) \cdot V_{DS} - \frac{V_{DS}^2}{2} \right]$$

Tale equazione opera (con V_{DS} piccoli) in una regione del mos detta RESISTIVA o LINEARE, nella quale è presente un canale conduttivo continuo tra D e S e una dipendenza lineare tra V_{DS} e I_D .

Transistor in modalità di saturazione



Aumentando V_{DS} fino a rendere $V_{DS} > V_{GS} - V_T$, si ha che la zona svuotata tende ad allargarsi ancora all'interno del canale, rendendo i portatori disponibili insufficienti a garantire il canale n \rightarrow PINCH-OFF: si crea una zona in cui non c'è più canale, si dice in questo caso che la corrente satura (non può più crescere).

In tale situazione, il ΔV tra il pinch-off e source rimane fisso a $V_{GS} - V_T$, e la corrente rimane costante a:

$$I_D = \frac{K_n}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2 \quad \rightarrow \text{non è più in funzione di } V_{DS} \text{ (si è strozzato il canale, non$$

c'è più continuità).

PS: la corrente non va a 0 perchè, una volta arrivati alla fine della zona svuotata, i portatori (e^-) trovano un campo elettrico favorevole al loro passaggio (più sul lato del D)

Modulazione larghezza-canale

I_D dipende inversamente dalla larghezza del canale, quindi più è stretto maggiore è la corrente che posso estrarre. Aumentando V_{DS} , il punto di pinch-off si sposta verso S e quindi il canale effettivo si stringe, per questo motivo, la corrente non satura mai completamente (continua sempre ad aumentare un pò).

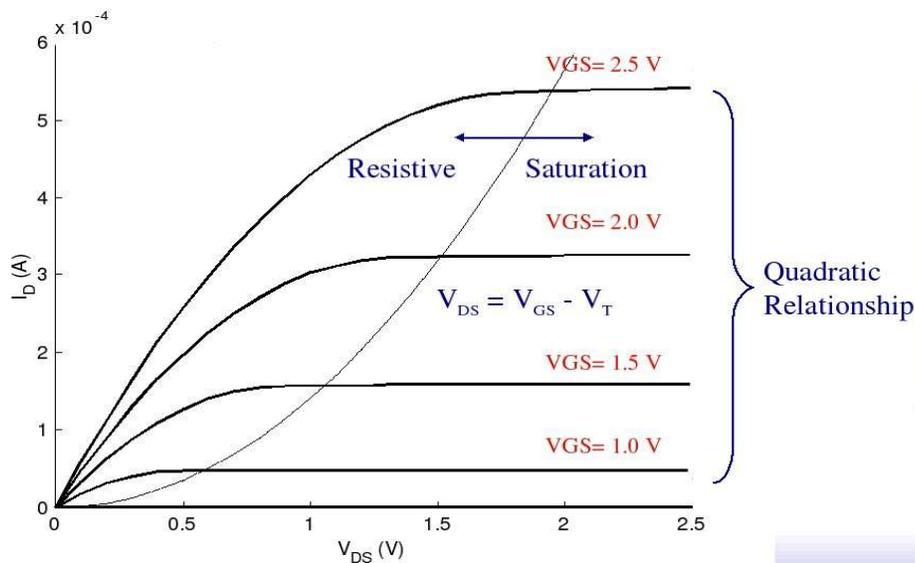
Inoltre, non è del tutto vero che I_D è indipendente da V_{DS} , infatti il valore reale I_D^1 è:

$$I_D^1 = I_D \cdot (1 + \lambda \cdot V_{DS})$$

dove λ risulta essere un parametro empirico detto MODULAZIONE LARGHEZZA-CANALE, che varia in base all'inverso della larghezza (in modo però non molto preciso)

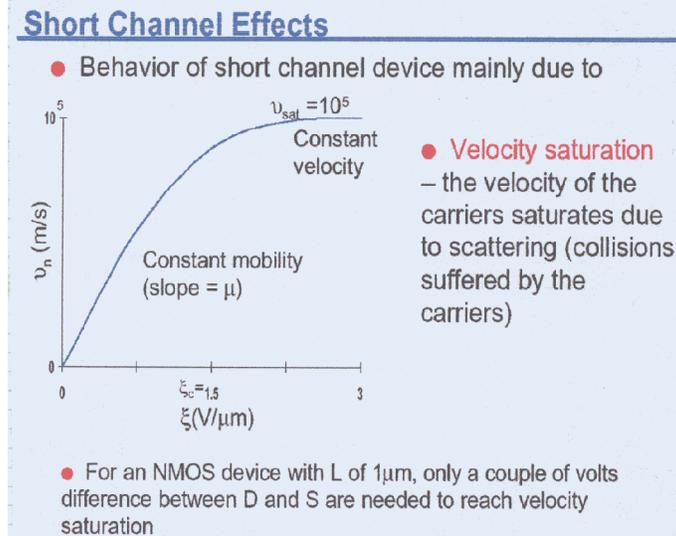
NB: quindi in definitiva, per un fissato V_{DS} , e $V_{GS} > V_T$, I_D è funzione di:

- 1) distanza D – S (L);
- 2) lunghezza del canale (W);
- 3) V_T ;
- 4) spessore ossido (T_{OX});
- 5) permittività dell'ossido (ϵ_{OX});
- 6) mobilità dei portatori (μ_n, μ_p);



Effetti di canale corto

L'utilizzo di un canale sempre più corto fa cambiare il comportamento del mos, soprattutto a causa della (conseguente) saturazione della velocità. La velocità è proporzionale al campo elettrico e alla mobilità, ma questo cessa di valere a valori di E molto alti -->



– In condizioni di linearità ($V_{DS} < V_{GS} - V_T$)

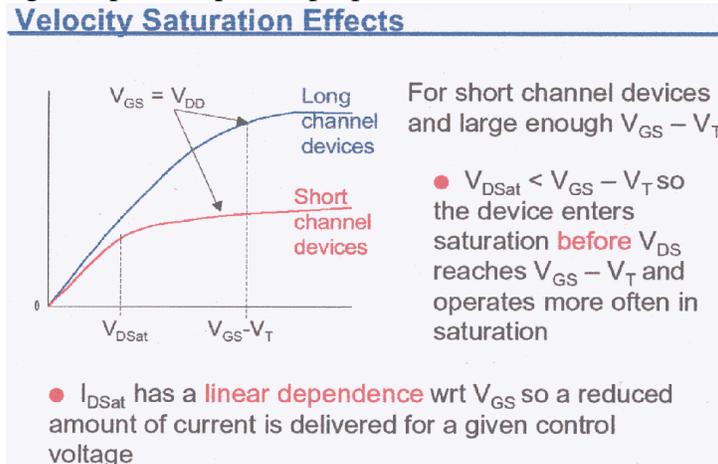
$$I_D = K(V_{DS}) \cdot K_n \cdot \frac{W}{L} \cdot \left[(V_{GS} - V_T) \cdot V_{DS} - \frac{V_{DS}^2}{2} \right]$$

– In condizioni di saturazione ($V_{DS} > V_{GS} - V_T$)

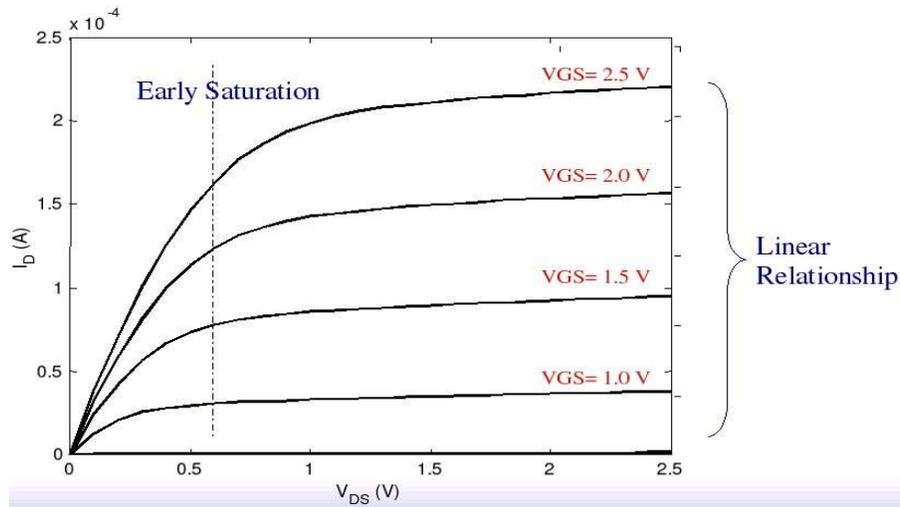
$$I_D = K(V_{DSat}) \cdot K_n \cdot \frac{W}{L} \cdot \left[(V_{GS} - V_T) \cdot V_{DSat} - \frac{V_{DSat}^2}{2} \right] \quad \text{con:}$$

$$K(V) = \frac{1}{1 + \frac{V}{E_0 \cdot L}} \quad \text{--> misura del grado di saturazione della velocità}$$

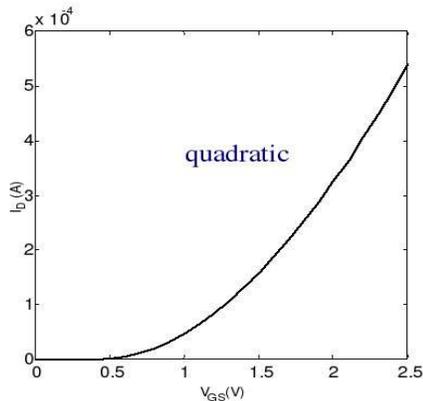
NB: nei dispositivi corti, $V_{DSat} < V_{GS} - V_T$ --> entrano in saturazione prima dei dispositivi con canale più lungo (e spesso separano proprio in condizioni di saturazione).



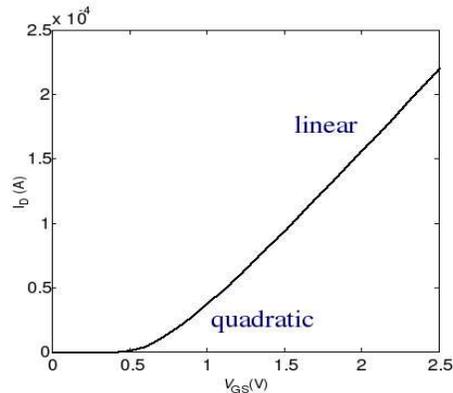
Inoltre, I_{DSat} è linearmente dipendente da V_{GS} (a differenza dei dispositivi a canale lungo, nei quali, tale dipendenza è quadratica).



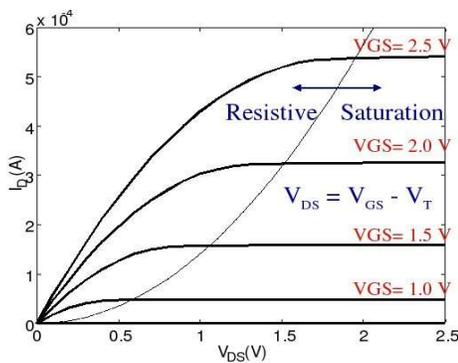
Confronto tra i2 dispositivi (a canale corto e lungo)



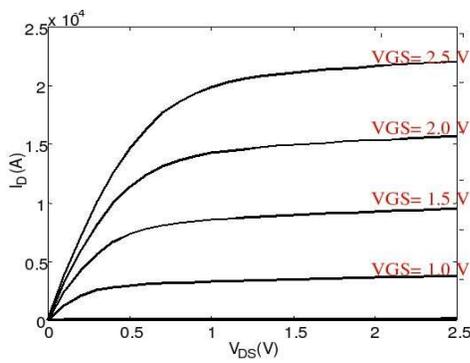
Long Channel



Short Channel



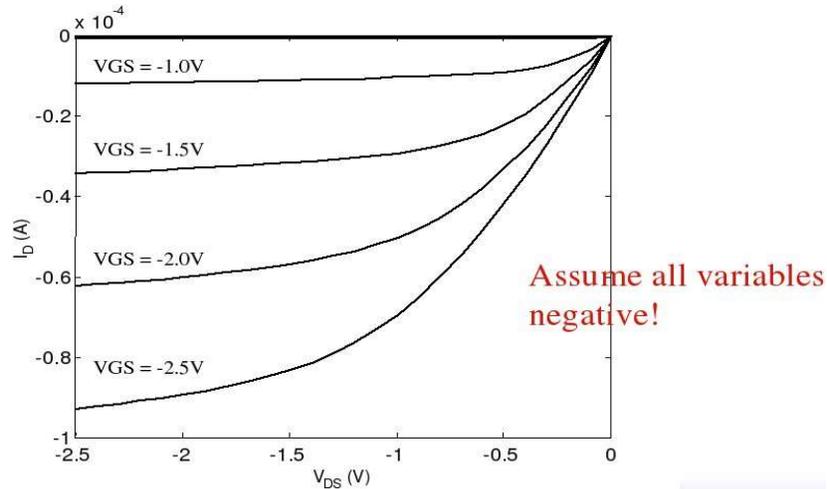
Long Channel



Short Channel

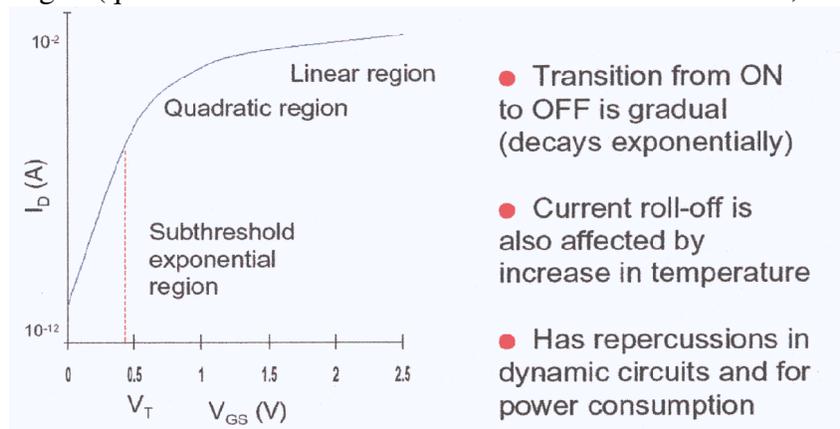
NB: la saturazione della corrente e della velocità sono l'effetto indesiderato che fa perdere i vantaggi che deriverebbero dall'impiego (rispettivamente) di una tensione V_{DS} maggiore della riduzione della lunghezza del canale. La riduzione delle dimensioni la paga dunque nelle caratteristiche del transistor (perdita in termini di controllo del deposito), ma non tanto da andare in "negativo", in quanto i vantaggi ricavati dall'aumento della corrente e dall'integrazione sono maggiori.

PS: per il dispositivo PMOS, naturalmente, valgono le stesse conclusioni, solamente con tutte le tensioni e le correnti invertite, cambia solo il fatto che, per ottenere la stessa corrente, devo farlo 2 volte e mezzo più largo, perchè le lacune sono 2,5 volte più veloci degli e⁻.



Conduzione Sottosoglia

La corrente non v^a istantaneamente a 0 con $V_{GS} = V_T$ (non si comporta quindi come un interruttore ideale), in quanto il mos conduce un p^o di corrente anche a potenziali minori di quello di soglia (quindi la transizione tra on e off del mos non \grave{e} brusca, ma graduale).



$$I_D = I_S \cdot e^{\frac{V_{GS}}{n \cdot K \cdot T}} \cdot \left(1 - e^{-\frac{V_{SD}}{K \cdot T}}\right) \quad \text{--> dipendenza esponenziale della } V_{GS}$$

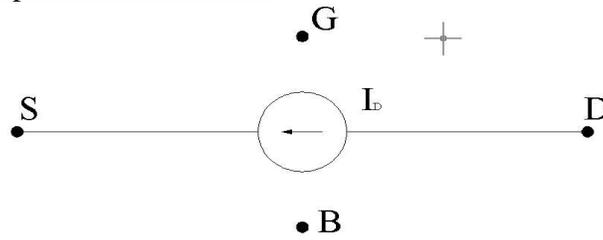
Tale effetto \grave{e} indesiderato, per cui si cerca di avere una transazione di corrente pi^u veloce possibile. Per questo motivo, la pendenza della curva nella regione sottosoglia \grave{e} un indice

della qualit^a del dispositivo: $S = n \cdot \left(\frac{K \cdot T}{q}\right) \cdot \log_e(10)$ --> fattore di pendenza, dove $n > 1$

nei dispositivi attuali.

PS: tale effetto \grave{e} favorito dalla diminuzione delle dimensioni e dall'aumento della temperatura.

Modello di transistor per l'analisi manuale



1) con I_D :

- 0 per $V_{GT} < 0$

$$- K_n \cdot \frac{W}{L} \cdot \left[(V_{GS} - V_T) \cdot V_{min} - \frac{V_{min}^2}{2} \right] \cdot (1 + \lambda \cdot V_{DS}) \quad \text{per } V_{GS} - V_T \geq 0$$

$$- V_{min} = \min(V_{GS} - V_T, V_{DS}, V_{Dsat}) \quad V_{GT} = V_{GS} - V_T$$

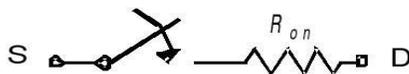
$$V_T = V_{T0} + \gamma \cdot (\sqrt{|-2 \cdot \phi_F + V_{SB}|} - \sqrt{|-2 \cdot \phi_F|})$$

Table 3.2 Parameters for manual model of generic 0.25 μm CMOS process (minimum length device).

	V_{T0} (V)	γ ($\text{V}^{0.5}$)	V_{Dsat} (V)	k' (A/V^2)	λ (V^{-1})
NMOS	0.43	0.4	0.63	115×10^{-6}	0.06
PMOS	-0.4	-0.4	-1	-30×10^{-6}	-0.1

2)

- $R_{off} = \infty, R_{on} = \text{finita}$;



- $R \propto \frac{W}{L}$;

- per $V_{DD} \gg V_T + \frac{V_{Dsat}}{2}$, R virtualmente

indipendente da V_{DD} ;

- quando $V_{DD} \rightarrow V_T$, R_{on} aumenta velocemente;

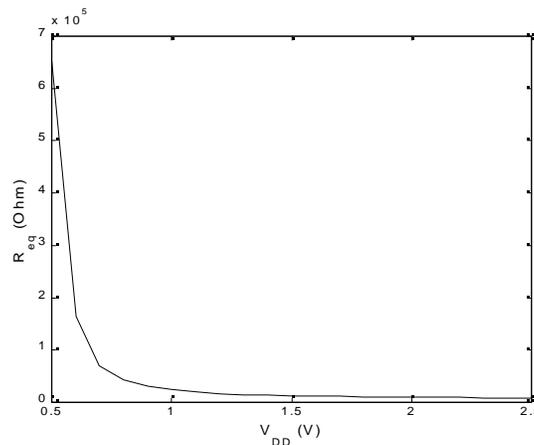
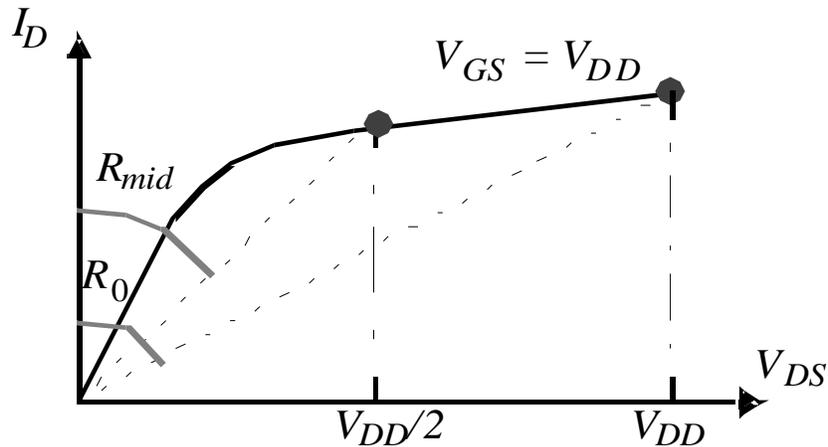


Table 3.3 Equivalent resistance R_{eq} ($W/L = 1$) of NMOS and PMOS transistors in 0.25 μm CMOS process (with $L = L_{min}$). For larger devices, divide R_{eq} by W/L .

V_{DD} (V)	1	1.5	2	2.5
NMOS (k Ω)	35	19	15	13
PMOS (k Ω)	115	55	38	31

--> NMOS ha valori di resistenza più bassi;

per $V_{GS} = V_{DD}$, $V_{DS} = V_{DD}$ --> $\frac{V_{DD}}{2}$ --> situazione particolare che ci interessa, in quanto $\frac{V_{DD}}{2}$ è il valore usato nell'analisi del delay nella commutazione.



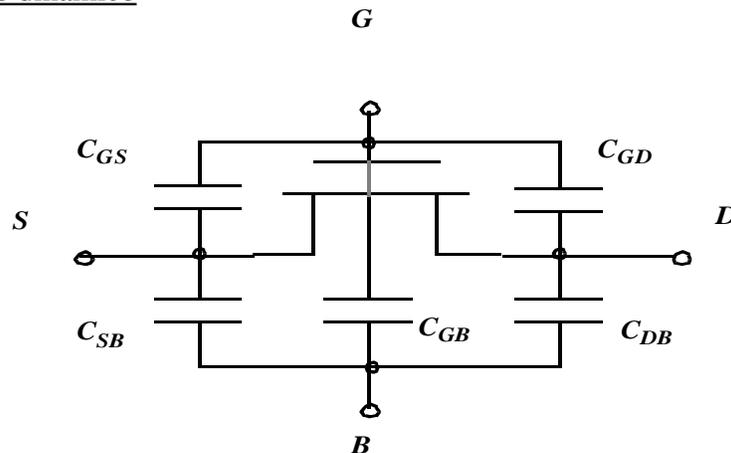
R_{eq} = resistenza media nell'intervallo $[\frac{V_{DD}}{2}, V_{DD}]$

$$R_{eq} \simeq \frac{3}{4} \cdot \frac{V_{DD}}{I_{Dsat}} \cdot \left(1 - \frac{5}{6} \cdot \lambda \cdot V_{DD}\right)$$

tale resistenza è dell'ordine dei K Ω , per cui riveste una

notevole importanza nell'analisi della velocità con il quale il gate esegue la commutazione.

Comportamento dinamico



tale comportamento è funzione del tempo impiegato dal gate a commutare, e quindi caricare/scaricare le sue capacità intrinseche e quelle dovute alle linee di

interconnessione. Tali capacità sono originate:

1) dalla struttura del mos; -->

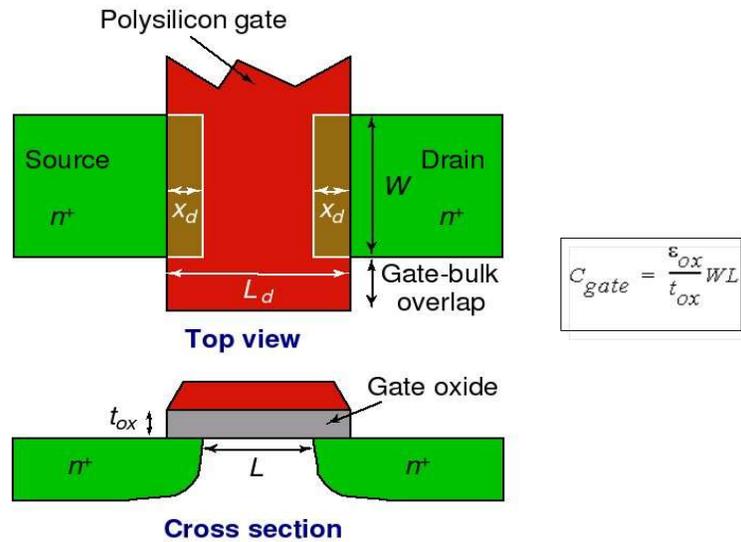
2) dalla carica del canale; --> $C_{gate} = \frac{\epsilon_{ox}}{T_{ox}} \cdot W \cdot L$

3) dalle regioni di svuotamento nelle giunzioni Pn in inversa (tra D e S);

PS: il gate è separato dal canale da un sottile strato di ossido di capacità $C_{ox} = \frac{\epsilon_{ox}}{T_{ox}}$

(strato che deve essere più largo o più sottile possibile). La capacità dell'intera struttura è costituita dai punti 1 e 2.

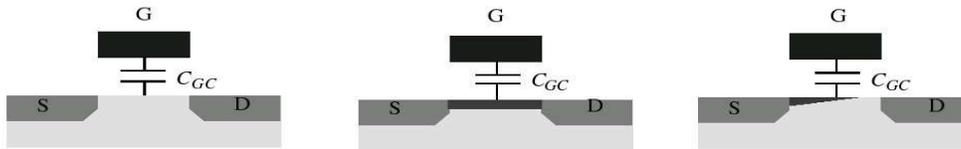
1) Capacità dovuta alla struttura del gate: in un transistor reale, D e S non terminano esattamente quando inizia il gate, ma (a causa della diffusione laterale) tendono ad estendersi anche sotto di esso, in modo tale che $L < L_d$.



$C_{GS_0} = C_{GD_0} = C_{ox} \cdot X_d \cdot W$ --> capacità di sovrapposizione

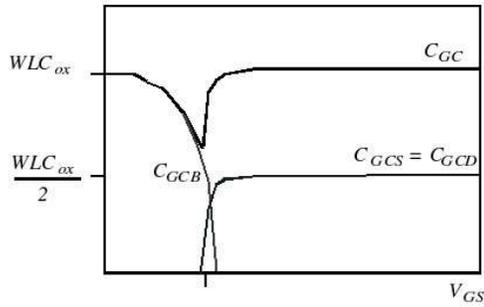
2) Capacità del canale: (C_{GC} --> gate-to-channel) è quella più significativa

C_{GC} è formata da 3 componenti: C_{GCS} , C_{GCD} , C_{GCB} .

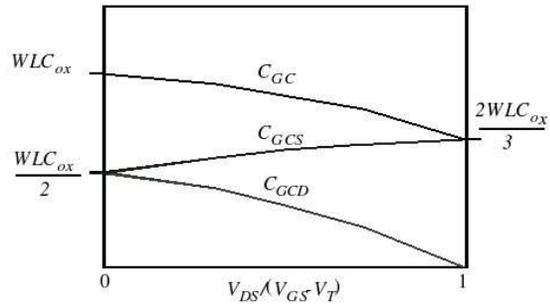


Operation Region	C_{gb}	C_{gs}	C_{gd}
Cutoff	$C_{ox}WL_{eff}$	0	0
Triode	0	$C_{ox}WL_{eff}/2$	$C_{ox}WL_{eff}/2$
Saturation	0	$(2/3)C_{ox}WL_{eff}$	0

Most important regions in digital design: saturation and cut-off



Capacitance as a function of VGS
(with VDS = 0)

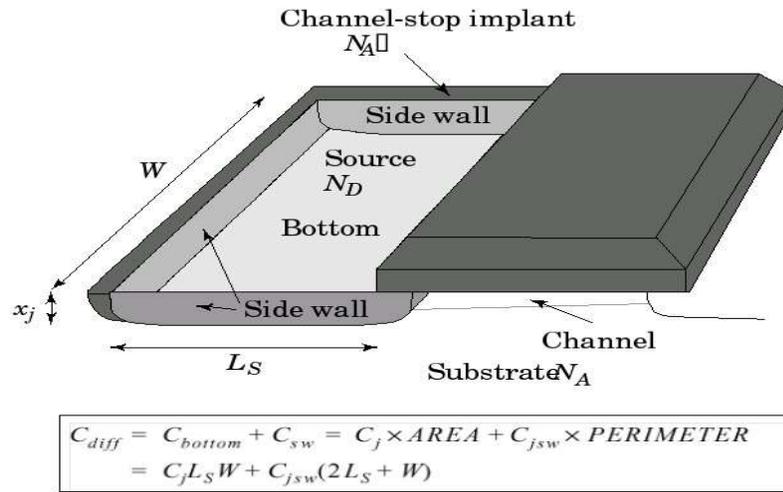


Capacitance as a function of the
degree of saturation

a) nel primo grafico ($V_{GS} = 0$), quando il transistor è spento, non c'è canale, e la capacità $e^- = C_{OX} \cdot W \cdot L$. Man mano che aumenta V_{GS} , si forma una regione di svuotamento che va a ridurre tale capacità, fino a quando il canale è formato: $C_{GCB} = 0$, e la capacità totale si ripartisce equivalentemente tra S e D (perchè il transistor è resistivo);

b) nel secondo grafico, a transistor on, la sua capacità dipende dal gate di saturazione $\frac{V_{DS}}{V_{GS} - V_T}$ --> variazione non lineare e dipendente della tensione.

3) Capacità di giunzione:



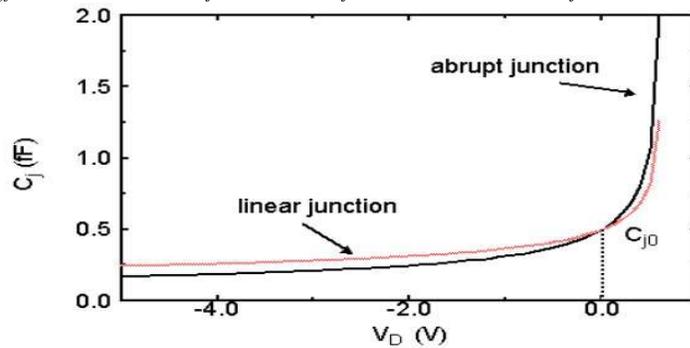
2 conduttori:

- uno legato alla piattaforma della giunzione (bottom), tra la parte inferiore della giunzione e il substrato (brusca) $C_b = C_j \cdot W \cdot L_S$;

- una legata alle pareti laterali (graduale) $C_{sw} = C_{jsw}^1 \cdot X_j \cdot (W + 2 \cdot L_S)$ dove $C_{jsw} = C_{jsw}^1 \cdot X_j$

PS: minore superficie in gioco, ma confina con zone molto più drogate (capacità specifica più elevata)

In totale: $C_{diff} = C_b + C_{SW} = C_j \cdot Area + C_{jsw} \cdot Perimetro = C_j \cdot L_S \cdot W + C_{jsw} \cdot (2 \cdot L_S + W)$



$$C_j = \frac{C_{j0}}{(1 - V_D / \phi_0)^m}$$

m = 0.5: abrupt junction
m = 0.33: linear junction

PS: si possono anche sostituire queste capacità non lineari con delle capacità lineari (a regime)

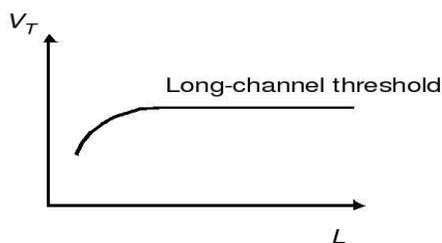
$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j \cdot V_{high} - Q_j \cdot V_{low}}{V_{high} - V_{low}} = K_{eq} \cdot C_{j0}$$

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low}) \cdot (1 - m)} \cdot [(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}]$$

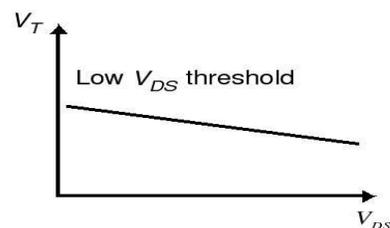
	C_{ox} (fF/ μm^2)	C_o (fF/ μm)	C_j (fF/ μm^2)	m_j	ϕ_b (V)	C_{jsw} (fF/ μm)	m_{jsw}	ϕ_{b0w} (V)
NMOS	6	0.31	2	0.5	0.9	0.28	0.44	0.9
PMOS	6	0.27	1.9	0.48	0.9	0.22	0.32	0.9

Effetti secondari a piccole dimensioni

Variazioni sottosoglia:



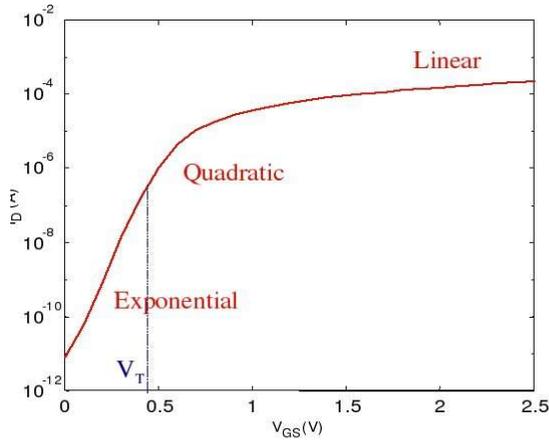
Threshold as a function of the length (for low V_{DS})



Drain-induced barrier lowering (for low L)

- V_T in teoria è in funzione solo delle caratteristiche tecnologiche e di V_{SB} , quindi costante. Tuttavia, alla diminuzione delle dimensioni, V_T diventa in funzione di L , W e V_{DS} ;
- V_T diminuisce alla diminuzione di L (a causa dello svuotamento);
- Aumentando V_{DS} , V_T diminuisce (effetto DIBL --> V_T in funzione delle tensioni a cui opera il transistor). Per V_{DS} abbastanza elevati, le 2 regioni di D e S possono anche entrare in contatto --> punch.through

Conduzione sottosoglia



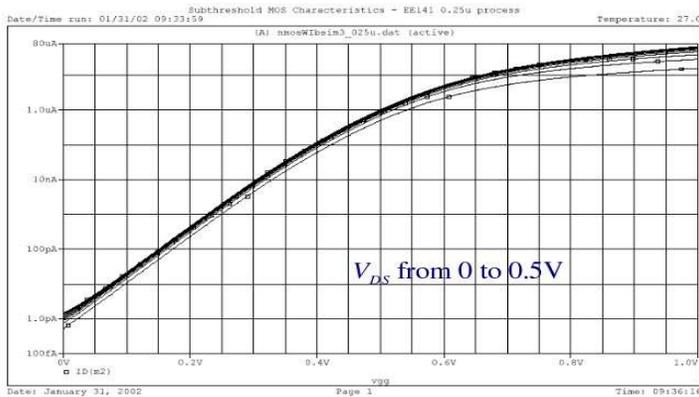
The Slope Factor

$$I_D \sim I_0 e^{\frac{qV_{GS}}{nKT}}, \quad n = 1 + \frac{C_D}{C_{ox}}$$

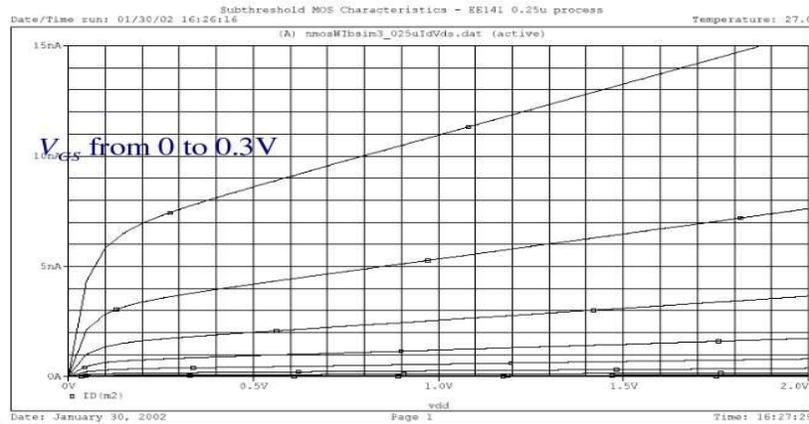
S is ΔV_{GS} for $I_{D2}/I_{D1} = 10$

$$S = n \left(\frac{kT}{q} \right) \ln(10)$$

Typical values for S :
60 .. 100 mV/decade



$$I_D = I_0 \cdot e^{\frac{q \cdot V_{GS}}{n \cdot K \cdot T}} \cdot \left(1 - e^{-\frac{q \cdot V_{DS}}{K \cdot T}} \right)$$

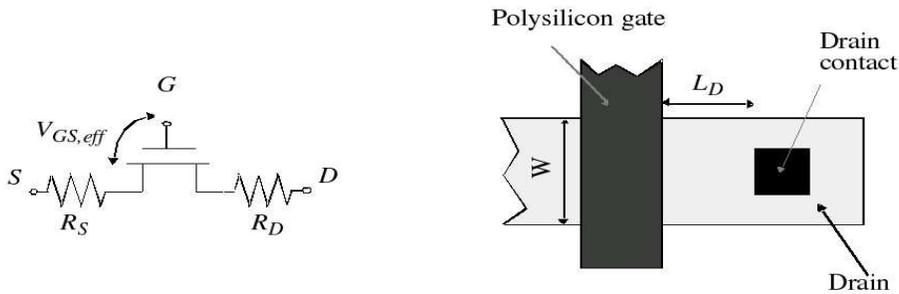


$$I_D = I_0 \cdot e^{\frac{q \cdot V_{GS}}{n \cdot K \cdot T}} \cdot \left(1 - e^{-\frac{q \cdot V_{DS}}{K \cdot T}} \right) \cdot (1 + \lambda \cdot V_{DS})$$

NB:

- Inversion forte: $V_{GS} > V_T$
 - Lineare (resistiva): $V_{DS} < V_{Dsat}$
 - Saturata (corrente costante): $V_{DS} \geq V_{Dsat}$
- Inversione debole: $V_{GS} \leq V_T$
 - esponenziale in V_{GS} , con dipendenza lineare in V_{DS} ,

Resistenze parassite



Resistenze in serie a S e D (effetto indesiderato amplificato dalla miniaturizzazione) dovuto al contatto metallico di conduzione

$$R_{S(D)} = \frac{L_{S(D)}}{W} \cdot R_* + R_C \quad \text{dove} \quad R_* = \frac{\rho}{X_i} \quad \text{è la resistenza di strato del source (drain) ed } R_C$$

è la resistenza di contatto.

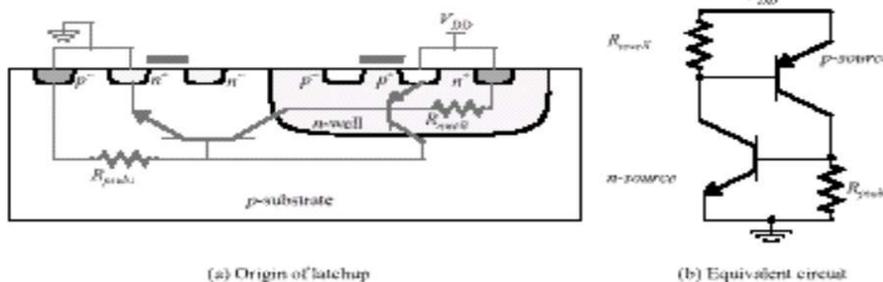
tali resistenze diminuiscono la performance del dispositivo in quanto riducono I_D che arriva da una certa tensione di controllo.

PS: per minimizzare tale effetto, di solito si ricopre S e D con materiali che hanno bassa resistività, come titanio o tungsteno.

Effetto hot-carrier

La miniaturizzazione ha aumentato il campo elettrico e quindi la velocità dei portatori di carica nel MOS. Acquistando abbastanza energia (con 10^4 V/m) possono entrare nel gate, aumentando V_T negli NMOS (e diminuendola nei PMOS) --> problemi di affidabilità dal punto di vista termico.

CMOS latch-up



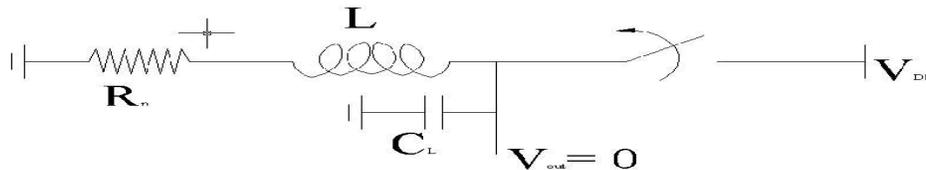
Nella struttura di un CMOS, vengono usate sequenze di zone drogate p-n-p-n, in cui il collettore di un CMOS è la base del successivo (una struttura che quindi, una volta accesa, non si dovrebbe più spegnere). Polarizzando in diretta ad esempio il primo terminale, si accende il successivo, il quale restituisce alla base del primo una corrente ancora maggiore (dato che le resistenze sono molto basse). Si arriva dunque alla saturazione, che porta poi alla rottura del CMOS --> si devono mettere delle protezioni per evitare i latch-up.

PS: con l'andare del tempo, la fisica dei dissipatori (che voleva mantenere standard la

tensione di alimentazione, a 5V prima, 3,3V poi) si è dovuta arrendere di fronte alla necessità di dover produrre dispositivi, e che quindi avessero diverse (e più basse) esigenze di alimentazione.

Circuiti ed effetti parassiti

Gli effetti parassiti hanno iniziato ad imporsi come fattore critico nei dispositivi all'aumentare dell'integrazione, alla diminuzione delle dimensioni e man mano che così aumentava la rete di interconnessioni, ossia quell'insieme di collegamenti metallici idealmente equipotenziali, ma che in realtà provocano ritardi nella propagazione del segnale.



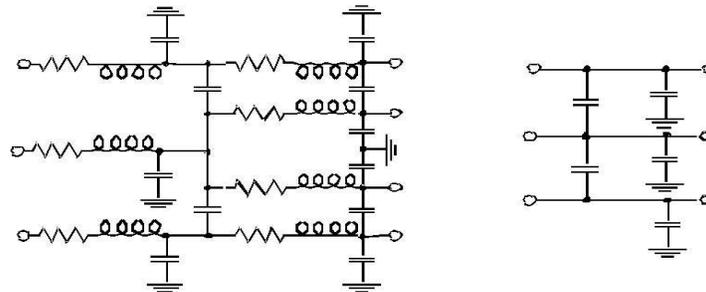
$$T_{pHL} = f(R_n, C_L, L) \text{ dal 50\% del segnale in ingresso al 50\% del segnale in uscita.}$$

Gli effetti parassiti aumentano al diminuire della grandezza del dispositivo e alla crescita della sua velocità (frequenza di lavoro). Le conseguenze degli effetti parassiti nell'interconnessioni sono:

- 1) aumento del ritardo di propagazione (resistenza) e caduta di performance;
- 2) aumento dell'energia dissipata, e quindi maggiori problemi di distribuzione della potenza (aumento della temperatura);
- 3) presenza di un maggior numero di fonti di rumore;

In fase di progetto è spesso impossibile tener conto di tutti i possibili effetti parassiti, per cui si ricorre a semplificazioni:

- gli effetti induttivi si possono trascurare se le variazioni di tensione non sono troppo brusche;
- si può modellare il circuito usando solo effetti capacitivi se i fili sono corti, oppure hanno una sezione grande, oppure hanno una bassa resistività;

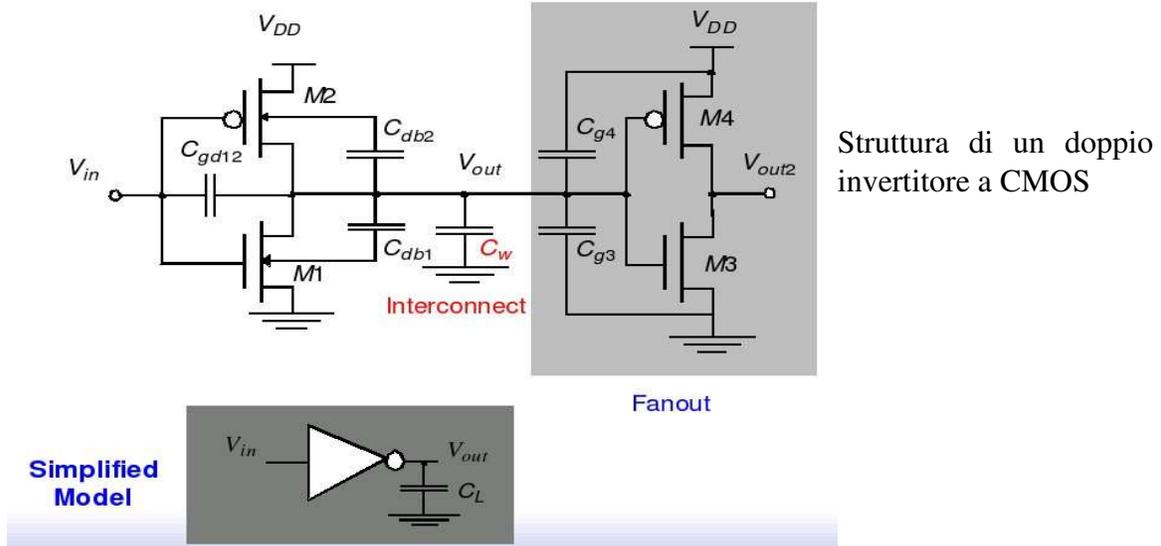


All-inclusive model

Capacitance-only

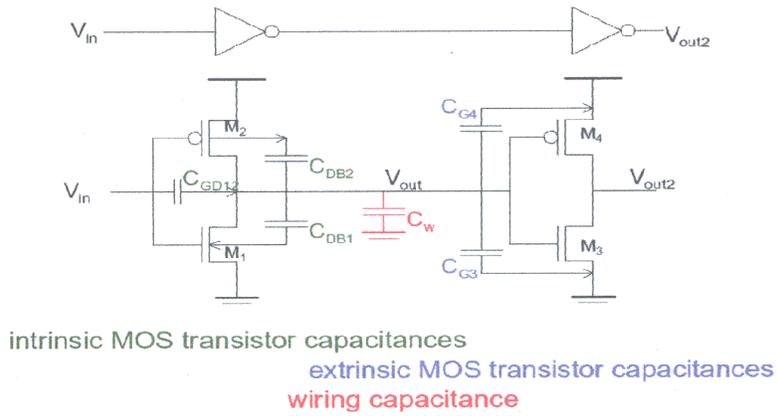
- se la distanza fra i conduttori è grande, o i fili viaggiano vicini per una breve distanza, le capacità di interconnessione possono essere ignorate, e tutte le altre capacità parassite possono essere modellate come capacità di massa.

Capacità parassite

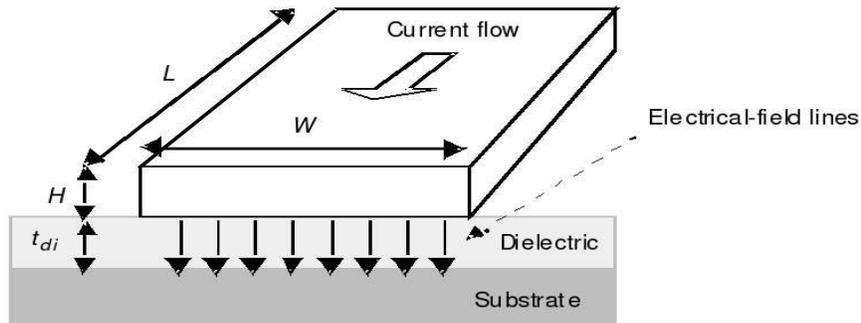


PS:

- nei CMOS ogni ingresso vede 2 transistor;
- il funzionamento dell'inverter sarà quello di caricare e scaricare C_w (capacità di carica);



Una semplice modellazione di capacità parassita può essere quella di un condensatore piano a facce parallele piane:



$$C_{int} = \frac{\epsilon_{di}}{t_{di}} \cdot W \cdot L \quad S_{C_{wire}} = \frac{S}{s \cdot S_L} = \frac{1}{S_L} \quad \text{dove: } \epsilon_{di} \text{ è la permittività del dielettrico e } t_{di} \text{ è lo spessore.}$$

Cioè, la capacità è proporzionale all'area di sovrapposizione tra i conduttori e inversamente proporzionale alla loro distanza.

PS: la permittività del silicio è 3,9 --> andrebbe meglio un materiale con un valore più basso, ma ci sono problemi costruttivi e tecnologici.

Material	ϵ_r
Free space	1
Aerogels	~1.5
Polyimides (organic)	3-4
Silicon dioxide	3.9
Glass-epoxy (PC board)	5
Silicon Nitride (Si ₃ N ₄)	7.5
Alumina (package)	9.5
Silicon	11.7

Questo modello è però troppo impreciso. Infatti per usare il modello a sole capacità devo avere R piccole --> per minimizzare R, bisogna mantenere più grande possibile il prodotto WL --> tuttavia servono valori piccoli di W per ottenere una maggiore densità delle interconnessioni, e quindi un'area occupata dal chip minore --> come conseguenza, col tempo, il rapporto W/H è arrivato ad essere < 1 (più alto che largo) --> per questo motivo, il modello a facce piane parallele è impreciso: diventa ora importante la capacità tra le 2 "pareti laterali" e il substrato (detta capacità di frangimento).

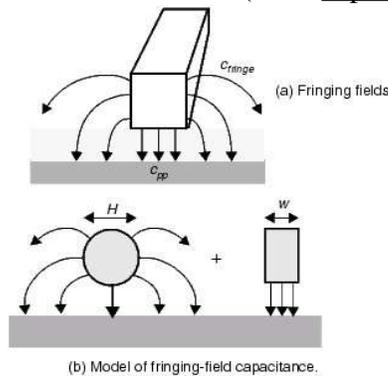
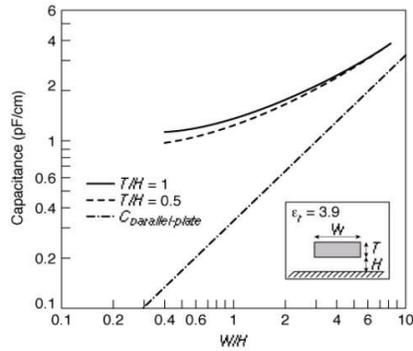


Figure 4.4 The fringing-field capacitance. The model decomposes the capacitance into two contributions: a parallel-plate capacitance, and a fringing capacitance, modeled by a cylindrical wire with a diameter equal to the thickness of the wire.

PS: la somma è anch'essa un' approssimazione
 somma strutture \simeq somma delle rispettive capacità

$$C_{wire} = C_{PP} + C_{fringe} = \frac{\bar{W} \cdot \epsilon_{di}}{t_{di}} + \frac{2 \cdot \pi \cdot \epsilon_{di}}{\log\left(\frac{t_{di}}{H}\right)} \quad \text{dove } \bar{W} \text{ risulta essere la larghezza del } C_{PP} \text{ che}$$

è uguale a $W - H/2$.

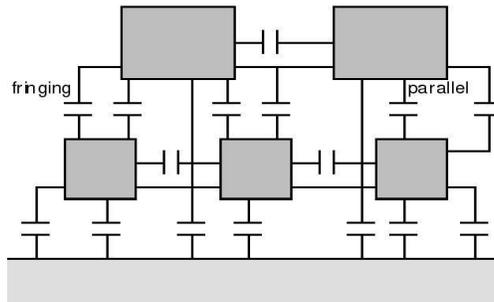


- per W/H grandi, C_{wire} si avvicinano a C_{PP} ;
- per $W/H < 1,5$ la C_{fringe} diventa dominante (è predominante il campo laterale);
- per W/H minori, $C_{\text{wire}} \simeq 1 \text{ pF/cm}$ (costante, non più in funzione di W);

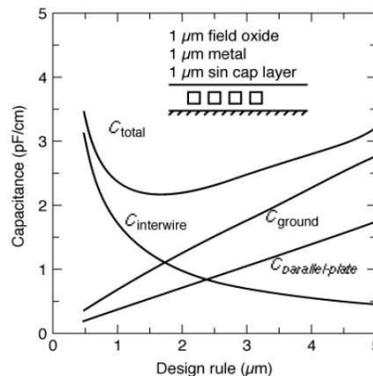
PS: $\frac{W}{H} \simeq \frac{W}{t_{di}}$ solo a causa delle attuali tecniche di planarizzazione.

Finora sono state prese in considerazione strutture planari di metallizzazione, in cui le capacità si trovavano solo con il substrato. Negli odierni dispositivi però ci sono diversi strati di metallizzazione e quindi ci saranno delle capacità parassite anche tra uno strato e l'altro (oltre che con il substrato).

PS: anzi, le capacità tra i livelli saranno maggiori di quelle con il substrato (quest'ultima sarebbe trascurabile se usassimo l'arsenuro di gallio, dotato di mobilità maggiore rispetto al silicio e appunto più isolante verso il substrato).



L'impatto di tale capacità aumenta poi con la diminuzione delle dimensioni:



Resistenza parassita

è direttamente proporzionale a L e inversamente proporzionale alla sezione del filo:

$$R = \frac{\rho \cdot L}{A} = \frac{\rho \cdot L}{H \cdot W} \quad \text{--> conduttore rettangolare}$$

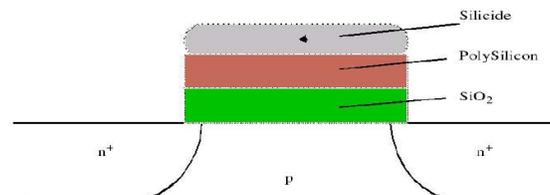
Material	ρ ($\Omega\text{-m}$)
Silver (Ag)	1.6×10^{-8}
Copper (Cu)	1.7×10^{-8}
Gold (Au)	2.2×10^{-8}
Aluminum (Al)	2.7×10^{-8}
Tungsten (W)	5.5×10^{-8}

Material	Sheet Resistance (Ω/\square)
n- or p-well diffusion	1000 – 1500
n^+ , p^+ diffusion	50 – 150
n^+ , p^+ diffusion with silicide	3 – 5
n^+ , p^+ polysilicon	150 – 200
n^+ , p^+ polysilicon with silicide	4 – 5
Aluminum	0.05 – 0.1

Per ridurre tali effetti, è necessario:

- usare una tecnologia di scaling selettiva;
- usare materiali per l'interconnessione migliori (e quindi in modo tale da ridurre la lunghezza media delle connessioni), come ad esempio siliciuri e rame;
- usare più strati di interconnessione (sempre per ridurre la lunghezza media dei fili);

PS: i siliciuri hanno conduttività a metà tra i metalli e il silicio policristallino. uso tuttavia il silicio policristallino per i gate poichè con questo posso creare strutture con il minimo di capacità parassite. Nella costruzione del dispositivo, inoltre, nella fase di impostazione, il silicio è sottoposto (tra i tanti) ad un processo termico d'attivazione del drogante: i siliciuri fonderebbero. Infine, il silicio policristallino ha una maggiore precisione e rischio meno capacità parassite dovute a tolleranze del processo produttivo.



Silicides: WSi_2 , TiSi_2 , PtSi_2 and TaSi

Conductivity: 8-10 times better than Poly

Vi sono poi anche resistenze di contatto, che si evidenziano nelle transizioni tra i diversi strati di instradamento: tali resistenze si riducono tentando di mantenere la rete dei segnali sullo stesso strato quanto più possibile, e facendo dei buchi larghi per i contatti (anche se questo provoca l'effetto di affollamento della corrente, che tende a concentrarla lungo il perimetro del buco).

PS: alcune resistenze di contatto:

- metallo-polysilicon: 5-20 Ω ;
- polysilicon/metallo - zone n⁺/p⁺: 5-20 Ω ;
- metallo-metallo: 1-5 Ω ;

c'è infine l'effetto pelle, che si manifesta a frequenze di lavoro molto alte (in cui $R = f(f)$): in tali condizioni, infatti, la corrente tende a passare sulla superficie del conduttore anzichè al suo interno.

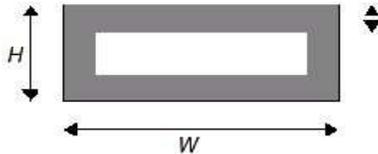


Figure 4.9 The skin-effect reduces the flow of the current to the surface of the wire.

La profondità della pelle (δ) è: $\delta = \sqrt{\frac{\rho}{\pi \cdot f \cdot \mu}}$ dove: ρ è la resistività, μ la permeabilità del dielettrico circostante e f è la frequenza.

PS: considerando $L = 2(W + H)\delta$, si ha che: $r(f) = \frac{\sqrt{\pi \cdot f \cdot \mu \cdot \rho}}{2 \cdot (W + H)}$ ad alte frequenze tale resistenza (per unità di lunghezza) in funzione della frequenza.

Induttanza Parassita

Ha importanza all'aumentare della frequenza di lavoro (sopra il Ghz). Provoca:

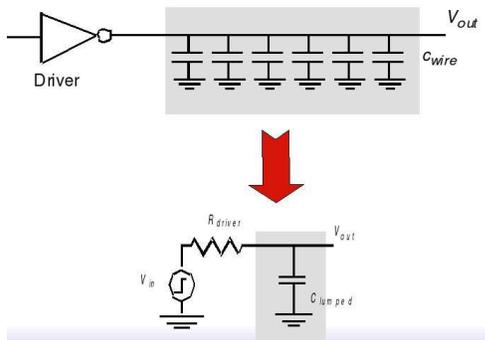
- effetti di ringing e overshoot;
- riflessione del segnale;
- accoppiamento induttivo;
- rumore di commutazione (a causa del fatto che $dV = L \cdot \frac{di}{dt}$)

Si può eliminare sfruttando la geometria del dispositivo, dato che:

$Cl = \mu\epsilon$, ma tale espressione vale solo se tale dielettrico è uniforme...

Modelli di interconnessione

- modello ideale: assenza di effetti parassiti, di ritardi nella propagazione (rete equipotenziale), commutazioni istantanee...
- modello concentrato (RC): le differenti parti degli effetti parassiti capacitivi e/o resistivi sono concentrate in singoli elementi --> il circuito è risolvibile con un'equazione differenziale ordinaria.



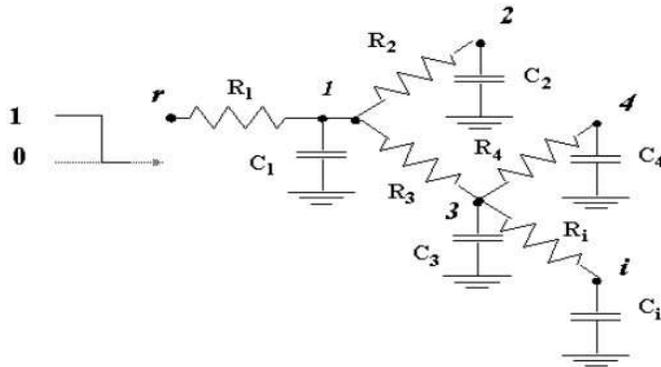
tale modello va bene se le componenti resistive e capacitive sono piccole e la frequenza di lavoro è bassa

$$C_{lumped} = \frac{dV_{out}}{dt} + \frac{V_{out} - V_{in}}{R_{driver}} = 0 \quad \text{da cui risulta, integrando da 0 a V:}$$

$$V_{out}(t) = (1 - e^{-\frac{t}{\tau}}) \cdot V$$

NB: modelli concentrati vanno bene per connessioni corte, ma sono pessimistici e imprecisi per le connessioni lunghe.

- modello RC distribuito: dovrei considerare ogni resistenza e capacità parassite, distribuite lungo la rete di interconnessioni --> servirebbero complesse equazioni differenziali e l'uso di derivate parziali. Per semplificare il calcolo, si usa la formula del ritardo di ELMORE:



- singolo nodo di input (r);
- tutte le capacità si trovano tra un nodo e la massa;
- la rete non contiene loop resistivi (è un'albero);

Ciò provoca la presenza di un cammino resistivo unico tra sorgente ed uno qualsiasi dei nodi della rete.

La resistenza totale di tale percorso è R_{ii} , ad esempio: $R_{44} = R_1 + R_3 + R_4$

R_{ik} , è invece detta resistenza del percorso condiviso (da r a k e da r a i):

$$R_{ik} = \sum R_j \Rightarrow R_j \in (\text{path}(r \rightarrow i) \cap \text{path}(r \text{ tende a } k)) \quad \text{ad esempio } R_{34} = R_3 + R_4, \quad R_{12} = R_1$$

Per trovare il ritardo da r a un nodo i, si usa l'approssimazione di ELMORE considerando il cammino condiviso:

$$\tau_{Di} = \sum_{k=1}^N C_k \cdot R_{ik} \quad \text{ad esempio il ritardo nel nodo 4 è:}$$

$$\tau_{D4} = R_1 \cdot C_1 + R_1 \cdot C_2 + (R_1 + R_3) \cdot C_3 + (R_1 + R_3) \cdot C_4 + \dots + (R_1 + R_3 + R_4) \cdot C_4$$

Caso particolare di questa approssimazione è il circuito RC in catena (scala a pioli)

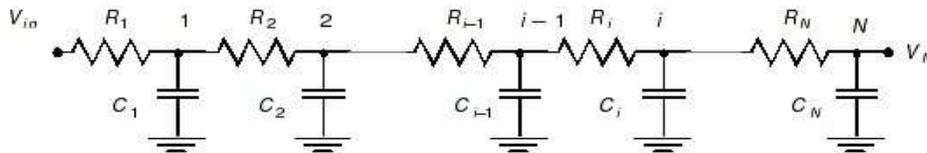


Figure 4.13 RC chain.

$$\tau_{Di} = \sum_{i=1}^N C_i \cdot \sum_{j=1}^i R_j = \sum_{i=1}^N C_i \cdot R_{ii} \quad \text{ad esempio il ritardo nel nodo i è:}$$

$$\tau_{Di} = R_1 \cdot C_1 + (R_1 + R_2) \cdot C_2 + \dots + (R_1 + R_2 + \dots + R_i) \cdot C_i$$

Se infine, anzichè considerare la resistenza globalmente, la considero parzialmente su N segmenti lunghi L/N , ottengo che:

$$\tau_{Di} = \left(\frac{L}{N}\right)^2 \cdot (rc + 2 \cdot rc + \dots + N \cdot rc) = (rcL)^2 \cdot \frac{N \cdot (N+1)}{2N^2} = RC \cdot \frac{N \cdot (N+1)}{2N^2} \quad \text{se } N \gg 1 \text{ (valori}$$

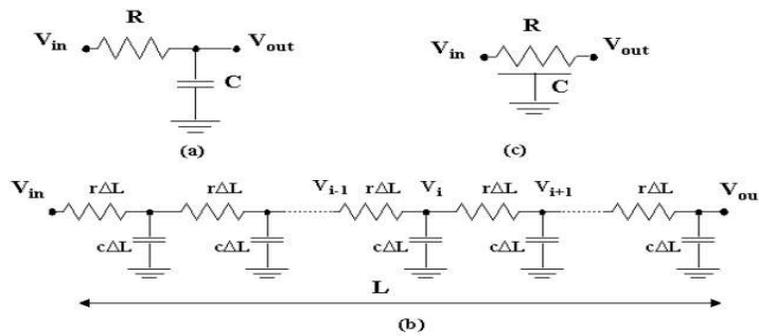
grandi) --> $\tau_{Di} = \frac{RC}{2} = \frac{rcL^2}{2}$

NB, osservazioni:

- 1) il ritardo è una funzione quadratica della lunghezza;
- 2) il ritardo di una linea RC distribuita è la metà del ritardo previsto con un modello RC concentrato (--> stima pessimistica, dato che $RC > (RC)/2$)

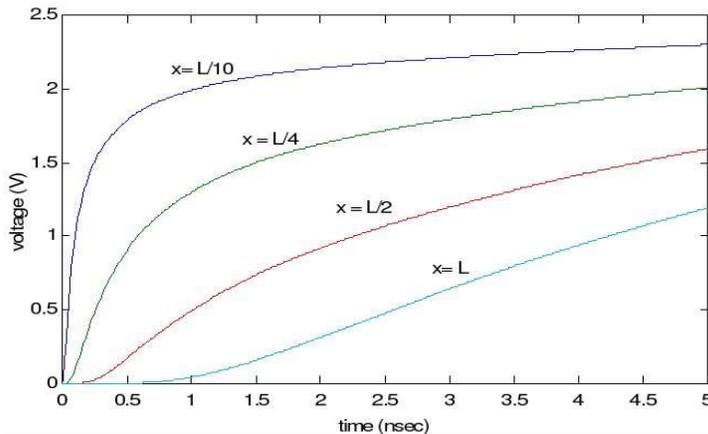
Se volessi invece risolvere il circuito in modo rigoroso, utilizzando le grandezze distribuite, mi dovrei appoggiare alle derivate parziali:

$$c \Delta l \cdot \frac{\partial V_i}{\partial t} = \frac{(V_{i+1} - V_i) + (V_{i+1} - V_i)}{r \Delta l}$$



Risolvendo e facendo tendere Δl a 0 --> $rc \cdot \frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}$ (equazione della diffusione),

dove x è la distanza tra la sorgente e un punto generico.



man mano che prosegue, l'impulso diventa sempre più lento --> importanza dell'avere linee corte.

NB: confronto tra modello concentrato e distribuito

Voltage Range	Lumped RC-network	Distributed RC-network
0→50% (t_p)	0.69 RC	0.38 RC
0→63% (τ)	RC	0.5 RC
10%→90% (t_r)	2.2 RC	0.9 RC

La linea di trasmissione

Quando i dispositivi sono sufficientemente buoni e le resistenze smettono d'avere grossa importanza, cresce l'influenza delle impedenze. I ritardi RC dovrebbero essere considerati solo quando:

- 1) $t_{pRC} \gg t_{pgate}$ quindi $t_{crit} \gg \sqrt{\frac{t_{pgate}}{0,38rc}}$;
- 2) $t_{rise(fall)} < RC$ ossia $t_{rise(fall)}$ comparabile con il tempo di viaggio del segnale lungo la linea.

modello di linea senza perdite di segnale

solo effetti capacitivi e induttivi (resistenza trascurabile)

$t_p = \sqrt{lc} \rightarrow Z_0 = \frac{1}{CV}$ dove Z_0 è l'impedenza caratteristica in funzione del dielettrico e della geometria del conduttore e dell'isolante, e V è la velocità di propagazione che risulta

essere: $V = \frac{1}{\sqrt{\epsilon \mu}}$