

Dispositivi elettronici integrati

Richiami di teoria dei semiconduttori

Veniamo, a questo inizio di trattazione, richiamati ai concetti di base della teoria dei semiconduttori "tradizionali" (trascurando per ora le eterostrutture).

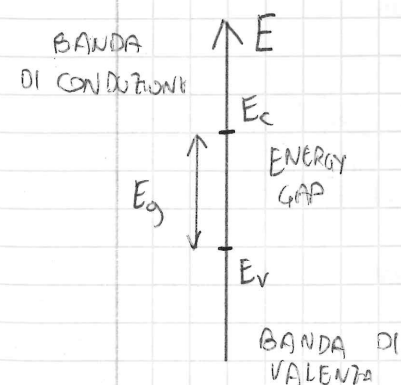
L'elemento fondamentale che permette il funzionamento di un semiconduttore è la sua struttura cristallina: i vari atomi di un semiconduttore sono infatti disposti "periodicamente" nello spazio.

Nel caso dei semiconduttori, la conduzione di corrente è basata sullo spostamento fisico di due tipi di carica elettrica; esse hanno valore uguale in modulo, opposto in segno: elettroni liberi e lacune libere. Si ricorda che $q = 1,6 \times 10^{-19}$ C.

Le condizioni che portano alla conduzione sono 2:

- Presenza di elettroni nella banda di conduzione del cristallo
- Presenza di lacune in banda di valenza.

Si tenga a mente che i posti disponibili sul guscio più esterno dell'atomo sono delocalizzati nel cristallo; per questo motivo, i "posti liberi" non sono singoli, ma dipendono da funzioni di distribuzione.



Nella teoria dei dispositivi a semiconduttore, si usa, come rappresentazione delle proprietà, l'asse delle energie E , ossia lo studio grafico dell'energia che gli elettroni possono assumere.

Si può collegare ciò che è stato detto prima con questo grafico, identificando tre regioni "importanti": la banda di conduzione, il cui minimo livello energetico è E_c , la banda di valenza, il cui massimo livello è E_v , e la "banda proibita", o "energy gap", di ampiezza E_g :

$$E_g = E_c - E_v.$$

Questa teoria si può usare proprio grazie alle proprietà cristalline del materiale.

Si possono collegare i "posti disponibili" della banda di valenza alle proprietà elettriche del cristallo: un elettrone in banda di valenza infatti è necessario in quanto rappresenta un legame covalente (da qua il nome "di valenza"). Se un elettrone sta invece in banda di conduzione, esso non partecipa al legame chimico, e si può muovere liberamente all'interno del cristallo.

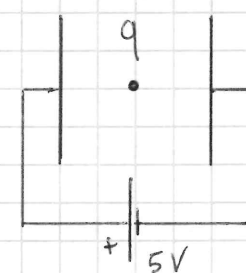
Si è parlato di "lacune": esse non sono "particelle reali", bensì "fittizie", che rappresentano l'assenza di un elettrone, o la presenza di un legame spezzato all'interno del cristallo.

Se la temperatura T è maggiore di 0 K , l'energia termica comporta il passaggio di parte degli elettroni in banda di conduzione, dunque si ha, per quanto poca, conduzione: applicando un campo elettrico, la carica si muove.

Con le lacune, il ragionamento è duale: dati dei posti vuoti, è come se "i posti vuoti si muovessero"; nella realtà sono sempre gli elettroni a muoversi, ma per semplificazione noi la vediamo in questo modo.

Il numero di coriche che si muove per eccitazione da campo elettrico permette di caratterizzare le proprietà elettriche del materiale in: conduttori, semiconduttori, isolanti. Si può trovare, in tal senso, una relazione tra l'ampiezza della banda proibita e la quantità di carica libera di muoversi, a una certa temperatura. Si ricorda che per ora la conduzione è dovuta solo alla rottura di legami chimici covalenti: se si alza molto la temperatura, si spezzano i legami, e il solido si fonde.

Si vuol definire una grandezza utile per la misura dell'energia:



l'elettronvolt. Esso è l'energia di un elettrone dovuta all'eccitazione di 1 V di ddp: $1\text{ eV} = 1,6 \cdot 10^{-19}\text{ J}$.
Nel caso della figura, la carica q avrà un'energia pari a $1,6 \cdot 10^{-19} \cdot 5\text{ V} = 8\text{ eV}$. stretta

Si potrebbe pensare che si vorrebbe una banda proibita larga, ~~al fine di facilitare la conduzione~~, ciò non è sempre vero: nei dispositivi optoelettronici, infatti, la banda larga può tornare utile, dal momento che, per l'optoelettronica, è necessario avere a che fare con lunghezze d'onda λ piccole, dunque con fotoni con λ piccolo, e " f " (frequenza) elevato, ma dunque anche con energia elevata.

Le caratteristiche elettriche dei semiconduttori possono essere modificate mediante i processi di drogaggio: mediante esso, si può intervenire, modificando il numero di coriche libere in banda di conduzione e/o di valenza. Drogare significa inserire nel

crystallo atomi non del semiconduttore, ma studiati apposta per migliorare le caratteristiche come, nel caso del Si, l'arsenico As, il boro B, ecc.

La disposizione di questi atomi va studiata bene, in modo da creare, per il silicio, i quattro legami covalenti che tengono assieme la materia, e introdurre o elettroni liberi in banda di conduzione (vicino ad E_c), o lacune libere in banda di valenza (vicino ad E_v). Si introduce una quantità di drogante piccola rispetto al silicio, in modo da non alterare troppo la struttura della materia. Si vede che i materiali usati per il drogaggio introducono i livelli energetici in prossimità della banda "utile"; si noti che essi sono nella banda proibita: questo non è importante, del momento che, in seguito all'introduzione di materiale "estaneo", la teoria delle bande risulta essere diversa, dunque la struttura ideale risulta essere perturbata.

E, nel caso del silicio, si introduce un elemento del V° gruppo della tavola periodica, si introduce (supponendo di aver scelto un buon materiale) un livello energetico in prossimità a E_c , in grado di "fornire elettroni" che devono effettuare un "salto energetico" molto inferiore rispetto all'energy gap. Ciò incrementa il numero di cariche in banda di conduzione, ed è detto drogaggio di tipo "n", o "donatore".

Esiste un processo duale, ma effettuato con un elemento del III° gruppo: il fatto che l'elemento abbia solo 3 elettroni

che possono partecipare ai legami covalenti con il silicio, lascia un posto libero per un quarto elettrone, che sarà derivare dalla banda di valenza; la banda di valenza dovrà mandare degli elettroni nel livello energetico creato dal drogante (si spara, livello prossimo a E_v , per facilitare il "salto"), dunque si formeranno lacune in banda di valenza, dovute a questo passaggio. Dal momento che ora il drogante "diventa a sé" gli elettroni invece si danno, esso è detto "accettore", o "p". Drogare può avere effetti di tipo diverso a seconda e del gruppo dell'atomo drogante, e della posizione del livello introdotto dal drogante in quella che era la banda proibita.

Esiste una condizione detta "drogaggio profondo", rispetto alla banda proibita, per cui i livelli "utili" sono molto "lontani" da E_c ed E_v . Questo può essere utile per introdurre dei "centri di ricombinazione", ma per creare zone dove si "ammazzi la conducibilità" del semiconduttore.

Equilibrio termodinamico.

L'equilibrio termodinamico è una condizione in cui non si ha scambio di energia con l'esterno, dunque isolamento rispetto a riscaldamento, illuminazione, campi o altro.

Non è dunque possibile avere scambi di flussi di corrente con l'esterno, ma non per questo non ci devono essere neanche spostamenti "interni" di elettroni: è sufficiente che ci sia una

compensazione dei flussi tali da avere flusso risultante nullo.

Si introduce il concetto di densità volumica di portiche in banda di conduzione e in banda di valenza: queste sono le bande "interessanti", dal momento che sono esse quelle che partecipano ai processi di conduzione. Data "n" la concentrazione di elettroni in banda di conduzione, "p" quella di lacune in banda di valenza, si può dire che:

$$n = \int_{E_c}^{+\infty} g_n(E) dE \quad ; \quad p = \int_{-\infty}^{E_v} g_p(E) dE$$

Per conoscere questa strada è facile, dal momento che le funzioni di densità, g , possono essere espresse in modo abbastanza semplice. Le funzioni di densità sono infatti il numero di portiche presenti tra un generico livello energetico E e $E+dE$; l'operatore di integrazione somma, dunque si considerano "n" e "p" come la somma di tutti i contributi dei vari livelli energetici rispettivamente in banda di conduzione e di valenza. Si approssimano con $\pm\infty$ i livelli energetici "minimi" e "massimi" delle rispettive bande.

Le funzioni di distribuzione possono essere espresse come prodotto di due termini: la densità e la probabilità di occupazione dei livelli energetici E :

$$N_n(E) = \gamma_n \sqrt{E - E_c} \quad , \quad \gamma_n = 4\pi h^{-3} (2m_n)^{\frac{3}{2}}$$

$$N_p(E) = \gamma_p \sqrt{E_v - E}$$

La probabilità di occupazione, dal momento che si parla di elettroni, è la Fermi-Dirac:

$$f(E) = \left[1 + e^{\frac{E - E_F}{k_B T}} \right]^{-1}$$

Nel caso delle lacune, è il complemento a 1 di $f(E)$ (probabilità di NON occupazione di un livello energetico). La cosa finita è la presenza di E_F (livello di Fermi / potenziale chimico): esso è incognito, almeno per ora. Date dunque $f(E)$, $N_n(E)$ e $N_p(E)$ saremmo in grado di calcolare n e p, ma ci mancano ancora alcuni dati. 1/4/05

Si noti che il numero degli stati disponibili è indipendente dal drogaggio; le informazioni sul drogaggio stanno dunque nella funzione di occupazione, o sia nella Fermi-Dirac; nella fattispecie, si può osservare, per esclusione, che l'unico parametro che può presentare la dipendenza dal drogaggio è E_F , o sia il livello di Fermi.

E_F è quel livello energetico in cui, indipendentemente dalla temperatura T , la probabilità di occupazione è pari a $\frac{1}{2}$: 50%. In altri termini, esso è, a 0 K, la massima energia che possa essere acquisita da un elettrone (all'equilibrio termodinamico). Sostituendo le espressioni in "n" e "p", si trovano i seguenti integrali:

$$n = \gamma_n \int_{E_c}^{+\infty} \frac{\sqrt{E - E_c}}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)} dE \quad ; \quad p = \gamma_p \int_{-\infty}^{E_v} \frac{\sqrt{E_v - E}}{1 + \exp\left(-\frac{E - E_F}{k_B T}\right)} dE$$

Questi integrali non ammettono primitiva esprimibile con funzioni elementari, a meno che $T = 0$ K; al fine di esprimerli, è ne-

possiamo usare le funzioni speciali di Fermi.

Slide 18 Esistono grafici relativi a queste funzioni, da cui, a vista, si può evincere che E_F sia al centro della banda proibita. Questo si può capire per il seguente ragionamento: le funzioni di distribuzione degli stati sono costanti e, dal momento che non si sta drogando, le aree sotto le curve dei grafici sono uguali. L'unico elemento variabile è il grafico centrale che, in questa situazione, sarà simmetrico. Conseguentemente, drogando tipo n, aumenta l'area "n" nel grafico conformato, ma dunque, poiché le distribuzioni sono costanti, il grafico della funzione di occupazione deve cambiare, e E_F alzarsi. Effetto duale si avrebbe, drogando tipo p.

Al fine di semplificare l'integrale, sarebbe bello poter "buttare via" il "1" al denominatore, in modo da rendere gli integrali gaussiani, dunque risolvibili in forma chiusa. È possibile far ciò, nel caso di semiconduttore non drogato, ~~ovvero se $E_c > E_F$~~ in tal caso l'esponente è maggiore di 1. Nella fattispecie, si può studiare il fatto che $E_c - E_F > 0$, e $E_F - E_v > 0$: in tal caso il livello di Fermi è all'interno della banda proibita, e il semiconduttore "non drogato", dunque l'approssimazione valida. Questa condizione è verificata, se il drogaggio non è eccessivamente introdotto nel semiconduttore, eccessivamente grande; nei dispositivi optoelettronici, per esempio, servono grandi drogaggi, dunque l'approssimazione non varia.

Allora dunque approssimiamo la Fermi-Dirac alla Maxwell-Boltzmann, ottenendo:

$$n \approx \gamma_n \exp\left(\frac{E_F}{k_B T}\right) \cdot \int_{E_c}^{\infty} \sqrt{E - E_c} \exp\left(-\frac{E}{k_B T}\right) dE =$$

$$= N_c \exp\left(\frac{E_F - E_c}{k_B T}\right) ; \quad N_c \propto T^{\frac{3}{2}}$$

Allo stesso modo, si dimostra che:

$$p = N_v \exp\left(\frac{-E_F + E_v}{k_B T}\right)$$

Dove N_c e N_v sono dette "densità efficaci degli stati", rispettivamente in banda di conduzione e in banda di valenza.

È a questo punto possibile introdurre la cosiddetta "legge dell'azione di massa": moltiplicando tra loro "n" e "p", si ottiene:

$$n \cdot p = N_c \cdot N_v \cdot \exp\left(-\frac{E_c - E_v}{k_B \cdot T}\right) = N_c \cdot N_v \cdot \exp\left(-\frac{E_g}{k_B \cdot T}\right)$$

Nel caso in cui non si abbia per i poteri del drogaggio, si sa che $n_i = p_i$ (pedice "i" sta per "intrinseco"), dunque:

$$n_i \approx N_c \exp\left(\frac{E_F - E_c}{k_B T}\right) ; \quad p_i \approx N_v \exp\left(\frac{-E_F + E_v}{k_B T}\right)$$

$$n_i \cdot p_i = n_i^2 = N_c N_v \exp\left(-\frac{E_g}{k_B T}\right)$$

Si tratta dello stesso risultato di prima, usando le concentrazioni intrinseche (dunque uguali tra loro) al posto di quelle generiche; questo ci dice che, indipendentemente dal drogaggio, il prodotto delle concentrazioni è costante.

$$n_i, @ 300 \text{ K}, \approx 1,45 \cdot 10^{10} \text{ atom/cm}^3$$

Questo numero, a temperatura costante, è idealmente costante.

Senza considerare il drogaggio (per semplicità), in un semiconduttore dunque intrinseco, il numero di cariche libere nasce dall'equilibrio tra due meccanismi: la generazione termica, e la ricombinazione. Questa coppia di meccanismi è la "porta" da cui si entra nel mondo dell'elettronica: essa si basa su meccanismi di questo tipo. Questo perché, al fine di rompere legami covalenti, è possibile usare eccitazione di tipo luminoso, ova fotoni. n_i è il risultato dell'equilibrio tra generazione termica e ricombinazione: tanto si genera quanto si ricombina. In tal processo, la dipendenza da T è più che esponenziale.

Si può calcolare la posizione energetica di E_{Fi} , ova del livello di Fermi del semiconduttore intrinseco. Supponendo che E_{Fi} sia nella banda proibita:

$$E_{Fi} = \frac{E_c + E_v}{2} - \frac{k_B T}{2} \ln \left(\frac{N_c}{N_v} \right) \approx \frac{E_c + E_v}{2}, \text{ se } N_c \approx N_v.$$

Approssimando il termine logaritmico a nullo, è possibile usare le equazioni di Boltzmann; si ricardi che:

$$n_i \approx N_c \exp \left(- \frac{E_c - E_{Fi}}{k_B T} \right); \quad p \approx N_v \exp \left(- \frac{E_v - E_{Fi}}{k_B T} \right)$$

Ricorrendo dalla prima N_c e sostituendo nella seconda, si ottiene:

$$n \approx n_i \exp \left(\frac{E_F - E_{Fi}}{k_B T} \right) \quad \text{Ovvero la 1ª eq. di Shockley;}$$

Dualmente:

$$p \approx n_i \exp \left(\frac{E_{Fi} - E_F}{k_B T} \right)$$

Si può rafforzare una terza, un'interpretazione già accennata in precedenza: il livello di Fermi può essere pensato come una sorta di "baricentro" delle concentrazioni di carica libere: più esso è verso una delle due bande, più essa sarà piena di cariche libere (se vicino alla banda di conduzione essa avrà molti elettroni liberi, viceversa lacune in banda di valenza).

A questo punto, abbiamo 2 equazioni in 3 incognite: n , p , E_F . Alliamo invece di una terza condizione: la neutralità globale?
 Supponendo di aver drogato il semiconduttore con N_D atomi donatori e N_A atomi accettori, di cui N_D^+ e N_A^- risultano essere ionizzati, si ha che:

$$n + N_A^- = p + N_D^+$$

Sinfatti, una volta ionizzati, gli atomi donatori (donato l'elettrone) diventano ioni carichi positivamente; d'altronde, gli atomi accettori, diventano ioni carichi negativamente. Volendo imporre la neutralità, si pone l'uguaglianza delle cariche: somma delle cariche positive uguale alla somma delle cariche negative.

Utilizzando le espressioni di n (Fermi-Dirac o Maxwell-Boltzmann) e di p (quella di N_A^- e N_D^+), è possibile, applicando la legge dell'azione di massa, ottenere l'andamento delle cariche libere al variare della temperatura T ; si può dunque indicare la presenza di tre regimi di funzionamento:

- 1) Di congelamento: la ionizzazione non è stata completata
- 2) Estrinseco: solo il drogante ha influenza sul comportamento

3) Intrinseco: a temperature elevate, il semiconduttore torna a lavorare in regime intrinseco.

Nota conclusiva: le equazioni di Boltzmann permettono una stima della validità del loro stesso uso: mediante esse, come visto, è possibile ricavare la posizione del livello di Fermi, e vedere "quanto il semiconduttore è non-degenero" osservando la sua posizione all'interno della banda proibita; se $N_0 \gg N_D$, $N_0 \gg N_A$, la condizione di non degenerazione è verificata, e il "logaritmo" del valore di E_F è tracciabile.

Modello matematico dei semiconduttori

Sino ad ora ci si era concentrati sullo stato di equilibrio termodinamico; si tratta di una situazione all'ostanza ideale. Dal momento che i dispositivi vengono usati non tanto all'equilibrio, quanto eccitati da segnali, al fine di elaborarli. L'obiettivo sarà dunque identificare strumenti che permettano di descrivere il comportamento elettrico del dispositivo anche fuori equilibrio, tipicamente sotto l'azione del ~~del~~^{campo} elettrico.

Quando si inserisce il semiconduttore in un campo elettrico, le leggi fondamentali della fisica dicono che agisce una forza sulle cariche libere; ciò le accelera, portando a ottenere un flusso di carica elettrica che si muove.

All'equilibrio termodinamico e con $T > 0$, la corrente era nulla; ciò non significa che gli elettroni fossero fermi: se la meccanica

quantistica permette di misurare la posizione e la velocità di ciascuna particella, si potrebbe osservare che quest'ultima, spesso, non solo non è nulla, ma anche piuttosto elevata.

Mediante un grosso numero di cariche mediante la meccanica statistica, si potrebbe ottenere una v_{media} , la quale rappresenta un parametro per la caratterizzazione della corrente a livello macroscopico per il semiconduttore.

Nel caso si introduca un campo elettrico, si ha una sorta di sovrapposizione degli effetti del moto termico e della forza elettrica, dunque il moto risultante è dato dalla sovrapposizione del moto termico (e della velocità v_{media}) e di questo effetto.

A livello macroscopico, sono interessanti principalmente il solo moto provocato dal campo elettrico.

Per quanto riguarda questo moto, si può dimostrare che la sua velocità media dipende da un parametro (due, in realtà), detto "mobilità degli portatori liberi":

$$\begin{cases} v_n = -\mu_n \cdot E \\ v_p = \mu_p \cdot E \end{cases} \quad \begin{array}{l} \text{I coefficienti } \mu_i \text{ indicano quanto i portatori} \\ \text{sono "disposti" a muoversi se eccitati dal campo.} \end{array}$$

Questi parametri caratterizzano il materiale in termini di massima frequenza di funzionamento: il canale di drain, per esempio, non potrà seguire variazioni più rapide rispetto al tempo di transito di un elettrone.

Si potrebbe pensare che, al crescere del campo elettrico, la velocità di trasinamento dei portatori continui a crescere; ciò non è del

tutto vero, dal momento che ci sono effetti che portano a una saturazione della velocità; il cristallo infatti non è una funzione perfettamente periodica dello spazio: ci sono imperfezioni simili a impurità non umore, o al drogaggio; anche gli atomi subiscono inoltre l'influenza della temperatura, le quali provocano vibrazioni reticolari (sotto forma di "fononi"). Si identificano per questo motivo due regioni di funzionamento:

- Regione di basso campo: la velocità cresce come si può immaginare;
- Regione a mobilità differenziale negativa: il drift velocity va meno veloce di quello che ci si potrebbe aspettare.

Anche per questo motivo si tende a ridurre le tensioni: evitare la regione di saturazione.

Si noti che, a seconda del materiale, μ_n può essere anche molto diversa da μ_p (come nel GaAs). Si osservi che la mobilità elettronica è una funzione della temperatura, del materiale, e del drogaggio totale (non solo di quello ionizzato).

Legge di Ohm microscopica

Si vuole a questo punto determinare una sorta di generalizzazione, per i semiconduttori, della legge di Ohm. Dato un campione infinitesimo di lunghezza ds , e sezione A , si ha:

$$I = \frac{dQ}{dt} \quad \text{Come noto dalla definizione di corrente.}$$

Considerando che Q è funzione della lunghezza di ~~drift~~ ^{semiconduttore} percorsa, si può usare il teorema delle funzioni composte, ottenendo:

$$\frac{dQ}{dt} = \frac{dQ}{ds} \cdot \frac{ds}{dt}$$

Moltiplicando e dividendo per la sezione A , si ottiene:

$$\frac{dQ}{ds} \cdot \frac{ds}{dt} \cdot \frac{A}{A} = \frac{dQ}{dV} \cdot A \cdot v_n \quad \left[\frac{ds}{dt} = v_n \right]$$

Ma la densità della carica nel volume è pari alla concentrazione di carica, per il valore di una singola carica; nel caso dell'elettrone, $-q = -1,6 \cdot 10^{-19} \text{ C}$:

$$\rightarrow I_n = -q \cdot n \cdot A \cdot v_n = q \cdot n \cdot A \cdot \mu_n \cdot E$$

Si può a questo punto definire la conducibilità elettronica, a partire dalla densità di corrente J :

$$J_n \stackrel{\Delta}{=} \frac{I}{A} = \sigma_n \cdot E = q \cdot n \cdot \mu_n$$

$$\rightarrow \sigma_n = q \cdot n \cdot \mu_n$$

Nello stesso modo, si può fare la stessa cosa per le lacune, ottenendo:

$$J_p = q \cdot p \cdot \mu_p \cdot E \quad \rightarrow \quad \sigma_p = q \cdot p \cdot \mu_p$$

La conducibilità totale sarà:

$$\sigma = \sigma_n + \sigma_p = q \cdot (n \cdot \mu_n + p \cdot \mu_p) \quad ; \quad \rho = \frac{1}{\sigma}$$

Si noti che questi sono ragionamenti statici: essi non introducono elementi quantificanti la velocità di variazione della corrente. Volendo rappresentare un grafico quantificante la resistività ρ al variare del drogaggio, esso sarebbe decrescente: dal momento che gli effetti del

drogaggio sono più tenui rispetto alla quantità di cariche libere, sarà essa a prendere.

Semiconduttore fuori equilibrio

Intendiamo a questo punto caratterizzare il comportamento dei semiconduttori, fuori dalla condizione di equilibrio. Nella fattispecie, siamo interessati alle concentrazioni di cariche libere, in funzione della loro posizione spaziale (x) del semiconduttore.

5 contributi possono derivare da diverse sorgenti:

- Variazione del flusso di corrente del semiconduttore
- Variazioni dirette del numero di elettroni/lacune.

Un fenomeno presente, ma che non verrà molto approfondito, riguarda lo spostamento dielettrico: esso è dipendente dalla derivata rispetto al tempo del vettore spostamento dielettrico; un esempio di corrente dovuta allo spostamento dielettrico è quella presente in un condensatore, per variazioni ad alta frequenza.

Diffusione

Il fenomeno della diffusione deriva dalla formazione di gradienti non nulli nella concentrazione delle cariche; esistono, nella fattispecie, le seguenti relazioni:

$$J_{ndiff} = q D_n \frac{\partial n}{\partial x} ; \quad J_{pdiff} = -q D_p \frac{\partial p}{\partial x}$$

Dove D_n e D_p sono coefficienti in grado di quantificare il fenomeno di diffusione. Essi sono esprimibili come:

$$D_n = V_T \cdot \mu_n ; \quad D_p = V_T \cdot \mu_p$$

Dove V_T è l'equivalente termico in tensione:

$$V_T = \frac{k_B T}{q}$$

Trasporto

La densità di corrente di trasporto è già stata determinata, ed è unitata essere quantificabile come:

$$J_{tr} = J_{trn} + J_{trp} = q n \mu_n E + q p \mu_p E$$

Concentrazioni in eccesso

Si usa definire le concentrazioni in eccesso, rispetto alla condizione di equilibrio termodinamico. Date n_0 e p_0 le concentrazioni di portatori con drogaggio tipo "n", n_p e p_p quelle con drogaggio tipo "p", e aggiungendo un ulteriore pedice "e" nel caso di equilibrio termodinamico, si definiscono, con l'apice "e", le concentrazioni in eccesso:

$$\text{Drogaggio tipo n: } \begin{cases} n'_e = n_n - n_{n0} \\ p'_e = p_n - p_{n0} \end{cases}$$

$$\text{Drogaggio tipo p: } \begin{cases} n'_e = n_p - n_{p0} \\ p'_e = p_p - p_{p0} \end{cases}$$

Quando n'_e e p'_e sono positive, si ha il fenomeno dell'iniezione; se negative, lo sottrazione.

Si introduce l'ipotesi di "quasi neutralità" del dispositivo se: $n'_e \approx p'_e$.

Per quanto riguarda l'iniezione, si distinguono due particolari situazioni:

- Basso livello di iniezione: l'iniezione di coppie elettrone-lacuna

altera sensibilmente i portatori minoritari, ma non i maggioritari.

- Alto livello di iniezione: l'iniezione di coppie elettrone-lacuna è tale da alterare sia i minoritari, sia i maggioritari.

Spesso si usa la prima condizione: essa è infatti favorevole, dal momento che permette di ricalcolare solo le concentrazioni di minoritari.

Si parla di iniezione ma... Cosa effettua il processo di iniezione?

Il processo di iniezione, come già accennato, riguarda l'iniezione di coppie elettrone-lacuna; essa può derivare da riscaldamento di temperatura o, più probabilmente, da eccitazione ottica. Questo è ciò che capita, per esempio, nei fotodetettori. Andando un poco nel dettaglio, esistono due sotto-tipi di iniezione.

- Meccanismi di iniezione diretti (a gap diretto): gli elettroni saltano dalla banda di valenza alla banda di conduzione, sfruttando il fatto che il minimo della banda di conduzione abbia la stessa posizione di fase del massimo della banda di valenza. Sono i migliori, per l'optoelettronica.
- Meccanismi indiretti: il salto avviene in livelli energetici intermedi, realizzati mediante l'introduzione di centri di ricombinazione. Se il "livello trappola" è prossimo al centro della banda proibita, questi meccanismi sono particolarmente efficienti per il silicio, (suoi centri di ricombinazione possono essere l'oro o il platino).

Generazione e Ricombinazione

Al fine di caratterizzare i processi di generazione e ricombinazione, si usano dei tassi, per unità di tempo.

Dati "G" il tasso di generazione, "R" il tasso di ricombinazione, specificando il pedice di quale tipo di portatore ci si sta occupando, si possono definire i tassi netti di ricombinazione come:

$$U_n = R_n - G_n \quad ; \quad U_p = R_p - G_p.$$

All'equilibrio, essi vanno a 0: vi è tanta generazione quanta ricombinazione!

Approssimando a funzioni lineari, si può dire che:

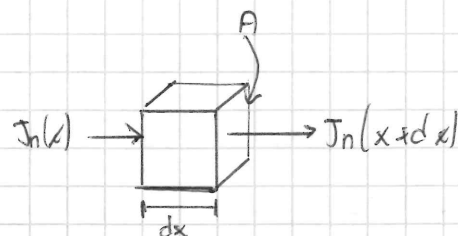
$$U_n \approx \frac{n - n_0}{\tau_n} = \frac{n'}{\tau_n} \quad ; \quad U_p \approx \frac{p - p_0}{\tau_p} = \frac{p'}{\tau_p}$$

Dove i τ_i sono i tempi di vita medi dei portatori. Se esprimiamo U_n e U_p sono in realtà molto più complicati, ma questa vuole essere solo una prima approssimazione. Questa approssimazione è molto scadente, ma ci permette di risolvere, almeno qualitativamente, diversi problemi numerici.

Abbiamo dunque visto cosa capita quando un semiconduttore scambia energia con l'esterno. Con l'optoelettronica la storia sarà più complicata: è necessario studiare segnali ottici ed elettrici; il fatto di dover trattare assieme queste due condizioni, questi due termini, complica le cose. Nella trattazione ci occuperemo solo dell'aspetto della parte elettrica. Si noti che le approssimazioni fatte per i tassi netti di ricombinazione valgono solo per $n', p' \gg 0$ (buchi).

Equazione di continuità

Si consideri il seguente elemento infinitesimo:



Si può dire che, nel volume, la variazione di concentrazione di elettroni sarà: $\frac{\partial n}{\partial t} dV = \frac{\partial n}{\partial t} A dx$

Questa variazione è dovuta a 4 fattori: gli elettroni entranti, quelli usanti, quelli generati termicamente, quelli ricombinati:

$$\frac{\partial n}{\partial t} A dx = \frac{J_n(x)}{-q} \cdot A - \frac{J_n(x+dx)}{-q} + G_n \cdot A dx - R_n \cdot A dx$$

Considerando, del termine in $x+dx$, lo sviluppo di Taylor linearizzato, si ha:

$$J_n(x+dx) \approx J_n(x) + \frac{\partial J_n}{\partial x} dx, \quad dx \rightarrow 0.$$

Si può approssimare dx con "0", ottenendo:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} - U_n$$

Allo stesso modo, si può dimostrare:

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} - U_p$$

Dove per U_n e U_p si può usare, al fine di "introdurre dei numeri", l'approssimazione "tempo di vita". Per quanto riguarda J_n e J_p , si può usare il modello a deriva/diffusione, usando:

$$J_n = q \mu_n n E + q D_n \frac{\partial n}{\partial x}; \quad J_p = q \mu_p p E - q D_p \frac{\partial p}{\partial x}$$

Equazione di Poisson

Manca ancora una relazione, al fine di completare il modello matematico: si parla di " n ", di " p ", di " E "; E , o il campo elettrico, non è di comune uso in ambito elettronico, dove si preferisce usare, storicamente, un'altra grandezza: il potenziale ϕ . Ciò che può chiudere il cerchio, introducendo una relazione con la differenza di potenziale ϕ , è la cosiddetta equazione di Poisson:

$$\frac{\partial E}{\partial x} = \frac{\rho}{\epsilon}; \quad E = -\frac{\partial \phi}{\partial x}$$

Dove ϵ è la costante dielettrica del materiale e ρ la densità di carica netta positiva:

$$\rho = q(p - n + N_D^+ - N_A^-).$$

Le due equazioni di continuità, completate dall'equazione di Poisson, formano il modello matematico dei semiconduttori, al quale poi si aggiungeranno condizioni al contorno. Il modello verrà da noi usato tenendo sempre conto delle seguenti approssimazioni:

- Mobilità elettronica/della lacuna costante con il campo (considerando dunque solo campi elettrici "bassi").
- Modello di Einstein di diffusione: $D = V_T \cdot \mu$
- Uso dell'approssimazione di tempo di vita
- Spazi di completa ionizzazione dei droganti (la più ragionevole).

A questo punto, sostituendo tra loro le varie equazioni, si può ottenere con una certa facilità:

$$\begin{cases} \frac{\partial n}{\partial t} = \mu_n \frac{\partial(n\mathcal{E})}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2} - \frac{n-n_0}{\tau_n} \\ \frac{\partial p}{\partial t} = -\mu_p \frac{\partial(p\mathcal{E})}{\partial x} + D_p \frac{\partial^2 p}{\partial x^2} - \frac{p-p_0}{\tau_p} \\ \frac{\partial^2 \phi}{\partial x^2} = -\frac{q}{\epsilon} (p - n + N_D - N_A) \end{cases} \quad \mathcal{E} = -\frac{\partial \phi}{\partial x}$$

Applicando, su questo modello, alcune condizioni, esso può ulteriormente semplificarsi, permettendo di modellare situazioni particolari per i semiconduttori, ricordando comunque i limiti di questo stesso modello di partenza imposte. Le situazioni possono essere descritte introducendo informazioni supplementari sugli elementi quali il campo elettrico, e considerando eventuali uniformità del substrato. Si analizzeranno alcune situazioni.

Ipotesi di quasi-neutralità

Un'ipotesi interessante per esempio è la cosiddetta ipotesi di quasi neutralità: nel caso si possa assumere $g=0$, per vari motivi che devono essere forniti, si può considerare nullo il termine dipendente da $\mathcal{E}$. Il modello si riduce a:

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} - U_n \quad ; \quad \frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} - U_p$$

Si noti che dove il campo elettrico è nullo può non essere sufficiente: nel caso si abbia a che fare con la presenza di un elevato campo di maggioranza, anche campi elettrici molto bassi possono risultare determinanti.

Questa affermazione ci porta a qualcosa di interessante: nel caso si sia all'equilibrio termodinamico, $g=0$ di mano. Nel caso si sia fuori equilibrio:

$$g = p_0 + p' - n_0 - n' + N_D - N_A$$

Lavorando su questa espressione si può determinare l'importanza di g e lavorare di conseguenza.

Regime stazionario

Un'altra possibile situazione che può tornare molto spesso utile è il caso di "regime stazionario": questo equivale a non tener conto di transienti, direttamente considerando i valori finali degli stati elettrici. "Stazionario" significa "costante nel tempo", dunque:

$$\begin{cases} 0 = \mu_n \frac{d(n\mathcal{E})}{dx} + D_n \frac{d^2 n}{dx^2} - U_n \\ 0 = -\mu_p \frac{d(p\mathcal{E})}{dx} + D_p \frac{d^2 p}{dx^2} - U_p \\ \frac{d^2 \phi}{dx^2} = -\frac{q}{\epsilon} (p - n + N_D - N_A) \end{cases} \quad \begin{array}{l} \text{Le derivate nel tempo sono nulle,} \\ \text{dal momento che il sistema} \\ \text{è a regime.} \end{array}$$

Quasi - livelli di Fermi

Le equazioni di Maxwell-Boltzmann o Shockley valgono, almeno fino ad adesso, solo in una condizione ben definita: l'equilibrio termodinamico. Il livello di Fermi è infatti stato precedentemente definito, ma solo a patto di essere proprio in condizione di equilibrio. Ora: le equazioni di Shockley ci tornerebbero molto utili

onde fuori dall'equilibrio, dove non si potrebbe parlare di livelli di energia di Fermi. Quello che si può fare fuori equilibrio, tuttavia, è parlare di concentrazioni: si definiscono i "quasi-livelli di Fermi" come quei livelli energetici fittizi che danno quella concentrazione a quella temperatura. In questo modo:

$$n = n_i \exp\left(\frac{E_{Fn} - E_{Fi}}{k_B T}\right) \quad ; \quad p = n_i \exp\left(\frac{E_{Fi} - E_{Fp}}{k_B T}\right)$$

Sappiamo che:

$$E_{Fi}(x) = -q \phi(x)$$

$$\hookrightarrow E(x) = -\frac{\partial \phi}{\partial x} = \frac{1}{q} \frac{\partial E_{Fi}}{\partial x}$$

Dalla precedente definizione:

$$\frac{\partial n}{\partial x} = \frac{n}{k_B T} \left(\frac{\partial E_{Fn}}{\partial x} - \frac{\partial E_{Fi}}{\partial x} \right) \quad \left[\text{Ricordare che il termine esponenziale è "nascosto" in "n"} \right]$$

Usando il modello di Einstein, la corrente di diffusione vale:

$$J_{\text{diff}} = q D_n \frac{\partial n}{\partial x} = n \cdot \mu_n \left(\frac{\partial E_{Fn}}{\partial x} - \frac{\partial E_{Fi}}{\partial x} \right)$$

Sostituendo questi risultati nel modello di drift-diffusion, si ha:

$$J_n = q n \mu_n E + q D_n \frac{\partial n}{\partial x} = n \mu_n \frac{\partial E_{Fn}}{\partial x}$$

$$J_p = q p \mu_p E - q D_p \frac{\partial p}{\partial x} = p \mu_p \frac{\partial E_{Fp}}{\partial x}$$

$$\rightarrow n \cdot p = n_i^2 \exp\left(\frac{E_{Fn} - E_{Fp}}{k_B T}\right)$$

Questa ultima espressione può essere molto utile, al fine di capire quale sia il comportamento dei quasi-livelli di Fermi al variare delle condizioni:

- In equilibrio termodinamico, i quasi-livelli di Fermi degenerano nei livelli di Fermi (come è ovvio che no);
- In caso di iniezione, $n_p > n_i^2$, dunque $E_{Fn} > E_{Fp}$
- In caso di smaltimento, $n_p < n_i^2$, dunque $E_{Fn} < E_{Fp}$

Una nota finale: può essere molto utile, nel caso di regime di iniezione, verificare il fatto che esso sia un "basso" regime di iniezione: in questa situazione i maggioritari sono invariati, dunque si possono rappresentare anche solo i quasi-livelli dei maggioritari, dal momento che i minoritari sono sostanzialmente all'equilibrio.

Il fatto che essi sono costanti nei lati neutri è dovuto al fatto che la corrente di maggioritari è l'oscillazione.

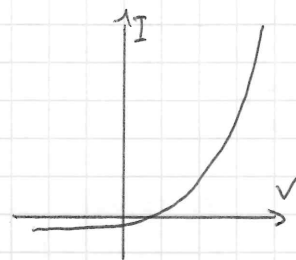
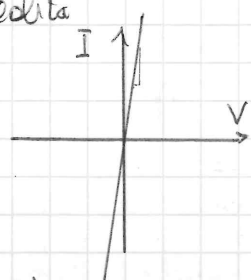
DR/FAR ALM

La giunzione metallo-semiconduttore

Si tratta di un dispositivo che si realizza unendo del metallo con del semiconduttore drogato (nel nostro esempio, tipo "n"; tipo "p" è quasi analogo).

A seconda del metallo utilizzato (caratterizzato mediante il lavoro di estrazione) e del drogaggio utilizzato, si possono ottenere due tipi di comportamenti:

- Ohmico: comportamento di un cavo circuito non ideale, dunque $R \approx 0$: vi è una componente parassita resistiva, dunque una non idealità.



- Rettificante: comportamento tipo "diodo a giunzione" (detto, in questo caso, "diode Schottky").

Questi comportamenti possono essere utili in molti campi, ad esempio a microonde.

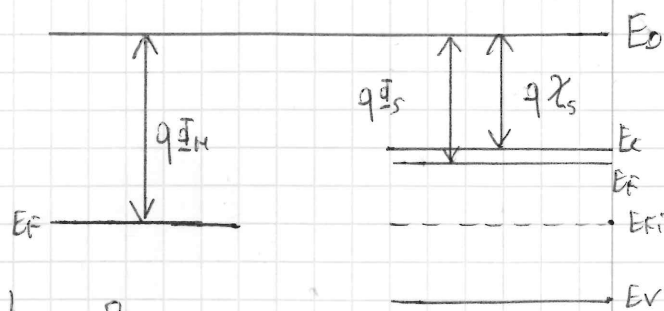
Ciò che discrimina questi comportamenti è la presenza di una barriera di energia in prossimità della giunzione. Si parla del di-

gramma a bande dei due di-

spendenti isolati; abbiamo due li-

velli di Fermi distinti, del momento che non

consideriamo per ora unite le parti. Il metallo è un conduttore ideale, dunque, se portato fuori equilibrio, resterà comunque sempre



in stato di basso regime di iniezione; ciascun metallo è caratterizzabile con il suo lavoro di estrazione: $q\Phi_H$. Per quanto riguarda il semiconduttore, esso è stato finora caratterizzato da E_c e E_v ; si ricordi che $E_0 - E_c = q\chi_s$, ma l'affinità elettronica. Dato semiconduttore tipo "n", il livello di Fermi è sulla semibanda superiore, dunque si può calcolare il lavoro di estrazione facendo: $q\Phi_s = q\chi_s + E_c - E_F$.

Questo, supponendo di avere completa ionizzazione e di essere in regione estrinseca.

Supponendo di aver conservato la periodicità del reticolo, dunque di avere un'ultima bandiera a disposizione, capita che, unendo i materiali, vi sarà una sorta di "transitorio" nel quale il dispositivo tende all'equilibrio termodinamico; vi sarà uno spostamento di particelle tale da far tendere all'annullamento dei flussi di particelle; al termine, il livello di Fermi dovrà essere costante: il materiale con più elettroni ne darà via a quello con meno elettroni, in modo da bilanciare i flussi.

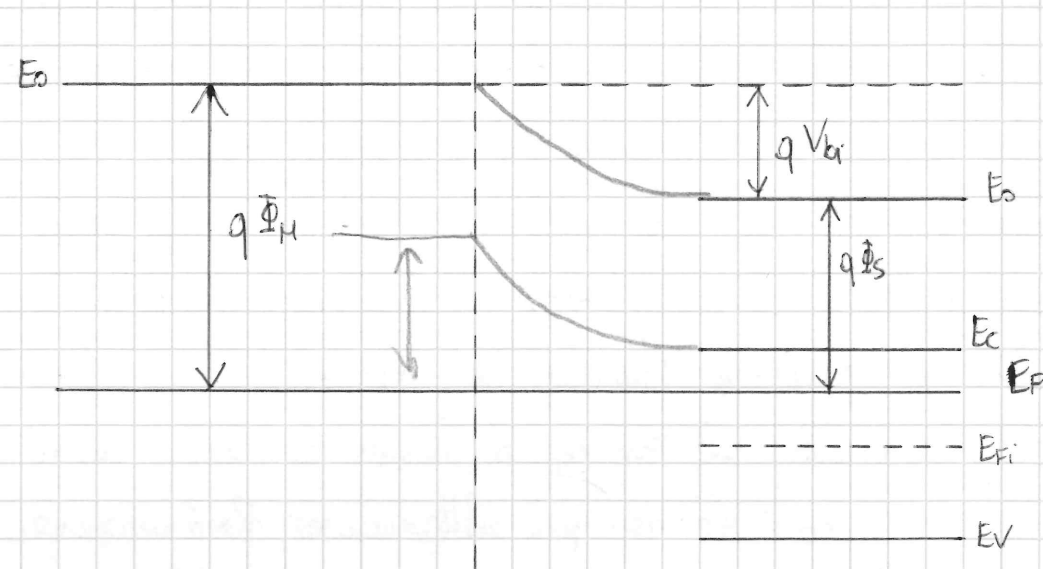
Si potrebbe scegliere un drogaggio tale da avere lo stesso E_F di un metallo, ma questo è un caso veramente raro.

Normalmente, si ha a che fare con lavori di estrazione diversi; a questo punto, 2 possibilità: lavoro di estrazione maggiore nel metallo o nel semiconduttore. Nel caso di drogaggio tipo n, "spesso" capita che $q\Phi_H > q\Phi_s$, dunque che il semiconduttore ceda elettroni al metallo; nel metallo vi sarà una regione di accumulo, di lunghezza simile alla lunghezza di Debye:

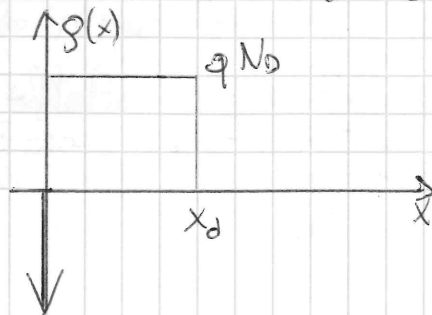
$$L_D = \sqrt{\frac{\epsilon V_T}{q n}} \approx 0,4 \text{ } \mu\text{m}$$

mentre nel semiconduttore una regione di smontamento. Essendo la lunghezza di Debye nel metallo molto ridotta, si può dire che l'accumulo di elettroni sia presente, ma solo sulla superficie: se la regione di accumulo sia in sostanza una δ di Dirac, centrata sul punto di giunzione.

Vogliamo a questo punto presentare il diagramma a bande all'equilibrio termodinamico, e la relativa costruzione.



Dato lo stato di equilibrio termodinamico, il punto di partenza è il livello di Fermi: esso è disegnato allo stato di equilibrio termodinamico, dunque costante. Lontani dalla discontinuità (giunzione) i materiali si comportano come omogenei, tenendo conto che tutti i livelli vanno traslati rispetto alla stessa traslazione del livello di Fermi, in modo da averlo costante. La domanda di cosa ha questo andamento (idealizzato): si avrà



una regione smontata, sulla quale si farà l'ipotesi di completo smontamento; al termine si ipotizza per semplicità la presenza di un salto brusco al termine della regione smontata, per dar spazio alla regione neutra.

Si vuole a questo punto proporre un'analisi quantitativa: come noto dalla fisica, tra potenziale elettrostatico ed energia vi è una ben nota relazione:

$$E = q \cdot V;$$

Dato il livello di Fermi intrinseco $E_{Fi}(x)$, funzione di x . Dal momento che esso non è costante, si può dire che:

$$E_{Fi}(x) = -q \cdot \phi(x) + E_{FiF}$$

Ormai, dato un riferimento ad-hoc E_{FiF} , si può dire che il potenziale elettrostatico $\phi(x)$, moltiplicato per la carica di un elettrone (dunque negativa), aggiungendo un opportuno riferimento E_{FiF} , sia uguale a $E_{Fi}(x)$. Vorremmo usare $E_{Fi}(x)$ e lo studio della sua variazione per determinare qV_{bi} ; per far ciò, è sufficiente vedere che E_{Fi} varia, a causa dell'equilibrio della giunzione, proprio di qV_{bi} ; possiamo dunque dire che:

$$qV_{bi} = E_{Fi}(0) - E_{Fi}(x_0)$$

Ormai, se sia uguale alla differenza di valore dei livelli di Fermi, nei due estremi. Si può vedere che:

$$E_{Fi}(0) - E_{Fi}(x_0) = [-q \phi(0) + E_{FiF}] - [-q \phi(x_0) + E_{FiF}] = q [\phi(x_0) - \phi(0)].$$

Dunque:

$$V_{bi} = \phi(x_0) - \phi(0) = \phi(x_0) - \phi(x_0)$$

non si divide per $-q$!

(*) Sostituendo $\rightarrow$ d. coduti $\rightarrow$ carica POSITIVA.

(1) da quando parte

La funzione di densità di carica è nota: sapendo che, in zona estinseca, si ha come concentrazione di porticelle N_D , la densità di carica sarà N_D per il valore della carica, dunque $q N_D$; da (*) qui, con l'equazione di Poisson, è possibile ottenere la funzione potenziale $\phi(x)$, integrando.

Si riprende il teorema di Gauss, ricordando, assieme a Poisson:

$$\frac{dE}{dx} = \frac{\rho}{\epsilon} \quad ; \quad \rightarrow E(x) = \int \frac{\rho}{\epsilon} dx + c =$$

$$= \int \frac{q N_D}{\epsilon} dx + c = \frac{q N_D}{\epsilon} x + c$$

Imponendo il fatto che, per $x = x_d$, $E = 0$, si ottiene:

$$E(x_d) = 0 = \frac{q N_D}{\epsilon} x_d + c \rightarrow c = -\frac{q N_D}{\epsilon} x_d$$

Dunque:

$$E(x) = \frac{q N_D}{\epsilon} [x - x_d]$$

Si integri nuovamente:

$$\phi(x) = K + \int E(x) dx = K + \int \frac{q N_D}{\epsilon} [x - x_d] dx = K + \int \left[\frac{q N_D}{\epsilon} x - \frac{q N_D}{\epsilon} x_d \right] dx =$$

$$= K + \left[\frac{q N_D}{2\epsilon} x^2 - \frac{q N_D}{\epsilon} x_d x \right]$$

Imponendo, per $x = x_d$, che $\phi(x) = 0$, si ha:

$$0 = K + \left[\frac{q N_D}{2\epsilon} x_d^2 - \frac{q N_D}{\epsilon} x_d^2 \right] = K + \frac{q N_D}{2\epsilon} x_d^2 = 0 \rightarrow K = -\frac{q N_D}{2\epsilon} x_d^2$$

Dunque, sostituendo K :

$$\phi(x) = \frac{q N_D}{2\epsilon} [x_d^2 - x^2]$$

Sn base alla teoria precedentemente introdotta:

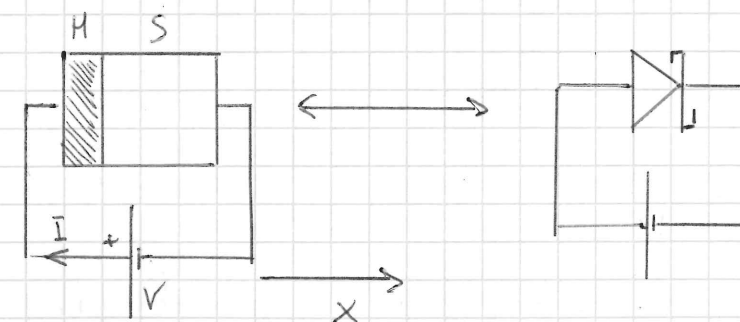
$$-(\phi(0) - \phi(x_d)) = \frac{q N_D}{2\epsilon} x_d^2 - 0 = V_{bi}$$

$$\rightarrow x_d = \sqrt{\frac{2\epsilon}{q N_D} V_{bi}}$$

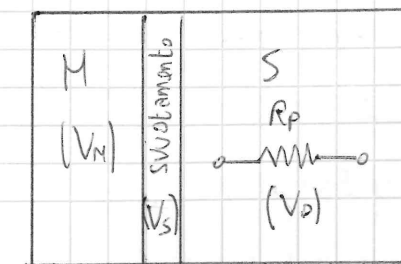
Da qui, volendo si potrebbe sostituire x_d in $E(x)$ e ricavare l'espressione del campo.

Caratteristica Statica

Si vuole a questo punto proporre un'analisi statica, fuori equilibrio, della giunzione metallo-semiconduttore. Si inizierà con un'analisi qualitativa, per poi quantificare.



La corrente entra nel metallo, attraversa la barriera, la regione di svuotamento, passa per la regione neutra, dunque esce dall'estremo opposto, chiudendo il circuito. Ci sono dunque 3 regioni: metallo, contatto, semiconduttore, in serie, dal momento che, su di esse, scorre la stessa corrente.



In ciascuna zona vi sarà una caduta di tensione: V_n , V_s , V_p la somma delle 3 tensioni deve essere pari a V_0 .

Ora: se il metallo è ideale, la tensione V_n sarà nulla; il semiconduttore neutro avrà al suo interno un comportamento resistivo, dunque:

$$V_p = R_p \cdot I$$

Supponendo, in prima approssimazione, di avere I molto piccolo, si considera $V_p \ll V_s$, con ciò $V \approx V_s$.

Poi: la corrente di giunzione, come noto, è una corrente di elettroni, dal momento che le lacune sono ininfluenti; i flussi di elettroni saranno 2: quelli che dal metallo vanno verso il semiconduttore, e quelli che dal semiconduttore vanno verso il metallo.

Dal metallo al semiconduttore, si ha esclusivamente un'emissione termionica: più la temperatura T è alta, più la coda della Fermi-Dirac permette di far passare elettroni ad alti livelli energetici, come suggerisce l'equazione di Richardson:

$$J_{in} = A_0 T^2 \exp\left(-\frac{q \Phi_B}{k_B T}\right), \quad A_0 = \left[\begin{array}{l} \text{costante di} \\ \text{Richardson} \end{array} \right]$$

(1) Si ha dipendenza da $q \Phi_B$: essa è la barriera data dal fatto che, per gli elettroni nel metallo, è necessario avere, per arrivare al semiconduttore, un'energia pari al E_0 conseguente allo smottamento. Come si può quantificare? Beh, sappiamo quanto vale $q \Phi_M$, e che E_0 data, energeticamente, $q \chi_s$ del livello E_0 . Per differenza:

$$q \Phi_B = q \Phi_M - q \chi_s.$$

Si può dunque ora, quantificare la corrente nel metallo, diretta verso il semiconduttore.

Introduciamo a questo punto un elemento di non-equilibrio: una tensione $V \neq 0$; date le approssimazioni fatte, V viene applicata solo sulla regione di smottamento del semiconduttore; distinguiamo a questo punto due casi:

- $V < 0$: in questo caso si aumenta la caduta di potenziale nella regione di smottamento del semiconduttore, allargandola; questa è detta "polarizzazione inversa"; la barriera $q \Phi_B$ è sempre uguale, dal momento che né $q \Phi_M$ né $q \chi_s$ cambiano il proprio valore, dunque la corrente di elettroni da metallo a semiconduttore non viene alterata; essendoci però elevata la differenza tra i livelli del metallo, diventando $q(V_0 - V)$,

il flusso di portiche dal semiconduttore si riduce, dal momento che serve un'energia maggiore per effettuare il "salto". Ora, ovviamente, siamo fuori dall'equilibrio termodinamico. Ci si può aspettare, per ora intuitivamente, che, con $V < 0$, si avranno correnti molto ridotte.

- $V > 0$: per quanto riguarda il flusso da metallo a semiconduttore nulla da dire; per quanto riguarda l'altro, da dire ce n'è: ora la barriera di energia è ridotta, dunque la larghezza della regione smottata si restringe. Si può dire che gli elettroni passino più facilmente, dal momento che il "salto energetico" è meno "faticoso".

Un'osservazione: se si alza troppo la tensione, secondo il modello, si può pensare di ottenere a un certo punto barriera nulla, e addirittura "negativa"; questo in realtà non capita, perché in questo caso il semiconduttore andrebbe in regione di accumulo, dunque verso l'alto livello di iniezione; in tal caso, I varrebbe essere "grande", e il termine $V_p = R_p \cdot I$ non più trascurabile. Il modello usato non varrebbe più essere valido. Volendo raffinare il modello per il caso di polarizzazione diretta forte, si avrebbe:

$$q(V_a - V) = q[V_a - (V_i - R_p I)] = q(V_a - V) + q R_p I$$

Si veda che I ha un andamento esponenziale e dunque, prima o poi, diventa dominante.

Analisi quantitativa

Si supponga, da ora, di essere in stato di basso regime di iniezione, date le precedenti enumerazioni; essendo i portatori solo elettroni, I_{ns} , ma la corrente provocata dal flusso di elettroni dal metallo verso il semiconduttore, dal momento che si muove nello stesso verso della corrente ma con carica negativa, sarà minore di 0; data I_{sn} la corrente duale (da semiconduttore a metallo), $I = I_{sn} - |I_{ns}|$.

Si suppone (usando le enumerazioni e le dimostrazioni precedenti) che I_{ns} sia costante rispetto a V .

Si vuole dunque quantificare la corrente "importante": I_{sn} .

Essa sarà proporzionale alla quantità di elettroni in $x=0$ dunque:

$$n(0) = N_c \exp\left(-\frac{q\Phi_B}{k_B T}\right) \quad \text{Ricordando che } q\Phi_B = E_c - E_F.$$

Si consideri a questo punto un punto $x_0 \gg x_d$, dunque in piena zona neutra; si può usare Boltzmann qui, e dire:

$$n(x_0) = N_c \exp\left(-\frac{E_c(x_0) - E_F}{k_B T}\right) = N_D$$

Quindi, considerando che i portatori derivano prevalentemente dal drogaggio. Ricavando N_c da questa espressione nota, si ottiene:

$$n(0) = N_D \exp\left(\frac{E_c(x_0) - E_F - q\Phi_B}{k_B T}\right) = N_D \exp\left(-\frac{qV_{bi}}{k_B T}\right)$$

Ricordando infatti che $q\Phi_B = q\Phi_n - q\Phi_p$, o che $q\Phi_p + E_c = -E_F$,
 $-qV_{bi} = q\Phi_p - q\Phi_n$.

Inoltre, si ha che $\frac{q}{k_B T} = \frac{1}{V_T}$, dunque:

$$\therefore \left(-|I_{ns}| + I_{sn} \right) = I_s \left(\exp\left(\frac{V}{V_T}\right) - 1 \right)$$

Dal momento che:

$$I_{sn} \propto n(0) = N_D \exp\left(-\frac{qV_{bi} - qV}{k_B T}\right)$$

e

$$I = I_{sn} - |I_{ns}| \propto N_D \exp\left(-\frac{qV_{bi} - qV}{k_B T}\right) - N_D \exp\left(-\frac{qV_{bi}}{k_B T}\right)$$

Questo, a meno di coefficienti di proporzionalità.

MA LA CORREZIONE COME SI FA?

La teoria appena presentata, detta "teoria di Schottky", è molto semplificata, dal momento che non tiene conto di un fatto: gli elettroni possono non solo "scandire" la barriera, ma anche passare attraverso, mediante l'effetto tunnel.

Se la barriera in questione è molto sottile, dunque se la regione smistata è molto sottile, cosa ottenibile con un semiconduttore fortemente drogato, aumenterà questa probabilità di attraversamento, fino a diventare il fenomeno prevalente (oltre tutto, in modo bidirezionale, non solo da semiconduttore a metallo!); questo è il principio su cui si basano i contatti ohmici dei semiconduttori.

Un "bonus" per il silicio: l'alluminio "Al" è un drogante tipo p per il silicio, dunque, metallizzando con esso il silicio, si ottiene "gratuitamente" un discreto contatto ohmico, con una zona molto drogata tipo p.

Effetti capacitivi

Per ora è stato applicato uno studio di tipo stazionario; nella realtà, come noto, i dispositivi funzionano dinamicamente; tenendo conto di ciò, il quadro subirà delle modifiche: una delle cause della latenza dei dispositivi sono le capacità parassite. Nei dispositivi, infatti, vi è almeno la presenza di una giunzione, di una regione di smistamento; le regioni smistate corrispondono alla presenza di carica accumulata in una parte della struttura. Dal momento che la regione di smistamento ha carica diversa da 0, l'applicazione di una tensione va a cambiare la

"base" del "rettangolo", dunque la carica presente: Q_d dunque è funzione della tensione $v(t)$, variabile nel tempo:

$Q_d = Q_d[v(t)]$, dove si ha: $Q_d[v(t)] = \sqrt{\frac{2\epsilon}{qN_D}} [V_{bi} - v(t)]$
Considerando il fatto che, nel lato trasversale, si ha una sezione A , una larghezza l , dunque la carica Q_d sarà:

$$Q_d = A q N_D x_d[v(t)]$$

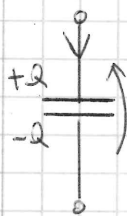
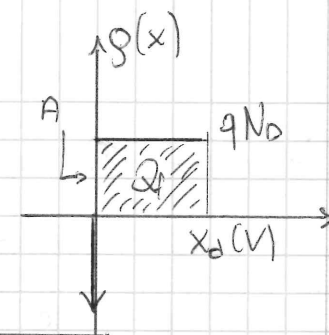
Si può dunque esprimere la corrente che passa nella struttura, a partire dalla variazione di questa carica nel tempo:

$$i(t) = -\frac{dQ_d}{dt} = -\frac{dQ_d}{dv} \cdot \frac{dv}{dt} = C_s[v(t)] \cdot \frac{dv}{dt}$$

La presenza del segno "-" dipende da questo fatto: secondo la convenzione di utilizzatore, si accumula, sul punto a potenziale positivo, una carica $+Q$, sull'altro $-Q$. Nel nostro caso però la carica accumulata sul punto a potenziale positivo è negativa, dunque si avrà inversione di segno, per rispettare la convenzione. La capacità di smistamento, associata alla regione smistata, sarà:

$$C_s[v(t)] = -\frac{dQ_d}{dv} = -q A N_D \frac{dx_d}{dv} = A \sqrt{\frac{q \epsilon N_D}{2 [V_{bi} - v(t)]}}$$

Questa espressione presenta un sintomo "fittizio": esso è tale dal momento che, a causa di V_{Rp} , non si sta includendo nel modello effetti parassiti, dunque, se si andasse in regione di alta iniezione, si dovrebbe raffinare a poco il modello usato.

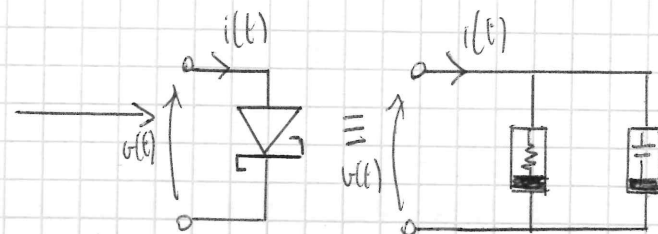


Modello di ampio segnale

Considerando ambo i contributi della caratteristica statica e della componente dinamica, si può, tenendo conto dell'applicazione di una tensione $v(t)$ ai capi del dispositivo, dire che la corrente ne è composta da due contributi:

$$i(t) = i_{dc}(t) + C_s[v(t)] \frac{dv(t)}{dt}$$

$$\text{Dove } i_{dc}(t) = I_s \left[\exp\left(\frac{v(t)}{V_T}\right) - 1 \right]$$



È possibile determinare questo modello circuitale non lineare parallelo, grazie al fatto che $i(t)$ è data dalla somma di due contributi. Questo modello di ampio segnale è valido, a meno che il segnale non sia tale da mandare in alta iniezione il dispositivo.

Modello di piccolo segnale

Si consideri a questo punto una linearizzazione del precedente modello: supponendo che sia tensione sia corrente siano uguali alla somma di due contributi, tali per cui le costanti con cui si linearizza siano pari alla caratteristica statica e al solo segnale, reperendo dunque il bias point e il punto di funzionamento istantaneo, se $v_{ss}(t)$ è molto minore di V_0 , possiamo sviluppare intorno a V_0 mediante Taylor, ottenendo:

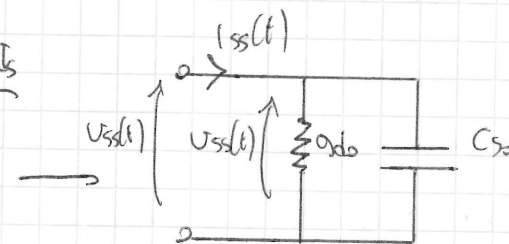
$$i_{dc}(v(t)) = i_{dc}[V_0 + v_{ss}(t)] \approx i_{dc}(V_0) + \left. \frac{\partial i_{dc}}{\partial v} \right|_{v=V_0} v_{ss}(t)$$

$$Q_f(v(t)) = Q_f[V_0 + v_{ss}(t)] \approx Q_f(V_0) + \left. \frac{\partial Q_f}{\partial v} \right|_{v=V_0} v_{ss}(t)$$

Si definiscono dunque i parametri del modello di piccolo segnale come:

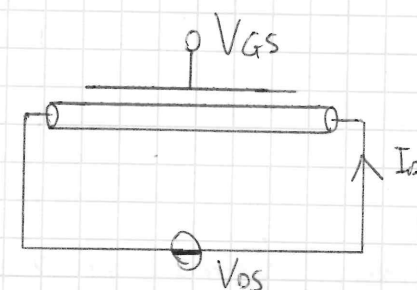
$$g_{do} = \left. \frac{\partial i_{dc}}{\partial v} \right|_{v=V_0} = \frac{I_s}{V_T} \exp\left(\frac{V_0}{V_T}\right) = \frac{I_0 + I_s}{V_T}$$

$$C_{so} = \left. \frac{\partial Q_f}{\partial v} \right|_{v=V_0} = A \sqrt{\frac{q \epsilon N_D}{2[V_{bi} - V_0]}}$$



Sistema MOS - transistor MOSFET

Un transistor a effetto di campo è un dispositivo in cui il controllo dell'intensità della corrente in uscita si ottiene modificando il numero di portatori liberi nel canale di conduzione.



La corrente I_D si modula con V_{GS} e V_{DS} ; la prima, nella fattispecie, è in grado di cambiare il numero di portatori liberi che il canale ha a disposizione.

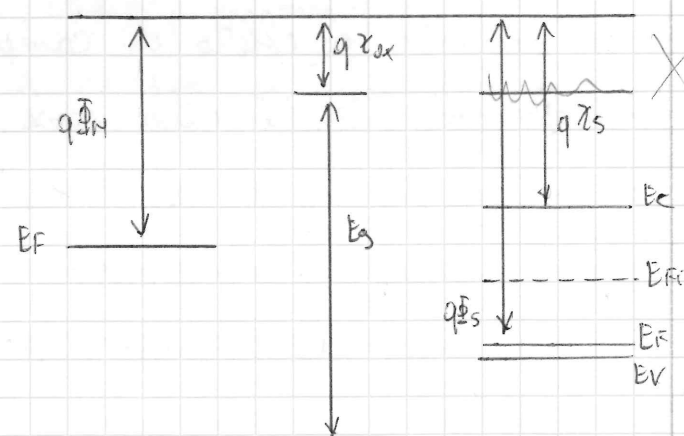
Il MOSFET è un'implementazione di questa idea, su di una struttura MOS: metallo-ossido-semiconduttore. M è un metallo, o, a volte, un polisilicio cristallino molto drogato; D è un

isolato, un tempo fondo di silicio, ora in via di sostituzione, a causa dell'effetto tunnel; S un semiconduttore drogato (nel caso se noi consideriamo prudentemente, drogato tipo p).

Quo che differenzia questa struttura da un band condensatore è la presenza di quest'ultimo strato semiconduttivo, detto "bulk"; il cando conduttivo sarà tra l'interfaccia ossido-bulk. Nel caso come "metallo" si ha il polinomio, esso deve essere talmente drogato da avere il livello di Fermi nella banda conduttiva, in modo da avere un comportamento analogo a quello di un metallo.

Si consideri il diagramma a bande della struttura non unita:

Questa è un'idea di come si trova, prima del "transitorio", il sistema; al termine di questo, si avrà equilibrio termodinamico, corrente 0, dunque livello di Fermi costante.



Al centro della struttura si ha l'ossido, il quale impedisce lo scambio di elettroni necessario a raggiungere l'equilibrio: la parte con lavoro di estrazione inferiore dovrebbe infatti cedere all'altra elettroni, in modo da agevolare i conti, ma l'ossido non lo permette; non essendoci dunque né correnti di trasnamento né di diffusione, il bilanciamento globale dei flussi sarà da attribuirsi ad una corrente di ricombinazione, modellabile mediante una corrente di spostamento; fisicamente, in caso di struttura non neutra,

la ricombinazione va a prendere sulla ~~generazione~~ ^{generazione} tecnica, permettendo la neutralità globale all'equilibrio.

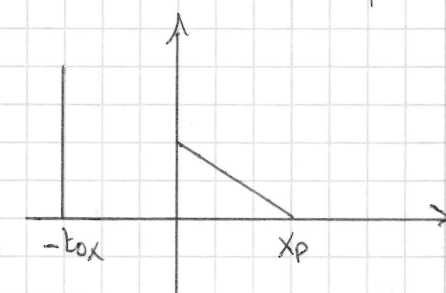
Nel semiconduttore si avrà una regione di smottamento, e nel metallo la S , causata dalla modellazione del metallo come una sorta di "semiconduttore molto drogato":

" $-t_{ox}$ " è il punto in cui l'onda termina, e il metallo inizia.

$$Q(-t_{ox}) = q N_A x_p S(t_{ox}).$$

Una curiosità: se aumenteremo il drogaggio del semiconduttore, la regione di smottamento si stringerebbe o allargherebbe, tendendo ad una $S(x)$; in questo caso la struttura degenererebbe in un band condensatore.

Si costruisca a questo punto $E(x)$ a partire da $S(x)$: per quanto riguarda il semiconduttore, nulla di nuovo; per quanto riguarda l'ossido, ne parleremo dopo; su $-t_{ox}$ si ha la S di Dirac, dunque campo discontinuo a sinistra di $-t_{ox}$, nullo. E l'ossido? Beh, nell'ossido, il campo elettrico sarà costante, dal momento che si può pensare come generato dalla S , ma... quanto vale?



Recuperando le nozioni di Campi Elettromagnetici, una delle condizioni al contorno che vengono ordinariamente applicate alle equazioni di Maxwell è:

$$\epsilon_1 \underline{E}_1 \cdot \hat{n}_1 = \epsilon_2 \underline{E}_2 \cdot \hat{n}_2 \rightarrow \underline{D}_1 \cdot \hat{n}_1 = \underline{D}_2 \cdot \hat{n}_2$$

Dunque, nel caso in cui sussista la presenza (come in questo caso) di una discontinuità di mezzi materiali, con coefficienti dielettrici relativi ϵ_1 e ϵ_2 , il vettore di spostamento dielettrico si mantiene costante: $\underline{D}_{ox} = \underline{D}_s$.

Integrando ancora, si può ricavare il potenziale $\phi(x)$ e da esso il diagramma a bande.

Si definisce a questo punto la "tensione di banda piatta"

V_{FB} come:

$$q|V_{FB}| = q\phi_s - q\phi_n$$

Da:

$$V_{FB} = \phi_n - \phi_s$$

Si noti che questa è una grandezza con segno; si può stimare che essa valga:

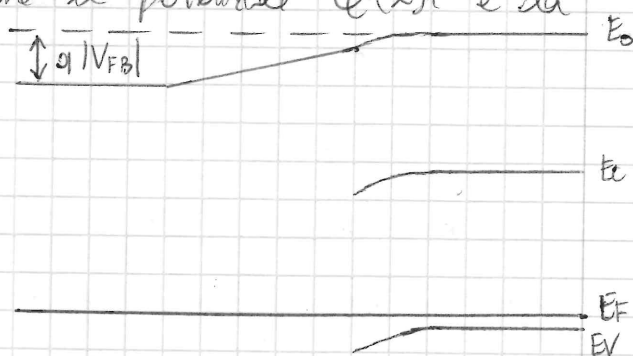
$$V_{FB} \approx -1V.$$

Esistono formulazioni in grado di approssimare questo valore.

Sistema MOS fuori equilibrio

Dal momento che si ha l'ossido, la corrente di gate (ovvia del metallo) sia zero; dunque, i quasi-livelli di Fermi della struttura devono essere costanti: i quasi-livelli derivano dal modello di deriva-diffusione e, essendo la loro densità nulla, essi devono essere costanti, nonostante siamo fuori equilibrio (la corrente è infatti dovuta alla ricombinazione).

Dal punto di vista dell'energia potenziale, del diagramma a



bande, il gate ha energia potenziale più bassa del semiconduttore; se applichiamo una tensione $V \geq 0$ al gate, la barriera si alza (si ricordi che l'energia varia proporzionalmente al potenziale cambiato di segno): $q(V_{GB} - V_{FB})$

Il motivo per cui V_{FB} è stata definita con segno, è il fatto che, se $V_{GB} = V_{FB}$, si ha la banda piatta.

Attenzione: ora, la situazione di "banda piatta" esiste, ed è possibile; perché? Beh, grazie all'ossido: dal momento che ora c'è l'ossido, non c'è più la corrente di prima andata a creare il termine V_p (nella giunzione metallo-semiconduttore), dunque non si hanno più le precedenti limitazioni di modello.

Parliamo di quasi-livelli di Fermi: la differenza tra quello del metallo e quelli del semiconduttore è coincidente al "salto" necessario per mettere allo stesso livello i livelli di Fermi. Perché però si ha $E_{Fn} - E_{Fp}$? Beh, ricordiamo che i quasi-livelli sono un'espressione delle particelle libere; in stato di banda piatta, la regione smodata è stata "riempita", dunque, per ogni x maggiore di 0, si ha:

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

Supponendo che $p = N_A$, dal momento che ci si immagina il fatto che è il drogante a prender parte alla conduzione, si può dire che, ricordando le espressioni di Shockley fuori equilibrio:

$$n \cdot p = \frac{n_i^2}{N_A \cdot N_A} = \left[n_i \exp\left(\frac{E_{Fn} - E_{Fp}}{kT}\right) \right] \cdot \left[n_i \exp\left(\frac{E_{Fp} - E_{Fn}}{kT}\right) \right]$$

Occhio: in banda piatta, $n = N_A$; si può ricavare n_i dallo due moltiplicam. Di regola guarda quad. e segni suggerendo.

$$\rightarrow n_i^2 = n_i^2 \exp\left(\frac{E_{Fn} - E_{Fp}}{kT}\right)$$

Perché l'equazione sia soddisfatta, $E_{Fn} = E_{Fp}$.

Presenza di carica nell'ossido

Nella realtà, a causa di ragioni tecnologiche, l'ossido potrebbe non essere perfetto, nel senso che la carica al suo interno potrebbe essere non nulla. È possibile infatti avere cariche localizzate, forse, le quali possono partecipare all'elettrostatica del problema. Data una Q_0 concentrata in un punto dell'ossido, si ha una "variazione della tensione di banda piatta": la presenza del campo elettrico generato da questa carica può essere modellata con una ΔV_{FB} , introducendo dunque qui l'effetto della carica dell'ossido.

Avere la situazione di banda piatta corrisponde ad avere $g=0$ in ogni punto, dunque neutralità globale ma anche locale. Se si ha questa Q_0 nel metallo, all'interfaccia con l'ossido, si avrà una carica di compensazione $-Q_0$, che porterà a una "banda piatta fittizia". Si noti che il diagramma a bande non sarà più piatto, a causa di questa carica, anche se l'effetto sarà presente solo nell'ossido.

Caso più complicato: presenza di carica localizzata distribuita:

$$g(y) = g_0(y) + Q_{int} \delta(y - t_{ox})$$

In questo caso, come si può lavorare? Beh: usando la linearità dell'equazione di Poisson, applicando dunque il principio di sovrapposizione degli effetti, e integrando i contributi. Quello che si può fare è considerare infiniti ottomigli infinitesimi di forma:

$$d(\Delta V_{FB}) = \frac{1}{\epsilon_{ox}} g(y) dy \cdot y$$

$$\rightarrow \Delta V_{FB} = \int \left[\frac{1}{\epsilon_{ox}} y g(y) \right] dy$$

È da qui si può determinare la nuova forma di ΔV_{FB} :

$$\Delta V_{FB} = -\frac{1}{\epsilon_{ox}} \int_0^{t_{ox}} y g_0(y) dy - \frac{t_{ox}}{\epsilon_{ox}} Q_{int}$$

Questo "trucco", usato "volontariamente", è utile per modificare "a proprio piacimento" la tensione di banda piatta del sistema MOS e quella che vedremo in seguito essere la "tensione di soglia".

Regioni di funzionamento del sistema MOS

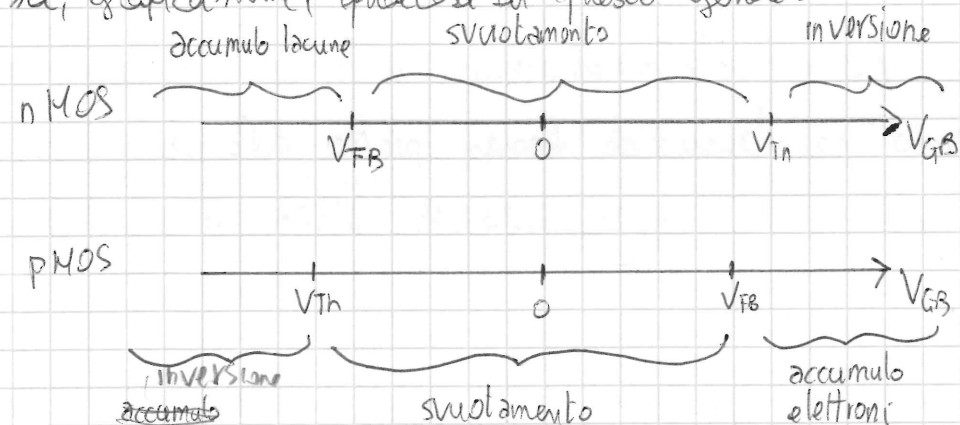
Sia per i sistemi MOS tipo "n" (dunque substrato tipo "p") è possibile identificare tre regioni di funzionamento, di cui una risulterà essere particolarmente interessante. Si consideri per ora come "riferimento" la tensione di banda piatta, V_{FB} (positiva nei pmos, negativa nei già noti nmos):

- Nel caso si applichi una tensione minore di V_{FB} , il semiconduttore funziona in regione di accumulo di maggioritari;
- In un intorno di $V_{FB}=0$, compreso tra la V_{FB} e la ancora non nota V_{th} (tensione di soglia), il semiconduttore lavora nella regione che per ora conosciamo meglio: quella di svuotamento. A questa regione appartiene anche lo stato di regime termodinamico (equilibrio).
- La regione di funzionamento più interessante è quella di inversione: quando la carica accumulata non è solo dovuta allo svuotamento,

ma soprattutto ad una carica di elettroni liberi sull'interfaccia, il "tubo" del nostro "modello" iniziale si riempie, e può condurre corrente, prima applicazione di una differenza di potenziale.

Nel sistema duali al nMOS, ora il pMOS (con substrato drogato tipo "n"), come già detto la tensione di banda porta è positiva mentre quella di soglia, come vedremo, negativa.

Si ha, graficamente, qualcosa di questo genere:



Passiamo alla quantificazione; scegliamo, come già fatto:

$$-q \phi(x) = E_F(x) - E_F(+\infty)$$

Dal momento che, guardando il dia-

gramma a bande, le linee sono incurvate

"verso il basso", la regione di funzionamento sarà o lo svuotamento o l'inversione. Come noto, il potenziale soddisfa l'equazione di Poisson:

$$\frac{d^2 \phi}{dx^2} = \frac{-q}{\epsilon_s} [p(x) - n(x) + N_D^+(x) - N_A^-(x)]$$

Se riusciamo a determinare $\phi(x)$, potremmo avere informazioni su E_F , dunque sulle concentrazioni di carica.

Si ricordi a questo punto che $p(x)$ e $n(x)$ possono essere espresse

come:

$$p(x) = n_i \exp\left(\frac{E_F(x) - E_{FP}}{k_B T}\right) \quad ; \quad n(x) = n_i \exp\left(\frac{E_{FP} - E_F(x)}{k_B T}\right)$$

Si considerino i numeratori degli esponenti (esponenti): sommando e sottraendo $E_F(+\infty)$, si ottiene:

$$E_F(x) - E_{FP} = E_F(x) - E_F(+\infty) + E_F(+\infty) - E_{FP} = -q \phi(x) + q \Phi_P$$

Sostituendo:

$$p(x) = n_i \exp\left(\frac{-q \phi(x) + q \Phi_P}{V_T}\right) \quad ; \quad n(x) = n_i \exp\left(\frac{q \phi(x) - q \Phi_P}{V_T}\right)$$

Ricordando che:

$$\frac{\Phi_P}{V_T} = \ln\left(\frac{N_A}{n_i}\right) \rightarrow \frac{N_A}{n_i} = \exp\left(\frac{\Phi_P}{V_T}\right)$$

Sostituendo:

$$p(x) = N_A \exp\left(\frac{-q \phi(x)}{V_T}\right) \quad ; \quad n(x) = \frac{n_i^2}{N_A} \exp\left(\frac{q \phi(x)}{V_T}\right)$$

Per quanto riguarda gli altri termini, $N_A^-(x)$ ecc., ricordando che siamo in regione neutra, si ha:

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

Dunque:

$$N_A - \frac{n_i^2}{N_A} = N_D^+ - N_A^- = 0 \quad \longrightarrow \quad N_D^+ - N_A^- = \frac{n_i^2}{N_A} = N_A$$

Si usa questa ora anche nella regione di svuotamento, supponendo che la ionizzazione sia completa e che non dipenda dalla tensione

applicata:

$$\frac{d^2 \phi}{dx^2} = -\frac{q}{\epsilon_s} \left[N_A \exp\left(-\frac{\phi(x)}{V_T}\right) - \frac{n_i^2}{N_A} \exp\left(\frac{\phi(x)}{V_T}\right) + \frac{n_i^2}{N_A} - N_A \right]$$

Si noti che le $[\]$ contengono una funzione $f(\phi)$, dipendente dal potenziale ϕ . Dunque:

$$\frac{d^2 \phi}{dx^2} = f(\phi); \quad \left[\begin{array}{l} \text{Moltiplichiamo a destra e a sinistra} \\ \text{per } \frac{d\phi}{dx} \end{array} \right].$$

$$\rightarrow \frac{d^2 \phi}{dx^2} \cdot \frac{d\phi}{dx} = f(\phi) \cdot \frac{d\phi}{dx} \quad \left[\begin{array}{l} \text{Integriamo ora ambo i} \\ \text{membri} \end{array} \right]$$

$$\rightarrow \int_x^{\infty} \frac{d^2 \phi}{dx^2} \frac{d\phi}{dx} dx = \int_x^{\infty} f(\phi) \frac{d\phi}{dx} dx$$

A "sinistra", si effettua una sostituzione del tipo:

$$\begin{aligned} z = \frac{d\phi}{dx} &\rightarrow dz = \frac{d^2 \phi}{dx^2} dx \Rightarrow \int_{\frac{d\phi}{dx}}^0 z dz = \\ &= -\frac{1}{2} \left(\frac{d\phi}{dx} \right)^2 = -\frac{1}{2} (-E)^2 = -\frac{1}{2} E^2 \end{aligned}$$

L'estremo di integrazione superiore è 0 perché la derivata del potenziale per $x \rightarrow \infty$ dunque in una regione a banda piatta, è nulla (pendenza nulla).

Il secondo membro è più semplice da trattare:

$$\Rightarrow \int_x^{\infty} f(\phi) \frac{d\phi}{dx} dx = \int_{\phi(x)}^{\phi(\infty)} f(\phi) d\phi$$

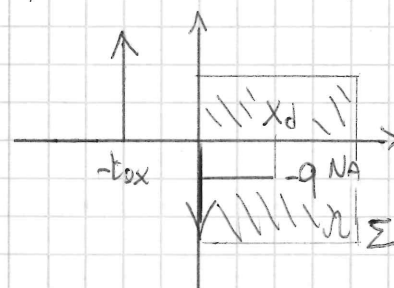
Bisogna calcolare l'integrale indefinito delle $[\]$, cosa fattibile

ma "lunga"; si propone il risultato finale:

$$\phi(x) = \mp \sqrt{2} \frac{V_T}{L_D} \left\{ \exp\left(-\frac{\phi(x)}{V_T}\right) + \frac{\phi(x)}{V_T} - 1 + \frac{n_i^2}{N_A^2} \left[\exp\left(\frac{\phi(x)}{V_T}\right) - \frac{\phi(x)}{V_T} - 1 \right] \right\}^{\frac{1}{2}}$$

Due soluzioni di $\phi(x)$ (l'equazione era di II° grado in E), di cui quella col "-" vale in regione di accumulo, quella col "+" in svuotamento e inversione. L_D è la lunghezza di Debye nel substrato. Da relazione ricavata unisce, lega, il campo elettrico E al potenziale ϕ .

Quello che si può fare a questo punto è cercare di legare la carica al campo, mediante il teorema di Gauss in forma integrale. Usando una "scatola" che racchiuda la regione svuotata del semiconduttore e parte di quella neutra, si potrebbe vedere che l'unico punto che dà un contributo non nullo al flusso del campo è un intorno del punto $x=0$; si sa che:



$$\oint_{\Sigma} \underline{D} \cdot \underline{n} d\sigma = \epsilon_s \oint_{\Sigma} \underline{E} \cdot \underline{n} d\sigma = \int_V \rho dV = Q_S$$

Il flusso del campo è dato dai vari contributi che il campo elettrico dà; si può dire che il campo elettrico $E(x)$ tenga conto di tutti i contributi da $-\infty$ a x , andando a ritroso (si parla di contributi di carica sommati, come suggerisce il teorema di Gauss in forma integrale prima riportata); il punto x che tenga conto di tutto ciò che sta dopo, dunque, è $x=0$, e dunque $E_{x=0}$ (inteso come risultato dell'integrazione).

Cono si vede dalla figura, i contributi di carica sono di due fonti: la $\delta(D-x) = \delta(x)$, di ampiezza Q_n , indicante la carica formata per inversione, o la Q_d , "depletion", o la somma di tutti i contributi della carica di spostamento. La carica totale nel semiconduttore Q_s sarà la somma di queste due:

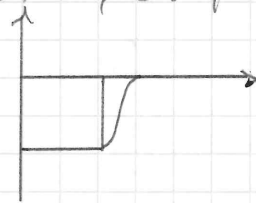
$$Q_s = Q_n + Q_d = -E_s E_{s0}$$

Questo ultimo passaggio è dovuto all'applicazione del teorema di Gauss, che lega la carica allo spostamento dielettrico. Considerando dunque il punto $x=0$, definito $V_s = \phi(0)$ come il potenziale che cade nel semiconduttore, si può dire che la carica nel semiconduttore, collegata alla precedente relazione usata tra E e ϕ , vale:

$$Q_s = \pm \sqrt{2} \frac{E_s V_T}{L_D} \left\{ \exp\left(-\frac{V_s}{V_T}\right) + \frac{V_s}{V_T} - 1 + \frac{n_i^2}{N_A^2} \left[\exp\left(\frac{V_s}{V_T}\right) - \frac{V_s}{V_T} - 1 \right] \right\}^{\frac{1}{2}}$$

Essendo il segno "-" nella relazione tra Q_s e E_s , ora "+" indica la regione di accumulo, "-" le regioni di spostamento e inversione. L'espressione di Q_s tiene conto del fenomeno di inversione, cosa che si fa sentire soprattutto per V_s elevate: il solo modello di spostamento suggerirebbe una crescita molto lenta della carica, ma quello che invece si ha è una crescita esponenziale, dovuta al prevalere dell'inversione sullo spostamento. Il punto "critico" è $V_s = 2\Phi_F$: questo, in prima approssimazione, è il punto di confine tra la debole e la forte inversione. Oltre questo punto dunque V_s rimane fissa a $2\Phi_F$, dal momento che prima di questo punto il semiconduttore riusciva a "prendersi in carico" bene il potenziale di contatto, e dopo questo "limite" lascia finire l'opera

all'angolo (tensione V_{ox}), in grado di sostenere, forte delle proprie caratteristiche dielettriche. Si aggiungerà eventualmente altro. Oltre a questo effetto, qualificando il confronto tra l'altitudine mobile e quello di spostamento, si può notare che, in un intorno di $V_s = 0$, si ha un discostamento; questo è dovuto al fatto che, per $V_s \approx 0$, cade l'approssimazione di completo spostamento, dunque la "netta transizione" a scapito di quella "mitigata", non è più una buona approssimazione, dal momento che la regione di transizione, per $V_s \approx 0$, diventa importante rispetto a quella di completo spostamento.



A questo punto, abbiamo un problema: non abbiamo una relazione tra le componenti, Q_n e Q_d , e soprattutto nulla che coinvolga la tensione nota: V_{GS} (è infatti solo nota V_s). Per le applicazioni del sistema MOS sarà necessaria una relazione utile tra la carica di inversione, Q_n , e la tensione tra gate e bulk, V_{GS} .

Si incominci dunque da un punto noto: il principio di conservazione della carica; la carica del semiconduttore, nulla (somata) a quella di ossido (nulla) e metallo, deve essere nulla:

$$Q_s + Q_t = Q_n + Q_d + Q_t = 0$$

La Q_t è interessante, in quanto presenta legami con V_{GS} . $V_{GS} - V_{FB}$ è la tensione tra gate e parte neutra del bulk, dunque potrebbe essere pensata come somma di due contributi: V_{sc}

Chiamo per 493

e V_{ox} : cadute di tensione su semiconduttore e ossido. Noi approssimiamo una tensione tra gate e bulk ed essa cadrà tra ossido e semiconduttore, a meno dell'intrinseca tensione di flat-band.

Considerando un ossido dielettrico lineare, si può dire che:
 $Q_t = -Q_s = C_{ox} V_{ox}$, $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$, t_{ox} spessore dell'ossido.

Che si può dunque fare? Siamo interessati a vedere qual è una legge tra carica di inversione e tensioni note; proviamo a ottenerlo, in questo modo: abbiamo informazioni sulle capacità del sistema, dunque definendo una capacità differenziale C_d , proviamo a collegare i vari pezzi:

$$C_d \triangleq \frac{dQ_t}{dV_{GB}} = \left[\frac{dQ_t}{dV_{GB}} \right]^{-1}$$

Usando dunque il teorema della funzione inversa, ricordando che:

$$V_{GB} - V_{FB} = V_s + V_{ox} \rightarrow V_{GB} = V_s + V_{ox} + V_{FB}$$

$$\rightarrow \frac{1}{C_d} = \frac{dV_{ox}}{dQ_t} + \frac{dV_s}{dQ_t} + \frac{dV_{FB}}{dQ_t} \quad \left[\begin{array}{l} \text{Si ricordi che } V_{FB} \\ \text{è una tensione costante} \end{array} \right] =$$

$$= \frac{1}{C_{ox}} + \frac{1}{C_s} \Rightarrow C_d = C_{ox} \oplus C_s \quad (\text{facendo il reciproco})$$

Ora, $\frac{1}{C_s}$ può anche esser vista come:

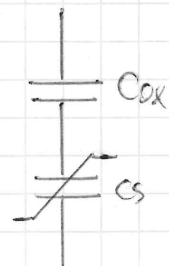
$$\frac{1}{C_s} = \frac{dV_s}{dQ_s} = - \frac{dV_s}{dQ_s} = - \left[\frac{dQ_s}{dV_s} \right]^{-1}$$

C_d è sostanzialmente nota: ma C_{ox} ma C_s

possono essere espresse in forma chiusa a partire

dai dati noti. Poiché la somma armonica di due

capacità è modellabile con una serie di due capacità, si può



dunque C_d è molto legata alla fisica del MOS: o come se si percorresse la strada dal gate al bulk, incontrando queste due capacità.

È possibile, come detto, esprimere $C_s(V_s)$ in forma chiusa, mediante derivazione da Q_s :

$$C_s =$$

A seconda delle regioni di funzionamento del sistema MOS, C_s assume diversi valori: tende ad essere mediamente grande in regione di accumulo, molto ridotta in regione di smaltimento, abbastanza importante (quasi) in regione di inversione forte.

Questo fatto si ripercuote su C_d : dal momento che essa è la ^{armonica} somma di due capacità, "vincerà sempre la minor". Normalmente, si segue il profilo di C_s fino alla regione di debole inversione compresa, per poi seguire C_d (dal momento che C_s è la più grande), a meno che V_s non sia un segnale ad alta frequenza: in tal caso, C_s non riesce a "seguire" le variazioni della tensione V_s , dunque C_s resta "piccolo", e sarà essa a prevalere nella somma armonica.

Struttura MOS a 3 terminali

Un primo passo avanti per introdurre strutture MOS più complicate è la struttura MOS a tre terminali.

Si introduce una regione drogata n+ che, in caso di inversione, risulta essere collegata al "bulk". Gli elettroni presenti

nel canale devono essere polarizzati; la polarizzazione viene portata al "tubo" attraverso regioni laterali, dunque vi deve essere un perfetto allineamento tra il fondo del gate e la regione drogata $n+$, in modo che, quando si va in inversione, si abbia contatto tra la regione drogata e il canale di inversione.

Quando si applica V_G , la V_G supera il "valore critico", dunque il contributo di V_G andrà via a " $n+$ ", cioè al canale ed esso collegato: tutti i punti si troveranno allo stesso potenziale.

Quello che capita "in vertice", è che, una volta che V_G è il potenziale di ogni punto del "tubo", i quasi-livelli di Fermi non saranno più uguali, dal momento che n è venuta a formare una differenza di tensione, $V_G \neq V_S$ ottenendo di fatto, in seguito all'introduzione di $n+$, una giunzione $p-n$; la distanza tra i quasi-livelli è pari alla tensione di polarizzazione della giunzione $p-n$, dunque a $V_{GS} = V_G - V_S$.

Si tenga ben presente il fatto che questa giunzione va polarizzata inversamente: noi non vogliamo avere un flusso di corrente sul bulk, dal momento che nel dispositivo finale che si vuole realizzare, il transistor sarebbe solo passivo: $V_B \ll V_G$, sempre, per ipotesi.

Una considerazione: supponendo di essere in stato di forte inversione, ora il potenziale non sarà più bloccato a $2\Phi_F$, ma a $2\Phi_F + V_{GS}$: $2\Phi_F$ deriva dalla "deflazione" di Φ_F , più quello sulla superficie che deve garantire la presenza di un numero di elettroni tale da essere in forte inversione: uno deriva dal occhio al quadrato di bulk qui

semplice effetto del drogaggio, l'altro dal fatto che il quasi-livello n è abbassato proprio di Φ_F , per introdurre gli elettroni nel canale.

Allora, si aggiunge V_{GS} .

Si era detto che:

$$n(x) = \frac{n_i^2}{N_A} \exp\left(\frac{q\phi(x)}{V_T}\right) \quad ; \quad \phi(0) \approx 2\Phi_F \quad \left[\begin{array}{l} \text{Approssimazione che "blocca"} \\ \text{il potenziale a } 2\Phi_F \end{array} \right]$$

Dunque:

$$n(0) = \frac{n_i^2}{N_A} \exp\left(\frac{2\Phi_F}{V_T}\right) = \frac{n_i^2}{N_A} \cdot \frac{N_A^2}{n_i^2} = N_A$$

A questo punto, si può capire quale sia la prima differenza rispetto al MOS a 2 terminali: ora V_S dipende dalla tensione applicata; questo fenomeno è detto "effetto substrate".

L'ampiezza della regione cambierà: x_d aumenterà, data la polarizzazione inversa della giunzione $p-n$, dunque anche la carica di spostamento:

$$x_d = \sqrt{\frac{2\epsilon_s}{qN_A} (2\Phi_F + V_{GS})} \quad ; \quad Q_d = -qN_A x_d = -\sqrt{2\epsilon_s q N_A (2\Phi_F + V_{GS})}$$

Conosciamo Q_d ; usiamo la legge di conservazione della carica, ottenendo:

$$C_{ox} V_{ox} = Q_t = -Q_n - Q_d$$

$$\rightarrow Q_n = -Q_t - Q_d = -C_{ox} V_{ox} - Q_d$$

C_{ox} è nota; V_{ox} è la tensione che cade sull'ossido, dunque: (considerando l'ipotesi di forte inversione)

$$V_{ox} = V_{GB} - V_{FB} - V_S \quad , \quad V_S = 2\Phi_F + V_{GS} \rightarrow V_{ox} = V_G - V_C - 2\Phi_F - V_{FB}$$

del momento che:

$$V_{GB} - V_{CB} = V_G - V_B - (V_C - V_B) = V_G - V_C = V_{GC}$$

Mettendo assieme i vari pezzi, possiamo calcolare la carica di immersione:

$$Q_n = -C_{ox} V_{ox} - Q_d = -C_{ox} (V_G - V_C - 2\Phi_p - V_{FB}) + \sqrt{2\epsilon_s q N_A (2\Phi_p + V_{CB})}$$

Da qua siamo più interessati a calcolare il valore della tensione di gate tale da annullare Q_n , in modo da ottenere il valore di "soglia" che crea lo stato di immersione. Dunque:

$$V_{th} = V_{FB} + V_C + 2\Phi_p + \frac{1}{C_{ox}} \sqrt{2\epsilon_s q N_A (2\Phi_p + V_{CB})}$$

Raccogliendo un po' di zola, si ottiene:

$$V_{th} = V_{FB} + V_C + 2\Phi_p + \gamma_B \sqrt{2\Phi_p + V_{CB}}, \quad [\gamma_B] = \sqrt{V}$$

Solitamente, si esprime V_{th} in un modo leggermente diverso: si considera V_{th0} come:

$$V_{th0} = V_{th}(V_C = V_B = 0) = V_{FB} + 2\Phi_p + \gamma_B \sqrt{2\Phi_p}$$

Dunque:

$$V_{th} = V_{th0} + V_C + \gamma_B \left(\sqrt{2\Phi_p + V_{CB}} - \sqrt{2\Phi_p} \right)$$

Sostituendo questa espressione in Q_n si trova la condotta "relazione di controllo di carica":

$$Q_n = -C_{ox} (V_G - V_{th}) = -C_{ox} \left[V_G - V_{th0} - V_C - \gamma_B \left(\sqrt{2\Phi_p + V_{CB}} - \sqrt{2\Phi_p} \right) \right]$$

Una osservazione: questa approssimazione (espressione approssimata) funziona bene ovunque, tranne nell'intorno della soglia; si sa che purtroppo molte applicazioni lavorano nell'intorno della soglia, per udine i consumi, dunque bisogna tener presente la cosa.

Transistore MOS

La struttura "finale" è quella a 4 piedini: il transistor MOS. Si nota, oltre alle 2 regioni drogate, due regioni di "ossido di campo": esse servono per fare in modo da aumentare la tensione, evitando la formazione di canali spuri.

1) Segno

Nel normale funzionamento si applica una tensione $V_{GS} > 0$, dunque con drain positivo rispetto al source. Il MOSFET è generalmente un dispositivo simmetrico, nel senso che per source si definisce semplicemente il piedino collegato a un potenziale n^+ a potenziale inferiore. Nel caso il MOSFET ha a 3 piedini, o per alcune applicazioni di potenza (che richiedono l'introduzione di un diodo), si possono avere delle dissimmetrie.

Si hanno, ad ogni modo, 2 giunzioni in antiserie: una verrà polarizzata, l'altra no.

Parliamo di condizioni statiche: $I_{gate} = 0$; I_D dipende da 2 tensioni: V_{GS} , ma anche V_{DS} , nel senso che essa deve essere tale da garantire la formazione del canale, ma non solo: essa avrà anche a canale formato influenza su I_D .

Il controllo del "tubo", del canale, è fatto dal campo elettrico, che porta una caduta di potenziale di $2\Phi_p$. Da qua l'acronimo FET.

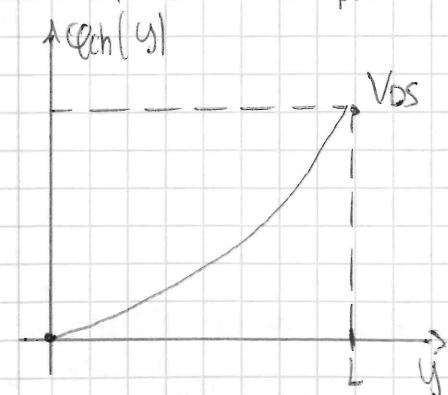
Si usano ossidi a elevata costante dielettrica perché, quando si applica al gate una tensione, si avrà il 2^o p. Il campo elettrico nell'ossido vale $E_{ox} = E_0 \frac{\epsilon_s}{\epsilon_{ox}}$; questo vuol dire che se l'ossido fosse troppo sottile, si avrebbe l'effetto tunnel, e noi non lo vogliamo: non vogliamo corrente sul gate.

Dobbiamo mantenere costante E ; la costante ϵ_s è quella che è; cambiando ϵ_{ox} possiamo avere ossidi più spessi, a parità di tensione di soglia, migliorando la schematura all'effetto tunnel.

Chiediamo di rendere più quantitativa l'analisi: dovremo introdurre il concetto di potenziale di canale, $\phi_{ch}(y)$.

Sappiamo che ci sono le regioni di source e drain; queste non sarebbero intrinsecamente funzionalmente necessarie, ma lo sono tecnologicamente, dal momento che è necessario, in qualche modo, portare la tensione al semiconduttore. Il loro comportamento è resistivo, e sono attraversate da I_D . Dalla corrente di drain più carica e drogata sono queste regioni, meglio è.

Il potenziale su di un generico punto y del canale, purtroppo, non è costante ed è il "potenziale di canale": esso varia a causa di I_D e del comportamento resistivo del dispositivo.



Per $y=0$, $\phi_{ch}(y)=0$; per $y=L$, con L lunghezza del canale, $\phi_{ch}(L) = V_{DS}$ e essendo crescente non meno che si avanza sul canale, ce sarà una funzione monotona crescente di y . Questo $\phi_{ch}(y)$

è la funzione di polarizzazione degli elettroni del canale, solo che

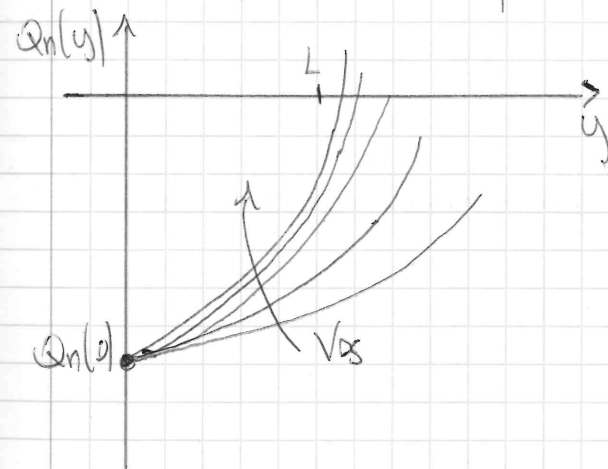
ora, a differenza del MOS a 3 terminali, non è costante.

Prendendo la formula di Q_n , essa può essere generalizzata, tenendo conto del potenziale di canale:

$$Q_n = -C_{ox} [V_{GS} - V_{th0} - \phi_{ch}(y)] + C_{ox} \gamma_B [\sqrt{2\Phi_F + \phi_{ch}(y) + V_S - V_D} - \sqrt{2\Phi_F}]$$

Questa espressione è fondamentale: essa ci dà le caratteristiche statiche della corrente nel FET; gli elettroni nel tubo saranno in numero dipendente dalla posizione nel medesimo, dunque varierà anche la polarizzazione. Trattandosi queste funzioni monotone crescenti, onde l'equivalente lo sarà.

Si avrà qualcosa di questo genere:



In $y=0$ si ha il massimo numero di elettroni, $Q_n(0)$; a L , dunque, il minimo numero.

A seconda di V_{DS} , il numero di elettroni decrescerà in velocità diversa: maggiore è V_{DS} , maggiore

è Q_n , dunque la sua influenza nella formula. Si ha un fenomeno di pinch-off (strozzamento): maggiore è V_{DS} , più grandi saranno le "derivate", prima dunque termineranno gli elettroni; può anche capitare che il canale finisca gli elettroni ancora prima di terminare, in un punto $L' < L$. Il termine dipendente da γ_B può essere trascurato, a seconda del drogaggio del substrato: si ricordi infatti che γ_B varia proporzionalmente a $\sqrt{N_A}$.

Considerando $\gamma_B \approx 0$, si può definire il valore di V_{GS} che inizia a "stirizzare" il canale. Si ha:

$$V_{GS} - V_{th} - \phi_{ch}(y) = 0$$

Dunque:

$$\phi_{ch}(y) = V_{GS} - V_{th}$$

Quello che capita in pratica è che il MOS si comporta da "esistenza"; per V_{GS} grossa, capita che la carica totale tende ad annullarsi, ottenendo l'equazione precedente: se ϕ_{ch} cresce troppo, infatti, come si evince dal modello a controllo di carica, Q_n si annulla. Dato il punto $L = L'$, ossia la condizione "limite" (il fatto che il canale, se si aggiunge ^{ancora} tensione, inizierebbe a esser ridotto), si ha:

$$\phi_{ch}(L) = V_{GS} - V_{th} = V_{DSS}$$

Il valore di $\phi_{ch}(L)$, quando $Q_n = 0$, è V_{DSS} , ossia il massimo valore di V_{GS} per cui non si ha il pinch-off.

Caratteristica statica

Quello che si vuol fare, a questo punto, è calcolare la caratteristica statica del dispositivo. Per fare ciò, sarà necessario introdurre alcune approssimazioni, che, per maggior chiarezza, verranno introdotte tra breve, durante la trattazione.

$$I_D = W \int_{\text{canale}} J_n dx$$

Si considera il MOS uniforme lungo l'asse x , e, nel prossimo paragrafo, si considera, anziché il modello drift-diffusion, solo una densità

Trascinamento

di corrente di ~~diffusione~~ trascinando i termini legati alla diffusione. Dunque:

$$J_n \equiv J_{n,drift} = q \mu_n E \cdot n$$

Ricordando, da Gauss/Poisson, che:

$$E \equiv E_y = - \frac{d\phi_{ch}}{dy} \quad \left(\text{"solarizzando" il gradiente} \right)$$

$$\rightarrow I_D = -Wq \int_{\text{canale}} \mu_n \cdot n(x,y) \cdot E_y(y) dx$$

A questo punto, un'altra approssimazione: considereremo da ora il fatto che i campi elettrici longitudinale e trasversale al canale siano indipendenti; dunque, E_y dipende solo da y , e non anche da x . Questa approssimazione è valida solo se il canale è lungo, ed è detta "approssimazione di canale graduale". Si considera inoltre il fatto che μ_n sia costante, dunque utilizzeremo la mobilità di basso campo: $\mu_n \approx \mu_{n0}$. Dunque:

$$\begin{aligned} I_D &= -Wq \mu_{n0} \frac{d\phi_{ch}(y)}{dy} \int_{\text{canale}} n(x,y) dx = \\ &= -Wq \mu_{n0} \frac{d\phi_{ch}(y)}{dy} N_{ch}(y) \end{aligned}$$

A questo punto, integriamo in dy :

$$\rightarrow \int_0^y I_D dy = \int_0^L \left[-Wq \mu_{n0} N_{ch}(y) \frac{d\phi_{ch}}{dy} \right] dy$$

Poiché stiamo facendo un cambio di variabili, ottenendo come variabile di integrazione finale $\phi_{ch}(y)$, avremo come $\phi_{ch}(0)$ il valore 0, mentre è noto che $\phi_{ch}(L) = V_{GS}$:

$$= -Wq\mu_{n0} \int_0^{V_{DS}} N_{ch}(y) d\phi_{ch}(y)$$

A sinistra, avremo $L \cdot I_0$; a destra:

$$L \cdot I_0 = -W\mu_{n0} \int_0^{V_{DS}} [V_{GS} - \phi_{ch} - V_{th}] d\phi_{ch}$$

Questo, dal momento che $qN_{ch}(y) = Q_n$, che viene approssimato con il solo termine lineare (o almeno penso)

Dunque:

$$I_0 = \frac{W\mu_{n0}C_{ox}}{L} \left[(V_{GS} - V_{th})V_{DS} - \frac{1}{2}V_{DS}^2 \right]$$

A questo punto, alcune osservazioni: ciò che abbiamo fatto è vero, a patto da avere il canale ^{lungo} ~~esteso~~: L dell'ordine dei μm

Volendo plottare la curva, si potrebbe vedere che, dalla curva "parabolica" in poi, si è in stato di pinch-off, e la corrente rimane quasi costante, al valore assunto alla tensione $V_{DS} = V_{DSS}$;

come noto, $V_{DSS} = V_{GS} - V_{th}$; sostituendo, dunque si ha:

$$I_{0,sat} = \beta_n \cdot \frac{1}{2} \cdot (V_{GS} - V_{th})^2$$

Dove β_n è il termine:

$$\beta_n = \frac{W\mu_{n0}C_{ox}}{L}$$

Cosa capita, dal punto di vista fisico? Beh, se il canale si interrompe, la corrente non è nulla, poiché sostenuta da

un fenomeno di iniezione; si tratta di qualcosa di simile al diodo a giunzione p-n polarizzato direttamente.

In verità, quello che capita è che la corrente aumenta un poco, dal momento che, se il canale si accorcia, la sua resistenza si riduce: essendo sempre presente V_{DS} , il rapporto tensione/resistenza aumenta.

Se L è molto grande, si può dire che $L' \approx L$, dunque la variazione è all'estrema ridotta, dunque l'aumento della corrente è molto debole.

Empiricamente, quello che si fa di solito è moltiplicare I_0 per $[1 + \lambda V_{DS}]$, sia in zona lineare sia in zona di saturazione; di solito, comunque, λ è molto piccolo.

Esiste un modello più completo, più complicato, che tiene conto dell'effetto di substrato; usando dunque l'espressione "completa" di $Q_n(\phi_{ch})$, si ottiene: dimostrare

$$I_0 = I_{0,lin} - \gamma_B \beta_n \left[\frac{2}{3} \left((2\Phi_p - V_{BS} + V_{DS})^{\frac{3}{2}} - (2\Phi_p - V_{DS})^{\frac{3}{2}} \right) - \sqrt{2\Phi_p} V_{DS} \right] V$$

Come si può vedere facilmente, il modello lineare sovrastima la corrente di saturazione; per il resto, sono differenze minime.

Quello che si può fare è una via di mezzo: si trascura la presenza del potenziale di canale nel termine "quadratico", sostituendo la tensione di soglia con la cosiddetta "tensione di soglia al source":

$$V_{th,s} = V_{th0} + \gamma_B \left[\sqrt{2\Phi_p - V_{BS}} - \sqrt{2\Phi_p} \right]$$

Nel caso di dispositivi scalati, a canale corto, l'approssimazione a canale graduale non è più verificata: V_{DS} è infatti applicata su di una regione piccola, dunque i campi elettrici sono più grandi, e l'effetto di saturazione non è più dovuto al pinch-off: al posto della mobilità dei portatori costante se ne dovrebbe avere una di alto campo, che tenderebbe a saturare, portando alla saturazione.

Si ricordi che tutto ciò che è stato detto vale anche per MOS a canale p: le termioni contengono segno, i vuoti di elettroni e lacune si invertono, e cambia anche la relazione a controllo di carica.

Comportamento dinamico

In un MOS sono presenti diversi effetti capacitivi: la capacità di gate, vista "nel metallo", la capacità di giunzione, delle varie giunzioni pn tra substrate, source e drain (si tratta come noto di giunzioni planari invertite, che dunque comportano la presenza di regioni di indebolimento), ^{infine} dunque capacità di overlap, dovute a un non perfetto allineamento tra gate, source e drain. Siamo interessati allo studio di modelli di piccolo segnale: separazione del punto di funzionamento in continua, e di segnale, in zona di saturazione. Si sa che:

$$I_G = I_{G0} + \frac{\partial I_G}{\partial V_{GS}} v_{GS} + \frac{\partial I_G}{\partial V_{DS}} v_{DS}$$

$$I_D = I_{D0} + \frac{\partial I_D}{\partial V_{GS}} v_{GS} + \frac{\partial I_D}{\partial V_{DS}} v_{DS}$$

$$\rightarrow \begin{bmatrix} I_G \\ I_D \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_{GS} \\ V_{DS} \end{bmatrix}$$

Perché conosciamo I_{D0} , costruiamo un modello limitato ad essa:

$$Y_{21} = \frac{\partial I_D}{\partial V_{GS}} = K_n (V_{GS} - V_{th}) [1 + \lambda V_{DS}] = \frac{2 I_{D0}}{V_{GS} - V_{th}} = g_m$$

$$Y_{22} = \frac{\partial I_{D0}}{\partial V_{DS}} = \frac{K_n}{2} (V_{GS} - V_{th})^2 \lambda = \frac{\lambda I_{D0}}{1 + \lambda V_{DS}} = \frac{1}{r_o}$$

involve?

disegno di semiconduttori e eterostrutture

Le leghe di semiconduttori sono molto utili per applicazioni avanzate, quindi quelle ad alta frequenza o per l'optoelettronica: il silicio, per queste cose, non sarà più idoneo.

In una lega di semiconduttori si realizza una crescita di tipo epitassiale, ma la crescita di uno strato di un materiale cristallino su di un altro materiale cristallino, ma diverso. Se due materiali sono differenti, dunque avranno diversi parametri reticolari; il nostro scopo è quello di capire come si comportano questi due materiali, una volta interfacciati.

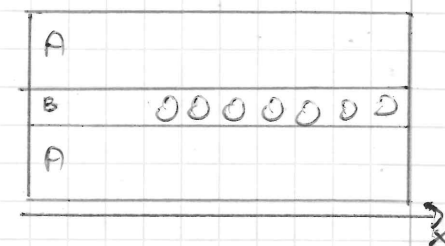
Quando si vuole realizzare ciò, ma quando si fanno strutture di tanti layer con diversi parametri reticolari, si tende a creare difetti all'interfaccia, introducendo degradazione anche del punto di vista elettronico: centri di trappola, di scattering, che riducono mobilità e tempo di vita. Ciò che fanno è studiare una stratificazione ideale di diversi materiali; a tal fine, si definiscono "eterostrutture" quelle strutture nelle quali, idealmente, non sono presenti difetti all'interfaccia. Ciò che si riesce a fare è:

- creare buche di potenziale che permettano di confinare parzialmente i portatori liberi di carica (1)
- introdurre discontinuità "valute" di indice di rifrazione: materiali con diverso energy gap E_g hanno anche diverso coefficiente di rifrazione. (2)

Vediamo perché:

(1)

- Considerando una struttura di questo tipo, dal momento che lo strato B ha energy gap più piccolo, gli elettroni potranno solo muoversi lungo "x";



- Si consideri questa stratificazione: il fatto di avere qualcosa di questo tipo permette di confinare il campo ottico nella regione B, ottenendo una guida d'onda dielettrica.



Se non si ha una perfetta conoscenza della costante reticolare, si ricorda che, con l'aumentare di E_g nei materiali, il passo reticolare tende a ridursi.

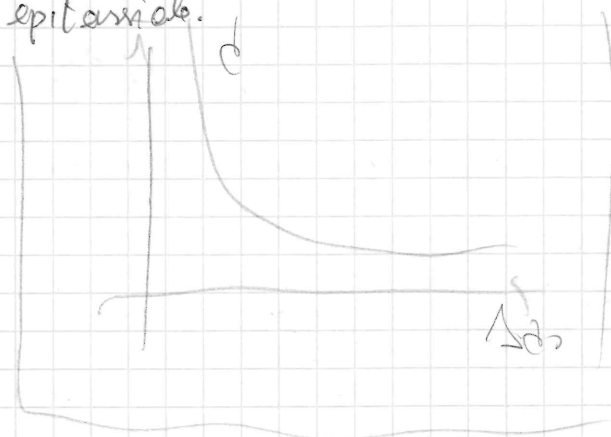
Si consideri un esempio terzo: volendo fare una lega di silicio (Si) e di germanio (G), si sa che il silicio ha energy gap più grande (1.12 eV contro ~0.6 eV), dunque il passo reticolare del germanio sarà maggiore. Si parla, in questi casi, di "disadattamento reticolare".

- Nel caso di substrato di Si e strato di G, ci si aspetta che il G, a passo reticolare maggiore, si adatti, ottenendo uno strato pseudomorfico (dato da una riorganizzazione dello strato, della topologia del suo cristallo), ottenendo un cristallo equivalente simile a quello del substrato; nel caso la crescita sia "non epitassiale", si ha che solo alcuni atomi di germanio (e non tutti) si legano al silicio, non formando tutti i legami, creando un problema all'interfaccia. Qui, tendenzialmente, il G non si adatta al Si.

- Stessa cosa può accadere con G substrato. Si stacca da depositare lo strato sono adattati, dunque sarà il silicio, nel caso di crescita epitassiale, a dover organizzare la propria struttura.

L'epitassia pseudomorfica purtroppo non si può fare in maniera indefinita: non man mano che si aumenta lo spessore dello strato depositato, si fa sempre più fatica a depositare "bene"; si ha uno spessore critico, di cui di sopra del quale si cade gradualmente nella struttura non epitassiale.

Si ha dunque il "problema del substrato": a seconda del tipo di applicazione si può avere una scelta limitata di substrati utilizzabili, e di questi comunque bisogna limitare lo spessore di deposizione. Questo grafico presenta ϵ_{eff} versus differenza dei passi reticolari.



Quello che si riesce dunque a fare è una sorta di "miscela" di 3 o 4 elementi, ottenendo delle proprietà "intermedie".

Si parla di:

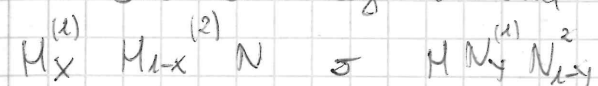
- leghe ternarie: 2 composti, 3 elementi;
- leghe quaternarie: 4 composti, 4 elementi.

Esistono regole di composizione per le leghe:

- i semiconduttori composti (leghe diione) son formate da un metallo M e da un non-metallo N;

- le leghe ternarie sono formate da due composti che han in comune un metallo e un non metallo
- le leghe quaternarie sono formate da quattro composti che condividono due metalli e due non metalli

Nel caso delle leghe ternarie, si trova qualcosa del tipo:



"x" e "y" sono la stessa cosa, nei due casi: esse sono dette "frazione molare"; si tratta di un concetto associato alla composizione stechiometrica degli elementi.

Nel caso delle leghe quaternarie si avrebbe comunque qualcosa di simile, con più parametri.

Le frazioni molari tornano molto utili, se i semiconduttori risultanti (e di portenza) seguono la "legge di Vegard": data una proprietà P di una lega, essa può essere ottenuta a partire da una combinazione lineare della stessa proprietà, nei vari materiali.

P può essere l'energy gap per esempio.

$$P = x P^{(1)} + (1-x) P^{(2)} \quad [\text{ternaria}]$$

$$P = x P^{(1)} + (1-x) P^{(2)} + y P^{(3)} + (1-y) P^{(4)} \quad [\text{quaternaria}]$$

Scegliendo "x" e "y", è possibile dunque "scegliere" il parametro E_g per esempio "ingegnerizzando" le bande.

Esistono correlazioni bilineari (legge di Aboles), ma esistono anche casi in cui non è possibile usare alcuna interpolazione (come, per esempio, il caso di passaggio da gap diretto a indiretto); dipende, sostanzialmente, dai materiali scelti.

Una nota: al di sopra di un certo valore di x , quello che può capitare è che oltre a cambiare il parametro interessante, ad esempio E_g , potrebbe cambiare anche per esempio il passo reticolare, a sua volta influente su E_g .

Altri problemi riguardano la crescita: se la costante reticolare è molto diversa, si deve usare un approccio metamorfico, usando strati buffer con passi reticolari "di adattamento" (graded layer). Vale il già fatto discorso, sullo spesso critico.

Affinity rule

Finora abbiamo avuto a che fare con materiali uguali (a meno dei MOS / MS): stessa affinità elettronica, stesso energy gap, stessi valori di E_c e E_v .

Al fine di studiare queste strutture, si fa qualcosa di abbastanza simile a ciò che si è fatto in precedenza: si disegnano i diagrammi a bande delle strutture separate, usando come riferimento il livello del vuoto, U_0 . Si può vedere che sono presenti salti, discontinuità, sia nella banda di conduzione sia in quella di valenza.

Definendo:

$$\Delta\chi = \chi_B - \chi_A \quad ; \quad \Delta E_g = E_{g,B} - E_{g,A}$$

Si possono ricavare le differenze tra $E_{c,A}$ e $E_{c,B}$ come:

$$\Delta E_c \hat{=} E_{c,B} - E_{c,A} = \dots$$

Si ricordi che è sempre buona cosa usare un riferimento; per noi, esso può essere il livello del vuoto, U_0 :

$$= (E_{c,B} - U_0) - (E_{c,A} - U_0) = -q\chi_B + q\chi_A = -q\Delta\chi$$

Per quanto riguarda ΔE_v :

$$\Delta E_v = E_{v,B} - E_{v,A} = (E_{c,B} - E_{g,B}) - (E_{c,A} - E_{g,A}) = -q\Delta\chi - \Delta E_g$$

A parità: la differenza, la discontinuità tra le bande di conduzione è quantificabile a partire dalle sole differenze di affinità elettronica, mentre, nel caso della discontinuità in banda di valenza, si deve anche tener conto della differenza degli energy gap. Questa è la regola dell'affinità elettronica.

Vi sono diversi tipi di eterostrutture, classificabili in base a come sono sovrapposte le bande dei materiali. Se ne presentano alcuni tipi:

- Tipo 1: un diagramma è completamente compreso nell'altro. Può essere utile per avere una buca di potenziale sia per gli elettroni, sia per le lacune, dal momento che, come si dirà meglio, è possibile realizzare delle sorte di "trappe".
- Tipo 2: parziale allacciamento delle bande. A seconda della stratificazione è possibile confinare solo uno dei due tipi di portatori. Può capitare (2a), che una delle due bande sia perfettamente allineata con l'altra.
- Tipo 3: due bande completamente disaccoppiate.

A questo punto ci chiediamo: come si studia l'interazione tra due materiali, una volta uniti? Come si può dunque raggiungere l'equilibrio termodinamico?

L'idea di base è abbastanza simile a quella dietro lo studio "tradizionale" dei dispositivi a semiconduttore: si deve,

a partire dal diagramma a bande delle strutture separate, capire come i materiali scambiano tra loro i portatori liberi.

Si ricordi che il livello di Fermi misura l'energia media degli stati occupati, e che:

- per gli elettroni, la barriera è "positiva": una "ostacolo" al moto, se è "dal basso verso l'alto"
- per le lacune, l'opposto: un ostacolo per le lacune è una barriera "dall'alto verso il basso".

Questo fatto è vero dal momento che i diagrammi a bande vengono disegnati rispetto agli elettroni, per le lacune non dunque letti "al contrario".

La costruzione del diagramma a bande è all'incirca simile alla precedente: dalla situazione fuori equilibrio, in cui il riferimento è il livello del vuoto U_0 , all'equilibrio termodinamico si trova uno livello di Fermi costante su tutto il dispositivo.

La differenza sostanziale che si può evincere è la presenza di discontinuità: ora si tratta al fatto che le grandezze sono eterogenee, dunque $E_{c1} \neq E_{c2}$, $E_{v1} \neq E_{v2}$. Si ricordi la seguente cosa:

- se E_c ha curvatura positiva ("sonda"), cede elettroni; altrimenti, se negativa ("è triste"), ne acquista; questo permette di capire quale delle strutture cede elettroni.
- dualmente, se E_v ha curvatura positiva ("sonda"), sta acquistando lacune; se negativa ("triste"), sta cedendo lacune.

I moti di carica porteranno alla creazione di carica mobile (data dai portatori iniettati) da una parte, e fissa (gli ioni) nel lato opposto.