

Dispositivi Elettronici

Alberto Tibaldi

25 febbraio 2008

Indice

1	Proprietà elettriche dei materiali	2
1.1	Metalli	2
1.2	Semiconduttori	5
1.2.1	Calcolo di n_i	9
1.3	Caratterizzazione dei semiconduttori drogati	10
1.3.1	Calcolo di n e p in un semiconduttore drogato	12
1.3.2	Equazioni di Shockley	17
2	Modello matematico delle correnti nei semiconduttori	20
2.1	Correnti di diffusione	25
2.2	Relazioni di Einstein ed Equazioni del Trasporto	26
2.3	Equazione di continuità	27
2.4	Equazione di Poisson	29
2.4.1	Legame Carica - Campo Elettrico	30
2.4.2	Dimostrazione qualitativa dell'Equazione di Poisson	31
2.5	Considerazioni su semiconduttori fuori equilibrio	32
2.6	Modello del Tasso di Ricombinazione Netto	33
2.7	Esercizio Pratico	35
2.7.1	Risoluzione	36
2.7.2	Ipotesi di lato corto	41
3	Giunzione p-n	44
3.1	Studio qualitativo del diagramma a bande della giunzione p-n	47
3.2	Studio elettrostatico della giunzione p-n	51
3.2.1	Equazione di neutralità globale	51
3.2.2	Analisi della barriera di potenziale	52
3.2.3	Analisi formale dell'andamento di campo e potenziale	53
3.2.4	Risoluzione del sistema	56
3.3	Applicazione di una tensione esterna alla giunzione	57
3.3.1	Capacità di svuotamento	61
3.4	Esercizio pratico	62

3.4.1	Risoluzione	62
3.5	Effetti di una tensione su di una giunzione p-n	65
3.5.1	Quasi-Livelli di Fermi	66
3.6	Leggi della Giunzione	68
3.7	Caratteristica statica della giunzione p-n	70
3.8	Studio delle correnti nella giunzione p-n	74
3.8.1	Polarizzazione diretta	74
3.8.2	Polarizzazione inversa: corrente inversa di saturazione	76
3.9	Esercizio Pratico	78
3.9.1	Risoluzione	79
3.10	Cenni ad altri modelli della giunzione	81
3.10.1	Modello di ampio segnale	81
3.10.2	Modello di piccolo segnale	83
3.11	Valutazione di V_γ	83
3.12	Meccanismi di rottura della giunzione	84
3.12.1	Effetto valanga	84
3.12.2	Effetto tunnel	85
3.12.3	Diodi Zener	85
3.13	Calcolo della capacità di diffusione	86
3.14	Modello a controllo di carica	87
3.15	Modello di piccolo segnale	89
3.16	Esercizio pratico	91
3.16.1	Risoluzione	91
4	Il transistorore bipolare	95
4.1	Analisi delle correnti	96
4.2	Calcolo delle correnti	99
4.3	Modello statico del transistorore bipolare	101
4.3.1	Equazioni di Ebers-Moll	104
4.4	Modello circuitale statico del transistorore bipolare	106
4.4.1	Andamento della caratteristica statica	108
4.4.2	Effetto Early	109
4.5	Esercizio Pratico	110
4.5.1	Risoluzione	111
5	Sistemi MOS	120
5.1	Analisi Elettrostatica	121
5.1.1	Polarizzazione di un sistema MOS con tensione negativa	124
5.1.2	Inversione di popolazione	125
5.2	Calcolo della carica di inversione	127
5.3	Determinazione della tensione di soglia di un sistema MOS	131

5.3.1	Calcolo della tensione di soglia	133
5.4	Il transistor MOSFET	134
5.5	Caratteristica statica del MOSFET: Analisi a canale graduale	135
5.6	Esercizio Pratico	143
5.6.1	Risoluzione	144

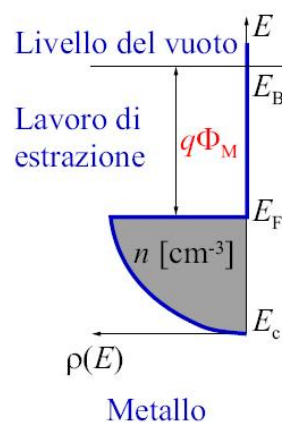
Capitolo 1

Proprietà elettriche dei materiali

Dalla fisica dello stato solido sappiamo che un solido è rappresentabile mediante un diagramma a bande; il diagramma a bande è un grafico in cui si riportano sulle ascisse una posizione (ossia una coordinata spaziale, nel reticolo diretto, o un vettore d'onda, nel reticolo reciproco), sulle ordinate un'energia E ; si noti che per buona parte della trattazione tuttavia si considereranno materiali isotropi, quindi non si considereranno in un primo tempo variazioni delle condizioni del solido nello spazio.

I due tipi di solidi che siamo interessati a studiare in dispositivi sono i metalli ed i semiconduttori; diamo alcuni cenni sul loro comportamento, a partire dallo studio del diagramma a bande.

1.1 Metalli



Indichiamo ora il significato dei diversi livelli energetici:

- E_C : Livello energetico in cui inizia la banda di conduzione;
- E_F : Livello di Fermi
- E_B : Livello del vuoto, ossia livello energetico da superare per estrarre il metallo
- $q\Phi_M$: Lavoro di estrazione del metallo;
- γ_n : Parametro della funzione di densità degli stati energetici.

Facciamo alcune osservazioni: l'inizio della banda di conduzione dei metalli, E_C , si può considerare al livello energetico $E = 0$, poichè l'energia dell'elettrone in un metallo è esclusivamente cinetica; il livello di Fermi si troverà dunque all'interno della banda di conduzione, all'incirca a 4,5 eV. Allo zero assoluto, tutti gli elettroni saranno al di sotto del livello di Fermi (come si può osservare mediante lo studio della distribuzione di Fermi-Dirac); si definisce dunque una funzione esprimente il lavoro di estrazione, $q\Phi_M$, come la minima energia in grado di estrarre, ad una temperatura pari a 0 K, un elettrone da un metallo. Gli elettroni più facili da estrarre, ad una temperatura di 0 K, saranno quelli in prossimità del livello di Fermi E_F .

I due ingredienti per poter esprimere l'occupazione di uno stato in un metallo saranno dunque la funzione di densità degli stati energetici $N(E)$, e la funzione di occupazione $f(E; T)$: poichè trattiamo lo studio sull'occupazione di stati da parte di elettroni, ossia fermioni, la funzione di occupazione in questione sarà la già citata distribuzione di Fermi-Dirac.

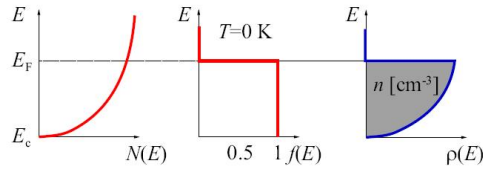
$$N(E) = \gamma_n E^{\frac{1}{2}}$$

$$f(E; T) = \frac{1}{1 + e^{\frac{E - E_F}{kT}}}$$

La funzione di densità degli stati effettivamente occupati, ρ_n , sarà data dal prodotto dei nostri due ingredienti: funzione di densità degli stati e funzione di occupazione degli stati. Ciò è meglio motivabile con un esempio classico: dato un numero di sedie vuote disponibili in un cinema, ed il numero di spettatori, il prodotto delle due fornirà il numero di sedie effettivamente occupate dai potenziali spettatori.

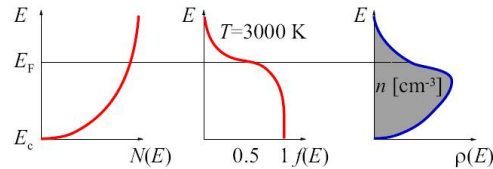
Esprimiamo questa nostra intuizione, formulando la funzione ρ_n :

$$\rho_n(E) = N(E)f(E; T)$$



In un metallo, accadrà graficamente qualcosa di simile:

Questo grafico raffigura una situazione che non considera effetti termici, ossia in cui la temperatura è pari a 0 K, e quindi la funzione di Fermi-Dirac è di fatto un impulso rettangolare; in una condizione di notevole influenza termica, con una temperatura sui 3000 K, la funzione di Fermi-Dirac cambierà sensibilmente, e di conseguenza la densità degli stati occupati.



Noi consideriamo però di trovarci ad una temperatura $T \simeq 300$ K, ossia una temperatura ragionevolmente confrontabile alla temperatura ambiente; a meno che non si parli di studi specifici riguardanti effetti termici particolarmente influenti, una temperatura come 300 K ci permetterà una buona approssimazione rispetto alla realtà.

Per sapere esattamente quanti elettroni vi saranno nel metallo in banda di conduzione, ad una certa temperatura, sarà sufficiente integrare la funzione di densità di popolazione da 0 (livello di inizio della banda di conduzione E_C) a $+\infty$ (in prima approssimazione, infatti, il modello di Bethe-Sommerfeld prevede una modellizzazione di un metallo come una buca di potenziale infinita).

$$n = \int_0^{+\infty} \rho_n(E) dE = \int_0^{+\infty} N(E) f(E; T) dE \sim 10^{22} e^- / \text{cm}^3$$

In un metallo, abbiamo quindi a disposizione circa 10^{22} elettroni per cm^3 .

Abbiamo già nominato il lavoro di estrazione, e la sua unità di misura, l'elettronvolt (eV): 1 eV è l'energia che un elettrone ha se sottoposto ad una differenza di potenziale pari ad 1 V; questa unità di misura è molto compatibile con i problemi che si trattano in dispositivi, e quindi molto comoda. Alcuni lavori di estrazione noti sono quello dell'alluminio Al, di 4,1 eV, o quello dell'oro Au, di 5 eV; cosa si intende però per lavoro di estrazione, precisamente? Abbiamo già dato una prima definizione; ora che

abbiamo parlato meglio di livello di Fermi, possiamo dire che il lavoro di estrazione è la distanza energetica tra il livello di Fermi E_F e il livello del vuoto, della buca modellizzante il metallo, E_B .

1.2 Semiconduttori

Il diagramma a bande di un semiconduttore è ben più articolato di quello di un metallo; analizziamone un modello semplificato, per poter iniziare una prima trattazione dei semiconduttori; gli elementi in esso contenuti saranno:

- E_C : Inizio della banda di conduzione;
- E_F : Livello di Fermi;
- E_V : Termine della banda di valenza;
- E_g : Energy Gap, ossia distanza energetica tra banda di conduzione e banda di valenza

Il livello di Fermi si trova, in un semiconduttore intrinseco, a metà tra il livello E_V ed il livello E_C . Si introduce una prima diversificazione delle notazioni, inserendo, quando si parla di semiconduttori intrinseci, ossia puri, regolari, e che non han subito trattamenti tecnologici di alcun tipo, un pedice aggiuntivo i : parlando di livello di Fermi intrinseco, si parlerà di E_{F_i} .

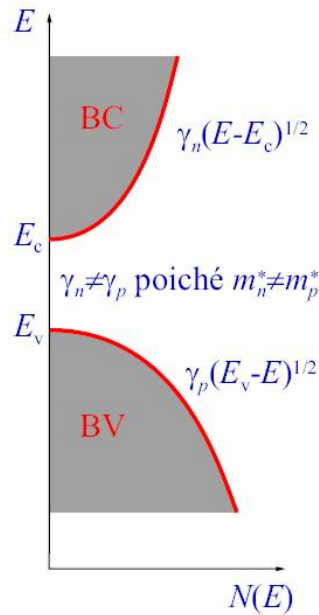
Abbiamo affrettato alcune conclusioni, per quanto esse siano corrette, parlando del livello di Fermi: E_F non può essere all'interno della banda di valenza, altrimenti gli elettroni potrebbero andarsene liberamente in banda di conduzione, e quindi il semiconduttore condurrebbe senza alcun problema, anche senza trattamenti particolari, anche allo zero assoluto; l'evidenza sperimentale mostra che allo zero assoluto, e in condizioni di intrinsecità, la banda di conduzione è vuota, e quella di valenza completamente piena di elettroni.

Anche parlando di semiconduttori, si introduce il concetto di densità degli stati disponibili, anche se sarà più complesso rispetto a quello nei metalli:

Le densità degli stati, in questo ambito, avranno forma:

$$N_{BC} = \gamma_n(E - E_C)^{\frac{1}{2}}$$

$$N_{BV} = \gamma_p(E_V - E)^{\frac{1}{2}}$$



Un'ulteriore convenzione: il pedice n identifica gli elettroni in banda di conduzione; al contrario, il pedice p evidenzia le lacune lasciate dagli elettroni, in banda di valenza. In un semiconduttore, vi saranno infatti due tipi di portatori di carica, ossia gli elettroni in banda di conduzione (come ovvio, anche dallo studio dei metalli), ma anche le lacune in banda di valenza, ossia gli ammanchi di elettroni causati dalla transizione di questi verso livelli energetici superiori ad E_V (quali ad esempio la banda di conduzione). La tecnica per trovare il numero effettivo di portatori, sarà la stessa di prima, solo che in questo caso, avendo due contributi, si dovranno risolvere due integrali:

$$n_i = \int_{E_C}^{+\infty} \gamma_n (E - E_C)^{\frac{1}{2}} f(E; T) dE$$

$$p_i = \int_{-\infty}^{E_V} \gamma_p (E_V - E)^{\frac{1}{2}} f(E; T) dE$$

Ma, come ci suggeriscono i pedici i , ci troviamo in un caso intrinseco, e quindi possiamo dire senza indugi che

$$n_i = p_i$$

Nel silicio, il semiconduttore per eccellenza nei nostri studi, $n_i = p_i = 1,45 \cdot 10^{10}$ elettroni per cm^3 , ad una condizione di temperatura di circa 300 K. Confrontando con i 10^{22} elettroni in un metallo, la conduzione è un evento possibile, ma molto improbabile.

Continuiamo la nostra analisi fisica del comportamento elettrico dei semiconduttori, proponendo due equazioni fondamentali, che saranno spesso usate nella trattazione:

$$n = n_i = \int_{E_C}^{+\infty} N_{BC}(E) f(E) dE = N_{C,EFF} e^{-\frac{E_C - E_F}{kT}}$$

$$p = p_i = \int_{-\infty}^{E_V} N_{BV}(E) f(E) dE = N_{V,EFF} e^{\frac{E_V - E_F}{kT}}$$

I due termini N_{EFF} sarebbero termini efficaci, che si riferiscono alla definizione

$$N_{ef} = \frac{(2m^* \pi kT)^{\frac{3}{2}}}{h}$$

Dove m^* è la massa efficace; sostituendo la massa efficace di elettroni o lacune, si trovano rispettivamente la prima e la seconda equazione. Queste due equazioni, si ottengono mediante un'approssimazione della distribuzione di Fermi-Dirac alla distribuzione di Boltzmann, integrandola.

Ci imponiamo ora di calcolare la posizione energetica del livello di Fermi, E_{F_i} : considerando N_C ed N_V valori efficaci, possiamo dire che

$$n_i = p_i = N_C e^{-\frac{E_C - E_{F_i}}{kT}} = N_V e^{-\frac{E_{F_i} - E_V}{kT}}$$

Sfruttiamo ora le condizioni di neutralità del semiconduttore intrinseco, e cerchiamo di risolvere tale equazione rispetto al parametro E_{F_i} :

$$e^{-\frac{E_C - E_{F_i}}{kT}} \cdot e^{\frac{E_{F_i} - E_V}{kT}} = \frac{N_V}{N_C}$$

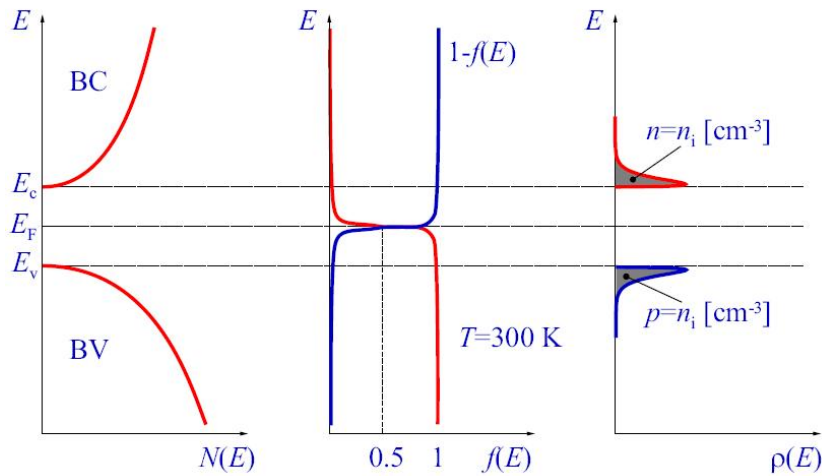
$$e^{-\frac{2E_{F_i}}{kT}} \cdot e^{-\frac{E_V + E_C}{kT}} = \frac{N_V}{N_C}$$

$$\implies E_{F_i} = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln \frac{N_V}{N_C}$$

Si noti tuttavia che le due concentrazioni N_V ed N_C differiscono esclusivamente per la massa efficace di rispettivamente lacune ed elettroni; possiamo dunque dire che, in caso di semiconduttori intrinseci quali quelli che stiamo attualmente trattando, il termine logaritmico corregga solo di pochi meV (al più una decina) il livello di Fermi:

$$\frac{kT}{2} \ln \frac{N_V}{N_C} \simeq 0 \implies E_{F_i} \simeq \frac{E_C + E_V}{2}$$

Ci è ora possibile quotare il livello di Fermi intrinseco, potendo dire con un'ottima approssimazione che esso si trova a metà del gap tra banda di valenza e banda di conduzione. Di fatto, il livello di Fermi intrinseco perde molto significato fisico, dal momento che raramente capita di aver a che fare con semiconduttori intrinseci; sarà però spesso usato come punto di riferimento per i nostri calcoli e le nostre congetture. Ora è veramente possibile quindi disegnare il diagramma a bande dettagliato e quotato di un semiconduttore, e nella fattispecie del silicio:



Valori fondamentali da ricordare, nel silicio, sono:

- E_g : Energy Gap (1,12 eV nel Silicio);
- $q\chi$: Affinità Elettronica (4,05 eV nel Silicio);
- $q\Phi_S$: Lavoro di estrazione.

Abbiamo definito anche nel semiconduttore il lavoro di estrazione, $q\Phi_S$; si noti che questo perde il significato fisico fortissimo che aveva nel metallo, in quanto al livello di Fermi intrinseco, E_{F_i} , non abbiamo elettroni, e dunque il lavoro necessario per estrarre elettroni sarà ben maggiore di quello che dovremmo compiere in un metallo. Affinità elettronica ed energy gap sono utili in quanto restano costanti, indipendentemente dai processi tecnologici che vengono attuati sul substrato semiconduttivo, quindi utili riferimenti per i calcoli che sarà necessario attuare.

1.2.1 Calcolo di n_i

Abbiamo parlato del numero di elettroni in banda di conduzione, lo abbiamo eguagliato al numero di lacune in banda di valenza per quanto riguarda semiconduttori intrinseci, e abbiamo accennato al fatto che il moto dei portatori di carica, ossia la corrente elettrica, sarà data sia dagli elettroni in banda di conduzione, che dalle lacune in banda di valenza. Studiando un semiconduttore intrinseco, i fenomeni che stabiliscono la conduzione sono principalmente due, uno in contrapposizione con l'altro:

- Generazione Termica (Gth): una temperatura non nulla muove un certo numero di elettroni dalla banda di valenza alla banda di conduzione, generando per l'appunto termicamente coppie elettrone-lacuna;
- Ricombinazione (R): gli elettroni, una volta separatisi dalle lacune, possono ritornare nella banda di valenza, in seguito ad un certo transitorio di tempo trascorso in banda di conduzione, per ricombinarsi con le lacune, facendo sparire una coppia elettrone-lacuna; si usa dire per questo che sia il fenomeno antagonista della generazione termica.

Possiamo qualitativamente dire che i termini n_i e p_i rappresentino il punto di equilibrio tra la ricombinazione termica e la generazione, siano cioè il numero di coppie elettrone-lacuna presenti ad una certa temperatura. Vogliamo però studiare espressioni che ci permettano di calcolare n_i , ed in seguito più generalmente n . I nostri punti di partenza, come abbiamo già detto in precedenza, sono:

$$n_i = N_C e^{-\frac{E_C - E_{F_i}}{kT}}$$

$$p_i = N_V e^{-\frac{E_{F_i} - E_V}{kT}}$$

Siamo stati soliti finora usare come condizione al contorno del nostro problema la neutralità; ci capiterà spesso però di studiare condizioni molto meno favorevoli, in cui la neutralità non sarà presente nel sistema; moltiplichiamo dunque n_i e p_i , tra loro, e vediamo che:

$$n_i \cdot p_i = n_i^2 = N_C N_V e^{-\frac{E_C - E_{F_i}}{kT}} \cdot e^{-\frac{E_{F_i} - E_V}{kT}} = N_C N_V e^{-\frac{E_V - E_C}{kT}}$$

Quindi:

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

Vediamo alcune cose: sappiamo da osservazioni precedentemente fatte che N_C ed N_V variano proporzionalmente a $T^{\frac{3}{2}}$; di conseguenza, la concentrazione dei portatori intrinseci varierà più che esponenzialmente: oltre al contributo dell'esponenziale, c'è quello appena citato di N_C ed N_V . Anche il gap energetico E_g è funzione della temperatura: aumentando T , E_g tende a diminuire; non esistono funzioni analitiche in grado di determinare un andamento del gap di energia in funzione della temperatura, a meno che non si ricavino da processi di statistica inferenziale (fitting), che studiano l'andamento del passo reticolare medio al variare della temperatura; all'aumentare della temperatura, aumenta il passo reticolare medio, e quindi il gap energetico tende a ridursi. Per questo motivo, il gap è molto influenzabile dalla temperatura, ma poichè noi consideriamo di trovarci sempre a 300 K, possiamo considerare il nostro sistema invariante rispetto a problemi di questo tipo.

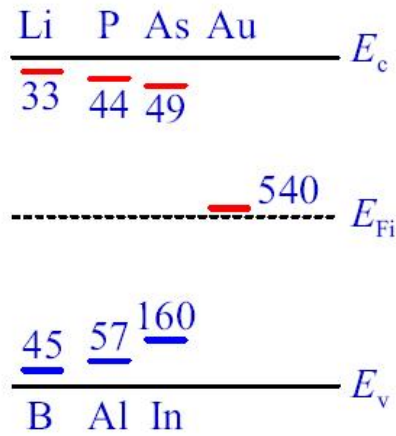
1.3 Caratterizzazione dei semiconduttori drogati

Abbiamo finora esaminato semiconduttori regolari, intrinseci, e quindi non trattati tecnologicamente in alcun modo; iniziamo ora a considerare lo studio della variazione delle caratteristiche elettriche di un semiconduttore, al variare di alcuni parametri tecnologicamente impiantati all'interno di un semiconduttore, considerando idealmente che essi vengano disposti in modo regolare, e omogeneo nel substrato semiconduttivo.

Un semiconduttore si può drogare in due modi: o tipo n, ossia incrementando gli elettroni in banda di conduzione, o tipo p, incrementando le lacune in banda di valenza. Concettualmente e a livello fisico i due processi presentano notevoli differenze, anche se il risultato è molto simile. Incominciamo la trattazione dei semiconduttori drogati, parlando di un drogaggio tipo n.

Definiamo N_D la concentrazione di drogante tipo n inserito nel semiconduttore. Droganti comunemente usati in processi di drogaggio del silicio sono l'arsenico As, o il fosforo P. Cosa significa però, precisamente, drogare un semiconduttore?

Drogare tipo n significa aggiungere stati artificiali nell'energy gap, introducendo livelli energetici particolari, detti livelli donatori E_D , in prossimità di E_C . Arsenico e fosforo sono ideali droganti per il silicio proprio perchè si andranno a posizionare tendenzialmente in prossimità dell'inizio della banda di conduzione (circa 40 meV l'arsenico, e pochi meV in più il fosforo). I drogaggi variano dall'introduzione di N_D dai 10^{14} ai 10^{19} atomi su cm^3 .



In un semiconduttore drogato, le fenomenologie che provocheranno conduzione saranno tre:

- Generazione termica;
- Ricombinazione;
- Ionizzazione.

Drogare significa introdurre nel semiconduttore un fenomeno di ionizzazione; ad una temperatura nulla, $T=0$, possiamo considerare i livelli energetici donatori E_D come delle sorte di gusci, contenenti elettroni. Questi gusci tenderanno ad aprirsi, all'aumentare della temperatura T , e più T aumenterà, più saranno i gusci che si apriranno, lasciando liberi gli elettroni, che tenderanno ad andare verso la banda di conduzione, vicinissima ai livelli E_D . Da un salto di 1,12 eV, essi dovranno solo saltare pochi meV, e dunque un effetto termico pressochè minimo permetterà una transizione in banda di conduzione.

Si noti che la ionizzazione non ha un fenomeno antagonista, a differenza della ricombinazione e della generazione termica che si elidono tra loro: ciò è dovuto sia dal fatto che a 300 K non vi è la possibilità fisica che un elettrone tenda a tornare verso il proprio guscio, sia e soprattutto dal fatto che gli E_D sono livelli energetici statici, localizzati, e non creano lacune. La ionizzazione dunque realizza stati energetici che non dispongono assolutamente delle stesse proprietà di trasporto di cui disponevano le lacune in precedenza descritte.

Per il drogaggio tipo p, il discorso è piuttosto simile, o meglio duale: gli ioni droganti saranno identificati come N_A , e daranno luogo a livelli accettori E_A , che tendono a disporsi in prossimità della banda di valenza, E_V . Un

ottimo drogante tipo p per il silicio è il boro B. Anche nel drogaggio tipo p avverrà un fenomeno di ionizzazione, che però agirà in senso diverso, fisicamente: gli atomi idonei per il drogaggio tipo p, sono in grado, a spesa di un minimo effetto termico, di aprirsi ed accettare elettroni dalla banda di valenza, come dei gusci che si aprono, e catturano elettroni, anzichè liberarli (a differenza di prima). Man mano che la temperatura aumenta, sempre più gusci si apriranno e cattureranno elettroni. Anche qua il gap da saltare per gli elettroni sarà infinitesimo, poichè il buon drogante posizionerà i propri livelli accettori in prossimità del livello E_V , e quindi daran vita a fenomeni di trasporto in banda di valenza, senza avere fenomeni duali in contrapposizione. Aldilà di queste osservazioni fisiche qualitative, le fenomenologie sono del tutto analoghe a quelle descritte nel semiconduttore tipo n.

A partire da ciò che abbiamo finora descritto, facciamo alcune puntualizzazioni onde evitare confusione:

- Non tutti gli atomi sono in grado di drogare un substrato semiconduttivo, ma solo certi, in base ad alcune caratteristiche che devono possedere. Nel caso del silicio, per drogare tipo n devono essere pentavalenti, e posizionare i propri livelli donatori in prossimità di E_C , per drogare tipo p devono essere trivalenti e posizionare i propri livelli accettori in prossimità di E_V ;
- L'effetto del drogaggio provoca una notevole alterazione della popolazione delle bande energetiche, e dunque del diagramma a bande;
- L'effetto del drogaggio non è puramente additivo: non siamo sicuri che, aggiungendo ad esempio 10^{10} atomi di arsenico, gli elettroni raddoppieranno (dalla loro condizione iniziale a 300 K di $1,45 \cdot 10^{10}$, intrinseci); siamo sicuri di un miglioramento nella conducibilità, per ora non quantificabile, a causa di fenomeni intrinseci di ricombinazione.

1.3.1 Calcolo di n e p in un semiconduttore drogato

Per poter determinare parametri come le concentrazioni dei portatori attivi, n e p, in un semiconduttore drogato (e quindi non più intrinseco), dobbiamo innanzitutto cercare una condizione al contorno idonea per potervi poi applicare le giuste equazioni. Supponiamo per ora di drogare tipo n, con concentrazione di drogante N_D . Il meccanismo di ionizzazione ha una forte dipendenza dalla temperatura, quindi di fatto degli N_D droganti ne sarà attiva, ionizzata solo una parte, che chiameremo N_D^+ . Per ognuno di questi droganti ionizzati, vi sarà un elettrone libero, in banda di conduzione. Dover studiare una dipendenza della ionizzazione alla temperatura è molto

pesante per i conti che ne deriverebbero, ma si può applicare una condizione interessantissima che ci accompagnerà spesso nella trattazione: l'ipotesi di completa ionizzazione. Studiando l'andamento della ionizzazione al variare della temperatura, si nota che per una temperatura intorno ai 100 K, tutte le impurità droganti saranno ionizzate, e dunque attive, dopodichè aumentando la temperatura si instaurerà un regime di saturazione, poichè non vi sono ulteriori impurità da ionizzare. Poichè noi dunque studiamo sistemi ad una temperatura $T \sim 300$ K, potremo sempre considerare valida l'ipotesi di completa ionizzazione, e quindi

$$N_D^+ = N_D$$

Il livello di Fermi E_F , nel drogaggio tipo n, si innalzerà rispetto alla posizione di partenza, nel semiconduttore intrinseco, E_{F_i} . Si può dire che infatti il livello di Fermi sia una sorta di baricentro dei livelli energetici, come vedremo tra breve; non allontaniamoci per ora dal problema di base, ossia il calcolo delle concentrazioni n e p. Supponiamo di conoscere E_F , ossia il livello di Fermi alterato dal drogaggio. Consideriamo n_n il numero di elettroni in banda di conduzione in seguito al drogaggio tipo n (come suggerisce il pedice), e p_n il numero di lacune in banda di valenza sempre in questo stato:

$$n_n = N_C e^{-\frac{E_C - E_F}{kT}}$$

$$p_n = N_V e^{-\frac{E_F - E_V}{kT}}$$

Consideriamo nuovamente, come prima, il prodotto delle due equazioni membro a membro:

$$n_n p_n = N_C N_V e^{-\frac{E_C - E_F}{kT}} \cdot e^{-\frac{E_F - E_V}{kT}} = N_C N_V e^{-\frac{E_g}{kT}} = n_i^2$$

Abbiamo appena fatto, in seguito a questo semplice conto, una scoperta interessantissima: qualsiasi siano le concentrazioni di portatori di carica nelle bande, il loro prodotto è una costante, e questa costante è il quadrato degli elettroni in banda di conduzione nel semiconduttore in stato intrinseco. Questo fenomeno è detto legge dell'azione di massa:

$$n_n \cdot p_n = n_i^2$$

Ciò ci fa rendere conto però anche di un problema gravissimo: tecnologicamente non è possibile crescere una delle due concentrazioni senza andare a scapito dell'altra: le due concentrazioni non possono crescere di pari passo, poichè il loro prodotto sarà sempre e comunque costante.

Consideriamo anche un altro fatto: la carica totale del semiconduttore è nulla: ogni lacuna in banda di valenza ha un elettrone in banda di conduzione corrispondente, ed ogni elettrone derivante dal drogaggio ha un livello (accettore in drogaggio tipo p, donatore nel drogaggio tipo n) ionizzato corrispondente. Esisterà dunque una condizione di neutralità del semiconduttore, ma globale: la somma di tutte le cariche del semiconduttore sarà nulla.

Applichiamo dunque la condizione di neutralità globale del semiconduttore:

$$q(p_n + N_D^+ - n_n) = 0 \implies p_n + N_D^+ - n_n = 0$$

Dalla legge dell'azione di massa, sappiamo che

$$p_n = \frac{n_i^2}{n_n} \implies \frac{n_i^2}{n_n} + N_D^+ - n_n = 0$$

Dunque, abbiamo un'equazione di secondo grado con variabile indipendente n_n ; risolvendola, e considerando solo la radice positiva, si ottiene:

$$n_n = \frac{N_D^+}{2} + \sqrt{\left(\frac{N_D^+}{2}\right)^2 + n_i^2}$$

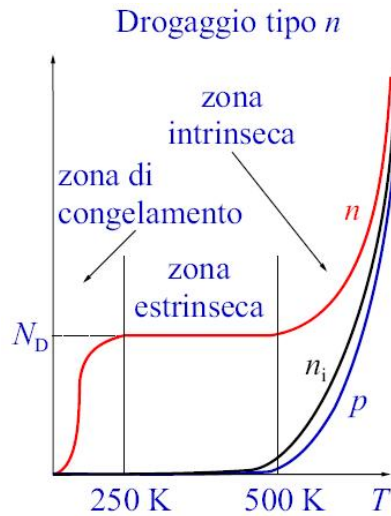
Consideriamo il fatto che le grandezze sono funzione della temperatura T :

$$n_n(T) = \frac{N_D^+(T)}{2} \left[\sqrt{1 + \left(\frac{2n_i(T)}{N_D^+(T)}\right)^2} + 1 \right]$$

Però $n_i(T)$ è nota, e di $N_D^+(T)$ conosciamo l'andamento. Possiamo pensare in modo molto qualitativo che l'andamento dei portatori maggioritari, ossia degli elettroni in banda di conduzione nel semiconduttore drogato tipo n, al variare della temperatura, sia il seguente:

Si possono distinguere tre regioni, associate a tre regimi particolari:

1. Regime di congelamento: per temperature inferiori ai 100 K, l'attivazione dei droganti non è ancora completa, la generazione termica è molto bassa a causa della scarsa temperatura e dunque scarsa energia termica del sistema, e quindi vi sarà una bassa conducibilità;
2. Regime estrinseco: per temperature superiori a 100 K, si può considerare l'ipotesi di completa ionizzazione, e quindi si ha un regime di saturazione: $N_D^+ = N_D$, e $n_n \simeq N_D$. Il fatto che si siano attivati completamente i droganti, ma comunque vi siano scarsi effetti termici, fa prevalere le interazioni di tipo elettrostatico su tutte le altre,



nella fattispecie su quelle di origine termica. Questo è il regime che tendenzialmente si utilizza, studiando dispositivi a semiconduttore;

3. Regime intrinseco: per temperature alte, le interazioni fononiche, ossia derivanti da effetti termici, prevalgono su quelle elettrostatiche, e così l'effetto del drogaggio tende a perdersi, per farsi sostituire da elettroni intrinseci del materiale; in questo regime, $n_n \simeq n_i(T)$. Questo tipo di comportamento è stato fatale per il germanio: il padre dei semiconduttori è infatti stato l'appena citato germanio, semiconduttore a gap estremamente piccolo. Il fatto che il gap fosse così ridotto, fa sì che il regime intrinseco si instauri con una temperatura molto inferiore rispetto a quella che lo farebbe instaurare nel silicio, di conseguenza n_i cresce molto più rapidamente, e quindi è scarsamente utilizzabile in regime estrinseco, ossia quello utile nella maggior parte dei dispositivi.

Abbiamo largamente discusso l'andamento della funzione $n_n(T)$; potremmo ora chiederci quale sia l'andamento della funzione dei minoritari, $p_n(T)$, ossia delle lacune in banda di valenza in un semiconduttore drogato tipo n . Sappiamo che

$$p_n(T) = \frac{n_i^2(T)}{n_n(T)}$$

Noi conosciamo bene sia l'andamento di $n_i(T)$ che di $n_n(T)$; la differenza fondamentale da quest'ultima è che i portatori minoritari non hanno una zona di saturazione, in un certo livello energetico, ma la concentrazione varierà in modo puramente intrinseco. Non è inutile lo studio di questo andamento,

poichè alcuni dispositivi si basano sullo studio dell'andamento dei minoritari, anche se ciò sarà molto meglio approfondito in seguito.

In un drogaggio tipo p, la situazione è del tutto duale a quella appena descritta:

$$q(p_p - N_A^- - n_p) = 0 \implies p_p - N_A^- - n_p = 0$$

Dall'azione di massa,

$$n_p = \frac{n_i^2}{p_p} \implies p_p^2 - N_A^- - n_i^2 = 0$$

Da ciò, ragionando dualmente a prima,

$$p_p(T) = \frac{N_A^-(T)}{2} \left[\sqrt{1 + \left(\frac{2n_i(T)}{N_A^-(T)} \right)^2} + 1 \right]$$

L'andamento è dunque del tutto analogo al precedente, esistono i tre regimi, con le stesse condizioni del drogaggio tipo n. Stessa cosa per la concentrazione dei minoritari $n_p(T)$, al variare della temperatura.

Potremmo a questo punto riproporre un problema precedentemente accennato: immaginiamo di drogare un substrato semiconduttivo sia tipo p che tipo n; supponiamo di essere tecnologicamente in grado di poter introdurre una concentrazione di droganti N_A uguale a quella di droganti N_D ; applicando la neutralità globale e la legge dell'azione di massa, si ottiene che:

$$\begin{cases} p \cdot n = n_i^2 \\ q(p - n + N_D^+ - N_A^-) = 0 \end{cases} \implies \begin{cases} N_D^+ = N_A^- \\ p - n = 0 \end{cases}$$

Si capisce che l'effetto del drogante N_D compensa quello di N_A , e viceversa; se i due drogaggi sono dunque uguali, il semiconduttore torna semplicemente ad essere intrinseco. Ciò ci induce a pensare che vi sia un meccanismo di compensazione dei droganti. Abbiamo dunque analizzato una casistica, ossia $N_D = N_A$; analizziamo le rimanenti due, al fine di poter studiare più profondamente questo meccanismo di compensazione.

- $N_D > N_A$: i donatori prevalgono sugli accettori, e così il drogante tipo donatore in totale sarà $N_D^{+'}$:

$$N_D^{+'} = N_D^+ - N_A^-$$

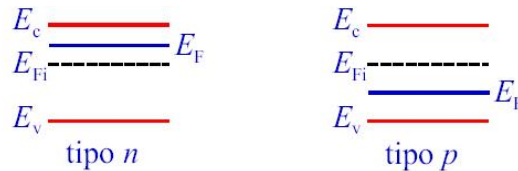
- $N_D < N_A$: situazione duale a prima, dove gli accettori prevalgono sui donatori, e quindi $N_A^{-'}$ sarà:

$$N_A^{-'} = N_A^- - N_D^+$$

Un esempio pratico di semiconduttore drogato sia n che p è il transistor bipolare (anche se sarà necessario introdurre molte altre nozioni prima di arrivare a trattarlo); aprendo finestre tipo p in un semiconduttore drogato tipo n, si può ottenere un prototipo di transistor bipolare.

1.3.2 Equazioni di Shockley

Partendo dalle nostre conoscenze, siamo ora in grado di disegnare e quotare il diagramma a bande dettagliato di un semiconduttore drogato tipo n:



I simboli usati sono sempre i soliti; si può considerare valida l'ipotesi di completa ionizzazione, e a partire da questa effettuare i nostri studi; sappiamo che:

$$n_n \sim N_D = N_C e^{-\frac{E_C - E_F}{kT}} \implies E_C - E_F = kT \ln \frac{N_C}{N_D}$$

Possiamo dunque calcolare $E_C - E_F$, e da qui quotare il lavoro di estrazione:

$$q\Phi_{S_n} = E_C - E_F + q\chi = q\chi + kT \ln \frac{N_C}{N_D}$$

A seconda del drogaggio, il lavoro di estrazione varrà dai 4,61 al 4,05 eV.

Il calcolo che però noi abbiamo effettuato di E_F non è ottimale: se dovessimo studiare il livello di Fermi in un semiconduttore drogato tipo p, dovremmo ridisegnare da zero il diagramma a bande, rifare da capo le nostre congetture, e considerare la presenza di dati che comunemente non si considerano, quali E_C o E_V . Utilizzando le relazioni appena studiate, in un semiconduttore drogato tipo p il lavoro di estrazione sarebbe:

$$q\Phi_{S_p} = q\chi + E_g - kT \ln \frac{N_V}{N_A}$$

Proponiamo ora qualcosa di diverso, ossia un metodo più globale, più semplice, più generale per il calcolo del livello di Fermi E_F , sia in semiconduttori tipo n che in semiconduttori tipo p. Partiamo dalle solite:

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

$$p = N_V e^{-\frac{E_F - E_V}{kT}}$$

Considerate in un semiconduttore intrinseco, esse saranno:

$$n_i = N_C e^{-\frac{E_C - E_{F_i}}{kT}}$$

$$p_i = N_V e^{-\frac{E_{F_i} - E_V}{kT}}$$

Ricaviamo ora da queste ultime due, N_C ed N_V , tenendo conto che in un semiconduttore intrinseco, $n_i = p_i$:

$$N_C = n_i e^{\frac{E_C - E_{F_i}}{kT}}$$

$$N_V = n_i e^{\frac{E_{F_i} - E_V}{kT}}$$

Sostituendo le espressioni appena trovate a partire dalle equazioni intrinseche, N_C ed N_V , nelle equazioni del semiconduttore non intrinseco, troviamo:

$$n = n_i e^{\frac{E_C - E_{F_i}}{kT}} \cdot e^{-\frac{E_C - E_F}{kT}}$$

$$p = n_i e^{\frac{E_C - E_{F_i}}{kT}} \cdot e^{-\frac{E_F - E_V}{kT}}$$

Ordinando e semplificando alcuni termini, si ricavano le cosiddette Equazioni di Shockley, in onore del fisico statunitense che le ha formulate per la prima volta:

$$n = n_i e^{\frac{E_F - E_{F_i}}{kT}}$$

$$p = n_i e^{\frac{E_{F_i} - E_F}{kT}}$$

Queste equazioni sono particolarmente interessanti in quanto in esse spariscono i parametri estremamente scomodi che nelle relazioni precedentemente utilizzate erano da considerare come noti; gli unici parametri richiesti sono

n_i , che ci è noto, e il livello di Fermi intrinseco, che si troverà esattamente a metà dell'energy gap: quando si parlava di utilizzo del livello di Fermi intrinseco come punto di riferimento, si voleva proprio arrivare a questo risultato: un'equazione in grado di fornire informazioni enormi a partire da dati minimi. A partire da queste, invertendole, sarà semplicissimo calcolare il livello di Fermi:

$$E_F - E_{F_i} = kT \ln \frac{N_D}{n_i}$$

$$E_{F_i} - E_F = kT \ln \frac{N_A}{n_i}$$

Dopo la formulazione delle Equazioni di Shockley, possiamo ritenere conclusa questa prima trattazione del comportamento fisico di un semiconduttore, e utilizzare i mezzi appresi per studiare una prima modellazione di un semiconduttore.

Capitolo 2

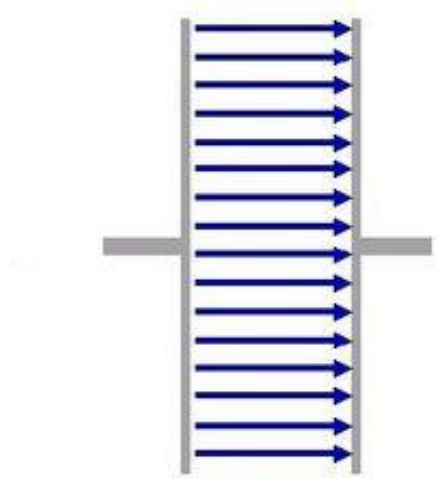
Modello matematico delle correnti nei semiconduttori

In un metallo, gli elettroni come già detto si trovano tutti in banda di conduzione, in continuo movimento, a causa di un moto di agitazione termica. Essi possono solo essere trascinati da un campo elettrico ε , che provoca una corrente di drift, cioè di trascinamento.

In un semiconduttore all'equilibrio termodinamico, le correnti sono nulle. Supponiamo di considerare un semiconduttore con contatti metallici (al fine di disperdere resistenze parassite), i quali sono collegati ad una batteria che fornisce una tensione V_a ; all'interno del semiconduttore vi sarà un campo elettrico direzionato nel verso dell'asse delle x da noi introdotto, $E_x = \varepsilon$. Ci poniamo a questo punto una domanda legittima: che effetto provoca l'applicazione di una tensione V_a , al semiconduttore? Come varierà il suo diagramma a bande in seguito all'applicazione di questa tensione esterna?

Supponiamo di studiare un condensatore, e di poter disporre di un voltmetro ideale, in grado di misurare la $\Delta V(x)$, ossia la differenza di potenziale al variare della posizione spaziale. Fissiamo l'origine del nostro sistema di riferimento cartesiano spaziale sulla faccia sinistra del condensatore, e l'asse delle x ortogonale alla faccia, verso positivo da sinistra a destra; posizioniamo uno dei due puntali del voltmetro a massa sulla faccia sinistra del condensatore, e muoviamo l'altro all'interno di questo sistema: avremo come risultato una misura della differenza di potenziale al variare della posizione, in riferimento alla prima faccia. La tensione avrà, al variare della x , un andamento lineare, ossia varierà come una retta, con pendenza positiva: man mano che ci si allontana dalla faccia cui è posta la massa, la tensione tra i due puntali continuerà ad aumentare. Giunti al punto d , ossia dopo aver raggiunto la seconda faccia del condensatore, la tensione sarà V_a , ossia la tensione ai capi del sistema (considerando condensatore a facce piane parallele, escludendo

eventuali effetti di bordo e quindi considerando un campo elettrico uniforme su tutta la superficie del condensatore).



Come varierà l'energia potenziale di un elettrone in questo condensatore? Noi sappiamo, dallo studio dell'elettrostatica, che:

$$\varepsilon = -\frac{V_a}{d}; F = -q\varepsilon = -q\left(-\frac{V_a}{d}\right)$$

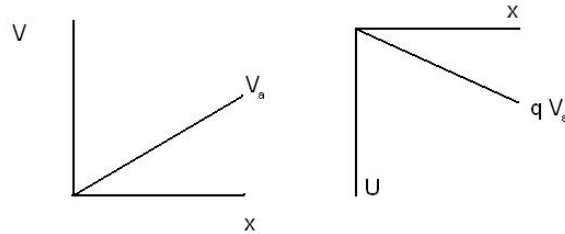
L'energia potenziale al variare della posizione spaziale x , $U(x)$, da 0 a x , sarà l'integrale di linea della forza elettrostatica del condensatore; poichè abbiamo posto il puntale a massa sulla faccia sinistra del condensatore, e lo 0 della posizione spaziale in prossimità della massa:

$$U(x) = U(0) - \int_0^x F dx = 0 - \int_0^x q \frac{V_a}{d} dx$$

Dunque, possiamo dire, risolvendo l'integrale, che:

$$U(x) = -\frac{qV_a}{d}x$$

Misurando l'energia in eV, il grafico coincide con il simmetrico rispetto all'asse delle x del grafico appena visto per la tensione:



Esiste dunque un forte legame tra tensione ed energia potenziale di un elettrone: poichè abbiamo trovato un'espressione dell'energia potenziale come retta a pendenza negativa che da 0 raggiunge un livello $-qV_a$, e poichè il diagramma a bande altri non è che il grafico dell'energia di un elettrone, una tensione esterna V_a applicata ad un diagramma a banda, tende a ruotare tutte le grandezze presenti all'interno del diagramma, di un certo angolo. Effettivamente sembra piuttosto stravagante come idea, ma del tutto sensata: poichè l'energia potenziale di un elettrone di fatto, in seguito all'applicazione di una tensione esterna, è una retta a pendenza negativa, allora anche il diagramma a bande dovrà subire una sorta di pendenza negativa, per poter tener conto degli effetti della tensione sull'energia dell'elettrone. Fingendo che ora al posto di un condensatore vi sia il nostro substrato semiconduttivo, il discorso appena fatto è perfettamente valido.

Questo diagramma inclinato rappresenta dunque il punto di vista degli elettroni: il campo elettrico provocato dalla tensione V_a farà spostare verso il basso (ossia in senso ad esso opposto) gli elettroni, e verso l'alto (ossia in verso ad esso concorde) le lacune. Possiamo quasi immaginare, per fare un confronto azzardato ma efficace, che l'applicazione di una polarizzazione V_a , inclinando il diagramma a bande, faccia risalire le lacune, come se fossero bollicine d'aria, e faccia scendere gli elettroni, come se fossero palline di ferro.

In seguito a questa premessa, che ci permette di capire in modo del tutto qualitativo il comportamento di un semiconduttore e la variazione del suo diagramma a bande in seguito all'applicazione di una tensione esterna, introduciamo alcuni nuovi concetti; si consideri sempre il fatto che il semiconduttore è drogato in modo del tutto uniforme, e che dunque vi è una condizione di neutralità sicuramente globale, ma anche locale: se infatti sezionassimo il semiconduttore, vedremmo che ogni singola fetta è localmente neutra, ossia che la somma di tutti gli eccessi di cariche in essa presenti è nulla. Questo perchè l'applicazione di un campo elettrico esterno non cambia in alcun modo la concentrazione degli elettroni in banda di conduzione. Possiamo immaginare che la differenza di potenziale V_a esterna sposti gli

elettroni verso destra, dove vi è il contatto metallico nel quale gli elettroni possono entrare liberamente; nel lato sinistro, capiterà qualcosa di duale: il semiconduttore a causa della tensione e del campo da essa derivante preleva dal contatto ohmico sinistro elettroni. Un elettrone, assorbito dalla lamina destra, lascia un buco, una lacuna, che sarà poi in seguito ricombinata dalla lamina sinistra, che fornisce elettroni al semiconduttore. Questo modello può intuitivamente spiegare così il trasporto di cariche.

Cerchiamo di formalizzare ora tutti questi concetti, partendo da una definizione: vogliamo cercare di capire esattamente che cosa sia una corrente.

Una densità di corrente J è un flusso di cariche, ossia di elettroni; per flusso possiamo intendere semplicemente il prodotto di una velocità per una densità. Date le ipotesi precedentemente illustrate, l'unica corrente che avremo a disposizione sarà quella di drift; la densità di corrente dovuta agli elettroni in banda di conduzione sarà data da:

$$J_- = \Phi_{x,e^-}(-q)$$

Il contributo delle correnti nel semiconduttore però sarà dato anche da una densità di corrente di lacune, formalizzata come:

$$J_+ = \Phi_{x,e^+}(+q)$$

In totale, possiamo dire che la totale densità di corrente sarà data da:

$$J = \Phi_{x,e^-}(-q) + \Phi_{x,e^+}(+q)$$

Abbiamo però detto qualche parola riguardo questo flusso, Φ , senza approfondire molto, dicendo che è una densità di elementi moltiplicato per una velocità di moto. Le ipotesi di neutralità ci spingono a considerare velocità v_d , ossia di drift. Partendo da ciò, potremo dire che, date le nostre ipotesi, le espressioni dei due flussi saranno:

$$\Phi_{x,e^-} = nv_{d,e^-}$$

$$\Phi_{x,e^+} = nv_{d,e^+}$$

Sostituendo nell'espressione della densità totale di corrente J ,

$$J = -qnv_{d,e^-} + qnv_{d,e^+}$$

Stiamo ora continuando a parlare della velocità di drift, senza però averne fornito alcuna informazione: essa è una velocità non istantanea, ma media, dell'intera popolazione: essa è un parametro fittizio che descrive, ba-

sandosi sullo studio statistico di tutta la popolazione, la velocità media di trascinamento di un singolo elettrone, tra un urto ed un altro con il reticolo.

Definiamo alcune nuove relazioni:

$$v_{d,e^-} = -\mu_n \varepsilon$$

$$v_{d,e^+} = \mu_p \varepsilon$$

Dove μ rappresenta la mobilità dei portatori di carica, e ε è il solito campo elettrico derivante dalla polarizzazione esterna. Mediante alcune osservazioni, si può dire che:

$$\mu_n = \frac{-q\tau}{m_n^*}$$

$$\mu_p = \frac{q\tau}{m_p^*}$$

$$\tau \propto \frac{\lambda}{N_{th}}$$

Il parametro λ ha una forte dipendenza dalla regolarità del reticolo cristallino, che noi consideriamo perfetto per ipotesi.

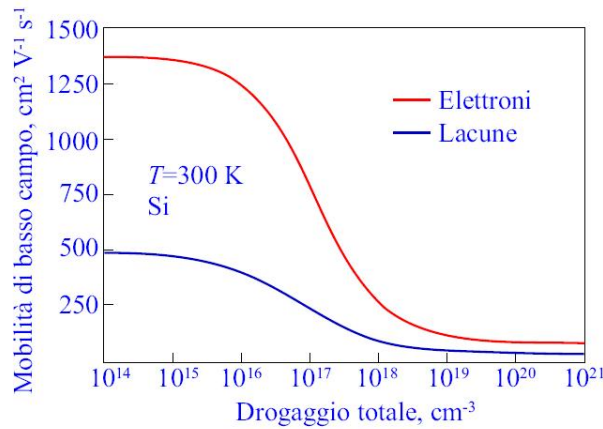
Riassumendo ciò che abbiamo appena detto, la mobilità degli elettroni è maggiore di quella delle lacune, poichè l'unico parametro variabile è la massa efficace m^* , notoriamente superiore nelle lacune che negli elettroni. La mobilità dei portatori risente di alcune caratteristiche:

- Maggiore sarà la temperatura, maggiore sarà l'interazione fononica coi nuclei, e dunque minore la mobilità dei portatori;
- Maggiore sarà il numero di imperfezioni del reticolo, minore sarà la mobilità dei portatori;
- Maggiore sarà il drogaggio, e dunque la somma delle concentrazioni ($N_A + N_D$), e minore sarà la mobilità dei portatori di carica, poichè vi sarà una popolazione maggiore.

E' possibile spesso trovare dei grafici dei degradamenti di μ_n e μ_p :

Ora che abbiamo introdotto anche la mobilità elettronica, e una sua stima grafica, possiamo completare la definizione della densità di carica, ed arrivare ad una prima definizione di conducibilità elettrica in un semiconduttore:

$$J_{n,drift} = (-q)n(-\mu_n)\varepsilon$$



$$J_{p,drift} = qp\mu_p\varepsilon$$

La densità di corrente di drift totale, J , sarà

$$J_d = (q\mu_n n + q\mu_p p)\varepsilon = \sigma\varepsilon$$

L'espressione operativa della conducibilità in un semiconduttore sarà dunque:

$$\sigma = \frac{1}{\rho} = q\mu_n n + q\mu_p p$$

In un metallo, la conducibilità è solo affidata al moto degli elettroni; in un semiconduttore, bisogna dunque considerare anche le lacune in banda di valenza, per avere un quadro completo delle correnti presenti.

2.1 Correnti di diffusione

Potrebbe capitare che, a causa di fenomeni particolari o per motivi elettrici/tecnologici, si abbia in un semiconduttore una concentrazione degli elettroni in banda di conduzione non uniforme; sezionando idealmente il semiconduttore, le fettine non avrebbero più carica complessiva nulla, e quindi la condizione di neutralità locale del semiconduttore sarebbe inapplicabile in questo ambito. In una situazione come questa (che molto spesso trova riscontro in applicazioni pratiche, come vedremo in seguito), avviene un fenomeno di diffusione degli elettroni, causato dal gradiente della concentrazione al variare della posizione in cui ci si trova. Questo perchè, qualitativamente parlando, gli elettroni tendono ad uniformarsi, e dunque ad occupare le sezioni del semiconduttore meno popolate. La diffusione degli elettroni causata

dal gradiente dunque provoca un flusso di elettroni, ed anche un flusso di lacune, poichè anche esse tenderanno all'omogeneità nel sistema. Il campo ε è del tutto ininfluenza in questo processo, o meglio il processo è indipendente da esso. Il campo elettrico provocava le correnti di drift precedentemente studiate, ora avremo però due nuove correnti, provocate dalla diffusione dei portatori:

$$J_{n,diff} = -q\Phi_{x,e^{-}diff}$$

$$J_{p,diff} = q\Phi_{x,e^{+}diff}$$

Possiamo dunque intuire ciò:

$$\Phi_{x,e^{-}diff} = -D_n \frac{\partial n}{\partial x}$$

$$\Phi_{x,e^{+}diff} = -D_p \frac{\partial p}{\partial x}$$

I termini D_n e D_p sono i coefficienti di diffusione rispettivamente di elettroni in banda di conduzione e di lacune in banda di valenza, per ora non ancora quantificati. Questi coefficienti, moltiplicati per la variazione nello spazio di concentrazione di portatori, ossia per la derivata parziale rispetto a x dei portatori rispettivi, definisce i due flussi di diffusione.

Possiamo considerare questo moto come confrontabile a quello di un gas; il gradiente della concentrazione di elettroni è negativo (secondo il sistema di riferimento precedentemente definito), ma quindi il flusso risulta essere positivo. Si noti che in un metallo questa situazione è assolutamente irrealizzabile: nel metallo gli elettroni sono come un mare, che assorbe facilmente qualsiasi perturbazione gli venga fornita. Tecnologicamente non è difficile creare correnti di diffusione in un semiconduttore, come vedremo in seguito parlando ad esempio della giunzione p-n.

2.2 Relazioni di Einstein ed Equazioni del Trasporto

Formalizziamo ora tutto ciò che abbiamo finora detto, partendo dalle ultime relazioni discusse: la densità totale di corrente di diffusione, J_{diff} , è data da:

$$J_{diff} = J_{n,diff} + J_{p,diff}$$

$$J_{n,diff} = qD_n \frac{\partial n}{\partial x}$$

$$J_{p,diff} = -qD_p \frac{\partial p}{\partial x}$$

Si possono ricavare i due coefficienti D_n e D_p , mediante le Relazioni di Einstein, e scoprire che:

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

$$D_p = \frac{kT}{q} \mu_p$$

Per semplicità, si definisce il termine $\frac{kT}{q}$ come equivalente in tensione della temperatura: esso infatti dimensionalmente è una tensione, chiamata convenzionalmente V_T :

$$V_T = \frac{kT}{q}$$

Il valore di V_T ad una temperatura di 300 K, è circa di 26 mV.

Mediante queste prime relazioni, abbiamo trovato i primi tasselli del puzzle rappresentante il modello matematico delle correnti in un semiconduttore: le Relazioni di Einstein, e le equazioni del trasporto che ora esporremo in forma ordinata:

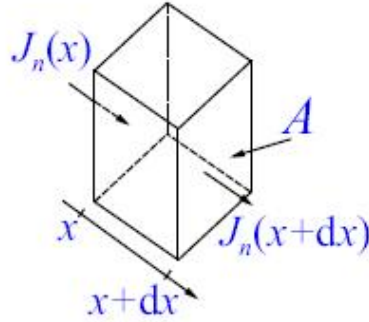
$$J_n = qn\mu_n\varepsilon + qD_n \frac{\partial n}{\partial x}$$

$$J_p = qp\mu_p\varepsilon - qD_p \frac{\partial p}{\partial x}$$

2.3 Equazione di continuità

Supponiamo di avere un cubo di lato infinitesimo dx , e sezione A ; vogliamo quantificare la carica in esso contenuta. Supponiamo che il cubo si trovi, sul nostro sistema di posizione cartesiano, nel punto x , e termini quindi nel punto $x + dx$.

Nel punto x entrerà una certa corrente, e quindi un certo flusso di cariche, mentre nel punto $x + dx$ ve ne sarà presumibilmente un'altra, non sappiamo se uguale o meno. Formalizzando, possiamo dire che:



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

Cosa abbiamo detto qui: la variazione del numero di elettroni nel tempo, all'interno del volume del cubo, è uguale alla densità di carica diviso la carica fondamentale, e quindi è uguale al flusso di cariche che circolano nel cubo. In realtà, sarebbe necessario considerare, nella transizione all'interno del cubo, altri tipi di fenomenologie, quali la generazione termica e la ricombinazione. Formalizzando ulteriormente, possiamo aggiungere che:

$$\frac{\partial n}{\partial t} A dx = \frac{J_n(x) A}{-q} - \frac{J_n(x+dx) A}{-q} + (Gth - R) A dx$$

Sviluppiamo ora in serie di Taylor troncando al primo ordine il termine $J_n(x+dx)$:

$$J_n(x+dx) \simeq J_n(x) + \frac{\partial J_n(x)}{\partial x} dx$$

Sostituendo nell'espressione appena trovata, si trova che:

$$\begin{aligned} \frac{\partial n}{\partial t} A dx &= \frac{J_n(x) A}{-q} - \left(\frac{J_n(x) A}{-q} + \frac{1}{-q} \frac{\partial J_n(x)}{\partial x} A dx \right) + (Gth - R) A dx \\ \implies \frac{\partial n}{\partial t} &= \frac{\partial J_n}{\partial x} \frac{1}{q} - U_n \end{aligned}$$

Dove si definisce U_n come il tasso medio di ricombinazione:

$$U_n = R - Gth$$

In modo del tutto analogo si può ricavare l'equazione di continuità delle lacune, partendo dalle stesse equazioni precedentemente usate, e sfruttando tecniche analitiche ed algebriche del tutto uguali, si ottiene:

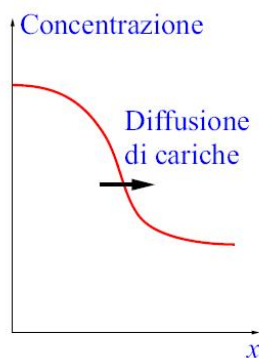
$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} - U_p$$

2.4 Equazione di Poisson

Il nostro modello matematico delle correnti nel semiconduttore è ancora molto carente: l'equazione di continuità è molto potente, tuttavia non siamo in grado di riferirci con precisione ai termini della suddetta, in quanto conosciamo molto poco riguardo ad essi.

Possiamo però sfruttare un'idea geniale: sostituire le due equazioni di continuità, all'interno delle equazioni di trasporto precedentemente ricavate: le soluzioni di questo sistema di equazioni sarà un'espressione analitica delle densità di portatori, in funzione di tempo e posizione, $n(x; t)$ e $p(x; t)$. Queste non sono però le uniche incognite del problema: leggendo le equazioni precedentemente introdotte, troviamo anche il campo elettrico ε , e il termine di ricombinazione U_n o U_p a seconda se si studiano rispettivamente elettroni o lacune.

Supponiamo dunque di avere, rispetto alla condizione di equilibrio, uno spostamento di cariche, un accumulo locale di cariche, in grado di instaurare gradienti di concentrazione e dunque un processo di diffusione mediante il generarsi di flussi di portatori. Formalizziamo, nel nostro modello matematico, questo moto di cariche.



Supponiamo di avere eccessi locali di cariche, uno negativo ed uno positivo: globalmente, il sistema sarà neutro. Gli eccessi di cariche daranno vita a campi elettrici nella zona compresa tra i due eccessi, non trascurabili. Incominciamo dunque a studiare la densità di distribuzione delle cariche, in tutti

gli eccessi, al variare della posizione x in cui osserviamo il semiconduttore in questo stato:

$$\rho(x) = +q(p(x; t) - n(x; t) + N_D^+(x) - N_A^-(x))$$

Data la distribuzione delle cariche, ora ci interessiamo del campo elettrico ε . Per far ciò, dobbiamo però aver ben presente il legame tra carica e campo elettrico, senza aver ancora introdotto l'Equazione di Poisson, risultato finale di questa sezione della trattazione del modello matematico. Apriamo una breve parentesi ad un primo studio del legame tra carica e campo elettrico.

2.4.1 Legame Carica - Campo Elettrico

Dato il solito modello semplificato di condensatore a facce piane parallele non dotato di effetti di bordo, applicata sulle facce una tensione $\Delta V = V_a$, osserviamo che il campo elettrico $\varepsilon(x)$ fuori dal condensatore è nullo; sappiamo dallo studio di Fisica riguardo l'elettrostatica, e dall'analisi dei campi vettoriali, che:

$$\varepsilon(x) = -\frac{\partial\Phi}{\partial x}$$

Dove Φ è il potenziale. Utilizzando il teorema fondamentale del calcolo integrale:

$$\Phi(x) = -\int_0^d \varepsilon(x)dx \iff \Phi(x) = \Phi(0) - \int_0^x \varepsilon_1 dx$$

Come sempre, consideriamo a 0 la massa del nostro sistema, dunque $\Phi(0) = 0$; si otterrà che:

$$\Phi(x) = -\varepsilon_1 x \implies \varepsilon_1 = -\frac{V_a}{d}$$

Tuttavia, parlando di condensatori, si introduce la capacità definita come il rapporto tra carica e tensione, ma anche come rapporto del dielettrico contenuto nel condensatore, ϵ_0 , e della distanza tra le facce:

$$C = \frac{Q}{V_a} = \frac{\epsilon_0}{d}$$

Possiamo dunque dire, invertendo l'espressione, che

$$V_a = \frac{Q}{C} = \frac{Q}{\epsilon_0 d}$$

Riprendendo il legame tra campo e potenziale, si arriva finalmente a dire che:

$$\varepsilon_1 = -\frac{Q}{\epsilon_0}$$

Abbiamo trovato in questo particolare caso un'espressione del campo ε_1 , esplicitato in funzione della carica, anche detto Legge di Gauss (in realtà essa racchiude significati ben più profondi di quello appena mostrato, ma che non sono utili alla trattazione). Questo è un caso molto particolare, che dovremo estendere per ottenere la più generale formulazione del legame tra una carica ed una differenza di potenziale, ossia l'Equazione di Poisson

2.4.2 Dimostrazione qualitativa dell'Equazione di Poisson

Ci preoccupiamo ora di mostrare l'equazione di Poisson, e di spiegarne il significato profondo, mediante una dimostrazione non rigorosa ma comunque efficace. L'equazione di Poisson avrà una forma del tipo:

$$\frac{\partial^2 \Phi}{\partial x^2} = -\frac{\rho(x)}{\epsilon_S}$$

Il termine ϵ_S è la costante dielettrica del semiconduttore: esso sarà il prodotto di ϵ_0 , ossia la costante dielettrica del vuoto, e di $\epsilon_{r,S}$, ossia la costante dielettrica relativa del semiconduttore. ε sarà il campo elettrico provocato dagli eccessi di carica descritti mediante la loro funzione di densità $\rho(x)$. Poichè sappiamo già che:

$$\varepsilon = -\frac{\partial \Phi}{\partial x}$$

Mediante l'applicazione dell'equazione di Poisson possiamo dire che:

$$\varepsilon(x) = \int \frac{\rho(x)}{\epsilon_S} dx$$

Dunque,

$$\varepsilon(x) = \frac{\rho(x)}{\epsilon_S} x$$

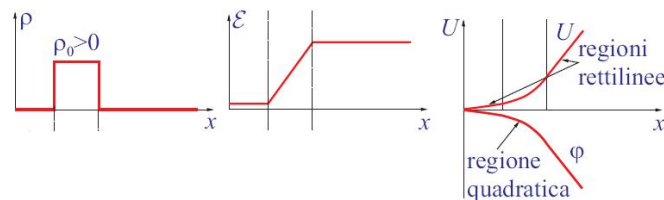
Il campo avrà una forma di questo tipo: da $-\infty$ a x_1 il campo è costantemente nullo; addentrandoci nella regione di cariche, il campo crescerà, fino all'uscita dalla regione, in x_2 . Il campo ora rimarrà costante, fino a quando non ci si addentrerà nella seconda regione di cariche, da x_3 , che ridurrà il

campo, fino ad annullarlo in x_4 . Andando da sinistra a destra, e considerando prima una densità positiva (per $x < 0$), poi una positiva (in $x > 0$).

Integrando poi $\varepsilon(x)$, si vedrà che:

$$\Phi(x) = - \int \varepsilon(x) dx$$

Considerando come massa $\Phi(x_1) = 0$. Vediamo dal precedente studio, sul campo, che esso è maggiore di 0 per qualsiasi posizione x si scelga nello spazio. L'integrale di tale funzione cambiata di segno sarà dunque negativo, ed avrà un andamento di questo tipo: per $x < x_1$, il potenziale sarà sempre 0, come fissato anche dalla massa. Poichè nella regione delle cariche, da x_1 a x_2 , abbiamo un campo ad andamento lineare, integrandolo ulteriormente esso diverrà quadratico, e quindi parabolico. Poichè, ribaltando l'andamento del campo, la retta avrà pendenza negativa, allora la concavità del tratto di parabola sarà rivolta verso il basso. A questo punto, da x_2 a x_3 , si trova un regime di campo costante, in cui il campo è positivo, ma quindi ribaltato negativo, e ciò darà vita ad un tratto di retta a pendenza negativa, dal punto x_2 al punto x_3 . Da x_3 a x_4 si ripeterà il ragionamento precedente, ma questa volta la distribuzione ribaltata del campo avrà pendenza positiva, poichè dovrà crescere fino a 0, quindi il tratto di parabola avrà concavità rivolta verso l'alto.



L'equazione di Poisson avrà un ruolo fondamentale nel modello matematico delle correnti in un semiconduttore, poichè sarà essa a fornirci le informazioni sul campo e sul potenziale, a partire dalla distribuzione di cariche fornita o studiata nel problema.

2.5 Considerazioni su semiconduttori fuori equilibrio

Prima di iniziare le nostre considerazioni sui semiconduttori fuori equilibrio, aggiorniamo la simbologia che utilizzeremo d'ora in poi: quando tratteremo quantità all'equilibrio, termodinamico, ottico o di qualunque altro tipo,

useremo il pedice aggiuntivo 0: esempi saranno n_{p0} , n_{n0} , n_{i0} , p_{n0} , p_{p0} , p_{i0} . Senza il pedice 0, i simboli finora utilizzati indicheranno una condizione di non equilibrio nel semiconduttore causata da un qualche fenomeno.

Ci chiediamo però: come possiamo creare, in un semiconduttore, una situazione di non-equilibrio?

Prendiamo un semiconduttore drogato tipo n, con un drogaggio N_D , ed esponiamolo ad un fascio di fotoni dotati di energia $\hbar\omega > E_g$, ossia dotati di un'energia superiore al gap di energia tra banda di valenza e banda di conduzione, in modo da facilitare le transizioni di elettroni.

Irradiamo una sola faccia del nostro semiconduttore con i fotoni $\hbar\omega$, in un punto che consideriamo come origine del sistema cartesiano; il fatto che la loro energia sia sufficiente a provocare molte transizioni di elettroni dalla banda di valenza a quella di conduzione, permette la generazione di diverse coppie elettrone-lacuna: esse sbilanceranno le distribuzioni dei portatori in prossimità del punto 0, $n_n(0)ep_n(0)$ le quali varieranno rispetto ai valori di equilibrio. Questo tipo di processo si può chiamare generazione termica (Gott), e lo consideriamo come un fenomeno puramente additivo: l'unico fenomeno ad essa contrapposto è la ricombinazione. Di fatto la generazione ottica inietta coppie elettrone-lacuna in un semiconduttore, ma esistono due sostanziali livelli di iniezione:

- Basso livello di iniezione: l'iniezione di coppie elettrone-lacuna altera i portatori minoritari, ma non è abbastanza elevata da poter variare sensibilmente i portatori maggioritari. Ad esempio, se $N_D = 10^{16}$, $n_{n0} = N_D$, $p_{n0} = 10^4$, i fotoni $\hbar\omega$ generano 10^{13} coppie elettrone-lacuna, $n_n \simeq n_{n0}$, poichè $10^{16} + 10^{13} \simeq 10^{16}$, ma $p_n \simeq G(\hbar\omega)$: infatti, $10^4 + 10^{13} \simeq 10^{13}$.
- Alto livello di iniezione: l'iniezione di coppie elettrone-lacuna è talmente elevata da riuscire ad alterare sensibilmente sia i minoritari che i maggioritari: supponiamo di avere una situazione come prima, ossia $N_D = 10^{16}$, $n_{n0} = N_D$, $p_{n0} = 10^4$, i fotoni $\hbar\omega$ generano però 10^{19} coppie elettrone-lacuna. A queste condizioni, $n_n \simeq p_n \simeq G(\hbar\omega)$: la generazione termica è così potente da compensare interamente il drogaggio N_D , e così annullarlo.

2.6 Modello del Tasso di Ricombinazione Netto

Cerchiamo ora di completare il nostro modello matematico delle correnti in un semiconduttore, introducendo un modello del tasso netto di ricombina-

zione: supponiamo, prima di iniziare con il formalismo, di avere un semiconduttore all'equilibrio, e dunque $n_n = n_{n0}$, e $p_n = p_{n0}$. In questa situazione, la generazione termica eguaglia la ricombinazione: per ogni generazione termica vi è la ricombinazione antagonista, che non permette il perdurare di un'eventuale variazione della situazione. Possiamo immaginare che questo sia un caso limite del nostro modello: vi dovrebbe essere un coefficiente di proporzionalità α che permetta di regolare la ricombinazione, in questo modo:

$$R = \alpha n_{n0} p_{n0}$$

In altre parole, aumentare uno dei due portatori, implicherebbe aumentare anche il fenomeno di ricombinazione nel nostro semiconduttore. Consideriamo questa proporzionalità come un assioma, e sviluppiamo il nostro modello; avevamo detto che, all'equilibrio termodinamico, $R = Gth$, e dunque:

$$R = \alpha n_{n0} p_{n0} = Gth$$

Ad un basso livello di iniezione, vi sarà una variazione dei minoritari, in questo ambito di p_n , ma non dei maggioritari: in altre parole, $n_n = n_{n0}$. Il livello di ricombinazione, da quello precedente, diventerà:

$$R = \alpha n_n p_n$$

Definendo U_n il tasso medio di ricombinazione degli elettroni, ma soprattutto U_p il tasso medio di ricombinazione delle lacune, minoritarie in questo ambito, vediamo che:

$$U_p = R - Gth = \alpha n_n p_n - \alpha n_{n0} p_{n0} = \alpha n_{n0} (p_n - p_{n0})$$

Ma poichè consideriamo sempre valida l'ipotesi di completa ionizzazione, possiamo tranquillamente dire che:

$$U_p = \alpha N_D (p_n - p_{n0})$$

Introduciamo dunque un'ulteriore definizione: il tempo di vita medio delle lacune, τ_p

$$\tau_p \triangleq \frac{1}{\alpha N_D}$$

τ_p è un tempo medio di vita delle lacune in una condizione a loro sfavorevole, detto in modo molto approssimativo: poichè ci troviamo in un semiconduttore drogato tipo n, il tempo di vita degli elettroni in quanto maggioritari sarà molto superiore a quello delle lacune, dunque di fatto la

loro vita media sarà inferiore, in quanto in proporzione saranno molto più colpite dalla ricombinazione degli elettroni (gli elettroni nati da ionizzazione infatti non hanno un antagonista). Possiamo dunque finalmente terminare il modello del tasso di ricombinazione, per le lacune, come:

$$U_p = \frac{p_n - p_{n0}}{\tau_p}$$

Cosa ci dice ciò: drogando un semiconduttore, a seconda di quanto più drogante vi immetteremo, faremo sempre più abbreviare la vita dei portatori minoritari, poichè aumenteremo la ricombinazione, che va principalmente a loro scapito. Considerando $p_n - p_{n0}$ un eccesso di lacune, si suol definire

$$p_n - p_{n0} = p'_n$$

Dove l'apice indica proprio fatto che si parla di un eccesso. In tali condizioni, il tasso di ricombinazione netto sarà:

$$U_p = \frac{p'_n}{\tau_p}$$

Chiaramente, parlare di p'_n e di n'_p ha senso solo fuori equilibrio, in seguito ad una perturbazione esterna, altrimenti la differenza tra il valore finale della concentrazione di minoritario e il valore all'equilibrio sarebbe banalmente nulla.

Le due relazioni che costituiranno il nostro modello di ricombinazione saranno dunque:

$$U_p = \frac{p_n - p_{n0}}{\tau_p}$$

$$U_n = \frac{n_p - n_{p0}}{\tau_n}$$

Si noti che questo è un modello di ricombinazione diretta, ossia basato su semiconduttori a gap diretto, e dove cioè le transizioni elettroniche da un livello energetico ad un altro sono di tipo diretto, e quindi non consideranti le interazioni fononiche. Esistono modelli che prevedono transizioni indirette, ma ciò non toglie l'efficacia di questo modello, nelle situazioni che lo riguardano.

2.7 Esercizio Pratico

1. Dato un campione di silicio, drogato con $N_A = 10^{16}$ atomi/cm³:

- (a) Disegnare diagramma a bande dettagliato all'equilibrio termodinamico quotando ogni dettaglio;
 - (b) Calcolare la resistenza sapendo che $l=1\text{mm}$; $A = 1 \text{ mm}^2$.
2. Successivamente, si droghi lo stesso campione, con $N_D = 10^{17} \text{ atomi/cm}^3$; a questo punto:
- (a) Disegnare diagramma a bande dettagliato all'equilibrio termodinamico quotando ogni dettaglio
 - (b) Calcolare la resistenza sapendo che $l=1\text{mm}$; $A = 1 \text{ mm}^2$.
3. Sapendo che il semiconduttore è illuminato con una radiazione luminosa generante 10^{13} coppie elettrone-lacuna sulla faccia posta in $x=0$, si calcoli come le lacune in eccesso si redistribuiscono nel semiconduttore con $\tau_p = 1 \text{ ms}$.
4. Supponendo di ripetere l'analisi del punto precedente, avendo applicato una tensione di polarizzazione $V_a = -1 \text{ V}$ al campione, se ne valuti l'effetto.

2.7.1 Risoluzione

Disegnare diagramma a bande dettagliato all'equilibrio termodinamico quotando ogni dettaglio

Cerchiamo di quotare il nostro diagramma a bande, e quindi di trovare tutti i dati che ci servono per avere un quadro completo della situazione. Siamo in un semiconduttore drogato tipo p, e quindi, poichè ci troviamo a 300 K, possiamo dire che la condizione di completa ionizzazione è applicabile, e non siamo in un regime intrinseco:

$$p_p(300) \simeq N_A$$

Calcoliamo dunque il livello di Fermi, partendo dall'equazione di Shockley:

$$p_p(300) \simeq N_A = n_i e^{-\frac{E_{F_i} - E_F}{kT}} \implies E_{F_i} - E_F = kT \ln \frac{N_A}{N_D}$$

Poichè come abbiam già detto, $kT \simeq 0,026\text{V}$, ed $N_A = 10^{16}$, facendo i conti troviamo che il livello di Fermi E_F vale 349 meV. Possiamo considerare noto il gap di energia E_g , e l'affinità elettronica $q\chi$.

Il lavoro di estrazione, dunque, varrà:

$$q\Phi_{S_p} = q\chi + \frac{E_g}{2} + E_{F_i} - E_F = 4,05 + 0,56 + 0,349 = 4,959 \text{ eV}$$

Calcolare la resistenza sapendo che $l=1\text{mm}$; $A = 1 \text{ mm}^2$

Abbiamo quotato tutto ciò che siamo in grado di fare, per questo possiamo passare al calcolo della resistenza del semiconduttore. Partendo dalla legge di Ohm, possiamo dire che:

$$R = \rho \frac{l}{A}$$

Disponiamo partendo dai dati del problema di l ed A , ma non abbiamo alcuna informazione riguardante ρ . L'unica cosa che sappiamo, è che:

$$\rho = \frac{1}{\sigma}$$

Abbiamo ampiamente studiato σ in precedenza, e siamo arrivati a dire che:

$$\sigma = q(\mu_n n + \mu_p p)$$

Noi disponiamo di p , ma ci mancano sia i minoritari n , che le mobilità relative ad n e p . Per queste ultime, è necessario consultare un grafico, contenente la degenerazione delle mobilità al variare della temperatura e della concentrazione di portatori $N_A + N_D$, e quindi possiamo dire che:

$$\mu_n = 1050; \mu_p = 350$$

Per quanto riguarda invece i minoritari, è sufficiente applicare la legge dell'azione di massa:

$$n_p \cdot p_p = n_i^2 \implies n_p = \frac{n_i^2}{p_p} = 2,1 \cdot 10^4$$

Il fatto che il numero di minoritari sia inferiore a quello di maggioritari di molti ordini di grandezza, ci permette di trascurarli, calcolando la conducibilità elettrica. Possiamo dunque dire che:

$$\sigma \simeq q\mu_p p_p = 0,746$$

Invertendo ora σ , troviamo ρ , e possiamo quindi calcolare R :

$$R = \frac{1}{\sigma} \frac{l}{A}$$

Abbiamo così risolto facilmente i primi due punti del problema; risolvere i secondi è del tutto inutile, in quanto il processo è del tutto analogo, tenendo semplicemente conto del fatto che vi è una compensazione del drogante tipo p, N_A , in seguito all'introduzione di N_D . Ricalcolando la nuova concentrazione di minoritari mediante la legge dell'azione di massa, e le nuove mobilità dei portatori, si può facilmente raggiungere lo stesso risultato appena trovato (chiaramente le cifre saranno diverse).

Calcolo lacune in eccesso

Trattiamo il terzo punto del problema, che sarà molto più interessante dei primi due. Talvolta, converrà fare divagazioni, che torneranno molto utili nello studio di dispositivi in un momento più avanzato della trattazione. Supponiamo dunque di avere 10^{13} coppie elettrone-lacuna, e dunque di avere un basso livello di iniezione. Ciò che capiterà, è che:

$$\begin{cases} p_{n0} \ll 10^{13} \\ n_{n0} \gg 10^{13} \end{cases}$$

Applichiamo per la prima volta il modello matematico, partendo dall'equazione di continuità:

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

Consideriamo il fatto che la generazione ottica non subisce variazioni nel tempo, e che dunque, al variare del tempo t , il numero di coppie sarà sempre lo stesso, e quindi la derivata nel tempo sarà nulla.

$$\implies -\frac{1}{q} \frac{\partial J_p}{\partial x} - U_p = 0$$

Introducendo il nostro modello di ricombinazione diretta,

$$-\frac{1}{q} \frac{\partial J_p}{\partial x} - \frac{p_n - p_{n0}}{\tau_p} = 0$$

Consideriamo l'equazione di trasporto della corrente, rappresentante la densità di corrente J_p , e quindi

$$J_p = q\mu_p p_n \varepsilon - qD_p \frac{\partial p_n}{\partial x}$$

Consideriamo però che il nostro semiconduttore è isolato rispetto all'esterno, e dunque non vi sono fenomeni di polarizzazione, e il campo elettrico dall'esterno risulterà essere nullo, poichè non abbiamo polarizzazioni esterne.

Potremmo pensare però ad eccessi di carica interni generati dall'iniezione ottica: sarebbe un pensiero fuorviante, poichè la generazione ottica provoca la nascita di coppie elettrone-lacuna, e quindi non viene perturbato l'equilibrio in questo senso, dunque possiamo considerare $\rho = 0$, ed $\varepsilon = \varepsilon_{ext} + \varepsilon_{int} = 0$. Il vero problema che potrebbe nascere è il seguente: la diffusione di elettroni e lacune avviene con una velocità differente, poichè la differenza delle masse efficaci di elettroni e lacune fa sì che queste seconde abbiano una mobilità inferiore, ergo potrebbero generarsi campi elettrici momentanei, in seguito a gradienti di diffusione temporanei. Al fine di prevenire questa possibilità, consideriamo un'ipotesi aggiuntiva: la quasi neutralità del semiconduttore: $\varepsilon \simeq 0$: infatti il campo elettrico effettivamente potrebbe agire, ma solo sulla corrente dei minoritari, che quindi non ci riguarda, in quanto del tutto trascurabile. Accettando tale ipotesi, l'equazione di trasporto si ridurrà semplicemente a:

$$J_p \simeq -qD_p \frac{\partial p_n}{\partial x}$$

Sostituendo l'equazione di trasporto in quella di continuità, e considerando solo gli eccessi di cariche dovuti alla ricombinazione, si otterrà:

$$0 = -\frac{1}{q}D_p \frac{\partial^2 p'_n}{\partial x^2} - \frac{p'_n}{\tau_p}$$

Questa di fatto rappresenta un'equazione omogenea di secondo ordine a coefficienti costanti:

$$\frac{\partial^2 p'_n}{\partial x^2} = \frac{1}{D_p \tau_p} p'_n$$

Considerando il polinomio caratteristico di quest'equazione, e riducendola ad un'equazione di secondo grado, si ottiene che gli autovalori del polinomio caratteristico avranno forma:

$$\lambda^2 = \frac{1}{D_p \tau_p} \implies \lambda_{1,2} = \pm \frac{1}{\sqrt{D_p \tau_p}}$$

Potremmo ora fare quest'osservazione: il termine $D_p \tau_p$ è una lunghezza quadratica, dimensionalmente parlando. Si definisce il parametro detto lunghezza di diffusione, relativo alle lacune, come:

$$L_p = \sqrt{D_p \tau_p}$$

Di fatto la lunghezza di diffusione è un parametro che ci permette di capire quanto velocemente i minoritari iniettati si disperdono, man mano

che ci si allontana dalla zona di iniezione: se la posizione che si analizza è lontana alcune lunghezze di iniezione, si troveranno pochissimi minoritari. Chiaramente, vale una relazione del tutto duale, per gli elettroni iniettati in un semiconduttore drogato p, ossia:

$$L_n = \sqrt{D_n \tau_n}$$

Tornando al nostro problema, come sappiamo dalla teoria della risoluzione delle equazioni differenziali a coefficienti costanti, le soluzioni dell'equazione differenziale prima esposta avrà forma del tipo:

$$p'_n(x) = Ae^{\lambda_1 x} + Be^{\lambda_2 x}$$

Nella fattispecie, la soluzione sarà:

$$p'_n(x) = Ae^{\frac{x}{L_p}} + Be^{-\frac{x}{L_p}}$$

Dobbiamo ora trovare le costanti di normalizzazione A e B: per far ciò serviranno due condizioni al contorno del problema:

1. Sappiamo che, lontano dalla zona di iniezione, le lacune si saranno interamente ricombinate, e quindi, considerando di trovarci in una zona molto lontana dalla zona di iniezione, possiamo dire che la ricombinazione avrà annullato tutti i minoritari, e quindi

$$p'_n(\infty) = 0 \implies A = 0$$

2. Abbiamo una condizione al contorno sulla posizione iniziale: noi sappiamo infatti esattamente quante coppie elettrone-lacuna vengono iniettate otticamente, e quindi possiamo dire che, in quella che consideriamo l'origine del sistema cartesiano, e l'origine dell'iniezione delle lacune,

$$p'_n(0) = 10^{13} \implies B = 10^{13}$$

Siamo arrivati, analizzando il nostro semiconduttore in condizione di lato lungo (che vedremo non essere sempre valida), ad esprimere una relazione in grado di descrivere l'andamento degli eccessi di portatori minoritari iniettati in un semiconduttore drogato. Ci chiediamo a questo punto, come studiare l'andamento delle correnti, data la funzione appena trovata. Ci chiediamo innanzitutto, se esistono correnti nel semiconduttore, a queste condizioni; la corrente totale infatti, indipendentemente dall'andamento della distribuzione appena espressa, dovrebbe essere nulla: se così non fosse, significherebbe che

vi sono accumuli nel tempo, che violerebbero la legge di conservazione della carica. Dato che trattiamo un semiconduttore lungo,

$$p'_n(x) = p'_n(0)e^{-\frac{x}{L_p}}$$

Quindi, possiamo ricavare la corrente derivando questa distribuzione; infatti:

$$J_{p,diff}(x) = -qD_p \frac{\partial p'_n}{\partial x} = \frac{qp'_n(0)D_p}{L_p} e^{-\frac{x}{L_p}}$$

La variazione di concentrazione dovrebbe provocare una corrente di questo tipo:



Ciò è in accordo con quanto abbiamo appena detto: allontanandoci dal punto di iniezione, abbiamo un comportamento soddisfacente in quanto la ricombinazione elimina tutte le correnti di minoritari, ma in un intorno del punto di iniezione la corrente è diversa da 0. Si può intuire che vi sia un flusso di elettroni che controbilancia la corrente in questo intorno di punti, in modo da garantire comunque la conservazione della carica. Questo problema sarà molto importante da studiare in seguito, trattando le correnti nella giunzione p-n, che ora inizieremo a studiare.

2.7.2 Ipotesi di lato corto

Permettiamoci una digressione sul problema appena studiato: esistono condizioni (molto comuni, parlando ad esempio di microelettronica o nanotecnologie), in cui un dispositivo elettronico ha una lunghezza confrontabile o addirittura inferiore alla lunghezza di diffusione, e quindi in cui il meccanismo di ricombinazione del semiconduttore di fatto non riesce, nonostante l'andamento esponenziale, a ridurre a zero il numero di lacune. Le condizioni al contorno che abbiamo utilizzato, non sono dunque più valide, in quanto

non è possibile assimilare la lunghezza del lato a ∞ . Chiaramente, i passaggi iniziali saranno del tutto analoghi, e la soluzione dell'equazione differenziale cui arriveremo sarà sempre della forma:

$$p'_n(x) = Ae^{\frac{x}{L_p}} + Be^{-\frac{x}{L_p}}$$

Al fine di poter ricavare nuove condizioni al contorno, sarà necessario proporre nuove ipotesi assuntive: il semiconduttore, per quanto corto, dovrà essere per esempio essere chiuso su di un metallo, mediante due placchette, al fine di poter ottenere un contatto ohmico. Costruendo tecnologicamente un dispositivo di questo tipo, la condizione al contorno che potremo applicare, sarà la totale ricombinazione nel punto estremo del semiconduttore, d : infatti, il contatto ohmico permette una totale ricombinazione con i minoritari non ancora ricombinati, all'interno del semiconduttore. Potremo dunque considerare, come condizioni al contorno,

$$p'_n(d) = 0$$

$$p'_n(0) = A + B$$

Partendo da ciò, troviamo che:

$$B = p'_n(0) - A$$

$$0 = Ae^{\frac{d}{L_p}} + (p'_n(0) - A)e^{-\frac{d}{L_p}}$$

Svolgiamo ora alcuni passaggi algebrici, al fine di esplicitare in questo ambito le costanti di normalizzazione A e B:

$$0 = A(e^{\frac{d}{L_p}} - e^{-\frac{d}{L_p}}) + p'_n(0)e^{-\frac{d}{L_p}}$$

Ma notiamo che $e^{\frac{d}{L_p}} - e^{-\frac{d}{L_p}} = 2 \sinh(\frac{d}{L_p})$; quindi:

$$A = \frac{-p'_n(0)e^{-\frac{d}{L_p}}}{2 \sinh(\frac{d}{L_p})}$$

Sostituiamo ora in B:

$$B = p'_n(0) + \frac{p'_n(0)e^{-\frac{d}{L_p}}}{2 \sinh(\frac{d}{L_p})} = 2 \sinh(\frac{d}{L_p})p'_n(0) + p'_n(0)e^{-\frac{d}{L_p}} =$$

$$= \frac{p'_n(0)e^{\frac{d}{L_p}} - p'_n(0)e^{-\frac{d}{L_p}} + p'_n(0)e^{-\frac{d}{L_p}}}{2 \sinh(\frac{d}{L_p})}$$

Quindi,

$$B = \frac{p'_n(0)e^{\frac{d}{L_p}}}{2 \sinh(\frac{d}{L_p})}$$

Sostituendo A e B nell'equazione di partenza, si ottiene che:

$$p'_n(x) = \frac{-p'_n(0)e^{-\frac{d}{L_p}}e^{\frac{x}{L_p}} + p'_n(0)e^{\frac{d}{L_p}}e^{-\frac{x}{L_p}}}{2 \sinh(\frac{d}{L_p})} = \frac{p'_n(0) \left(e^{\frac{d-x}{L_p}} - e^{-\frac{d-x}{L_p}} \right)}{2 \sinh(\frac{d}{L_p})} = \frac{\sinh(\frac{d-x}{L_p})}{\sinh(\frac{d}{L_p})}$$

Siamo dunque arrivati a dire che, dato un semiconduttore il cui lato è confrontabile con la lunghezza di diffusione, ossia $d \sim L_p$, la distribuzione dell'eccesso di minoritari (nel nostro caso, con un semiconduttore drogato tipo n, $p'_n(x)$,

$$p'_n(x) = \frac{\sinh(\frac{d-x}{L_p})}{\sinh(\frac{d}{L_p})}$$

Esiste un caso ancora più estremo: se il lato del semiconduttore è inferiore alla lunghezza di diffusione, e quindi $d \ll L_p$, possiamo dire che $d \ll x$, e quindi possiamo riutilizzare la precedente espressione, considerandola per un x tendente a 0. Ciò però significa che siamo in grado di semplificare notevolmente la nostra espressione, sviluppandola in polinomio di Taylor:

$$p'_n(x) \sim \frac{p'_n(0) \cdot \frac{d-x}{L_p}}{\frac{d}{L_p}} = p'_n(0) \left(1 - \frac{x}{d} \right)$$

Utilizzando gli andamenti degli eccessi di minoritari ora ricavati, sarà possibile calcolare le correnti facilmente, anche con ipotesi diverse da quella di lato lungo, sfruttando sempre le stesse conoscenze utilizzate precedentemente.

Capitolo 3

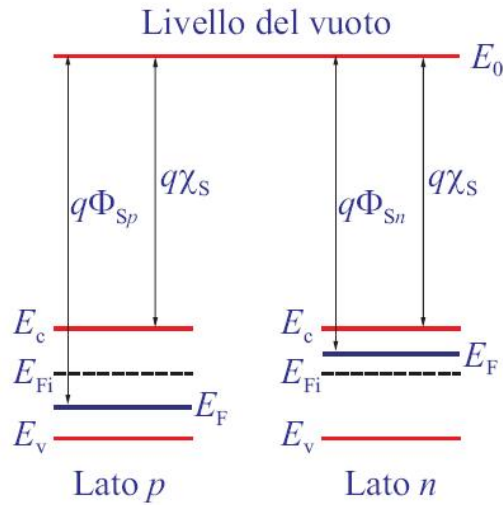
Giunzione p-n

Siamo riusciti a fornire un modello soddisfacente dell'andamento delle correnti in un semiconduttore drogato, e abbiamo fatto un esempio applicativo di come calcolarle. Ora vogliamo sfruttare queste conoscenze, in un dispositivo usato molto comunemente, e alla base dell'elettronica: la giunzione p-n.

Dati due semiconduttori (comunemente noi useremo il silicio), drogati uno tipo n, ed uno tipo p, essi avranno un diagramma a bande abbastanza diverso sotto alcuni punti di vista, anche se con le conoscenze finora apprese siamo in grado di motivarlo senza problemi.

Immaginiamo didatticamente di unire i due semiconduttori; i potenziali dei due, in principio separati, una volta uniti, avranno lo stesso sistema di riferimento. Consideriamo come origine dei tempi l'istante in cui si uniscono i due semiconduttori; in un tempo 0^+ , le due barriere di potenziale del vuoto, le E_B , saranno coincidenti. Anche gli altri livelli, quali E_C , E_V , E_{F_i} , saranno tutti coincidenti, tranne il livello di Fermi reale dei semiconduttori, E_F , che presenterà una discontinuità in prossimità del punto di giunzione. Volessimo dunque tracciare un diagramma a bande del semiconduttore in un istante 0^+ , esso avrebbe l'andamento appena descritto.

Consideriamo, in questo sistema, come origine degli assi, il punto di giunzione dei due semiconduttori. Convenzionalmente, si considera a sinistra il lato p, a destra il lato n. Possiamo dunque dire che, a destra della giunzione, avremo molti elettroni in banda di conduzione (maggioritari) e poche lacune in banda di valenza (minoritari), e viceversa a sinistra i portatori maggioritari saranno le lacune in banda di valenza, ed i minoritari gli elettroni in banda di conduzione. Di fatto, vi saranno accumuli di carica, che provocheranno flussi di portatori: dal lato p al lato n di lacune, dal lato n al lato p, di elettroni (sempre parlando di un istante $t = 0^+$). Questi flussi di cariche sussisteranno, fino a quando sussisteranno le discontinuità nei livelli energetici, e nella fatiscie la discontinuità del livello di Fermi. Per un certo transitorio



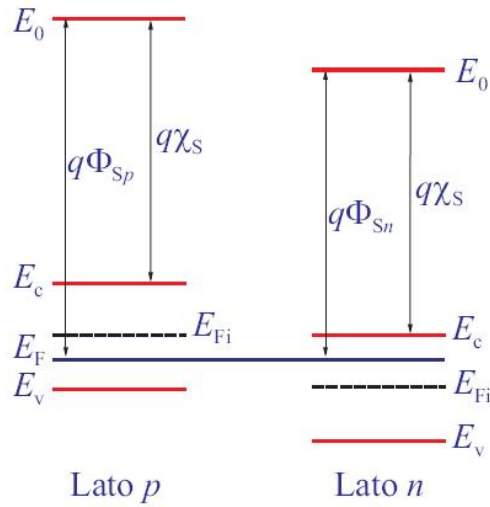
di tempo, dunque, a partire dall'istante 0, si avranno due flussi netti, ossia due flussi che prevarranno su tutti gli altri:

1. $\Phi_{x,e^- BC} \quad n \longrightarrow p$
2. $\Phi_{x,e^+ BV} \quad p \longrightarrow n$

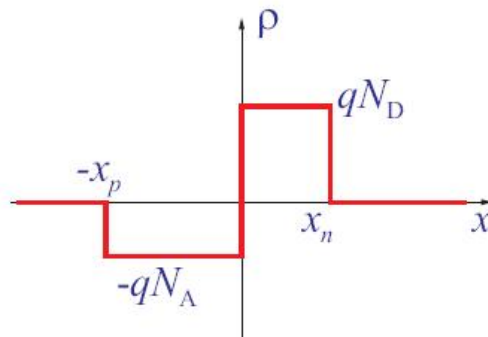
Vogliamo cercare però di capire come può instaurarsi, a partire da questo stato transitorio, una condizione di equilibrio. Possiamo dire, partendo dalle informazioni sui due flussi, che vi siano due canali di scambio di portatori, separati: generazione termica e ricombinazione rispetto a questi scambi saranno fenomeni molto lenti e poco efficaci, e quindi trascurabili in un primo tempo. Se per un certo intervallo di tempo si ha uno spostamento di portatori di carica, capiterà che, venendo a mancare elettroni di banda di conduzione nel lato n, e lacune in banda di valenza nel lato p, si perderà la condizione di neutralità locale: in banda di conduzione il semiconduttore p si sarà spopolato di lacune nelle regioni prossime al contatto con il lato n, e viceversa il semiconduttore lato n si sarà spopolato di elettroni nelle regioni prossime al contatto con il lato p.

Gli ioni positivi nel lato n dunque non avranno, in prossimità della giunzione, un elettrone in grado di neutralizzarli localmente, e idem gli ioni negativi nel lato p, non avranno lacune in grado di neutralizzarli localmente.

Per lo studio della giunzione dovremo attenerci ad un modello leggermente semplificato: considereremo, come ipotesi, che in prossimità della giunzione avvenga un completo svuotamento: le zone svuotate per ipotesi non conterranno portatori maggioritari dei rispettivi lati (o comunque una quantità del



tutto trascurabile rispetto al numero di ioni). Gli unici elementi di carica che considereremo in gioco, saranno gli ioni dei livelli energetici introdotti mediante il drogaggio. Sarà dunque semplice rappresentare la distribuzione spaziale della carica della giunzione:



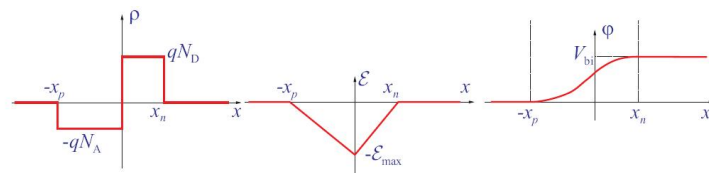
Fino ad un certo punto $-x_p$, la distribuzione di ioni che non dispongono di un maggioritario che lo neutralizza localmente sarà del tutto nulla, per poi avere un picco nel suddetto punto. La distribuzione di carica, negativa, resterà costante fino all'origine, dove vi sarà un ulteriore salto: vi sarà una distribuzione di carica positiva dal punto 0 al punto x_n , anch'essa costante. Inoltre, considerando l'ipotesi di completo svuotamento in coincidenza con l'ipotesi di completa ionizzazione, potremo dire che la densità di carica coinciderà con la concentrazione di drogante introdotta nel substrato semiconduttivo, quindi potremo dire che l'altezza del rettangolo di sinistra sarà $-N_A$, e l'altezza del rettangolo destro sarà N_D .

Potremmo ora chiederci alcune domande: ma qual è la posizione del salto? Coincide veramente con l'origine degli assi, è cioè in prossimità della giunzione? Dove le distribuzioni di cariche iniziano ad annullarsi? La risposta è la seguente: dove il livello di Fermi cessa di essere discontinuo.

Vogliamo capire e quantificare l'estensione delle regioni di carica, studiarne la geometria, quantificare i punti $-x_p$ ed x_n , e soprattutto i loro valori una volta raggiunto l'equilibrio, $-x_{p0}$ ed x_{n0} . Questi valori saranno ottenuti in seguito ad una crescita temporale dei valori fuori equilibrio $-x_p$ ed x_n , rispetto al loro stato iniziale. Nel punto di equilibrio, i flussi determinati dai gradienti di concentrazione di carica saranno eguagliati dalla ricombinazione e da flussi opposti, e quindi saranno nulli.

3.1 Studio qualitativo del diagramma a bande della giunzione p-n

Facendo una prima analisi qualitativa del campo (senza considerare calcoli analitici di alcun tipo, che verranno ripresi in seguito), possiamo vedere che poichè a sinistra la carica è negativa, si avrà, integrando, una retta decrescente da $-x_p$ fino a 0, e crescente da 0 fino a x_n . Il campo sarà dunque un triangolo, interamente negativo, con il vertice posto in 0. Integrando ulteriormente, si ottiene un insieme di due contributi parabolici, il primo con concavità verso l'alto, il secondo con concavità verso il basso (poichè si inverte il segno del campo per l'equazione di Poisson).



Ribaltando il grafico del potenziale rispetto all'asse x , si troverà una funzione monotona decrescente, trattabile come due parabole, la prima fino a 0 con concavità verso il basso, la seconda fino a x_n con concavità verso l'alto. Questo sarà il grafico dell'energia potenziale valutata in eV, ma sarà soprattutto utile se pensata come parte di un diagramma a bande: il grafico dell'energia potenziale, come già accennato, è il diario di bordo dell'elettrone, ossia è il suo diagramma a bande. Ciò che si trova con questa sorta di doppio andamento parabolico, è semplicemente il raccordo tra i livelli energetici dei due lati, una volta raggiunto l'equilibrio: infatti all'equilibrio il livello di Fermi reale del sistema diventa continuo, confrontabile con una retta, mentre i vari E_B , E_C , E_V , E_{F_i} , raggiungono livelli energetici diversi, ma non

presentano discontinuità, bensì un raccordo come quello appena descritto qualitativamente (e che in seguito verrà calcolato analiticamente). Possiamo dire che il lato n tenderà a scendere rispetto al lato p, che salirà relativamente ad n. Lo spostamento di carica, determinante l'equilibrio, ha determinato il curvarsi dei livelli energetici, nell'intorno della zona della giunzione. Il livello di Fermi si sarà riallineato, e tutti gli altri curvati con un andamento come quello appena descritto.

Notiamo che si viene a formare una barriera di potenziale, man mano che si instaura l'equilibrio: più passa il tempo, più portatori tenderanno a transire la barriera, e più diventerà difficile raggiungere l'altro lato, poichè il campo continuerà ad aumentare. Al livello di equilibrio, i flussi sono del tutto bilanciati: il flusso di elettroni in banda di conduzione dal lato n a quello p equivale e bilancia quello di elettroni in banda di conduzione dal lato p al lato n. Stesso discorso per le lacune. Possiamo esprimere con un formalismo migliore, il fatto che si hanno in totale quattro flussi:

1. $\Phi_{x,e^-BC} \quad n \longrightarrow p$
2. $\Phi_{x,e^-BC} \quad p \longrightarrow n$
3. $\Phi_{x,e^+BV} \quad p \longrightarrow n$
4. $\Phi_{x,e^+BV} \quad n \longrightarrow p$

$$\Phi_{x,e^-BC} \quad n \longrightarrow p = \Phi_{x,e^-BC} \quad p \longrightarrow n$$

$$\Phi_{x,e^+BV} \quad p \longrightarrow n = \Phi_{x,e^+BV} \quad n \longrightarrow p$$

Quindi, si può dire che:

$$\begin{cases} 1 = 2 \\ 3 = 4 \end{cases}$$

Le barriere di potenziale causate dalla transizione di elettroni, e quindi dalla non-neutralizzazione degli atomi droganti ionizzati, si modulano fino a bilanciare questi flussi. Ciò coincide esattamente con il bilanciare il livello di Fermi. Questo fatto finora espresso facilmente però non è banale, e vediamo come sia realmente possibile, mediante una semplice dimostrazione:

Dati due materiali, 1 e 2, essi saranno caratterizzati da una propria distribuzione degli stati energetici, e da una funzione di occupazione degli stati (in questo caso poichè si parla sempre di elettroni/lacune e quindi di fermioni,

la Fermi-Dirac), caratterizzata proprio dal livello di Fermi del materiale. I due materiali saranno dunque così caratterizzati:

$$\begin{cases} g_1(E) \\ f_1(E; T) \Leftrightarrow E_{F1} \end{cases}$$

$$\begin{cases} g_2(E) \\ f_2(E; T) \Leftrightarrow E_{F2} \end{cases}$$

Supponiamo ora che sia già stato raggiunto l'equilibrio termodinamico, in seguito al contatto. Ciò significa che per ogni generico livello energetico E , bisogna garantire un egual flusso di portatori, e cioè i flussi di lacune e di elettroni da un lato ad un altro devono essere tra loro bilanciati. Consideriamo anche soltanto gli elettroni: è sufficiente come condizione per la nostra dimostrazione: la nostra ipotesi dunque sarà:

$$\Phi_{x,e-1} \longrightarrow 2 = \Phi_{x,e-2} \longrightarrow 1$$

Quindi, esplicitando le funzioni caratterizzanti i flussi di portatori:

$$\Phi_{x,e-1} \longrightarrow 2 = [g_1(E) \cdot f_1(E; T)] \cdot [g_2(E) \cdot (1 - f_2(E; T))]$$

$$\Phi_{x,e-2} \longrightarrow 1 = [g_2(E) \cdot f_2(E; T)] \cdot [g_1(E) \cdot (1 - f_1(E; T))]$$

Ossia, perchè i flussi di carica siano efficaci, deve essere presente un elettrone dal lato 1 in grado di effettuare la transizione, e vi deve essere uno spazio libero (regolato dal complementare della funzione di occupazione) nel lato 2; viceversa per il flusso da 2 a 1. Svolgendo i conti, si ottiene che:

$$g_1 f_1 g_2 - g_1 f_1 f_2 g_2 = g_2 f_2 g_1 - g_1 f_1 f_2 g_2 \implies f_1 = f_2$$

Poichè supponiamo di trovarci ad una temperatura di 300 K come nostro solito, e consideriamo un qualunque livello energetico E , l'unica variabile in gioco fissate queste rimane il livello di Fermi: infatti, dire che le due funzioni di occupazione sono uguali, implica dire che:

$$\frac{1}{1 + e^{\frac{E - E_{F1}}{kT}}} = \frac{1}{1 + e^{\frac{E - E_{F2}}{kT}}} \implies E_{F1} = E_{F2}$$

Questo, fissata una E qualsiasi.

Tentiamo ora di meglio definire ciò che capita, stabilendo alcune osservazioni in grado di guidarci nel rappresentare un diagramma a bande (per ora qualitativo) di una giunzione p-n. In seguito effettueremo calcoli rigorosi per

il calcolo dei campi dovuti agli atomi ionizzati, e per la quantificazione della barriera di potenziale.

Se tentassimo di costruire il diagramma a bande della giunzione, con le nostre attuali competenze, dovremmo tener conto di queste osservazioni:

1. In una condizione di equilibrio, $E_F(x)$ è una funzione costante, e dunque $E_F(x) = E_F$:
2. Lontano dalla giunzione, i materiali si comportano come se la giunzione non esistesse, poichè la ricombinazione elimina tutti i portatori iniettati dall'altra parte. Si ha dunque una neutralità globale, ma anche locale;
3. I livelli energetici sono quantità tutte continue, e non presentano dunque salti di alcun tipo;
4. L'affinità elettronica è una caratteristica intrinseca del materiale, non è modificabile, e quindi si può sempre considerare invariata;
5. Il gap di energia tra livello E_C e livello E_V è invariante, in una giunzione p-n.

Proprietà extra, abbastanza scontata, è la seguente: se il semiconduttore dispone di una carica positiva, il diagramma a banda in quella zona sarà convesso; se ha carica negativa, sarà concavo.

Mediante queste cinque proprietà, si disporrà di un metodo per disegnare il diagramma a bande, in modo qualitativo.

Possiamo ora pensare ad un'idea: potremmo applicare sul livello di Fermi intrinseco, anch'esso variante a causa della giunzione, e quindi funzione della posizione, $E_{F_i}(x)$, e su E_F , le equazioni di Shockley; considerando un punto molto distante dalla giunzione come $-\infty$ nel lato p, e $+\infty$ molto distante nel lato n; nel lato n, consideriamo dunque gli andamenti dei portatori in punti distanti:

$$n_n(+\infty) = n_i e^{\frac{E_F - E_{F_i}(+\infty)}{kT}}$$

$$P_p(-\infty) = n_i e^{\frac{E_{F_i}(-\infty) - E_F}{kT}}$$

Le equazioni di Shockley sono utili in quanto, mediante la loro applicazione, si potrebbe cercare immediatamente il numero di elettroni e lacune in funzione della posizione, in ogni punto della giunzione.

Data ora questa prima impostazione qualitativa dello studio della giunzione, vorremmo passare ad un'analisi più quantitativa, e quindi poter determinare un'espressione che ci permetta di calcolare alcuni parametri: $-x_{p0}$, x_{n0} , e la barriera di potenziale Φ_i .

3.2 Studio elettrostatico della giunzione p-n

Iniziamo ora uno studio dettagliato e quantitativo della giunzione p-n, in modo da poter determinare i valori dei parametri mancanti, e poterne quotare ogni grandezza del diagramma a bande.

Le tre incognite sono le seguenti:

- $-x_{p0}$: ampiezza zona di svuotamento del lato p;
- x_{n0} : ampiezza zona di svuotamento del lato n;
- Φ_i : potenziale interno tra il lato p ed il lato n

Poichè abbiamo tre incognite, avremo bisogno di tre condizioni al contorno, e quindi di tre equazioni tra esse indipendenti, in modo da poter avere un problema ben posto. Le tre idee che ora affronteremo al fine di poter ultimare un modello quantitativo elettrostatico della giunzione, sono le seguenti:

1. L'applicazione della neutralità globale della giunzione p-n;
2. L'analisi della barriera come differenza tra i due lavori di estrazione dei lati;
3. L'analisi formale dell'andamento di campo e potenziale della giunzione in ogni zona.

3.2.1 Equazione di neutralità globale

In forma generale, l'equazione di neutralità globale ha forma:

$$-\int_{-x_p}^0 \rho(x)dx = \int_0^{x_n} \rho(x)dx$$

Nel nostro caso specifico, date le ipotesi di semiconduttore omogeneo, di completa ionizzazione dei droganti e di completo svuotamento, l'equazione si ridurrà semplicemente a:

$$qN_A x_{p0} = qN_D x_{n0}$$

Ci mancano ora altre due equazioni. Lavoriamo in tale senso, al fine di poter completare il sistema.

3.2.2 Analisi della barriera di potenziale

Data la barriera di potenziale Φ_i , possiamo pensarla, come già anticipato, come differenza dei lavori di estrazione del lato p, e del lato n:

$$\Phi_i = q\Phi_{S_p} - q\Phi_{S_n}$$

Questa differenza si può pensare in termini della semplice differenza dei soli livelli di Fermi intrinseci, in punti distanti dalla giunzione:

$$\Phi_i = E_{F_i}(-\infty) - E_{F_i}(+\infty)$$

Algebricamente, possiamo aggiungere e togliere lo stesso numero ad un'equazione ed essa rimarrà comunque invariata. Possiamo dunque aggiungere e sottrarre la quantità E_F :

$$\Phi_i = E_{F_i}(-\infty) - E_F + E_F - E_{F_i}(+\infty)$$

Ma ora è possibile ricondurre alle equazioni di Shockley queste espressioni:

$$\begin{aligned} (E_{F_i}(-\infty) - E_F) + (E_F - E_{F_i}(+\infty)) &= kT \ln \frac{N_A}{n_i} + kT \ln \frac{N_D}{n_i} = \\ &= kT \ln \frac{N_A N_D}{n_i^2} = q\Phi_i \end{aligned}$$

La barriera è l'energia positiva così definita: $q\Phi_i$. Si noti che la convenzione è particolare: parlando quantitativamente di potenziale, avevamo preso come riferimento, ossia come massa, il punto $-x_{p0}$; ora, parlando di energia, consideriamo come punto di riferimento (ossia come una sorta di massa, parlando però di energia sarebbe più corretto parlare di una sorta di energia iniziale) il punto x_{n0} : per questo, i segni risultano comunque essere positivi, anche se in effetti il potenziale espresso in V è semplicemente l'opposto dell'energia potenziale espressa in eV. Riassumendo dunque la convenzione:

- $q\Phi_i = U(-\infty) - U(+\infty)$
- $\Phi_i = \Phi(+\infty) - \Phi(-\infty)$

Ora, per terminare la formalizzazione, potremmo considerare la definizione corretta di potenziale di contatto della giunzione p-n, non utilizzando più la convenzione a $\pm\infty$, ma utilizzando al posto di $-\infty$ il valore $-x_{p0}$, e al posto di $+\infty$ il valore x_{n0} :

$$\Phi_i = \Phi(x_{n0}) - \Phi(-x_{p0})$$

Abbiamo così trattato anche la seconda condizione del nostro sistema. Ci manca ancora una condizione, ossia l'applicazione dell'equazione di Poisson.

3.2.3 Analisi formale dell'andamento di campo e potenziale

Partiamo dall'equazione di Poisson, e consideriamo quindi le due equazioni:

$$\varepsilon(x) = \int \frac{\rho(x)}{\epsilon_S} dx$$

$$\Phi(x) = - \int \varepsilon(x) dx$$

Partendo dalla densità di carica $\rho(x)$ studiamo l'andamento del campo elettrico. Dobbiamo però definire innanzitutto la funzione di densità di carica $\rho(x)$ come una funzione a tratti:

$$\rho(x) = \begin{cases} 0 & -\infty < x < -x_{p0} \\ -qN_A & -x_{p0} < x < 0 \\ +qN_D & 0 < x < x_{n0} \\ 0 & x_{n0} < x < +\infty \end{cases}$$

Considerando che abbiamo a che fare con una funzione definita a tratti, possiamo calcolare il campo elettrico integrando ogni singolo intervallo: basterà sfruttare l'equazione di Poisson ed il teorema fondamentale del calcolo integrale, nel seguente modo: considerando l'integrazione in un intervallo $[a; x]$,

$$\varepsilon(x) = \varepsilon(a) + \int_a^x \frac{\rho(x)}{\epsilon_S}$$

Si consideri inoltre che il campo deve ovviamente essere continuo, e che le espressioni trovate devono essere tutte riferite alla stessa massa. Per questo, sarà necessario calcolare il valore finale del campo in un intervallo, e porlo uguale al valore iniziale nell'intervallo successivo. Tenendo conto di queste cose, studiamo ora ciascuna delle quattro zone:

- Per $-\infty < x < -x_{p0}$, $\rho(x) = 0$, quindi il campo sarà nullo nella regione; si noti che il valore finale di quest'espressione sarà 0, poichè la funzione integranda sarà sempre 0, quindi possiamo tranquillamente dire anche senza calcoli che $\varepsilon(-x_{p0}) = 0$.

- Per $-x_{p0} < x < 0$, abbiamo che $\rho(x) = +qN_D$; applicando l'espressione prima proposta,

$$\varepsilon(x) = \varepsilon(-x_{p0}) + \int_{x_{p0}}^x -\frac{qN_A}{\epsilon_S} dx = -\frac{qN_A}{\epsilon_S}(x + x_{p0})$$

Questo poichè il valore iniziale $\varepsilon(-x_{p0}) = 0$. Calcoliamo il valore finale di quest'espressione, ossia nel punto $x = 0$, che diventerà il valore iniziale della prossima zona:

$$\varepsilon(0) = -\frac{qN_A}{\epsilon_S}x_{p0}$$

- Per $-x_{p0} < x < 0$, $\rho = +qN_D$, quindi calcoliamo che:

$$\varepsilon(x) = \varepsilon(0) + \int_0^x \frac{qN_D}{\epsilon_S} dx = -\frac{qN_A}{\epsilon_S}x_{p0} + \frac{qN_D}{\epsilon_S}x$$

Il valore finale di questo intervallo varrà:

$$\varepsilon(x_{n0}) = -\frac{qN_A}{\epsilon_S}x_{p0} + \frac{qN_D}{\epsilon_S}x_{n0}$$

- Per $x_{n0} < x < +\infty$, potremmo riapplicare tutti i ragionamenti ed i risultati analitici finora ricavati, per poter ricavare il valore del campo elettrico $\varepsilon(x)$ nella regione. In realtà, possiamo ragionare in un modo molto più qualitativo ma altrettanto efficace: applicando la neutralità, se il campo elettrico è nullo a sinistra della regione di carica negativa, se non fosse nullo a destra della regione di carica positiva, l'ipotesi di neutralità sarebbe impossibile da applicare nella giunzione p-n. Applicando in modo molto leggero l'ipotesi di neutralità, possiamo dunque dire che

$$\varepsilon(x) = 0$$

Riassumendo, il valore del campo elettrico $\varepsilon(x)$ sarà:

$$\varepsilon(x) = \begin{cases} 0 & -\infty < x < -x_{p0} \\ -\frac{qN_A}{\epsilon_S}(x + x_{p0}) & -x_{p0} < x < 0 \\ -\frac{qN_A}{\epsilon_S}x_{p0} + \frac{qN_D}{\epsilon_S}x & 0 < x < x_{n0} \\ 0 & x_{n0} < x < +\infty \end{cases}$$

A questo punto, disponiamo di una funzione a tratti del campo elettrico $\varepsilon(x)$: possiamo ripetere un ragionamento analogo al precedente, utilizzando la relazione:

$$\Phi(x) = \Phi(a) - \int_a^x \varepsilon(x) dx$$

Consideriamo dunque ciascuna delle quattro zone:

- Per $-\infty < x < -x_{p0}$, $\varepsilon(x) = 0$, dunque l'integrale varrà 0 in ogni punto di questo primo intervallo. Anche il valore finale, $\Phi(-x_{p0})$ sarà 0.
- Per $-x_{p0} < x < 0$, abbiamo che $\varepsilon(x) = -\frac{qN_A}{\epsilon_S}(x + x_{p0})$. Integrando:

$$\Phi(x) = \Phi(-x_{p0}) - \int_{-x_{p0}}^x -\frac{qN_A}{\epsilon_S}(x + x_{p0}) dx = \frac{qN_A}{\epsilon_S}(x + x_{p0})^2$$

Il valore finale di quest'espressione, ossia il potenziale valutato in 0, sarà:

$$\Phi(0) = \frac{qN_A}{\epsilon_S} x_{p0}^2$$

- Per $0 < x < x_{n0}$, $\varepsilon(x) = -\frac{qN_A}{\epsilon_S} x_{p0} + \frac{qN_D}{\epsilon_S} x$. Integrando la precedente espressione:

$$\Phi(x) = \Phi(0) - \int_0^x \left(-\frac{qN_A}{\epsilon_S} x_{p0} + \frac{qN_D}{\epsilon_S} x \right) dx = \frac{qN_A}{2\epsilon_S} x_{p0}^2 + \frac{qN_A}{\epsilon_S} x_{p0} x - \frac{qN_D}{2\epsilon_S} x^2$$

Il valore finale di questa espressione, ossia il potenziale valutato in x_{n0} , sarà:

$$\Phi(x_{n0}) = \frac{qN_A}{2\epsilon_S} x_{p0}^2 + \frac{qN_A}{\epsilon_S} x_{p0} x_{n0} - \frac{qN_D}{2\epsilon_S} x_{n0}^2 = \Phi_i$$

- Per $x_{n0} < x < +\infty$, applicando la neutralità come nel caso del campo elettrico, vediamo che nulla può modificare l'andamento del potenziale, che rimane dunque costante, e dunque pari al valore del potenziale di barriera Φ_i .

Avremo un'espressione a tratti del potenziale:

$$\Phi(x) = \begin{cases} 0 & -\infty < x < -x_{p0} \\ \frac{qN_A}{\epsilon_S}(x + x_{p0})^2 & -x_{p0} < x < 0 \\ \frac{qN_A}{2\epsilon_S}x_{p0}^2 + \frac{qN_A}{\epsilon_S}x_{p0}x - \frac{qN_D}{2\epsilon_S}x^2 & 0 < x < x_{n0} \\ \frac{qN_A}{2\epsilon_S}x_{p0}^2 + \frac{qN_A}{\epsilon_S}x_{p0}x_{n0} - \frac{qN_D}{2\epsilon_S}x_{n0}^2 & x_{n0} < x < +\infty \end{cases}$$

3.2.4 Risoluzione del sistema

Abbiamo già applicato nel nostro sistema una prima volta l'equazione di neutralità, per semplificare alcuni conti; l'applicazione introdotta è stata tuttavia molto leggera; per poter risolvere il sistema, sarà necessario fare di più, mettendo a sistema tra loro tutte le tre condizioni, ottenendo risultati precisi. Dallo studio del potenziale con l'equazione di Poisson, possiamo dire che:

$$\Phi_i = \frac{qN_A}{2\epsilon_S}x_{p0}^2 + \frac{qN_D}{2\epsilon_S}x_{n0}^2 = \Phi_p + \Phi_n$$

Ossia si può distinguere la barriera come somma di due contributi, dovuti uno al lato p, ed uno al lato n. Le tre equazioni che abbiamo ricavato dalle nostre congetture sono le seguenti:

1.

$$qN_Ax_{p0} = qN_Dx_{n0}$$

2.

$$\Phi_i = \frac{kT}{q} \ln \frac{N_A N_D}{n_i^2}$$

3.

$$\Phi_i = \frac{qN_A}{2\epsilon_S}x_{p0}^2 + \frac{qN_D}{2\epsilon_S}x_{n0}^2$$

Mettendo a sistema la prima e la terza equazione, possiamo ricavare x_{n0} o x_{p0} :

$$x_{p0} = x_{n0} \frac{N_D}{N_A} \implies \Phi_i = \frac{qN_A}{2\epsilon_S}x_{n0} \frac{N_D}{N_A} + \frac{qN_D}{2\epsilon_S}x_{n0}^2 = \frac{qN_D}{2\epsilon_S} \left(\frac{N_A + N_D}{N_A} \right) x_{n0}^2$$

Poichè la seconda equazione ci fornisce un'espressione operativa della barriera di potenziale Φ_i , possiamo facilmente ricavare che:

$$x_{n0} = \sqrt{\frac{2\epsilon_S}{qN_D} \Phi_i \frac{N_A}{N_A + N_D}}$$

Possiamo ora dire di aver trovato ogni dato necessario per quotare il diagramma a bande della giunzione p-n: ci imponiamo solo di mettere un poco ordine nelle espressioni, esprimendo l'ultimo risultato trovato in modo più saggio: anzichè considerare una singola ampiezza, consideriamo la totale ampiezza (come somma delle due ampiezze) delle regioni di svuotamento, x_{d0} :

$$x_{d0} = x_{n0} + x_{p0}$$

Partendo da qua, scriviamo un'espressione di x_{d0} :

$$x_{d0} = x_{n0} + \frac{N_D}{N_A} x_{n0} = x_{n0} \left(\frac{N_A + N_D}{N_A} \right) = \sqrt{\left(\frac{N_A + N_D}{N_A} \right)^2 \frac{N_A}{N_D + N_A} \frac{2\epsilon_S}{qN_D} \Phi_i};$$

$$x_{d0} = \sqrt{\left(\frac{N_A + N_D}{N_A} \right) \frac{2\epsilon_S}{qN_D} \Phi_i}$$

Esprimendo ora la funzione con il concetto di drogaggio equivalente N_{eq} , definito come somma armonica dei due drogaggi:

$$N_{eq} = \frac{1}{\frac{1}{N_A} + \frac{1}{N_D}} = \frac{N_A N_D}{N_A + N_D}$$

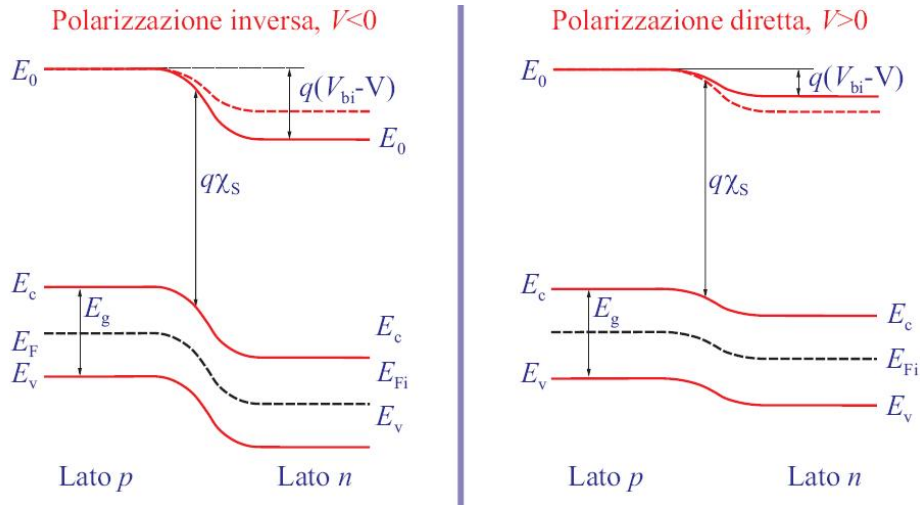
Si ottiene che:

$$x_{d0} = \sqrt{\frac{2\epsilon_S}{qN_{eq}} \Phi_i}$$

3.3 Applicazione di una tensione esterna alla giunzione

Abbiamo finora studiato la giunzione p-n in un caso di equilibrio. Potremmo ora pensare a casi in cui non vi è equilibrio, poichè esso viene perturbato dall'esterno, mediante l'applicazione di una batteria di tensione V_a alla giunzione.

Consideriamo la seguente convenzione: il + della batteria va collegato al lato p, il - della batteria al lato n. I contatti di accesso alla giunzione



saranno metallici (supponiamo ad esempio di oro Au): un accesso ad un semiconduttore (e quindi ad una giunzione) deve essere mediato da un contatto ohmico, al fine di poter eliminare resistenze parassite. Avremo a che fare quindi sostanzialmente con tre interfacce, e tre potenziali di contatto. Potremmo dunque ridisegnare il diagramma a bande del nostro sistema Au-Si-Au: supponendo di trovarci in uno stato di equilibrio, il livello di Fermi sarà ovunque costante. Il metallo non potrà, a causa della natura fluida del suo mare di elettroni, presentare concavità in prossimità dei contatti, e quindi solo il diagramma a bande dei semiconduttori potrà subire curvature. Lontano dalle interfacce, i materiali si comporteranno come isolati.

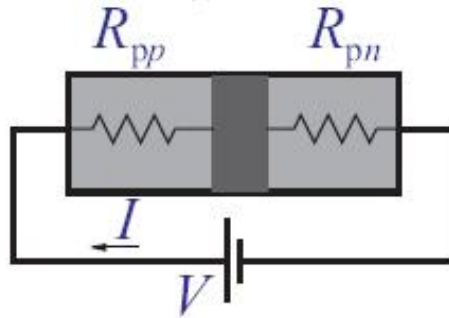
Cosa possiamo osservare: il metallo Au ha un livello di estrazione superiore a quello del silicio; come già detto, il raccordo sarà solo da parte dei semiconduttori poichè la natura fluida del mare di elettroni non permette modifiche al materiale. Oltre alla barriera Φ_i , avremo altre due barriere di potenziale, causate dalla differenza tra i lavori di estrazione. Ogni tensione di contatto si può circuitalmente rappresentare mediante una batteria: il contatto Au-Si_p sarà una tensione positiva rispetto a p su n, poichè è dall'alto verso il basso. Il contatto Si_n-Au sarà una tensione dal basso verso l'alto rispetto a p, e quindi sarà negativa. Circuitalmente, si può pensare al sistema come una maglia di batterie.

L'equazione di questa maglia sarà la seguente:

$$\Phi_{Au,p} - \Phi_i + \Phi_{n,Au} = 0$$

Da ciò è facilmente ricavabile la tensione Φ_i :

$$\Phi_i = \Phi_{n,Au} - \Phi_{Au,p}$$



Supponiamo di poter avere un voltmetro ideale: esso non potrebbe misurare niente, poichè i potenziali di contatto bilanciano la tensione interna, che quindi sarà nulla.

Collegando finalmente la tensione esterna V_a , supponendo che i contatti metallici siano dei buoni contatti, ossia a bassissima resistenza, e che quindi provocano una bassa caduta di tensione per le correnti che vi entrano, possiamo pensare che le due resistenze serie dei contatti ohmici siano nulle, e che quindi la tensione alla giunzione subirà una variazione, raggiungendo un valore V_j , a noi incognito. La maglia di batterie sarà modificata, ed avrà un'equazione associata di questo tipo:

$$-V_a - \Phi_{Au,p} - V_j + \Phi_{n,Au} = 0$$

$$V_j = \Phi_{n,Au} - \Phi_{Au,p} - V_a$$

Ma noi abbiamo prima detto che $\Phi_i = \Phi_{n,Au} - \Phi_{Au,p}$; dunque:

$$V_j = \Phi_i - V_a$$

Quando applico una tensione positiva sul lato p, si abbassa la tensione sulla giunzione. A questo punto, in seguito a questa notevole scoperta, introduciamo due definizioni:

- Una giunzione polarizzata con una tensione esterna $V_a > 0$ è detta polarizzata direttamente;
- Una giunzione polarizzata con una tensione esterna $V_a < 0$ è detta polarizzata inversamente.

La tensione provoca alcune interessanti conseguenze: disegniamo, in un grafico, la distribuzione della tensione all'equilibrio, con una polarizzazione diretta, e con una polarizzazione inversa: la concavità della funzione rimarrà

sempre costante, ma quella che varierà sarà l'ampiezza delle regioni di svuotamento, e quindi volendo dire in modo poco elegante, il punto di inizio ed il punto di fine della funzione del potenziale (nonchè ovviamente l'altezza della barriera, come già dimostrato mediante il calcolo delle equazioni delle maglie): se si polarizza direttamente la giunzione, l'ampiezza della zona di svuotamento x_d diminuisce rispetto all'ampiezza di equilibrio x_{d0} , e la barriera di potenziale, a queste condizioni, diminuisce; al contrario, con una polarizzazione inversa, l'ampiezza x_d aumenta rispetto a quella di equilibrio, e la barriera di potenziale aumenta. Trascurando come da ipotesi cadute resistive sulle interfacce metalliche, capita ciò: supponendo che le distribuzioni di carica siano: $-qN_Ax_p$ (negative), e $+qN_Dx_n$ (positive); le tensioni, data l'ipotesi di buoni contatti, dipendono esclusivamente dagli ioni, ricavando a partire dalla loro funzione di densità l'espressione operativa del campo e del potenziale. Se la tensione esterna modula la barriera, allora essa modula potenziale (tensione), campo e carica. Con una V_a positiva ridurremo la regione di svuotamento: l'applicazione di una tensione esterna però non è in grado di modificare il drogaggio del semiconduttore, nella fattispecie le concentrazioni N_A ed N_D , e tantomeno il valore della carica fondamentale q (quantizzata come la carica di un protone o di un elettrone); V_a quindi potrà solo variare l'ampiezza della regione di svuotamento x_d , ma i parametri delle curve, ossia la pendenza delle rette che formano il triangolo (nel campo elettrico), o la concavità del doppio andamento parabolico (nel potenziale), resteranno sempre costanti, e in questo senso avremo sempre le stesse funzioni, con però un'ampiezza di zona di svuotamento diversa.

Potremmo ora chiederci come calcolare, a partire dalla variazione di Φ_i con l'applicazione della tensione esterna V_a , la variazione della zona d'ampiezza x_d . L'idea è semplicemente sostituire al posto della barriera di potenziale Φ_i nell'espressione di x_d la tensione finale sulla giunzione V_j :

$$x_d(V_a) = \sqrt{\frac{2\epsilon_S}{qN_{eq}}(\Phi_i - V_a)}$$

In realtà, l'effetto della tensione può essere molto più interessante di una semplice variazione di x_d : la tensione $V_a > 0$ riduce la barriera di potenziale, e così il flusso dei portatori verrà modificato, e nello specifico verranno variare le concentrazioni di portatori che attraverseranno la barriera, quindi generate correnti. Dal punto di vista elettrostatico una tensione esterna fa variare le regioni di svuotamento, e quindi la carica positiva Q^+ e quella negativa Q^- .

3.3.1 Capacità di svuotamento

Un effetto molto interessante della tensione è quello di rendere visibile una sorta di capacità del semiconduttore: dato che V_a modifica $qN_D x_n(V_a)$ e $-qN_A x_p(V_a)$, la situazione potrebbe ricordare quello che in elettrostatica si studia come modello di un condensatore: rivediamolo.

Data una carica Q , si può definire una capacità C come:

$$C = \left| \frac{dQ}{dV} \right|$$

Consideriamo come tensione la nostra V_a , e come carica ad esempio quella positiva, Q^+ . Ci servirà una Q^+ funzione di V_a , al fine di poterne calcolare la derivata in V_a . Ma noi abbiamo già visto che:

$$Q^+ = qN_D x_n(V_a)$$

$$x_n(V_a) = \sqrt{\frac{2\epsilon_S N_A}{qN_D(N_A + N_D)} (\Phi_i - V_a)}$$

$$Q^+(V_a) = \sqrt{\frac{2\epsilon_S N_A q^2 N_D^2}{qN_D(N_A + N_D)} (\Phi_i - V_a)} = \sqrt{2q\epsilon_S N_{eq} (\Phi_i - V_a)}$$

Calcolando il modulo della derivata rispetto a V_a , si trova che:

$$\begin{aligned} C &= \left| \frac{dQ^+(V_a)}{dV_a} \right| = \frac{2q\epsilon_S N_{eq}}{2\sqrt{2\epsilon_S q N_{eq}}} = \sqrt{\frac{q^2 \epsilon_S^2 N_{eq}^2}{2\epsilon_S q N_{eq} (\Phi_i - V_a)}} = \\ &= \sqrt{\frac{q\epsilon_S N_{eq}}{2(\Phi_i - V_a)}} = C(V_a) \end{aligned}$$

Si noti che la capacità è una funzione della tensione V_a : il condensatore con il quale stiamo modellando la regione di svuotamento è dunque comandabile in tensione, ossia la sua capacità varia con il variare della tensione che gli viene applicata.

Cerchiamo ora però di ragionare in un modo diverso, al fine di ottenere un risultato più semplice; abbiamo detto che:

$$x_d(V_a) = \sqrt{\frac{2\epsilon_S}{qN_{eq}} (\Phi_i - V_a)} \implies C(V_a) = \frac{\epsilon_S}{x_d(V_a)}$$

Abbiamo scoperto una cosa molto interessante: un semiconduttore si può comportare come un condensatore a facce piane parallele, di costante dielettrica ϵ_S , e distanza x_d tra le armature. Un semiconduttore svuotato dunque è equivalente ad un materiale dielettrico.

3.4 Esercizio pratico

Data una giunzione p-n in silicio con i seguenti drogaggi: $N_A = 10^{16}$; $N_D = 5 \cdot 10^{16}$:

1. Disegnare il diagramma a bande dettagliato all'equilibrio.
2. Calcolare il campo elettrico massimo all'equilibrio.
3. Calcolare il valore della tensione esterna V_a tale per cui il campo massimo raddoppi.
4. Calcolare il valore di capacità allo stato della giunzione nel punto precedente.
5. Si metta a punto un metodo di misura sperimentale che tramite misura di capacità permetta una valutazione del potenziale di contatto Φ_i

3.4.1 Risoluzione

Disegnare il diagramma a bande dettagliato all'equilibrio

Il primo punto è già stato risolto una volta, tuttavia lo discutiamo brevemente: ricordando le proprietà dei semiconduttori e dei metalli, e il fatto che il livello di Fermi deve essere sempre costante nel semiconduttore, una raffigurazione qualitativa è immediata. Per però poter quotare ogni singolo punto, mancano alcuni dati: le posizioni delle regioni di svuotamento, l'ampiezza della barriera, e i contributi di ogni lato alla barriera.

Si ricava immediatamente che:

$$\Phi_i = \frac{kT}{q} \ln \frac{N_A N_D}{n_i^2}$$

L'ampiezza della banda di svuotamento sarà:

$$x_{d0} = x_{n0} + x_{p0} = \sqrt{\frac{2\epsilon_S}{N_{eq}} \Phi_i}$$

Sappiamo che $\epsilon_S = \epsilon_0 \epsilon_{rS} = 11,7 \cdot 8,854 \cdot 10^{-14}$. $N_{eq} = 8,33 \cdot 10^{15}$.

Svolgendo i conti,

$$x_{d0} = 3,4 \cdot 10^{-7}$$

Potremmo ora ragionare in due modi diversi: uno è quello di considerare le espressioni operative per il calcolo delle singole ampiezze x_{p0} e x_{n0} . Un altro modo, più intelligente, è quello di considerare il fatto che le distribuzioni di carica sono rettangolari, e quindi che:

$$\frac{N_A}{N_D} = \frac{x_{n0}}{x_{p0}}$$

Si può ricavare dunque che:

$$x_{n0} = 58nm; \quad x_{p0} = 292nm$$

Dunque, i valori dei contributi della barriera saranno:

$$\Phi_n = \frac{qN_D}{2\epsilon_S} x_{n0}^2 = 0,12V$$

$$\Phi_p = \frac{qN_A}{2\epsilon_S} x_{p0}^2 = 0,65V$$

Abbiamo quotato tutte le grandezze utili al diagramma a bande.

Calcolare il campo elettrico massimo all'equilibrio

Passiamo al punto successivo: il calcolo del campo elettrico massimo nella giunzione. Anche qua ci vengono aperte notevoli strade, ma la teoria ci insegna che il massimo del campo è in prossimità dell'origine degli assi, ergo del punto di giunzione, e che equivale all'area del triangolo. Possiamo dunque calcolarlo così:

$$\varepsilon(0) = -\frac{qN_A}{\epsilon_S} x_{p0} = \frac{-qN_D}{\epsilon_S} = 44kV/cm$$

Calcolare il valore della tensione esterna V_a tale per cui il campo massimo raddoppi

Risolto il punto 2, tentiamo di risolvere in modo altrettanto rapido il punto 3: anche qua la teoria potrebbe spingerci a conti molto elaborati, ma in realtà si tratta esclusivamente di risolvere un problema in cui abbiamo un triangolo la cui area raddoppia, al raddoppiare della base. Quindi:

$$(\Phi_i - V_a) = \frac{2x_d \varepsilon(0) \cdot 2}{2} \implies V_a = 2,2eV$$

Calcolare il valore di capacità allo stato della giunzione nel punto precedente

Il quarto punto potrebbe nuovamente trarci in inganno, ma anch'esso è decisamente semplice: esiste un sistema banalissimo di calcolare la capacità di svuotamento del semiconduttore, considerandolo (come la teoria ci può suggerire) come un condensatore a facce piane parallele. Quindi:

$$C_{DEP} = \frac{\epsilon_S}{x_d} = 15nF/cm^2$$

Si metta a punto un metodo di misura sperimentale che tramite misura di capacità permetta una valutazione del potenziale di contatto Φ_i

Il quinto punto è molto più interessante di quelli passati finora. Si noti come, sfruttando le relazioni giuste e con qualche osservazione intelligente, i problemi diventino molto più semplici. Questo punto però effettivamente richiede osservazioni più importanti: vogliamo trovare un metodo per misurare, in buona approssimazione, la tensione di contatto Φ_i . Dalla teoria, sappiamo che è impossibile realizzare un voltmetro capace di ciò, quindi dovremmo affidarci a misure di tipo indiretto. La traccia dell'esercizio suggeriva di sfruttare effetti capacitivi al fine di realizzare questo metodo. Dalla teoria sappiamo che:

$$C_{DEP} = \sqrt{\frac{q\epsilon_S N_{eq}}{2(\Phi_i - V_a)}}$$

Vorremmo ora una funzione facile da studiare, a partire da questa, e che non presenti il seguente problema: un'idea è prendere l'espressione di C_{DEP} , e considerare C_{DEP}^{-2} :

$$\frac{1}{C_{DEP}^2} = \frac{2(\Phi - V_a)}{q\epsilon_S N_{eq}}$$

Quest'espressione è molto interessante da studiare in quanto di fatto non è altri che una retta: la proprietà interessante è che l'intersezione con l'asse delle ascisse di questa retta non è altri che Φ_i , e quindi questo metodo di misura ci fornisce implicitamente due misure: Φ_i , ma anche N_{eq} , poichè la pendenza della retta $f(V_a)$, e quindi è possibile calcolare con un'ottima

precisione entrambe le grandezze. Possiamo dunque considerare terminato questo esercizio, e tornare a studiare gli effetti pratici dell'applicazione di una tensione esterna, su di una giunzione p-n.

3.5 Effetti di una tensione su di una giunzione p-n

Vorremmo ora comprendere, aldilà degli effetti puramente elettrostatici, le conseguenze dell'applicazione di una tensione esterna V_a ad una giunzione. Soprattutto, vorremmo studiare cosa capita ai flussi di portatori, in seguito alla modulazione della barriera provocata dalla tensione. Introduciamo una notazione che ci permetta di identificare univocamente i quattro flussi che entrano in gioco nella giunzione p-n:

1. Elettroni da lato n a lato p;
2. Elettroni da lato p a lato n;
3. Lacune da lato p a lato n;
4. Lacune da lato n a lato p.

Se applichiamo alla giunzione una tensione $V_a > 0$, e quindi siamo in un caso di polarizzazione diretta, il lato p scenderà e quello n salirà, in modo che la barriera sarà ridotta, il flusso 1 e il flusso 3 cresceranno, mentre gli altri rimarranno inalterati. Il nostro obiettivo è ora quello di quantificare la transizione dei portatori, ponendoci alcune domande specifiche: vorremmo poter quantificare la concentrazione dei portatori minoritari iniettati, e calcolare il livello di iniezione; V_a infatti determina l'iniezione di elettroni nel lato p, e di lacune nel lato n, dunque un'iniezione elettrica. Studiando le concentrazioni dei minoritari, partiamo da una condizione di equilibrio: possiamo dire, mediante le leggi dell'azione di massa (o mediante le equazioni di Shockley) che:

$$n_p(-x_p)|_{EQ} = \frac{n_i^2}{p_p(-x_p)} = \frac{n_i^2}{N_A} = n_{p0}$$

$$p_n(x_n)|_{EQ} = \frac{n_i^2}{n_n(x_n)} = \frac{n_i^2}{N_D} = p_{n0}$$

Possiamo intuire che, polarizzando la giunzione, $n_p(-x_p)$ e $p_n(-x_n)$ crescono notevolmente. Siamo dunque in grado di studiare l'equilibrio, ma non

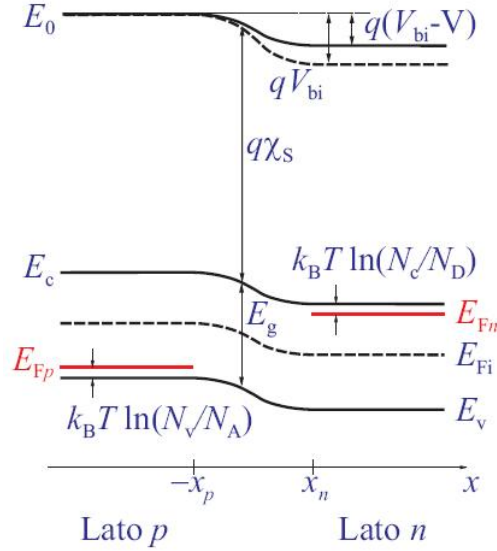
una condizione fuori equilibrio. Possiamo qualitativamente dire, come avevamo già affermato prima, che per una polarizzazione diretta l'ampiezza della regione di svuotamento si riduce, e che $V_j < \Phi_i$. Lontano dalla giunzione sappiamo che non abbiamo effetto della barriera, e che quindi molto a sinistra della giunzione il semiconduttore sarà tipo p neutro, molto a destra sarà tipo n neutro. Lontano dalla giunzione, di fatto, vi saranno le stesse condizioni che si trovano in una condizione di equilibrio. Volendo dunque calcolare le concentrazioni lontano dalla barriera, possiamo usare come al solito, le equazioni di Shockley. Dalle parti della barriera, in un suo intorno, nascono i nostri problemi: il livello di Fermi presenta una discontinuità, come abbiamo visto nel precedente grafico qualitativo. La discontinuità sarà esprimibile come un salto di ampiezza qV_a tra i due livelli E_F . In regioni neutre dunque sappiamo esattamente come comportarci, ma in prossimità della giunzione, e quindi in una zona fuori equilibrio, non sappiamo come procedere.

3.5.1 Quasi-Livelli di Fermi

L'idea vincente è la seguente: il livello di Fermi, quando si studiano problemi di equilibrio, è sempre la base da cui partire per determinare le soluzioni: con le equazioni di Shockley si possono infatti studiare, in funzione del livello di Fermi E_F , le concentrazioni di portatori. Come possiamo però descrivere un semiconduttore in funzione del livello di Fermi, fuori equilibrio? Consideriamo un caso di basso livello di iniezione: in un semiconduttore drogato tipo n, con N_D , ed un'iniezione di portatori, capita che i maggioritari restano di fatto inalterati, e che i minoritari subiscono una variazione sensibile. Ma cosa capita al livello di Fermi? Poichè aumentano i minoritari, esso subirà una variazione, poichè si ha una variazione dell'equilibrio finale, causata dall'iniezione. L'idea che permette di studiare casi fuori equilibrio mediante il livello di Fermi, è l'introduzione del concetto di quasi-livello di Fermi: in prossimità di una zona di non equilibrio, sdoppiamo il livello di Fermi in due livelli distinti: quello riferito ai maggioritari, nel nostro caso $E_{F_n} = E_F$, coinciderà con il livello all'equilibrio, poichè siamo in un regime di basso livello di iniezione, e l'equilibrio dei maggioritari non viene alterato dall'iniezione. Nei minoritari, si considererà un quasi livello di Fermi E_{F_p} , diverso da E_F , che considera il fatto che l'iniezione ha notevolmente alterato l'equilibrio dei minoritari, e quindi il loro livello di Fermi di riferimento.

Nello sdoppiamento dei livelli energetici consideriamo ad esempio per i maggioritari il quasi-livello di Fermi $E_{F_n} = E_F$ e quindi una condizione coincidente con l'equilibrio, mentre per i minoritari il quasi-livello di Fermi sarà E_{F_p} . Distinguendo questi due quasi-livelli di Fermi (quello dei minoritari si distingue sempre in quanto, nel diagramma a bande, è più spostato verso il

Polarizzazione diretta, $V > 0$



livello di Fermi intrinseco, E_{Fi}), possiamo calcolare sia n_n che p_n , mediante una banale applicazione delle equazioni di Shockley. Facendo le stesse osservazioni per un semiconduttore drogato p con drogaggio N_A , possiamo dunque introdurre le equazioni di Shockley:

$$N_{D0} : \begin{cases} n_{n0} = n_i e^{\frac{E_F - E_{Fi}}{kT}} \\ p_{n0} = n_i e^{-\frac{E_F - E_{Fi}}{kT}} \end{cases}$$

$$N_{A0} : \begin{cases} p_{p0} = n_i e^{\frac{E_{Fi} - E_F}{kT}} \\ n_{p0} = n_i e^{-\frac{E_{Fi} - E_F}{kT}} \end{cases}$$

In questi casi in equilibrio, la legge dell'azione di massa vale:

$$\begin{cases} N_{D0} = n_{n0} p_{n0} = n_i^2 \\ N_{A0} = n_{p0} p_{p0} = n_i^2 \end{cases}$$

Nei casi fuori equilibrio, le equazioni di Shockley saranno:

$$N_D : \begin{cases} n_n = n_i e^{\frac{E_{Fn} - E_{Fi}}{kT}} \\ p_n = n_i e^{-\frac{E_{Fn} - E_{Fi}}{kT}} \end{cases}$$

$$N_A : \begin{cases} p_p = n_i e^{\frac{E_{Fi} - E_{Fp}}{kT}} \\ n_p = n_i e^{-\frac{E_{Fi} - E_{Fp}}{kT}} \end{cases}$$

In questi casi fuori equilibrio, vediamo che le leggi dell'azione di massa subiranno una correzione (esponenziale) rispetto a quelle classiche in casi di equilibrio:

$$\begin{cases} N_D : n_n p_n = n_i^2 e^{\frac{E_{F_n} - E_{F_p}}{kT}} \\ N_A : n_p p_p = n_i^2 e^{\frac{E_{F_n} - E_{F_p}}{kT}} \end{cases}$$

Mediante questi strumenti teorici, ossia i quasi-livelli di Fermi, è possibile studiare un semiconduttore con le equazioni di Shockley anche in condizioni fuori equilibrio.

3.6 Leggi della Giunzione

Tornando allo studio della giunzione, cerchiamo di applicare quanto appena introdotto, e di utilizzare dunque sulle regioni di svuotamento le equazioni di Shockley. Ipotizziamo che lo svuotamento di queste induca un basso livello di iniezione, e quindi che il vecchio livello di Fermi descriva senza problemi la concentrazione dei maggioritari. Usando le equazioni di Shockley si vede che la distanza $E_{F_n} - E_{F_i}$ al variare di x , nel lato n , è costante fino a $x = x_n$: a partire dal punto x_n , ossia a partire da quando si entra nella zona di svuotamento, la differenza dei due livelli energetici crolla bruscamente, al variare di x , man mano che ci si avvicina alla giunzione, per essere precisi con un andamento esponenziale (come suggeriscono le equazioni di Shockley):

$$n_n(x) = n_i e^{\frac{E_F - E_{F_i}(x)}{kT}}$$

Quando $x = x_n$, $n_n(x) = n_{n0}$; per $x < x_n$, vi è una diminuzione esponenziale degli elettroni: questo perchè prima della zona di svuotamento la neutralità è garantita dagli elettroni derivanti dagli ioni positivi (iniettati come droganti). Per $x > x_n$ infatti, $p_n \simeq 0$, e quindi $N_D^+ - n_n \simeq 0$ (grazie all'ipotesi di completa ionizzazione che si può applicare sempre).

Via via che ci si inoltra nella zona di svuotamento, $n_n(x)$ diminuisce velocemente, poichè abbiamo la dipendenza dall'esponenziale della differenza $E_F - E_{F_i}(x)$, quindi, dalla condizione di neutralità in cui $n_{n0} \simeq N_D$, e $\rho = 0$, si arriva a:

$$n_n \ll n_{n0}; \quad n_n \ll N_D; \quad \rho \simeq N_D$$

Questa è una verifica a posteriori della validità dell'ipotesi di completo svuotamento. Si noti che nella regione di svuotamento esistono moltissimi portatori, ma che sono in numero estremamente trascurabile rispetto a N_D .

Stesso ragionamento, duale, per il lato p. Dire che non vi sono minoritari è un errore gravissimo: dire che però l'ipotesi di completo svuotamento è attuabile, può nascondere il fatto che i minoritari esistano, ma con una concentrazione inferiore di diversi ordini di grandezza rispetto a quella dei maggioritari.

La transizione delle concentrazioni avverrà con una certa gradualità; partendo dall'ipotesi di basso livello di iniezione, e dal fatto che i maggioritari nel lato n sono descritti da $E_{F_n} = E_F$, mentre nel lato p da $E_{F_p} = E_F$, possiamo architettare le nostre congetture:

1. In equilibrio:

$$n_{p0}(-x_p) = n_i e^{-\frac{E_{F_i}(-x_p) - E_F}{kT}}$$

2. Fuori equilibrio:

$$n_p(-x_p) = n_i e^{-\frac{E_{F_i}(-x_p) - E_{F_n}}{kT}}$$

Ma, nel lato p all'equilibrio, $E_F = E_{F_p}$, poichè trattiamo i maggioritari e il basso livello di iniezione. Ricaviamo dunque dalla prima espressione n_i , tenendo conto di quest'osservazione:

$$n_{p0} = n_i e^{-\frac{E_{F_i}(-x_p) - E_F}{kT}} \implies n_i = n_{p0} e^{\frac{E_{F_i}(-x_p) - E_F}{kT}}$$

Sostituendo nella seconda espressione, si ottiene che:

$$n_p(-x_p) = n_{p0} e^{-\frac{E_{F_i}(-x_p) - E_{F_n}}{kT}} e^{\frac{E_{F_i}(-x_p) - E_F}{kT}} = n_{p0} e^{-\frac{E_{F_p} - E_{F_n}}{kT}}$$

Dualmente, nel lato n, considerando il punto di inizio della regione di svuotamento $x = x_n$, mediante gli stessi passaggi, si ricava che:

$$p_n(x_n) = p_{n0} e^{\frac{E_{F_n} - E_{F_p}}{kT}}$$

Ma possiamo osservare che

$$\frac{E_{F_n} - E_{F_p}}{kT} = \frac{V_a}{V_T}$$

Abbiamo così dimostrato in modo formale, a partire dai quasi-livelli di Fermi, due equazioni fondamentali per lo studio della giunzione p-n:

$$n_p(-x_p) = n_{p0} e^{\frac{V_a}{V_T}} = \frac{n_i^2}{N_A} e^{\frac{V_a}{V_T}}$$

$$p_n(x_n) = p_{n0} e^{\frac{V_a}{V_T}} = \frac{n_i^2}{N_D} e^{\frac{V_a}{V_T}}$$

Cerchiamo di comprendere l'importanza ed il significato di queste due equazioni così fondamentali: l'esigenza iniziale era quella di capire quale concentrazione di carica si originasse, ossia quanti elettroni andassero dal lato n verso il lato p, e quante lacune dal lato p verso il lato n, in seguito ad una modulazione della barriera, in prossimità delle zone di svuotamento, ossia per $x = x_n$ e $x = -x_p$. Siamo arrivati a capire, mediante la formulazione delle leggi della giunzione, che la popolazione si sposta esponenzialmente con la barriera, ossia man mano che ci si addentra nelle regioni di svuotamento, si vede che si ha uno svuotamento con un andamento esponenziale. Al fine di arrivare alle leggi della giunzione, abbiamo introdotto i quasi-livelli di Fermi: possiamo immaginare che questi, all'aumentare della distanza dalla giunzione, tendano asintoticamente ad unirsi, fino a coincidere nel livello di Fermi all'equilibrio E_F .

La legge della giunzione infesterà con gli andamenti esponenziali quasi tutte le equazioni di funzionamento dei dispositivi elettronici. Ciò deriva dall'applicazione delle equazioni di Shockley (necessarie, quindi altre applicazioni non cambierebbero comunque questo tipo di andamento), che a loro volta derivano dalla prima approssimazione fatta, partendo dalla distribuzione di Fermi-Dirac, a quella di Boltzmann.

3.7 Caratteristica statica della giunzione p-n

Vogliamo ora studiare, nell'ambito della giunzione p-n, l'espressione della corrente I in funzione della tensione di polarizzazione V_a : $I(V_a)$. Cerchiamo, a questo scopo, di riformulare tutte le nozioni finora apprese, applicando le leggi della giunzione. All'equilibrio, i minoritari saranno:

$$n_{p0} = \frac{n_i^2}{N_A}$$

$$p_{n0} = \frac{n_i^2}{N_D}$$

Polarizzando direttamente la giunzione, a sinistra del punto $-x_p$ ed a destra del punto x_n , le concentrazioni dei minoritari saranno costanti; in questi punti però il valore della concentrazione di portatori minoritari aumenterà esponenzialmente con la tensione, poichè vi è un abbassamento delle barriere di potenziale; i minoritari sul lato p (nel punto $-x_p$) e sul lato n (in x_n) saranno rispettivamente:

$$n_p(-x_p) = n_{p0} e^{\frac{V_a}{V_T}}$$

$$p_n(x_n) = p_{n0} e^{\frac{V_a}{V_T}}$$

Cosa succede in parole povere: mediante l'applicazione della tensione, vengono iniettati elettricamente portatori minoritari nel lato: dal lato n gli elettroni (che in n sono maggioritari) vengono iniettati nel lato p (dove diventano minoritari), e viceversa da p ad n le lacune da maggioritarie divengono minoritarie. Studiamo dunque, lato per lato, l'iniezione di portatori, partendo dallo studio delle funzioni $p_n(x)$ ed $n_p(x)$, in posizioni generiche. Per studiare le lacune iniettate dal lato p al lato n, dovremo studiare l'andamento dell'eccesso di lacune nel lato n al variare della posizione x , ossia la differenza tra le lacune presenti all'equilibrio e quelle iniettate dal lato p a quello n:

$$p'_n(x) = p_n(x_n) - p_{n0} = p_{n0} e^{\frac{V_a}{V_T}} - p_{n0} = p_{n0} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Avevamo accennato a p'_n studiando la generazione ottica in un esempio pratico precedentemente analizzato: anche in questo caso avviene un fenomeno di iniezione, ma con cause e modalità differenti. Cerchiamo di meglio definire la funzione $p'_n(x)$, sfruttando il modello matematico del semiconduttore: partendo dall'equazione di continuità, e dal modello di ricombinazione diretta:

$$0 = -\frac{1}{q} \frac{\partial J_p}{\partial x} - \frac{p_n - p_{n0}}{\tau_p}$$

Consideriamo valida l'ipotesi di quasi neutralità del semiconduttore: $\varepsilon \simeq 0$. La corrente di drift sarà dunque circa uguale a 0, ed avremo una corrente puramente diffusiva. Quindi:

$$J_p = J_{diff} = -q \frac{\partial p_n}{\partial x} D_p$$

Sostituendo nell'equazione di continuità, si ottiene, considerando l'eccesso $p'_n = p_n - p_{n0}$:

$$0 = D_p \frac{\partial^2 p'_n}{\partial x^2} - \frac{p'_n}{\tau_p}$$

Considerando l'espressione della lunghezza di diffusione $L_p = \sqrt{D_p \tau_p}$, la forma della soluzione dell'equazione differenziale sarà:

$$p'_n(x) = A e^{\frac{x}{L_p}} + B e^{-\frac{x}{L_p}}$$

Il semiconduttore si può considerare nell'ipotesi di lato lungo: la lunghezza del lato n rispetto alla lunghezza di diffusione L_p è molto elevata. Possiamo dunque confondere la lunghezza del lato W_n con $+\infty$, e poter studiare le solite condizioni al contorno:

$$p'_n(W_n) = 0 \implies A = 0$$

$$p'_n(x_n) = B e^{-\frac{x_n}{L_p}} \implies B = p'_n(x_n) e^{\frac{x_n}{L_p}}$$

Sostituendo nell'equazione i coefficienti A e B , si ottiene:

$$p'_n(x) = p'_n(x_n) e^{\frac{x_n}{L_p}} e^{-\frac{x}{L_p}} = p'_n(x_n) e^{-\frac{x-x_n}{L_p}}$$

Quindi $p'_n(x)$ dipende esclusivamente da un fenomeno di diffusione, subisce un fenomeno di ricombinazione, e dopo poche lunghezze di diffusione si potrà considerare nulla. Ciò avvalorava a posteriori le nostre congetture sul quasi-livello di Fermi: esso ci ha permesso di calcolare la concentrazione delle lacune sul lato n; poichè l'esponentiale ha l'esponente normalizzato di L_p , dopo che l'esponente sarà multiplo di L_p , il decadimento esponenziale di portatori minoritari al variare della posizione dovuto alla ricombinazione annullerà o quasi la presenza dei suddetti minoritari. Per questo il quasi-livello di Fermi coinciderà con il livello di Fermi, poichè le lacune saranno tutte ricombinate, come ci suggerisce l'espressione appena trovata. Considerando il fatto che, come prima abbiamo visto, abbiamo un'espressione dei minoritari nel punto x_n :

$$p'_n(x_n) = p_{n0} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Sostituendo nella nostra espressione ottenuta dal modello matematico avremo che:

$$p'_n(x) = p_{n0} \left(e^{\frac{V_a}{V_T}} - 1 \right) e^{-\frac{x-x_n}{L_p}}$$

Sul lato p, si può dimostrare con gli stessi passaggi che capita qualcosa di completamente duale:

$$0 = \frac{1}{q} \frac{\partial J_n}{\partial x} - \frac{n'_p}{\tau_n}$$

$$J_n = J_{diff} = q D_n \frac{\partial n_p}{\partial x}$$

$$\implies \frac{1}{q} \frac{\partial^2 n'_p}{\partial x^2} - \frac{n'_p}{\tau_n}$$

Data $L_n = \sqrt{D_n \tau_n}$:

$$n'_p(x) = Ae^{\frac{x}{L_n}} + Be^{-\frac{x}{L_n}}$$

Questa volta ci interesserà confondere con $-\infty$ la lunghezza del lato $-W_p$: ad annullarsi sarà questa volta però il termine B , nella prima condizione al contorno, poichè sostituendo nell'esponenziale $x \rightarrow -\infty$, il termine di A tenderà a 0 da solo, mentre quello di B no.

$$n'_p(W_p) = 0 \implies B = 0$$

$$n'_p(-x_p) = Ae^{-\frac{x_p}{L_n}} \implies A = n'_p(-x_p)e^{\frac{x_p}{L_n}}$$

Sostituiamo nella soluzione il coefficiente A e ricaviamo, dualmente a prima:

$$n'_p(x) = n_{p0} \left(e^{\frac{V_a}{V_T}} - 1 \right) e^{\frac{x+x_p}{L_n}}$$

Avendo svolto i passaggi più rapidamente poichè esattamente duali a prima.

Riassumiamo ora ciò che abbiamo ricavato da questo studio dell'andamento dei portatori mediante il fenomeno di diffusione in seguito all'iniezione:

- Lato p: i portatori saranno gli elettroni iniettati dal lato n, e dunque i relativi valori deriveranno solo dalla diffusione di questi:

$$n'_p(x) = n'_p(-x_p)e^{\frac{x+x_p}{L_n}}$$

$$n'_p(-x_p) = n_{p0} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

$$J_n = J_{n,diff}(x) = -qD_n \frac{\partial n'_p}{\partial x} = \frac{qD_n}{L_n} n'_p(-x_p)e^{\frac{x+x_p}{L_n}}$$

- Lato n: i portatori saranno le lacune iniettate dal lato p, e dunque i relativi valori deriveranno solo dalla diffusione di queste:

$$p'_n(x) = p'_n(x_n)e^{-\frac{x-x_n}{L_p}}$$

$$p'_n(x_n) = p_{n0} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

$$J_p = J_{p,diff}(x) = qD_p \frac{\partial p'_n}{\partial x} = \frac{qD_p}{L_p} p'_n(x_n) e^{-\frac{x-x_n}{L_p}}$$

Le correnti avranno un andamento come quello appena descritto.

3.8 Studio delle correnti nella giunzione p-n

Riformuliamo ora meglio l'andamento delle correnti; possiamo infatti scriverle, nel seguente modo:

$$J_{p,diff}(x) = \frac{qD_p p_{n0}}{L_p} \left(e^{\frac{V_a}{V_T}} - 1 \right) e^{-\frac{x-x_n}{L_p}}$$

$$J_{n,diff}(x) = \frac{qD_n n_{p0}}{L_n} \left(e^{\frac{V_a}{V_T}} - 1 \right) e^{\frac{x-x_p}{L_n}}$$

Abbiamo informazioni su di un intorno della giunzione; non sappiamo però niente sulle correnti lontane da essa. Il fatto che vi sia una variazione di corrente in prossimità della giunzione, garantisce il fatto che la corrente deve essere globale, in tutto il semiconduttore, poichè altrimenti vi dovrebbero essere infiniti accumuli di carica che permettono il mantenersi della corrente solo localmente, e ciò va contro le leggi di conservazione della carica.

3.8.1 Polarizzazione diretta

Tentiamo di capire cosa capita nel nostro modello, quando alla giunzione è collegata una batteria che le fornisce una polarizzazione diretta. Abbiamo da sinistra una corrente di lacune che va verso destra, passa per la giunzione (dove consideriamo che non avvengono fenomeni di alcun tipo, in prima approssimazione), e quindi viene ricombinata nel lato in cui diventa minoritaria; dualmente da sinistra a destra vi è una corrente di elettroni che passa per la giunzione, e viene ricombinata.

Ci chiediamo però da dove nascano queste due correnti: poichè abbiamo escluso per conservazione della carica che vi sia corrente solo nella giunzione o nelle sue prossimità, nella fatiscie che le cariche che permettono la conducibilità siano tutte nella zona di svuotamento $[-x_p; x_n]$, ci chiediamo cosa generi queste correnti. Analizzando la situazione, vediamo che dal lato p giungono lacune, che vanno in n mediante un meccanismo diffusivo, prima descritto quantitativamente. Ma cosa trasporta fino alla giunzione le cariche che poi verranno inviate dall'altro lato della giunzione, mediante diffusione? In altre parole, cosa permette la corrente?

In una giunzione polarizzata direttamente, vi è un campo elettrico dovuto alla tensione applicata ai capi del dispositivo, che non è in grado di modificare la mobilità dei minoritari, ma è in grado di muovere senza problemi una corrente di maggioritari. Dal lato neutro della giunzione dunque proverrà una corrente di drift, di trascinamento, di portatori maggioritari, che, giunta in prossimità della zona di svuotamento, si dividerà in due sezioni: una si preoccuperà di ricombinare la corrente arrivata dall'altro lato, un'altra sarà inviata all'altro lato: dualmente capiterà nell'altro lato la stessa cosa, e in questo modo sarà motivabile sia la presenza della corrente, che la presenza della ricombinazione a partire della zona di svuotamento: il meccanismo chiave della corrente in polarizzazione diretta è questa corrente di drift portata dal campo, che ricombina la corrente giunta per diffusione dall'altro lato.

Riassumiamo ciò che abbiamo appena detto: abbiamo in precedenza quantificato gli eccessi di carica di minoritari iniettati in un lato dall'altro, seguendo un modello completamente diffusivo (in quanto parliamo di minoritari). Matematicamente si è verificato che, poichè la corrente di diffusione ha un modello di decadimento esponenziale, dopo alcune lunghezze di diffusione si potrà considerare sparita tutta la corrente iniettata: ciò è dovuto ad un fenomeno di ricombinazione con portatori maggioritari del lato trasportati mediante il campo elettrico della polarizzazione. I minoritari verranno saturati, e così si verificherà il moto di cariche globale in tutta la giunzione, e quindi la corrente. La somma delle due correnti esterne alla giunzione sarà la corrente totale in essa contenuta e mossa:

$$J_{tot} = J_{p,diff}(x_n) + J_{n,diff}(-x_p) = \left(\frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n} \right) \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Questa corrente totale J_{tot} sarà anche coincidente con la somma delle due correnti di drift, lontane dalla giunzione. In questo modo, studiando semplicemente le correnti di diffusione, è possibile calcolare le correnti di drift in un punto qualsiasi del sistema lontano dalla giunzione. Introducendo l'effetto del campo e della ricombinazione abbiamo così esteso le correnti ovunque nella giunzione p-n.

Questo è il meccanismo alla base di un diodo a giunzione (o almeno, metà di esso, poichè stiamo solo considerando la corrente in polarizzazione diretta); studiamolo ora sotto il punto di vista di altri fenomeni, considerando la corrente nel diodo, I : data l'area della giunzione A_j , consideriamo la corrente I nel diodo come:

$$I = qn_i^2 A_j \left(\frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n} \right) \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Passiamo ora allo studio della seconda metà del modello del nostro diodo a giunzione, studiandolo in regime di polarizzazione inversa.

3.8.2 Polarizzazione inversa: corrente inversa di saturazione

Data l'espressione appena analizzata, viene definito il termine I_0 come corrente inversa di saturazione:

$$I_0 = qn_i^2 A_j \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right)$$

Questa è la corrente che sperimentalmente si può misurare applicando una tensione esterna sufficientemente negativa al diodo: data infatti una giunzione, essa, se polarizzata con una $V_a \ll 0$, e misurata la corrente passante in essa, avrà il valore I_0 . Il meccanismo fisico dietro questo fenomeno è differente da quello che abbiamo considerato in polarizzazione diretta, anche se anche qua si può scorgere una dualità tra i due fenomeni.

Supponiamo dunque di polarizzare inversamente il nostro diodo: se disegnassimo il diagramma a bande della giunzione, essa dovrebbe presentare il lato p molto rialzato rispetto al lato n. La barriera di potenziale derivante dalla polarizzazione sarà molto elevata, e dunque i maggioritari non avranno la possibilità di attraversarla. Saranno tuttavia possibili transizioni di minoritari, che non avrebbero fatica ad attraversarla, poichè la barriera per i maggioritari rappresenta una sorta di scivolo per i minoritari. Ciò ci fa intuire che la corrente inversa di saturazione è provocata da un moto di portatori minoritari, e dunque che abbia una forte dipendenza dalla temperatura (come meglio diremo dopo però): si inizi a considerare il fatto che $I_0 \propto n_i^2$, ma che n_i varia sensibilmente con la temperatura: esiste una regola empirica che afferma che in un intorno di 300 K, per ogni incremento di 10 K la corrente raddoppi.

Anche in questo caso, in prossimità della giunzione, non consideriamo effetti di alcun tipo: potremmo considerare ricombinazione o generazione termica, che aumenterebbero o ridurrebbero la corrente. Ma l'efficacia di questi due fenomeni dipende dalla temperatura: una volta che viene generata una coppia elettrone-lacuna all'interno della giunzione, questa viene immediatamente separata dal campo, che allontana dalla giunzione i portatori. Altri fenomeni quali la generazione ottica potrebbero comunque aumentare la corrente inversa di saturazione, sempre generando nuove coppie elettrone-lacuna. Ma noi non ci preoccupiamo di questi aspetti, per studiare quindi solamente cosa capita a livello di correnti. Fenomeni additivi non verranno

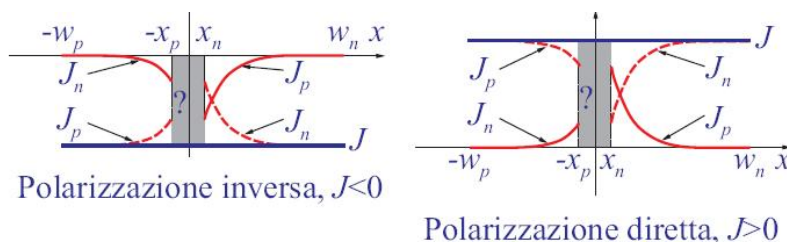
considerati per ora nel nostro modello, anche se dispositivi quali fotodiodi o celle fotovoltaiche si basano proprio su di essi.

Abbiamo detto che i minoritari sono in grado di transire la barriera, divenendo maggioritari nel lato in cui arrivano; applicando la legge della giunzione, sappiamo che:

$$n'_p(-x_p) = n_{p0} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

$$p'_n(x_n) = p_{n0} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Gli eccessi di minoritari, man mano che ci si allontana dalla giunzione, tendono ad azzerarsi. Le correnti saranno tutte negative, in quanto consideriamo correnti di minoritari. Potremo dire che avremo, derivando i portatori rispetto alla variabile spaziale x (applicando nuovamente dunque un modello diffusivo), delle correnti, ma di segno opposto alle precedenti. Potremmo raffigurare di nuovo un diagramma delle correnti diffuse, al fine di studiare la corrente inversa di saturazione I_0 :



Il campo di polarizzazione non è in grado di agire sui minoritari: se fosse possibile, allora la corrente sarebbe opposta rispetto al verso che effettivamente risulta avere. Giunti dall'altro lato, i minoritari diverranno maggioritari. La situazione è del tutto duale alla precedente, ossia alla polarizzazione diretta: prima i maggioritari divenivano minoritari, ora i minoritari divengono maggioritari. Prima il fenomeno che permetteva la corrente era la ricombinazione delle correnti arrivate per diffusione unite ad un effetto del campo; ora l'effetto del campo sarà sempre presente (anche se in modo diverso), ma il fenomeno chiave del meccanismo sarà la generazione termica: mediante generazione termica verranno a crearsi coppie elettrone-lacuna che vengono immediatamente separate: i maggioritari rimarranno per qualche tempo nel lato p, mentre i minoritari saranno iniettati nell'altro lato. Ad evitare accumuli di carica, di maggioritari, che andrebbero contro le leggi di conservazione, è il campo elettrico: poichè il semiconduttore è polarizzato inversamente, la tensione tenderà ad allontanare il flusso di maggioritari dalla barriera. Poichè questo eccesso di carica è esiguo, dal momento che la

sola sorgente di cariche lontano dalla giunzione è la generazione termica, in questo ambito, I_0 sarà esigua, a meno che non ci si trovi in alte temperature o regimi di generazione ottica. La tensione di polarizzazione inoltre non è in grado di favorire il processo, poichè non varia la produzione di coppie elettrone-lacuna con il variare della tensione. Al contrario iniettando dunque termicamente od otticamente coppie elettrone-lacuna, le correnti di generazione aumenteranno, e quindi anche la I_0 , come proposto precedentemente in un breve cenno.

3.9 Esercizio Pratico

Data una giunzione p-n simmetrica brusca, drogata p $N_A = 10^{16}$, n $N_D = 5 \cdot 10^{16}$, sapendo che i tempi di vita dei portatori sono τ_n nel lato p, τ_p nel lato n, $\tau_n(N_D) = 0,5\mu s$, $\tau_p(N_A) = 0,3\mu s$:

1. Si valuti la corrente inversa di saturazione I_0 , e si disegni, quotandole, le distribuzioni delle correnti, all'applicazione di una tensione di polarizzazione $V_a = 0,5$ V, con area della giunzione $A_j = 1mm^2$, $W_n = W_p = 1mm$.
2. Si valutino i valori delle resistenze serie nel lato p e nel lato n e si discuta il loro effetto sulla caratteristica statica del diodo, $I(V_a)$

Una breve divagazione, prima di risolvere l'esercizio: cosa si intende precisamente per caratteristica statica? La risposta è semplice: la caratteristica statica è una funzione della tensione V_a che non considera fenomeni di transitorio: si considera, a tal fine, che la tensione V_a vari molto lentamente, in modo di poter vedere nel grafico della caratteristica statica esclusivamente i valori di regime della corrente, e non eventuali picchi di corrente transitori.

Una giunzione simmetrica è una giunzione in cui i livelli di drogaggio dei due lati sono confrontabili: per confrontabili, si intende una differenza di qualche ordine di grandezza. Se vi fosse un drogaggio tipico da una parte (circa 10^{16} per esempio), e nell'altro lato un drogaggio dell'ordine di 10^{19} , la giunzione avrebbe una forte asimmetria, con effetti che più avanti analizzeremo. Si arriva a parlare, in casi come quello appena accennato, di semiconduttore degenerare, poichè con un drogaggio elevato come 10^{19} il livello di Fermi del semiconduttore arriva a coincidere con il livello E_C della banda di conduzione.

Una giunzione si definisce brusca quando il drogaggio cambia rapidamente, ossia presenta un salto, come in tutte le giunzioni finora studiate; i tecnologi sono soliti definire una funzione del drogaggio al variare della posizione x , $N(x)$:

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

Se $N(x)$ presenta una discontinuità (che sarà ovviamente di tipo salto) la giunzione è detta brusca; se al contrario $N(x)$ è una funzione continua, si parla di giunzione graduale. Lo studio del grafico di $N(x)$ ci può ricordare molto l'andamento delle densità di carica: mediante $N(x)$ sarà infatti facile dunque studiare le grandezze ρ , ε , Φ .

3.9.1 Risoluzione

Si valuti la corrente inversa di saturazione I_0

Il calcolo della corrente I_0 si può considerare fattibile come somma di due contributi: il contributo del lato n, ed il contributo del lato p:

$$I_0 = I_{0n} + I_{0p}$$

Possiamo dunque semplicemente calcolare in questo modo la corrente inversa di saturazione:

$$I_{0n} = qn_i^2 A_j \frac{D_n}{L_n N_A}$$

$$I_{0p} = qn_i^2 A_j \frac{D_p}{L_p} N_D$$

Calcoliamo dunque i coefficienti di diffusione, mediante le relazioni di Einstein:

$$D_n = \frac{kT}{q} \mu_n(N_A); \mu_n(N_A) \sim 1250 \implies D_n = 32,5 \text{ cm}^2/\text{s}$$

$$D_p = \frac{kT}{q} \mu_p(N_D); \mu_p(N_D) \sim 300 \implies D_p = 7,8 \text{ cm}^2/\text{s}$$

$$L_n = \sqrt{D_n \tau_n} = 40 \mu_m$$

$$L_p = \sqrt{D_p \tau_p} = 15 \mu_m$$

$$I_{0n} = 2,73 \cdot 10^{-13} \text{ A}$$

$$I_{0p} = 0,35 \cdot 10^{-13} \text{ A}$$

$$I_0 = 3,08 \cdot 10^{-13} A$$

Ora che abbiamo calcolato il termine della corrente di saturazione inversa, quotare le correnti sarà molto semplice, in quanto possiamo usare l'espressione della corrente del diodo:

$$I_{p,diff} = I_{0p} \left(e^{\frac{V_a}{V_T}} - 1 \right) = 8 \mu A$$

$$I_{n,diff} = I_{0n} \left(e^{\frac{V_a}{V_T}} - 1 \right) = 61 \mu A$$

In toto, la corrente sul diodo con la tensione fornita dall'esercizio sarà:

$$I = I_{p,diff} + I_{n,diff} = 69 \mu A$$

Si valutino i valori delle resistenze serie nel lato p e nel lato n

Passiamo ora alla soluzione del secondo punto del nostro esercizio: il calcolo delle resistenze serie di accesso alla giunzione. Possiamo dire che, per arrivare al punto di giunzione, attraverseremo infatti due resistenze, opposte dalla lunghezza del lato p, e dalla lunghezza del lato n. Dalla legge di Ohm, sappiamo che, per resistenze piccole e correnti piccole, vi è una caduta di tensione molto piccola, e dunque si potrebbe considerare che tutta la tensione V_a vada alla giunzione, per ottenere una tensione risultante V_j . Vogliamo però comunque calcolare le resistenze opposte dai lati, al fine di capire effettivamente quanta tensione vi sia ai capi della giunzione, e comprendere in modo più quantitativo l'effetto della presenza del semiconduttore nel circuito. Possiamo considerare che:

$$R_{Sp} = \rho_p \frac{W_p - x_p}{A_j} = \frac{1}{q\mu_p p_{pv}} = \frac{1}{q\mu_p N_A} = 1,78 \cdot 10 = 17,8 \omega$$

Con un ragionamento del tutto analogo,

$$R_{Sn} = \rho_n \frac{W_n - x_n}{A_j} = \frac{1}{q\mu_n n_{nv}} = \frac{1}{q\mu_n N_D} = 0,114 \cdot 10 = 1,14 \omega$$

Il modello elettrico rappresentante il circuito sarà dunque quello del nostro diodo, con in serie le due resistenze R_{Sp} ed R_{Sn} collegate, in serie, alla batteria V_a . Si verifica un problema nello studio della caratteristica statica: si dovrà considerare, per alte tensioni, una V'_a considerata come la tensione sul solo diodo, I_D una corrente elevata sul diodo, al di sopra del quale, le resistenze

avranno un effetto non indifferente in quanto provocheranno una caduta di tensione sempre meno trascurabile. La tensione sul diodo, sarà uguale a:

$$V'_a = V_a - (R_{Sp} + R_{Sn})I_D$$

La corrente sul diodo avrà, basandoci sull'espressione della corrente del diodo ricavata a partire delle leggi della giunzione, un'espressione del tipo:

$$I_D = I_0 \left(e^{\frac{V_a - (R_{Sp} + R_{Sn})I_D}{V_T}} - 1 \right)$$

Si noti che quest'espressione è fortemente non lineare, in quanto la corrente I_D appare sia come variabile dipendente che come argomento dell'esponenziale, e dunque sarebbe necessario sfruttare metodi numerici al fine di determinare le soluzioni di quest'equazione in funzione di I_D . Intuitivamente, possiamo però fare un ragionamento molto più fine, pensando alla natura del circuito: di base avremmo una corrente che varia esponenzialmente con la tensione, ma le resistenze, facendo cadere tensione, al di sopra di un certo valore, faranno stabilizzare il valore del rapporto $\frac{I}{V}$, facendolo diventare da esponenziale a lineare per grossi valori di V_a . Al di sopra di questo valore di V_a , l'andamento della caratteristica statica sarà quello di una retta. In altre parole, in queste condizioni, più V_a sarà elevata, e sempre meno corrente in proporzione percorrerà il diodo.

3.10 Cenni ad altri modelli della giunzione

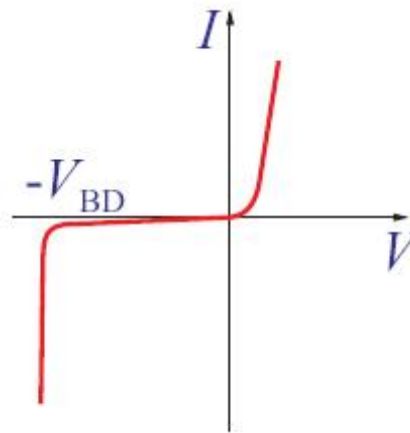
Continuiamo a discutere la modellizzazione del diodo a giunzione; finora abbiamo considerato un modello statico, basato cioè sul non considerare fenomeni di transitorio, effettuando lievi e lente variazioni di tensione, aspettando che i valori si stabilizzassero ad un regime. Il vero problema del nostro modello statico però è il seguente: abbiamo descritto in modo abbastanza valido il diodo a giunzione in zona diretta (anche se alcuni aspetti sono ancora da chiarire), ma molto male in zona inversa; introduciamo dunque alcune idee nuove, per poi approfondirle in seguito, in modo da estendere il nostro modello statico, accostandovi modelli più generali, quali quello di ampio segnale, di piccolo segnale, ed altre osservazioni.

3.10.1 Modello di ampio segnale

Considerando un diodo, in polarizzazione diretta, esso presenta una cosa particolare: considerando una scala meno sensibile di quella che ci permetterebbe di visualizzare chiaramente la corrente di saturazione inversa: consideriamo

ad esempio la scala del mA. Volendo ora disegnare un grafico della caratteristica statica, vedremmo che essa vale 0 fino ad un certo valore che noi chiameremo V_γ , ossia la tensione di accensione del diodo, leggermente positiva. Questa è la tensione in cui l'esponenziale inizia ad attivarsi, e a curvare la caratteristica statica, prima di stabilizzarsi ad una retta. Un veteroelettronico, ossia un elettronico esperto, dotato di manualità, affermerebbe di getto che se $V_a < V_\gamma$, il diodo sarebbe spento, e modellabile con un circuito aperto; se $V_a > V_\gamma$, il diodo si modellerebbe con un generatore di tensione V_γ ; questo valore, V_γ , sarà in seguito argomento di discussione, in quanto ora bisogna accennare a qualcosa di molto più interessante.

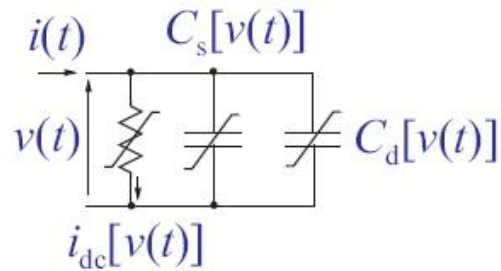
Se consideriamo una tensione di polarizzazione inversa $V_a \ll 0$, molto inferiore a 0, al di sotto di un certo valore, che noi chiameremo V_Z , o detta anche tensione Zener (nel disegno seguente indicata con V_{BD}), si avrà un'ingente corrente negativa: il diagramma corretto della caratteristica $I(V_a)$ avrebbe in realtà una forma del tipo:



Potremmo chiederci da dove provengono i valori V_γ e V_Z ; per rispondere a ciò, si è introdotto il modello di ampio segnale, che racchiude in sé tutte queste fenomenologie fisiche per noi ancora inspiegabili. Alla base del modello di ampio segnale, vi è un concetto già precedentemente introdotto, di elettrostatica: la capacità di svuotamento del semiconduttore C_{DEP} . In parallelo col diodo, potremo infatti immaginare un condensatore comandato in tensione, a capacità variabile. A ciò, si dovrà associare però un'ulteriore capacità variabile, sempre in parallelo al diodo: la capacità di diffusione C_{DEP} . Possiamo anticipare che essa è una capacità legata a fenomeni di diffusione di minoritari, in uno stato di polarizzazione inversa del diodo.

Il modello circuitale del modello di ampio segnale sarà il seguente:

Come già detto, il diodo a giunzione è in parallelo alle capacità (coman-



date in tensione) di svuotamento C_{DEP} e di diffusione C_{DEP} , il tutto in serie con le resistenze serie di accesso del semiconduttore.

Questo modello ci presenterà molti problemi, in quanto esso è un circuito fortemente non lineare, e dunque impossibile da studiare se non mediante simulatori numerici, che effettivamente fino a vent'anni fa non esistevano. Esistevano tecniche ingegneristiche in grado di eliminare o quantomeno accantonare la non linearità del circuito, al fine di poterlo ugualmente studiare; ora vedremo qualche cenno su questo.

3.10.2 Modello di piccolo segnale

La forte non linearità del circuito rende impossibile l'applicazione dell'elettrotecnica, e delle sue tecniche di analisi circuitale. La strategia vincente sarà quella di considerare un preciso punto di funzionamento del circuito, e di linearizzarlo in un intorno di questo valore, mediante una sorta di sviluppo in serie troncato. Piccolo segnale deriva proprio dal fatto che questo metodo funziona però solo localmente, a seconda del punto di lavoro che si utilizza.

Ricominciamo ora a trattare argomenti più dettagliati, e di risolvere alcuni dubbi introdotti in questi cenni.

3.11 Valutazione di V_γ

Come già detto, V_γ è la tensione di accensione del diodo; il diodo sarà in stato di polarizzazione diretta, e con una tensione sufficientemente elevata da far attivare l'esponenziale che controlla la crescita della caratteristica del diodo. Supponiamo di stabilire a 0,1 mA la minima corrente sensibile da uno strumento di misura: V_γ sarà la tensione tale per cui avremo garanzia di poter osservare questa minima corrente (scelta in realtà arbitrariamente), I_{0min} . La tensione minima in questione sarà data da:

$$V_\gamma = V_T \ln \frac{I_{0,min}}{I_0} \sim 0,6V$$

La scelta di I_{0min} che noi abbiamo ipotizzato di 0,1 mA per fare un esempio in realtà è poco influente, poichè il logaritmo attenua molto questo fattore. Si tratta solo di una scelta di corrente minima osservabile con un valore sensato. Si noti che il valore che abbiamo osservato di V_γ è circa a metà del valore dell'energy gap (1,12 eV). Questa è una regola empirica, ma sensata, in quanto gli ingegneri di vent'anni fa potevano solo sfruttare tali regole empiriche per semplificare i propri modelli. L'importanza storica di V_γ è la sua caratteristica di rappresentare un limite, al di sotto del quale non vi è crescita di corrente, al variare di V_a , ma dopo il quale la crescita è esponenziale, ossia quasi verticale potremmo dire. Poi, però, si stabilizza, e per questo motivo un diodo è modellabile con una batteria: al di sopra di un certo valore, la tensione resta sempre circa costante, poichè le resistenze serie di accesso al diodo rendono circa costante la tensione.

Passiamo ora all'analisi più dettagliata del diodo a giunzione in stato di polarizzazione inversa, analizzando i fenomeni di rottura della giunzione.

3.12 Meccanismi di rottura della giunzione

Se consideriamo una polarizzazione inversa con tensione $V_a \ll 0$, e quindi una tensione molto negativa, ci aspettiamo, dal modello statico, una corrente inversa di saturazione, e quindi costante. In realtà, al di sotto della tensione Zener V_Z , il diodo va in rottura (breakdown): vi è una rottura elettrostatica del diodo a giunzione.

Si noti che questo fenomeno è reversibile: la rottura è un fenomeno puramente elettrostatico, a meno che con una corrente troppo alta, si manifesti un effetto Joule non indifferente che produca un forte innalzamento termico del dispositivo provocando danni fisici, che però non dipendono dalle caratteristiche elettriche dispositivo, quanto dalla sua resistenza al calore che verrebbe dissipato dalle resistenze. Esistono due fondamentali meccanismi di breakdown della giunzione, che analizzeremo in modo però puramente qualitativo:

- Effetto Valanga
- Effetto Tunnel

3.12.1 Effetto valanga

Data una giunzione polarizzata inversamente, il lato p risulta essere molto rialzato rispetto al lato n: il campo elettrico, a causa della forte tensione, sarà molto elevato, come già detto. Nella zona di svuotamento, ci saranno,

come di consueto, fenomeni di generazione termica, che regoleranno la corrente di saturazione inversa: questi provocheranno la nascita di una coppia elettrone-lacuna. Il forte campo elettrico potrà accelerare verso sinistra le lacune (secondo la nostra usuale convenzione), e verso destra gli elettroni. L'urto con il reticolo alla velocità acquisita dall'elettrone causata dall'intensità del campo elettrico sarà così violento da riuscire a provocare effetti termici non trascurabili con esso, che daranno luogo a due nuove coppie elettrone lacuna (una per la lacuna scontratasi, una per l'elettrone scontratosi): a questo punto il campo separerà ed accelererà questi due nuovi elementi, che causeranno a loro volta nuove generazioni, che a loro volta causeranno nuove generazioni, e in questo modo vi sarà un effetto a valanga che provocherà una corrente non trascurabile nel diodo, e inversa, a causa del campo di polarizzazione.

3.12.2 Effetto tunnel

A causa della polarizzazione inversa, abbiamo una barriera di potenziale molto elevata. La banda di valenza, nel lato p, è ricchissima di elettroni (per definizione); poichè la barriera si alza, ma anche restringe, quando assume uno spessore sufficientemente ridotto (alcuni \AA), aumentano le probabilità dell'effetto tunnel, ossia dal fatto che la funzione d'onda degli elettroni contenga una buona probabilità dell'elettrone di superare la barriera nonostante essa abbia un potenziale molto più elevato di quello dell'elettrone. Con una barriera sufficientemente stretta, dunque, il tunnelling diventerà possibile e anzi probabile.

3.12.3 Diodi Zener

I diodi che funzionano prevalentemente in zona di breakdown sono detti diodi Zener. Il loro simbolo è quello di un diodo, con due sorte di alette aggiuntive. Essi vengono modellati, al di sotto del raggiungimento della zona Zener della tensione, con una batteria. Si noti che la convenzione di utilizzatore del diodo Zener è opposta rispetto a quella dei comuni diodi, poichè esso lavora per l'appunto in zona inversa. Questo tipo di dispositivo è utilizzato molto comunemente, in quanto con una certa corrente, esso fornirà sempre la stessa tensione, rappresentando così un ottimo punto di riferimento in un circuito.

Cosa ci garantisce un diodo a tunnel piuttosto che un diodo a valanga? Esistono diversi aspetti, ma il più importante è il drogaggio: un diodo a giunzione molto drogato, avrà una tensione Zener V_Z molto piccola, e quindi sarà a effetto tunnel; al contrario, un diodo a giunzione poco drogato, con una tensione Zener abbastanza elevata, sarà a valanga.

3.13 Calcolo della capacità di diffusione

Parlando del nostro modello di ampio segnale, abbiamo detto che in parallelo al diodo vi sono due capacità variabili: la capacità di svuotamento C_{DEP} e la capacità di diffusione C_{DIF} , ma non abbiamo ancora analizzato e considerato minimamente quest'ultima. Sappiamo solo che essa è legata all'accumulo di carica dei minoritari nei due lati.

Le distribuzioni dei minoritari saranno:

$$p'_n(x) = p'_n(x_n)e^{-\frac{x-x_n}{L_p}}$$

$$n'_p(x) = n'_p(-x_p)e^{\frac{x+x_p}{L_n}}$$

Le cariche Q'_p e Q'_n sono gli integrali delle distribuzioni nello spazio in cui agiscono gli eccessi di carica:

$$Q'_p = \int_{x_n}^{+\infty} qp'_n(x_n)e^{-\frac{x-x_n}{L_p}} dx = -qp'_n(x_n)(-L_p)e^{-\frac{x-x_n}{L_p}} \Big|_{x_n}^{+\infty} = qp'_n(x_n)L_p = \frac{qn_i^2 L_p}{N_D} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

$$Q'_n = \int_{-\infty}^{-x_p} qn'_p(-x_p)e^{\frac{x+x_p}{L_n}} dx = -qn'_p(-x_p)L_n e^{\frac{x+x_p}{L_n}} \Big|_{-\infty}^{-x_p} = -qn'_p(-x_p)L_n = \frac{-qn_i^2 L_n}{N_A} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Vedendo un legame tra carica e tensione, potremmo voler introdurre a questo punto una capacità. Per questo si definisce la capacità di diffusione, come la somma dei moduli delle derivate delle cariche rispetto alla tensione V_a :

$$C_{DIF}(V_a) = \left| \frac{\partial Q'_p}{\partial V_a} \right| + \left| \frac{\partial Q'_n}{\partial V_a} \right|$$

Da ciò, si ricava che:

$$C_{DIF}(V_a) = \frac{qn_i^2 L_p}{N_D} e^{\frac{V_a}{V_T}} + \frac{qn_i^2 L_n}{N_A} e^{\frac{V_a}{V_T}}$$

Si noti che se la giunzione è fortemente asimmetrica, il termine a drogaggio superiore sarà trascurabile, poichè la zona di svuotamento sarà infinitesima.

Anche questa capacità è una quantità fortemente non lineare: essa dipende esponenzialmente dalla tensione esterna V_a .

Torniamo alle espressioni appena ricavate degli eccessi di carica Q'_p e Q'_n : prendiamo ad esempio nello specifico la Q'_p :

$$Q'_p = \frac{qn_i^2 L_p}{N_D} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Si può facilmente notare una similitudine con la corrente di diffusione nel punto x_n , di lacune:

$$J_p(x_n) = \frac{qD_p n_i^2}{N_D L_p} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Moltiplicando e dividendo il termine L_p all'espressione di Q'_p , si ottiene:

$$Q'_p \cdot \frac{L_p}{L_p} = \frac{qn_i^2 L_p^2}{N_D L_p} \left(e^{\frac{V_a}{V_T}} - 1 \right) = \frac{qn_i^2 D_p \tau_p}{N_D L_p}$$

Dunque, si può vedere facilmente che:

$$J_{p,diff}(x_n) = \frac{Q'_p}{\tau_p}$$

In modo del tutto duale, si può verificare facilmente che:

$$J_{n,diff}(-x_p) = \frac{Q'_n}{\tau_n}$$

Poichè la corrente totale nel diodo è la somma dei due contributi di carica, possiamo dire che:

$$J_{tot} = J_{p,diff} + J_{n,diff} = \frac{Q'_p}{\tau_p} + \frac{Q'_n}{\tau_n}$$

Questa è la corrente totale nella giunzione in condizioni stazionarie. Ciò ci può far intuire che la corrente nel diodo è controllata sostanzialmente da una ricombinazione: questa è la base del modello del diodo a controllo di carica. Manca però un fattore fondamentale: la dipendenza dal tempo. Dobbiamo dunque meglio definire il nostro modello a controllo di carica, introducendo una variazione della corrente nel tempo.

3.14 Modello a controllo di carica

Abbiamo finora considerato nel nostro diodo condizioni stazionarie; per poter studiare il modello a controllo di carica, dovremo risolvere l'equazione di continuità, in condizioni non stazionarie, e dunque con derivata temporale della densità di carica non nulla. Considerando ad esempio la corrente dovuta dalle sole lacune, nel lato n, vediamo che:

$$\frac{\partial p'_n}{\partial t} = -\frac{1}{q} \frac{\partial}{\partial x} J_p(x) - \frac{p'_n(x)}{\tau_p}$$

Risulta essere sempre e comunque valida l'ipotesi di quasi neutralità, per quanto riguarda i portatori minoritari, e dunque possiamo dire che la corrente nel diodo in un intorno della giunzione è puramente diffusiva:

$$J_p = -qD_p \frac{\partial p'_n}{\partial x} \implies D_p \frac{\partial^2 p'_n}{\partial x^2} = \frac{\partial}{\partial t} p'_n(x) + \frac{p'_n(x)}{\tau_p}$$

Integriamo dunque quest'espressione sull'asse delle ascisse spaziali x , entrambi i membri, a partire dal punto x_n in cui inizia la ricombinazione, fino ad una generica posizione spaziale $x > x_n$:

$$D_p \left[\frac{\partial p'_n}{\partial x} \Big|_x - \frac{\partial p'_n}{\partial x} \Big|_{x_n} \right] = \frac{\partial}{\partial t} \int_{x_n}^x p'_n(x) dx + \frac{1}{\tau_p} \int_{x_n}^x p'_n(x) dx$$

Moltiplicando ambo i membri per la carica fondamentale q , si ottiene che il termine con derivata temporale diventa:

$$\frac{\partial}{\partial t} \int_{x_n}^x qp'_n(x) dx + \frac{1}{\tau_p} \int_{x_n}^x qp'_n(x) dx = \frac{\partial}{\partial t} Q'_p(x) + \frac{1}{\tau_p} Q'_p(x)$$

Al primo membro, invece, per $x \rightarrow +\infty$, il primo termine si annulla, poichè la ricombinazione avrà annullato tutta la corrente di minoritari, e quindi, considerando l'ipoteso di diodo lungo, potremo annullare il primo termine, mentre il secondo sarà:

$$-qD_p \frac{\partial p'_n}{\partial x} \Big|_{x_n} = J_{p,diff}(x_n)$$

Dualmente, effettuando lo stesso processo con la corrente di elettroni derivanti dal lato n verso il lato p, si ottiene:

$$J_{n,diff}(-x_p) = \frac{\partial}{\partial t} Q'_n + \frac{1}{\tau_n} Q'_n$$

Poichè la corrente totale nel diodo J_{tot} sarà data dalla somma dei due contributi di corrente, il modello a controllo di carica sarà determinato dall'equazione:

$$J_{tot} = \frac{\partial}{\partial t} Q'_p + \frac{\partial}{\partial t} Q'_n + \frac{Q'_p}{\tau_p} + \frac{Q'_n}{\tau_n}$$

Terminiamo ora la trattazione di questo modello, cambiando argomento, e parlando di un altro modello del diodo a giunzione.

3.15 Modello di piccolo segnale

Il modello a controllo di carica è un sottocaso del modello ad ampio segnale, che non considera effetti elettrostatici (quali le capacità prima discusse); il modello ad ampio segnale è il modello che più interamente caratterizza un dispositivo, nella fattispecie il diodo a giunzione nel nostro caso, però come già accennato è molto difficile da utilizzare, a causa della sua forte non linearità. L'idea del modello di piccolo segnale è quella di studiare solo un preciso punto di lavoro del dispositivo, linearizzandolo in un intorno di quel punto, mediante una sorta di sviluppo in serie di Taylor troncata.

Dovremo studiare una funzione $I(V_a)$ per una certa $V_a = V_{op} + v_{SS}$: V_{op} (op sta per Operating Point) è la tensione del punto di lavoro a cui ci portiamo, al fine di sviluppare l'espressione della I ; v_{SS} è la rappresentazione dell'intorno, del piccolo segnale (SS sta per Small Signal), ed è la cosiddetta tensione di piccolo segnale.

Partiamo dalla corrente del diodo, dicendo che:

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

Poichè consideriamo $V_a = V_{op} + v_{SS}$, possiamo sviluppare fino al primo ordine, e quindi linearmente in un intorno del punto di lavoro la funzione della caratteristica come:

$$I(V_{op} + v_{SS}) = I_0 \left(e^{\frac{V_{op}}{V_T}} - 1 \right) + \frac{\partial}{\partial V_a} \left[I_0 e^{\frac{V_a}{V_T}} - 1 \right]$$

Di fatto la tensione V_{op} è una componente costante, mentre v_{SS} il suo intorno in cui possiamo considerare valida la linearizzazione effettuata. La derivata della corrente I_0 nel punto di lavoro V_{op} sarà:

$$\frac{\partial}{\partial V_a} \left[I_0 e^{\frac{V_a}{V_T}} - 1 \right] \Big|_{V_{op}} = \frac{I_0 e^{\frac{V_{op}}{V_T}}}{V_T} = g_d$$

Il termine g_d è anche detto conduttanza differenziale; ad essa associata è la resistenza differenziale, come il suo reciproco:

$$r_d = \frac{1}{g_d}$$

Spesso gli elettronici scambiano tra loro il termine $I_0 e^{\frac{V_{op}}{V_T}}$ con I_{op} . Considerando questa definizione della corrente I_{op} come corrente del punto di lavoro V_{op} , sarà:

$$I(V(t)) = I(V_{op} + v_{SS}(t)) \simeq I_{op} + g_d v_{SS}(t) = I(V_{op}) + i_{SS}(t)$$

In un intorno di V_{op} , è possibile dunque linearizzare il diodo, sostituendolo con la resistenza differenziale r_d : questa non sarà altro che la pendenza della retta osculatrice $I(V_a)$ nel punto di lavoro V_{op} , con un intorno di piccolo segnale v_{SS} . Il termine i_{SS} sarà dato dall'approssimazione di $I(V_a)$ al punto di lavoro V_{op} , $I(V_{op}) = I_{op}$; possiamo dunque definire ora la corrente di piccolo segnale come il semplice intorno-ordinata della caratteristica, ossia l'intera caratteristica considerata in $V_a = V_{op} + v_{SS}$ sottraendo la componente di offset $I(V_{op}) = I_{op}$:

$$i_{SS}(t) = I_{tot}(V_a) - I_{op} = g_d v_{ss}$$

Abbiamo così trattato in modo assolutamente completo il diodo, analizzato sotto il punto di vista del metodo di piccolo segnale. Cerchiamo ora di trattare brevemente allo stesso modo le capacità di svuotamento e di diffusione C_{DEP} e C_{DIF} , considerandole sempre ad un punto di funzionamento $V_a = V_{op} + v_{SS}$; possiamo dire che:

$$Q_{DEP}(V_{op} + v_{SS}(t)) \simeq Q_{DEP}(V_{op}) + \left. \frac{\partial Q_{DEP}}{\partial V_a} \right|_{V_{op}} \cdot v_{SS}$$

$$Q_{DIF}(V_{op} + v_{SS}(t)) \simeq Q_{DIF}(V_{op}) + \left. \frac{\partial Q_{DIF}}{\partial V_a} \right|_{V_{op}} \cdot v_{SS}$$

Utilizzando le espressioni operative già ricavate in precedenza delle cariche di svuotamento e di diffusione, si possono facilmente ricavare le espressioni delle capacità di svuotamento e di diffusione.

Il modello di piccolo segnale riduce dunque, in un intorno del punto di lavoro V_{op} , il circuito di ampio segnale ad un resistore (resistenza differenziale r_d) in parallelo a due capacità: questo, in serie con le resistenze serie dei lati della giunzione (che non hanno bisogno di essere linearizzate in quanto sono quantità costanti). Si noti che, per $V_a < 0$, ossia in polarizzazione inversa, la r_d avrà su di sé pochissima corrente, e diventerà così dell'ordine dei $M\Omega$, assimilabile dunque con un circuito aperto; la diffusione sarà pressochè nulla, poichè tutti i minoritari saranno solo nella zona di svuotamento, e la capacità relativa alla diffusione sarà circa nulla: $C_{DIF} \simeq 0$. Possiamo dunque ulteriormente semplificare il modello di piccolo segnale, per una tensione $V_a < 0$, con un condensatore (nella fattispecie, il condensatore con cui si modellano le zone di svuotamento, C_{DEP}), in serie alle due resistenze dovute ai lati del semiconduttore. Ovviamente, non si parla di una tensione di lavoro

negativa paragonabile a quella Zener, altrimenti la corrente è molto elevata e questa semplificazione non è fattibile.

3.16 Esercizio pratico

Data una giunzione p-n così descritta: $W_n = 200\mu m$, $W_p = 2\mu m$, $A_j = 100\mu m^2$, $\mu_p(N_A) = 300$, $\mu_n(N_A) = 100$, $\mu_n(N_D) = 600$, $\mu_p(N_D) = 150$, $\tau_p = 0,2\mu s$, $\tau_n = 1\mu s$, $V_a = 0,2V$, $N_D = 10^{18}$, $N_A = 10^{16}$, considerata con la convenzione inversa: lato n a sinistra, lato p a destra:

1. Disegnare il diagramma a bande quotato (ampiezza regioni di svuotamento, quasi-livelli di Fermi, etc.);
2. Disegnare e quotare il diagramma delle correnti;
3. Calcolare il modello di ampio segnale del diodo a giunzione;
4. Calcolare il modello di piccolo segnale considerando V_a come punto di lavoro.

3.16.1 Risoluzione

Disegnare il diagramma a bande quotato

Per il primo punto del problema potremmo adottare le solite strategie di risoluzione, oppure ragionare in un modo più fine: osservando i dati del problema, notiamo che $W_n \gg W_p$, e che $N_D \gg N_A$. Questi dati ci permettono di introdurre da un lato l'ipotesi di diodo corto, e dall'altro lato la forte differenza di drogaggio fa praticamente ridurre la zona di svuotamento ad una δ (idealizzando, poichè l'ampiezza della regione di svuotamento del lato n sarà molto inferiore a quella nel lato p). Possiamo trascurare nei nostri conti x_n , e dunque dire che l'ampiezza della regione di svuotamento sia:

$$x_d \simeq x_a = \sqrt{\frac{2\epsilon_S(\Phi_i - V_a)}{qN_A}}$$

Poichè ci manca per ora la barriera Φ_i , calcoliamola, considerando questa convenzione: poichè il lato n è un semiconduttore n^+ , ossia molto drogato, possiamo dire in buona approssimazione che il livello di Fermi coincida con il livello di partenza della banda di conduzione E_C . Dunque:

$$q\Phi_{S_n^+} \simeq q\chi$$

Confondiamo dunque senza problemi il lavoro di estrazione con l'affinità elettronica del silicio n^+ (che è un parametro costante, e noto).

Possiamo dunque calcolare la barriera nel lato p, confrontando il livello di Fermi intrinseco in zona di neutralità, con il quasi-livello di Fermi, aggiungendo al conto metà dell'energy gap:

$$q\Phi_i = \frac{E_g}{2} + E_{F_i}(+\infty) - E_{F_p} = \frac{E_g}{2} + kT \ln \frac{N_A}{N_i}$$

Si può dunque trascurare il termine al lato n, e ripetere il ragionamento geometrico fatto sul diagramma a bande della giunzione già esposto nella teoria.

Disegnare e quotare il diagramma delle correnti

Passiamo al punto successivo: il calcolo delle correnti nella giunzione. Abbiamo una situazione molto particolare, poichè abbiamo a che fare con un diodo metà lungo, e metà corto; calcoliamo dunque, in entrambi i lati, la distribuzione dei minoritari, al fine di calcolare le correnti di diffusione come derivata della distribuzione. La distribuzione di lacune iniettate nel lato n (ricordando che la nostra convenzione abituale è invertita) sarà:

$$p'_n(x) = p'_n(-x_n) e^{\frac{x+x_n}{L_p}}$$

Per l'altro lato le cose si fanno più difficili, in quanto le condizioni al contorno abituali per la risoluzione dell'equazione di continuità non valgono. Da un altro esercizio teorico abbiamo però già ricavato il risultato fondamentale considerante le condizioni al contorno al fine di studiare il semiconduttore corto, e quindi possiamo dire che:

$$n'_p(x) = n'_p(x_p) \frac{\sinh \frac{L-x}{L_n}}{\sinh \frac{L}{L_n}}$$

Poichè nel nostro caso consideriamo il fatto che la lunghezza effettiva del lato parta aldilà della regione di svuotamento, e quindi $L = W_p - x_p$, possiamo dire che, poichè $x \gg x_p$ generalmente,

$$n'_p(x) = n'_p(x_p) \frac{\sinh \frac{W_p-x}{L_n}}{\sinh \frac{W_p-x_p}{L_n}}$$

Partendo dalle due espressioni delle correnti ora ricavate possiamo, mediante la loro derivazione, ricavare le espressioni operative delle correnti:

$$J_{p,diff}(x) = -qD_p \frac{\partial p'_n(x)}{\partial x} = \frac{-qD_p n_i^2}{N_D L_p} \left(e^{\frac{V_a}{V_T}} - 1 \right) e^{\frac{x+x_n}{L_p}}$$

$$J_{n,diff}(x) = -qD_n \frac{\partial n'_p(x)}{\partial x} = \frac{-qD_n n_i^2}{N_A L_n} \left(e^{\frac{V_a}{V_T}} - 1 \right) \frac{-\cosh \frac{W_p-x}{L_n}}{\sinh \frac{W_p-x_p}{L_n}}$$

Ciò che si potrebbe notare facendo il grafico delle due correnti, è che $J_{p,diff}$ è una corrente del tutto trascurabile; inoltre, la decrescita che si potrebbe analizzare osservando l'andamento di $J_{n,diff}$ è altrettanto trascurabile, quindi di fatto la corrente totale nel diodo è regolata dalla sola $J_{n,diff}$, che è una corrente praticamente costante. Possiamo a questo punto chiederci quale sia l'andamento della caratteristica statica del diodo; dal momento che abbiamo usato per esperienza (e tornerà utile in seguito) questa convenzione opposta della giunzione, invertiamo il senso di percorrenza degli assi, per comodità, considerando $-I_{tot}$, ed i segni $-$ nelle espressioni potranno essere annullati:

$$-I_{tot} = A_j (J_{p,diff}(-x_n) + J_{n,diff}(x_p))$$

$$I = A_j \left(\frac{-qD_p n_i^2}{N_D L_p} + \frac{qD_n n_i^2}{N_A L_n} \frac{\cosh \frac{W_p-x_p}{L_n}}{\sinh \frac{W_p-x_p}{L_n}} \frac{W_p-x_p}{L_n} \right) \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

In realtà, però, il lato è corto, quindi possiamo sviluppare seno e coseno iperbolico, come rispettivamente argomento del seno e 1:

$$n'_p(x) \simeq n'_p(x_p) \frac{W_p-x_p}{L_n}$$

In un dispositivo di questo tipo, si annulla l'usuale andamento della ricombinazione: l'andamento è approssimabile a lineare, ed anzi costante, poichè il lato corto annulla gli effetti ricombinativi. Nel lato n invece l'unica particolarità introdotta nell'esercizio è un forte drogaggio, che di fatto non modifica il fenomeno di ricombinazione (se non rendendolo più rapido). Nel transistor bipolare, l'uso del lato corto sarà proprio ciò che servirà per gestire le correnti nel modo più corretto, in modo da aumentare l'effetto transistor.

Calcolare il modello di ampio segnale del diodo a giunzione

Terminiamo l'esercizio calcolando i modelli di ampio e piccolo segnale: sappiamo che le nostre ipotesi di lato corto ci permettono di dire che:

$$C_{DEP} = \frac{\epsilon_S}{x_d} A_j \simeq \frac{\epsilon_S}{x_p(V_a)} A_j$$

La capacità di diffusione si potrà calcolare mediante il modello a controllo di carica, calcolando la derivata parziale rispetto alla variazione di tensione di polarizzazione esterna V_a ; in questo caso l'ipotesi semplificatrice sarà quella di grosso drogaggio: poichè il drogaggio è molto più elevato al lato n che al lato p, possiamo dire che il contributo della carica Q'_p sarà trascurabile.

$$C_{DIF} \simeq A_j \left| \frac{\partial Q'_n}{\partial V_a} \right|$$

Possiamo inoltre, come già detto, approssimare la distribuzione degli elettroni ad una retta, a causa dello sviluppo in polinomio di Taylor del seno iperbolico; possiamo dunque ulteriormente semplificare il modello a controllo di carica, calcolando l'area del triangolo il cui lato è il lato del diodo, e l'altezza la carica in questione è la distribuzione di carica; quindi:

$$Q'_n = \frac{-qn_i^2(W_p - x_p)}{2N_A} \left(e^{\frac{V_a}{V_T}} - 1 \right)$$

Considerando il modulo della derivata parziale di quest'espressione rispetto alla tensione V_a , si otterrà:

$$\left| \frac{\partial Q'_n}{\partial V_a} \right| = \frac{qn_i^2(W_p - x_p)}{2N_A V_T} e^{\frac{V_a}{V_T}}$$

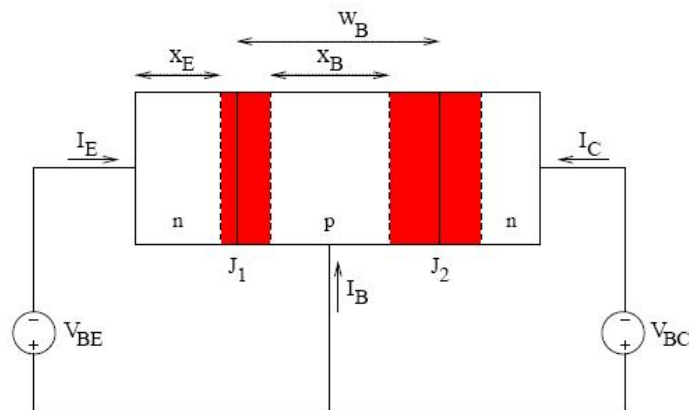
Il modello di piccolo segnale sarà una semplice linearizzazione di questo, considerando il punto di lavoro scelto (ossia la V_a).

Capitolo 4

Il transistor bipolare

Il transistor bipolare è una geniale applicazione della giunzione p-n: esso è composto da una doppia giunzione n-p-n (oppure p-n-p); ma qual è il senso di fare una giunzione doppia?

Sappiamo che, per poter aumentare la corrente inversa di saturazione I_0 , servirebbe un'iniezione di portatori mediante via termica, od ottica, oppure utilizzare iniettori di altro tipo: la giunzione p-n. Abbiamo infatti studiato che una giunzione è un iniettore di cariche, poichè è possibile iniettare portatori mediante essa, e quindi incrementare la corrente inversa di saturazione. Possiamo dunque suddividere in tre principali zone questo dispositivo: emettitore, base, collettore.



Consideriamo come emettitore per ora un semiconduttore drogato n, come base un semiconduttore drogato p, come collettore un altro semiconduttore drogato n. L'emettitore emette elettroni nel lato p ad esso adiacente, ossia alla base, e questo mediante l'applicazione di una polarizzazione diretta. Da p a n, vi sarà una polarizzazione inversa, ma gli elettroni, giunti in abbondanza

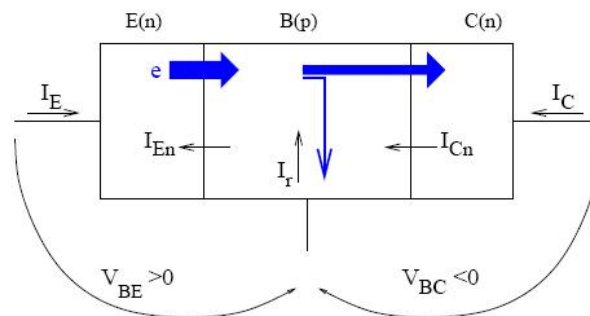
dall'emettitore, andranno verso il collettore passando dalla base, mediante una corrente di tipo diffusivo. Riassumendo, la barriera bassa dall'emettitore E alla base B permette il passaggio di elettroni; dalla base B al collettore C questi scivoleranno dalla barriera, arrivando senza problemi alla meta. Ma qual è il senso dell'utilizzo di questo dispositivo?

Si noti che la corrente di elettroni dall'emettitore E, alla base B è regolabile dalla tensione sulla prima giunzione; ciò che poi capita, è che si possa controllare la corrente che viene iniettata dalla base al collettore (poichè essa dipende da quanta ne è stata iniettata dalla base, e questa è regolabile dalla tensione tra E e B), ma senza doverla riprendere tutta: di solito dobbiamo pagare tutta la carica inviata da un semiconduttore ad un altro, con altrettanta carica dal secondo semiconduttore verso il primo, come abbiamo visto nei meccanismi alla base del diodo a giunzione; di fatto, non sarà necessaria una grossa contropartita di corrente dalla base all'emettitore, o meglio ne sarà necessaria una quantità minima, come vedremo: vi sarà una grossa corrente da E a B, ma infinitesima da B ad E (in un buon transistor, alle condizioni che introdurremo). L'astuzia si basa sul fatto che, tendenzialmente, gli elettroni derivanti dall'emettitore E verso la base B non dovranno ricombinarsi, ma andare verso il collettore C, dove saranno comunque maggioritari (poichè il collettore è un semiconduttore drogato tipo n).

Riassumendo, il transistor bipolare permette di controllare una corrente tra emettitore e collettore senza dover pagare con una corrente troppo ingente tra base ed emettitore, bensì solo con una minima corrente inversa.

4.1 Analisi delle correnti

Cerchiamo ora di rappresentare in modo qualitativo il modello delle correnti in un transistor bipolare; consideriamo la giunzione da emettitore E a base B polarizzata direttamente, e da base B a collettore C polarizzata inversamente: $V_{BE} > 0$, $V_{BC} < 0$.



Nel dispositivo in studio, esistono cinque correnti; analizziamole una per una:

1. I_{nE} è una corrente di diffusione di elettroni che vengono iniettati dall'emettitore E alla base B;
2. I_{nC} è la frazione di I_{nE} che non si è ricombinata nella base; essa in un dispositivo ben funzionante deve essere più elevata possibile, poichè la ricombinazione in base deve essere minima, e poichè essa sarà destinata a terminare nel collettore; I_{nE} ed I_{nC} sono le due correnti fondamentali ai fini del funzionamento del transistor bipolare;
3. I_{rB} è una corrente di diffusione di lacune iniettata nella base a partire dall'emettitore, che va a ricombinarsi interamente: è una frazione di I_{nE} , che deve essere limitata, in quanto essa limita la corrente principale I_{nC} ;
4. I_{pB} è una corrente di diffusione di lacune iniettate nell'emettitore a partire dalla base; questa corrente deve essere piccola, in un buon dispositivo, poichè dipende dalla ricombinazione di I_{nE} , che deve essere minore possibile per garantire un buon funzionamento;
5. I_{pC} è una corrente inversa di lacune diretta dal collettore C verso la base B, molto ridotta, e spesso trascurabile nello studio dei transistori.

Ragioniamo ora su queste correnti: la corrente di ingresso nell'emettitore dalla batteria, I_E , è data da:

$$I_E = I_{nE} + I_{pB}$$

La corrente I_E è la somma delle due correnti I_{nE} e I_{pB} , la seconda derivante dalla base. Poichè una delle due correnti principali del transistor è quella totale che va dall'emettitore verso la base, la corrente da base a emettitore, I_{pB} , sarà da limitare. Si definisce sull'emettitore un parametro γ_F che determina l'efficienza del transistor, all'emettitore (F sta per Forward):

$$\gamma_F = \frac{I_{nE}}{I_{nE} + I_{pB}}$$

Il parametro γ_F è dunque detto efficienza di emettitore, e in un buon dispositivo deve tendere a 1, poichè I_{pB} dovrà tendere a 0; ciò è tecnologicamente fattibile, aumentando il drogaggio del lato emettitore n (come si può facilmente verificare dalla legge della giunzione). Possiamo dunque dire che:

$$I_{nE} = \gamma_F I_E$$

Ci interessa ora sapere, di questa $\gamma_F I_E$, quanta finirà nel collettore, e quindi quanta non verrà ricombinata. Possiamo ricavare facilmente che:

$$I_{nC} = I_{nE} - I_{rB}$$

Possiamo dunque definire il parametro α_T come:

$$\alpha_T = \frac{I_{nE} - I_{rB}}{I_{nE}}$$

Anche in questo caso, in un buon dispositivo, α_T deve tendere a 1. In questo caso, tecnologicamente il trucco per far tendere a 1 il coefficiente α_T , è quello di usare una base corta: se la lunghezza del semiconduttore $W_B \ll L_n$, la ricombinazione non avrà il tempo di agire efficacemente sulla carica, e dunque $\alpha_T \rightarrow 1$.

La corrente $I_{pC} \simeq 0$ e quindi possiamo considerarla ora come ora trascurabile.

Definiamo α_F il prodotto dei due coefficienti di efficienza:

$$\alpha_F = \gamma_F \alpha_T$$

La corrente del collettore, I_C , sarà pari a:

$$I_C = -\alpha_F I_E$$

Possiamo considerare il transistor come un nodo generalizzato, e dunque studiare mediante le leggi di Kirchhoff le tre correnti su emettitore, collettore, base: anche la base infatti avrà una corrente, che andrà ridotta mediante i metodi tecnologici accennati, in un dispositivo valido:

$$I_E + I_B + I_C = 0$$

Considerando l'espressione precedentemente vista in funzione di I_E , sostituiamo nella relazione appena trovata:

$$I_E = -\frac{I_C}{\alpha_F} \implies -\frac{I_C}{\alpha_F} + I_C + I_B = 0$$

Quindi:

$$I_B = \frac{I_C(1 - \alpha_F)}{\alpha_F} \implies \frac{I_C}{I_B} = \frac{\alpha_F}{1 - \alpha_F} = \beta_F$$

Questo parametro, β_F , è detto guadagno in corrente. Abbiamo visto che α_F è il prodotto di due numeri vicini ad 1, e dunque β_F potrebbe essere un numero molto elevato (da 100 a 300 o anche molto più in un buon dispositivo).

Possiamo dire che:

$$I_C = \beta_F I_B$$

Con il prezzo di una corrente di base di soli pochi mA, è possibile controllare correnti molto elevate, modulabili mediante la tensione tra emettitore e base, V_{BE} . Possiamo considerare I_B come un prezzo da pagare, al fine di poter liberamente modulare mediante la tensione la corrente del transistor. Passiamo ora ad un calcolo più quantitativo, al fine di studiare le correnti.

4.2 Calcolo delle correnti

Consideriamo alcune ipotesi preliminari: scegliamo come origine del nostro sistema di assi cartesiani rappresentante la posizione spaziale sul dispositivo il punto di inizio della regione di svuotamento sulla base, alla giunzione con l'emettitore E; consideriamo inoltre la larghezza della base W_B uguale alla distanza tra le due regioni di svuotamento. Questa seconda ipotesi è molto approssimativa, in quanto di fatto la distanza tra le due regioni di svuotamento varia con le regioni di svuotamento stesse, le quali sono modulabili mediante le tensioni esterne che controllano il dispositivo; poichè però utilizziamo l'ipotesi di lato corto, possiamo considerare come non troppo errata questa approssimazione, anche se di fatto confondiamo una grandezza variabile con la tensione con una costante.

I quattro punti del dispositivo interessanti per il nostro studio in questo ambito saranno: $-x_{dE}$, 0, W_B , x_{dC} , ossia il punto di inizio della zona di svuotamento della giunzione tra emettitore e base, dal lato della base, l'origine prima definita, la posizione della giunzione con il lato n (collettore), confusa con l'inizio della relativa zona di svuotamento, e il punto di inizio della regione di svuotamento nel collettore. Studiamo dunque le leggi delle giunzioni, applicandole nei suddetti punti; consideriamo N_{DE} come la concentrazione di drogaggio tipo n N_D nell'emettitore E, e N_{AB} come la concentrazione di drogaggio tipo p N_A nella base B:

$$p'_n(-x_{dE}) = \frac{n_i^2}{N_{DE}} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

$$n'_p(0) = \frac{n_i^2}{N_{AB}} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

$$n'_p(W_B) = \frac{n_i^2}{N_{AB}} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

In un transistor, i fenomeni di interesse avvengono soprattutto nella base: in essa, ci dedicheremo allo studio dettagliato dei portatori di carica. Supponiamo di conoscere come condizioni al contorno i portatori nei punti 0 e W_B : essi saranno le grandezze appena ricavate mediante le leggi della giunzione. Applichiamo ora alla base il modello matematico, ricordando che ci troviamo in ipotesi di lato corto.

Consideriamo dunque le equazioni di continuità, nella base; si noti che le nostre due equazioni al contorno sono però entrambe non nulle, quindi avremo a che fare con conti algebrici di difficoltà non indifferente.

Dal modello matematico, sappiamo che:

$$D_n \frac{\partial^2 n'_p(x)}{\partial x^2} = \frac{n'_p(x)}{\tau_n}$$

$$D_p \frac{\partial^2 p'_n(x)}{\partial x^2} = \frac{p'_n(x)}{\tau_p}$$

La soluzione generale delle equazioni differenziali avrà come al solito una forma del tipo (consideriamo la diffusione dentro alla base di elettroni, e quindi l'eccesso di elettroni $n'_p(x)$):

$$n'_p(x) = Ae^{\frac{x}{L_n}} + Be^{-\frac{x}{L_n}}$$

Applichiamo ora le due condizioni al contorno, che abbiamo precedentemente studiato:

$$n'_p(0) = \frac{n_i^2}{N_{AB}} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \implies A + B$$

$$n'_p(W_B) = \frac{n_i^2}{N_{AB}} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) \implies Ae^{\frac{W_B}{L_n}} + Be^{-\frac{W_B}{L_n}}$$

Ora, possiamo ricavare dalla prima delle due condizioni:

$$B = n'_p(0) - A$$

Sostituendo nella seconda, otteniamo che:

$$n'_p(W_B) = Ae^{\frac{W_B}{L_n}} + (n'_p(0) - A)e^{-\frac{W_B}{L_n}}$$

Da ciò, possiamo ricavare un'espressione del coefficiente A:

$$A = \frac{n'_p(W_B) - n'_p(0)e^{-\frac{W_B}{L_n}}}{e^{\frac{W_B}{L_n}} - e^{-\frac{W_B}{L_n}}}$$

Sostituiamo ora l'espressione di A nella prima equazione, trovando il valore di B:

$$B = n'_p(0) - A = \frac{n'_p(0)e^{\frac{W_B}{L_n}} - n'_p(0)e^{-\frac{W_B}{L_n}} - n'_p(W_B) + n'_p(0)e^{-\frac{W_B}{L_n}}}{e^{\frac{W_B}{L_n}} - e^{-\frac{W_B}{L_n}}} =$$

$$= \frac{n'_p(0)e^{\frac{W_B}{L_n}} - n'_p(W_B)}{2 \sinh \frac{W_B}{L_n}}$$

Ora, considerando la soluzione generale dell'equazione differenziale, inseriamo le condizioni al contorno ricavate mediante i nostri artifici, e ricaviamo:

$$n'_p(x) = \frac{n'_p(W_B)e^{\frac{x}{L_n}} - n'_p(0)e^{-\frac{W_B+x}{L_n}} + n'_p(0)e^{\frac{W_B-x}{L_n}} - n'_p(W_B)e^{-\frac{x}{L_n}}}{2 \sinh \frac{W_B}{L_n}}$$

$$n'_p(x) = n'_p(0) \frac{\sinh \frac{W_B-x}{L_n}}{\sinh \frac{W_B}{L_n}} + n'_p(W_B) \frac{\sinh \frac{-x}{L_n}}{\sinh \frac{W_B}{L_n}}$$

Derivando l'espressione dei portatori così ricavata, possiamo studiare i flussi delle correnti nella base. Si noti che le correnti che raggiungono il collettore, lo faranno mediante un meccanismo diffusivo, come già detto, regolato dalle espressioni appena ricavate.

4.3 Modello statico del transistor bipolare

Abbiamo già detto che il transistor bipolare è un dispositivo dominato dalle correnti di diffusione di minoritari in base; consideriamo d'ora in avanti una convenzione semplificata: dal momento che non abbiamo a che fare con altro che regioni di svuotamento, consideriamo $x_{dE} = x_E$, 0, e il punto x_C come punto di fine della regione di svuotamento nel collettore.

Le equazioni dei portatori nelle tre regioni, come abbiamo già ricavato, saranno: nell'emettitore E:

$$p'_{n,E}(x) = p'_n(-x_E)e^{\frac{x+x_E}{L_p}} = \frac{n_i^2}{N_{DE}} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) e^{\frac{x+x_E}{L_p}}$$

Nella base B:

$$n'_{p,B}(x) = n'_p(0) \frac{\sinh \frac{W_B-x}{L_n}}{\frac{W_B}{L_n}} + n'_p(W_B) \frac{\sinh \frac{x}{L_n}}{\frac{W_B}{L_n}}$$

Nel collettore C:

$$p'_{n,C}(x) = \frac{n_i^2}{N_{DC}} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) e^{-\frac{x-x_C}{L_p}}$$

Partendo dalle precedenti espressioni, interessandoci nella fatispecie di quelle concernenti la base B e l'emettitore E, ricaviamo le correnti $I_{p,diff}(x)$ e $I_{n,diff}(x)$, che verranno tra poco riutilizzate in un'applicazione pratica; ricordando le relazioni di Einstein per il calcolo delle correnti a partire dai portatori:

$$I_n = qAD_n \frac{\partial n}{\partial x}$$

$$I_p = -qAD_p \frac{\partial p}{\partial x}$$

Ricaviamo:

$$I_{nB,diff}(x) = \frac{-qA_E n_i^2 \cosh \frac{W_B - x}{L_n}}{N_{AB} L_n \sinh \frac{W_B}{L_n}} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + \frac{qA_E n_i^2 \cosh \frac{x}{L_n}}{N_{DE} L_n \sinh \frac{W_B}{L_n}} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

$$I_{pE,diff}(x) = \frac{-qA_E n_i^2 e^{\frac{x-x_E}{L_p}}}{L_p N_{DC}} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

Per quanto riguarda invece la corrente di lacune nel collettore, C possiamo dire che:

$$I_{pC,diff}(x) = \frac{qA_E n_i^2 e^{\frac{x-x_C}{L_p}}}{N_{DC} L_p} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

Queste espressioni delle correnti, al variare della posizione spaziale x , ci torneranno presto utili: dal punto di vista dei soli emettitore o collettore, il transistor non avrebbe infatti molte differenze da una comunissima giunzione; poichè la base è già un caso diverso, in quanto è un lato corto, e si comporta in modo molto anomalo per una giunzione p-n. Tenendo conto di questi fatti, vogliamo determinare un modello statico per il transistor bipolare; sarà più complesso rispetto al modello della giunzione, in quanto questa volta si ha a che fare con tre correnti (di cui possiamo semplicemente studiarne due, in quanto comunque la terza è data dalla somma delle altre due, cambiata di segno):

$$\begin{cases} I_C(V_{BE}; V_{BC}) \\ I_E(V_{BE}; V_{BC}) \end{cases}$$

Dal momento che, come detto:

$$I_B = -I_C - I_E$$

Ricavando le due caratteristiche indicate alla parentesi graffa, il modello statico del transistor bipolare sarà di fatto completo. Come nel caso della giunzione p-n, possiamo considerare, al posto delle correnti in zone lontane dalla giunzione, le correnti valutate ai punti di inizio della regione di svuotamento: sappiamo infatti che, lontano dalla giunzione, la corrente sarà del tutto equivalente, in quanto sarà la corrente in grado di ricombinare i portatori giunti dalle regioni di svuotamento, e quindi esattamente le correnti valutate nei punti di inizio della regione di svuotamento, punti in cui non sono di fatto ancora iniziati fenomeni di ricombinazione. Per quanto riguarda la caratteristica I_E :

$$I_E = I_{p,diff}(-x_E) + I_{n,diff}(0)$$

Studiando la caratteristica statica della giunzione p-n, avevamo fatto lo stesso ragionamento: usando le espressioni delle correnti appena ricavate come derivazione da quelle dei portatori, otterremo:

$$\begin{aligned} I_E &= A_E(-q)D_p \left. \frac{\partial p'_n(x)}{\partial x} \right|_{x_E} + A_E(+q)D_n \left. \frac{\partial n'_p(x)}{\partial x} \right|_0 = \\ &= \frac{-qA_ED_p n_i^2}{N_{DE}L_p} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) e^{\frac{-x_E+x_E}{V_T}} + \frac{-qA_ED_n n_i^2}{N_{AB}L_n} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \frac{\cosh \frac{W_B}{L_n}}{\sinh \frac{W_B}{L_n}} + \\ &\quad + \frac{-qA_ED_n n_i^2}{N_{AB}L_n} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) \frac{\cosh(0)}{\sinh \frac{W_B}{L_n}} \end{aligned}$$

Ordinando l'espressione:

$$\begin{aligned} I_E &= \left(\frac{-qA_ED_p n_i^2}{N_{DE}L_p} + \frac{-qA_ED_n n_i^2}{N_{AB}L_n} \operatorname{cotgh} \frac{W_B}{L_n} \right) \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + \\ &\quad + \frac{qA_ED_n n_i^2}{N_{AB}L_n \sinh \frac{W_B}{L_n}} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) \end{aligned}$$

In maniera del tutto analoga, si può effettuare lo studio di I_C : possiamo infatti vedere facilmente che:

$$I_C = I_{n,diff}(W_B) + I_{p,diff}(x_C)$$

Riprendendo anche ora le espressioni delle correnti appena ricavate, possiamo ottenere che:

$$\begin{aligned}
I_C &= A_E(+q)D_n \left. \frac{\partial n'_p(x)}{\partial x} \right|_{W_B} + A_E(-q)D_p \left. \frac{\partial p'_n(x)}{\partial x} \right|_{x_C} = \\
&= qA_E D_n \frac{n_i^2}{N_{AB}} \frac{1}{\sinh \frac{W_B}{L_n}} \left(-\frac{\cosh \frac{W_B - W_B}{L_n}}{L_n} \right) \left(e^{\frac{v_{BE}}{V_T}} - 1 \right) + \\
&+ qA_E D_n \frac{n_i^2}{N_{AB}} \frac{\cosh \frac{W_B}{L_n}}{\sinh \frac{W_B}{L_n}} \frac{1}{L_n} \left(e^{\frac{v_{BC}}{V_T}} - 1 \right) + \frac{qA_E D_p n_i^2}{N_{DE} L_p} \left(e^{\frac{v_{BC}}{V_T}} - 1 \right)
\end{aligned}$$

Ordinando:

$$\begin{aligned}
I_C &= \frac{qA_E D_n n_i^2}{N_{AB} \sinh \frac{W_B}{L_n} L_n} \left(e^{\frac{v_{BE}}{V_T}} - 1 \right) + \\
&+ \left(\frac{-qA_E D_n n_i^2}{N_{AB} L_n} \operatorname{cotgh} \frac{W_B}{L_n} + \frac{-qA_E D_p n_i^2}{N_{DC} L_p} \right) \left(e^{\frac{v_{BC}}{V_T}} - 1 \right)
\end{aligned}$$

Confrontando le due espressioni, vediamo che esse, ossia il modello statico del transistor bipolare, presentano tra loro alcune similitudini.

4.3.1 Equazioni di Ebers-Moll

Le equazioni appena ricavate del modello statico di un transistor bipolare sono le cosiddette equazioni di Ebers-Moll, in onore agli scopritori del modello statico del dispositivo in trattazione. Si è soliti esprimere queste equazioni nella seguente forma:

$$\begin{aligned}
I_E &= a_{1,1} \left(e^{\frac{v_{BE}}{V_T}} - 1 \right) + a_{1,2} \left(e^{\frac{v_{BC}}{V_T}} - 1 \right) \\
I_C &= a_{2,1} \left(e^{\frac{v_{BE}}{V_T}} - 1 \right) + a_{2,2} \left(e^{\frac{v_{BC}}{V_T}} - 1 \right)
\end{aligned}$$

I quattro coefficienti in questione sono:

$$\begin{aligned}
a_{1,1} &= \frac{-qA_E D_p n_i^2}{N_{DE} L_p} - \frac{qA_E D_n n_i^2}{N_{AB} L_n} \operatorname{cotgh} \frac{W_B}{L_n} \\
a_{1,2} &= \frac{qA_E D_n n_i^2}{N_{AB} L_n \sinh \frac{W_B}{L_n}}
\end{aligned}$$

$$a_{2,1} = \frac{qA_E D_n n_i^2}{N_{AB} L_n \sinh \frac{W_B}{L_n}}$$

$$a_{2,2} = \frac{qA_E D_n n_i^2}{N_{AB} L_n} \operatorname{cotgh} \frac{W_B}{L_n} - \frac{qA_E D_p n_i^2}{N_{DC} L_p}$$

Si noti che questo modello risulti essere corretto a meno di una variazione di W_B (che abbiamo ipotizzato coincidente con la distanza tra le due zone di svuotamento nella base); questo modello trascura dunque l'effetto delle tensioni sulla lunghezza del lato di base B, ossia sull'appena citato W_B . Si noti però un'altra cosa: abbiamo iniziato a sviluppare questo modello considerando esclusivamente il caso di $V_{BE} > 0$, $V_{BC} < 0$; abbiamo però ricavato equazioni che risultano essere valide indipendentemente dalle due tensioni (a meno di particolari fenomeni quali il breakdown delle giunzioni, o cose di questo tipo), esattamente come nello studio del diodo a giunzione: avevamo infatti al tempo considerato la giunzione in stato di polarizzazione diretta, ma ottenuto equazioni valide anche per la polarizzazione inversa. Siamo dunque in grado di studiare, mediante le equazioni di Ebers-Moll, qualsiasi coppia di tensione ($V_{BE}; V_{BC}$). Consideriamo dunque le quattro (che poi vedremo diventare cinque) zone di tensione:

- $V_{BE} > 0$, $V_{BC} < 0$: è la zona che abbiamo sinora studiato: il transistor provoca fenomeni di guadagno della corrente, regolati dai vari coefficienti γ_F , α_T , α_F , β_F , dove la F significa Forward: italianizzando, possiamo tradurre con Diretta. La regione, che è la regione di funzionamento più tipica del transistor bipolare, è detta regione attiva diretta.
- $V_{BE} < 0$, $V_{BC} > 0$: di fatto, il collettore diventa l'emettitore, e l'emettitore diventa il collettore. Fisicamente, il significato, il funzionamento del dispositivo sarà del tutto uguale a prima, a meno del seguente fatto: il transistor è progettato sia geometricamente che sotto il punto di vista dell'iniezione di impurità drogante in modo da funzionare con l'emettitore e il collettore in grado di svolgere i propri relativi compiti nativi, quindi il funzionamento vi sarà, ma il guadagno in corrente nella zona, β_R (R da Reverse, italianizzando Inversa), sarà una decina di volte inferiore rispetto al guadagno in zona diretta. Questa zona è detta regione attiva inversa.
- $V_{BE} < 0$, $V_{BC} < 0$: nella base arriverà dall'emettitore un contributo di corrente infinitesimo, ossia la corrente di polarizzazione inversa. Dal

momento che anche la tensione che modula le correnti è negativa, dunque, il dispositivo sarà di fatto su OFF: questa zona è detta regione di interdizione, e nel transistor non passa corrente.

- $V_{BE} > 0, V_{BC} > 0$: è la zona più complicata da studiare, delle quattro (anche se vedremo che essa si sdoppierà): essa è detta regione di saturazione: entrambe le giunzioni sono polarizzate direttamente, ma in questo caso esistono due ulteriori sottopossibilità:
 - $V_{BE} > V_{BC}$: il diagramma a bande ricorda quello della regione attiva diretta, ma senza avere un guadagno in corrente: poichè la tensione tra base ed emettitore è superiore a quella tra base e collettore, l'iniezione in base dalla parte dell'emettitore partirà da un punto a potenziale più elevato rispetto a quella da base a collettore, quindi di fatto avremo una corrente, nella nostra convenzione, da sinistra verso destra.
 - $V_{BE} < V_{BC}$: si ha una situazione duale a quella appena descritta: il potenziale tra base e collettore è maggiore a quello tra base ed emettitore, dunque si può dire che la corrente vada da destra verso sinistra, e si ha comunque un guadagno pressochè nullo (caso duale a prima, in tutto e per tutto, come già accennato)

Esiste un guadagno in corrente minimo, però non ha senso definirlo, sia per il fatto che esso è veramente trascurabile (e quindi si può approssimare ad un regime di saturazione), non presentando un effetto transistor, sia per motivi che analizzeremo in seguito.

In regione attiva (diretta o inversa che sia) il transistor bipolare si comporta come un amplificatore di corrente: il guadagno, definibile e modificabile mediante la tensione di modulazione, è costante, e permette di utilizzare l'effetto transistor; in zona di interdizione o di saturazione il transistor si usava, un tempo, come interruttore OFF/ON (una sorta di diodo); la prima elettronica digitale si basava sul transistor bipolare, sfruttando anche queste zone di interdizione e saturazione; ora la tecnologia MOS ha soppiantato la tecnologia BJT, a causa dei progressi tecnologici fatti in tal senso.

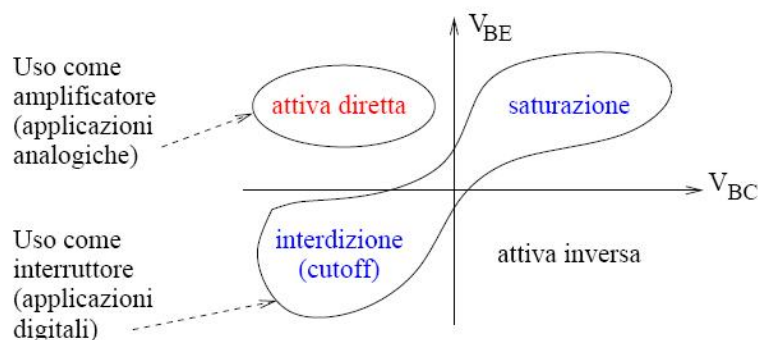
4.4 Modello circuitale statico del transistor bipolare

Abbiamo parlato delle equazioni di Ebers-Moll, rappresentanti il modello statico del transistor bipolare. Cerchiamo ora di estrarre da esse, mediante

semplificazioni in ciascuna zona, un modello circuitale del dispositivo elettronico, partendo dall'idea che esso è interpretabile come un nodo generalizzato in cui convergono tre lati: Base, Emittitore, Collettore. Abbiamo quattro regioni sostanzialmente da distinguere: occupiamoci di ciascuna di esse.

- Regione attiva diretta: abbiamo un nodo generalizzato con B, C, ed E. Sappiamo che, a partire dalla corrente di base I_B , si ha un guadagno in corrente. Circuitualmente, l'elettrotecnica ci insegna che un buon modello per rappresentare un guadagno di questo genere è il generatore pilotato di corrente: considerando il guadagno β_F , possiamo dire che tra collettore C ed emittitore E vi sarà una corrente $I_C = \beta_F I_B$, esprimibile come un generatore pilotato dal pilota I_B , corrente che dalla base B va verso l'emittitore E. Dall'altra parte, si può considerare la totale analogia con una comune giunzione p-n polarizzata direttamente, e quindi con un diodo; ricordiamo che, in questo modello, si introduceva un generatore di tensione V_γ , dove in questo caso la nostra tensione di attivazione del transistor coincide con V_{BE} . Questo modello presenta alcune debolezze, che però possiamo per ora tralasciare.
- Regione attiva inversa: situazione del tutto uguale alla precedente, considerando semplicemente le seguenti variazioni: al posto dell'emittitore E vi sarà il collettore C, al posto del collettore C vi sarà l'emittitore E, il guadagno in corrente sarà β_R , ed in questo caso la giunzione assimilabile come diodo polarizzato direttamente sarà quella tra base e collettore, dunque il nostro generatore V_γ sarà la tensione di attivazione V_{BC} .
- Regione di interdizione: non essendovi contatto tra le tre regioni, possiamo immaginare i tre nodi separati tra loro a due a due.
- Regione di saturazione: abbiamo due giunzioni polarizzate direttamente, e quindi di fatto avremo due diodi in stato ON, con le relative V_γ . Una sarà la tensione tra base ed emittitore, V_{BE} , l'altra tra emittitore e collettore, $V_{CE} = V_{BE} - V_{BC}$.

Possiamo dunque dire, per quanto scontato sia, che i modelli circuitali valgono se e solo se ci troviamo nelle opportune condizioni di tensione: si tratta di quattro casi singoli, che però devono essere considerati nel punto in cui sono stati definiti.

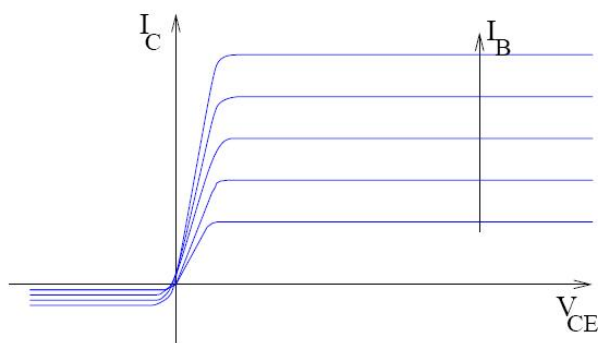


4.4.1 Andamento della caratteristica statica

Abbiamo terminato la descrizione del funzionamento del transistor bipolare nelle quattro regioni, ci dedichiamo ora a studiarne l'andamento grafico, facendo alcune osservazioni che ci mancano.

Avendo diverse opzioni da rappresentare, ne scegliamo una fondamentale: l'uscita di emettitore comune, $I_C(V_{CE})$, dove $V_{CE} = V_{BE} - V_{BC}$. Algebricamente, possiamo fare un'osservazione: sappiamo che $V_{BC} < 0$, e $V_{BE} \sim V_\gamma$; consideriamo inoltre un parametro variabile, I_B , ossia la corrente di base, che viene amplificata dal guadagno in corrente β_F . Si otterrà di fatto una serie di curve parametriche, in funzione della tensione V_{CE} . Possiamo considerare in base ad osservazioni algebriche questo fatto: la caratteristica statica avrà due zone di interesse, ossia $V_{CE} < V_\gamma$ e $V_{CE} > V_\gamma$: se infatti $V_{CE} > V_\gamma$, poichè sappiamo che il valore di $V_{BE} \sim V_\gamma$, allora di sicuro la tensione V_{BC} sarà sufficientemente negativa da poterci permettere di essere sicuri di trovarci in una regione attiva diretta. Secondo il modello statico, derivante dalle equazioni di Ebers-Moll, la caratteristica statica del transistor si presenta come una retta parallela all'asse delle x (dunque una funzione costante), all'aumentare di V_{CE} , poichè il guadagno in corrente resta costante, al di sopra della tensione V_{BE} . Per valori di V_{CE} inferiori alla tensione V_γ , si ha invece una funzione decrescente a 0, e si noti il seguente dettaglio: al variare del parametro I_B , le curve, da un certo punto di funzionamento V_{CE} in indietro verso l'origine degli assi, tendono asintoticamente allo stesso andamento: da tutte parallele tra loro, convergono tutte a zero, asintoticamente, allo stesso modo:

Questo andamento del grafico ci potrebbe far pensare al seguente fatto: se tutte le curve parametriche di I_B al variare di V_{CE} , parallele tra loro, sono tutte costanti dal punto V_γ in poi, potremmo dire che mediante un transistor bipolare si può realizzare un generatore di corrente pilotato ideale: poichè il guadagno secondo il modello di Ebers-Moll è una costante, poichè la caratteristica è una funzione costante, potremmo pensare di aver ottenuto



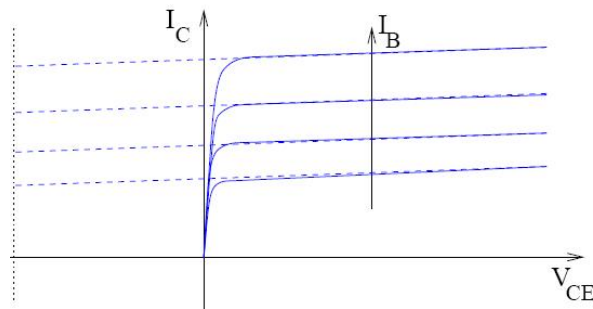
un componente ideale.

Osservando il grafico a sinistra del valore V_γ , osserviamo che questo andamento asintotico di tutte le curve parametriche al variare di I_B , tendono asintoticamente allo stesso modo ad andare a 0. Possiamo vedere algebricamente che, se $V_{CE} < V_\gamma$, allora significa che se $V_{BE} \sim V_\gamma$, dunque $V_{BC} > 0$, e ci troviamo in una zona di saturazione. L'andamento asintotico di tutte le curve parametriche ci fa intuire il fatto che non ha senso definire un guadagno in corrente in zona di saturazione non solo per il fatto che sarebbe minimo, ma anche perchè non è possibile identificare univocamente la curva presente in un determinato punto della suddetta zona: lo stesso livello di corrente, può equivalere a diversi valori di I_B , poichè se le curve parametriche passano tutte per lo stesso punto di corrente alla stessa tensione, allora significa che il guadagno non è univoco, come effettivamente lo è in regione attiva diretta, dove le curve non si intersecano (in quanto tutte parallele tra loro).

4.4.2 Effetto Early

Le vere cattive notizie arrivano però su di un altro fronte: tutto ciò che abbiamo finora descritto, è l'andamento della corrente I_C previsto dal modello statico, e, come abbiamo detto, sembrerebbe in grado di simulare l'andamento di un generatore pilotato ideale: al variare della tensione, corrente costante. Qua vengono però evidenziati i veri limiti del modello statico di Ebers-Moll del transistor bipolare: se analizzassimo in laboratorio la caratteristica statica del transistor, osserveremmo in realtà il fatto che la pendenza delle caratteristiche in regione attiva diretta è non nulla, ma positiva di poco:

Questo dipende dal fatto che l'effetto statico si fonda sulla famosa ipotesi di considerare l'ampiezza della regione di svuotamento in base non variabile, ma assimilabile con la lunghezza della base W_B : la tensione V_{CE} però è in grado di modulare l'ampiezza di svuotamento (poichè essa dipende da V_{BC} , quindi di fatto, variando la regione di svuotamento, gli effetti ricombinativi



diventano sempre meno trascurabili, e quindi di fatto si vedrà una pendenza positiva. Il modello come generatore pilotato è dunque fallimentare, ma nella cattiva notizia c'è una buona notizia: man mano che aumenta la tensione, il transistor continua a guadagnare sempre più corrente, potremmo dire che funziona sempre meglio, e quindi pur non avendo l'idealità, siamo riusciti a costruire con questo dispositivo un buon amplificatore di corrente.

L'effetto che tiene conto di questa variazione delle regioni di svuotamento è detto Effetto Early: esso prevede l'aumentare di facilità dei portatori ad attraversare la barriera di potenziale, grazie all'aumento della regione di svuotamento provocato dall'aumento della tensione V_{CE} . Si noti che anche per il transistor bipolare esiste un fenomeno di breakdown: esso è legato al fatto che, alzando molto la tensione V_{CE} (o, dualmente, abbassandola molto), è possibile far toccare, sovrapporre le due regioni di svuotamento, ed in questo modo di fatto eliminare del tutto fattori ricombinativi, far coincidere i lati, e perdere il controllo del dispositivo, la cui caratteristica schizzerebbe aumentando improvvisamente di pendenza.

Analizziamo ora un esercizio pratico, riguardante il transistor bipolare, nel quale faremo alcuni approfondimenti e tratteremo alcuni aspetti non ancora analizzati.

4.5 Esercizio Pratico

Cominciamo ancora prima di fornire i dati dell'esercizio con una piccola divagazione: dato un transistor bipolare, vale sempre la disuguaglianza a 3:

$$N_{DE} > N_{AB} > N_{DC}$$

Questo perchè:

- $N_{DE} > N_{AB}$: perchè il coefficiente α_F tenda a 1, un buon drogaggio permette di migliorare l'efficienza di emettitore del transistor;

- $N_{AB} > N_{DC}$: aumentando il drogaggio in base rispetto a quello in collettore, si fa sì che l'ampiezza della regione di svuotamento in base sia più difficile da incrementare, in questo modo si riesce ad arginare l'effetto Early ed anche (e soprattutto) il breakdown.

Considerati dunque, come dati del transistor bipolare: $N_{DE} = 10^{18}$; $N_{AB} = 5 \cdot 10^{16}$; $N_{DC} = 10^{16}$; $W_B = 0,2\mu m$; $W_E = 0,5\mu m$; $A_E = 100\mu m$, $\tau_p = 100ns$, $\mu_p(N_{DE}) = 150$, $\tau_n = 1\mu s$, $\mu_n(N_{AB}) = 1200$

1. Si determini il diagramma a bande all'equilibrio
2. Supponendo che l'ampiezza della base sia quella all'equilibrio, si valutino i parametri γ_F , α_T , α_F , β_F
3. Supponendo una tensione di polarizzazione $V_{BE} = 0,6$ V tra base ed emettitore, e $V_{CE} = 1$ V tra emettitore e collettore, si valuti il modello di piccolo segnale del transistor bipolare.

4.5.1 Risoluzione

Si determini il diagramma a bande all'equilibrio

In questo esercizio affronteremo diversi modi per calcolare alcuni parametri del transistor, utilizzando idee di base completamente diverse. Incominciamo con il calcolare, in modo classico, i valori di estrazione, in modo da poter stimare le barriere di potenziale, all'equilibrio.

Poichè l'emettitore del transistor bipolare è un semiconduttore estremamente drogato, quasi degenere, possiamo assumere alcune ipotesi semplificative: dal momento che in esso il livello di Fermi tende a coincidere con il livello energetico E_C , possiamo dire che il lavoro di estrazione coincida con l'affinità elettronica del silicio, nell'emettitore E:

$$q\Phi_{S_{n+E}} \simeq q\chi = 4,05eV$$

Nella base B:

$$q\Phi_{S_pB} \simeq q\chi + \frac{E_g}{2} + kT \ln \frac{N_{AB}}{n_i} \simeq 5eV$$

Nel collettore C:

$$q\Phi_{S_nC} \simeq q\chi + \frac{E_g}{2} - kT \ln \frac{N_{DC}}{n_i} \simeq 4,26eV$$

Che senso cercare in questo ambito le quotature del diagramma a bande? Aldilà del mero esercizio, possiamo renderci conto di un fatto importante: all'equilibrio le regioni di svuotamento permettono una zona neutra sufficientemente estesa; e nella base? Analizziamole: considerando che l'emettitore è estremamente drogato, nel calcolo si potrà considerare $N_{eq} \simeq N_{AB}$, per le proprietà della somma armonica di due termini: un termine molto grande e uno molto piccolo, sommati armonicamente, daranno vita ad un termine molto vicino a quello piccolo.

$$x_{dEB} \simeq x_{pEB} = \sqrt{\frac{2\epsilon_S \Phi_{i, BE}}{q N_{AB}}} = 0,15 \mu m$$

$$x_{dBC} = \sqrt{\frac{2\epsilon_S \Phi_{i, BC}}{q N_{eq}}} = 0,34 \mu m$$

Poichè della regione di svuotamento tra base ed emettitore sappiamo che essa è decisamente più ampia nella base, ma non abbiamo informazioni aggiuntive su quella di base-collettore, consideriamo la proporzione dei drogaggi, con la proporzione delle ampiezze delle regioni di svuotamento: le ampiezze delle regioni di svuotamento sono inversamente proporzionali alla concentrazione di drogaggio:

$$\frac{x_{nBC}}{x_{pBC}} = \frac{N_{AB}}{N_{BC}} \implies x_{pBC} = \frac{1}{6} x_{dBC} = 0,057 \mu m$$

Si noti che se sommiamo le due ampiezze di svuotamento, otteniamo:

$$0,15 + 0,057 = 0,207 \mu m$$

Dal momento che $W_B = 0,2 \mu m$, già all'equilibrio abbiamo una condizione di breakdown: le regioni di svuotamento si toccano, in questo modo il transistor ha una condizione di rottura e non vi è sostanzialmente base.

Scegliamo di modificare il testo del problema, considerando $W_B = 0,4 \mu m$: in questo modo, pur avendo allungato il lato della base, e quindi peggiorato il fattore di efficienza, avremo un lato di base neutro pari a $400 - 207$ nm.

Si valutino i parametri γ_F , α_T , α_F , β_F

Consideriamo dopo queste prime osservazioni e decisioni, il calcolo dei parametri richiesti dal problema. Partiamo dall'efficienza di emettitore, γ_F :

$$\gamma_F = \frac{I_{nE}}{I_{nE} + I_{pB}} = \frac{I_{ndiff}|_0}{I_{ndiff}|_0 + I_{pB}|_{-x_E}}$$

Studiamo dunque queste espressioni, ricavate in precedenza:

$$I_{pdiff}(-x_E) = -\frac{qA_E D_p n_i^2}{N_{DE} L_p} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

Possiamo semplificare la nostra espressione, osservando il fatto che, dal momento che ci troviamo in polarizzazione diretta, $V_{BE} > 0$, $e^{\frac{V_{BE}}{V_T}} \gg 1$; sappiamo che la lunghezza di diffusione L_p è così definita:

$$L_p = \sqrt{D_p \tau_p} = 6\mu m$$

Osserviamo che anche il lato di emettitore è corto: $W_E \ll L_p$.

Ciò può essere un bene in un transistor bipolare, in quanto lato corto implica riduzione della resistenza di accesso al transistor, e quindi siamo anche in grado di dire che $W_E \sim L_p$:

$$I_{pdiff}|_{-x_E} = \frac{-qA_E D_p n_i^2}{N_{DE} W_E} e^{\frac{V_{BE}}{V_T}}$$

Semplifichiamo a questo punto anche l'espressione della corrente di elettroni in 0:

$$I_{ndiff}|_0 = \frac{-qA_E D_n n_i^2}{N_{AB} L_n} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \operatorname{cogh} \frac{W_B}{L_n} + \frac{qA_E D_n n_i^2}{N_{AB} L_n \sinh \frac{W_B}{L_n}} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

Il secondo termine è trascurabile rispetto al primo; inoltre, nel primo termine, oltre a semplificare in precedenza il -1 , siamo in grado di fare dei ragionamenti simili al precedente, sul termine W_B : poichè:

$$L_n = \sqrt{D_n \tau_n} = 56\mu m \gg W_B$$

Possiamo dire di nuovo di trovarci in ipotesi di lato corto (come è ovvio che sia la base), e dunque possiamo sviluppare il coseno iperbolico come 1, ed il seno iperbolico come il proprio argomento:

$$\cosh \frac{W_B}{L_n} \simeq 1; \sinh \frac{W_B}{L_n} \simeq \frac{W_B}{L_n}$$

Possiamo dunque dire che:

$$I_{ndiff}(0) \simeq -\frac{qA_E D_n n_i^2}{N_{AB} L_n \frac{W_B}{L_n}} e^{\frac{V_{BE}}{V_T}} = -\frac{qA_E D_n n_i^2}{N_{AB} W_B} e^{\frac{V_{BE}}{V_T}}$$

Possiamo dunque ora riprendere la definizione di efficienza di emettitore, e, sostituendo, si ottiene che:

$$\gamma_F = \frac{\frac{-qA_E D_p n_i^2}{N_{DE} W_E} e^{\frac{V_{BE}}{V_T}}}{\frac{-qA_E D_p n_i^2}{N_{DE} W_E} e^{\frac{V_{BE}}{V_T}} + \frac{-qA_E D_n n_i^2}{N_{AB} W_B} e^{\frac{V_{BE}}{V_T}}}$$

$$\gamma_F = \frac{1}{1 + \frac{N_{AB} D_p W_B}{N_{DE} D_n W_E}}$$

Possiamo calcolare che $\gamma_F = 0,9976$.

Abbiamo calcolato mediante questa nuova formula l'efficienza di emettitore, semplificando notevolmente i nostri conti. Passiamo ora al calcolo di un'espressione in grado di fornire il valore del parametro α_T , ossia del fattore di trasporto; sappiamo che:

$$\alpha_T = \frac{I_{ndiff}(0) - I_{RB}}{I_{ndiff}(0)} = \frac{I_{ndiff}(W_B)}{I_{ndiff}(0)}$$

Abbiamo in precedenza effettuato alcune approssimazioni nel termine di corrente di diffusione di elettroni in base, nel punto $x = 0$. Queste approssimazioni sono da rivedere, per questo motivo: vediamo che in questo ambito, è coinvolta anche la grandezza $I_{n,diff}(W_B)$, che coinvolge un $\cosh(0)$, che vale esattamente 1, tende e non solo a 1, come precedentemente. Di fatto, non sarà più possibile approssimare il coseno iperbolico a 1, considerando $W_B \ll L_n$, in quanto otterremo un'approssimazione troppo grossolana; questa volta un coseno iperbolico tenderà naturalmente a 1, l'altro sarà sviluppato in serie di Taylor, ma troncato al secondo ordine; consideriamo dunque le espressioni in gioco:

$$I_{ndiff}(0) = \frac{-qA_E D_n n_i^2}{N_{AB} L_n} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \operatorname{cogh} \frac{W_B}{L_n} \simeq \frac{-qA_E D_n n_i^2 \cosh \frac{W_B}{L_n}}{N_{AB} L_n \sinh \frac{W_B}{L_n}} e^{\frac{V_{BE}}{V_T}}$$

Abbiamo dunque considerato come unica approssimazione, per ora, l'effetto dell'esponenziale che prevale enormemente sul termine -1 in parentesi; consideriamo ora l'espressione della corrente di diffusione di elettroni in base, nel punto W_B : in questo ambito, come già anticipato, il coseno iperbolico avrà argomento algebricamente 0, e non solo tendente a 0:

$$I_{ndiff}(W_B) = \frac{-qA_E D_n n_i^2 \cosh(0)}{N_{AB} \sinh \frac{W_B}{L_n} L_n} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \simeq \frac{-qA_E D_n n_i^2}{N_{AB} \sinh \frac{W_B}{L_n} L_n} e^{\frac{V_{BE}}{V_T}}$$

Consideriamo ora il rapporto tra le due funzioni, ai fini di calcolare il parametro α_T :

$$\alpha_T = \frac{\frac{-qA_E D_n n_i^2}{N_{AB} \sinh \frac{W_B}{L_n} L_n} e^{\frac{V_{BE}}{V_T}}}{\frac{-qA_E D_n n_i^2 \cosh \frac{W_B}{L_n} e^{\frac{V_{BE}}{V_T}}}{N_{AB} L_n \sinh \frac{W_B}{L_n}}} = \frac{1}{\cosh \frac{W_B}{L_n}}$$

Solo ora sfruttiamo il fatto che ci troviamo in semiconduttore corto, considerando il fatto che $W_B \ll L_n$, ma, come anticipato, lo sviluppo considerato sarà quello troncato non al primo, ma al secondo ordine:

$$\alpha_T = \frac{1}{1 + \frac{1}{2} \frac{W_B^2}{L_n^2}} = 1 - \frac{W_B^2}{2L_n^2} = 0,999994$$

Abbiamo ottenuto, anche per il fattore di trasporto α_T , un'espressione operativa molto più utile (in caso di un transistor bipolare come quello proposto dall'esercizio, ma che comunque rappresenta il transistor ottimale: emettitore corto e molto drogato, base corta), e facile da utilizzare, rispetto alla definizione. Possiamo ora, partendo da ciò, calcolare il guadagno in corrente del transistor:

$$\alpha_F = \gamma_F \alpha_T \simeq 0,997$$

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F} \simeq 413$$

Permettiamoci ora una divagazione, ragionando in modo diverso sul coefficiente di trasporto, α_T : possiamo infatti importare anche nel transistor bipolare il modello a controllo di carica precedentemente studiato nella giunzione p-n. Ciò risulta molto interessante, soprattutto poichè è possibile linearizzare la concentrazione degli elettroni in base, $n'_p(x)$:

$$n'_p(x) \simeq \frac{n_i^2}{N_{AB}} e^{\frac{V_{BE}}{V_T}} \frac{W_B - x}{L_n} = \frac{n_i^2}{N_{AB}} e^{\frac{V_{BE}}{V_T}} \left(1 - \frac{x}{W_B}\right)$$

Abbiamo dunque ridotto ad una semplice retta la distribuzione di carica, sfruttando l'ipotesi di lato corto, la distribuzione dei portatori minoritari in base. Potremmo chiederci come mai non abbiamo effettuato in precedenza una semplificazione di questo tipo: linearizzare una funzione, ossia ridurla a retta, implica ridurre la sua derivata ad una costante, o addirittura a 0, perdendo informazioni su di essa. Il modello a controllo di carica però non richiede la derivazione delle distribuzioni di portatori, bensì l'integrazione, al fine di ottenere la carica totale in eccesso nel semiconduttore. Poichè siamo in un caso di retta decrescente, la totale corrente di ricombinazione sarà

assimilabile al calcolo dell'area di un triangolo: considerando la lunghezza del lato di base come base, e la concentrazione di carica come altezza:

$$Q'_n = \frac{n_i^2}{N_{AB}} e^{\frac{V_{BE}}{V_T}} \frac{W_B}{2} (-qA_E)$$

Da ciò, sappiamo semplicemente che la corrente di ricombinazione in base vale:

$$I_{RB} = \frac{Q'_n}{\tau_n} = \frac{n_i^2}{N_{AB}} e^{\frac{V_{BE}}{V_T}} \frac{-qA_E W_B}{2\tau_n}$$

Da ciò, è possibile calcolare il coefficiente di trasporto come:

$$\alpha_T = \frac{I_{n,diff}(0) - I_{RB}}{I_{n,diff}(0)} = 1 - \frac{I_{RB}}{I_{n,diff}(0)} = 1 - \frac{\frac{-qA_E W_B n_i^2}{2N_{AB}\tau_n} e^{\frac{V_{BE}}{V_T}}}{\frac{-qA_E D_n n_i^2}{2N_{AB}W_B} e^{\frac{V_{BE}}{V_T}}} = 1 - \frac{W_B^2}{2L_n}$$

Abbiamo ottenuto lo stesso risultato espresso precedentemente, mediante un processo completamente diverso. In realtà, esiste una via ulteriore, e ancora più comoda: partendo dalle equazioni di Ebers-Moll, si può vedere che:

$$I_E = a_{1,1} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + a_{1,2} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

$$I_C = a_{2,1} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + a_{2,2} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

Poichè ci troviamo in regione attiva diretta, $V_{BC} < 0$, possiamo dire che il secondo termine sia in questo ambito del tutto trascurabile, per entrambe le equazioni:

$$I_E \simeq a_{1,1} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

$$I_C \simeq a_{2,1} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

Da ciò, si può semplicemente ricavare che:

$$I_E \simeq \frac{a_{1,1}}{a_{2,1}} I_C \implies \alpha_F = -\frac{a_{1,1}}{a_{2,1}}$$

Possiamo terminare questa divagazione, ed essere soddisfatti per tutte le tecniche finora acquisite per il calcolo dei parametri del transistor bipolare: abbiamo visto che, mediante ipotesi e tecniche completamente diverse, si arriva a ottenere risultati assolutamente analoghi.

Si valuti il modello di piccolo segnale del transistor bipolare

Parliamo ora del terzo punto del problema, ossia lo studio di un modello di piccolo segnale del transistor bipolare. Come dati abbiamo che $V_{BE} = 0,6$ V, che $V_{CE} = 1$ V, ma poichè $V_{CE} = V_{BE} - V_{BC}$, sappiamo che $V_{BC} = -0,4$ V. Usiamo quello che viene detto il modello a π semplificato, utile in quanto tornerà utile in seguito nello studio di sistemi MOS. Nella regione attiva diretta, I_E dipende molto da V_{BE} , mentre I_C è rappresentabile come:

$$I_C = -\alpha_F I_E$$

Il modello di piccolo segnale si basa, come abbiamo già visto precedentemente, su piccolissime variazioni di tensione; per analizzare la variazione di corrente I_C , varieremo la tensione che la modula, indirettamente: V_{BE} (si discuterà meglio sul significato della parola 'indirettamente').

Abbiamo, nel transistor bipolare, un dispositivo che dal mondo esterno si vede come un diodo, ed un collettore che di fatto non fa altro che raccogliere la corrente. Consideriamo, a partire dalle equazioni di Ebers-Moll, che, per la stessa ipotesi considerata precedentemente,

$$I_C(V_{BE}; V_{BC}) \simeq I_C(V_{BE})$$

$$I_C = a_{2,1} e^{\frac{V_{BE}}{V_T}}$$

Si noti che, poichè adottiamo il modello statico introdotto da Ebers-Moll, consideriamo trascurato l'effetto Early. Possiamo al solito trascurare i termini dipendenti dalla tensione tra base e collettore, V_{BC} , poichè siamo in regione attiva diretta. Linearizziamo dunque il nostro modello, considerando una variazione infinitesima della corrente I_C rispetto alla tensione di controllo, V_{BE} :

$$\left. \frac{\partial I_C}{\partial V_{BE}} \right|_{V_{BEop}} = \frac{a_{2,1}}{V_T} e^{\frac{V_{BEop}}{V_T}}$$

Definiamo dunque la corrente del punto di lavoro, I_{Cop} , come:

$$I_{Cop} \triangleq \left. \frac{\partial I_C}{\partial V_{BE}} \right|_{V_{BEop}}$$

Essendo noi in un modello di piccolo segnale, modificando infinitesimalmente la tensione V_{BEop} con un piccolo segnale v_{be} , possiamo vedere che:

$$\frac{I_{C_{op}}}{V_T} = g_m$$

Dove questa g_m rappresenta una conduttanza analoga a quella precedentemente studiata nel diodo a giunzione. Si noti il seguente fatto: togliendo gli offset verticali ed orizzontali $I_{C_{op}}$ e $V_{BE_{op}}$, si vedrebbe che g_m rappresenta una retta, e dunque una resistenza, o almeno potremmo pensarlo. In realtà, V_{BE} come già detto modula solo indirettamente il valore di g_m , poichè è la tensione di controllo, ma è la tensione sulla giunzione tra base e collettore, mentre g_m si trova tra base ed emettitore; si parla dunque di g_m come di una transconduttanza, ossia, come di un parametro che, moltiplicato per il pilota v_{be} , permette di ottenere I_C , anche se non vi è una relazione diretta.

Ci manca un ulteriore parametro: ciò che collega la base B all'emettitore E. Potremmo ripetere le operazioni precedenti, considerando I_B :

$$I_B = \frac{I_C}{\beta_F} \implies \frac{1}{\beta_F} \left. \frac{\partial I_C}{\partial V_{BE}} \right|_{V_{BE_{op}}} = \frac{g_m}{\beta_F}$$

Possiamo dunque definire la resistenza del nostro modello π , r_π , come:

$$r_\pi = \frac{\beta_F}{g_m}$$

Il modello di piccolo segnale del transistor bipolare sarà semplicemente rappresentato da un nodo generalizzato, in cui tra B ed E sta la r_π , tra E e C il generatore pilotato di valore $I_C = g_m v_{be}$, dove v_{be} , ossia il piccolo segnale del nostro modello, rappresenta la tensione sulla r_π .

Divaghiamo, accennando al modello di ampio segnale del transistor bipolare: consideriamo il fatto che ci troviamo semplicemente in una doppia giunzione, e dunque dovremo considerare capacità di svuotamento e capacità di diffusione per ognuna delle giunzioni. Dal momento che il punto di funzionamento del transistor interessante è la regione attiva diretta, e quindi in cui $V_{BE} > 0$, e $V_{BC} < 0$, possiamo dire che un parametro sarà trascurabile: $C_{DIF}(V_{BC}) \simeq 0$ poichè ci troviamo in una zona di polarizzazione inversa della giunzione base-collettore, e quindi possiamo trascurare.

Possiamo calcolare i parametri linearizzati, a partire dal modello di ampio segnale, al fine di terminare del tutto il modello di piccolo segnale:

$$g_m = \frac{I_{C_{op}}}{V_T}; \quad I_{C_{op}} = a_{2,1} e^{\frac{V_{BE_{op}}}{V_T}}; \quad a_{2,1} = \frac{q A_E D_n n_i^2}{L_n N_{AB} \sinh \frac{W_B}{L_n}} \simeq \frac{q A_E D_n n_i^2}{N_{AB} W_B}$$

$$r_\pi = \frac{\beta_F}{g_m}; \quad C_{DEP}(V_{BEop}) = \sqrt{\frac{q\epsilon_S N_A}{2(\Phi_i - V_{BEop})}}; \quad C_{DEP,BC} = \sqrt{\frac{q\epsilon_S N_{eq}}{2(\Phi_{BC} - V_{BCop})}}$$

Possiamo infine linearizzare la carica totale, considerando come al solito le ipotesi di lato corto. Poichè inoltre il lato dell'emettitore è estremamente drogato (cosa già sfruttata nel calcolo di $C_{DEP}(V_{BEop})$), possiamo dire che il contributo di carica di diffusione in base sia trascurabile. Calcoliamo dunque nella base B la carica totale come area del triangolo, per poi derivare rispetto alla tensione e così trovare la capacità di diffusione:

$$Q' = Q'_p + Q'_n \simeq Q'_n = -qA_E n'_p(0) \frac{W_B}{2} = -\frac{qn_i^2 W_B A_E}{2N_{AB}} e^{\frac{V_{BE}}{V_T}}$$

Derivando:

$$C_{DIF} = \left| \frac{\partial Q'_n}{\partial V_{BE}} \right| = \frac{qn_i^2 W_B A_E}{2N_{AB} V_T} e^{\frac{V_{BE}}{V_T}}$$

Capitolo 5

Sistemi MOS

L'acronimo MOS significa Metallo-Ossido-Semiconduttore (o meglio in inglese Metal-Oxide-Semiconductor); per costruire dispositivi basati sulla tecnologia MOS, si usano:

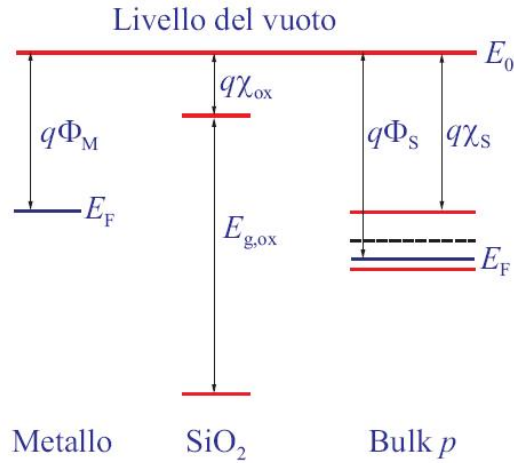
- Come metalli: in principio Alluminio (Al), poi Silicio (Si) policristallino molto drogato;
- Come ossido: biossido di silicio (SiO_2), comunemente detto quarzo (notoriamente isolante);
- Come semiconduttore: Silicio (Si-p o Si-n).

Per ora, fino a quando non si affermerà il contrario, si considererà una struttura di alluminio, quarzo, e silicio drogato tipo p.

Siamo in grado di tracciare il diagramma a bande del dispositivo appena unito, ancora fuori dall'equilibrio, considerando come facemmo in precedenza il livello del vuoto E_0 continuo, e tutti gli altri livelli discontinui.

Unendo i tre materiali, i livelli di Fermi dovrebbero tendere a parificarsi, per poter ottenere una condizione di equilibrio. Il problema che fa variare questa situazione rispetto a quelle precedentemente studiate, è la presenza dell'ossido, che è un isolante, e dunque impedirà il passaggio di carica al suo interno, a causa del suo enorme gap di energia (e non solo come vedremo tra pochissimo). Poichè si raggiunga l'equilibrio si può ipotizzare, con le nostre attuali conoscenze, che serva un tempo infinito. Supponendo che l'equilibrio sia già stato raggiunto, il diagramma a bande sarà di questo tipo:

Il metallo a causa della propria natura non può subire modifiche o pieghe dei livelli energetici; il semiconduttore, poichè ha lavoro di estrazione più elevato, si svuota parzialmente, di lacune. Poichè del dielettrico sappiamo

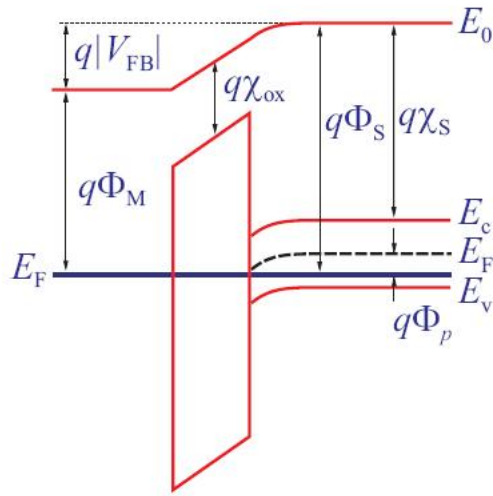


che esso è in mezzo ai due, possiamo assimilare il sistema ad un condensatore, il cui campo è dunque costante ovunque, ed il potenziale varia in modo intuitivamente lineare. Cosa si è originata dunque? Una struttura dotata di barriere di potenziale all'interfaccia con l'ossido: conosciamo il lavoro di estrazione del metallo, e del semiconduttore, e sappiamo che nell'ossido $q\Phi_{OX} \simeq q\chi$, quindi possiamo dire che $E_0 - q\chi = E_C$; ma E_C non è costante, e anzi in prossimità delle interfacce con l'ossido presenta discontinuità, per essere precisi di 3,3 eV (con il metallo), e di 3,25 eV (con il semiconduttore), come si può calcolare semplicemente mediante l'approssimazione appena presentata del lavoro di estrazione dell'ossido con la sua affinità elettronica. Una barriera di potenziale così elevata non può essere saltata, e nei nostri casi neanche attraversata per effetto tunnel, poichè i dispositivi che consideriamo hanno strati di ossido il cui spessore $T_{OX} \simeq 100$ amstrong.

5.1 Analisi Elettrostatica

Introduciamo un primo discorso sullo studio del sistema MOS; alcuni dettagli per ora trascurati, saranno comunque in seguito ripresi ed approfonditi. Consideriamo nel sistema una convenzione che prevede l'asse x con l'origine nell'interfaccia metallo-ossido, e con verso positivo da sinistra verso destra. Dalle equazioni di Maxwell si può studiare che, in un qualsiasi sistema, il prodotto della costante dielettrica ϵ e del campo elettrico nel medesimo punto è una funzione continua (teorema dello spostamento dielettrico). L'ossido sarà privo di cariche in principio, il semiconduttore p in questione avrà invece una carica pari a $\rho = -qN_Ax_{d0}$.

Il metallo per definizione non potrà avere eccessi locali di carica, ma,

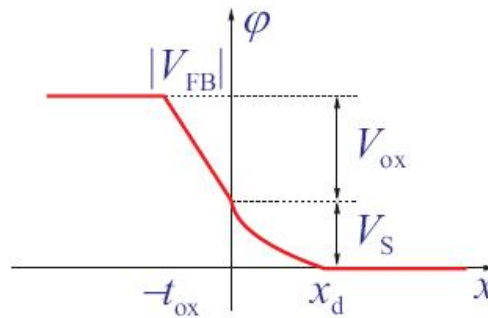


Equilibrio termodinamico

poichè deve essere rispettata la neutralità globale nel semiconduttore, si può pensare che, in prossimità dell'interfaccia metallo-ossido, vi sia una δ di carica che permetta un impulso di carica pari a qN_Ax_{d0} , e cioè in grado di neutralizzare la carica di svuotamento del semiconduttore. Poichè vale la legge dello spostamento dielettrico, possiamo dire che:

$$\epsilon_{OX}\epsilon_{OX} = \epsilon_S(0)\epsilon_S$$

Poichè $\epsilon_{OX} = 3,9$, e $\epsilon_S = 11,7$, vi sarà una barriera, come già accennato. Il campo presenterà una discontinuità nella posizione 0 dell'asse delle x , affinché lo spostamento dielettrico sia verificato; applicando l'equazione di Poisson, integriamo rapidamente la carica di svuotamento, ottenendo:



$$\epsilon_{OX} = \frac{\epsilon_S}{\epsilon_{OX}}\epsilon_S(0) = \frac{\epsilon_S}{\epsilon_{OX}} \frac{qN_A}{\epsilon_S} x_{d0}$$

Per il calcolo del potenziale, introduciamo un'ulteriore convenzione: consideriamo la massa del sistema in $+\infty$, e quindi:

$$\Phi(+\infty) = 0; \quad \Phi = - \int \varepsilon dx$$

Qualitativamente, il potenziale avrà un andamento quadratico dall'inizio della zona di svuotamento all'interfaccia con l'ossido, per poi crescere linearmente, e quindi stabilizzarsi nel metallo. Si definisce, nel metallo, il potenziale Φ_{MS} come il potenziale di contatto tra metallo e semiconduttore: questa definizione è priva di senso fisico in quanto non vi è un contatto, impedito dall'ossido, ma tornerà estremamente utile come definizione.

Il potenziale tra silicio e metallo, sarà:

$$\Phi_{MS} = \frac{q\Phi_M - q\Phi_{Sp}}{-q} \sim 0,9V$$

Il potenziale di contatto ha due contributi: V_S , ossia quello provocato dalla zona di svuotamento del semiconduttore, e V_{OX} , ossia quello provocato dall'appena citato andamento lineare del potenziale nell'ossido. Per quotare tutto il diagramma a bande, mancano ancora alcuni parametri: x_{d0} , V_S , V_{OX} .

Consideriamo una sorta di equazione alla maglia:

$$\Phi_{MS} = V_{OX} + V_S$$

Possiamo anche esprimere le due tensioni come funzioni di x_{d0} :

$$V_S = \frac{qN_A}{2\varepsilon_S} x_{d0}^2$$

$$V_{OX} = \frac{qN_A}{\varepsilon_{OX}} x_{d0} T_{OX}$$

Si noti che però possiamo identificare, come già detto, lo spazio compreso tra metallo e semiconduttore come un condensatore, e dunque identificare nell'ossido una capacità, ossia con un condensatore a facce piane parallele:

$$C_{OX} = \frac{\varepsilon_{OX}}{T_{OX}} \implies V_{OX} = \frac{qN_A x_{d0}}{C_{OX}}$$

Sostituendo, l'equazione alla maglia diverrà:

$$\Phi_{MS} = \frac{qN_A x_{d0}}{C_{OX}} + \frac{qN_A}{2\varepsilon_S} x_{d0}^2 \implies x_{d0}^2 + \frac{2\varepsilon_S}{C_{OX}} x_{d0} - \Phi_{MS} \frac{2\varepsilon_S}{qN_A} = 0$$

Risolviendo l'equazione in funzione dell'incognita x_{d0} , e considerando solo la radice positiva:

$$x_{d0} = -\frac{\epsilon_{OX}}{C_{OX}} + \sqrt{\left(\frac{\epsilon_S}{C_{OX}}\right)^2 + \frac{2\epsilon_S\Phi_{MS}}{qN_A}}$$

Tipicamente, la zona di svuotamento x_{d0} ha un valore intorno ai $0,1 \mu\text{m}$; nelle condizioni che stiamo studiando ora (equilibrio), Φ_{MS} sarà per metà dovuta a tensioni dell'ossido V_{OX} , e per metà alla V_S del semiconduttore. Possiamo dire di aver quotato il diagramma a bande all'equilibrio di questa prima struttura MOS.

Passiamo ora al considerare il semiconduttore MOS in situazioni fuori equilibrio, e quindi consideriamo l'applicazione sul metallo M e sul semiconduttore S di una tensione esterna V_a , considerando, in coerenza con la precedente convenzione introdotta, il $-$ sul verso del semiconduttore, e il $+$ verso il metallo.

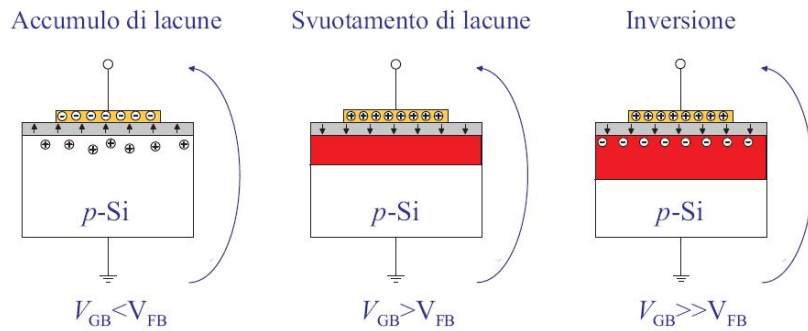
Abbiamo sostanzialmente due possibilità da studiare, ossia $V_a < 0$, e $V_a > 0$.

5.1.1 Polarizzazione di un sistema MOS con tensione negativa

Se consideriamo una $V_a < 0$, dalle nostre attuali conoscenze sappiamo esclusivamente che i lati del semiconduttore saranno abbassati: con una tensione di polarizzazione esterna negativa infatti i livelli energetici dei lati si alzeranno rispetto al lato semiconduttivo, la curvatura del semiconduttore si ridurrà, e con essa la regione di svuotamento. Per una determinata tensione, precisamente per $V_a = -\Phi_{MS}$, la tensione esterna compenserà interamente il potenziale di contatto, e così il livello di barriera E_0 sarà allo stesso livello per i tre materiali, e quindi avremo interamente compensato il potenziale di contatto Φ_{MS} (come suggerisce anche la definizione della tensione). Questa situazione si dice di Flat Band (ossia condizione di banda piatta); questa è una condizione che non abbiamo ancora incontrato in precedenza, infatti il fatto che l'ossido separi metallo e semiconduttore, impedisce il raggiungimento della giunzione di non possibile raddrizzamento dei livelli, a causa dell'irrefrenabile flusso di portatori di carica, che in questo caso vengono bloccati dalla barriera di potenziale dell'ossido. Si è soliti dunque definire la tensione di banda piatta, V_{FB} , come:

$$V_{FB} \triangleq -\Phi_{MS}$$

Abbassando ulteriormente la tensione, e quindi per $V_a < V_{FB}$, vi è il sorpasso del livello E_0 del metallo rispetto a quello del semiconduttore: ciò indurrà una carica positiva nel semiconduttore drogato tipo p (come si potrebbe anche vedere dal diagramma a bande), dando vita ad una regione duale a quella di svuotamento: la regione di accumulo. In questo caso, la δ di carica del metallo diventerebbe negativa, poichè dovrebbe compensare la carica di accumulo positiva nel semiconduttore. Non ci interessa ad ogni modo lo studio del sistema MOS in condizioni di accumulo, quindi questo cenno sulla condizione di $V_a < 0$ non rappresenterà il punto di funzionamento dei dispositivi basati su tecnologie di questo tipo, o almeno non di quelle da noi trattate.



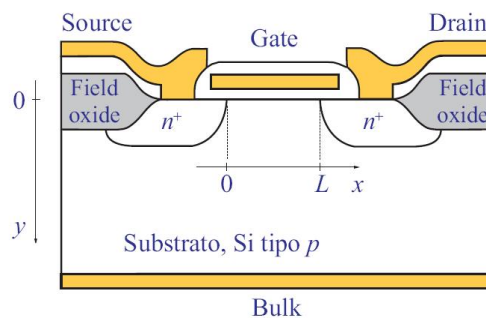
5.1.2 Inversione di popolazione

Considerando una $V_a > 0$, rispetto alla condizione di equilibrio, aumenterà la flessione del livello energetico del semiconduttore, che si innalzerà ulteriormente rispetto a quello del metallo e dell'ossido. Progressivamente con l'aumentare di V_a , aumenterà progressivamente l'ampiezza della regione di svuotamento, ma vi saranno fenomeni che non siamo in grado di prevedere finora: lontani dal punto x_d si ha la neutralità, la più totale normalità, come in uno stato di equilibrio; avvicinandoci però all'ossido, notiamo che la distanza $E_{F_i}(x) - E_F$ continua a diminuire, addirittura, a certe tensioni V_a , diventando negativa; mediante le equazioni di Shockley, si può di fatto studiare (come faremo tra non molto) fenomeni interessanti legati a ciò. Sappiamo per ora che per $x \rightarrow +\infty$, $n_p(+\infty) \rightarrow n_{p0}$, e che $p_p(+\infty) \rightarrow N_A$; la distanza varia, come già detto in precedenza per la giunzione, solo in regione di svuotamento. Continuiamo a proporre definizioni e osservazioni, prima di passare a calcoli rigorosi, definendo il parametro:

$$q\Phi_p \triangleq E_{F_i}(+\infty) - E_F$$

Possiamo qualitativamente osservare che, se la V_a flette a sufficienza le bande al punto da far sì che $E_{F_i}(0) - E_F = -q\Phi_p$, allora nel punto 0, $n_p(0) = p_p(+\infty)$. Sostituendo i dati nelle equazioni di Shockley si può avere riscontro formale di questo fatto, straordinario: dato un drogaggio $N_A = 10^{17}$, in un semiconduttore in cui quindi gli elettroni sono minoritari, si avrebbe una densità di elettroni pari a N_A , in prossimità dell'interfaccia con l'ossido. Questo meccanismo è detto inversione di popolazione, ed è interessantissimo per creare dispositivi molto particolari: la densità enorme di inversione $n_p(x)$ si viene a formare in uno spazio ridottissimo, in una nicchia provocata dall'estremo curvamento dei livelli energetici causato dalla tensione esterna V_a : questa è di fatto una zona confinata stracolma di elettroni (zona di inversione), e dunque fortemente conduttiva. Questa zona è alla base del funzionamento del transistor MOSFET (Metal-Oxide-Semiconductor Field-Effect-Transistor), dispositivo così costruito: su di un substrato drogato p (seguendo il nostro caso, poi ovviamente è possibile invertire i drogaggi), vengono impiantate due zone, due pozzetti fortemente drogati tipo n^+ , uno detto source, l'altro detto drain. Metteremo a contatto ossido e metallo con questi pozzetti, in modo che il canale scorra, per così dire, sotto l'ossido. Il canale è proprio questa zona di inversione, questa nicchia creata dalla forte tensione di cui abbiamo parlato. Vi saranno due tensioni alla base del funzionamento di un MOSFET:

- V_{GS} : la tensione del gate metallico, ossia quella che finora abbiamo chiamato V_a : essa è la tensione che permette la creazione di un canale di elettroni, nel semiconduttore;
- V_{DS} : è la tensione positiva tra drain D e source S, che sposta gli elettroni, passando nel canale che si viene a creare per inversione permessa dalla V_{GS} , appunto prendendoli dal source e spostandoli verso il drain (poichè il + è sul drain, gli elettroni andranno in questo senso).



5.2 Calcolo della carica di inversione

Consideriamo d'ora in poi un'analisi più quantitativa: per poter calcolare eventuali correnti infatti è necessario conoscere la concentrazione complessiva della carica di inversione (di cui ora non abbiamo un'espressione operativa). Per fare ciò, avremo a disposizione mezzi matematici molto ridotti e basilari: equazione di continuità, equazioni di Shockley, equazione di Poisson.

Sappiamo che:

$$\rho(x) = q(p_p(x) - n_p(x) - N_A)$$

Vorremmo esplicitare, in funzione della tensione complessiva del semiconduttore, le due concentrazioni di portatori, $p_p(x)$ e $n_p(x)$; partiamo dalle equazioni di Shockley:

$$p_p(x) = n_i e^{\frac{E_{F_i}(x) - E_F}{kT}}; \quad n_p(x) = n_i e^{-\frac{E_{F_i}(x) - E_F}{kT}}$$

Definiamo per comodità il parametro $q\Phi(x) \triangleq E_{F_i}(x) - E_F$:

$$\implies p_p(x) = n_i e^{\frac{q\Phi(x)}{kT}}; \quad n_p(x) = n_i e^{-\frac{q\Phi(x)}{kT}}$$

Sappiamo che, molto lontani dall'interfaccia con l'ossido:

$$p_p(+\infty) = p_{p0} = n_i e^{\frac{q\Phi_p}{kT}}; \quad n_p(+\infty) = n_{p0} = n_i e^{-\frac{q\Phi_p}{kT}}$$

Ricaviamo, ora, da queste ultime due espressioni, n_i , per poi sostituirlo nelle equazioni di partenza:

$$n_i = p_{p0} e^{-\frac{q\Phi_p}{kT}}; \quad n_i = n_{p0} e^{\frac{q\Phi_p}{kT}}$$

Tornando finalmente alle equazioni di Shockley iniziali:

$$p_p(x) = p_{p0} e^{\frac{q\Phi(x) - q\Phi_p}{kT}}; \quad n_p(x) = n_{p0} e^{-\frac{q\Phi(x) - q\Phi_p}{kT}}$$

Possiamo, ora, per comodità, definire una funzione che studi il potenziale, con riferimento a $+\infty$:

$$\Psi(x) = \frac{q\Phi(x) - q\Phi_p}{-q}$$

In un certo senso potremmo dire che $\Psi(x)$ rappresenta il nostro voltmetro ideale: fissata la massa in un punto di neutralità (come un ideale $+\infty$), spostandoci in diverse x avremo la tensione nel punto interessato. Si noti che $\Psi(0) = V_S$ (poichè misuriamo semplicemente in questo modo il contributo del solo semiconduttore al potenziale di contatto);

Possiamo dunque riscrivere ordinatamente le equazioni di Shockley, nel seguente modo:

$$p_p(x) = p_{p0} e^{-\frac{\Psi(x)}{V_T}}$$

$$n_p(x) = p_{p0} e^{\frac{\Psi(x)}{V_T}}$$

In condizioni di banda piatta, $\Psi(x) = 0$ a tappeto, per qualsiasi x ; in inversione, invece, $\Psi(x)$ sarà una funzione crescente, muovendoci da destra verso sinistra. Possiamo dunque dire che essa ben modella ciò che intendevamo modellizzare.

Torniamo ora a parlare della ρ ; possiamo dire, poichè lontano dall'interfaccia con l'ossido vi è neutralità, che $\rho(+\infty) = 0$. Possiamo però quindi anche dire che:

$$\rho(+\infty) = 0 \implies q(p_p(+\infty) - n_p(+\infty) - N_A) = 0 \implies p_{p0} - n_{p0} - N_A = 0$$

Possiamo dunque ricavare N_A :

$$N_A = p_{p0} - n_{p0}$$

Sostituendo ciò nell'espressione della densità di carica, otteniamo che:

$$\rho(x) = +q \left(p_{p0} e^{-\frac{\Psi(x)}{V_T}} - n_{p0} e^{\frac{\Psi(x)}{V_T}} - p_{p0} + n_{p0} \right)$$

Quindi, ordinando:

$$\rho(x) = q \left[p_{p0} \left(e^{-\frac{\Psi(x)}{V_T}} - 1 \right) - n_{p0} \left(e^{\frac{\Psi(x)}{V_T}} - 1 \right) \right]$$

Ora che abbiamo un'espressione della carica più interessante, possiamo considerare l'equazione di Poisson con la funzione $\Psi(x)$ come tensione variabile. Nella fatispicie, il nostro scopo sarà risolvere l'equazione differenziale:

$$\frac{\partial^2 \Psi(x)}{\partial x^2} = -\frac{q}{\epsilon_S} \left[p_{p0} \left(e^{-\frac{\Psi(x)}{V_T}} - 1 \right) - n_{p0} \left(e^{\frac{\Psi(x)}{V_T}} - 1 \right) \right]$$

Vogliamo dunque capire in che situazione ci verremo a trovare, in seguito all'integrazione. Incominciamo rappresentando qualitativamente carica e campo elettrico: l'effetto della tensione applicata aumenta infatti progressivamente l'ampiezza della zona svuotata; se V_a è abbastanza positiva, si formerà, interfacciata con l'ossido, una linguetta di carica, la famosa nicchia che abbiamo prima nominato, ossia la zona di inversione, sottile alcuni nm.

Partendo da un generico $x_0 > x_p$, sappiamo per certo, avendo noi fissato la nostra massa, che $\varepsilon(x_0) = 0$, e $\Psi(x_0) = 0$. Possiamo dunque applicare la legge di Poisson, per ottenere il campo come integrale della densità di carica ρ ; consideriamo un cambio di variabili, ed una moltiplicazione membro a membro per il fattore $\frac{\partial\Psi(x)}{\partial x}$, il cui differenziale sarà $\partial y = \frac{\partial^2\Psi(x)}{\partial x^2}\partial x$, al fine di risolvere l'integrale per sostituzione:

$$y = \frac{\partial\Psi}{\partial x}; \quad dy = \frac{\partial^2\Psi}{\partial x^2}\partial x$$

Notiamo che dunque, sostituendo, abbiamo semplicemente, al primo membro, ydy . Integrando dunque ambo i membri, incominciamo con il discutere il primo:

$$ydy \implies \int_{-\varepsilon(x_0)}^{-\varepsilon(x)} ydy = \frac{y^2}{2} = \frac{(-\varepsilon(x))^2}{2}$$

Gli intervalli di integrazione considerano il fatto che, per l'equazione di Poisson,

$$y = \frac{\partial\Psi}{\partial x} = -\varepsilon(x)$$

Risolto questo dettaglio, passiamo al secondo membro, integrato:

$$\begin{aligned} & \int_{\Psi(x_0)}^{\Psi(x)} -\frac{q}{\epsilon_S} \left[p_{p0} \left(e^{-\frac{\Psi(x)}{V_T}} - 1 \right) - n_{p0} \left(e^{\frac{\Psi(x)}{V_T}} - 1 \right) \right] \partial\Psi = \\ & = -\frac{q}{\epsilon_S} \left[p_{p0} \left(-V_T e^{-\frac{\Psi(x)}{V_T}} + V_T \right) - p_{p0}\Psi(x) - n_{p0} \left(V_T e^{\frac{\Psi(x)}{V_T}} - V_T \right) + n_{p0}\Psi(x) \right] = \\ & = -\frac{q}{\epsilon_S} \left[p_{p0}V_T \left(1 - e^{-\frac{\Psi(x)}{V_T}} \right) + n_{p0}V_T \left(1 - e^{\frac{\Psi(x)}{V_T}} \right) + \Psi(x)(n_{p0} - p_{p0}) \right] = \frac{(-\varepsilon(x))^2}{2} \end{aligned}$$

Abbiamo ottenuto un risultato molto importante: abbiamo espresso il campo elettrico $\varepsilon(x)$ in funzione del potenziale $\Psi(x)$: infatti:

$$\varepsilon(x) = \pm \sqrt{\frac{2q}{\epsilon_S} \left[p_{p0}V_T \left(1 - e^{-\frac{\Psi(x)}{V_T}} \right) + n_{p0}V_T \left(1 - e^{\frac{\Psi(x)}{V_T}} \right) + \Psi(x)(n_{p0} - p_{p0}) \right]}$$

Si noti però un fatto: in precedenza, avevamo già calcolato, velocemente, il campo nel punto $x = 0$, sfruttando le nostre competenze elettrostatiche

riguardanti lo studio di un semiconduttore. Come mai abbiamo rieflettuto questa operazione, complicandola oltretutto notevolmente? In effetti, le nostre precedenti competenze erano insufficienti ad esprimere interamente la fenomenologia nascosta dietro ciò che capita nel semiconduttore drogato tipo p che stiamo studiando: il precedente studio, teneva esclusivamente conto della carica di svuotamento, ignorando totalmente la linguetta, la carica di inversione. Questa avrà un'influenza assolutamente non trascurabile, in quanto provocherà un picco del campo ε , in un intorno del punto $x=0$.

Consideriamo il seguente fatto: se applicassimo il teorema di Gauss al sistema in studio, appoggiando un cilindro sull'ossido, vedremmo che sulla superficie laterale del cilindro, e sulla base sospesa, il campo sarebbe nullo. Di fatto, l'unica zona a campo elettrico non nullo è proprio la faccia appoggiata sull'ossido. Applicando dunque anche matematicamente il teorema, vediamo che la carica totale del semiconduttore S, è esprimibile come:

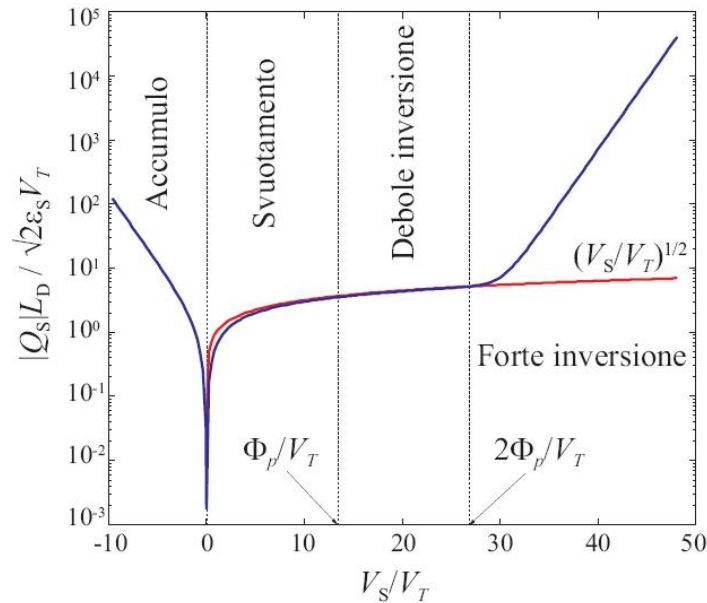
$$Q_S = -\varepsilon(0)\epsilon_S$$

Q_S è una funzione della carica totale nel semiconduttore, in funzione di V_S , rappresentante la carica su cm^2 . Ricordando che abbiamo precedentemente detto che $\Psi(0)$ rappresenta V_S , ossia il contributo del semiconduttore al potenziale di contatto, possiamo dire che il campo in 0 vale:

$$\varepsilon(0) = f(\Psi(0)) = f(V_S) = \pm \sqrt{\frac{q}{\epsilon_S} \left[p_{p0} V_T \left(1 - e^{-\frac{V_S}{V_T}} \right) + n_{p0} V_T \left(1 - e^{\frac{V_S}{V_T}} \right) + V_S (n_{p0} - p_{p0}) \right]}$$

Vedendo l'andamento del campo, vediamo che esso è funzione della tensione V_S , ma dunque anche Q_S sarà una funzione di V_S . Si è soliti rappresentare a questo punto il modulo della funzione Q_S , fondamentale per lo studio di un sistema MOS:

Quando $V_S = 0$, ci troviamo in zona di banda piatta. Quando $V_S < 0$, siamo in regione di accumulo, come abbiamo precedentemente introdotto. La parte interessante verrà dopo, quando $V_S > 0$: quando $V_S \sim 0,45$ V siamo in una zona di equilibrio. Man mano che andiamo avanti, da 0 in poi, avremo prima un regime governato da $\sqrt{V_S}$. Crescendo però la tensione V_S , si attivano, dentro la radice, gli esponenziali, che instaurano il regime di inversione dapprima debole, per poi stabilizzarsi nel regime di forte inversione. L'esponenziale è talmente potente che cresce praticamente verticalmente, da una certa tensione in poi, e in questo modo blocca la tensione: aldilà di un certo valore, si avrebbe una Q_S elevatissima, senza controllo, quindi abbiamo un blocco della tensione di controllo in questione. Il valore in cui la tensione



si blocca è, nella fattispecie, $V_S = 2\Phi_p$. Questa è l'informazione fondamentale che ci fornisce lo studio della funzione di Q_S .

Dall'andamento del grafico di Q_S possiamo dunque affermare che se un sistema MOS entra in stato di forte inversione (SI, Strong Inversion), possiamo considerare costante la tensione V_S , e uguale a $2\Phi_p$. Questo accade perchè, fino a questo limite, il semiconduttore riesce di fatto a contribuire al potenziale di contatto, prendendosi a carico metà della sua tensione, mentre l'altra metà andrà a carico dell'ossido (la V_{OX}). Aldilà del valore limite da noi descritto, il semiconduttore continuerà a prendere in carico esclusivamente $V_S = 2\Phi_p$, mentre continuerà a crescere la tensione a carico dell'ossido, V_{OX} , in grado di sostenerla, grazie alle proprie caratteristiche dielettriche.

5.3 Determinazione della tensione di soglia di un sistema MOS

Abbiamo parlato largamente della tensione V_S , ma non abbiamo idea di come studiare il suo andamento, al variare della tensione applicata al gate metallico, ossia al variare della nostra V_a esterna. Abbiamo infatti finora parlato dettagliatamente di tensioni interne al semiconduttore, e al dispositivo MOS in generale, ma senza creare collegamenti con l'esterno. Vogliamo dunque studiare la tensione di soglia nominale, ossia la minima tensione che porta un sistema MOS in stato di strong inversion.

L'idea che possiamo sfruttare è la seguente: studiare la distribuzione delle tensioni in un MOS partendo dal punto di equilibrio, per arrivare al valore critico di $V_S = 2\Phi_p$. Graficamente, possiamo immaginare che continui ad alzarsi il potenziale del metallo rispetto a quello del semiconduttore, e che al di sopra di una certa tensione, il semiconduttore non sia più in grado di sostenere V_S , e dunque V_a andrà interamente a ripartirsi sull'ossido, mentre a rimanere ferma a $2\Phi_p$ sul semiconduttore.

Possiamo scrivere un'equazione alla maglia di questo dispositivo:

$$V_a - V_{FB} = V_{OX} + V_S$$

Da qui, possiamo dire che:

$$V_{OX} = V_a - V_{FB} - 2\Phi_p$$

Come abbiamo già affermato in precedenza, l'ossido funziona come un condensatore a facce piane parallele dotato di un dielettrico tra le armature; la carica sul metallo, è di fatto come già affermato una Q_d , di carica uguale ed opposta a quella del semiconduttore (al fine di mantenere la neutralità globale).

Possiamo dividere il contributo della carica totale al semiconduttore, Q_S , come somma di due contributi: cariche dovute allo svuotamento, Q_d , e cariche dovute all'inversione, Q_n . Possiamo dire che:

$$Q_S = -(Q_d + Q_n) = C_{OX}V_{OX} \implies Q_S = C_{OX}(V_a - V_{FB} - 2\Phi_p)$$

Sappiamo che:

$$C_{OX} = \frac{\epsilon_{OX}}{T_{OX}}$$

Dunque, possiamo dire che la carica di inversione, Q_n , valga:

$$Q_n = -C_{OX}(V_a - V_{FB} - 2\Phi_p) - Q_d$$

Ma Q_d in questo ambito, è la carica di svuotamento, in stato di strong inversion. Se $V_S = 2\Phi_p$, ed è bloccata, allora possiamo dire che anche la variazione di ampiezza di regione di svuotamento, x_{dSI} , si blocchi:

$$x_{dSI} = \sqrt{\frac{2\epsilon_S V_S}{qN_A}}$$

In generale, è possibile esprimere la tensione V_S come area del triangolo rappresentante il campo elettrico nel semiconduttore:

$$V_S = \frac{qN_A x_d^2}{2\epsilon_S}$$

Possiamo dunque dire che la carica di svuotamento in queste condizioni valga:

$$Q_{dSI} = -qN_A x_{dSI} = -\sqrt{2\epsilon_S q N_A 2\Phi_p}$$

E quindi, la carica di inversione Q_n sarà:

$$Q_n = -C_{OX}(V_a - V_{FB} - 2\Phi_p) + \sqrt{2\epsilon_S q N_A 2\Phi_p}$$

5.3.1 Calcolo della tensione di soglia

Prima di terminare il calcolo della tensione di soglia, forniamo la definizione precisa di tensione nominale di soglia, V_T (nella fatispecie, si parlerà di V_{Tn} in caso di MOS a zona di inversione contenente elettroni, e dunque con un semiconduttore drogato p, e di V_{Tp} in un MOS con zona di inversione contenente lacune (e dunque semiconduttore drogato tipo n): V_T è la tensione che, applicata ad un sistema MOS, porta il sistema in forte inversione pur potendo considerare ancora quasi nulla la carica di inversione.

Cosa significa questa definizione? Partendo dall'equazione appena ricavata, ed imponendo una carica di inversione nulla, quindi $Q_n = 0$, vediamo che:

$$V_{Tn} = V_{FB} + 2\Phi_p + \frac{1}{C_{OX}} \sqrt{2q\epsilon_S N_A 2\Phi_p}$$

Questa è la tensione di soglia nominale: il limite che separa la strong inversion e la weak inversion.

Possiamo dunque riscrivere l'espressione della carica di inversione come:

$$Q_n = -C_{OX}(V_a - V_{Tn})$$

Cosa abbiamo scoperto, da quest'ultima riscrittura della carica di inversione? Di fatto il sistema MOS si comporta come un condensatore molto particolare: esso richiede una tensione di attivazione, pari alla tensione di soglia V_{Tn} , dopodichè è assimilabile a tutti gli effetti ad una capacità; nella fatispecie, $(V_a - V_{Tn})$ sarà proprio la tensione presente nel dispositivo MOS funzionante.

Alcune precisazioni sulla tensione di soglia: la tensione

$$V_{Tn} = V_{FB} + 2\Phi_p + \frac{1}{C_{OX}} \sqrt{2q\epsilon_S N_A 2\Phi_p}$$

è di norma positiva, altrimenti il MOS sarebbe costruito male. Se abbiamo una tensione $V_a < V_{Tn}$, non si ha uno strato di inversione. Valori tipici della tensione V_{Tn} variano da 0,2 a 1 V.

Abbiamo finora trattato solo ed esclusivamente n-MOS, ossia MOS con semiconduttore drogato tipo p, e dunque canale di inversione pieno di elettroni. Esiste una situazione del tutto duale, ossia il p-MOS, in cui la tensione di soglia varrà:

$$V_{Tp} = V_{FB} - 2\Phi_n - \frac{1}{C_{OX}} \sqrt{2q\epsilon_S N_D 2\Phi_n}$$

$$\Phi_n = \frac{kT}{q} \ln \frac{N_D}{n_i}$$

Attualmente, vengono utilizzate sia tecnologie n-MOS che p-MOS, in una tecnologia unica chiamata CMOS (Complementary MOS): si possono realizzare dispositivi MOS cioè con inversioni sia n che p: tecnologicamente, basandosi su di un unico substrato semiconduttivo debolmente drogato o comunque tendenzialmente intrinseco, si impiantano vaschette p-Well (per gli n-MOS), o n-Well (per i p-MOS): qua si costruiranno i transistori, mediante processi tecnologici.

Abbiamo parlato abbastanza in dettaglio del sistema MOS; descriviamo da ora in poi una trattazione della sua più interessante applicazione, ossia il transistor a effetto di campo basato sulla tecnologia MOS, il MOSFET.

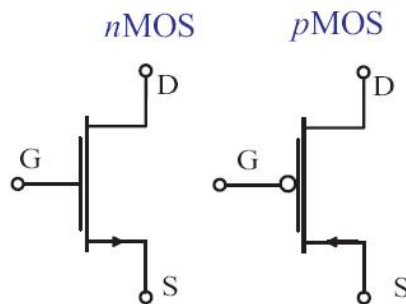
5.4 Il transistor MOSFET

Abbiamo già sommariamente parlato del transistor MOSFET, dicendo che esso rappresenta un sistema MOS il cui canale di inversione è adiacente a due pozzetti fortemente drogati (tipo n nel caso dell'n-MOS che stiamo studiando), uno di source ed uno di drain. La tensione V_a che abbiamo finora studiato è la tensione di gate, V_{GS} ; si prenderà in considerazione inoltre un'altra tensione, V_{DS} , ossia la tensione tra source e drain, positiva verso il drain. Una prima differenza con il transistor bipolare che si nota, è il fatto che i due pozzetti, source e drain, sono esattamente identici, a parte il potenziale in cui si trovano. In tutto il dispositivo, il source è il punto a potenziale più basso. La tensione V_{DS} è proprio la tensione:

$$V_{DS} = V_D - V_S$$

Dove normalmente però V_S si può anche considerare come punto a potenziale 0. La tensione V_{DS} sarà dunque sempre positiva: $V_{DS} \geq 0$. Ulteriore

fatto da considerare, è che tutti i transistori MOS che studieremo, avranno una tensione $V_{BS} = 0$: questa è una semplificazione che noi presentiamo, circuitalmente introducendo un corto circuito tra il terminale B (di Bulk, o Body), ed il terminale S del source. Circuitalmente come già detto S sarà collegato ad una massa (anche se qua sotto non lo indichiamo, consideriamolo come ipotesi).



V_{DS} sarà entrante verso il drain, e dunque gli elettroni saranno spinti dal source verso il drain: $V_{DS} > 0$, $I_{DS} \geq 0$.

Trascuriamo abbastanza nella nostra trattazione dei MOS i transistori p-MOS: possiamo velocemente dire che in un dispositivo di questo tipo, la regola operativa è invertire le tensioni: poichè il canale di inversione è infatti popolato da lacune, al fine di poter effettuare la transizione di carica, le tensioni saranno invertite, ed invertite saranno anche le convenzioni:

$$V_{DSp} < 0; V_{Tp} \leq 0; I_{DSp} \leq 0$$

La trattazione si baserà sul trattare generalmente un sistema n-MOS, talvolta citando il p-MOS.

5.5 Caratteristica statica del MOSFET: Analisi a canale graduale

Vogliamo trovare un'espressione operativa in grado di fornirci la caratteristica statica del MOSFET, nel caso di n-MOS, ossia l'andamento della sua corrente principale al variare delle tensioni. La corrente che interessa il dispositivo MOS sarà la corrente di canale di inversione, tra drain e source, I_{DS} . Possiamo dire che questa dipenda di fatto da due tensioni: V_{GS} , ossia la tensione che regola l'apertura del canale di inversione, e V_{DS} , ossia la tensione che permette il moto dei portatori all'interno del canale di inversione, dal source verso il drain.

$$I_{DS} = f(V_{GS}; V_{DS})$$

In realtà vi sarebbe anche una dipendenza della corrente dalla tensione tra bulk e source, V_{BS} , che noi abbiamo per semplificazione annullato mediante un cortocircuito tra i due terminali. Spesso ciò non è vero, quindi si ricordi che ci troviamo in un'ipotesi semplificativa, ma non sempre verificabile. Non consideriamo infatti, così facendo, un'eventuale polarizzazione del substrato.

Sappiamo che la corrente nel canale di un MOS c'è solo con $V_{GS} > V_{Tn}$: bisogna infatti far sì che il canale di inversione sia creato nel dispositivo, e perchè ciò sia sicuramente verificato, la tensione al gate deve essere almeno pari alla tensione di soglia, al fine di poter indurre lo stato di strong inversion nel sistema MOS. Cosa capita ora? Vorremmo capire cosa succede con una $V_{DS} > 0$, per vedere e studiare il moto degli elettroni nel canale. Possiamo intuire facilmente che il fatto che una tensione muova degli elettroni, significhi che la corrente è una corrente di trascinamento, di drift, e che quindi la tensione generi la corrente in questo modo. Si noti però un dettaglio non trascurabile: il canale di inversione è un canale resistivo, e con una resistenza non trascurabile, in quanto R è dell'ordine dei $K\Omega$. Il fatto che si abbia a che fare con resistenze così elevate, ci fa capire che una corrente di qualche mA, sia in grado di provocare una caduta di tensione dell'ordine dei V. Questo problema inoltre è reso ancora più difficile da analizzare dal fatto che la resistenza del canale non è uniforme.

Consideriamo ora il nostro sistema in due dimensioni: poichè l'asse delle x è già stato occupato dallo spostamento sul gate per andare fino al termine del bulk di silicio, consideriamo che la distanza tra i due canali sia su di un asse ortogonale ad x , e chiamiamolo y . Possiamo considerare nel punto $y=0$ il source, e ad una distanza L , nel punto $y=L$, il drain. Sappiamo solo dire una cosa, della funzione del potenziale del canale, Φ_{CH} : essa in 0 vale $\Phi_{CH}(0) = 0$, mentre in L , punto in cui si localizza il drain, $\Phi_{CH}(L) = V_{DS}$; non abbiamo la benchè minima idea di come vari però la tensione tra i due punti, quindi consideriamo per ora incognita la funzione $\Phi_{CH}(y)$.

Faremo quella che si dice analisi a canale graduale del dispositivo MOSFET: supponiamo di applicare una $V_{GS} > V_{Tn}$, al fine di trovarci in stato di strong inversion e dunque di aver la certezza di avere il canale di inversione aperto. Consideriamo una tensione positiva tra source e drain, e quindi $V_{DS} > 0$. Tentando di rappresentare in tre dimensioni il grafico, noteremo ciò: per $y=0$, e cioè in prossimità del source, avremmo il nostro andamento del potenziale già noto. Al variare di y , però, il potenziale aumenta sempre di più, poichè abbiamo applicato una tensione V_{DS} tra drain e source. Capitano alcuni effetti interessanti: se la tensione massima sopportabile al source

è $2\Phi_p$ per $y=0$, questa man mano cresce spostandoci nelle y , e, arrivando a $x=0$, $y=L$, si avrebbe $2\Phi_p + V_{DS}$. Una tensione da sopportare per il semiconduttore si può interpretare come aumento dell'ampiezza della regione di svuotamento, che quindi avrà un'ampiezza variabile con y .

L'andamento di Φ_{CH} influenza pesantemente tutto il grafico.

Ci interesserebbe dunque esprimere un'equazione dell'andamento del potenziale, ad una generica posizione y . Possiamo dire che:

$$V_{GS} - V_{FB} = V_{OX}(y) + V_S(y)$$

Dall'analisi a canale graduale, sappiamo che:

$$V_S(y) = 2\Phi_p + \Phi_{CH}(y)$$

Sostituendo nella precedente:

$$V_{GS} - V_{FB} = V_{OX}(y) + 2\Phi_p + \Phi_{CH}(y) \implies V_{OX}(y) = V_{GS} - V_{FB} - 2\Phi_p - \Phi_{CH}(y)$$

Possiamo tentare di introdurre, dal momento che conosciamo esattamente la capacità dell'ossido interpretato come dielettrico in un condensatore, la carica totale nel semiconduttore, considerabile come somma di carica di svuotamento Q_d e di carica di inversione Q_n :

$$-(Q_d(y) + Q_n(y)) = C_{OX}V_{OX}$$

Da qui, sostituendo:

$$-(Q_d(y) + Q_n(y)) = C_{OX}(V_{GS} - V_{FB} - 2\Phi_p - \Phi_{CH}(y))$$

La carica di inversione avrà un'espressione del tipo:

$$Q_n(y) = -C_{OX}(V_{GS} - V_{FB} - 2\Phi_p + \Phi_{CH}(y)) - Q_d(y)$$

Mediante dunque queste equazioni alla maglia, abbiamo trovato un'espressione della carica di inversione in una generica posizione y . Questa però è ancora incompleta: non siamo ancora riusciti ad esplicitare la carica di svuotamento nella generica posizione y , $Q_d(y)$.

Riflettiamo un attimo: le cariche di svuotamento sono strettamente legate alle zone di svuotamento, del dispositivo. Come sono fatte queste, nello stato che stiamo studiando? La tensione V_{DS} è positiva dal source S al drain D, e quindi vedremo una regione di svuotamento crescente nel canale da S a D, ed una zona molto estesa al drain, poichè una polarizzazione positiva su di un semiconduttore drogato tipo n provoca un allargamento della sua tensione di

svuotamento, poichè si effettua una polarizzazione di tipo inverso. La regione di svuotamento come abbiamo detto sarà dunque crescente, per culminare nel drain. Spostiamo dunque temporaneamente l'attenzione dalle cariche di svuotamento, all'ampiezza delle regioni di svuotamento: vogliamo ora determinare un'espressione di $x_d(y)$, ossia una funzione in grado di fornirci, in una generica y , l'ampiezza della regione di svuotamento in quel punto. Possiamo dire che x_d sia una funzione non direttamente di y , quanto della funzione del potenziale sul canale: come sappiamo il potenziale del canale è un valore aggiunto alla tensione totale nel dispositivo, e che effettivamente l'ampiezza della regione di svuotamento ne risenta molto:

$$x_d(y) \simeq x_d(\Phi_{CH}(y)) = \sqrt{\frac{2\epsilon_S(2\Phi_p - \Phi_{CH}(y))}{qN_A}}$$

Questo risultato è semplicemente un'applicazione dei classici studi elettrostatici effettuati sul semiconduttore, precedentemente.

Possiamo dire che la carica di svuotamento al variare di y sia semplicemente l'area del rettangolo, il prodotto tra l'altezza (carica in un punto) e di ampiezza della regione di svuotamento, utilizzando l'ipotesi di completo svuotamento:

$$Q_d(y) = -qN_A x_d(y) = -\sqrt{2q\epsilon_S N_A(2\Phi_p + \Phi_{CH}(y))}$$

Possiamo, dunque, ora che abbiamo ottenuto un'espressione operativa della $Q_d(y)$, ricavare la Q_n , sostituendo nell'equazione ricavata in precedenza. Ricordandola, dunque:

$$Q_n(y) = -C_{OX}(V_{GS} - V_{FB} - 2\Phi_p - \Phi_{CH}(y)) - Q_d$$

Vediamo:

$$Q_n(y) = -C_{OX}(V_{GS} - V_{FB} - 2\Phi_p - \Phi_{CH}(y)) + \sqrt{2q\epsilon_S N_A(2\Phi_p + \Phi_{CH}(y))}$$

Esiste un modo alternativo di esprimere quest'equazione, mediante alcuni artifici algebrici: se sommiamo e sottraiamo il termine $\sqrt{2q\epsilon_S N_A}$, si ottiene che:

$$Q_n(y) = -C_{OX}(V_{GS} - V_{FB} - \Phi_{CH}(y)) + \sqrt{2q\epsilon_S N_A} \left(\sqrt{2\Phi_p + \Phi_{CH}(y)} - \sqrt{2\Phi_p} \right)$$

Ora che abbiamo studiato dettagliatamente le cariche, vogliamo finalmente passare al calcolo delle correnti, e nella fattispecie di I_{DS} . Come già detto,

la corrente in un MOSFET è prevalentemente di trascinamento; in condizioni statiche deve esserci in ogni punto del dispositivo, la stessa corrente. La corrente che consideriamo parte dal source ed arriva al drain, quindi procede in verso positivo sulla direzione dell'asse delle y . Consideriamo dunque tra source e drain un punto generico y , ed un punto ad esso vicinissimo, $y + dy$. Poichè la corrente in questa situazione è di drift, possiamo quantificarla in questo punto dicendo che:

$$I_{drift}(y) = -q\mu_n n_{INV}(y)\varepsilon(y)WT_{INV} = -I_{DS}$$

Cosa significa ciò che abbiamo fatto: il numero di cariche di inversione $n_{INV}(y)$ nel generico punto y moltiplicato per la carica fondamentale q e per la mobilità μ_n , con trascinamento dovuto al campo elettrico nel punto y , $\varepsilon(y)$, moltiplicato per la sezione attraversata, ossia il prodotto del lato W e dello spessore dello strato di inversione T_{INV} , ci fornisce, con una convenzione di segno opposto, la corrente desiderata.

Quest'espressione contiene due termini che non abbiamo mai discusso o nominato, ossia lo spessore dello strato di inversione, e il numero di elettroni nello strato di inversione, rispettivamente T_{INV} e n_{INV} . Possiamo però dire che

$$-qT_{INV}(y)n_{INV}(y) = Q_n(y)$$

Infatti la funzione $Q_n(y)$, precedentemente ricavata, è proprio la carica totale di inversione nello straterello, per centimetro quadrato. Di fatto, non conosciamo le due funzioni richieste, ma conosciamo il loro prodotto, e così possiamo sostituirvi l'espressione della carica totale per centimetro quadro precedentemente introdotta, al fine di calcolare la corrente:

$$-I_{DS} = \mu_n Q_n(y)\varepsilon(y)$$

Sappiamo anche un'altra cosa: il campo elettrico trascinante gli elettroni è la derivata del potenziale di canale rispetto alla generica posizione y :

$$\varepsilon(y) = -\frac{\partial\Phi_{CH}(y)}{\partial y}$$

Possiamo dunque dire una cosa: sappiamo, qualitativamente, che il potenziale nel canale è una funzione monotona crescente, e che quindi la derivata sarà una funzione sempre positiva. Possiamo dunque mantenere i segni come abbiamo fatto, poichè rispetteranno la convenzione utilizzata.

Sostituiamo tutto, e vediamo che:

$$-I_{DS} = \mu_n W (-Q_n(y)) \left(-\frac{\partial \Phi_{CH}(y)}{\partial y} \right)$$

Integrando ambo i membri per tutta la lunghezza del canale di inversione, e quindi da 0 ad L:

$$\int_0^L I_{DS} \partial y = - \int_{\Phi_{CH}(0)}^{\Phi_{CH}(L)} Q_n(y) \partial \Phi_{CH}(y)$$

Ma noi sappiamo che $\Phi_{CH}(0) = 0$, e che $\Phi_{CH}(L) = V_{DS}$. Possiamo dunque dire che:

$$I_{DS} = -\frac{\mu_n W}{L} \int_{\Phi_{CH}(0)}^{\Phi_{CH}(L)} Q_n(y) \partial \Phi_{CH}(y)$$

Abbiamo tutti i dati: la funzione $Q_n(y)$ infatti sarebbe:

$$Q_n(y) = -C_{OX}(V_{GS} - V_{FB} - \Phi_{CH}(y)) + \sqrt{2q\epsilon_S N_A} \sqrt{2\Phi_p + \Phi_{CH}(y) - \sqrt{2\Phi_p}}$$

La letteratura classica di dispositivi è solita trascurare, nell'integrazione, tutti i termini sotto radice, al fine di semplificare i conti. Questa è un'approssimazione molto brutale, e molto grossolana, ma che semplifica a tal punto i conti da essere comunemente usata. Un simulatore numerico per circuiti, quale ad esempio SPICE, considera la formula nella sua forma più completa. Considerando ad ogni modo esclusivamente i termini polinomiali:

$$I_{DS} = \frac{\mu_n C_{OX} W}{L} \left[(V_{GS} - V_{Tn}) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

Questa è la caratteristica statica del transistor MOSFET, ed è il risultato che utilizzeremo in applicazioni pratiche. A partire da qua, possiamo trovare una grandissima differenza con il transistor bipolare: il MOSFET ha il gate isolato, ossia in cui non scorre una corrente. Nella base del BJT, invece, vi è una minima corrente.

Ci chiediamo ora, leggendo l'equazione della caratteristica statica di un n-MOS, quale sia il suo andamento. Poichè non abbiamo un'equivalente della I_B , non avremo un parametro in corrente da regolare. Lo studio della caratteristica statica, si può effettuare in questo modo: la variazione di I_{DS} , al variare della tensione V_{DS} . Se applichiamo diverse tensioni di gate, tutte in grado di indurre uno stato di strong inversion, ad esempio:

$$V_{Tn} < V_{GS1} < V_{GS2} < V_{GS3}$$

Possiamo vedere che per piccole V_{DS} , il MOS si comporta all'incirca come una resistenza. Si ha, crescendo la V_{DS} , una zona detta regione quadratica, dove l'andamento della corrente è parabolico, ossia raffigurante una parabola con coefficiente quadratico negativo (parabola rivolta verso il basso). Sappiamo che quindi la funzione dovrebbe essere monotona crescente fino al massimo, per poi divenire decrescente. Calcoliamo dunque il massimo della funzione, derivandola rispetto alla variabile V_{DS} , e ponendola nulla:

$$\frac{\partial I_{DS}}{\partial V_{DS}} = 0 \implies \frac{\mu_n W C_{OX}}{L} [(V_{GS} - V_{Tn}) - V_{DS}] = 0$$

Questo si verifica quando:

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

Sostituendo dunque l'espressione del massimo appena ricavata nella funzione di I_{DS} , si ottiene:

$$I_{DS,Max}(V_{DS}) = \frac{\mu_n W C_{OX}}{L} \left[(V_{GS} - V_{Tn})(V_{GS} - V_{Tn}) - \frac{(V_{GS} - V_{Tn})^2}{2} \right];$$

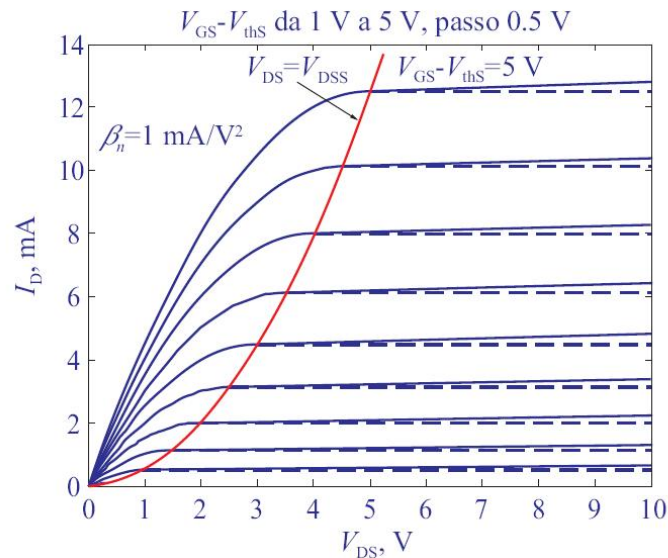
$$I_{DS,Max}(V_{DS}) = \frac{\mu_n W C_{OX}}{2L} (V_{GS} - V_{Tn})^2$$

A seconda della tensione V_{GS} applicata sul gate, abbiamo dunque un diverso valore del massimo della corrente, quantificato. Dopo il massimo però, la caratteristica dovrebbe essere decrescente.

Sperimentalmente, si verifica un andamento molto simile a quello che effettivamente avveniva nel BJT:

Le caratteristiche crescono tutte, al variare del parametro della tensione di gate V_{GS} , e si passa così dalla regione quadratica, ad una regione detta di saturazione (si noti che non centra niente con la regione di saturazione del BJT: quella che qua è la regione di saturazione, corrisponde alla regione attiva diretta del BJT). Possiamo dunque dire che, per $V_{DS} > V_{DSMax}$, $I_{DS,Sat} = I_{DS,Max}$. Possiamo dunque riassumere il modello statico, esprimendolo in modo purtroppo non elegante per quanto corretto (a parte per la grossolana approssimazione precedentemente fatta e criticata), mediante una funzione a tratti:

- Per $V_{GS} < V_{Tn}$, la $I_{DS} = 0$, ed il transistor MOSFET è OFF;
- Per $V_{GS} > V_{Tn}$, e $V_{DS} < V_{GS} - V_{Tn}$, il transistor è in zona quadratica, e:



$$I_{DS} = \frac{\mu_n C_{OX} W}{L} \left[(V_{GS} - V_{Tn}) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

- Per $V_{GS} > V_{Tn}$, e $V_{DS} > V_{GS} - V_{Tn}$, il transistorore è in zona di saturazione, e dunque:

$$I_{DS} = I_{DS,Sat} = \frac{\mu_n W C_{OX}}{2L} (V_{GS} - V_{Tn})^2$$

Possiamo ora cercare di rispondere all'ultimo dei nostri dubbi, prima di considerare del tutto terminata la trattazione della caratteristica statica in un MOSFET: perchè la corrente aldilà del valore di massimo diviene costante? (Anche se si ha un effetto duale dell'effetto Early, che accenneremo).

Proviamo a pensare alla distribuzione delle cariche nel sistema n-MOS che stiamo studiando, esaminando le varie zone che abbiamo espresso nella definizione della funzione a tratti della caratteristica statica.

- Date $V_{GS} = 0$, $V_{DS} = 0$, non abbiamo variazioni delle zone di svuotamento rispetto alla condizione di equilibrio, quindi possiamo pensare che la carica sia interamente di svuotamento poichè non abbiamo uno stato di strong inversion;
- Data $V_{GS} > V_{Tn}$, $V_{DS} = 0$, capita che il sistema n-MOS va in stato di strong inversion, vi è un aumento dello spessore del canale, dal momento che si induce la carica di inversione Q_n (precedentemente nulla), e

si ha un aumento generale della tensione, che porta anche all'aumento della zona di svuotamento. In prossimità dei pozzetti drain e source però non si ha una variazione sostanziale della zona di svuotamento.

- Se $V_{GS} > V_{Tn}$, $V_{DS} < V_{GS} - V_{Tn}$, capita che si entra nella zona di conduzione del canale, le cariche si spostano dal source al drain, e così si ha il dispositivo in zona di saturazione. Vediamo che la zona di svuotamento tenderà molto a crescere andando dal source verso il drain, per effetto della tensione di canale Φ_{CH} , o se vogliamo della V_{DS} che polarizza il drain. Ma ciò che sarà interessante non sarà solo la crescita della carica di svuotamento (associata alla crescita dell'ampiezza della regione di svuotamento), bensì anche la diminuzione della carica di inversione: man mano che la tensione aumenta, la carica di inversione diminuisce, quindi, man mano che si procede dal source verso il drain, la carica di inversione tenderà a diminuire. Questo perchè capita il seguente fatto: aumentando la tensione sul drain, questo per reggerla deve ridurre la carica di inversione, e così in prossimità del drain, l'andamento della carica andrà a strozzarsi. Per un caso limite, $V_{GS} > V_{Tn}$, $V_{DS} = V_{GS} - V_{Tn}$, la carica di inversione si annullerà sul drain, e così si avrà una sorta di triangolo con base al source e vertice al drain. Considerando una tensione ancora più elevata, $V_{GS} > V_{Tn}$, $V_{DS} > V_{GS} - V_{Tn}$, avviene quello che si chiama Channel Length Modulation (modulazione di ampiezza di canale): questo è il duale dell'effetto Early del BJT: il punto ottenuto dalla precedente osservazione alla tensione limite, tende ad arretrare, così che la lunghezza del canale resistivo anzichè L sarà L' , e la zona ($L-L'$) restante dalla parte del drain sarà una zona ad alto campo elettrico, causato dalla forte tensione applicata sul drain. Gli elettroni giunti in questa zona verranno sparati nel drain da questo fortissimo campo provocato dal forte potenziale applicato, tuttavia la corrente non aumenterà. Questo modello si può assimilare ad un modello idrodinamico: data una cascata la cui pendenza aumenta, ipotizziamo come la nostra tensione, arrivata ad un certo limite si fermerà, e non potrà più aumentare la corrente d'acqua. Questo avviene perchè l'entità della corrente viene regolata dalla zona resistiva, e, al di sopra di un certo limite, il flusso di portatori non aumenterà.

5.6 Esercizio Pratico

Dato un sistema MOS costituito da un gate di polisilicio drogato n^+ , un ossido di spessore 300 Å di quarzo, ed un substrato di silicio drogato

$5 \cdot 10^{16}$, con un lato L di $1\mu m$, $W = 5\mu m$, ed una mobilità elettronica $\mu_n = 400$:

1. Si determini il diagramma a bande all'equilibrio quotato, e si stabilisca in quale condizione si trova il semiconduttore.
2. Si calcoli la tensione di soglia.
3. Supponendo di disporre di una tensione di alimentazione pari a 5 V, si calcoli la massima corrente erogabile dal dispositivo MOSFET.
4. Costruire un modello di ampio segnale del dispositivo MOSFET.
5. Si valuti il modello di piccolo segnale alle condizioni del punto 3.

5.6.1 Risoluzione

Diagramma a bande all'equilibrio quotato

Partiamo con una considerazione: al posto del gate metallico, ne abbiamo uno di silicio policristallino. Ciò non ci provoca alcun problema, in quanto esso è del tutto analogo ad un metallo, sotto ogni punto di vista, a parte per l'affinità elettronica, che al posto di 4,1 varrà 4,05. Possiamo immediatamente valutare il lavoro di estrazione partendo da $q\Phi_p$, come:

$$q\Phi_p = E_{F_i}(\infty) - E_F = kT \ln \frac{N_A}{n_i} = 0,39eV$$

$$q\Phi_{S_p} = q\Phi_p + q\chi + \frac{E_g}{2} = 5eV$$

Possiamo dunque dire che il potenziale di contatto, ossia la differenza tra i potenziali di metallo e semiconduttore, valga:

$$q\Phi_{S_p} - q\Phi_M = 0,95eV$$

Sappiamo che il potenziale di contatto allora, analogamente, Φ_{MS} , varrà 0,95 eV.

Calcoliamo la capacità dell'ossido, C_{OX} :

$$C_{OX} = \frac{\epsilon_{OS}}{T_{OX}}$$

Possiamo ora calcolare l'ampiezza della regione di svuotamento nel semiconduttore, come:

$$x_{d0} = \frac{2\epsilon_S}{C_{OX}} + \sqrt{\left(\frac{2\epsilon_S}{C_{OX}}\right)^2 - \frac{\Phi_{MS}2\epsilon_S}{qN_A}} = 90nm$$

Sappiamo che:

$$V_S = \frac{qN_A x_{d0}^2}{2\epsilon_S} = 0,32V; \quad V_{OX} = \Phi_{MS} - V_S = 0,63V$$

Ci troviamo in una zona circa di equilibrio, quindi di fatto siamo in ottime condizioni: il nostro dispositivo MOS è costruito ottimamente.

Si calcoli la tensione di soglia

Possiamo passare al secondo punto, ossia al calcolo della tensione di soglia. Questo è semplicemente una banale applicazione della teoria; ricordando che:

$$V_{Tn} = V_{FB} + 2\Phi_p + \frac{1}{C_{OX}} \sqrt{2q\epsilon_S N_A 2\Phi_p}$$

Essa si può anche esprimere come:

$$V_{Tn} = V_{FB} + 2\Phi_p + \gamma_B \sqrt{2\Phi_p}$$

Dove il termine γ_B è il body factor, ed è uguale a:

$$\gamma_B = \frac{\sqrt{2q\epsilon_S N_A}}{C_{OX}}$$

Il significato del body factor può tornare molto utile, in quanto abbiamo calcolato la tensione di soglia, nella teoria, considerando $V_{BS} = 0$ (per il solito discorso del cortocircuito tra body/bulk e source. Di fatto, il body factor considera anche eventuali tensioni tra body e source, quindi possiamo tenere conto di questa relazione, che potrebbe aiutarci in casistiche generali. Usando una qualsiasi delle due relazioni, si può facilmente verificare che:

$$V_{Tn} = 4,18V$$

Data alimentazione di 5 V, si calcoli la massima corrente erogabile dal dispositivo MOSFET

Passando al terzo punto, il problema ci fornisce come tensione massima 5 V. Dalla teoria, sappiamo che, più elevata è V_{GS} , più potremo ottenere correnti elevate. Possiamo dunque porre senza molti dubbi, $V_{GS} = 5 V$.

Per quanto riguarda invece la tensione tra drain e source, V_{DS} , è sufficiente che essa sia maggiore o uguale del massimo:

$$V_{DS} \geq V_{GS} - V_{Tn} = 5 - 0,82 = 4,18$$

Possiamo dunque considerare per comodità $V_{DS} = 4,18V$, e calcolare la corrente in saturazione, come il massimo della regione quadratica:

$$I_{DS} = \frac{\mu_n C_{OX} W}{2L} (V_{GS} - V_T)^2 = 230 \mu A V^{-2}$$

$$I_{D,Sat} = 2mA$$

Costruire un modello di ampio segnale del dispositivo MOSFET

Passiamo al quarto punto, più teorico, dell'esercizio: la costruzione di un modello di ampio segnale del MOSFET. Il modello di ampio segnale come sappiamo deve saper prevedere le condizioni del dispositivo considerando qualsiasi variazione di tensione con qualsiasi rapidità. Il modello di ampio segnale deve essere il più generale di un dispositivo e anche in grado di prevedere comportamenti statici. Da un punto di vista logico, tutto il modello si baserà su di un generatore di corrente I_{DS} , attorno al quale inseriremo ulteriori elementi. Conosciamo già l'andamento di questo, considerando staticità. In realtà abbiamo però effetti capacitivi, e due diodi (dovuti ai pozzetti di source e drain). Esiste dunque una capacità di svuotamento sia per il source ($C_{DEP}(0)$), che per il drain ($C_{DEP}(-V_{DS})$). Si noti che si usa $-V_{DS}$ poichè, dal momento che il pozzetto di drain è drogato n+, e la tensione su di esso è positiva, allora avremo una polarizzazione inversa sul drain.

Non abbiamo ancora fatto considerazioni sul gate: sul gate abbiamo infatti un condensatore a capacità variabile, poichè anche il gate ha la sua capacità: dal gate si vedono due dielettrici: se il transistor è spento, per V_S in zona di equilibrio, siamo in zona di svuotamento nel grafico della Q_S , ossia della carica del semiconduttore. Man mano che si aumenta la V_S , si arriva in zona di weak e poi di strong inversion, e così varia radicalmente la carica contenuta nel semiconduttore.

Sappiamo che la capacità dell'ossido è costante, e che del semiconduttore conosciamo l'andamento della carica complessiva, $Q_S(V_S)$. Possiamo dunque calcolare la capacità variabile del semiconduttore, come:

$$C_S = \frac{\partial Q_S}{\partial V_S}$$

Semplicemente, sarà sufficiente mettere la capacità C_S in serie con la capacità C_{OX} , per completare il modello di ampio segnale.

Si valuti il modello di piccolo segnale alle condizioni del punto 3

Passiamo al modello di piccolo segnale, alle condizioni del terzo punto del problema: in quelle condizioni, il transistor era saturo, e quindi il modello di piccolo segnale è sensato. Scegliamo di considerare come segnale variabile per il modello, la V_{GS} : al fine di modulare la corrente I_{DS} , scegliamo semplicemente di modificare l'apertura del canale di inversione. Possiamo dunque dire che:

$$V_{GS}(t) = V_{GSop} + v_{gs}(t)$$

Sappiamo che:

$$I_{DS} = \frac{\beta_n}{2}(V_{GS} - V_T)^2$$

Da ciò, linearizzando:

$$I_{DS} \simeq \frac{\beta_n}{2}(V_{GSop} - V_T)^2 + \left. \frac{\partial I_{DS}}{\partial V_{GS}} \right|_{V_{GSop}} \cdot v_{gs}(t)$$

Possiamo dire che $i_{ds}(t) = I_{DS}(V_{GS}) - I_{DSop}$; da ciò, possiamo definire il parametro del generatore che useremo in un modello a π analogo o quasi a quello del BJT, ma con una capacità tra gate e source, al posto di una resistenza.