

## 1) I solidi nei circuiti integrati

I circuiti integrati (CI) sono delle reti di dispositivi attivi e passivi allo stato solido realizzati sullo stesso chip di silicio (*silicon* in inglese). I singoli dispositivi a loro volta sono costituiti da materiali (elementi o composti) allo stato solido, quali ad esempio Si, Al, SiO<sub>2</sub>, W, TiN, Si<sub>3</sub>N<sub>4</sub> ecc. (silicio, alluminio, biossido di silicio, tungsteno, nitrato di titanio, nitrato di silicio ecc.). I solidi citati differiscono fra loro per le proprietà elettriche, termiche, chimiche, meccaniche nonché per la loro struttura atomica che le determina.

Distingueremo i solidi principalmente in base alle proprietà elettriche, anche se, come vedremo, le altre caratteristiche sono importanti per la fabbricazione ed il funzionamento dei dispositivi elettronici e dei CI.

Dal punto di vista del trasporto di corrente i solidi possono essere: conduttori (Al, Cu, Au...), semiconduttori (Si, Ge, GaAs...), isolanti (SiO<sub>2</sub>, Si<sub>3</sub>N<sub>4</sub>...). La conducibilità  $\sigma$  per un metallo è molto elevata ( $\simeq 10^5 \Omega^{-1}\text{cm}^{-1}$ ) e, in prima approssimazione, indipendente dalla temperatura; per un semiconduttore può coprire un range molto esteso (lo stesso semiconduttore può essere quasi un *metallo* o quasi un *isolante* a seconda del contenuto di impurezze) e dipende inoltre dalla temperatura; per un isolante è teoricamente nulla a tutte le temperature.

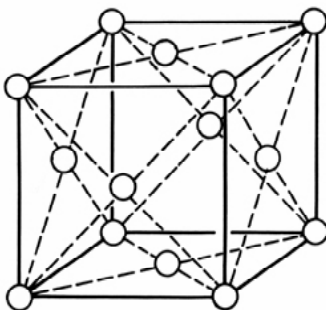
La struttura dei solidi ne determina le proprietà elettriche. Quando gli atomi che costituiscono un solido sono disposti in modo periodico nello spazio si parla di *struttura cristallina*. La più semplice è la struttura cubica (che peraltro non è adottata da alcun solido): un atomo posto in ogni punto di un reticolo cubico. Il lato del cubo  $a$  costituisce la *costante reticolare* e il suo valore è di qualche Å.

Per ottenere un solido nella forma di monocristallo sono necessarie tecniche particolari. I solidi possono avere anche struttura policristallina o amorfa.

In un monocristallo la periodicità della struttura si ha in tutto il volume occupato dal solido; in un policristallo questo ordine si ha solo localmente in volumi più o meno ampi; in un solido amorfo invece non esiste alcuna struttura regolare e ripetitiva.

## 2) Metalli

L'Al adotta la struttura cubica a facce centrate (FCC): si tratta di una struttura cubica con in più un atomo *al centro* di ogni faccia.



A partire dalla conoscenza di  $a$  per l'Al ( $a = 4.04 \text{ \AA}$ ) si può calcolare la sua densità atomica  $N_{Al}$ , considerando quanti atomi appartengono ad un cubo di lato  $a$ . Si ottiene (tenendo conto che un atomo al vertice del cubo è a comune con altre otto celle, mentre un atomo al centro della faccia è a comune con due sole celle)

$$N_{Al} = \frac{1 + 3}{(4.04 \times 10^{-10})^3} = 6.07 \times 10^{28} \text{ atomi/m}^3 = \text{m}^{-3}.$$

Le densità atomiche di tutti i solidi sono di quest'ordine di grandezza:  $10^{28} \text{ m}^{-3} = 10^{22} \text{ cm}^{-3}$ . Nell'Al, per effetto delle interazioni fra gli atomi, ogni atomo cede al volume occupato dal solido un elettrone, che diviene così libero di muoversi all'interno del cristallo. La concentrazione di elettroni di *conduzione*  $n$  sarà dunque pari a  $6.07 \times 10^{22} \text{ cm}^{-3}$ .

A partire da questo dato si può facilmente calcolare la conducibilità  $\sigma$  dell'Al. Come è noto

$$\sigma = \frac{ne^2\tau}{m_0}$$

dove  $\tau$  è il tempo di rilassamento degli elettroni di conduzione ( $2.2 \times 10^{-14} \text{ s}$  per l'Al) e  $m_0$  è la massa dell'elettrone libero ( $9.1 \times 10^{-31} \text{ Kg}$ ).

$$\sigma = \frac{6.07 \times 10^{28} \times (1.6 \times 10^{-19})^2 \times 2.2 \times 10^{-14}}{9.1 \times 10^{-31}} = 3.76 \times 10^7 \text{ } \Omega^{-1} \text{ m}^{-1} = 3.76 \times 10^5 \text{ } \Omega^{-1} \text{ cm}^{-1}$$

La resistività  $\rho = \frac{1}{\sigma}$  vale dunque  $\frac{1}{3.76 \times 10^5} = 2.66 \times 10^{-6} \text{ } \Omega \cdot \text{cm}$ . Tutti i metalli hanno resistività dell'ordine del  $\mu\Omega \cdot \text{cm}$ .

### Equazione della neutralità di carica

Un campione metallico è normalmente neutro. E' costituito, per quanto visto, da ioni positivi fissi nelle posizioni reticolari (atomi che hanno ceduto un elettrone) e da un gas di elettroni liberi. Si avrà dunque

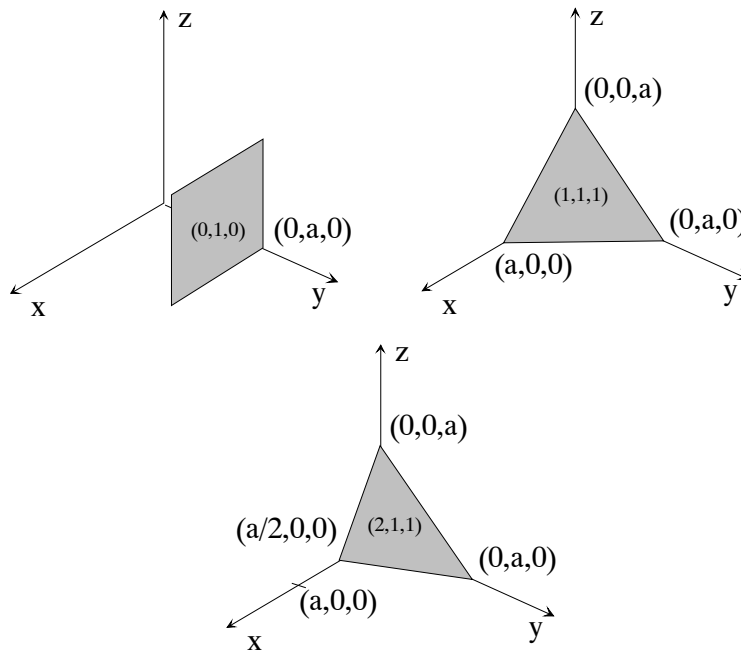
$$n = N_{Al}^+$$

o, in termini di carica netta nell'unità di volume

$$-qn + qN_{Al}^+ = 0.$$

Ovviamente  $N_{Al}^+ = N_{Al}$ .

-----  
Le strutture cristalline sono caratterizzate da un certo grado di simmetria. E' importante individuare piani e direzioni notevoli nello spazio occupato dal cristallo. Allo scopo servono gli indici di Miller, che si ottengono, con riferimento ai piani della figura, come segue:



1) si individuano le intercette del piano con gli assi del sistema di riferimento  $x, y, z$  scegliendo come unità di misura la costante reticolare;

2) si invertono i numeri trovati;

3) se risultano delle frazioni si riducono al minimo comune denominatore, che viene poi ommesso.

In una struttura cubica i piani con determinati indici sono ortogonali a direzioni con gli stessi indici.

Notazione

$(100)$ : un insieme di piani paralleli;  $[100]$ : la direzione ad essi ortogonale;

$\{100\}$ : piani di tipo  $(100)$ , ovvero i piani  $(100), (010), (001), (\bar{1}00), (0\bar{1}0), (00\bar{1})$

$\langle 100 \rangle$ : direzioni di tipo  $[100]$ , ovvero le  $[100], [010], [001], [\bar{1}00], [0\bar{1}0], [00\bar{1}]$ .

$\bar{1}$  sta per  $-1$ .

-----

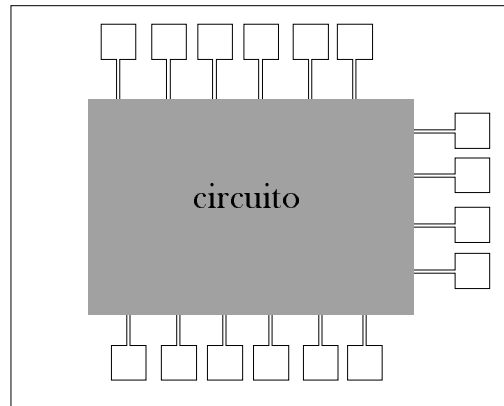
Si pone adesso il problema dell'impiego dei metalli nei CI. A questo scopo consideriamo un film di Al con uno spessore  $t = 3000 \text{ \AA}$  ( $0.3 \mu\text{m}$ ) e calcoliamo la resistenza di un resistore con la forma di un parallelepipedo rettangolo di lunghezza  $L$  e larghezza  $W$ . Si ha

$$R = \rho \frac{L}{Wt} = \frac{\rho}{t} \frac{L}{W} = R_{\square} \frac{L}{W}$$

in cui  $R_{\square}$  è la resistenza di strato, che rappresenta la resistenza di un quadrato con lato  $W$ , indipendentemente dal valore di  $W$ .  $\frac{L}{W}$  è il numero di quadrati che

costituiscono il rettangolo  $L \times W$ . Numericamente  $R_{\square} = \frac{2.66 \times 10^{-6} (\Omega \text{cm})}{0.3 \times 10^{-4} (\text{cm})} = 8.87 \times 10^{-2} \Omega/\square$ . Se si suppone il parallelepipedo costituito da 100 quadrati otteniamo un valore di  $R$  pari a soli 8.87  $\Omega$ . E' evidente che i metalli non possono essere usati per la realizzazione di resistori integrati, ma piuttosto come collegamenti fra i vari dispositivi (linee di metal o *metal*). Nei moderni CI, soprattutto in quelli ad altissima densità di integrazione e ad alta velocità, si tende a sostituire l'Al con il rame (Cu), in quanto quest'ultimo presenta una resistività inferiore (dell'ordine di 1.7  $\mu\Omega \cdot \text{cm}$ ).

Nota la resistenza di strato, per stabilire il valore di resistenza di una certa metal, basta moltiplicare  $R_{\square}$  per il numero dei quadrati di cui è formata. Le metal che costituiscono le connessioni all'interno del CI terminano alla periferia del chip stesso con delle piazzole quadrate (*pad*), di dimensioni  $\simeq 100 \mu\text{m} \times 100 \mu\text{m}$ , sulle quali verranno poi saldati i fili che consentono al CI di colloquiare con il mondo esterno.

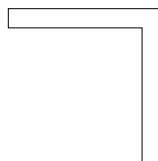


In prima approssimazione la resistività di un metallo non dipende dalla temperatura  $T$ , dato che  $n$  è indipendente da  $T$ .

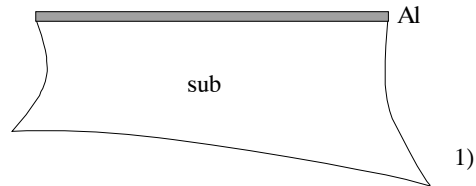
### 3) Introduzione alla fotolitografia

Si usa comunemente l'Al come materiale delle metal perché presenta un'ottima adesione sia sul Si che sull' $\text{SiO}_2$  (Au e Cu aderiscono invece molto male). Prima di passare a descrivere il processo mediante il quale si realizzano le metal di un CI, si osservi che sia la larghezza che lo spessore sono generalmente submicrometrici (poche migliaia di  $\text{\AA}$ ).

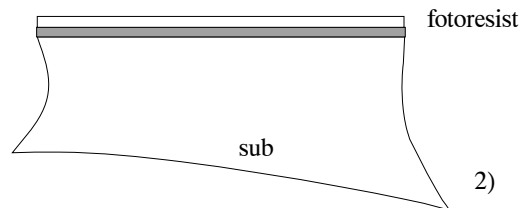
Per definire una linea di metal, come quella del disegno, si procede come segue.



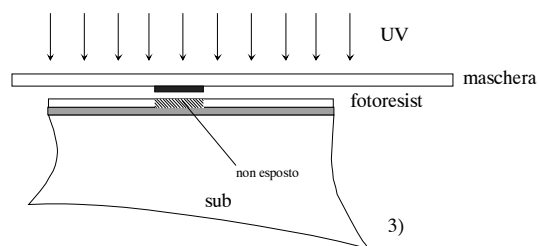
1) Su un substrato (diametro  $\simeq 20$  cm, spessore  $0.7$  mm =  $700$   $\mu\text{m}$ ) si depone (vedremo in seguito come) un film *sottile* di Al (ad esempio  $200$  nm). Per film sottile si intende uno strato uniforme di materiale con spessore  $\leq 1$   $\mu\text{m}$ .



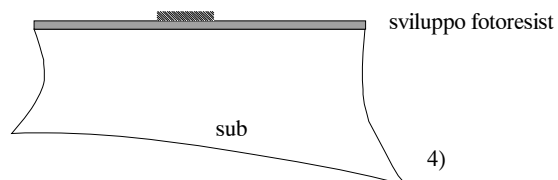
2) Il film viene poi ricoperto con uno strato di fotoresist. Il fotoresist è un materiale sensibile alla radiazione con lunghezza d'onda inferiore ai  $400$  nm (UV). Ciò significa che quando il fotoresist viene esposto all'UV cambia le sue proprietà e diviene solubile in una soluzione di sviluppo. Quanto detto definisce il comportamento di un f. *positivo*. Un f. *negativo* si comporta in modo complementare.



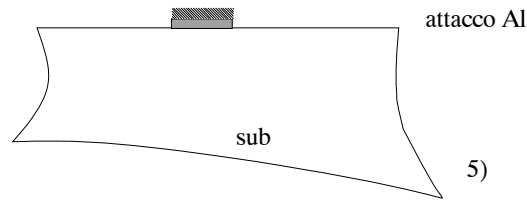
3) Si espone il f. attraverso una maschera. Per maschera si intende una lastra di quarzo (trasparente all'UV) su cui si hanno delle zone *non trasparenti*.



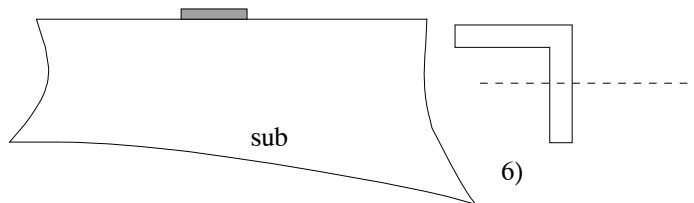
4) Si sviluppa il fotoresist. L'operazione rimuove completamente il f. esposto lasciando scoperto il film di Al.



5) Si immerge il substrato in una soluzione che attacca l'Al (*etching solution* o, più brevemente, *etching*) nelle zone non protette.



6) Si elimina il fotoresist non esposto. Restano quindi sul substrato le porzioni del film di Al che durante l'attacco erano state protette dal fotoresist.



Risulta altresì evidente che con un'unica sequenza di passi (processo fotolitografico) si possono definire sul substrato metal di Al con geometrie diverse. Il processo fotolitografico è l'intermediario fra un disegno (un file di dati) e l'oggetto reale sul substrato.

Il processo descritto, schematicamente per il momento, è lo stesso per tutti i processi che portano alla realizzazione dei CI, anche quelli più complessi.

Gli attacchi dei vari materiali che compongono un CI sono di due tipi: *wet* (in soluzione) o *dry* (al plasma). Entrambi i tipi di attacco possono essere isotropi o anisotropi. In un attacco isotropo la velocità di attacco è identica in tutte le direzioni. Si definisce l'anisotropia  $A$  dell'attacco mediante la relazione

$$A = 1 - \frac{v_o}{v_v}$$

in cui  $v_o$  è la velocità di attacco nelle direzioni  $yz$  e  $v_v$  quella in direzione  $x$ .

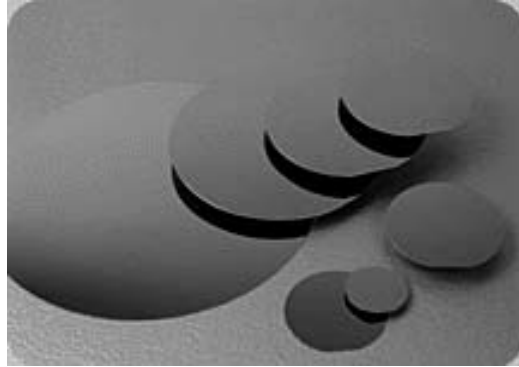
## 4) Semiconduttori

Esistono numerosi materiali semiconduttori (sc), sia semplici (Si, Ge), che composti (GaAs, InP, GaAlAs...). La resistività di un sc dipende fortemente da  $T$  ed inoltre può essere cambiata di ordini di grandezza introducendo nel sc una quantità controllata di impurezze opportune (drogaggio). Il sc di riferimento sarà d'ora in poi, salvo menzione esplicita, il silicio (Si).

Il Si ha numero atomico 14 e appartiene al IV gruppo del sistema periodico, a cui appartengono anche C e Ge.

Senza il Si e l'SiO<sub>2</sub>, l'elettronica integrata non esisterebbe, almeno nella forma attuale.

Dal lingotto vengono ricavati i wafer, che tipicamente hanno uno spessore di  $600 \div 700 \mu\text{m}$ . Come già detto i wafer più comunemente usati nella fabbricazione dei CI sono quelli (100) e (111). I piani di facile rottura nel Si sono gli  $\{111\}$ .



Per rendere ragione delle

## 5) Proprietà elettriche dei solidi

è necessario conoscere come sono distribuiti in energia all'interno di un solido gli elettroni di conduzione. Lo stato di un elettrone è definito quando si conoscano le tre componenti della sua quantità di moto e l'orientamento dello spin. Descrivendo il comportamento dell'elettrone con il formalismo della meccanica quantistica (MQ) sussiste la seguente uguaglianza fra le componenti della quantità di moto e le componenti del vettore d'onda

$$p_x = m_0 v_x = \hbar k_x, \quad p_y = m_0 v_y = \hbar k_y, \quad p_z = m_0 v_z = \hbar k_z$$

per cui l'energia cinetica dell'elettrone si potrà scrivere come

$$E = \frac{p^2}{2m_0} = \frac{\hbar^2 k^2}{2m_0} = \frac{\hbar^2}{2m_0} (k_x^2 + k_y^2 + k_z^2).$$

La relazione precedente stabilisce che si possono avere più stati con la stessa energia  $E$ . D'ora in avanti intenderemo per stato una terna di valori  $k_x, k_y, k_z$  (un vettore  $k = \frac{2\pi}{\lambda}$ ).

La costante di Planck  $h$  vale  $6.62 \times 10^{-34}$  joule·s.

$\hbar$  è per definizione  $= h/2\pi$ .

Uno dei risultati più importanti della MQ è la quantizzazione dei valori permessi di  $k$  e quindi dell'energia. Questo accade quando l'elettrone è confinato in uno spazio finito e quindi per un elettrone di conduzione di un solido: *non tutti i valori di  $k$  sono permessi*.

Introduciamo il concetto di densità di stati  $D(E)$ :  $D(E)$  rappresenta il numero di stati per unità di intervallo di energia e di volume. Dimensionalmente  $E^{-1}l^{-3}$ .

Se con  $n(E)$  indichiamo il numero di elettroni per unità di volume con una certa energia  $E$ , vale la seguente relazione fondamentale

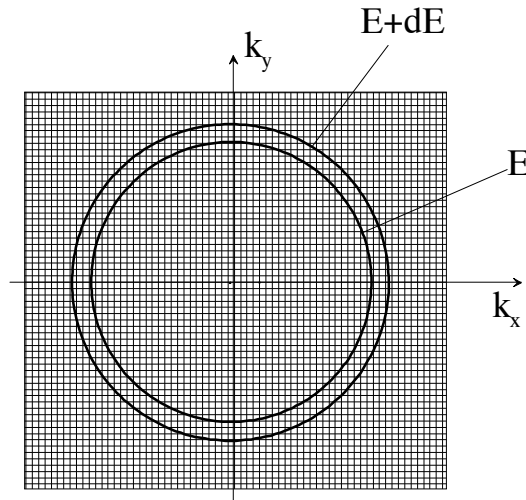
$$n(E)dE = 2D(E)f(E)dE :$$

il numero di elettroni per unità di volume che hanno energia compresa fra  $E$  e  $E+dE$  è dato dal prodotto della densità di stati *per la probabilità che uno stato con energia  $E$  sia occupato ad una temperatura assoluta  $T$* . 2 tiene conto ovviamente del fatto che ciascuno stato (terna di valori  $k_x, k_y, k_z$ ) può essere occupato da due elettroni con spin opposto. Prima di ricavare la  $D(E)$  e di vedere quale espressione abbia la  $f(E)$  osserviamo che

$$n = \int_0^\infty n(E)dE = 2 \int_0^\infty D(E)f(E)dE :$$

la concentrazione totale  $n$  di elettroni di conduzione si ottiene integrando su tutte le possibili energie. La relazione precedente costituisce il punto di partenza per tutte le considerazioni sul comportamento elettrico dei solidi.

### La densità di stati



Consideriamo uno spazio dei  $k$  e in questo spazio una sfera di energia  $E$  di equazione  $E = \frac{\hbar^2}{2m_0} (k_x^2 + k_y^2 + k_z^2)$ ; concentrica a questa un'altra sfera di energia  $E + dE$ . Il volume  $d\Omega$  (nello spazio dei  $k$ ) compreso fra le due sfere vale

$$d\Omega = 4\pi k^2 dk.$$

Se è noto il volume  $d\omega$  occupato da uno stato, il rapporto  $d\Omega/d\omega$  fornisce il numero di stati permessi compresi fra  $E$  ed  $E + dE$ . Non essendo questa la sede in cui trattare diffusamente questi argomenti, si da per buono il risultato seguente: in un



solido con dimensioni  $L_x, L_y, L_z$  due valori (contigui) permessi delle componenti di  $k$  differiscono fra loro di  $2\pi/L_i$  (la quantizzazione cresce al diminuire delle dimensioni del solido che confina al suo interno gli elettroni).

I valori permessi delle terne  $k_x, k_y, k_z$  formano nello spazio dei  $k$  un reticolo di punti a ciascuno dei quali compete un volume  $d\omega = (2\pi)^3 / (L_x L_y L_z) = (2\pi)^3 / V$ , in cui  $V$  è il volume del solido in questione. Ponendo  $V = 1 \Rightarrow d\omega_1 = (2\pi)^3$ . Poiché  $D(E)dE$  rappresenta il numero di stati nell'intervallo  $dE$  sarà

$$D(E)dE = \frac{d\Omega}{d\omega_1} = \frac{4\pi k^2 dk}{(2\pi)^3}.$$

Utilizzando la relazione che lega  $E$  a  $k$

$$E = \frac{\hbar^2 k^2}{2m_0}$$

e differenziando

$$dE = \frac{\hbar^2 2k dk}{2m_0} = \frac{\hbar^2 k dk}{m_0}$$

si ottiene

$$D(E)dE = 2\pi \left( \frac{2m_0}{\hbar^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}} dE$$

e infine

$$D(E) = 2\pi \left( \frac{2m_0}{\hbar^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}}$$

che è la densità di stati cercata. Si osservi che la quantizzazione dei  $k$  implica necessariamente la quantizzazione dei valori permessi di  $E$ .

L'energia degli elettroni nei solidi si misura solitamente in

### **elettronvolt (eV)**

definito come l'energia che acquista la carica di un elettrone quando viene accelerato dalla ddp di un volt

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ joule.}$$

### **La Fermi-Dirac**

La funzione che segue è detta funzione di distribuzione di Fermi-Dirac

$$f(E) = \frac{1}{1 + \exp\left(\frac{E-E_F}{kT}\right)}$$

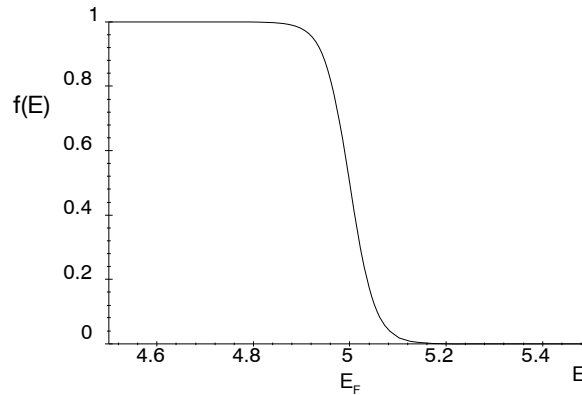
e rappresenta la probabilità che uno stato con energia  $E$  sia occupato alla temperatura  $T$ .

$k$  è la costante di Boltzmann che vale  $8.633 \times 10^{-5} \text{ eVK}^{-1}$ .

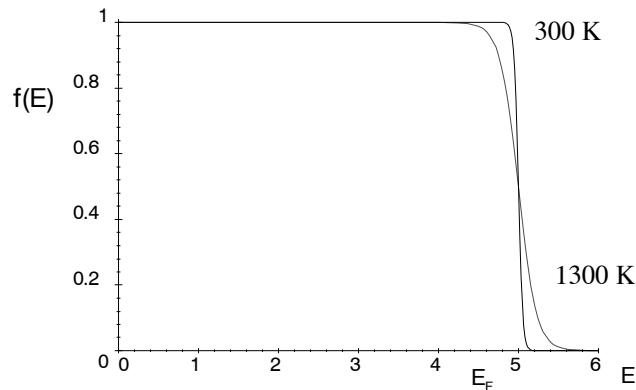
$kT$  a  $300 \text{ K} = 0.0259 \text{ eV}$ .

Il primo grafico mostra la FD nei pressi di  $E_F$  per  $T = 300 \text{ K}$  ed  $E_F = 5 \text{ eV}$ . Valori dell'ordine degli eV sono tipici dei metalli. Si noti che solo su una scala espansa si può chiaramente apprezzare l'andamento della FD nei pressi di  $E_F$ .

In effetti l'argomento dell'esponenziale a denominatore è la variabile  $x = \frac{E-E_F}{kT}$  e quindi, poiché si ha  $\frac{1}{1+\exp(4)} = 1.78 \times 10^{-2}$  e  $\frac{1}{1+\exp(-4)} = 0.982$ , la FD differisce da 0 ed 1, rispettivamente, solo per  $|E - E_F| \leq 4kT$ ; complessivamente in un intervallo di  $8kT$ , pari a  $0.208 \text{ eV}$ .



Il secondo grafico è relativo alla  $f(E)$  a due temperature notevolmente diverse ( $300$  e  $1300 \text{ K}$ ); si noti la deformazione della funzione nei pressi di  $E_F$ : la probabilità che uno stato con  $E > E_F$  sia occupato cresce al crescere di  $T$ .

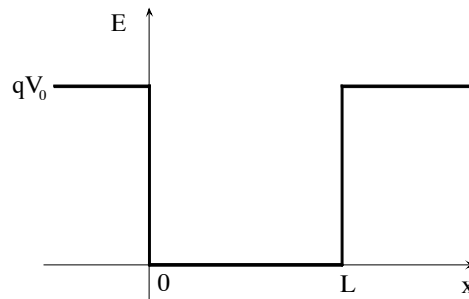


Se  $f(E)$  è la probabilità che uno stato sia occupato, ovviamente  $1 - f(E)$  è la probabilità che uno stato sia vuoto.  $f(E)$  vale  $0.5$  quando  $E = E_F$  e per ogni  $T$ ; la definizione di livello di Fermi come quel livello energetico i cui stati hanno probabilità  $0.5$  di essere occupati ad ogni temperatura viene mantenuta anche allo zero assoluto.

Resta il problema di sapere rispetto a quale valore di riferimento viene misurata l'energia  $E_F$ . Si introduce a questo scopo il

### Modello dell'elettrone libero

che consiste nel considerare ciascuno degli elettroni del metallo in interazione con un potenziale costante, generato dagli  $N$  atomi della struttura, che hanno perso un elettrone (ioni positivi fissi nelle posizioni reticolari), e dai rimanenti  $N - 1$  elettroni. Un singolo elettrone vede dunque un potenziale positivo, che gli impedisce normalmente di abbandonare il metallo. Una rappresentazione schematica monodimensionale del modello è quella della figura. L'energia potenziale è assunta uguale a zero nell'intervallo  $(0, L)$  dell'asse  $x$ , mentre all'esterno vale  $qV_0$  (profondità della *buca di potenziale*). E' rispetto al fondo della buca che vengono misurate le energie cinetiche dell'elettrone e quindi anche  $E_F$ .



Per elettrone libero si intende un elettrone non soggetto a forze: all'interno della buca questo è sicuramente vero, non lo è tuttavia in  $0$  ed  $L$ ; non entreranno comunque nel merito delle condizioni al contorno in  $0$  e in  $L$ . Il valore di  $E_F$  si calcola supponendo di essere alla temperatura di  $0$  K. La

$$n = 2 \int_0^{\infty} D(E) f(E) dE$$

diviene, in questa ipotesi

$$n = 2 \int_0^{E_F} 2\pi \left( \frac{2m_0}{h^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}} dE = \frac{8}{3} \pi \left( \frac{2m_0}{h^2} \right)^{\frac{3}{2}} E_F^{\frac{3}{2}}$$

da cui

$$E_F = \left( \frac{3n}{8\pi} \right)^{\frac{2}{3}} \frac{h^2}{2m_0}$$

Dato che per l'Al  $n = 6.07 \times 10^{28} \text{ m}^{-3}$

$$\begin{aligned} E_F &= \left( \frac{3 \times 6.07 \times 10^{28}}{8\pi} \right)^{\frac{2}{3}} \frac{(6.62 \times 10^{-34})^2}{2 \times 9.1 \times 10^{-31}} = 9.07 \times 10^{-19} \text{ joule} \\ &= \frac{9.07 \times 10^{-19}}{1.6 \times 10^{-19}} = 5.67 \text{ eV.} \end{aligned}$$

Poiché  $n$  in un metallo è indipendente da  $T$  lo è anche  $E_F$ .

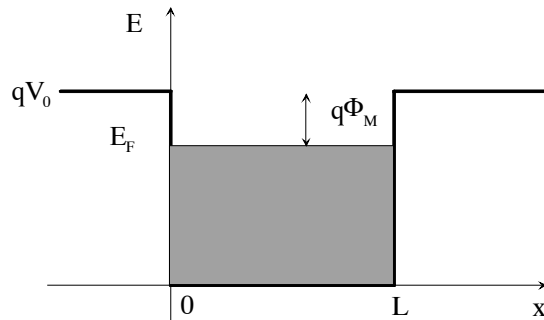
Un elettrone che ha energia pari ad  $E_F$  avrà una velocità (all'equilibrio) che si ottiene dalla relazione

$$E_F = \frac{1}{2}m_0v_F^2$$

$$v_F = \sqrt{\frac{2E_F}{m_0}} = \sqrt{\frac{2 \times 5.67 \times 1.6 \times 10^{-19}}{9.1 \times 10^{-31}}} = 1.4 \times 10^6 \text{ m/s} = 1400 \text{ Km/s}$$

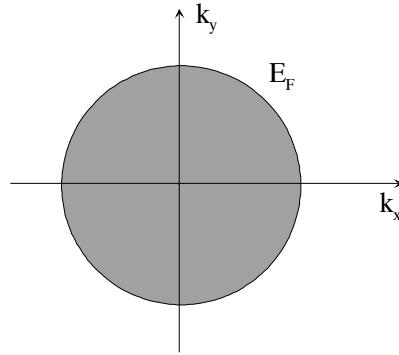
1/200 circa della velocità della luce.

Facendo sempre riferimento allo spazio dei  $k$ , alla temperatura di 0 K cominceremo a disporre gli  $n$  elettroni di conduzione a partire dagli stati con energia più bassa, ma poiché per ogni livello energetico sono in numero finito, una volta sistemati tutti gli elettroni avremo raggiunto un certo valore di energia, che è appunto  $E_F$ . Il modello dell'elettrone libero è sintetizzato nella figura.  $q\Phi_M$  rappresenta la funzione di lavoro del metallo, ovvero l'energia minima che bisogna fornire ad un elettrone per estrarlo dal metallo stesso. Le funzioni di lavoro dei metalli sono comprese nell'intervallo 3÷6 eV.



### Il passaggio di corrente nei metalli

La figura mostra lo spazio degli stati permessi agli elettroni (spazio dei  $k$ , inversi delle lunghezze  $\equiv$  spazio *reciproco*).



Si osservi che a 0 K e all'*equilibrio* tutti gli stati con energia  $\leq E_F$  sono occupati. Ricordando che  $m_0 v = \hbar k$  e dato che la sfera che rappresenta il volume relativo a questi stati è centrata nell'origine,  $\sum_i k_i = 0 \Rightarrow \sum_i v_i = 0$  : l'insieme degli elettroni ha una velocità complessivamente nulla e quindi, attraverso una qualunque superficie all'interno del metallo, non si ha passaggio netto di elettroni (corrente nulla). L'aumento della temperatura non cambia la situazione; infatti la distribuzione degli elettroni sugli stati permessi conserva la simmetria precedente.

Se ad un campione metallico viene applicata una ddp, al suo interno si avrà un campo elettrico costante che agisce su ciascun elettrone. Supponiamo che sia parallelo all'asse  $x$ . Utilizzando l'equazione classica del moto

$$-q\mathcal{E} = m_0 \frac{dv_x}{dt} = \hbar \frac{dk_x}{dt}$$

$$k_x(t) = -\frac{q\mathcal{E}}{\hbar}t + k_{0x}$$

la componente secondo  $x$  di ciascun vettore d'onda varia nel tempo. Se l'elettrone fosse effettivamente libero il suo moto sarebbe uniformemente accelerato e quindi in contrasto con l'evidenza sperimentale: ddp costante  $\Rightarrow$  campo costante  $\Rightarrow$  velocità costante  $\Rightarrow$  corrente costante. In realtà gli elettroni non si muovono nello spazio libero ma in un mezzo (il reticolo formato dagli ioni) con il quale subiscono collisioni. Il motivo del  $\tau$  (tempo di rilassamento) che compare nella conducibilità (e nella mobilità)

$$\sigma = \frac{ne^2\tau}{m_0}$$

è proprio legato al fatto che, dopo un tempo che vale mediamente  $\tau$ , l'elettrone subisce una collisione a seguito della quale perde memoria dell'effetto del campo. Dopo ogni collisione  $k_{ix}(t) = -\frac{q\mathcal{E}}{\hbar}t + k_{ix}$ . Il risultato complessivo è che gli elettroni acquistano una velocità di insieme, che prende il nome di velocità di drift (trascinamento), che è proporzionale al campo

$$\Delta k_x = -\frac{q\mathcal{E}}{\hbar}\tau$$

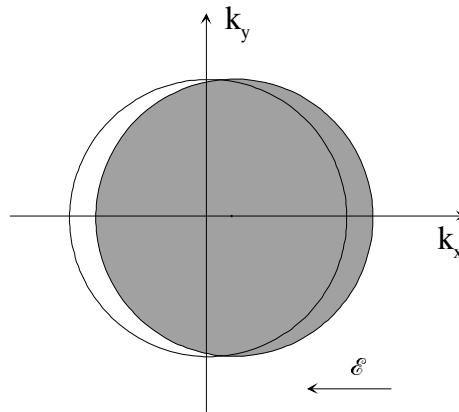
$$\hbar\Delta k_x = m_0\Delta v_x = -q\mathcal{E}\tau$$

da cui

$$\Delta v_x = v_{drift} = -\frac{q\tau}{m_0}\mathcal{E} = -\mu_n\mathcal{E}.$$

La costante  $\mu_n$  prende il nome di mobilità e usualmente si misura in  $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ .

La sfera di Fermi che rappresenta lo stato degli elettroni all'equilibrio si sposterà nello spazio dei  $k$  di una quantità  $\Delta k_x = -q\mathcal{E}\tau/\hbar$  e l'insieme degli stati occupati dagli elettroni non sarà più simmetrico (sfera di F. centrata in zero).



Avremo dunque  $\sum_i k_i \neq 0 \Rightarrow \sum_i v_i \neq 0$  e quindi passaggio di corrente.

Per l'Al  $\mu_n = \frac{q\tau}{m_0} = \frac{1.6 \times 10^{-19} \times 2.2 \times 10^{-14}}{9.1 \times 10^{-31}} = 3.87 \times 10^{-3} \text{ m}^2\text{V}^{-1}\text{s}^{-1} = 38.7 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ .

-----

Legge di Ohm (l'intensità di corrente riguarda per convenzione un moto di cariche positive)

$$v_{drift} = -\mu_n\mathcal{E} = \mu_n \frac{dV}{dx}$$

se il campo è costante

$$v_{drift}L = \mu_n V_{app}$$

dove  $L$  rappresenta la lunghezza di un filo metallico di sezione  $S$  costante con una ddp  $V_{app}$  ai suoi estremi.

La densità di corrente è per definizione la carica mobile che passa nell'unità di tempo attraverso l'unità di superficie

$$J = qnv_{drift} = qn\frac{\mu_n V_{app}}{L}$$

e l'intensità  $I$

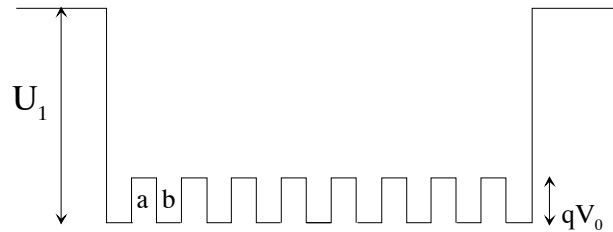
$$I = JS = Sq n \frac{\mu_n V_{app}}{L} = qn\mu_n \frac{S}{L} V_{app} = \sigma \frac{S}{L} V_{app} = \frac{V_{app}}{R}.$$

-----

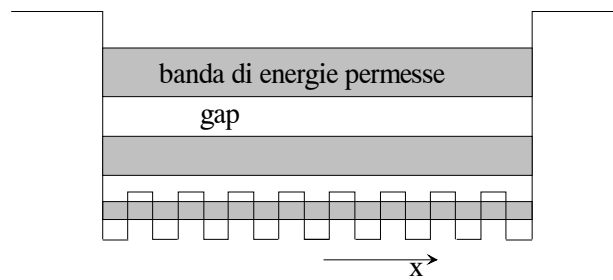
Non è ovviamente possibile descrivere tutti i solidi con questo modello, dato che si comporterebbero tutti come metalli. Un modello, ancora schematico, ma più aderente alla realtà è il

### Modello di Kronig e Penney

che differisce dal precedente in quanto l'energia potenziale all'interno della buca non è costante, ma varia periodicamente con la posizione.



Risolviendo l'equazione di Schroedinger per un elettrone in interazione con un potenziale come quello della figura (i parametri che definiscono il modello sono:  $a$ ,  $b$ ,  $qV_0$ ) si ottiene che, a parte la quantizzazione dovuta al confinamento dell'elettrone, i livelli di energia permessi sono raggruppati in *bande* separate fra loro da intervalli di energie proibite (*gap*).



Questi intervalli assumono valori dell'ordine degli eV. Vediamo invece quale sia il  $\Delta E$  relativo alla quantizzazione.

Per un metallo, nel caso in cui tutte le componenti di  $k$  sono uguali:

$$E_1 = \frac{\hbar^2 k^2}{2m_0} = \frac{\hbar^2}{2m_0} (k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2}{2m_0} \left( \left( \frac{n_1 2\pi}{L} \right)^2 + \left( \frac{n_1 2\pi}{L} \right)^2 + \left( \frac{n_1 2\pi}{L} \right)^2 \right) = \frac{3\hbar^2 n_1^2}{2m_0} \left( \frac{2\pi}{L} \right)^2 ;$$

$$E_2 = \frac{\hbar^2}{2m_0} \left( \left( \frac{(n_1 + 1) 2\pi}{L} \right)^2 + \left( \frac{(n_1 + 1) 2\pi}{L} \right)^2 + \left( \frac{(n_1 + 1) 2\pi}{L} \right)^2 \right) = \frac{3\hbar^2 (n_1 + 1)^2}{2m_0} \left( \frac{2\pi}{L} \right)^2$$

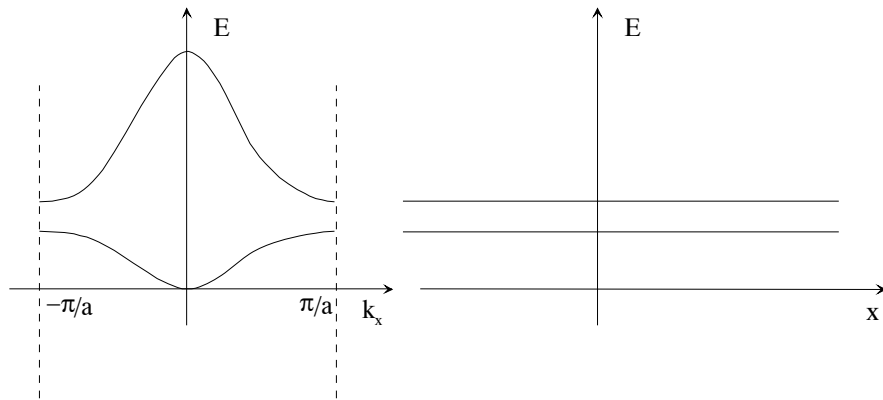
$$\begin{aligned} \Delta E &= \frac{3\hbar^2}{2m_0} \left( \frac{2\pi}{L} \right)^2 ((n_1 + 1)^2 - n_1^2) \\ &= \frac{3\hbar^2}{2m_0} \left( \frac{2\pi}{L} \right)^2 (2n_1 + 1) \\ &= \frac{3 \times (6.62 \times 10^{-34})^2}{8 \times \pi^2 \times 9.1 \times 10^{-31}} \left( \frac{2\pi}{10^{-2}} \right)^2 (2n_1 + 1) \\ &= 7.2 \times 10^{-33} (2n_1 + 1) \text{ joule} = \frac{7.2 \times 10^{-33}}{1.6 \times 10^{-19}} (2n_1 + 1) \\ &= 4.5 \times 10^{-14} (2n_1 + 1) \text{ eV.} \end{aligned}$$

avendo assunto  $L = 1 \text{ cm}$ . Se supponiamo che  $E_1 = 1 \text{ eV}$ , otteniamo per  $n_1$

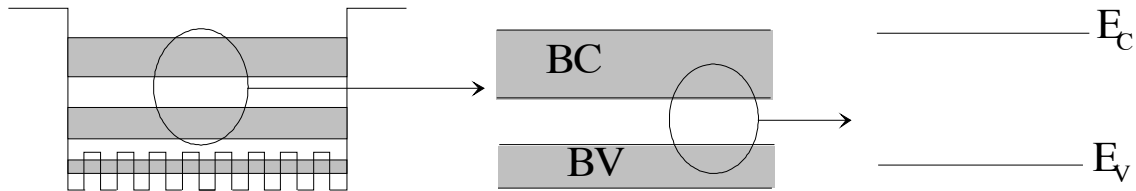
$$\begin{aligned} n_1 &= \sqrt{E_1 \frac{2m_0}{3\hbar^2} \left( \frac{L}{2\pi} \right)^2} \\ &= \sqrt{1.6 \times 10^{-19} \frac{2 \times 9.1 \times 10^{-31}}{3 \times \left( \frac{6.62 \times 10^{-34}}{2\pi} \right)^2} \times \left( \frac{10^{-2}}{2\pi} \right)^2} = 4.7 \times 10^6 \end{aligned}$$

e quindi  $\Delta E = 4.5 \times 10^{-14} \times 2 \times 4.7 \times 10^6 = 4.23 \times 10^{-7} \text{ eV}$ , che è assolutamente trascurabile rispetto al valore assunto dai gap.





I grafici mostrano la struttura a bande  $E = E(x)$  ed  $E = E(k)$  per un solido 1D. Diamo qui solo le conclusioni salienti del modello.



- 1) Ogni solido è caratterizzato da una struttura a bande.
- 2) Le bande possono essere completamente occupate dagli elettroni; la più alta in energia, immediatamente successiva all'ultima completamente occupata, può essere occupata per metà.
- 3) I metalli o non hanno gap (l'ultima occupata e quella successiva vuota si sovrappongono) oppure l'ultima è occupata per metà.
- 4) I semiconduttori e gli isolanti hanno le ultime due bande (quella piena  $\equiv$  banda di valenza e quella vuota  $\equiv$  banda di conduzione) separate da un gap; il fatto che si comportino da semiconduttori o da isolanti dipende dal valore del gap  $E_g$ . Per il Si  $E_g \simeq 1$  eV; per l'SiO<sub>2</sub>  $E_g \simeq 9$  eV.
- 5) In un metallo  $E_F$  si trova fra i livelli permessi; in un semiconduttore generalmente  $E_F$  è nel gap.
- 6) nella banda di valenza a 0 K trovano posto tutti e solo gli elettroni di valenza (la banda è completamente occupata).

## 6) I portatori di carica nei semiconduttori

Le proprietà dei semiconduttori, e in particolare la concentrazione dei portatori all'equilibrio, sono descrivibili a partire dalla struttura a bande utilizzando la  $f(E)$  e la  $D(E)$ , già definite. Ricordiamo che il riferimento per la  $f(E)$  è il livello di Fermi, a cui la  $f(E)$  è, per così dire, vincolata. Le bande di interesse

sono quella di valenza (BV) e quella di conduzione (BC), all'interno delle quali gli stati sono caratterizzati da una densità  $D(E)$  che ha una dipendenza dall'energia analoga a quella dei metalli, salvo l'origine delle energie, che per la BC è il suo valore di energia più basso  $E_C$  e per la BV il suo valore più alto  $E_V$ . Avremo dunque per la BC

$$D(E) = 2\pi \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} (E - E_C)^{\frac{1}{2}}$$

e analoga per la BV. In questo caso, come sarà più chiaro nel prosieguo, i portatori sono delle particelle cariche positivamente (*lacune, holes* in inglese) e l'asse delle energie è orientato in verso opposto rispetto a quello degli elettroni della BC.  $m_{ed}^*$  prende il nome di massa efficace per la densità degli stati.

Supponiamo di essere alla temperatura assoluta di 0 K e di aver a che fare con un campione di Si ideale, ossia privo di impurezze (Si *intrinseco*). Tutti gli elettroni di valenza (quattro per ogni atomo di Si) sono nella BV: la  $f(E)$  vale 1 nella BV e 0 nella BC. E' ragionevole supporre che  $E_F$  si trovi a metà del gap. All'aumentare di  $T$  alcuni elettroni acquistano energia sufficiente per passare nella BC; questo trasferimento è promosso dall'energia di agitazione termica e quindi la concentrazione  $n$  di elettroni in BC sarà funzione di  $T$ :  $n = n(T)$ . Poiché ogni elettrone in BC lascia libero uno stato in BV, se indichiamo con  $p$  la concentrazione di questi stati liberi, sarà anche necessariamente  $p = p(T)$ . E inoltre  $n(T) = p(T)$ , nel caso in questione (Si intrinseco).

### Calcolo di $n$ nel caso generale

Definiamo un semiconduttore *non - degenere* quando  $E_F$  si trova all'interno del gap, ovvero nell'intervallo di energie in cui non ci sono stati permessi. Cerchiamo un'espressione che ci consenta di calcolare  $n$  noto  $E_F$  e viceversa. Il procedimento che segue vale per la BC. Per la BV il procedimento sarà identico, salvo alcune modifiche che verranno precisate in seguito.

Scriviamo la

$$n = \int_0^\infty n(E) dE = 2 \int_0^\infty D(E) f(E) dE$$

per la BC:

$$n = 2 \int_{E_C}^\infty 2\pi \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} (E - E_C)^{\frac{1}{2}} \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} dE$$

(la densità di stati si annulla per  $E = E_C$ ). L'integrale non è valutabile in generale; lo è solo a patto di fare delle approssimazioni. Nel caso in questione l'approssimazione è che  $E_C - E_F > kT$ , ovvero il livello di Fermi è all'interno del gap.

$$n = kT \times 2 \int_{E_C}^\infty 2\pi \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} (E - E_C)^{\frac{1}{2}} \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} d\left(\frac{E - E_C}{kT}\right)$$

$$n = kT \times (kT)^{\frac{1}{2}} \times 2 \int_{E_C}^{\infty} 2\pi \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} \left( \frac{E - E_C}{kT} \right)^{\frac{1}{2}} \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} d\left(\frac{E - E_C}{kT}\right)$$

$$n = (kT)^{\frac{3}{2}} \times 2 \int_0^{\infty} 2\pi \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} (x)^{\frac{1}{2}} \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} dx$$

avendo posto  $\frac{E - E_C}{kT} = x$ . Aggiungendo e togliendo  $E_C$  all'argomento dell'esponenziale nella  $f(E)$

$$\frac{1}{1 + \exp\left(\frac{E - E_F + E_C - E_C}{kT}\right)} = \frac{1}{1 + \exp\left(\frac{E - E_C - (E_F - E_C)}{kT}\right)} = \frac{1}{1 + \exp(x - y)}$$

con  $x = \frac{E - E_C}{kT}$  e  $y = \frac{E_F - E_C}{kT}$ . Dopo queste posizioni si ha

$$n = 4\pi (kT)^{\frac{3}{2}} \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} \int_0^{\infty} (x)^{\frac{1}{2}} \frac{1}{1 + \exp(x - y)} dx.$$

Nell'ipotesi che  $y < -1$  ( $E_C - E_F > kT$ ) il termine  $\exp(x - y)$  al denominatore è sicuramente  $> 1$  per ogni  $x$  e quindi

$$\begin{aligned} n &= 4\pi (kT)^{\frac{3}{2}} \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} \exp(y) \int_0^{\infty} (x)^{\frac{1}{2}} \exp(-x) dx \\ &= 4\pi (kT)^{\frac{3}{2}} \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} \exp(y) \frac{\sqrt{\pi}}{2} \\ &= 2 \left( \frac{2\pi m_{ed}^* kT}{h^2} \right)^{\frac{3}{2}} \exp(y) = N_C \exp\left(-\frac{E_C - E_F}{kT}\right) \\ n &= N_C \exp\left(-\frac{E_C - E_F}{kT}\right), \end{aligned}$$

che rappresenta la relazione fondamentale per il calcolo della concentrazione degli elettroni in BC, nota che sia la posizione di  $E_F$ .  $N_C$  prende il nome di densità equivalente di stati (nella BC) e, per il Si, vale  $2.8 \times 10^{19} \text{ cm}^{-3}$  a  $T = 300 \text{ K}$ . L'espressione è valida nell'ipotesi iniziale,  $E_C - E_F > kT$ . Per le lacune in BV si trova analogamente (asse delle energie verso il basso, FD delle lacune =  $1 - f(E)$ )

$$p = N_V \exp\left(-\frac{E_F - E_V}{kT}\right),$$

con  $N_V = 10^{19} \text{ cm}^{-3}$  a  $T = 300 \text{ K}$ . Definiamo a questo punto il concetto di

### Massa efficace

Il punto di partenza è il comportamento di un elettrone in interazione con una struttura periodica. Il modello di KP dà luogo alla struttura a bande; all'interno di ciascuna banda si ha un andamento dell'energia in funzione del vettore d'onda dato dalla funzione  $E = E(k)$ . Si osservi che gli elettroni nella BC, ad esempio, occupano con probabilità non nulla solo gli stati con energie molto prossime ad  $E_C$ . Si può dunque sviluppare in serie la funzione  $E = E(k)$ , nell'intorno di un minimo o di un massimo, dove la derivata prima è nulla:

$$E(k) = E(k_0) + \left. \frac{d^2 E}{dk^2} \right|_{k=k_0} \frac{(k - k_0)^2}{2} + TOS;$$

l'andamento di  $E$  è quadratico con  $k$ , come nel modello dell'elettrone libero, con la differenza che in questo caso la dipendenza quadratica vale solo nell'intorno di un punto di massimo (o di minimo). Ricordando che per un metallo

$$E = \frac{\hbar^2 k^2}{2m_0}$$

si può stabilire un'analogia formale

$$E(k) - E(k_0) = \left. \frac{d^2 E}{dk^2} \right|_{k=k_0} \frac{(k - k_0)^2}{2} = \frac{\hbar^2 (k - k_0)^2}{2m_e^*}$$

pur di porre

$$m_e^* = \frac{\hbar^2}{\left. \frac{d^2 E}{dk^2} \right|_{k=k_0}}.$$

$m_e^*$  prende il nome di massa efficace dell'elettrone (per la mobilità) nella BC ed è in generale diversa da  $m_0$ . E' un parametro che tiene conto di tutte le complesse interazioni che subisce un elettrone in un potenziale periodico. In altre parole, quando si studiano i problemi legati al movimento dei portatori di carica soggetti ad un campo elettrico, la massa efficace consente di tenere in conto *soltanto* del campo elettrico applicato dall'*esterno*.

Dato che nella definizione compare la derivata seconda,  $m_e^*$  può essere positiva o *negativa*, a seconda che lo sviluppo in serie sia stato effettuato nei pressi di un minimo o di un *massimo*, rispettivamente. Il fatto che si possa avere una massa (efficace) negativa non deve turbare più di tanto: si tratta solo di un parametro atto a descrivere il comportamento dinamico dell'elettrone. Ovviamente la massa gravitazionale dell'elettrone vale sempre  $m_0$ . Rimane tuttavia il problema del trasporto di corrente in un semiconduttore nel quale (supponiamolo intrinseco) a una certa temperatura  $T > 0$ , ci sono sia elettroni sul fondo della BC (con massa e. positiva) che stati liberi alla sommità della BV (nei quali la massa e. dell'elettrone sarebbe stata negativa).

Supponiamo di essere allo zero assoluto: la BV è completamente piena e la BC completamente vuota. Se applichiamo un campo elettrico nessun elettrone della BV è in grado di incrementare il suo vettore d'onda (e la sua energia) dato che nessuno stato è disponibile e quindi la corrente elettrica risultante sarà pari a zero. Potremo dunque affermare che il contributo  $C(N)$  di *tutti* gli  $N$  elettroni della BV è nullo.

$$C_{BV}(N) = 0.$$

Osserviamo ora che di tutti gli elettroni della BV, alcuni hanno massa e. positiva, altri negativa, mentre per la maggior parte di essi non è neppure definibile. Possiamo tuttavia scrivere, estraendo il contributo di un singolo elettrone che si trovi alla sommità della BV (e quindi destinato a passare in BC quando  $T$  diviene  $> 0$ )

$$C_{BV}(N - 1) + C_{BV}(-m_e^*, -q) = 0$$

in cui  $C(-m_e^*, -q)$  rappresenta il contributo di un elettrone con massa e. negativa e carica (ovviamente) negativa. Proseguendo in questa trattazione formale

$$C_{BV}(N - 1) = -C_{BV}(-m_e^*, -q)$$

ovvero

$$C_{BV}(N - 1) = -C_{BV}(-m_e^*, -q) + C_{BV}(m_h^*, q) - C_{BV}(m_h^*, q)$$

in cui abbiamo sommato e sottratto il contributo di una particella fittizia di massa e. positiva ( $m_h^*$ ) e carica positiva (lacuna). Nel secondo membro dell'uguaglianza precedente

$$-C_{BV}(-m_e^*, -q) - C_{BV}(m_h^*, q) = 0$$

in quanto si tratta del contributo di una coppia di particelle con massa e. e carica pari a zero. Rimane dunque

$$C_{BV}(N - 1) = C_{BV}(m_h^*, q) :$$

*il contributo di  $N - 1$  elettroni al trasporto di corrente è equivalente a quello di una particella, la lacuna, con massa efficace positiva  $|-m_e^*| = m_h^*$  e carica positiva.*

Riassumendo:

- 1) elettroni nei pressi del minimo della BC: massa e.  $m_e^*$ , carica  $-q$ ;
  - 2) stati vuoti nei pressi della sommità della BV: massa e.  $m_h^*$ , carica  $q$ .
- $m_e^*$  e  $m_h^*$  prendono il nome di masse e. *per la conducibilità*.

Dovrebbe a questo punto risultare evidente che una banda completamente occupata non dà alcun contributo al passaggio di corrente; quindi un sc alla  $T$  di 0 K si comporta come un isolante. La differenza fra sc ed isolanti sta nel valore del gap  $E_g$ . L'SiO<sub>2</sub>, che è un isolante a tutte le temperature accessibili, ha un gap di

circa 9 eV; questo significa che qualunque sia la sua temperatura la Fermi-Dirac non è in grado di mandare elettroni nella BC.

Il Si ha un gap di 1.08 eV alla temperatura di 300 K (il gap dipende debolmente dalla temperatura).

### Mobilità e conducibilità

In un sc in cui siano presenti entrambi i tipi di portatori si avrà:

$$\mu_n = \frac{e\tau}{m_e^*}$$

$$\mu_h = \frac{e\tau}{m_h^*}$$

$$\sigma = \sigma_e + \sigma_h = e(n\mu_n + p\mu_h)$$

### $n$ e $p$ nel Si intrinseco

Abbiamo visto che per un sc non degenera

$$n = N_C \exp\left(-\frac{E_C - E_F}{kT}\right)$$

$$p = N_V \exp\left(-\frac{E_F - E_V}{kT}\right).$$

Per il Si intrinseco  $n = p = n_i$ , quindi facendo il prodotto

$$np = n_i^2 = N_C N_V e^{\left(-\frac{E_C - E_F}{kT}\right)} e^{\left(-\frac{E_F - E_V}{kT}\right)} = N_C N_V e^{\left(-\frac{E_C - E_V}{kT}\right)} = N_C N_V e^{\left(-\frac{E_g}{kT}\right)}$$

$$n_i = \sqrt{N_C N_V} e^{\left(-\frac{E_g}{2kT}\right)}.$$

Il risultato è *indipendente* dalla posizione del livello di Fermi, purché ci si trovi all'equilibrio. La relazione  $np = n_i^2$  prende il nome di *legge dell'azione di massa*. Il suo significato verrà chiarito in seguito.

-----

Esempio

Per il Si a T ambiente

$$n_i = \sqrt{2.8 \times 10^{19} \times 10^{19}} \times e^{\left(-\frac{1.08}{2 \times 8.633 \times 10^{-5} \times 300}\right)} = 1.47 \times 10^{10} \text{ cm}^{-3}$$

che usualmente si approssima con  $1.5 \times 10^{10} \text{ cm}^{-3}$  ( $n_i^2 = 2.25 \times 10^{20} \text{ cm}^{-6}$ ).

-----

Invertendo le due relazioni

$$kT \ln \left( \frac{N_C}{n} \right) = E_C - E_F$$

$$kT \ln \left( \frac{N_V}{n} \right) = E_F - E_V$$

e facendo la differenza

$$E_C - 2E_F + E_V = kT \ln \left( \frac{N_C}{N_V} \right)$$

$$E_F = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln \left( \frac{N_V}{N_C} \right);$$

il livello di Fermi si trova a metà del gap solo alla temperatura di 0 K. Esplicitando  $N_C$  ed  $N_V$

$$E_F = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln \left( \frac{N_V}{N_C} \right) = \frac{E_g}{2} + \frac{0.026}{2} \times \ln \left( \frac{10^{19}}{2.8 \times 10^{19}} \right) = \frac{E_g}{2} - 1.34 \times 10^{-2}$$

si vede che il livello di Fermi ad ogni  $T$  si trova spostato, sia pur di poco, verso la BV e questo è ovvio dato che per il Si intrinseco

$$n = p = n_i$$

$$N_V \exp \left( -\frac{E_F - E_V}{kT} \right) = N_C \exp \left( -\frac{E_C - E_F}{kT} \right)$$

con  $N_C > N_V$ .

Si osservi ancora che

$$\ln n_i = \frac{1}{2} \ln (N_C N_V) - \frac{E_g}{2kT} = \frac{3}{2} \ln A + \frac{3}{2} \ln T - \frac{E_g}{2kT}$$

$$\ln n_i = B - \frac{3}{2} \ln \frac{1}{T} - \frac{E_g}{2kT} = B - 1.5 \ln \frac{1}{T} - 6.26 \times 10^3 \frac{1}{T}$$

e quindi  $\ln n_i$  è sostanzialmente una funzione lineare di  $1/T$  con pendenza  $-\frac{E_g}{2k}$ .

Vengono riportati di seguito alcuni valori di  $n_i$  a varie temperature calcolati con  $E_g = 1.08$  eV.

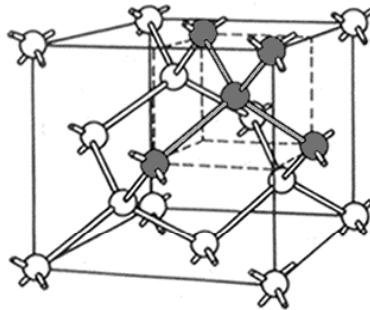
$T$ (K)	$n_i$ (cm <sup>-3</sup> )
1400	$1.93 \times 10^{18}$
1200	$7.29 \times 10^{17}$
1000	$1.95 \times 10^{17}$
700	$7.85 \times 10^{15}$
500	$1.33 \times 10^{14}$
400	$4.16 \times 10^{12}$

### La struttura cristallina e il drogaggio del Si

Il Si intrinseco non è generalmente di grande utilità per quanto riguarda la realizzazione di dispositivi. Infatti  $n_i$  dipende fortemente dalla temperatura. Tuttavia l'introduzione controllata di impurezze nel Si ne cambia drasticamente le proprietà elettriche e segnatamente la concentrazione e il tipo dei portatori. Si parla di *drogaggio* del Si. I droganti usuali sono P, As, Sb (gruppo V) per ottenere Si di tipo  $n$ , B (gruppo III) per ottenere Si di tipo  $p$ .

La concentrazione di atomi di Si in un monocristallo si può calcolare a partire dalla conoscenza della sua struttura cristallina.

Date due strutture FCC coincidenti si immagini di separarle facendo scorrere rigidamente una delle due in una direzione  $\langle 111 \rangle$  per un quarto della lunghezza della diagonale principale del cubo (vedi figura). Quella che risulta è la struttura del Si (che è la stessa del diamante).



Si individuino per esercizio nella figura gli atomi che appartengono ad una delle due strutture.

Caratteristica fondamentale di questa struttura è che ogni atomo di Si si trova al centro di un tetraedro regolare ed è quindi legato ad altri quattro da un legame covalente: il Si ha quattro elettroni di valenza e ne mette in comune uno con un altro atomo contiguo. Un legame covalente è formato dunque da due elettroni. La costante reticolare della struttura vale  $5.428 \text{ \AA}$  (lato della FCC). La concentrazione di atomi di Si vale

$$N_{\text{Si}} = \frac{8}{(5.428 \times 10^{-8})^3} = 5.0 \times 10^{22} \text{ cm}^{-3}.$$

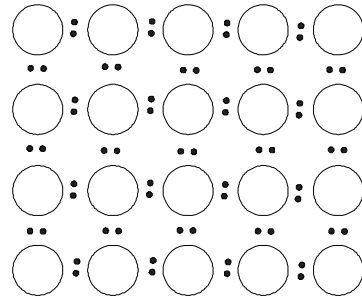
Per le strutture cristalline si può introdurre il modello a sfere rigide, che consiste nell'immaginare ciascun atomo come una sfera con raggio pari alla metà della distanza fra atomi *primi vicini*. Per il silicio si indica questa quantità con  $r_t$  (raggio tetraedrico):

$$r_t = \frac{a\sqrt{3}}{8} = \frac{5.428 \times \sqrt{3}}{8} = 1.175 \text{ \AA}.$$

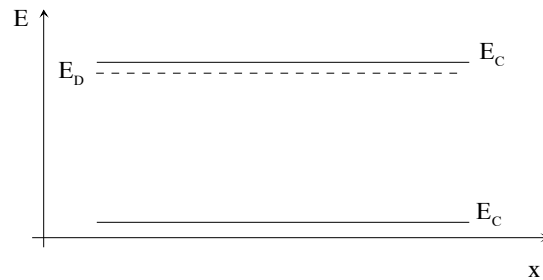


Il motivo di questa schematizzazione è che quando il Si viene drogato si introducono nella struttura atomi estranei che hanno in generale  $r_t$  diverso e ciò provoca effetti di deformazione locale del reticolo.

Per motivi di praticità grafica si suole schematizzare la struttura del Si con quella bidimensionale della figura seguente



Supponiamo di aver sostituito un atomo di Si con un atomo di P (si parla di impurezza *sostitutiva*) e di essere a 0 K. Dato che il P ha un elettrone esterno in più, il numero complessivo di elettroni di valenza della struttura è aumentato di un'unità; non può essere quindi sistemato nella BV che, per quanto visto, è completamente occupata, né può andare in BC dato che la Fermi-Dirac non lo consente. In effetti se la struttura energetica del solido non cambiasse non ci sarebbe nessuna possibilità per l'elettrone in più. Ma il solido non è evidentemente lo stesso di prima perché è formato da  $N - 1$  atomi di silicio e da *un* atomo di fosforo. Quello che accade nella sua struttura energetica è mostrato dalla figura:



per ogni atomo di P si ha uno *stato localizzato* all'*interno* del gap. Questo stato è molto vicino energeticamente alla BC (dell'ordine del centesimo di eV).

L'elettrone eccedente dell'atomo di P è vincolato al suo atomo di origine dall'attrazione elettrostatica, ma poiché si trova in un mezzo, il silicio, con costante dielettrica relativa

$$\epsilon_{\text{Si}} = 11.8$$

questa attrazione è ridotta rispetto al caso di atomo di P isolato. In altri termini per liberare questo elettrone è necessaria un'energia molto minore di quella che ci vorrebbe per l'atomo isolato. Per l'elettrone nel solido essere libero significa trovarsi nella BC. Lo stato su cui si trova l'elettrone a 0 K è individuato nella

figura dalle coordinate di posizione (asse delle ascisse) e dal valore dell'energia  $E_D$ . E' evidente che ci sono nel silicio tanti stati localizzati quanti sono gli atomi di P introdotti. Per il P ad esempio  $E_C - E_D = 0.045$  eV. Basta quindi un piccolo aumento di temperatura per portare l'elettrone in BC. Il P prende il nome di atomo donatore o, più brevemente, donatore.

La BC viene adesso rifornita di elettroni di conduzione sia dalla BV, distante in energia circa 1 eV, che dagli stati donatori la cui distanza in energia è molto minore.

Se indichiamo con  $N_D$  la *concentrazione* in  $\text{cm}^{-3}$  di atomi di P che possono essere introdotti nel Si

$$10^{14} \lesssim N_D \lesssim 10^{20};$$

come si vede un intervallo molto esteso. Il Si drogato con atomi di P (o di As, o di Sb) prende il nome di Si di tipo  $n$ . Si pone il problema, in questa situazione mutata, di conoscere la funzione

$$n = n(T).$$

### Si non-degenere di tipo $n$

L'equazione della neutralità di carica si scrive

$$n = N_D^+ + p;$$

la concentrazione  $n$  degli elettroni nella BC è data dalla somma della concentrazione di elettroni provenienti dagli stati donatori (= alla frazione  $N_D^+$  di donatori che hanno perso il loro elettrone) più la concentrazione di elettroni provenienti dalla BV (=  $p$ ). Si noti che i donatori ionizzati costituiscono una *carica fissa*.

Dato che si tratta di un Si non-degenere  $n = N_C \exp\left(-\frac{E_C - E_F}{kT}\right)$

$$N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_D^+ + p;$$

se supponiamo poi di essere in un intervallo di  $T$  tali che  $p$  sia trascurabile (basse  $T$ ) l'equazione precedente diviene

$$N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_D^+$$

in cui l'incognita è  $E_F$ . E' necessario ora esprimere  $N_D^+$  in funzione di  $E_F$ . La probabilità che uno stato con energia  $E_D$  sia occupato da un elettrone è data da

$$f(E_D) = \frac{1}{1 + \frac{1}{2} \exp\left(\frac{E_D - E_F}{kT}\right)}$$

e quindi la probabilità che sia vuoto è

$$1 - f(E_D) = 1 - \frac{1}{1 + \frac{1}{2} \exp\left(\frac{E_D - E_F}{kT}\right)} = \frac{\frac{1}{2} \exp\left(\frac{E_D - E_F}{kT}\right)}{1 + \frac{1}{2} \exp\left(\frac{E_D - E_F}{kT}\right)} = \frac{1}{2 \exp\left(\frac{E_F - E_D}{kT}\right) + 1};$$

l'equazione  $N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_D^+$  diviene infine

$$N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = \frac{N_D}{2 \exp\left(\frac{E_F - E_D}{kT}\right) + 1}.$$

Ponendo  $e^{\frac{E_F}{kT}} = x$

$$\begin{aligned} x N_C \exp\left(-\frac{E_C}{kT}\right) &= \frac{N_D}{2x \exp\left(-\frac{E_D}{kT}\right) + 1} \\ 2x^2 N_C \exp\left(-\frac{E_C + E_D}{kT}\right) + x N_C \exp\left(-\frac{E_C}{kT}\right) - N_D &= 0 \\ x^2 + x \frac{N_C \exp\left(-\frac{E_C}{kT}\right)}{2N_C \exp\left(-\frac{E_C + E_D}{kT}\right)} - \frac{N_D}{2N_C \exp\left(-\frac{E_C + E_D}{kT}\right)} &= 0 \\ x^2 + x \frac{\exp\left(\frac{E_D}{kT}\right)}{2} - \frac{N_D \exp\left(\frac{E_C + E_D}{kT}\right)}{2N_C} &= 0 \end{aligned}$$

che ha per soluzioni

$$\begin{aligned} x &= -\frac{\exp\left(\frac{E_D}{kT}\right)}{4} \pm \sqrt{\frac{\exp\left(\frac{2E_D}{kT}\right)}{16} + \frac{N_D \exp\left(\frac{E_C + E_D}{kT}\right)}{2N_C}} \\ &= \frac{\exp\left(\frac{E_D}{kT}\right)}{4} \left(-1 \pm \sqrt{1 + \frac{8N_D \exp\left(\frac{E_C - E_D}{kT}\right)}{N_C}}\right); \end{aligned}$$

scartando il segno  $-$  nella parentesi

$$x = \frac{\exp\left(\frac{E_D}{kT}\right)}{4} \left(-1 + \sqrt{1 + \frac{8N_D \exp\left(\frac{E_C - E_D}{kT}\right)}{N_C}}\right).$$

L'espressione trovata ha validità nell'intervallo di basse  $T$  (a partire da 0 K). Il termine sotto radice dipende da  $T$  (si ricordi che  $N_C = N_C(T)$ ). Al crescere di  $T$  il termine  $\frac{8N_D \exp\left(\frac{E_C - E_D}{kT}\right)}{N_C}$  diminuisce; quando diviene  $\ll 1$  si può sviluppare in serie la radice ( $\sqrt{1 + y} \simeq 1 + \frac{y}{2}$ ) e ottenere

$$x = \frac{\exp\left(\frac{E_D}{kT}\right)}{4} \left(-1 + 1 + \frac{4N_D \exp\left(\frac{E_C - E_D}{kT}\right)}{N_C}\right) = \frac{N_D}{N_C} \exp\left(\frac{E_C}{kT}\right);$$

ricordando che  $e^{\frac{E_F}{kT}} = x$

$$e^{\frac{E_F}{kT}} = \frac{N_D}{N_C} e^{\left(\frac{E_C}{kT}\right)}$$

$$N_C e^{\left(-\frac{E_C - E_F}{kT}\right)} = N_D$$

$$n = N_D :$$

*c'è un intervallo di temperature nel quale la concentrazione di elettroni nella BC è uguale alla concentrazione dei donatori mentre la concentrazione di lacune è trascurabile. Questo significa che il trasporto di corrente è dovuto esclusivamente agli elettroni.  $n$  può essere fissata mediante il drogaggio nell'intervallo  $10^{14} \div 10^{20} \text{ cm}^{-3}$ . Gli elettroni sono i portatori *maggioritari*, le lacune i *minoritari*.*

Se non può esser fatta l'approssimazione  $n = N_D^+$  (alte  $T$ ) è necessario risolvere l'equazione di neutralità completa

$$n = N_D^+ + p.$$

Utilizzando la legge dell'azione di massa  $np = n_i^2$

$$n = N_D + \frac{n_i^2}{n}$$

( $N_D^+$  si scrive ora  $N_D$  in quanto tutti i donatori sono certamente ionizzati)

$$n^2 - nN_D - n_i^2 = 0$$

che ha per soluzioni

$$n = \frac{N_D}{2} \pm \sqrt{\frac{N_D^2}{4} + n_i^2}$$

e quindi

$$n = \frac{N_D}{2} + \sqrt{\frac{N_D^2}{4} + n_i^2}.$$

Questa soluzione deve coincidere, per le basse  $T$  di questo intervallo, con quella precedente; infatti se  $\frac{N_D^2}{4} \gg n_i^2(T)$  si ha

$$n = N_D.$$

Per alte  $T$  invece:  $\frac{N_D^2}{4} \ll n_i^2(T) \Rightarrow$

$$n = n_i :$$

il Si, pur essendo drogato, è tornato, ad alte  $T$ , a comportarsi come se fosse intrinseco. Ciò vuol dire che  $n$  e  $p$  sono adesso confrontabili. Come vedremo la  $T$  necessaria al verificarsi di questa circostanza dipende dal drogaggio.

E' importante però notare fin da adesso che i principali processi tecnologici avvengono a temperature dell'ordine dei 1000 °C, alle quali il Si può comportarsi da intrinseco o meno a seconda del drogaggio e questo fatto influenza in modo decisivo i processi stessi.

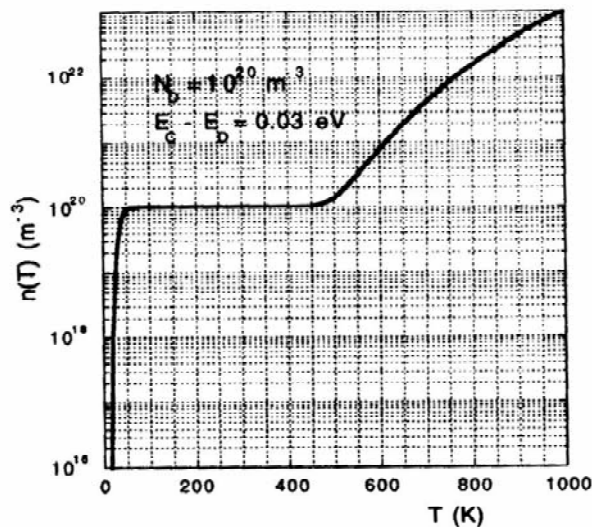
L'intervallo di temperature per il quale  $n = N_D$  e  $p$  è trascurabile prende il nome di intervallo di temperature di svuotamento. All'aumentare di  $T$  la BV è in grado di mandare sempre più elettroni in BC cosicché il suo contributo non può più essere trascurato. Dovremo perciò scrivere

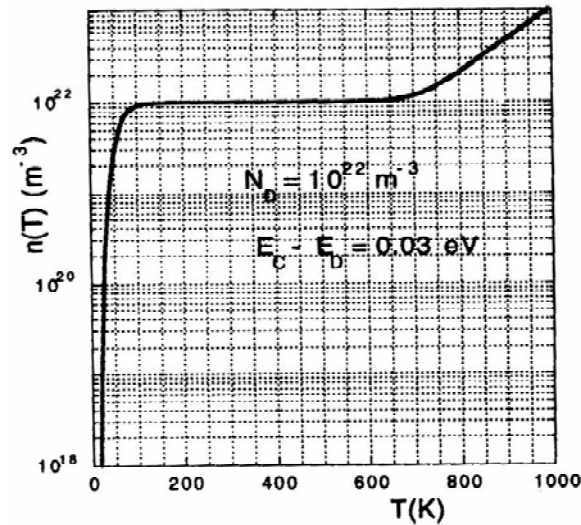
$$n = N_D + p.$$

Quando  $p$  diviene uguale ad  $N_D$  la concentrazione di elettroni sarà raddoppiata; se ad esempio  $N_D = 10^{15} \text{ cm}^{-3}$  sarà  $n = 2 \times 10^{15}$  e  $p = 10^{15}$ : il Si in queste condizioni si comporta praticamente da intrinseco. La temperatura  $T^*$  alla quale questo avviene può essere ricavata dalla relazione

$$pn = n_i^2 = 2N_D^2 = N_C N_V \exp\left(-\frac{E_g}{kT^*}\right)$$

ma non esplicitamente in quanto sia  $N_C$  che  $N_V$  dipendono da  $T$ . I grafici mostrano l'andamento  $n = n(T)$  per due diversi valori di drogaggio.





E' chiaramente individuabile l'intervallo di temperature di svuotamento.

### Si non-degenere di tipo $p$

La procedura è la stessa utilizzata per il Si di tipo  $n$ . Il risultato notevole è, come nel caso precedente, che nell'intervallo delle  $T$  di svuotamento,

$$p = N_A.$$

Qualche considerazione sul drogaggio di tipo  $p$ . Il boro è trivalente e quindi quando è in posizione sostituzionale ha un legame incompleto. Per completare questo legame può intervenire un elettrone del legame Si-Si, ma per far questo deve acquistare un pò di energia. Si hanno due effetti: 1) si libera parzialmente un legame Si-Si e questo corrisponde alla formazione di una lacuna nella BV; 2) l'elettrone che si è trasferito a completare il legame Si-B resta bloccato in quella posizione formando così uno ione fisso negativo. Il tutto viene descritto brevemente dicendo che gli atomi di B introducono degli *stati localizzati accettori* a poca distanza dalla BV. La distanza  $E_C - E_F$  ( $E_F - E_V$ ) si ottiene invertendo le

$$N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_D$$

$$N_V \exp\left(-\frac{E_F - E_V}{kT}\right) = N_A$$

$$E_C - E_F = kT \ln\left(\frac{N_C}{N_D}\right)$$

$$E_F - E_V = kT \ln\left(\frac{N_V}{N_A}\right).$$

Ricordiamo che queste relazioni sono valide se  $E_C - E_F > kT$  ( $E_F - E_V > kT$ ) (sc non-degenere).

$$E_C - E_F @ 10^{14} = 0.326 \text{ eV}$$

$$E_C - E_F @ 10^{15} = 0.266 \text{ eV}$$

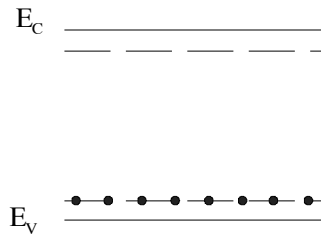
$$E_C - E_F @ 10^{17} = 0.146 \text{ eV.}$$

Si noti che il limite di validità dell'espressione utilizzata si situa nei pressi del valore  $N_D (N_A) = 10^{19} \text{ cm}^{-3}$ .

Il Si,  $p$  o  $n$ , con drogaggi superiori a  $10^{19} \text{ cm}^{-3}$  viene indicato rispettivamente con  $n^+$  e  $p^+$ .

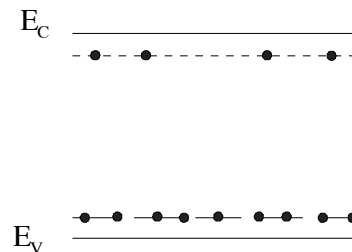
### Drogaggio netto e silicio compensato

Un campione di Si è drogato con accettori e donatori in modo che  $N_D = N_A$ . La struttura a bande è quella della figura, nella quale è stato messo in evidenza che gli elettroni dei donatori a 0 K hanno trovato posto sugli stati accettori.



All'aumentare della temperatura il Si, dal punto di vista della concentrazione dei portatori, si comporterà *come se* fosse intrinseco, dato che gli elettroni disponibili sono tutti in BV o molto vicini ad essa. Il Si in queste condizioni prende il nome di Si *compensato*.

Nella figura successiva è illustrato il caso  $N_D > N_A$ . E' immediato vedere che il Si si comporta come se fosse di tipo  $n$  con una concentrazione netta di donatori pari a  $N_D - N_A$ .



### Profilo dei droganti e profilo di drogaggio

Per profilo dei droganti si intende la concentrazione degli atomi donatori  $N_D(x)$  (accettori  $N_A(x)$ ) in funzione di  $x$ ; per profilo di drogaggio si intende la differenza  $N_A - N_D$  in funzione di  $x$ . Questa differenza determina il tipo di conducibilità del Si. Se  $N_A - N_D > 0$  il Si è di tipo  $p$ , se  $N_A - N_D < 0$  il Si è di tipo  $n$ . Solitamente, dato che la concentrazione dei droganti varia di diversi ordini di grandezza, i profili vengono rappresentati su un grafico semilogaritmico.

Il profilo del boro, dopo l'operazione di drogaggio di un wafer di tipo  $n$ , è dato dalla funzione ( $N_{AS} = N_A(0)$  e  $a$  una costante)

$$N_A(x) = N_{AS}e^{-\frac{x^2}{a^2}}$$

mentre il profilo del drogaggio netto è la differenza

$$N(x) = N_A(x) - N_D$$

riportata in grafico come

$$N(x) = \ln(|N_A(x) - N_D|)$$

Dall'uguaglianza

$$N_{AS}e^{-\frac{x_i^2}{a^2}} = N_D$$

si ottiene immediatamente il valore della coordinata  $x_i$ ,

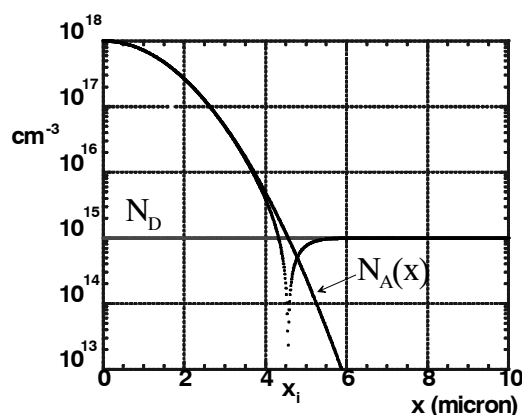
$$x_i = a\sqrt{\ln\left(\frac{N_{AS}}{N_D}\right)}$$

in cui la differenza  $N_A - N_D = 0$ . Il luogo dei punti in cui  $N_A - N_D = 0$  è la superficie della giunzione, che separa la zona di tipo  $p$  da quella di tipo  $n$ . Abbiamo qui supposto che la diffusione avvenga solo secondo  $x$ . In realtà la diffusione avviene anche secondo direzioni nel piano  $yz$  (diffusione laterale).

Esempio

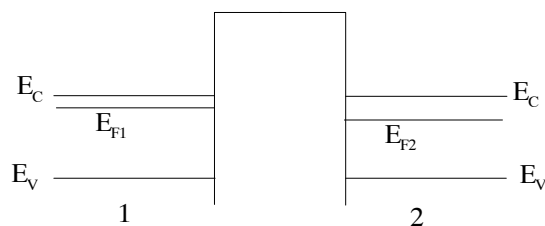
$N_D$  (costante) =  $10^{15}$  cm $^{-3}$ ,  $N_{AS} = 10^{18}$  cm $^{-3}$ ,  $a^2 = 3$  μm $^2$ . Dalla relazione precedente si ottiene immediatamente  $x_i = 4.55$  μm. Il grafico che segue mostra i vari profili prima e dopo l'operazione di drogaggio.





### 7) Solidi diversi a contatto

Consideriamo due campioni di Si *n* con drogaggio uniforme, ma con  $N_{D1} > N_{D2}$ ; la struttura a bande prima del contatto è quella della figura.



È evidente che il livello di Fermi è posizionato diversamente rispetto al comune riferimento del vuoto. In altre parole i due solidi *non hanno la stessa funzione di lavoro*. Se la distanza che separa le due superfici diviene dell'ordine delle distanze interatomiche si avrà come risultato un solido unico. Il problema che si pone è la posizione del livello di Fermi all'*equilibrio*. Consideriamo la superficie di separazione. Attraverso questa si dovrà avere un flusso netto di portatori uguale a zero; stessa cosa dicasi per il flusso netto di energia. La probabilità che un elettrone con energia  $E$  passi da 1 a 2 sarà data da

$$f_1(E)D_1(E) ((1 - f_2(E)) D_2(E))$$

dove  $f_1(E)D_1(E)$  è la concentrazione in 1 di elettroni con energia  $E$  e  $(1 - f_2(E)) D_2(E)$  la concentrazione in 2 di stati vuoti alla stessa energia. Per il passaggio da 2 a 1

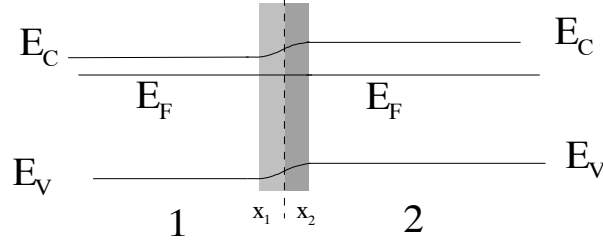
$$f_2(E)D_2(E) ((1 - f_1(E)) D_1(E))$$

e, in definitiva all'equilibrio

$$f_1(E)D_1(E) ((1 - f_2(E)) D_2(E)) = f_2(E)D_2(E) ((1 - f_1(E)) D_1(E))$$

$$f_1(E) (1 - f_2(E)) = f_2(E) (1 - f_1(E))$$

che è verificata solo se  $f_1 = f_2 \Rightarrow E_{F1} = E_{F2}$  : il livello di Fermi nei solidi a contatto è costante con la posizione:  $E_F(x) = C$ . Per disegnare il diagramma a bande presso il piano di giunzione, disegniamo prima il livello di Fermi, poi le bande a destra e sinistra della giunzione e finalmente raccordiamo  $E_C$  ed  $E_V$  tenendo presente che il gap del Si deve rimanere costante.



Prima del contatto c'è un gradiente di concentrazione di elettroni fra i due lati; a contatto avvenuto gli elettroni tendono a diffondere da 1 a 2. Come si può vedere  $E_C$ , fra le due linee tratteggiate, varia in funzione di  $x$ . Ciò significa che la concentrazione di elettroni, data all'equilibrio dalle relazioni

$$n_{1,2} = N_C e^{-\frac{E_C - E_{F1,2}}{kT}}$$

è dopo il contatto funzione di  $x$

$$n(x) = N_C e^{-\frac{E_C(x) - E_F}{kT}}$$

fra  $-x_1$  e  $x_2$ , e precisamente varia fra  $n_{n01}$  (in  $-x_1$ ) a  $n_{n02}$  (in  $x_2$ ). Per un qualunque valore  $x$  compreso fra 0 e  $-x_1$   $n(x)$  è minore di  $n_{n01}$ ; questo significa che per  $-x_1 \leq x \leq 0$  non c'è più la neutralità di carica come prima del contatto. Stessa cosa accade nell'intervallo  $0 \leq x \leq x_2$ .

Lontano dal piano di giunzione le concentrazioni rimangono quelle che si avevano all'equilibrio prima del contatto.

Invertiamo la relazione precedente

$$kT \ln \left( \frac{n(x)}{N_C} \right) = -E_C(x) + E_F,$$

deriviamo rispetto ad  $x$ :

$$kT \frac{1}{n(x)} \frac{dn(x)}{dx} = -\frac{dE_C(x)}{dx}$$

$$kT \frac{1}{n(x)} \frac{dn(x)}{dx} = q \frac{dV(x)}{dx} = -q\mathcal{E}(x)$$

$$kT \frac{1}{n(x)} \frac{dn(x)}{dx} = -q\mathcal{E}(x)$$

e moltiplichiamo entrambi i membri per  $\mu_n n(x)$

$$kT\mu_n \frac{dn(x)}{dx} + q\mu_n n(x)\mathcal{E}(x) = 0;$$

nel termine  $J_n(x) = q\mu_n n(x)\mathcal{E}(x)$  riconosciamo una densità di corrente di trascinamento, dovuta al campo presente nella zona di carica spaziale compresa fra  $-x_1$  e  $x_2$ . Ne consegue che anche il termine  $kT\mu_n \frac{dn(x)}{dx}$  rappresenterà una densità di corrente, la cui causa è il *gradiente* di concentrazione dei portatori  $\frac{dn(x)}{dx}$ ; dato che siamo all'equilibrio la somma delle due densità deve essere zero (la densità di corrente  $-q\mu_n n(x)\frac{dV(x)}{dx}$  è causata da un *gradiente* di potenziale).  $kT\mu_n \frac{dn(x)}{dx} = J_{diff}$  prende il nome di densità di corrente di diffusione.

Nella zona compresa fra  $-x_1$  ed  $x_2$  non si ha più neutralità di carica ed è quindi presente un campo elettrico  $\mathcal{E}(x)$ .

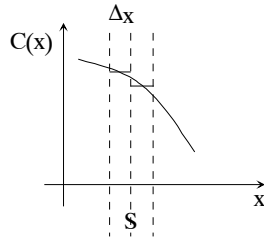
Uguagliando l'espressione trovata per  $J_{diff}$  con quella più comunemente nota

$$J_{diff} = kT\mu_n \frac{dn(x)}{dx} = qD_n \frac{dn(x)}{dx}$$

si ottiene

$$\frac{kT}{q} = \frac{D_n}{\mu_n}$$

che costituisce la relazione di Einstein fra coefficiente di diffusione  $D_n$  e mobilità.



Nella figura è rappresentata una concentrazione di particelle funzione di  $x$  (caso 1D). A causa delle collisioni ciascuna di esse sarà caratterizzata da un tempo libero  $\tau$  e da un cammino libero medio  $l$ . Se la particella si muove in una certa direzione, dopo l'urto avrà uguale probabilità di muoversi nella stessa direzione o in direzione opposta. Dette  $C(x)$  e  $C(x + \Delta x)$  le concentrazioni a sinistra e a destra della superficie  $S$  indicata, il numero netto di particelle che l'attraversa nell'unità di tempo sarà dato da

$$\frac{C(x)S\Delta x}{2\tau} - \frac{C(x + \Delta x)S\Delta x}{2\tau};$$

dividendo per  $S$  e moltiplicando e dividendo per  $\Delta x$  si ottiene il flusso  $F$

$$F = \frac{C(x)\Delta x^2}{2\tau\Delta x} - \frac{C(x + \Delta x)\Delta x^2}{2\tau\Delta x} = -\frac{\Delta x^2}{2\tau} \frac{dC(x)}{dx} = -\frac{l^2}{2\tau} \frac{dC(x)}{dx} = -D \frac{dC(x)}{dx}.$$

$D$  è il coefficiente di diffusione che dimensionalmente è  $l^2 t^{-1}$ . Il segno meno significa che il flusso ha verso contrario al gradiente di concentrazione (le particelle diffondono dalle zone a concentrazione più alta verso quelle a concentrazione più bassa).

-----

1) Il livello di Fermi in una catena di solidi diversi all'equilibrio è costante. Per rendere possibile questa condizione è necessario che si abbia un trasferimento di carica da un solido all'altro in modo da modificare l'energia potenziale.

2) Nella zona dove non sussiste più la neutralità di carica, ed è quindi presente un campo elettrico e un gradiente di concentrazione dei portatori, per ciascun tipo di portatori all'equilibrio si scriverà

$$J_{TOT} = J_{diff} + J_{drift} = 0$$

3) La differenza di energia dovuta all'allineamento dei livelli di Fermi si traduce in una ddp di contatto  $V_0$  che è data in modulo da

$$V_0 = \left| \frac{E_{F1} - E_{F2}}{q} \right|$$

E' possibile ottenere analiticamente l'andamento di  $V(x)$  e quindi del campo elettrico?

L'equazione di Poisson

$$\frac{d^2 V(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_s}$$

si scrive, ad esempio fra 0 e  $x_2$

$$\frac{d^2 V(x)}{dx^2} = -\frac{qN_{D2} - qn(x) + qp(x)}{\epsilon_s},$$

trascurando i minoritari, le lacune in questo caso

$$\frac{d^2 V(x)}{dx^2} = -\frac{qN_{D2} - qN_C e^{-\frac{E_C(x) - E_F}{kT}}}{\epsilon_s} = -\frac{q \left( N_{D2} - N_C e^{\frac{qV(x) + E_F}{kT}} \right)}{\epsilon_s}.$$

la densità di carica, la cui conoscenza è necessaria per determinare  $V(x)$ , è a sua volta funzione di  $V(x)$ . La risposta è quindi negativa. Il problema può essere risolto solo con opportune approssimazioni.

Si osservi che il caso trattato non è che la semplificazione della situazione che frequentemente si incontra nei dispositivi integrati, ovvero il drogaggio funzione della posizione.

### 8) Generazione e ricombinazione

$n_i$  è funzione di  $T$ ; più in generale  $n$  e  $p$  sono funzioni di  $T$ . Al crescere di  $T$  si ha *generazione* di coppie elettrone-lacuna, quando invece  $T$  diminuisce si ha *ricombinazione*. Indichiamo con  $G$  e con  $R$  le velocità di generazione e ricombinazione, rispettivamente.  $G$  rappresenta il numero di coppie che si generano nell'unità di volume nell'unità di tempo. Dimensionalmente:  $\text{cm}^{-3}\text{s}^{-1}$ . Analogamente per  $R$ .

La probabilità che un elettrone si ricombini nell'unità di tempo è proporzionale alla concentrazione di lacune  $p$ ; segue che

$$R = rnp;$$

se il sc è intrinseco

$$R = rn_i^2.$$

All'equilibrio deve essere evidentemente  $R = G$  e quindi  $G(T) = rn_i^2(T)$ .

Per un semiconduttore drogato ( $n \neq p$ )  $G$  rimane la stessa  $\Rightarrow$  all'equilibrio

$$G = R, \quad n_i^2(T) = np.$$

La *legge dell'azione di massa* ( $n_i^2 = np$ ) ha questo significato: all'*equilibrio* le velocità di generazione e ricombinazione sono *uguali*.

Notazione:

- $p_{n0}$  : conc. di minoritari all'equilibrio nel Si  $n$ ;
- $n_{n0}$  : conc. di maggioritari all'equilibrio nel Si  $n$ ;
- $n_{p0}$  : conc. di minoritari all'equilibrio nel Si  $p$ ;
- $p_{p0}$  : conc. di maggioritari all'equilibrio nel Si  $p$ ;
- $p_n$  : conc. di minoritari nel Si  $n$ ;
- $n_n$  : conc. di maggioritari nel Si  $n$ ;
- $n_p$  : conc. di minoritari nel Si  $p$ ;
- $p_p$  : conc. di maggioritari nel Si  $p$ .

#### 1) Illuminazione uniforme

Un campione di Si  $n$  ( $n_{n0} = N_D$ ,  $p_{n0} = n_i^2/N_D$ ) viene illuminato uniformemente in modo che si abbia una generazione uniforme di coppie  $G_{OP}$  in tutto il volume. A regime (non siamo più all'equilibrio)

$$\begin{aligned} n_n &= n_{n0} + \delta n \\ p_n &= p_{n0} + \delta p \end{aligned}$$

con  $\delta n = \delta p$ , che prendono il nome di concentrazione in eccesso di portatori maggioritari e minoritari, rispettivamente. Fondamentale per tutte le considerazioni che seguiranno è l'*ipotesi di bassa iniezione*:

$$\delta n = \delta p \ll n_{n0}$$

$$\delta n = \delta p \gg p_{n0}$$

Se, ad esempio,  $n_{n0} = 10^{15} \text{ cm}^{-3}$  e  $\delta n = \delta p = 10^{11} \text{ cm}^{-3}$  l'hp precedente è verificata. Si noti che

$$n_n p_n = (p_{n0} + \delta p)(n_{n0} + \delta n) > n_i^2.$$

### 1.1) Transitorio di spegnimento

Quando la sorgente luminosa viene spenta e quindi  $G_{OP}$ , che è una causa esterna di generazione, viene meno, come evolve nel tempo la concentrazione dei minoritari in eccesso?

$$\frac{dp_n}{dt} = G - R + [G_{OP} = 0]$$

$$\frac{dp_n}{dt} = r n_i^2 - r n p$$

$$\frac{dp_n}{dt} = r n_i^2 - r (p_{n0} + \delta p)(n_{n0} + \delta n)$$

e, per l'hp di bassa iniezione,

$$\frac{dp_n}{dt} = -r n_{n0} \delta p$$

che può anche essere scritta ( $\delta p = p_n - p_{n0}$ )

$$\frac{d\delta p(t)}{dt} = -r n_{n0} \delta p(t)$$

L'equazione differenziale ha per soluzione

$$\delta p(t) = \delta p(0) e^{-r n_{n0} t} = \delta p(0) e^{-\frac{t}{\tau}} :$$

quando cessa la causa l'eccesso dei portatori si ricombina con una costante di tempo  $\tau$ .

Fra gli istanti  $t$  e  $t + \delta t$  (al tempo  $t$  a meno di infinitesimi) si ricombinano  $\delta p(t) - \delta p(t + \delta t)$  portatori; il tempo medio di vita si calcola come  $\left( \frac{\delta p(t) - \delta p(t + \delta t)}{\delta t} = -\delta' p(t) \right)$

$$\langle t \rangle = \frac{1}{\delta p(0)} \int_0^\infty t (\delta p(t) - \delta p(t + \delta t)) = -\frac{1}{\delta p(0)} \int_0^\infty t \delta' p(t) dt$$

$$\langle t \rangle = - \int_0^{\infty} t d \left( e^{-\frac{t}{\tau}} \right) = \tau :$$

$\tau$  rappresenta dunque il *tempo di vita medio* dei portatori minoritari.

1.2) Transitorio di accensione

$$\frac{dp_n}{dt} = G - R + G_{OP}$$

$$\frac{d\delta p}{dt} = G_{OP} - \frac{\delta p}{\tau};$$

a regime

$$0 = G_{OP} - \frac{\delta p}{\tau} \Rightarrow$$

$$\delta p = G_{OP}\tau.$$

Supponiamo che  $\tau = 10^{-5}$  s e  $G_{OP} = 10^{15}$  cm<sup>-3</sup>s<sup>-1</sup>,  $\delta p$  vale allora  $10^{10}$  cm<sup>-3</sup>. Il transitorio di accensione si ottiene da

$$\frac{d\delta p}{dt} = G_{OP} - \frac{\delta p}{\tau}$$

$$\frac{d\delta p}{dt} = \frac{\tau G_{OP} - \delta p}{\tau};$$

ponendo  $\delta^*p = \tau G_{OP} - \delta p$

$$\frac{d\delta^*p}{dt} = -\frac{\delta^*p}{\tau}$$

la cui soluzione è

$$\delta^*p = \delta^*p(0) e^{-\frac{t}{\tau}}$$

$$\tau G_{OP} - \delta p = (\tau G_{OP}) e^{-\frac{t}{\tau}}$$

e in definitiva

$$\delta p(t) = \tau G_{OP} - \tau G_{OP} e^{-\frac{t}{\tau}} = \tau G_{OP} \left( 1 - e^{-\frac{t}{\tau}} \right)$$

Come vedremo in seguito  $\tau$  rappresenta una quantità di fondamentale importanza per il funzionamento dei dispositivi a stato solido, integrati e non.

Nel Si l'espressione di  $\tau$  è la seguente

$$\tau = \frac{1}{v_{th} N_t \sigma_c}$$

dove  $N_t$  rappresenta la concentrazione dei difetti *interstiziali* e  $\sigma_c$  è la sezione di cattura dei difetti in questione.

-----

In un sc non-degenere  $E_F$  si trova nel gap e la FD in BC è approssimabile con una funzione di distribuzione di tipo classico

$$f(E) = \exp\left(-\frac{E - E_F}{kT}\right);$$

gli elettroni (bassa  $n(E)$  in rapporto alla  $D(E)$ ) si comportano come particelle classiche e quindi

$$\frac{1}{2}m_e^*v_{th}^2 = \frac{3}{2}kT.$$

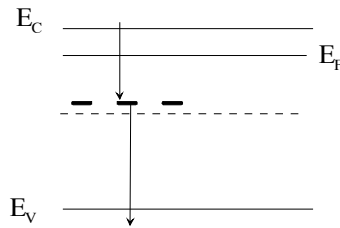
-----

$v_{th} = \sqrt{\frac{3}{2}kT \frac{2}{m_e^*}} = \sqrt{\frac{3 \times 8.63 \times 10^{-5} \times 1.6 \times 10^{-19} \times 300}{0.26 \times 9.1 \times 10^{-31}}} = 2.3 \times 10^5 \text{ ms}^{-1}$  (si noti il valore della massa efficace degli elettroni).

Un difetto tipico, che si usa anche intenzionalmente per stabilire il valore di  $\tau$ , è l'Au; quando  $N_t(\text{Au}) = 10^{14} \text{ cm}^{-3}$   $\tau$  vale  $\simeq 2 \times 10^{-7} \text{ s}$ .

Si può adesso calcolare  $\sigma_c = \frac{1}{v_{th}N_t\tau} = \frac{1}{2.3 \times 10^7 \times 10^{14} \times 2 \times 10^{-7}} = 2.17 \times 10^{-15} \text{ cm}^2$ .

Il ruolo di questo tipo di difetti si spiega sapendo che introducono all'interno del gap e nei pressi del livello di Fermi intrinseco degli stati (accettori e/o donatori) che fungono da intermediari per la ricombinazione delle coppie elettrone-lacuna.



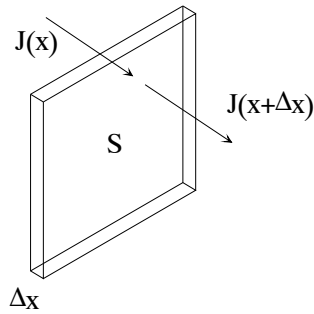
Sono noti come *deep levels* o trappole. Nel Si, che è un semiconduttore a gap indiretto, i processi di ricombinazione diretta banda-banda non sono permessi.

Se è possibile con opportuni processi rendere  $N_t$  molto basso si può avere un Si con un tempo di vita dei minoritari molto lungo (es. di applicazione: CCD (*Charge Coupled Devices*) → fotocamere digitali)

## 9) L'equazione di continuità

La variazione nel tempo del numero dei portatori nel volume indicato è data dal flusso netto ( $F(x) - F(x + \Delta x)$ ) più quello che si genera internamente al volume meno quello che si ricombina; prendendo in considerazione le lacune





$$\frac{\partial}{\partial t} (pS\Delta x) = \frac{[J_h(x) - J_h(x + \Delta x)]}{q} S + (G - R) S\Delta x$$

ovvero

$$\frac{\partial p}{\partial t} = \frac{[J_h(x) - J_h(x + \Delta x)]}{q\Delta x} + (G - R)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_h(x)}{\partial x} + (G - R)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial (q\mu_h p \mathcal{E} - qD_h \frac{\partial p}{\partial x})}{\partial x} + (G - R)$$

Nel termine  $q\mu_h p \mathcal{E} - qD_h \frac{\partial p}{\partial x}$  sia  $\mu_h$  che  $D_h$  dipendono in generale da  $x$ ; supponendo tuttavia per semplicità che siano costanti

$$\frac{\partial p}{\partial t} = -\left( \mu_h p \frac{\partial \mathcal{E}}{\partial x} + \mu_h \mathcal{E} \frac{\partial p}{\partial x} - D_h \frac{\partial^2 p}{\partial x^2} \right) + (G - R)$$

che non può essere risolta analiticamente se non in casi particolari quali l'illuminazione uniforme prima considerata. In quel caso infatti  $p(x) = \text{cost.}$  ed  $\mathcal{E} = 0$  e quindi l'equazione di continuità si riduceva a

$$\frac{dp_n}{dt} = -\frac{p_n - p_{n0}}{\tau} + G_{OP}$$

N.B.: Il termine  $G - R$  si può scrivere come  $-\frac{p_n - p_{n0}}{\tau}$  solo nell'hp di bassa iniezione.

-----

Dipendenza della mobilità dalla posizione.

$\mu_{e,h} = \frac{e\tau}{m^*}$  dove  $\tau$  è il tempo di rilassamento.  $\frac{1}{\tau}$  rappresenta dunque il numero medio di collisioni al secondo. Una causa importante di scattering sono le impurezze (ionizzate e non). All'aumentare del drogaggio cresce il numero di collisioni nell'unità di tempo e quindi la mobilità diminuisce. In un campione in cui il drogaggio è funzione di  $x$  anche la mobilità è funzione di  $x$ .

Nella tabella sono riportati i valori della mobilità dei portatori nel Si per un ampio range di valori di drogaggio.

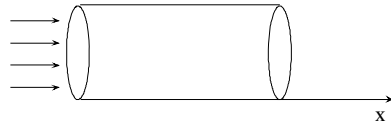
$N_{D,A}$ ( $\text{cm}^{-3}$ )	$\mu_n$ ( $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ )	$\mu_h$ ( $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ )
$10^{14}$	1350.0	495.0
$2 \times 10^{14}$	1345.0	495.0
$4 \times 10^{14}$	1335.0	495.0
$6 \times 10^{14}$	1320.0	495.0
$8 \times 10^{14}$	1310.0	495.0
$10^{15}$	1300.0	491.1
$2 \times 10^{15}$	1248.0	487.3
$4 \times 10^{15}$	1200.0	480.1
$6 \times 10^{15}$	1156.0	473.3
$8 \times 10^{15}$	1115.0	466.9
$10^{16}$	1076.0	460.9
$2 \times 10^{16}$	960.0	434.8
$4 \times 10^{16}$	845.0	396.5
$6 \times 10^{16}$	760.0	369.2
$8 \times 10^{16}$	720.0	348.3
$10^{17}$	675.0	331.5
$2 \times 10^{17}$	524.0	279.0
$4 \times 10^{17}$	385.0	229.8
$6 \times 10^{17}$	321.0	203.8
$8 \times 10^{17}$	279.0	186.9
$10^{18}$	252.0	178.0
$2 \times 10^{18}$	182.5	130.0
$4 \times 10^{18}$	140.6	90.0
$6 \times 10^{18}$	113.6	74.5
$8 \times 10^{18}$	99.5	66.6
$10^{19}$	90.5	61.0
$2 \times 10^{19}$	86.9	55.0
$4 \times 10^{19}$	83.4	53.7
$6 \times 10^{19}$	78.8	52.9
$8 \times 10^{19}$	71.6	52.4
$10^{20}$	67.8	52.0
$2 \times 10^{20}$	52.0	50.8
$4 \times 10^{20}$	35.5	49.6
$6 \times 10^{20}$	23.6	48.9
$8 \times 10^{20}$	19.0	48.4
$10^{21}$	17.8	48.0

Si osservi che la mobilità dipende dalla concentrazione *totale* di droganti e

quindi dalla somma  $N_D + N_A$  e non dal drogaggio netto (differenza  $|N_D - N_A|$ ).

-----  
 2) Illuminazione da un estremo

Il caso illuminazione uniforme ha consentito di analizzare il comportamento temporale dell'eccesso dei portatori. Il caso in questione (non risolvibile analiticamente se non a patto di approssimazioni) darà indicazioni sul comportamento spaziale dell'eccesso (Si di tipo  $n$ ). L'illuminazione, e quindi la generazione, avviene solo alla superficie del wafer. Si ha dunque eccesso solo in  $x = 0$ .



L'equazione di continuità diviene, in condizioni di regime e nell'ipotesi preliminare  $\mathcal{E} = 0$ ,

$$0 = D_h \frac{\partial^2 p}{\partial x^2} - \frac{p_n - p_{n0}}{\tau} = D_h \frac{\partial^2 \delta p}{\partial x^2} - \frac{\delta p}{\tau}$$

che ha per soluzione

$$\delta p(x) = A \exp\left(-\frac{x}{\sqrt{D_h \tau}}\right) + B \exp\left(\frac{x}{\sqrt{D_h \tau}}\right);$$

$B = 0$ , dato che questo termine della soluzione è privo di significato fisico

$$\delta p(x) = \delta p(0) \exp\left(-\frac{x}{\sqrt{D_h \tau}}\right)$$

in cui  $\delta p(0)$  rappresenta l'eccesso in  $x = 0$  (sulla superficie illuminata) e  $\sqrt{D_h \tau} = L_h$  è la lunghezza di diffusione.

L'illuminazione stabilisce un eccesso di portatori (sia maggioritari che minoritari) in  $x = 0$ , che poi tendono a diffondere verso l'interno del campione. Per ogni  $x > 0$  si avrà  $np > n_i^2$  e quindi i portatori si ricombinano nel tentativo di riportare la concentrazione al valore di equilibrio.

La velocità con la quale i portatori diffondono si calcola utilizzando l'espressione della densità di corrente di diffusione delle lacune ed uguagliandola all'espressione di densità di corrente valida in generale

-----  
 Un flusso  $F$  di particelle con densità  $\rho$  e velocità  $v$  è dato da  $\rho v$ , che rappresenta il numero di particelle contenute in un cilindro che ha per base l'unità di superficie e per altezza lo spazio percorso nell'unità di tempo:

$$F = \rho v;$$

moltiplicando per la carica si ha la densità di corrente  $J$

$$J = q\rho v$$

---


$$J_h = -qD_h \frac{dp(x)}{dx} = qp(x)v_{diff}(x);$$

dato che ciò che diffonde è l'eccesso

$$J_h = -qD_h \frac{d\delta p(x)}{dx} = q\delta p(x)v_{diff}(x)$$

$$J_h = -qD_h \frac{d}{dx} \left[ \delta p(0) \exp\left(-\frac{x}{L_h}\right) \right] = q \left[ \delta p(0) \exp\left(-\frac{x}{L_h}\right) \right] v_{diff}(x)$$

da cui

$$\frac{D_h}{L_h} \delta p(0) \exp\left(-\frac{x}{L_h}\right) = \delta p(0) \exp\left(-\frac{x}{L_h}\right) v_{diff}$$

e infine

$$\frac{D_h}{L_h} = v_{diff}$$

$$\frac{D_h \tau_h}{L_h \tau_h} = \frac{L_h^2}{L_h \tau_h} = \frac{L_h}{\tau_h} = v_{diff}$$

relazione che lega fra loro il tempo di vita medio e la lunghezza media di diffusione.

Esempio

In un Si  $n \approx 10^{15} \text{ cm}^{-3}$   $\mu_h = 490 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ . Utilizzando la relazione di Einstein e assumendo  $\tau_h = 10^{-6} \text{ s}$

$$D_h = \frac{kT}{q} \mu_h = 0.026 \times 490 = 12.74 \text{ cm}^2\text{s}^{-1}$$

$$L_h = \sqrt{D_h \tau_h} = \sqrt{12.74 \times 10^{-6}} = 3.57 \times 10^{-3} \text{ cm} = 35.7 \text{ } \mu\text{m}$$

$$v_{diff} = \frac{L_h}{\tau_h} = \frac{3.57 \times 10^{-3}}{10^{-6}} = 3.57 \times 10^3 \text{ cm/s.}$$

Per gli *elettroni* invece  $\mu_n = 1300 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$

$$D_n = \frac{kT}{q} \mu_n = 0.026 \times 1300 = 33.8 \text{ cm}^2\text{s}^{-1}$$

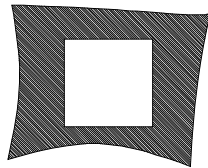
$$L_n = \sqrt{D_n \tau_n} = \sqrt{33.8 \times 10^{-6}} = 5.8 \times 10^{-3} \text{ cm} = 58 \text{ } \mu\text{m}$$

$$v_{diff} = \frac{L_n}{\tau_n} = \frac{5.8 \times 10^{-3}}{10^{-6}} = 5.8 \times 10^3 \text{ cm/s.}$$

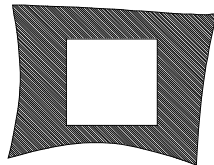
La lunghezza di diffusione degli elettroni in eccesso è *maggiore* di quella delle lacune. Dato che la generazione delle coppie avviene in  $x = 0$  e dato che gli elettroni tendono a diffondere più in profondità, a regime si stabilirà un campo elettrico, dovuto alla differenza dei profili di diffusione, che impedisce agli elettroni di assumere il loro profilo esponenziale. Per ottenere il vero profilo di elettroni e lacune sarebbe necessario risolvere l'equazione di continuità tenendo conto anche del campo che ha origine dalle diverse lunghezze di diffusione e questo non può essere fatto per via analitica. Se tuttavia la differenza fra le i due profili ( $\delta p(x)$  e  $\delta n(x)$ ) non è molto grande, anche la densità di carica risultante e il campo associato saranno quasi trascurabili: si parla in tal caso di *ipotesi di quasi-neutralità*. In tutte le considerazioni che seguiranno utilizzeremo sempre il profilo esponenziale.

## 12) La giunzione pn

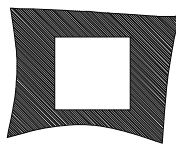
Una giunzione *pn* si ottiene dunque cambiando il drogaggio di una zona di silicio cristallino *p* o *n*. Di seguito sono mostrate le maschere necessarie per realizzare una giunzione *pn* integrata ed il layout relativo.



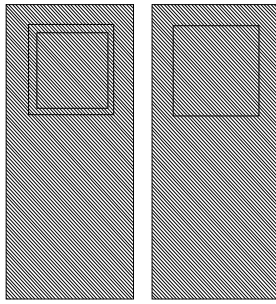
1) Diffusione *p* su wafer *n*; se c'è stato un drive-in in atmosfera ossidante questa finestra si è richiusa automaticamente.



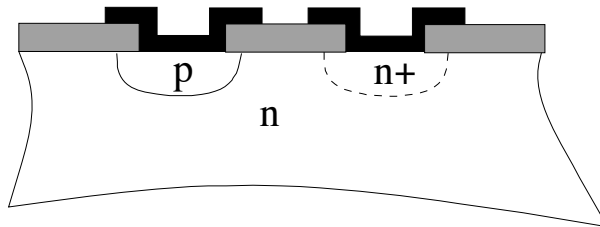
2) Diffusione  $n^+$ ; è necessaria per rendere ohmico il contatto Al/Si (*n*).



3) Apertura di una finestra sulla diffusione *p*.



4) Maschera per la definizione della metal.



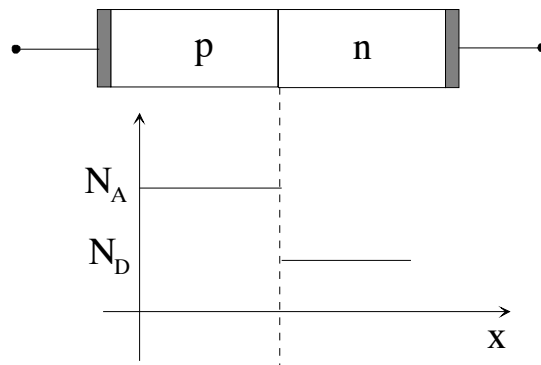
Si noti che tutti i diodi del wafer hanno a comune la parte  $n$ . Si dovrà pertanto provvedere a isolarli elettricamente uno dall'altro.

La sequenza di maschere mostrata deve essere considerata solamente come un esercizio.

Nell'attraversare la giunzione non si ha nessuna interruzione della periodicità del cristallo. È evidente da quanto esposto che il drogaggio, a destra e sinistra del piano di giunzione, non può in generale essere considerato costante. Tuttavia, per motivi di semplicità considereremo inizialmente la giunzione ideale  $pn$  brusca.

### La giunzione $pn$ all'equilibrio

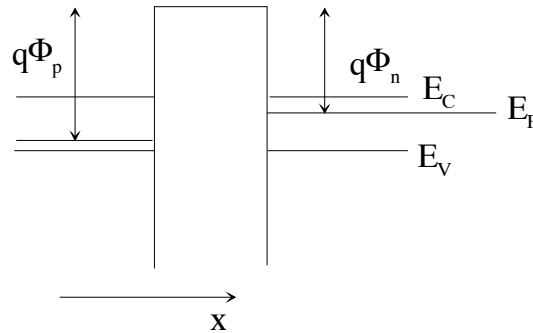
La figura mostra la giunzione  $pn$  brusca (ideale) così definita:



- 1) i drogaggi  $N_A$  ed  $N_D$  sono costanti;
- 2) i contatti metallo-semiconduttore alle estremità non presentano apprezzabile resistenza di contatto (contatti ohmici);

3) il dispositivo è monodimensionale; ciò significa che tutte le quantità che descrivono il comportamento della giunzione (concentrazione di elettroni, di lacune, campo elettrico, densità di corrente ecc.) sono funzione soltanto di  $x$ . L'asse  $x$  è ortogonale al piano della giunzione ( $x = 0$ ) ed è orientato dalla zona  $p$  alla zona  $n$ .

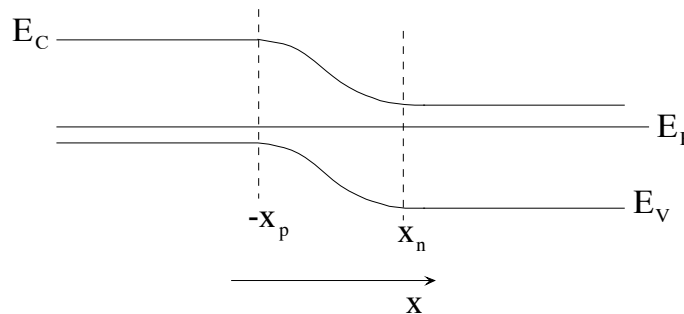
Per comprendere ciò che accade nei pressi della giunzione *immaginiamo* che le due parti  $p$  ed  $n$  siano inizialmente separate. Il diagramma a bande prima del contatto è mostrato nella figura.



È evidente che le funzioni di lavoro del Si  $p$  ( $q\Phi_p$ ) e del Si  $n$  ( $q\Phi_n$ ) sono diverse:

$$q\Phi_p > q\Phi_n.$$

Sappiamo peraltro, che a contatto avvenuto e all'equilibrio, il livello di Fermi deve essere costante con la posizione,  $E_F(x) = C$ . Mettiamo a contatto le due parti e, per disegnare il diagramma a bande della giunzione nel suo complesso, disegniamo prima il livello di Fermi, poi le bande a destra e sinistra della giunzione e finalmente raccordiamo  $E_C$  ed  $E_V$  tenendo presente che il gap del Si deve rimanere costante.



Prima del contatto c'è un gradiente di concentrazione di elettroni e di lacune fra il lato  $n$  e il lato  $p$ . A contatto avvenuto gli elettroni tendono a diffondere verso sinistra e le lacune verso destra.

Come si può vedere  $E_C$ , fra le due linee tratteggiate, varia in funzione di  $x$ . Ciò significa che la concentrazione di elettroni, data all'equilibrio dalla relazione

$$n(x) = N_C e^{-\frac{E_C(x) - E_F}{kT}}$$

è dopo il contatto funzione di  $x$  fra  $x_n$  e  $-x_p$ , e precisamente varia fra  $n_{n0}$  (in  $x_n$ ) a  $n_{p0}$  (in  $-x_p$ ). Per un qualunque valore  $x$  compreso fra 0 e  $x_n$  (zona  $n$ ),  $n(x)$  è minore di  $n_{n0}$ ; questo significa che per  $0 \leq x \leq x_n$  non c'è più la neutralità di carica come prima del contatto.

Un elettrone passa, per diffusione, dalla zona  $n$  alla zona  $p$  e lascia dietro di sé una carica positiva fissa (donatore ionizzato).

Stesse considerazioni si possono ripetere per le lacune nell'intervallo  $0 \geq x \geq -x_p$ .

### La differenza di potenziale di contatto $V_0$

Sono queste cariche fisse (positive nella zona  $n$ , negative nella zona  $p$ ) che fanno assumere alle due parti costituenti la giunzione un potenziale diverso e consentono quindi lo spostamento sulla scala delle energie potenziali necessario all'allineamento dei livelli di Fermi.

Fra i due lati della giunzione, ovvero fra  $x_n$  e  $-x_p$ , si è dunque stabilita all'equilibrio una ddp di contatto  $V_0$ . Per calcolare il suo valore consideriamo in dettaglio ciò che accade fra  $x_n$  e  $-x_p$ .

Abbiamo già visto che  $n$  e  $p$  sono adesso funzioni di  $x$ : gli elettroni tendono a diffondere verso sinistra e le lacune verso destra. Si hanno dunque, fra  $x_n$  e  $-x_p$ , due densità di corrente di diffusione, una di elettroni

$$J_n = qD_n \frac{dn(x)}{dx}$$

e una di lacune

$$J_h = -qD_h \frac{dp(x)}{dx}$$

Poiché siamo all'equilibrio deve essere  $J_{nTOT} = 0$  e  $J_{hTOT} = 0$ , ovvero

$$J_{nTOT} = qD_n \frac{dn(x)}{dx} + q\mu_n n(x)\mathcal{E} = 0$$

$$J_{hTOT} = -qD_h \frac{dp(x)}{dx} + q\mu_h p(x)\mathcal{E} = 0$$

dove  $\mathcal{E}$  è il campo elettrico fra  $x_n$  e  $-x_p$ . La componente diffusiva della densità di corrente viene annullata punto per punto dalla densità di corrente di trascinamento. L'andamento del campo  $\mathcal{E} = \mathcal{E}(x)$  resta da determinare.



Consideriamo ad esempio la

$$qD_n \frac{dn(x)}{dx} + q\mu_n n(x)\mathcal{E} = 0$$

Avremo

$$qD_n \frac{dn(x)}{dx} - q\mu_n n(x) \frac{dV}{dx} = 0$$

$$\frac{D_n}{\mu_n} \frac{dn(x)}{dx} = n(x) \frac{dV}{dx}$$

$$\frac{kT}{q} \frac{dn(x)}{n(x)} = dV$$

La relazione precedente può essere integrata fra  $n_{n0}$  e  $n_{p0}$

$$\frac{kT}{q} \int_{n_{p0}}^{n_{n0}} \frac{dn(x)}{n(x)} = \int_0^{V_0} dV$$

dove è stata scelta come origine del potenziale ( $V = 0$ ) la coordinata  $-x_p$  (dove la concentrazione degli elettroni vale  $n_{n0}$ ). Si ottiene

$$\frac{kT}{q} \ln \left( \frac{n_{n0}}{n_{p0}} \right) = V_0$$

che è proprio la ddp di contatto cercata.

Ricordando che  $n_{n0} = N_D$  e che  $N_A n_{p0} = n_i^2$  (legge dell'azione di massa) l'espressione precedente può essere scritta

$$\frac{kT}{q} \ln \left( \frac{N_D N_A}{n_i^2} \right) = V_0$$

che fornisce la ddp di contatto per una giunzione brusca, nell'ipotesi che per entrambi i lati della giunzione si possa scrivere  $n_{n0} = N_D$  e  $n_{p0} = N_A$  (intervallo di temperature di svuotamento). Si noti che quando i due lati della giunzione tendono a diventare intrinseci ( $N_D = n_i$ ,  $N_A = n_i$ ), per effetto di un aumento di  $T$ , la ddp di contatto tende a zero, com'è ovvio dato che la giunzione si trasforma in un resistore di silicio intrinseco.

Se supponiamo  $N_D = 10^{15} \text{ cm}^{-3}$ ,  $N_A = 10^{17} \text{ cm}^{-3}$ , si ottiene immediatamente

$$V_0 = 0.026 \times \ln \left( \frac{10^{32}}{2.25 \times 10^{20}} \right) = 0.697 \text{ V.}$$

### Andamento del campo e del potenziale fra $-x_p$ e $x_n$

$\mathcal{E}(x)$  e  $V(x)$  si ottengono risolvendo l'equazione di Poisson fra  $-x_p$  e  $x_n$

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_s}$$

una volta nota la densità di carica  $\rho(x)$ .

Contribuiscono alla  $\rho(x)$  fra 0 e  $x_n$ :

- i) i donatori ionizzati con concentrazione  $N_D$  (carica positiva);
- ii) gli elettroni con concentrazione  $n(x) < N_D$  (carica negativa);
- iii) le lacune con concentrazione  $p(x)$  (carica positiva).

Avremo dunque

$$\rho(x) = qN_D + qp(x) - qn(x) = qN_D + qN_V e^{-\frac{E_F - E_V(x)}{kT}} - qN_C e^{-\frac{E_C(x) - E_F}{kT}}$$

dove  $E_C(x) = -qV(x)$ . Nella  $\rho(x)$  non compare esplicitamente la variabile  $x$  per cui non è possibile ottenere direttamente, mediante integrazione, l'andamento di  $\mathcal{E}(x)$  e  $V(x)$ .

È possibile tuttavia rendere integrabile l'equazione di Poisson ricorrendo a delle approssimazioni, la cui validità potrà essere verificata a posteriori.

Approssimazione di *svuotamento completo*.

Si suppone che nell'intervallo  $(-x_p, x_n)$  la concentrazione dei portatori mobili,  $p(x)$  ed  $n(x)$ , sia nulla. In questo caso l'equazione di Poisson diviene, nell'intervallo  $(0, x_n)$

$$\frac{d^2V(x)}{dx^2} = -\frac{qN_D}{\epsilon_s}$$

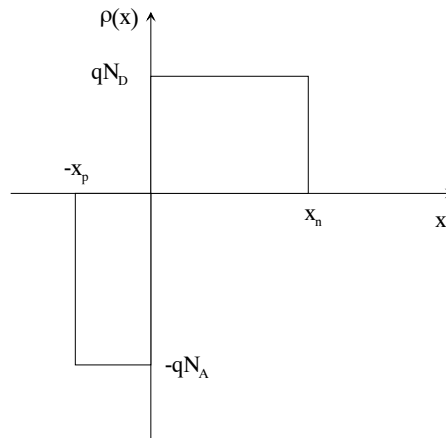
e

$$\frac{d^2V(x)}{dx^2} = +\frac{qN_A}{\epsilon_s}$$

nell'intervallo  $(0, -x_p)$ .

Che si tratti di un'approssimazione è evidente. Infatti, ad esempio,  $n(x)$  non è mai nulla nell'intervallo  $(-x_p, x_n)$ , ma è compresa fra i limiti  $n_{n0} = N_D$  (in  $x_n$ ) e  $n_{p0} = n_i^2/N_A$  (in  $-x_p$ ). Tuttavia, poiché  $n(x)$  è una funzione esponenziale con argomento  $-(E_C(x) - E_F)/kT$  l'approssimazione  $\rho(x) = qN_D$ , nell'intervallo  $(0, x_n)$ , sarà tanto migliore quanto più ci si avvicina ad  $x = 0$ .

La  $\rho(x)$  con l'approssimazione di svuotamento completo è rappresentata nella figura.



Per ottenere l'andamento del campo elettrico si integra una prima volta l'eq. di Poisson nei due intervalli:

$$\frac{dV(x)}{dx} = -\frac{qN_D}{\epsilon_s}x + C_1$$

$$\frac{dV(x)}{dx} = \frac{qN_A}{\epsilon_s}x + C_2$$

Le costanti  $C_1$  e  $C_2$  si determinano imponendo, come condizioni al contorno, che il campo elettrico ( $\mathcal{E}(x) = -dV(x)/dx$ ), sia nullo in  $x_n$  e  $-x_p$ .

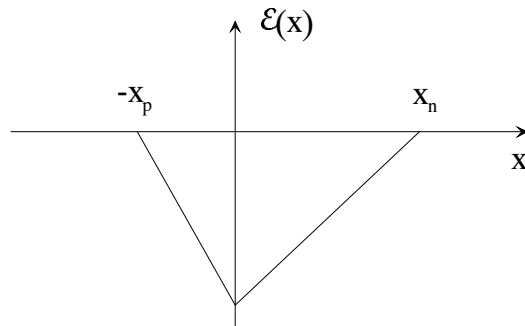
Si ha pertanto, nella zona  $n$

$$-\frac{dV(x)}{dx} = \mathcal{E}(x) = \frac{qN_D}{\epsilon_s}x - \frac{qN_D}{\epsilon_s}x_n$$

e, nella zona  $p$

$$-\frac{dV(x)}{dx} = \mathcal{E}(x) = -\frac{qN_A}{\epsilon_s}x - \frac{qN_A}{\epsilon_s}x_p$$

L'andamento del campo è mostrato nella figura.



In  $x = 0$  il campo è continuo (non c'è infatti discontinuità nella costante dielettrica nel passaggio dalla zona  $n$  alla zona  $p$ ) ed assume il suo massimo valore.

Dalla continuità del campo in  $x = 0$  si ottiene immediatamente

$$|\mathcal{E}_{MAX}| = \frac{qN_D}{\epsilon_s} x_n = \frac{qN_A}{\epsilon_s} x_p$$

ovvero

$$N_D x_n = N_A x_p :$$

la zona di svuotamento si estende maggiormente nella parte meno drogata. Calcoliamo adesso quanto vale tutta la zona di svuotamento  $W = x_n + x_p$ . Dalla relazione

$$-\frac{dV(x)}{dx} = \mathcal{E}(x)$$

otteniamo

$$\int_0^{V_0} dV(x) = - \int_{-x_p}^{x_n} \mathcal{E}(x) dx = \text{area del triangolo}$$

$$V_0 = \frac{W |\mathcal{E}_{MAX}|}{2} = W \frac{|\mathcal{E}_{MAX}|}{2} = W \frac{qN_D}{2\epsilon_s} x_n$$

che, una volta espressa  $W$  come

$$W = x_n + x_p = x_n + \frac{N_D}{N_A} x_n = x_n \left( 1 + \frac{N_D}{N_A} \right)$$

diventa

$$V_0 = W^2 \frac{qN_D}{2\epsilon_s} \frac{N_A}{N_D + N_A}$$

da cui infine

$$W = \sqrt{\frac{2\epsilon_s V_0}{q} \left( \frac{N_D + N_A}{N_A N_D} \right)}$$

Esempio

$$N_D = 10^{15} \text{ cm}^{-3}, N_A = 10^{17} \text{ cm}^{-3}.$$

Calcolo di  $W$

$$W = \sqrt{\frac{2 \times 11.8 \times 8.85 \times 10^{-12} \times 0.697}{1.6 \times 10^{-19}} \left( \frac{10^{21} + 10^{23}}{10^{44}} \right)} = 0.958 \times 10^{-6} \text{ m} = 0.958 \text{ } \mu\text{m}$$

mentre  $x_n$

$$x_n = W \left( \frac{N_A}{N_A + N_D} \right) = 0.958 \times \left( \frac{10^{23}}{10^{21} + 10^{23}} \right) = 0.948 \text{ } \mu\text{m}$$

Come si vede la zona di svuotamento si estende prevalentemente nel lato meno drogato.

Calcolo di  $|\mathcal{E}_{MAX}|$

$$|\mathcal{E}_{MAX}| = \frac{qN_D}{\epsilon_s} x_n = \frac{1.6 \times 10^{-19} \times 10^{21}}{11.8 \times 8.85 \times 10^{-12}} \times 0.948 \times 10^{-6} = 1.45 \times 10^6 \text{ V/m}$$

Se poi  $N_D \gg N_A$ , avremo una giunzione brusca *unilaterale*  $n^+p$  per la quale  $W$  si trova quasi tutta nella zona  $p$

$$W \simeq \sqrt{\frac{2\epsilon_s V_0}{q} \frac{1}{N_A}}$$

mentre se  $N_A \gg N_D$

$$W \simeq \sqrt{\frac{2\epsilon_s V_0}{q} \frac{1}{N_D}}$$

(giunzione brusca *unilaterale*  $p^+n$ ).

Per semplicità faremo spesso riferimento a questo tipo di giunzioni.

Nel caso di giunzione brusca unilaterale, ad esempio  $p^+n$ , avremo per  $|\mathcal{E}_{MAX}|$

$$|\mathcal{E}_{MAX}| = \frac{qN_D}{\epsilon_s} x_n \simeq \frac{qN_D}{\epsilon_s} W = \frac{qN_D}{\epsilon_s} \sqrt{\frac{2\epsilon_s V_0}{q} \frac{1}{N_D}} = \sqrt{\frac{2V_0 q N_D}{\epsilon_s}}$$

il valore massimo del campo aumenta con il drogaggio.

Per ottenere l'andamento di  $V(x)$  basta integrare la

$$\frac{dV(x)}{dx} = -\frac{qN_D}{\epsilon_s} x + \frac{qN_D}{\epsilon_s} x_n$$

nella zona  $n$  ottenendo

$$V(x) = -\frac{qN_D}{2\epsilon_s} x^2 + \frac{qN_D}{\epsilon_s} x_n x + D_1$$

Come condizione al contorno si impone che in  $x = x_n$  il potenziale sia uguale a  $V_0$  (ricordiamo che abbiamo assunto come origine del potenziale la coordinata  $-x_p$ ):

$$V_0 = -\frac{qN_D}{2\epsilon_s}x_n^2 + \frac{qN_D}{\epsilon_s}x_n^2 + D_1$$

da cui

$$V_0 - \frac{qN_D}{2\epsilon_s}x_n^2 = D_1$$

ed infine

$$V(x) = V_0 - \frac{qN_D}{2\epsilon_s}(x - x_n)^2$$

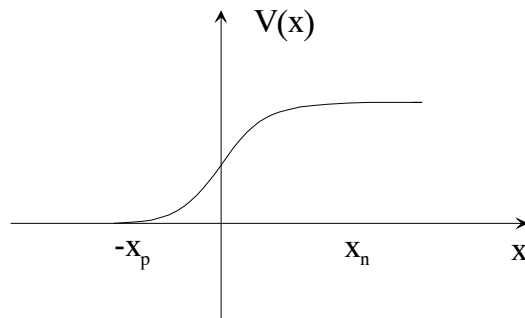
Per ottenere l'andamento di  $V(x)$  nella zona  $p$  si procede analogamente integrando la

$$\frac{dV(x)}{dx} = \frac{qN_A}{\epsilon_s}x + \frac{qN_A}{\epsilon_s}x_p$$

e imponendo come condizione al contorno che per  $x = -x_p$   $V(x) = 0$ . Si ottiene

$$V(x) = \frac{qN_A}{2\epsilon_s}(x + x_p)^2$$

L'andamento di  $V(x)$  è mostrato nella figura. Si osservi che il valore  $V_0/2$  si ha per  $x = 0$  (giunzione metallurgica) solo nel caso  $N_A = N_D$ .



### Rimozione dell'approssimazione di svuotamento completo

La ddp di contatto  $V_0$  data dall'espressione

$$V_0 = \frac{kT}{q} \ln \left( \frac{N_A N_D}{n_i^2} \right)$$

si può ottenere, come è noto, in due modi diversi: i) calcolando la differenza fra i livelli di Fermi prima del contatto, ii) imponendo l'annullarsi della densità di corrente totale di lacune (o di elettroni) nella zona di carica spaziale. Si tratta in ogni caso di una quantità caratteristica della giunzione.

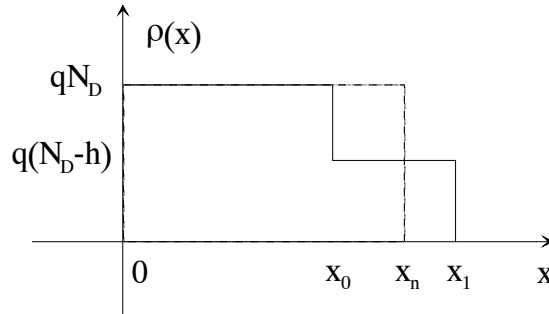
La successiva integrazione dell'equazione di Poisson,

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_s}$$

che fornisce l'andamento di  $\mathcal{E}(x)$  e di  $V(x)$ , ci ha consentito di determinare una relazione semplice fra  $V_0$  e  $W$ , ampiezza della zona di svuotamento.

È importante osservare che il risultato dell'integrazione, e quindi l'andamento di  $\mathcal{E}(x)$  e di  $V(x)$ , *dipende dall'approssimazione usata per la  $\rho(x)$* .

Consideriamo ora una giunzione  $p^+n$  e rimuoviamo l'approssimazione di svuotamento completo, ipotizzando il seguente andamento di  $\rho(x)$ : fra 0 e  $x_0$  svuotamento completo ( $\rho(x) = qN_D$ ) e fra  $x_0$  e  $x_1$  carica minore di  $qN_D$  ma costante ( $\rho(x) = qN_D - qh$ ). Questo andamento rende ragione, molto schematicamente, del fatto che al confine della zona di carica spaziale la densità di carica è funzione decrescente di  $x$ . La linea tratteggiata mostra invece l'approssimazione di svuotamento completo.



Si tratta di risolvere l'equazione di Poisson nei due intervalli dell'asse  $x$ , imponendo la continuità del campo in  $x_0$ .

$$\frac{d^2V(x)}{dx^2} = -\frac{qN_D}{\epsilon_s}, \text{ nell'intervallo } 0 \leq x < x_0$$

$$\frac{d^2V(x)}{dx^2} = -\frac{q(N_D - h)}{\epsilon_s}, \text{ nell'intervallo } x_0 < x \leq x_1$$

Si ottiene

$$\frac{dV(x)}{dx} = -\frac{qN_D}{\epsilon_s}x + C_1, \quad (0 \leq x < x_0) \quad (1)$$

$$\frac{dV(x)}{dx} = -\frac{q(N_D - h)}{\epsilon_s}x + C_2, \quad (x_0 < x \leq x_1); \quad (2)$$

la costante  $C_2$  viene determinata imponendo che in  $x = x_1$  sia  $dV/dx = 0$

$$\frac{dV(x)}{dx} = -\frac{q(N_D - h)}{\epsilon_s}x + \frac{q(N_D - h)}{\epsilon_s}x_1 \quad (3)$$

mentre per la  $C_1$  si deve imporre la continuità del campo in  $x = x_0$

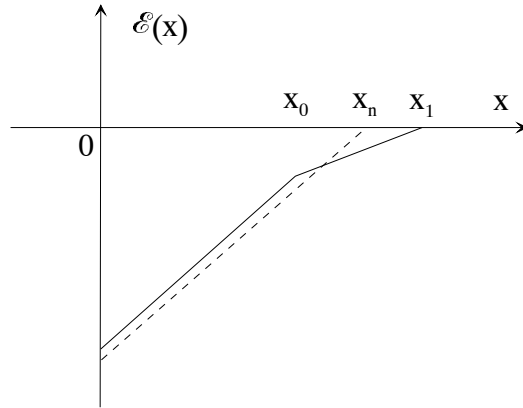
$$-\frac{qN_D}{\epsilon_s}x_0 + C_1 = -\frac{q(N_D - h)}{\epsilon_s}x_0 + \frac{q(N_D - h)}{\epsilon_s}x_1$$

$$C_1 = \frac{qN_D}{\epsilon_s}x_0 + \frac{q(N_D - h)}{\epsilon_s}(x_1 - x_0)$$

La (1) diviene

$$\frac{dV(x)}{dx} = -\frac{qN_D}{\epsilon_s}x + \frac{qN_D}{\epsilon_s}x_0 + \frac{q(N_D - h)}{\epsilon_s}(x_1 - x_0) \quad (4)$$

La (3) e la (4), cambiate di segno, descrivono completamente il comportamento del campo il cui andamento è mostrato nella figura.



Per confronto è riportato anche l'andamento del campo che si ottiene con l'approssimazione di svuotamento completo (linea tratteggiata).

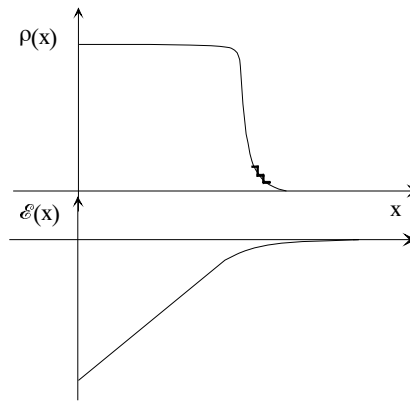
Poiché l'integrale del campo è la ddp  $V_0$ , in accordo con la relazione

$$-\int_0^{V_0} dV = \int_0^{x_i} \mathcal{E}(x)dx$$

( $i = n$  oppure 1 a secondo dell'approssimazione usata per  $\rho(x)$ ) e  $V_0$  viene determinata indipendentemente, si può notare che il valore del campo all'interfaccia e l'ampiezza della zona di svuotamento dipendono dall'approssimazione usata (l'area sottesa dalle due curve deve essere sempre uguale a  $V_0$ ).

Un andamento più realistico della  $\rho(x)$ , mostrato nella figura insieme al campo che ne risulta, può essere schematizzato come una successione di tratti di lunghezza  $dx$  all'interno dei quali la funzione  $\rho(x)$  è costante.





### La giunzione $pn$ a profilo di drogaggio lineare

Abbiamo già visto che una giunzione  $pn$  si ottiene drogando  $p$  ( $n$ ) un substrato a drogaggio costante  $n$  ( $p$ ). Profili di drogaggio reali hanno un andamento del tipo

$$N_A(x) = N_{AS} \exp\left(-\frac{x^2}{a^2}\right)$$

per cui il drogaggio netto a partire dalla superficie si ottiene dalla differenza

$$N(x) = N_{AS} \exp\left(-\frac{x^2}{a^2}\right) - N_D$$

Sviluppando in serie la funzione  $N(x)$  in un intorno di  $x_i$  si ottiene

$$N(x) = -\frac{N_{AS} e^{-\frac{x_i^2}{a^2}} 2x_i}{a^2} (x - x_i) = -k_0(x - x_i)$$

Questa approssimazione prende il nome di giunzione a gradiente (*linear graded junction*). Spostando l'origine dell'asse  $x$  in  $x_i$ ,  $N(x)$  può essere scritta

$$N(x) = -k_0x$$

Per ottenere campo, potenziale e ampiezza della zds si deve, come al solito integrare l'equazione di Poisson fra  $-L$  ed  $L$

$$\frac{d^2V}{dx^2} = -\frac{qk_0x}{\varepsilon_s}$$

nell'approssimazione di svuotamento completo imponendo che il campo si annulli in  $x = L$

$$\frac{dV}{dx} = -\frac{qk_0x^2}{2\varepsilon_s} + \frac{qk_0L^2}{2\varepsilon_s}$$

$$\mathcal{E} = +\frac{qk_0x^2}{2\varepsilon_s} - \frac{qk_0L^2}{2\varepsilon_s}.$$

Per il potenziale

$$V(x) = -\frac{qk_0x^3}{6\varepsilon_s} + \frac{qk_0L^2}{2\varepsilon_s}x + C_1$$

con la condizione  $V(x) = 0$  in  $x = -L$ :

$$0 = +\frac{qk_0L^3}{6\varepsilon_s} - \frac{qk_0L^3}{2\varepsilon_s} + C_1$$

$$V(x) = -\frac{qk_0x^3}{6\varepsilon_s} + \frac{qk_0L^2}{2\varepsilon_s}x + \frac{q2k_0L^3}{6\varepsilon_s}.$$

$L$ , semiampiezza della zds, si trova sapendo che in  $x = L$   $V(x)$  è uguale a  $V_0$

$$V_0 = -\frac{qk_0L^3}{6\varepsilon_s} + \frac{3qk_0L^3}{6\varepsilon_s} + \frac{2qk_0L^3}{6\varepsilon_s} = \frac{2qk_0L^3}{3\varepsilon_s}$$

$$L = \left(\frac{3\varepsilon_s V_0}{2k_0q}\right)^{\frac{1}{3}}$$

### La capacità differenziale della zona di svuotamento

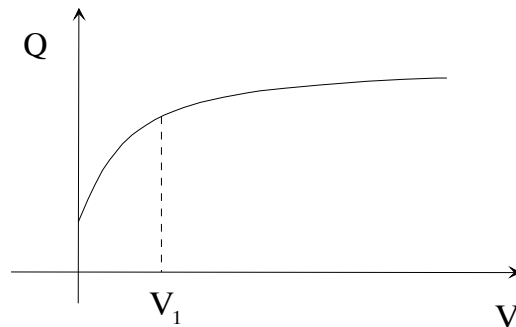
In un condensatore a facce piane e parallele la carica  $Q$  è funzione lineare della ddp applicata

$$Q = CV.$$

In generale si scriverà

$$Q = f(V)$$

ovvero la carica immagazzinata è una funzione della ddp, non necessariamente lineare.



Sviluppando in serie in un intorno di  $V_1$  (punto di riposo) si ottiene

$$Q(V) = f(V_1) + \frac{df(V)}{dV} \Big|_{V=V_1} (V - V_1) = Q(V_1) + C_1 dV$$

e quindi, per piccole variazioni  $dV$  della tensione applicata intorno a  $V_1$

$$Q(V) - Q(V_1) = dQ = C_1 dV$$

relazione che definisce la capacità differenziale come la derivata della funzione  $f(V)$  nel punto di riposo. Assumiamo d'ora in avanti, salvo menzione esplicita, come *giunzione di riferimento* una giunzione brusca unilaterale  $p^+n$  definita da:

- 1) sezione  $S$ :  $100 \times 100 \mu\text{m}^2 = 10^{-4} \text{ cm}^2$ ;
- 2)  $N_D = 10^{15} \text{ cm}^{-3}$ ;  $N_A = 10^{19} \text{ cm}^{-3}$ ;
- 3)  $\mu_{h_n} = 490 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;  $\mu_{h_p} = 60 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;  
 $\mu_{n_p} = 90 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;  $\mu_{n_n} = 1300 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;
- 4)  $\tau = 10^{-5} \text{ s}$ ;
- 5)  $V_0 = 0.026 \times \ln \left( \frac{10^{15} \times 10^{19}}{2.25 \times 10^{20}} \right) = 0.817 \text{ V}$ .

Supponiamo che la giunzione sia polarizzata inversamente: la ddp applicata fra zona  $p$  e zona  $n$  è negativa. Come è noto (e come vedremo) la corrente è trascurabile; si può ammettere allora che tutta la tensione cada ai capi della zona di svuotamento, la cui espressione è all'equilibrio

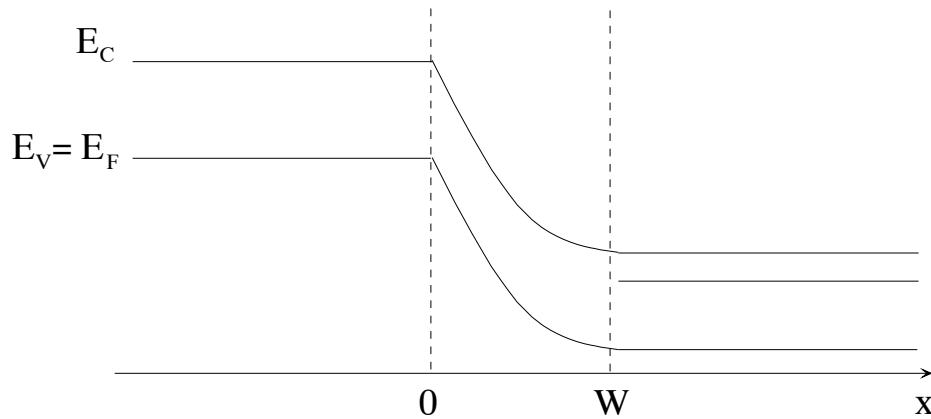
$$W = \sqrt{\frac{2\varepsilon_s V_0}{qN_D}}$$

Ne consegue che:

- i) il campo dovuto alla ddp applicata si somma a quello interno;
- ii) l'ampiezza della zds aumenta

$$W = \sqrt{\frac{2\varepsilon_s (V_0 - V)}{qN_D}}$$

iii) aumenta l'altezza di barriera vista dai maggioritari e i livelli di Fermi si separano di una quantità pari a  $qV$  (vedi figura).



In un semiconduttore degenere,  $n^+$  o  $p^+$ , si assume che il livello di Fermi coincida con  $E_C$  o  $E_V$ , rispettivamente.

All'equilibrio la carica totale compresa fra 0 e  $W$  è data da ( $S$  è la sezione del diodo)

$$Q = SqN_D W = SqN_D \sqrt{\frac{2\varepsilon_s V_0}{qN_D}} = S\sqrt{2\varepsilon_s V_0 qN_D}$$

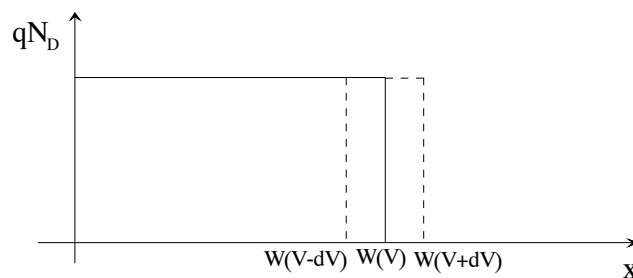
mentre in polarizzazione inversa

$$Q = S\sqrt{2\varepsilon_s (V_0 - V) qN_D} = f(V)$$

per cui

$$C_W(V) = \left| \frac{dQ}{dV} \right| = S \frac{2\varepsilon_s qN_D}{2\sqrt{2\varepsilon_s (V_0 - V) qN_D}} = S\sqrt{\frac{\varepsilon_s qN_D}{2(V_0 - V)}}.$$

L'ampiezza della zds viene modulata dalla variazione (piccola) di tensione come è mostrato nel disegno.



Se si scrive

$$C_W(V) = \varepsilon_s \frac{S}{W(V)}$$

in modo da renderla formalmente analoga alla capacità di un condensatore piano, si ottiene la stessa espressione:

$$C_W = S \sqrt{\frac{\varepsilon_s q N_D}{2(V_0 - V)}}.$$

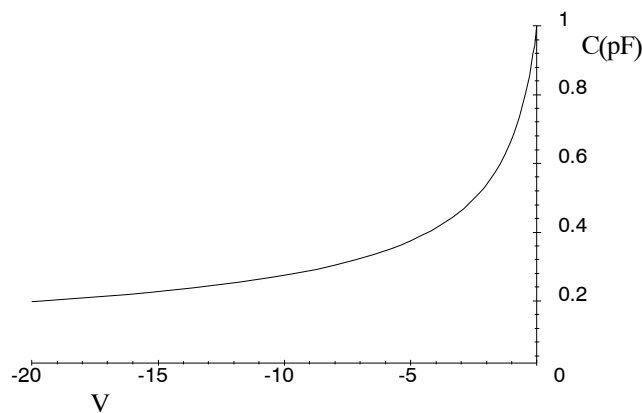
Si noti che l'espressione della  $C_W$  è determinata dalla particolare dipendenza di  $W$  da  $V$ ; in questo caso quella relativa ad una giunzione brusca. La capacità a 0 volt vale

$$C_W(0) = 10^{-8} \times \sqrt{\frac{11.8 \times 8.85 \times 10^{-12} \times 1.6 \times 10^{-19} \times 10^{21}}{2 \times 0.817}} = 1.01 \times 10^{-12} \text{ F};$$

la figura mostra il grafico di  $C_W = \frac{C_W(V)}{C_W(0)}$  per tensioni comprese fra 0 e  $-20$  V.

Alcuni valori sono riportati nella tabella

V	$C_W$
-2 V	0.544 pF
-5 V	0.379 pF
-10 V	0.278 pF
-20 V	0.200 pF



L'espressione  $C_W(V) = \varepsilon_s \frac{S}{W(V)}$ , che si potrebbe dimostrare essere vera per ogni profilo di drogaggio, è molto utile nelle considerazioni sulle capacità parassite dei dispositivi integrati: *in ogni dispositivo ci sono sempre delle giunzioni pn.*

Si osservi ancora che in base alla  $C_W = S \sqrt{\frac{\varepsilon_s q N_D}{2(V_0 - V)}}$  più elevato è il drogaggio del lato  $n$  maggiore la capacità differenziale della zona di svuotamento.

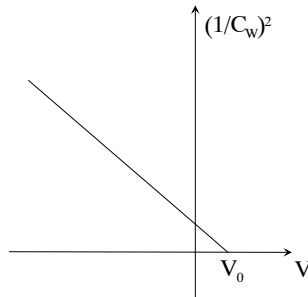
### La misura del drogaggio

1) Eleviamo al quadrato e 2) invertiamo l'espressione di  $C_W$ :

$$1) C_W^2 = S^2 \frac{\varepsilon_s q N_D}{2(V_0 - V)}$$

$$2) \frac{1}{C_W^2} = \frac{1}{S^2} \frac{2(V_0 - V)}{\varepsilon_s q N_D} = \frac{1}{S^2} \left( \frac{2V_0}{\varepsilon_s q N_D} - \frac{2}{\varepsilon_s q N_D} V \right)$$

si ottiene una dipendenza lineare da  $V$  la cui derivata è

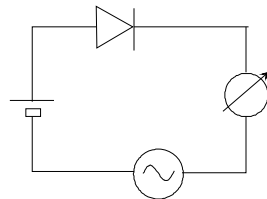


$$\frac{2}{S^2 \varepsilon_s q N_D}$$

Se l'area della giunzione è nota con precisione, da misure  $C_W = C_W(V)$  e dalla successiva elaborazione si ottiene il drogaggio (costante) del lato  $n$ .

$V$	$C_W$	$C_W^2$	$\frac{1}{C_W^2}$
-2 V	$0.544 \times 10^{-12}$	$2.9594 \times 10^{-25}$	$3.3791 \times 10^{24}$
-5 V	$0.379 \times 10^{-12}$	$1.4364 \times 10^{-25}$	$6.9618 \times 10^{24}$
-10 V	$0.278 \times 10^{-12}$	$7.7284 \times 10^{-26}$	$1.2939 \times 10^{25}$
-20 V	$0.200 \times 10^{-12}$	$4.0 \times 10^{-26}$	$2.5 \times 10^{25}$

Il circuito schematico per la misura è quello seguente:



un generatore di tensione continua stabilisce il punto di riposo (ad es. -3 V); il generatore di segnale ( $v(t) = v_m \text{sen}(\omega t)$ ;  $f = 1 \text{ Mhz}$ ,  $v_m = 10 \text{ mV}$ ) provvede alla variazione; l'amperometro misura la corrente nel circuito.

Si calcoli la corrente che passa nel circuito quando  $V = 0 \text{ V}$ .

In modulo

$$I = v_m \omega C_W(0) = 10^{-2} \times 2 \times \pi \times 10^6 \times 10^{-12} = 6.28 \times 10^{-8} \text{ A.}$$

Dal punto di vista delle *variazioni* il diodo si comporta come una capacità  $C_W(V)$ .

E inoltre: estrapolando la retta fino ad intercettare l'asse delle tensioni, ovvero quando

$$\frac{1}{C_W^2} = 0 = \frac{1}{S^2} \frac{2(V_0 - V)}{\epsilon_s q N_D}$$

si ottiene il valore della ddp di contatto nel punto di intersezione  $V = V_0$ .

### 13) La giunzione *pn* polarizzata

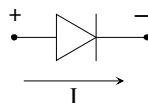
#### Polarizzazione diretta

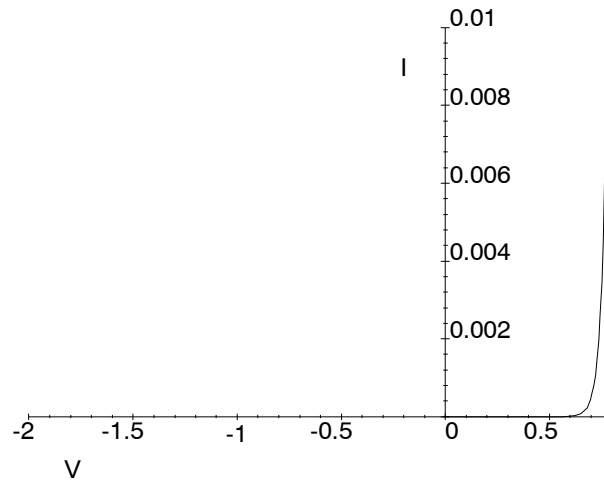
La caratteristica  $I = f(V)$  della giunzione *pn* è data dall'espressione

$$I = I_0 \left( e^{\frac{qV}{kT}} - 1 \right) = I_0 \left( e^{\frac{V}{V_T}} - 1 \right)$$

dove  $V$  è la ddp fra zona *p* e zona *n*; abbiamo posto  $V_T = kT/q = 26 \text{ mV}$ .

Il simbolo del diodo, i riferimenti per tensione e corrente e la caratteristica sono mostrati nelle figure (la scala della corrente è in A)





Il diodo è un dispositivo unidirezionale che fa passare corrente solo quando  $V$  è positiva. Per  $V$  negative la corrente è trascurabile.

► Polarizzazione diretta:  $V \geq 100$  mV

$$I = I_0 \left( e^{\frac{150}{26}} - 1 \right) = I_0 (320.3 - 1) \simeq I_0 e^{\frac{150}{26}} = I_0 e^{\frac{V}{V_T}}$$

► Polarizzazione inversa:  $V \leq -100$  mV

$$I = I_0 \left( e^{-\frac{150}{26}} - 1 \right) = I_0 (3.1 \times 10^{-3} - 1) \simeq -I_0$$

$I_0$  prende il nome di corrente inversa di saturazione.

I limiti di validità dell'espressione  $I_0 \left( e^{\frac{V}{V_T}} - 1 \right)$  verranno messi di volta in volta in evidenza nel corso della trattazione che useremo per ricavarla.

Si consideri la struttura (schematica) della giunzione  $pn$ .



C'è una zona di svuotamento e due zone neutre, una  $p$  e una  $n$ . Inoltre, agli estremi, ci sono dei contatti metallo/semiconduttore che, per ipotesi, hanno resistenza trascurabile. Se si ammette ulteriormente che anche le zone  $n$  e  $p$  neutre presentino una resistenza trascurabile rispetto alla zona svuotata, la ddp applicata fra i terminali del diodo cadrà interamente ai capi della zona di svuotamento.

► I<sup>a</sup> ipotesi: la ddp applicata ai terminali esterni del diodo cade interamente ai capi della zona di svuotamento.

Ricordiamo che all'equilibrio

$$W = \sqrt{\frac{2\epsilon_s V_0}{q} \left( \frac{N_D + N_A}{N_A N_D} \right)}$$



Come conseguenza della I<sup>a</sup> ipotesi avremo

$$W = \sqrt{\frac{2\epsilon_s (V_0 - V)}{q} \left( \frac{N_D + N_A}{N_A N_D} \right)}:$$

l'ampiezza della zona di svuotamento diminuisce in condizioni di polarizzazione diretta ( $V$  positive) e aumenta, come abbiamo già visto, in condizioni di polarizzazione inversa ( $V$  negative).

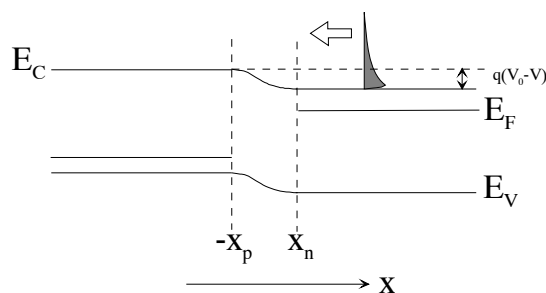
E inoltre: l'altezza di barriera  $qV_0$  diminuisce in diretta e aumenta in inversa. In altre parole il campo elettrico applicato dall'esterno, orientato da sinistra verso destra, fa diminuire il campo elettrico interno. Poiché all'equilibrio ( $I = 0$ ) si aveva

$$J_{nTOT} = qD_n \frac{dn(x)}{dx} + q\mu_n n(x)\mathcal{E} = 0$$

$$J_{hTOT} = -qD_p \frac{dp(x)}{dx} + q\mu_p p(x)\mathcal{E} = 0$$

è evidente che la diminuzione del campo fa prevalere la componente diffusiva rispetto alla componente di trascinamento, con il risultato che adesso  $J_{nTOT} \neq 0$  e  $J_{hTOT} \neq 0$ .

Con riferimento alla figura,



quando la giunzione è polarizzata direttamente sia gli elettroni che le lacune vedono una barriera di altezza ridotta  $q(V - V_0)$ . Prendiamo in considerazione il comportamento delle lacune; per gli elettroni si può ripetere lo stesso ragionamento.

Il prevalere della componente diffusiva della densità di corrente

$$J_{hdiff} = -qD_p \frac{dp(x)}{dx}$$

fa sí che un maggior numero di lacune possano diffondere da sinistra verso destra. Una volta entrate nella zona  $n$ , queste lacune vanno ad incrementare la

concentrazione di lacune minoritarie. Siamo in presenza di *iniezione di portatori minoritari*.

Nella zona  $n$  dunque, e precisamente in  $x_n$

$$p_n(x_n) = p_{n0} + \delta p(x_n)$$

A questo punto è necessaria la

►II<sup>a</sup> ipotesi: la riduzione dell'altezza di barriera provocata dalla polarizzazione è tale che sia sempre

$$\delta p \ll n_{n0}$$

(*ipotesi di bassa iniezione*), dove  $\delta p$  è l'eccesso di minoritari iniettati.

Questa ipotesi pone evidentemente dei limiti alla tensione applicabile alla giunzione. Ciò non vuol dire che alla giunzione non è possibile applicare una tensione tale che  $\delta p$  sia confrontabile con  $n_{n0}$ ; significa semplicemente che la trattazione che stiamo sviluppando non sarebbe, in quest'ultimo caso, più valida.

All'equilibrio

$$p_{p0}(-x_{p0}) = p_{n0}(x_{n0})e^{\frac{V_0}{V_T}}$$

$$n_{p0}(-x_{p0}) = n_{n0}(x_{n0})e^{-\frac{V_0}{V_T}}$$

In condizioni di polarizzazione

$$p_p(-x_p) = p_n(x_n)e^{\frac{(V_0-V)}{V_T}}$$

$$n_p(-x_p) = n_n(x_n)e^{-\frac{(V_0-V)}{V_T}}$$

►Osservazione. Le ultime due sono state dedotte dalle prime due, che sono valide all'equilibrio. L'estensione, ad una situazione di non equilibrio, di relazioni valide *solo* all'equilibrio, non è in generale lecita. In questo caso tuttavia il loro utilizzo è giustificato dal fatto che la deviazione dalla situazione di equilibrio è trascurabile.

Per il prodotto delle concentrazioni in  $x_n$  avremo

$$p_n(x_n)n_n(x_n) = p_p(-x_p)n_{n0}e^{-\frac{(V_0-V)}{V_T}}$$

Poiché la zona  $n$  in  $x_n$  era all'equilibrio neutra, la carica positiva in eccesso iniettata richiama un ugual numero di elettroni in  $x_n$  per ristabilire la neutralità.

$$n_n(x_n) = N_D + \delta n$$

con  $\delta p = \delta n$ . Ma in virtù dell'ipotesi di bassa iniezione

$$n_n(x_n) \simeq N_D = n_{n0};$$

le stesse considerazioni possono essere fatte per  $p_p(-x_p)$  (la concentrazione dei maggioritari resta di fatto inalterata in entrambi i lati della giunzione). In definitiva la si può scrivere

$$p_n(x_n)n_n(x_n) = p_p(-x_p)n_{n0}e^{-\frac{(V_0-V)}{V_T}} = p_{p0}n_{n0}e^{-\frac{(V_0-V)}{V_T}} = p_{p0}n_{n0}e^{-\frac{V_0}{V_T}}e^{\frac{V}{V_T}} = p_{p0}n_{p0}e^{\frac{V}{V_T}}$$

$$p_n(x_n)n_n(x_n) = n_i^2 e^{\frac{V}{V_T}}$$

L'ultima relazione costituisce *la legge di Shockley della giunzione*.

Per quanto riguarda la concentrazione delle lacune in  $x_n$

$$p_n(x_n) = p_{n0}e^{\frac{V}{V_T}} = p_{p0} + \delta p(x_n)$$

da cui

$$\delta p(x_n) = p_{n0} \left( e^{\frac{V}{V_T}} - 1 \right)$$

che fornisce, in funzione di  $V$ , l'eccesso dei minoritari iniettati. Si confronti questa relazione con l'espressione della caratteristica  $I - V$  del diodo.

Per gli elettroni in  $-x_p$  vale un'analogha relazione

$$\delta n(-x_p) = n_{p0} \left( e^{\frac{V}{V_T}} - 1 \right)$$

Abbiamo visto che l'iniezione di lacune richiama in  $x_n$  un ugual eccesso di elettroni per mantenere la neutralità di carica.

Le lacune in eccesso diffondono verso destra e si ricombinano. Stessa cosa faranno gli elettroni in eccesso.

Per semplificare supponiamo che la giunzione sia  $p^+n$ . In base alla relazione di Shockley

$$n_p(-x_p) = n_n(x_n)e^{-\frac{V_0-V}{V_T}} \simeq n_{n0}e^{-\frac{V_0-V}{V_T}} = n_{p0}e^{\frac{V}{V_T}}$$

il contributo degli elettroni iniettati nel lato  $p$  è del tutto trascurabile  $\left( \frac{n_i^2}{N_A} \ll \frac{n_i^2}{N_D} \right)$ .

L'equazione di continuità nella zona  $n$  ha per soluzione (poniamo l'origine dell'asse  $x$  in  $x_n$ )

$$\delta p(x) = p_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) e^{-\frac{x}{L_h}}$$

cui corrisponde una corrente di diffusione

$$I_h = -qSD_h \frac{d\delta p(x)}{dx}$$

dove  $S$  è la sezione del diodo.

$$I_h(x) = \frac{qSD_h}{L_h} p_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) e^{-\frac{x}{L_h}}$$

Notiamo che *la corrente di lacune è funzione di  $x$ .*

Ora è evidente che la corrente che passa in ogni sezione del diodo deve essere costante. Chiamiamo questa corrente  $I_{TOT}$ . Nella zona  $p$  ( $x \leq -x_p$ ) la concentrazione di portatori è uniforme (eccesso di elettroni iniettati trascurabile): la corrente è solo corrente di trascinamento di lacune. Tuttavia, non è possibile calcolare questa corrente dato che abbiamo supposto che tutta la ddp cada ai capi della zds ( $I^a$  ipotesi).

È necessaria a questo punto la

►III<sup>a</sup> ipotesi: non si ha ricombinazione di portatori nella zona di svuotamento.

Ne consegue che tutte le lacune che attraversano la sezione del diodo in  $-x_p$  attraverseranno anche il piano in  $x_n$ . Per  $x \geq x_n$  le lacune divengono portatori minoritari e la corrente una corrente di diffusione. Quindi

$$I_{TOT} = I_{diff}(0) = \frac{qSD_h}{L_h} p_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) = I_0 \left( e^{\frac{V}{V_T}} - 1 \right)$$

*La corrente  $I_{TOT}$  nella giunzione  $p^+n$  si calcola valutando la corrente di diffusione di lacune in  $x = 0$ .*

Esempio

Calcoliamo  $I_0$  per la giunzione di riferimento.

- 1) sezione  $S$ :  $100 \times 100 \mu\text{m}^2 = 10^{-4} \text{ cm}^2$ ;
- 2)  $N_D = 10^{15} \text{ cm}^{-3}$ ;  $N_A = 10^{19} \text{ cm}^{-3}$ ;
- 3)  $\mu_{h_n} = 490 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;  $\mu_{h_p} = 60 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;  
 $\mu_{n_p} = 90 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;  $\mu_{n_n} = 1300 \text{ cm}^2\text{s}^{-1}\text{V}^{-1}$ ;
- 4)  $\tau = 10^{-5} \text{ s}$ ;
- 5)  $V_0 = 0.026 \times \ln \left( \frac{10^{15} \times 10^{19}}{2.25 \times 10^{20}} \right) = 0.817 \text{ V}$ .

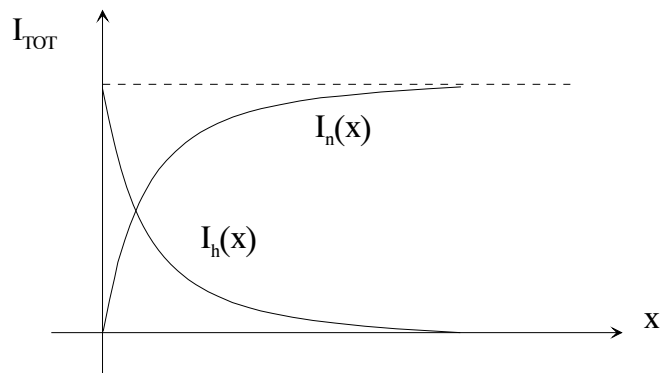
$$I_0 = \frac{qSD_h}{L_h} p_{n0}$$

$$I_0 = \frac{1.6 \times 10^{-19} \times 10^{-8} \times 0.026 \times 490 \times 10^{-4}}{\sqrt{0.026 \times 490 \times 10^{-4} \times 10^{-5}}} \times \frac{2.25 \times 10^{32}}{10^{21}} = 4.1 \times 10^{-15} \text{ A.}$$

Per ogni  $x > 0$ , poiché  $I_{hdiff}(x) < I_{hdiff}(0)$ , avremo ( $I_{TOT}$  deve essere costante)

$$I_{TOT} = I_{hdiff}(x) + I_n(x) = I_{hdiff}(x) + I_{ndiff}(x) + I_{ndrift}(x)$$

La figura mostra l'andamento di  $I_{hdiff}(x)$  e di  $I_n(x)$ .



Ci si può domandare a questo punto perché nella  $I_{TOT}$  non sia stata considerata anche una componente  $I_{hdrift}(x)$ . Ricordiamo la I<sup>a</sup> ipotesi: *tutta* la tensione cade ai capi della zona di svuotamento. Se l'ipotesi è ragionevolmente vera allora la caduta nel silicio  $n$  (zona quasi neutra) deve essere piccola. Ma per la II<sup>a</sup> ipotesi siamo in condizione di bassa iniezione e quindi  $p_n(x)$  è trascurabile in confronto a  $n_n(x)$ . Le due circostanze rendono il valore di  $I_{hdrift}(x)$  trascurabile.

Discutiamo il comportamento di  $I_n(x)$ . Le lacune iniettate in  $x_n$  richiamano un'uguale concentrazione di elettroni per mantenere la neutralità di carica. Entrambi i tipi di portatori diffondono verso destra dando luogo a correnti di diffusione. Per gli elettroni

$$I_n = qAD_n \frac{d\delta n(x)}{dx}$$

questa componente non solo scorre in verso opposto alla corrente di diffusione di lacune ma è anche maggiore in modulo ( $D_n > D_h$ ).

Per far sí che  $I_{TOT}$  sia costante è necessario che la componente di trascinamento di elettroni abbia l'andamento indicato in figura (le scale degli assi  $x$  e  $y$  sono del tutto arbitrarie).

Per  $x$  tali che  $I_{hdiff}(x) = 0$  la  $I_{TOT}$  si riduce ad una sola componente di trascinamento di elettroni.

Riepilogo del procedimento.

► *La ddp applicata ai terminali del diodo cade interamente ai capi di  $W$ .*

Che questa sia un'approssimazione è evidente. Infatti, a grande distanza da  $x_n$  e  $-x_p$  (qualche lunghezza di diffusione), le zone  $p$  ed  $n$  si comportano come semplici resistori in cui la corrente  $I_{TOT} = I$  provoca una caduta, la cui entità dipende proprio da  $I$ . Indicando con  $R$  tutti gli effetti resistivi del diodo si dovrà scrivere

$$I = I_0 \left( e^{\frac{V - RI}{V_T}} - 1 \right)$$

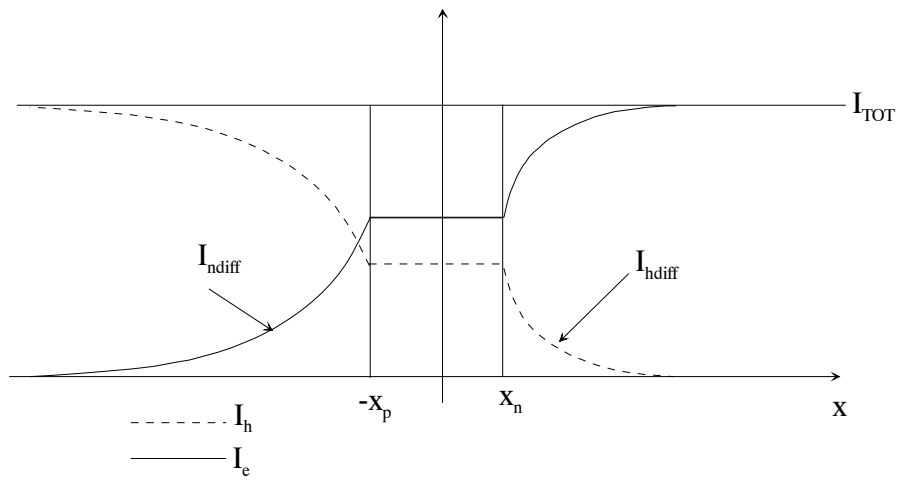
Ritorniamo più avanti sull'effetto che ha  $R$  sulla caratteristica del diodo.

► *La riduzione della barriera provoca l'iniezione di portatori minoritari nei due lati della giunzione.*

► *La valutazione della concentrazione dei portatori in eccesso viene fatta utilizzando espressioni valide all'equilibrio, nell'ipotesi che le deviazioni dall'equilibrio stesso siano trascurabili. Essenziale a questo scopo è l'ipotesi di bassa iniezione.*

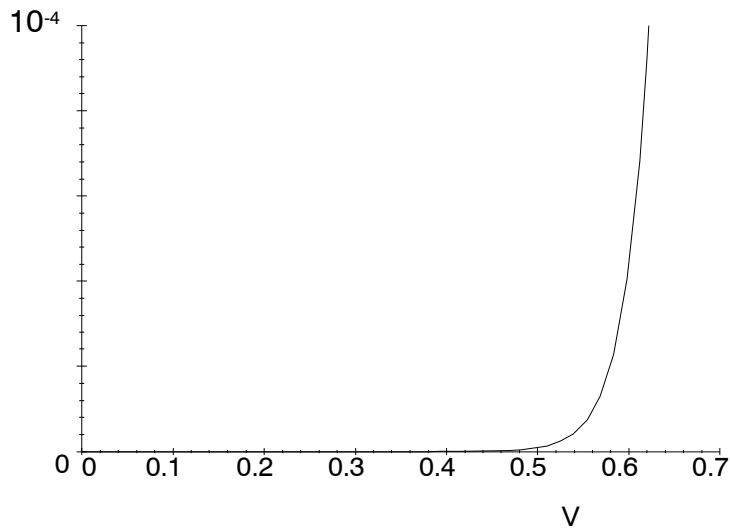
► *La corrente del diodo si valuta calcolando le correnti di diffusione dei minoritari in  $x_n$  e  $-x_p$ .*

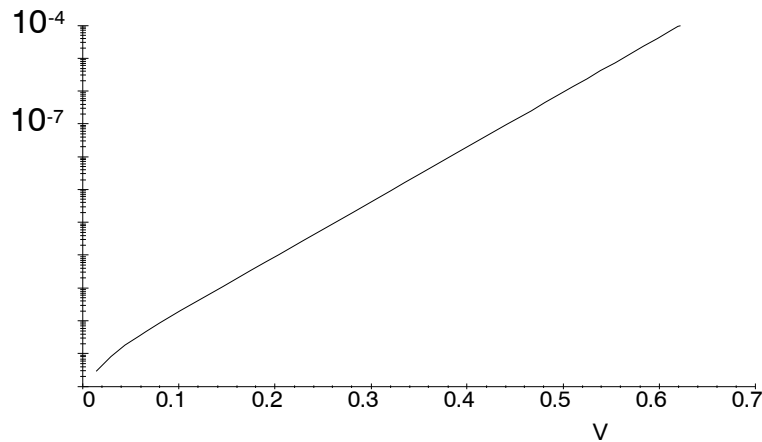
Si riporta qui di seguito il grafico delle intensità di corrente di elettroni e di lacune per una giunzione non unilaterale, per la quale l'iniezione di entrambi i tipi di portatori è rilevante per il calcolo di  $I$ .



Esempio di caratteristica  $I - V$

$$I = I_0 \left( e^{\frac{V}{V_T}} - 1 \right) = 4.1 \times 10^{-15} \times \left( e^{\frac{V}{0.026}} - 1 \right)$$





I valori della corrente per  $V \leq 0.5$  V ( $I \leq 2 \times 10^{-8}$  A) non sono leggibili sulla scala lineare, per cui generalmente la caratteristica è rappresentata in forma semi-logaritmica (si noti che per valori di tensione inferiori a 0.05 V non si può più trascurare il  $-1$ ).

### Dipendenza della corrente dalla temperatura

Una delle caratteristiche fondamentali dei semiconduttori è la dipendenza delle proprietà elettriche dalla temperatura. Ne consegue che nei dispositivi a semiconduttore le caratteristiche sono funzione della temperatura  $T$ . Questa dipendenza, generalmente non desiderabile, ha una notevole influenza anche sulle soluzioni da adottare in fase di progettazione circuitale.

La corrente di saturazione per una giunzione qualsiasi ha per espressione

$$I_0 = \frac{qSD_h}{L_h} p_{n0} + \frac{qSD_n}{L_n} n_{p0}$$

che può essere scritta

$$I_0 = \frac{qSD_h}{L_h} \frac{n_i^2}{N_D} + \frac{qSD_n}{L_n} \frac{n_i^2}{N_A}$$

dove le quantità che dipendono dalla temperatura sono  $D_{h,n}$ ,  $L_{h,n}$ ,  $n_i^2$ . È noto che

$$n_i^2 = N_C N_V e^{-\frac{E_g}{kT}}$$

Ammettiamo per il momento che  $D_h$ ,  $L_h$  siano indipendenti da  $T$ . Segue che

$$I_0 = \left( \frac{qSD_h}{L_h N_D} + \frac{qSD_n}{L_n N_A} \right) n_i^2 = B n_i^2$$

in cui  $N_C$  e  $N_V$  dipendono da  $T$ :



$$N_C = 2 \left( \frac{2\pi m_{ed}^* kT}{h^2} \right)^{\frac{3}{2}}$$

$$N_V = 2 \left( \frac{2\pi m_{hd}^* kT}{h^2} \right)^{\frac{3}{2}};$$

il loro prodotto dipende da  $T^3$ . In definitiva

$$I_0 = B'T^3 e^{-\frac{E_g}{kT}}$$

in cui si può notare che la dipendenza più importante dalla temperatura è quella esponenziale. Assumiamo ulteriormente che

$$I_0 \propto e^{-\frac{E_g}{kT}};$$

questa approssimazione non altera la sostanza delle considerazioni che stiamo per fare.

Supponiamo che la giunzione sia polarizzata direttamente e che  $V \gg V_T$ :

$$I \simeq I_0 e^{\frac{V}{V_T}} \propto e^{-\frac{E_g}{kT}} e^{\frac{V}{V_T}} = e^{\frac{(qV - E_g)}{kT}}$$

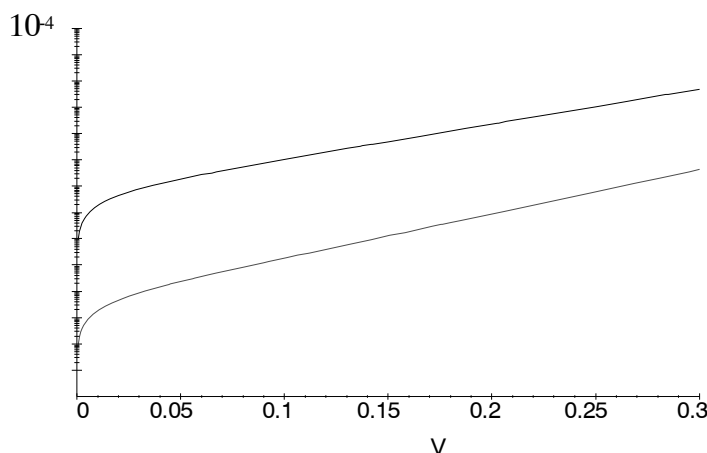
Segue che, per una data  $V$  applicata al diodo,  $I$  aumenta, al crescere di  $T$ , se  $qV < E_g$  e diminuisce se  $qV > E_g$ . L'ultima possibilità è da escludere. Infatti come abbiamo già visto, indipendentemente dalle approssimazioni usate per ricavare la caratteristica del diodo, la tensione che cade ai capi della zona di svuotamento non è indipendente dalla corrente

$$I = I_0 \left( e^{\frac{V - RI}{V_T}} - 1 \right):$$

al crescere di  $I$  diminuisce la frazione di tensione che cade ai capi di  $W$ . Per questo motivo si può senz'altro affermare che è sempre  $qV < E_g$ .

► Per ogni  $V$  la corrente  $I$  aumenta all'aumentare di  $T$ .

La figura mostra in modo qualitativo due caratteristiche (dirette) per due diversi valori di  $T$  (300 e 380 K).



### Modello del controllo di carica

Consideriamo la solita giunzione  $p^+n$  e calcoliamo la carica totale in eccesso dovuta alle lacune iniettate in condizioni di stazionarietà;  $Q$  è uguale all'integrale del profilo di eccesso moltiplicato per  $q$  e per la sezione  $S$

$$Q = qS \int_0^{\infty} \delta p(x) dx = qSp_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) \int_0^{\infty} e^{-\frac{x}{L_h}} dx = qSp_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) L_h.$$

Se portiamo la ddp a zero questo eccesso si ricombina in un tempo che vale mediamente  $\tau$ ; in altri termini se la condizione di stazionarietà deve permanere, è necessario che una carica pari a  $Q$  venga rifornita ogni tempo  $\tau$  mediante iniezione dal lato  $p^+$ . La corrente  $I$  del diodo sarà quindi

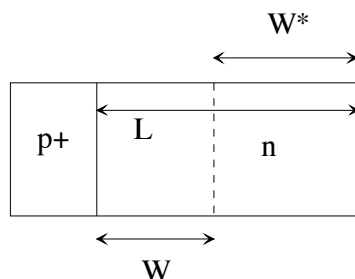
$$I = \frac{Q}{\tau} = \frac{qSp_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) L_h}{\tau} = \frac{qSp_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) L_h}{\tau} \frac{L_h}{L_h} = \frac{qSp_{n0} \left( e^{\frac{V}{V_T}} - 1 \right) D_h}{L_h}$$

che è la stessa ottenuta calcolando la derivata del profilo in zero e moltiplicandola per  $-qD_h$ .

L'utilizzo del modello del controllo di carica si rivela utile nella trattazione di alcuni aspetti del comportamento dei dispositivi, soprattutto bipolari.

### Il diodo a base corta

Con riferimento alla figura, che rappresenta schematicamente la struttura di una giunzione  $p^+n$  brusca, supponiamo che  $W^* \ll L_h$ . Il diodo è polarizzato direttamente.



Si tratta di risolvere la solita equazione di continuità con condizioni al contorno diverse. La soluzione della

$$D_h \frac{d^2 \delta p(x)}{dx^2} = \frac{\delta p(x)}{\tau_h}$$

si scrive

$$\delta p(x) = A e^{-\frac{x}{L_h}} + B e^{\frac{x}{L_h}}$$

nell'intervallo  $0 \leq x \leq W^*$

Poiché abbiamo assunto  $W^* \ll L_h$  sarà sicuramente  $x \ll L_h$  in tutto l'intervallo.

Sviluppando in serie gli esponenziali avremo

$$\delta p(x) \simeq A \left(1 - \frac{x}{L_h}\right) + B \left(1 + \frac{x}{L_h}\right) = A^* + B^* \frac{x}{L_h}$$

In  $x = 0$  niente è cambiato rispetto al caso precedente (diode a base lunga); quindi

$$\delta p(0) = p_{no} \left( e^{\frac{qV}{kT}} - 1 \right)$$

In  $x = W^*$  facciamo l'ipotesi che il contatto ohmico sia in grado di far ricombinare tutte le lacune in eccesso, per cui

$$\delta p(W^*) = 0.$$

La soluzione, con le condizioni al contorno in 0 e  $W$ , diviene

$$\delta p(x) = p_{no} \left( e^{\frac{qV}{kT}} - 1 \right) \left( 1 - \frac{x}{W^*} \right) = \delta p(0) \left( 1 - \frac{x}{W^*} \right)$$

Ricordando che

$$J_{hdiff} = -q D_h \frac{d\delta p(x)}{dx}$$

(si noti che adesso la derivata è costante per ogni  $x$ ) si ottiene

$$I = \frac{SqD_h p_{no}}{W^*} (e^{\frac{qV}{kT}} - 1) = I_0 (e^{\frac{qV}{kT}} - 1).$$

Il confronto con il diodo a base lunga rivela che adesso la  $I_0$  è determinata da  $W^*$  invece che da  $L_h$ .

► Osservazione 1

La  $\delta p(x) = \delta p(0)(1 - \frac{x}{W^*})$  non è soluzione dell'equazione di continuità, a meno di non porre  $\tau_h \rightarrow \infty$ .

► Osservazione 2

Se la densità corrente di diffusione di lacune è costante per ogni  $x$ , ciò significa che le lacune, nel transitare da 0 a  $W^*$ , non hanno avuto modo di ricombinarsi. In questo senso  $\tau_h \rightarrow \infty$ .

► Osservazione 3

Se si applica il modello del controllo di carica al caso presente, si ottiene una corrente che non ha niente a che fare con quella ottenuta calcolando la derivata del profilo.

Infatti dividendo per  $\tau_h$  la carica totale in eccesso sottesa dal profilo si ottiene:

a) zero, se si pone  $\tau_h = \infty$

b)

$$\frac{Sp_{no}(e^{\frac{qV}{kT}} - 1)W^*}{2\tau} = \frac{Sp_{no}W^*}{2\tau} (e^{\frac{qV}{kT}} - 1)$$

se si utilizza il valore di  $\tau$  caratteristico del tratto di silicio  $n$ , che è palesemente diversa dalla  $\frac{SqD_h p_{no}}{W^*} (e^{\frac{qV}{kT}} - 1)$ .

L'apparente incongruenza si risolve se si considera che il modello del controllo di carica stabilisce che:

*Dato un certo profilo di minoritari in eccesso, la condizione di stazionarietà viene mantenuta se la diminuzione nel tempo dei portatori in eccesso, dovuta a una qualsivoglia causa, viene compensata mediante l'afflusso di nuovi portatori.*

Nel caso del diodo a base lunga l'eccesso dei minoritari tendeva a diminuire per effetto della ricombinazione e quindi doveva essere ripristinato mediamente ogni  $\tau_h$  secondi; nel diodo a base corta l'eccesso tende a diminuire perché i portatori vengono portati via al contatto ohmico. Il profilo deve quindi essere ripristinato ogni  $\tau_t$  secondi, dove  $\tau_t$  è il tempo di transito, ovvero il tempo che mediamente impiega una lacuna per passare da 0 a  $W^*$ .

Resta dunque da calcolare  $\tau_t$ .

$$J_{hdiff} = -qD_h \frac{d}{dx} [\delta p(0)(1 - \frac{x}{W^*})] = \frac{qD_h \delta p(0)}{W^*}$$

che può essere posta formalmente uguale a

$$J_{hdiff} = \rho(x)v_{diff}$$

dove  $\rho(x)$  è la densità di carica associata ai portatori coinvolti nel trasporto. Uguagliando le due si ottiene la velocità di diffusione

$$v_{diff}(x) = \frac{D_h}{W^* \left(1 - \frac{x}{W^*}\right)}.$$

Si può adesso calcolare il tempo di transito dalla relazione ( $v_{diff}(x) = \frac{dx}{dt}$ )

$$\int_0^{\tau_t} dt = \int_0^{W^*} \frac{dx}{v_{diff}(x)} = \int_0^{W^*} \frac{W^* \left(1 - \frac{x}{W^*}\right) dx}{D_h}$$

che risulta quindi essere uguale a

$$\frac{W^{*2}}{2D_h} = \tau_t$$

Il modello del controllo di carica

$$\frac{qSp_{no} \left(e^{\frac{qV}{kT}} - 1\right) W^*}{2\tau_t} = \frac{qAp_{no} D_h}{W^*} \left(e^{\frac{qV}{kT}} - 1\right) = I$$

fornisce ora lo stesso risultato ottenuto calcolando la corrente come corrente di diffusione.

Esempio

a) Base lunga

Sappiamo che per la giunzione di riferimento

1) sezione  $S$ :  $100 \times 100 \mu\text{m}^2 = 10^{-4} \text{ cm}^2$ ;

2)  $N_D = 10^{15} \text{ cm}^{-3}$ ;  $N_A = 10^{19} \text{ cm}^{-3}$ ;

3)  $\mu_{h_n} = 490 \text{ cm}^2 \text{ s}^{-1} \text{ V}^{-1}$ ;  $\mu_{h_p} = 60 \text{ cm}^2 \text{ s}^{-1} \text{ V}^{-1}$ ;

$\mu_{n_p} = 90 \text{ cm}^2 \text{ s}^{-1} \text{ V}^{-1}$ ;  $\mu_{n_n} = 1300 \text{ cm}^2 \text{ s}^{-1} \text{ V}^{-1}$ ;

4)  $\tau = 10^{-5} \text{ s}$ ;

5)  $V_0 = 0.026 \times \ln \left( \frac{10^{15} \times 10^{19}}{2.25 \times 10^{20}} \right) = 0.817 \text{ V}$ .

$$I_0 = \frac{qSD_h}{L_h} p_{n0} =$$

$$\frac{1.6 \times 10^{-19} \times 10^{-8} \times 0.026 \times 490 \times 10^{-4}}{\sqrt{0.026 \times 490 \times 10^{-4} \times 10^{-5}}} \times \frac{2.25 \times 10^{32}}{10^{21}} = 4.1 \times 10^{-15} \text{ A}.$$

b) Base corta.

Supponiamo che  $W^* = 3 \mu\text{m}$ . Si deve valutare  $L_h$ .

$$L_h = \sqrt{D_h \tau} = \sqrt{0.026 \times 490 \times 10^{-5}} = 1.1287 \times 10^{-2} \text{ cm} = 112.87 \mu\text{m}$$

quindi  $L_h \gg W^*$ .

$$I_0 = \frac{SqD_h p_{n0}}{W^*} = 4.1 \times 10^{-15} \frac{L_h}{W^*} = 4.1 \times 10^{-15} \times \frac{112.87}{3} = 1.54 \times 10^{-13} \text{ A.}$$

Si noti che  $W^*$  non è indipendente dalla tensione applicata; infatti  $W^* = L - W(V)$  in cui  $W$ , ampiezza della zds è, come noto, funzione della ddp applicata.

### Effetto delle resistenze parassite sulla caratteristica diretta del diodo

Prima di affrontare l'argomento in oggetto, consideriamo il circuito della figura in cui  $R$  è una resistenza posta in serie al diodo. Dato il generatore  $V_1$  quanto vale la tensione ai capi del diodo? Si scrive per le tensioni

$$V_1 = V_D + RI = V_D + RI_0 \left( e^{\frac{V_D}{V_T}} - 1 \right)$$

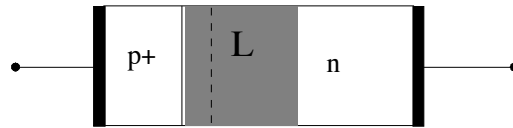
che non può essere risolta per  $V_D$ . Per la corrente  $I$  che scorre sia nel diodo che in  $R$  si ha

$$I = \frac{V_1}{R} - \frac{V_D}{R}$$

che può essere graficata sul piano  $I - V_D$  insieme alla caratteristica della giunzione ( $R = 10 \Omega$ ,  $V_1 = 4 \text{ V}$ )

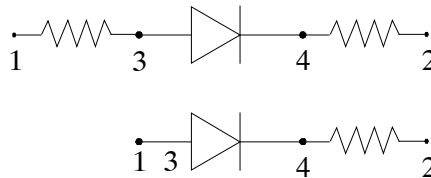
$$10^{-15} \left( e^{\frac{x}{0.026}} - 1 \right), \frac{4}{10} - \frac{x}{10}$$

Il punto di intersezione delle due curve rappresenta il *punto di lavoro* del diodo ( $I^*$ ,  $V_D^*$ ). Il metodo grafico descritto prende il nome di metodo della *retta di carico* e, come vedremo, può essere utilizzato in generale per trovare il punto di lavoro di due dispositivi in serie, note che siano le caratteristiche.



La figura rappresenta schematicamente un diodo. L'effetto diodo si manifesta nella zona di carica spaziale (zona di svuotamento) e nelle zone quasi neutre, indicate complessivamente con  $L$  nella figura. All'esterno di  $L$  sia il silicio  $p$  che quello  $n$  si comportano semplicemente come dei resistori. C'è inoltre da tener conto che i contatti  $Al/p^+$  e  $Al/n$  presenteranno delle resistenze di contatto (affronteremo in seguito il problema dei contatti  $Al/Si$ ). Tutti questi effetti possono essere descritti da un'unica resistenza in serie al diodo.

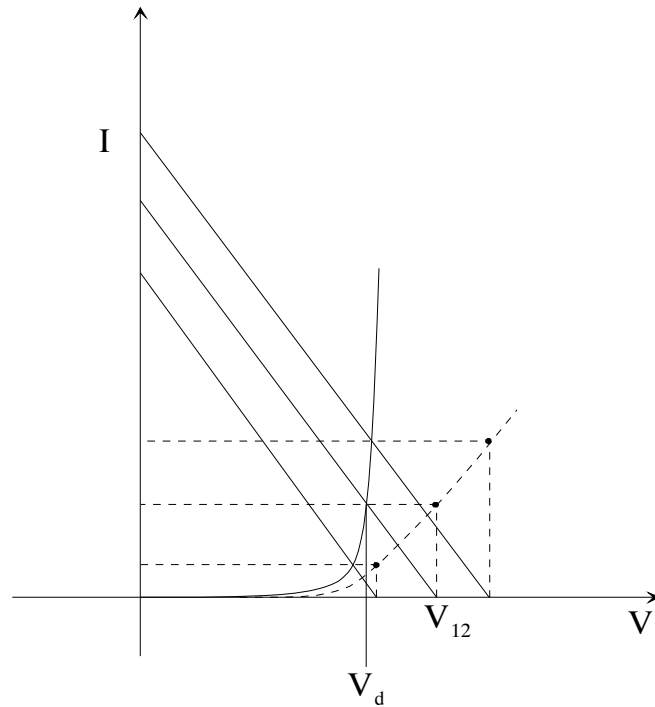
Quando si misura la caratteristica  $I - V$  si impone ai terminali 1 e 2 una tensione  $V_{12}$  e si misura una corrente  $I$ . Ma a causa della presenza della resistenza serie, la tensione  $V_{12}$  che si impone dall'esterno non è quella che cade ai capi del diodo (il diodo "ideale" si trova fra i terminali 3 e 4, inaccessibili dall'esterno). Il circuito equivalente, visto fra i terminali 1 e 2 è quindi il seguente:



Il punto di lavoro del diodo, la cui caratteristica fra i terminali 3 e 4 è rappresentata col tratto spesso nella figura, si trova con il metodo della retta di carico.

Dall'esterno si misura  $V_{12}$  e  $I$ , mentre la tensione che cade effettivamente ai capi del diodo è  $V_d$ . La caratteristica del diodo "ideale" è  $I = f(V_d)$ , mentre quella misurata è  $I = f^*(V_{12})$ , che si ottiene unendo i punti della figura.

Nell'esempio è stato fortemente esagerato l'effetto della resistenza serie per rendere più chiaro il procedimento grafico.



Le resistenze parassite del diodo, e segnatamente quelle dovute ai contatti, *dipendono criticamente dai processi*. Per quello che riguarda le resistenze di contatto metallo/Si, dato che si tratta un tipico fenomeno di interfaccia, l'espressione della  $R_c$  non può essere ovviamente quella consueta  $R = \rho L/S$ , ma piuttosto

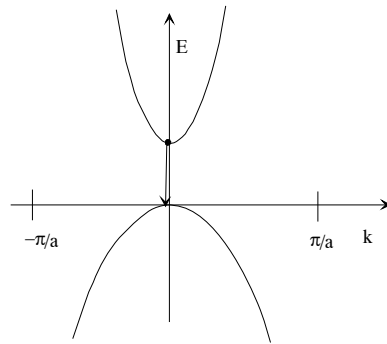
$$R_c = \frac{r_c}{S} :$$

la resistenza di contatto è inversamente proporzionale all'area di contatto  $S$ ; la costante di proporzionalità  $r_c$  prende il nome di resistenza specifica di contatto e si misura in  $\Omega \times \text{cm}^2$ .

## Il LED (Ligth Emitting Diode)

Il principio di funzionamento dei LED è basato sulla ricombinazione dei portatori che avviene in una giunzione  $pn$  in condizioni di polarizzazione diretta. Se il semiconduttore di cui è formato il diodo è a gap diretto (vedi figura) le ricombinazioni avvengono prevalentemente mediante emissione di un fotone con energia pari a quella del gap.





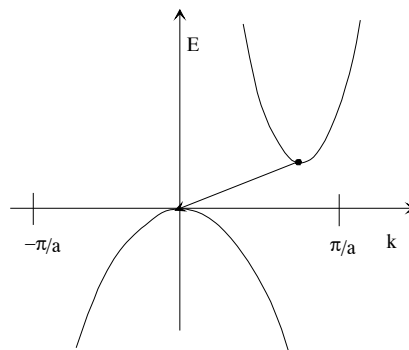
Calcoliamo la quantità di moto di un fotone con lunghezza d'onda pari a 700 nm.

$$p_{ph} = \hbar k = \frac{h}{\lambda} = \frac{6.62 \times 10^{-34}}{7 \times 10^{-7}} = 9.45 \times 10^{-28} \text{ mKgs}^{-1}$$

e confrontiamola con quella di un elettrone con  $k = \pi/2a$  ( $a$  è la costante reticolare pari, in questo esempio, a 4 Å)

$$p_{el} = \hbar k = \frac{h}{2\pi} \frac{\pi}{2a} = \frac{6.62 \times 10^{-34}}{4 \times 4 \times 10^{-10}} = 4.13 \times 10^{-25} \text{ mKgs}^{-1};$$

è evidente che  $p_{ph} \ll p_{el}$ . Un elettrone della BC, come quello della figura può eseguire la transizione indicata con conservazione dell'energia (emissione di un fotone con  $h\nu = E_g$ ) e della quantità di moto, la cui variazione (piccola), nel caso di semiconduttore a gap diretto, è dovuta al fotone stesso. Nel caso invece del Si, il gap è indiretto e la ricombinazione non può avvenire nel modo precedente, perché la variazione (grande) della quantità di moto dell'elettrone non può essere attribuita al fotone.



In un LED la potenza ottica emessa si calcola utilizzando il modello a controllo di carica. Infatti

$$I = \frac{Q}{\tau}$$

e il numero di ricombinazioni nell'unità di tempo si ottiene dividendo per  $q$

$$N = \frac{Q}{q\tau};$$

nell'ipotesi che ogni ricombinazione dia luogo all'emissione di un fotone con energia  $h\nu$ , la potenza ottica emessa è semplicemente

$$P_{OP} = N h\nu = \frac{Q}{q\tau} h\nu = \frac{I}{q} E_g.$$

Un LED polarizzato con  $I = 10$  mA e che emetta a  $7000 \text{ \AA}$ , ha una  $P_{OP}$

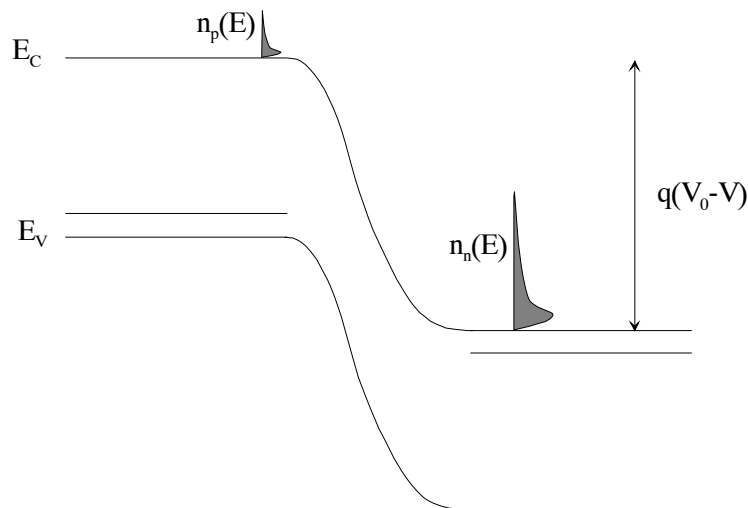
$$P_{OP} = \frac{10^{-2}}{1.6 \times 10^{-19}} \frac{6.62 \times 10^{-34} \times 3 \times 10^8}{7 \times 10^{-7}} = 1.77 \times 10^{-2} \text{ W}.$$

Esempio

Un LED rosso commerciale è realizzato in GaAlAs (semiconduttore composto *ternario* a gap diretto).

### Polarizzazione inversa

L'ampiezza della zona di svuotamento e l'altezza di barriera vista dai maggioritari aumenta.



La distribuzione  $n_n(E)$  viene spostata verso il basso fino a che nessuno degli elettroni  $\in n_n(E)$  riesce a superare la barriera. Affinché si verifichi questa condizione basta uno spostamento verso il basso di qualche  $kT$ .

Solo i minoritari (l'integrale della  $n_p(E)$ ) passano dal lato  $p$  a quello  $n$  per effetto del campo (che aumenta in polarizzazione inversa). E' evidente a questo

punto che la giunzione  $pn$  è un dispositivo il cui funzionamento è basato sui portatori minoritari.

Ricordiamo che

$$n(E) = 2D(E)f(E)$$

per il Si non-degenere

$$n(E) = 2 \cdot 2\pi \left( \frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} (E - E_C)^{\frac{1}{2}} e^{-\frac{E-E_F}{kT}}$$

Mentre in polarizzazione diretta si aveva un'iniezione, adesso abbiamo un'estrazione di portatori minoritari.

La soluzione dell'equazione di continuità è sempre la stessa

$$\delta p(x) = \delta p(0) e^{-\frac{x}{L_h}}$$

ma con condizioni al contorno diverse. Se il campo in  $x_n$  porta a zero la concentrazione dei minoritari bisogna imporre che  $\delta p(0) = -p_{n0}$  ( $\delta p(0) = p_{n0} \left( e^{\frac{V}{V_T}} - 1 \right)$ ), per cui

$$\delta p(x) = -p_{n0} e^{-\frac{x}{L_h}};$$

il profilo dei minoritari a destra di  $x_n$ , assunto come origine delle coordinate, sarà

$$p_n(x) = p_{n0} \left( 1 - e^{-\frac{x}{L_h}} \right).$$

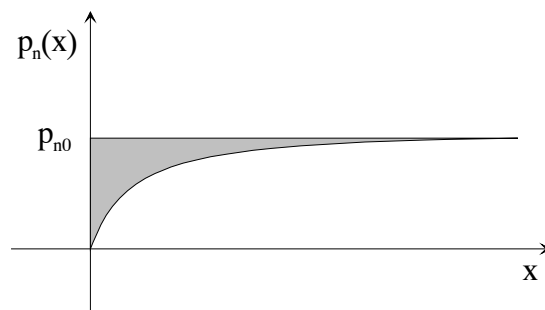
Noto il profilo il calcolo della corrente può essere effettuato:

1)

$$I = -I_0 = -qSD_h \frac{d\delta p(x)}{dx} \Big|_{x=0} = -\frac{qSD_h p_{n0}}{L_h}$$

oppure

2) utilizzando il metodo del controllo di carica.



L'eccesso è ora *negativo* (vedi figura), per cui

$$I = -\frac{Q}{\tau} = -\frac{1}{\tau} q S p_{n0} \int_0^{\infty} e^{-\frac{x}{L_h}} dx = -\frac{q S p_{n0} L_h}{\tau} = -\frac{q S D_h p_{n0}}{L_h}.$$

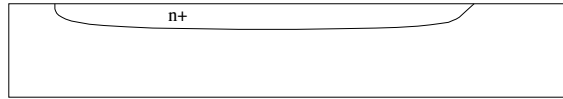
Utilizzeremo ancora in seguito questo metodo per il calcolo delle correnti nei BJT.

Nei pressi di  $x_n$  si ha uno svuotamento di portatori minoritari per cui  $np < n_i^2$ ; sono i processi di generazione che continuano a rifornire i portatori necessari al mantenimento della corrente inversa di saturazione.

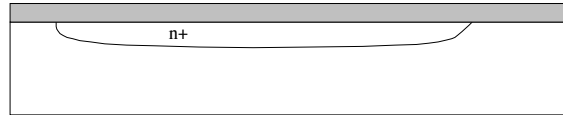
2) misura della capacità.

Un condensatore integrato la cui capacità sia indipendente dalla tensione si realizza come segue.

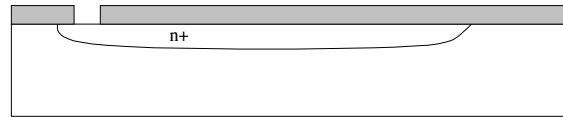
a) Una diffusione di tipo  $n^+$  (vedremo poi su quale substrato);



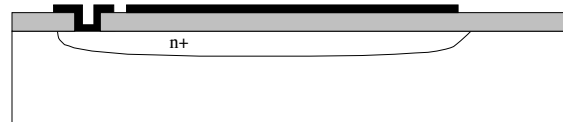
b) la crescita di uno strato di ossido;



c) l'apertura di una finestra nell'ossido;



d) la deposizione di una metal e la sua definizione.



La diffusione  $n^+$ , contattata dalla metal, costituisce l'armatura inferiore del condensatore.

La sua capacità vale

$$C = \varepsilon_{ox} \frac{S}{x_{ox}}.$$

Se  $S = 100 \times 100 \mu\text{m}^2$  e  $x_{ox} = 50 \text{ nm}$

$$C = 3.9 \times 8.85 \times 10^{-12} \frac{10^{-8}}{50 \times 10^{-9}} = 6.90 \text{ pF};$$

come si vede una capacità piccola che occupa una superficie notevole.

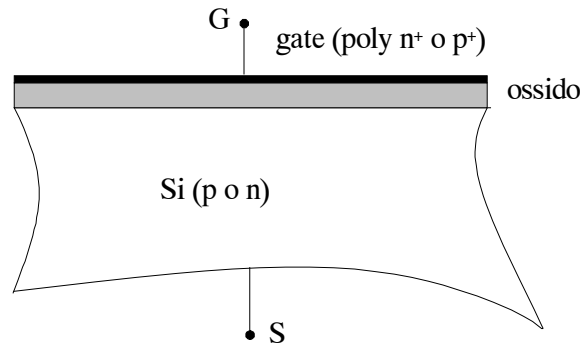
La costante dielettrica relativa dell' $\text{SiO}_2$  vale 3.9.

Da una misura di capacità si risale banalmente allo spessore. Tuttavia, come vedremo, per la misura dello spessore è più semplice utilizzare una struttura un pò diversa, ovvero il condensatore MOS.

## 15) La struttura MOS

### La struttura MOS ideale

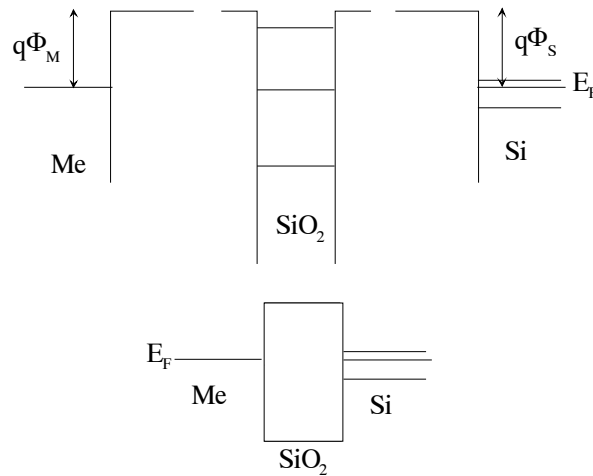
Il comportamento elettrico dei transistori ad effetto di campo (MOSFETs: *Metal Oxide Field Effect Transistors*) è strettamente legato alle proprietà della struttura MOS, che costituisce la parte attiva del dispositivo. Il condensatore MOS è poi indispensabile come strumento per la caratterizzazione elettrica dell'ossido ed a loro volta, parametri propri dell'isolante, come la densità di carica fissa e mobile, sono da un lato legati alle tecnologie di fabbricazione e dall'altro condizionano la qualità dei dispositivi realizzati.



Le ipotesi che stanno alla base della trattazione della struttura MOS *ideale* sono:

- 1) la differenza  $q\Phi_{MS}$  tra le funzioni di lavoro del metallo (gate) e del semiconduttore è uguale a zero;
- 2) la resistività dell'isolante è infinita;
- 3) non ci sono cariche nell'ossido (né fisse né mobili).

Con le ipotesi semplificative di cui sopra la struttura a bande in assenza di tensione applicata è mostrata nella figura (prima e dopo il contatto). Si noti che a contatto avvenuto l'interesse, come in generale accade, si sposta sulla struttura a bande *interna*.



L'ipotesi di uguaglianza delle funzioni di lavoro fra il gate, costituito da un certo metallo, e il Si con un dato drogaggio non è più vera, pur lasciando invariato il materiale metallico, se si cambia il valore del drogaggio.

### La tensione di soglia $V_{TH}$

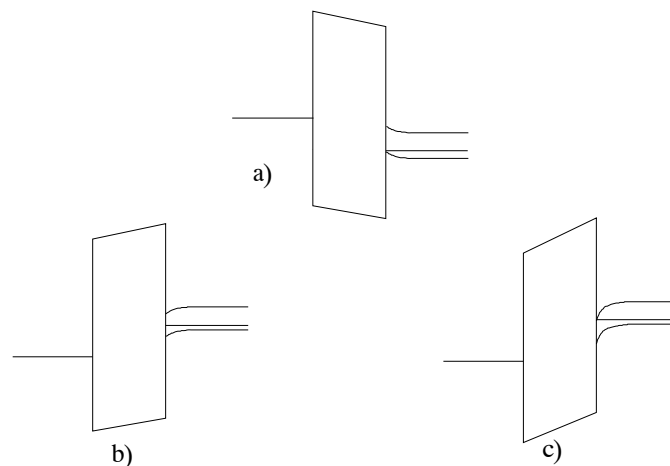
Consideriamo la struttura MOS ideale.

Poiché  $q\Phi_{MS}$  è uguale a zero non c'è caduta di potenziale interna (condizione di *bande piatte*).

In presenza di polarizzazione esterna si possono avere le tre condizioni di:

- a) *accumulazione* ( $V_{GS} < 0$ ),
- b) *svuotamento* ( $V_{TH} > V_{GS} > 0$ ),
- c) *inversione* ( $V_{GS} > V_{TH}$ ),

come è mostrato nella figura relativa ad un substrato di tipo *p*.  $V_{GS}$  è per definizione la ddp fra il gate e il Si (substrato).

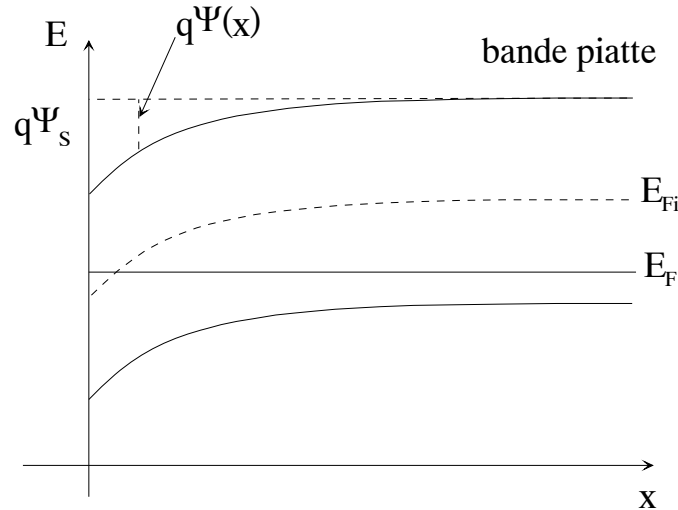


NB: nei primi dispositivi il gate era di Al, attualmente è costituito di poly ( $p^+$  o  $n^+$ ). La tendenza attuale è di utilizzare materiali metallici con alto punto di fusione.

La differenza di potenziale imposta dall'esterno si ripartisce fra isolante e semiconduttore. Sarà quindi

$$V_{GS} = V_{ox} + \Psi_S$$

dove  $\Psi_S$  è definito come la caduta totale nel semiconduttore (vedi figura).



Il comportamento del condensatore MOS è descritto a partire dall'equazione di Poisson nella zona di carica spaziale del semiconduttore; deve essere risolta senza l'approssimazione di svuotamento completo (siamo infatti interessati proprio agli elettroni di conduzione della zona di inversione).

La condizione di *inversione* si ha quando :

$$n_s \geq p_{po}$$

dove  $n_s$  è la concentrazione di elettroni in  $x = 0$ ;  $p_{po}$  vale

$$p_{po} = n_i e^{\frac{E_{Fi} - E_F}{kT}}$$

-----  
Infatti ricordiamo che

$$p_{po} = N_V e^{-\frac{E_F - E_V}{kT}}$$

ma  $p_i = N_V e^{-\frac{E_{Fi} - E_V}{kT}}$  per cui

$$p_{po} = p_i e^{\frac{E_{Fi} - E_V}{kT}} e^{-\frac{E_F - E_V}{kT}} = p_i e^{\frac{E_{Fi} - E_F}{kT}}$$

-----

Per determinare  $n_p$  in funzione di  $\Psi$  si osservi che la densità di corrente di elettroni è nulla ( $E_F$  costante nel Si).

$$j_n = q\mu_n n_p \mathcal{E} + qD_n \left( \frac{dn_p}{dx} \right) = 0$$

$$\mu_n n_p \left( \frac{d\Psi}{dx} \right) = D_n \left( \frac{dn_p}{dx} \right)$$

$$d\Psi(x) = \left( \frac{kT}{q} \right) \left( \frac{dn_p}{n_p} \right)$$

$$n_p = A \exp \left( \frac{q\Psi}{kT} \right)$$

e, per  $\Psi = 0$ , si ha  $n_p = n_{p0}$ ; quindi

$$n_p = n_{p0} \exp \left( \frac{q\Psi}{kT} \right)$$

All'inversione deve essere

$$n_s = n_{p0} \exp \left( \frac{q\Psi_s}{kT} \right) = p_{p0}$$

$$\left( \frac{n_i^2}{p_{p0}} \right) \exp \left( \frac{q\Psi_s}{kT} \right) = n_i \exp \left( \frac{E_{Fi} - E_F}{kT} \right)$$

da cui

$$n_i^2 \exp \left( \frac{q\Psi_s}{kT} \right) = p_{p0} n_i \exp \left( \frac{E_{Fi} - E_F}{kT} \right) = n_i^2 \exp \left( \frac{2(E_{Fi} - E_F)}{kT} \right);$$

segue che

$$\Psi_S = \left( \frac{2(E_{Fi} - E_F)}{q} \right) = 2\Psi_B$$

Quindi  $\Psi_S = 2\Psi_B$  è la caduta di tensione nel silicio necessaria per l'instaurarsi inversione.  $\Psi_B$  è funzione del drogaggio secondo la relazione:

$$\Psi_B = \left( \frac{kT}{q} \right) \ln \left( \frac{N_A}{n_i} \right)$$

Esempio

Se  $N_A = 10^{15} \text{ cm}^{-3}$

$$2\Psi_B = 2 \times 0.026 \times \ln \left( \frac{10^{15}}{1.5 \times 10^{10}} \right) = 0.577 \text{ V.}$$



L'equazione di Poisson nel silicio (zona di carica spaziale) è:

$$\frac{d^2 \Psi}{dx^2} = -\frac{\rho(x)}{\varepsilon_s}$$

$$\rho(x) = q (-N_A^- + p_p - n_p)$$

senza l'approssimazione di svuotamento completo. Poiché nel volume (lontano dall'interfaccia dove le bande sono piatte) c'è neutralità di carica, avremo:

$$p_{p0} - n_{p0} = N_A^-$$

$$\rho(x) = q (n_{p0} - p_{p0} + p_p - n_p)$$

e, per un generico valore di  $\Psi$  (e quindi di  $x$ ):

$$\rho(x) = q (n_{p0} - p_{p0} + p_{p0} e^{-\beta\Psi(x)} - n_{p0} e^{\beta\Psi(x)})$$

dove  $\beta = \frac{q}{kT}$ .

L'equazione da risolvere è :

$$\frac{d^2 \Psi}{dx^2} = -\frac{q}{\varepsilon_s} [p_{p0} (e^{-\beta\Psi(x)} - 1) - n_{p0} (e^{\beta\Psi(x)} - 1)]$$

Utilizzando l'uguaglianza

$$\frac{1}{2} \frac{d}{d\Psi} \left( \frac{d\Psi}{dx} \right)^2 = \frac{d^2 \Psi}{dx^2}$$

si ottiene:

$$d \left( \frac{d\Psi}{dx} \right)^2 = -\frac{2q}{\varepsilon_s} [p_{p0} (e^{-\beta\Psi(x)} - 1) - n_{p0} (e^{\beta\Psi(x)} - 1)] d\Psi.$$

Integrando

$$\left( \frac{d\Psi}{dx} \right)^2 = -\frac{2q}{\beta\varepsilon_s} [-p_{p0} e^{-\beta\Psi} - p_{p0} \beta\Psi - n_{p0} e^{\beta\Psi} + n_{p0} \beta\Psi] + C;$$

poiché il campo è nullo dove le bande sono piatte ( $-\frac{d\Psi}{dx} = 0$  per  $\Psi = 0$ ), dopo aver determinato  $C$  si ha:

$$\left( \frac{d\Psi}{dx} \right)^2 = -\frac{2q}{\beta\varepsilon_s} [p_{p0} (-e^{-\beta\Psi} - \beta\Psi + 1) + n_{p0} (-e^{\beta\Psi} + \beta\Psi + 1)]$$

che mette in relazione  $\mathcal{E}^2$  e  $\Psi$ ; ponendo in evidenza  $p_{po}$  e cambiando segno si ottiene:

$$\mathcal{E}^2 = \frac{2qp_{po}}{\beta\varepsilon_s} \left[ (e^{-\beta\Psi} + \beta\Psi - 1) + \frac{n_{po}}{p_{po}} (e^{\beta\Psi} - \beta\Psi - 1) \right]$$

Posto

$$F\left(\beta\Psi; \frac{n_{po}}{p_{po}}\right) = \left[ (e^{-\beta\Psi} + \beta\Psi - 1) + \frac{n_{po}}{p_{po}} (e^{\beta\Psi} - \beta\Psi - 1) \right]^{\frac{1}{2}}$$

il campo è dato da

$$\mathcal{E} = -\frac{d\Psi}{dx} = \pm \sqrt{\frac{2qp_{po}}{\beta\varepsilon_s}} F\left(\beta\Psi; \frac{n_{po}}{p_{po}}\right)$$

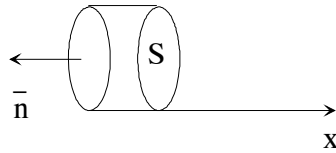
Si noti che il rapporto  $\frac{n_{po}}{p_{po}}$  è un numero molto piccolo. Infatti se  $p_{po} = 10^{15} \text{ cm}^{-3}$

$$\frac{n_{po}}{p_{po}} = \frac{2.25 \times 10^5}{10^{15}} = 2.25 \times 10^{-10}.$$

Il campo elettrico all'interfaccia Si/SiO<sub>2</sub> sarà:

$$\mathcal{E}_s = \pm \sqrt{\frac{2qp_{po}}{\beta\varepsilon_s}} F\left(\beta\Psi_s; \frac{n_{po}}{p_{po}}\right)$$

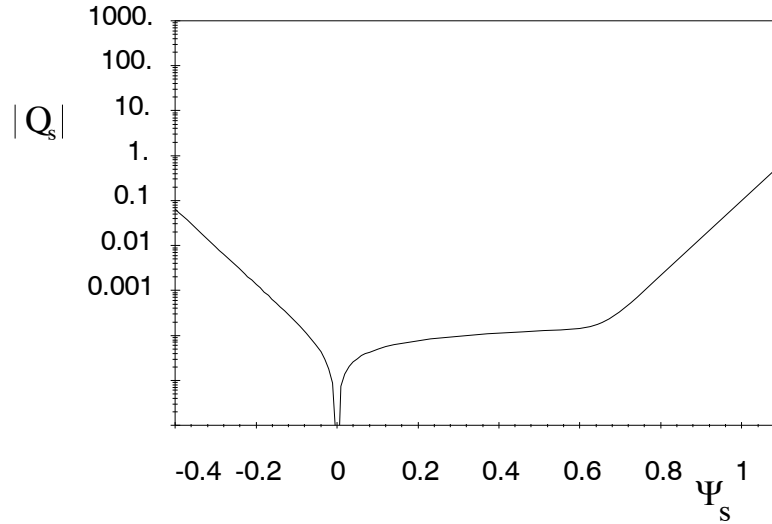
Se si applica il teorema di Gauss ad un cilindro di base unitaria, con base all'interfaccia silicio /ossido, e altezza parallela all'asse  $x$  con valore maggiore o uguale della profondità della zona di carica spaziale, si ottiene



$$Q_s = -\varepsilon_s \mathcal{E}_s = \mp \varepsilon_s \sqrt{\frac{2qp_{po}}{\beta\varepsilon_s}} F\left(\beta\Psi_s; \frac{n_{po}}{p_{po}}\right)$$

$$Q_s = \mp \sqrt{\frac{2qp_{po}\varepsilon_s}{\beta}} \left[ (e^{-\beta\Psi_s} + \beta\Psi_s - 1) + \frac{n_{po}}{p_{po}} (e^{\beta\Psi_s} - \beta\Psi_s - 1) \right]^{\frac{1}{2}}$$

L'andamento di  $|Q_s|$ , carica per unità di superficie, è illustrato nella figura in funzione di  $\Psi_s$ , sia per valori positivi che negativi. Il drogaggio del Si è  $N_A = 10^{15} \text{ cm}^{-3}$ .



Si noti che per  $\Psi_S > 2\Psi_B$ ,  $|Q_s|$  cresce esponenzialmente con  $\Psi_S$ .

Osserviamo che la carica per unità di superficie del gate ( $Q_M$ ) deve essere uguale al valore assoluto della carica totale nel semiconduttore, che è costituita in generale sia da accettori ionizzati ( $Q_W$ ) che da elettroni ( $Q_n$ ):

$$Q_M = |Q_n + qN_AW| = |Q_s|$$

La differenza di potenziale applicata al condensatore è

$$V_{GS} = V_{ox} + \Psi_S.$$

Non essendoci cariche nell'ossido,  $V_{ox} = \mathcal{E}_{ox}t_{ox}$ ; imponendo all'interfaccia la continuità del vettore induzione

$$\mathcal{E}_{ox}\varepsilon_{ox} = \mathcal{E}_s\varepsilon_s$$

si ha

$$\mathcal{E}_{ox} = \frac{\mathcal{E}_s\varepsilon_s}{\varepsilon_{ox}}$$

e

$$V_{ox} = \mathcal{E}_{ox}t_{ox} = \mathcal{E}_s\varepsilon_s \frac{t_{ox}}{\varepsilon_{ox}} = -\frac{Q_s}{C_{ox}}$$

dove  $C_{ox}$  è la capacità *per unità di superficie* dell'isolante.

La tensione di soglia, ovvero la tensione necessaria per ottenere l'inversione, sarà quindi:

$$V_{TH} = -\frac{Q_s(2\Psi_B)}{C_{ox}} + 2\Psi_B$$

in cui  $Q_s(2\Psi_B)$  è dato da

$$Q_s(2\Psi_B) = -\sqrt{\frac{2qp_{po}\varepsilon_s}{\beta}} \left[ (e^{-2\beta\Psi_B} + 2\beta\Psi_B - 1) + \frac{n_{po}}{p_{po}} (e^{2\beta\Psi_B} - 2\beta\Psi_B - 1) \right]^{\frac{1}{2}}$$

Di tutti i termini sotto radice resta soltanto  $2\beta\Psi_B = \frac{2q\Psi_B}{kT}$  (si ricordi che  $\frac{n_{po}}{p_{po}}$  è molto piccolo) e quindi

$$Q_s(2\Psi_B) = -\sqrt{\frac{2qp_{po}\varepsilon_s}{\beta}} \sqrt{2\beta\Psi_B} = -\sqrt{2\varepsilon_s 2\Psi_B q N_A}$$

Si osservi che nell'espressione precedente

$$Q_s = -(2\varepsilon_s q N_A 2\Psi_B)^{\frac{1}{2}}$$

è la carica per unità di superficie dovuta ai soli accettori ionizzati. Infatti se si risolve l'equazione di Poisson con l'approssimazione di *svuotamento completo*:

$$\frac{d^2\Psi}{dx^2} = \frac{qN_A}{\varepsilon_s}$$

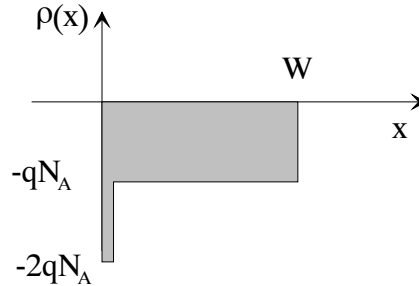
si ottiene quando  $\Psi_s = 2\Psi_B$

$$W = \left( \frac{2\varepsilon_s 2\Psi_B}{qN_A} \right)^{\frac{1}{2}}$$

ovvero:

$$Q_s = -qN_A W = -(2\varepsilon_s q N_A 2\Psi_B)^{\frac{1}{2}}$$

Quindi per  $V = V_{TH}$  la carica è costituita principalmente dagli accettori. In effetti all'inversione  $n_s = p_{po} = N_A$ ; poiché lo spessore dello strato di inversione è trascurabile ( $n(x)$  decresce esponenzialmente con  $\Psi(x)$ ) rispetto a  $W$  si può, *alla soglia*, trascurare  $Q_n$  rispetto a  $qN_A W$ .



L'espressione della *tensione di soglia della struttura MOS ideale* è

$$V_{TH_i} = \frac{(2\varepsilon_s q N_A 2\Psi_B)^{\frac{1}{2}}}{\varepsilon_{ox}} t_{ox} + 2\Psi_B$$

ed è funzione crescente del drogaggio e dello spessore dell'ossido.

Supponiamo che  $N_A = 10^{15} \text{ cm}^{-3}$ ; abbiamo già visto che  $2\Psi_B = 0.577 \text{ V}$ . Ponendo  $t_{ox} = 50 \text{ nm}$  si ottiene  $V_{THi} = 0.779 \text{ volt}$ .

La relazione fondamentale per lo studio del transistor MOS è

$$V_{GS} = -\frac{Q_n + Q_w}{C_{ox}} + 2\Psi_B$$

dove, ricordiamo,  $Q_n$  è la carica per unità di superficie relativa agli elettroni.

Per quanto riguarda quest'ultima, in condizioni di forte inversione la

$$Q_s = \sqrt{\frac{2qp_{po}\varepsilon_s}{\beta}} \left[ (e^{-\beta\Psi_s} + \beta\Psi_s - 1) + \frac{n_{po}}{p_{po}} (e^{\beta\Psi_s} - \beta\Psi_s - 1) \right]^{\frac{1}{2}}$$

può essere approssimata da

$$Q_n = \sqrt{\frac{2qp_{po}\varepsilon_s}{\beta}} \left[ \frac{n_{po}}{p_{po}} e^{\beta\Psi_s} \right]^{\frac{1}{2}} = [2kT\varepsilon_s n_{po} e^{\beta\Psi_s}]^{\frac{1}{2}} = [2kT\varepsilon_s n_s]^{\frac{1}{2}};$$

ma  $Q_n$  è data da

$$Q_n = q \int_0^\infty n_p(x) dx$$

in cui l'andamento di  $n_p(x)$  non è noto. Supponiamo tuttavia che

$$n_p(x) = n_s e^{-\frac{x}{L}}$$

(il confronto con le simulazioni ci dice che l'andamento della funzione  $n_p(x)$  è qualitativamente simile); si può così ottenere un'espressione di  $Q_n$

$$Q_n = qn_s \int_0^\infty e^{-\frac{x}{L}} dx = qn_s L$$

da cui si può ricavare  $L$

$$L = \frac{[2kT\varepsilon_s n_s]^{\frac{1}{2}}}{qn_s} = \left[ \frac{2kT\varepsilon_s}{q^2 n_s} \right]^{\frac{1}{2}};$$

$L$  rappresenta lo spessore caratteristico dello strato di inversione ed è una funzione decrescente di  $n_s$ . Diamone una stima supponendo che  $n_s = 10^{19} \text{ cm}^{-3}$  (il valore è stato ottenuto dalle simulazioni in condizione di alta inversione)

$$L = \left[ \frac{2 \times 0.026 \times 11.8 \times 8.85 \times 10^{-12}}{1.6 \times 10^{-19} \times 10^{25}} \right]^{\frac{1}{2}} \simeq 18 \text{ \AA}$$

il che conferma che lo strato di inversione si estende poco (in confronto a  $W$ ) al disotto dell'interfaccia. Riprendiamo in esame la

$$V_{GS} = -\frac{Q_n + Q_W}{C_{ox}} + 2\Psi_B.$$

E' da notare che anche in condizioni di inversione, più o meno elevata ( $Q_n \gg Q_W$ ), il termine  $2\Psi_B$  è rimasto costante.

Si tratta evidentemente di un'approssimazione, giustificata dal fatto che  $Q_W$  aumenta con la radice di  $\Psi_S$ , mentre  $Q_n$  cresce esponenzialmente: sono sufficienti piccole variazioni di  $\Psi_S$  per aumentare notevolmente la  $Q_n$  dello strato di inversione. D'ora in avanti supporremo sempre  $2\Psi_B$  costante quando  $V_{GS} > V_{THi}$ .

## 16) Il processo LOCOS per l'integrazione di NMOS

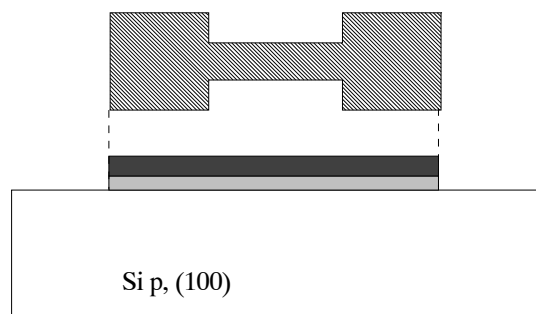
L'acronimo LOCOS sta per *LOCAl Oxidation of Silicon*.

Si parte da un substrato  $p$  poco drogato ( $10^{15} \text{ cm}^{-3}$  o meno) con orientamento (100). Il motivo di questa scelta è dovuta alla carica fissa positiva che inevitabilmente si ha all'interfaccia silicio/ossido; nel caso di substrati (100) è minore che nel caso di substrati (111). Questa carica influisce sulla tensione di soglia, come vedremo.

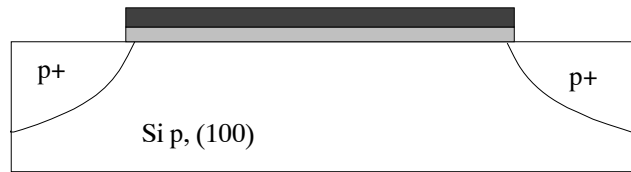
1) Si cresce un sottile (qualche nm) strato di ossido termico (*pad oxide*) e su questo si depone uno strato sottile di  $\text{Si}_3\text{N}_4$ . L' $\text{Si}_3\text{N}_4$  è generalmente in condizioni di stress e il pad oxide serve a scaricare parzialmente al suo interno queste tensioni. Il coefficiente di diffusione dell' $\text{O}_2$  nell' $\text{Si}_3\text{N}_4$  è, agli effetti pratici, nullo.



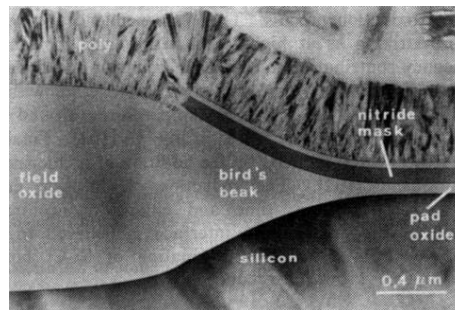
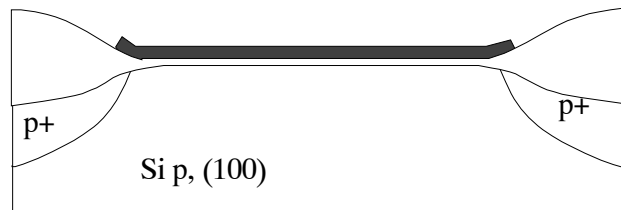
2) Si usa la prima maschera (area attiva) e si eliminano con un attacco dry (da vedere) il nitruro e il pad oxide all'esterno della zona definita dalla maschera. D'ora in avanti verranno evidenziate con tratteggio solo le geometrie *non* trasparenti. Non verrà indicata quindi la parte scura all'esterno delle geometrie trasparenti.



3) Si esegue un drogaggio di *channel stop* ( $p^+$ ). Lo scopo è quello di aumentare la tensione di soglia locale.

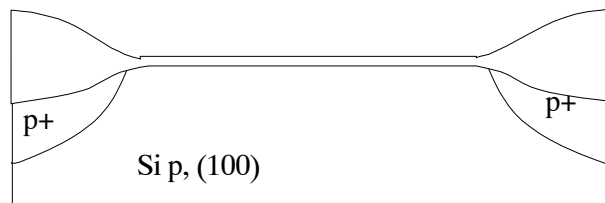


4) Si cresce uno strato di ossido spesso (*field oxide*). L'ossido cresce solo dove il substrato non è protetto dal nitruro. Tuttavia durante l'ossidazione si ha penetrazione della specie ossidante sotto il nitruro; il profilo del field oxide è mostrato in figura e dà luogo al fenomeno dei becchi d'uccello (*bird's beaks*) che è non trascurabile nel caso di lunghezze di canale submicrometriche.

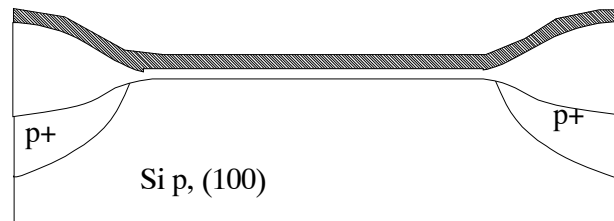


5) Si eliminano nitruro e pad oxide e si cresce un sottile strato di ossido sacrificale, che ha lo scopo di eliminare la difettosità dalla superficie libera del Si.

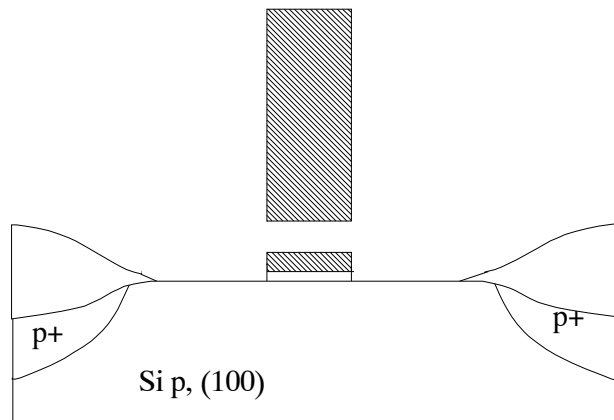
6) Si asporta l'ossido sacrificale (immersione rapida in HF) e si cresce l'ossido di gate. Si tratta un passo molto delicato. L'ossidazione è ovviamente dry.



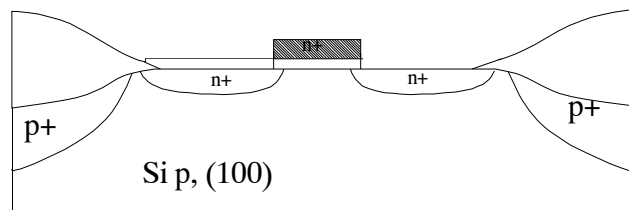
7) Si stende uno strato di poly (come nel caso dell' $\text{Si}_3\text{N}_4$  si tratta di una tecnica CVD).



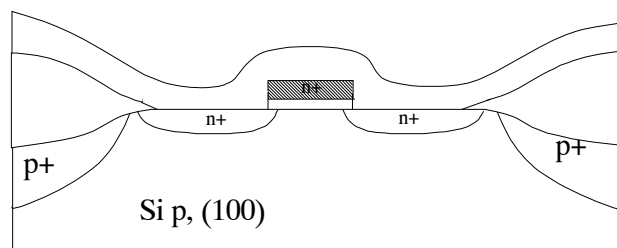
8) Si utilizza la seconda maschera per definire il gate.



9) Si droga  $n^+$ . Durante questa operazione vengono formati i pozzetti di source (S) e drain (D) e contemporaneamente si rende  $n^+$  anche il poly di gate.

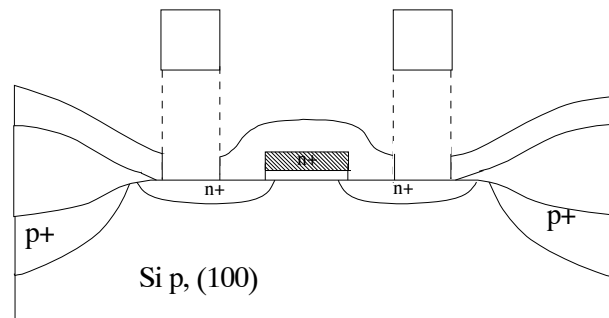


10) Si deposita uno strato di isolante.

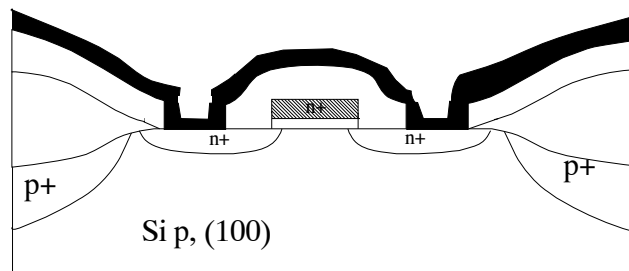


11) Si utilizza la terza maschera per aprire delle finestre in corrispondenza di S e D.

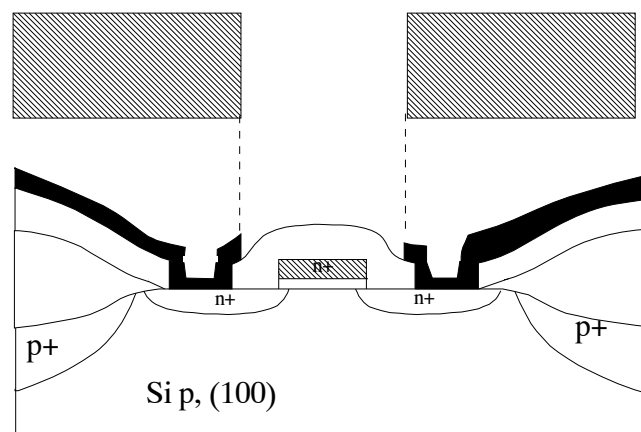




12) Si deposita un film di Al.

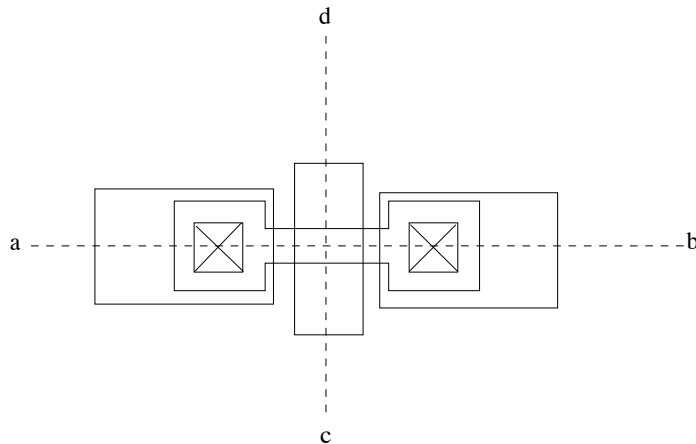


13) Si usa una quarta maschera per la definizione della metal di Al.

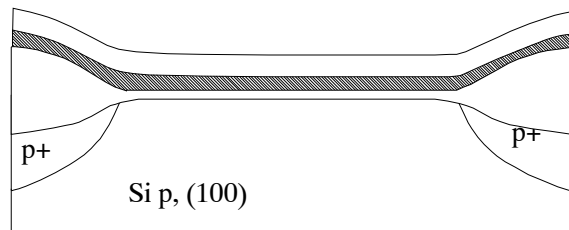


A questo stadio è completamente definita la struttura dell'NMOS, ma non è completo il processo che darà luogo al CI finale. Ritourneremo su questo aspetto in seguito.

Il layout del dispositivo è mostrato nella figura.

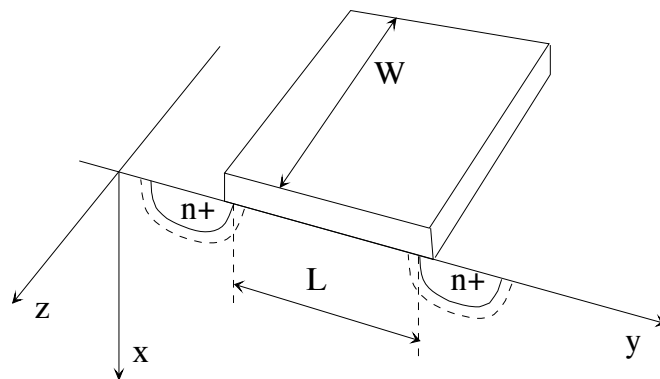


Tutto il processo è stato descritto facendo riferimento ad una sezione secondo la retta *ab*. Consideriamo una sezione *dc*.



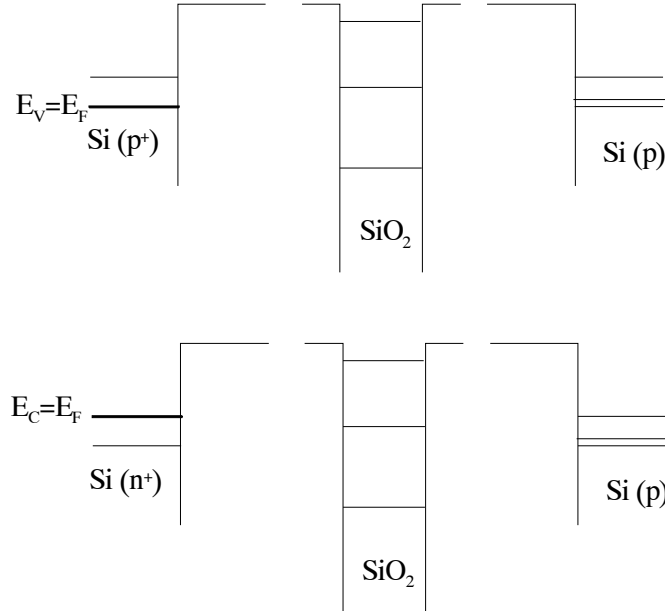
Il gate, che è equipotenziale, ha la funzione di invertire il silicio solo quando è sopra l'ossido di gate. In corrispondenza dell'ossido di campo la tensione di soglia del complesso *gate/ossido di campo/channel stop* deve essere maggiore della massima tensione  $V_{GS}$  che può essere applicata al gate, in modo da non correre il rischio che all'esterno del canale (regione definita dall'incrocio dell'area attiva con la maschera di gate) si possa verificare l'inversione.

La figura mostra una vista prospettica schematica del dispositivo in cui sono definiti i riferimenti e le dimensioni che serviranno per la trattazione dell'NMOS.



### 17) La struttura MOS reale.

La prima ipotesi da rimuovere è quella dell'uguaglianza delle funzioni di lavoro del gate e del substrato. Facciamo riferimento ai due materiali più comuni utilizzati per il gate, ovvero il poly  $p^+$  e il poly  $n^+$ . Le strutture a bande prima del contatto sono quelle seguenti.



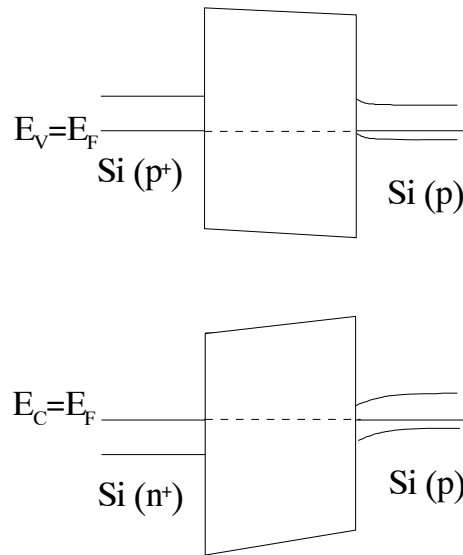
A contatto avvenuto la ddp di contatto fra gate e Si si ripartirà fra ossido e substrato. Nel primo caso (gate  $p^+$ ) questa ddp vale evidentemente  $E_F - E_V$  nel substrato e quindi

$$\Phi_{MS} = V_{FB} = \frac{kT}{q} \ln \left( \frac{N_V}{N_A} \right) > 0;$$

nel secondo (gate  $n^+$ )

$$\Phi_{MS} = V_{FB} = \frac{q\chi - \left( q\chi + E_g - kT \ln \left( \frac{N_V}{N_A} \right) \right)}{q} = -\frac{E_g}{q} + \frac{kT}{q} \ln \left( \frac{N_V}{N_A} \right) < 0.$$

Dalla struttura a bande all'equilibrio dopo il contatto



si vede che nel caso  $p^+$  il Si è in condizioni di accumulazione (si è allontanato dalla condizione di inversione), mentre nel caso  $n^+$  è avvicinato a quella condizione. In conclusione nel primo caso la tensione di soglia è aumentata, nel secondo diminuita della quantità  $V_{FB}$  ( $V_{FB}$  sta per tensione di *flat band*, e rappresenta quella tensione che si dovrebbe applicare dall'esterno per ripristinare la condizione di bande piatte).

$$V_{TH} (\text{gate } p^+) = V_{THi} + \Phi_{MS} \text{ con } \Phi_{MS} > 0$$

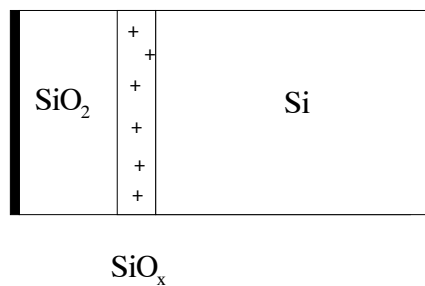
$$V_{TH} (\text{gate } n^+) = V_{THi} + \Phi_{MS} \text{ con } \Phi_{MS} < 0.$$

Il termine  $\Phi_{MS}$  nel caso di gate  $n^+$  vale

$$\Phi_{MS} = -1.08 + 0.026 \times \ln \left( \frac{10^{19}}{10^{15}} \right) = -0.84 \text{ V}$$

che è un valore considerevole se confrontato con le tensioni di soglia di dispositivi moderni (a canale corto e ossido di gate sottile), che sono dell'ordine di 0.3 V.

Resta il problema della carica fissa interfacciale, dovuta ad incompleta ossidazione. Se lo strato è molto sottile si può definire una carica fissa *positiva* per unità di superficie  $Q_{ox}$  che va a sottrarsi dalla  $Q_W$  e quindi il suo contributo alla  $V_{TH}$  sarà banalmente



$$V_{TH} = -\frac{Q_W}{C_{ox}} + 2\Psi_B + \Phi_{MS} - \frac{Q_{ox}}{C_{ox}}$$

Questa è l'espressione completa della tensione di soglia da utilizzare per valutarla numericamente caso per caso. Manca il valore di  $Q_{ox}$ , che è un tipico parametro di processo, in quanto dipende dalle condizioni di ossidazione e dai trattamenti post-ossidazione. In assenza di tali trattamenti  $Q_{ox}$  è dell'ordine di  $1.6 \times 10^{-19} \times 10^{11} = 1.6 \times 10^{-8} \text{ C}\cdot\text{cm}^{-2}$  ed quindi in grado di provocare un decremento della tensione di soglia pari a (nel caso in cui  $t_{ox} = 50 \text{ nm}$ )

$$\Delta V_{TH} = -\frac{1.6 \times 10^{-4}}{3.9 \times 8.85 \times 10^{-12}} \times 50 \times 10^{-9} = -0.23 \text{ V.}$$

Per ridurre la  $Q_{ox}$  si esegue un trattamento di annealing ( $\simeq 400 \text{ }^\circ\text{C}$ ) in forming gas ( $\text{N}_2 + \text{H}_2$ ) che riduce la densità di cariche positive da  $10^{11}$  a  $10^{10} \text{ cm}^{-2}$ .

## 18) I contatti metallo/semiconduttore

La descrizione del processo LOCOS ha messo in evidenza problemi ancora non risolti, come quello dei contatti metallo/semiconduttore. Come ormai è d'uso tratteremo prima la situazione ideale.

### I contatti M/S ideali

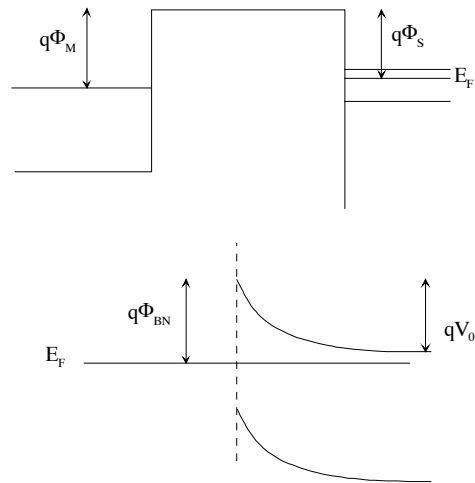
Sono quattro i casi possibili, in relazione alla differenza fra le funzioni di lavoro del metallo e del semiconduttore.

- 1) Si( $n$ )  $q\Phi_M > q\Phi_S$
- 2) Si( $n$ )  $q\Phi_M < q\Phi_S$
- 3) Si( $p$ )  $q\Phi_M > q\Phi_S$
- 4) Si( $p$ )  $q\Phi_M < q\Phi_S$

Interfaccia M/S ideale è quella per la quale le bande del Si erano piatte fino alla superficie prima che avvenisse il contatto.

- 1) Si ( $n$ )  $q\Phi_M > q\Phi_S$ .

La struttura a bande prima e dopo il contatto è mostrata in figura. Si è avuto trasferimento di elettroni dal Si al metallo. La ddp di contatto  $V_0$  è pari a  $\Phi_M - \Phi_S$  e l'ampiezza della zds si calcola risolvendo l'equazione di Poisson come nel caso della giunzione  $pn$ .

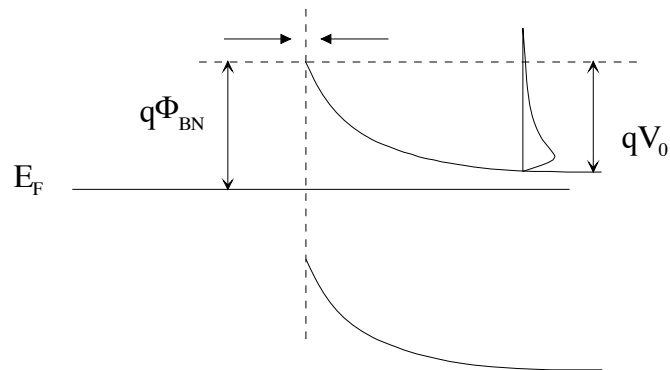


Quindi

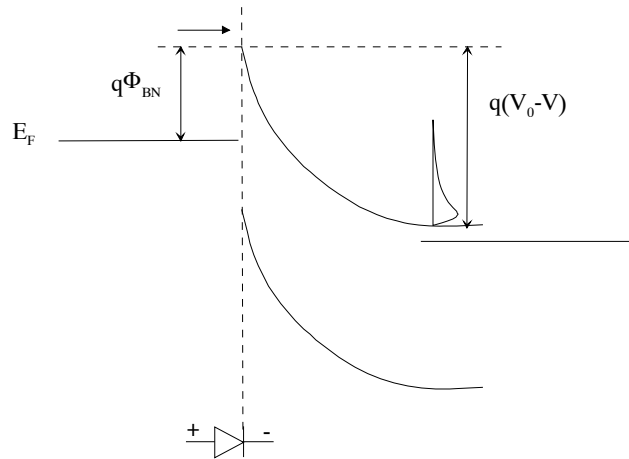
$$W = \sqrt{\frac{2\varepsilon_s V_0}{qN_D}}$$

e si estende tutta nel Si. E, come nel caso della  $pn$ , tutta la ddp si localizza ai capi della zds.

All'equilibrio gli elettroni possono passare nel metallo solo se hanno energia sufficiente a superare la barriera di potenziale; stessa cosa dicasi per gli elettroni del metallo. All'equilibrio la  $J_{TOT}$  di elettroni (che sono i maggioritari) attraverso l'interfaccia è nulla.



In condizioni di polarizzazione inversa (simbolo e riferimenti del diodo nella figura) la  $n(E)$  relativa agli elettroni del Si viene traslata verso il basso:



l'interfaccia è attraversata solo dagli elettroni che dal metallo vanno al semiconduttore (corrente inversa di saturazione  $I_0$ ). In polarizzazione diretta la  $n(E)$  viene traslata verso l'alto e il comportamento, per quanto riguarda i maggioritari del lato  $n$ , è simile a quello della giunzione  $pn$ . Con una rimarchevole differenza tuttavia: gli elettroni che passano nel metallo continuano a comportarsi da maggioritari e nel metallo non si verifica nessun effetto di iniezione etc. Il diodo SCHOTTKY, questo è il nome del dispositivo, basa il suo funzionamento sui portatori maggioritari e proprio in questo differisce profondamente dalla giunzione  $pn$ .

La caratteristica ideale è identica a quella della giunzione  $pn$

$$I = I_0(e^{\frac{V}{V_T}} - 1);$$

la corrente inversa di saturazione, in base alla teoria termoionica, è data da

$$I_0 = SA^*T^2 e^{-\frac{q\Phi_{BN}}{kT}}$$

in cui  $q\Phi_{BN}$  è l'altezza di barriera vista dal metallo e  $A^*$  è la costante di Richardson che vale, nel caso del Si,  $31.2 \text{ A}\cdot\text{cm}^{-2}\text{K}^{-2}$ .

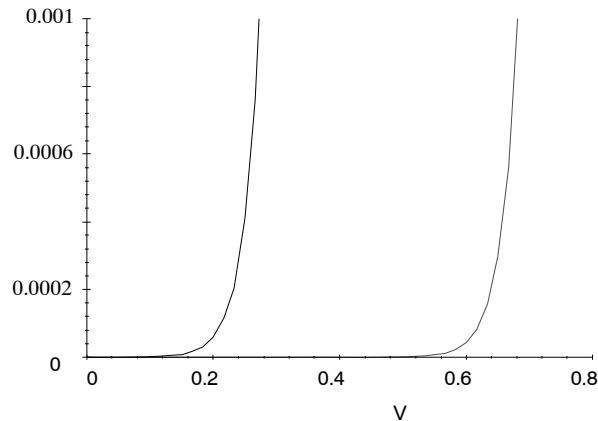
Valutiamo  $I_0$ , a  $T$  ambiente, per un diodo Schottky con  $S = 10^{-8} \text{ m}^2$  (la stessa della giunzione  $pn$  di riferimento), supponendo che  $q\Phi_{BN} = 0.6 \text{ eV}$ :

$$I_0 = 10^{-8} \times 31.2 \times 10^4 \times 300^2 \times e^{-\frac{0.6}{0.026}} = 2.67 \times 10^{-8} \text{ A}$$

un valore molto maggiore di quello della giunzione  $pn$  di riferimento

$$(I_{0pn} = 4.1 \times 10^{-15} \text{ A}).$$

Nel grafico sono rappresentate le due caratteristiche su scala lineare.

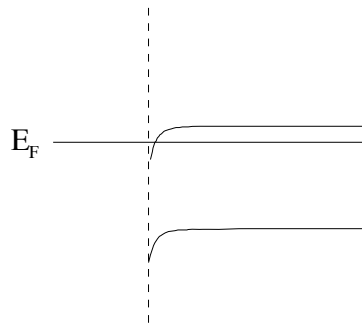


In termini circuitali si può dire, con riferimento alla figura, che la  $V_\gamma$  del diodo Schottky è minore della  $V_\gamma$  della giunzione  $pn$  di pari sezione.

N.B. La  $V_\gamma$  non ha alcun significato fisico: non compare in effetti nella caratteristica  $I - V$  dei due tipi di diodo. E' errore abbastanza frequente attribuire a  $V_\gamma$  il significato di tensione di soglia, nonostante  $I$  sia uguale a zero solo quando  $V = 0$ .

2) Si ( $n$ )  $q\Phi_M < q\Phi_S$ .

In questo caso si ha trasferimento di elettroni dal metallo al semiconduttore con conseguente formazione di uno strato di accumulazione nel Si. Come si vede l'altezza della barriera non è molto grande e di fatto non costituisce un impedimento al passaggio di elettroni nei due sensi.



Inoltre, in questa situazione la ddp applicata cade lungo la parte del Si lontana dall'interfaccia, in cui si comporta semplicemente come un resistore: non ci sono dunque gli effetti rettificanti legati alla presenza di una zds.

Come regola generale: zds  $\Rightarrow$  contatto rettificante, strato di accumulazione  $\Rightarrow$  contatto ohmico.

## Gli stati di superficie e i contatti ohmici

La caratteristica dello Schottky

$$I = I_0(e^{\frac{V}{V_T}} - 1)$$

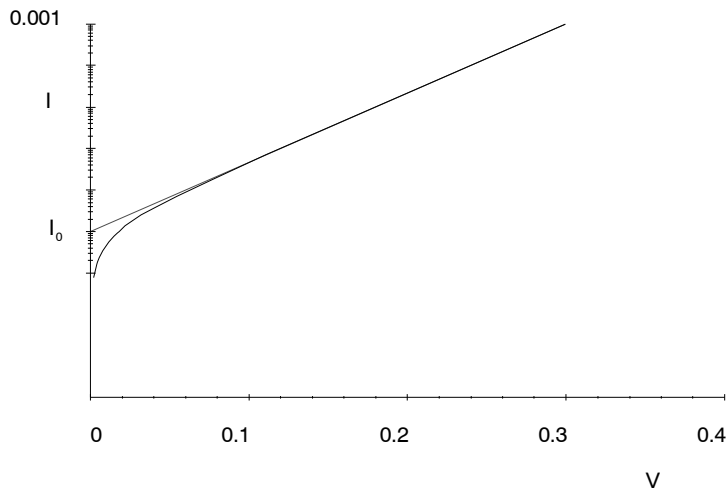


si riscrive più precisamente

$$I = I_0(e^{\frac{V}{nV_T}} - 1)$$

in cui  $n > 1$  prende il nome di *fattore di idealità* e rende ragione della struttura fine dell'interfaccia, che può molto difficilmente essere presa in considerazione da un modello di validità generale. In un buon diodo  $n$  è dell'ordine di  $1.02 \div 1.05$ .

La caratteristica  $I - V$ , rappresentata su scala semilog, può essere utilizzata per la misura di  $I_0$ . Estrapolando la validità del comportamento  $I = I_0 e^{\frac{V}{nV_T}}$  fino a  $V = 0$  si ottiene, come intercetta sull'asse delle correnti e per  $V = 0$ , il valore di  $I_0$ .



Poiché la sua espressione è nota

$$I_0 = SA^*T^2 e^{-\frac{q\Phi_{BN}}{kT}}$$

si ha immediatamente

$$\ln I_0 = \ln(SA^*T^2) - \frac{q\Phi_{BN}}{kT}$$

e quindi si può risalire al valore di  $q\Phi_{BN}$ .

Consideriamo due metalli diversi:

Al ( $q\Phi_M = 4.2$  eV) e Au ( $q\Phi_M = 5.22$  eV). Nel caso ideale

$$q\Phi_{BN} = q(\Phi_M - \chi)$$

$$q\Phi_{BN_{Al}} = 4.2 - 4.05 = 0.15 \text{ eV}$$

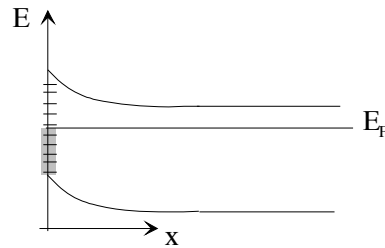
$$q\Phi_{BN_{Au}} = 5.22 - 4.05 = 1.17 \text{ eV}$$

mentre le misure forniscono

$$q\Phi_{BN_{Al}} = 0.68 \text{ eV}$$

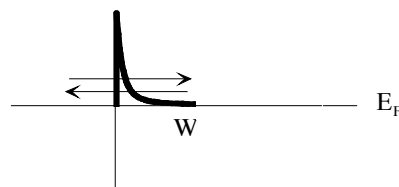
$$q\Phi_{BN_{Au}} = 0.79 \text{ eV.}$$

I valori sperimentali sono quasi indipendenti dalla funzione di lavoro del metallo. La spiegazione va cercata nella struttura energetica della superficie del Si. Per vari motivi (interruzione della periodicità del cristallo, contaminazione superficiale) sono presenti nel gap, alla superficie, degli stati localizzati che si possono comportare come accettori o donatori.



La densità  $D(E)$  ( $\text{cm}^{-2}\text{eV}^{-1}$ ) di questi stati, che è determinabile sperimentalmente, stabilisce, in assenza del contatto, lo stato di carica della superficie. Con riferimento alla figura, se gli stati sotto  $E_F$  sono accettori, saranno occupati da elettroni provenienti dalla BC, con conseguente piegamento delle bande verso l'alto. Questo strato localizzato di elettroni ha un effetto simile a quello di una superficie metallica e scherma quindi il Si dagli effetti della successiva deposizione del contatto.

Nel caso del Si un contatto metallo/Si( $n$ ) è *sempre rettificante*. Per renderlo ohmico si deve drogare pesantemente ( $n^+$ ) il Si nei pressi dell'interfaccia.



In questo caso la struttura a bande presenta una  $W$  molto ridotta, tale che la barriera evidenziata in figura, possa essere attraversata agevolmente nei due sensi per effetto tunnel. La resistenza specifica di contatto è dell'ordine di  $10^{-6} \div 10^{-7} \Omega \cdot \text{cm}^2$ , per un buon contatto ohmico.

Se per esempio un contatto di drain ha un'area di  $16 \mu\text{m}^2$  presenterà una  $R_c = \frac{10^{-7}}{16 \times 10^{-8}} = 0.625 \Omega$ . I contatti ohmici hanno un'importanza fondamentale nel caso di dispositivi submicrometrici.

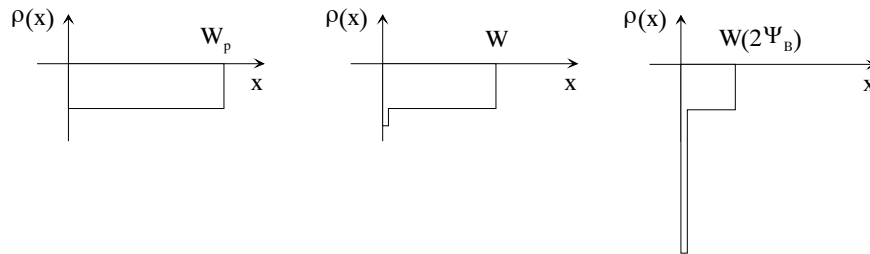
## 19) La struttura MOS in commutazione e i CCD

### La struttura MOS durante il transitorio

Nelle sezioni precedenti abbiamo studiato in dettaglio il comportamento della struttura MOS in condizioni stazionarie per ogni tensione  $V_{GS}$  applicata ai suoi capi. Supponiamo adesso che  $V_{GS}$  possa assumere solo due valori:  $V_0 = 0$  e  $V^0 \gg V_{TH}$ . A regime si avrà la condizione corrispondente a  $V_{FB}$  nel primo caso e a forte inversione nel secondo.

Esaminiamo il comportamento della struttura durante il transitorio, nell'ipotesi che la tensione  $V = V(t)$  abbia un andamento a gradino. All'istante  $t_{0+}$  la carica  $Q_G$  per unità di superficie sul gate dovrà essere compensata da una carica  $Q_S$ , uguale in modulo, nel semiconduttore.

Poiché la carica dello strato di inversione è dovuta a un processo di generazione dalla banda di valenza, che è mediato da livelli profondi nel gap del silicio, un intervallo di tempo  $\Delta t$  intercorre fra l'istante  $t_{0+}$  e l'istante in cui lo strato di inversione si è completamente formato. La figura mostra, in sequenza temporale e in modo schematico, l'evolversi della densità di carica nel semiconduttore.



Ci sarà dunque un intervallo di tempo  $\Delta t^* < \Delta t$  durante il quale la carica  $Q_G$  sul gate sarà compensata soltanto dagli accettori ionizzati: il semiconduttore si trova in una condizione di *svuotamento profondo*. Scriveremo per la tensione di gate  $V_{GS}$

$$V_{GS} = -\frac{Q_W(\Psi_S)}{C_{ox}} + \Psi_S = \frac{qN_A W_p}{C_{ox}} + \Psi_S = \frac{\sqrt{2qN_A \varepsilon_s \Psi_S}}{C_{ox}} + \Psi_S$$

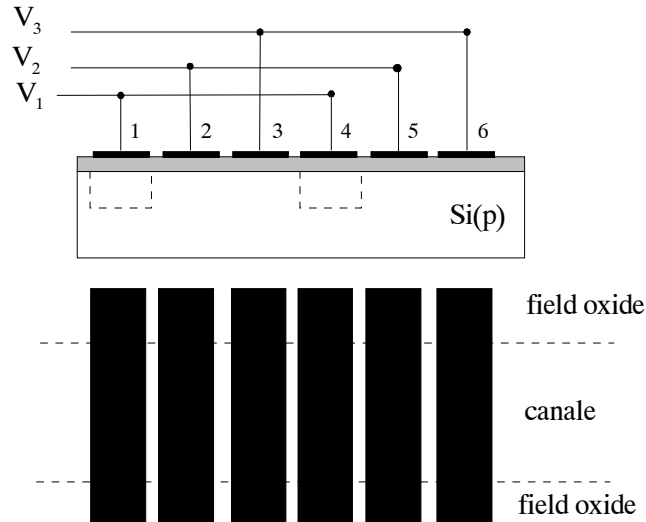
con  $\Psi_S > 2\Psi_B$ ;

$W_p$ , ampiezza della zona di svuotamento (profondo), è maggiore di  $W$ , ampiezza della zona di svuotamento a regime, in quanto  $qN_A W_p$  deve compensare da sola  $Q_G$ .

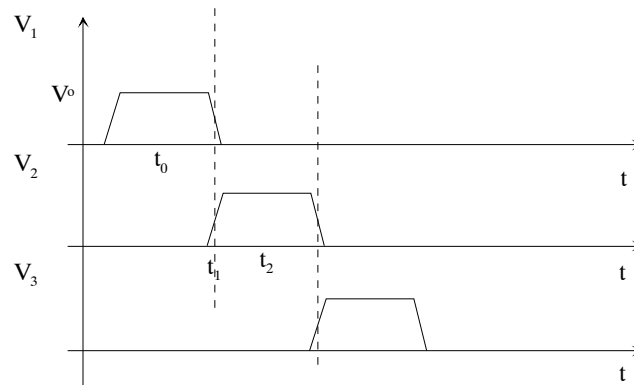
Durante l'intervallo  $\Delta t^*$  l'andamento della BC del silicio presso l'interfaccia costituisce una buca di potenziale nella quale possono essere iniettati elettroni dall'esterno, che costituiranno o un bit di informazione digitale o una parte di un segnale analogico. L'informazione contenuta nel pacchetto di elettroni si conserverà per tutto l'intervallo  $\Delta t^*$ ; poi verrà oscurata dalla generazione dello strato di inversione, a meno che non si provveda a trasferirla sotto un altro gate che si trovi nella condizione iniziale del transitorio.

## Principio di funzionamento dei CCD

Nella figura sono mostrati, in pianta e in sezione, sei gate che sono parte di una catena molto più lunga;

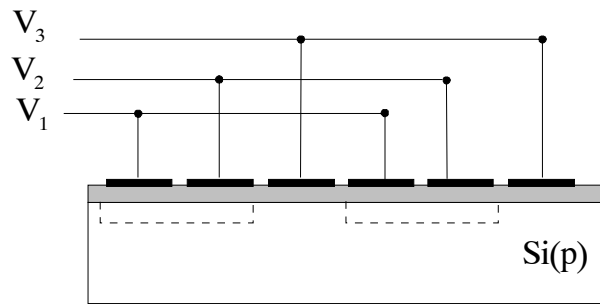


la distanza fra i gate, per un motivo che verrà chiarito fra poco, deve essere inferiore al  $\mu\text{m}$ . La linea tratteggiata sotto il gate indica la profondità della buca di potenziale all'interfaccia ( $q\Psi_s$ ) o l'estensione della zds. L'andamento delle tensioni  $V_1$ ,  $V_2$  e  $V_3$  in funzione del tempo è tale che ad un istante

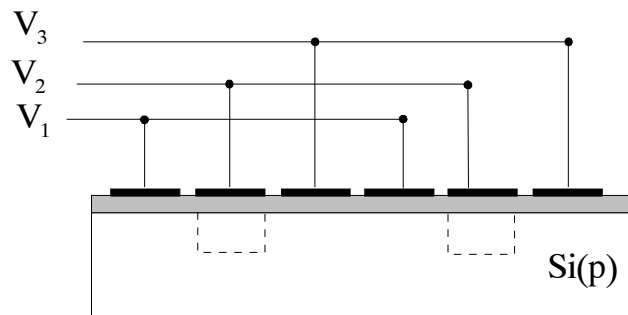


$t = t_0$ :  $V_1 = V_0$  ( $\gg V_{TH}$ ),  $V_2 = 0$ ,  $V_3 = 0$ , le informazioni si trovano sotto i gate 1 e 4;

-----  
 Si immagina il substrato illuminato dal retro. Verranno generate coppie elettrone-lacuna; gli elettroni, se la loro lunghezza di diffusione è sufficientemente elevata, raggiungeranno la zds sotto i gate 1 e 4 e verranno intrappolate all'interfaccia  $\text{Si/SiO}_2$ . Il numero di elettroni dipende dall'intensità luminosa locale e dal tempo di esposizione.  
 -----



all'istante  $t = t_1$ :  $V_1 = \frac{V^o}{2}$  e  $V_2 = \frac{V^o}{2}$ , le informazioni si distribuiscono sotto i gate 1, 2, 4 e 5; le coppie 1, 2 e 4, 5, in questa situazione, si comportano ciascuna come un unico gate;



$t = t_2$ :  $V_1 = 0$  e  $V_2 = V^o$ , le informazioni sono state trasferite sotto i gate 2 e 5.

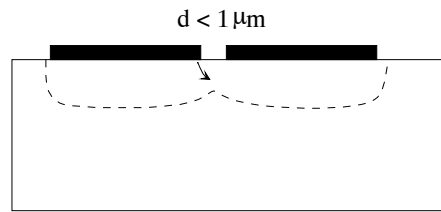
Risulta evidente che la funzione dei gate 3 e 6, durante la sequenza temporale esaminata, è quella di impedire che le informazioni, contenute inizialmente sotto i gate 1 e 4, si mescolino fra di loro. Hanno dunque la funzione, poiché la loro tensione rimane zero quando  $V_1$  e  $V_2$  sono diverse da zero, di elemento separatore fra due celle di memoria adiacenti. Quindi: una cella di memoria è costituita da tre gate contigui, se lo spessore dell'ossido è costante lungo tutta la struttura.

Se l'andamento nel tempo delle tensioni viene ripetuto per i gate 2, 3, 5 e 6, mentre  $V_1$  rimane al valore zero, le informazioni sono trasferite sotto i gate 3 e 6.

Le informazioni in ciascuna cella di memoria possono essere introdotte:

- 1) in modo seriale con un opportuno dispositivo in ingresso;
- 2) contemporaneamente in ciascuna cella mediante illuminazione di tutta l'area, se si fa operare una catena (o una matrice di gate) come un rivelatore di immagini.

La distanza fra gate adiacenti deve essere inferiore al  $\mu\text{m}$  se si vuole un effettivo accoppiamento elettrico fra le regioni sottostanti i gate, ad esempio quando i loro potenziali sono contemporaneamente al valore alto. La figura mostra che l'accoppiamento è dovuto agli effetti di bordo del campo elettrico (componenti del campo secondo  $y$ ).



Quindi un passo di processo assai critico nella costruzione di un CCD è la definizione di una spaziatura interelettrodica, costante con la lunghezza del gate, uguale o inferiore al  $\mu\text{m}$ . Si noti che un qualunque cortocircuito fra gate adiacenti dà luogo ad un dispositivo non funzionante.

## 20) Le curve C-V della struttura MOS

Polarizzando la struttura MOS ideale con una qualunque  $V_{GS}$  si stabiliscono le condizioni di carica del substrato, che si traducono nella conoscenza della funzione

$$Q_s = Q_s(\Psi_s)$$

dove  $\Psi_s$  è la caduta di tensione e  $Q_s$  è la carica per unità di superficie nel Si.

D'altra parte  $C_D$ , capacità della zona di carica spaziale nel Si per unità di superficie, è per definizione

$$C_D = \left| \frac{dQ_s}{d\Psi_s} \right|;$$

quindi derivando  $Q_s = Q_s(\Psi_s)$  rispetto a  $\Psi_s$  si ottiene la  $C_D$  in funzione di  $\Psi_s$ .

Differenziando la relazione

$$V_{GS} = -\frac{Q_s}{C_{ox}} + \Psi_s$$

e considerando il valore assoluto

$$dV_{GS} = \frac{dQ_s}{C_{ox}} + d\Psi_s$$

questa può essere scritta

$$\begin{aligned} dV_{GS} &= \frac{dQ_s}{C_{ox}} + \frac{dQ_s}{C_D} \\ \frac{dV_{GS}}{dQ_s} &= \frac{1}{C_{ox}} + \frac{1}{C_D} \end{aligned}$$

da cui

$$\frac{1}{C_{TOT}} = \frac{1}{C_{ox}} + \frac{1}{C_D}$$

dove  $C_{TOT}$  è la capacità differenziale di tutta la struttura MOS, definita da

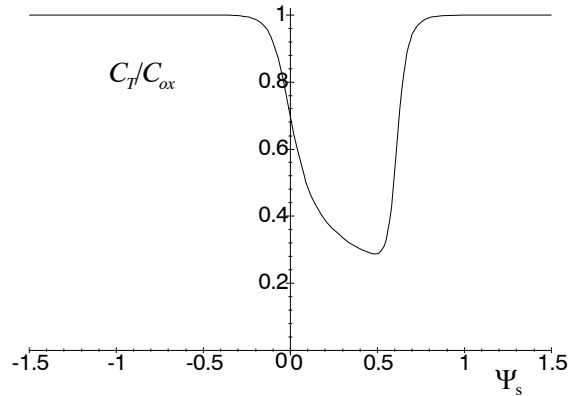
$$C_{TOT} = \frac{dQ_s}{dV_{GS}} = \frac{dQ_G}{dV_{GS}}.$$

La capacità per unità di superficie dell'ossido (costante rispetto a  $V_{GS}$ ) e quella del Si sono in serie.

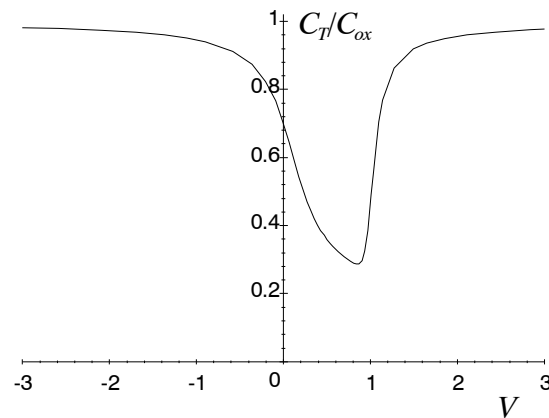
L'andamento di  $C_{TOT} = C_{TOT}(\Psi_s)$  si ottiene immediatamente dalla

$$C_{TOT}(\Psi_s) = \left( \frac{1}{C_{ox}} + \frac{1}{C_D(\Psi_s)} \right)^{-1};$$

il grafico relativo è quello della figura che segue.



Se poi si tiene in considerazione la relazione che lega  $\Psi_s$  a  $V_{GS}$  si ottiene l'andamento cercato.



Si deve a questo punto notare che quando si deriva la  $Q_s$  rispetto a  $\Psi_s$  tutti i termini di cui questa è costituita ( $Q_n$  e  $Q_W$ ) subiscono una variazione, indipendentemente dal fatto che questa variazione possa avvenire o meno.

Lo strato di inversione nella struttura MOS è costituito da elettroni che possono essere forniti soltanto dai processi di generazione termica, la cui costante di tempo caratteristica è il tempo di vita medio  $\tau$ .

Le quantità da prendere in considerazione sono il periodo  $T$  del segnale di misura e  $\tau$ . Si hanno due casi limite:

1)  $T \gg \tau$  : durante un semiperiodo, nel quale  $V_{GS}$  e quindi  $\Psi_s$  aumentano o diminuiscono, gli elettroni dello strato di inversione seguono il segnale (bassa frequenza del segnale di misura);

2)  $T \ll \tau$  : lo strato di inversione non può seguire le variazioni di  $\Psi_s$  (alta frequenza).

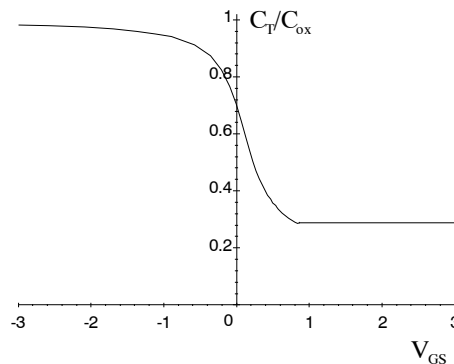
La curva in *bassa frequenza*, riportata nella figura precedente, si interpreta come segue:

$V_{GS} < 0$  : c'è un'alta concentrazione di portatori mobili (lacune) all'interfaccia. Il MOS si comporta come un condensatore con due armature metalliche: il Si è infatti in condizioni di accumulazione  $\Rightarrow C_{TOT} = C_{ox}$ .

$V_{GS} < V_{TH}$  : siamo in condizioni di svuotamento, le variazioni di  $Q_W$  sono dovute alla modulazione di  $W = W(\Psi_s)$ ; la capacità del Si è la  $C_W$  che è in serie alla  $C_{ox}$ .

$V_{GS} > V_{TH}$  : in inversione le variazioni della  $Q_n$  sono molto maggiori della  $Q_W$  (una variazione del tipo  $e^{\beta\Psi_s}$  a confronto con una di tipo  $\sqrt{\Psi_s}$ ). Ci troviamo in un caso simile all'accumulazione con i segni invertiti dei due strati di carica mobile. Per le curve in *alta frequenza* bisogna considerare che  $Q_n$  non segue il segnale, quindi l'unico contributo alla capacità del Si viene dalla zona di svuotamento, il cui spessore da  $V_{TH}$  in poi rimane pressoché costante ( $\Psi_s$  bloccato a  $2\Psi_B$ ). Per  $V_{GS} > V_{TH}$  quindi  $C_{TOT}$  rimane costante e pari a

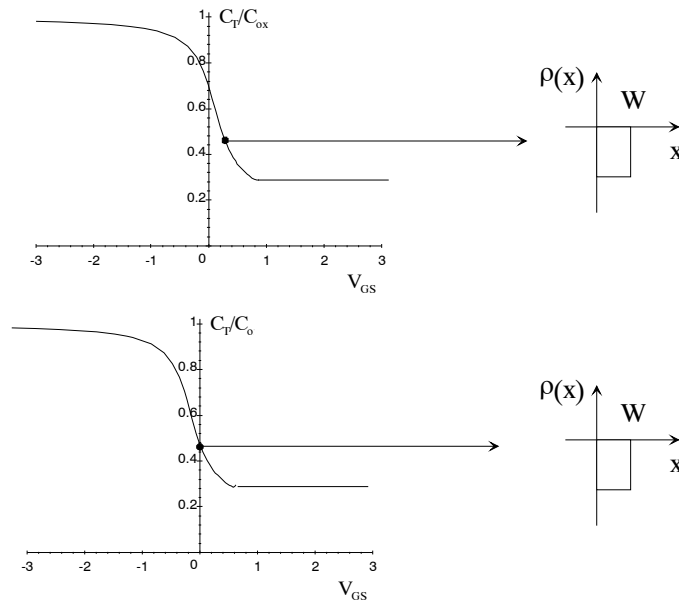
$$C_{TOT} = \left( \frac{1}{C_{ox}} + \frac{W(2\Psi_B)}{\varepsilon_s} \right)^{-1}.$$



I due valori limite della  $C_{TOT}$  si indicano generalmente con  $C_{MAX}$  e  $C_{MIN}$ .

Come dovrebbe essere evidente da quanto esposto al valore di  $C_{TOT}$  contribuisce la  $C_{ox}$ , che è costante, e la  $C_D$ , che dipende invece dallo stato di carica del Si (piegamento delle bande), indipendentemente da cosa sia stato causato questo piegamento. Quindi se una certa condizione, come quella della figura, è dovuta alla  $V_{GS}$  applicata, oppure ha origine da una ddp interna, dovuta ad esempio alla  $q\Phi_{MS}$ , il valore della  $C_{TOT}$ , che è una capacità differenziale, non cambia. E' cambiato soltanto il valore di  $V_{GS}$  per il quale si ha questo valore.





Nella prima figura  $0 < V_{GS} < V_{TH}$ , nella seconda  $V_{GS} = 0$ . Le curve  $C - V$  ideali traslano rigidamente lungo l'asse  $x$  di una quantità pari alla  $V_{FB}$ .

### 21) Fenomeni dipendenti dal tempo nella giunzione pn

La caratteristica  $I - V$  DC della giunzione è stata ricavata dall'equazione di continuità indipendente dal tempo. Quella dipendente dal tempo sarà utilizzata per descrivere il comportamento del diodo quando corrente e tensione variano in funzione di  $t$ .

$$\frac{\partial \delta p(x, t)}{\partial t} = -\frac{1}{q} \frac{\partial J_h(x, t)}{\partial x} - \frac{\delta p(x, t)}{\tau}$$

L'equazione precedente non è di facile soluzione, per cui è necessario ridurla ad una forma più conveniente; moltiplichiamola per  $qS$  e integriamo in  $x$  fra 0 e  $\infty$ , ad un generico istante  $t$ .

$$qS \frac{\partial}{\partial t} \int_0^\infty \delta p(x, t) dx = -S \int_0^\infty \partial J_h(x, t) - \frac{qS}{\tau} \int_0^\infty \delta p(x, t) dx$$

Il primo membro è  $\frac{dQ(t)}{dt}$ , essendo  $Q(t)$  la carica totale in eccesso immagazzinata nella zona quasi-neutra all'istante  $t$ . Il primo termine del secondo membro

$$-S [J_h(x, t)]_0^\infty = i(t)$$

rappresenta la corrente della giunzione dato che  $J_h(\infty, t) = 0$ ; avremo

$$i(t) = \frac{dQ(t)}{dt} + \frac{Q(t)}{\tau}$$

Questa equazione verrà utilizzata tutte le volte che è necessario trattare i dispositivi in condizioni di dipendenza temporale.

### Il circuito equivalente del diodo $pn$ per piccoli segnali

Consideriamo, come al solito, la giunzione di riferimento. A regime e in polarizzazione diretta  $Q$  è data da

$$Q = qS \int_0^{\infty} \delta p(x) dx$$

in cui

$$\delta p(x) = \delta p(0) e^{-\frac{x}{L_h}}.$$

In una diversa condizione di regime

$$\delta p^*(x) = \delta p^*(0) e^{-\frac{x}{L_h}}$$

cui corrisponde una  $Q^*$ . Nella figura sono mostrati due diversi profili, corrispondenti a due tensioni diverse. La differenza fra le due curve è l'eccesso di carica che deve essere aggiunto (o eliminato) nel passare dall'una all'altra condizione e viceversa. La scala delle concentrazioni è arbitraria.

Se si cambia la tensione del diodo è evidente che non si può passare istantaneamente da  $Q$  a  $Q^*$ , perché i minoritari iniettati (o estratti) devono avere il tempo di diffondere e/o di ricombinarsi.  $Q$  d'altra parte è nota in funzione di  $t$  solo se è noto  $\delta p(x, t)$ . Assumeremo come valida la seguente approssimazione: *anche in regime dipendente dal tempo il profilo dei portatori in eccesso è ad ogni istante un profilo di regime, ovvero*

$$\delta p(x, t) = \delta p(0, t) e^{-\frac{x}{L_h}}.$$

Se supponiamo che al diodo, oltre alla continua, sia applicata anche una tensione variabile periodicamente nel tempo, l'hp precedente significa che il periodo del segnale è molto maggiore del tempo caratteristico di diffusione (tempo di vita medio).

La tensione ai capi del diodo sia

$$v(t) = V_1 + v_m \sin \omega t$$

con  $v_m \ll V_1$ . L'equazione

$$i(t) = \frac{dQ(t)}{dt} + \frac{Q(t)}{\tau}$$

se a  $Q(t)$  si sostituisce la sua espressione, determinata dall'ipotesi precedente, consente di esplicitare tutti i termini di cui è formata la corrente  $i(t)$ .

$$\begin{aligned} i(t) &= qS \frac{d}{dt} \int_0^\infty \delta p(x, t) dx + \frac{qS}{\tau} \int_0^\infty \delta p(x, t) dx \\ i(t) &= qS \frac{d}{dt} \int_0^\infty \delta p(0, t) e^{-\frac{x}{L_h}} dx + \frac{qS}{\tau} \int_0^\infty \delta p(0, t) e^{-\frac{x}{L_h}} dx \\ i(t) &= qSL_h \frac{d}{dt} \delta p(0, t) + \frac{qSL_h}{\tau} \delta p(0, t). \end{aligned}$$

La relazione che collega l'eccesso dei portatori in 0 e la tensione applicata ( $V_1 + v(t)$ ) consente di scrivere

$$i(t) = qSL_h \frac{d}{dt} p_{n0} \left( e^{\frac{V_1 + v(t)}{V_T}} - 1 \right) + \frac{qSL_h}{\tau} p_{n0} \left( e^{\frac{V_1 + v(t)}{V_T}} - 1 \right)$$

o anche, se  $V_1 \gg V_T$ ,

$$i(t) = qSL_h \frac{d}{dt} p_{n0} \left( e^{\frac{V_1}{V_T}} e^{\frac{v(t)}{V_T}} \right) + \frac{qSL_h}{\tau} p_{n0} \left( e^{\frac{V_1}{V_T}} e^{\frac{v(t)}{V_T}} \right).$$

Con l'ulteriore hp che  $v(t) < V_T$  si può sviluppare in serie  $e^{\frac{v(t)}{V_T}}$

$$i(t) = qSL_h \frac{d}{dt} p_{n0} \left( e^{\frac{V_1}{V_T}} \left( 1 + \frac{v(t)}{V_T} \right) \right) + \frac{qSL_h}{\tau} p_{n0} \left( e^{\frac{V_1}{V_T}} \left( 1 + \frac{v(t)}{V_T} \right) \right).$$

Il secondo termine del secondo membro è costituito da

$$I_1 = \frac{qSL_h}{\tau} p_{n0} e^{\frac{V_1}{V_T}}$$

che rappresenta la componente continua della corrente e da

$$i_R(t) = \frac{qSL_h p_{n0} e^{\frac{V_1}{V_T}}}{\tau V_T} v(t) = \frac{v(t)}{r_d}$$

che è invece la parte ohmica e variabile della corrente.  $r_d$  è la resistenza differenziale, che si può calcolare linearizzando la caratteristica in un intorno del punto di lavoro fissato da  $V_1$ .

$$r_d = \frac{V_T}{I_1}$$

Infatti la costante, come si riconosce moltiplicando e dividendo per  $L_h$

$$\frac{qSL_h p_{n0} e^{\frac{V_1}{V_T}} L_h}{\tau} \frac{L_h}{L_h} = I_0 e^{\frac{V_1}{V_T}}$$

rappresenta la componente DC della corrente

$$i(t) = qSL_h \frac{d}{dt} p_{n0} \left( e^{\frac{V_1}{V_T}} \left( 1 + \frac{v(t)}{V_T} \right) \right) + \frac{v(t)}{r_d} + I_1$$

Il primo termine del secondo membro

$$\frac{qSL_h p_{n0} e^{\frac{V_1}{V_T}}}{V_T} \frac{dv(t)}{dt}$$

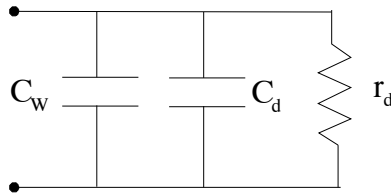
può essere scritto

$$C_d \frac{dv(t)}{dt} = \frac{qSL_h p_{n0} e^{\frac{V_1}{V_T}}}{V_T} \frac{dv(t)}{dt}$$

e rappresenta la corrente che scorre nella capacità di diffusione  $C_d$

$$C_d = \frac{qSL_h p_{n0} e^{\frac{V_1}{V_T}}}{V_T} = \frac{I_1 \tau}{V_T} = \frac{\tau}{r_d}$$

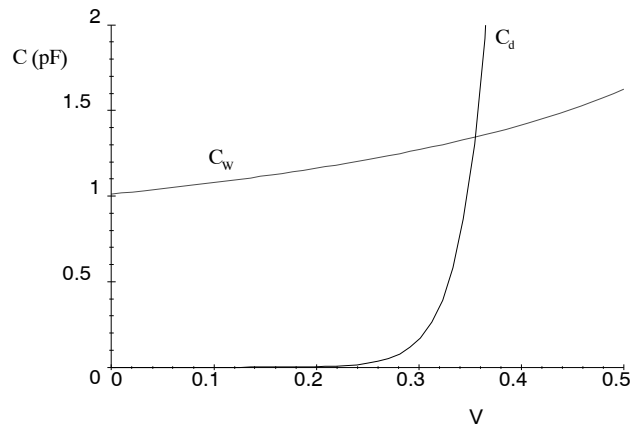
Il nome ha origine dal fatto che alla zona quasi-neutra deve essere aggiunta o sottratta una quantità  $dQ$  di carica in eccesso e questa carica deve diffondere affinché si formi il nuovo profilo. Il circuito equivalente completo della giunzione  $pn$  per piccoli segnali è mostrato in figura.



Si noti la presenza anche della capacità  $C_W$  della zds.

Confrontiamo adesso  $C_d$  e  $C_W$  per la giunzione di riferimento quando  $V > 0$ .

- 1) sezione  $S$ :  $100 \times 100 \mu\text{m}^2 = 10^{-4} \text{ cm}^2$ ;
- 2)  $\tau = 10^{-5} \text{ s}$ ;
- 3)  $V_0 = 0.026 \times \ln \left( \frac{10^{15} \times 10^{19}}{2.25 \times 10^{20}} \right) = 0.817 \text{ V}$ ;
- 4)  $I_0 = 4.1 \times 10^{-15}$



Come si vede per bassi valori di  $V$  prevale  $C_W$ , mentre per valori di  $V > 0.38$  V prevale nettamente la  $C_d$  e quindi  $C_W$  può essere trascurata.

Nel diodo Schottky polarizzato direttamente l'unica capacità di cui tener conto è la  $C_W$ , dato che non si hanno fenomeni di immagazzinamento di carica in eccesso.

### La giunzione $pn$ in commutazione

Se il diodo è pilotato da un generatore di corrente che commuta, all'istante 0, da  $I$  a  $I + \Delta I$ , è evidente che la tensione non può seguire la corrente. Anche in questo caso si suppone che il profilo dei minoritari sia ad ogni istante un profilo di regime. Confronteremo alla fine il risultato ottenuto utilizzando questa approssimazione con quello che deriva da una soluzione esatta. Si tratta ora, noto l'andamento della  $i(t)$ , di trovare come evolve nel tempo la  $Q(t)$ , che è legata alla tensione  $v(t)$  dalla relazione (nell'ipotesi fatta)

$$Q(t) = qSL_n \delta p(0, t) = qSL_n p_{n0} \left( e^{\frac{v(t)}{V_T}} - 1 \right).$$

L'equazione di continuità

$$i(t) = \frac{dQ(t)}{dt} + \frac{Q(t)}{\tau}$$

ha come soluzione

$$Q(t) = A + B e^{-\frac{t}{\tau}}$$

con le condizioni al contorno

$$t = 0^-, Q = I\tau$$

$$t = \infty, Q = (I + \Delta I)\tau$$

$$A + B = I\tau$$

$$A = (I + \Delta I)\tau$$

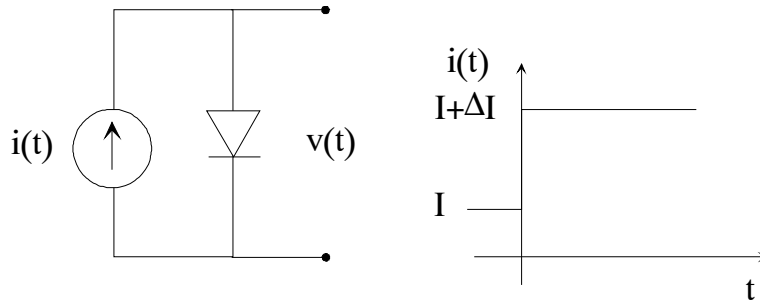
e infine

$$Q(t) = (I + \Delta I)\tau - \Delta I\tau e^{-\frac{t}{\tau}} = I\tau + \Delta I\tau \left(1 - e^{-\frac{t}{\tau}}\right).$$

Avremo dunque

$$qSL_h p_{n0} \left( e^{\frac{v(t)}{V_T}} - 1 \right) = I\tau + \Delta I\tau \left(1 - e^{-\frac{t}{\tau}}\right)$$

da cui si può ottenere l'andamento di  $v(t)$ .

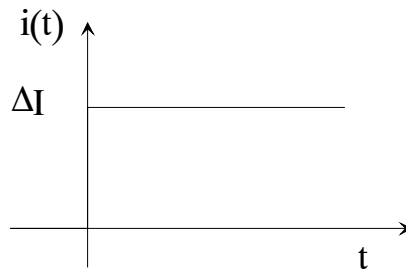


$$\frac{qSL_h p_{n0}}{\tau} \left( e^{\frac{v(t)}{V_T}} - 1 \right) = I + \Delta I \left(1 - e^{-\frac{t}{\tau}}\right)$$

$$I_0 \left( e^{\frac{v(t)}{V_T}} - 1 \right) = I + \Delta I \left(1 - e^{-\frac{t}{\tau}}\right)$$

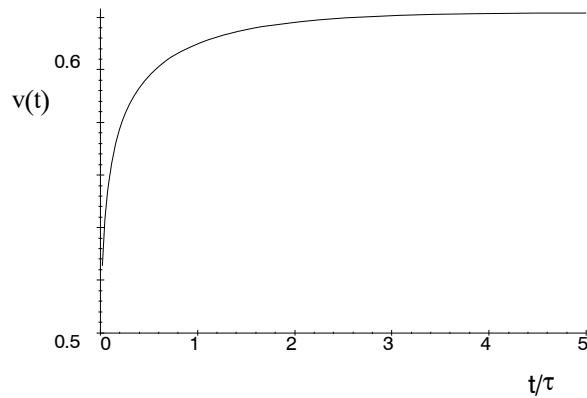
$$v(t) = V_T \ln \left( \frac{I}{I_0} + \frac{\Delta I}{I_0} \left(1 - e^{-\frac{t}{\tau}}\right) + 1 \right)$$

Se  $I = 0$  e  $\Delta I = 0.1 \text{ mA}$  si ha , per la giunzione di riferimento



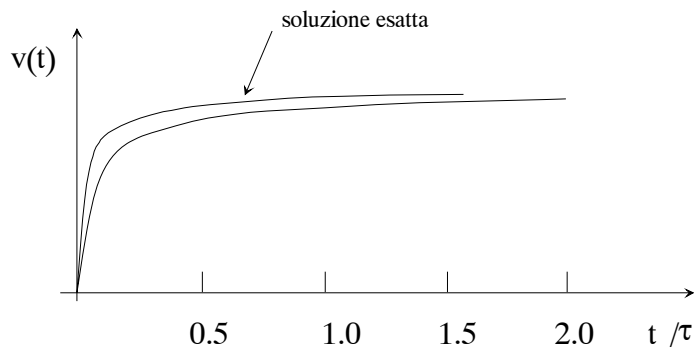
$$v(t) = V_T \ln \left( \frac{\Delta I}{I_0} \left(1 - e^{-\frac{t}{\tau}}\right) + 1 \right)$$

$$v(t) = 0.026 \times \ln \left( \frac{10^{-4}}{4.1 \times 10^{-15}} \left(1 - e^{-\frac{t}{10^{-5}}}\right) + 1 \right)$$



Come si vede, dopo tre costanti di tempo il transitorio è praticamente finito.

Si riporta per confronto nel grafico che segue anche la soluzione non approssimata.

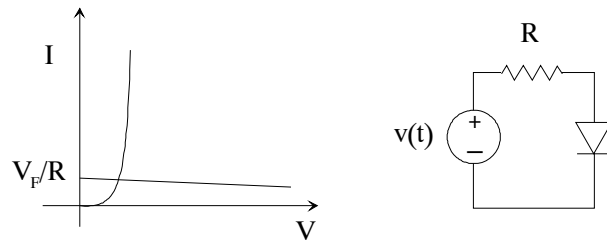


Resta da spiegare come la corrente nel diodo possa commutare da  $I$  a  $I + \Delta I$ , dato che istantaneamente non può essere apportata alcuna carica aggiuntiva alla zona quasi-neutra; infatti

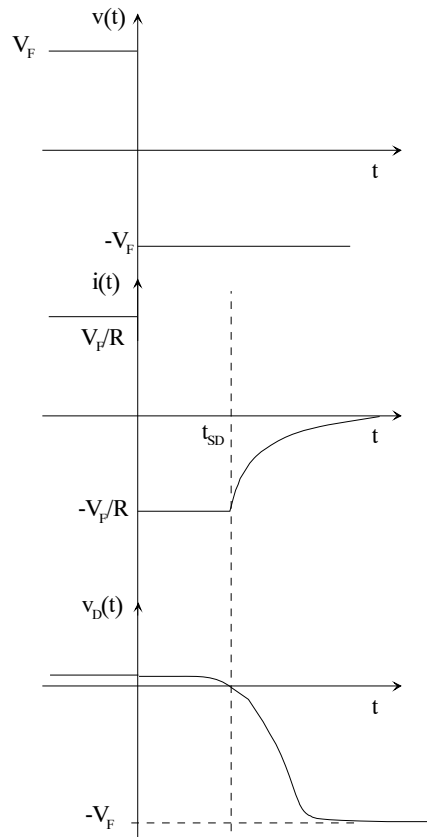
$$\Delta Q = \int_{0^-}^{0^+} \Delta I dt = 0.$$

Ricordando che la corrente del diodo è in  $x = 0$  una corrente di diffusione e dipende dalla derivata in 0 del profilo, la variazione di corrente imposta dal generatore può avere luogo perché istantaneamente cambia la derivata del profilo in zero, in modo che la corrente possa assumere a  $t_{0+}$  il valore  $I + \Delta I$ .

Quando si cerca di far commutare il diodo dalla polarizzazione diretta a quella inversa intervengono a ritardare il raggiungimento della nuova condizione i fenomeni di immagazzinamento (*storage*) dell'eccesso di carica, che come vedremo, sono importanti anche nella commutazione del BJT.

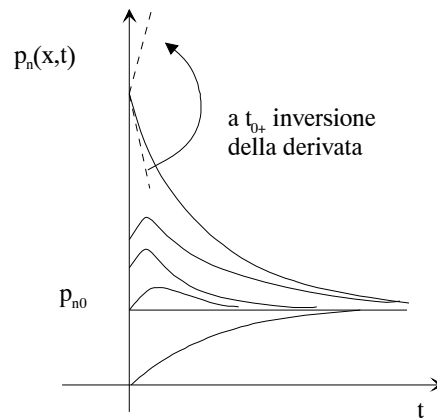


Con riferimento alla figura nel diodo scorre, prima dell'istante di commutazione, una corrente  $I_F \simeq \frac{V_F}{R}$ ; il profilo dei minoritari è quello solito.



Quando il generatore si inverte la tensione del diodo istantaneamente non può cambiare a causa dell'inerzia del profilo di eccesso. D'altra parte la corrente dovrà essere  $I_R = -\frac{V_F}{R}$ , per cui, come nel caso precedente, cambia istantaneamente la derivata del profilo per adattare il profilo alla situazione  $I = -I_R$ . Questa si mantiene fino a che la tensione ai capi del diodo non è uguale a zero: da questo istante in poi la tensione tende a  $-V_F$  e la corrente a  $-I_0$ .





Se si osserva che durante l'intervallo  $(0, t_{SD})$  il diodo vede un generatore di corrente, il valore di  $t_{SD}$  si può valutare mediante l'espressione, ricavata nel caso di pilotaggio da parte di un generatore di corrente

$$v(t) = V_T \ln \left( \frac{I}{I_0} + \frac{\Delta I}{I_0} (1 - e^{-\frac{t}{\tau}}) + 1 \right)$$

imponendo che per  $t = t_{SD}$  la tensione ai capi del diodo si sia ridotta a zero

$$0 = V_T \ln \left( \frac{I}{I_0} - \frac{2I}{I_0} (1 - e^{-\frac{t_{SD}}{\tau}}) + 1 \right);$$

segue

$$1 = \frac{I}{I_0} - \frac{2I}{I_0} (1 - e^{-\frac{t_{SD}}{\tau}}) + 1$$

e con qualche passaggio si ottiene

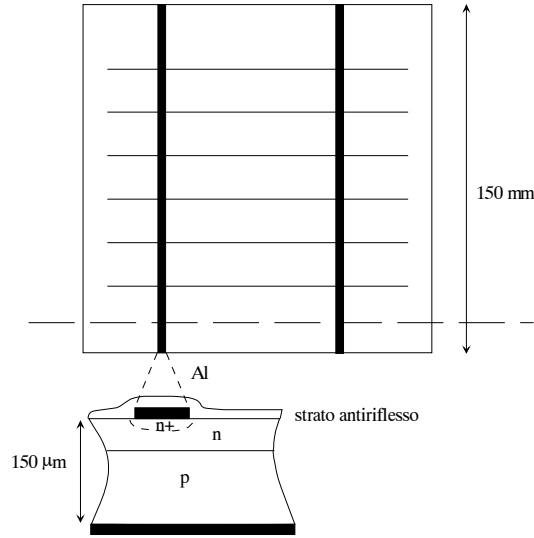
$$t_{SD} = \tau \ln 2.$$

Come si vede  $\tau$  è una caratteristica molto importante del Si utilizzato (ricordiamo che  $\tau$  può essere fissato ad un valore stabilito "drogando" con Au la zona in cui vengono realizzati i dispositivi).

## 22) La cella solare

Un'applicazione di una certa importanza della giunzione  $pn$  è la conversione fotovoltaica dell'energia solare.

La struttura di una singola cella è la seguente (sezione e pianta).



La parte frontale è quella che riceve l'illuminazione. La geometria dei contatti è tale da consentire la maggior esposizione possibile alla luce.

Per ricavare la caratteristica  $I - V$  della giunzione in condizioni di illuminazione, si supponga che la parte  $p$  sia  $p^+$  e che la zona  $n$  sia illuminata uniformemente in modo che in tutto il volume ci sia una generazione ottica  $G_{OP}$ .

L'equazione di continuità delle lacune iniettate nella zona  $n$  in queste condizioni si scrive

$$D_h \frac{d^2 \delta p(x)}{dx^2} = \frac{\delta p(x)}{\tau} - G_{OP}$$

$$D_h \frac{d^2 \delta p(x)}{dx^2} = \frac{\delta p(x) - G_{OP}\tau}{\tau};$$

definendo  $\delta^* p(x) = \delta p(x) - G_{OP}\tau$  si ha

$$D_h \frac{d^2 \delta^* p(x)}{dx^2} = \frac{\delta^* p(x)}{\tau}$$

che ha per soluzione

$$\delta^* p(x) = \delta^* p(0) e^{-\frac{x}{L_h}}$$

$$\delta p(x) - G_{OP}\tau = (\delta p(0) - G_{OP}\tau) e^{-\frac{x}{L_h}}$$

$$\delta p(x) = \delta p(0) e^{-\frac{x}{L_h}} + G_{OP}\tau \left(1 - e^{-\frac{x}{L_h}}\right).$$

La concentrazione dei minoritari nella zona  $n$  è, in condizioni di illuminazione,  $n_{p0} + G_{OP}\tau \simeq G_{OP}\tau$ .

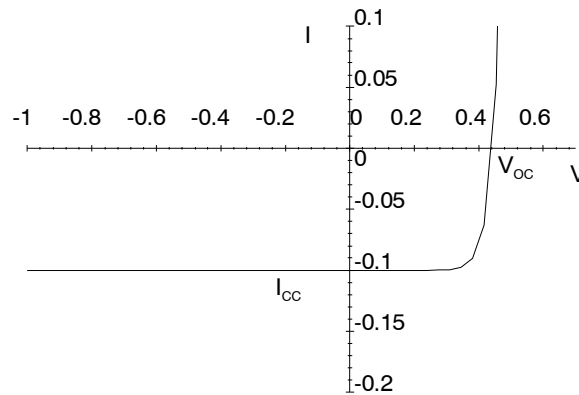
La corrente si trova come al solito utilizzando la

$$I = -qSD_h \frac{d}{dx} \delta p(x) \text{ in } x = 0$$

$$I = -qSD_h \frac{d}{dx} \left( \delta p(0) e^{-\frac{x}{L_h}} + G_{OP\tau} \left( 1 - e^{-\frac{x}{L_h}} \right) \right)$$

$$I = \frac{qSD_h p_{n0}}{L_h} \left( e^{\frac{V}{V_T}} - 1 \right) - \frac{qSD_h G_{OP\tau}}{L_h} = I_0 \left( e^{\frac{V}{V_T}} - 1 \right) - I_{CC}$$

La caratteristica della figura non è che la caratteristica  $I - V$  del diodo traslata verso il basso di una quantità  $-I_{CC}$ .



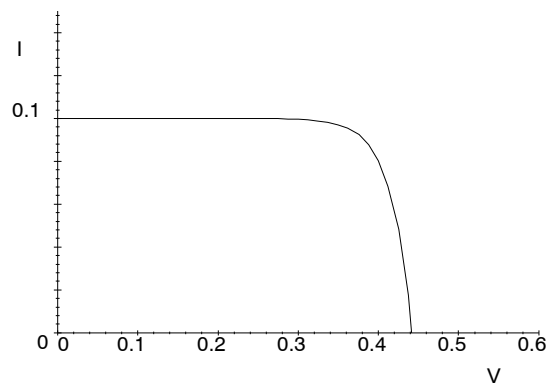
Per  $I = 0$  si trova una tensione  $V_{OC}$  di circuito aperto data da

$$0 = I_0 \left( e^{\frac{V_{OC}}{V_T}} - 1 \right) - I_{CC}$$

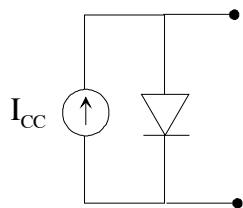
$$V_{OC} = V_T \ln \left( \frac{I_{CC}}{I_0} + 1 \right)$$

e per  $V = 0$  una corrente di corto circuito pari a  $-I_{CC}$ . La corrente  $I_{CC}$  è costituita dai portatori fotogenerati che vengono raccolti dal campo della zds.

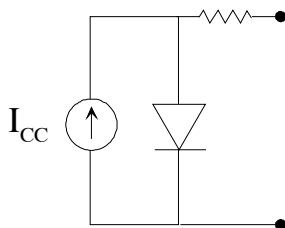
La cella, per  $I < 0$  e  $V > 0$  si comporta come un generatore la cui caratteristica viene solitamente rappresentata come segue



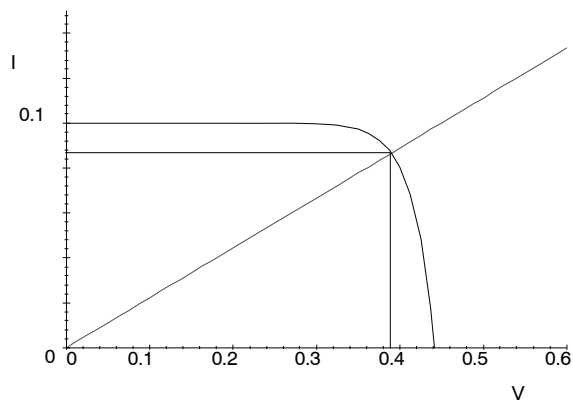
Il circuito equivalente DC semplificato è un generatore di corrente in parallelo con la giunzione.



L'elemento parassita più importante di cui tener conto è la resistenza serie complessiva, il cui contributo principale proviene dalla metal del contatto superiore.



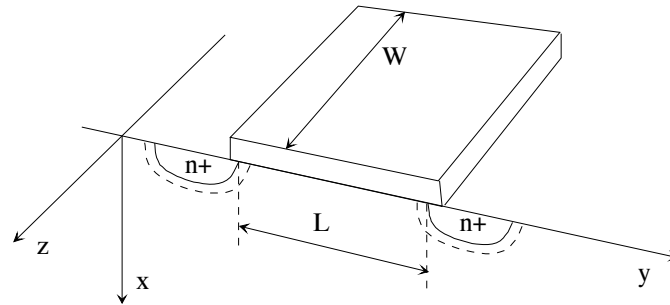
Se si chiude il generatore su un carico resistivo si ottiene il punto di lavoro;



L'area del rettangolo rappresenta la potenza effettiva erogata al carico. Come è facilmente intuibile esiste un valore del carico che rende massima la potenza utilizzata.

L'efficienza massima teorica di una cella al Si è dell'ordine del 30%. Celle commerciali hanno rendimenti che vanno dal 15 al 18 %.

### 23) Il transistore NMOS



Le ipotesi alla base del modello che verrà utilizzato per ricavare le caratteristiche  $I_{DS} = f(V_{DS}; V_{GS})$  sono:

- 1) la struttura MOS è ideale;
- 2) la mobilità degli elettroni del canale (strato di inversione) è costante (indipendente da  $V_{DS}$  e  $V_{GS}$ );

3) il source è in corto con il substrato;

4)  $\mathcal{E}_x \gg \mathcal{E}_y$ , dove  $\mathcal{E}_y$  è la componente del campo dovuta alla  $V_{DS}$ . In base a questa ipotesi, che è quasi vera nel caso di canale lungo, ma non lo è nel caso di canale corto, si possono utilizzare nella trattazione i risultati ottenuti per la struttura MOS, nel corso della quale è stato utilizzato il teorema di Gauss per ottenere la  $Q_S$ . In quel caso  $\mathcal{E}_y = 0$ ; in questo caso è trascurabile.

Canale lungo:  $L \gg$  dell'ampiezza delle zone di svuotamento  $n^+/\text{sub}$ .

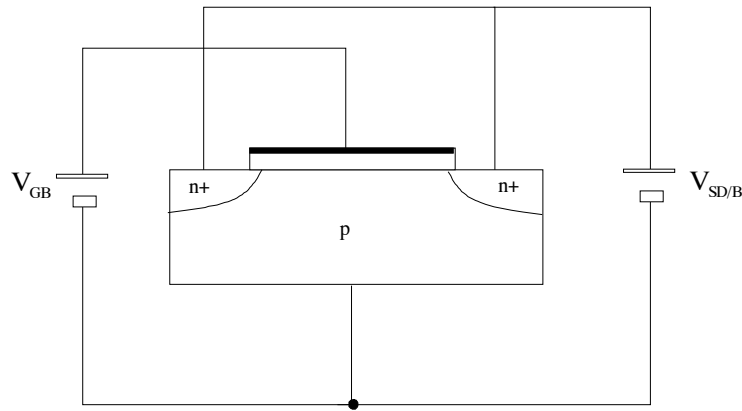
Canale corto:  $L$  comparabile con l'ampiezza delle zone di svuotamento  $n^+/\text{sub}$ .

5) La corrente nel canale è solo una corrente di trascinamento.

Conviene a questo punto prendere in considerazione la

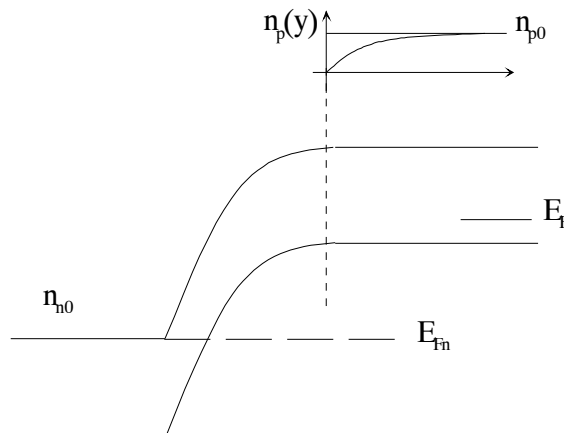
### Struttura MOS a tre terminali

definita dalla figura.



Questa si riduce alla struttura MOS già trattata se  $V_{SD/B} = 0$ . Il terminale di substrato verrà indicato d'ora in poi con B, per non confonderlo con il source S.

Supponiamo  $V_{GB} = 0$ . Se  $V_{SD/B} > 0$  la giunzione  $n^+/p$  è polarizzata inversamente. La sua struttura a bande secondo  $y$  in  $x = 0$  è quella seguente (ne viene mostrata solo una metà);

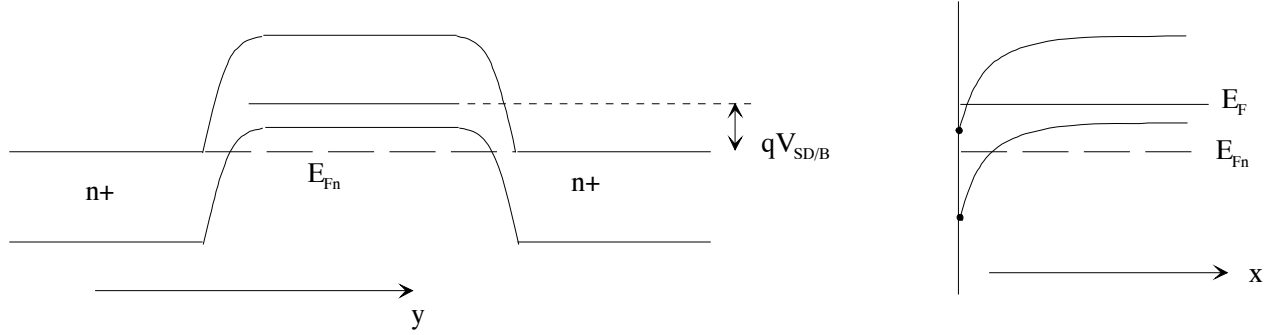


non siamo all'equilibrio e quindi il livello di Fermi si separa nei due quasi-livelli  $E_{Fp}$  e  $E_{Fn}$ . A noi qui interessa l'andamento di  $E_{Fn}$ , che determina la concentrazione degli elettroni. Il suo andamento è quello della linea tratteggiata. A destra della zds, dato che siamo in inversa, la concentrazione di elettroni è sicuramente minore di  $n_{p0}$ :  $E_{Fn}$  deve essere quindi più in basso di  $E_F$ , perché  $E_F$  rende ragione di una concentrazione  $n_{p0}$ , che è maggiore, nel volume, di quella a destra della zds entro almeno una lunghezza di diffusione. In  $x = 0$  e in presenza di una  $V_{SD/B} > 0$  è a  $E_{Fn}$  che bisogna riferirsi per stabilire quale sia la concentrazione di elettroni.

La figura che segue rappresenta la struttura a bande secondo  $y$  in  $x = 0$  (all'interfaccia), prima e dopo l'applicazione fra gate e substrato della tensione di soglia  $V_{THi}$ . In questo caso la  $V_{SD/B} = 0$  e il livello di Fermi è costante.

Le bande del substrato sono state piegate in modo da ottenere l'inversione all'interfaccia, come dovrebbe essere evidente dalla seconda parte della figura che è l'andamento secondo  $x$  nella sezione rappresentata dalla linea tratteggiata.

Se invece  $V_{SD/B} > 0$  la figura assume l'aspetto seguente:



Per ottenere l'inversione le bande all'interfaccia (indicate con i due punti in evidenza) devono assumere adesso rispetto ad  $E_{Fn}$  la posizione che prima dell'applicazione della  $V_{SD/B} > 0$  avevano rispetto ad  $E_F$ . In altre parole la caduta nel Si deve essere  $2\Psi_B + V_{SD/B}$  invece di  $2\Psi_B$ . La tensione di soglia sarà allora

$$V_{THi} = \frac{\sqrt{2qN_A\epsilon_s(2\Psi_B + V_{SD/B})}}{C_{ox}} + 2\Psi_B + V_{SD/B}$$

Se adesso immaginiamo di scollegare il S dal D e di applicare una  $V_{DS} > 0$ , nel canale si avrà nel punto  $(x = 0, y)$  una tensione  $V(y)$  per cui per una generica  $V_{GS}$  si scriverà

$$\begin{aligned} V_{GS} &= -\frac{Q_n(y) + Q_W(2\Psi_B + V(y))}{C_{ox}} + 2\Psi_B + V(y) \\ &= -\frac{Q_n(y)}{C_{ox}} - \frac{Q_W(2\Psi_B + V(y))}{C_{ox}} + 2\Psi_B + V(y) \\ &= -\frac{Q_n(y)}{C_{ox}} + V_{THi}(y) \end{aligned}$$

avendo messo in evidenza che la tensione di soglia locale, in presenza della  $V_{DS}$ , è funzione di  $y$ . La caduta nel Si è infatti bloccata in  $(0, y)$ , per  $V_{GS} > V_{THi}$ , al valore  $2\Psi_B + V(y)$ .

### Le caratteristiche $I_{DS} = f(V_{DS}; V_{GS})$

Dalla

$$V_{GS} = -\frac{Q_n(y) + Q_W(2\Psi_B + V(y))}{C_{ox}} + 2\Psi_B + V(y)$$

si può ottenere l'espressione della  $Q_n(y)$ , carica per unità di superficie relativa allo strato di inversione

$$Q_n(y) = -[C_{ox}V_{GS} + Q_W(2\Psi_B + V(y)) - 2\Psi_B C_{ox} - V(y)C_{ox}]$$

e il suo valore assoluto

$$|Q_n(y)| = [C_{ox}V_{GS} + Q_W(2\Psi_B + V(y)) - 2\Psi_B C_{ox} - V(y)C_{ox}].$$

Un elemento  $dR$  di resistenza del canale si scrive

$$dR = \bar{\rho}(y) \frac{dy}{Wx_i}$$

in cui  $\bar{\rho}(y)$  è la resistività media del canale in  $y$  definita da

$$\frac{1}{\bar{\rho}(y)} = \overline{\sigma(y)} = \frac{q\mu_n}{x_i} \int_0^{x_i} n(x, y) dx = \mu_n \frac{Q_n(y)}{x_i};$$

$x_i$  rappresenta la coordinata per la quale il substrato diviene nuovamente intrinseco.  $dR$  si può quindi esprimere in funzione di  $Q_n(y)$  e precisamente

$$dR = \frac{1}{\mu_n \frac{Q_n(y)}{x_i}} \frac{dy}{Wx_i} = \frac{1}{\mu_n W} \frac{dy}{Q_n(y)}.$$

In condizioni di stazionarietà la corrente  $I_{DS}$  provoca su  $dR$  una caduta  $dV$

$$I_{DS} dR = dV$$

$$\frac{I_{DS}}{\mu_n W} \frac{dy}{Q_n(y)} = dV$$

$$I_{DS} dy = \mu_n W Q_n(y) dV$$

che può essere integrata in  $y$  fra 0 ed  $L$  e in  $V$  fra 0 e  $V_{DS}$ .

$$I_{DS} = \frac{\mu_n W}{L} \int_0^{V_{DS}} Q_n(y) dV$$

$$I_{DS} = \frac{\mu_n W}{L} \int_0^{V_{DS}} [C_{ox}V_{GS} + Q_W(2\Psi_B + V(y)) - 2\Psi_B C_{ox} - V(y)C_{ox}] dV$$



$$I_{DS} = \frac{\mu_n C_{ox} W}{L} \int_0^{V_{DS}} \left[ V_{GS} + \frac{Q_W(2\Psi_B + V(y))}{C_{ox}} - 2\Psi_B - V(y) \right] dV$$

Posto  $k = \frac{\mu_n C_{ox} W}{L}$  e esplicitando l'espressione di  $Q_W(2\Psi_B + V(y))$  si ottiene

$$I_{DS} = k \left( V_{GS} V_{DS} - \frac{2\sqrt{2qN_A\epsilon_s}}{3C_{ox}} \left[ (2\Psi_B + V_{DS})^{\frac{3}{2}} - (2\Psi_B)^{\frac{3}{2}} \right] - 2\Psi_B V_{DS} - \frac{V_{DS}^2}{2} \right) \quad (A1)$$

che costituisce l'espressione delle caratteristiche relativa al modello con canale graduale.

Il fattore  $k = \frac{\mu_n C_{ox} W}{L}$  è costituito da due termini:

a) il fattore geometrico  $W/L$ , che non è nient'altro se non il rapporto fra le dimensioni del canale;

b) il parametro di processo  $\mu_n C_{ox}$ , che contiene lo spessore dell'ossido e la mobilità degli elettroni nel canale.

Il valore di  $\mu_n$  dipende da come è stata eseguita l'ossidazione di gate e dai trattamenti di annealing ed è inferiore alla corrispondente mobilità del bulk. Il suo valore è infatti compreso fra 750 e 850  $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ .

Per trasformare l'espressione in una più facilmente usabile manualmente si può trascurare la dipendenza di  $Q_W$  da  $V$ . Ponendo infatti  $Q_W \simeq \sqrt{2qN_A\epsilon_s 2\Psi_B}$

$$I_{DS} = k \int_0^{V_{DS}} \left[ V_{GS} - \frac{\sqrt{2qN_A\epsilon_s 2\Psi_B}}{C_{ox}} - 2\Psi_B - V(y) \right] dV$$

integrando

$$I_{DS} = k \left( V_{GS} V_{DS} - \frac{\sqrt{2qN_A\epsilon_s 2\Psi_B}}{C_{ox}} V_{DS} - 2\Psi_B V_{DS} - \frac{V_{DS}^2}{2} \right)$$

e raccogliendo i termini con  $V_{DS}$

$$I_{DS} = k \left( V_{DS} \left( V_{GS} - \frac{\sqrt{2qN_A\epsilon_s 2\Psi_B}}{C_{ox}} - 2\Psi_B \right) - \frac{V_{DS}^2}{2} \right)$$

$$I_{DS} = k \left( (V_{GS} - V_{THi}) V_{DS} - \frac{V_{DS}^2}{2} \right) \quad (A0)$$

in cui  $V_{THi}$  è la tensione di soglia della struttura MOS 1D, che è stato possibile esplicitare nelle caratteristiche solo con l'approssimazione  $Q_W \simeq \sqrt{2qN_A\epsilon_s 2\Psi_B}$ .

Esempio

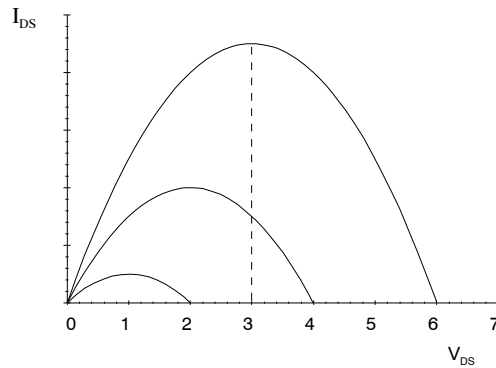
$$t_{ox} = 50 \text{ nm}, N_A = 5 \times 10^{15} \text{ cm}^{-3}, W/L = 4, \mu_n = 800 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}.$$

$$C_{ox} = \frac{3.9 \times 8.85 \times 10^{-12}}{50 \times 10^{-9}} = 6.9 \times 10^{-4} \text{ F/m}^2;$$

$$k = 6.9 \times 10^{-4} \times 800 \times 10^{-4} \times 4 = 2.21 \times 10^{-4} \simeq 2.0 \times 10^{-4} \text{ A/V}^2$$

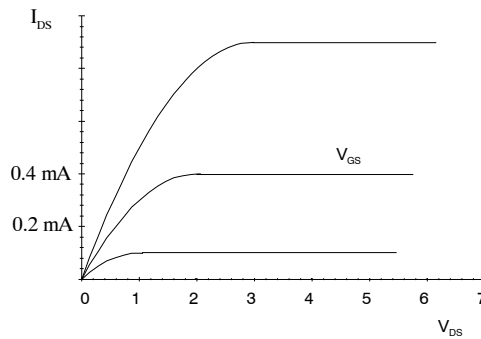
$$V_{THi} \simeq 1 \text{ V}.$$

La corrente  $I_{DS}$  raggiunge un massimo per  $V_{DSAT} = V_{GS} - V_{THi}$



e poi, secondo il modello, decresce fino ad annullarsi e ad assumere valori negativi. Nella figura i valori di  $V_{GS}$  sono 2, 3, e 4 V. Sperimentalmente si verifica invece che da  $V_{DSAT}$  in poi la corrente si mantiene costante al suo valore massimo  $I_{DSAT}$ .  $I_{DSAT}$  ha l'espressione

$$I_{DSAT} = \frac{k}{2} (V_{GS} - V_{THi})^2.$$



In effetti il modello presuppone una caduta  $dV$  sull'elemento di resistenza  $dR$  e quindi descrive un resistore variabile con la  $V_{DS}$  (a  $V_{GS}$  costante).

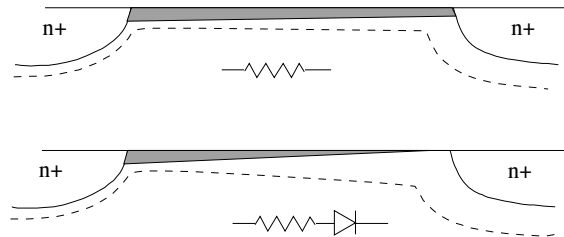
Consideriamo l'espressione della carica mobile per unità di superficie in  $L$

$$Q_n(L) = [C_{ox} V_{GS} + Q_W(2\Psi_B) - 2\Psi_B C_{ox} - V_{DS} C_{ox}];$$

per  $V_{DS} = V_{GS} - V_{THi}$

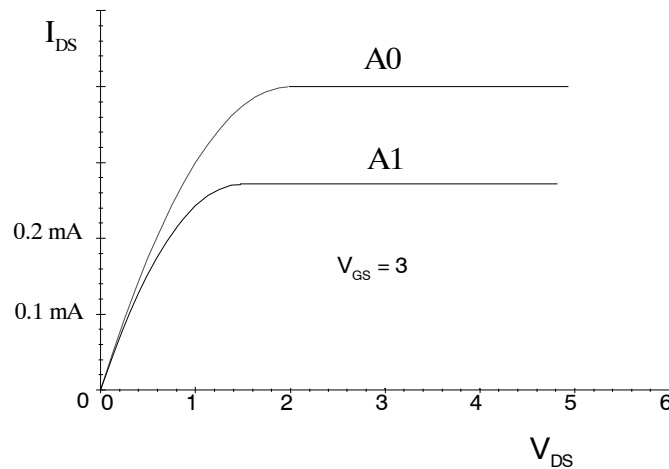
$$Q_n(L) = C_{ox} \left[ V_{GS} + \underbrace{\frac{Q_W(2\Psi_B)}{C_{ox}} - 2\Psi_B}_{-V_{THi}} - (V_{GS} - V_{THi}) \right]$$

$Q_n(L)$  è nulla (il termine indicato dalla graffa è  $-V_{THi}$ ): il canale si è chiuso (strozzamento o *pinch-off*). Nella figura è schematizzata la situazione, fissata una certa  $V_{GS}$ , per  $V_{DS} < V_{DSAT}$  e per  $V_{DS} > V_{DSAT}$ .



Nel primo caso lo strato di inversione connette il S e il D, mentre nel secondo il D è scollegato. Dopo il pinch-off si ha il canale e un diodo, la giunzione sub/D, polarizzato inversamente, che il modello non può descrivere. Il motivo della corrente di saturazione sta nel fatto che la giunzione sub/D ha una corrente inversa di saturazione costituita dai minoritari, che in questo caso sono gli elettroni che arrivano nel punto di strozzamento attraverso il canale.

Per completezza si riportano insieme le caratteristiche, con  $V_{GS} = 3$ , ottenute con le espressioni A0 e A1.



Come si vede l'espressione approssimata sovrastima il valore della corrente e della tensione di saturazione.

### La misura della $V_{TH}$

Osservando l'espressione A1 si vede non è possibile definire una  $V_{TH}$  come per la A0. Si ricordi anche che il modello è comunque approssimato e certamente non valido per dispositivi a canale corto. E' in generale opportuno poter misurare la tensione di soglia.

Per valori di  $V_{DS}$  molto piccoli sia la A0 che la A1 si riducono alla relazione lineare

$$I_{DS} = k (V_{GS} - V_{TH}) V_{DS}$$

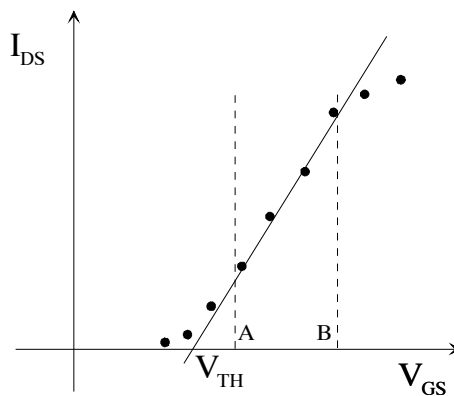
Fissiamo dunque un valore molto piccolo di  $V_{DS} = V_{DS0}$  (ad es. 100 mV) e misuriamo  $I_{DS}$  in funzione di  $V_{GS}$ . Si otterranno dei dati sperimentali come quelli mostrati nella figura.

Si può osservare che:

- 1)  $I_{DS}$  è diversa da zero anche per valori di  $V_{GS} < V_{TH}$ ;
- 2)  $I_{DS}$  si discosta dalla linearità, prevista dalla relazione precedente, per valori di  $V_{GS}$  elevati.

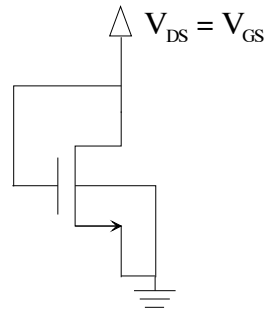
Per ciò che riguarda il punto 1), notiamo che l'esistenza dello strato di inversione non è determinata in modo on-off dalla  $V_{TH}$ . Per valori di  $V_{GS}$  di poco inferiori alla  $V_{TH}$  l'interfaccia Si/SiO<sub>2</sub> si trova in uno stato di debole inversione (ricordiamo che la condizione di inversione è definita dall'uguaglianza  $n_s = p_{po}$ ); ne consegue una  $I_{DS}$  sottosoglia non nulla e non sempre trascurabile. Si noti ancora che in condizioni di debole inversione la concentrazione dei portatori liberi è piccola; non è più possibile quindi trascurare la componente diffusiva della corrente.

Interpolando linearmente i dati nell'intervallo (A, B) si ottiene, per  $I_{DS} = 0$  un valore sperimentale di  $V_{TH}$ . Questa è la tensione di soglia in zona lineare.



Inoltre, la pendenza della retta è data da  $k = \frac{\mu_n C_{ox} W}{L}$  da cui, noti  $W$ ,  $L$  e  $C_{ox}$  si ricava  $\mu_n$ .

Un altro modo per misurare  $V_{TH}$  consiste nel collegare l'NMOS in modo che il gate sia in corto con il drain;

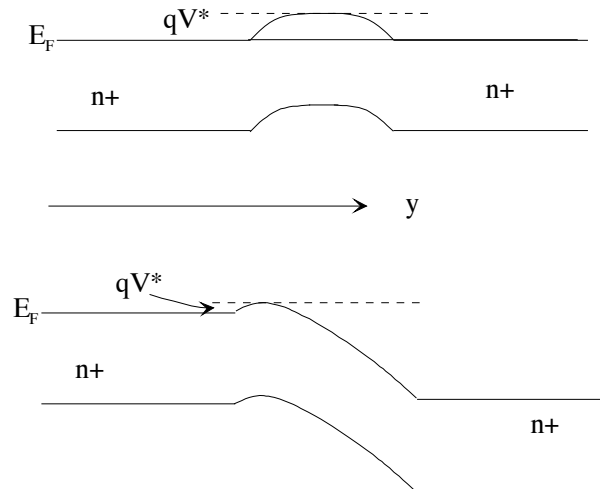


in questo caso, dato che l'NMOS è in saturazione quando  $V_{DS} \geq V_{GS} - V_{TH}$ , la corrente è data da

$$I_{DSAT} = \frac{k}{2} (V_{GS} - V_{TH})^2$$

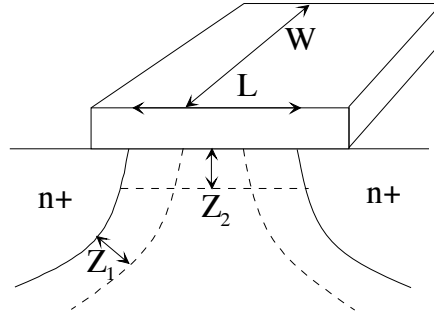
e si annulla per  $V_{GS} = V_{DS} = V_{TH}$ . Riportando in grafico  $\sqrt{I_{DS}}$  in funzione di  $V_{GS}$  e interpolando come nel caso precedente, si ottiene un valore di  $V_{TH}$  che è in generale diverso, soprattutto per i dispositivi a canale corto, da quello determinato in zona lineare. Si parla in questo caso di tensione di soglia in zona saturazione.

$V_{THsat}$  è minore di  $V_{THlin}$  e questo fatto può essere spiegato se si adotta un diverso punto di vista per rendere ragione del significato della tensione di soglia.



Nella parte superiore della figura è rappresentata la struttura a bande in  $x = 0$  quando  $V_{DS}$  è molto piccola: la BC è a una distanza tale da  $E_F$  che la barriera  $qV^*$  vista dagli elettroni del S è facilmente superabile (il D e il S sono in comunicazione). Per valori elevati di  $V_{DS}$  (oltre la saturazione), la  $z_{ds}$  del diodo D/sub polarizzato inversamente si estende per gran parte del canale e la barriera risulta ridotta. Questo modo di considerare le cose è noto come *Drain Induced Barrier Lowering* (DIBL).

Un ulteriore punto di vista è quello rappresentato dai cosiddetti *charge sharing models*.



Nel caso di canale corto, illustrato in figura, dato che per arrivare all'inversione è necessario prima indurre una carica fissa  $Q_W$  ( $C/cm^2$ ), alla soglia questa carica esiste già, almeno parzialmente, ed è dovuta alle zds circostanti il S e il D. Un possibile modello geometrico si ottiene calcolando la  $Q_{WTOT}$  che dovrebbe essere indotta sotto il gate e attribuendone in parte l'origine alle giunzioni di source e drain.

$$Q_{WTOT} = qN_A W L Z_2 - qN_A W (m Z_2 Z_1)$$

in cui si è supposto che i due parallelogrammi siano dei rettangoli e che  $m$  sia la frazione della carica sotto il gate dovuta alle zds. Poiché parte della  $V_{TH}$  è la  $Q_W$  ( $C/cm^2$ ) sarà

$$Q_W = \frac{Q_{WTOT}}{WL} = -\frac{qN_A W L Z_2 - qN_A W (m Z_2 Z_1)}{WL} = -\left( qN_A Z_2 - \frac{qN_A (m Z_2 Z_1)}{L} \right) :$$

se  $Z_1$  è molto minore di  $L$  ci si riduce al solito contributo, altrimenti la tensione di soglia si scrive

$$V_{TH} = \left( \frac{qN_A Z_2}{C_{ox}} - \frac{qN_A (m Z_2 Z_1)}{L C_{ox}} \right) + 2\Psi_B + V_{FB}$$

che diminuisce all'aumentare del rapporto  $Z_1/L$ .

## Modulazione della lunghezza di canale

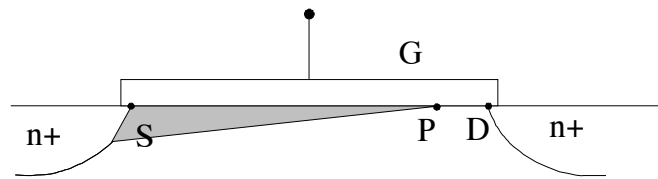
Un altro effetto di canale corto è l'accorciamento della lunghezza di canale  $L$  che si verifica per tensioni  $V_{DS} > V_{DSAT}$ . Ricordiamo che il canale si chiude in un punto P definito da  $Q_n(L) = 0$ . In P il canale è alla soglia ovvero

$$V_G - V_P = V_{GS} - V_{PS} = V_{TH}$$

da cui

$$V_{GS} - V_{TH} = V_{PS} = V_{DSAT} :$$

ai capi del canale, definito dalla distanza fra P ed S, cade la tensione  $V_{DSAT}$ ;

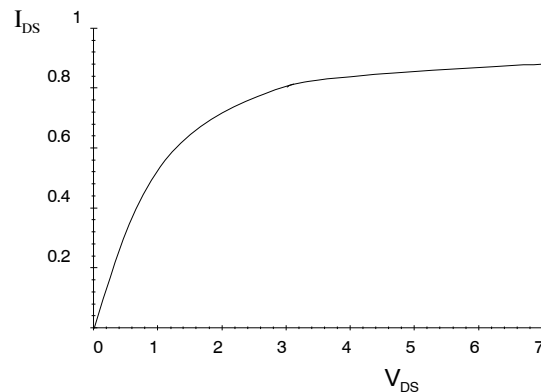


la differenza  $V_{DS} - V_{DSAT}$  è localizzata sulla giunzione drain/P. Ora è chiaro come non sia possibile separare gli effetti combinati del gate e del drain su questa  $zds$ , essendo la situazione evidentemente 2D. Si può tuttavia comprendere come il punto P arretri al crescere della differenza  $V_{DS} - V_{DSAT}$ , separando le due cause e considerando solo quello che accade alla giunzione drain/P. La lunghezza effettiva di canale  $L_{eff}$  alla saturazione può essere ad esempio espressa come

$$L_{eff} = L - \sqrt{\frac{2\varepsilon_s V_0}{qN_A}}$$

Quello che qui interessa è far vedere qualitativamente come si comporta  $I_{DS}$  per  $V_{DS} > V_{DSAT}$ .

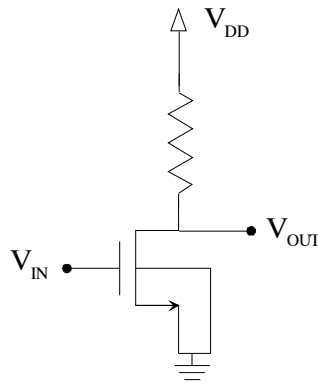
$$I_{DSAT}(V_{DS}) = I_{DSAT0} \frac{L_{eff}}{L - \sqrt{\frac{2\varepsilon_s (V_0 + (V_{DS} - V_{DSAT}))}{qN_A}}}$$



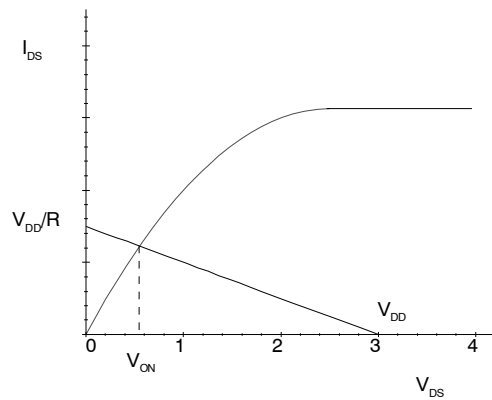
La corrente è in unità arbitrarie,  $V_{DSAT} = 3$  V.

## Introduzione alle porte logiche a NMOS

Nel circuito della figura la  $V_{GS} = V_{IN}$  può assumere due valori: 0 e  $V_{DD}$ . I rispettivi punti di lavoro (tensione  $V_{DS} = V_{OUT}$  ai capi dell'NMOS) si trovano



all'intersezione della retta di carico con le due caratteristiche con  $V_{GS} = 0$  (corrente nulla) e  $V_{GS} = V_{DD}$ .  $V_{GS}$  rappresenta la tensione di ingresso  $V_{IN}$ .



L'NMOS può essere considerato come un interruttore comandato dalla  $V_{GS}$ .

Quando  $V_{GS} = V_{DD}$  la  $V_{OUT}$  assume il valore  $V_{ON}$ , che è tanto più basso quanto più alto il valore di  $R$ . L'NMOS lavora in zona lineare e quindi può essere definita una  $R_{ON}$  ovvero

$$R_{ON} = \frac{V_{DS}}{I_{DS}} = \frac{1}{k(V_{GS} - V_{TH})}$$

Se ad esempio  $k = 2.0 \times 10^{-4} \text{ A/V}^2$   $R_{ON} = \frac{1}{2.0 \times 10^{-4} \times (3 - 0.5)} = 2 \text{ k}\Omega$ , avendo supposto  $V_{DD} = 3 \text{ V}$  e  $V_{TH} = 0.5 \text{ V}$ . La tensione bassa di uscita vale

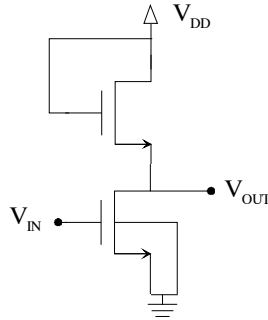
$$V_{ON} = \frac{V_{DD} R_{ON}}{R_{ON} + R}$$

e, se supponiamo che debba essere  $0.25 \text{ V}$ , si ha  $R = 22 \text{ k}\Omega$ . In un processo LOCOS  $R$  potrebbe essere realizzata in poly  $n^+$ . Supponendo per il poly una  $R_{\square}$  dell'ordine di  $50 \text{ }\Omega/\square$ ,  $R$  dovrebbe essere costituita da 440 quadrati e quindi occuperebbe un'area troppo grande, se il circuito deve essere integrato, rispetto a quella occupata dall'NMOS.



La  $R_{\square}$  dipende ovviamente dallo spessore, dal drogaggio, dalla dimensione dei grani.

La soluzione consiste nell'utilizzare come carico un altro NMOS, come mostrato dalla figura.

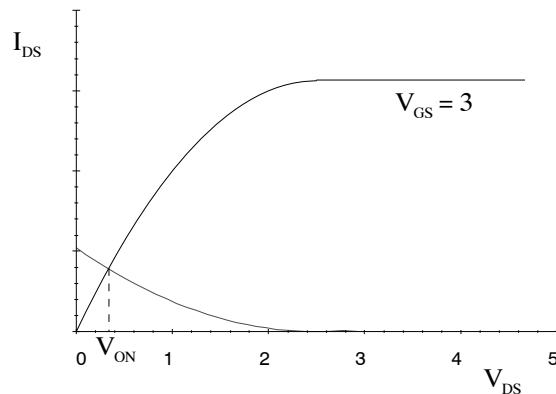


In questo caso il carico è un dispositivo non lineare, la cui caratteristica come abbiamo già visto, è data da

$$I_{DS2} = \frac{k}{2} (V_{DS2} - V_{TH})^2$$

Per trovare il punto di lavoro si osserva che  $I_{DS1} = I_{DS2}$  e che  $V_{DD} = V_{DS1} + V_{DS2}$ . Si possono dunque riportare sullo stesso grafico le caratteristiche di entrambi, pur di scrivere per  $I_{DS2}$

$$I_{DS2} = \frac{k}{2} (V_{DD} - V_{DS1} - V_{TH})^2.$$

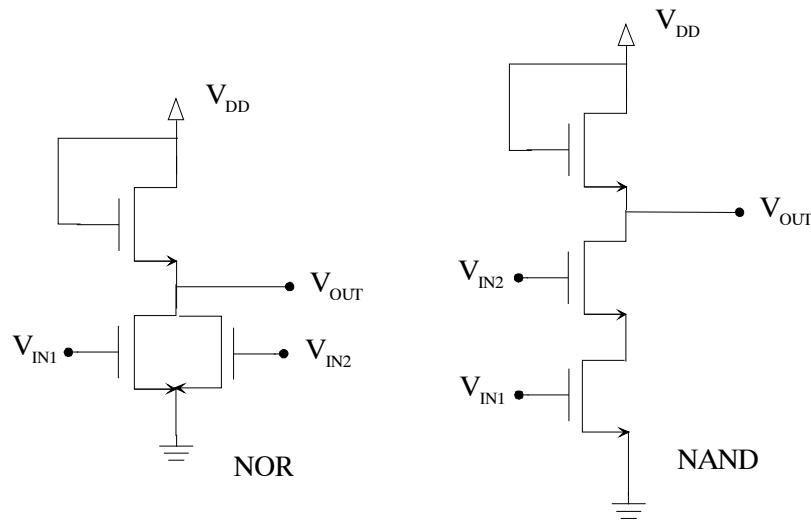


Invece della retta di carico si ha adesso la caratteristica dell'NMOS in saturazione.

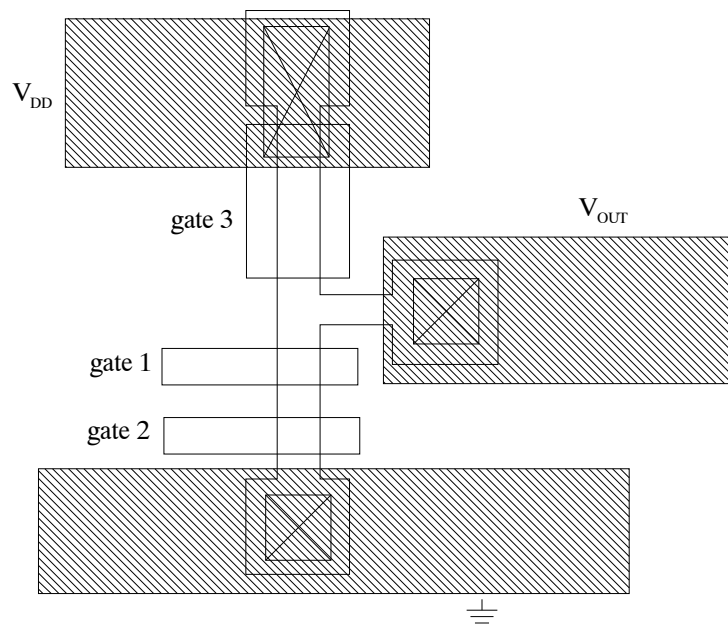
I dispositivi che hanno dato luogo alla figura hanno le larghezze di canale nel rapporto  $W_1/W_2 = 4$ ; è evidente che il carico occupa un'area paragonabile a quella del transistor di ingresso.

Il circuito costituisce un inverter. Infatti con  $V_{IN} = 0$  l'uscita è a  $V_{DD} - V_{TH}$ , mentre con l'ingresso a  $V_{DD}$  l'uscita è a  $V_{ON}$ . Se poniamo in serie un altro inverter uguale, questo vedrà come tensione bassa di ingresso la  $V_{ON}$ , e non più zero, come abbiamo supposto all'inizio. Di qui la necessità che la  $V_{ON}$  abbia un valore che sia almeno la metà della  $V_{TH}$ . Dal punto di vista della porta logica  $V_{ON}$  rappresenta zero e  $V_{DD} - V_{TH}$  uno.

Nella figura sono mostrate le porte NOR e NAND con due ingressi. Si osservi che per la NOR quando sia  $V_{IN1}$  che  $V_{IN2}$  sono 1, i due NMOS, entrambi accesi, equivalgono ad un solo NMOS con larghezza di canale doppia; per la NAND la stessa considerazione vale per la lunghezza  $L$ . Non è mostrato il terminale di substrato.

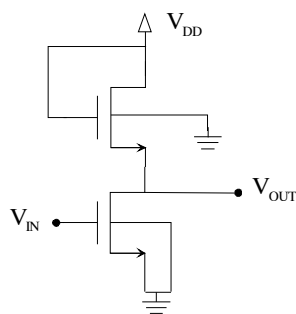


Quello che segue è un layout schematico per la NAND (processo LOCOS).

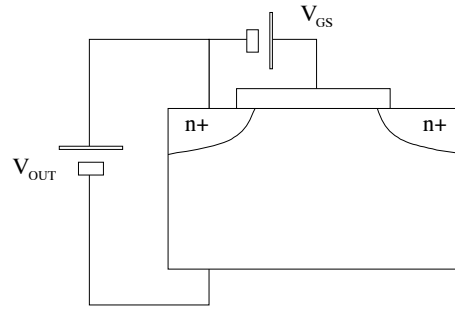


### L'effetto body

Mentre nell'NMOS di ingresso il source è collegato al substrato, il source dell'NMOS di carico si trova ad una tensione  $V_{OUT}$  rispetto al substrato. Questo fatto ha effetto sulla  $V_{TH}$  del carico.



Si può schematizzare la ddp fra source e substrato come una batteria  $V_{OUT}$ ; se il source è in corto con il drain la situazione è identica a quella della struttura MOS a tre terminali, per cui la tensione di soglia fra gate e substrato sarà



$$V_{TH}^* = \frac{\sqrt{2qN_A\epsilon_s(2\Psi_B + V_{OUT})}}{C_{OX}} + 2\Psi_B + V_{OUT},$$

mentre la  $V_{TH}$  che compare nelle caratteristiche, che è una  $V_{GS}$ , ed è data da

$$V_{TH} = V_{TH}^* - V_{OUT} = \frac{\sqrt{2qN_A\epsilon_s(2\Psi_B + V_{OUT})}}{C_{OX}} + 2\Psi_B$$

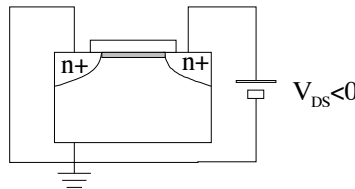
e quindi è aumentata, rispetto alla situazione source a massa, di una quantità

$$\Delta V_{TH} = \frac{\sqrt{2qN_A\epsilon_s(2\Psi_B + V_{OUT})}}{C_{OX}} - \frac{\sqrt{2qN_A\epsilon_s 2\Psi_B}}{C_{OX}}.$$

Questo incremento della tensione di soglia (effetto *body*) può essere verificato sperimentalmente misurando la  $V_{TH}$ , con uno dei metodi prima descritti, di un NMOS con il source tenuto ad una tensione  $V_{OUT}$  rispetto al sub.

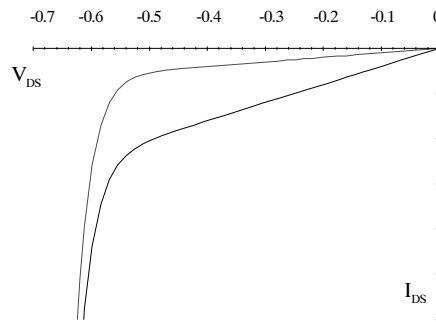
### Le caratteristiche con $V_{GS} > 0$ e $V_{DS} < 0$

E' immediato rendersi conto dalla figura che quando  $V_{DS}$  diviene negativa



il circuito equivalente in continua è costituito da un diodo (la giunzione drain/sub) polarizzato direttamente e da un resistore, costituito dal canale.  $I_{DS}$  sarà data da

$$I_{DS} = -I_0 \left( e^{\frac{v_{DS}}{V_T}} - 1 \right) - \frac{V_{DS}}{R}.$$



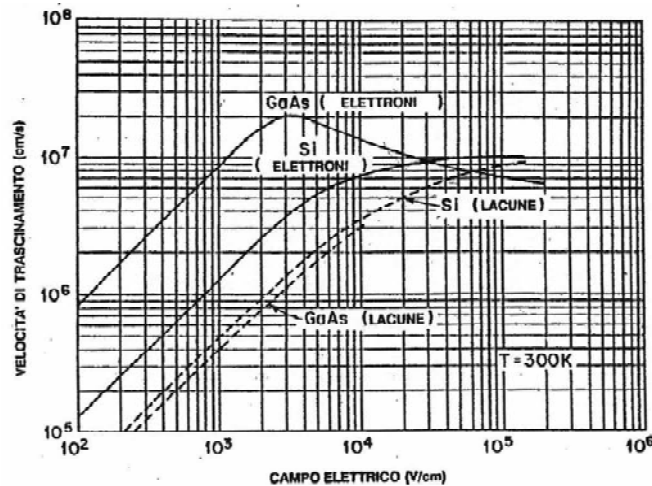
### L'NMOS a campi elevati

Quando si riducono le dimensioni dei dispositivi (*scaling down*), al fine di occupare meno spazio ed avere una maggiore velocità di risposta, le tensioni non possono essere ridotte dello stesso fattore  $K > 1$  utilizzato per scalare le geometrie. Se  $W$  ed  $L$  diventano  $W/K$  e  $L/K$  nella nuova versione del dispositivo, la tensione di alimentazione  $V_{DD}$  sarà in generale maggiore di  $V_{DD}/K$ . Dato che una stima del valore della componente del campo secondo  $y$  è il rapporto  $V_{DS}/L$ , è evidente che il campo secondo  $y$  è maggiore nel dispositivo con dimensioni ridotte. Questo fatto comporta un certo numero di conseguenze; una di queste è il fenomeno di saturazione della velocità di drift degli elettroni nel canale.

Normalmente (a campi non troppo elevati) vale la nota relazione

$$v_{drift} = \mu_n \mathcal{E};$$

quando il valore del campo si avvicina al valore  $\mathcal{E}_C$  (dell'ordine di  $10^4$  V/cm per il Si) la velocità comincia a saturare, come mostrato nel grafico.



Ad esempio per un NMOS con  $V_{DD} = 3.3$  V e  $L = 0.35 \mu\text{m}$  la stima del campo dà il valore  $9.4 \times 10^4$  V/cm e quindi i portatori sono in pieno regime di saturazione.

L'andamento della  $v_d$  in funzione del campo è descritto da una relazione empirica del tipo

$$v_d = \frac{\mu_o \mathcal{E}}{\left(1 + \left(\frac{\mathcal{E}}{\mathcal{E}_C}\right)^n\right)^{\frac{1}{n}}}$$

( $n = 2$  per gli elettroni e  $n = 1$  per le lacune) in cui  $\mathcal{E}_C$  è il campo critico. Per  $\mathcal{E} \gg \mathcal{E}_C$  la  $v_d$  assume il suo valore di saturazione

$$v_{dsat} = \mu_o \mathcal{E}_C.$$

Per valutare quale sia l'effetto sulle caratteristiche supponiamo che  $n = 1$  e consideriamo la

$$dV = \frac{I_{DS} dy}{W \mu_n Q_n(y)}$$

che può essere scritta

$$\frac{dV}{dy} = \frac{I_{DS}}{W \mu_n Q_n(y)}$$

$$\mathcal{E}(y) W \mu_n Q_n(y) = I_{DS}$$

e infine

$$I_{DS} = \mu_n \mathcal{E}(y) W Q_n(y) = \frac{\mu_o \frac{dV}{dy}}{1 + \left(\frac{\frac{dV}{dy}}{\mathcal{E}_C}\right)} W Q_n(y).$$

$$I_{DS} \left(1 + \left(\frac{\frac{dV}{dy}}{\mathcal{E}_C}\right)\right) = \mu_o \frac{dV}{dy} W Q_n(y)$$

$$I_{DS} = \frac{dV}{dy} \left(-\frac{I_{DS}}{\mathcal{E}_C} + \mu_o W Q_n(y)\right)$$

e integrando fra i soliti limiti (anche qui nell'ipotesi che  $Q_W$  non dipenda da  $V(y)$ ) si ottiene

$$I_{DS} = \mu_o C_{OX} \frac{W}{L \left(1 + \frac{V_{DS}}{\mathcal{E}_C L}\right)} \left( (V_{GS} - V_{TH}) V_{DS} - \frac{V_{DS}^2}{2} \right).$$

Si noti che il canale di lunghezza  $L$  appare più lungo quando  $V_{DS}/L$  diviene comparabile con  $\mathcal{E}_C$ . Nel caso  $V_{DS}/L \ll \mathcal{E}_C$  l'espressione è identica a quella usuale.

La tensione di saturazione si trova come al solito imponendo che

$$\frac{dI_{DS}}{dV_{DS}} = 0;$$

si ottiene, con qualche calcolo,

$$V_{DSAT} = \mathcal{E}_C L \left( \sqrt{1 + \frac{2(V_{GS} - V_{TH})}{\mathcal{E}_C L}} - 1 \right)$$

che per lunghezze di canale molto ridotte ( $L \rightarrow 0$ ) diviene

$$V_{DSAT} = \sqrt{2(V_{GS} - V_{TH}) \mathcal{E}_C L}.$$

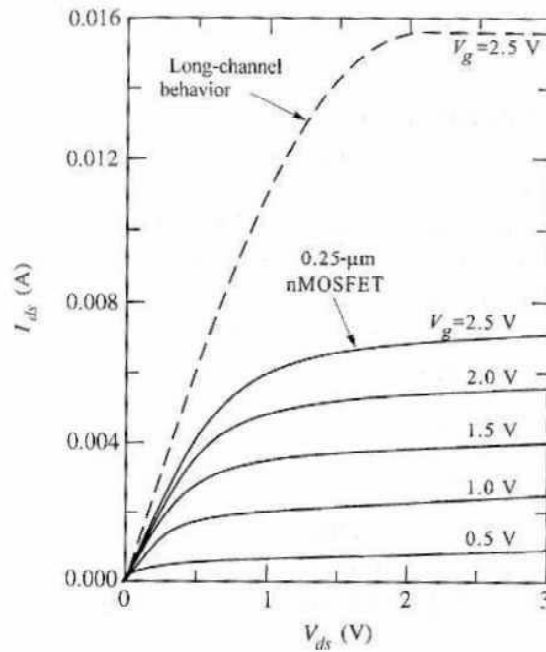
L'espressione della  $I_{DSAT}$  sarà dunque

$$I_{DSAT} = \mu_o C_{OX} \frac{W}{L \left(1 + \frac{V_{DSAT}}{\mathcal{E}_C L}\right)} \left( (V_{GS} - V_{TH}) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right)$$

e, per  $L \rightarrow 0$

$$\begin{aligned} I_{DSAT} &= \mu_o C_{OX} \frac{W \mathcal{E}_C}{V_{DSAT}} \left( (V_{GS} - V_{TH}) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right) \\ &= \mu_o C_{OX} W \mathcal{E}_C \left( (V_{GS} - V_{TH}) - \frac{\sqrt{2(V_{GS} - V_{TH}) \mathcal{E}_C L}}{2} \right) \\ &\simeq \mu_o C_{OX} W \mathcal{E}_C (V_{GS} - V_{TH}) : \end{aligned}$$

in regione di saturazione  $I_{DSAT}$  dipende *linearmente* da  $V_{GS} - V_{TH}$ , a differenza di quanto accadeva per il canale lungo. Si riportano nella figura le caratteristiche di un NMOS a canale corto ( $0.25 \mu\text{m}$ ), confrontate con quelle di un NMOS con la stessa lunghezza di canale, ma nel quale non è stato preso in considerazione l'effetto di saturazione della velocità.



### Capacità di ingresso e $f_T$

L'espressione della carica totale dello strato di inversione si può ottenere dalla relazione

$$Q_{nTOT} = W \int_0^L Q_n(y) dy.$$

Mediante  $Q_{nTOT}$  si calcola:

1) la corrente di piccolo segnale fra gate e source

$$i_{gs} = \frac{dQ_{nTOT}}{dt} = C_{GS} \frac{dv_{gs}}{dt}$$

dove  $C_{GS}$  rappresenta la capacità di ingresso del circuito equivalente per piccoli segnali.

2) il tempo di transito fra S e D, se si applica il modello del controllo di carica. Infatti

$$I_{DS} = \frac{Q_{nTOT}}{\tau_t}.$$

Con l'ipotesi grazie alla quale sono state ottenute le caratteristiche approssimate ( $Q_W$  indipendente da  $V(y)$ )  $Q_n(y)$  è data da

$$Q_n(y) = C_{ox} [(V_{GS} - V_{TH}) - V(y)].$$

Per potere esplicitare  $Q_{nTOT}$  è necessario conoscere l'andamento di  $V(y)$ , che si ricava ricordando che

$$I_{DS} = \frac{\mu_n C_{ox} W}{L} \left( (V_{GS} - V_{TH}) V_{DS} - \frac{V_{DS}^2}{2} \right)$$

è il risultato di una integrazione fra 0,  $L$  e fra 0 e  $V_{DS}$ . Se si integra invece fra 0 e  $y$  e fra 0 e  $V(y)$

$$I_{DS} = \frac{\mu_n C_{ox} W}{y} \left( (V_{GS} - V_{TH}) V(y) - \frac{V(y)^2}{2} \right)$$

Dato che per applicazioni analogiche si opera in regime di saturazione poniamo  $I_{DS} = I_{DSAT}$

$$I_{DSAT} = \frac{\mu_n C_{ox} W}{y} \left( (V_{GS} - V_{TH}) V(y) - \frac{V(y)^2}{2} \right)$$

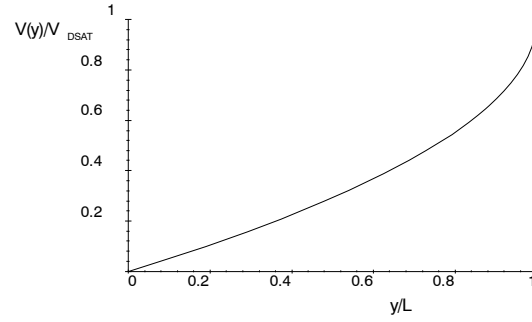
ovvero

$$\frac{\mu_n C_{ox} W}{2L} (V_{GS} - V_{TH})^2 = \frac{\mu_n C_{ox} W}{y} \left( (V_{GS} - V_{TH}) V(y) - \frac{V(y)^2}{2} \right)$$

che costituisce un'equazione di secondo grado in  $V(y)$ , che risolta dà

$$V(y) = (V_{GS} - V_{TH}) \left( 1 - \sqrt{1 - \frac{y}{L}} \right)$$





il cui andamento è mostrato in figura.

Siamo in grado ora di ottenere  $Q_{nTOT}$

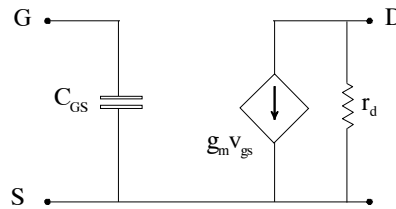
$$\begin{aligned}
 Q_{nTOT} &= W \int_0^L C_{ox} [(V_{GS} - V_{TH}) - V(y)] dy \\
 &= W \int_0^L C_{ox} \left[ (V_{GS} - V_{TH}) - (V_{GS} - V_{TH}) \left( 1 - \sqrt{1 - \frac{y}{L}} \right) \right] dy \\
 &= WC_{ox} \int_0^L (V_{GS} - V_{TH}) \left( \sqrt{1 - \frac{y}{L}} \right) dy \\
 &= \frac{2}{3} WLC_{ox} (V_{GS} - V_{TH}) = \frac{2}{3} C_{ox}^* (V_{GS} - V_{TH})
 \end{aligned}$$

in cui abbiamo indicato con  $C_{ox}^*$  la capacità totale del gate.

Per piccoli segnali

$$i_{gs} = \frac{dQ_{nTOT}}{dt} = \frac{2}{3} C_{ox}^* \frac{d(V_{GS} + v_{gs} - V_{TH})}{dt} = C_{GS} \frac{dv_{gs}}{dt}$$

e quindi la capacità di ingresso in saturazione vale  $\frac{2}{3} C_{ox}^*$ .



### Tempo di transito e $f_T$

Il tempo di transito in saturazione

$$\tau_t = \frac{Q_{nTOT}}{I_{DSAT}} = \frac{\frac{2}{3} C_{ox} W L (V_{GS} - V_{TH})}{\frac{\mu_n C_{ox} W (V_{GS} - V_{TH})^2}{2L}} = \frac{4}{3} \frac{L^2}{\mu_n (V_{GS} - V_{TH})}$$

dipende dalla lunghezza di gate al quadrato.

La frequenza di taglio  $f_T$  è definita dalla condizione di amplificazione di corrente unitaria. Ricordando che il generatore controllato di corrente in uscita è  $i_{ds} = g_m v_{gs}$  in cui  $g_m$  è la transconduttanza  $\left( g_m = \left[ \frac{\partial i_{DS}}{\partial v_{GS}} \right]_{V_{DS0}} \right)$

$$A(I) = \frac{g_m v_{gs}}{v_{gs} \omega_T C_{GS}} = 1$$

$$\frac{g_m}{2\pi f_T C_{GS}} = 1;$$

$g_m$  in saturazione è

$$g_m = \frac{\mu_n C_{ox} W (V_{GS} - V_{TH})}{L}$$

per cui

$$f_T = \frac{1}{2\pi \frac{2}{3} C_{ox} W L} \frac{\mu_n C_{ox} W (V_{GS} - V_{TH})}{L}$$

$$f_T = \frac{1}{\pi \frac{4}{3}} \frac{\mu_n (V_{GS} - V_{TH})}{L^2} = \frac{1}{\pi \tau_t}.$$

Risulta quindi evidente la necessità di avere un canale corto e/o alta mobilità nei dispositivi per applicazioni ad alta frequenza.

### Il breakdown della giunzione drain/sub

Uno dei problemi conseguenti ai campi elevati negli NMOS a canale corto è il breakdown delle giunzione drain/substrato, che può essere schematizzata come una giunzione brusca  $n^+p$  nel caso dell'NMOS e  $p^+n$  nel caso del PMOS.

Nella  $p^+n$  il valore massimo del campo all'equilibrio è dato da

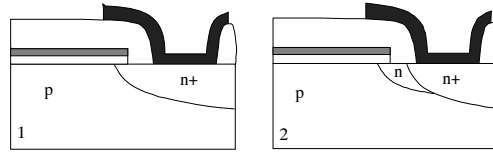
$$\mathcal{E}_{MAX} = \frac{qN_D W}{\epsilon_s} = \frac{qN_D}{\epsilon_s} \sqrt{\frac{2\epsilon_s V_0}{qN_D}} = \sqrt{\frac{2qN_D V_0}{\epsilon_s}}$$

ed è quindi funzione crescente del drogaggio del lato  $n$ ; abbiamo supposto  $W \simeq x_n$ . Se invece poniamo  $W = x_n + x_p$   $\mathcal{E}_{MAX}$  è dato da

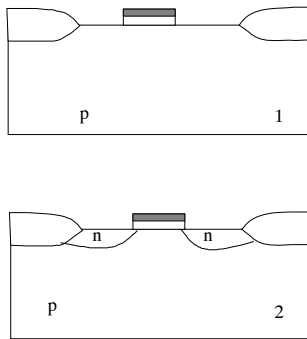
$$\mathcal{E}_{MAX} = \frac{qN_D x_n}{\epsilon_s} = \frac{qN_D}{\epsilon_s} \frac{N_A}{N_A + N_D} \sqrt{\frac{2qV_0}{\epsilon_s} \left( \frac{1}{N_A} + \frac{1}{N_D} \right)} = \frac{q}{\epsilon_s} \sqrt{\frac{2qV_0}{\epsilon_s} \frac{N_D}{1 + N_D/N_A}}$$

che, con  $N_D$  costante, è funzione decrescente del drogaggio del lato  $p^+$ . Risulta quindi conveniente, se si vuole diminuire il valore di  $\mathcal{E}_{MAX}$ , ridurre il drogaggio anche del lato più drogato della giunzione. Questo contrasta con l'esigenza di avere un buon contatto ohmico fra metal e pozzetto di drain. La soluzione comunemente

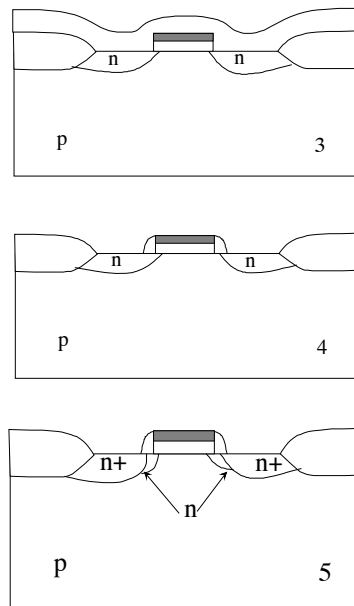
adottata e sempre presente in ogni dispositivo NMOS o CMOS è la LDD (*Lightly Doped Drain*). Si vuole trasformare la struttura del contatto da 1 a 2, mantenendo così una bassa  $R_c$  e diminuendo al tempo stesso il valore del campo all'interfaccia fra drain e sub.



Il processo LOCOS viene modificato come segue. Subito dopo la definizione del gate (1) viene eseguita un'impiantazione  $n$  con drogaggio non troppo elevato (2);



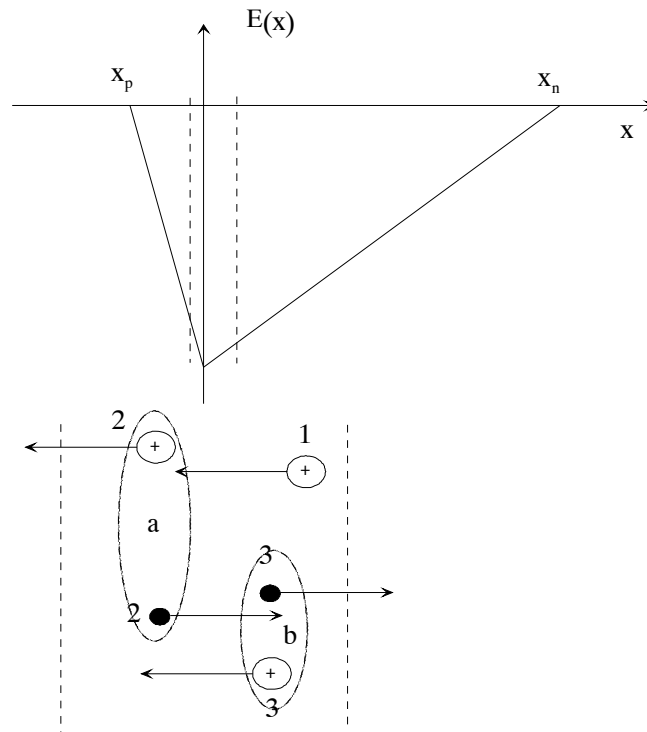
si depone poi un isolante CVD (3) e un successivo attacco dry anisotropo, che agisce solo in direzione verticale, elimina (4) l'isolante fino a scoprire le regioni di S e D, che tuttavia restano protette ai lati del gate da due *spacer*.



Da questo punto in poi si prosegue come al solito: impiantazione  $n^+$  (5) (che non può arrivare sotto il gate perché mascherata dagli spaziatori), deposizione CVD di isolante, apertura delle finestre dei contatti ecc.

Il fenomeno del breakdown è dovuto ai portatori minoritari che, nell'attraversare la zds, vengono accelerati dal campo elettrico. Se l'energia che acquistano dal campo fra una collisione e l'altra è sufficiente possono, durante un urto, creare una coppia elettrone-lacuna. Ciò che avviene può essere descritto nel modo semplificato seguente, con riferimento alla giunzione  $p^+n$ .

### Breakdown a valanga



Una lacuna (1) che attraversa la zds collide con il reticolo (evento  $a$ ) e produce una coppia elettrone2-lacuna2. L'elettrone2 a sua volta, nell'evento  $b$ , produce un'altra coppia e3-l3 e così via. Detta  $P$  la probabilità che venga generata una coppia, sia da una lacuna che da un elettrone, avremo per il flusso di lacune uscente da una superficie ortogonale ad  $x_p$

$$F_{out} = F_{in} + PF_{in} + P^2F_{in} + \dots + P^nF_{in} + \dots$$

$$\frac{F_{out}}{F_{in}} = 1 + P + P^2 + \dots + P^n + \dots = \sum_{n=0}^{\infty} P^n = \frac{1}{1-P};$$

il rapporto  $\frac{F_{out}}{F_{in}}$  costituisce evidentemente il fattore  $M$  di moltiplicazione della densità di corrente. Quindi

$$M = \frac{1}{1 - P}.$$

Se si ammette poi che  $P$  dipenda dalla tensione inversa applicata  $V$  come

$$P = \left( \frac{V}{V_{BD}} \right)^\alpha$$

avremo

$$M = \frac{1}{1 - \left( \frac{V}{V_{BD}} \right)^\alpha}$$

che diverge quando  $V \rightarrow V_{BD}$ .

L'andamento di  $M$  con  $V_{BD} = 10$  e  $\alpha = 3$  è mostrato nella figura.

Si osservi che se  $N$  lacune attraversano nell'unità di tempo una superficie in  $x_p$ , la stessa superficie in  $x_n$  sarà attraversata da  $N - 1$  elettroni e da una lacuna.

Per ciò che riguarda  $V_{BD}$ , in condizioni di polarizzazione inversa il valore massimo del campo per il quale si ha il breakdown della giunzione è dato da

$$\mathcal{E}_{\max BD} = \sqrt{\frac{2qN_D (V_0 - V_{BD})}{\epsilon_s}}$$

Se  $V_0 \ll |V_{BD}|$

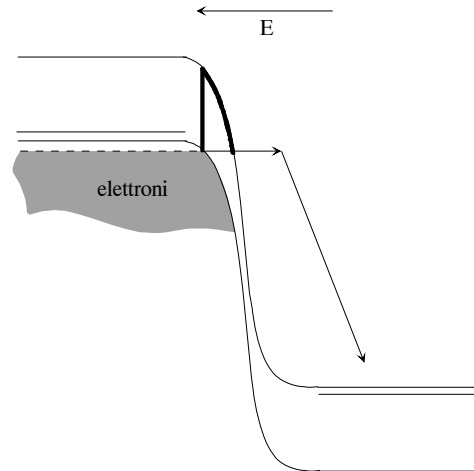
$$\mathcal{E}_{\max BD} \simeq \sqrt{\frac{2qN_D V_{BD}}{\epsilon_s}}$$

$$\frac{\epsilon_s \mathcal{E}_{\max BD}^2}{2qN_D} = V_{BD}$$

e, se supponiamo che per avere l'innescò del fenomeno, è necessario che nella zds si abbia un campo che valga  $\mathcal{E}_{\max BD}$ , allora la  $V_{BD}$  decresce all'aumentare del drogaggio.

### Breakdown Zener

Se il drogaggio dei due lati della giunzione diviene elevato la struttura a bande in polarizzazione inversa è la seguente:



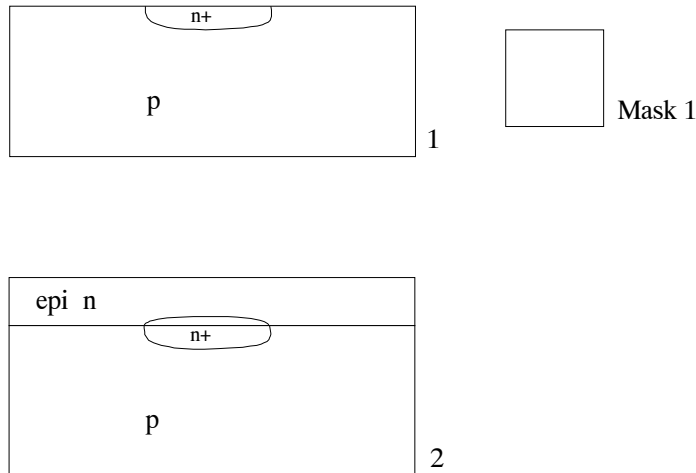
Gli elettroni della BV possono attraversare la barriera evidenziata (di altezza  $E_g$ ) per effetto tunnel, se questa ha una larghezza sufficientemente ridotta, ed entrare così nella BC del lato  $n$ . Ad un elettrone che compie questo passaggio corrisponde una lacuna nella BV del lato  $p$ ; in definitiva la generazione di una coppia.

## 24) IL TRANSISTORE BIPOLARE

### Il processo SBC

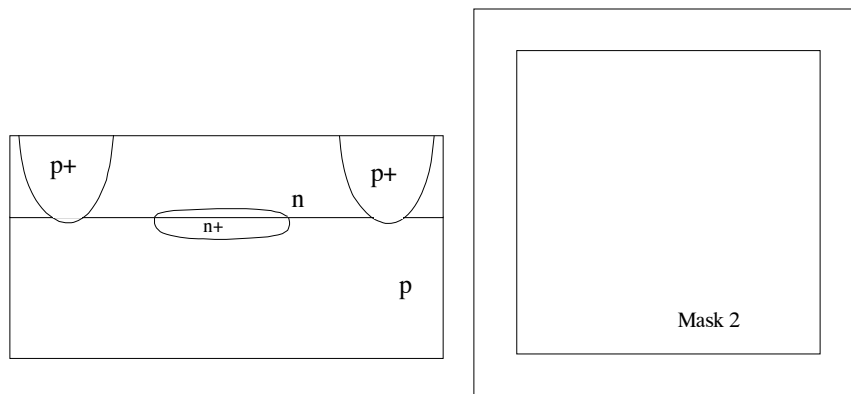
SBC sta per *Standard Buried Collector* ed è il processo tradizionale per integrare BJT *npn*. E' uno dei tanti possibili, ed è abbastanza versatile.

1) Si parte da un sub *p* poco drogato ((100) o (111)) e con la maschera 1 si diffonde lo strato sepolto (*buried layer*). Viene utilizzato l'arsenico perché è un diffusore lento e ha lo stesso raggio tetraedrico del Si.

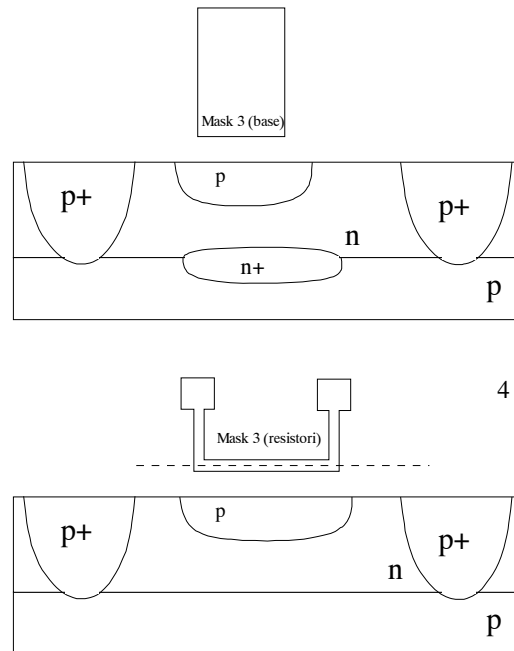


2) Si cresce uno strato epitassiale di Si *n* ( $\simeq 10^{15} \text{ cm}^{-3}$ ), il cui spessore dipende dalle applicazioni (per esempio la massima  $V_{CB}$ ). Si noti che il buried layer diffonde anche verso l'alto durante la crescita dell'epi.

3) Con la maschera 2 si esegue la diffusione di isolamento. Se lo strato epitassiale è spesso la diffusione richiede tempi lunghi e occupa molta area, a causa della diffusione laterale.



4) Con la maschera 3 viene diffusa la base. In un pozzetto isolato a parte si esegue contemporaneamente la diffusione dei resistori, che hanno dunque lo stesso profilo di drogaggio della base.

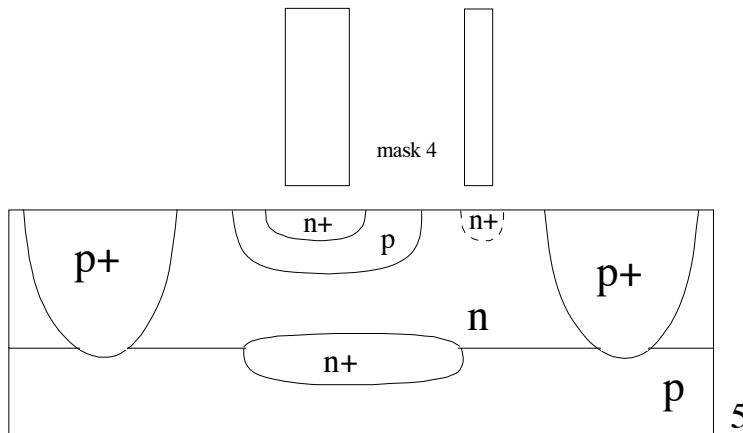


La resistenza di strato del resistore diffuso è data da

$$R_{\square} = \frac{\bar{\rho}}{x_j} = \frac{1}{\bar{\sigma}x_j} = \frac{1}{x_j \frac{q}{x_j} \int_0^{x_j} \mu_h(N_A(x))N_A(x)dx}$$

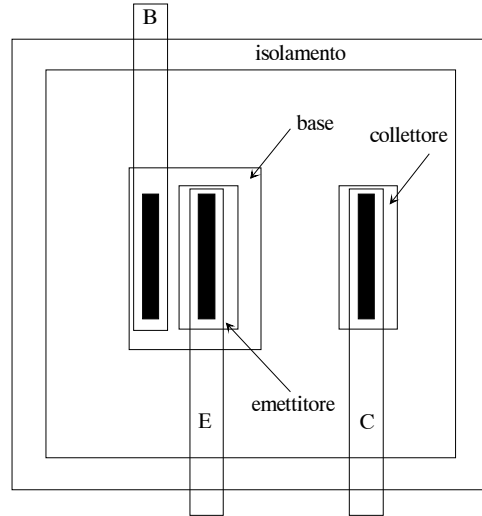
dove evidentemente è stata usata l'approssimazione di quasi-neutralità ( $p(x) \simeq N_A(x)$ ). In uno stesso pozzetto vengono integrati più resistori, tipicamente quelli che sono contigui anche nello schema circuitale. Si noti ancora che un resistore forma con lo strato epi una giunzione e che quindi si dovrà provvedere a che questa non entri in conduzione.

5) Viene realizzato l'emettitore e il contatto di collettore (maschera 4). Sia la diffusione di base che quella di emettitore sono critiche perché determinano la larghezza della base e quindi il  $\beta$  del BJT.





Il BJT è completato. Bisogna provvedere ai contatti e ai collegamenti che si ottengono aprendo delle finestre nell'ossido in corrispondenza di E, B e C. Il layout schematico del dispositivo, completo di metal, è mostrato nella figura seguente.



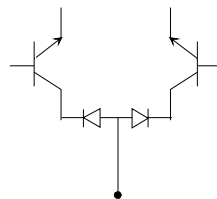
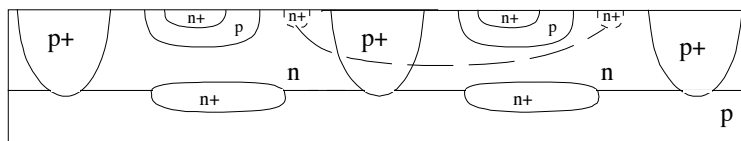
Insieme all'E viene diffuso il contatto  $n^+$  nel pozzetto dei resistori. Se questo contatto è tenuto al potenziale più positivo di tutto il circuito le giunzioni costituite dai resistori e dall'epi  $n$  non entrano mai in conduzione.

6) Il circuito viene passivato con una deposizione CVD di isolante ( $\text{SiO}_2$ ,  $\text{Si}_3\text{N}_4$ ) e con l'ultima maschera si aprono i contatti dei pad.

Resta da spiegare la funzione della diffusione di isolamento e dello strato sepolto.

i) Strato sepolto: costituisce un percorso a bassa resistenza fra collettore e base e contribuisce quindi a diminuire la resistenza parassita di collettore.

ii) Diffusione di isolamento.

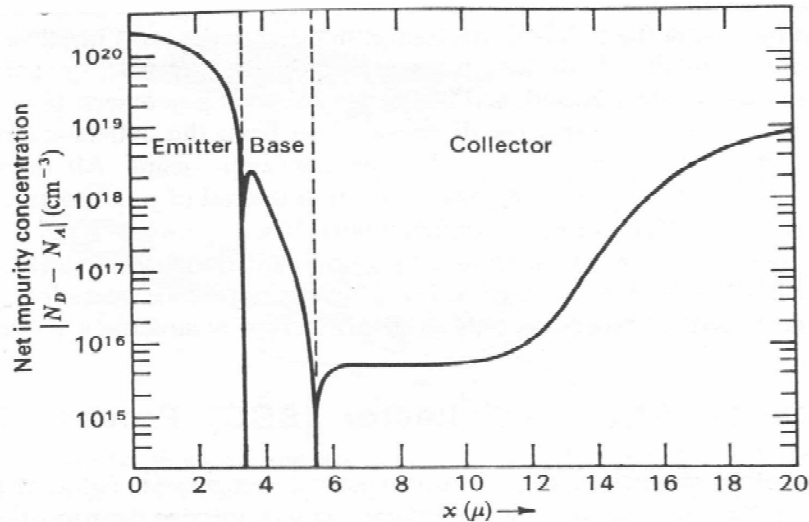


Se consideriamo due BJT in pozzetti separati, il collegamento fra i collettori è quello mostrato in figura dal percorso tratteggiato. Polarizzando il terminale

di substrato con la tensione più negativa di tutto il CI i collettori comunicano attraverso due diodi back-to-back e quindi di fatto sono isolati fra loro.

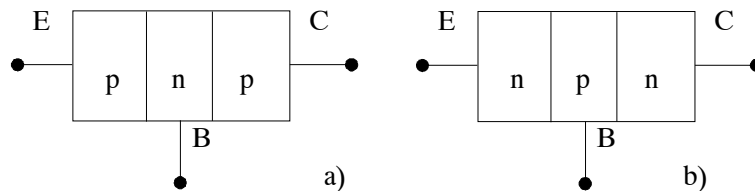
In generale un dato processo viene ottimizzato per il componente più importante (in questo caso l'*nnp*); tutti gli altri elementi circuitali (resistori, diodi, condensatori) sono ottenuti sfruttando i passi di processo disponibili.

Il profilo di drogaggio di un tipico BJT integrato SBC è riportato nella figura seguente. Si noti l'aumento di drogaggio del collettore dovuto allo strato sepolto.



### Struttura schematica e principio di funzionamento

Il transistor bipolare è un dispositivo a tre terminali costituito dall'unione di due giunzioni *pn*, che possono avere a comune la parte *n* (BJT *pnp*), ovvero la parte *p* (BJT *nnp*).



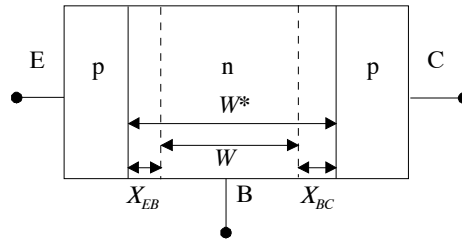
Affinché la struttura possa effettivamente funzionare come transistor è indispensabile che la zona a comune (la *base* B del BJT) abbia la dimensione  $W$  (lunghezza effettiva di base) molto minore della lunghezza di diffusione  $L$  dei portatori minoritari ( $L_h$  se la base è di tipo *n*,  $L_n$  se la base è di tipo *p*). Gli altri due terminali sono l'emettitore E e il collettore C.

La funzione del BJT è il controllo della corrente di collettore  $I_C$  mediante un'altra corrente, quella di emettitore  $I_E$  o quella di base  $I_B$ .

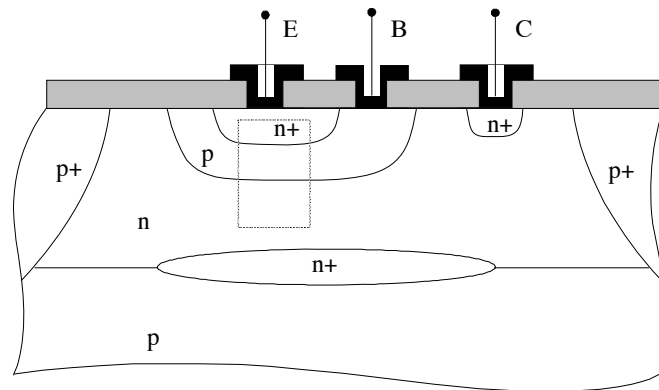
$W$  è definita come la distanza fra i confini delle zone di svuotamento, inevitabilmente presenti nella base.

$W^*$  è la larghezza metallurgica di base;  $W^* = W + X_{EB} + X_{BC}$ .

Si noti fin da adesso che  $X_{EB}$  e  $X_{BC}$  dipendono dalla polarizzazione delle giunzioni EB e BC.



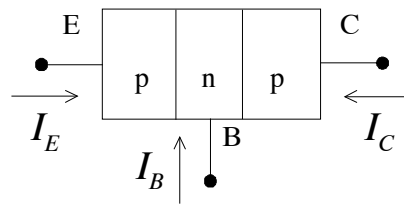
Le strutture delle figure a) e b) rappresentano schematicamente il BJT e possono essere considerate come una parte di un BJT reale integrato. Struttura schematica e struttura effettiva sono mostrate nella figura per un *npn*.



Il dispositivo ideale (BJT didattico) oggetto del presente studio è così definito:

- 1) i drogaggi di E, B e C sono *costanti*;
- 2) i drogaggi di E e C sono molto maggiori del drogaggio della base;
- 3) non ci sono elementi parassiti (resistenze di contatto, resistenze delle zone di E, B, C);
- 4) la struttura è monodimensionale.

Consideriamo inizialmente un BJT  $p^+np^+$  e fissiamo come verso positivo per le correnti  $I_E$ ,  $I_B$  e  $I_C$  quello entrante. Si ha per le correnti del nodo, costituito dal BJT



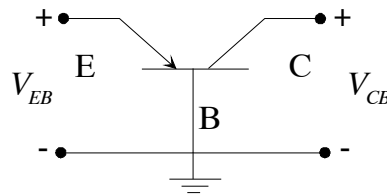
$$I_E + I_B + I_C = 0$$

e quindi, note due correnti, la terza risulta univocamente determinata.  $I_E$ ,  $I_B$  e  $I_C$  sono correnti misurabili ai terminali (correnti *esterne*) e devono essere messe in relazione con i fenomeni che avvengono all'*interno* della struttura.

Per quanto riguarda le tensioni, si può scegliere come terminale di riferimento o la base (configurazione a base comune, CB) ed è questo il caso più semplice, oppure l'emettitore (configurazione a emettitore comune, CE).

Consideriamo preliminarmente il caso CB. Nella figura è riportato il simbolo del dispositivo, ai cui terminali si possono applicare indipendentemente due tensioni  $V_{EB}$  e  $V_{CB}$ ; la terza risulta quindi determinata dalla relazione

$$V_{EB} + V_{BC} + V_{CE} = 0$$



Poiché le due tensioni  $V_{EB}$  e  $V_{CB}$  sono indipendenti, per il dispositivo sono possibili quattro condizioni di funzionamento:

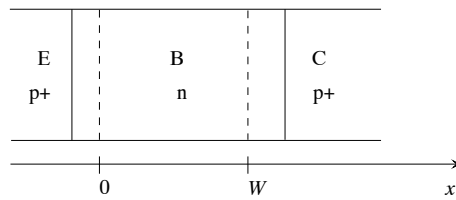
1)  $V_{EB} > 0$  (giunzione EB polarizzata direttamente) e  $V_{CB} < 0$  (giunzione CB polarizzata inversamente): zona di funzionamento *attiva diretta*.

2)  $V_{EB} < 0$  (giunzione EB polarizzata inv.) e  $V_{CB} > 0$  (giunzione CB polarizzata dir.): condizione di funzionamento *attiva inversa*. Si noti che in questo caso, data la simmetria della struttura, E e C possono essere scambiati fra di loro. Nel transistor reale questo non è possibile.

3)  $V_{EB} > 0$  e  $V_{CB} > 0$ : condizione di *saturazione*.

4)  $V_{EB} < 0$  e  $V_{CB} < 0$ : condizione di *interdizione* (*cut-off*).

Prendiamo in considerazione la zona di funzionamento attiva diretta ( $V_{EB} > 0$ ,  $V_{CB} < 0$ ) e stabiliamo un sistema di riferimento (asse  $x$ ) orientato da E a C e con origine posta alla fine della zona di svuotamento nella base, dalla parte dell'E (dato che la struttura è per ipotesi  $p^+np^+$  le zone di svuotamento delle giunzioni EB e BC si estendono prevalentemente nella base).



Poiché  $V_{EB} > 0$ , la giunzione EB è polarizzata direttamente e l'eccesso dei portatori minoritari nella base (in questo caso lacune) è determinato dalla nota relazione

$$\delta p_n(0) = p_{n0} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

mentre l'eccesso in  $W$  (giunzione BC) è dato dall'analogha espressione

$$\delta p_n(W) = p_{n0} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right)$$

Si ricordi ora che  $V_{CB} < 0$ ; se poi è anche  $|V_{CB}| \gg V_T$  avremo

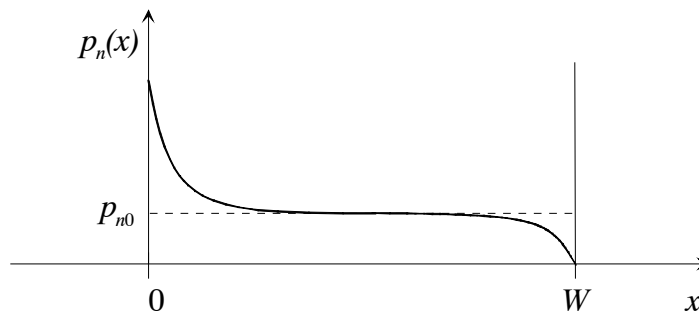
$$\delta p_n(W) = -p_{n0}$$

ovvero la concentrazione dei minoritari in  $W$  vale zero, come è logico per una giunzione  $pn$  polarizzata inversamente.

Abbiamo dunque stabilito il valore della concentrazione dell'eccesso dei portatori agli estremi della zona (neutra) di base.

Facciamo adesso alcune considerazioni su ciò che accade nella base. La giunzione  $p^+n$  EB, polarizzata direttamente, inietta lacune in eccesso nella base  $n$ . Questi portatori in eccesso diffondono e si ricombinano. Ma poiché per ipotesi  $W \ll L_n$ , solo poche lacune avranno la possibilità di ricombinarsi (si ricordi il diodo a base corta); la maggior parte raggiungerà per diffusione la giunzione BC, polarizzata inversamente. Questo è possibile perché i due diodi sono fortemente accoppiati ( $W \ll L_n$ ).

Per apprezzare meglio questo fatto, esaminiamo quello che accadrebbe nel caso opposto ( $W \gg L_n$ ). Nella figura è riportato l'andamento del profilo delle lacune in questo caso.



Come si vede la giunzione EB inietta lacune, la cui concentrazione decade esponenzialmente con costante  $L_h$ ; c'è poi una zona in cui la concentrazione dei portatori nella base è quella di equilibrio ( $p_{n0}$ ) e infine una zona svuotata di minoritari (giunzione BC polarizzata inversamente). I due diodi sono indipendenti: le condizioni di polarizzazione di una delle giunzioni non hanno effetto sul comportamento dell'altra. La struttura non si comporta come un transistoro.

Dobbiamo adesso collegare ciò che accade all'interno del dispositivo con le quantità misurabili esternamente (correnti e tensioni) e ricordiamo che nel caso di giunzioni  $pn$  il calcolo della corrente, in condizioni stazionarie, si riduce alla conoscenza del profilo di eccesso dei minoritari.

Nel nostro caso (BJT  $p^+np^+$ ) la corrente  $I_E$ , conseguenza della polarizzazione diretta della giunzione EB, si può considerare costituita tutta dalle lacune (giunzione brusca unilaterale) per cui

$$I_E \simeq I_{Eh}$$

mentre in generale

$$I_E = I_{Eh} + I_{En}.$$

Definiamo l'*efficienza di emettitore*  $\gamma$  come il rapporto fra la corrente dovuta ai minoritari iniettati nella base e la corrente totale di emettitore

$$\gamma = \frac{I_{Eh}}{I_E} = \frac{I_{Eh}}{I_{Eh} + I_{En}}$$

In generale  $\gamma$  è minore di uno.

Assumiamo, come già detto,  $\gamma = 1$ : il calcolo di  $I_E$  si riduce al calcolo di  $I_{Eh}$ , e dunque alla conoscenza del profilo delle lacune in eccesso nella base. Infatti

$$I_{Eh} = -qD_h S \frac{d\delta p_n(x)}{dx}$$

Ricordiamo che due valori notevoli di  $\delta p_n(x)$  sono già noti e determinati dalle condizioni di polarizzazione in cui ci siamo posti.

L'andamento del profilo delle lacune in eccesso si ottiene risolvendo l'equazione di continuità nella base

$$D_h \frac{d^2 \delta p_n(x)}{dx^2} = \frac{\delta p_n(x)}{\tau_h}$$

la cui soluzione generale è

$$\delta p_n(x) = A e^{-\frac{x}{L_h}} + B e^{+\frac{x}{L_h}}$$

Osserviamo che, essendo  $W \ll L_h$ , il profilo delle lacune in eccesso può essere scritto, sviluppando in serie gli esponenziali e arrestandosi al prim'ordine

$$\delta p_n(x) = A\left(1 - \frac{x}{L_h}\right) + B\left(1 + \frac{x}{L_h}\right) = A^* + B^* \frac{x}{L_h}$$

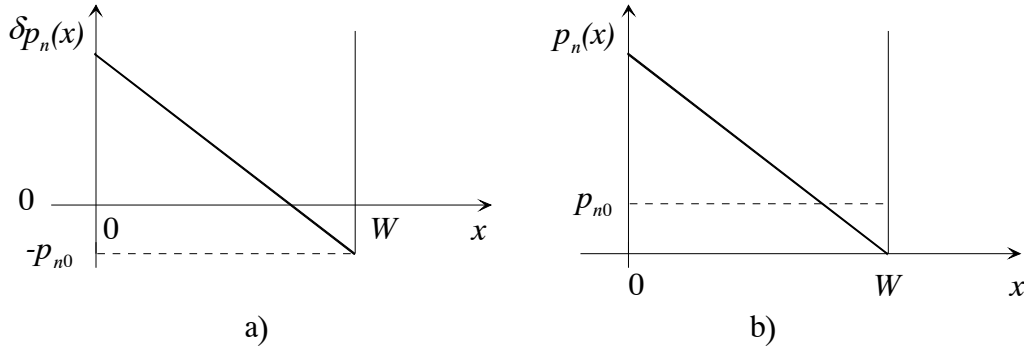
da cui, imponendo le condizioni al contorno in 0 e  $W$ , si ottiene

$$\delta p_n(x) = \delta p_n(0)\left(1 - \frac{x}{W}\right) + \delta p_n(W)\frac{x}{W}$$

In zona attiva avremo

$$\delta p_n(x) = p_{n0}\left(e^{\frac{V_{EB}}{V_T}} - 1\right)\left(1 - \frac{x}{W}\right) - p_{n0}\frac{x}{W}$$

il cui andamento è mostrato nella figura (a) insieme con  $p_n(x)$ , concentrazione di lacune nella base (b).



In prima approssimazione  $\delta p_n(x)$  è una retta. La corrente  $I_E$  sarà dunque

$$I_E = -qD_hS \frac{d\delta p_n(x)}{dx} = \frac{qSD_h p_n(0)}{W}$$

Avendo approssimato il profilo di eccesso con una retta, è evidente che la corrente di lacune in  $W$ , ovvero la  $I_C$ , risulterà uguale in modulo alla  $I_E$ . In questo caso

$$I_E = -I_C$$

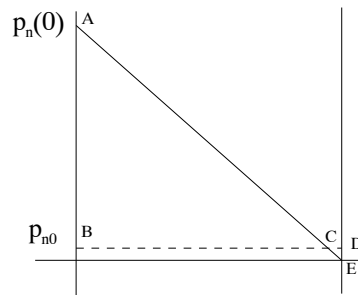
e quindi  $I_B = 0$ .

Riprendiamo in considerazione ciò che accade all'interno della base: i) le lacune iniettate dal contatto EB diffondono verso il contatto CB; ii) durante il percorso una certa frazione di lacune (piccola poiché  $W \ll L_h$ ) si ricombina con elettroni, il cui rifornimento è assicurato dal terminale di base.  $I_B$  non può dunque essere nulla; ne consegue che le derivate del profilo, calcolate in 0 e  $W$  rispettivamente, devono essere diverse, sia pure di poco in quanto

$$I_E = -I_C - I_B.$$

$I_E$  è positiva (lacune che entrano dal terminale di E verso il BJT),  $I_C$  è negativa (lacune che escono dal BJT verso il terminale di C),  $I_B$  è negativa (elettroni che entrano dal terminale di base verso il BJT).

Con riferimento alla figura e tenuto conto che  $I_B$  è una corrente di ricombinazione una sua stima può essere data utilizzando il modello del controllo di carica



$$I_B = -\frac{Q_B}{\tau_h}$$

dove  $Q_B$  è la carica totale in eccesso immagazzinata nella base.

Per il calcolo di  $Q_B$  si osservi che:

- l'area  $ABC$  moltiplicata per  $qS$  ( $q$  è la carica dell'elettrone,  $S$  la sezione del BJT) rappresenta un eccesso positivo di carica;
- l'area  $CDE$  moltiplicata per  $qS$  rappresenta un eccesso negativo;

$$Q_B = qS(ABC - CDE)$$

Notiamo tuttavia che

$$\blacktriangleright AB = p_{n0} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

$$\blacktriangleright DE = p_{n0}$$

e quindi, se  $e^{\frac{V_{EB}}{V_T}} \gg 1$  segue che  $AB \gg DE$ ; tenendo conto poi che il profilo effettivo differisce di poco da quello lineare si commetterà un errore trascurabile utilizzando per il calcolo di  $Q_B$  l'espressione seguente

$$Q_B \simeq \frac{qS\delta p_n(0)W}{2}$$

A proposito di  $I_C$  si osservi che le lacune provenienti dall'emettitore vengono raccolte dalla giunzione CB, in questo caso polarizzata inversamente, e considerate



da questa come portatori minoritari. A formare la  $I_C$  contribuisce dunque anche la corrente inversa di saturazione della giunzione CB.

Nella configurazione a base comune la corrente di ingresso  $I_E$ , causata una tensione  $V_{EB}$  (diretta) di entità inferiore ad 1 V, trasferisce in uscita una corrente  $I_C$ , di poco inferiore, ad una tensione  $V_{CB}$  (inversa) che può variare in un range di alcune decine di volt.

### Le equazioni di Ebers e Moll

Dobbiamo adesso risolvere il problema di dare forma analitica alle caratteristiche di ingresso e di uscita del BJT. Nella configurazione a base comune le caratteristiche di uscita

$$I_C = f(V_{CB}; I_E)$$

descrivono l'andamento di  $I_C$  (corrente di uscita) in funzione di  $V_{CB}$  (tensione di uscita) con  $I_E$  (corrente di ingresso) come variabile di controllo.

Le caratteristiche di ingresso

$$I_E = g(V_{EB}; V_{CB})$$

descrivono l'andamento di  $I_E$  (corrente di ingresso) in funzione di  $V_{EB}$  (tensione di ingresso) con  $V_{CB}$  come parametro.

Riprendiamo in esame la soluzione  $\delta p_n(x)$  dell'equazione di continuità nella base.

Le costanti  $A$  e  $B$  si determinano imponendo le condizioni al contorno in  $x = 0$  e  $x = W$

$$\delta p_n(0) = A + B$$

e

$$\delta p_n(W) = Ae^{-\frac{W}{L_h}} + Be^{+\frac{W}{L_h}}$$

da cui si ottiene

$$A = \frac{\delta p_n(0)e^{\frac{W}{L_h}} - \delta p_n(W)}{e^{\frac{W}{L_h}} - e^{-\frac{W}{L_h}}}$$

$$B = \frac{\delta p_n(W) - \delta p_n(0)e^{-\frac{W}{L_h}}}{e^{\frac{W}{L_h}} - e^{-\frac{W}{L_h}}}$$

in cui, ricordiamo

$$\delta p_n(0) = p_{n0} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

e

$$\delta p_n(W) = p_{n0} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right)$$

sono indipendentemente determinati da  $V_{EB}$  e  $V_{CB}$ .

$I_E$  ed  $I_C$  sono correnti di diffusione determinabili dalle relazioni

$$I_E = -qD_h S \frac{d\delta p_n(x)}{dx} \Big|_{x=0}$$

$$I_C = qD_h S \frac{d\delta p_n(x)}{dx} \Big|_{x=W}$$

nell'ipotesi che  $\gamma = 1$ .

Il segno positivo nell'espressione della  $I_C$  tiene conto del fatto che  $I_C$  è una corrente uscente.

Dalla

$$\delta p_n(x) = A e^{-\frac{x}{L_h}} + B e^{+\frac{x}{L_h}}$$

si ottiene

$$I_E = \frac{qD_h S}{L_h} (A - B)$$

e

$$I_C = \frac{qD_h S}{L_h} (B e^{\frac{W}{L_h}} - A e^{-\frac{W}{L_h}})$$

Poiché  $A$  e  $B$  sono funzioni lineari di  $\delta p(0)$  e  $\delta p(W)$  anche per  $I_E$  ed  $I_C$  si potrà scrivere

$$I_E = a_{11} \delta p_n(0) + a_{12} \delta p_n(W)$$

$$I_C = a_{21} \delta p_n(0) + a_{22} \delta p_n(W)$$

dove

$$a_{11} = a_{22} = \frac{qD_h S}{L_h} \coth\left(\frac{W}{L_h}\right)$$

$$a_{21} = a_{12} = -\frac{qD_h S}{L_h} \operatorname{csch}\left(\frac{W}{L_h}\right)$$

e quindi

$$\begin{aligned} I_E &= \frac{qD_h S}{L_h} \coth\left(\frac{W}{L_h}\right) \delta p_n(0) - \frac{qD_h S}{L_h} \operatorname{csch}\left(\frac{W}{L_h}\right) \delta p_n(W) \\ I_C &= -\frac{qD_h S}{L_h} \operatorname{csch}\left(\frac{W}{L_h}\right) \delta p_n(0) + \frac{qD_h S}{L_h} \coth\left(\frac{W}{L_h}\right) \delta p_n(W) \end{aligned}$$

ovvero, esplicitando la dipendenza da  $V_{EB}$  e  $V_{CB}$

$$\begin{aligned} I_E &= A_{11} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) + A_{12} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right) \\ I_C &= A_{21} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) + A_{22} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right) \end{aligned}$$

con

$$\begin{aligned} A_{11} &= A_{22} = p_{n0} \frac{qD_h S}{L_h} \coth\left(\frac{W}{L_h}\right) \\ A_{21} &= A_{12} = -p_{n0} \frac{qD_h S}{L_h} \operatorname{csch}\left(\frac{W}{L_h}\right) \end{aligned}$$

Le correnti  $I_E$  ed  $I_C$  dipendono sia da  $V_{EB}$  che da  $V_{CB}$  e le costanti  $A_{12}$  e  $A_{21}$  sono le costanti di accoppiamento fra le due giunzioni.

Poiché

$$\coth\left(\frac{W}{L_h}\right) = \frac{e^{\frac{W}{L_h}} + e^{-\frac{W}{L_h}}}{e^{\frac{W}{L_h}} - e^{-\frac{W}{L_h}}}$$

$$\operatorname{csch}\left(\frac{W}{L_h}\right) = \frac{2}{e^{\frac{W}{L_h}} - e^{-\frac{W}{L_h}}}$$

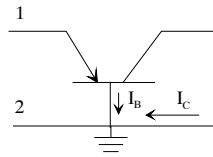
se  $W \gg L_h$   $\coth\left(\frac{W}{L_h}\right) \rightarrow 1$  e  $\operatorname{csch}\left(\frac{W}{L_h}\right) \rightarrow 0$ ;  $I_E$  e  $I_C$  si riducono a

$$I_E = \frac{qD_h S}{L_h} p_{n0} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

$$I_C = \frac{qD_h S}{L_h} p_{n0} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right)$$

In questo caso le costanti di accoppiamento sono nulle e le due giunzioni si comportano quindi come due diodi indipendenti.

Cortocircuitiamo ora il collettore con la base ( $V_{CB} = 0$ ) e misuriamo sia  $I_E$  che  $I_C$ .



Otteniamo

$$I_E = A_{11} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

$$I_C = A_{21} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

La prima è la caratteristica del diodo emettitore-base quando il collettore è in corto con la base; segue che  $A_{11} = I_{ES}$  è la corrente di saturazione di questa giunzione

$$I_E = I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right)$$

Tuttavia il fatto più interessante è che anche la  $I_C$  è non nulla: la maggior parte della corrente iniettata dall'emettitore arriva al terminale 2 passando attraverso il terminale di collettore, mentre una minima parte esce dal terminale di base.

Definiamo il guadagno di corrente a base comune

$$\alpha_F = -\frac{I_C}{I_E} \Big|_{V_{CB}=0} = -\frac{A_{21}}{A_{11}}$$

dove il pedice F sta per forward, in quanto il BJT sta funzionando in modo diretto, con l'emettitore che inietta portatori e il collettore che li raccoglie. Si ricordi che il BJT didattico che stiamo considerando è simmetrico (l'E e il C possono essere scambiati fra di loro), mentre un BJT reale non lo è. Fin da adesso comunque faremo distinzione fra F (Forward) e R (Reverse).

Nel nostro caso  $\alpha_F$  può essere calcolato; in generale, dato che la struttura di un BJT reale, come vedremo, è più complicata, può risultare più conveniente misurarlo.

$$\alpha_F = -\frac{A_{21}}{A_{11}} = \frac{\operatorname{csch}\left(\frac{W}{L_h}\right)}{\operatorname{coth}\left(\frac{W}{L_h}\right)} = \frac{1}{\operatorname{cosh}\left(\frac{W}{L_h}\right)}$$

Si è soliti porre

$$\alpha_F = \frac{1}{1 + \frac{1}{2} \left( \frac{W}{L_h} \right)^2}$$

Esempio

$W = 4.5 \mu\text{m}$ ,  $L_h = 36.7 \mu\text{m}$ ,  $\alpha_F = 0.99253$ ;

$W = 3.5 \mu\text{m}$ ,  $L_h = 36.7 \mu\text{m}$ ,  $\alpha_F = 0.99547$ :  $\alpha_F$  aumenta al diminuire di  $W$  e tende al valore limite 1.

La relazione

$$I_C = -\alpha_F I_E$$

ci consente di scrivere

$$I_C = -\alpha_F I_{ES} (e^{\frac{V_{EB}}{V_T}} - 1)$$

Le equazioni di Ebers e Moll (EEM) divengono

$$\begin{aligned} I_E &= I_{ES} (e^{\frac{V_{EB}}{V_T}} - 1) + A_{12} (e^{\frac{V_{CB}}{V_T}} - 1) \\ I_C &= -\alpha_F I_{ES} (e^{\frac{V_{EB}}{V_T}} - 1) + A_{22} (e^{\frac{V_{CB}}{V_T}} - 1) \end{aligned}$$

Ripetiamo il procedimento cortocircuitando E con B ( $V_{EB} = 0$ ). Avremo

$$\begin{aligned} I_E &= A_{12} (e^{\frac{V_{CB}}{V_T}} - 1) \\ I_C &= A_{22} (e^{\frac{V_{CB}}{V_T}} - 1) \end{aligned}$$

la seconda delle quali è la caratteristica della giunzione CB quando E è cortocircuitato con B. Quindi

$$I_C = I_{CS} (e^{\frac{V_{CB}}{V_T}} - 1)$$

Definendo

$$\alpha_R = -\frac{I_E}{I_C} \Big|_{V_{EB}=0} = -\frac{A_{12}}{A_{22}}$$

otteniamo

$$\begin{aligned} I_E &= -\alpha_R I_{CS} (e^{\frac{V_{CB}}{V_T}} - 1) \\ I_C &= I_{CS} (e^{\frac{V_{CB}}{V_T}} - 1) \end{aligned}$$

e, dato che il BJT è simmetrico,  $\alpha_R = \alpha_F$  e  $I_{CS} = I_{ES}$ . Ricordiamo ancora una volta che i BJT reali non sono simmetrici.

In conclusione possiamo scrivere

$$\begin{aligned} I_E &= I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1) - \alpha_R I_{CS}(e^{\frac{V_{CB}}{V_T}} - 1) \\ I_C &= -\alpha_F I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1) + I_{CS}(e^{\frac{V_{CB}}{V_T}} - 1) \end{aligned}$$

che costituiscono le equazioni di Ebers-Moll del BJT *pn*p. Queste equazioni e la relazione

$$I_E + I_B + I_C = 0$$

descrivono completamente il comportamento del BJT.

In modo del tutto analogo si ottengono le equazioni di Ebers-Moll per l'*n*pn :

$$\begin{aligned} I_E &= -I_{ES}(e^{\frac{V_{BE}}{V_T}} - 1) + \alpha_R I_{CS}(e^{\frac{V_{BC}}{V_T}} - 1) \\ I_C &= \alpha_F I_{ES}(e^{\frac{V_{BE}}{V_T}} - 1) - I_{CS}(e^{\frac{V_{BC}}{V_T}} - 1) \end{aligned}$$

nelle quali, quando  $V_{BC} = 0$ ,  $I_E$  è negativa (uscite dall'E) se  $V_{BE} > 0$  (giunzione BE polarizzata direttamente). Come regola mnemonica basta ricordare di cambiare i segni sia alle tensioni che alle correnti nelle equazioni valide per il *pn*p.

### Caratteristiche di uscita CB: $I_C = f(V_{CB}; I_E)$

Le caratteristiche di uscita a base comune  $I_C = f(V_{CB}; I_E)$  si ottengono dalla seconda delle equazioni di Ebers-Moll sostituendo al posto di  $I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1)$  l'espressione, ottenuta dalla prima

$$I_E + \alpha_R I_{CS}(e^{\frac{V_{CB}}{V_T}} - 1) = I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1)$$

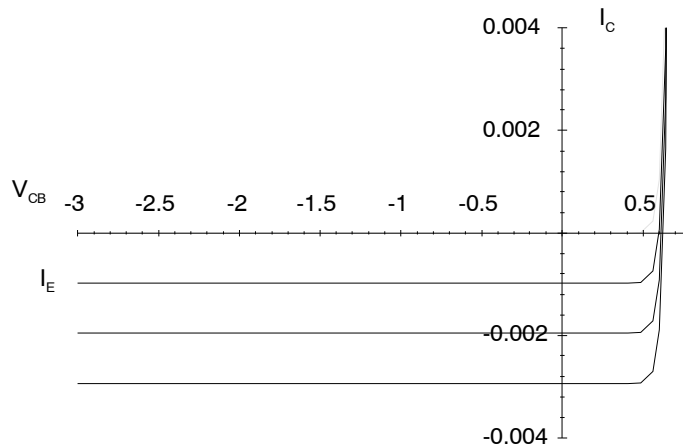
Avremo

$$\begin{aligned} I_C &= -\alpha_F \left[ I_E + \alpha_R I_{CS}(e^{\frac{V_{CB}}{V_T}} - 1) \right] + I_{CS}(e^{\frac{V_{CB}}{V_T}} - 1) \\ &= -\alpha_F I_E + I_{CS}(1 - \alpha_F \alpha_R)(e^{\frac{V_{CB}}{V_T}} - 1) \\ &= -\alpha_F I_E + I_{CBO}(e^{\frac{V_{CB}}{V_T}} - 1) \end{aligned}$$

Quando  $I_E = 0$

$$I_C = I_{CBO}(e^{\frac{V_{CB}}{V_T}} - 1)$$

Imponendo poi valori positivi crescenti alla  $I_E$  con incremento costante otteniamo la famiglia di caratteristiche di uscita della figura.



Nella zona attiva il BJT ( $I_E > 0$ ,  $V_{CB} < 0$ ) si comporta come un generatore di corrente ( $I_C$ ) controllato in corrente ( $I_E$ ).

### L'effetto Early

Nella maggior parte della zona attiva, dato che  $|V_{CB}| \gg V_T$

$$I_C = -\alpha_F I_E - I_{CBO}$$

e quindi, apparentemente,  $I_C$  è indipendente da  $V_{CB}$ .

Bisogna tuttavia tenere conto del fatto che, all'aumentare della polarizzazione inversa della giunzione CB l'ampiezza della zona di svuotamento  $X_{BC}$  aumenta e quindi  $W$  diminuisce. Nel caso presente (BJT  $p^+np^+$  con drogaggi costanti)

$$X_{BC} = \sqrt{\frac{2\varepsilon_s(V_0 - V_{CB})}{qN_D}}$$

Al diminuire di  $W$  la frazione di portatori iniettati che si ricombinano in base diminuisce:  $I_B$  diminuisce e  $I_C$  aumenta. Poiché  $I_E$  è costante (imposta dal generatore di corrente in ingresso)  $\alpha_F$  aumenta all'aumentare in modulo di  $V_{CB}$ . Nella zona attiva

$$I_C = -\alpha_F(V_{CB})I_E - I_{CBO}$$

Il fenomeno descritto prende nome di effetto Early.

Consideriamo ora quello che accade quando  $V_{CB} > 0$ . La giunzione CB è polarizzata direttamente e invece di estrarre portatori li inietta anch'essa nella base: la corrente di collettore è data dalla differenza

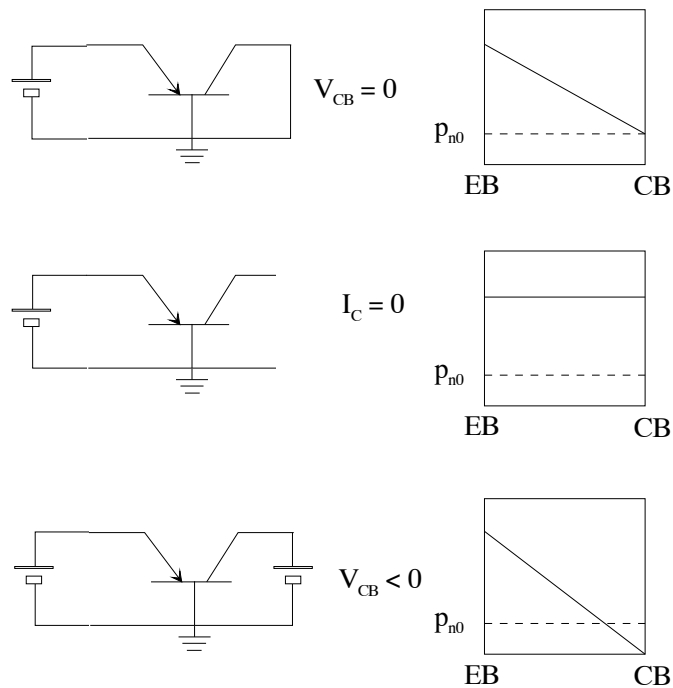
$$I_C = -\alpha_F I_E + I_{CBO} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right)$$

Inizialmente tende a diminuire in valore assoluto, si annulla per

$$V_{CB} = V_T \ln \left( \frac{\alpha_F I_E + I_{CBO}}{I_{CBO}} \right)$$

e infine cambia di segno. Quando entrambe le giunzioni sono polarizzate direttamente si dice che il BJT si trova in condizioni di saturazione.

Il profilo dei minoritari è mostrato in tre casi significativi  $V_{CB} = 0$ ,  $I_C = 0$ ,  $V_{CB} < 0$ ; in tutti  $V_{EB} > 0$ .



Il solo caso  $I_C = 0$  merita qualche considerazione. La corrente di collettore è nulla. D'altra parte la corrente di diffusione attraverso la base, per cui  $V_{CB}$  deve essere maggiore di zero per annullare  $I_C$ . Dato poi che  $I_E \simeq I_C$  è evidente che  $V_{CB} \simeq V_{EB}$ . Si noti tuttavia che mentre la derivata del profilo è nulla in  $x = W$ , non lo può ovviamente essere in  $x = 0$  perché  $I_E$  è diversa da zero. Si ricordi a questo proposito che il profilo lineare costituisce un'approssimazione che non sempre può essere utilizzata. Il caso  $I_C = 0$  poteva essere trattato anche partendo dalle EEM.

A conclusione notiamo che le prime due configurazioni costituiscono dei diodi ( $I_E$ ,  $V_{EB}$ ), con caratteristiche diverse a seconda dei collegamenti. In effetti nei processi bipolari i diodi necessari sono ottenuti dai BJT, collegando fra loro due terminali.



**Caratteristiche di ingresso a base comune:**  $I_E = f(V_{EB}; V_{CB})$ 

È immediato ricavarle dalla prima delle equazioni di Ebers-Moll

$$I_E = I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1) - \alpha_R I_{CS}(e^{\frac{V_{CB}}{V_T}} - 1)$$

Quando  $V_{CB} = 0$  abbiamo, come già visto, la caratteristica della giunzione EB. Per  $V_{CB} < 0$

$$I_E = I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1) + \alpha_R I_{CS}$$

Il BJT è simmetrico:  $\alpha_R = \alpha_F$  e  $I_{CS} = I_{ES}$  e quindi

$$I_E = I_{ES}(e^{\frac{V_{EB}}{V_T}} - 1) + \alpha_F(V_{CB}) I_{ES} :$$

quando  $V_{CB} < 0$  la caratteristica non passa per l'origine.

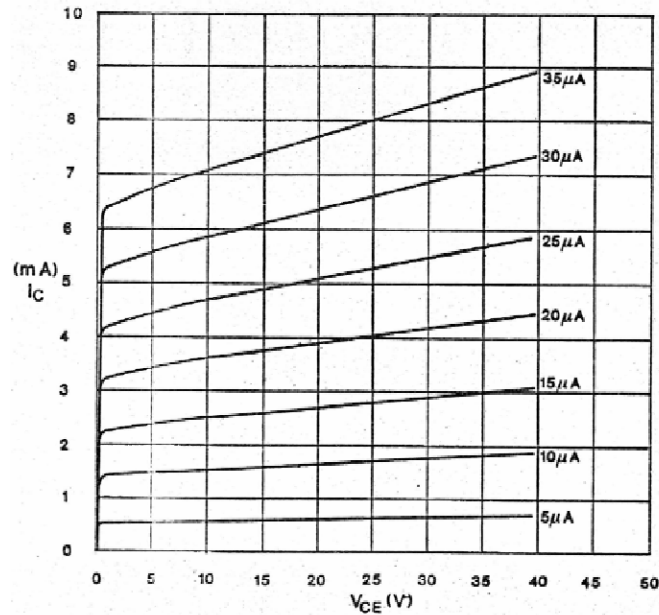
►NB: si può dimostrare che per ogni tipo di BJT vale l'uguaglianza

$$\alpha_R I_{CS} = \alpha_F I_{ES}$$

indipendentemente dal fatto che sia simmetrico o meno. La relazione prende il nome di *relazione di reciprocità* del BJT.

**Caratteristiche di uscita a CE:**  $I_C = g(V_{CE}; I_B)$ 

Nella configurazione a base comune la variabile di controllo ( $I_E$ ) ha un valore dello stesso ordine della variabile controllata ( $I_C$ ). Se si sceglie come terminale a cui riferire le tensioni l'emettitore, la variabile di controllo in ingresso diviene  $I_B$ , che ha un valore molto minore della variabile controllata e le caratteristiche di uscita a emettitore comune (CE) presentano l'andamento mostrato nella figura (*nnp BC 107*).



Il problema di ricavare le caratteristiche di uscita CE è un pò più complicato del caso CB. Il punto di partenza sono le caratteristiche di uscita CB

$$I_C = -\alpha_F I_E + I_{CBO} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right)$$

da cui, tenuto conto che è  $I_E = -I_C - I_B$ , si ottiene

$$\begin{aligned} I_C &= -\alpha_F (-I_C - I_B) + I_{CBO} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right) \\ &= \alpha_F I_C + \alpha_F I_B + I_{CBO} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right) \end{aligned}$$

e ancora ( $V_{CB} = V_{EB} + V_{CE}$ )

$$\begin{aligned} I_C (1 - \alpha_F) &= \alpha_F I_B + I_{CBO} \left( e^{\frac{V_{EB} + V_{CE}}{V_T}} - 1 \right) \\ I_C &= \frac{\alpha_F}{1 - \alpha_F} I_B + \frac{I_{CBO}}{1 - \alpha_F} \left( e^{\frac{V_{EB} + V_{CE}}{V_T}} - 1 \right) \\ I_C &= \beta_F I_B + (\beta_F + 1) I_{CBO} \left( e^{\frac{V_{EB} + V_{CE}}{V_T}} - 1 \right) \end{aligned}$$

In zona attiva la giunzione CB è polarizzata inversamente ( $V_{CE} < 0$ ) e quella BE direttamente ( $V_{EB} > 0$ ). Dalla relazione

$$V_{CB} = V_{EB} + V_{CE}$$

si deduce che affinché  $V_{CB} < 0$  è necessario che  $V_{CE}$  sia negativa e in valore assoluto maggiore di  $V_{EB}$ . Se poi la somma  $V_{EB} + V_{CE}$  è in valore assoluto  $\gg V_T$  avremo

$$I_C = \beta_F I_B - (\beta_F + 1) I_{CBO} = \beta_F I_B - (\beta_F + 1) I_{CBO}$$

Nella zona attiva, se si trascura il termine  $(\beta_F + 1) I_{CBO} = I_{CEO}$ ,

$$I_C \simeq \beta_F I_B :$$

il BJT si comporta come un generatore di corrente  $I_C$  controllato dalla corrente  $I_B$ .

►  $\beta_F = \frac{I_C}{I_B}$ , calcolato ad una fissata  $V_{CE}$  è il *guadagno di corrente a emettitore comune*.

Poichè  $\alpha_F$  è molto prossimo ad uno,  $\beta_F$  assume valori dell'ordine di alcune centinaia.

$\beta_F$  è inoltre molto sensibile alle variazioni di  $\alpha_F$ : l'effetto Early

$$I_C \simeq \beta_F(V_{CE})I_B$$

è molto più pronunciato nella configurazione CE.

In generale si deve eliminare la dipendenza da  $V_{EB}$  delle caratteristiche

$$I_C = \beta_F I_B + I_{CEO} \left( e^{\frac{V_{EB}}{V_T}} e^{\frac{V_{CE}}{V_T}} - 1 \right). \quad (\text{CE1})$$

Dato che  $I_B = -I_E - I_C$ , dalle EEM si ha:

$$I_B = -I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) + \alpha_R I_{CS} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right) + \alpha_F I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) - I_{CS} \left( e^{\frac{V_{CB}}{V_T}} - 1 \right)$$

$$I_B = -I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) + \alpha_R I_{CS} \left( e^{\frac{V_{EB}}{V_T}} e^{\frac{V_{CE}}{V_T}} - 1 \right) + \alpha_F I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) - I_{CS} \left( e^{\frac{V_{EB}}{V_T}} e^{\frac{V_{CE}}{V_T}} - 1 \right)$$

da cui si può ottenere il termine  $e^{\frac{V_{EB}}{V_T}}$ .

$$I_B = e^{\frac{V_{EB}}{V_T}} \left( -I_{ES} + \alpha_R I_{CS} e^{\frac{V_{CE}}{V_T}} + \alpha_F I_{ES} - I_{CS} e^{\frac{V_{CE}}{V_T}} \right) + I_{ES} - \alpha_R I_{CS} - \alpha_F I_{ES} + I_{CS}$$

$$e^{\frac{V_{EB}}{V_T}} = \frac{I_B + I_{ES} (\alpha_F - 1) + I_{CS} (\alpha_R - 1)}{\left( I_{ES} (\alpha_F - 1) + I_{CS} e^{\frac{V_{CE}}{V_T}} (\alpha_R - 1) \right)}$$

Sostituendo nella CE1 si ottiene l'espressione cercata della  $I_C$ , che dipende solo da  $I_B$  e da  $V_{CE}$ .

$$I_C = \beta_F I_B + I_{CEO} \left( e^{\frac{V_{CE}}{V_T}} \left( \frac{I_B + I_{ES} (\alpha_F - 1) + I_{CS} (\alpha_R - 1)}{I_{ES} (\alpha_F - 1) + I_{CS} e^{\frac{V_{CE}}{V_T}} (\alpha_R - 1)} \right) - 1 \right)$$

Nel caso simmetrico

$$I_C = \beta_F I_B + I_{CEO} \left( e^{\frac{V_{CE}}{V_T}} \left( \frac{I_B + 2I_{ES}(\alpha_F - 1)}{I_{ES}(\alpha_F - 1) \left( e^{\frac{V_{CE}}{V_T}} + 1 \right)} \right) - 1 \right).$$

In zona attiva

$$I_C = \beta_F I_B - I_{CEO} \simeq \beta_F I_B$$

Quando  $V_{CE} = 0$  si ha

$$I_C = \beta_F I_B + I_{CEO} \left( \left( \frac{I_B + 2I_{ES}(\alpha_F - 1)}{2I_{ES}(\alpha_F - 1)} \right) - 1 \right) \quad (\text{CE2})$$

e quindi le caratteristiche non passano per l'origine. Si verifichi che nel caso simmetrico  $I_C = -I_B/2$ .

Esempio

Si definisce la struttura del BJT assegnando drogaggi, dimensioni geometriche, tempo di vita dei minoritari:  $N_A = 10^{19} \text{ cm}^{-3}$ ,  $N_D = 10^{16} \text{ cm}^{-3}$ , etc.

$$\beta_F = 2 \left( \frac{L_h}{W} \right)^2$$

con

$$W = W^* - X_{EB} - X_{BC} = W^* - \sqrt{\frac{2\varepsilon_s(V_0 - V_{EB})}{qN_D}} - \sqrt{\frac{2\varepsilon_s(V_0 - (V_{EB} + V_{CE}))}{qN_D}}$$

in cui, con  $I_B$  fissata,  $V_{EB}$  è costante.

$$W = W_K^* - \sqrt{\frac{2\varepsilon_s}{qN_D}} (V_0 - (V_{EB} + V_{CE}))^{\frac{1}{2}}.$$

$\beta_F$  si scrive quindi

$$\beta_F = \frac{2L_h^2}{\left( W_K^* - \sqrt{\frac{2\varepsilon_s}{qN_D}} (V_0 - (V_{EB} + V_{CE}))^{\frac{1}{2}} \right)^2}$$

Per graficare le caratteristiche scriviamo

$$I_C = \beta_F I_B + I_{CEO} \left( e^{\frac{V_{CE}}{V_T}} \left( \frac{I_B + 2I_{ES}(\alpha_F - 1)}{I_{ES}(\alpha_F - 1) \left( e^{\frac{V_{CE}}{V_T}} + 1 \right)} \right) - 1 \right)$$

in cui si può trascurare  $2I_{ES}(\alpha_F - 1)$  rispetto a  $I_B$ .

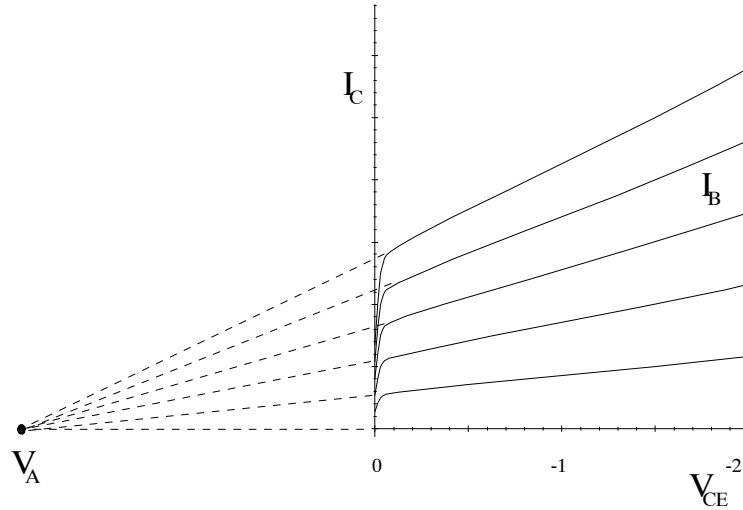
$$I_C = \beta_F I_B + I_{CEO} \left( e^{\frac{V_{CE}}{V_T}} \left( \frac{I_B}{I_{ES}(\alpha_F - 1) \left( e^{\frac{V_{CE}}{V_T}} + 1 \right)} \right) - 1 \right)$$

Dato che  $\beta_F + 1 \simeq \beta_F$

$$I_C = \beta_F I_B + \beta_F I_{ES} (1 - \alpha_F^2) \left( e^{\frac{V_{CE}}{V_T}} \left( \frac{I_B}{I_{ES}(\alpha_F - 1) \left( e^{\frac{V_{CE}}{V_T}} + 1 \right)} \right) - 1 \right)$$

$$I_C = \frac{2L_h^2}{\left( W_K^* - \sqrt{\frac{2\epsilon_s}{qN_D}} (V_0 - (V_{EB} + V_{CE}))^{\frac{1}{2}} \right)^2} \times \left( I_B + I_{ES} (1 - \alpha_F^2) \left( e^{\frac{V_{CE}}{V_T}} \left( \frac{I_B}{I_{ES}(\alpha_F - 1) \left( e^{\frac{V_{CE}}{V_T}} + 1 \right)} \right) - 1 \right) \right)$$

Assegnando ad  $I_B$  valori crescenti in valore assoluto ( $I_B$  è negativa) si ottengono le caratteristiche della figura.



$I_C$  è negativa. Come si vede l'effetto Early è molto evidente; il prolungamento delle caratteristiche incontra l'asse delle  $V_{CE}$  in un punto  $V_A$  che prende il nome di tensione di Early.

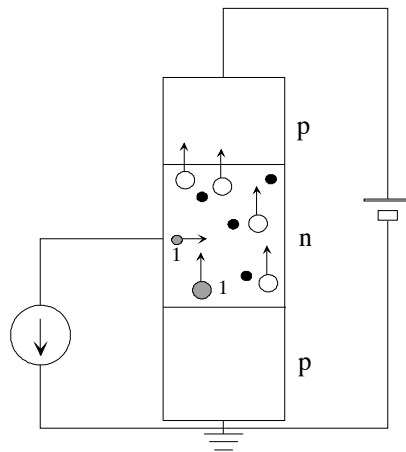
La corrente di collettore comincia a diminuire in valore assoluto nell'intorno di  $V_{CE} \simeq -0.2 \div -0.3$  V. Infatti dato che

$$V_{CE} = V_{BE} + V_{CB}$$

e  $V_{EB} \simeq 0.6$  volt (imposti dal generatore  $I_B$ ), quando  $V_{CE}$  diventa dell'ordine di  $-0.6$  V si avrà  $-0.6 \simeq -0.6 + V_{CB}$  e quindi la giunzione CB risulterà, per valori di  $V_{CE}$  prossimi a zero, polarizzata direttamente: il BJT entra in zona di saturazione e la corrente di collettore diminuisce per effetto dell'iniezione di corrente dovuta alla giunzione CB.

### Interpretazione con il modello a controllo di carica

Nella configurazione CB la corrente di base costituisce un effetto, mentre in quella CE la  $I_B$  costituisce la variabile di controllo ed è quindi una causa. Con riferimento alla figura,



in una situazione di stazionarietà la base è neutra. Un aumento  $\Delta I_B$  (ad esempio: ingresso in base di un elettrone in più) provoca una condizione di non neutralità che viene ripristinata mediante l'iniezione di una lacuna da parte dell'emettitore. Questa lacuna attraversa la base per diffusione in un tempo  $\tau_t = \frac{W^2}{2D_h}$ , che è molto minore del tempo di vita medio  $\tau_h$ .

Esempio

Se  $W = 2 \mu\text{m}$  e  $D_h = 0.038 \times 0.026 = 9.88 \times 10^{-5} \text{ m}^2\text{s}^{-1}$

$$\tau_t = \frac{(2 \times 10^{-6})^2}{2 \times 9.88 \times 10^{-5}} = 2.0 \times 10^{-8} \text{ s}$$

Prima che una delle lacune iniettate in conseguenza di  $\Delta I_B$  si ricombini deve trascorrere un tempo  $\tau_h$ , durante il quale vengono iniettate  $\tau_h/\tau_t$  lacune, che sono in grado di essere raccolte dal collettore.  $I_C$  risulta amplificata di un fattore  $\tau_h/\tau_t$  e quindi

$$I_C = \frac{\tau_h}{\tau_t} I_B = \beta_F I_B$$

Ricordando che

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F}$$

e che

$$\alpha_F = \frac{1}{1 + \frac{1}{2} \left( \frac{W}{L_h} \right)^2}$$

si ha

$$\beta_F = 2 \left( \frac{L_h}{W} \right)^2 = \frac{2L_h^2}{2D_h\tau_t} = \frac{\tau_h}{\tau_t}.$$

Se, nell'esempio precedente,  $\tau_h = 10^{-6}$  s,  $\beta_F = \frac{10^{-6}}{2.0 \times 10^{-8}} = 50$ .

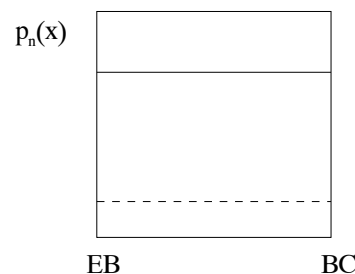
### Caratteristiche di ingresso a emettitore comune:

$$I_B = f(V_{BE}; V_{CE})$$

Prendiamo in esame per i terminali di uscita le due condizioni : i)  $V_{CE} = 0$ ,  
ii)  $V_{CE} < 0$ , facendo riferimento al profilo dei minoritari nella base.

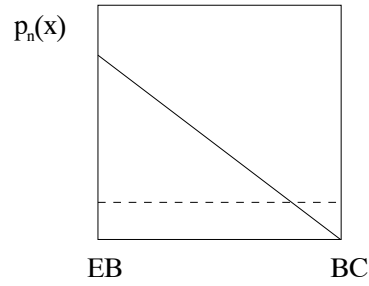
i)  $V_{CE} = 0$ .

In questo caso  $V_{CB} = V_{EB}$  e quindi il profilo è quello seguente



la corrente  $I_B$  si calcola con il modello del controllo di carica. Si noti che le derivate in 0 e  $W$  non possono essere nulle.

ii)  $V_{CE} < 0$ .



Se  $V_{CE}$  è tale che anche  $V_{CB} < 0$ , il profilo dà evidentemente luogo ad una corrente di base che è la metà di quella precedente.

### Effetto del campo nella base

In un BJT integrato il drogaggio nella base non è costante. Al profilo dei droganti è associato (all'equilibrio) un campo elettrico dovuto ai portatori *maggioritari*, il cui andamento si trova imponendo, come al solito, che la densità di corrente totale dei *maggioritari* sia nulla. Facciamo riferimento a un *npn* e schematizziamo il profilo dei droganti nella base con un esponenziale

$$N_A(x) = N_A(0)e^{-\frac{a}{W}x}$$

decescente; il parametro  $a$  è compreso in questo esempio fra 0.01 e 7 e quantifica quanto rapidamente varia il drogaggio fra 0 e  $W$ .

Il campo si ottiene da

$$J_h = q\mu_h p(x)\mathcal{E} - qD_h \frac{dp(x)}{dx} = 0$$

e, con l'approssimazione di quasi-neutralità

$$q\mu_h N_A(x)\mathcal{E} - qD_h \frac{dN_A(x)}{dx} = 0$$

$$\mathcal{E} = -\frac{aV_T}{W}$$

che tende a zero per  $a \rightarrow 0$ . Il campo è orientato da C a B e quindi favorisce il trasporto dei minoritari (elettroni) attraverso la base. La densità di corrente  $J_n$  nella base sarà data da

$$J_n = q\mu_n n(x)\mathcal{E} + qD_n \frac{dn(x)}{dx} = -q\mu_n n(x) \frac{aV_T}{W} + qD_n \frac{dn(x)}{dx}$$

che può essere ritenuta costante se si trascura la ricombinazione. Si ottiene l'equazione

$$J_n = -q\mu_n n(x) \frac{aV_T}{W} + qD_n \frac{dn(x)}{dx} = qD_n \left( \frac{dn(x)}{dx} - n(x) \frac{a}{W} \right)$$



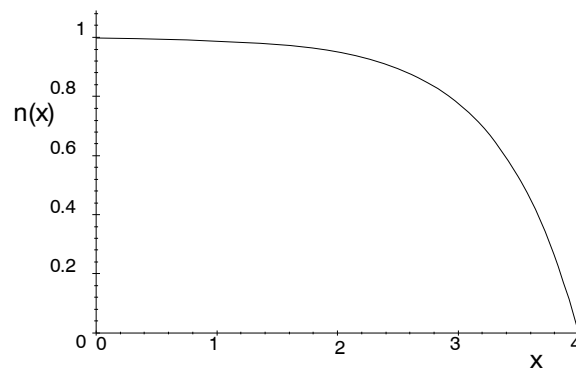
che ha per soluzione

$$n(x) = Ae^{\frac{ax}{W}} - \frac{J_n}{qD_n} \frac{W}{a}.$$

Per determinare la costante  $A$  sappiamo che in zona attiva  $n(W) = 0$  da cui segue

$$n(x) = -\frac{J_n}{qD_n} \frac{W}{a} (1 - e^{-a} e^{\frac{ax}{W}}).$$

Un esempio di profilo ( $W = 4 \mu\text{m}$  e  $a = 6$ ) è mostrato nella figura, in cui la scala delle ordinate è in unità arbitrarie. Come si vede nel primo tratto della base la concentrazione dei portatori minoritari è quasi costante e quindi la componente diffusiva della corrente molto piccola.



Più precisamente: sappiamo che la densità di corrente di diffusione è data da

$$J_{ndiff} = qD_n \frac{dn(x)}{dx}$$

per cui possiamo valutarla, dato che il profilo è noto, e confrontarla con quella totale  $J_n$ . Avremo

$$J_{ndiff} = J_n e^{-a} e^{\frac{ax}{W}}$$

e

$$\frac{J_{ndiff}}{J_n} = e^{-a} e^{\frac{ax}{W}}.$$

1)  $a = 0$  (drogaggio costante):

$$\frac{J_{ndiff}}{J_n} = 1$$

la densità di corrente è tutta di diffusione, come era da attendersi.

2)  $a \neq 0$ ;

2i) in  $x = 0$

$$\frac{J_{ndiff}}{J_n} = e^{-a}$$

ovvero per  $a$  grande la diffusione è trascurabile; il trasporto avviene prevalentemente per drift.

2ii) in  $x = W$

$$\frac{J_{ndiff}}{J_n} = 1 \quad \forall a :$$

la densità di corrente è tutta di diffusione.

Il tempo di transito si trova scrivendo che

$$J_n = qn(x)v(x)$$

e quindi

$$v(x) = \frac{dx}{dt} = \frac{J_n}{qn(x)} = \frac{D_n a}{W (1 - e^{-a} e^{\frac{ax}{W}})};$$

integrando fra 0 e  $W$  e fra 0 e  $\tau_t$

$$\int_0^W \frac{W (1 - e^{-a} e^{\frac{ax}{W}}) dx}{D_n a} = \int_0^{\tau_t} dt$$

si ottiene

$$\tau_t = \frac{W (W - \frac{W}{a} e^{-a} (e^a - 1))}{D_n a}$$

che, per  $a$  grande diviene

$$\tau_t = \frac{(W^2 - \frac{W^2}{a})}{D_n a}.$$

Nel caso di drogaggio costante, come abbiamo già visto nell'esempio precedente

$$\tau_t = \frac{W^2}{2D_n} = \frac{(2 \times 10^{-6})^2}{2 \times 9.88 \times 10^{-5}} = 2.0 \times 10^{-8} \text{ s.}$$

( $W = 2 \mu\text{m}$  e  $D_n = 0.038 \times 0.026 = 9.88 \times 10^{-5} \text{ m}^2\text{s}^{-1}$ ).

Se invece, ad esempio,  $a = 7$  avremo

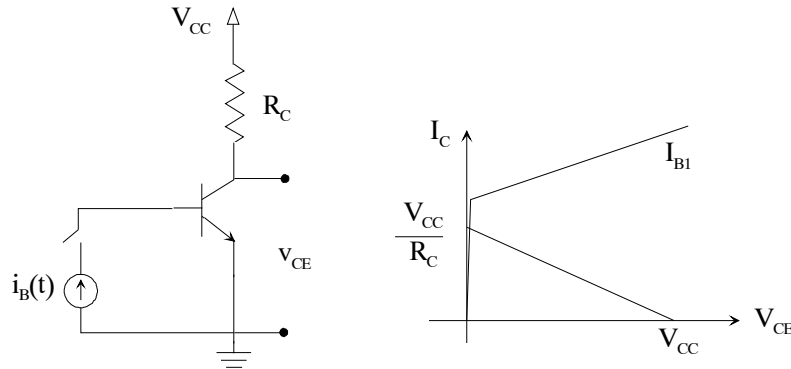
$$\tau_t' = \frac{\left( (2 \times 10^{-6})^2 - \frac{(2 \times 10^{-6})^2}{7} \right)}{9.88 \times 10^{-5} \times 7} = 4.96 \times 10^{-9} \text{ s;}$$

a parità di  $\tau_n$  il rapporto fra i  $\beta_F$  è

$$\frac{\beta_F'}{\beta_F} = \frac{2.0 \times 10^{-8}}{4.96 \times 10^{-9}} = 4.$$

## La commutazione del BJT

Con riferimento al circuito di figura,



si vuole studiare l'andamento temporale della  $V_{out} = V_{CE}$  quando il generatore di corrente  $i_B(t)$  impone una brusca variazione della corrente di base. I due valori di  $V_{CE}$  a regime, per  $I_B = 0$  (interruttore aperto) e per  $I_B = I_{B1}$  (interruttore chiuso) sono immediatamente ricavabili dalle caratteristiche alle quali è stata sovrapposta la retta di carico:  $I_B = 0, V_{CE} = V_{CC}$ ;  $I_B = I_{B1}, V_{CE} < V_{CESAT}$ .

Per quel che concerne il transitorio si deve ricorrere al modello del controllo di carica.

Per un *npn* in zona attiva *diretta*:

$$I_B = \frac{Q_F}{\tau_n}, \quad I_C = \frac{Q_F}{\tau_t}$$

da cui

$$I_E = -\frac{Q_F}{\tau_n} - \frac{Q_F}{\tau_t};$$

in zona attiva *inversa* (E scambiato con C):

$$I_B = \frac{Q_R}{\tau_n}, \quad I_E = \frac{Q_R}{\tau_t'}$$

e quindi

$$I_C = -\frac{Q_R}{\tau_n} - \frac{Q_R}{\tau_t'}.$$

Sovrapponendo gli effetti per una condizione qualsiasi:

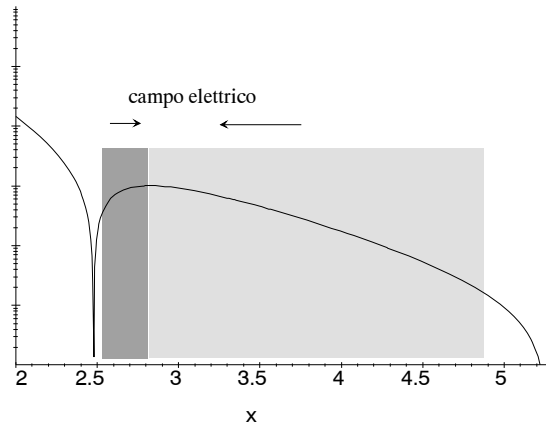
$$I_B = \frac{Q_F}{\tau_n} + \frac{Q_R}{\tau_n}$$

e

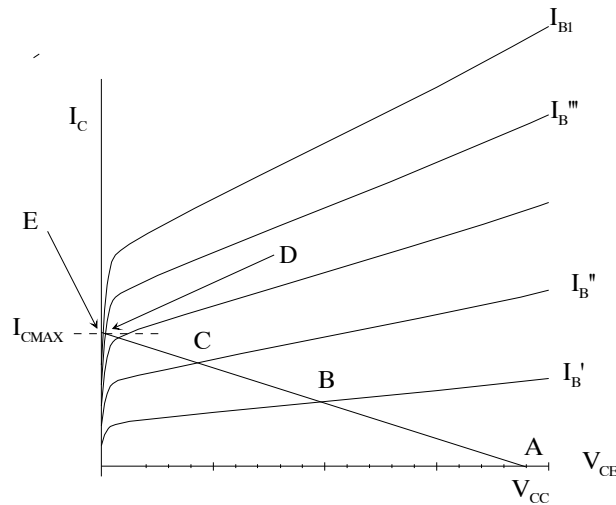
$$I_C = \frac{Q_F}{\tau_t} - \frac{Q_R}{\tau_n} - \frac{Q_R}{\tau_t'}.$$

Queste due ultime relazioni legano fra loro gli eccessi di carica in base e le correnti  $I_B$  e  $I_C$ .

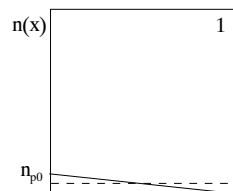
Si noti che nella zona attiva inversa il tempo di transito è  $\tau_t'$ ; in effetti il profilo di drogaggio in base è asimmetrico e quindi diversi sono gli effetti del campo sui minoritari a seconda che questi vengano iniettati dall'E o dal C. La figura, che rappresenta il profilo  $N_A(x)$  nella base, chiarisce quanto detto.



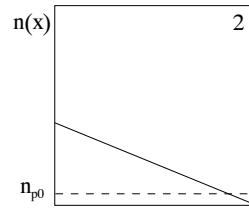
Seguendo il punto di lavoro statico lungo la retta di carico si hanno le condizioni notevoli indicate dai punti A, B, C, D, E.



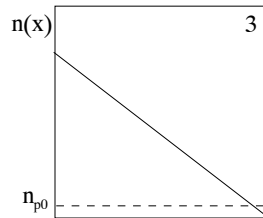
A)  $I_B = 0$ ; il BJT è interdettato. Il profilo dei minoritari è quello della figura 1; l'eccesso in base è nullo.



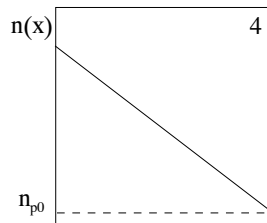
B)  $I_B = I'_B$ ; zona attiva (figura 2);



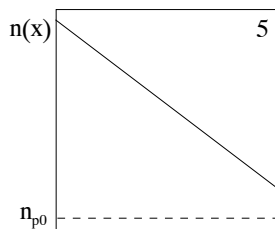
C)  $I_B = I''_B$ ; zona attiva (figura 3);



D)  $I_B = I'''_B$ ; inizio saturazione (figura 4);



D)  $I_B = I_{B1}$ ; saturazione profonda (figura 5);



la corrente di collettore è fissata al valore  $I_{C\text{MAX}}$  dalla retta di carico

$$I_{C\text{MAX}} \simeq \frac{V_{CC}}{R_C}$$

e non può più aumentare qualunque sia il valore di  $I_B$ ; aumenta tuttavia l'eccesso di carica in base, come risulta dalla figura 5. Dalla situazione indicata in figura 4, il segmento di retta che rappresenta il profilo si sposterà parallelamente a se stesso ( $I_C$  costante,  $I_B$  che aumenta).

Conoscere il transitorio della  $V_{CE}$  significa conoscere l'andamento temporale della  $I_C$ ; infatti

$$v_{CE}(t) = V_{CC} - R_C i_C(t).$$

Si deve utilizzare l'equazione di continuità per la corrente di base

$$i_B(t) = \frac{Q_B}{\tau_n} + \frac{dQ_B}{dt}$$

in cui  $Q_B = Q_F + Q_R$ . In effetti anche quando  $I_C$  ha raggiunto il suo limite  $I_{CMAX}$  l'eccesso nella base continua ad aumentare fino al valore  $I_{B1}\tau_n$ . Sappiamo che la soluzione del transitorio di accensione è

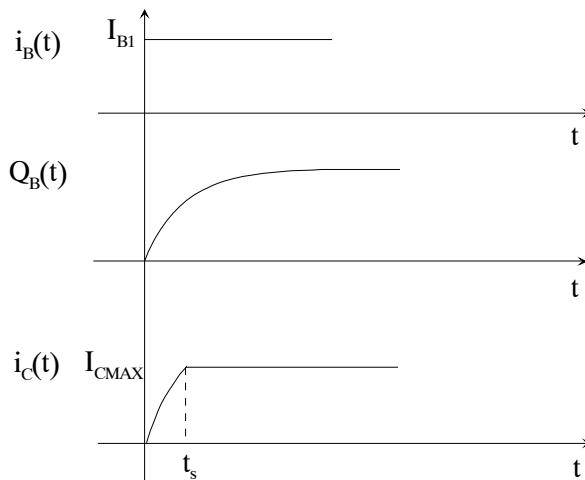
$$Q_B(t) = I_{B1}\tau_n \left(1 - e^{-\frac{t}{\tau_n}}\right)$$

e quindi la corrente di collettore sarà data da

$$i_C(t) = \frac{Q_B(t)}{\tau_t} = \frac{Q_F(t)}{\tau_t}$$

in quanto, dopo l'instaurarsi della condizione di saturazione, anche se  $Q_B(t)$  continua a crescere, l'eccesso che riguarda  $i_C(t)$  è appunto solo quello fino alla saturazione.

La figura mostra l'andamento temporale del generatore di ingresso, della  $Q_B(t) = Q_F(t) + Q_R(t)$  e della  $i_C(t)$  che, per quanto detto, è dato da  $i_C(t) = \frac{Q_F(t)}{\tau_t}$ .



Il tempo  $t_s$  che impiega la corrente di collettore ad andare a regime si trova imponendo che

$$i_C(t_s) = \frac{I_{B1}\tau_n \left(1 - e^{-\frac{t_s}{\tau_n}}\right)}{\tau_t} = I_{CMAX}$$

da cui

$$t_s = \tau_n \ln \left( \frac{1}{1 - \frac{I_{CMAX}}{I_{B1}\beta_F}} \right)$$

in cui si è utilizzata la relazione  $\beta_F = \frac{\tau_n}{\tau_t}$ .

Per ridurre  $t_s$  :

- i) si può ridurre  $\tau_n$ ;
- ii) utilizzare un BJT con un  $\beta_F$  maggiore;
- iii) fare in modo che sia  $I_{B1}\beta_F \gg I_{CMAX}$ , ovvero sovrasaturare la base.

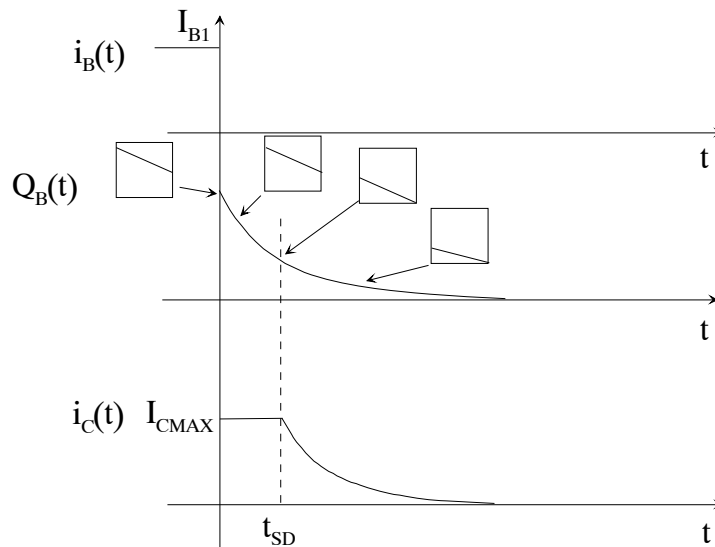
Il fatto che la base sia in saturazione di carica dopo il transitorio di accensione influisce, com'è ovvio, sul successivo transitorio di spegnimento, per il quale sono possibili due casi.

- 1) La corrente di base commuta a zero.

Il transitorio della  $Q_B(t)$

$$Q_B(t) = I_{B1}\tau_n e^{-\frac{t}{\tau_n}}$$

è mostrato nella figura. Si deve osservare tuttavia che, prima che la corrente di collettore possa cominciare a diminuire, è necessario che la base esca dalla saturazione e si riporti all'inizio della zona attiva. Durante questo intervallo di tempo  $I_C$  resta costante al suo valore  $I_{CMAX}$ . L'intervallo  $t_{SD}$  è lo *storage delay time*.



$t_{SD}$  è determinato dalla relazione

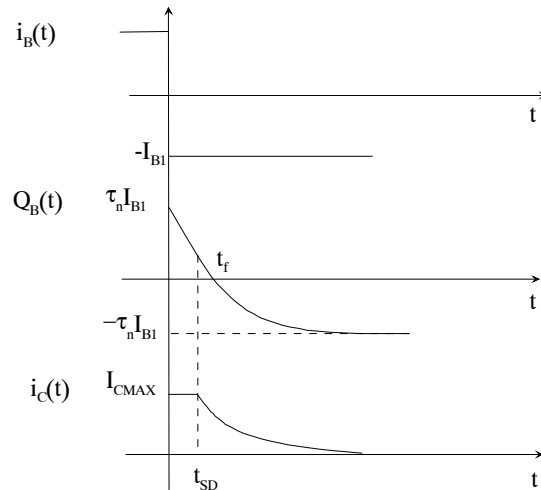
$$I_{CMAX}\tau_t = I_{B1}\tau_n e^{-\frac{t_{SD}}{\tau_n}}$$

segue

$$t_{SD} = \tau_n \ln \left( \frac{I_{B1}\beta_F}{I_{CMAX}} \right);$$

come era da aspettarsi la sovrasaturazione della base, di cui al punto iii), peggiora le cose durante il transitorio di spegnimento.

2) La corrente di base commuta ad un valore negativo (ad esempio  $-I_{B1}$ ); di conseguenza l'eccesso in base tende *virtualmente* ad un eccesso negativo di pari entità; il transitorio della carica in base si esaurisce però all'istante  $t = t_f$  e il BJT esce dalla saturazione in un tempo  $t_{SD}$  minore che nel caso precedente.



La soluzione dell'equazione di continuità

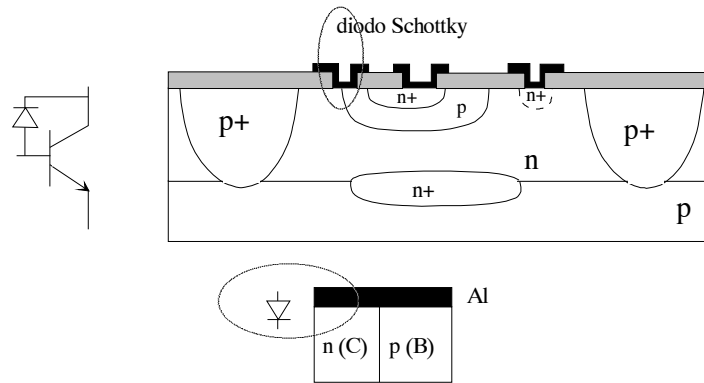
$$Q_B(t) = I_{B1} \tau_n \left( 2e^{-\frac{t}{\tau_n}} - 1 \right)$$

si trova imponendo come condizioni al contorno:

$$\begin{aligned} t &= 0, & Q_B &= I_{B1} \tau_n \\ t &= \infty, & Q_B &= -I_{B1} \tau_n. \end{aligned}$$

Al fine di prevenire la saturazione della base si può mettere in parallelo fra base e collettore un diodo Schottky disposto nello stesso verso. Poiché la  $V_\gamma$  dello Schottky è minore di quella della giunzione BC, la tensione  $V_{BC}$  risulta bloccata a questo valore (quando il BJT entra in regione di saturazione), impedendo in tal modo la formazione di un'eccessiva carica nella base. Il dispositivo, di cui è mostrata la realizzazione integrata, prende il nome di BJT Schottky.





### Il breakdown nei bipolari

Due sono le tensioni di breakdown nei BJT: la  $BV_{CBO}$  e la  $BV_{CEO}$ . La prima riguarda la configurazione a base comune quando l'emettitore è *Open*. Si tratta evidentemente di un diodo (caratteristica CB con  $I_E = 0$ ) e la  $BV_{CBO}$  determina il fattore di moltiplicazione a valanga

$$M = \frac{1}{1 - \left(\frac{V_{CB}}{BV_{CBO}}\right)^n}$$

nello stesso modo già visto per il diodo; non ci sono portatori iniettati dall'E.

Nel caso della configurazione CE la  $BV_{CEO}$  riguarda la caratteristica con la base *Open*.

In questo caso  $I_C = -I_E$  e i portatori che arrivano al C sono quelli che provengono dall'E. Dalle equazioni di Ebers e Moll (*pn*p simmetrico), tenendo conto che la giunzione CB è polarizzata inversamente

$$\begin{aligned} I_E &= I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) + \alpha_R I_{CS} \\ I_C &= -\alpha_F I_{ES} \left( e^{\frac{V_{EB}}{V_T}} - 1 \right) - I_{CS} \end{aligned}$$

si ottiene

$$\begin{aligned} I_C &= -\alpha_F (I_E - \alpha_R I_{CS}) - I_{CS} \\ I_C &= -\alpha_F I_E + I_{CS} (\alpha_F \alpha_R - 1) \end{aligned}$$

e, poiché  $I_C = -I_E$

$$I = \alpha_F I + I_{CS} (\alpha_F \alpha_R - 1).$$

Quando la tensione è tale da innescare il BD si avrà

$$\begin{aligned} I_C &= M \alpha_F I_C + M I_{CS} (\alpha_F \alpha_R - 1) \\ I_C (1 - M \alpha_F) &= M I_{CS} (\alpha_F \alpha_R - 1) \\ I_C &= \frac{M I_{CS} (\alpha_F \alpha_R - 1)}{1 - M \alpha_F} \end{aligned}$$

che diverge per  $M \alpha_F \rightarrow 1$ , con

$$M = \frac{1}{1 - \left( \frac{V_{CB}}{BV_{CBO}} \right)^n}$$

da cui si ottiene la  $V_{CB}$  per la quale questa condizione è verificata

$$\alpha_F = 1 - \left( \frac{V_{CB}}{BV_{CBO}} \right)^n$$

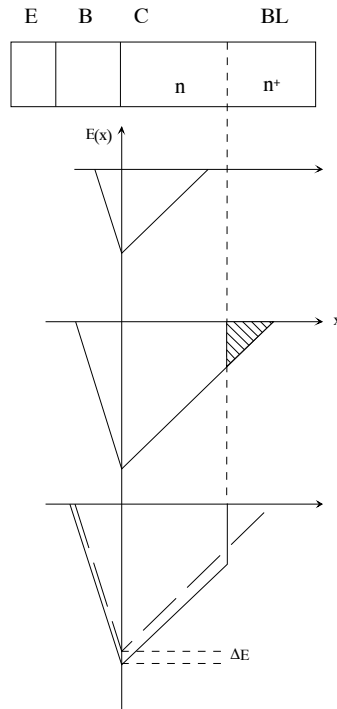
$$V_{CB} = BV_{CBO} \sqrt[1 - \alpha_F]{1 - \alpha_F}.$$

Questa  $V_{CB}$  di BD non è la  $BV_{CEO}$ , ma ne differisce soltanto per la  $V_{EB}$  che è dell'ordine di 0.7 V.

Se per esempio  $BV_{CBO} = 30$  V risulta che, se  $\alpha_F = 0.96$ ,  $BV_{CEO} = \sqrt{1 - 0.96} \times 30 = 6.0$  V. Questo può sembrare strano, ma ha la sua spiegazione nella configurazione CE nella quale ogni carica nella base, comunque prodotta, viene amplificata.

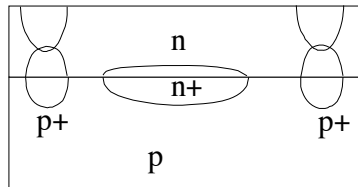
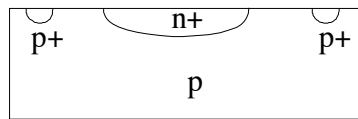
### Il reach-through e i processi high voltage

La struttura della figura rappresenta schematicamente E, B, C e strato sepolto  $n^+$  di un  $nnp$  integrato. La tensione inversa  $V_{BC}$  provoca un'estensione della zona di svuotamento nella base (effetto Early), ma anche nel collettore. L'andamento del campo è quello della figura.



Se lo strato epi avesse uno spessore sufficiente la zona di svuotamento non raggiungerebbe mai lo strato sepolto per qualunque valore di  $V_{BC}$ . Ma uno strato epi troppo spesso pone evidentemente dei problemi quando si deve effettuare la diffusione di isolamento che, come è noto, deve attraversare tutto lo strato  $n$ . Se, per effetto della  $V_{BC}$ , la  $zds$  raggiunge lo strato sepolto, l'andamento del campo risulta modificato come in figura. Poiché l'area sottesa dal campo non è nient'altro se non la  $V_{BC}$ , dato che ora viene a mancare il contributo della parte tratteggiata è necessario che il campo massimo (alla giunzione) aumenti; l'effetto risultante è una diminuzione della  $V_{BD}$  della giunzione BC.

Se in un processo bipolare si vuole ottenere una  $V_{BD}$  elevata si deve dunque aumentare lo spessore dello strato epi, il che richiederebbe tuttavia, tempi di diffusione molto elevati e una eccessiva occupazione di area. Si ricorre all'isolamento *dall'alto e dal basso*. Prima della crescita epi vengono impiegate *due* maschere: quella del buried layer e una seconda, con la stessa forma di quella di isolamento, che ha la funzione di predeporre boro. Come si vede dal disegno quando si esegue la diffusione di isolamento l'incontro dei due drogaggi  $p$  avviene a metà strada con conseguente risparmio di tempo e di area occupata (diffusione laterale).



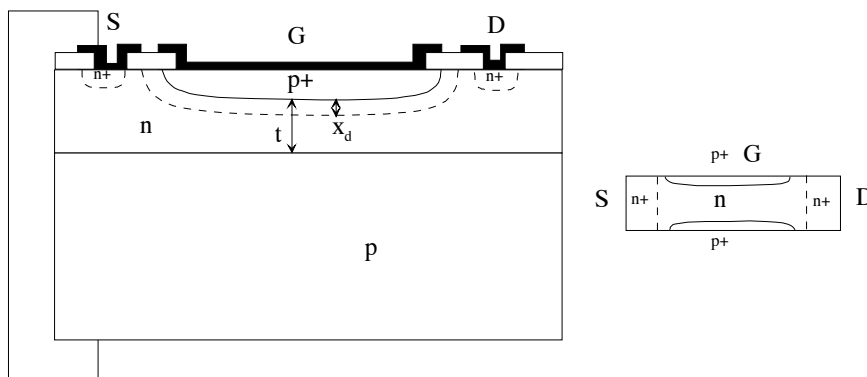
Ricordando che il profilo di drive-in è dato da

$$N(x) = \frac{Q}{\sqrt{\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

in cui la lunghezza di diffusione è  $\lambda = 2\sqrt{Dt}$ , dimezzare questa lunghezza significa ridurre ad un quarto il tempo di diffusione.

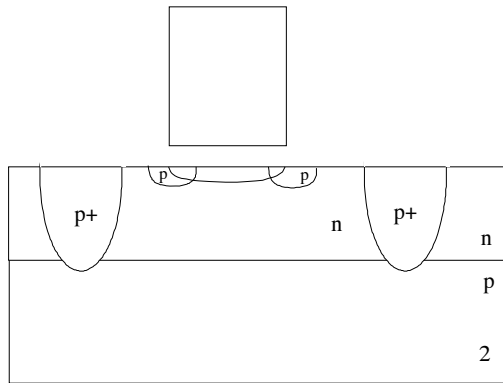
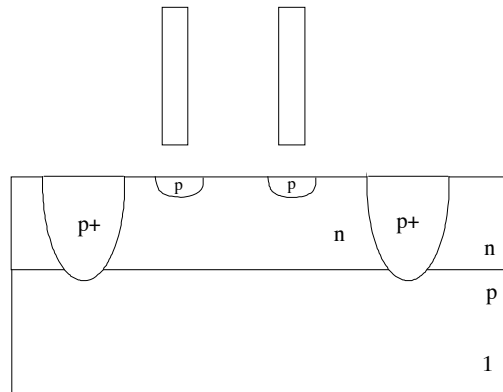
## 25) IL JFET

JFET sta per *Junction Field Effect Transistor*. Una sezione del dispositivo è mostrata nella figura. La giunzione  $p^+n$  polarizzata inversamente fa da gate. La corrente  $I_{DS}$  scorre nel canale  $n$  fra D e S nel materiale bulk e quindi la mobilità degli elettroni è determinata dal drogaggio e non da fenomeni interfacciali come nei MOSFET. Il JFET può essere schematizzato come un resistore, terminato con due contatti ohmici, in cui una delle dimensioni è controllata dalla  $V_{GS}$ .

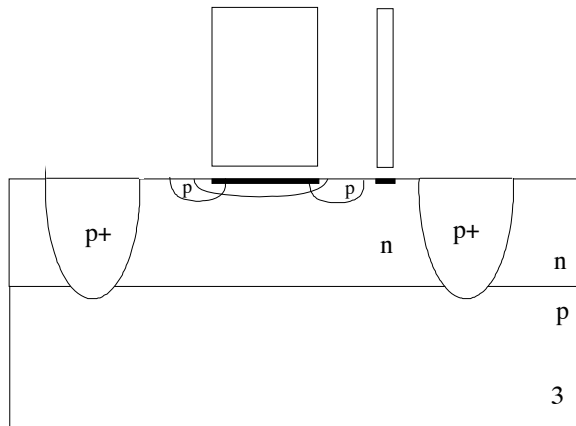


La figura in sezione non rende ragione della struttura effettiva del dispositivo. Infatti, anche quando il canale è completamente chiuso ( $x_d = t$ ), la corrente può aggirare lateralmente il gate. Viene descritto un processo per integrare un JFET a canale  $p$  all'interno di un processo SBC. Utilizzando la diffusione di base vengono

formati, con la maschera 1, i contatti di S e D. Con la maschera 2 si realizza il canale di tipo  $p$  fra S e D (questa è una maschera aggiuntiva al processo SBC).



Finalmente con la maschera 3 si realizza il gate (in questo caso  $n^+$ ; si può quindi usare il drogaggio dell'emettitore). Notare la doppia presa di gate che consente lo svuotamento del canale sia dall'alto che dal basso.



Per piccoli valori di  $V_{DS}$  la resistenza di canale può essere scritta

$$R = \rho \frac{L}{S} = \rho \frac{L}{W(t - x_d)}$$

con

$$x_d = \sqrt{\frac{2\varepsilon_s(V_0 - V_{GS})}{qN_D}}$$

se si considera la giunzione gate/canale come brusca. Altrimenti  $x_d$  dipenderà dalla  $V_{GS}$  nel modo appropriato al tipo di giunzione.

$$\begin{aligned} R &= \rho \frac{L}{S} = \rho \frac{L}{W \left( t - \sqrt{\frac{2\varepsilon_s(V_0 - V_{GS})}{qN_D}} \right)} \\ &= \rho \frac{L}{Wt \left( 1 - \sqrt{\frac{2\varepsilon_s(V_0 - V_{GS})}{t^2 q N_D}} \right)} = \frac{1}{G_0 \left( 1 - \sqrt{\frac{2\varepsilon_s(V_0 - V_{GS})}{t^2 q N_D}} \right)} \end{aligned}$$

in cui  $G_0 = \sigma \frac{Wt}{L}$  è una conduttanza.

$R$  diviene infinita quando

$$\left( 1 - \sqrt{\frac{2\varepsilon_s(V_0 - V_{GS})}{t^2 q N_D}} \right) = 0$$

da cui

$$V_{GS}^* = V_0 - \frac{t^2 q N_D}{2\varepsilon_s}$$

che prende il nome di tensione di interdizione (*turn-off*).

Esempio

$$N_A = 10^{19} \text{ cm}^{-3}, N_D = 10^{15} \text{ cm}^{-3}, t = 2.5 \text{ } \mu\text{m}$$

$$V_{GS}^* = 0.877 - \frac{(2.5 \times 10^{-6})^2 \times 1.6 \times 10^{-19} \times 10^{21}}{2 \times 11.8 \times 8.85 \times 10^{-12}} = -3.91 \text{ V.}$$

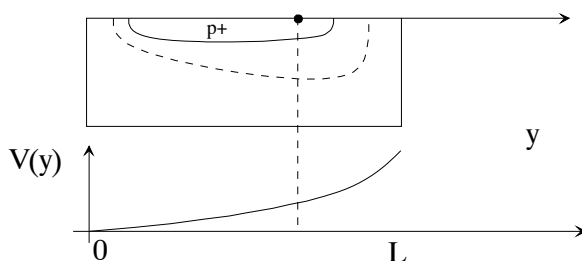
Le caratteristiche  $I_{DS} = f(V_{DS}; V_{GS})$  si ottengono in modo analogo a quelle del MOSFET, utilizzando l'approssimazione di canale graduale, già nota. Si scrive

$$dV = I_{DS} dR = I_{DS} \frac{dy}{Wq\mu_n N_D t \left( 1 - \sqrt{\frac{2\varepsilon_s(V_0 - V_{GS} + V(y))}{t^2 q N_D}} \right)} \quad (1)$$

in cui il termine

$$\sqrt{\frac{2\varepsilon_s (V_0 - V_{GS} + V(y))}{t^2 q N_D}}$$

rappresenta l'ampiezza della zona di svuotamento in presenza di una  $V_{DS}$  non più trascurabile. Infatti la  $I_{DS}$  provoca una caduta  $V(y)$  lungo il canale e quindi la ddp locale (in  $y$ ) fra gate e canale vale  $-V_{GS} + V(y)$ . Si consideri la situazione con  $V_{GS} = 0$  e  $V_{DS} > 0$ .



Nel punto indicato la ddp fra gate e canale vale

$$V_G - V_S = V_{GS} = 0 - V(y)$$

e quindi la giunzione  $p^+n$  è polarizzata inversamente, nonostante  $V_{GS}$  sia uguale a zero. Un'ovvia conseguenza è che l'ampiezza della zds dipende dalla posizione lungo il canale e inoltre che lo strozzamento dello stesso avviene in prossimità del D. Lo stesso ragionamento si può ripetere per una qualunque  $V_{GS} < 0$ .

Si tratta adesso di integrare la (1) fra 0 e  $V_{DS}$  e fra 0 ed  $L$ .

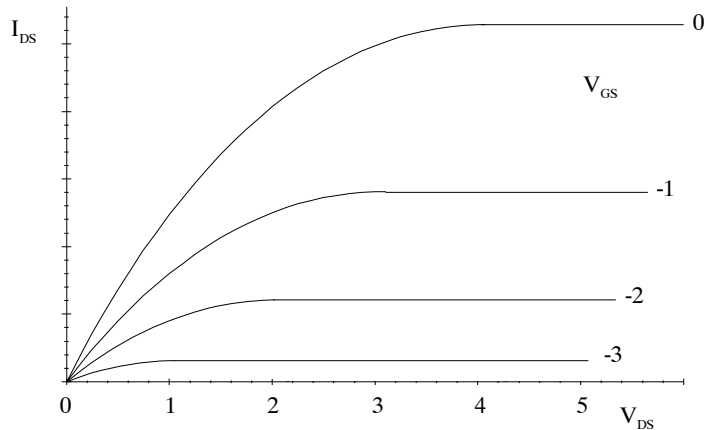
$$\frac{Wq\mu_n N_D t}{L} \int_0^{V_{DS}} \left( 1 - \sqrt{\frac{2\varepsilon_s (V_0 - V_{GS} + V(y))}{t^2 q N_D}} \right) dV = I_{DS}$$

$$G_0 \left( V_{DS} - \frac{2}{3} \left( \frac{2\varepsilon_s}{t^2 q N_D} \right)^{\frac{1}{2}} \left( (V_0 - V_{GS} + V_{DS})^{\frac{3}{2}} - (V_0 - V_{GS})^{\frac{3}{2}} \right) \right) = I_{DS}$$

Le caratteristiche della figura sono relative al seguente

Esempio

$$W/L = 1, N_A = 10^{19} \text{ cm}^{-3}, N_D = 10^{15} \text{ cm}^{-3}, t = 2.5 \text{ } \mu\text{m}, \mu_n = 1300 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}, G_0 = 5.2 \times 10^{-5} \text{ } \Omega^{-1}$$



Anche in questo caso si ha saturazione. La  $V_{DSAT}$  si trova, come al solito, imponendo che

$$\frac{\partial I_{DS}}{\partial V_{DS}} = 0$$

$$G_0 \left( 1 - \left( \frac{2\varepsilon_s}{t^2 q N_D} \right)^{\frac{1}{2}} (V_0 - V_{GS} + V_{DSAT})^{\frac{1}{2}} \right) = 0$$

$$\left( \frac{2\varepsilon_s}{t^2 q N_D} \right)^{\frac{1}{2}} (V_0 - V_{GS} + V_{DSAT})^{\frac{1}{2}} = 1$$

$$V_{DSAT} = V_{GS} - V_0 + \frac{t^2 q N_D}{2\varepsilon_s} = V_{GS} - V_{GS}^*;$$

ricordiamo che  $V_{GS}^*$  nel caso in questione vale -3.91 V.

La transconduttanza

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} \Big|_{V_{DS0}},$$

qualunque sia il punto di lavoro, dipende da  $G_0$  che ha per espressione

$$G_0 = \frac{W q \mu_n N_D t}{L} \propto \mu_n N_D;$$

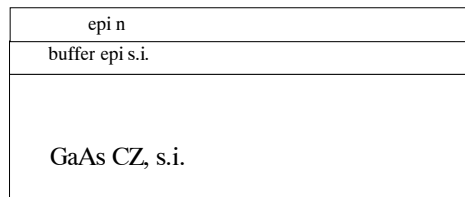
se si vuole aumentare  $G_0$ , a parità di rapporto  $W/L$ , si dovrebbe aumentare  $N_D$ . Ma questo aumento peggiora la situazione delle capacità parassite dovute alle zds e inoltre causa un deterioramento della mobilità. Infatti, ricordiamo che la mobilità è funzione decrescente del drogaggio. La soluzione si può trovare considerando dispositivi analoghi al JFET, ma realizzati con semiconduttori composti.



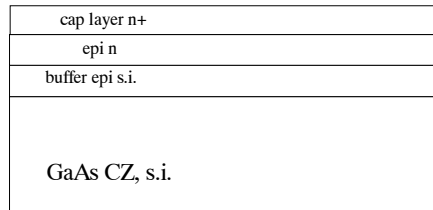
## 26) IL MESFET

Si descrive preliminarmente il processo per la fabbricazione di un MESFET (MEtal Semiconductor FET) in GaAs.

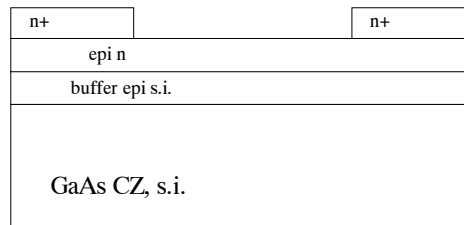
Il substrato di partenza è un CZ intrinseco (semi-isolante), che tuttavia presenta un'alta concentrazione di difetti (dislocazioni, difetti di punto). Per questo motivo è necessario crescere uno strato epi (*buffer*) semi-isolante, per separare il sub dal successivo strato attivo, di tipo  $n$ . La struttura risultante è quella della figura.



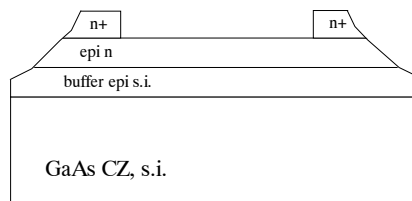
Su questa si depone ancora uno strato  $n^+$  (*cap layer*), che servirà per i contatti ohmici di S e D.



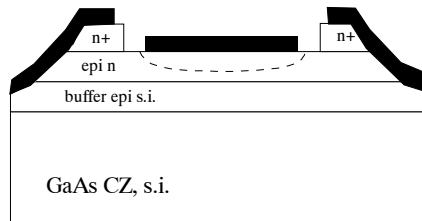
1) Si esegue un attacco che rimuove il cap layer in una zona definita dalla maschera.



2) Con un'altra maschera si delimita il dispositivo mediante un attacco MESA.



3) Si depone un film metallico, che è costituito nella realtà da un multilayer, ma che qui indicheremo semplicemente con Au, e si definiscono i contatti di S, D, G i quali, proseguendo fuori della zona rilevata, scorrono sul GaAs semi-isolante.



La  $z_{ds}$  indicata è quella del diodo Schottky Au/GaAs che fa da gate.

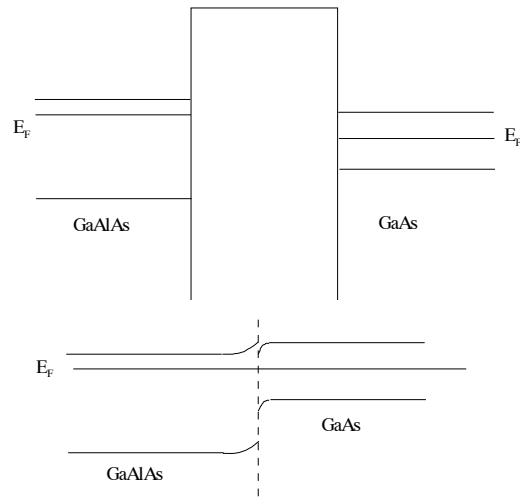
La tecnologia del GaAs e, in generale, quella dei semiconduttori composti è per certi aspetti diversa e più complicata di quella del Si, *in primis* perché non è possibile ottenere per ossidazione termica un ossido di buona qualità.

Il modello descritto per il JFET è evidentemente ancora valido, anche perché in questo caso la giunzione è sicuramente brusca (drogaggio costante). La differenza sta in  $G_0$ ; infatti, a parità di dimensioni del canale e di drogaggio, la mobilità degli elettroni nel GaAs è molto più elevata, dato che la massa efficace degli elettroni della BC vale  $0.068 \cdot m_0$ . Anche in questo caso tuttavia, il trasporto di carica avviene nel canale, dove sono inevitabilmente presenti i droganti che, come sappiamo, limitano la mobilità.

Si ricorre allora ad una struttura che consente di disaccoppiare il canale (i portatori) dal drogaggio (i centri di scattering). A questo scopo è necessario introdurre le

## Eterogiunzioni

che sono giunzioni fra semiconduttori diversi. Il primo problema che si presenta in queste strutture è la differenza di costante reticolare (*mismatch*), che comporta problemi non indifferenti nella realizzazione della giunzione. C'è tuttavia almeno un caso fortunato, ed è quello costituito dal GaAs e dal GaAlAs. Il  $\text{Ga}_x\text{Al}_{1-x}\text{As}$  è un composto ternario ( $x$  costituisce la percentuale di Ga) in cui Ga e Al sono trivalenti. In dipendenza dal valore di  $x$  il gap può essere più o meno ampio, i limiti essendo ovviamente il gap del GaAs (1.43 eV) e quello dell'AlAs (2.16 eV). GaAs e  $\text{Ga}_x\text{Al}_{1-x}\text{As}$  hanno, per tutti i valori di  $x$ , la stessa costante reticolare e quindi la deposizione epi di uno sull'altro non comporta problemi di mismatch. Si consideri la struttura a bande del GaAs (intrinseco) e del GaAlAs di tipo  $n$  prima del contatto e dopo.



L'allineamento dei livelli di Fermi all'equilibrio (c'è stato un travaso di elettroni dal GaAlAs al GaAs) ha fatto sì che si venisse a formare all'interfaccia e nel GaAs uno strato di accumulazione di elettroni che *costituisce un canale* in direzione ortogonale al piano della figura. Il risultato notevole è che il canale esiste in una zona di semiconduttore in cui il drogaggio è assente e quindi la mobilità degli elettroni non viene degradata dalle impurezze ionizzate. Dispositivi con  $f_T$  elevatissima (centinaia di GHz) vengono realizzati con vari tipi di sc composti e con strutture di solito abbastanza complesse. La loro sigla è HEMT (*High Electron Mobility Transistor*), ma la loro descrizione esula dal programma di questo corso.

### 31) Costanti e grandezze di uso frequente

carica dell'elettrone	$e, q$	$= 1.6 \times 10^{-19} \text{ C}$
massa dell'elettrone libero	$m_0$	$= 9.1 \times 10^{-31} \text{ Kg}$
costante di Planck	$h$	$= 6.62 \times 10^{-34} \text{ J} \cdot \text{s}$
costante di Boltzmann	$k$	$= 8.63 \times 10^{-5} \text{ eV} \cdot \text{K}^{-1}$
c. dielettrica del vuoto	$\varepsilon_0$	$= 8.85 \times 10^{-12} \text{ F/m}$
c. diel. relativa del Si	$\varepsilon_s$	$= 11.8$
c. diel. relativa dell'SiO <sub>2</sub>	$\varepsilon_{ox}$	$= 3.9$
costante di Richardson	$A$	$= 120 \text{ A} \cdot \text{cm}^{-2} \text{K}^{-2}$

gap del Si a 300 K	$E_g$	= 1.08 eV
densità eq. di stati in BC a 300 K	$N_C$	= $2.8 \times 10^{19} \text{ cm}^{-3}$
densità eq. di stati in BV a 300 K	$N_V$	= $1.0 \times 10^{19} \text{ cm}^{-3}$
conc. portatori int. nel Si a 300 K	$n_i$	= $1.5 \times 10^{10} \text{ cm}^{-3}$
affinità elettronica del Si	$\chi$	= 4.05 eV
gap dell'SiO <sub>2</sub>		≈ 9 eV
gap del GaAs	$E_g$	= 1.43 eV
massa efficace per la cond. nella BC (Si)	$m_e^*$	= $0.26 \cdot m_0$
massa efficace per la cond. nella BC (GaAs)	$m_e^*$	= $0.068 \cdot m_0$

temperatura di fusione dell'Al	$T_{FAl}$	= 660°C
temperatura di fusione dell'SiO <sub>2</sub>	$T_{Fox}$	≈ 1710°C
temperatura di fusione del Si	$T_{FSi}$	= 1421°C
resistività dell'Al	$\rho$	= $2.65 \mu\Omega \cdot \text{cm}$
resistività del Cu	$\rho$	= $1.7 \mu\Omega \cdot \text{cm}$
funzione di lavoro dell'Al	$\Phi_M$	= 4.25 eV

**32) Bibliografia essenziale**

- 1) B. G. Streetman: "*Solid State Electronic Devices*", Prentice-Hall International Editions
- 2) G. Ghione: "*Dispositivi per la Microelettronica*", McGraw-Hill
- 3) R. S. Muller, T. I. Kamins. "*Device Electronics for Integrated Circuits*", J. Wiley & Sons
- 4) S. Wolf, R. N. Tauber: "*Silicon Processing for the VLSI Era*", vol. 1, Lattice Press
- 5) S. M. Sze: "*Dispositivi a Semiconduttore*", Hoepli
- 6) P. S. Kireev: "*Semiconductor Physics*", Mir Publishers
- 7) J. F. Gibbons: "*Semiconductor Electronics*", McGraw-Hill
- 8) Y. Taur, T. H. Ning: "*Fundamentals of Modern VLSI Devices*", Cambridge University Press
- 9) Y. P. Tsividis: "*Operation and Modeling of the MOS Transistor*", McGraw-Hill