

A questo punto è possibile introdurre alcune caratteristiche fondamentali di eterostruttura; si fa riferimento alle slide

Eterogiunzione tra materiali intrinseci

- Tipo 1: il livello di Fermi del materiale B, a destra, è maggiore di quello in A; questo significa che B cede elettroni a A, dove ci sono livelli energetici inferiori disponibili. Dunque, mentre le lacune da A andranno in B. Il potenziale di built-in, qV_{bi} , si può calcolare come:

$$qV_{bi} = |(E_{FA} - U_0) - (E_{FB} - U_0)| = -q\Phi_A - (-q\Phi_B) = \\ = \left| -\left(q\chi_A + \frac{E_{gA}}{2}\right) - \left(-q\chi_B - \frac{E_{gB}}{2}\right) \right| = \left| (-q\chi_A + q\chi_B) + \left(-\frac{E_{gA}}{2} + \frac{E_{gB}}{2}\right) \right|$$

- Tipo 2: A cede elettroni, B cede lacune, come si può osservare dal diagramma a bande.

Eterogiunzione nn

Consideriamo solo il "tipo 1": ora abbiamo due lati drogati; un fatto interessante è il seguente: l'accumulo di elettroni in A comporta la formazione di una regione in cui il livello di Fermi si trova dentro la banda di conduzione, ottenendo un elemento "quasi degenero". A sinistra il potenziale è più "positivo", dunque l'energia necessaria per andare "a sinistra" più elevata, dunque questi elettroni si trovano di fatto confinati in una buca di potenziale: si ha una regione fortemente popolata da elettroni, che però possono solo muoversi "ortogonalmente al foglio", dal momento che compresi tra due barriere di potenziale.

Il fatto che si abbia questo possibile movimento solo in questa direzione, o comunque solo non lungo la direzione perpendicolare alla giunzione, genera il cosiddetto "gas bidimensionale di elettroni".
 Ciò è utile al fine di realizzare particolari transistori a effetto di campo. Il potenziale di built-in sarà legato anche al drogaggio.

Eterogiunzione PP

All'estremità simile a prima, anche se ora forse conviene incominciare a ragionare sulle lacune. Si consideri la struttura tipo "2" (per esempio): il materiale B ha $E_F < E_{VB}$, dunque le lacune vanno da B verso A. In B si ha una regione di svuotamento di lacune (carica fissa e negativa), in A una regione di accumulo di lacune.

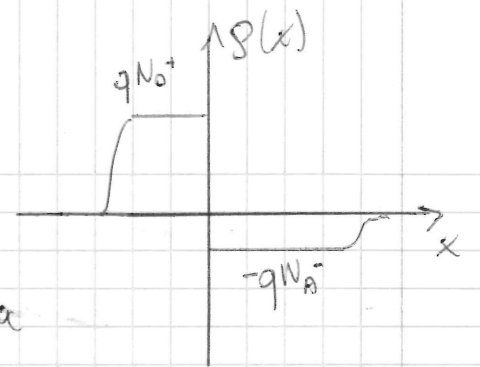
In A si crea una regione che fa da buca di potenziale per le lacune, fortemente popolata da esse (si noti che E_F , durante a prima, sta in banda di valenza): ricordandosi il punto di vista delle lacune, si può vedere che esse sono "intrappolate" in una buca di potenziale, un po' come gli elettroni nel caso precedente.

In questo caso si ha un gas bidimensionale di lacune.

Eterogiunzione p-n

Si considera a questo punto l'eterogiunzione p-n di tipo 1.

Di essa ci sono due sottocasi: livello di Fermi superiore in A, o in B. Si consideri il primo caso: livello di Fermi maggiore nel lato "n", ossia "A". Gli elettroni tendono dunque a muoversi da A a B (e le lacune da B a A), come nella meno sofisticata giunzione p-n "classica". In ciascuno dei due lati si formeranno regioni svotate relative ai maggioritari rispettivi ceduti.



Considerando il lato p meno drogato (in questo specifico esempio), si avrà una distribuzione di carica di questo tipo.

Calcoliamo il potenziale di built-in:

$$qV_{bi} = q\Phi_B - q\Phi_A;$$

$$q\Phi_B = q\chi_B + E_{gB} - [E_F - E_{vB}] = q\chi_B + E_{gB} - k_B T \ln\left(\frac{N_{A,B}}{N_{v,B}}\right);$$

$$q\Phi_A = q\chi_A - [E_F - E_c] = q\chi_A - k_B T \ln\left(\frac{N_{D,A}}{N_{c,A}}\right);$$

$$qV_{bi} = q\Delta\chi_{AB} + E_{gB} + k_B T \ln\left(\frac{N_{A,B} N_{D,A}}{N_{v,B} N_{c,A}}\right)$$

Si può verificare la validità dell'espressione, effettuando un confronto con il caso della "omogiunzione p-n": con la giunzione "classica", infatti,

$$q\Delta\chi = 0;$$

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

Dunque:

$$qV_{bi} = \underbrace{-E_{gA} + k_B T \ln(N_{cA} N_{vA})}_{= k_B T \ln(n_i^2)} - k_B T \ln(N_A N_D) =$$

$$= -k_B T \ln\left(\frac{N_A N_D}{n_i^2}\right) \quad \left[\begin{array}{l} \text{Espressione molto simile alla} \\ \text{precedente!} \end{array} \right]$$

Aggiunger altro?!

Confinamento spaziale di carica libera

È stato precedentemente visto qualche esempio in cui si ha la formazione di buche di potenziale in banda di conduzione σ di valenza, confinando i portatori.

Se la buca è molto piccola, oltre a limitare i momenti delle cariche, si può avere una quantizzazione dei livelli energetici: se la buca è piccola i livelli energetici, a causa del principio di esclusione di Pauli, diventano discreti come negli atomi, ottenendo una quantizzazione dei portatori.

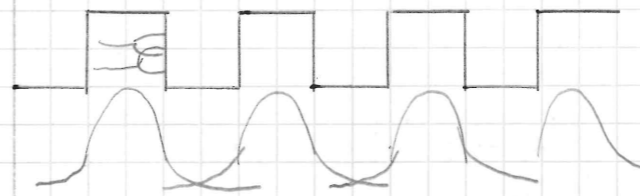
Normalmente, le informazioni sulla natura quantistica dei portatori sono racchiuse nella massa efficace, ma, dal momento in cui si creano queste buche di potenziale, se esse sono abbastanza sottili, bisogna tener conto di questo ulteriore effetto di discretizzazione.

Il confinamento può essere bidimensionale, monodimensionale, puntuale:

- monodimensionale: si toglie 1 grado di libertà, dunque i portatori sono liberi di muoversi in 2 direzioni; questa è una "buca quantica", o "quantum well".
- bidimensionale: si tolgono due gradi di libertà, permettendo il movimento solo su di un "filo quantico" (quantum wire).
- tridimensionale: gli elettroni sono confinati in un "quantum dot", e non possono muoversi.

A causa della discretizzazione degli stati, la funzione di densità efficace degli stati non va più con la radice quadrata, ma con un gradino che segue l'inviluppo della radice quadrata.

La cosa bella è che però ora è possibile determinare le soluzioni analitiche (senza passare dagli integrali alle sommatorie); trattando per esempio il gas bidimensionale di elettroni, però, ora la soluzione sarà una concentrazione per unità di area. Ci porremo a questo punto una domanda: cosa capita quando si affiancano diversi quantum well?



La probabilità che, per la meccanica quantistica, gli elettroni stiano dentro o fuori dalla buca, è simile a una funzione "cosinusoidale con coda esponenziale".

Se gli elettroni stanno in buche diverse con elevata probabilità, nessun problema: possono anche occupare gli stessi livelli energetici, senza contraddire Pauli; questo capita se le buche sono ben separate, e le "code" ben attenuate.

Se le buche sono vicine, le code si sovrappongono "presto", dunque la probabilità che gli elettroni siano nello stesso livello energetico e nella stessa buca aumentano; Pauli non permette una cosa del genere, dunque i livelli energetici delle buche vengono perturbati, splittandosi.

A seconda di quante buche si mettono vicine, i livelli si splittano molte volte, tendendo a degenerare in una banda (spettro discreto molto fine). Con queste strutture si può "idealizzare" il

semiconduttore, rendendo "semplici" da stabilire i parametri. Queste strutture sono dette "super-reticoli".

Eterostruttura a modulazione di drogaggio

Si consideri la seguente struttura:

Gli elettroni vanno da A a B, dunque in B si ha una regione arricchita di elettroni, in altre di potenziale "triangolare".
Se essa è fatta a senso, si introduce una quantizzazione dei livelli energetici perenni.

Dato dall'altra parte un materiale intrinseco, l'interazione con il materiale drogato fa cambiare la concentrazione di portatori, come se fosse drogato; quello che si ha è un grosso numero di elettroni, senza però avere ioni donatori, che introducono impurità nel reticolo; in questo modo non si hanno le interazioni coulombiane, che degraderebbero la mobilità elettronica (la quale infatti diminuisce al crescere del drogaggio!). Si hanno dunque elettroni con una grossa mobilità, dal momento che non ci sono donatori che la riducono.

JFET e MESFET

Siamo sul punto di studiare le prime strutture "importanti" a eterostruttura: i FET a eterostruttura.

Come noto, i FET sono dispositivi in cui si ha un canale conduttivo la cui conducibilità viene modulata attraverso la tensione sul terminale di gate. Esistono due modi per realizzare ciò:

- 1) Modificare il numero di portatori liberi presenti nel canale
- 2) Modificare la lunghezza della sezione trasversale del canale.

Nei MOSFET, come visto in precedenza, il controllo di corrente è causato da un cambio del numero di portatori, mediante una variazione del campo elettrico, il quale, attraverso il teorema di Gauss, è collegato alla carica presente nel canale.

I FET a giunzione e i MESFET sono basati sul secondo principio: variando il campo elettrico si varia la sezione trasversale del canale.

Si hanno:

- JFET: molto utili nelle applicazioni a basso rumore;
- MESFET: utili perché spesso realizzati con GaAs, dunque molto di moda qualche anno fa, anche se oggi si tende a preferire MOSFET o eterogiunzioni.

Qual è l'idea di base? Beh, al solito, si hanno source e drain (contatti ohmici su strato epitassiale), collegati da un canale conduttivo, in prossimità di una giunzione. Come nei MOS non si vede una corrente di gate, dunque per isolare serve o un substrato p poco drogato e polarizzato inversamente, o un semi-isolante, o un materiale

con E_g sufficientemente elevato.

L'idea è la seguente: si "suota" il canale conduttivo, al fine di "chiuderlo", mediante una giunzione p-n polarizzata inversamente: si mette una regione p+ sotto al gate, ma molto drogata, in modo da avere una regione molto estesa di svuotamento nel canale, che sarà meno drogata. Applicando una tensione all'estensione negativa la regione di svuotamento si "mangia" il canale.

Nel MESFET cambia solo il modo con cui si ottiene lo svuotamento: mediante un contatto metallo-semiconduttore non ohmico in polarizzazione inversa.

Nella realtà, c'è un po' di differenza: nella giunzione Schottky, come visto, i minoritari non hanno importanza, dunque questo dispositivo è intrinsecamente più veloce del FET a giunzione. Usando poi il GaAs, le prestazioni migliorano ulteriormente.

Altra differenza con il JFET, immaginare una struttura "complementare" è abbastanza facile; nel MESFET è peggio, perché, negli elementi dei gruppi 3 e 5, mentre la mobilità degli elettroni è molto alta, quella delle lacune è molto bassa, rendendo inutili strutture di questo tipo.

Caratteristica statica

L'andrea della caratteristica statica è all'estensione simile a quella del MOSFET: anche ora, data I_0 nel canale, si ha una caduta di potenziale distribuita lungo di esso: $\phi_{ch}(y)$; questa cosa è piuttosto fastidiosa.

Come per il MOSFET, $\phi_{ch}(y)$ è una funzione monotona crescente dal source verso il drain: essa tende a polarizzare sempre più inversamente la giunzione, in ciascun punto, man mano che ci si avvicina al drain. Aumentando $\phi_{ch}(y)$ aumenta dunque lo svuotamento, senza però toccare il numero di elettroni nel canale. Si può calcolare facilmente che:

$$x_d(y) = \sqrt{\frac{2\epsilon_s}{qN_D} [V_{bi} - (V_{GS} - \phi_{ch}(y))]}$$

Per un certo valore di V_{GS} , dunque, il canale sarà strozzato; si tenga conto che la fenomenologia è diversa da quella del MOSFET, dal momento che ora si ha non una riduzione di carica, bensì un reale strozzamento, determinato dalla modulazione della regione di svuotamento.

Si potrebbe ri-ricavare lo stesso modello a controllo di carica del MOSFET: si tenga conto del fatto che ora, per nuova Q_n , basta ricordare che dove non si ha strozzamento gli elettroni sono N_D , dunque tener conto nell'integrale del fatto che da "d" la sezione si riduce a "d - x_d ":

$$Q_n(\phi_{ch}) = -q \int_0^{l+\infty} n(x,y) dy = -q N_D [d - x_d(y)]$$

Sostituendo ciò nei calcoli già visti col MOSFET, si può ricavare I_0 ; il modello sarà poi linearizzato, e così via.

Come per il MOSFET, inoltre, è possibile definire una "tensione di soglia"

$$V_{th0} \triangleq V_{bi} - \frac{q N_D}{2 \epsilon_s} d^2$$

Se non si ha almeno la tensione "critica" al gate, la curva si annulla. Questo è tutto vero, a meno della modulazione di canale: se il canale si "storce" per $L < L'$, si definisce una V_{DS} di saturazione, V_{DSS} , come:

$$V_{DSS} = \phi_{ch}(L')$$

La corrente I_D dovrà essere calcolata, in tal caso, come:

$$I_D = -\mu_n \frac{W}{L'} \int_0^{V_{DSS}} Q_n(\phi_{ch}) d\phi_{ch}$$

A questo punto, si prende Q_n , la si sviluppa e tronca al 1° ordine $\left[\sqrt{1-x} \approx 1 - \frac{x}{2} \right]$, ottenendo:

$$Q_n(\phi_{ch}) \approx -C_{ch} [V_{DS} - V_{th} - \phi_{ch}(y)]$$

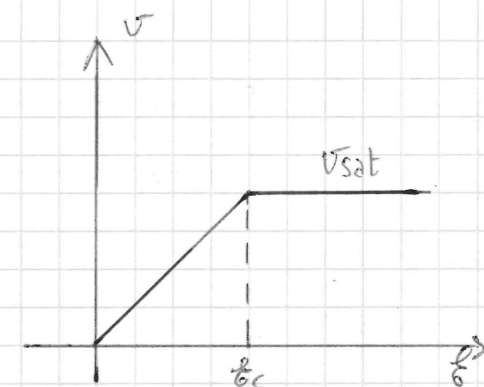
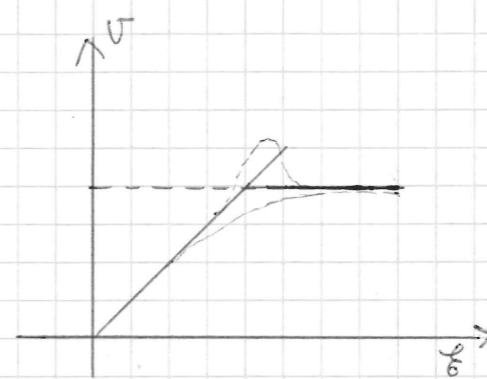
Dove:

$$C_{ch} \triangleq \frac{\epsilon_s}{a} \quad \left[\text{capacità di canale per unità di superficie} \right]$$

a è lo spessore nominale dello strato epitassiale, dunque C_{ch} non è realmente una capacità, ma la si tiene/definisce così (non ha un forte significato fisico). Si può ricorrere su I_D ma molto simile a quella del MOSFET.

Saturazione della velocità

Tutto ciò che è stato finora fatto, usava come ipotesi la mobilità costante. In realtà la velocità di trasamento non cresce in modo lineare dal momento che la curva velocità/campo elettrico fa qualcosa di diverso. Per valori elevati di campo infatti, si ha una saturazione della velocità.



A seconda del materiale, si hanno comportamenti diversi: nel silicio monotono, nel GaAs presenta un'oscillazione, ma può anche essere, per altri materiali, ben più complesso.

In dispositivi piccoli, a parità di tensioni il campo è più elevato: man mano che il canale si accorcia, dunque, si devono considerare valori di μ più approssimati.

Come il disegno suggerisce, quello che si può fare è una violenta approssimazione: anziché analizzare il complesso andamento, usare solo due tratti di retta: approssimazione brutale, ma facile da utilizzare!

Si definisce E_c come il "campo elettrico critico": è il valore di campo per cui si ha il cambio di pendenza. Si ha che:

$$-\mu_n E_c = v_{sat}$$

E dunque il campo cresce sopra E_c , V_{DS} non modula più la velocità, la quale satura, facendo anche saturare I_D .

Il canale viene dunque separato in due regioni: una per cui $E < E_c$, una per cui $E > E_c$, di lunghezze L' e L'' .

Si ha un valore di campo, ma non di tensione: supponendo che il punto "limite", di inizio del fenomeno, sia $L' = L$, $L'' = 0$, ma il "cambio di zona" coincide con la fine del canale fisico,

si può definire una tensione critica per la velocità, V_{ossv} .
Come si può fare? Beh, la corrente satura, ma non per straramento;
si fa così:

$$I_{os} = -W Q_n(L) V_{sat} = W C_{ch} [V_{gs} - V_{th0} - V_{ossv}] V_{sat}$$

Questo, usando il modello lineare a controllo di carica.

Ma I_{os} , quanto vale? Beh, se si ricorda che alla situazione limite il campo elettrico è minore del campo critico, I_{os} può essere calcolata ancora con il modello a mobilità costante; ponendo uguali I_o di questo modello e l'espressione sovrastante, V_{ossv} è l'unica incognita dell'equazione. Si ottiene in questo modo:

$$V_{ossv} = V_{gs} - V_{th0} + L E_c - \sqrt{(V_{gs} - V_{th0})^2 + (L E_c)^2}$$

FET a eterostruttura

Si propone a questo punto un'idea un po' diversa: studiamo un dispositivo "intermedio", tra JFET/MEFET e MOSFET; l'idea alla base di ciò che stiamo per fare è usare, come canale conduttivo, una buca quantica. Una buona lega è il AlGaAs; essendo GaAs e AlAs matchati in termini di passo atomico, è la più "facile" delle eterostrutture.

Si vuole realizzare una eterostruttura tra un materiale e un altro a banda proibita più grande, in modo da avere un ΔE_c che permetta la formazione di una buca di potenziale. Questa struttura risultante è di tipo I. La discontinuità deve essere tale da essere stretta abbastanza da avere un gas bidimensionale di elettroni.

Il gate è in sostanza un contatto Schottky, che deve essere in grado di smuovere anche la concentrazione di elettroni, nel canale conduttivo; lo spessore dello strato superficiale deve dunque essere sufficientemente sottile... ? Bah Magari risolvibile.

ci sono due modalità:

- Se si droga tipo "n" (ad esempio) il substrato, si parla di "FET a eterostruttura";
- Più complicato ma molto meglio, è usare la modulazione di drogaggio, "iniettando" portatori liberi senza introdurre atomi donatori, riducendo lo scattering dovuto alle impurità.

In quest'ultimo caso, il dispositivo è detto HEMT (o MODFET o TEGFET).

Come capita spesso nei dispositivi non-di-potenza, i contatti sono sulla superficie; si deve dunque in teoria avere un trasporto di elettroni lungo l'eterostruttura, ma quello che si fa in pratica è introdurre impianti n^+ che permettono di trasportare carica "direttamente" al canale.

Data questa introduzione, arriviamo al "nodo" del problema: il nostro obiettivo è determinare la caratteristica statica, cosa fattibile, nei FET, passando per la legge a controllo di carica. Trattandosi di un dispositivo molto diverso dai precedenti FET, dovremo esaminare il meccanismo su cui si basa la modulazione della carica del canale. La complessità sta nel fatto che il canale è composto da un gas di elettroni bidimensionale; la III^a direzione risulta dunque essere "vincolata", dal momento che le autofunzioni dei livelli di-

scati tendono a 0 all'estremità rapidamente.

Quò che si dovrebbe fare è risolvere l'equazione di Schrödinger, ottenendo formalmente la quantizzazione dei livelli. Se i livelli sono discreti, si sa che, data la densità g_n :

$$g_n = -q \cdot n$$

n può essere espresso sovrapposendo le popolazioni di elettroni nei vari livelli energetici della buca di potenziale:

$$n = \sum_{l=1}^{\infty} n_{sl} |\psi_l(x)|^2$$

Dove "l" è l'indice del livello energetico nella buca, e $\psi_l(x)$ è la funzione d'onda. Di essa sappiamo, in sostanza, che al centro della buca è massima, e man mano che ci si discosta dal centro tende a ridursi, con una certa rapidità. Altra proprietà: il suo modulo quadro è, a tutti gli effetti, una densità di probabilità.

Una considerazione: come tutti i FET, anche in questo si varrebbe corrente di gate nulla; il quasi-livello di Fermi della struttura nell'asse "x" deve essere considerato costante anche fuori equilibrio, dunque si ha la possibilità di usare anche fuori equilibrio il diagramma a bande, avendo come barriera $q(V_G - V_G)$ (modificando con V_G l'elettrostatica del problema).

Si noti la presenza di un sottile strato di AlGaAs non drogato: esso è detto "spacer", e viene introdotto al fine di evitare di disturbare la mobilità degli elettroni con le impurità droganti, in prossimità del canale conduttivo. Questa regione deve essere sottile, in modo che le cariche possano "cascare nel canale".

3 criteri per il dimensionamento dello spacer sono:

- Sottile all'estremità da permettere la popolazione del canale;
- All'estremità stretta da rendere insignificante lo scattering con gli atomi droganti.

Come già visto in precedenza, se la buca è sufficientemente stretta, il numero di elettroni per unità di superficie al livello l è:

$$n_{sl} = \frac{4\pi m_n^* k_B T}{h^2} \log \left[1 + \exp \left(\frac{E_F - E_l}{k_B T} \right) \right]$$

A questo punto, vogliamo calcolare Q_n , per definire il modello a controllo di carica, si sa che:

$$Q_n = -q n_s = \int_0^{+\infty} g_n(x) dx$$

Ma g dipende da n , e in n l'unica espressione funzione di x è $\psi(x)$; essendo essa una densità di probabilità, l'integrale sulla semiretta reale sarà 1:

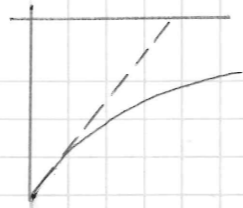
$$\int_0^{+\infty} |\psi(x)|^2 dx = 1$$

A questo punto, n_s può essere ridotto a:

$$n_s = \sum_{l=1}^{\infty} n_{sl} = \frac{4\pi m_n^* k_B T}{h^2} \sum_{l=1}^{\infty} \log \left[1 + \exp \left(\frac{E_F - E_l}{k_B T} \right) \right]$$

3 nostri problemi devono ancora cominciare: volendo ottenere tutti i parametri utili, c'è da ricordare che, risolvendo l'equazione di Schrödinger, sembrerebbe gli E_l i quali però sono funzione della carica n_s ; abbiamo dunque in mano una relazione implicita tra n_s e $E_l(n_s)$. Quò che si deve fare a questo punto, è introdurre delle approssimazioni.

l'idea che si fa è approssimare la buca, la cui forma è complessa, con una buca di forma triangolare; a questo punto, usando l'approssimazione WKB, l'equazione di Schrödinger può essere usata in forma esatta, scrivendo le posizioni dei livelli energetici della buca.



Dato $E_{s0} = E_c(0^+)$, si ottiene, passando per dimostrazioni:

$$E_l - E_c(0^+) \approx \frac{(q\hbar)^{2/3}}{(2m_n^*)^{2/3}} \left[\frac{3\pi(4l+3)}{8} \right]^{3/2} E_{s0}^{2/3}$$

Un'osservazione: si ha una dipendenza dal campo elettrico \mathcal{E} , sulla superficie! Questa relazione nasce dal fatto che E_c è "parente" del potenziale, dal momento che:

$$-q \frac{\partial \phi}{\partial x} = \mathcal{E}$$

A meno di qualche costante moltiplicativa, dunque, si può dire che \mathcal{E} sia la "pendenza del diagramma a bande" in corrispondenza di $x=0^+$, o sia la pendenza della "retta" che forma il triangolo con cui si approssima la buca di potenziale.

A questo punto, vorremmo "chiusare il cerchio", mediante il teorema di Gauss, come nel caso del MOSFET:

$$E_{well} \cdot \epsilon_{s0} = -Q_n = q n_s$$

Dove E_{well} è la costante dielettrica del materiale in cui vi è la buca quantica; il GAs. Si può esplicitare la formula, sommando e sottraendo all'esponente $E_c(0^+)$, ottenendo:

$$n_s \approx \frac{4\pi m_n^* k_B T}{h^2} \sum_{l=1}^{\infty} \log \left[1 + \exp \left(\frac{E_f - E_c(0^+) - \gamma l n_s^{2/3}}{k_B T} \right) \right]$$

Dal momento che:

$$E_l - E_c(0^+) \approx \gamma l n_s^{2/3} \quad [\text{si approssima la funzione con una retta}]$$

Questo risultato lega n_s e $E_f - E_c(0^+)$, il quale a sua volta è collegato con V_g . Dal momento che il nostro obiettivo finale è legare V_g a I_n , possiamo sentirci abbastanza vicini.

Si noti un altro fatto: di solito, scegli $l \in [1, \infty)$, si considerano solo $l=1, l=2$.

Sotto queste osservazioni, passiamo alla prossima parte: leghiamo $E_f - E_c(0^+)$ a V_g . Quello che si deve fare è studiare lo strato superficiale, mediante l'equazione di Poisson.

Un'osservazione: al crescere di V_g cresce anche E_{s0} , ma non in maniera indefinita; per ora, consideriamo una crescita strettamente monotona, per questo dopo si specificerà che il modello è incompleto. Dato ϵ_{top} la costante dielettrica dello strato superficiale, si sceglie che:

$$-q \phi(x) = E_c(x) \rightarrow \phi(x) = -\frac{E_c(x)}{q}$$

Supponendo che lo strato superficiale sia strutturato di portatori liberi (ipotesi, come vedremo, discutibile), si avrà:

$$\frac{d\phi}{dx} = \frac{\rho}{\epsilon_{top}} = \begin{cases} \frac{q N_D}{\epsilon_{top}} & -d_L \times L - d_s \quad [\text{dove si ha drogaggio } n, \text{ si ha anche svuotamento e carica fissa positiva}] \\ 0 & -d_s \times L \times L_0 \quad [\text{dove lo strato è intrinseco}] \end{cases}$$

Integrando:

$$\phi(x) = \frac{1}{q} \frac{dE_c}{dx} = \begin{cases} \frac{q N_D}{\epsilon_{top}} x + K_1 & -d_L \times L \times d_s \\ K_2 & -d_s \times L \times L_0 \end{cases}$$

Le costanti K_1 e K_2 vanno calcolate, definendo due condizioni al contorno:

1) Il vettore di spostamento dielettrico \underline{D} deve essere continuo nella strato; essendo il dielettrico sempre uguale, dunque, si ha:

$$\epsilon(-d_s^-) = \epsilon(-d_s^+)$$

2) Lo spostamento dielettrico deve essere costante anche all'interfaccia:

$$\epsilon_{top} \epsilon(0^-) = \epsilon_{well} \epsilon(0^+)$$

Da qui, sostituendo si ricava facilmente che:

$$K_2 = \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so} \quad ; \quad K_1 = \frac{q N_0}{\epsilon_{top}} d_s + \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so}$$

Passiamo al secondo step: collegare V_G (nota) a $E_F - E_C(0)$. Si sa che $q(V_{bi} - V_G)$ è pari al salto di potenziale; collegiamo tutto, integrando:

[una volta sostituito K_1 e K_2]

$$E_C = \begin{cases} \frac{q^2 N_0}{2 \epsilon_{top}} (x + d_s)^2 + q \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so} x + c_1 & -d_L \leq x \leq -d_s \\ q \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so} x + c_2 & -d_s \leq x \leq 0 \end{cases}$$

Le condizioni al contorno ora sono:

1) La continuità di E_C tra strato drogato e spacer:

$$E_C(-d_s^-) = E_C(-d_s^+)$$

2) In $x=0^-$, ossia all'interfaccia al lato dello spacer, E_C deve avere un salto; dunque, dato ΔE_C il salto:

$$E_C(0^-) = E_C(0^+) + \Delta E_C$$

Questo ΔE_C è un parametro dipendente del tipo di eterostruttura da cui è usata. Si può dimostrare che:

$$c_1 = c_2 = E_C(0^+) + \Delta E_C$$

Sostituendo c_1 e c_2 , si ottiene:

$$E_C(x) - E_C(0^+) - \Delta E_C = \begin{cases} \frac{q^2 N_0}{2 \epsilon_{top}} (x + d_s)^2 + q \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so} x & -d_L \leq x \leq -d_s \\ q \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so} x & -d_s \leq x \leq 0 \end{cases}$$

A questo punto, si osservi il diagramma a bande della struttura: nel punto $-d$, si ha che:

$$E_C(-d) - E_{Fn} = q(V_{bi} - V_G)$$

Valutando nel punto $x = -d$ il modello, si trova:

$$q(V_{bi} - V_G) = \frac{q^2 N_0}{2 \epsilon_{top}} (-d + d_s)^2 - q \frac{\epsilon_{well}}{\epsilon_{top}} \epsilon_{so} d + E_C(0^+) + \Delta E_C - E_{Fn}$$

Scelto V_{bi} , anche gli altri parametri dell'equazione sono, come si può evincere dall'equazione, costanti o univocamente definiti (ϵ_{so}).

Si può anche a questo punto, calcolare la tensione di soglia.

Si consideri la precedente equazione usata nel seguente modo:

$$\epsilon_{so} = \frac{q N_0}{2 \epsilon_{well} d} (d - d_s)^2 + \frac{\epsilon_{top}}{\epsilon_{well} q d} [E_C(0^+) - E_{Fn} + \Delta E_C - q(V_{bi} - V_G)]$$

Alla tensione di soglia, capitano due casi: dal momento che $n_s = 0$,

1) $\epsilon_{so} = 0$: non si ha carica, ma dunque neanche campo elettrico

2) $E_C(0^+) - E_{Fn} = 0$: il diagramma a bande non si è invertito ancora abbastanza da permettere la formazione di carica, ma si è in una situazione per cui un minimo campo di V_G permetterebbe la formazione di carica, dunque E_C che "entra" in E_F .

Da ciò si ricava facilmente, data $V_G = V_{th}$:

$$V_{th} = V_{bi} - \frac{q N_0}{2 \epsilon_{top}} (d - d_s)^2 - \frac{1}{q} \Delta E_C$$

Più comunemente, l'espressione si trova nella forma:

$$q(V_{bi} - V_G) = q(V_{bi} - V_{bi0}) + E_c(0^+) - E_{Fn} - q \frac{E_{well}}{E_{top}} \epsilon_{so} d$$

A questo punto, uniamo i pezzi e studiamo il cerchio! Il e^- quando esprime n_s (la sommatoria di logaritmi) può essere espresso, linearizzando come:

$$E_F - E_c(0^+) \approx q \Delta n_s, \quad \Delta = 9.125 \times 10^{-19} \frac{V}{m^2}; \quad \text{dato } \Delta d, \text{ per}$$

Gauss si ha che:

$$Q_n = -q n_s = -\epsilon_{so} E_{well}$$

Si può scrivere, unendo e sostituendo:

$$V_G - V_{bi0} = \frac{1}{E_{top}} \left(d + \epsilon_{top} \frac{\Delta}{q} \right) q n_s = \frac{d + \Delta d}{E_{top}} q n_s$$

Dove:

$$\Delta d = \epsilon_{top} \frac{\Delta}{q} \approx 80 \text{ \AA}$$

Si ricordi che Δd , a dispetto di come certi libri affermano, è un parametro semplicemente derivante dall'equazione di Poisson applicata sullo strato superficiale. Si eriti di cercar per esso significati finiti particolari. Smentendo si può trovare la relazione di controllo di carica:

$$Q_n = -q n_s = -C_{ch}(V_G - V_{bi0}), \quad C_{ch} \triangleq \frac{\epsilon_{top}}{d + \Delta d}$$

Come detto in precedenza, questa relazione non è esattissima: al crescere di V_G , la differenza tende a ridursi, dunque la banda di potenziale si riduce, e il minimo della banda di conduzione del AlGaAs si avvicina al quasi-livello di Fermi, fino a toccarlo.

Quando si toccano (E_c e E_{Fn}), si forma nello strato una concentrazione di elettroni. Questa è la pregiatura dell'HEMT: quando V_G cresce, aumenta il numero di elettroni in questa "buca" (della "spuria"), ma solo più in essa, e non nel canale, a causa di questa sorta di "schermo".

Il nostro errore è stato considerare "totale" lo spostamento, ignorando, su " qN_0 ", una quantità che tenga conto della presenza di elettroni liberi; ciò non è stato fatto perché una correzione del genere renderebbe il problema non risolvibile in forma chiusa.

Ciò che si può fare formalmente è uno studio del caso "limite": $\min\{E_c\} = E_{Fn}$.

La V_G critica sarà quella che porterà proprio a questa condizione.

Smentitatta, si deve determinare la posizione spaziale del minimo: essa può essere calcolata annullando la derivata di E_c (nota da prima):

$$\left. \frac{dE_c}{dx} \right|_{x=-d_0} = 0 = \frac{q^2 N_0}{E_{top}} (-d_s + d_s) + q \frac{E_{well}}{E_{top}} \epsilon_{so, sat}$$

$$\xrightarrow{\text{invertito}} d_0 = d_s + \frac{E_{well}}{q N_0} \epsilon_{so, sat}$$

Si valuti E_c in $x = -d_0$:

$$E_c(-d_0) = \frac{q^2 N_0}{2 E_{top}} (d_s - d_0)^2 - q \frac{E_{well}}{E_{top}} \epsilon_{so, sat} + E_c(0^+) + \Delta E_c = E_{Fn}$$

Il " $= E_{Fn}$ " è la condizione "limite": $E_c(-d_0) = E_{Fn}$.

Sostituendo in d_0 l'espressione estesa, ricordando poi che:

$$E_{Fn} - E_c(0^+) = qd n_{s,sat}$$

$$q n_{s,sat} = E_{well} \epsilon_{so,sat}$$

$$\Delta d \cong E_{top} \frac{d}{q}$$

Si ottiene:

$$n_{s,sat}^2 - 2(d_s + \Delta d) N_0 n_{s,sat} - 2 \frac{N_0 E_{top}}{q} \frac{\Delta E_c}{q} = 0$$

Da qui, risolvendo questa equazione e considerando la radice positiva (dal momento che le concentrazioni sono quantità positive!)

si ha:

$$n_{s,sat} = - (d_s + \Delta d) N_0 + \sqrt{(d_s + \Delta d)^2 N_0^2 + \frac{N_0 E_{top}}{q} \frac{\Delta E_c}{q}}$$

Da qui si vede cosa si può fare per cambiare la condizione di saturazione sulla tensione: si può cambiare N_0 , ma pure il ΔE_c ; in questi modi si può far lavorare l'HEMT a V_0 più elevate senza incappare nella saturazione.

Si noti un ultimo fatto: quando il HEMT va in saturazione, la presenza degli elettroni nel "canale" spinge ferma il condotto "MESFET parassita", che ha prestazioni orionante molto più scarse dell'HEMT, e peggiora quelle del dispositivo principale.

Riprendendo la relazione a controllo di carica (lineare), considerando che a $V_g = V_{gsat}$ si ha $n_{s,sat}$, si può dire che:

$$V_{gsat} = V_{tho} + \frac{q n_{s,sat}}{C_{ch}}$$

Se fatto che la concentrazione saturi ha conseguenze importanti: se

satura n_s satura anche $q n$, ma dunque anche I_D !

$$I_{D,sat} = W V_{gsat} q n_{s,sat}$$

Infatti, saturando n_s , satura anche V_n (si ricorda che V_n si definisce come $-\mu_n E$), dunque complementemente l'intera corrente di drain tende a saturare, per quanto il modello lineare a controllo di carica ci suggerisca che, a V_{gsat} , si abbia una saturazione.

Dal momento che I_D tende a saturare, g_m , definita al solito come:

$$g_m = \frac{\partial I_D}{\partial V_{gs}}$$

Tende a 0. Essa può comunque essere valutata prima della saturazione, calcolando la derivata dell'espressione totale; considerando L piccolo rispetto agli altri parametri, si può dire che:

$$g_m = \frac{W}{L} \mu_{no} C_{ch} \frac{V_{gs} - V_{tho}}{\sqrt{\left(\frac{V_{gs} - V_{tho}}{L \epsilon_c}\right)^2 + \tau_c}}; \quad \lim_{L \rightarrow 0} g_m(L) = W C_{ch} V_{gsat}$$

A questo punto, ci chiediamo: vi sono vantaggi "sostanziali", dovuti all'introduzione dell'HEMT, rispetto ai "vecchi" MESFET? Vediamo:

- Si sa che la corrente è proporzionale a g_m ; nel caso del MESFET, essa è controllata dallo spessore dello strato, il quale però regola anche la tensione di soglia: un parametro ne influenza contemporaneamente altri 2. Con l'HEMT si introduce un ulteriore grado di libertà: il salto di banda! Regolandolo, è possibile controllare anche un parametro alla volta!
- Si ha un miglior confinamento del canale conduttivo, dal momento che la buca quantica ha una distanza ben definita dal gate!

• Si usa la modulazione di drogaggio, cosa molto utile, del momento che si cambia il numero di portatori liberi senza introdurre scattering!

Alcune osservazioni: noi abbiamo solo analizzato l'HENT "convenzionale"; in realtà ve ne sono di altri tipi! Le principali varianti prevedono l'uso di una buca quantica, realizzata mediante 2 eterostrutture, le quali migliorano il confinamento del canale.

Altra variante è l'uso di un drogaggio tipo "delta": anziché distribuire il drogaggio, lo si concentra in un singolo punto spaziale, migliorando il problema della saturazione.

Con gli HENT si ha un altro grado di libertà, in realtà uno: il materiale! Cambiando il materiale si possono "ingegnerizzare" le bande, ottenendo quello che si vuole.

Due parole infine sul modello di p-nolo seguente: esso è allosteranza simile a quello degli altri FET, anche se si vedono nuovi elementi parassiti, quali induttanze. Non è da esse non esistono negli altri dispositivi, ma c'è da tener conto di un fatto: questi sono dispositivi progettati per le altissime frequenze, dunque alcune non ideali risultano più evidenti che non in frequenze come $100 \text{ MHz} \div 1 \text{ GHz}$.

Prv 2 parole / disegno!

3 dispositivi optoelettronici

L'optoelettronica è una branca molto importante dell'elettronica, che si pone il seguente problema da risolvere: passare dal dominio dei segnali elettrici a quello dei segnali ottici, e viceversa.

3 principali dispositivi sono i fotodivelatori, i LED e i LASER; essi trovano applicazione in ambito di telecomunicazioni, storage, o altro.

L'esempio forse più completo di sistema optoelettronico è un sistema di comunicazione ottico; esso è composto da 3 blocchi sostanziali:

- Trasmettitore
- Canale
- Ricevitore

Il canale deve garantire la propagazione guidata (mediante una fibra ottica); se la tratta è lunga, bisognerà però introdurre blocchi di rigenerazione del segnale: su lunghe tratte, il segnale tende a degradarsi; è dunque fondamentale l'uso di amplificatori ottici, realizzati su semiconduttore o in fibra ottica.

Ma a monte ma a valle servono due categorie di sottoblocchi: a monte serve un generatore della portante ottica, e un meccanismo che permetta di tradurre il segnale elettrico in ottico, e di trasferirlo sulla portante ottica (modulatore). La sorgente luminosa assorbe potenza elettrica e la restituisce (con l'efficienza più alta possibile) otticamente, mediante un LED o un LASER.

Questi possono essere realizzati con qualcosa di molto simile a una giunzione p-n, in cui si cerca di massimizzare i meccanismi

di generazione termica/ricombinazione; la "giunzione" va polarizzata direttamente.

Al ricevitore, come al trasmettitore, vi sarà una parte elettronica e una optoelettronica: la seconda dovrà prendere la portante ottica e convertirla in un segnale elettrico: un fotodivelatore.

I semiconduttori sono molto interessanti, sotto il punto di vista dell'optoelettronica, dal momento che un segnale ottico è a tutti gli effetti un'onda elettromagnetica, a frequenza molto elevata. Come già detto, si usa come fenomeno di trasduzione la generazione/ricombinazione, ossia la formazione di corrente dovuta alla nascita di coppie elettrone-lacuna, per la promozione di elettroni da E_v a E_c .

Il segnale ottico, come all'esplosione di de Broglie, può essere pensato in termini particellari, dunque come flusso di fotoni, per cui $E = h\nu$.

Si deve chiedere alla frequenza del segnale ottico di essere almeno uguale all' E_g ; dal momento che $h\nu$, per le frequenze dello spettro ottico, è dell'ordine delle decine di eV fino a qualche eV, l'uso dei semiconduttori risulta perfetto!

Si vuole ottenere un sistema di propagazione guidata, dunque usare guide d'onda; esse potrebbero essere studiate mediante le equazioni di Maxwell e il formalismo di Hertz e Schwinger, ma anche in modo

diverso: ricordando che l'indice di rifrazione n è uguale a $\sqrt{\epsilon_r}$, dunque, usando due strati, con $n_1 > n_2$, si riesce a confinare, nello strato con n minore, il segnale ottico;

	E_2
	E_1

questo, si può studiare anche con la sola ottica geometrica e con le leggi di Snell, senza scendere Maxwell!

Soltanto le guide più usate sono quelle circolari, come le già citate fibre ottiche. Questo ragionamento, basato sull'ottica geometrica, possono essere applicati, a patto che la lunghezza d'onda dei segnali ottici sia trascurabile rispetto alle dimensioni spaziali in gioco.

Una cosa più semplice che si possa fare per realizzare fibre ottiche è usare del berillio di silicio drogato: il drogaggio cambia la costante dielettrica relativa ϵ_r , e con essa il coefficiente di rifrazione n .

A seconda delle dimensioni finché, sono possibili diversi comportamenti:

- fibra monomodale: si propaga un solo modo;
- multimodale: diversi modi si propagano assieme nello guida.

Le fibre hanno un grosso problema: quasi tutta la perdita di efficienza deriva dalla difficoltà di accoppiare fibre e dispositivi a semiconduttore, ottenendo quasi "perdite di accoppiamento" all'interfaccia.

Di solito si usano fibre monomodali, al fine di ridurre la presenza di perdite. Tuttavia, al fine di garantire una buona trasmissione monomodale, e l'andamento della spettro; i numeri di fase devono essere minimi.

Nel trasmettitore sono presenti due tipi di dispositivi optoelettronici, come già detto: LED/LASER, e il modulatore. Esistono due modi per realizzare la modulazione:

- Interna: se si manda una corrente non costante nel tempo ma a una certa frequenza, se la corrente viene modulata nel tempo allora si modula in ampiezza l'uscita della sorgente luminosa. In tal caso, se non si ha un modulatore "separato" rispetto al

generatore di segnale ottico, la modulazione è sotto "interna"; facile da vedere, anche l'introduzione di altri dispositivi, ma la banda è limitata a 20 GHz circa (poca, contro i THz!).

- Esterna: si usa anche un modulatore esterno, ossia un dispositivo terzo, mentre la sorgente di portante ha corrente costante.

8 il ricettore, cosa deve fare? Beh, deve convertire il segnale ottico (solitamente modulato in AM) in un segnale elettrico, mediante la generazione/ricombinazione, con un fotodivelatore.

5 parametri caratteristici di un ricettore sono:

- sensibilità: la minima potenza ottica rivelabile!
- larghezza di banda di modulazione: essere in grado di tradurre segnali a elevato bitrate, dunque rapidi.

Assorbimento e emissione

Dal punto di vista del fotone, dunque del segnale ottico, i meccanismi collegati alla generazione/ricombinazione sono l'emissione e l'assorbimento di fotoni.

Si sa che la generazione/ricombinazione è "forte", se i semiconduttori sono a gap diretto; ciò è un po' brutto, dal momento che il nostro semiconduttore preferito, il silicio, è a gap indiretto; tipicamente ^{usati sono} il GaAs, InP sono a gap diretto.

Importante parametro è quello che caratterizza l'assorbimento dei fotoni: il coefficiente di assorbimento α ; $[\alpha] = \text{cm}^{-1}$; si può dimostrare che la potenza ottica incidente decresce esponenzialmente, nello spazio, con una costante di decadimento pari a $\frac{1}{\alpha}$. Se la potenza

ottica è molto ridotta, dunque significa che i fotoni sono stati assorbiti, e dunque che le coppie elettrone-lacuna sono state generate. Maggiore è α , e minore sarà la lunghezza caratteristica di assorbimento. α è dipendente, ovviamente, da λ !

Si nota che in corrispondenza di $E = E_g$, non si ha assorbimento e dunque in un grafico si vedrebbe degli assenti verticali!

Una nota positiva: anche se Si e Ge sono a gap indiretto, hanno un α abbastanza elevato, dunque, per quanto in applicazioni di basso livello, essi sono utilizzabili! Ormai, nel caso di applicazioni di alta qualità, il GaAs o il InP sono senza dubbio molto più indicati.

Ripetiamo ancora una volta il concetto di base: il punto di contatto tra il modello dei semiconduttori e il dominio ottico è il meccanismo di generazione/ricombinazione. Valutare la potenza ottica è abbastanza facile: procedendo nel materiale, da un punto x al punto $x+dx$, si ha:

$$P_{\text{op}}(x+dx) - P_{\text{op}}(x) \quad L \quad 0$$

Ovvero: più si va avanti in x , più i fotoni sono stati assorbiti, e dunque la "potenza ottica" si riduce. Andando avanti, ricorrendo che l'assorbimento segue una legge esponenziale, si avrà:

$$P_{\text{op}}(x+dx) - P_{\text{op}}(x) = P_{\text{op}}(0) \exp(-\alpha(x+dx)) - P_{\text{op}}(0) \exp(-\alpha x)$$

Si sviluppa mediante Taylor e troncando al primo ordine:

$$P_{\text{op}}(x+dx) - P_{\text{op}}(x) \approx -\alpha P_{\text{op}}(x) dx \quad \left[P_{\text{op}}(x) = P_{\text{op}}(0) \exp(-\alpha x) \right]$$

Definire il tasso di generazione ottica è semplice: dato $P_{op}(x)$, dividendo per l'energia di un singolo fotone, si vede quante coppie elettrone/lacuna saranno generate:

$$G_{op}(x) = -\frac{dP_{op}(x)}{h\nu} = 2\Phi_0(x)$$

Dove $\Phi_0(x)$ è il flusso di fotoni incidenti su x .

Per generare un fotone, invece, è necessario avere la morte di una coppia elettrone-lacuna, in modo da poter generare il fotone; questo processo è detto "emissione spontanea".

Esiste un altro processo di emissione: l'emissione stimolata. Esso è un fenomeno di emissione di fotoni, ma dove la ricombinazione avviene in presenza di altri fotoni; la particolarità di questo fenomeno è il fatto che il nuovo fotone sarà coerente col precedente. Questo fenomeno è alla base dei LASER, e si realizza mediante una reazione positiva, a partire da un ambiente con già molti fotoni.

Trattiamo più nel dettaglio questi dispositivi.

LED

La struttura di un LED è all'estrema semplice: una giunzione p-n in polarizzazione diretta. Il punto chiave è far in modo che, nella regione di giunzione, nel caso di polarizzazione diretta dunque una zona in cui prevale un comportamento ricombinativo, la ricombinazione sia di tipo ottico; ovviamente, molto favoriti in tal senso sono i materiali a gap diretto. Di solito questi materiali sono α -GaAs o InP; il materiale è scelto in modo da avere una finestra, nel

grafico delle perdite, in modo da avere adattamento con la fibra ottica.

Il meccanismo prevalente è l'emissione spontanea; conseguenza di ciò è una scarsa purezza spettrale, del momento che lo spettro di emissione è dell'ordine di $k_B T$: $\Delta\lambda \approx 2\mu\text{m}$; in realtà si possono anche realizzare LED più puri di questo; i LASER, anche i GaAs , con 100 kHz di banda: molto meglio!

Relazione facile da ricavare è quella che fornisce un'idea dell'ordine di grandezza della potenza ottica emessa:

$$P_{op} = h\nu \frac{n_{\text{fotoni}}}{t} = h\nu \frac{\text{coppie el/lacuna}}{t} \approx h\nu \frac{I}{q} \approx I E_g$$

Dove I è la corrente che permette la emissione. Si ricordi che questo è solo un valore di massima, a efficienza unitaria. Si può anche stimare la banda, come:

$$B = \frac{1}{2\pi\tau_r}, \quad [\tau_r = \text{tempo di vita radiativo}]$$

Esistono due tipi di strutture, per i LED:

- LED verticali: la direzione da cui i fotoni vengono estratti è tale da avere il flusso di fotoni ortogonale rispetto ai "piani" della struttura multi-strato. Il quantum well sarà il materiale da cui si vogliono prelevare i fotoni; al fine di non far attraversare tutti ~~gli~~ strati ai fotoni, si realizza uno scavo, a cui si collega direttamente la fibra ottica, in modo da evitare perdite.
- LED orizzontali: in questa struttura i flussi di corrente e di fotoni

sono ortogonali tra loro. Questa struttura di fotoni non è molto diversa dai LASER. La "comparsa" indica il fatto che i fotoni sono soprattutto localizzati dove sono generati.

È possibile che sia presente anche un po' di emissione stimolata; in tal caso, si parla di LED super-radianti.

LASER

Nei LASER si rende prevalente l'emissione stimolata su quella spontanea: questo è ottenuto perché da un lato si eliminano le impurità spettrali, dall'altro si riduce il tempo di vita radiativa, aumentando la banda.

La concentrazione dei fotoni viene aumentata, inserendo il LED in una cavità risonante per i fotoni, la quale funge da "specchio". Si ha dunque un sistema risonante, in cui la corrente totale si scompone in due contributi: una parte alimenta l'emissione spontanea, l'altra l'emissione stimolata. In pratica, la risonanza agisce sul segnale ottico: si generano dapprima fotoni non coerenti, che però non saranno emessi, a meno che non si raggiunga la frequenza di risonanza della cavità, f_0 . Mediante la meccanica quantistica si può dimostrare che dunque l'emissione stimolata finisce per prevalere su quella spontanea.

Si ha un meccanismo di soglia: è infatti necessario introdurre un minimo valore di corrente, al fine di garantire l'emissione stimolata.

Il laser a cavità risonante di Fabry-Perot è un tipo di laser a cavità risonante mediante doppia eterostruttura; storicamente è stato il primo a essere realizzato. Il primo miglioramento fu un miglior

confinamento della corrente, atto ad attenuare i modi trasversali, incrementando l'efficienza. Il passo successivo fu l'eliminazione del semiconduttore, sostituendo la cavità risonante con una sorta di "specchio distribuito", ottenuto mediante un processo di coniugazione; questo garantisce maggior purezza spettrale. Passo "finale", è l'introduzione di un reticolo di Bragg esterno.

Fotorivelatori

Studiamo a questo punto i dispositivi diodi: assorbire fotoni, e generare coppie elettrone-lacuna. Serve un fotodiodo, per avere un primo esempio di fotorivelatore: si tratta di un sistema di un materiale idoneo ad assorbire la radiazione luminosa, e varare la propria resistività con i fotoni; applicando agli estremi una tensione, la corrente sarà proporzionale alla resistività.

Si può usare anche una giunzione p-n; volendo far ora prevalere la generazione, al contrario di prima, si dovrà polarizzare inversamente, modulando così la corrente di saturazione inversa. Trattando circuiti piccoli, ora la sensibilità sarà buona. Una variante è quella di aumentare la regione di movimento, per aumentare di conseguenza la regione di assorbimento: ciò si può fare introducendo, nella struttura, uno strato di semiconduttore intrinseco, ottenendo la giunzione p-i-n.

Ancora un modo è il fotorivelatore a valanga: si separano spazialmente le due regioni p drogata, e con un campo elettrico si polarizza le coppie generate nell'assorbitore, "tirandole" verso

la giunzione p-n; questa dovrà essere plasmata molto imprecisamente, in modo da riprendere la rottura elettrolitica. Questo dispositivo funziona, ma è molto rumoroso, ottenendo meno sensitivity. Al fine di caratterizzare un fotorelettor, parametro fondamentale è la "responsivity":

$$R = \frac{I}{P_{\text{opt}}} = \frac{q \frac{\text{coppie e/h}}{\text{tempo}}}{h\nu \frac{\text{fotoni}}{\text{tempo}}} = \frac{q}{h\nu} \frac{\text{num. coppie e/h}}{\text{num. fotoni}} = \frac{q}{h\nu} \eta$$

Si tratta di una cifra di merito; η è l'efficienza quantica.

Si noti che la corrente I è associata al segnale utile, dunque la formula è coerente: non basta infatti che un fotone venga assorbito, dal momento che la coppia elettrone-lacuna deve "entrare in circolo", in modo da dare un contributo utile alla corrente.

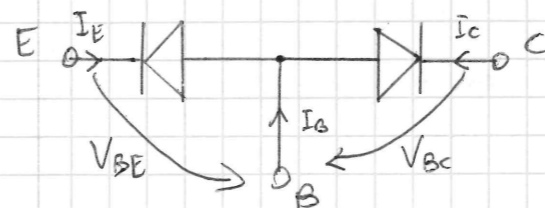
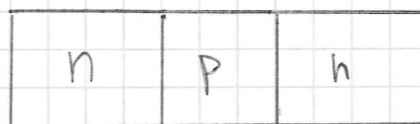
Nel fotodiodo a valanga, η sarà molto elevata.

grafico responsivity?!

5 transistori bipolari

L'obiettivo di questo capitolo della trattazione è dapprima quello di riprendere alcuni concetti riguardanti il transistor bipolare a omogiunzione, per poi vedere come le eterostrutture possono migliorarlo. Come noto, esso è dato dall'unione di due giunzioni p-n: uno dei due lati è "in comune", quello "centrale", mentre gli altri sono drogati alla stessa maniera (o quasi): n-p-n, o p-n-p; si ha dunque la possibilità di avere due strutture "complementari", duali.

L'obiettivo è quello di avere un dispositivo in grado di realizzare l'effetto transistor: controllare la corrente di uscita tramite un terminale di ingresso, controllato con una qualche grandezza elettrica. Per quanto risulta comodo studiare Vce a partire dalla corrente di base I_B (usata come controllo), di solito si studia lo stato elettrico a partire da una tensione di controllo anziché da una corrente: è più intuitivo.



Analizziamo questa struttura (il p-n-p ~~X~~ è abbastanza simile: essa sembrerebbe intrinsecamente simmetrica, anche se in realtà ciò non è vero: al fine di ottimizzare la realizzazione dell'effetto transistor, si usa un E molto più drogato di B (base) e B più drogato di C (collettore). Il modello "a diodi" ci dà un'idea del funzionamento, per quanto molto frammentaria: se infatti fosse veramente tutto ideale,

di B (base) e B più drogato di C (collettore). Il modello "a diodi" ci dà un'idea del funzionamento, per quanto molto frammentaria: se infatti fosse veramente tutto ideale,

questo circuito sarebbe un circuito aperto: solo uno dei due "diodi" accesi porterebbe corrente nulla, non corrente controllabile.

Volendo controllare mediante le tensioni V_{BE} e V_{BC} , ci sono le regioni di funzionamento:

- saturazione: se entrambe le giunzioni sono in polarizzazione diretta, ossia le tensioni al catodo son inferiori rispetto all'anodo, $V_{BE} > 0$, $V_{BC} > 0$, si ha la "regione di saturazione"; questo significa che la tensione sull'uscita, V_{CE} , è pari alla differenza $V_{BE} - V_{BC}$; questo, dal momento che le giunzioni son polarizzate direttamente, porta ad avere la differenza delle V_f , numero vicino a 0V. La corrente sarà allora grande. Il transistor è "simile a un corto circuito".
- interdizione: se entrambe le giunzioni sono polarizzate inversamente, si applicano tensioni negative, e dunque anche con V_{CE} elevata, si ha una corrente quasi nulla; è simile al circuito aperto.
- regione attiva diretta: quando $V_{BE} > 0$, $V_{BC} < 0$, si ha la rad, la regione più usata per l'effetto transistor: controllo della corrente di collettore a partire da quella di base.
- regione attiva inversa: $V_{BE} < 0$, $V_{BC} > 0$; è come scambiare i ruoli di elettroni e lacune. Se il transistor fosse sul senso simmetrico, la rad e la rad sarebbero veramente uguali, a meno di un cambio di segno della corrente.

Consideriamo da ora soprattutto la rad; in realtà il modello

circuito per ora introdotto rende sì l'idea, ma niente più: ciò che differenzia un BJT dal collegamento di due diodi in antiseno è ciò che capita nella base; i semiconduttori sono infatti dominati da fenomeni di tipo diffusivo; se si ha una base di lunghezza fisica corta rispetto alla lunghezza di diffusione dei minoritari, si può ottenere l'effetto transistor. Questo significa che la porta di uscita (collettore) deve avere una corrente molto maggiore rispetto a quella di base, ossia a quella di controllo.

Questa "richiesta" si traduce nel seguente modo: dalla base al collettore vi può essere una corrente solo di elettroni, come si spiegherà meglio tra poco; essi devono essere portati da qualche posto o in qualche modo. Al fine di studiare ciò è prima necessario introdurre un certo numero di "flussi" di portatori: F_{nEB} , F_{pBE} , F_{nBC} , F_{pCB} , F_{pCB} . Date aree di giunzione unitarie ($A=1$), una densità corrente può essere definita come:

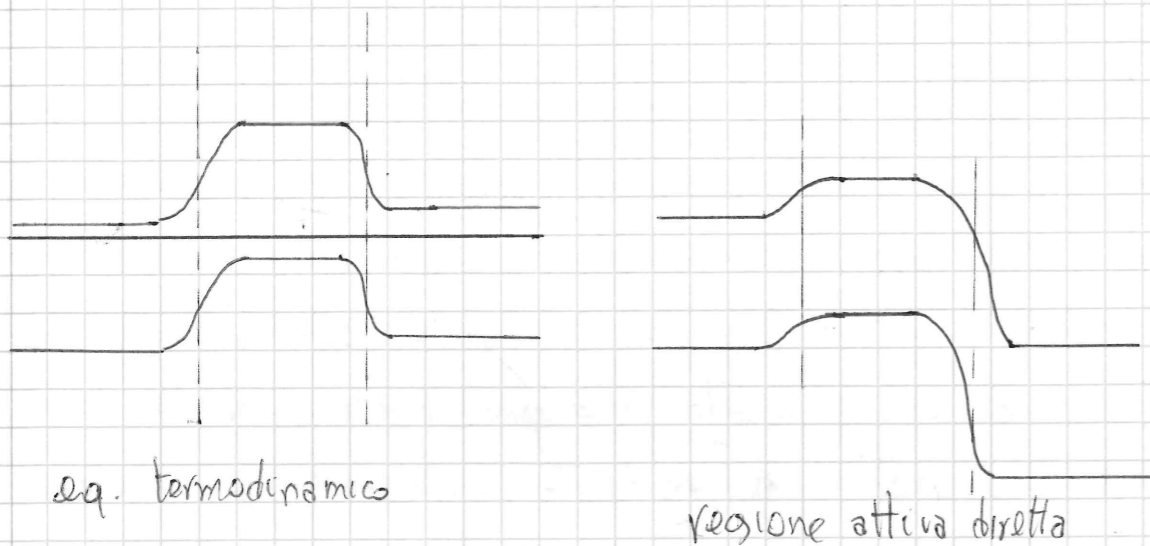
$$I = \pm q F_d A = \pm q F_d \quad \left[\text{Dove "+" e "-" si distinguono dalla presenza} \right. \\ \left. \text{rispettivamente di lacune o elettroni} \right]$$

Si definiscano la corrente di emettitore I_E e di collettore I_C come:

$$I_E \triangleq -q F_{nEB} - (+q F_{pBE}); \quad I_C \triangleq q F_{nBC} + q F_{pCB} + q F_{pCB}$$

Ma I_E sia I_C si considerano entranti nel dispositivo, da qui le scelte dei segni (I_C , definita "da destra a sinistra positiva", ha il flusso di lacune da C a B, meno i flussi di elettroni da B a C).

Studiamo rapidamente i diagrammi a bande all'equilibrio termodinamico e in rad.



A questo punto l'idea dovrebbe essere banale: la barriera da E a B si è ridotta, così che vi sia un flusso di elettroni ingente verso la base; in base, in seguito a fenomeni di ricombinazione più o meno ingenti, vi sarà un grosso flusso di minoritari verso il collettore, ma i minoritari per la base sono maggioritari per il collettore!

Ora: se la base è lunga, i fenomeni di ricombinazione attenuano lentamente il flusso di elettroni che provengono dall'emettitore, in modo che quasi solo per I_B influenzi I_C , ma questo è il comportamento dei "diodi in antiserio": non ce ne frega nulla. La corrente è dunque costante, ma prossima a 0! Per avere una corrente grande si deve ridurre i fenomeni di ricombinazione, ma questo può essere fatto solo accorciando la base. E in base vi è poca ricombinazione, del flusso F_{EB}^{EB} si perde ben poco, dunque modificando la tensione V_{BE} si controlla

F_{EB}^{EB} , che, si spera, sarà simile a F_{BC} .

La corrente sarà, in rad, senza dubbio indipendente dalla tensione:

$$V_{CE} = V_{BE} - V_{BC}$$

Ma: V_{BE} è pari a V_f ; V_{BC} è negativo, ma cambiare V_{BC} non cambia il valore della corrente, dunque si ha un "generatore quasi ideale" e "pilato", da I_B . E la base è corta, inoltre, è possibile controllare la quantità di corrente. Questo è l'effetto transistor.

Il nostro obiettivo è fare in modo che F_{EB} prevalga su F_{BE} , e sia simile a F_{BC} ; per far ciò si dovrà di più lenificare rispetto alla base, aumentando l'efficienza di iniezione prima da E a B, poi, drogando più B di C, quella da B a C.

Si possono distinguere tre flussi di corrente da B a C: uno è quello "intrinseco", uno di lacune, e un altro di elettroni: esso è indipendente dall'iniezione introdotta in B, ed è il flusso di minoritari

intrinsecamente presenti.

Vale la legge di Kirchhoff:

$$I_B = -I_C - I_E$$

A partire da qua, e dalle ipotesi precedentemente fatte, ossia che:

$$F_{BE} \approx 0; F_{EB} \approx F_{BC}; I_{CB} \approx 0$$

Si può quantificare la qualità della struttura realizzata. Esistono a tal fine due parametri, che ci dicono quanto il BJT si discosti dalla struttura ideale.

Il primo parametro è l'efficienza di iniezione di emettitore:

$$\gamma \triangleq \frac{F_{NEB}}{F_{NEB} + F_{PBC}}$$

Un altro è la quantità degli elettroni in base che escono nel collettore:

$$b \triangleq \frac{F_{NBC}}{F_{NEB}} \quad (\equiv \Delta T)$$

Concludiamo ora di capire a che servono: si può dire che:

$$I_E = -q \frac{F_{NEB}}{\gamma}$$

Inoltre:

$$I_C = q b F_{NBC} + I_{CBO} = -\gamma b I_E + I_{CBO}$$

Si definisce dunque Δ_F come:

$$\Delta_F \triangleq \gamma b$$

questo è il "fattore di amplificazione della corrente a base comune", anche se in realtà è minore di 1. Data $I_B = -I_C - I_E$, sostituendo e ordinando, si può definire:

$$\beta_F = \frac{\Delta_F}{1 - \Delta_F} \gg 0$$

Cosa bella: β_F è un numero grosso; cosa brutta: la grandezza deriva dal fatto che si divide per un numero prossimo a 0;

piccole variazioni di Δ_F ne provocano di enormi di β_F ! Considerando

$$\Delta \beta_F = \frac{\partial \beta_F}{\partial \Delta_F} \Big|_{\Delta_F} \Delta \Delta_F = \frac{1}{(1 - \Delta_F)^2} \Delta \Delta_F \approx \beta_F^2$$

Ottenendo dunque un fattore di moltiplicazione dell'errore simile a β_F^2 .

Ultima nota: si può esprimere I_C funzione di I_B :

$$I_C = \frac{\Delta_F}{1 - \Delta_F} I_B + \frac{I_{CBO}}{1 - \Delta_F}$$

Il secondo termine ricorda la "corrente di collettore a base aperta", ma per $I_B = 0$, supponendo che, per $I_B = 0$, π sia in rad; spesso ciò, in questa situazione, non accade: non basta dire che la base sia "aperta", ma bisogna assicurarsi di essere in rad.

Proviamo a questo punto a introdurre un maggior formalismo, entrando nei dettagli di questa struttura, in modo da provare a quantificare I_C e I_E .

Ciò che si può fare è sfruttare ciò che si sa della giunzione p-n, ossia la "teoria di Shockley", per il calcolo delle correnti singole; I_B , grazie a Kirchhoff, può essere calcolata come somma algebrica delle due. Si può sfruttare il fatto di essere in basso livello di iniezione, dunque usare la "legge della giunzione", calcolando in ciascun lato le concentrazioni di minoritari, come fatto di solito. Si sfrutta il fatto che, dato questo regime, il dispositivo è fondamentalmente controllato da meccanismi di diffusione, da qui la necessità di calcolare le concentrazioni; valgono infatti le seguenti relazioni:

$$I_E = I_{p,diff}(x_{EB}) + I_{n,diff}(x_{EB}) = -q A D_{pE} \frac{dp'}{dx} \Big|_{x_{EB}} + q A D_{nB} \frac{dn'}{dx} \Big|_{x_{BE}}$$

$$I_C = -I_{n,diff}(x_{EC}) - I_{p,diff}(x_{EC}) = -q A D_{nB} \frac{dn'}{dx} \Big|_{x_{EC}} + q A D_{pC} \frac{dp'}{dx} \Big|_{x_{CB}}$$

A questo punto, assunto il suddetto regime di iniezione, e l'ipotesi che non vi sia in ciascuna regione di smontamento né generazione né ricombinazione, al fine di quantificare le grandezze si devono quindi tipicare le varie concentrazioni nei vari punti. Questo si può fare combinando le informazioni fornite dalla legge della giunzione (la quale ci fornisce le concentrazioni ai bordi), e quelle sulla lunghezza del lato.

Nell'emettitore e nel collettore tutto è molto semplice: da una parte si ha iniezione di portatori minoritari, dall'altra un contatto ohmico; l'iniezione si assorbe nello spazio, andando verso il contatto (nel quale si ha eccesso nullo), o linearmente (se il lato è corto) o esponenzialmente (se il lato è lungo). Spesso l'emettitore è corto e il collettore lungo.

Ora: i minoritari nell'emettitore si assorbono linearmente. In realtà si ha un eccesso di portatori minoritari che dalla giunzione EB decresce linearmente nell'emettitore; questo è indice del fatto che la giunzione BE è polarizzata direttamente; il fatto che si ha un andamento "negativo" di lacune nel collettore è indice del fatto che la giunzione BC è invece polarizzata inversamente; dalle

leggi della giunzione, infatti, si sa che:

$$p'(x_{CB}) = \frac{n_i^2}{N_{BC}} \left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1 \right]; \quad p'(x_{EB}) = \frac{n_i^2}{N_{BE}} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right]$$

In realtà, $V_{BC} < 0$, $V_{BE} > 0$, dunque $p'(x_{EB})$ sarà molto grande, mentre $p'(x_{CB})$ prossimo a "-1". Questo termine andrà poi moltiplicato per gli esponenti derivanti dalla soluzione del modello matematico,

ottenendo:

$$p'(x) = p'(x_{CB}) \exp\left(-\frac{x - x_{CB}}{L_{PC}}\right) \quad \text{in C}$$

$$p'(x) = p'(x_{EB}) \cdot \frac{1}{W_E} [x - (x_E - W_E)] \quad \text{in E}$$

Derivando, si possono ottenere 2 dei 4 termini precedentemente richiesti, risolvendo il problema per emettitore e collettore.

Parliamo della base; date le precedenti ipotesi, si può calcolare le concentrazioni ai bordi mediante la legge della giunzione; ora i minoritari diffusi saranno elettroni, dunque:

$$n'(x_{BE}) = \frac{n_i^2}{N_{AB}} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right]; \quad n'(x_{BC}) = \frac{n_i^2}{N_{AB}} \left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1 \right]$$

Si ha qualcosa di un po' più complesso: da entrambi le parti ora si ha iniezione; usando un modello stazionario, di sola diffusione (quasi neutralità), si avrà una combinazione lineare di esponenziali come soluzione del modello:

$$n'(x) = A \exp\left(\frac{x}{L_{nB}}\right) + B \exp\left(-\frac{x}{L_{nB}}\right)$$

Applicando le opportune condizioni al contorno, si trova qualcosa del tipo:

$$n'(x) = n'(x_{BE}) \frac{\sinh\left(\frac{x_{BC} - x}{L_{nB}}\right)}{\sinh\left(\frac{W_B'}{L_{nB}}\right)} + n'(x_{BC}) \frac{\sinh\left(\frac{x - x_{BE}}{L_{nB}}\right)}{\sinh\left(\frac{W_B'}{L_{nB}}\right)}$$

Andò in questo caso, in pratica, si ha a che fare con combinazioni lineari di esponenziali, parlando di leggi di diffusione;

linearizzando, si trova un andamento ragionevole per la concentrazione; si può notare che essa all'inizio è molto elevata, alla fine molto bassa; questo è importante, dal momento che la distribuzione è una retta; dal momento che la derivata della retta è una costante, comunque, ci va benissimo perché significa che nella base non si hanno grandi problemi o riduzioni. Prima di passare al calcolo effettivo delle correnti, una nota: si trascura, spesso, la lunghezza di "iniettamento" rispetto a quella "fisica", in quanto influente.

Modello di Ebers-Moll

Sotto tutto ciò che è stato fatto, con tutte le approssimazioni introdotte, si può proporre il modello di Ebers-Moll del transistor bipolare:

$$I_E = \alpha_{11} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right] + \alpha_{12} \left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1 \right]$$

$$I_C = \alpha_{21} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right] + \alpha_{22} \left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1 \right]$$

Dove:

$$\alpha_{11} = -\frac{q A D_{PE} n_i^2}{N_{DE} \xi_E} - \frac{q A D_{NB} n_i^2}{N_{AB} L_{NB}} \coth\left(\frac{W_B}{L_{NB}}\right)$$

$$\alpha_{12} = \alpha_{21} = \frac{q A D_{NB} n_i^2}{N_{AB} L_{NB}} \left[\sinh\left(\frac{W_B}{L_{NB}}\right) \right]^{-1} \quad \text{Dove } \xi_a = \begin{cases} W_a, & \text{di corto} \\ L_{pa}, & \text{di lungo.} \end{cases}$$

$$\alpha_{22} = -\frac{q A D_{PC} n_i^2}{N_{DC} \xi_C} - \frac{q A D_{NB} n_i^2}{N_{AB} L_{NB}} \coth\left(\frac{W_B}{L_{NB}}\right)$$

Questo è il modello statico di ampio segnale controllato in tensione del BJT (costituito da un doppio bipolo non lineare comandato in tensione).

Si noti che i termini misti sono uguali tra loro; questo indica la reciprocità del modello di Ebers-Moll. "E" serve a tener conto della possibilità di avere a che fare con un lato lungo o corto. Il fatto che si chiami "di ampio segnale" fa intuire che esso valga anche al di fuori della rad: in rad poi, tenendo conto del segno delle tensioni, alcuni esponenziali possono essere trascurati, semplificando il modello.

Se trascuri gli esponenziali in BC, puoi esplicitare il fatto che:

$$I_E \approx -\frac{q A D_{PE} n_i^2}{N_{DE} W_E} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right] - \frac{q A D_{NB} n_i^2}{N_{AB} L_{NB}} \coth\left(\frac{W_B}{L_{NB}}\right) \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right]$$

Dunque, si può dire che:

$$F_{PBE} \approx \frac{D_{PE} n_i^2}{N_{DE} W_E} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right]; \quad F_{NEB} \approx \frac{D_{NB} n_i^2}{N_{AB} W_B} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right], \quad \frac{W_B}{L_{NB}} \ll 1$$

Alla stessa modo, da I_C , si può ricavare alla stessa maniera che:

$$F_{NBC} \approx \frac{D_{NB} n_i^2}{N_{AB} L_{NB}} \sinh^{-1}\left(\frac{W_B}{L_{NB}}\right) \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right]$$

A questo punto è possibile quantificare i fattori di efficienza di iniezione γ , nonché poi β , e valutare come devono comportarsi le grandezze in gioco per far tendere γ a 1; si può vedere che:

$$\gamma = \frac{1}{1 + \frac{D_{PE} N_{AB} W_B}{D_{NB} N_{DE} W_E}} \rightarrow \gamma \approx 1, \text{ se } N_{AB} \ll L_{NB} N_{DE}$$

Nulla di nuovo, anche se finalmente ne abbiamo una moti-
vazione formale.

Senza delle solite relazioni, si ricava da:

$$b = \left[\cosh\left(\frac{W_B}{L_{nB}}\right) \right]^{-1} \approx 1 - \frac{W_B^2}{2L_{nB}^2}$$

Per avere $b \approx 1$, $W_B \ll L_{nB}$; ci serve la base corta!

Ricordando le usuali definizioni, infine si può ricavare da:

$$\beta_F \approx \left(\frac{W_B^2}{2L_{nB}^2} + \frac{D_{PE} N_{AB} W_B}{D_{NB} N_{DE} W_E} \right)^{-1}$$

Spesso tuttavia il primo termine si trascura, date le basi molto
corte che si usano a realizzare. Questa formula ci fa intuire
un fatto: ciò che più fa decadere e ridurre β_F è il rapporto dei
drogaggi.

C'è da parlare di un fenomeno aggiuntivo, piuttosto recente: il
bandgap narrowing. Quando il drogaggio è tendenzialmente alto
(ma anche non altissimo: non si considerano degenerazioni del
semiconduttore), si può riscontrare un assottigliamento della
banda proibita. Questo fenomeno è anche piuttosto importante:
con $N = 10^{19}$, si ha un $\Delta E_g \approx 100$ meV, dunque decisamente
elevato! Questo fatto provoca un aumento della concentrazione
intrinseca n_i , la quale diventa dunque di fatto una funzione
del drogaggio.

Questo fatto a sua volta provoca una variazione imprevista
nel modello di Ebers-Moll: a_{11}, a_{12} infatti subiscono delle
variazioni. Quello che si può fare in pratica è tener conto di

questo fatto nell'espressione di β_F :

$$\beta_F = \frac{n_{iEB}^2}{n_{iRE}^2} \frac{D_{NB}}{D_{PE}} \frac{N_{DE} W_E}{N_{AB} W_B} = \frac{G_E}{G_B}$$

Questa semplificazione si può caratterizzare mediante i "numeri
di Gummel" per la base B e per l'emettitore E: G_B e G_E .

Il lato più influenzato dal fenomeno sarà l'emettitore, in quanto
il più drogato; dal momento che n_i^2 è proporzionale all'esponente
di E_g , ciò che si avrà alla fine è un'occasione dell'amplificazione
di corrente.

Quel che dovremo fare per ridurre ΔE_g è tener basso il dro-
gaggio nell'emettitore, ma dunque anche quello in base e collettore
(si ricordi che $N_{DE} > N_{AB} > N_{CE}$); in pratica, si userà in altra
maniera (mediante le eterogiunzioni).

Il modello di Ebers-Moll può essere presentato anche
circuitualmente; se si manipolano le equazioni precedente-
mente ricavate, si può trovare il seguente modello
circuitale:

$$I_E = -I_F + d_R I_R$$

$$I_C = -I_R + d_F I_F$$

Manipolando le equazioni si può ricavare questo modello,
e legare i parametri delle vecchie equazioni a questi. Una
sola nota: si ha a che fare con d_R e d_F ; d_F è già noto,

dal momento che esso è l'amplificazione di corrente a base comune, supponendo di essere in rad; I_E è un parametro assolutamente analogo, ma per la r_{e1} ! Esso interviene quando si scambiano i ruoli di emettitore e di collettore. Dal momento che i drogaggi sono "scalati", $I_E \approx I_C$. Si parla di "modello di imitazione". Da queste equazioni è anche possibile spiegare le caratteristiche del transistor; di solito esse vengono espresse "a emettitore comune": si presenta la corrente di collettore, parametrizzata da V_{CE} , in funzione della corrente di base.

L'andamento è noto: se V_{CE} è abbastanza grande, siamo in rad, e la corrente di uscita è costante con V_{CE} .

Dualmente si può studiare $I_B(V_{BE})$, vedendo che essa, per V_{BE} sufficientemente elevata, diventa simile a una retta verticale; se V_{CE} è maggiore di qualche decina di volt, la caratteristica diventa pressoché indipendente da V_{CE} .

Esistono andamenti approssimati di quelli noti, mediante rette orizzontali o verticali, in modo da avere andamenti con curve lorde ma molto semplici.

Non idealità

Esistono alcuni effetti di non idealità del transistor che il modello finora studiato non è in grado di prevedere; vediamo. Nella realtà, la corrente di uscita dipenderà da V_{CE} , dal momento che sussiste il cosiddetto "effetto Early": la lunghezza della regione neutra di base è dipendente da V_{CE} , dal momento che,

al crescere di V_{CE} , V_{BE} sta fissa, ma V_{BC} diventa sempre più negativa, rendendo la giunzione base-collettore polarizzata sempre più negativamente, allungando la regione di svuotamento, quindi aumentando la pendenza della retta risultante l'incremento dei minoranti in base; aumentando la pendenza, anche la corrente aumenta, grazie a V_{CE} , dunque si può dire che la corrente sia una funzione crescente di V_{CE} . Quello che cade è l'ipotesi di trascurare la lunghezza di svuotamento di base; si potrebbe tentare di usarla, ma ciò complicerebbe molto i conti; quello che si fa dunque è introdurre una corrente "empirica", come nei MOS.

Si dimostra che le varie rette della caratteristica finiscono per incontrarsi in un punto / tensione; essa è detta "tensione di Early": più essa è alta, meglio è. Questo fatto porta ad avere nel modello finché una resistenza di uscita non infinita: tanto più grande è la pendenza causata dall'effetto Early, tanto minore (dunque più influente) la resistenza di uscita vista nel collettore.

Vi è un altro effetto negativo: aumentando troppo V_{CE} , o rendendo troppo negativa V_{BC} , si può finire in breakdown, o in rottura elettrostatica, facendo esplodere la corrente di collettore. Tensioni eccessive elevate portano potenza dissipata elevata, che potrebbe anche portare a danni nel transistor.

Ora: come si può minimizzare l'effetto Early? Beh, un'idea sarebbe quella di avere base larga, cosa che ridurrebbe l'effetto

transistore, oppure far spostare più sul collettore che sulla base, riducendo N_A ; quest'ultima è più ragionevole, anche se introdurrebbe maggiore resistenza serie.

Il modello di ampio segnale è semplicemente Ebers-Moll, con l'aggiunta dei vari effetti capacitivi di diffusione e di smottamento. In rad però si sa che predomina da una parte solo la capacità di diffusione mentre in polarizzazione inversa quella di smottamento.

Tempo di transito: è il tempo impiegato per attraversare la base; esso determina il "limite della dinamica" (come nei FET): è il tempo impiegato per andar dall'emettitore al collettore.

Indicando con Q_B la carica in base, si definisce in prima approssimazione la variazione di carica di base dovuta al transito come:

$$\left. \frac{dQ_B}{dt} \right|_t = - \frac{Q_B}{\tau_t}$$

È possibile in rad identificare i flussi dei portatori e trovare la relazione per γ e per b , determinando β_F ; così da si vede che per massimizzare b serve un τ_t piccolo.

Transistore bipolare a eterostruttura (HBT)

Abbiamo visto che ci serve un τ_t basso e un drogaggio di collettore alto, al fine di aumentare γ , e dunque β ; dal momento che si ha il bandgap narrowing, si hanno modifiche tra queste richieste, "interazioni": si modificano entrambe.

Da un lato aumentare il drogaggio di emettitore non si può fare, per non clipare l'amplificazione, ma dall'altro lato allora

tenere basso quello di base, creando problemi per l'alta frequenza.

Ciò che si può fare è sincronizzare il drogaggio di emettitore e quello di base, ottenendo un γ elevato senza avere particolari vincoli di drogaggio; ciò si può fare, grazie a una eterostruttura. Il suo uso, inoltre, permette di avere elevata efficienza di iniezione, anche senza avere grossi drogaggi di emettitore/base.

L'arricchimento di E_c permette inoltre di avere elettroni a elevata energia che entrano in base, elettroni dunque molto veloci, che minimizzano il tempo di transito.

Qua ha fatto poco. Ne parlo?

3 dispositivi per l'elettronica di potenza.

Nelle applicazioni di potenza, i dispositivi più importanti sono gli interruttori controllati; nella trattazione che presenteremo, abbastanza discorsiva, prima parleremo di alcuni parametri utili a caratterizzare i dispositivi, poi di alcuni dispositivi.

Per quasi tutte le applicazioni senza, come detto, interruttori controllati; un esempio sono gli alimentatori a commutazione (switching); poniamo classificare gli interruttori in diversi modi:

In base alla modalità di controllo della commutazione:

- passivi (tipo diodi); in sostanza, dei rettificatori;
- attivi (tipo transistori).

In base ai possibili versi della corrente in em:

- unidirezionali (la corrente va solo in una direzione, come nei diodi o nei BJT/IGBT)
- bidirezionali: la corrente può scorrere in ambo le direzioni (MOSFET).

Altro criterio è il segno della tensione di blocco, ossia della massima tensione che si può tollerare in stato di circuito aperto; si può avere:

- blocco a tensione positiva o negativa (senza di solito a meglio transistor)
- indipendentemente dal segno della tensione (a più dispositivi).

I principali fattori di merito che si usano per caratterizzare gli interruttori sono 3:

- velocità di commutare (tempo di commutazione), parametro dinamico;

esso è soprattutto dipendente dal tempo di vita dei minoritari: la dinamica di emi dipende dalla presenza e dallo smaltimento dei minoritari (la cosa sarà dunque negativa per BJT/giunzione p-n; nei MOS o nei diodi Schottky, dipendenti non basati sui minoritari, questo parametro sarà migliore).

- Per quanto riguarda la "qualità del stato aperto", ossia la caratteristica d'isolamento, si usa come parametro R_{on} , ossia la resistenza posta dal canale conduttivo quando l'interruttore conduce. Al contrario di prima, i dispositivi bipolari si comportano meglio degli altri, sotto questo punto di vista.

- Massima tensione applicabile: quando si applica l'arco inavvertitamente una tensione, essa va, prima o poi, in breakdown; questo porta ad avere gravi conseguenze nel dispositivo, dal momento che da queste correnti derivano grosse dissipazioni, dunque il dispositivo finisce per essere distrutto. Se E supera un certo valore critico si va in effetto avalanche/rever. Anche sotto questo punto di vista i dispositivi bipolari si comportano meglio.

Volendo "abbassare il campo" E , ci sono due soluzioni:

- a) Abbassare le tensioni (nulla, per dispositivi di potenza)
- b) Fare in modo che il campo si "svilga" in una regione lunga, in modo da poter applicare tensioni elevate (da qua si capisce perché i dispositivi di potenza sono così larghi).

Per quanto riguarda questi ultimi aspetti, ci sono anche limitazioni intrinseche dei materiali: ogni materiale ha un proprio massimo valore di E accettabile, prima del quale si ottiene nec-

Caratteristiche di generazione/ricombinazione in controllabili; questo è detto "capacità dielettrica". Per questo vogliamo materiali a capacità dielettrica più alta, ma essi sono anche materiali ad alta banda proibita (come carbonio di silicio o nitruro di gallio). Per il breakdown, dato un materiale, dovremo cercare il punto da fissare il valore di picco, dunque limitarlo.

Esistono diverse varianti sulla struttura del dispositivo, al fine di migliorare la tensione di breakdown. Se pensiamo a una giunzione p-n, il picco è proprio alla giunzione, dunque sarà essa il punto da tenere sotto controllo. Problema: le strutture reali non sono monodimensionali! Quello che si fa è avere un contatto metallico (es. in un diodo Schottky) che occupa parte della superficie e, quando si applica una ΔV tale da polarizzare innescando la giunzione MS, il metallo ha tutti i punti equipotenziali, ma a causa della presenza di spigoli, si hanno discontinuità di E , dunque gli spigoli diventano, nella realtà, i punti critici. Questi per noi sono i veri punti critici! Quello che si cerca di fare è far in modo da "abbassare" questi picchi, mediante la progettazione delle terminazioni!

Vediamo qualche esempio

- Field plate: si interpongono tra metallo e semiconduttore dell'isolante, in modo da sgombrare nell'angolo parte del campo allentando lo spigolo.
- Ci sono varianti, per esempio mettere dei ^{spigoli} di metallo su strati di ossidi e tali da mettere regioni p^+ in modo da distribuire la caduta di potenziale su spigoli più elevati. L'"anello" p^+ circonda tutto, con bordi arrotondati. La regione p^+ circonda tutti i punti della

giunzione. Ci sono strutture con anodi più simili di geometria. Si usano con questo termine a volere ^{diodi} corpi che reggono anche 120 V, e corpi simili alla capacità dielettrica del silicio.

Dunque, ciò che vogliamo fare è abbassare il campo elettrico. C'è ancora una possibilità, diversa: allargare la regione di movimento, allungando il campo; questo si può fare, passando da una giunzione p-n a una giunzione p-i-n (con in mezzo una regione intrinseca). Il campo elettrico, in questo modo, ha un andamento trapezoidale anziché triangolare, e una conseguente altezza molto ridotta: area costante, base aumentata, dunque altezza ridotta! L'unico problema di questa struttura è: essa va bene ora, ma in conduzione la regione intrinseca fa aumentare R_{on} .

Transistori di potenza

Double-diffusion MOSFET

Parliamo di transistori per applicazioni di potenza, e partiamo da quelli MOS; la prima particolarità sta nel fatto che le lunghezze di gate sono intorno al μm , dunque elevate. La struttura è però fondamentalmente diversa da quella dei MOSFET tradizionali: siamo abituati a pensare ai MOS come a strutture planari, in cui source e drain sono sulla superficie, e la corrente scorre parallelamente alla superficie. Qui cambia tutto: i transistori di potenza sono strutture verticali, esse in cui il flusso di corrente è diretto ortogonalmente alla superficie.

Questo è utile perché strutture di questo tipo sono facilmente parallelizzabili: basta mettere vicino tante celle, il flusso entra verso la superficie e il canale permette che esso scenda così.

Ciò è legato dunque alla comodità di parallelizzazione.

Uno tra source o drain sta in superficie, il drain dalla parte opposta. Il canale si fa così: in fondo c'è il drain, dunque la sua regione n_1 e una regione n_2 , una poco drogata (per aumentare il breakdown); le varie drogate tipo p servono per fare i bordi, il substrato del MOS. Il MOS sta nel gate omologato nell'ombra, a contatto con la regione p. Le regioni "n" sono contatti di source; il canale si forma all'interfaccia tra semiconduttore e il canale (ovvero, si ha tanti canali!). Questa struttura è facilmente parallelizzabile. La corrente entra dal drain, va attraverso la n_2 , entra nelle regioni p, esce dal source. Problema: queste varie drogate fanno da body, fanno anche una giunzione p-n, dunque bisogna polarizzare in verso opportuno, garantendo $V_{DS} > 0$: il drain va tenuto a potenziale più elevato.

Stato OFF: se la tensione è sotto soglia, si deve ricordare di polarizzare il "diode di body", dunque più la regione n_2 è ampia e meno è drogata, più essa si comporta come vogliamo, distribuendo il campo e facendosi così allargare il suo pice (di fatto, questa è una giunzione p-n).

Stato ON: si hanno i canali conduttivi; data $V_{DS} > 0$, si parte (agli elettroni) dal source, attraversano in corrente il body, raggiungono la zona n_2 e così arrivano al drain. La regione n_2 è poco drogata,

dunque avrà R_{ON} non indifferente; bisognerà tirare un trade-off tra R_{ON} e $V_{breakdown}$, lavorando su drogaggio e ampiezza di n_2 . Queste strutture non sono ideali, ma tanto sono dedicate alla potenza. Doppia diffusione: si parte dal substrate, si diffonde nel body, e nel source.

Transistoro bipolare.

Ande qua vogliamo elevata corrente, dunque parallelizzare, dunque struttura verticale. Il collettore sarà una regione $n_1 n_2$ (in modo da spalmare il campo elettrico). La struttura poi presenta la base, e sulla superficie si impiantano vari emettitori (n).

Questa struttura è meglio in termini di R_{ON} , perché in altissimo esso è in saturazione, dunque il lato p mette a disposizione lavoro nella regione intrinseca, aumentando la conducibilità elettrica.

Problema: qua, a differenza del MOS abbiamo i minoratori, che non quelli che mandano l'effetto precedente; il problema è che quando si spinge il transistor i minoratori n_2 devono assorbire, secondo il modello di ricombinazione, ottenendo tempi lunghi.

Simulated Gate Bipolar Transistor (IGBT)

Si tratta di una specie di "misto": l'IGBT è molto simile a un DMOSFET, ma con una variante: al posto della regione n fondo, si ha una regione p. Ciò lo rende simile al DMOSFET come costruzione, ma molto diverso come funzionamento: al posto della regione $n_1 n_2$ si ha una struttura p n_1 p, ovvero un transistoro bipolare! Si ha il canale conduttivo del MOS al solito punto, e poi il pnp. Creativamente si impara

un Darlington!

La porta ed avere una veloce transizione, ma allo stesso tempo una R_{on} piccola e $V_{brakdown}$ alta!

Il BJT ha infatti una tensione di breakdown più alta nel MOS, dal momento che nel bipolar si può permettere di aumentare la regione n_1 perché l'effetto di iniezione delle lacune quando si va in condizione allora R_{on} ; a parità di R_{on} la parte n può essere fatta più ampia, dunque $V_{brakdown}$ essere più alta.