

UNIVERSITA' DEGLI STUDI DI PISA

Facoltà di Ingegneria

Corso di Laurea in Ingegneria Elettronica

TESI DI LAUREA

**Simulazione di Gate Stack Alternativi per Applicazioni
di Logica e Memorie non Volatili**

Relatori:

Prof. Giuseppe Iannaccone

Prof. Bruno Pellegrini

Candidato:

Andrea Campera

Anno Accademico 2003/2004

*A Silvia, che mi è stata vicina
in tutti questi mesi di lavoro
e mi ha sempre sostenuto
come solo lei poteva fare.
Alla mia famiglia tutta
che mi ha permesso di arrivare
fino a questo momento.
A tutti i miei amici
che hanno rallegrato e reso
più felici questi anni di studio.
Infine una dedica all'ultimo
arrivato: Mattia, spero
che tu possa avere un futuro
bellissimo. Vi voglio bene.*

Andrea

Indice

Introduzione	1
1 Corrente Tunnel	4
1.1 Introduzione	4
1.2 Trattazione semiclassica	5
1.2.1 Struttura a bande in tre dimensioni	6
1.3 Trattazione quantistica	16
1.4 Matrici di trasferimento	29
1.4.1 Matrice di trasferimento elementare	35
1.4.2 Matrice di trasferimento per un potenziale localizzato	37
1.5 Corrente tunnel	45
1.5.1 Densità di stati per unità di energia longitudinale e round trip time	51
1.6 Approssimazione quasi classica (metodo WKB)	58
1.7 Tunneling diretto e di Fowler Nordheim	61
2 Simulazione di strutture MOS con dielettrici ad alto k	67
2.1 Introduzione	67
2.2 Misure di capacità	70
2.3 Effetto delle trappole interfacciali sulla capacità	71
2.4 Capacità delle trappole interfacciali	74
2.5 Calcolo della capacità degli stati interfacciali	76
2.6 Capacità misurata ad alta frequenza	80
2.7 Fermi level pinning	81
2.8 Simulazioni con ossido di afnio	88
2.9 Passaggio dall'ossido di afnio al silicato di afnio	100
2.10 Listato dei principali file d'ingresso	112

3 Memorie a nanocristalli	115
3.1 Introduzione	115
3.2 Modello fisico	119
3.3 Master equation	127
3.4 Soluzione della master equation nel dominio del tempo	132
3.5 Programmi utilizzati per le simulazioni	135
3.6 Risultati delle simulazioni	141
3.7 Listato dei principali file d'ingresso	164
Appendice A	
Teoria delle perturbazioni	170
A.1 Perturbazioni indipendenti dal tempo	170
A.2 Teoria delle perturbazioni con degenerazione	174
A.3 Perturbazioni dipendenti dal tempo: transizioni quantistiche	176
Appendice B	
Discretizzazione delle equazioni del modello	181
Conclusioni	188
Bibliografia	191

Introduzione

In questo lavoro di tesi abbiamo cercato di approfondire alcuni aspetti delle moderne tecnologie nanoelettroniche, ossia di quelle tecnologie che utilizzano o producono dispositivi con dimensioni caratteristiche di alcuni nanometri.

Ci siamo rivolti in particolare a due aspetti molto importanti: il primo argomento che abbiamo trattato è stato quello dei gate stack alternativi, il secondo le memorie a nanocristalli.

Il problema dei gate stack alternativi al classico biossido di silicio (che è sempre stato utilizzato per fare gli ossidi di gate dei MOS e che ha fatto la fortuna del silicio nella moderna elettronica) nasce dal continuo scaling delle dimensioni dei dispositivi. Quando si scende a dimensioni troppo piccole dell'ossido di silicio, più precisamente intorno a qualche nanometro, le correnti di perdita attraverso l'ossido diventano sempre più grandi e hanno un impatto significativo sulla dissipazione di potenza di un circuito integrato. Si prevede quindi di sostituire il classico ossido di gate con degli stack di materiali diversi, ossia pile di strati di materiali diversi. La ITRS prevede infatti che a partire dall'anno 2010 si debbano introdurre gate stack alternativi nei MOS fabbricati con la tecnologia a 45 nm; è prevista l'introduzione, tra l'altro, di gate metallici in sostituzione del classico poly-silicio. Senza l'introduzione di gate stack con dielettrici *high-k* le correnti di perdita raggiungerebbero un valore troppo elevato; ecco perché molte tra le più importanti industrie del settore (STM, MOTOROLA, PHILIPS per citarne alcune) si sono mosse in anticipo svolgendo una ricerca fortissima sui dielettrici *high-k*. L'IMEC, in Belgio, è uno dei più grandi centri di ricerca del mondo sull'argomento dei dielettrici alternativi; i nostri

studi si sono concentrati su tre strutture interamente realizzate e caratterizzate in IMEC. Il materiale che più di ogni altro è stato studiato è l'afnio; l'afnio infatti si è dimostrato, nei vari composti, come il migliore dei materiali con cui costruire gli stack di dielettrici e si pensa possa entrare molto presto in produzione, in special modo gli stack con silicati di afnio. Le nostre simulazioni sono state mirate ad estrarre i parametri fisici di questi composti; non esiste attualmente un processo stabile che garantisca il valore di questi parametri. Per far questo abbiamo confrontato i risultati sperimentali con quelli delle nostre simulazioni; per queste ultime abbiamo utilizzato un codice realizzato e sviluppato presso il Dipartimento di Ingegneria dell'Informazione dell'Università di Pisa; per i risultati sperimentali sono stati utilizzati i lavori svolti dal Prof. Crupi dell'Università della Calabria presso l'IMEC. Al Prof. Crupi va un sentito ringraziamento anche per i consigli che mi ha fornito durante questi mesi di lavoro.

Allo stato attuale dell'arte i silicati di afnio si presentano come il più promettente dielettrico alternativo e si pensa possano entrare in produzione entro pochi anni. Questi, ad differenza di altri composti, come l'ossido di afnio, risolvono in parte uno dei problemi di molti dielettrici alternativi: la degradazione della mobilità degli elettroni nel canale. Nei silicati questa raggiunge l'80% della curva universale. Le cause della degradazione della mobilità devono ancora essere comprese appieno, così come, per quel che riguarda l'aspetto tecnologico, devono essere studiati processi stabili e a bassa densità di difetti.

Nella seconda parte del nostro lavoro abbiamo studiato le memorie a nanocristalli, anch'esse introdotte negli ultimi anni e che promettono di risolvere molti problemi delle memorie Flash convenzionali in futuro molto prossimo. Le memorie a nanocristalli hanno una robustezza ai difetti dell'ossido e alle SILC nettamente superiore; un percorso

conduttivo verso il substrato, infatti, comporta in un caso la perdita di informazione associata all'intero gate flottante, mentre nell'altro la perdita di informazione associato ad un solo nanocristallo. Si possono inoltre utilizzare ossidi più sottili, questo a tutto vantaggio delle tensioni di programmazione (leggi potenze dissipate) che risulteranno più basse e delle velocità di programmazione della cella, anch'esse nettamente inferiori.

L'introduzione sul mercato delle memorie a nanocristalli è possibile in un tempo relativamente breve, anche perché la tecnologia necessaria è addirittura più semplice del processo dual poly che si usa per le memorie a gate flottante.

Il nostro lavoro si è rivolto, in particolare, ad una modellizzazione del comportamento del singolo nanocristallo (detto anche dot) con un modello originariamente sviluppato per lo studio delle SILC (Stress Induced Leakage Current); tramite questo modello, opportunamente modificato, abbiamo cercato di ricavare le curve sperimentali a nostra disposizione, cercando di capire il meccanismo di intrappolamento dell'elettrone nel singolo dot.

CAPITOLO 1

Corrente tunnel

1.1 Introduzione

In questo capitolo descriveremo il meccanismo di conduzione per effetto tunnel attraverso una barriera di potenziale, generalmente costituita da un ossido, con una descrizione, utilizzando l'approssimazione WKB, della corrente tunnel di tipo diretto e di tipo Fowler-Nordheim.

Verrà inoltre presentato l'approccio numerico che il programma di simulazione utilizza per risolvere la sola equazione di Poisson, quindi per ottenere soluzioni semiclassiche, oppure le equazioni di Poisson e Schrödinger accoppiate per soluzioni di tipo quantistico.

Il tunneling di elettroni attraverso una barriera di potenziale fu sviluppato inizialmente da Fowler e Nordheim nel 1928 per spiegare l'emissione di campo di elettroni da un elettrodo di tungsteno nel vuoto. Ulteriori studi sull'argomento vennero fatti da Zener nel 1934 per spiegare la corrente di breakdown nei diodi pn in inversa; per giunzioni in diretta fortemente drogate, con resistenza negativa, i concetti del tunneling vennero applicati da Esaki nel 1957. Il tunneling è stato poi applicato a strutture metallo-ossido-semiconduttore negli anni 50 quando vennero presentati i primi dispositivi ad effetto di campo. Nel seguito faremo principalmente riferimento al tunneling di elettroni ma questo non significa che non sia presente anche un fenomeno di tunneling di lacune.

Tenendo conto della minore massa efficace degli elettroni rispetto a quella delle lacune e del fatto che l'altezza della barriera che devono superare le lacune all'interfaccia ossido-silicio è di 4.25 eV mentre l'altezza della barriera per gli elettroni è di 3.13 eV circa, possiamo trascurare la corrente tunnel di lacune.

1.2 Trattazione semiclassica

Il nostro scopo è risolvere l'equazione di Poisson $\nabla \cdot (\varepsilon_0 \varepsilon_r \nabla \varphi(\underline{r}, t)) = -\rho(\underline{r}, t)$ che lega la densità di carica al potenziale elettrico. In una dimensione l'equazione di Poisson diventa:

$$\frac{\partial}{\partial x} \left(\varepsilon_0 \varepsilon_r \frac{\partial \Psi}{\partial x} \right) = -\rho(\varphi) \quad (1.2.1)$$

con $\varepsilon_0 = 8.854 \cdot 10^{-12} F/m$ permittività dielettrica del vuoto, ε_r costante dielettrica relativa, φ potenziale elettrico e ρ densità di carica per unità di volume. In un materiale semiconduttore la densità di carica assume la particolare forma:

$$\rho(\varphi) = q \cdot \{ p(\varphi) - n(\varphi) + N_D^+(\varphi) - N_A^-(\varphi) \} + \rho_f \quad (1.2.2)$$

dove $q = 1.602 \cdot 10^{-19} C$ è la carica dell'elettrone, $p(\varphi)$ la concentrazione di lacune, $n(\varphi)$ la concentrazione di elettroni, $N_D^+(\varphi)$ la concentrazione di donatori ionizzati, $N_A^-(\varphi)$ la concentrazione di accettori ionizzati e ρ_f la densità di carica fissa. Per poter risolvere l'equazione di Poisson si devono trovare i termini precedenti in funzione di φ oppure di x . Cominciamo col ricavare la concentrazione di elettroni in banda di conduzione; essa è data da:

$$n = 2 \int_{E_C}^{\infty} N(E) f(E, E_F) dE \quad (1.2.3)$$

dove E_C è il valore di energia corrispondente al fondo della banda di conduzione, $N(E)$ è la densità di stati per unità di volume e per unità di energia e $f(E, E_F)$ è la funzione di distribuzione di Fermi-Dirac, data da:

$$f(E, E_F) = \frac{1}{1 + \exp\left(\frac{E - E_F}{k_B \cdot T}\right)} \quad (1.2.4)$$

dove E_F è l'energia corrispondente al livello di Fermi, k_B è la costante di Boltzmann ($k_B = 1.38 \cdot 10^{-23} J / K$) e T è la temperatura assoluta.

Il fattore 2 tiene conto del fatto che ogni stato può essere occupato da due elettroni con spin opposto.

Per il calcolo della densità di elettroni in banda in conduzione dobbiamo calcolare $N(E)$, densità di stati, e per farlo dobbiamo prima parlare della struttura a bande in tre dimensioni.

1.2.1 Struttura a bande in tre dimensioni

La struttura a bande in una dimensione, come noto, si ottiene schematizzando l'andamento del potenziale periodico e risolvendo l'equazione di Schrödinger (modello di Kronig-Penney, vedi per esempio [1]). La soluzione ci fornisce l'andamento dell'energia delle varie bande in funzione di k_x come mostrato in figura 1.1.

L'intervallo $-\pi/a < k_x < \pi/a$ (con a passo reticolare del cristallo) contiene tutti i k_x non equivalenti e costituisce quella che viene detta prima zona di Brillouin.

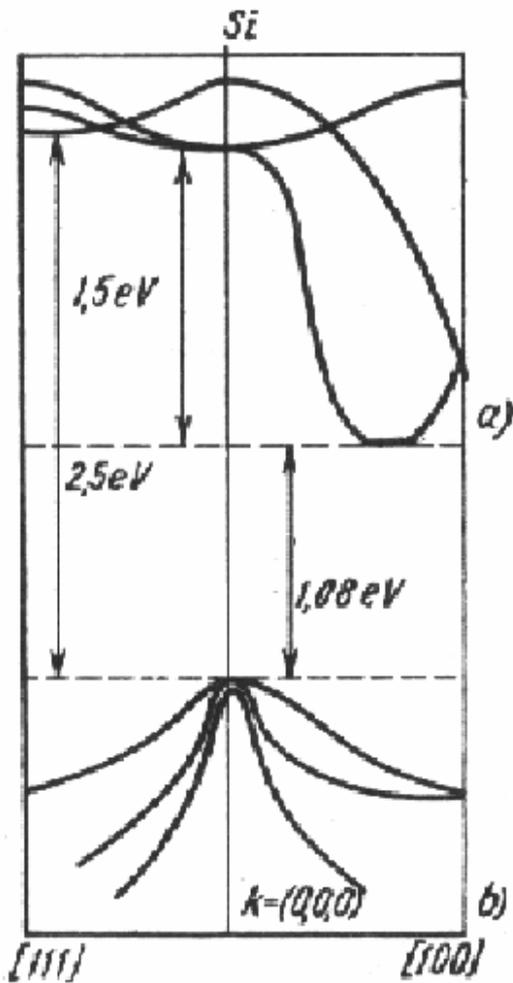


fig 1.1.a

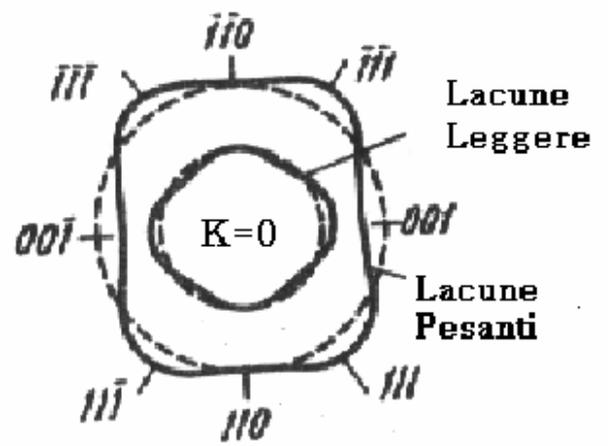


fig 1.1.b

Figura 1.1: struttura a bande del silicio (1.a) e superfici equienergetiche relative alle lacune nell'intorno del massimo della banda di valenza (1.b).

In 3D non è possibile rappresentare graficamente la struttura a bande; ci dobbiamo limitare a caratterizzare il materiale fornendo l'andamento di E in funzione di k nelle varie direzioni di simmetria e la forma delle superfici equipotenziali.

La funzione $E = E(k_x, k_y, k_z)$ può essere sviluppata in serie nell'intorno di un massimo o di un minimo, che indichiamo con \underline{k}_0 ; otteniamo:

$$E(\underline{k}) = E(\underline{k}_0) + (\underline{k} - \underline{k}_0) \left. \frac{dE}{d\underline{k}} \right|_{\underline{k}=\underline{k}_0} + \frac{(\underline{k} - \underline{k}_0)^2}{2} \left. \frac{d^2E}{d\underline{k}^2} \right|_{\underline{k}=\underline{k}_0} + TS \quad (1.2.5)$$

dove TS indica i termini di ordine superiore al secondo.

Il termine $(\underline{k} - \underline{k}_0) \left. \frac{dE}{d\underline{k}} \right|_{\underline{k}=\underline{k}_0}$ è nullo (siamo in un massimo o in un minimo e quindi la

derivata prima si annulla) mentre il termine $(\underline{k} - \underline{k}_0)^2 \left. \frac{d^2E}{d\underline{k}^2} \right|_{\underline{k}=\underline{k}_0}$ è costituito da nove

derivate seconde parziali fatte rispetto a k_x, k_y, k_z .

$$\frac{dE}{d\underline{k}} = \left(\frac{\partial E}{\partial k_x}, \frac{\partial E}{\partial k_y}, \frac{\partial E}{\partial k_z} \right) \quad (1.2.6)$$

$$\frac{d^2E}{d\underline{k}^2} = \left(\frac{d}{d\underline{k}} \frac{\partial E}{\partial k_x}, \frac{d}{d\underline{k}} \frac{\partial E}{\partial k_y}, \frac{d}{d\underline{k}} \frac{\partial E}{\partial k_z} \right) = \begin{pmatrix} \frac{\partial^2 E}{\partial k_x^2} & \frac{\partial^2 E}{\partial k_y \partial k_x} & \frac{\partial^2 E}{\partial k_z \partial k_x} \\ \frac{\partial^2 E}{\partial k_x \partial k_y} & \frac{\partial^2 E}{\partial k_y^2} & \frac{\partial^2 E}{\partial k_z \partial k_y} \\ \frac{\partial^2 E}{\partial k_x \partial k_z} & \frac{\partial^2 E}{\partial k_y \partial k_z} & \frac{\partial^2 E}{\partial k_z^2} \end{pmatrix} \quad (1.2.7)$$

Le quantità $\frac{\partial^2 E}{\partial k_i \partial k_j}$, con $i, j = x, y, z$, costituiscono le componenti di un tensore simmetrico

che può essere reso diagonale scegliendo opportunamente gli assi di riferimento.

Con una tale scelta di assi lo sviluppo in serie diventa:

$$E(\underline{k}) = E(\underline{k}_0) + \frac{1}{2}(k_x - k_{0x})^2 \frac{\partial^2 E}{\partial k_x^2} + \frac{1}{2}(k_y - k_{0y})^2 \frac{\partial^2 E}{\partial k_y^2} + \frac{1}{2}(k_z - k_{0z})^2 \frac{\partial^2 E}{\partial k_z^2} + TS \quad (1.2.8)$$

quindi una superficie ad $E = \text{costante}$ sarà rappresentata nello spazio dei k da un ellissoide.

Infatti possiamo riscrivere l'espressione precedente come:

$$E(\underline{k}) = E(\underline{k}_0) + \hbar^2 \left(\frac{(k_x - k_{0x})^2}{2m_x^*} + \frac{(k_y - k_{0y})^2}{2m_y^*} + \frac{(k_z - k_{0z})^2}{2m_z^*} \right) \quad (1.2.9)$$

se si definiscono le masse efficaci, in generale diverse, mediante le relazioni

$$m_i^* = \frac{\hbar^2}{\frac{\partial^2 E}{\partial k_i^2}} \quad (1.2.10)$$

e si trascurano i termini di ordine superiore al secondo (TS).

Abbiamo sviluppato $E(k)$ in serie di Taylor fino al secondo ordine rispetto al minimo della banda di conduzione E_c , perciò si parla di approssimazione parabolica.

Osserviamo che le masse efficaci degli elettroni hanno segno positivo se calcolate nell'intorno di un minimo, segno negativo se calcolate nell'intorno di un massimo; inoltre ricordiamo brevemente che l'atomo di silicio ha 14 elettroni con configurazione $1s^2 2s^2 2p^6 3s^2 3p^2$ e che la struttura cristallina è quella del diamante. Quindi il reticolo è di tipo FCC e la zona di Brillouin è un BCC la cui cella primitiva è un ottaedro troncato.

La figura 1.1 mostra l'andamento della struttura a bande del Silicio lungo le direzioni (100), indicata con ΓX , e (111), indicata con ΓL .

Come possiamo notare il gap è di tipo indiretto e il minimo della banda di conduzione si ha per un valore di \underline{k} diverso da 0; inoltre data la simmetria della struttura si hanno sei minimi equivalenti; nell'intorno di tali minimi le superfici equipotenziiali sono ellissoidi di rotazione.

Una superficie equipotenziale è descritta dalla seguente equazione:

$$E(\underline{k}) = \hbar^2 \left(\frac{(k_x - k_{0x})^2 + (k_y - k_{0y})^2}{2m_t^*} + \frac{(k_z - k_{0z})^2}{2m_l^*} \right) \quad (1.2.11)$$

dove m_t^* è la massa efficace trasversale e m_l^* è la massa efficace longitudinale.

Consideriamo una superficie ad energia costante in banda di conduzione che, come abbiamo visto, è costituita da sei ellissoidi; per calcolare la densità degli stati in BC consideriamo uno degli ellissoidi (alla fine dovremo quindi moltiplicare per sei) la cui equazione è:

$$\begin{aligned} E(\underline{k}) &= E(\underline{k}_0) + \hbar^2 \left(\frac{(k_x - k_{0x})^2}{2m_x^*} + \frac{(k_y - k_{0y})^2}{2m_y^*} + \frac{(k_z - k_{0z})^2}{2m_z^*} \right) = \\ &= E_C + \hbar^2 \left(\frac{(k_x - k_{0x})^2}{2m_x^*} + \frac{(k_y - k_{0y})^2}{2m_y^*} + \frac{(k_z - k_{0z})^2}{2m_z^*} \right) \end{aligned} \quad (1.2.12)$$

che possiamo riscrivere come

$$1 = \frac{(k_x - k_{0x})^2}{a^2} + \frac{(k_y - k_{0y})^2}{b^2} + \frac{(k_z - k_{0z})^2}{c^2} \quad (1.2.13)$$

ovvero l'equazione di un ellissoide in forma canonica, avendo posto

$$a^2 = \frac{2(E - E_C)m_x^*}{\hbar^2}, \quad b^2 = \frac{2(E - E_C)m_y^*}{\hbar^2}, \quad c^2 = \frac{2(E - E_C)m_z^*}{\hbar^2}.$$

Adesso, per calcolare la densità degli stati, consideriamo il volume di un ellissoide dato da

$$V_E = \frac{4}{3} \pi abc = \frac{4\pi}{3\hbar^3} (8m_x^*m_y^*m_z^*)^{\frac{1}{2}} (E - E_C)^{\frac{3}{2}} \quad (1.2.14)$$

e il volume infinitesimo compreso tra due superfici ad energia E ed $E + dE$

$$dV_E = \frac{2\pi}{\hbar^3} (8m_x^*m_y^*m_z^*)^{\frac{1}{2}} (E - E_C)^{\frac{1}{2}} dE \quad (1.2.15)$$

Dividendo dV_E per il volume occupato, nello spazio dei k , da un singolo stato si ottiene il numero di stati compresi tra E ed $E + dE$ per un singolo ellissoide.

Per calcolare il volume occupato da un singolo stato consideriamo una regione di dimensioni macroscopiche $L_x L_y L_z$ schematizzabile come una buca di potenziale per l'elettrone.

Scriviamo in questa regione l'equazione di Schrödinger:

$$-\frac{\hbar^2}{2} \left[\frac{\partial}{\partial x} \left(\frac{1}{m_x} \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{1}{m_y} \frac{\partial}{\partial y} \right) + \frac{\partial}{\partial z} \left(\frac{1}{m_z} \frac{\partial}{\partial z} \right) \right] \Psi_i(x, y, z) + E_C(x, y, z) \Psi_i(x, y, z) = E_i \Psi_i(x, y, z)$$

dove $\Psi_i(x, y, z)$ sono le autofunzioni dell'hamiltoniano relative agli autovalori E_i i quali corrispondono alle energie permesse per l'elettrone. Risolviamo tale equazione con le condizioni al contorno di Dirichlet: se l'altezza della buca è molto maggiore dell'energia E dell'elettrone, possiamo approssimare la buca reale con una buca di altezza infinita per il calcolo dei vettori d'onda permessi. In questo caso la condizione al contorno si riduce ad imporre che la funzione d'onda si annulli sulle pareti esterne, cioè per $x=0$ e $x=L_x$, $y=0$ e $y=L_y$, $z=0$ e $z=L_z$.

Si trovano quindi le componenti del vettore d'onda permesse per l'elettrone, date da:

$$k_i = \frac{\pi}{L_i} n_i \text{ con } n_i = 1, 2, \dots, \infty \text{ e } i = x, y, z.$$

Come si vede i vettori d'onda sono quantizzati e questo comporta anche una quantizzazione dell'energia che risulta pari a:

$$E = E_C + \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z} \right) = E_C + \frac{\hbar^2 \pi^2}{2} \left(\frac{n_x^2}{m_x L_x} + \frac{n_y^2}{m_y L_y} + \frac{n_z^2}{m_z L_z} \right) \quad (1.2.16)$$

Dato che il volume occupato da uno stato nello spazio degli impulsi è $d\Gamma = \frac{\pi^3}{L_x L_y L_z}$,

dividendo dV_E per $d\Gamma$ si ottiene:

$$\frac{dV_E}{d\Gamma} = L_x L_y L_z \frac{2}{\hbar^3 \pi^2} (8m_x^* m_y^* m_z^*)^{\frac{1}{2}} (E - E_C)^{\frac{1}{2}} dE \quad (1.2.17)$$

Dobbiamo adesso considerare che:

- $L_x L_y L_z$ è il volume del semiconduttore
- sono permessi soltanto i valori di k per $n > 0$ e quindi appartenenti ad un solo ottante
- si hanno M_C minimi equivalenti

l'espressione della densità di stati per unità di volume e di energia è data da:

$$N(E) = M_C \frac{1}{\sqrt{2} \cdot \hbar^3 \pi^2} (m_x^* m_y^* m_z^*)^{\frac{1}{2}} (E - E_C)^{\frac{1}{2}} \quad (1.2.18)$$

e ponendo $m_{de}^3 = m_x^* m_y^* m_z^*$ si ottiene

$$N(E) = M_C \frac{1}{\sqrt{2} \pi^2} \frac{(E - E_C)^{\frac{1}{2}}}{\hbar^3} (m_{de})^{\frac{3}{2}} \quad (1.2.19)$$

Dal procedimento esposto si comprende come la massa efficace per la densità degli stati sia una grandezza che ha origine dalla forma delle superfici equipotenziali e non abbia relazione con il comportamento dinamico dell'elettrone, ossia come risposta a campi elettrici.

Per il Silicio abbiamo $m_x^* = m_y^* = m_t^* = 0.19m_0$ e $m_z^* = m_l^* = 0.98m_0$ quindi $m_{de}^* = 1.08m_0$,

valore da usare per il calcolo di N_C .

La struttura a bande per la BV è approssimabile, in prossimità di E_V , con due parabole di diversa curvatura. Ne segue l'esistenza di lacune con masse efficaci diverse, le cosiddette lacune *pesanti* (m_{lh}) e le cosiddette lacune *leggere* (m_{hh}), come si vede dalla figura 1.1.

Nei semiconduttori con il reticolo del diamante o della zincoblenda la banda di valenza presenta tre rami degeneri in $\underline{k} = 0$, ciascuno dei quali è doppiamente degenero, nel senso che può ospitare due elettroni di spin opposto. Lo stato di degenerazione delle bande è cambiato da un effetto quantistico noto come interazione spin-orbitale. Tale interazione coinvolge per $\underline{k} \approx 0$ uno dei due rami degeneri di lacune pesanti e il ramo delle lacune leggere e produce due nuovi rami che, nell'intorno di $\underline{k} = 0$, posseggono la stessa curvatura del ramo delle lacune leggere; uno dei due, però, è spostato verso il basso di un'energia Δ e prende il nome di banda di split-off. Si ha quindi, in definitiva, una coppia di bande di lacune leggere e pesanti, ciascuna delle quali porta due stati a spin opposto, che sono degeneri per $\underline{k} = 0$, e una banda di split-off a due stati a spin opposto. I valori di Δ sono molto diversi a seconda del materiale; per il silicio vale $\Delta = 0.044 \text{ eV}$.

La banda di split-off entra in gioco solo in particolari processi nell'ambito dei semiconduttori, per cui viene spesso trascurata nella rappresentazione della banda di valenza; per quanto riguarda le lacune, il massimo della BV si trova al centro della zona di Brillouin.

Procedendo in modo analogo a quanto fatto per gli elettroni si ottiene la densità degli stati per le lacune in banda di valenza:

$$\begin{aligned}
 N_h(E) &= \frac{1}{\sqrt{2} \cdot \pi^2 \hbar^3} (m_{lh})^{\frac{3}{2}} (E_V - E)^{\frac{1}{2}} + \frac{1}{\sqrt{2} \cdot \pi^2 \hbar^3} (m_{hh})^{\frac{3}{2}} (E_V - E)^{\frac{1}{2}} = \\
 &= \frac{1}{\sqrt{2} \cdot \pi^2 \hbar^3} (m_{dh})^{\frac{3}{2}} (E_V - E)^{\frac{1}{2}}
 \end{aligned} \tag{1.2.20}$$

avendo imposto che $m_{dh} = \left(m_{lh}^{\frac{3}{2}} + m_{hh}^{\frac{3}{2}} \right)^{\frac{2}{3}}$.

Adesso abbiamo la densità di stati e la funzione di distribuzione quindi possiamo passare a calcolare la concentrazione di elettroni in banda di conduzione. Come avevamo visto si ha:

$n = 2 \int_{E_C}^{\infty} N(E) f(E, E_F) dE$ e sostituendo otteniamo

$$n = N_C \frac{2}{\sqrt{\pi}} F_{1/2} \left(\frac{E_F - E_C}{k_B T} \right) \quad (1.2.21)$$

dove $N_C = 2 \left(\frac{2\pi m_{de} k_B T}{h^2} \right)^{\frac{3}{2}}$ è la densità efficace di stati in banda di conduzione e $F_{1/2}$ è

l'integrale di Fermi definito come segue:

$$F_{1/2}(\eta) = \frac{1}{\left(\frac{1}{2}\right)!} \int_0^{\infty} \frac{x^{\frac{1}{2}}}{1 + e^{x-\eta}} dx \quad (1.2.22)$$

Analogamente si ottiene la concentrazione di lacune in banda di valenza

$$p = N_V \frac{2}{\sqrt{\pi}} F_{1/2} \left(\frac{E_V - E_F}{k_B T} \right) \quad (1.2.23)$$

dove $N_V = 2 \left(\frac{2\pi m_{dh} k_B T}{h^2} \right)^{\frac{3}{2}}$ è la densità efficace di lacune in banda di valenza e

$m_{dh} = \left(m_{lh}^{\frac{3}{2}} + m_{hh}^{\frac{3}{2}} \right)^{\frac{2}{3}}$ è la massa efficace delle lacune per la densità di stati, con m_{lh} e m_{hh} la

masse delle lacune leggere e pesanti rispettivamente.

Tutte queste quantità sono note una volta che si conosca il legame tra E_C (e quindi E_V una volta noto il gap del materiale) e φ . Tale relazione si ottiene da semplici considerazioni sulla struttura a bande del materiale ed è:

$$E_C = (\Phi_m)_{rif} + (E_F)_{rif} - \chi - \varphi \quad (1.2.24)$$

dove $(\Phi_m)_{rif}$ e $(E_F)_{rif}$ sono rispettivamente il valore della funzione di lavoro e il valore dell'energia di Fermi nel punto assunto come riferimento per il potenziale ($\varphi = 0$) e χ è l'affinità elettronica del materiale.

Adesso dobbiamo esplicitare i termini $N_D^+(\varphi)$ e $N_A^-(\varphi)$; gli atomi droganti introducono livelli energetici nel gap del silicio che possono essere occupati da un solo elettrone con spin up o down.

La concentrazione degli atomi droganti ionizzati si può fare una volta nota la probabilità di occupazione sia dei livelli donatori che dei livelli accettori. Tale probabilità è fornita dalla statistica di Fermi-Dirac, che deve però essere modificata; nel caso di atomi donatori la funzione di distribuzione è data da:

$$f_D(E_D, E_F) = \frac{1}{1 + \frac{1}{g_D} \exp\left(\frac{E_D - E_F}{k_B T}\right)} \quad (1.2.25)$$

in cui g_D tiene conto della degenerazione dello spin, pari a due nel caso del silicio, e E_D indica il livello di energia introdotto dall'impurezza nella banda proibita. Il termine di degenerazione deriva dal fatto che il livello dell'impurezza può essere occupato da un solo elettrone, a differenza degli stati in banda di conduzione per i quali infatti g_D vale 1.

Nel caso degli accettori la funzione di distribuzione è data da:

$$f_A(E_A, E_F) = \frac{1}{1 + g_A \exp\left(\frac{E_F - E_A}{k_B T}\right)} \quad (1.2.26)$$

in cui E_A indica il livello energetico degli accettori e g_A tiene conto della degenerazione della banda di valenza per $\underline{k} = 0$ e del fatto che uno stato accettore, come gli stati donatori, può essere occupato da una lacuna alla volta con qualsiasi spin; nel caso del silicio g_A vale 4.

La densità di donatori ionizzati è quindi data da:

$$N_D^+ = N_D [1 - f_D(E_D, E_F)] = \frac{N_D}{1 + g_D \exp\left(\frac{E_F - E_D}{k_B T}\right)} \quad (1.2.27)$$

con N_D concentrazione di atomi donatori. Otteniamo un risultato analogo per gli atomi accettori, tenendo conto che stavolta bisogna considerare i livelli pieni e non vuoti di elettroni:

$$N_A^- = N_A f_A(E_A, E_F) = \frac{N_A}{1 + g_A \exp\left(\frac{E_F - E_A}{k_B T}\right)} \quad (1.2.28)$$

con N_A concentrazione di atomi accettori.

1.3 Trattazione quantistica

Il modello semiclassico che abbiamo sviluppato fino ad ora non è sufficiente per comprendere le caratteristiche di una struttura MOS perché non tiene conto di importanti fenomeni quantistici che avvengono nella struttura stessa. Consideriamo ad esempio il caso di un condensatore MOS (o più genericamente una struttura MIS se l'isolante non è un

ossido) al quale venga applicata una tensione positiva di gate. In queste condizioni all'interfaccia ossido-silicio, se la tensione è maggiore della tensione di soglia, si ha l'inversione del semiconduttore ossia la concentrazione di portatori minoritari supera quella di maggioritari vicino all'interfaccia e si ha la formazione di un canale. Gli elettroni, supponiamo di utilizzare un substrato drogato p, nel canale, sono confinati lungo la direzione x perpendicolare all'interfaccia mentre sul piano parallelo all'interfaccia (in questo caso yz) non si ha nessun tipo di confinamento. Si forma quindi un gas bidimensionale di elettroni la cui densità è diversa da quella di bulk perché il confinamento lungo x introduce una quantizzazione delle energie in tale direzione che modifica gli stati disponibili per l'elettrone.

Per ricavare la densità di stati dobbiamo prima ricavare la densità di stati bidimensionale in direzione yz ; consideriamo una regione di dimensioni $L_y L_z$:

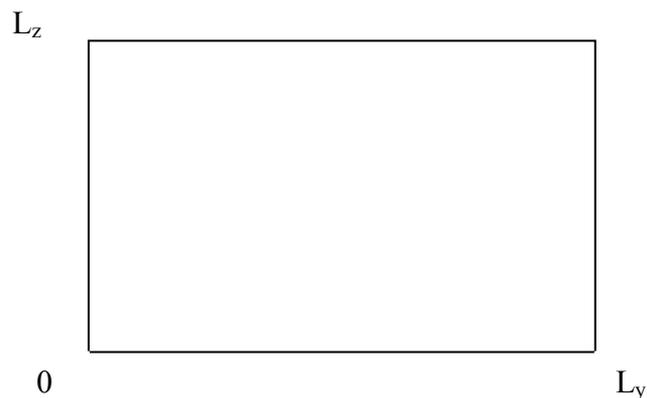


Figura 1.2: regione dello spazio bidimensionale dove risolvere l'equazione di Schrödinger.

Consideriamo tale zona come una buca di potenziale e risolviamo l'equazione di Schrödinger al suo interno. Tale equazione nel caso bidimensionale è:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(y,z) = E\Psi(y,z) \quad (1.3.1)$$

Cerchiamo una autofunzione del tipo $\Psi(y,z) = Y(y)Z(z)$; utilizzando la tecnica di separazione delle variabili è facile mostrare che le funzioni $Y(y), Z(z)$ devono soddisfare alle equazioni:

$$\begin{aligned} -\frac{\hbar^2}{2m}\frac{d^2Y(y)}{dy^2} &= E_1Y(y) \\ -\frac{\hbar^2}{2m}\frac{d^2Z(z)}{dz^2} &= E_2Z(z) \end{aligned} \quad (1.3.2)$$

con l'energia totale data da $E = E_1 + E_2$ e dove abbiamo supposto, per semplicità, che la massa efficace sia costante e pari a m .

Ciascuna delle equazioni 1.3.2 ammette soluzioni generali del tipo:

$$\begin{aligned} Y(y) &= A_y e^{ik_y y} + B_y e^{-ik_y y} \\ Z(z) &= A_z e^{ik_z z} + B_z e^{-ik_z z} \end{aligned} \quad (1.3.3)$$

con A_y, B_y, A_z, B_z costanti.

Tali soluzioni devono soddisfare le condizioni al contorno nei punti di discontinuità del potenziale (continuità dell'autofunzione e della sua derivata). Se l'altezza della buca è molto maggiore dell'energia E dell'elettrone, possiamo approssimare la buca reale con una buca di altezza infinita per il calcolo dei vettori d'onda permessi. In questo caso la condizione al contorno si riduce ad imporre che la funzione d'onda si annulli sulle pareti esterne, cioè per $y = 0$ e $y = L_y$, $z = 0$ e $z = L_z$.

Imponendo tali condizioni alle equazioni 1.3.3 si trova che le ampiezze A_i e B_i ($i = y, z$) sono uguali e opposte e che le componenti dei vettori d'onda permessi devono soddisfare le relazioni:

$$k_y = \frac{n_y \pi}{L_y} \quad \text{e} \quad k_z = \frac{n_z \pi}{L_z} \quad n_y, n_z = 1, 2, \dots, \infty.$$

La funzione d'onda risulta, quindi:

$$\Psi(y, z) = A \sin k_y y \sin k_z z \quad (1.3.4)$$

L'energia associata a tali stati è data da $E = \frac{\hbar^2}{2} \left(\frac{k_y^2}{m_y} + \frac{k_z^2}{m_z} \right)$ che nello spazio reciproco è

l'equazione di un'ellisse. Riscrivendola in forma canonica otteniamo:

$$\frac{\hbar^2 k_y^2}{2m_y E} + \frac{\hbar^2 k_z^2}{2m_z E} = 1 \quad (1.3.5)$$

e possiamo immediatamente trovare la lunghezza dei semiassi che è

$$a = \sqrt{\frac{2m_y E}{\hbar^2}} \quad \text{e} \quad b = \sqrt{\frac{2m_z E}{\hbar^2}}.$$

Possiamo quindi scrivere l'area dell'ellisse che è semplicemente $S = \pi ab$ e poi ricavare l'area della regione compresa tra due ellissi di energia E ed $E + dE$ che sarà:

$$dS = \pi \frac{2 \cdot \sqrt{m_y m_z}}{\hbar^2} dE \quad (1.3.6)$$

L'area occupata da uno stato nel reticolo reciproco è data da $\frac{\pi^2}{L_y L_z}$ e quindi il numero di

stati con energia compresa tra E ed $E + dE$ è $dN_s = \frac{2 \sqrt{m_y m_z}}{\pi \hbar^2} L_y L_z dE$; si deve poi

moltiplicare per 2 per tener conto dello spin e dividere per 4 visto che sono permessi solo le componenti di \underline{k} aventi $n > 0$.

A questo punto abbiamo la densità di stati per unità di energia e per unità di superficie che è:

$$\rho_{2D} = \frac{\sqrt{m_y m_z}}{\pi \hbar^2} \quad (1.3.7)$$

che, come noto, è indipendente dall'energia.

Siamo adesso in grado di trovare la densità di stati totale. Nel caso di confinamento quantistico lungo x gli stati permessi nello spazio dei k hanno la struttura di figura 1.3.

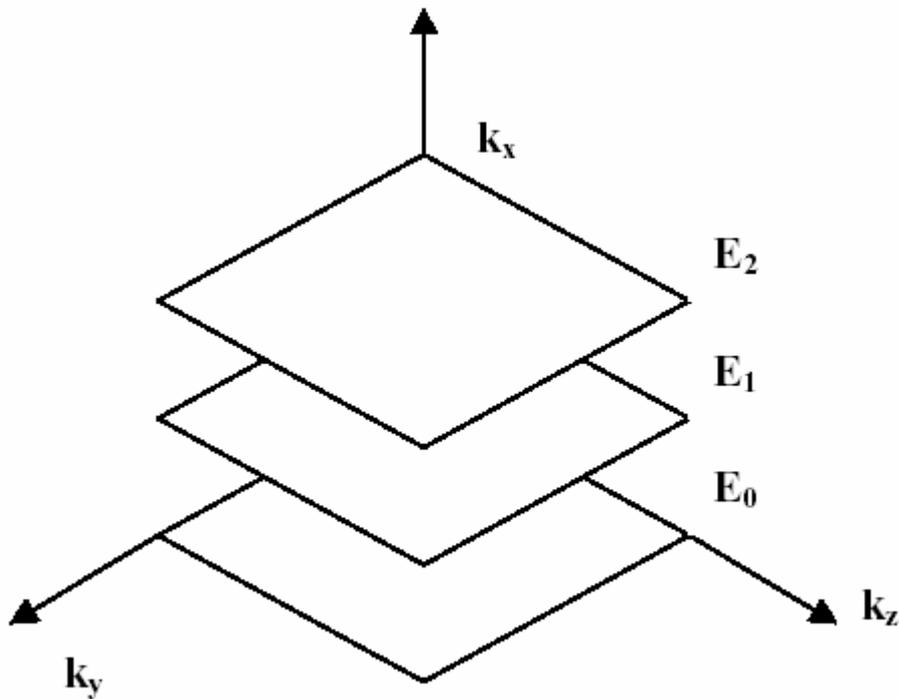


Figura 1.3: quantizzazione degli stati lungo la direzione di confinamento e loro suddivisione in sottobande.

Come possiamo notare abbiamo stati molto fitti sui piani paralleli all'asse yz e stati più radi in direzione x ; per ogni piano distinto si parla di sottobanda che quindi indica l'insieme degli stati appartenenti ad un piano parallelo al piano $k_y k_z$. I valori di k_x (o equivalentemente dell'energia longitudinale E_l) sono dati dagli autovalori della equazione di Schrödinger e tutti gli stati di una sottobanda sono associati allo stesso k_x .

Abbiamo ricavato in precedenza la ρ_{2D} , densità di stati per unità di energia e per unità di superficie, che ci serve adesso per trovare la densità di stati totale. Come possiamo vedere dalla figura per $E < E_0$ non ci sono stati permessi quindi la densità superficiale di stati è $\rho_s(E) = 0$ mentre non appena $E = E_0$ la superficie energetica equipotenziale è tangente al

primo piano. Quando $E_0 < E < E_1$ allora possiamo scrivere $E = \frac{\hbar^2}{2} \left(\frac{k_y^2}{m_x} + \frac{k_z^2}{m_z} \right) + E_0$ e visto

che non ci sono stati permessi tra E_0 ed E_1 la densità degli stati è ancora quella della prima

sottobanda e quindi quella bidimensionale $\rho_s(E) = \rho_{2D} = \frac{\sqrt{m_y m_z}}{\pi \hbar^2}$.

Come si può intuire per $E_1 < E < E_2$ gli stati permessi saranno quelli relativi alla prima e alla seconda sottobanda e la densità di stati sarà doppia rispetto al caso precedente, quindi

$$\rho_s(E) = \frac{2\sqrt{m_y m_z}}{\pi \hbar^2} \quad (1.3.8)$$

L'andamento generale è quindi una gradinata e più precisamente ogni volta che l'energia

longitudinale E raggiunge una nuova sottobanda la densità di stati aumenta di $\frac{\sqrt{m_y m_z}}{\pi \hbar^2}$

quindi possiamo scrivere:

$$\rho_s(E) = \sum_{i=1}^{\infty} \frac{\sqrt{m_y m_z}}{\pi \hbar^2} u(E - E_i) \quad (1.3.9)$$

dove $u(E - E_i)$ è la funzione di Heavyside che vale 1 se $E > E_i$, 0 altrimenti. Dalla $\rho_s(E)$ si ricava facilmente la densità di stati per unità di volume e di energia

$$\rho_V(E, x) = \sum_{i=1}^{\infty} \frac{\sqrt{m_y m_z}}{\pi \hbar^2} |\Psi_i(x)|^2 u(E - E_i) \quad (1.3.10)$$

dove $\Psi_i(x)$ è l'autofunzione relativa all'autovalore E_i .

Adesso procediamo in modo analogo al caso semiclassico per il calcolo della concentrazione di elettroni, possiamo infatti scrivere:

$$\begin{aligned} n(x) &= \int \rho_V(E, x) f(E, E_F) dE = \\ &= \sum_{i=1}^{\infty} \frac{\sqrt{m_y m_z}}{\pi \hbar^2} |\Psi_i(x)|^2 \cdot \int_0^{+\infty} \frac{u(E - E_i)}{1 + \exp\left(\frac{E_i + E_T + E_F}{k_B T}\right)} dE_T \end{aligned} \quad (1.3.11)$$

dove $E_T = \frac{\hbar^2 k_y^2}{2m_y} + \frac{\hbar^2 k_z^2}{2m_z}$; stavolta manca il fattore 2 dello spin fuori dall'integrale (come

nell'equazione 1.2.3), ma dobbiamo ricordarci che l'abbiamo già inserito nella ρ_{2D} e quindi è già compreso in $\rho_V(E, x)$.

Considerando valori di $E > E_i$, allora $\forall i$ la funzione di Heavyside vale 1 e si ottiene,

moltiplicando numeratore e denominatore dell'integrando per $\exp\left(-\frac{E_i + E_T - E_F}{k_B T}\right)$, la

seguinte espressione:

$$\begin{aligned}
 n(x) &= \sum_{i=1}^{\infty} \frac{\sqrt{m_y m_z}}{\pi \hbar^2} |\Psi_i(x)|^2 \int_0^{+\infty} \frac{\exp\left(-\frac{E_i + E_T - E_F}{k_B T}\right)}{1 + \exp\left(-\frac{E_i + E_T - E_F}{k_B T}\right)} dE_T \\
 &= \frac{k_B T}{\pi \hbar^2} \sqrt{m_y m_z} \sum_{i=0}^{\infty} |\Psi_i(x)|^2 \cdot \left[-\ln \left[1 + \exp\left(-\frac{E_i + E_T - E_F}{k_B T}\right) \right] \right]_0^{+\infty} \\
 &= \frac{k_B T}{\pi \hbar^2} \sqrt{m_y m_z} \sum_{i=0}^{\infty} |\Psi_i(x)|^2 \ln \left[1 + \exp\left(-\frac{E_i - E_F}{k_B T}\right) \right]
 \end{aligned} \tag{1.3.12}$$

Nel nostro caso bisogna considerare che il silicio ha sei minimi equivalenti in banda di conduzione e bisogna vedere a cosa corrisponde il termine $\sqrt{m_y m_z}$ per ciascuna superficie equienergetica. Guardando la figura 1.4 si può osservare che lungo k_x si ha $m_x = m_l$, $m_y = m_t$, $m_z = m_t$ quindi si ha $\sqrt{m_y m_z} = m_t$; lungo k_y si ha $m_x = m_t$, $m_y = m_l$, $m_z = m_t$ quindi si ha $\sqrt{m_y m_z} = \sqrt{m_l m_t}$; lungo k_z si ha $m_x = m_t$, $m_y = m_t$, $m_z = m_l$ quindi si ha $\sqrt{m_y m_z} = \sqrt{m_t m_l}$.

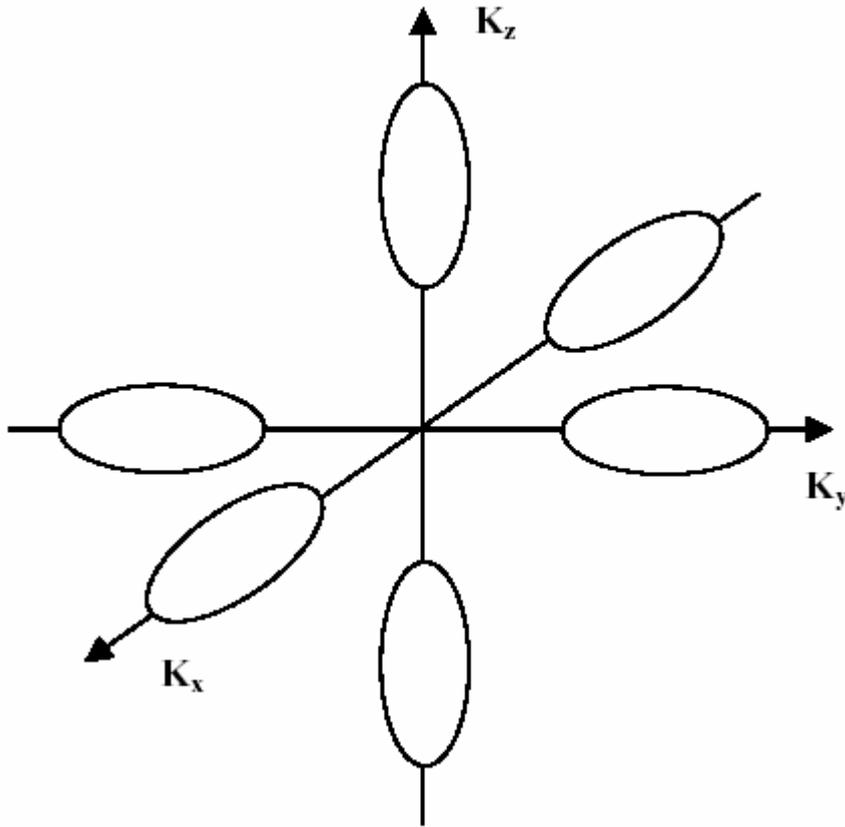


Figura 1.4: superfici equienergetiche del silicio nell'intorno del minimo della banda di conduzione.

Adesso per calcolare $n(x)$ dobbiamo risolvere l'equazione di Schrödinger per determinare le autofunzioni e in particolare dovremo risolvere due volte l'equazione di Schrödinger, una volta per i due minimi longitudinali, con massa efficace m_l ed un'altra volta per i quattro minimi trasversali, con massa efficace m_t .

$$-\frac{\hbar^2}{2} \frac{\partial}{\partial x} \left(\frac{1}{m_l} \frac{\partial \Psi_{il}}{\partial x} \right) + E_C(x) \Psi_{il} = E_{il} \Psi_{il} \quad (1.3.13)$$

$$-\frac{\hbar^2}{2} \frac{\partial}{\partial x} \left(\frac{1}{m_t} \frac{\partial \Psi_{it}}{\partial x} \right) + E_C(x) \Psi_{it} = E_{it} \Psi_{it} \quad (1.3.14)$$

Otteniamo alla fine l'espressione per $n(x)$ che viene usata dal programma di simulazione:

$$n(x) = 2 \frac{k_B T}{\pi \hbar^2} m_t \sum_{i=0}^{\infty} |\Psi_{it}(x)|^2 \ln \left[1 + \exp \left(-\frac{E_{it} - E_F}{k_B T} \right) \right] + 4 \frac{k_B T}{\pi \hbar^2} \sqrt{m_l m_t} \sum_{i=0}^{\infty} |\Psi_{it}(x)|^2 \ln \left[1 + \exp \left(-\frac{E_{it} - E_F}{k_B T} \right) \right] \quad (1.3.15)$$

Tutto il procedimento può essere portato avanti anche per le lacune e alla fine si ottiene:

$$p(x) = \frac{k_B T}{\pi \hbar^2} m_{dh} \sum_{i=0}^{\infty} |\Psi_i(x)|^2 \ln \left[1 + \exp \left(-\frac{E_F - E_i}{k_B T} \right) \right] \quad (1.3.16)$$

Adesso diamo una breve descrizione di come opera il programma di simulazione per calcolare le quantità che abbiamo trovato fino a questo punto.

Il programma di simulazione è costituito da tre routine principali: *griglia*, *struttura* e *principale* (il *principale* può assumere nomi diversi come *ifresh* per lo studio della corrente tunnel diretta o *prnano* nello studio delle memorie a nanocristalli; le modifiche sono piccolissime e riguardano principalmente i file di uscita che vengono generati). La routine *griglia* viene eseguita per prima ed ha il compito di discretizzare il dominio della soluzione; le specifiche per tale discretizzazione le vengono fornite dal file d'ingresso *grild.dat*: questo file è costituito dal numero di zone in cui si vuole suddividere il dominio e, per ogni zona, dall'ascissa iniziale, dall'ascissa finale, dal passo iniziale e dal passo finale. La routine calcola la ragione con cui varia il passo della griglia e, se necessario, modifica il passo iniziale e quello finale in modo che il numero di passi per coprire la regione risulti un numero intero. La routine genera due file di uscita: il file *grigliax.out* che contiene le ascisse dei vari punti della griglia, e *punti.out* che contiene il numero di punti della griglia (tale numero non deve essere superiore a 256).

La seconda routine è *struttura* che riceve in ingresso i file di uscita generati da *griglia* e un file che descrive i vari strati della struttura e che si chiama *str1d.dat*.

C'è un altro file che *struttura* legge ed è *input.material*; spiegheremo per prima la forma di *str1d.dat*. Gli strati in cui suddividiamo la struttura sono strati fisici e non necessariamente coincidono con le zone fissate in *gr1d.dat*, anzi ne differiscono quasi sempre. Nel file d'ingresso *str1d.dat* vengono specificati, oltre al numero di strati in cui è suddivisa la struttura, anche parametri esterni come la temperatura (espressa in Kelvin), il livello di Fermi e la funzione lavoro sia del gate di controllo che del contatto di bulk, il coefficiente di sottorilassamento e la tolleranza per la norma euclidea e il flag *flagpot* la cui funzione sarà spiegata più avanti. Per ciascuno strato di cui è costituito il materiale, vengono specificati il tipo di materiale, le ascisse iniziale e finale, la concentrazione di atomi accettori e donatori e la carica fissa; inoltre viene fissato il valore di due flag, *flagn* e *flagp*, che abilitano (1) o meno (0) la presenza di elettroni e lacune nello strato.

Il file *input.material* contiene i parametri fisici dei vari materiali che si possono inserire nella struttura; vengono forniti il valore dei livelli energetici introdotti dagli atomi accettori e donatori, la massa efficace (longitudinale e trasversale) per gli elettroni, la massa efficace delle lacune, il gap a $T = 0$ K e i parametri alfa e beta per il calcolo del gap (il gap ad una generica temperatura T è calcolato come $E_G(T) = E_G(0) - \alpha T^2 (T + \beta)^{-1}$) a qualsiasi temperatura, la costante dielettrica e l'affinità elettronica per vari materiali (Si, SiO₂, HfO₂, etc.).

La routine *struttura* associa ad ogni punto della griglia i vari parametri fisici a seconda dello strato al quale il punto appartiene e per ogni parametro genera un file d'uscita; avremo quindi molti file d'uscita come *mel.out* (che contiene i valori della massa efficace

longitudinale dell'elettrone nei vari punti della griglia), *eps.out* (che contiene il valore della costante dielettrica nei vari punti della griglia) e così via per tutti i parametri fisici di interesse. Oltre a questi *struttura* genera i file *lavoro1.out*, *lavoro2.out* e altri file di appoggio per *principale*: il primo contiene il numero di strati nei quali si vuole risolvere l'equazione di Schrödinger e, per ciascuno di essi, l'ascissa iniziale e finale e, mediante il flag *flags* il tipo di portatori da trattare quantisticamente (se *flags* vale 1 vengono trattati quantisticamente gli elettroni, se vale 2 vengono trattate quantisticamente solo le lacune, mentre se *flags* vale 3 viene fatto il calcolo quantistico sia per gli elettroni che per le lacune). Il secondo (*lavoro2.out*), contiene il numero delle regioni da considerare barriere di potenziale e, per ciascuna di esse, l'ascissa iniziale e finale ed il valore di τ_{relax} per il calcolo della densità longitudinale di stati e del round trip time nelle zone adiacenti alla barriera nelle quali si ha il confinamento dei portatori.

Prima dell'esecuzione della routine *principale* viene utilizzata un'altra routine che serve a definire il drogaggio del substrato; la routine *drogaggio* riscrive il file *ennea.out* (o *ennedi.out* se il substrato è di tipo n) in modo da simulare la distribuzione di drogaggio che si ottiene per impiantazione ionica o diffusione. Tramite il file d'ingresso *drog1d.dat* vengono passati alla routine vari parametri che servono a definire la forma della gaussiana con cui modelliziamo il processo di drogaggio: tali parametri sono la dose, il range proiettato, la varianza e le ascisse iniziale e finale della regione in cui prevediamo questo drogaggio.

La routine *principale* è quella che risolve le equazioni di Poisson e Schrödinger ed è quindi in grado di ricavare l'andamento delle bande, la concentrazione dei portatori e la corrente tunnel sia in assenza che in presenza di trappole nell'ossido. *principale* legge tutti

i file di uscita prodotti da *struttura* e da *griglia*, dopodiché si crea un potenziale di riferimento supponendo che la tensione del gate di controllo cada tutta tra il gate stesso e il punto di mezzo della struttura. Il potenziale di partenza è questo se *flagpot* letto sul file *lavoro.out* (generato anch'esso da *struttura*) vale 1; l'utilizzo di tale potenziale di partenza può richiedere un numero troppo elevato di iterazioni dato che, per la sua estrema semplicità, molto probabilmente si scosterà di molto dal potenziale effettivo. Se il valore di *flagpot* è 0 il potenziale di partenza viene letto nel file *p1d.out*; tale file può essere generato con una routine che si chiama *defpot* se è noto, almeno approssimativamente, l'andamento del potenziale. Il valore del flag *flagpot* viene messo a zero anche nel caso si mandi la routine *principale* in esecuzione più volte, con diversi valori della tensione di gate, per esempio per ottenere una caratteristica I-V. Alla fine di ogni esecuzione la routine scrive in *p1d.out* l'andamento del potenziale che, se la variazione della differenza di potenziale è piccola, non si discosterà molto dalla soluzione della simulazione successiva, la quale, per questo motivo, verrà portata a termine con un numero di iterazioni molto ridotto. Una volta scelta la stima iniziale per il potenziale, la routine *principale* non fa altro che iterare il procedimento, come viene spiegato nell'appendice B sulla discretizzazione e risoluzione numerica delle equazioni di Poisson e di Schrödinger.

Prima di passare al calcolo della corrente tunnel dobbiamo parlare di una tecnica molto importante che ci servirà per risolvere l'equazione di Schrödinger e per calcolare il coefficiente di trasmissione di una barriera di potenziale; si tratta del metodo delle matrici di trasferimento.

1.4 Matrici di trasferimento

Il metodo delle matrici di trasferimento è uno dei più usati e comodi per risolvere l'equazione di Schrödinger monodimensionale e può essere usato in presenza di qualsiasi tipo di barriera di potenziale. Questo è utilizzato dal programma di simulazione per il calcolo del coefficiente di trasmissione attraverso la barriera di potenziale della struttura, che può essere di forma arbitraria.

Supponiamo che $V(x)$, energia potenziale, sia una funzione reale e ad un sol valore di x , tale che:

$$V(x) = \begin{cases} V_1 & a < x < b \\ V_2 & c < x < d \end{cases} \quad (1.4.1)$$

con $-\infty < b \leq c < +\infty$; si può prendere ad esempio la $V(x)$ mostrata in figura 1.5.

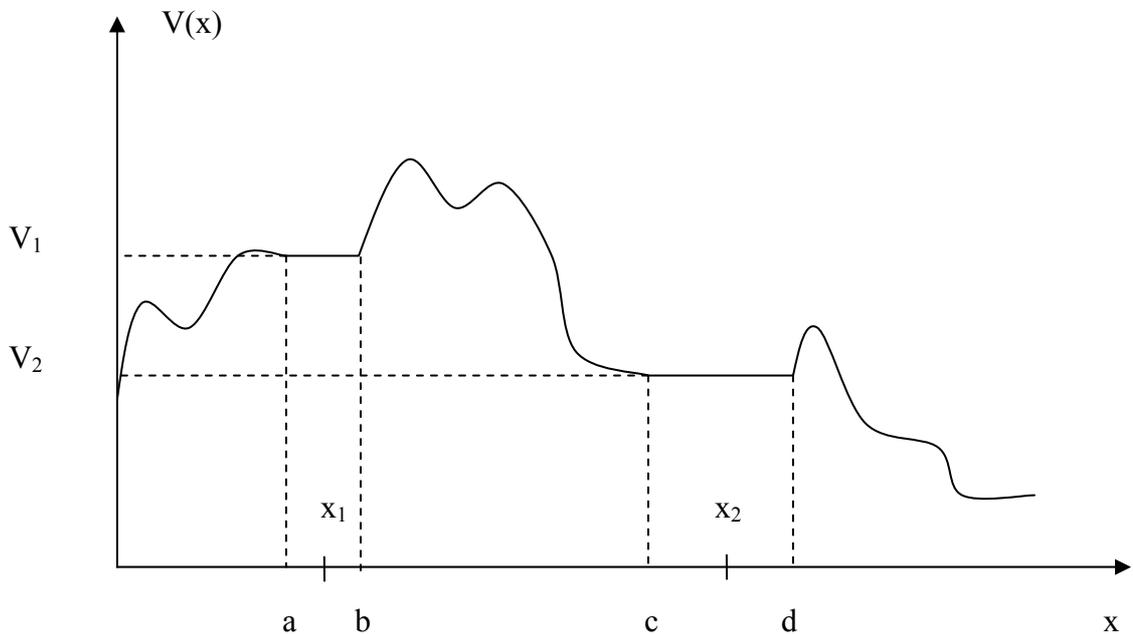


Figura 1.5: andamento dell'energia potenziale caratterizzato dal fatto che $V(x)=V_1$ per $a < x < b$; $V(x)=V_2$ per $c < x < d$.

Scriviamo l'equazione di Schrödinger monodimensionale:

$$-\frac{\hbar^2}{2} \frac{d}{dx} \left(\frac{1}{m} \frac{d\Psi(x)}{dx} \right) + V(x)\Psi(x) = E\Psi(x) \quad (1.4.2)$$

La soluzione si trova semplicemente nelle regioni $a < x < b$ e $c < x < d$ ed è scrivibile come sovrapposizione di due onde piane, una viaggiante verso le x positive ed una viaggiante verso le x negative. Se scegliamo x_1 ed x_2 interni agli intervalli $[a, b]$ e $[c, d]$ rispettivamente; abbiamo, in generale, per $i = 1, 2$:

$$\Psi(x) = A_i e^{ik_i(x-x_i)} + B_i e^{-ik_i(x-x_i)} \quad \text{per } a < x < b \text{ e } c < x < d \quad (1.4.3)$$

dove k_i è in generale complesso.

In particolare per k_i vale:

- se $E > V_i$ allora k_i è reale positivo, $k_i = \frac{\sqrt{2m_i(E - V_i)}}{\hbar}$
- se $E < V_i$ allora k_i è immaginario puro con segno positivo

Per trovare il valore delle costanti basta fissare Ψ e $\frac{1}{m} \frac{d\Psi}{dx}$ in un punto dell'asse x dato

che l'equazione è lineare del secondo ordine.

Dalla linearità dell'equazione di Schrödinger segue una dipendenza di A_2 e B_2 da A_1 e B_1 tramite una matrice 2×2 ad elementi complessi, e viceversa. Possiamo quindi scrivere:

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M_{x_1, x_2} \begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \quad (1.4.4)$$

La matrice M_{x_1, x_2} così definita è la *matrice di trasferimento* per l'intervallo (x_1, x_2) .

In generale non è semplice ricavare la matrice di trasferimento per un generico andamento di $V(x)$; si può però procedere, anche in vista di un implementazione numerica del metodo, ad una approssimazione di $V(x)$ con un andamento a gradino (*multistep potenzial approximation*).

Possiamo dividere l'intervallo x_1, x_2 in N parti e in ogni intervallo considerare la $V(x)$ costante; chiaramente maggiore sarà il numero di intervalli in cui dividiamo l'intervallo x_1, x_2 , migliore sarà l'approssimazione che facciamo.

In ogni intervallo utilizziamo una $V_{approx}(x) = V\left(\frac{x_i + x_{i+1}}{2}\right)$ per $x_i \leq x \leq x_{i+1}$.

Nella figura seguente vediamo come viene approssimato un generico potenziale con un potenziale a gradino. Le varie ascisse della griglia sono equispaziate in questo esempio, tuttavia si può pensare, più generalmente, che non lo siano, ossia possiamo scegliere una griglia più fitta dove serve maggiore definizione o dove le grandezze variano più velocemente, e una griglia più ampia dove le grandezze variano più lentamente.

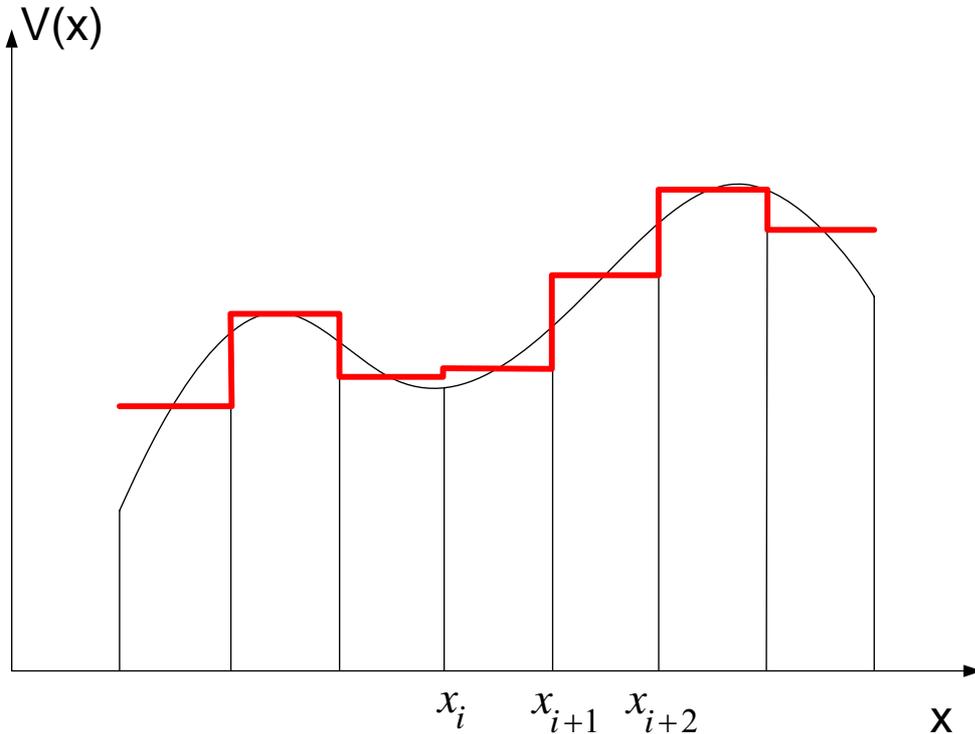


Figura1.6: generico andamento del potenziale e sua approssimazione con andamento a gradini

Tra x_i e x_{i+1} posso scrivere:

$$\Psi(x) = A_i e^{ik_i(x-x_i)} + B_i e^{-ik_i(x-x_i)} \quad (1.4.5)$$

mentre tra x_{i+1} e x_{i+2} scrivo:

$$\Psi(x) = A_{i+1} e^{ik_{i+1}(x-x_{i+1})} + B_{i+1} e^{-ik_{i+1}(x-x_{i+1})} \quad (1.4.6)$$

Imponendo la continuità di $\Psi(x)$ e $\frac{1}{m} \frac{d\Psi(x)}{dx}$ in x_{i+1} , trovo A_i, B_i in funzione di

A_{i+1}, B_{i+1} e così via per ogni i .

Utilizzando una tale approssimazione per il potenziale, le matrici di trasferimento “infinitesime” (ossia relative all’intervallo $dx = x_{i+1} - x_i$) possono essere facilmente moltiplicate tra loro per trovare la matrice totale. Infatti se $x_1 < x_2 < x_3$ allora

$M_{x_1 x_3} = M_{x_1 x_2} M_{x_2 x_3}$; il vantaggio di tale metodo consiste nel fatto che il potenziale con cui calcoliamo le singole matrici è facile da trattare e la matrice finale si ottiene come prodotto delle matrici “infinitesime”.

Una proprietà importante delle matrici di trasferimento è l’indipendenza dalle particolari condizioni al contorno; queste, infatti, dipendono solo dall’andamento di $V(x)$ per $x_1 \leq x \leq x_2$ e dalla E scelta. Se infatti fosse $E < V_2$ la $\Psi(x)$ nell’intervallo $c < x < d$ sarebbe:

$$\Psi(x) = A_2 e^{ik_2(x-x_2)} + B_2 e^{-ik_2(x-x_2)} \quad (1.4.7)$$

con $k_2 = \frac{\sqrt{2m(E-V_2)}}{\hbar}$ e come prima si può costruire una matrice di trasferimento che legghi i vari coefficienti.

Supponiamo di avere due soluzioni dell’equazione di Schrödinger Ψ' e Ψ'' associate a condizioni al contorno diverse; i coefficienti relativi a Ψ' siano A_1', A_2', B_1', B_2' mentre quelli relativi a Ψ'' siano $A_1'', A_2'', B_1'', B_2''$.

Allora possiamo scrivere:

$$\begin{pmatrix} A_1' \\ B_1' \end{pmatrix} = M'_{x_1 x_2} \begin{pmatrix} A_2' \\ B_2' \end{pmatrix} \quad (1.4.8)$$

$$\begin{pmatrix} A_1'' \\ B_1'' \end{pmatrix} = M''_{x_1 x_2} \begin{pmatrix} A_2'' \\ B_2'' \end{pmatrix} \quad (1.4.9)$$

Dalla linearità dell’equazione di Schrödinger segue che se c' e c'' sono due numeri complessi qualunque e se Ψ' e Ψ'' sono due soluzioni dell’equazione, allora anche

$\Psi = c'\Psi' + c''\Psi''$ è una soluzione. Chiamando A_1, A_2, B_1, B_2 i coefficienti di Ψ possiamo scrivere:

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = c' \begin{pmatrix} A_1' \\ B_1' \end{pmatrix} + c'' \begin{pmatrix} A_1'' \\ B_1'' \end{pmatrix} \quad (1.4.10)$$

$$\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} = c' \begin{pmatrix} A_2' \\ B_2' \end{pmatrix} + c'' \begin{pmatrix} A_2'' \\ B_2'' \end{pmatrix} \quad (1.4.11)$$

Inoltre possiamo scrivere per Ψ che :

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M_{x_1 x_2} \begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \quad (1.4.12)$$

Inserendo la 1.4.8 e 1.4.9 nella 1.4.10 otteniamo:

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = c' M'_{x_1 x_2} \begin{pmatrix} A_2' \\ B_2' \end{pmatrix} + c'' M''_{x_1 x_2} \begin{pmatrix} A_2'' \\ B_2'' \end{pmatrix} \quad (1.4.13)$$

mettendo la 1.4.11 nella 1.4.12 invece otteniamo:

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = c' M_{x_1 x_2} \begin{pmatrix} A_2' \\ B_2' \end{pmatrix} + c'' M_{x_1 x_2} \begin{pmatrix} A_2'' \\ B_2'' \end{pmatrix} \quad (1.4.14)$$

Perché la 1.4.13 e la 1.4.14 siano valide per ogni valore di c', c'' è necessario che sia:

$$M_{x_1 x_2} = M'_{x_1 x_2} = M''_{x_1 x_2} \quad (1.4.15)$$

1.4.1 Matrice di trasferimento elementare

Adesso vogliamo ricavare la matrice di trasferimento elementare; noi faremo il calcolo per un caso particolare di andamento del potenziale, ma l'espressione finale è la

stessa per ogni tipo di andamento; dovremo avere l'accortezza di sostituire ai vari k_i le giuste espressioni a seconda del potenziale e dell'energia E .

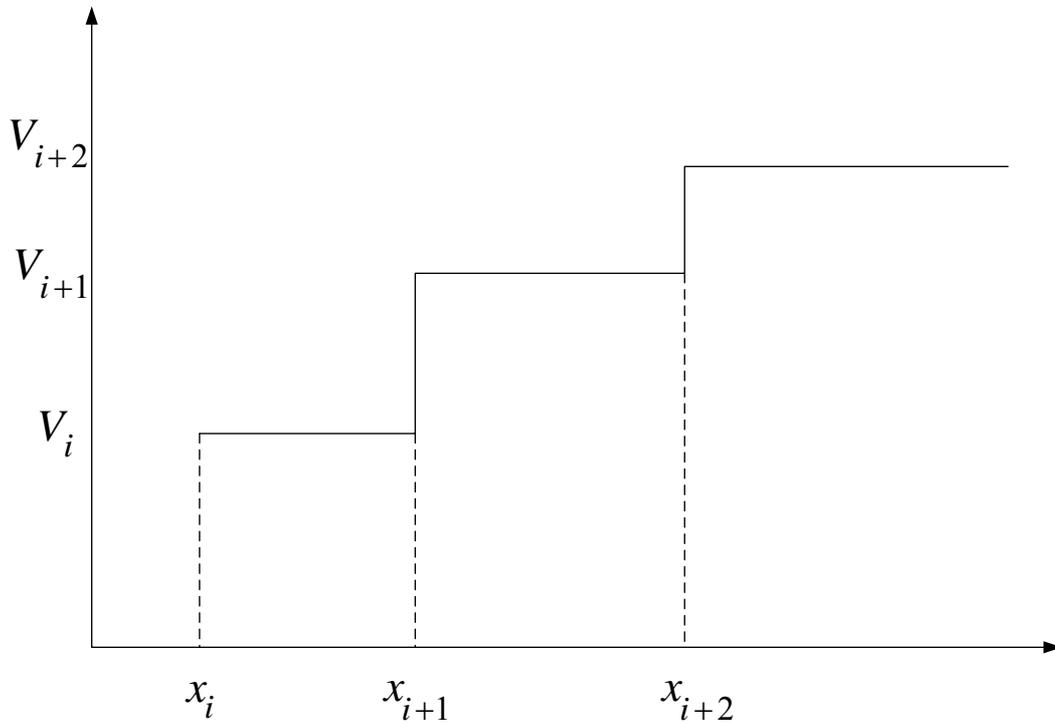


Figura 1.7: andamento del potenziale per il calcolo della matrice di trasferimento elementare.

Prendiamo il potenziale di figura 1.7 e calcoliamo la matrice di trasferimento elementare; il potenziale di figura sarà l'approssimazione di un potenziale qualsiasi.

V(x)

La soluzione dell'equazione di Schrödinger è data da:

$$\Psi(x) = \begin{cases} A_i e^{ik_i(x-x_i)} + B_i e^{-ik_i(x-x_i)} & x_i \leq x < x_{i+1} \\ A_{i+1} e^{ik_{i+1}(x-x_{i+1})} + B_{i+1} e^{-ik_{i+1}(x-x_{i+1})} & x_{i+1} \leq x < x_{i+2} \end{cases} \quad (1.4.16)$$

dove $k_i = \frac{\sqrt{2m_i(E-V_i)}}{\hbar}$; con m_i abbiamo indicato la massa efficace dell'elettrone tra x_i e

x_{i+1} .

Per ricavare la matrice di trasferimento bisogna imporre le condizioni di continuità per Ψ

$$\text{e } \frac{1}{m_i} \frac{d\Psi}{dx} \text{ in } x = x_{i+1}.$$

Si ottiene in questo modo:

$$\begin{pmatrix} A_i \\ B_i \end{pmatrix} = M(x_{i+1}) \begin{pmatrix} A_{i+1} \\ B_{i+1} \end{pmatrix} \quad (1.4.17)$$

dove con $M(x_{i+1})$ abbiamo indicato la matrice di trasferimento per l'interfaccia in x_{i+1} :

imponiamo quindi le due condizioni di continuità:

$$A_i e^{ik_i(x_{i+1}-x_i)} + B_i e^{-ik_i(x_{i+1}-x_i)} = A_{i+1} + B_{i+1} \quad (1.4.18)$$

$$\frac{1}{m_i} \left(A_i ik_i e^{ik_i(x_{i+1}-x_i)} - B_i ik_i e^{-ik_i(x_{i+1}-x_i)} \right) = \frac{1}{m_{i+1}} (A_{i+1} ik_{i+1} - B_{i+1} ik_{i+1}) \quad (1.4.19)$$

Dalla seconda equazione possiamo scrivere:

$$\left(A_i e^{ik_i(x_{i+1}-x_i)} - B_i e^{-ik_i(x_{i+1}-x_i)} \right) = \frac{1}{ik_i} \frac{m_i}{m_{i+1}} (A_{i+1} ik_{i+1} - B_{i+1} ik_{i+1}) \quad (1.4.20)$$

Sommando la 1.4.18 e la 1.4.20 otteniamo:

$$2A_i e^{ik_i(x_{i+1}-x_i)} = A_{i+1} \left(1 + \frac{k_{i+1} m_i}{k_i m_{i+1}} \right) + B_{i+1} \left(1 - \frac{k_{i+1} m_i}{k_i m_{i+1}} \right) \quad (1.4.21)$$

Sottraendo, invece, la 1.4.20 dalla 1.4.18 otteniamo:

$$2B_i e^{-ik_i(x_{i+1}-x_i)} = A_{i+1} \left(1 - \frac{k_{i+1} m_i}{k_i m_{i+1}} \right) + B_{i+1} \left(1 + \frac{k_{i+1} m_i}{k_i m_{i+1}} \right) \quad (1.4.22)$$

Otteniamo infine:

$$A_i = \frac{1}{2} e^{-ik_i(x_{i+1}-x_i)} \left[A_{i+1} \left(1 + \frac{k_{i+1} m_i}{k_i m_{i+1}} \right) + B_{i+1} \left(1 - \frac{k_{i+1} m_i}{k_i m_{i+1}} \right) \right] \quad (1.4.23)$$

$$B_i = \frac{1}{2} e^{ik_i(x_{i+1}-x_i)} \left[A_{i+1} \left(1 - \frac{k_{i+1}m_i}{k_i m_{i+1}} \right) + B_{i+1} \left(1 + \frac{k_{i+1}m_i}{k_i m_{i+1}} \right) \right] \quad (1.4.24)$$

Infine troviamo la matrice di trasferimento elementare:

$$M(x_{i+1}) = \frac{1}{2} \begin{bmatrix} \left(1 + \frac{k_{i+1}m_i}{k_i m_{i+1}} \right) e^{-ik_i(x_{i+1}-x_i)} & \left(1 - \frac{k_{i+1}m_i}{k_i m_{i+1}} \right) e^{-ik_i(x_{i+1}-x_i)} \\ \left(1 - \frac{k_{i+1}m_i}{k_i m_{i+1}} \right) e^{ik_i(x_{i+1}-x_i)} & \left(1 + \frac{k_{i+1}m_i}{k_i m_{i+1}} \right) e^{ik_i(x_{i+1}-x_i)} \end{bmatrix} \quad (1.4.25)$$

1.4.2 Matrice di trasferimento per un potenziale localizzato

Diremo localizzato un potenziale $V(x)$ che sia una funzione reale ed a un sol valore di x e caratterizzata da :

$$V(x) = \begin{cases} V_1 & x < x_1 \\ V_2 & x < x_2 \end{cases} \quad (1.4.26)$$

con $E > V_1, V_2$ e $-\infty < x_1 \leq x_2 < \infty$; un esempio è mostrato in figura 1.8. La soluzione dell'equazione di Schrödinger è data da:

$$\Psi(x) = \begin{cases} A_1 e^{ik_1(x-x_1)} + B_1 e^{-ik_1(x-x_1)} & x \leq x_1 \\ A_2 e^{ik_2(x-x_2)} + B_2 e^{-ik_2(x-x_2)} & x \geq x_2 \end{cases} \quad (1.4.27)$$

con $k_1 = \frac{\sqrt{2m(E-V_1)}}{\hbar}$, $k_2 = \frac{\sqrt{2m(E-V_2)}}{\hbar}$ e dove abbiamo supposto, per semplicità, che

le masse efficaci dell'elettrone per $x \leq x_1$ e per $x \geq x_2$ siano uguali.

Si possono fissare due coefficienti a piacere, in base a considerazioni sulla natura del problema, visto che gli altri due sono fissati dalla matrice di trasferimento.

Supponiamo di considerare il caso di un elettrone proveniente da sinistra; fissiamo quindi

$A_1 = 1, B_2 = 0$ visto che a destra della barriera non c'è onda riflessa ma solo un'onda

trasmessa; B_1 è ovviamente il coefficiente dell'onda riflessa mentre A_2 sarà il coefficiente dell'onda trasmessa verso destra dopo la barriera. Indichiamo questi due coefficienti rispettivamente con r e t ; possiamo riscrivere la funzione d'onda come:

$$\Psi(x) = \begin{cases} e^{ik_1(x-x_1)} + re^{-ik_1(x-x_1)} & x \leq x_1 \\ te^{ik_2(x-x_2)} & x \geq x_2 \end{cases} \quad (1.4.28)$$

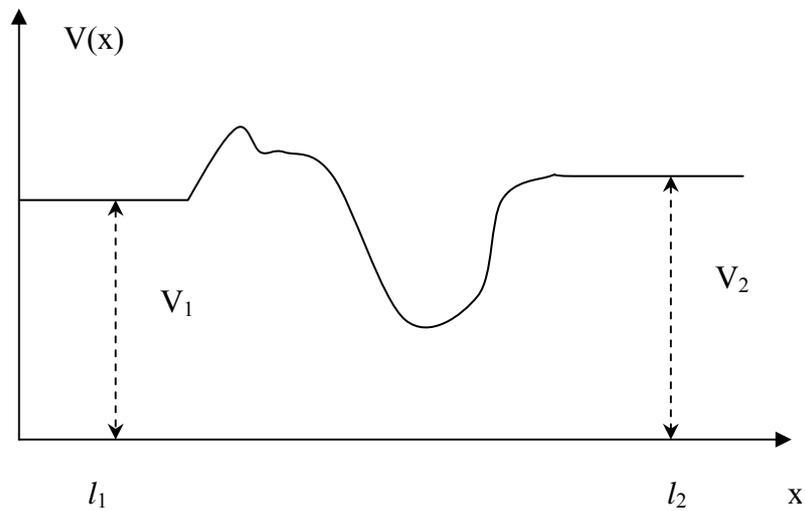


Figura 1.8: esempio di potenziale localizzato.

Possiamo scrivere la matrice di trasferimento in funzione di r e t

$$\begin{pmatrix} 1 \\ r \end{pmatrix} = M_{x_1, x_2} \begin{pmatrix} t \\ 0 \end{pmatrix} \quad (1.4.29)$$

Scriviamo adesso la soluzione per un elettrone proveniente da destra:

$$\Psi(x) = \begin{cases} t_d e^{-ik_1(x-x_1)} & x \leq x_1 \\ r_d e^{ik_2(x-x_2)} + e^{-ik_2(x-x_2)} & x \geq x_2 \end{cases} \quad (1.4.30)$$

e la corrispondente matrice di trasferimento

$$\begin{pmatrix} 0 \\ t_d \end{pmatrix} = M_{x_1 x_2} \begin{pmatrix} r_d \\ 1 \end{pmatrix} \quad (1.4.31)$$

Dalla 1.4.22 e dalla 1.4.24 si ricavano i termini della matrice di trasferimento che, ricordiamo, dipende solo dalla forma del potenziale ed è quindi la stessa per i due casi. Indicheremo con t_s e r_s i coefficienti della soluzione per un elettrone proveniente da sinistra.

$$M_{x_1 x_2} = \begin{bmatrix} \frac{1}{t_s} & -\frac{r_d}{t_s} \\ \frac{r_s}{t_s} & t_d - \frac{r_s r_d}{t_s} \end{bmatrix} \quad (1.4.32)$$

Per ottenere un'espressione migliore osserviamo che poiché l'equazione di Schrödinger è a coefficienti reali, se Ψ è soluzione allora lo è anche la sua complessa coniugata Ψ^* .

Adesso scriviamo le due espressioni per le matrici di trasferimento:

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M_{x_1 x_2} \begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \quad (1.4.33)$$

$$\begin{pmatrix} B_1^* \\ A_1^* \end{pmatrix} = M_{x_1 x_2} \begin{pmatrix} B_2^* \\ A_2^* \end{pmatrix} \quad (1.4.34)$$

Osserviamo che l'operazione di coniugazione comporta l'inversione della direzione del moto delle onde piane e quindi l'inversione dei coefficienti A e B. Dalle ultime due equazioni, che devono essere soddisfatte contemporaneamente, ricaviamo che:

$$\begin{aligned} \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \Rightarrow \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \begin{pmatrix} A_2 x_{11} + B_2 x_{12} \\ A_2 x_{21} + B_2 x_{22} \end{pmatrix} \\ \begin{pmatrix} B_1^* \\ A_1^* \end{pmatrix} &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{pmatrix} B_2^* \\ A_2^* \end{pmatrix} \Rightarrow \begin{pmatrix} B_1^* \\ A_1^* \end{pmatrix} = \begin{pmatrix} B_2^* x_{11} + A_2^* x_{12} \\ B_2^* x_{21} + A_2^* x_{22} \end{pmatrix} \end{aligned} \quad (1.4.35)$$

Questo sistema di quattro equazioni in quattro incognite può essere risolto con i metodi tradizionali, ma per ricavare la forma generale della $M_{x_1x_2}$ si può procedere nel modo seguente. Riscriviamo il sistema nel modo seguente:

$$\begin{pmatrix} A_1 \\ B_1^* \end{pmatrix} = \begin{bmatrix} A_2 & B_2 \\ B_2^* & A_2^* \end{bmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \quad (1.4.36)$$

$$\begin{pmatrix} B_1 \\ A_1^* \end{pmatrix} = \begin{bmatrix} A_2 & B_2 \\ B_2^* & A_2^* \end{bmatrix} \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} \quad (1.4.37)$$

Da questa si deduce che:

$$\begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = \begin{bmatrix} A_2 & B_2 \\ B_2^* & A_2^* \end{bmatrix}^{-1} \begin{pmatrix} A_1 \\ B_1^* \end{pmatrix} \quad (1.4.38)$$

$$\begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} = \begin{bmatrix} A_2 & B_2 \\ B_2^* & A_2^* \end{bmatrix}^{-1} \begin{pmatrix} B_1 \\ A_1^* \end{pmatrix} \quad (1.4.39)$$

La prima equazione fornisce gli elementi della prima riga di $M_{x_1x_2}$ che possiamo chiamare genericamente a e b ; si può allora notare che la seconda riga, fornita dalla seconda equazione è il prodotto della stessa matrice per una colonna che, rispetto alla colonna della prima riga, ha gli elementi invertiti e coniugati. La generica forma della seconda riga sarà quindi b^* e a^* ; la matrice di trasferimento ha quindi la seguente forma generica:

$$M_{x_1x_2} = \begin{bmatrix} a & b \\ b^* & a^* \end{bmatrix} \quad (1.4.40)$$

In seguito a questa considerazione possiamo riscrivere la matrice di trasferimento per un potenziale localizzato, che avevamo ricavato precedentemente, come:

$$M_{x_1 x_2} = \begin{bmatrix} \frac{1}{t} & \frac{r^*}{t^*} \\ r & \frac{1}{t^*} \end{bmatrix} \quad (1.4.41)$$

Dalla forma generale della matrice di trasferimento e dalla sua forma particolare in funzione di r_s, t_s, r_d, t_d si può ricavare la seguente espressione, che si ricava uguagliando i termini x_{12} e x_{22} :

$$r_d = -\frac{r^* t}{t^*} \quad \text{e} \quad t_d = t \frac{1 - |r|^2}{|t|^2} \quad (1.4.42)$$

Per ricavare un'espressione più significativa per t_d imponiamo la costanza lungo x della densità di corrente di probabilità in condizioni stazionarie.

$$J = \frac{\hbar}{m} \text{Im} \left\{ \Psi^* \frac{d\Psi}{dk} \right\} \quad (1.4.43)$$

La densità di corrente di probabilità dovrà essere uguale, in particolare, in $x = x_1$ e $x = x_2$.

Sostituiamo nell'espressione di J il valore di Ψ dato dalla 1.4.20 e uguagliamo $J(x_1)$ e $J(x_2)$ ottenendo:

$$k_1 (|A_1|^2 - |B_1|^2) = k_2 (|A_2|^2 - |B_2|^2) \quad (1.4.44)$$

Imponiamo questa condizione e scriviamo $M_{x_1 x_2}$ nella forma 1.4.33, risulta allora:

$$a^2 - b^2 = \frac{k_2}{k_1} \quad (1.4.45)$$

Utilizzando le 1.4.35, 1.4.37 e 1.4.38 si ottengono, dopo semplici passaggi algebrici, le seguenti relazioni:

$$\begin{aligned} r_d &= -r_s \\ t_d &= \frac{k_2}{k_1} t_s \\ 1 - |r_s|^2 &= \frac{k_2}{k_1} |t_s|^2 \end{aligned} \tag{1.4.46}$$

Una barriera di potenziale è un caso particolare di potenziale localizzato; applicheremo adesso la teoria delle matrici di trasferimento ad una barriera di potenziale, come esempio di applicazione.

Come sappiamo la probabilità di trasmissione attraverso la barriera di potenziale è data da $T = |t|^2$ e quella di riflessione è $R = |r|^2$.

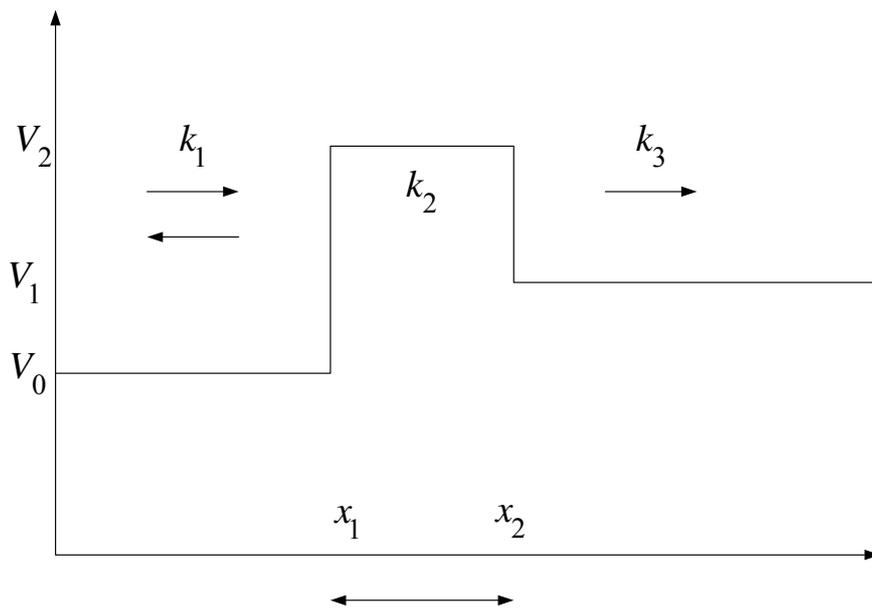


Figura 1.9: esempio di barriera di potenziale rettangolare.

Calcoliamo la matrice di trasferimento per la barriera di potenziale di figura 1.9; supporremo per semplicità che la massa efficace dell'elettrone sia dovunque la stessa.

Indicando con M_{bar} la matrice di trasferimento della barriera, possiamo osservare che la matrice totale è scrivibile come prodotto di due matrici di trasferimento elementari:

$$M_{bar} = M_1 M_2 \quad (1.4.47)$$

Difatti potremo scrivere che $\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M_1 \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}$ e $\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} = M_2 \begin{pmatrix} A_3 \\ B_3 \end{pmatrix}$; dunque otteniamo

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M_1 M_2 \begin{pmatrix} A_3 \\ B_3 \end{pmatrix} \quad (1.4.48)$$

dove M_1, M_2 sono matrici di trasferimento elementari del tipo 1.4.25.

Cerchiamo di trovare il coefficiente di trasmissione della barriera per un elettrone proveniente da sinistra; in tal caso nella regione 3 (quella per $x \geq x_2$) c'è solo un onda trasmessa, quindi $B_3 = 0$. Per calcolare il coefficiente di trasmissione dovremo calcolare il rapporto A_3/A_1 che sarà dato da:

$$A_1 = [M_1 M_2]_{11} A_3 \quad (1.4.49)$$

Sostituendo le espressioni generali 1.4.25 nella equazione sopra, possiamo valutare il termine 1,1 della matrice prodotto, che risulta, supponendo $x_0 = 0$:

$$[M_1 M_2]_{11} = \frac{1}{4} \left\{ \left(1 + \frac{k_2 m_1}{k_1 m_2} \right) e^{-ik_1 x_1} \cdot \left(1 + \frac{k_3 m_2}{k_2 m_3} \right) e^{-ik_2 (x_2 - x_1)} + \left(1 - \frac{k_2 m_1}{k_1 m_2} \right) e^{-ik_1 x_1} \cdot \left(1 - \frac{k_3 m_2}{k_2 m_3} \right) e^{ik_2 (x_2 - x_1)} \right\}$$

Fissiamo $\alpha_2 = ik_2$ e $w = x_2 - x_1$; con semplici passaggi algebrici si dimostra inoltre che:

$\left(1 - i \frac{\alpha_2 m_1}{k_1 m_2}\right) \left(1 + i \frac{k_3 m_2}{\alpha_2 m_3}\right)$ è il complesso coniugato di $\left(1 + i \frac{\alpha_2 m_1}{k_1 m_2}\right) \left(1 - i \frac{k_3 m_2}{\alpha_2 m_3}\right)$, tale termine

lo indicheremo con c . Allora

$$A_1 = \frac{1}{4} e^{-ik_1 x_1} \left[c \cdot e^{-\alpha_2 w} + c^* e^{\alpha_2 w} \right] A_3 \quad (1.4.50)$$

Trascurando adesso il termine $e^{-\alpha_2 w}$ rispetto a $e^{\alpha_2 w}$, ossia considerando barriere fortemente attenuanti, troviamo, scrivendo $c = |c| e^{-i\varphi}$:

$$A_3 = \frac{4e^{ik_1 x_1}}{c^* e^{\alpha_2 w}} A_1 = \frac{4e^{ik_1 x_1}}{|c| e^{i\varphi} e^{\alpha_2 w}} A_1 \quad (1.4.51)$$

Il rapporto tra la densità di probabilità di corrente trasmessa $j_3 = \frac{\hbar k_3}{m_3} |A_3|^2$ e incidente

$j_1 = \frac{\hbar k_1}{m_1} |A_1|^2$, ossia il coefficiente di trasmissione T , vale

$$T = \frac{j_3}{j_1} = \frac{k_3 m_1}{k_1 m_3} \frac{16}{\left(1 + \frac{k_3 m_1}{k_1 m_3}\right)^2 + \left(\frac{k_3 m_2}{\alpha_2 m_3} - \frac{\alpha_2 m_1}{k_1 m_2}\right)^2} e^{-2\alpha_2 w} \quad (1.4.52)$$

Se supponiamo le masse efficaci tutte uguali, ossia se $m_1 = m_2 = m_3$ si trova il coefficiente di trasmissione “classico” per una barriera di potenziale, come è possibile vedere, ad esempio su Kane [2].

Nel programma di simulazione viene calcolata la matrice di trasferimento di una barriera tramite la routine *tranmat*; per prima cosa viene generata la matrice di trasferimento elementare 1.4.25 e poi vengono fatte le moltiplicazioni in successione delle matrici, come descritto precedentemente.

1.5 Corrente tunnel

Adesso abbiamo tutti gli strumenti per calcolare la corrente tunnel che attraversa una barriera di potenziale costituita, nei nostri casi, da uno strato di ossido tra due lati di semiconduttore, ossia da una struttura MOS. Un elettrone può attraversare tale barriera per effetto tunnel, un effetto che non ha analogo classico e che può essere spiegato solo nell'ambito della meccanica quantistica. Se lo spessore degli ossidi di gate si riduce troppo, la corrente tunnel può diventare elevata.

Ci sono anche altri possibili meccanismi per i quali gli elettroni possono superare la barriera; uno di questi è l'emissione termoionica dovuta al fatto che gli elettroni non sono in equilibrio con il reticolo (si parla di elettroni caldi) ed hanno energia sufficiente per superare la barriera di potenziale.

Una volta risolta l'equazione di Poisson conosciamo l'andamento della struttura a bande e da questa possiamo partire per calcolare la corrente tunnel.

Supportremo nel resto del paragrafo di considerare un nMOS ossia considereremo un substrato di tipo p. Se la tensione di gate è maggiore della tensione di soglia, la regione di Silicio a contatto con l'ossido è invertita e il confinamento degli elettroni comporta la quantizzazione dei livelli energetici longitudinali. Usando una visione classica del problema, possiamo pensare che gli elettroni rimbazzino avanti e indietro nella buca di potenziale; ogni volta che incontrano la barriera tunnel la possono attraversare con una certa probabilità. Il fatto che gli elettroni abbiano una probabilità finita di superare la barriera di potenziale fa sì che essi si trovino in stati detti quasi – legati; infatti gli elettroni permangono su questi stati per un tempo medio finito dopodiché superano la barriera. I

modelli quantomeccanici computano i tempi di vita medi degli stati quasi – legati per calcolare la probabilità di tunneling.

Indicheremo con $\tau_{rt} = \tau_{rt}(E_l)$ il round trip time ossia il tempo che un elettrone di energia E_l impiega a percorrere avanti e indietro la buca di potenziale all'interfaccia ossido – bulk.

La buca di potenziale in condizioni di tensioni di gate fortemente positiva, può essere approssimata con una buca di forma triangolare: per tale tipo di buca esistono soluzioni analitiche ottenute col metodo WKB che fanno uso delle funzioni di Airy; il nostro programma di simulazione però non utilizza l'approssimazione WKB ma fa un calcolo esatto della corrente tunnel.

Per barriere fortemente attenuanti il coefficiente di trasmissione della barriera è piccolo, come si vede dall'equazione 1.4.46, a causa dell'esponenziale negativo; questo fa sì che il tunneling possa essere studiato con la teoria delle perturbazioni al primo ordine, come propose, per primo, Bardeen (per una spiegazione del metodo, vedi, ad esempio, [2]). Invece di introdurre stati che sono soluzione di un hamiltoniana approssimata, egli introdusse soluzioni approssimate dell'hamiltoniana esatta.

Se è applicabile la teoria perturbativa al primo ordine, la probabilità di attraversamento della barriera da parte di un elettrone può essere ricavata a partire dalla regola d'oro di Fermi che ci fornisce la probabilità per unità di tempo che un elettrone passi da uno stato iniziale, chiamiamolo Ψ_{s_1} , ad uno finale, chiamiamolo Ψ_{s_2} .

Consideriamo a tal proposito un generico diagramma a bande di una struttura MOS alla quale è stata applicata una tensione positiva, come nella figura 1.10.

Nella transizione dallo stato Ψ_{s_1} allo stato Ψ_{s_2} si conservano lo spin, l'energia longitudinale e la componente trasversale del vettore d'onda; la probabilità per unità di

tempo che un elettrone con energia E_1 nella regione 1, passi in uno stato vuoto, con la stessa energia, nella regione 2 è:

$$P_{1 \rightarrow 2} = \frac{2\pi}{\hbar} |M_{12}|^2 \delta(E - E_1) \quad (1.5.1)$$

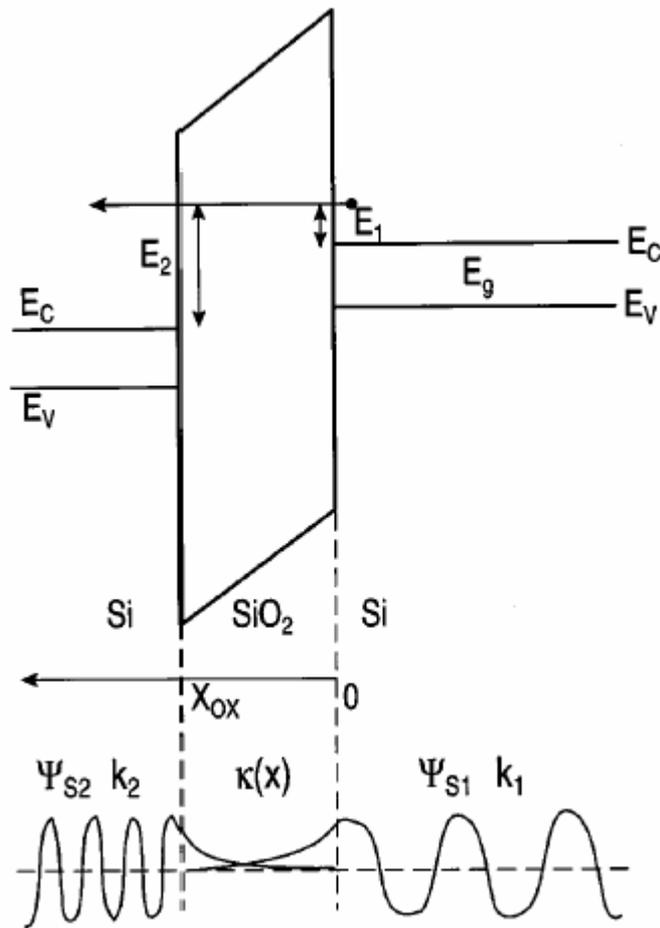


Figura 1.10: diagramma a bande e funzioni d'onda nello stato iniziale (destra) e finale (sinistra) di un processo di transizione per effetto tunnel. k_1 e k_2 ($k_2 > k_1$) sono i numeri d'onda reali fuori dalla barriera, mentre $\kappa(x)$ e il numero d'onda immaginario all'interno della barriera ($0 < x < X_{ox}$). La figura è tratta da un'importante articolo di Sah sulla corrente tunnel diretta [3].

Il tasso di transizione totale si trova integrando su tutti gli stati vuoti della regione 2 e su tutti gli stati occupati della regione 1, ossia:

$$P_{1 \rightarrow 2} = \frac{2\pi}{\hbar} \iint |M_{12}|^2 \rho_1 \rho_2 f_1 (1 - f_2) d\Psi_{s1} d\Psi_{s2} \quad (1.5.2)$$

dove M_{12} è la matrice della transizione, ρ_1 (ρ_2) è la densità di stati nella regione 1 (2) e f_1 (f_2) è la funzione di Fermi – Dirac nella regione 1 (2) che fornisce la probabilità di occupazione degli stati.

Ogni stato, in una trattazione monodimensionale in cui non si considerano i momenti angolari, è caratterizzato dalla sua componente longitudinale di energia, dalla componente trasversa del vettore d'onda e dallo spin; la transizione attraverso la barriera conserva queste tre grandezze. L'elemento di matrice può essere espresso come [4]:

$$|\langle \Psi_{s2} | M | \Psi_{s1} \rangle|^2 = |M_{12}|^2 = \hbar^2 v_d(E_l) v_s(E_l) T(E_l) \delta(E_2 - E_1) \quad (1.5.3)$$

dove $v_d(E_l)$ e $v_s(E_l)$ sono le cosiddette frequenze di fuga per lo stato Ψ_{s1} e Ψ_{s2} rispettivamente e $T(E_l)$ è la probabilità di trasmissione e della barriera

Dalle 1.5.2 e 1.5.3, definendo $\rho_{ls}(E_l)$ e $\rho_{ld}(E_l)$ come le densità longitudinali a sinistra e a destra della barriera rispettivamente, e $\rho_T(\underline{k}_T)$ la densità di stati trasversali, abbiamo:

$$J_{\leftarrow} = 2\pi q \hbar \int dE_l \int d\underline{k}_T v_d(E_l) v_s(E_l) T(E_l) \rho_{ls}(E_l) \rho_{ld}(E_l) \rho_T(\underline{k}_T) f_1 (1 - f_2) \quad (1.5.4)$$

dove abbiamo già eseguito l'integrazione sugli spin visto che di tale grandezza ne abbiamo tenuto di conto nella densità di stati. Osserviamo che $\rho_T(\underline{k}_T) d\underline{k}_T$ è il numero di stati disponibili con componente trasversale del vettore d'onda compresa tra \underline{k}_T e $\underline{k}_T + d\underline{k}_T$ ed è uguale a $\rho_T(E_T) dE_T$ (numero di stati compresi tra E_T e $E_T + dE_T$); portando fuori la

densità di stati trasversali, che è costante, se espressa in funzione dell'energia, e pari a

$\frac{\sqrt{m_y m_z}}{\pi \hbar^2}$, otteniamo:

$$J_{\leftarrow} = 2\pi q \hbar \frac{\sqrt{m_y m_z}}{\pi \hbar^2} \int dE_l \int dE_T v_d(E_l) v_s(E_l) T(E_l) \rho_{ls}(E_l) \rho_{ld}(E_l) f_1(1-f_2) \quad (1.5.5)$$

Integriamo dapprima su E_T e troviamo:

$$J_{\leftarrow} = 2\pi q \hbar \frac{\sqrt{m_y m_z}}{\pi \hbar^2} \int T(E_l) \frac{\rho_{ls}(E_l)}{\tau_{r_s}(E_l)} \frac{\rho_{ld}(E_l)}{\tau_{r_d}(E_l)} dE_l \cdot \int_0^{+\infty} \frac{1}{1 + \exp\left(\frac{E_T}{k_B T} + \frac{E_l - E_{Fd}}{k_B T}\right)} \cdot \frac{\exp\left(\frac{E_T}{k_B T} + \frac{E_l - E_{Fs}}{k_B T}\right)}{1 + \exp\left(\frac{E_T}{k_B T} + \frac{E_l - E_{Fs}}{k_B T}\right)} dE_T \quad (1.5.6)$$

dove abbiamo tenuto conto che $v(E_l) = \frac{1}{\tau_{r_t}(E_l)}$; possiamo riscrivere quindi:

$$J_{\leftarrow} = 2\pi q \hbar \frac{\sqrt{m_y m_z}}{\pi \hbar^2} \int T(E_l) \frac{\rho_{ls}(E_l)}{\tau_{r_s}(E_l)} \frac{\rho_{ld}(E_l)}{\tau_{r_d}(E_l)} dE_l \cdot \int_0^{+\infty} \frac{A}{1 + \exp\left(\frac{E_T}{k_B T} + \frac{E_l - E_{Fd}}{k_B T}\right)} + \frac{B}{1 + \exp\left(\frac{E_T}{k_B T} + \frac{E_l - E_{Fs}}{k_B T}\right)} dE_T \quad (1.5.7)$$

$$\text{con } A = -B = \frac{\exp\left(\frac{E_l - E_{Fs}}{k_B T}\right)}{\exp\left(\frac{E_l - E_{Fs}}{k_B T}\right) - \exp\left(\frac{E_l - E_{Fd}}{k_B T}\right)} = \frac{\exp\left(\frac{E_{Fd} - E_l}{k_B T}\right)}{\exp\left(\frac{E_{Fd} - E_l}{k_B T}\right) - \exp\left(\frac{E_{Fs} - E_l}{k_B T}\right)}$$

Per semplificare la scrittura delle equazioni fissiamo i seguenti parametri:

$$a_d = \exp\left(\frac{E_{Fd} - E_l}{k_B T}\right) \quad (1.5.8)$$

$$a_s = \exp\left(\frac{E_{Fs} - E_l}{k_B T}\right) \quad (1.5.9)$$

Per il calcolo della densità di corrente tunnel dobbiamo ricordarci che ci sono sei minimi equivalenti in banda di conduzione per i quali abbiamo:

$$\sqrt{m_y m_z} = m_t \quad \text{per i due minimi longitudinali}$$

$$\sqrt{m_y m_z} = \sqrt{m_l m_t} \quad \text{per i quattro minimi trasversali}$$

Siamo adesso in grado di calcolare la densità di corrente tunnel complessiva da destra a sinistra, data da:

$$\begin{aligned} J_{\leftarrow} = & 4\pi q \hbar k_B T \frac{m_t}{\pi \hbar^2} \int \frac{T_l(E_l)}{\tau_{dl}(E_l)} \rho_{ld_l}(E_l) \frac{\rho_{ls_l}(E_l)}{\tau_{sl}(E_l)} \frac{a_d}{a_s - a_d} \{\ln[1 + a_s] - \ln[1 + a_d]\} dE_l \\ & + 4 \frac{\sqrt{m_l m_t}}{\pi \hbar^2} k_B T 2\pi q \hbar \int \frac{T_l(E_l)}{\tau_{dt}(E_l)} \rho_{ld_l}(E_l) \frac{\rho_{ls_t}(E_l)}{\tau_{st}(E_l)} \frac{a_d}{a_s - a_d} \{\ln[1 + a_s] - \ln[1 + a_d]\} dE_l \end{aligned} \quad (1.5.10)$$

dove per semplicità di notazione si è indicato con T_l (T_t) e con τ_{il} (τ_{it}), con $i = s, d$, rispettivamente il coefficiente di trasmissione da destra a sinistra ed il round trip time per i minimi longitudinali (trasversali).

In maniera simile si ricava la densità di corrente da sinistra a destra e infine si può scrivere la densità di corrente totale data da:

$$J = J_{\leftarrow} - J_{\rightarrow} \quad (1.5.11)$$

Possiamo notare che se i livelli di Fermi a destra e a sinistra sono uguali è nulla la densità di corrente totale ma non le singole componenti da destra a sinistra e da sinistra a destra.

Il calcolo dei coefficienti di trasmissione viene fatto usando il metodo delle matrici di trasferimento che è stato esposto nel paragrafo precedente.

Nell'espressione della densità di corrente tunnel scritta sopra dobbiamo ancora esplicitare la densità longitudinale di stati; nel caso di quantizzazione delle energia longitudinali permesse all'elettrone, dobbiamo approfondire meglio il concetto di densità di stati e, in conseguenza di questo, dell'altro parametro ancora da definire con precisione: il round trip time.

1.5.1 Densità di stati per unità di energia longitudinale e round trip time

Studiamo dapprima quello che succede in assenza di quantizzazione; per fare questo consideriamo una buca di potenziale monodimensionale come mostrato nella figura seguente.

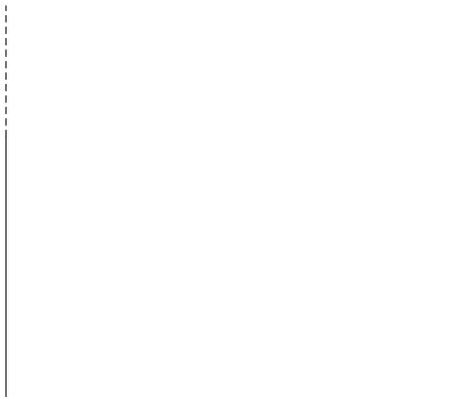


Figura 1.11: buca di potenziale monodimensionale all'interno della quale si risolve l'equazione di Schrödinger.

Risolvendo l'equazione di Schrödinger con le condizioni al contorno di Dirichlet, si trovano i valori permessi per il vettore d'onda, che sono $k = n \frac{\pi}{L}$ con n intero positivo. Ciò vuol dire che ogni $\frac{\pi}{L}$ c'è uno stato disponibile, quindi scriviamo

$$\rho(K) dk = \frac{L}{\pi} dk \quad (1.5.12)$$

Sappiamo inoltre che $E = \frac{\hbar^2 k^2}{2m}$ quindi $dE = \frac{\hbar^2 k}{m} dk$ da cui infine

$$\rho(E) dE = \frac{L}{\pi} \cdot \frac{m}{\hbar^2 k} dE \quad (1.5.13)$$

Adesso calcoliamo il round trip time, ossia il tempo che un elettrone impiega per percorrere avanti e indietro la buca di lunghezza L ; possiamo scrivere che:

$$\tau = 2 \frac{L}{v} \quad (1.5.14)$$

dove v è la velocità dell'elettrone che sappiamo essere pari a

$$v = \frac{\hbar k}{m} \quad (1.5.15)$$

A questo punto otteniamo la relazione che lega il round trip time alla densità longitudinale di stati, nel caso di assenza di quantizzazione di stati :

$$\frac{\rho(E_l)}{\tau_r(E_l)} = \frac{1}{2\pi\hbar} \quad (1.5.16)$$

che riscritta fornisce anche la relazione:

$$2\pi\hbar \frac{\rho(E_l)}{\tau_r(E_l)} = 1 \quad (1.5.17)$$

Si potrebbe erroneamente pensare che la relazione trovata dipenda dalle condizioni al contorno che abbiamo imposto per risolvere l'equazione di Schrödinger; in particolare noi abbiamo imposto condizioni al contorno di Dirichelet ossia abbiamo ipotizzato che l'energia dell'elettrone all'interno della buca sia molto più piccola dell'altezza della buca stessa così da considerare una buca di altezza infinita. In questo modo le condizioni al

contorno impongono che la funzione d'onda si annulli agli estremi della zona, quindi in $x=0$ e $x=L$. Se adesso imponiamo, per esempio, condizioni periodiche al contorno, ossia supponiamo che la funzione d'onda sia la stessa agli estremi della buca ($\Psi(x=0) = \Psi(x=L)$), otteniamo una quantizzazione dei vettori d'onda diversa da prima, in particolare otteniamo:

$$k = n \frac{2\pi}{L} \text{ con } n \text{ intero positivo.}$$

D'altronde il round trip time è dimezzato perché imporre queste ultime condizioni equivale a pensare che l'elettrone arrivi in $x=L$ e che immediatamente dopo si ritrovi in $x=0$;

quindi il round trip time è $\tau = \frac{L}{v}$ e il rapporto $\frac{\rho}{\tau}$ continua a valere $\frac{1}{2\pi\hbar}$.

Come abbiamo detto più volte, nelle regioni in cui gli elettroni sono confinati quantisticamente, come, per esempio, l'interfaccia ossido-silicio, le energie longitudinali permesse sono quantizzate; i valori di tali energie si trovano risolvendo l'equazione di Schrödinger e sono gli autovalori associati alle varie funzioni d'onda. In realtà tali valori non sono esatti perché a seguito di urti, elastici o anelatici, gli elettroni possono passare da uno stato ad un altro e gli stati occupati sono distribuiti simmetricamente rispetto all'autovalore E_i .

L'ampiezza della distribuzione di energia attorno ad un livello E_i è legata al tempo di permanenza sullo stato dalla relazione di Heisenberg:

$$\Gamma = \frac{\hbar}{\tau_{relax}} \tag{1.5.18}$$

La distribuzione viene presa, in accordo con [5] come una somma di Lorentziane e assume quindi la forma:

$$\rho_l(E_l) = \frac{1}{\pi} \sum_i \frac{\frac{\Gamma}{2}}{(E_l - E_i)^2 + \left(\frac{\Gamma}{2}\right)^2} \quad (1.5.19)$$

Una distribuzione di questo tipo è mostrata nella figura seguente.

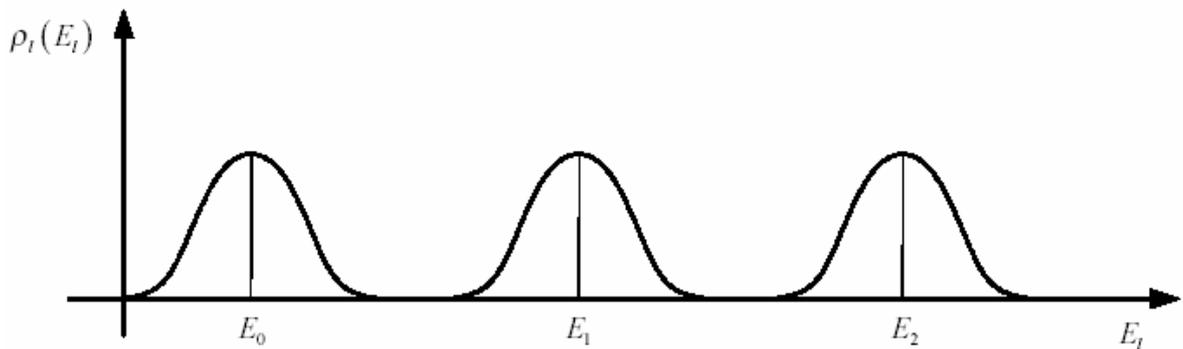


Figura 1.12: densità longitudinale di energia a lorentziane.

Una volta calcolata la densità longitudinale di stati, possiamo passare a calcolare il round trip time. Per fare questo consideriamo un generico andamento del potenziale $V(x)$ come in figura.



Figura 1.13: andamento (6.a) e deformazione (6.b) di un generico potenziale.

A seguito della deformazione effettuata il potenziale $V(x)$ è costante nella zona A, quindi in tale zona possiamo risolvere l'equazione di Schrödinger; la soluzione $\Psi(x)$ può essere scritta come un'onda piana stazionaria:

$$\Psi(x) = Ae^{ikx} + Be^{-ikx} \quad \text{con } A = B^* = Ce^{i\gamma} \quad C, \gamma \in \mathbb{R}.$$

Ovviamente i coefficienti A e B dipendono dall'andamento di $\Psi(x)$ nelle altre zone; k è

il vettore d'onda associato all'elettrone e vale $k = \frac{\sqrt{2m(E - E_c)}}{\hbar}$. Una tale

approssimazione sarà tanto migliore quanto più la zona A sarà piccola e, al limite per A che diventa infinitesima, si recupera l'andamento continuo.

Applicheremo tale tecnica alla nostra struttura a bande all'interfaccia ossido-silicio, in modo da calcolare il round trip time.

Il tempo di vita medio per il tunneling τ_n può essere definito, per ciascun livello energetico n , come:

$$\frac{1}{\tau_n} = T_n \nu_n \tag{1.5.20}$$

dove T_n è il coefficiente di trasmissione per l'n-esimo livello e ν_n è la frequenza di fuga per il livello stesso; al posto della frequenza di fuga si può utilizzare il flusso di probabilità

incidente. Il round trip time è l'inverso della frequenza di fuga, ossia $\tau_n = \frac{1}{\nu}$ (valida per

ciascun livello energetico).

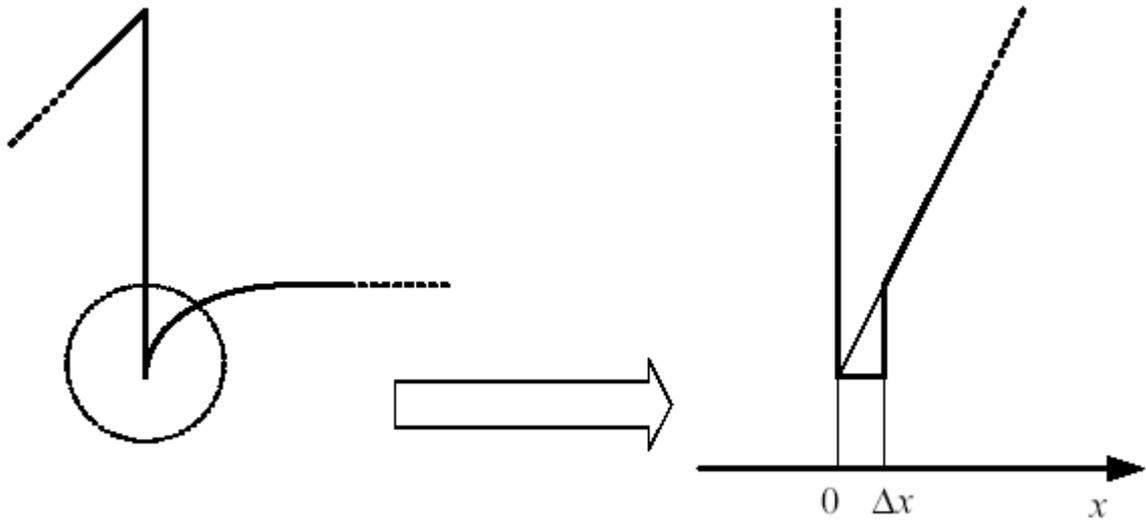


Figura 1.14: andamento dell'energia potenziale per gli elettroni e sua deformazione in un tratto infinitesimo in prossimità dell'interfaccia silicio-ossido.

Possiamo quindi scrivere che:

$$\frac{1}{\tau_n} = T_n \frac{1}{\tau_{rt}} \Rightarrow \tau_{rt} = T_n \tau_n \quad (1.5.21)$$

Se ci riferiamo al flusso di probabilità (manterremo per esso la stessa indicazione J_n), il tempo di vita medio può essere riscritto come:

$$\frac{1}{\tau_n} = T_n J_n \Rightarrow J_n = \frac{1}{T_n \tau_n} = \frac{1}{\tau_{rt}} \quad (1.5.22)$$

Il tempo di vita medio dipende, ovviamente, dalla forma della buca all'interfaccia ossido-silicio e dalle caratteristiche della barriera tunnel (altezza della barriera, campo elettrico nell'ossido e spessore dell'ossido stesso). Come abbiamo detto si parla di approssimazione di stati quasi-legati e tutto questo comporta grandi semplificazioni nella soluzione del

problema. Innanzitutto risolviamo l'equazione di Schrödinger stazionaria e non quella dipendente dal tempo e inoltre la descrizione del tunneling è ridotta alla conoscenza di un solo parametro, il tempo di vita medio, per ciascuna sottobanda.

Torniamo adesso al nostro potenziale approssimato e calcoliamo la densità di corrente di probabilità lungo x , associata all'onda incidente sulla barriera; la forma per un'onda piana del flusso è:

$$J_x = A^2 \frac{\hbar k}{m} \quad (1.5.23)$$

dove A^2 è l'ampiezza dell'onda e $\frac{\hbar k}{m}$ la velocità. Per determinare τ_r è quindi sufficiente determinare A (oppure C) e per fare questo andiamo a valutare la funzione d'onda in $x = 0$ e in $x = \Delta x$.

Riscriviamo la funzione d'onda:

$$\Psi(x) = Ce^{i(kx+\gamma)} + Ce^{-i(kx+\gamma)} = 2C \cos(kx + \gamma) \quad (1.5.24)$$

e andiamo a calcolarla nei punti stabiliti:

$$\begin{aligned} \Psi(0) &= 2C \cos \gamma \\ \Psi(\Delta x) &= 2C \cos(k\Delta x + \gamma) = 2C \cos(k\Delta x) \cos \gamma - 2C \sin(k\Delta x) \sin \gamma \end{aligned} \quad (1.5.25)$$

Facciamo il rapporto tra la prima e la seconda equazione ed otteniamo:

$$\begin{aligned} \frac{\Psi(\Delta x)}{\Psi(0)} &= \cos(k\Delta x) - \sin(k\Delta x) \operatorname{tg}(\gamma) \\ \cos(k\Delta x) - \frac{\Psi(\Delta x)}{\Psi(0)} &= \sin(k\Delta x) \operatorname{tg}(\gamma) \\ \operatorname{tg}(\gamma) &= \frac{\cos(k\Delta x) \Psi(0) - \Psi(\Delta x)}{\sin(k\Delta x) \Psi(0)} \end{aligned} \quad (1.5.26)$$

Dalla trigonometria sappiamo che vale la seguente relazione $\cos^2 \gamma = \frac{1}{1 + tg^2 \gamma}$, quindi possiamo scrivere ancora, dalla prima equazione a sistema, che:

$$C^2 = \frac{\Psi^2(0)}{4 \cos^2 \gamma} = \frac{\Psi^2(0)}{4} (1 + tg^2 \gamma) = \frac{\Psi^2(0)}{4} + \frac{1}{4} \left[\frac{\cos(k\Delta x) \Psi(0) - \Psi(\Delta x)}{\sin(k\Delta x)} \right]^2 \quad (1.5.27)$$

e quindi (visto che $A^2 = C^2$)

$$\tau_r = \left\{ \frac{\hbar k}{4m} \left\{ \Psi^2(0) + \left[\frac{\cos(k\Delta x) \Psi(0) - \Psi(\Delta x)}{\sin(k\Delta x)} \right]^2 \right\} \right\}^{-1} \quad (1.5.28)$$

Tutto quello che abbiamo fatto finora vale per un solo livello energetico; se indichiamo con E_i i livelli energetici corrispondenti agli autovalori dell'hamiltoniano, e tenendo conto della densità di stati longitudinali, l'espressione finale per il round trip time è:

$$\tau_r = \frac{\sum_i \left\{ \frac{\hbar k}{4m} \left\{ \Psi^2(0) + \left[\frac{\cos(k\Delta x) \Psi(0) - \Psi(\Delta x)}{\sin(k\Delta x)} \right]^2 \right\} \right\}^{-1} \frac{1}{\pi} \frac{\frac{\Gamma}{2}}{(E - E_i)^2 + \left(\frac{\Gamma}{2}\right)^2}}{\sum_i \frac{1}{\pi} \frac{\frac{\Gamma}{2}}{(E - E_i)^2 + \left(\frac{\Gamma}{2}\right)^2}} \quad (1.5.29)$$

Il round trip time totale è quindi dato dalla media pesata dei round trip time dei vari livelli energetici, con i pesi dati dalle lorentziane corrispondenti.

1.6 Approssimazione quasi classica (metodo WKB)

Riportiamo in questo paragrafo una breve descrizione del metodo Wentzel-Kramers-Brillouin, detto metodo WKB, utile per la soluzione approssimata dell'equazione di Schrödinger e che permette di trovare le espressioni per le correnti tunnel di tipo diretto

e di tipo Fowler-Nordheim. L'equazione di Schrödinger si trasforma, al limite per $\hbar \rightarrow 0$, con continuità nell'equazione di Hamilton-Jacobi della meccanica classica. È questo un risultato che si trova in meccanica quantistica e che in questa breve esposizione verrà considerato noto e quindi non sarà dimostrato; sappiamo anche che in virtù del principio di corrispondenza i risultati ottenuti con l'applicazione della condizione di quantizzazione di Bohr-Sommerfeld, nel caso di numero quantici grandi, si trasformano con continuità in risultati classici.

Scriviamo l'equazione unidimensionale, per semplicità, di Schrödinger:

$$\frac{d^2\Psi}{dx^2} + \frac{2m}{\hbar^2}(E - U(x))\Psi = 0 \quad (1.6.1)$$

Indichiamo con l'apice le derivate rispetto a x , quindi:

$$\hbar^2\Psi'' + 2m(E - U(x))\Psi = 0 \quad (1.6.2)$$

Cercheremo la soluzione nella forma $\Psi = e^{i\frac{S(x)}{\hbar}}$ dove $S(x)$ ha le dimensioni di una azione e indica un legame con l'equazione classica di Hamilton-Jacobi. Dopo la sostituzione otteniamo:

$$i\hbar S'' - S'^2 + 2m(E - U(x)) = 0 \quad (1.6.3)$$

Tale equazione è esatta ma ne cerchiamo una approssimazione sotto forma di serie rispetto ad \hbar :

$$S(x) = S_0(x) + \hbar S_1(x) + \hbar^2 S_2(x) + \dots \quad (1.6.4)$$

La serie è mal convergente quindi ne cerchiamo l'approssimazione fino al primo ordine ossia assumiamo:

$$S(x) \cong S_0(x) + \hbar S_1(x) \quad (1.6.5)$$

Sostituendo tale $S(x)$ nella $\Psi = e^{i\frac{S(x)}{\hbar}}$ e poi mettendola nell'equazione di Schrödinger abbiamo:

$$2m(E - U(x)) - S_0'^2 + \hbar(iS_0'' - 2S_0'S_1') = 0 \quad (1.6.6)$$

Affinché tale equazione sia verificata devono annullarsi separatamente i termini di ordine \hbar^0 e \hbar^1 :

$$2m(E - U(x)) - S_0'^2 = 0 \quad (1.6.7)$$

$$iS_0'' - 2S_0'S_1' = 0 \quad (1.6.8)$$

Vediamo la prima equazione che corrisponde all'approssimazione di ordine zero:

$$S_0' = \pm\sqrt{2m(E - U(x))} \quad (1.6.9)$$

Ma $\sqrt{2m(E - U(x))}$ è l'impulso classico p , quindi

$$S_0' = \pm\sqrt{2m(E - U(x))} = \pm p \quad (1.6.10)$$

$$S_0 = \pm \int_{x_0}^x p dx \quad (1.6.11)$$

dove x_0 è la coordinata di un punto fisso della retta lungo la quale avviene il moto;

l'approssimazione nulla è una soluzione puramente classica.

Dalla seconda equazione si trova $S_1(x)$:

$$S_1' = \frac{i}{2} \frac{S_0''}{S_0'} = \frac{i}{2} \left[\ln(S_0') \right]' \quad (1.6.12)$$

Integrando troviamo

$$S_1 = \frac{i}{2} \ln S_0' = \frac{i}{2} \ln p \quad (1.6.13)$$

a meno della costante di integrazione.

Quindi $S(x)$ al primo ordine vale:

$$S(x) = \pm \int p dx + i\hbar \ln \sqrt{p} \quad (1.6.14)$$

e sostituendola in $\Psi = e^{\frac{iS(x)}{\hbar}}$ troviamo

$$\Psi = \frac{C_1}{\sqrt{p}} e^{\frac{i}{\hbar} \int p dx} + \frac{C_2}{\sqrt{p}} e^{-\frac{i}{\hbar} \int p dx} \quad (1.6.15)$$

Le soluzioni di questo tipo si chiamano soluzioni WKB e sono molto usate per risolvere semplicemente alcuni problemi come l'attraversamento di una barriera triangolare o trapezoidale

1.7 Tunneling diretto e di Fowler Nordheim

L'effetto tunnel può essere studiato per via analitica con l'approssimazione WKB per avere delle espressioni relativamente semplificate, adatte allo studio qualitativo del tunneling.

Si parla di tunneling di tipo Fowler-Nordheim se l'effetto tunnel avviene attraverso una barriera triangolare; l'elettrone passa direttamente dalla banda di conduzione del silicio alla banda di conduzione nell'ossido.

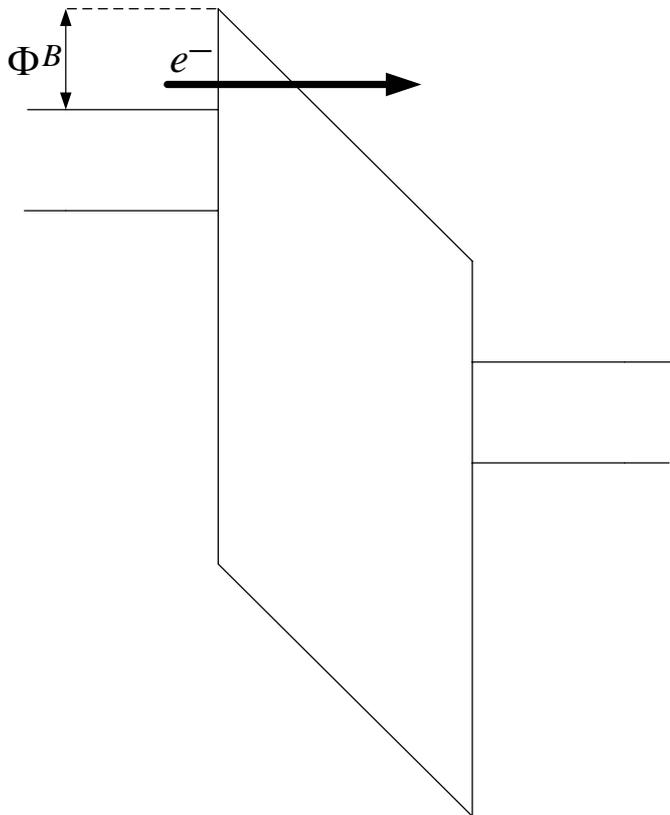


Figura 1.15: struttura a bande di un nMOS a cui è stata applicata una forte tensione di gate negativa, siamo quindi nel caso di tunnel Fowler-Nordheim; si nota l'elettrone che passa direttamente in banda di conduzione dell'ossido di silicio.

L'espressione che si ricava per la densità di corrente tunnel è la seguente:

$$J_{FN} = \frac{q^3}{8\pi h \Phi_B} \frac{m}{m^*} E_{OX}^2 e^{-\frac{4\sqrt{2m^*}\Phi_B^2}{3hqE_{OX}}} \quad (1.7.1)$$

dove q è la carica dell'elettrone, Φ_B è l'altezza della barriera all'interfaccia tra poly e ossido, m^* è la massa efficace degli elettroni nell'ossido e E_{OX} è il campo elettrico all'interno dell'ossido stesso.

Si

Per quanto riguarda l'altezza della barriera di potenziale vista dagli elettroni all'interfaccia silicio-ossido c'è da tenere di conto della forza immagine che abbassa il valore di tale barriera; tale diminuzione è data da:

$$\Delta\Phi = \left(\frac{q^3 \cdot E}{4\pi\epsilon_0\epsilon_r} \right)^{1/2} \quad (1.7.2)$$

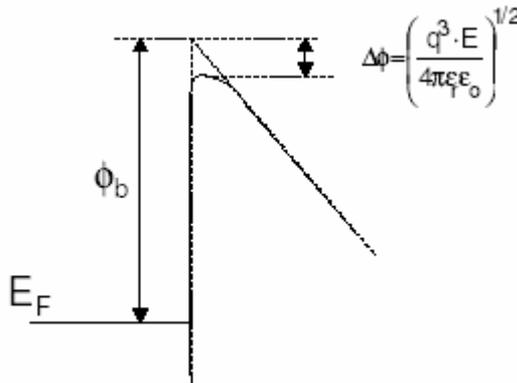


Figura 1.16: *abbassamento della barriera dovuto all'effetto della forza immagine.*

Abbiamo indicato con E il campo elettrico nell'ossido e con ϵ_r la costante dielettrica relativa dello stesso.

Per spiegare brevemente come nasce la cosiddetta forza immagine supponiamo di considerare un elettrone in una regione di spazio compresa tra due elettrodi paralleli (il nostro caso effettivo sarà un elettrone nell'ossido tra il gate e il substrato); l'elettrone polarizza entrambi gli elettrodi. Come risultato di ciò gli elettrodi influenzano il potenziale dell'elettrone che viaggia nell'area compresa tra i due elettrodi. L'effetto del potenziale

immagine è di arrotondare gli angoli e di ridurre l'altezza della barriera; tale potenziale può essere ottenuto usando il cosiddetto metodo delle immagini, ben noto dalla elettrostatica.

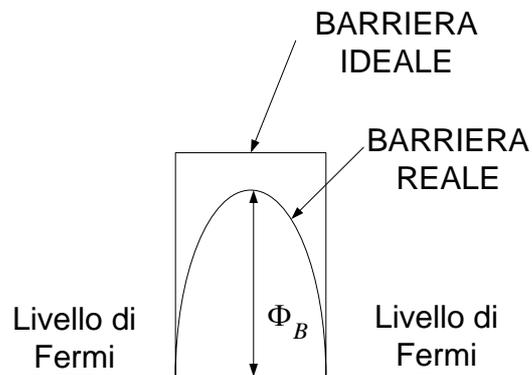


Figura 1.17: andamento della barriera di potenziale reale e rispettivo andamento reale modificato dalla forza immagine.

Il tunneling diretto è invece il tunneling attraverso una barriera trapezoidale tra due bande di conduzione del silicio. In questo caso non c'è quindi il trasporto nella banda di conduzione dell'ossido e il tunnel avviene attraverso il gap proibito dell'ossido di silicio.

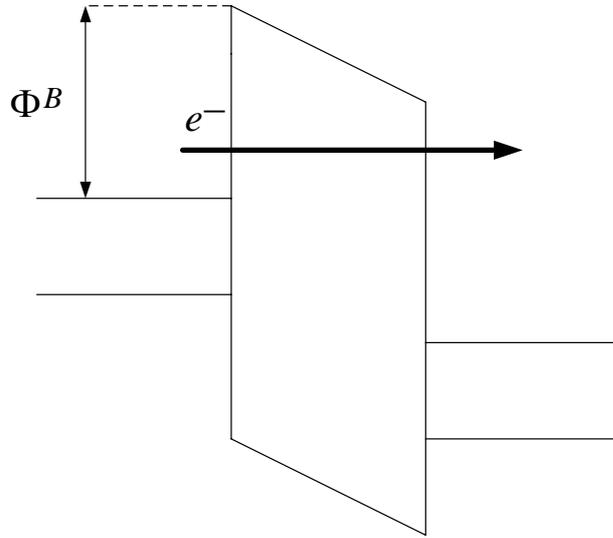


Figura 1.18: *diagramma a bande della struttura MOS nel caso di tunnel di tipo diretto.*

L'espressione della densità di corrente per il tunnel diretto si può trovare, come per il tunnel Fowler-Nordheim, risolvendo l'equazione di Schrödinger con l'approssimazione WKB e fornisce il seguente risultato:

$$J_{DT} = \frac{AE_{OX}^2}{\left[1 - \left(\frac{\phi_B + qV_{OX}}{\phi_B}\right)^{\frac{1}{2}}\right]^2} \exp\left[-\frac{B}{E_{OX}} \frac{\phi_B^{\frac{3}{2}} - (\phi_B - qV_{OX})^{\frac{3}{2}}}{\phi_B^{\frac{3}{2}}}\right] \quad (1.7.3)$$

dove $A = \frac{q^3}{8\pi h \Phi_B} \frac{m}{m^*}$ e $B = \frac{4\sqrt{2m^*} \Phi_B^{\frac{3}{2}}}{3hq}$ sono due coefficienti introdotti per comodità.

Per dare un esempio dell'incremento della corrente diretta, causata dalla diminuzione dello spessore dell'ossido, osserviamo che a 1.5 V la corrente incrementa di 5 ordini di grandezza per una diminuzione di 1.25 nm di spessore.

Le espressioni per la corrente tunnel di tipo diretto o di tipo Fowler-Nordheim sopra proposte, valgono solo per ossidi non troppo sottili; al di sotto dei 7 nm bisogna ricorrere

al metodo delle matrici di trasmissione che abbiamo esposto in precedenza e che viene usato dal nostro programma di simulazione. I risultati ottenuti con le simulazioni ci servono per determinare lo spessore dell'ossido e la massa efficace degli elettroni al suo interno (in genere tale grandezza viene usata come parametro di fitting); infatti confrontando i dati ottenuti sperimentalmente e i risultati delle simulazioni si possono estrarre i parametri sopra citati e anche altri come per esempio la costante dielettrica relativa da misure C-V. Per quanto riguarda l'estrapolazione dello spessore dell'ossido c'è da notare che con le tradizionali tecniche ellissometriche si determina lo spessore dell'ossido con un'incertezza di 0.5 nm ; la corrente tunnel ha una dipendenza forte dallo spessore dell'ossido e un errore su di esso (percentualmente rilevante visto che operiamo con ossidi di pochi nm) comporta un grande errore sulla corrente. In particolare un errore di 0.5 nm comporta un errore di due ordini di grandezza sul valore di corrente. Facciamo un'ulteriore osservazione sull'estrapolazione dei valori di spessore dell'ossido: l'accordo tra dati sperimentali e simulazioni va ricercato per alti valori di campo nell'ossido dove il tunneling è sicuramente di tipo Fowler-Nordheim; infatti per bassi campi possono intervenire altri meccanismi di conduzione come la SILC (acronimo di Stress Induced Leakage Current) che aumentano il valore della corrente anche in assenza di stress, a causa delle cosiddette trappole native.

CAPITOLO 2

Simulazione di strutture MOS con dielettrici ad alto k

2.1 Introduzione

Ci proponiamo adesso di spiegare il metodo seguito per la simulazione di condensatori MOS nei quali l'ossido di gate è sostituito da uno stack di dielettrici. La sostituzione del biossido di silicio come ossido di gate è iniziata, ancora a livello sperimentale, da alcuni anni, a causa delle grandi correnti di perdita che esso presenta per dispositivi al di sotto del nodo tecnologico dei 70 nm.

I materiali che possiamo pensare di sostituire al biossido di silicio devono possedere una serie di requisiti come la citata bassa corrente di perdita, l'alta costante dielettrica k , una bassa densità di trappole interfacciali D_{it} (simile quindi a quella dell' SiO_2 o dell' SiO_2 nitrato); devono inoltre presentare una barriera per la diffusione di droganti o impurezze, per ovvie questioni tecnologiche.

Anche la stabilità termica rappresenta ovviamente un requisito importante, ed inoltre tali ossidi devono essere stabili in contatto con Si, poly-Si e con gate metallici che dovranno essere impiegati nel prossimo futuro.

Noi analizzeremo due diversi tipi di approcci al problema, un approccio cosiddetto *high-k* ossia facente uso di ossidi ad alto k (come l'ossido di afnio - HfO_2) e un altro approccio denominato *medium-k* (che si basa su silicati di afnio). Per ossidi *high-k* intendiamo quelli con costante dielettrica superiore a 15, mentre con dielettrici *medium-k*, intenderemo quelli

con costante dielettrica intermedia tra quella dell'ossido di silicio (3.9) e quella dei dielettrici *high-k*. Al momento sembra che si possano ottenere migliori risultati con i silicati di afnio ma noi riporteremo anche gli studi fatti sull'ossido di afnio; questi ci faranno capire dove e perché i silicati sono migliori dell'ossido di afnio.

Le nostre simulazioni si sono basate su misure effettuate all'IMEC delle tre strutture MOS che riportiamo nelle figure seguenti:

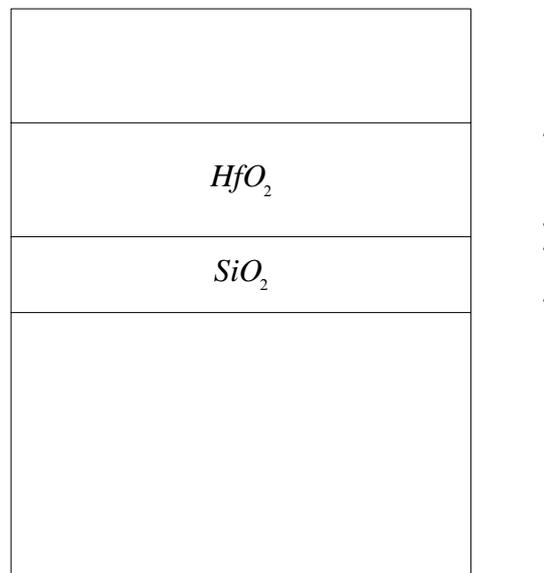


Figura 2.1: prima struttura utilizzata per le nostre simulazioni; si vede la presenza dello stack di dielettrici formato da 0.8 nm di SiO_2 e 4.5 nm di ossido di afnio.

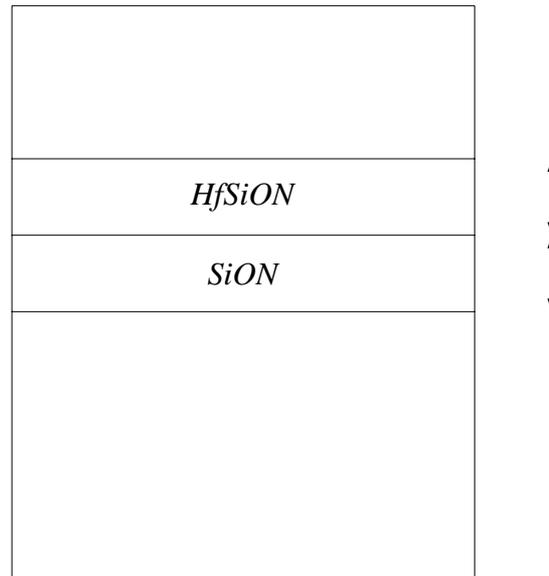


Figura 2.2: seconda struttura utilizzata per le simulazioni: questa struttura è composta da 1 nm di *SiON* e 1 nm di silicato di afnio nitrato, *HfSiON*.

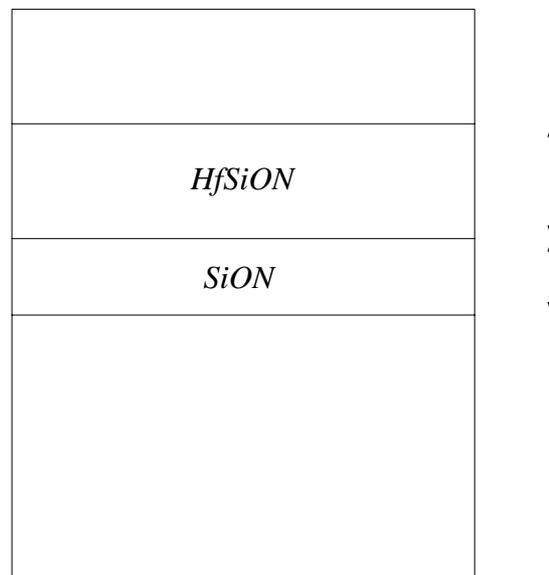


Figura 2.3: terza struttura utilizzata per le simulazioni; la differenza rispetto alla struttura precedente consiste nello spessore del silicato di afnio nitrato, in questo caso 2 nm.

Lo studio dei dispositivi è stato fatto su curve Capacità–Tensione e su curve Corrente–Tensione; dopo aver visto nel capitolo 1 come si ricava la corrente tunnel diamo una descrizione delle misure C-V e degli effetti che possono modificarle.

2.2 Misure di capacità

Il condensatore MOS è molto importante per la caratterizzazione elettrica dell'ossido; da questo si possono estrarre informazioni sulla costante dielettrica dell'ossido, sulla carica intrappolata nell'ossido, sulle trappole interfacciali e molto altro ancora. Il programma di simulazione risolve l'equazione di Poisson e fornisce l'andamento del potenziale nella struttura. Da questo si può risalire al campo elettrico nell'ossido, derivando il potenziale rispetto ad x . La routine *carica.f* calcola la carica applicando il teorema di Gauss ad un cilindro di area di base unitaria con base all'interfaccia SiO_2 /bulk.

La carica può infatti essere scritta come:

$$Q = -\frac{dV}{dx} \varepsilon_0 \varepsilon_r \cdot S \quad (2.2.1)$$

dove S è la superficie del condensatore MOS utilizzato (tutte le misure sono state fatte su condensatori MOS di area $70\mu m \times 70\mu m$), ε_0 è la permittività dielettrica del vuoto, ε_r la costante dielettrica relativa e V il potenziale nell'ossido.

2.3 Effetto delle trappole interfacciali sulla capacità

Gli stati interfacciali sono dovuti a difetti posizionati sull'interfaccia o immediatamente vicino ad essa; l'origine di questi stati va ricercata nei legami incompleti (ad esempio all'interfaccia $Si-SiO_2$ ci sono legami incompleti del tipo SiO_x con $x < 2$). Inoltre durante il funzionamento del dispositivo i difetti possono essere generati dalla rottura dei legami dovuta, ad esempio, all'iniezione di portatori o a radiazioni provenienti dall'esterno. La densità di stati può essere ridotta, in fase di produzione, con varie tecniche tra le quali menzioniamo l'annealing in ambiente ricco di idrogeno o di azoto, che va a saturare i legami incompleti o rotti.

Gli stati interfacciali si presentano come livelli energetici nel gap del Si e possono scambiare carica con il silicio stesso, sia come accettori sia come donatori. L'occupazione di questi stati dipende dal potenziale superficiale del silicio.

Gli stati interfacciali sono caratterizzati dalla densità di stati interfacciali, che rappresenta il numero di stati per unità di superficie e di energia: indicheremo tale densità con $D_{it}(E)$.

Questi stati hanno effetto sulla caratteristica C-V, sulla transconduttanza, sulla corrente di perdita e sul livello di rumore; lo studio delle trappole interfacciali avviene con diverse tecniche tra le quali citiamo le misure C-V, G-V, misure di rumore e la tecnica Charge Pumping che attualmente è quella maggiormente utilizzata. Infatti questa tecnica fornisce molte informazioni come il valor medio di D_{it} ma anche la distribuzione spaziale ed energetica delle trappole stesse.

Vediamo adesso l'effetto sulle caratteristiche C-V: la differenza tra le cariche intrappolate nel bulk e quelle interfacciali è che le prime portano, come noto, ad uno shift della V_T (perché sono cariche indipendenti da V) mentre le trappole interfacciali comportano una

minore pendenza della caratteristica nella fase di passaggio dall'accumulazione all'inversione a causa della dipendenza del riempimento degli stati dalla tensione di gate.

Gli stati di legame all'interfaccia tra due materiali sono anche chiamati trappole interfacciali per cui useremo la dizione stati interfacciali o trappole interfacciali equivalentemente.

Le misure di capacità rivelano la presenza di trappole interfacciali dallo "stiramento" delle curve misurate, rispetto a quelle ideali, lungo l'asse delle tensioni e da un termine di capacità aggiuntivo dovuto alle trappole interfacciali stesse. Stati interfacciali accettori sono carichi negativamente se occupati e neutri se vuoti, mentre stati interfacciali donatori sono neutri quando occupati e positivi se vuoti. Nello stesso dispositivo possono convivere entrambi i tipi di stati interfacciali; esistono metodi per determinare quale tipo è dominante e per ossidi cresciuti termicamente, ad esempio, si sa che sono predominanti gli stati donatori. Lo "stiramento" delle curve C-V dipende dal fatto che l'occupazione delle trappole interfacciali varia al variare della tensione applicata sul gate; per spiegarlo pensiamo a come cambia la curvatura delle bande per due strutture MOS, una con trappole interfacciali e una senza, quando la carica sul gate viene cambiata di δQ_G supponendo che entrambe raggiungano l'equilibrio dopo questa variazione.

Nel MOS senza trappole interfacciali, la neutralità di carica fa sì che venga richiamata, alla superficie del silicio, una carica δQ_S tale che $\delta Q_G + \delta Q_S = 0$; la curvatura delle bande cambia in modo da determinare questo bilancio.

Nel MOS con trappole interfacciali c'è un cambiamento della densità di carica interfacciale δQ_{it} quindi il bilancio soddisfa la relazione $\delta Q_G + \delta Q_S + \delta Q_{it} = 0$. Visto che in questo caso c'è il termine δQ_{it} , la variazione δQ_S richiesta sarà minore. Quindi la curvatura

delle bande in questo caso è minore visto che la variazione di carica richiesta è più piccola. Così per passare dall'accumulazione all'inversione è richiesta una maggiore variazione della carica di gate in presenza di trappole interfacciali.

Tra la carica di gate e la tensione di gate sussiste la relazione $dQ_G = C\delta V_G$, il che significa che con trappole interfacciali c'è bisogno anche di una maggiore variazione della tensione di gate per passare dall'accumulazione all'inversione, ossia la curva C-V è "distesa" lungo l'asse delle tensioni.

Come noto dalla trattazione elementare dei MOS, la carica nel silicio si trova applicando il teorema di Gauss ad un cilindro con base all'interfaccia tra silicio e ossido; si ottiene quindi:

$$C_{OX} (V_G - \Psi_s) = -Q_{it} (\Psi_s) - Q_s (\Psi_s) \quad (2.3.1)$$

dove C_{OX} è la capacità per unità di area dell'ossido, V_G è la tensione di gate, Ψ_s è il potenziale superficiale del silicio, Q_{it} è la carica interfacciale per unità di area e Q_s è la carica superficiale del silicio per unità di area.

Cambiando la tensione di gate di un dV_G si ottiene un corrispondente cambiamento del potenziale superficiale $d\Psi_s$ dato da:

$$C_{OX} dV_G = [C_{OX} + C_{it} (\Psi_s) + C_s (\Psi_s)] d\Psi_s \quad (2.3.2)$$

dove

$$C_{it} (\Psi_s) \equiv -\frac{dQ_{it}}{d\Psi_s} \quad (2.3.3)$$

e

$$C_s (\Psi_s) = -\frac{dQ_s}{d\Psi_s} \quad (2.3.4)$$

sono le capacità per unità di area delle trappole interfacciali e del silicio rispettivamente. Ancora una volta vediamo come una variazione dV_G comporti una variazione $d\Psi_s$ minore in caso di presenza di trappole interfacciali, facendo così distendere le curve C-V.

2.4 Capacità delle trappole interfacciali

La capacità a bassa frequenza del condensatore MOS è definita come

$$C_{LF} \equiv \frac{dQ_T}{dV_G} \quad (2.4.1)$$

dove Q_T è la carica totale per unità di area, ed è data da

$$Q_T = -(Q_s + Q_{it}) \quad (2.4.2)$$

Dalla 2.2.2 e dalla 2.3.1 otteniamo

$$\begin{aligned} C_{LF} &= \frac{dQ_T}{d\Psi_s} \frac{d\Psi_s}{dV_G} \\ &= (C_s + C_{it}) \frac{C_{OX}}{C_{OX} + C_s + C_{it}} \end{aligned} \quad (2.4.3)$$

il che significa che in aggiunta al distendimento delle curve, si ha anche un aumento della C_{LF} misurata di C_{it} ; se volessimo fare un circuito equivalente potremmo pensarlo così fatto:

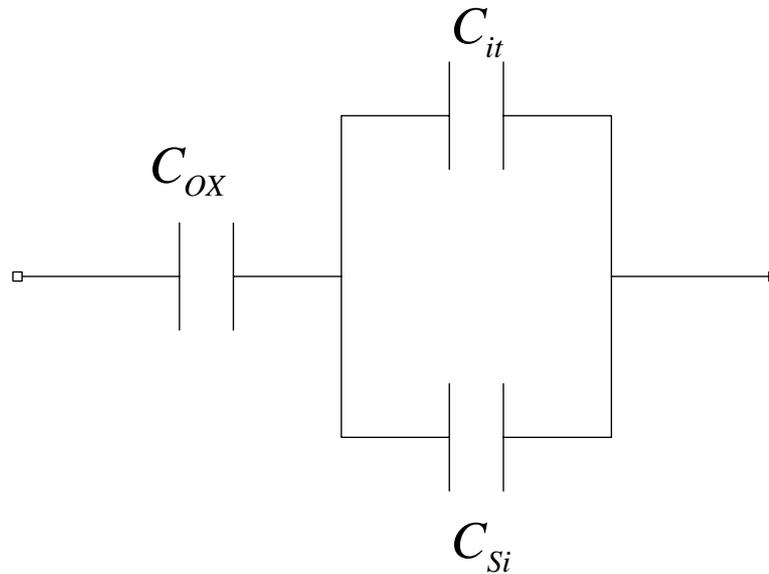


Figura 2.4: circuito equivalente dello stack dielettrico; la capacità totale è data dalla serie di C_{ox} con il parallelo di C_{si} e C_{it} .

La capacità 2.4.3 è misurata alle basse frequenze; se abbiamo una piccola variazione $dQ_T(f)$ in risposta ad una $dV_G(f)$, non tutte le trappole interfacciali risponderanno a questa variazione. Solo quelle che possono catturare elettroni nel periodo del segnale variabile $V_G(f)$ risponderanno. La C_{it} è dunque una funzione della frequenza e per $f \rightarrow \infty$ nessuna trappola risponderà più, cosicché $C_{it} \rightarrow 0$. Per alte frequenze la capacità misurata, per un dato piegamento delle bande, è identica a quella di un capacitore MOS che differisce solo nella densità di trappole interfacciali; il distendimento è l'unico effetto dato dalle trappole interfacciali per misure ad alta frequenza. Per misure a bassa frequenza c'è da considerare un termine aggiuntivo alla capacità dato dalla capacità delle trappole interfacciali. Da queste considerazioni si possono estrarre tre metodi per la determinazione della densità di stati interfacciali:

- Misura dello stiramento della C-V ad alta frequenza rispetto alla curva teorica

- Misura della capacità degli stati interfacciali per sottrazione della capacità calcolata da una curva C-V a bassa frequenza
- Confronto delle curve C-V a bassa e alta frequenza senza bisogno di curve teoriche

Non discuteremo i tre metodi, chi volesse approfondire l'argomento può consultare i libri di Sah [6] e Nicollian [7] riportati in bibliografia, ma dobbiamo cercare di capire come sono legate la capacità degli stati interfacciali e la densità di stati interfacciali, che abbiamo indicato, rispettivamente, con C_{it} e D_{it}

2.5 Capacità degli stati interfacciali

Sia $D_{it}(E_S)$ la probabilità per unità di area che una trappola interfacciale sia presente, con energia compresa tra E_S e $E_S + dE_S$; $D_{it}(E_S)$ è comunemente chiamata densità di trappole interfacciali per unità di area e di energia.

E_S denota una energia nel gap, misurata a partire dal livello di Fermi intrinseco alla superficie del silicio; nella figura seguente si mostra il diagramma a bande e le relative grandezze che utilizzeremo:

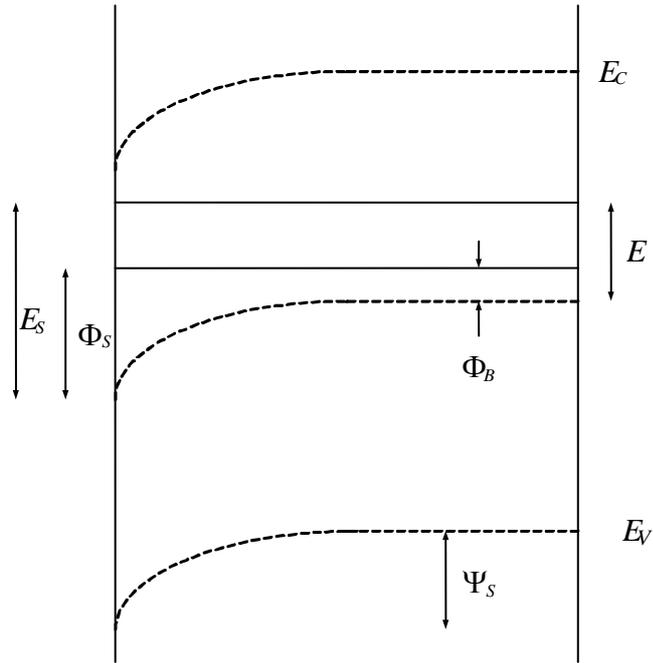


Figura 2.5: diagramma a bande nel substrato; il piano di sinistra definisce l'interfaccia Si/SiO_2 , mentre il piano sulla destra delimita la zona del substrato oltre la quale non si hanno più significative variazioni delle grandezze.

Il livello di Fermi è a distanza Φ_B dal livello intrinseco nel bulk; è indicato anche un livello di energia arbitrario a distanza E dal livello intrinseco, a cui corrisponde una certa energia E_S all'interfaccia. Ci si può facilmente convincere che $E_S = E + \Psi_S$. Per $E_S = \Phi_S$ il livello di energia interfacciale coincide con il livello di Fermi.

Adesso scriviamo la densità di stati interfacciali in funzione di $E + \Psi_S$ per calcolarci la carica $Q_{it}(\Psi_S)$; quest'ultima dipende dalla natura delle trappole interfacciali ossia dipende dal fatto che le trappole siano donatori o accettori.

Per le trappole donatrici, otteniamo

$$Q_{it}^d(\Psi_S) = q \int_{E_V - \Psi_S}^{E_C - \Psi_S} [1 - f(E - \Phi_B)] D_{it}^d(E + \Psi_S) dE \quad (2.5.1)$$

dove, come si vede dalla figura, Φ_B è la differenza tra il livello di Fermi e il livello di Fermi intrinseco nel bulk mentre $E_C - \Psi_S$ e $E_V - \Psi_S$ indicano i limiti del gap all'interfaccia.

Per trappole di tipo accettore, che ricordiamo essere neutre se occupate e negative se vuote, vale

$$Q_{it}^a(\Psi_S) = q \int_{E_V - \Psi_S}^{E_C - \Psi_S} f(E - \Phi_B) D_{it}^a(E + \Psi_S) dE \quad (2.5.2)$$

La carica interfacciale intrappolata netta per unità di area $Q_{it}(\Psi_S)$ è quindi

$$\begin{aligned} Q_{it}^a(\Psi_S) &= Q_{it}^d(\Psi_S) - Q_{it}^a(\Psi_S) = \\ &= q \int_{E_V - \Psi_S}^{E_C - \Psi_S} D_{it}^d(E + \Psi_S) dE - q \int_{E_V - \Psi_S}^{E_C - \Psi_S} f(E - \Phi_B) [D_{it}^a(E + \Psi_S) + D_{it}^d(E + \Psi_S)] dE \end{aligned} \quad (2.5.3)$$

Cambiando variabile di integrazione da E a $E_S = E + \Psi_S$ troviamo

$$\begin{aligned} Q_{it}^a(\Psi_S) &= Q_{it}^d(\Psi_S) - Q_{it}^a(\Psi_S) = \\ &= q \int_{E_V}^{E_C} D_{it}^d(E_S) dE_S - q \int_{E_V}^{E_C} f(E_S - \Phi_B - \Psi_S) [D_{it}^a(E_S) + D_{it}^d(E_S)] dE_S \end{aligned} \quad (2.5.4)$$

Possiamo adesso calcolarci la capacità degli stati interfacciali per variazioni lente di Ψ_S

$$\begin{aligned} \frac{dQ_{it}}{d\Psi_S} &= q \int_{E_V}^{E_C} \frac{1}{k_B T} f(E_S - \Phi_B - \Psi_S) [1 - f(E_S - \Phi_B - \Psi_S)] \cdot \\ &\quad \cdot [D_{it}^a(E_S) + D_{it}^d(E_S)] dE_S \end{aligned} \quad (2.5.5)$$

La variazione di carica interfacciale è determinata dalla somma delle probabilità per trappole interfacciali di tipo accettore e donatore; possiamo quindi definire la densità di trappole interfacciali come

$$D_{it}(E_S) = D_{it}^a(E_S) + D_{it}^d(E_S) \quad (2.5.6)$$

Non è possibile distinguere trappole interfacciali di tipo donatore da quelle di tipo accettore dalla misura di δQ_{it} visto che questa dipende dalla somma delle due diverse densità di stati interfacciali.

Nella derivazione dell'argomento dell'integrale 2.5.5, abbiamo usato l'identità

$$\frac{\partial f(E_S - \Phi_B - \Psi_S)}{\partial E_S} \equiv -\frac{1}{k_B T} f(1-f) \quad (2.5.7)$$

La funzione $f(1-f)$ ha un picco molto pronunciato quando si annulla l'argomento, ossia per $E_S = \Phi_B + \Psi_S \equiv \Phi_S$ che corrisponde ad una energia di fronte al livello di Fermi alla interfaccia del silicio. La larghezza energetica di tale picco è circa $\frac{k_B T}{q}$, quindi $D_{it}(E_S)$ è

circa costante, in un range energetico di $\frac{k_B T}{q}$, e pari a $D_{it}(\Phi_S)$.

Tale termine può allora essere portato fuori dall'integrale, rimane allora

$$C_{it}(\Psi_S) = -q D_{it}(\Phi_S) \int_{E_V}^{E_C} \frac{1}{k_B T} f(1-f) dE_S \quad (2.5.8)$$

L'integrale vale

$$\int_{E_V}^{E_C} -\frac{1}{k_B T} f(1-f) dE_S = \int_{E_V}^{E_C} \frac{\partial f(E_S - \Psi_S - \Phi_B)}{\partial E_S} dE_S = -1 \quad (2.5.9)$$

visto che $f(E_C - \Psi_S - \Phi_B) = 0$ e $f(E_V - \Psi_S - \Phi_B) = 1$ per tensioni che non portino il livello di Fermi non troppo vicino ai limiti di banda.

Quindi rimane

$$C_{it}(\Psi_S) \cong q D_{it}(\Phi_S) \quad (2.5.10)$$

2.6 Capacità misurata ad alta frequenza

La densità di trappole interfacciali può essere estratta anche da misure ad alta frequenza; la frequenza è fissata in modo che le trappole non siano in grado di rispondere. Sebbene le trappole interfacciali non seguono le rapide variazioni della tensione attorno ad un valore costante, esse cambiano il proprio stato dall'accumulazione all'inversione, ossia cambiano al variare del livello costante di tensione applicata.

In questo caso non è però presente il contributo di capacità dovuto agli stati interfacciali, come avveniva per misure a bassa frequenza; rimane però l'effetto di distensione delle curve lungo l'asse delle tensioni.

Le trappole interfacciali non causano una traslazione parallela delle curve, come invece fanno la carica fissa nell'ossido, la differenza delle funzioni di lavoro o il pinning del livello di Fermi.

Per alte frequenze la capacità totale è data dalla serie della capacità dell'ossido e di quella del silicio, infatti la $C_{it}(\omega) = 0$ per frequenze tali da impedire la risposta delle trappole interfacciali, quindi

$$C_{HF} = \frac{C_s C_{OX}}{C_s + C_{OX}} \quad (2.6.1)$$

Per poter comprendere appieno i dati ottenuti dalle simulazioni di capacità che abbiamo fatto, dobbiamo parlare di un altro fenomeno che non influenza i MOS tradizionali con biossido di silicio, ma che influenza molto le strutture con un ossido di un metallo, tipo l' HfO_2 o l' $HfSiON$.

2.7 Fermi level pinning

Il contenuto di questo paragrafo si basa su un importante lavoro fatto da un gruppo di ricercatori della Motorola, con prima firma di Christopher Hobbs, intitolato “Fermi-Level Pinning at the Polysilicon/Metal Oxide Interface” suddiviso in due parti [8] e [9].

Con l'introduzione dei composti a base di afnio al posto del tradizionale biossido di silicio si sono ottenute significative riduzioni delle correnti di perdita, ma, specie con l'utilizzo dell' HfO_2 , rimangono due problemi ancora irrisolti:

- 1) L'alta tensione di soglia V_t
- 2) Lo svuotamento del polysilicio

Il problema della tensione di soglia non può essere risolto aggiustando l'impianto di canale; lo shift della V_t riportato per processi CMOS con biossido di afnio (HfO_2), silicato di afnio ($HfSi_xO_y$), ossinitruro di afnio (HfO_xN_y) e silicato di afnio nitrato ($HfSi_xO_yN_z$) seguono lo stesso trend.

Lo shift della V_t per un nMOS con biossido di afnio è in direzione positiva, di circa $0.3 \div 0.4 V$ relativamente all' SiO_2 di controllo; molte pubblicazioni attribuiscono lo shift della V_t a carica fissa positiva nell' HfO_2 . Questa ipotesi però non spiega perché ci sia differenza negli shift per un pMOS e un nMOS, visto che la carica fissa sposta la V_t nella stessa direzione sia che sia un nMOS che un pMOS.

Tipicamente tra lo strato ad alto k , come ad esempio l' HfO_2 di questo caso, e il substrato di silicio è presente un layer interfacciale, quindi lo shift della V_t potrebbe dipendere dall'intera struttura dello stack e non dal singolo strato di ossido. Lo shift della V_t dipende

quindi da molti fattori che provengono da diverse zone dello stack, la cui struttura è riportata per comodità in figura seguente, nel caso di HfO_2 .

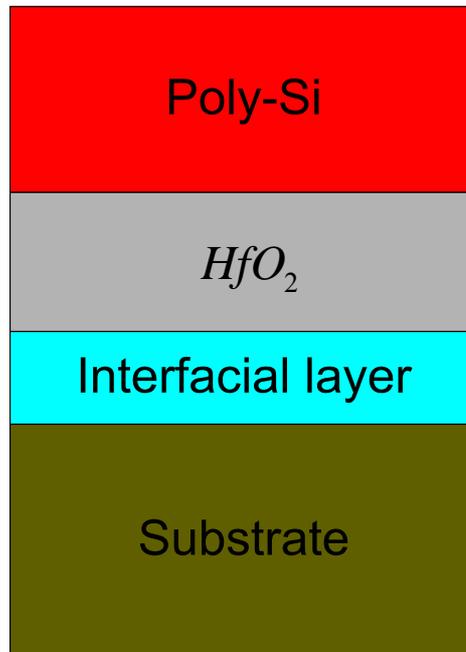


Figura 2.6: schema della struttura analizzata nel lavoro di Hobbs.

Vedremo che all'interfaccia superiore lo shift è dato dal Fermi level pinning; all'interfaccia poly-Si/ $MeOx$ ci sono sia legami silicio-metallo che legami silicio-ossigeno-metallo. Una superficie interfacciale poly-Si/ SiO_2 è invece principalmente composta da legami silicio-ossigeno-silicio; come influiscono i legami interfacciali sullo shift della V_t ?, e come influiscono sullo svuotamento del poly-Si?

Evidenze sperimentali dimostrano come un pMOS con dielettrico a base di afnio abbia uno svuotamento del poly-Si maggiore che nel caso di SiO_2 come dielettrico. Queste considerazioni modificano l'impostazione avuta sino ad ora sullo scaling dei MOS; infatti invece di ridurre l'EOT (equivalent oxide thickness) si dovrebbe pensare a ridurre il CET

(capacitance oxide thickness). Lo spessore equivalente per la capacità è infatti dato dallo spessore equivalente dell'ossido più la distanza dalle interfacce dei baricentri di carica. Pensiamo ad un condensatore a facce piane e parallele; lo spessore da usare nel calcolo è quello che individua la distanza tra le cariche sulle due facce. Semiclassicamente la carica è localizzata all'interfaccia ma se andiamo a fare un calcolo più accurato si vede che i baricentri delle cariche sono spostati rispetto all'interfaccia di alcune frazioni di nanometro; su spessori dell'ossido di pochi nanometri questo scostamento può divenire importante e se ne deve tener conto nel calcolo dello spessore equivalente per la capacità. Si può ridurre l'EOT con nuovi materiali ma se l'aumento dello svuotamento del poly-Si è grande si perdono questi vantaggi perché alla fine lo spessore effettivo sarà maggiore.

La tensione di bande piatte per una capacità MOS può essere espressa come:

$$V_{fb} = (\Phi_M - \Phi_S) - \sum_x \frac{Q_x}{C_x} \quad (2.7.1)$$

dove abbiamo definito Φ_M la funzione di lavoro del metallo (poly-Si nel nostro caso), Φ_S la funzione di lavoro del substrato, Q_x la carica nell'ossido a distanza x dall'interfaccia poly-Si/ MeO_x e C_x la capacità a distanza x dalla stessa interfaccia. La sommatoria in 2.7.1 rappresenta lo shift di V_t a causa della carica presente nello stack. Nel caso della nostra struttura, assumeremo che la carica nel biossido di afnio sia localizzata all'interfaccia HfO_2/SiO_2 e la carica nel biossido di silicio sia localizzata all'interfaccia $SiO_2/bulk$. Fatte queste ipotesi la sommatoria diventa:

$$\sum_x \frac{Q_x}{C_x} = \left(\frac{Q_{Hf} + Q_{Hf}}{C_{Hf}} \right) + \left(\frac{Q_{IL} + Q_{IL}}{C_{eq}} \right) \quad (2.7.2)$$

dove abbiamo indicato con Q_{Hf} la carica nell' HfO_2 , con Q_{IL} la carica del piano interfacciale (ossia dell' SiO_2), con $Q_{Hf/IL}$ la carica all'interfaccia Hf/IL , con $Q_{IL/Si}$ la carica all'interfaccia $SiO_2/bulk$, con C_{Hf} la capacità del piano di HfO_2 e con C_{eq} la capacità dello stack HfO_2/SiO_2 . Si possono separare sperimentalmente questi termini, ad esempio variando lo spessore dell' IL si può studiare l'effetto dei termini Q_{IL} e $Q_{IL/Si}$, mentre variando lo spessore dell' HfO_2 si può studiare l'effetto di Q_{Hf} e $Q_{Hf/IL}$. L'impatto sulle misure di Φ_M e Φ_S si studia variando il drogaggio di gate e substrato rispettivamente.

In una struttura convenzionale poly-Si/ $SiO_2/bulk$ è lecito trascurare gli stati interfacciali mentre nel nostro caso questo non è più corretto e dobbiamo tenere conto del Fermi level pinning. Daremo adesso una breve descrizione qualitativa di questo fenomeno.

Supporremo, per semplicità, di avere una interfaccia metallo/dielettrico; la figura mostra i due materiali subito prima di venire a contatto:

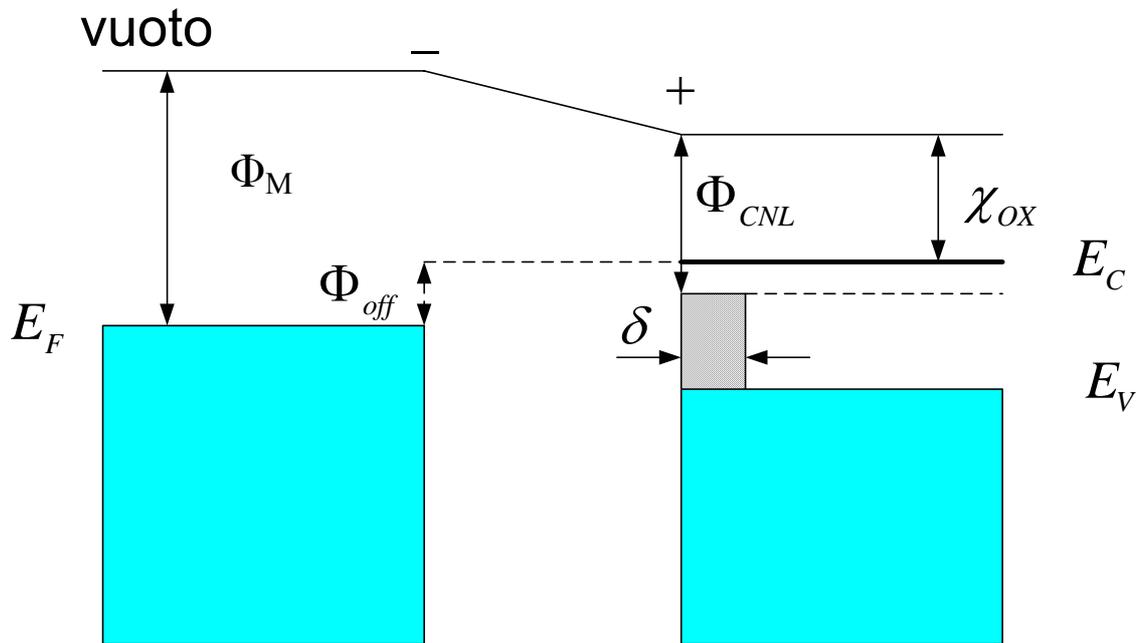


Figura 2.7 : interfaccia tra il gate e l'ossido di afnio; le grandezze riportate sono dettagliate nel seguito del paragrafo.

Nella figura sono indicate le seguenti grandezze: Φ_{off} altezza della barriera, Φ_M funzione di lavoro del metallo, χ_{OX} affinità elettronica del dielettrico, Φ_{CNL} il livello di neutralità di carica e con δ abbiamo indicato la profondità di penetrazione degli stati superficiali.

Per definizione il livello di neutralità di carica è il più alto stato superficiale occupato; si può pensare come un livello di Fermi locale nel senso che pensando ad una superficie neutra, gli stati superficiali con energia minore di Φ_{CNL} sono occupati, quelli ad energia maggiore sono vuoti. Come abbiamo già detto gli stati superficiali possono nascere da legami rotti o difetti che si formano nel processo tecnologico e sono stati localizzati con energia nel gap del silicio. Il valore di Φ_{CNL} è ricavabile solo da complessi calcoli superficiali, in cui non ci addentreremo; per approfondire l'argomento si può far riferimento, ad esempio, a [10].

Quando i due materiali vengono “messi in contatto” il livello di Fermi del metallo si sposta verso Φ_{CNL} attraverso uno scambio di carica dalla superficie di contatto. Lo shift di Φ_M verso Φ_{CNL} è relativo e riflette la discontinuità del livello del vuoto all’interfaccia creata dal dipolo di carica che si forma. La funzione di lavoro del metallo rimane invariata lontano dall’interfaccia.

Il gruppo di lavoro della Motorola ha eseguito una serie di misure, di seguito riportiamo quelle di nostro interesse; la figura seguente infatti mostra le varie curve C-V per lo stack descritto in precedenza.

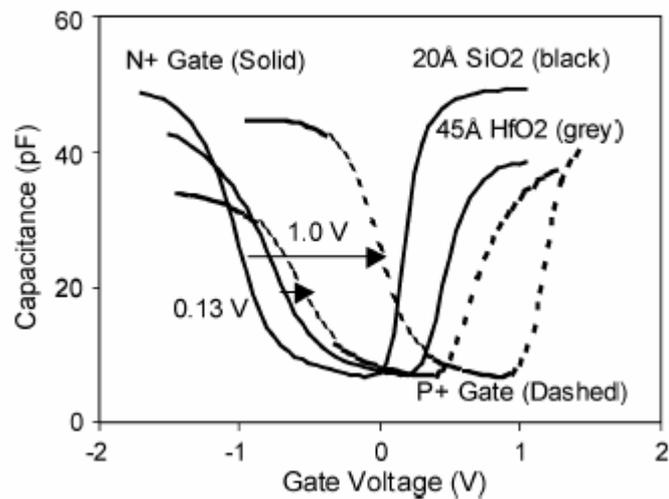


Figura 2.8: la figura è stata presa dall’articolo di Hobbs et al. citato in bibliografia.

Come si può notare passare da poly- $n+$ a poly- $p+$ fa shiftare molto poco le curve, a differenza di quanto avviene per un semplice ossido di silicio dove lo shift è di circa 1V ossia il gap del silicio; nel MOS con stack HfO_2/SiO_2 non c’è molta differenza tra gate $n+$ e $p+$; in figura è riportato un valore di 0.13 V.

Da osservare che anche la presenza di uno stato interfacciale cresciuto prima dell'ossido di afnio, lascia inalterato lo shift della V_t ; visto che l'impatto della carica nell'ossido è funzione di quanto vicino è al gate, l'indipendenza dalla riduzione dello spessore dell'ossido esclude che Q_{IL} e $Q_{IL/Si}$ contribuiscano in maniera determinante alla traslazione di V_t . Così come fatto per studiare la carica nel piano interfacciale, il gruppo di ricerca della Motorola ha fatto lo stesso con l'ossido di afnio per studiare l'effetto della carica nell'ossido di afnio e all'interfaccia poly-Si/ossido di afnio. Questi studi hanno portato alla conclusione che la carica fissa non contribuisce in maniera rilevante al fenomeno di shift di V_t ; il pesante spostamento della tensione di soglia può essere spiegato solo tenendo in considerazione il Fermi level pinning all'interfaccia poly-Si/ $MeOx$ dovuto ai legami interfacciali. Riepilogando, lo shift della V_t è stato studiato prendendo in esame tutte le parti dello stack che costituisce l'ossido di gate ed è stato dimostrato che il maggior contributo deriva dall'interfaccia poly-Si/ $MeOx$.

Nella seconda parte dell'articolo viene eseguito uno studio specifico dell'interfaccia; i risultati che citiamo in questo paragrafo sono quelli più significativi.

Innanzitutto viene dimostrato che basta anche un submonolayer di HfO_2 cresciuto sopra dell' SiO_2 termico, per provocare uno shift della V_t . Questo dimostra ancora una volta che basta una superficie di contatto (o anche meno di una superficie completa, come in questo caso) per avere il Fermi level pinning.

L'altro importante risultato è che l'ossido di afnio termina all'interfaccia con un piano di Hf , che è responsabile dello svuotamento del poly. L'idea è che questo strato diventi parte del poly facendolo comportare come un quasi metallo. Questo non elimina completamente

lo svuotamento del poly perché il numero di atomi di Hf all'interfaccia non costituisce un piano sufficientemente spesso per agire come un metallo.

2.8 Simulazioni con ossido di afnio

In questa sezione descriveremo le simulazioni che abbiamo effettuato sulle strutture descritte in precedenza; cominceremo dalla prima struttura introdotta, quella con l'ossido di afnio.

Il lavoro svolto su questa struttura è iniziato cercando di trovare il miglior fitting per le curve di capacità; da queste simulazioni abbiamo estratto la costante dielettrica dell' HfO_2 che è risultata essere circa 25 (facciamo presente che la costante dielettrica dell' HfO_2 dipende dal processo con il quale questo viene cresciuto; per approfondimenti si veda ad esempio [11]).

Un parametro importante da definire è lo spessore equivalente dell'ossido indicato con EOT (Equivalent Oxide Thickness); questo è lo spessore di SiO_2 che da un valore di capacità uguale a quello della serie della capacità dell'ossido di afnio e dell'ossido di silicio azotato. Possiamo esprimere tutto ciò nel modo seguente:

$$\frac{1}{C_{EOT}} = \left(\frac{1}{C_{SiON}} + \frac{1}{C_{HfO_2}} \right)^{-1} \quad (2.8.1)$$

Da questa possiamo ricavare

$$\frac{EOT}{\epsilon_{SiO_2}} = \frac{t_{SiON}}{\epsilon_{SiON}} + \frac{t_{HfO_2}}{\epsilon_{HfO_2}} \quad (2.8.2)$$

Nella figura 2.10 possiamo vedere la curva misurata e quella simulata inizialmente; come possiamo notare le due curve sono traslate e hanno forma diversa. Il Fermi level pinning

estratto da tali curve risulta circa 0.3 V in accordo con i valori riportati nell'articolo di Hobbs *et al.*

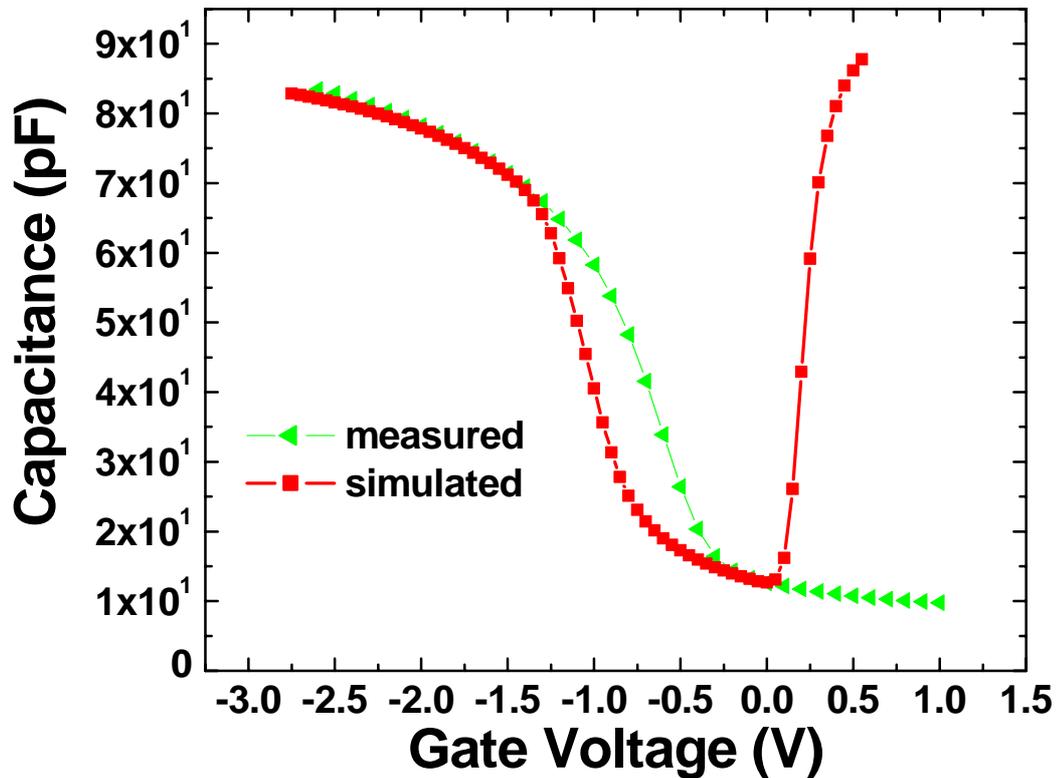


Figura 2.9: curva C-V simulata e misurata senza Fermi level pinning.

Si può notare anche la forte presenza di trappole interfacciali, che distorcono di molto la forma delle C-V; per la simulazione delle curve C-V ci siamo avvalsi , oltre alle routine principali prima descritte, di una routine *carica.f* che calcola la carica nello strato di inversione con il teorema di Gauss. Una volta calcolata la carica il programma *der.c* ne effettua la derivata in funzione della tensione per trovare la capacità che si trova nel file *dqv.out*.

Tutto il processo viene iterato dalla script Perl *cv.pl* che non fa altro che scandire tutto l'asse delle tensioni di interesse per il fitting. A questo proposito dobbiamo specificare che le curve misurate sono ad alta frequenza mentre il programma estrae curve a bassa frequenza; la zona da utilizzare per il fitting della costante dielettrica è quella di accumulazione dove la capacità è data dalla capacità dello stack di dielettrici. La zona del passaggio da inversione a accumulazione è invece utilizzata per trovare il Fermi level pinning, che viene variato fino a far coincidere le due curve. La parte di forte inversione invece è fortemente dipendente dal drogaggio del substrato, in tale zona la capacità è data dalla serie della capacità dello stack e del silicio e quest'ultima è fortemente dipendente dal drogaggio. Il profilo di drogaggio viene definito dalla routine *drogaggio.f* che viene eseguita dopo struttura e prima di *ifresh.f*, il profilo di drogaggio è fissato dal file di ingresso *drog1d.dat*; in questo file sono specificati l'ascissa iniziale e finale della zona interessata dal drogaggio (nelle simulazioni tutta la zona di bulk viene interessata dal drogaggio), il range proiettato R_p , la varianza della distribuzione dR_p e la dose dell'impianto. La routine *drogaggio.f* somma al drogaggio di substrato N_a il profilo ottenuto con la gaussiana e riscrive il file *ennea.out* che finisce in ingresso al programma principale.

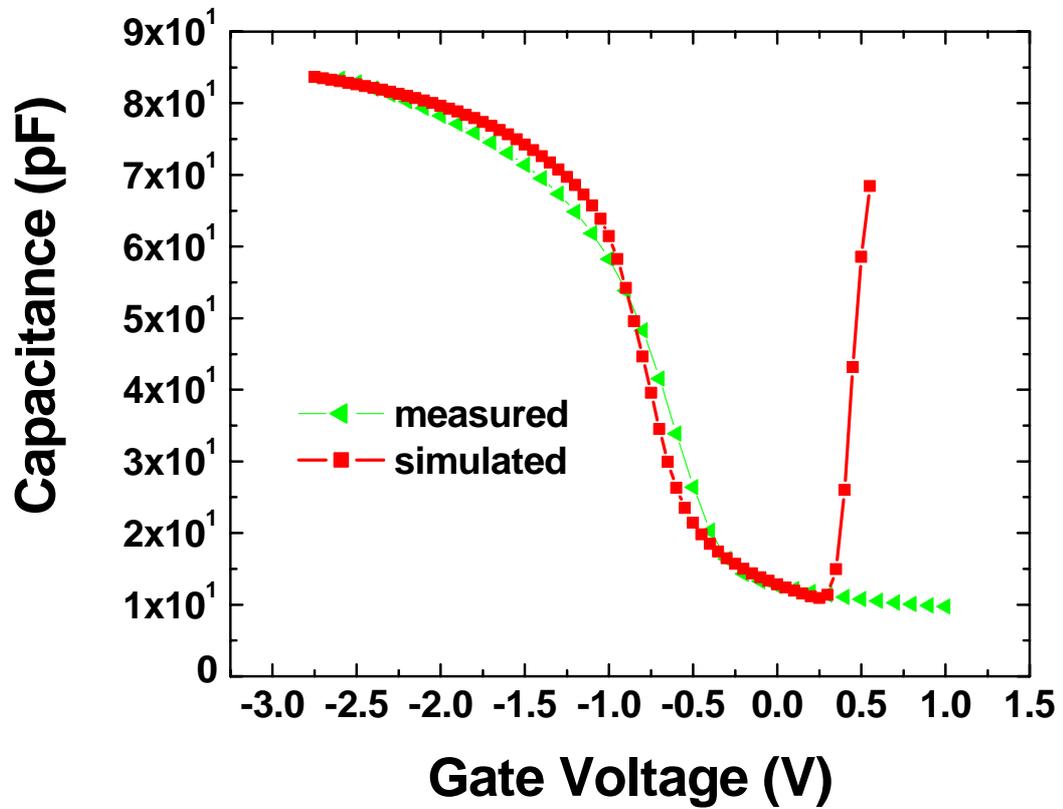


Figura 2.10: curva C-V simulata e misurata in cui è stato inserito un Fermi level pinning di 0.3V.

Aumentando il valore del drogaggio di substrato le curve si spostano verso l'alto, come mostrato in figura 2.11, dove viene mostrato l'effetto del drogaggio di substrato sulle curve C-V.

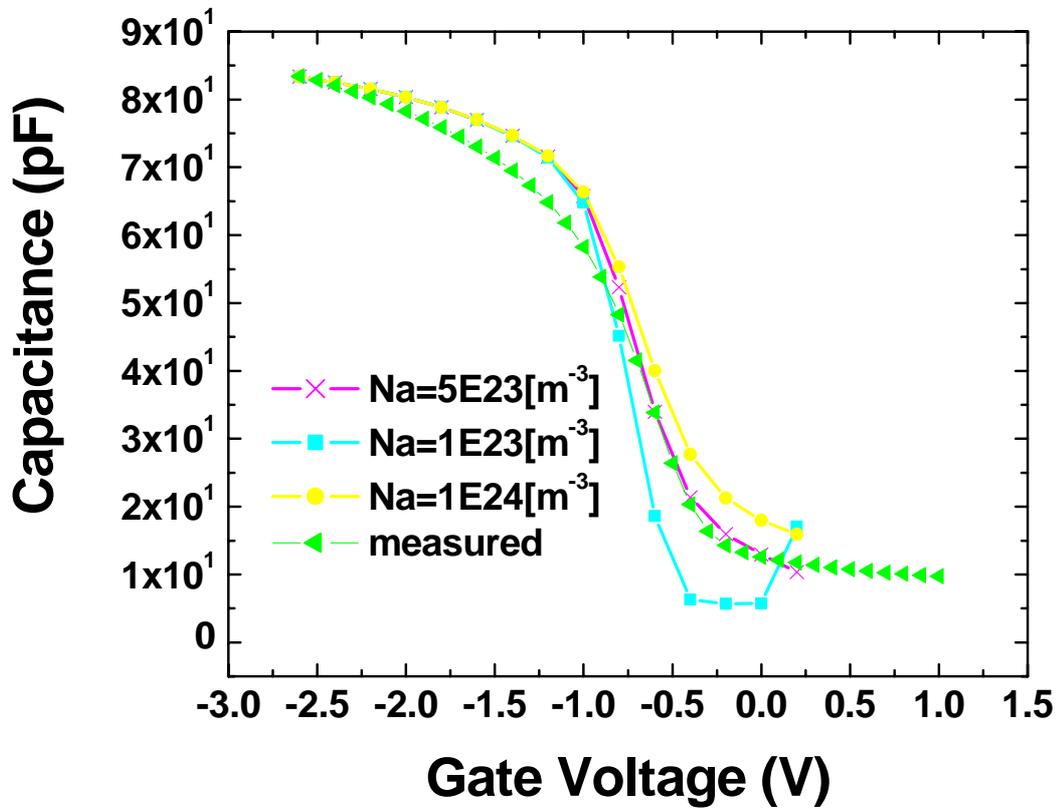


Figura 2.11: andamento delle curva C-V al variare del drogaggio di substrato; nelle nostre simulazioni abbiamo inserito il valore di 5E23 che è quello che meglio fa concordare le curve.

Nella figura 2.12 viene mostrato il profilo di drogaggio utilizzati per le simulazione della struttura con HfO_2 ; i valori di range proiettato, varianza e dose sono rispettivamente 30 nm, 20 nm e $8\text{E}24 \text{ m}^{-2}$.

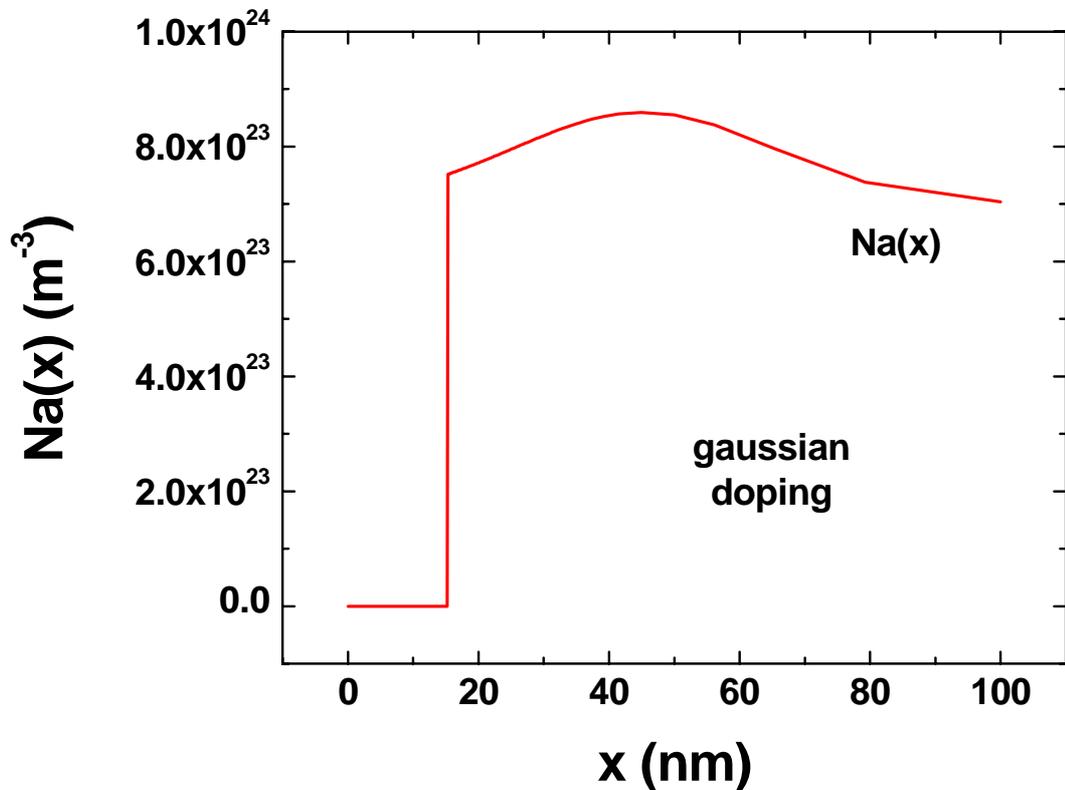


Figura 2.12: profilo di drogaggio utilizzato per le simulazioni.

Lo stesso comportamento si riscontra anche per un aumento della dose e della varianza. Il valore dello spessore degli strati di ossido è un parametro molto delicato; lo spessore dello strato di SiO_2 è molto più controllato e si può ragionevolmente supporre che non si discosti molto dal nanometro. Per quanto riguarda il valore dello strato di ossido di afnio dobbiamo tener presente che le misure di spessore sono fatte con il TEM, quindi la tolleranza è abbastanza alta (si può arrivare fino al mezzo nm). Lo spessore influisce anche sulle curve I-V quindi è molto complicato cercare di accordare le due misure. Il migliore accordo con tutte le simulazioni lo abbiamo ottenuto con uno spessore di 4.5 nm per lo

strato di ossido di afnio e di 0.8 nm per lo strato di ossido di silicio; lo strato di *SiON* è ottenuto con un processo standardizzato che garantisce uno spessore molto più preciso di quello dell'ossido di afnio.

Il parametro che abbiamo estratto dalle misure di capacità e quindi la costante dielettrica dell'*HfO₂* che è risultata circa 25; inoltre, in combinazione con le curve di corrente, abbiamo estratto i valori di spessori che sono risultati 4.5 nm e 0.8 nm.

Passiamo adesso alle misure di corrente tunnel attraverso lo stack di dielettrici; per prima cosa dobbiamo dire che c'è la necessità di usare due griglie e due file di ingresso input.form diversi. Esiste infatti una differenza netta tra le simulazioni di corrente per tensioni di gate positive e per tensioni di gate negative; Per dare solo un esempio, pensiamo che per tensioni positive il profilo di bande all'interfaccia ossido/bulk varia molto rapidamente, mentre nel caso di tensioni di gate negative la variazione è molto minore. Dovremo quindi prevedere una griglia molto fitta all'interfaccia ossido silicio per tensioni positive e nella zona di inversione considerare la quantizzazione dei livelli energetici; il tempo di rilassamento inserito è un valore ragionevole ed è di $1 \cdot 10^{-13}$ s .

Lo stesso discorso si può fare per lo svuotamento del polySi; infatti per tensioni di gate positive non c'è grande variazione, mentre per tensioni negative dobbiamo fare all'interfaccia polySi/*HfO₂* una griglia più fitta e dobbiamo tenere conto della quantizzazione dei livelli energetici.

In figura 2.13 riportiamo, per comodità, il diagramma a bande della struttura per una tensione positiva di 3 V, visto che ci servirà a chiarire alcuni aspetti.

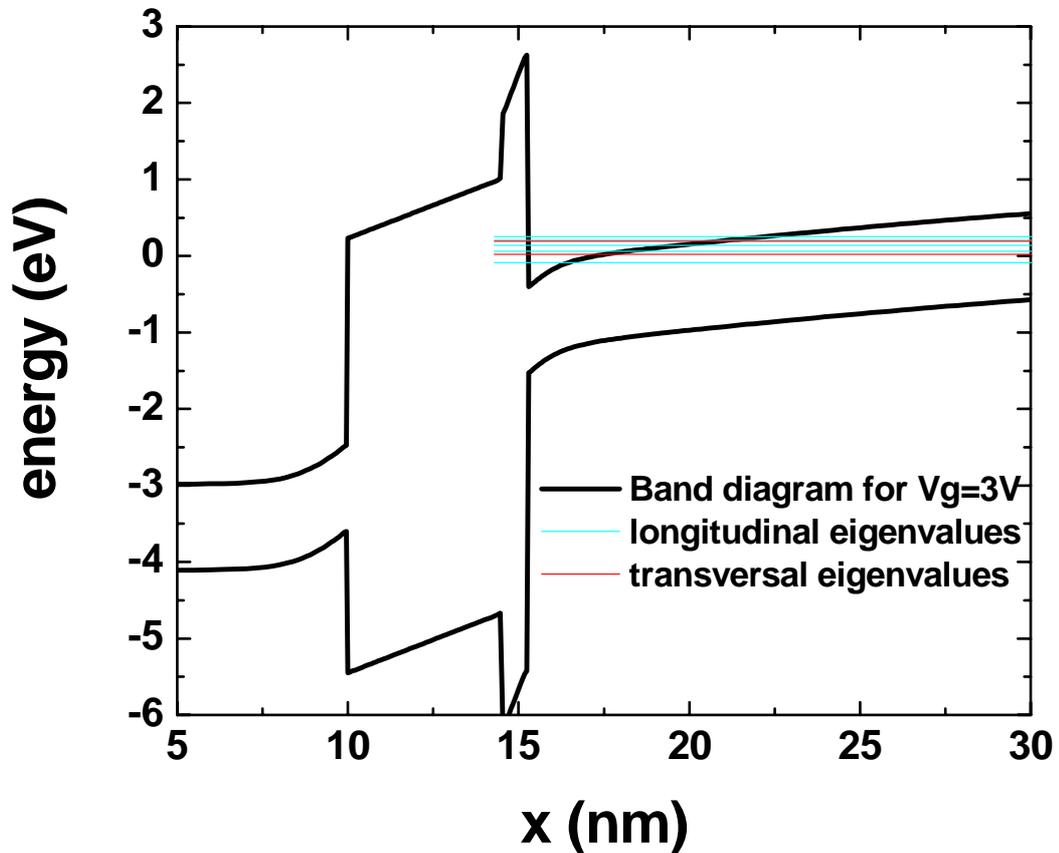


Figura 2.13: diagramma a bande della struttura; si possono vedere anche gli autovalori trasversi e quelli longitudinali all'interfaccia SiO_2 / Si .

La caratteristica I-V viene simulata dalle script `diretton.pl` e `direttop.pl` che simulano rispettivamente la corrente per tensioni negative e quella per tensioni positive. Le correnti di uscita si trovano rispettivamente nei file `corneg.out` e `corrpos.out`; da tenere presente che le curve da noi simulate danno una densità di corrente in A/m^2 mentre le curve misurate sono in A . Dovremo dividere per l'area, che è $5\mu\text{m}$ per $10\mu\text{m}$, per ottenere una densità di corrente. Di seguito viene riportato il confronto tra le curve misurate e simulate, in cui si può vedere il buon comportamento delle curve simulate per alte tensioni, come

deve essere. Infatti solo per alte tensioni la corrente è una corrente tunnel mentre a basse tensioni ci sono correnti di perdita dovute a trappole e a stati interfacciali che fanno aumentare il valore della corrente. Per tenere conto del Fermi level pinning abbiamo inserito al posto del polysilicio semplice, uno strato di materiale con le stesse caratteristiche del silicio ma con affinità elettronica di 0.3 eV maggiore:

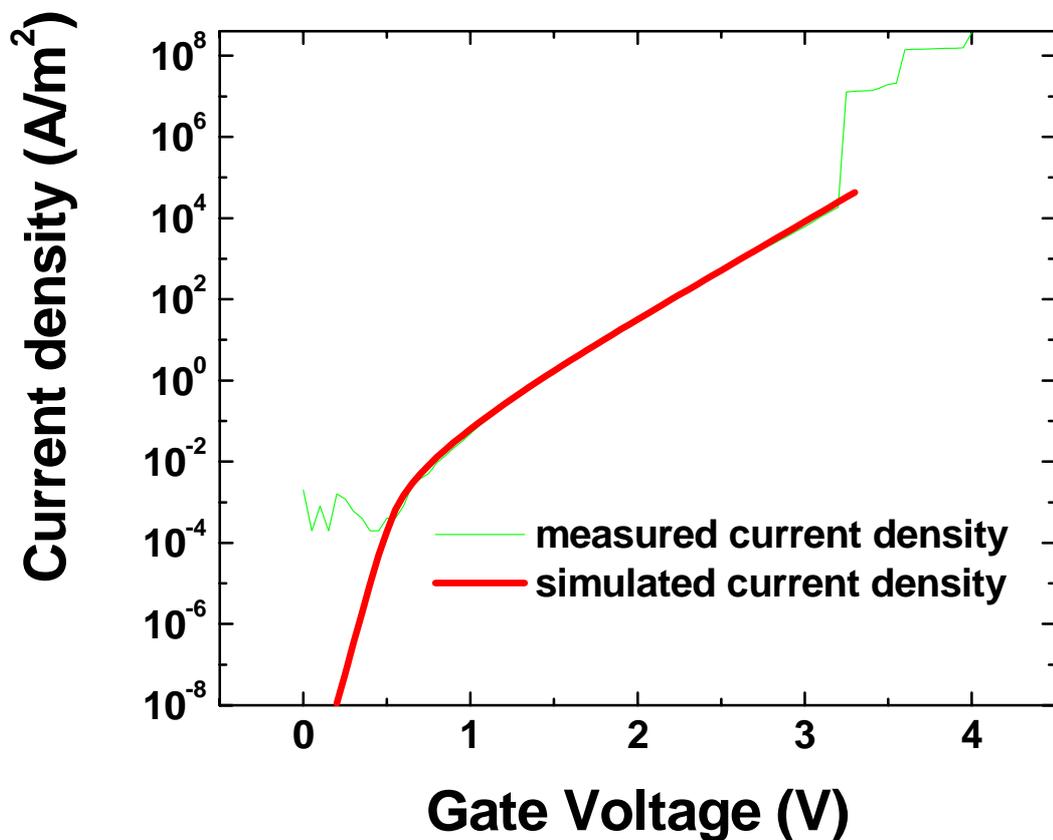


Figura 2.14: confronto tra le curve I-V sperimentale e simulata; si può notare l'accordo anche per basse tensioni di gate. Questo significa che non ci sono correnti di perdita consistenti, l'andamento della corrente misurata per bassissime tensioni può essere dovuto alle sensibilità dello strumento di misura o a correnti di perdita dovute a trappole interfacciali.

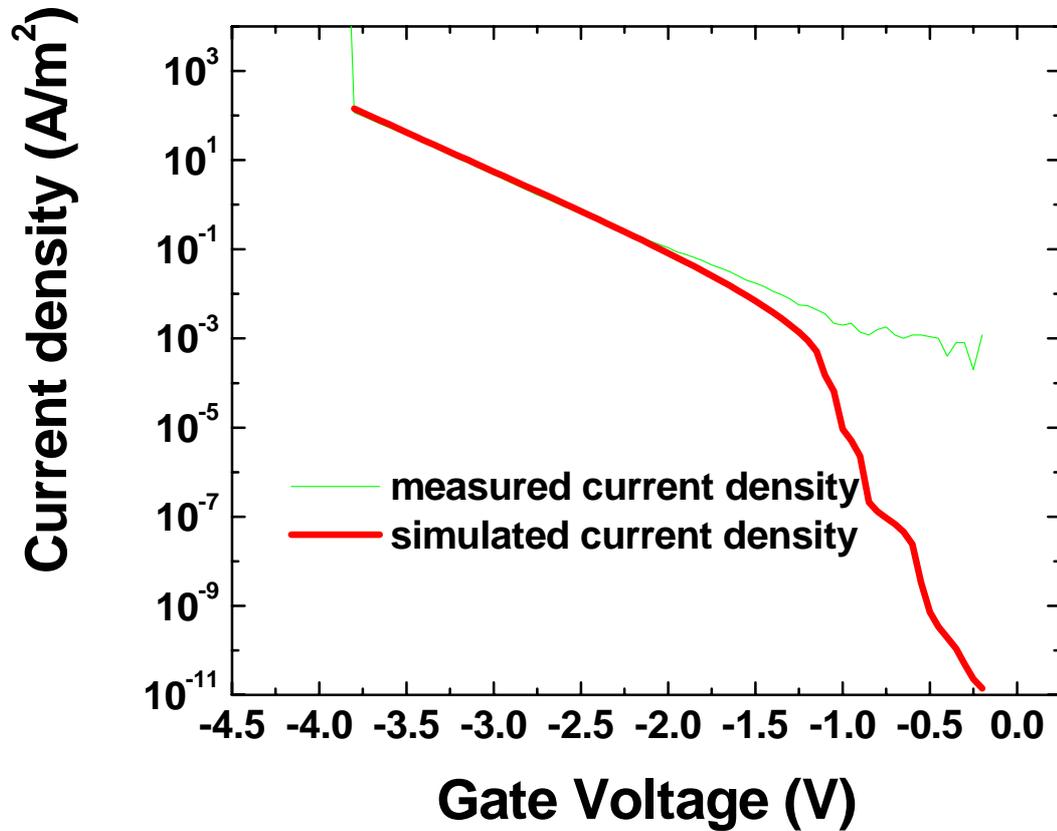


Figura 2.15: confronto tra le curve I-V sperimentale e simulata; per correnti negative si osserva una maggiore corrente di perdita rispetto a tensioni di polarizzazione positive. Questa corrente è dovuta a trappole native nell'ossido di afnio e, per tensioni piccole, a trappole interfacciali.

Come si vede per tensioni positive non sono presenti correnti di perdita; questo è ragionevole visto che per tensioni positive lo strato attraversato è 0.8 nm di SiO_2 che è molto sottile e non è in grado di introdurre forti correnti di perdita; per tensioni negative invece lo strato attraversato è composto da 4.5 nm di afnio e da 0.8 nm di SiO_2 quindi più spesso e probabilmente più difettoso a causa della presenza dell'afnio. I valori estratti dal fitting delle curve sono i seguenti: affinità elettronica 1.75 eV, massa efficace degli elettroni 0.08. I parametri sono in sostanziale accordo con quelli forniti in letteratura,

soprattutto per quel che riguarda la massa efficace (nell'articolo di Zhu e in altri ancora viene riportato il valore di 0.1, [12]); l'affinità elettronica è in disaccordo con quella fornita nell'articolo di Zhu (dove viene misurata in un valore di 2.9) ma dobbiamo ricordare che, come per la costante dielettrica anche per l'affinità elettronica è ipotizzabile una differenza in base al processo. Inoltre in altri articoli si trovano valori inferiori dello stesso parametro. Dobbiamo dire che i valori estratti possono rientrare in un vasto range in quanto c'è una seppur minima possibilità di variare i parametri e continuare a trovare un buon fitting. Fondamentalmente variare l'affinità elettronica del materiale cambia la pendenza e aumenta o diminuisce la corrente mentre la variazione di massa efficace tende solamente a traslare la curva. Nelle due figure seguenti 2.16 e 2.17 viene mostrato l'effetto dell'affinità e della massa efficace sulle curve I-V; vengono presentati i valori ottenuti per tensioni di gate negative in quanto l'effetto è identico anche per le curve ottenute con tensioni di gate positive.

Possiamo aggiungere che la massa efficace e l'affinità elettronica dell'afnio variano maggiormente la corrente negativa mentre gli stessi parametri dell'*SiON* cambiano di più la corrente positiva. Questo perché in positiva la barriera attraversata è quella di *SiON*; attraversata questa l'elettrone si trova in banda di conduzione dell'afnio e prosegue di moto balistico verso il gate quindi i parametri dell'*SiON* lo influenzano molto di più di quelli dell'*HfO₂*.

Per tensioni negative invece lo strato di afnio è il primo ad essere attraversato e quello di spessore maggiore che quindi influenza molto la corrente negativa.

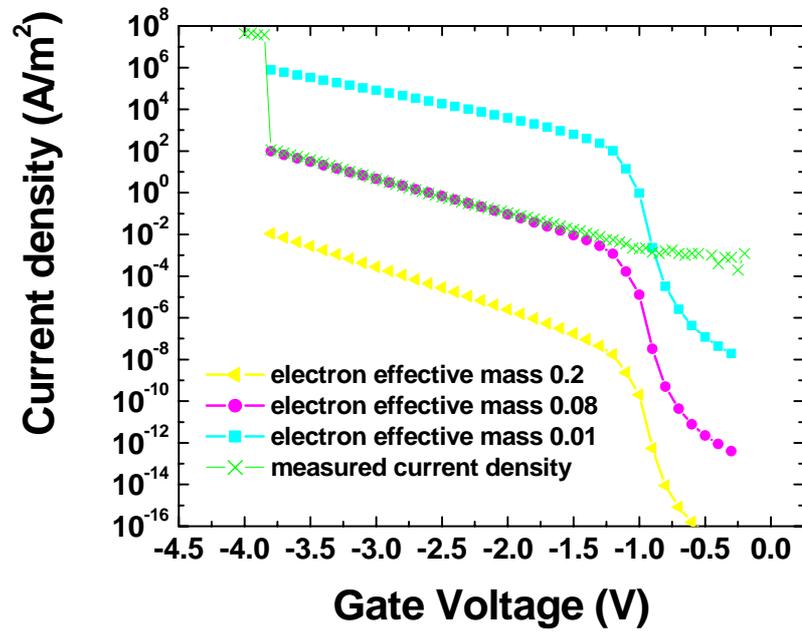


Figura 2 .16: effetto della massa efficace degli elettroni nell'ossido di afnio sulle curve I-V; i valori della massa efficace sono riportati in figura.

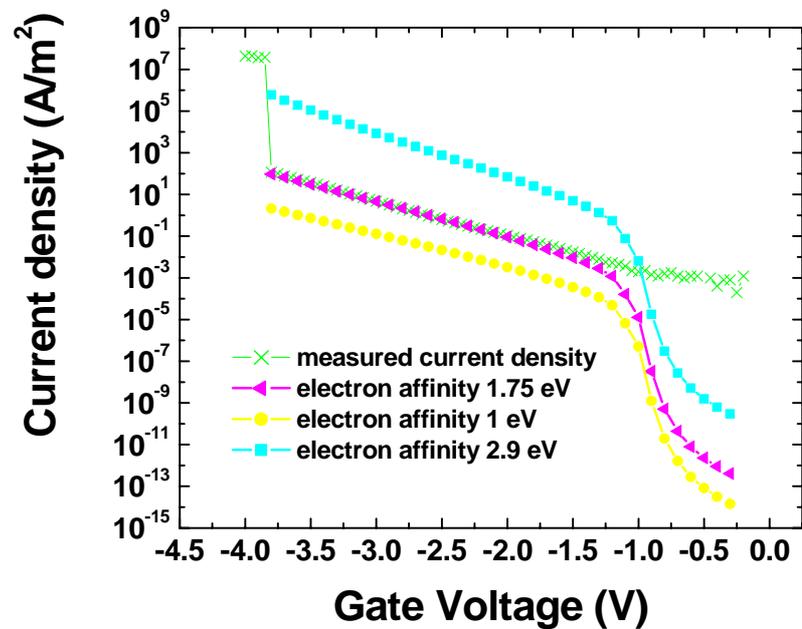


Figura 2 .17: effetto dell'affinità elettronica dell'ossido di afnio sulle curve I-V; i valori dell'affinità elettronica sono riportati in figura.

Non considerando la corrente di lacune il parametro della massa efficace delle stesse rimane ininfluenza e quindi non lo abbiamo preso in considerazione; lo abbiamo fissato a 0.5 e non c'è nessuna maniera di prevedere quanto vale senza informazioni, ad esempio, sulla soppressione di rumore shot all'interno del dispositivo. Il gap dell'ossido di afnio è di circa 5.68 V e lo abbiamo supposto costante con la temperatura dato che non sono disponibili valori dei parametri α e β necessari per il calcolo del gap a qualsiasi temperatura. Anche la temperatura è stata mantenuta costante e pari a 300 K ma sulla temperatura dobbiamo fare un breve commento. In una struttura con ossido di afnio infatti, è presente un meccanismo di conduzione di tipo Poole-Frenkel che non è previsto nel codice di simulazione; pur avendo a disposizione le curve I-V di corrente al variare della temperatura non le abbiamo utilizzate per il fitting visto che per temperature superiori a 300 K il meccanismo di conduzione diverso cambia la forma delle curve.

2.9 Passaggio dall'ossido di afnio al silicato di afnio

Il silicato di afnio è un dielettrico *medium-k* con struttura amorfa (fino a circa 1100 °C), con basse correnti di perdita, bassa densità di stati interfacciali, vicina a quella dell' SiO_2 o dell' SiO_2 nitrato. L'approccio *medium-k* permette di mantenere i vantaggi dell'ossido di silicio, ma la costante dielettrica viene aumentata introducendo un metallo, come lo zirconio o l'afnio, nella matrice di ossido. Il silicato è anche una buona barriera contro la diffusione di dopanti; l' HfO_2 , al contrario, è un materiale permeabile ai dopanti come, ad esempio, il Boro. Inoltre il silicato di afnio ha una ottima stabilità termica, cosa importantissima per i passaggi tecnologici ad alta temperatura che deve affrontare il materiale durante la costruzione. Un dielettrico amorfo è preferibile perché ha una ruvidità

interfacciale minore degli ossidi cristallini, inoltre la densità di trappole interfacciali è molto ridotta per la crescita in ambiente ricco di azoto e per natura stessa dei materiali amorfi. In tali materiali non si presentano neanche i fenomeni dovuti ai bordi di grano o alle dislocazioni; un ulteriore vantaggio nell'utilizzo dei silicati di afnio consiste nella maggiore mobilità degli elettroni rispetto all'ossido di afnio. Allo stato attuale MOSFET con silicato di afnio presentano prestazioni comparabili, se non superiori, a quelli con *SiON*. Pertanto, al momento, i silicati di afnio rappresentano il dielettrico più promettente e potrebbe essere il primo *high-k* (o meglio *medium-k*) ad entrare in produzione. In tali dispositivi inoltre la corrente varia molto poco con la temperatura, quindi il meccanismo di conduzione è tunneling puro.

Per le simulazioni con il silicato di afnio abbiamo usufruito di misure effettuate all'IMEC dal Prof. Felice Crupi dell'Università della Calabria, che ha eseguito misure di corrente e di capacità su due diverse strutture. Queste strutture hanno a comune lo strato di *SiON* cresciuto sul substrato, di circa 1nm di spessore; sopra a questo strato è stato cresciuto (o depositato) in un caso un layer di 2nm di silicato di afnio nitrato (*HfSiON*) e nell'altro 1 nm dello stesso silicato. Nel seguito indicheremo queste strutture, rispettivamente, come struttura 2-1 e 1-1.

I risultati migliori in termini di fitting sono stati ottenuti dalla struttura 1-1 per quel che riguarda le correnti, mentre le curve di capacità sono risultate migliori per la struttura 2-1. Questo è dovuto al fatto che le misure di capacità per la struttura 1-1 sono risultate attendibili fino a -1 V e quindi non è stato possibile ricavare informazioni circa la costante dielettrica, che comunque ricordiamo dipendere fortemente dal processo ed in particolare dalla quantità di azoto incorporata nell'ossido. Riportiamo comunque la simulazione per la

capacità della struttura 1-1 nella figura seguente; abbiamo comunque estratto il valore del Fermi level pinning che è risultato in accordo con quello estratto dall'altra struttura e pari a circa 0.13 V.

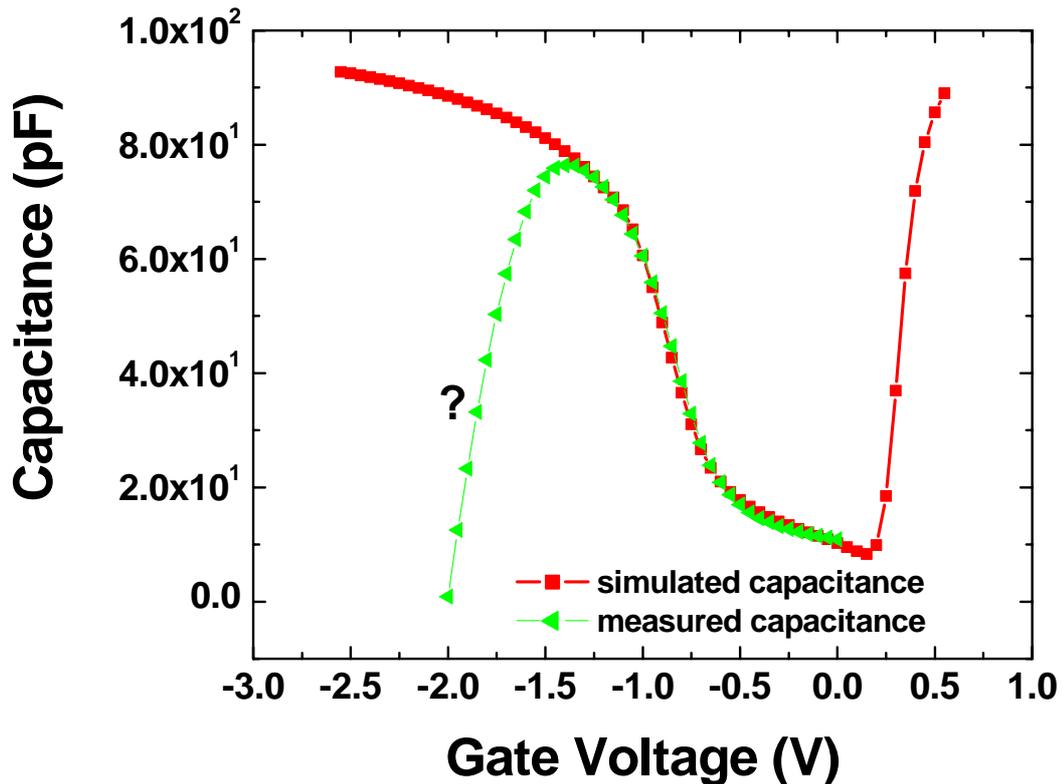


Figura 2.18: confronto tra la curva sperimentale e la curva simulata per la struttura 1-1; si può vedere che per tensioni inferiori a -1 V la curva misurata perde di significato. Non sappiamo fornire le motivazioni di questo ma abbiamo utilizzato la parte buona per ricavare il valore di Fermi level pinning.

Le misure di corrente per la struttura 2-1, invece, sono risultate leggermente in disaccordo con i dati sperimentali, soprattutto per tensioni di gate negative; tutto ciò può essere dipeso dalla tolleranza sullo spessore dello strato di silicato e dalla variabilità di molti altri parametri di processo, comunque mostreremo i risultati ottenuti per fare un paragone tra gli stessi e tra gli stessi e le strutture con ossido di afnio.

Presenteremo quindi per primi i risultati della misura di capacità della struttura 2-1 e poi le simulazioni sulla corrente per la struttura 1-1, visto che, senza perdere di generalità, potremo utilizzare questi due risultati per cercare di evidenziare i vantaggi dei silicati sui MOSFET con HfO_2 .

Cominciamo col presentare il confronto tra la curva C-V simulata e quella misurata; in figura si vedono la curva sperimentale e quella ottenuta dal simulatore senza alcuna correzione di Fermi level pinning.

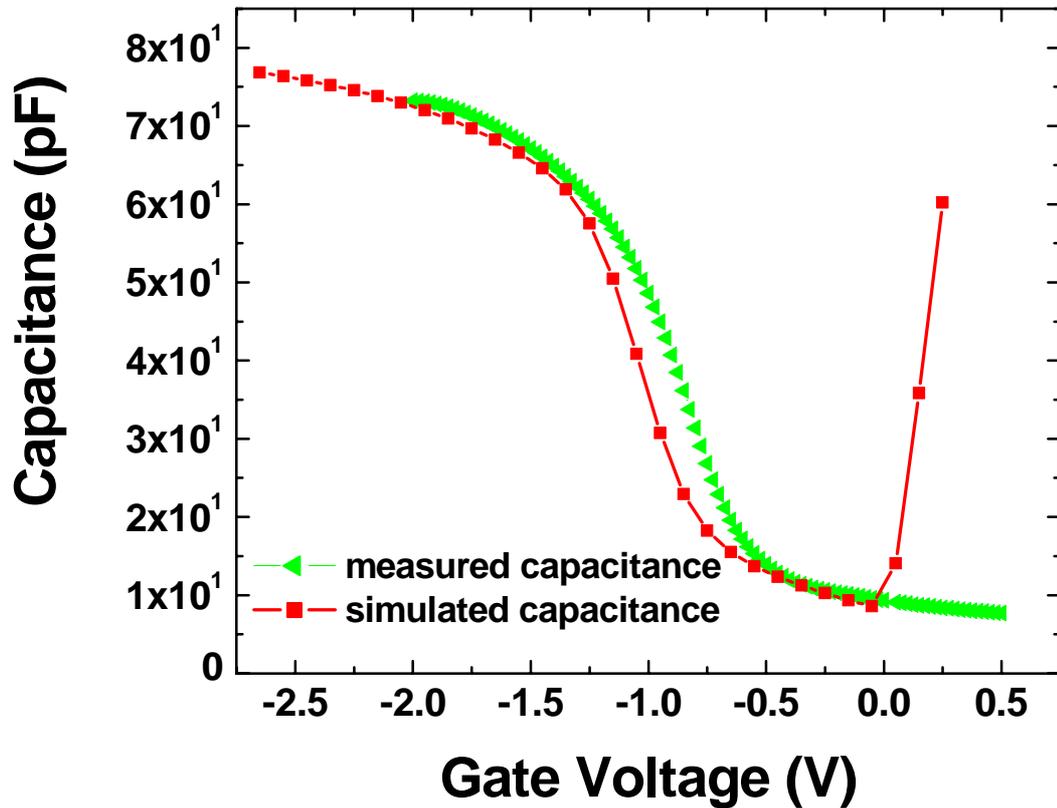


Figura 2.19: curva ottenuta dalla simulazione della struttura 2-1 senza Fermi level pinning; si può notare il minor livello dello stesso e la minor presenza di trappole interfacciali.

La prima osservazione da fare è che questo effetto è molto minore che nell'ossido di afnio, raggiungendo valori di $0.1 \div 0.2$ V contro gli $0.3 \div 0.4$ V dell'ossido di afnio. Si nota subito anche una minore presenza di trappole interfacciali, deducibile dalla minore distensione della curva C-V ; il valore della costante dielettrica estratta per il silicato e per l' $SiON$ è, rispettivamente, di 11 e 5.5 che sono in accordo con i valori riportati in letteratura.

La figura seguente mostra invece le curve C-V in cui abbiamo inserito un Fermi level pinning do 0.13 V.

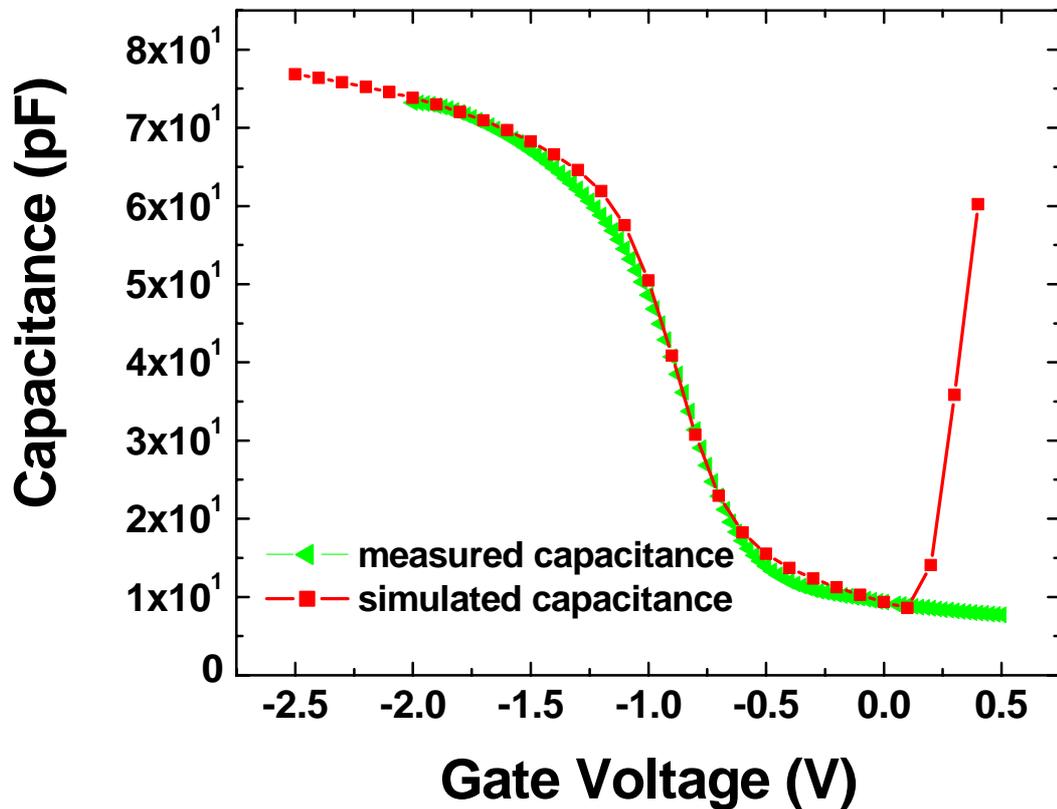


Figura 2.20: confronto tra la curva C-V simulata e quella sperimentale per la struttura 2-1; la presenza di una minore densità di stati interfacciali si vede immediatamente.

Il problema della scarsa conoscenza dei parametri di questi nuovi materiali è molto importante perché rende il problema di difficile soluzione; con ossido di silicio i valori sono molto conosciuti, mentre con l' *SiON* il valore dei parametri dipende fortemente dal processo e in particolare dalla quantità di azoto incorporata nell'ossido. Questa percentuale di ossido determina il valore della costante dielettrica, che può variare tra 3.9 e 6.5, della massa efficace, che può passare da 0.55 a 0.25, e dell'affinità elettronica, che può passare da 1 a 2, vedi [13]. La misura C-V è una misura di capacità totale mentre il programma fornisce una capacità per unità di area; la misura è stata effettuata su un condensatore di area $70\mu m$ per $70\mu m$ quindi dividendo per la superficie $S = 4.9 \cdot 10^{-9}$ troviamo la capacità totale.

Il drogaggio del substrato è ottenuto dalla esecuzione della routine drogaggio, di seguito riportiamo il profilo.

Con la variazione del drogaggio dobbiamo cercare di recuperare l'andamento delle C-V per la zona di inversione. Lo spessore equivalente dell'ossido, per questa struttura è di 1.58 nm, quindi molto basso.

Estraendo la costante dielettrica dei due materiali siamo consapevoli che una molteplice combinazione delle due possa dare lo stesso valore di capacità in accumulazione, ma il valore della costante dielettrica dell' *SiON* è in parte estraibile dai valori di massa efficace e di affinità elettronica dello stesso. Infatti, in dipendenza della concentrazione di azoto nella matrice di biossido di silicio, i valori della costante dielettrica, della massa efficace e dell'affinità elettronica variano con una certa continuità dai valori che questi parametri assumono nell' *SiO₂* puro (rispettivamente 3.9, 0.55 e 1 circa), fino a 7.5, 0.25 e 2. Come vedremo dalle simulazioni di corrente, i valori della massa efficace e dell'affinità

elettronica portano a pensare che il valore di 5 sia ragionevole; tutto ciò viene rafforzato dalle misure di costante dielettrica fatte sui silicati di afnio, che indicano un valore di 11.

Un'altra considerazione che ci porta a credere che il valore di 5 sia ragionevole è il fatto che anche la pendenza delle curve I-V dipenda, seppur lievemente, dalla costante dielettrica; nella figura seguente è rappresentato l'effetto della costante dielettrica sulle I-V e si vede come il valore di 5 sia quello che fornisce i risultati migliori.

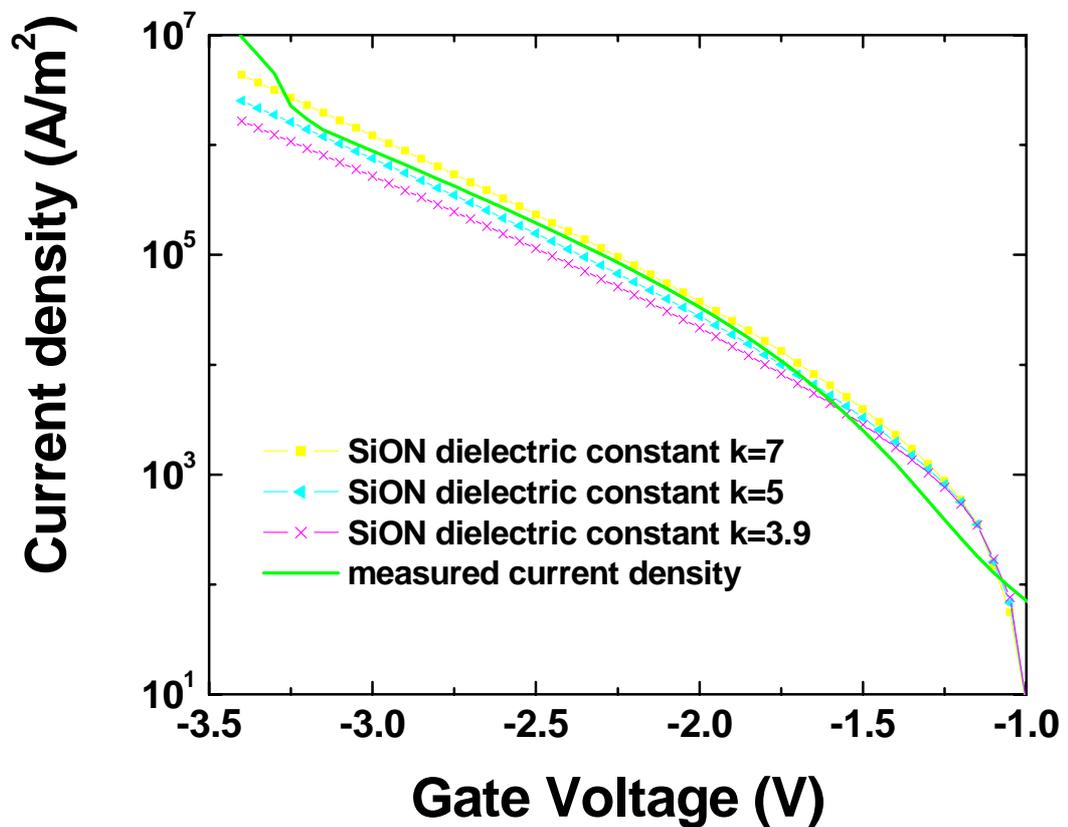


Figura 2.21: Andamento delle curve I-V al variare della costante dielettrica dell' SiON ; l'effetto è molto piccolo ma si può notare come il valore di 5 sia quello che fornisce i migliori risultati.

Anche per il silicato di afnio i parametri sono variabili con la concentrazione di azoto nella crescita, possiamo quindi pensare che anche per esso valgano le osservazioni sulla variabilità dei parametri fatta per l' *SiON* .

Vogliamo ricordare che c'è una dipendenza, seppur lieve, della pendenza delle curve di corrente in funzione della tensione, dalla costante dielettrica quindi il fitting delle C-V e delle I-V insieme è una buona verifica dei valori estratti, sempre ricordando che i parametri sconosciuti e variabili sono moltissimi.

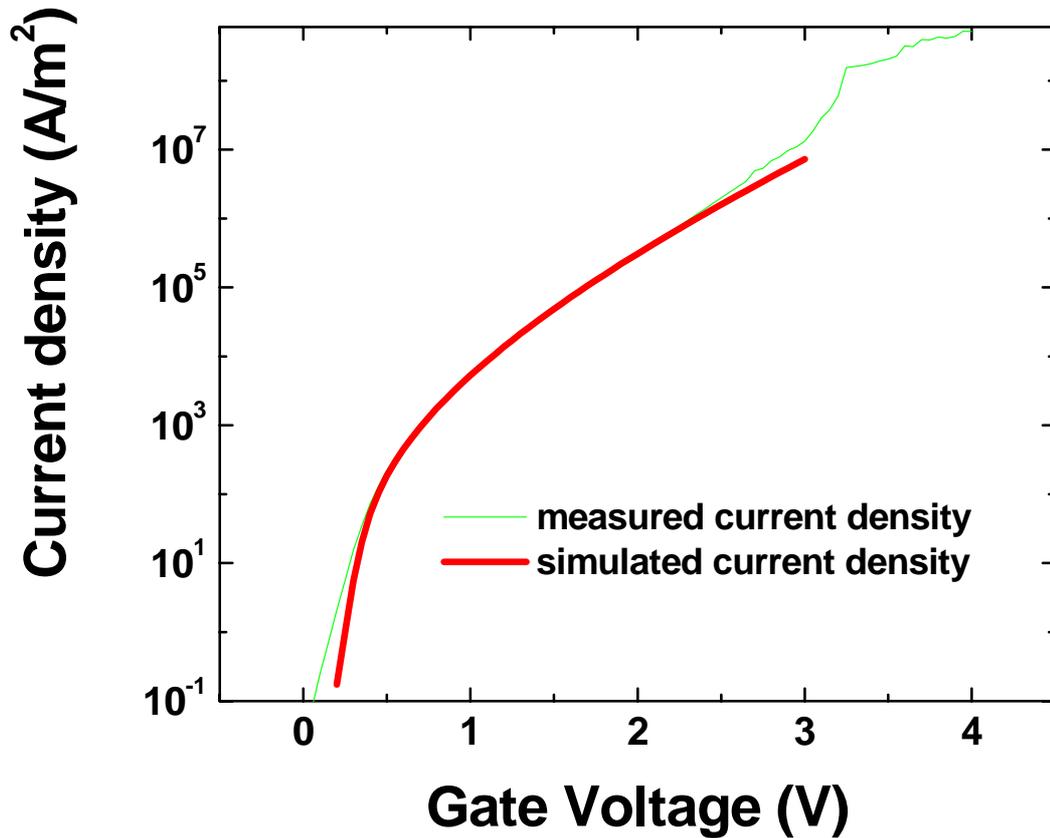


Figura 2.22: confronto tra la curva simulata e quella misurata per la struttura 1-1; si vede che non ci sono correnti di perdita significative, come ci si deve aspettare per tensioni di gate positive.

Nelle figure 2.22 e 2.23 è mostrato il confronto tra le correnti simulate e misurate per la struttura 1-1; come si può notare il fitting è molto buono tranne una piccola discordanza intorno a -2 V dove la curva misurata ha un comportamento anomalo, formando una specie di buca.

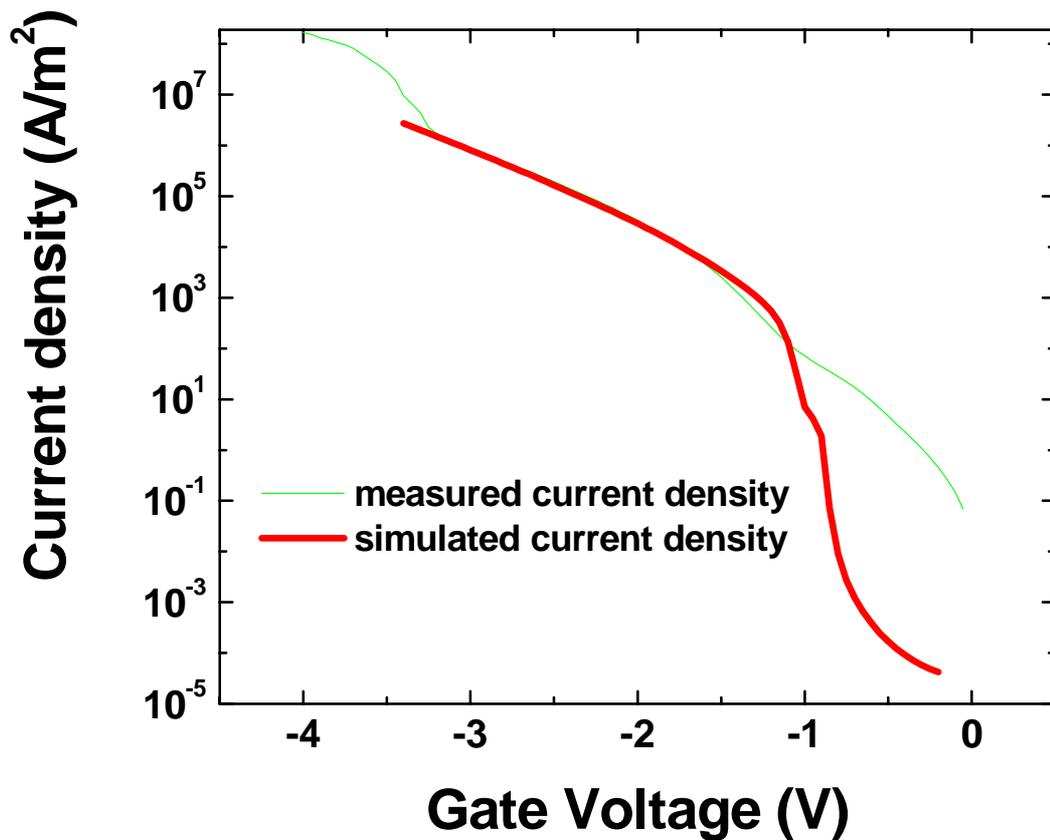


Figura 2.23: Confronto tra la corrente simulata e quella sperimentale, per tensioni di gate negative, relativo alla struttura 1-1; si nota la presenza di una corrente di perdita per basse tensioni dovuta, presumibilmente, a trappole native all'interno della struttura o a trappole interfacciali.

Per questa struttura abbiamo estratto un valore dell' affinità elettronica del silicato di afnio pari a 1.97, mentre per l' $SiON$ si ricava un valore di circa 1.27; le masse efficaci sono, rispettivamente, 0.24 per il silicato e 0.45 per l' $SiON$.

Si può notare come anche per queste strutture, ci siano correnti di perdita basse solo per piccole tensioni negative e questo per lo stesso motivo che abbiamo discusso per la struttura con ossido di silicio.

Infine vogliamo riportare i risultati sulle simulazioni per le correnti della struttura 2-1; come si vede dal confronto tra le curve simulate e quelle sperimentali, il fitting non è buono. La forma è leggermente diversa e anche i valori di corrente sono più bassi; bisogna tener presente che le curve sotto riportate sono state ottenute con gli stessi valori estratti per la struttura 1-1 e che sono stati riportati in precedenza. C'è la possibilità che il processo tecnologico determini valori dei parametri diversi da quelli ottenuti per la struttura 1-1, anche se il processo di crescita dello strato interfacciale di *SiON* è molto sicuro e affidabile; possiamo pensare che in questa struttura intervengano fenomeni di conduzione diversi oppure ancora che gli spessori riportati siano diversi da quelli effettivi. In mancanza di ulteriori informazioni ci siamo limitati alle simulazioni delle curve C-V, che sono risultate migliori di quelle della struttura 1-1, e alla conferma dei parametri di costante dielettrica fatti dal confronto tra le due strutture.

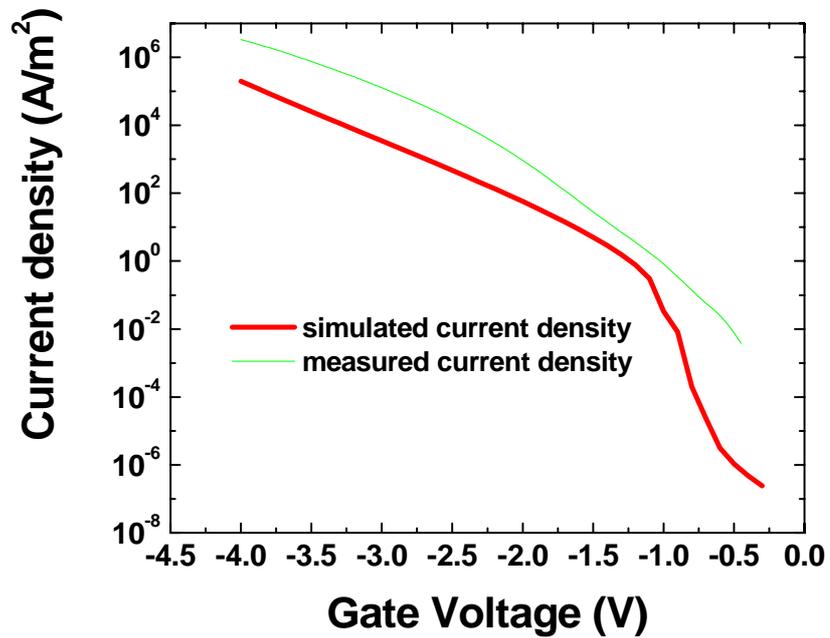


Figura 2.24: confronto tra le curve I-V per la struttura 2-1; si nota il disaccordo tra le due.

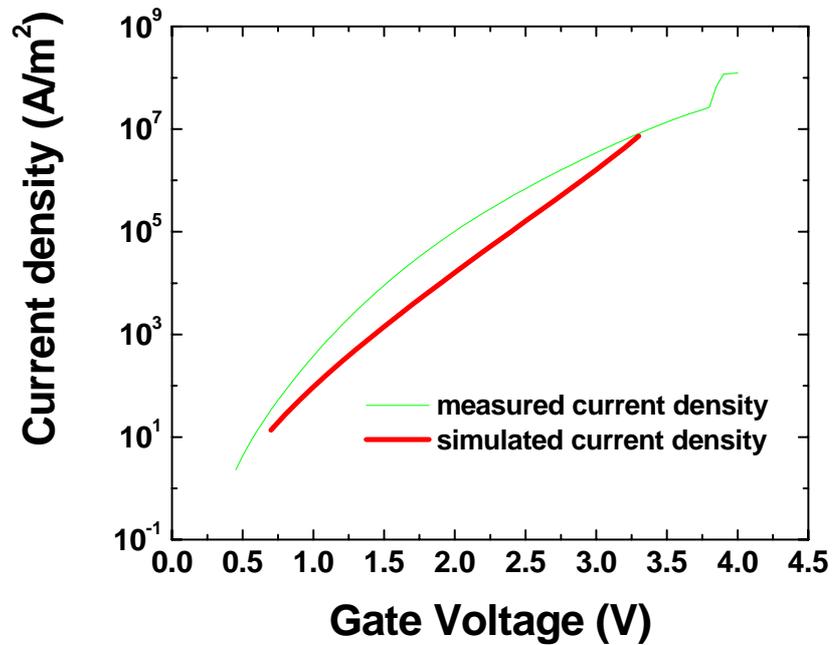


Figura 2.25: confronto tra le curve I-V per la struttura 2-1; si nota il disaccordo tra le due.

Alla fine di questo paragrafo verranno riportati i listati dei file di ingresso utilizzati nelle nostre simulazioni; non verranno riportati, per ovvie ragioni di spazio, i listati dei programmi utilizzati. L'inserimento dei listati viene fatto per aiutare chiunque debba utilizzare i risultati di questa tesi a riprodurli facilmente senza inutili perdite di tempo per cercare i migliori parametri per il fitting. Riporteremo quindi i file di ingresso `inputn.form` in cui viene specificata la struttura passata al programma principale (solo quella per tensioni di gate negative; per passare a quella per tensioni positive basta cambiare la zona da trattare quantisticamente e mettere il tempo di rilassamento a destra della barriera invece che a sinistra), il file di ingresso che contiene la griglia (sempre per tensioni negative; si può mantenere lo stesso file anche per le simulazioni delle tensioni positive, altrimenti conviene infittire la griglia nella zona a destra della barriera e farla più rada a sinistra, a scapito però del tempo necessario a completare la simulazione che è proporzionale al numero di punti interni alla zona di calcolo quantistico) e il file `drog1d.dat` che contiene i parametri del profilo di drogaggio. Non verranno inseriti i file per tutte le strutture considerate ma solo quelli per l'ossido di silicio, ossia la prima delle strutture analizzate; l'estensione alle altre strutture è immediata visto che i problemi incontrati nella struttura presentata sono gli stessi per tutte e tre le strutture.

Il valore di 4.45 eV che si vede nel file `inputn.form` è stato inserito perché, considerando il Fermi level pinning, l'affinità elettronica del polySi è stata aumentata di 0.3 eV diventando 4.45 eV . Se avessimo considerato l'affinità elettronica dell'alluminio, quella che viene generalmente utilizzata per le simulazioni, di 4.05 eV ci sarebbe stato uno svuotamento all'interfaccia.

2.10 Listato dei file d'ingresso

DROG1D.DAT

```
15.3      xiniziale
100       xfinale
30        range proiettato
20        varianza
8e24     dose
```

INPUTN.FORM

```
300      temperatura[K]
{vg}     Efermi gate di controllo(-V applicata)[eV]
4.45     Funzione lavoro gate di controllo[eV]
0        Efermi contatto di bulk[eV]
4.1      Funzione lavoro contatto di bulk[eV]
{ril}    coefficiente sottorilassamento
{nor}    tolleranza per la norma2[V]
{fp}     flagpot(se 1 usa il pot.rif. altrimenti sol.precedente)
4        numero di regioni
1        regione1
9        tipo di materiale(poly pinning)
1        flagn
1        flagp
0        Xiniziale[nm]
10       Xfinale[nm]
0        concentrazione Na[m**-3]
{ndg}    concentrazione Nd[m**-3]
0        carica fissa
1        flagf(se 1 Ef=Ef control gate altrimenti Ef=Ef bulk)
2        regione2
8        tipo di materiale(HfO2)
0        flagn
0        flagp
10       xiniziale
14.5     xfinale
0        concentrazione Na[m**-3]
0        concentrazione Nd[m**-3]
0        carica fissa
0        flagf
3        regione3
2        tipo di materiale (ossido)
0        flagn
0        flagp
14.5     xiniziale
15.3     xfinale
0        concentrazione Na[m**-3]
0        concentrazione Nd[m**-3]
0        carica fissa
0        flagf
```

```
4           regione4
1           tipo di materiale (bulk)
1           flagn
1           flagp
15.3       xiniziale[nm]
100        xfinale[nm]
{na}       concentrazione Na[m**-3]
0          concentrazione Nd[m**-3]
0          carica fissa
0          flagf
1          regioni dove calcolare Schroedinger
1          region1
1          fc
0          xiniziale
10.1       xfinale
1          barriere di potenziale
1          region1
10         xiniziale
15.3       xfinale
0          flagzs
1e-13      taus
1          flagzd
0          flagrandom
```

GRIID.DAT

```
10         numero di regioni
1          region1
0          xiniziale
5          xfinale
1          delta iniziale
0.5        delta finale
2          regione2
5          xiniziale
9          xfinale
0.5        delta iniziale
0.1        delta finale
3          regione3
9          xiniziale
10         xfinale
0.1        deltainiziale
0.05       deltafinale
4          regione4
10         xiniziale
12         xfinale
0.05       delta iniziale
0.1        delta finale
5          regione5
12         xiniziale
14         xfinale
0.1        deltainiziale
0.1        deltafinale
6          regione6
```

14	xiniziale
15.3	xfinale
0.1	deltainiziale
0.06	deltafinale
7	regione7
15.3	xiniziale
17	xfinale
0.06	deltainiziale
0.1	deltafinale
8	regione8
17	xiniziale
35	xfinale
0.1	deltainiziale
0.5	deltafinale
9	regione9
35	xiniziale
50	xfinale
0.5	deltainiziale
5	deltafinale
10	regione10
50	xiniziale
100	xfinale
5	deltainiziale
20	deltafinale

CAPITOLO 3

Memorie a nanocristalli

3.1 Introduzione

Le memorie a nanocristalli sono una particolare realizzazione delle memorie flash; queste fanno parte della categoria delle memorie ROM (read only memory), ovvero memorie che possono mantenere l'informazione anche in mancanza di alimentazione.

Le flash sono fatte con transistori a gate flottante (FLOTOX) che contengono, all'interno dello strato di ossido, un gate detto, per appunto, flottante. Nelle memorie a nanocristalli il gate flottante viene sostituito da un piano di nanocristalli immerso in un layer di ossido di silicio; questo permette di avere a disposizione una carica distribuita in maniera puntuale invece che stratificata, comportando così una capacità di accoppiamento più piccola.

La carica viene immagazzinata in maniera distribuita su un insieme discreto di nanocristalli di silicio (detti dot) mutuamente isolati. Ciascun dot immagazzinerà un numero limitato di elettroni; collettivamente le cariche immagazzinate controllano la conducibilità del canale della struttura .

I vantaggi che comporta l'immagazzinamento di carica nei nanocristalli sono molteplici tra cui il principale è di poter utilizzare ossidi tunnel più sottili senza sacrificare la caratteristica di non volatilità. Questa risulta essere l'elemento tecnologico di svolta perché offre la possibilità di operare con tensioni più basse e con velocità maggiori. La tensione

utilizzata per programmare la memoria è di pochi Volt applicati sull'elettrodo di gate; in questa maniera vengono iniettati elettroni da substrato verso i nanocristalli. Gli elettroni vengono intrappolati nei nanocristalli e l'alto spessore dell'ossido di controllo impedisce il tunneling verso il gate. La presenza dell'effetto Coulomb Blockade e l'elevato contenimento quantistico cui è sottoposto l'elettrone nel dot, consente la definizione di un processo autolimitante di carica che comporta la presenza di un numero di elettroni per dot che dipende fundamentalmente dall'entità della tensione applicata al dispositivo.

Questo si può osservare dalle caratteristiche I_D / V_G nelle quali si nota una discontinuità di corrente quando un elettrone viene intrappolato nel dot; l'intrappolamento comporta un incremento della tensione di soglia quindi l'informazione può essere letta valutando la corrente di saturazione ad una tensione più piccola di quella utilizzata per la programmazione. L'operazione di cancellazione si ottiene applicando una tensione negativa che consente l'espulsione degli elettroni dai nanocristalli nel canale.

Il processo di crescita di nanocristalli deve consentire una distribuzione spaziale e una forma degli stessi la più uniforme possibile; questo processo è inoltre meno complicato del processo dual-poly necessario per fabbricare una cella a gate flottante.

Inoltre le memorie a nanocristalli sono caratterizzate da un'eccellente immunità alle correnti di perdita (SILC) dovute allo stress dell'ossido, causato dai continui cicli di scrittura/cancellazione; anche i difetti dell'ossido sono meno dannosi grazie alla natura dell'immagazzinamento di carica nel layer di nanocristalli e al fatto che la tensione applicata è di solito minore di quella necessaria per indurre SILC.

Nelle memorie a nanocristalli si ottengono prestazioni soddisfacenti anche in termini di retention time, anche se non confrontabili con quelle di una flash tradizionale, grazie alla

possibilità di ridurre la perdita di carica laterale grazie alla presenza di ossido che separa i nanocristalli stessi .

Quando un elettrone è all'interno di un nanocristallo, l'energia del sistema si incrementerà di un termine relativo al caricamento elettrostatico $e^2/2C$, dove C è la capacità del nanocristallo. Inoltre, l'energia relativa al confinamento quantistico cui è sottoposto l'elettrone all'interno di un nanocristallo, diventa di entità rilevante quando la dimensione del dot è dell'ordine dei nanometri, definendo un incremento energetico della banda di conduzione nel nanocristallo.

La perdita di carica si verifica o lateralmente a causa della conduzione tra i nanocristalli, o verso la superficie del silicio con un eventuale percorso che porta verso le regioni n+ di drain e source.

Diverse sono le tecniche utilizzate per la deposizione dei nanocristalli, alcune delle quali sono basate sulla deposizione CVD (chemical vapor deposition) dei nanocristalli su un substrato di silicio ossidato, aerosol deposition o impiantazione ionica di atomi di silicio all'interno di SiO_2 .

In generale si formerà un layer di nanocristalli con una certa densità e caratteristiche geometriche; se supponiamo che i nanocristalli abbiano una forma sferica ci sarà una distribuzione dei raggi dei nanocristalli contraddistinta da un valor medio e da una certa varianza.

Un esempio dell'utilizzo della tecnica CVD è riportato nell'articolo di I. Crupi, G. Ammendola, G. Iannaccone [14], in cui i dot vengono realizzati mediante la deposizione CVD di silano (SiH_4) con idrogeno H_2 come gas portante e successivo annealing termico rapido a 1000 °C per la cristallizzazione. I nanocristalli ottenuti presentano un raggio il cui

valore medio è di circa 2 nm ed una densità di $3 \cdot 10^{11}$ dot/ cm^2 , come si vede nella figura 3.1.

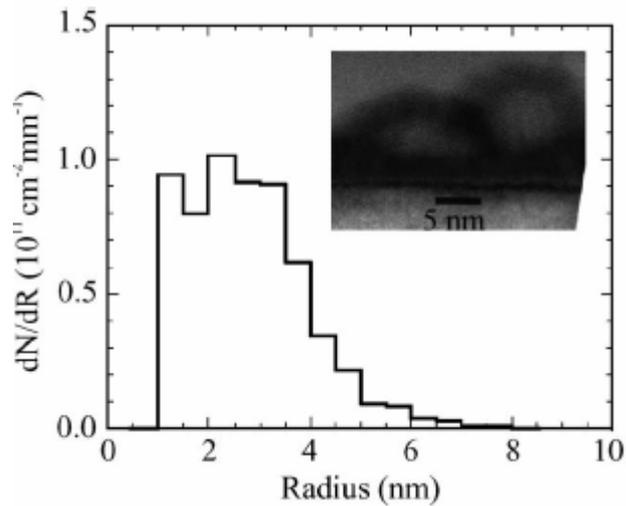


Figura 3.1: distribuzione del raggio dei nanocristalli all'interno del layer di ossido di silicio.

I fattori geometrici dei nanocristalli come il grande rapporto tra la superficie e il volume, elevata ruvidezza superficiale, non perfetta cristallinità e altri, portano ad un elevata densità di trappole interfacciali e difetti che possono agire come centri di intrappolamento di carica.

Dopo che l'elettrone è iniettato all'interno del nanocristallo, decade in uno stato relativo ad una trappola attraverso il meccanismo di scattering. Durante la fase di cancellazione gli elettroni vengono estratti dalle trappole termicamente e quindi back-scatterate verso il canale. Per quanto detto, nel processo di iniezione successivo all'applicazione di un potenziale positivo, gli elettroni andranno ad occupare gli stati vuoti relativi ai livelli profondi introdotti dalle trappole e quindi gli stati più superficiali.

3.2 Modello fisico

Per lo studio del modello utilizzato consideriamo la seguente struttura semiconduttore-isolante-semiconduttore-isolante-semiconduttore, in cui il layer di silicio presente tra i due strati di ossido di silicio rappresenta l'insieme dei nanocristalli immersi nell' SiO_2 . Il modello fa riferimento a quello riportato nell'articolo [15] G. Iannaccone, F. Crupi, S. Lombardi, che era stato originariamente proposto per studiare la corrente tunnel assistita da trappole attraverso l'ossido.

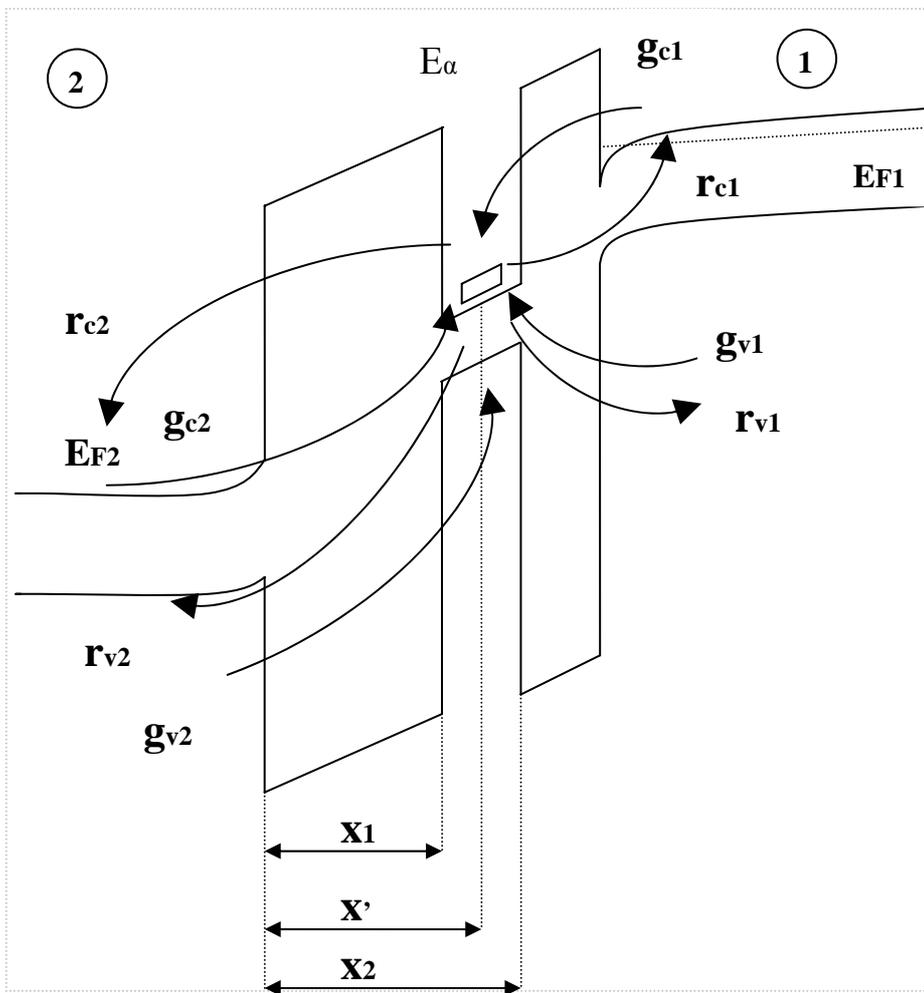


Figura 3.2: rappresentazione delle grandezze utilizzate nel modello proposto per lo studio delle memorie a nanocristalli; si notano i coefficienti di generazione e ricombinazione utilizzati.

Indichiamo con $d = X_1 - X_2$ la dimensione della regione occupata dai nanocristalli e con E_α il livello energetico rappresentante uno dei sei minimi nella banda di conduzione del silicio, con il quale si supporrà, da qui in avanti, siano realizzati i dot.

In questo modo vengono definiti dodici stati disponibili per l'elettrone nel dot, in corrispondenza dello stato fondamentale, dei quali sei sono contraddistinti da una componente di spin $+1/2$ e gli altri sei da una componente $-1/2$.

Indicheremo con $|\alpha\rangle$ lo stato in cui si trova l'elettrone quando risulta essere confinato nel dot e con $|\beta\rangle$ quello che caratterizza l'elettrone presente in uno dei due semiconduttori, substrato o gate.

Utilizzeremo il formalismo dei processi di generazione e ricombinazione, quindi definiamo il tasso di generazione g come il numero di transizioni di un elettrone, da un elettrodo verso il nanocristalli, nell'unità di tempo e , corrispondentemente, il tasso di ricombinazione r come il numero di transizioni di un elettrone, dal nanocristallo verso uno degli elettrodi, nell'unità di tempo.

Come abbiamo già visto nel calcolo della corrente tunnel, la regola d'oro di Fermi ci fornisce la probabilità per unità di tempo $\nu_{\beta \rightarrow \alpha}$ che un elettrone ha di passare dallo stato $|\beta\rangle$ allo stato $|\alpha\rangle$. Se ipotizziamo che la transizione avvenga senza perdita di energia, ossia supponiamo che la transizione sia elastica, abbiamo:

$$\nu_{\beta \rightarrow \alpha} = \frac{2\pi}{\hbar} |M(\beta, \alpha)|^2 \delta(E_\beta - E_\alpha) \quad (3.2.1)$$

dove abbiamo indicato con $M(\beta, \alpha)$ la matrice di transizione tra gli stati $|\alpha\rangle$ e $|\beta\rangle$, mentre con E_α ed E_β abbiamo indicato le energie associate ai corrispondenti stati ed \hbar è la costante di Planck ridotta.

Se consideriamo transizioni anelastiche, ovverosia transizioni accompagnate da perdita di energia, che può avvenire in modi diversi, dobbiamo modificare la regola d'oro di Fermi in modo che tenga conto della perdita di energia. Il termine che "conferisce" l'elasticità alla regola d'oro è il delta di Dirac $\delta(E_\beta - E_\alpha)$; possiamo pensare di sostituirlo con una lorentziana h_Γ di semiampiezza Γ , in questo modo ottenendo

$$h_\Gamma(E_\beta - E_\alpha) = \frac{1}{\pi} \cdot \frac{\Gamma}{(E_\beta - E_\alpha)^2 + \Gamma^2} \quad (3.2.2)$$

Sostituendo al delta di Dirac la lorentziana, possiamo ridefinire la probabilità di transizione nel caso anelastico:

$$v_{\beta \rightarrow \alpha} = \frac{2\pi}{\hbar} |M(\beta, \alpha)|^2 h_\Gamma(E_\beta - E_\alpha) \quad (3.2.3)$$

Si può notare come, facendo tendere Γ a zero, h_Γ tende ad una delta di Dirac recuperando la relazione scritta per le transizioni elastiche. Spesso si preferisce utilizzare un'altra espressione al posto della (3.1.3) svincolandosi dalla conoscenza della matrice di transizione ed in particolare si giunge alla seguente espressione :

$$v_{\beta \rightarrow \alpha} = \sigma_{\beta, \alpha} J(\beta, x') = \sigma_{\beta, \alpha} T(E_l) v(E_l) \quad (3.2.4)$$

e con

$$\sigma_{\beta, \alpha} = k \cdot h_\Gamma(E_\beta - E_\alpha) \quad (3.2.5)$$

dove k è una costante e il tasso di transizione viene messo in relazione con la densità di corrente di probabilità $J(\beta, x')$ dello stato $|\beta\rangle$ sul piano x' , dove il dot è posto, mediante la sezione di cattura $\sigma_{\beta,\alpha}$. $T(E_t)$ è la probabilità di trasmissione dell'elettrone attraverso una barriera monodimensionale che separa l'elettrone dal dot, funzione dell'energia longitudinale dell'elettrone stesso, mentre $\nu(E_t)$ è la frequenza dei tentativi di attraversamento della barriera stessa. Facendo riferimento alla figura 3.1, possiamo definire dei tassi di generazione e ricombinazione parziali; il concetto di tassi di generazione e ricombinazione parziali verrà specificato meglio nel seguito, per adesso diciamo che questi coefficienti vengono calcolati prescindendo dallo stato di carica effettivo del dot:

- $g_{1c} (g_{2c})$: tasso di generazione dalla banda di conduzione associata all'elettrodo 1 (2) al dot;
- $g_{1v} (g_{2v})$: tasso di generazione dalla banda di valenza associata all'elettrodo 1 (2) al dot;
- $r_{1c} (r_{2c})$: tasso di ricombinazione dal dot alla banda di conduzione associata all'elettrodo 1 (2);
- $r_{1v} (r_{2v})$: tasso di ricombinazione dal dot alla banda di valenza associata all'elettrodo 1 (2);

Il generico tasso di generazione g può essere valutato con la seguente espressione:

$$g = 2 \int_{\beta} \nu_{\beta \rightarrow \alpha} \rho_{\beta} f_{\beta} d\beta \quad (3.2.6)$$

dove abbiamo indicato con ρ_β la densità degli stati in una delle bande, di valenza o di conduzione, di uno dei due elettrodi (a seconda di quale fra i quattro tassi di generazione vogliamo considerare) e f_β è la probabilità che questi stati siano occupati; il fattore 2 tiene conto del fatto che sono possibili entrambi gli stati di spin. Analogamente per il tasso di ricombinazione è possibile scrivere:

$$r = \int_{\beta} v_{\alpha \rightarrow \beta} \rho_{\beta} (1 - f_{\beta}) d\beta \quad (3.2.7)$$

dove abbiamo inserito il termine $(1 - f_{\beta})$ per considerare la probabilità della presenza di stati vuoti nella banda di arrivo. Lo stato $|\beta\rangle$ è definito dalla sua energia longitudinale E_l , la sua energia nel piano trasverso E_T ($E_{\beta} = E_l + E_T$) e il suo spin.

Il tasso di generazione g_{1c} è ottenuto dall'applicazione della (3.1.6) considerando come stato di partenza uno stato nella banda di conduzione del substrato:

$$\begin{aligned} g_{1c} &= 2 \int_{\beta} \sigma_{\beta, \alpha} T(E_l) v(E_l) \rho_{\beta}(E) f_{\beta}(E) d\beta = \\ &= 2 \int_{E=E_l+E_T} k \cdot h_T(E_l + E_T - E_{\alpha}) T(E_l) v(E_l) \rho_1(E_l) \rho_T f_1(E_l + E_T) dE_l dE_T = \quad (3.2.8) \\ &= 2 \int_{E_l}^{\infty} k \cdot T(E_l) v(E_l) \rho_1(E_l) \rho_T dE_l \int_0^{\infty} h_T(E_l + E_T - E_{\alpha}) f_1(E_l + E_T) dE_T \end{aligned}$$

Il tasso di ricombinazione r_{1c} ha un'espressione molto simile, con la differenza che l'integrale deve essere esteso a tutti gli stati non occupati nella banda di conduzione del substrato, con lo stesso spin dell'elettrone intrappolato :

$$\begin{aligned}
 r_{1c} &= \int_{\beta} \sigma_{\beta,\alpha} T(E_l) \nu(E_l) \rho_{\beta}(E) (1 - f_{\beta}(E)) d\beta = \\
 &= \int_{E=E_l+E_T} k \cdot h_T(E_l + E_T - E_{\alpha}) T(E_l) \nu(E_l) \rho_l(E_l) \rho_T(E_T) (1 - f_1(E_l + E_T)) dE_l dE_T = \quad (3.2.9) \\
 &= \int_{E_l}^{+\infty} k \cdot T(E_l) \nu(E_l) \rho_l(E_l) \rho_T dE_l \int_0^{+\infty} h_T(E_l + E_T - E_{\alpha}) (1 - f_1(E_l + E_T)) dE_T
 \end{aligned}$$

A questo punto le espressioni degli altri tassi di transizione possono essere ricavate facilmente, andando a considerare un intervallo di integrazione che considera la sommità della banda di valenza invece del fondo della banda di conduzione.

Applicando una tensione alla struttura presa in considerazione, in condizioni stazionarie i processi di generazione dovranno bilanciare, statisticamente, i processi di ricombinazione.

Se indichiamo con f' il fattore di occupazione della trappola, tale bilancio si può scrivere come:

$$(1 - f') \cdot (g_1 + g_2) = f' \cdot (r_1 + r_2) \quad (3.2.10)$$

avendo raggruppato i vari tassi nel seguente modo:

$$\begin{aligned}
 g_1 &= g_{1c} + g_{1v} & r_1 &= r_{1c} + r_{1v} \\
 g_2 &= g_{2c} + g_{2v} & r_2 &= r_{2c} + r_{2v}
 \end{aligned}$$

Dalla 3.1.10 si ricava immediatamente:

$$f' = \frac{g_1 + g_2}{g_1 + g_2 + r_1 + r_2} \quad (3.2.11)$$

E' necessario effettuare una considerazione, ovvero bisogna considerare le espressioni precedenti dei tassi di generazione e ricombinazione, nel caso in cui la generazione avvenga col dot privo di elettroni, e che la ricombinazione faccia riferimento all'unico elettrone che occupa il dot stesso.

Se nel dot sono presenti n elettroni, possiamo definire un tasso di generazione vero che differisce da quello calcolato con l'equazione 3.1.6 perché tiene conto degli stati effettivamente disponibili per una transizione

$$g' = \frac{(12-n)}{2} g \quad (3.2.12)$$

Infatti se nel dot sono presenti n elettroni, il numero di stati disponibili per una nuova generazione si riduce da 12 a $(12-n)$, considerando che soltanto il livello fondamentale sia interessato da nuovi processi (questo è plausibile visto che il numero di elettroni nel dot non supera mai poche unità); il fattore 2 tiene conto del fatto che nel calcolo di g si moltiplicava per due per considerare entrambe le componenti dello spin.

Un'analogia osservazione si può fare per il coefficiente di ricombinazione, infatti se nel dot non ci sono elettroni il coefficiente di ricombinazione effettivo sarà nullo in quanto non ci sono elettroni che si possono ricombinare; nel calcolo di r dell'equazione 3.1.7 non si tiene conto di tutto questo ma si ipotizza che la ricombinazione sia corrispondente ad un elettrone presente nel livello E_α .

Generalizzando quanto abbiamo appena detto possiamo pensare che il coefficiente di ricombinazione vero sia pari a

$$r' = nr \quad (3.2.13)$$

infatti se nel dot ci sono n elettroni il tasso di ricombinazione va moltiplicato per il numero di elettroni, considerando che i vari processi siano indipendenti tra loro (in questo modo si possono moltiplicare i tassi, ovvero sia le probabilità di ricombinazione, per ottenere il tasso di ricombinazione totale).

Per sviluppare il nostro modello dobbiamo fare alcune ipotesi semplificative; innanzi tutto il processo tecnologico associato definisce una distribuzione casuale di nanocristalli che, per le nostre valutazioni attuali, porta ad una densità media di circa 10^{12}cm^{-2} . In questo studio considereremo nanocristalli di tipo sferico il cui raggio presenta un valore medio di 2nm.

La prima ipotesi che dobbiamo fare è che i nanocristalli siano tutti uguali e distribuiti uniformemente; tenendo conto della monodimensionalità del nostro modello, i dot verranno considerati come un unico strato di silicio di 4 nm di spessore compreso tra due strati di ossido: ossido tunnel e ossido relativo al control gate.

Il dot per l'elettrone rappresenta una trappola con elevate proprietà di confinamento, al cui interno l'elettrone può occupare uno dei possibili livelli energetici a disposizione, il cui valore è valutabile risolvendo l'equazione di Schrödinger. In particolare se consideriamo l'elettrone confinato all'interno di una struttura cubica di spigolo $L_x = L_y = L_z = 4 \text{ nm}$, otteniamo una espressione dei livelli energetici del tipo:

$$E_\alpha = \frac{\hbar^2}{2m_x} \left(\frac{n\pi}{L_x} \right)^2 + \frac{\hbar^2}{2m_y} \left(\frac{n\pi}{L_y} \right)^2 + \frac{\hbar^2}{2m_z} \left(\frac{n\pi}{L_z} \right)^2 \quad (3.2.14)$$

Per $n = 1$ otteniamo il livello fondamentale di energia rispetto il fondo della banda di conduzione, che risulta $E_0 = 0.273 \text{ eV}$.

Le masse efficaci da utilizzare sono quelle che abbiamo calcolato nel capitolo 1 per il calcolo della corrente tunnel, che ricordiamo essere $m_x = m_l = 0.91m_0$ e $m_y = m_z = 0.19m_0$ con $m_0 = 9.1 \cdot 10^{-31} \text{ Kg}$.

La seconda ipotesi che faremo è che i nanocristalli non interagiscano tra loro, ossia che ogni dot non influisca sullo stato di carica dei dot vicini

In realtà questa risulta essere una ipotesi piuttosto forte, in quanto la presenza di carica in un dot, determina sicuramente un aumento del livello del potenziale dei dot adiacenti.

3.3 Master Equation

Vogliamo scrivere una equazione che ci permetta di studiare lo stato di carica del dot, per fare questo introduciamo alcune grandezze che saranno utilizzate nel seguito.

Indicheremo con $g(i)$ il tasso di generazione dell' i -esimo elettrone, ossia la probabilità per unità di tempo che venga generato l' i -esimo elettrone in un dot in cui sono già presenti $i-1$ elettroni; analogamente definiremo $r(i)$ il tasso di ricombinazione dell' i -esimo elettrone, ovverosia la probabilità per unità di tempo che si ricombini l' i -esimo elettrone di un dot in cui un istante prima c'erano i elettroni.

Quindi con i indicheremo lo stato di carica del dot; vogliamo allora calcolarci la probabilità $P(i)$ che nel nanocristallo siano presenti i elettroni e valutarne l'evoluzione temporale.

Possiamo scrivere:

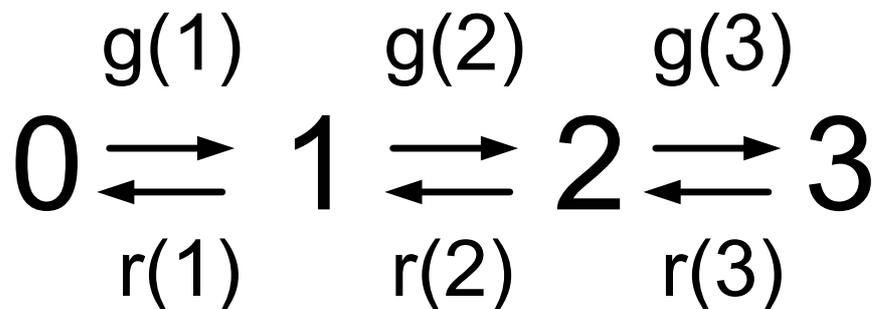
$$\frac{\partial P(N,t)}{\partial t} = r(N+1)P(N+1,t) + g(N)P(N-1,t) - P(N,t)[g(N+1) + r(N)] \quad (3.3.1)$$

La variazione temporale della probabilità che ci siano N elettroni all'istante t si calcola considerando tutti i termini che contribuiscono ad un aumento e sottraendo i termini che riducono il numero di elettroni nel dot.

Analizzando i vari termini osserviamo che:

- $g(N)P(N-1,t)$ tiene conto della probabilità che nel dot ci siano $N-1$ e che si verifichi la generazione dell' N -esimo elettrone
- $r(N+1)P(N+1,t)$ tiene conto della probabilità che nel dot ci siano $N+1$ elettroni e che si presenti un evento di ricombinazione
- $r(N)P(N,t)$ tiene conto della probabilità che nel dot ci siano N elettroni e che si presenti un evento di ricombinazione, in questo caso dopo la ricombinazione ci sarebbero $N-1$ elettroni
- $g(N+1)P(N,t)$ tiene conto della probabilità che nel dot ci siano N elettroni e che si verifichi un evento di generazione, a seguito del quale ci sarebbero $N+1$ elettroni.

Possiamo rappresentare quanto abbiamo esposto con un semplice diagramma:



Dalla soluzione della master equation, avendo precedentemente calcolato i valori dei $g(N)$ e $r(N)$, si ottengono gli andamenti temporali delle $P(N)$; calcolate le varie $P(N,t)$ è possibile ottenere l'andamento temporale del valore medio del numero di elettroni che occupano i dot. In questo modo sarà possibile avere una valutazione della

variazione della tensione di soglia e quindi dei tempi di programmazione: infatti si potrà valutare il tempo necessario affinché ΔV_T raggiunga il suo valore massimo, oppure si potrebbe valutare, per un fissato tempo di programmazione, la tensione necessaria per raggiungere il massimo valore di ΔV_T .

Il valor medio del numero di elettroni che occupano il dot e il valor medio della tensione di soglia possono essere scritte, rispettivamente, come:

$$\langle N(t) \rangle = \frac{\sum_{N=0}^{\infty} N \cdot P(N, t)}{\sum_{N=0}^{\infty} P(N, t)} \quad (3.3.2)$$

$$\langle \Delta V_T \rangle = k \langle N(t) \rangle \quad (3.3.3)$$

Nella espressione di ΔV_T si nota la presenza di un fattore k , che tiene conto della traslazione subita dalla tensione di soglia per ogni elettrone introdotto nel dot. Una ulteriore valutazione che possiamo effettuare è quella di regime in cui le variazioni temporali delle $P(N)$ sono nulle; anche in questo caso possiamo calcolare, noti i valori di $g(N)$ e $r(N)$, le diverse probabilità normalizzate rispetto al valore incognito di $P(0)$. Quest'ultimo sarà definito dalla condizione di normalizzazione della probabilità, secondo cui la somma di tutte le $P(N)$ deve essere uguale a uno. Note le $P(N)$ si potranno individuare il valore medio di N e quindi di ΔV_T .

A regime ovviamente si avrà $\frac{\partial P(N, t)}{\partial t} = 0$; possiamo allora scrivere:

$$0 = r(N+1)P(N+1, t) + g(N)P(N-1, t) - P(N, t)[g(N+1) + r(N)] \quad (3.3.4)$$

Consideriamo il dot privo di elettroni, ossia $N = 0$, otteniamo allora, sostituendo nella 3.2.4 abbiamo:

$$0 = r(1)P(1) - g(1)P(0) \quad (3.3.5)$$

$$\frac{P(1)}{P(0)} = \frac{g(1)}{r(1)} \quad (3.3.6)$$

dove abbiamo tenuto conto che $g(0) = r(0) = 0$, come ovvio che sia. Quanto fatto per zero elettroni può essere facilmente generalizzato al caso di N elettroni, ottenendo:

$$P(N) = P(0) \prod_{i=1}^N \frac{g(i)}{r(i)} \quad (3.3.7)$$

Ricordiamo inoltre che la condizione di normalizzazione della probabilità richiede che sia

$$\sum_{i=0}^{\infty} P(i) = 1.$$

Una volta note le $P(N)$ possiamo individuare i valori medi (a regime) del numero di elettroni per dot e di ΔV_T :

$$\langle N \rangle = \sum_{N=0}^{\infty} N \cdot P(N) \quad (3.3.8)$$

$$\langle \Delta V_T \rangle = k \langle N \rangle \quad (3.3.9)$$

La risoluzione della master equation si basa sulla conoscenza dei fattori di generazione e ricombinazione per le varie tensioni di gate a cui è sottoposta la memoria.

In particolare sono state prese in considerazione due strutture :3-4-7 e 4-4-8. Le tre cifre rappresentano la dimensione in nanometri rispettivamente dell'ossido di gate, della regione occupata dai dots e dell'ossido sottostante il gate, al di sopra dei dot stessi.

La 3.2.7 ci fornisce la soluzione di regime, o meglio fornisce il valore delle varie probabilità in funzione della $P(0)$, che dovremo calcolare. Generalizzando la 3.2.6 possiamo scrivere il rapporto tra due probabilità consecutive:

$$\frac{P(N+1)}{P(N)} = \frac{g(N+1)}{r(N+1)} = a_{N+1} \quad (3.3.10)$$

Sarà quindi necessario valutare l'entità dei rapporti g/r , in modo da individuare il numero di elettroni a partire dal quale l'occupazione del dot da parte di un ulteriore elettrone sarà poco probabile; quando i rapporti g/r diverranno piccoli, diciamo minori di un millesimo, la probabilità successiva sarà piccolissima e quindi trascurabile. Per un calcolo più preciso si deve valutare anche l'entità del rapporto a_N/a_{N-1} , in quanto indica il numero di ordini di grandezza che separa le due probabilità consecutive e quindi due termini che sono significativamente diversi, e la possibilità di trascurare $P(N)$ rispetto a $P(N-1)$. Una volta definito l'ordine "N", si può procedere al calcolo di $P(0)$; per la normalizzazione della probabilità sappiamo che:

$$\sum_{i=0}^{\infty} P(i) = 1 \quad (3.3.11)$$

inoltre abbiamo ricavato precedentemente che:

$$P(N) = P(0) \prod_{i=1}^N a_i \quad (3.3.12)$$

Dalla 3.2.11 possiamo riscrivere che:

$$P(0) + \sum_{i=1}^{\infty} P(i) = 1 \quad (3.3.13)$$

Supponendo che N sia il numero massimo di elettroni che con probabilità non trascurabile può occupare il dot, abbiamo:

$$P(0) + \sum_{i=1}^N P(i) = 1 \quad (3.3.14)$$

$$P(0) + \sum_{i=1}^N P(0) \prod_{j=1}^i a_j = 1 \quad (3.3.15)$$

quindi ricaviamo la probabilità che ci siano zero elettroni nel dot, che vale:

$$P(0) = \frac{1}{1 + \sum_{i=1}^N \prod_{j=1}^i a_j} \quad (3.3.16)$$

A questo punto è possibile calcolare tutti i valori delle probabilità $P(N)$, avendo calcolato $P(0)$, e avendo a disposizione gli a_i ottenuti dalle simulazioni che saranno iterate per un numero di volte pari all'ordine "N" individuato secondo la metodologia precedentemente esposta. Ovviamente ogni volta che si procede con una nuova simulazione, è indispensabile incrementare di uno il numero di elettroni nel dot, mantenendo costante la tensione di gate; il numero di elettroni nel dot è scritto nel file *caricad.out* che viene aggiornato durante l'esecuzione della script.

3.4 Soluzione della master equation nel dominio del tempo

La soluzione della master equation nel tempo ci serve per poter conoscere dopo quanto tempo la varie grandezze arrivano al valore di regime, importante per conoscere i tempi di programmazione della cella. Possiamo esprimere la derivata temporale in termini di rapporto incrementale:

$$\frac{dP(N,t)}{dt} = \frac{P(N,t+\Delta t) - P(N,t)}{\Delta t} \quad (3.4.1)$$

Ogni $P(N,t)$ è una funzione indipendente dalle altre e dipendente unicamente dal tempo.

Sostituendo nella master equation otteniamo:

$$P(N,t+\Delta t) = P(N,t) \left[1 - \Delta t (g(N+1) + r(N)) \right] + \Delta t \left[r(N+1)P(N+1,t) + g(N)P(N-1,t) \right] \quad (3.4.2)$$

Per ogni N è possibile valutare l'andamento temporale delle probabilità $P(N,t)$ facendo variare t secondo multipli dell'intervallo Δt di discretizzazione, ossia calcoliamo la probabilità agli istanti $i \cdot \Delta t$ con $i = 0, 1, 2, \dots$

Dato lo stato della trappola (N), è possibile valutare l'andamento temporale delle probabilità facendo variare t secondo multipli dell'intervallo Δt di discretizzazione : $t = i \cdot \Delta t$ con $i = 0, 1, 2, \dots$

Il Δt è legato ai coefficienti di generazione e ricombinazione; fissata un valore della tensione di gate la routine provvederà alla valutazione del valore del Δt , ovvero dell'unità temporale di riferimento utilizzata per la scansione temporale dell'evoluzione della cella:

$$\Delta t = \frac{1}{c} \min \left\{ \frac{1}{g}, \frac{1}{r} \right\} \quad (3.4.3)$$

Nella 3.3.3 si fa riferimento al massimo valore presente tra tutti i fattori di generazione e ricombinazione forniti dai files d'ingresso. Il fattore c presente al denominatore rappresenta una riduzione dell'intervallo di scansione dell'asse temporale che consente di individuare un Δt piccolo a sufficienza da garantire un comportamento delle curve indipendente dall'ampiezza di Δt stesso. Ovviamente il valore di c è individuato per

prove, andando a rappresentare le curve ottenute con l'esecuzione della routine *timeprob* e verificandone la loro sovrapposizione.

A questo punto, affinché il problema sia ben definito, è indispensabile fornire le condizioni iniziali, ovvero lo stato del dot all'istante $t = 0$ per ogni N :

$$\begin{cases} P(N, 0) = 0 & \text{per ogni } N \geq 1 \\ P(0, 0) = 1 \end{cases} \quad (3.4.4)$$

Il programma di simulazione, per risolvere la master equation nel dominio del tempo, genera una matrice con un numero di righe dato dal massimo numero di elettroni che occupano il dot e da un numero di colonne pari al numero di intervalli di tempo Δt compresi nell'intervallo di analisi T .

A questo punto è facile calcolare l'andamento temporale del numero medio di elettroni nel dot, che risulta essere pari a

$$\langle N(t) \rangle = \sum_{i=0}^N i \cdot P(i, t) \quad (3.4.5)$$

e di conseguenza la corrispondente traslazione temporale della tensione di soglia $\Delta V_T(t)$:

$$\Delta V_T(t) = k \cdot N(t) \quad (3.4.6)$$

dove il fattore k tiene conto della traslazione subita dalla tensione di soglia a causa dell'incremento del numero di elettroni nel dot. Il suo valore dipende anche dalle dimensioni del dispositivo che si considera e deriva matematicamente dall'integrazione monodimensionale della concentrazione del numero di portatori all'interno della regione di canale, ovvero dove è stato realizzato il calcolo della equazione di Schrödinger .

3.5 Programmi utilizzati per le simulazioni

Le routine utilizzate per la simulazione delle memorie a nanocristalli sono le stesse che abbiamo utilizzato per la simulazione dei MOSFET con gate stack alternativi, con alcune aggiunte che permettono di risolvere la master equation. In particolare la routine *principale* (che adesso si chiama *prnano*) aggiunge alcuni file in uscita e ne riceve alcuni in più in ingresso. Infatti questa riceve il file *dot.out* che serve a determinare la posizione spaziale della carica intrappolata nel nanocristallo; questa viene fissata nel punto centrale dello strato di nanocristalli .

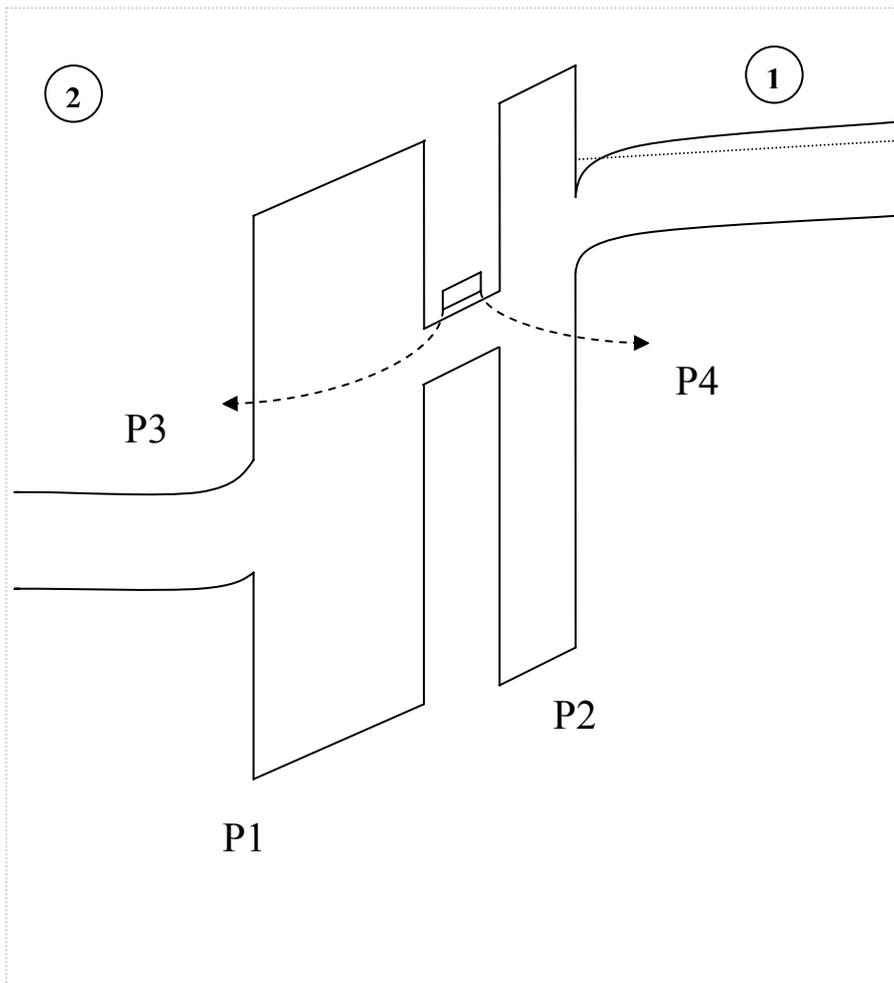


Figura 3.3: schema della struttura utilizzata in cui vengono mostrati i punti significativi per determinare la posizione della carica nel nanocristallo.

Viene calcolata successivamente la probabilità di trasmissione per effetto tunnel degli elettroni intrappolati nei nanocristalli attraverso le barriere di potenziale che gli stessi vedono verso il gate e verso il substrato.

In particolare il calcolo dei coefficienti di trasmissione è affidato alla procedura *tranmat*, che scandisce tutta la regione compresa tra $p3$ e $p4$ (punti che definiscono lo strato dei nanocristalli) mediante la variabile P_{int} , considerando per ogni passo di integrazione le barriere: $p1-P_{int}$ e $P_{int}-p2$.

La posizione energetica del livello di trappola nel nanocristallo viene definita dal file *enertrap.dat* in cui si forniscono la distanza, in termini energetici, del livello della trappola rispetto al fondo della banda di conduzione ($e1$ ed $e2$), la sezione di cattura (σ_{zero}), il delta che definisce la lorentziana (δ) e il numero di punti in cui vengono suddivisi l'intervallo energetico occupato dalle trappole e l'intervallo energetico totale; una ulteriore chiarificazione viene fornita dalla seguente figura.

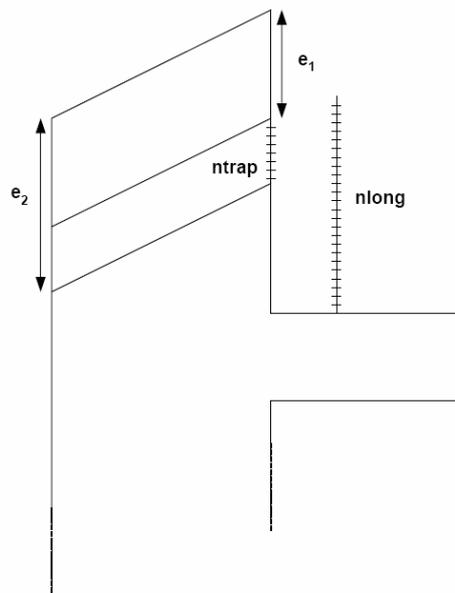


Figura 3.4: generica barriera di potenziale che serve per esemplificare alcune grandezze utilizzate dal file *enertrap.dat*.

Tra i file di uscita di *prnano* spiccano in termini di importanza operativa i seguenti files: *ai.out* relativo al rapporto tra la somma dei fattori di generazione $g_1 + g_2$ e la somma dei fattori di ricombinazione $r_1 + r_2$ ed i files *generation.out* e *recombination.out* che contengono rispettivamente la somma dei fattori di generazione ($g_1 + g_2$) e la somma dei fattori di ricombinazione ($r_1 + r_2$) relativi ad una simulazione.

Questi dati vengono utilizzati per la risoluzione della master equation; quest'ultima viene risolta in condizioni di regime dalla script Perl *prob.pl* e in condizioni dinamiche dalla script Perl *probtime.pl*. Queste ultime si basano sull'iterazione di due routine Fortran, rispettivamente *probabilità* e *timeprob* che risolvono la master equation per una determinata tensione di gate, saranno poi le script ad iterare il processo per analizzare l'intervallo di tensioni di interesse.

Questi dati sono necessari per le seguenti routine.

La routine *probabilita.f* ha in ingresso i seguenti files:

- *resoconto.out* che fornisce il valore della tensione di gate ed il numero di elettroni che occupano il nanocristallo (N);
- *aitot.out* che contiene i rapporti tra il fattore di generazione (g_1) e quello di ricombinazione (r_2) ottenuti dall'esecuzione di *Prnano* un numero di volte pari ad N , ovvero pari alla stima del numero di elettroni che occuperanno i dots con una certa tensione di gate.

Una volta estratti i dati di cui sopra, la routine procede innanzitutto con il calcolo della probabilità $P(0)$, che indica la probabilità che non ci siano elettroni nei dot rispettando l'espressione 3.2.16.

Immediatamente dopo viene effettuato il calcolo delle probabilità relative alla presenza di un numero di elettroni compresi tra 1 ed N , come previsto dall'equazione 3.2.12; successivamente viene calcolato il numero medio di elettroni e il valore della traslazione della tensione di soglia:

$$N = \sum_{n=0}^{\infty} n \cdot P(n) \quad (3.5.1)$$

$$\Delta V_T = k \cdot N \quad (3.5.2)$$

Come files di uscita si ottengono:

- *ennemedio.out*, che contiene il numero medio di elettroni che occupano i nanocristalli in corrispondenza della tensione di gate considerata;
- *deltav.out*, che contiene il valore della traslazione subita dalla tensione di soglia a causa della presenza di un numero di elettroni pari ad N all'interno dei nanocristalli;
- *probtot.out*, che contiene i valori delle probabilità calcolate nella routine.

E' importante sottolineare il ruolo assunto dalla script *prob.pl* che definisce i parametri d'ingresso di probabilità, ovvero il numero di elettroni che con probabilità non trascurabile possono, mediante i processi di generazione e ricombinazione, andare ad occupare i nanocristalli. Il principio su cui si basa l'algoritmo di definizione del parametro N è immediato in quanto tiene conto che la probabilità relativa alla presenza dell'elettrone i -esimo $P(i)$ differisce da quella che la precede di un fattore coincidente con a_m quindi nel momento in cui tale fattore sarà sufficientemente minore di uno (0.001 è una buona valutazione), il valore della probabilità considerato sarà sicuramente trascurabile rispetto a tutti gli altri valutati precedentemente; considerando che al crescere del numero di

elettroni, il valore degli a_m generati nelle successive simulazioni diminuiscono in valore. Si ricorda che i termini calcolati in prima istanza, ovvero con la sola conoscenza degli a_m , saranno valori probabilistici normalizzati rispetto a $P(0)$, in quanto quest'ultimo non è ancora noto. Ovviamente questo non toglie validità a quanto ora espresso.

La risoluzione della master equation nel dominio del tempo è affidata alla routine *timeprob.f*. I files d'ingresso di questa routine sono:

- *generation.out*, che viene generato dall'esecuzione di *prmano* all'interno della script *prob.pl* per la soluzione della master equation a regime e contenente i fattori di generazione calcolati per un dato valore della tensione di gate. Ricordiamo che il numero dei fattori contenuti è pari al numero di elettroni che con probabilità non trascurabile occuperanno i nanocristalli (N).
- *recombination.out*, che ha le stesse caratteristiche del file *generation.out* ma contenente i fattori di ricombinazione.
- *righe.dat* che definisce il numero di righe della matrice che verrà utilizzata nella routine per il calcolo delle probabilità $P(N,t)$; in particolare sarà pari ad N .
- *intervallo.dat* che contiene il numero degli intervalli di tempo Δt che compongono l'intervallo di tempo di analisi del comportamento della cella.

L'esecuzione di *timeprob* fa riferimento ad un valore della tensione di gate, di conseguenza è necessaria una script *proptime.pl* che provvede ad iterare l'esecuzione della routine per tutto l'intervallo di tensioni di interesse.

Il calcolo delle funzioni $P(N,t)$ viene effettuato considerando una matrice formata da N righe e con un numero di colonne che è pari ad un intero b , contenuto all'interno del file

intervallo.dat. Per poter avere una valutazione dell'intera evoluzione del comportamento temporale della cella dagli istanti iniziali sino alla condizione di regime, può essere necessario considerare un numero di colonne molto elevato. Siccome non è possibile realizzare matrici con un numero di colonne troppo elevato, è necessario utilizzare una matrice di appoggio in cui è possibile effettuare il calcolo relativamente ad un numero di colonne p e poi riportare solo l'ultima calcolata nella matrice principale, in modo tale da potervi riportare una colonna ogni p . La variabile che gestisce questo intero p viene fissata nella script *probtme.pl* e si chiama *zum* in quanto svolge una vera e propria funzione di zoom della curva lungo l'asse dei tempi. In questo modo sarà anche possibile ottenere uno snellimento della matrice di calcolo a parità di ampiezza massima dell'intervallo d'analisi considerato, oppure, a parità di colonne, aver la possibilità di approssimarsi meglio alla condizione di regime.

Durante la costruzione della matrice principale, che avverrà per colonne grazie alla particolare espressione della master equation è necessario verificare la condizione di convergenza dell'algoritmo; per ogni nuovo elemento di una colonna, la routine calcola la somma di tutti gli elementi della colonna stessa aggiornando la somma di volta in volta col nuovo elemento. La condizione da verificare è che tale somma non superi mai l'unità, per le proprietà delle probabilità; se questo si verifica l'algoritmo viene bloccato e viene consigliata la scelta di un Δt più piccolo.

Successivamente la routine procede col calcolo del numero medio di elettroni e della traslazione della tensione di soglia conseguente (entrambe le grandezze sono dipendenti dal tempo), come prescritto dalle equazioni 3.3.5 e 3.3.6.

Come files di uscita *timeprob* presenta:

- *VTtemp.out*, che contiene l'andamento temporale della traslazione della tensione di soglia relativa ad un valore della tensione di gate.
- *ennetemp.out*, che contiene l'andamento temporale del numero di elettroni intrappolati nei nanocristalli a seguito dell'applicazione di una tensione di gate.
- *sigma.out*, che contiene l'andamento temporale dello scarto quadratico medio del numero di elettroni rispetto al valore medio. Questo file fornisce come informazione la variabilità delle caratteristiche temporali del dispositivo e quindi dà la possibilità di valutare le prestazioni della cella nel caso peggiore e migliore. Da qui derivano i files relativi agli scostamenti del numero medio di elettroni e della traslazione della tensione di soglia: *VT+sigma.out*, *VT-sigma.out*, *enne+sigma.out*, *enne-sigma.out*.

3.6 Risultati delle simulazioni

Le simulazioni che abbiamo effettuato sulle memorie a nanocristalli si sono svolte su due strutture con uno strato di nanocristalli di raggio medio 4 nm e densità $1 \cdot 10^{12} \text{ dot/cm}^2$; le due strutture differiscono per lo spessore dell'ossido tunnel e dell'ossido di controllo. Una struttura, che indicheremo con 3-4-7, presenta un ossido tunnel di 3 nm e un ossido di controllo di 7 nm, l'altra struttura, che indicheremo con 4-4-8, presenta un ossido tunnel di 4 nm e un ossido di controllo di 8 nm. Nella figura seguente vengono riportate le due strutture utilizzate:

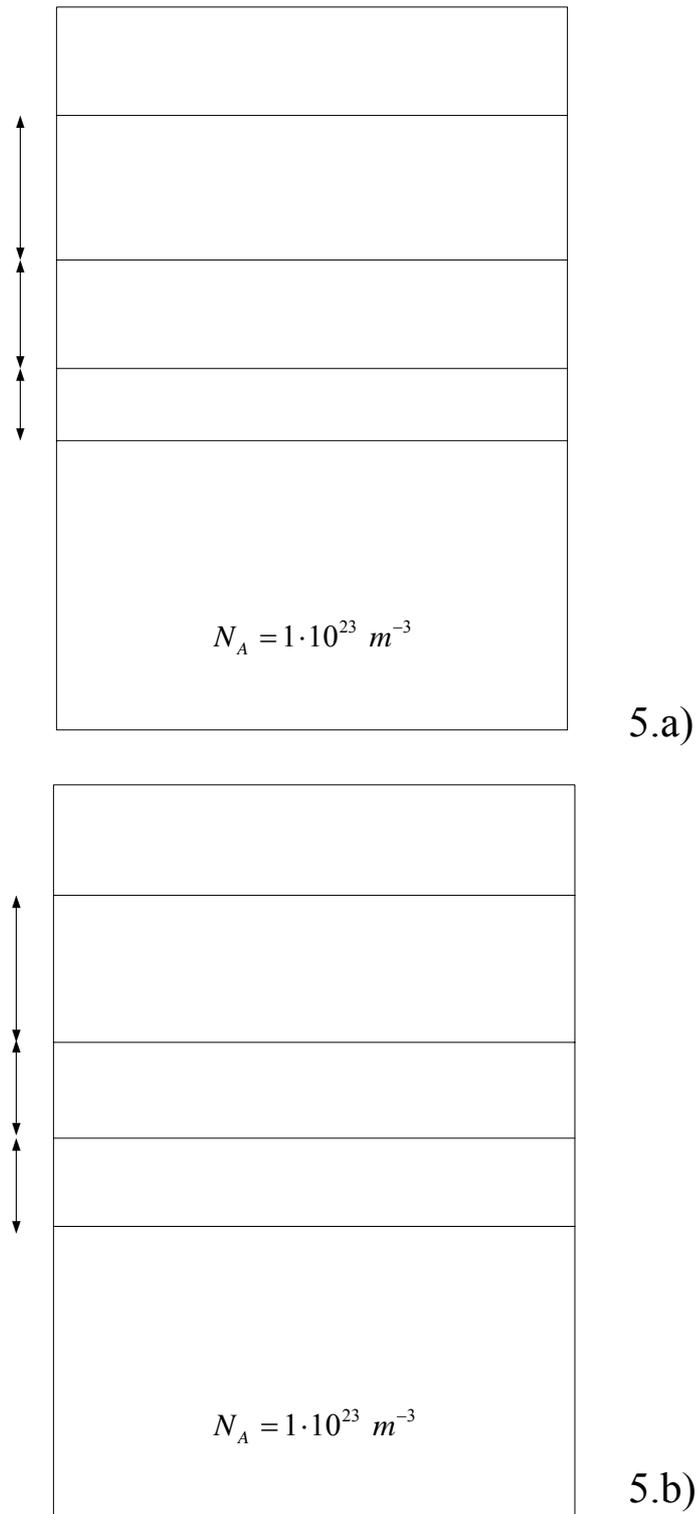


Figura 3.5: strutture utilizzate per le simulazioni delle memorie a nanocristalli; sono evidenziati gli spessori dei vari strati e il drogaggio del substrato.

Lo strato di 4 nm in cui sono immersi i nanocristalli è in realtà composto anche da un forte percentuale di ossido di silicio; per tener conto di questo fatto nel nostro modello 1D abbiamo variato la costante dielettrica dello strato. Possiamo pensare, in primissima approssimazione, che la costante dielettrica vari linearmente dalla costante dielettrica dell'ossido di silicio (3.9) a quella del silicio (11.8), in funzione della percentuale di silicio nello strato.

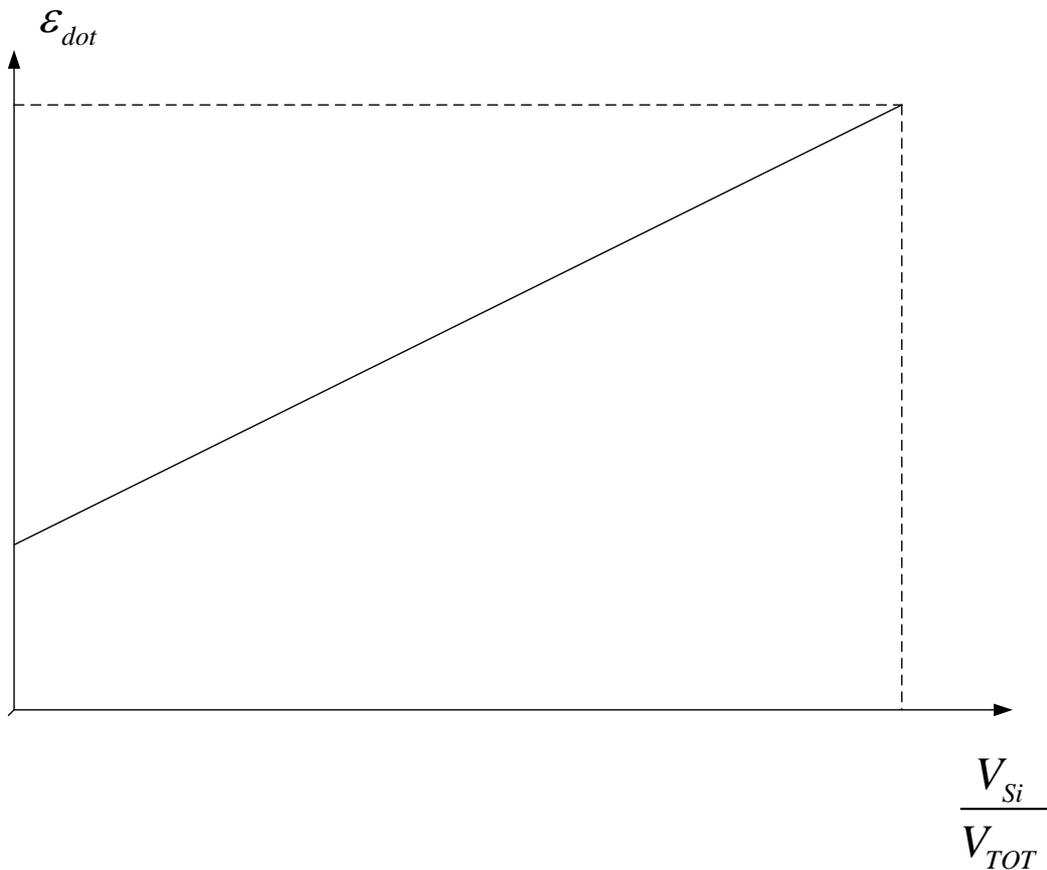


Figura 3.6: andamento della costante dielettrica dello strato di nanocristalli al variare della percentuale di silicio nello strato di nanocristalli.

Calcoliamo adesso il rapporto tra il volume totale e quello occupato dai nanocristalli per 1 cm^2 , supponendo gli stessi di forma sferica con raggio di 2 nm.

$$V_{dot} = \frac{4}{3} \pi r^3 \cdot 10^{12} \text{ dot/cm}^2 \quad (3.6.1)$$

$$V_{TOT} = 1 \text{ cm}^2 \cdot 4 \text{ nm} \quad (3.6.2)$$

Possiamo ricavarci allora la costante dielettrica dello strato di nanocristalli, come

$$\varepsilon_{strato} = \varepsilon_{SiO_2} + (\varepsilon_{Si} - \varepsilon_{SiO_2}) \frac{V_{Si}}{V_{TOT}} = 4.56 \quad (3.6.3)$$

Tutte le altre caratteristiche fisiche dello strato di nanocristalli sono state assunte uguali al silicio puro.

Un altro parametro importantissimo, o meglio il più importante, per la simulazione del nanocristallo è la densità di carica da aggiungere allo strato di nanocristalli per ogni elettrone nel dot. Classicamente questa sarebbe uguale a

$$\rho = \frac{\sigma}{d} = \frac{1 \cdot 10^{-16} \text{ dot/m}^2}{4 \cdot 10^{-9} \text{ m}} = 2.5 \cdot 10^{24} \text{ m}^{-3} \quad (3.6.4)$$

Quella introdotta è una densità di carica non moltiplicata per q perché questo è il valore da passare in ingresso al file *inpup.form*, la moltiplicazione viene eseguita all'interno della routine *struttura*. Dovremo avere anche l'accortezza di moltiplicare per -1 visto che la moltiplicazione in *struttura* avviene per $1.602 \cdot 10^{-19}$.

Il valore così ottenuto di $-2.5 \cdot 10^{24}$ non è però corretto perché non tiene conto dei fenomeni quantistici che avvengono nel dot. Abbiamo preso come dato affidabile una simulazione 3D sull'aumento del potenziale elettrochimico del dot in seguito all'inserimento di un elettrone; il valore che abbiamo preso è di $-4.3 \cdot 10^{24}$ perché è il valore di carica fissa che aumenta di 0.28 eV, ossia l'aumento del potenziale

elettrochimico, l'energia della banda di conduzione quando si passa da zero ad un elettrone per $V_G = 2 \text{ V}$.

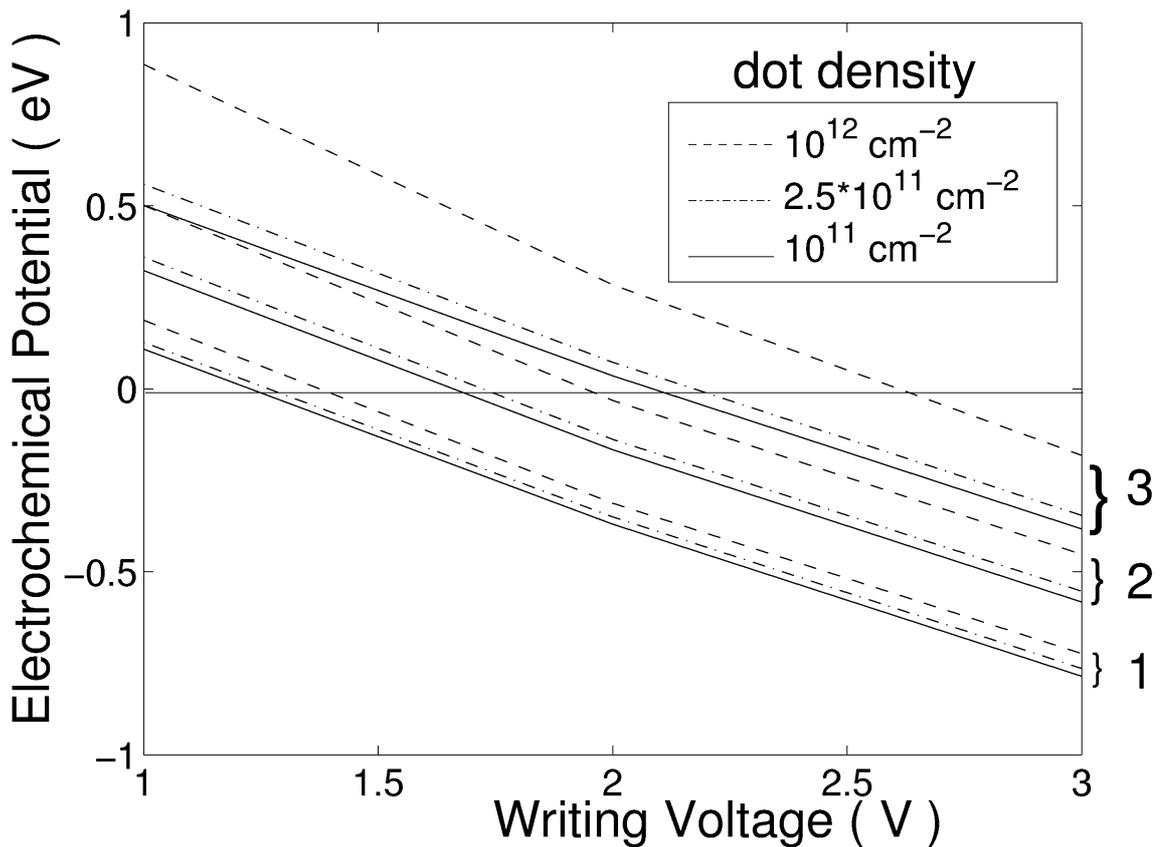


Figura 3.7: andamento del potenziale elettrochimico in funzione della tensione di gate applicata al variare del numero di elettroni nel dot; il nostro riferimento è la curva con densità 10^{12} cm^{-2} ad una tensione di 2 V.

Nella figura seguente possiamo vedere l'effetto dell'iniezione di elettroni nel dot andando a vedere come si modifica il diagramma a bande della struttura; in figura si vede l'effetto dell'iniezione di cinque elettroni. Le bande sono riferite alla struttura 448 e per una tensione di 3V.

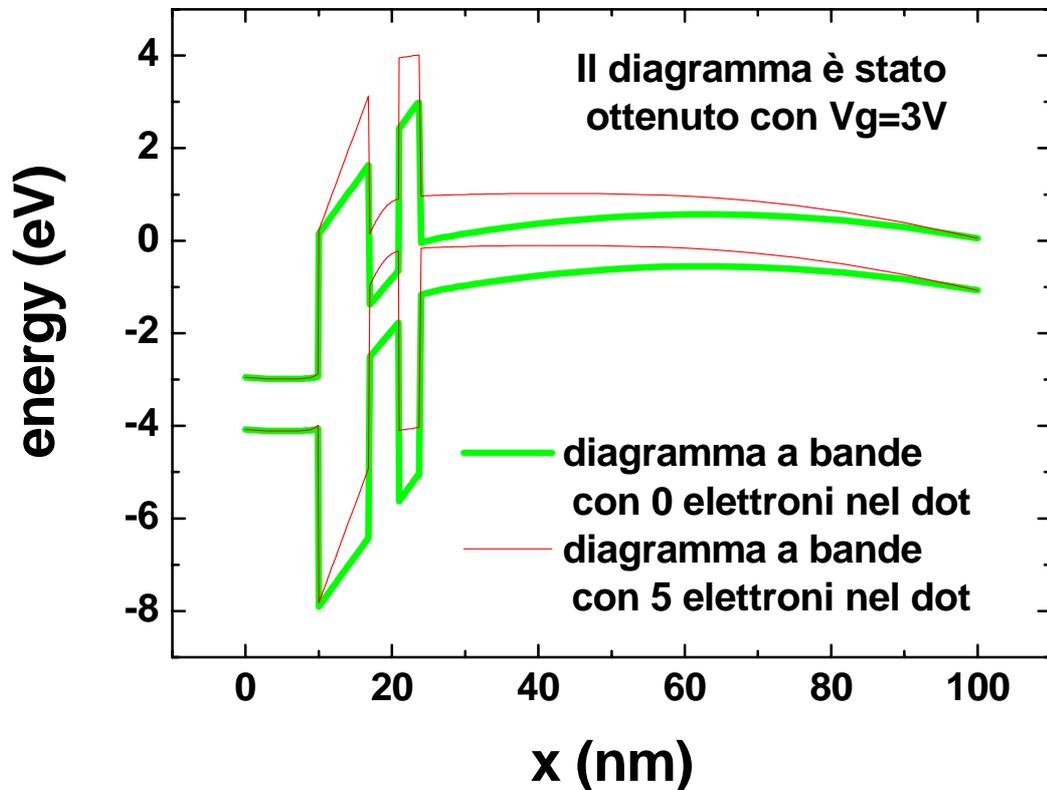


Figura 3.8: *effetto dell'iniezione di cinque elettroni nella struttura a bande di una memoria a nanocristalli.*

Abbiamo quindi impostato i parametri più importanti, tranne il fattore k di scaling per trovare la tensione di soglia media; il valore di k per le due strutture sarà diverso in quanto gli ossidi tunnel hanno dimensioni differenti e quindi ci sarà una diversa entità dello screening della tensione di gate da parte nanocristalli carichi. Matematicamente k deriva dall'integrazione monodimensionale della concentrazione del numero di portatori all'interno della regione di canale, ovvero dove è stato realizzato il calcolo della equazione di Schrödinger. I valori che si trovano sono di 0.396 V per elettrone per la struttura 3-4-7, mentre per la struttura 4-4-8 si trova un valore di 0.4523 V per elettrone.

Le nostre simulazioni sono iniziate calcolando il numero medio di elettroni nel dot per le varie tensioni di gate applicate; il risultato di queste simulazioni di regime dipende fortemente dalla densità di carica inserita nel dot per ogni elettrone e soprattutto dalla posizione del livello energetico della trappola. Se posizioniamo la trappola sul livello fondamentale del dot, cioè 0.2732 eV sopra la banda di conduzione del silicio, otteniamo un massimo della curva per tensioni basse rispetto ai dati sperimentali.

Inizieremo il nostro studio con la struttura 3-4-7; mostriamo quindi il numero medio di elettroni nel dot per tensioni che vanno da 0 a 6V, ottenuto dalla script *prob.pl* con i parametri che abbiamo indicato sopra.

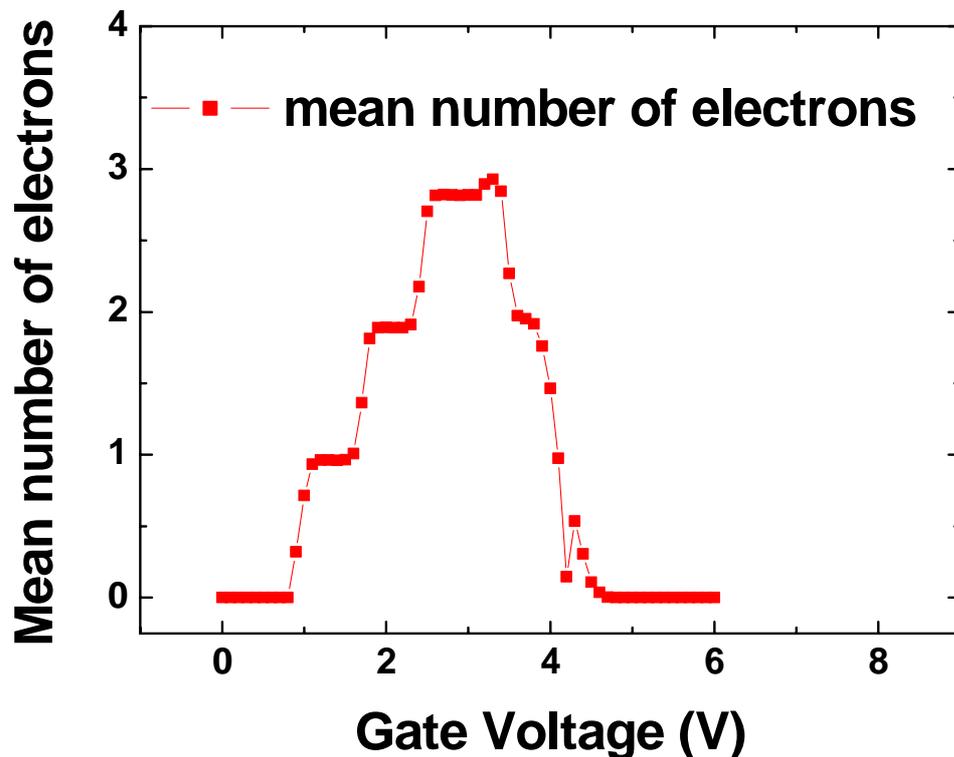


Figura 3.8: andamento del numero medio di elettroni nella struttura 3-4-7; la simulazione è relativa al livello fondamentale del dot, ossia per una energia 0.2732 eV sopra la banda di conduzione dello strato di nanocristalli.

La tensione di soglia media si ottiene semplicemente dalla moltiplicazione del numero medio di elettroni per la costante k , che per la struttura 3-4-7 vale 0.396 Volt per elettrone.

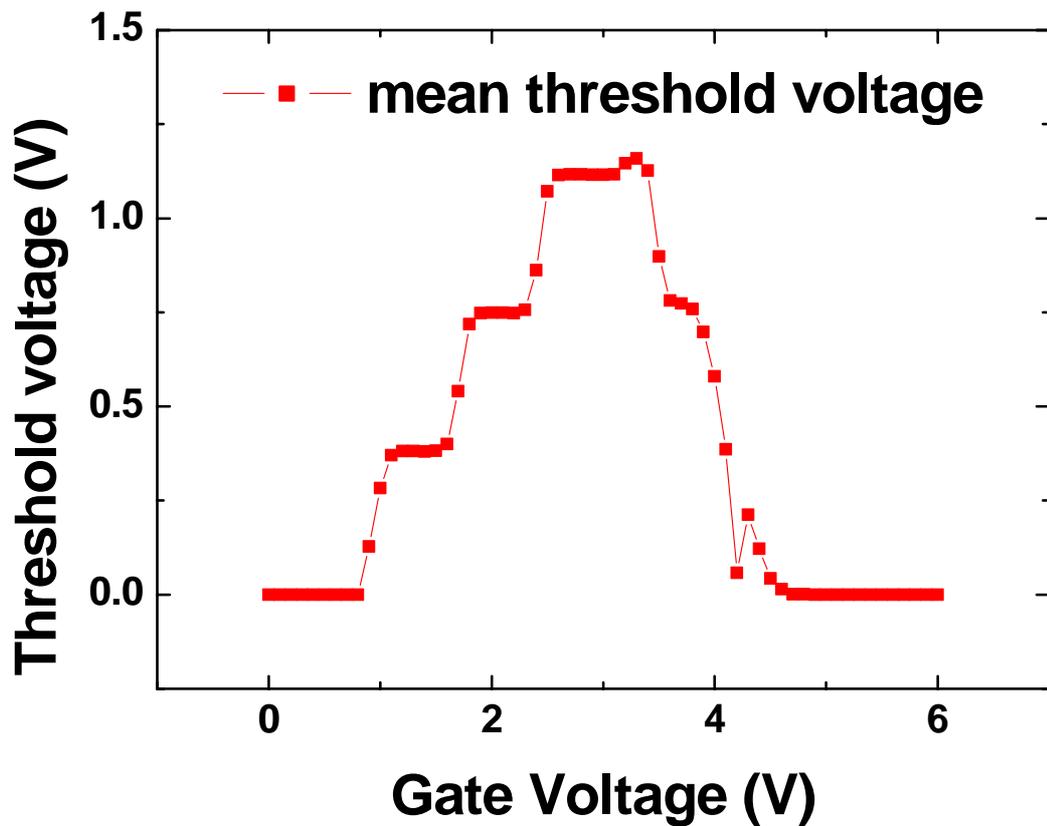


Figura 3.9: andamento della tensione di soglia media nella struttura 3-4-7; la simulazione è relativa al livello fondamentale del dot, ossia per una energia 0.2732 eV sopra la banda di conduzione dello strato di nanocristalli.

A questo punto è necessario far riferimento ai dati sperimentali, per dare una valutazione ai risultati ottenuti mediante le simulazioni; faremo riferimento alla seguente figura che fa capo all'articolo [16] di I. Crupi, D. Corso, G. Ammendola et al.

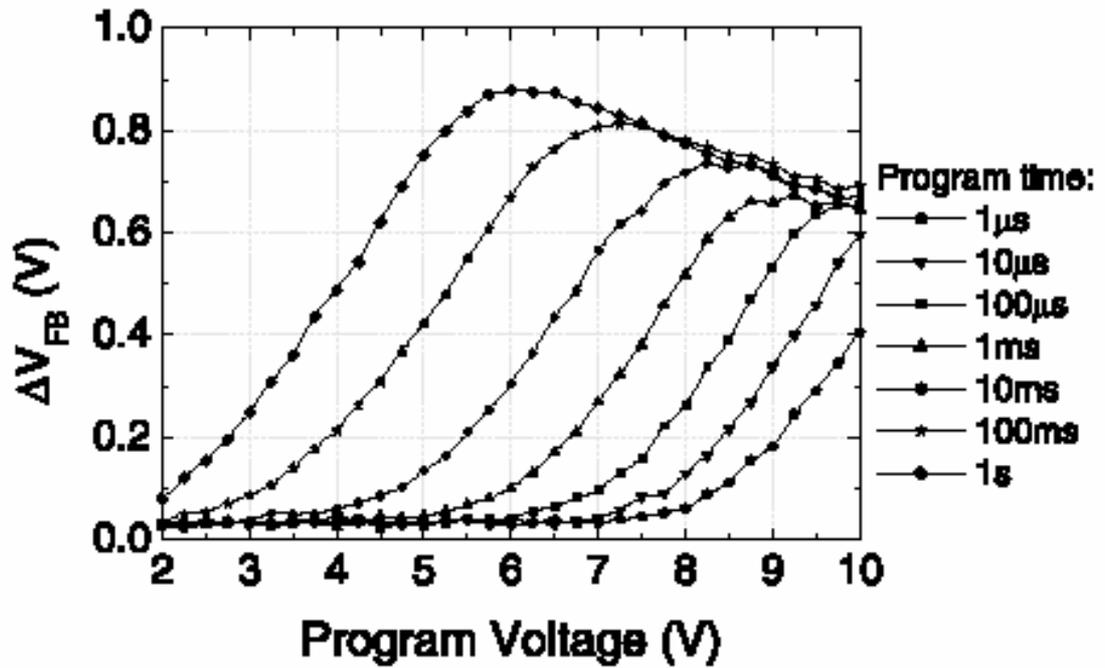


Figura 3.10: valutazione sperimentale della tensione di soglia per una struttura 3-4-7 con densità di nanocristalli di $1 \cdot 10^{12}$ dot/cm².

I risultati sperimentali si riferiscono a tempi diversi, per adesso ci limiteremo alla caratteristica di regime, ossia quella misurata per $t = 1\ \text{s}$; in seguito studieremo anche le altre curve quando passeremo allo studio del transitorio.

La prima simulazione è stata realizzata considerando un livello energetico di $0.273\ \text{eV}$, ovvero quello relativo al livello fondamentale introdotto dal dot, ottenuto mediante calcoli quantistici rappresentando ciascun nanocristallo come un elemento cubico di spigolo pari a 4 nm. Una differenza fondamentale è rappresentata dal valore della tensione di gate che individua il massimo della curva; in particolare nelle valutazioni sperimentali si individua per una tensione di 6 V, mentre la simulazione lo individua in corrispondenza di una tensione che è di poco superiore a 3 V.

Queste prime considerazioni hanno suggerito l'ipotesi che gli elettroni possano occupare livelli profondi, ovvero nel gap del silicio, invece del livello fondamentale introdotto dal dot. Infatti si nota un forte incremento delle curve con l'abbassamento del livello energetico responsabile della cattura gli elettroni; i livelli profondi nel gap del silicio sono pensabili come trappole superficiali dovute, per esempio, alla ruvidità superficiale del nanocristallo o a legami non saturi, come si verifica ad ogni interfaccia silicio/ossido di silicio.

Un'altra differenza che emerge con le curve sperimentali è la pendenza con cui decrescono le curve subito dopo aver raggiunto il massimo; nelle nostre simulazioni questo avviene molto velocemente, indice di una forte ricombinazione, mentre le curve sperimentali decrescono più lentamente. Nella figura seguente viene mostrata una curva sperimentale in cui si può notare una ricombinazione più forte rispetto a quella della figura 3.10, molto più simile alla pendenza con cui decrescono le nostre curve simulate[17].

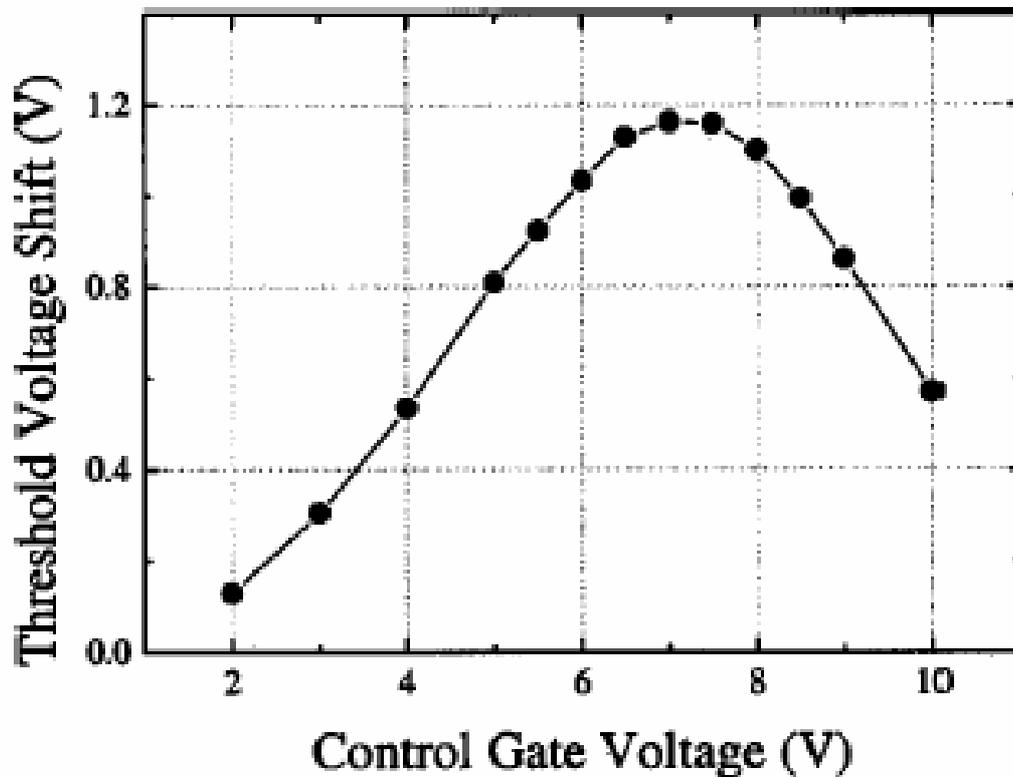


Figura 3.11: andamento della tensione di soglia a regime; la figura è tratta dal riferimento [1] e mostra un'a ricombinazione più forte dopo il massimo.

Dobbiamo tener conto che le curve dipendono in maniera critica da molti parametri di processo, come gli spessori dei due strati di ossido, oppure la concentrazione di nanocristalli, per non dimenticare che l'ipotesi di sfericità dei nanocristalli sarà difficilmente verificata nella pratica. Tutti questi fattori di non idealità tolgono un minimo di consistenza ai nostri risultati quantitativi, senza però togliere generalità ai risultati quantitativi che vogliamo ottenere.

Il passo successivo è quindi quello di provare a considerare dei livelli energetici profondi, ovvero livelli che energeticamente si posizionano all'interno del gap del silicio, oppure addirittura in banda di valenza. Ci aspettiamo, a parità di struttura e a parità di tensione di

gate applicata, che il massimo della tensione di soglia, ovverosia del numero medio di elettroni che occupano il dot, si sposti verso tensioni più alte a causa del fatto che livelli più profondi comportano un minor tasso di ricombinazione verso il gate. Infatti quest'ultimo è quello preponderante per alte tensioni di programmazione ed è proporzionale al coefficiente di trasmissione della barriera che l'elettrone incontra per andare nel gate; ovviamente per livelli profondi aumenta l'altezza della barriera e quindi diminuisce il tasso di ricombinazione.

Per basse tensioni, invece, il fattore di ricombinazione principale è quello verso il substrato, ovvero per basse tensioni gli elettroni vengono iniettati dal substrato e verso lo stesso tendono a ricombinarsi, senza però rimanere nel dot. Quindi anche per basse tensioni aumenterà il numero di elettroni intrappolati nel dot visto che una volta iniettati, trovano una barriera di altezza maggiore per ricombinarsi verso il substrato.

Per quanto riguarda il fattore di ricombinazione, possiamo pensare che gli elettroni vengano generati sempre e soltanto dal substrato verso il dot, mentre possiamo trascurare, senza commettere grossi errori, il fattore di generazione dal gate verso il dot.

Le figure seguenti mostrano, rispettivamente, numero medio di elettroni nel dot e il valore medio della tensione di soglia, sempre per la struttura 3-4-7.

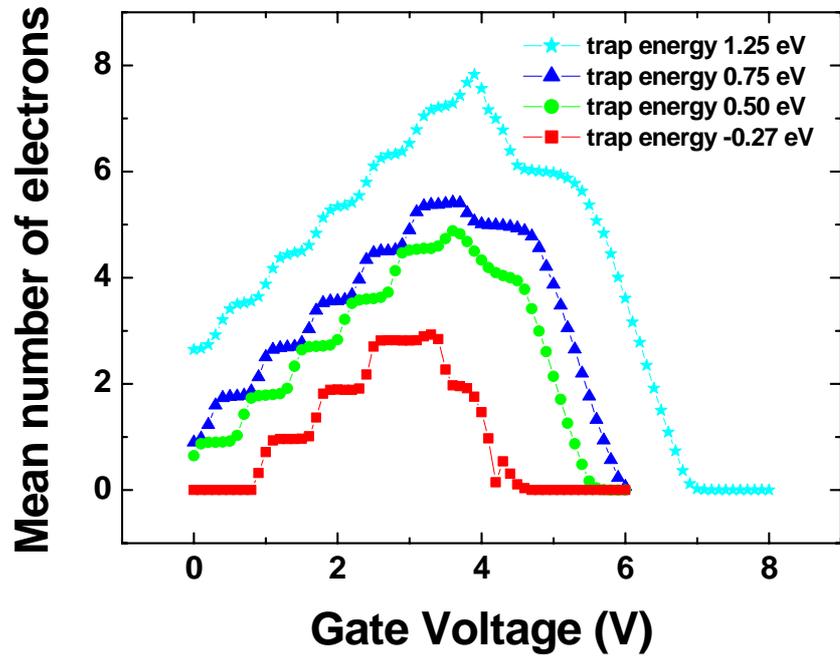


Figura 3.12: andamento del numero medio di elettroni nel dot al variare dell'energia della trappola.

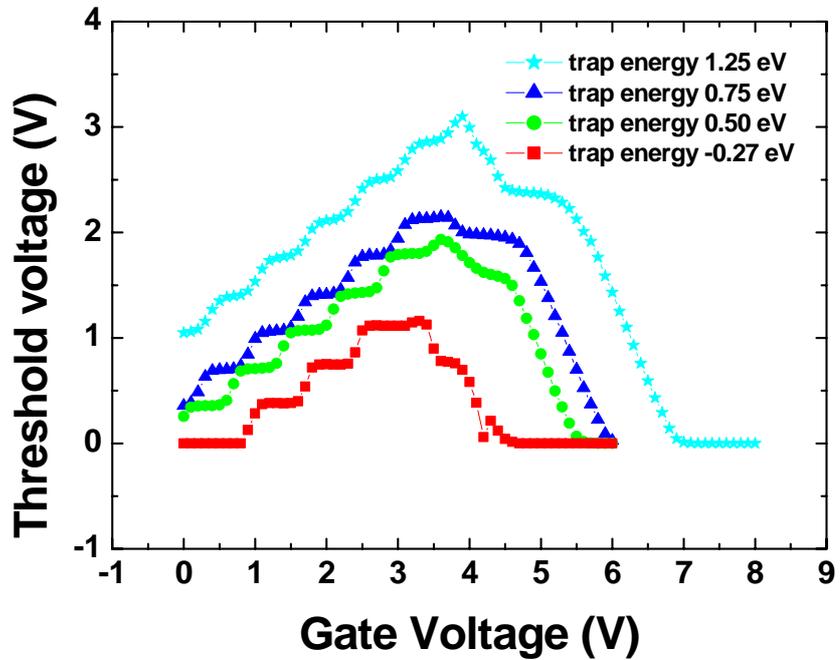


Figura 3.13 andamento della tensione di soglia media nel dot al variare dell'energia della trappola.

Nelle figure mostrate sino ad ora delle nostre simulazioni, si nota la presenza di plateau molto accentuata; in molti lavori, vedi ad esempio De Blauwe [18], si è misurata e ipotizzata la presenza di plateau nella tensione di soglia. In realtà questo comportamento è sicuramente verificato all'interno del singolo nanocristallo, che può contenere un numero intero di elettroni, ma dobbiamo pensare che in una memoria reale ci saranno una serie di nanocristalli, con raggi che non sono tutti uguali, ma che saranno distribuiti con un certo valor medio e una certa varianza. Inoltre la forma dei nanocristalli non sarà perfettamente sferica, così come altre caratteristiche non saranno perfettamente rispettate in tutti i nanocristalli. Questo comporta la scomparsa dei plateau nelle curve misurate, visto che la curva finale la possiamo pensare come la sovrapposizione di tante curva ognuna relativa ad un nanocristallo diverso. È certo che una memoria “perfetta” potrebbe realmente presentare un comportamento fatto da una serie di plateau in sequenza.

Un altro fattore importante di cui dobbiamo tener conto è la natura della trappola che presumibilmente catturerà l'elettrone; se le nostre simulazioni dipendessero in maniera critica dalle caratteristiche della trappola, come per esempio il delta che definisce la lorentziana in energia (vedi 3.15), non potremmo ottenere risultati generali ma solo applicabili ad un certo tipo di trappole; lo stesso si potrebbe pensare per la sezione di cattura che in realtà non influisce sui risultati di regime e su quelli dinamici. Infatti i rapporti g/r che definiscono le probabilità di occupazione non dipendono dalla sezione di cattura visto che compare a moltiplicare sopra e sotto, quindi nel rapporto sparisce. Anche nella master equation nel dominio del tempo la sezione di cattura finisce a moltiplicare tutti i fattori di generazione e ricombinazione, ossia determina soltanto una traslazione delle

curva tempo-varianti. Questa dipendenza può essere sfruttata per l'estrazione della sezione di cattura tramite confronto con le curve sperimentali.

La figura seguente mostra la dipendenza del numero medio di elettroni nel dot dal delta della trappola, per una struttura 3-4-7 e per una energia della trappola di 0.50 eV, che significa che la trappola è a metà del gap del silicio (le energie della trappola sono positive se sono inferiori al limite della banda di conduzione del silicio e negative se prese sopra di essa, come nel caso del livello fondamentale ottenuto da considerazioni quantistiche, che risulta -0.273).

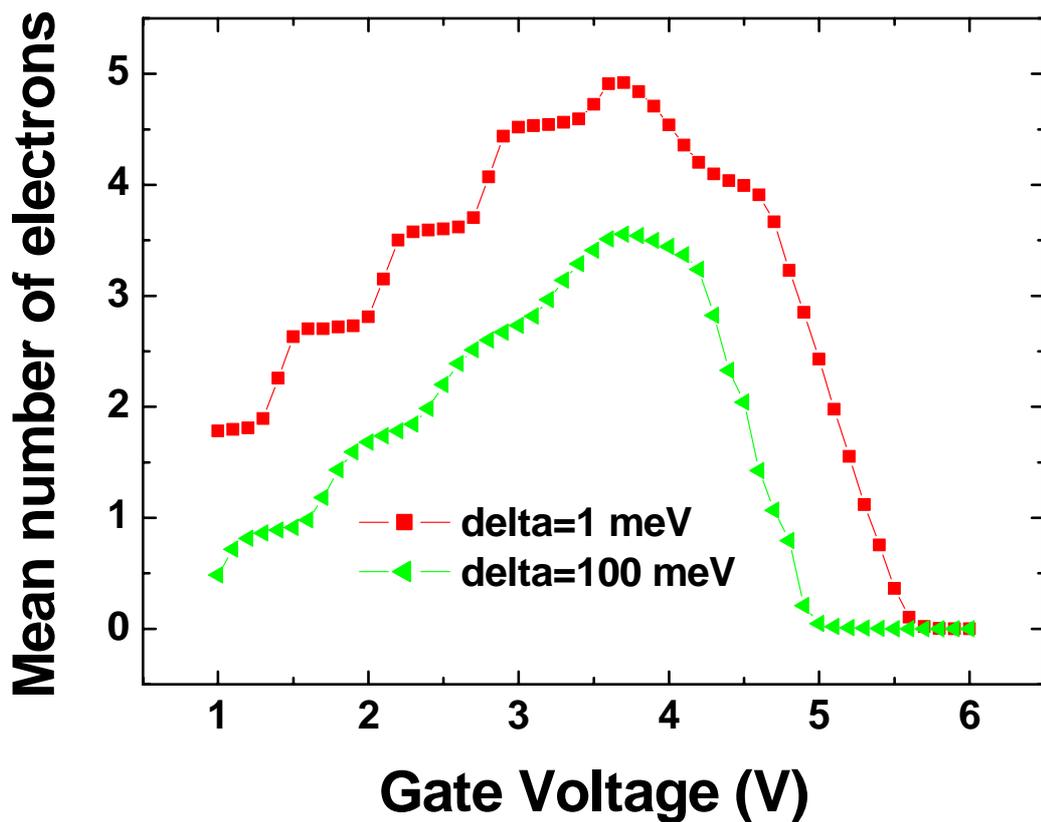


Figura 3.14: dipendenza del numero medio di elettroni nel dot dal delta della trappola.

Possiamo dire, generalmente, che basta scegliere un delta inferiore all'intervallo che si usa per schematizzare il livello energetico della trappola: nel nostro caso il livello della trappola è stato fissato tra 0.5 e 0.501 quindi al di sotto di 1 meV non abbiamo variazioni delle curve, come invece avviene per delta superiori. Il delta maggiore dell'intervallo energetico rappresentante la trappola comporta un diverso intervallo di energie interessate ai fenomeni di generazione e ricombinazione, infatti la lorentziana diventa più larga dell'intervallo energetico ed è quindi questa, e non il livello della trappola, a influire maggiormente sui tassi di generazione e ricombinazione.

Per chiarire maggiormente quanto abbiamo detto poco sopra riguardo alla dipendenza delle curve dai parametri geometrici della struttura, come, per esempio, gli spessori dell'ossido di controllo e dell'ossido tunnel, oppure del raggio dei nanocristalli, mostriamo di seguito le simulazioni effettuate al variare di questi parametri.

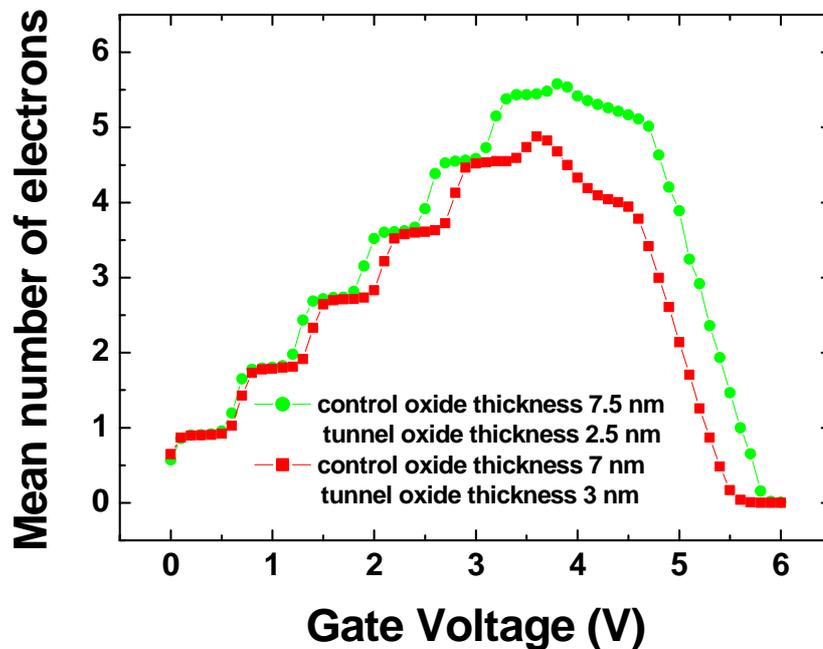


Figura 3.15: dipendenza del numero medio di elettroni dallo spessore degli strati di ossido.

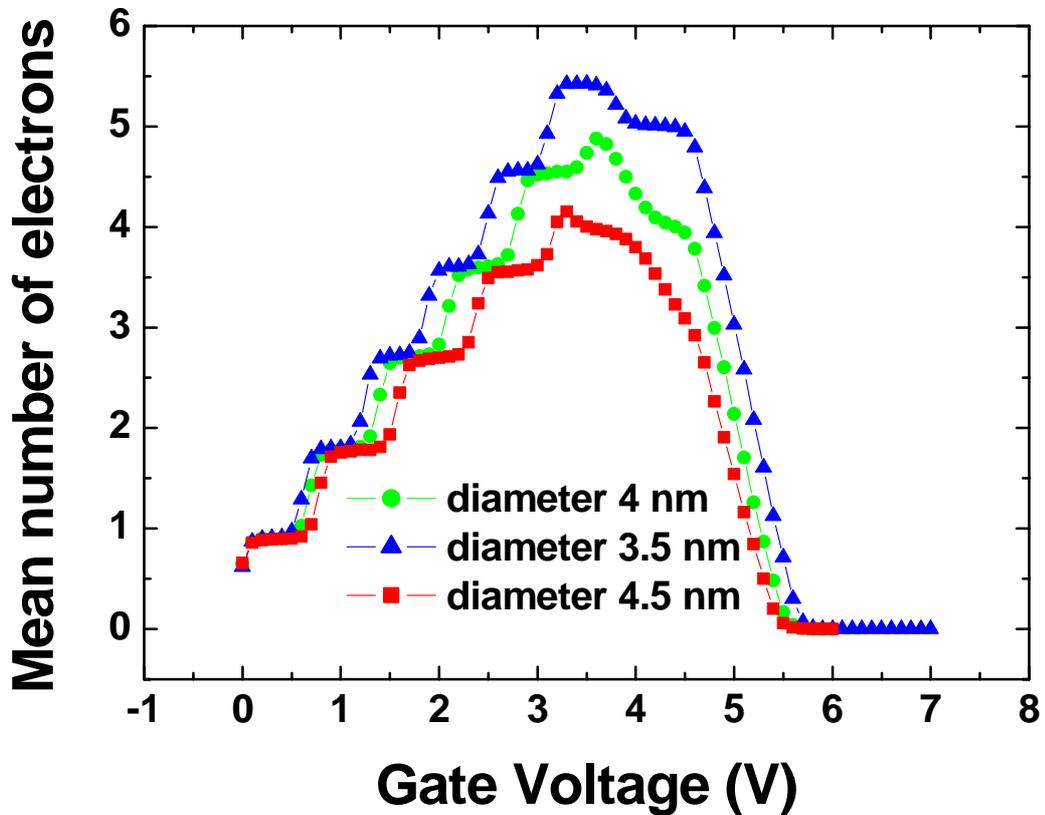


Figura 3.16: dipendenza del numero medio di elettroni nel dot dal diametro dei nanocristalli; in pratica variare il diametro dei nanocristalli equivale a cambiare la costante dielettrica dello strato di nanocristalli come indicato nell'equazione 3.5.3

Le nostre simulazioni si sono estese anche alla struttura 4-4-8 che presenta, rispetto alla struttura 3-4-7, un minor numero di elettroni intrappolati nel dot a causa degli spessori di ossido maggiori. Il comportamento qualitativo è comunque lo stesso, con il classico andamento a “campana” ossia un primo tratto lentamente crescente in cui si notano i plateau e un tratto discendente ripidamente, più di quanto non si riscontri nei risultati sperimentali evidenziati precedentemente.

Senza troppe esitazioni presentiamo il numero medio di elettroni nel dot e la tensione di soglia media per la struttura 4-4-8.

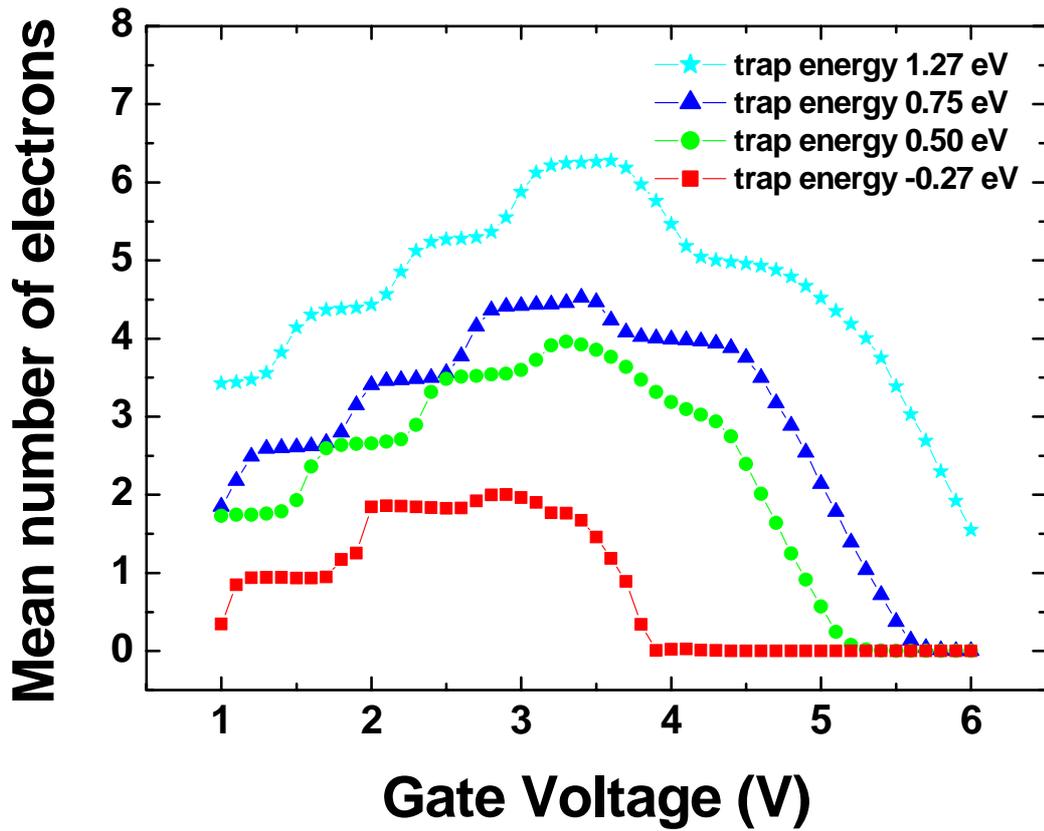


Figura 3.17: numero medio di elettroni nel dot per la struttura 4-4-8; si vede che, in confronto alla struttura 3-4-7, si ha un minor numero di elettroni intrappolati nel dot anche se le forme qualitative sono identiche.

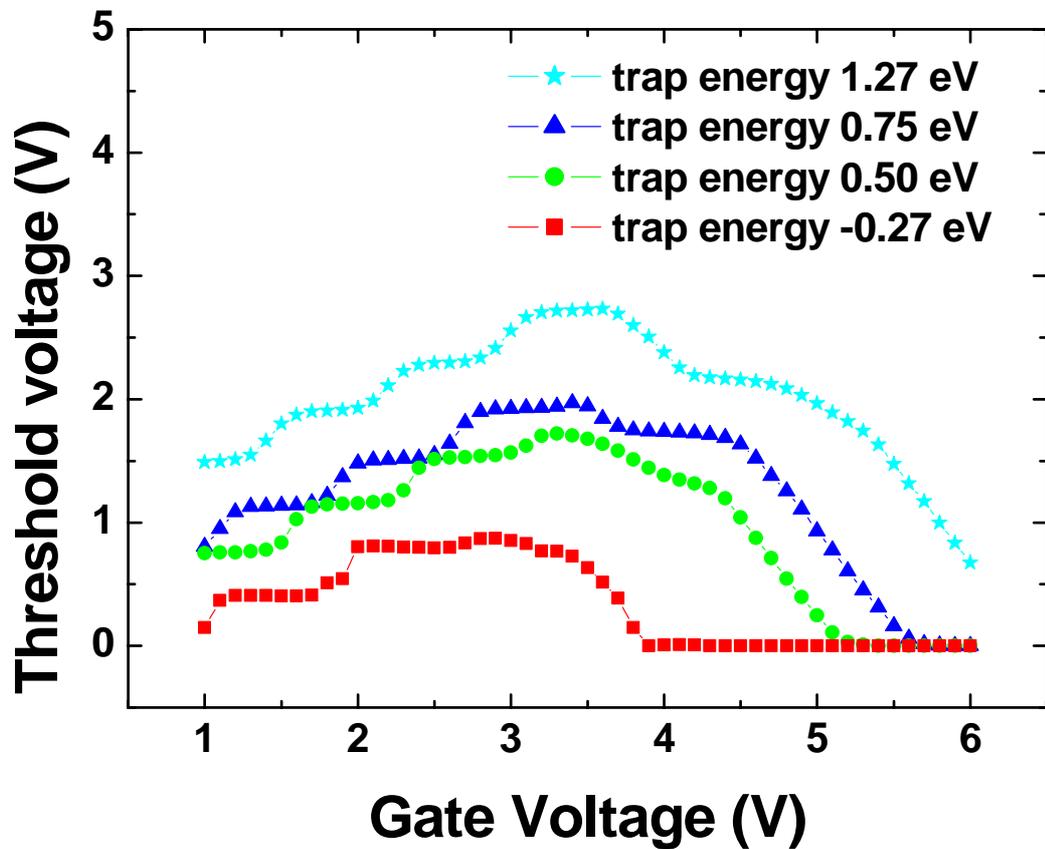


Figura 3.18: *valor medio della tensione di soglia per la struttura 4-4-8; si vede che, in confronto alla struttura 3-4-7, si ha una tensione di soglia più bassa.*

Passiamo adesso a presentare le ultime simulazioni che abbiamo effettuato sulle memorie a nanocristalli, che riguardano le variazioni temporali delle varie grandezze coinvolte, in particolare il numero medio di elettroni e il valore della tensione di soglia. Non riporteremo entrambe le simulazioni visto che le curve sono le stesse scalate del fattore k quindi riporteremo le curve per le tensioni di soglia che si possono estendere, *mutatis mutandis*, al valor medio degli elettroni nel dot.

Nella figura seguente si può vedere l'andamento del valor medio della tensione di soglia al variare del tempo, che va considerato in unità arbitrarie visto che non abbiamo informazioni sulla sezione di cattura del livello energetico nel dot.

La simulazione è relativa ad una struttura 3-4-7 con livello energetico della trappola posto a $-0.273 eV$ sopra la banda di conduzione del silicio.

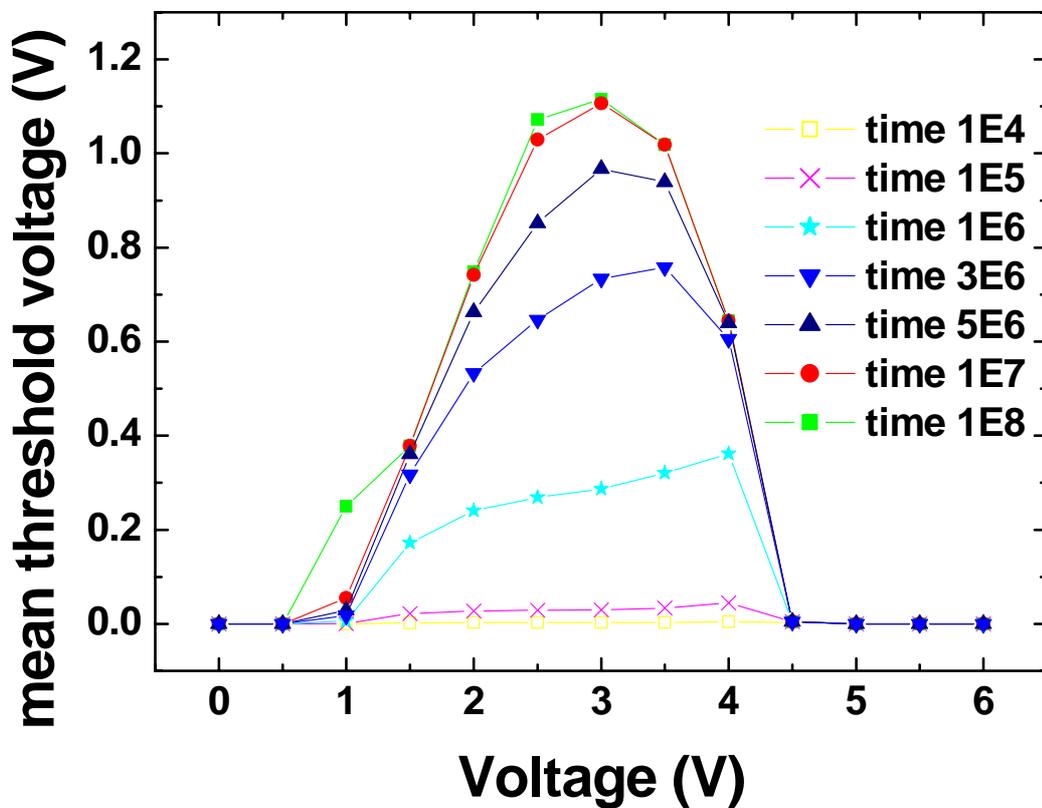


Figura 3.19: andamento del valor medio della tensione di soglia per diversi istanti di programmazione della cella; al tempo (in unità arbitrarie) di $1E13$ siamo in condizioni di regime.

Le stesse simulazioni sono state fatte anche per la struttura 4-4-8, considerando lo stesso livello energetico, ossia $0.273 eV$ sopra la banda di conduzione del silicio; si può notare

come i tempi di programmazione siano significativamente superiori a causa dei maggiori spessori dell'ossido tunnel e di quello di controllo.

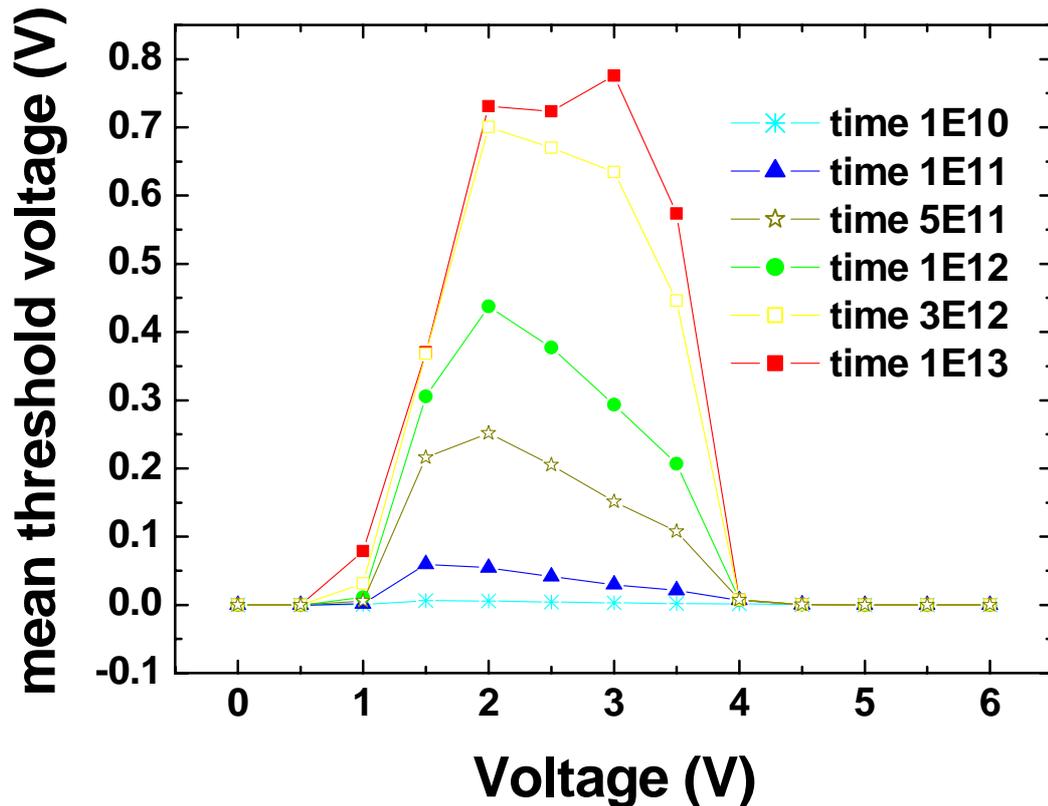


Figura 3.20: andamento del valor medio della tensione di soglia al variare del tempo di programmazione.

Una volta individuati gli andamenti temporali delle grandezze medie quali il numero di elettroni nel dot e la variazione della tensione di soglia, è utile considerare l'andamento temporale dello scostamento delle stesse dai rispettivi valori medi. In particolare si deve valutare lo scarto quadratico medio che, per il numero di elettroni intrappolati, assume la seguente forma:

$$\sigma(t)_n = \sqrt{\sum_0^{n'} (n - \langle n(t) \rangle)^2 \cdot P(n,t)} \quad (3.6.5)$$

in cui ogni valore istantaneo di $\sigma(t)_n$ è ricavato dalla somma su tutti gli n che vanno da zero fino a N_{MAX} (valore che contraddistingue il numero di righe della matrice di calcolo), e dove i $P(n,t)$ sono gli elementi della nostra matrice, in particolare i termini relativi ad una stessa colonna. Per lo scarto quadratico medio rispetto alla tensione di soglia c'è da moltiplicare per la costante k , quindi si ottiene:

$$\sigma(t)_{V_T} = k \cdot \sigma(t)_n \quad (3.6.6)$$

L'informazione che otteniamo dal calcolo delle curve di scostamento dal valore medio è il tempo minimo, medio e massimo che ci vuole per programmare la cella, corrispondenti, rispettivamente, al caso migliore ($\Delta V_T + \sigma(t)_{V_T}$), al caso medio (ΔV_T) e al caso peggiore ($\Delta V_T - \sigma(t)_{V_T}$).

Il valore di $\sigma(t)_{V_T}$ dovrebbe essere il più piccolo possibile in modo da minimizzare l'incertezza sui tempi di programmazione; nella figura seguente si riporta una simulazione effettuata per $V_G = 3 V$ su una struttura 4-4-8, che ci serve per dare un'idea del valore dello scarto quadratico medio.

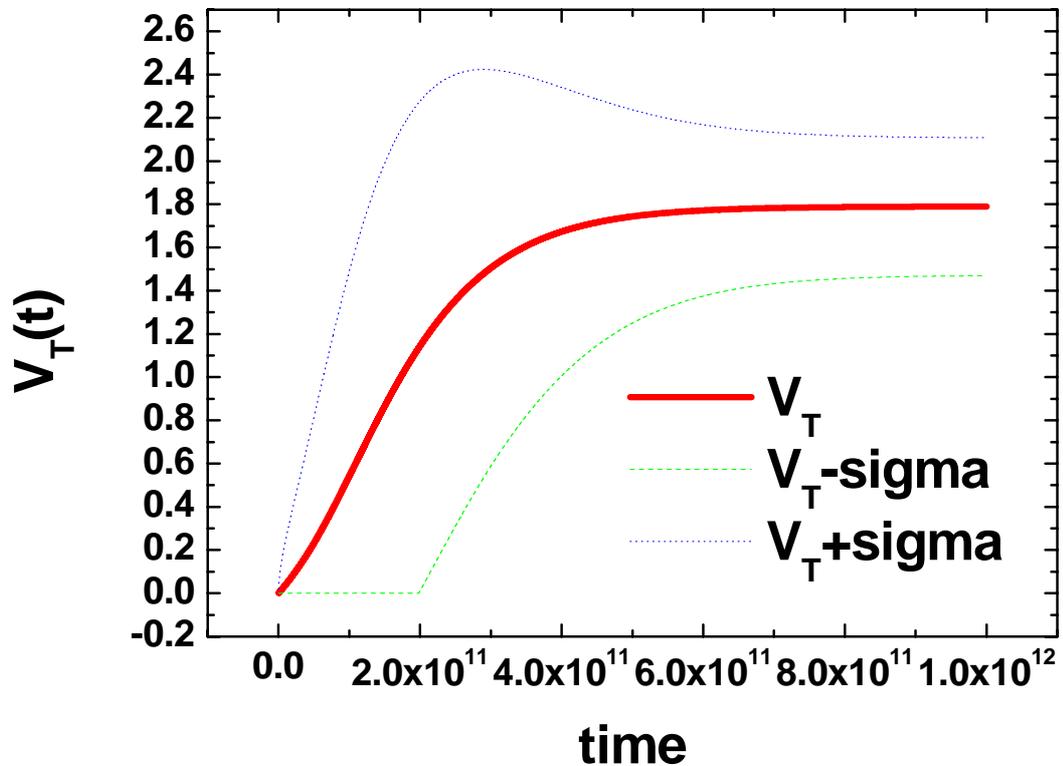


Figura 3.21: andamento nel tempo della tensione di soglia e curve relative allo scarto quadratico medio rispetto ad essa.

Per concludere questo capitolo riportiamo i listati dei principali file d'ingresso utilizzati per le simulazioni con lo scopo di facilitare il lavoro a chiunque debba studiare e simulare queste strutture, così come abbiamo fatto per i gate stack nel capitolo 2.

Riportiamo i listati del file *grild.dat*, di *inputp.form* e della script perl *prob.pl* che chiarifica quanto abbiamo detto a proposito della soluzione della master equation in condizioni stazionarie.

3.7 Listati dei principali file d'ingresso

GRI1D.DAT

10	numero di regioni
1	regione1
0	xiniziale
5	xfinale
3	delta iniziale
0.7	delta finale
2	regione2
5	xiniziale
10	xfinale
0.7	delta iniziale
0.1	delta finale
3	regione3
10	xiniziale
15	xfinale
0.1	delta iniziale
0.5	delta finale
4	regione4
15	xiniziale
17	xfinale
0.5	delta iniziale
0.2	delta finale
5	regione5
17	xiniziale
18.5	xfinale
0.2	deltainiziale
0.05	deltafinale
6	regione6
18.5	xiniziale
19.5	xfinale
0.05	deltainiziale
0.05	deltafinale
7	regione7
19.5	xiniziale
21	xfinale
0.05	deltainiziale
0.1	deltafinale
8	regione8
21	xiniziale
24	xfinale
0.1	deltainiziale
0.3	deltafinale
9	regione9
24	xiniziale
50	xfinale
0.3	deltainiziale
2	deltafinale
10	regione10
50	xiniziale
100	xfinale
2	deltainiziale
10	deltafinale

INPUTP.FORM

```

300      temperatura[K]
{vg}     Efermi gate di controllo(-V applicata)[eV]
4.1      Funzione lavoro gate di controllo[eV]
0        Efermi contatto di bulk[eV]
4.1      Funzione lavoro contatto di bulk[eV]
{ril}    coefficiente sottorilassamento
{nor}    tolleranza per la norma2[V]
{fp}     flagpot(se 1 usa il pot.rif. altrimenti sol. precedente)
5        numero di regioni
1        regione1
1        tipo di materiale(poly)
1        flagn
1        flagp
0        Xiniziale[nm]
10       Xfinale[nm]
0        concentrazione Na[m**.-3]
{ndg}    concentrazione Nd[m**.-3]
0        carica fissa
1        flagf(se 1 Ef=Ef control gate altrimenti Ef=Ef bulk)
2        regione2
2        tipo di materiale(SiO2)
0        flagn
0        flagp
10       xiniziale
17       xfinale
0        concentrazione Na[m**.-3]
0        concentrazione Nd[m**.-3]
0        carica fissa
0        flagf
3        regione3
12       tipo di materiale (Si per nanocristallo)
0        flagn
0        flagp
17       xiniziale
21       xfinale
0        concentrazione Na[m**.-3]
0        concentrazione Nd[m**.-3]
{ro}     carica fissa
0        flagf
4        regione4
2        tipo di materiale (SiO2)
0        flagn
0        flagp
21       xiniziale[nm]
24       xfinale[nm]
0        concentrazione Na[m**.-3]
0        concentrazione Nd[m**.-3]
0        carica fissa

```

Capitolo 3: Memorie a nanocristalli

0	flagf
5	regione5
1	tipo di materiale (bulk)
1	flagn
1	flagp
24	xiniziale[nm]
100	xfinale[nm]
{na}	concentrazione Na[m** ⁻³]
0	concentrazione Nd[m** ⁻³]
0	carica fissa
0	flagf
1	regioni dove calcolare Schroedinger
1	regione1
1	fc
23.8	xiniziale
50	xfinale
1	barriere di potenziale
1	barriera1
10	xiniziale
24	xfinale
1	flagzs
1	flagzd
0	flagrandom

PROB.PL

```
#!/usr/bin/perl
#SIMULAZIONE DI REGIME PER UNA MEMORIA A NANOCRISTALLI
chdir('/home/andrea/tesi/nanocristalli') or die "errore directory: $?";
use File::Copy;
#vengono assegnati i valori di tensione di gate
$sup=1;
$inf=1;
$step=0.5;
#viene assegnato il drogaggio del gate
$ndg=1e26;
#viene assegnato il drogaggio del substrato
$na=5e23;
#viene assegnato il coefficiente di sottorilassamento
$sj=0.03;
#viene assegnata la tolleranza per la norma euclidea
$sk=0.00001;
#viene imposto l'uso di un potenziale esterno come valore di partenza
$z=0;
unlink('resocontotot.out');
unlink('ennemtot.out');
unlink('vfb.out');
#il file resoconto contiene per ogni tensione il numero di elettroni
open(u,'>>resocontotot.out') or die "non apre resocontotot.out:$?";
#ennem.out è aperto in append e conterrà il valore di vg affiancato da quello di #ennemedio
open(x,'>> ennemtot.out') or die "non apre ennemtot.out:$?";
#il file vfb.out conitene il valore di vg affiancato dalla variazione della tensione di #soglia presente
#nel file deltav.out
open(dd,'>>vfb.out') or die "non apre vfb.out:$?";
```

Capitolo 3: Memorie a nanocristalli

```
#esecuzione di griglia
system('griglia')==0 or die "errore griglia: $?";
$vg=$inf;
#ciclo sulle tensioni di gate
#open(f,'>>generic.out') or die "non apre generic.out:$?";
#open(v,'>>integra.out') or die "non apre integra.out:$?";
while ($vg<=$sup) {
    $delta=0;
    $dot=0;
    $i=0;
    $alert=0;
    unlink('aitot.out');
    unlink('generation.out');
    unlink('recombination.out');
    unlink('resoconto.out');
    open(as,'>>aitot.out') or die "non apre aitot.out:$?";
    open(abc,'>>generation.out') or die "non apre generation.out:$?";
    open(ac,'>>recombination.out') or die "non apre recombination.out:$?";
#esecuzione di struttura e prnano
    while ($delta==0) {
        open(d,'>str1d.dat') or die "non apre str1d.dat: $?";
        open(c,'<inputp.form') or die "non apre inputp.form: $?";
        while($r=<c>) {
            $r=~s/{vg}/-$vg/;
            $r=~s/{ril}/$j/;
            $r=~s/{nor}/$k/;
            $r=~s/{fp}/$z/;
            $r=~s/{ndg}/$ndg/;
            $r=~s/{ro}/$dot/;
            $r=~s/{na}/$na/;
            print d $r;
        }
        close(c);
        close(d);
        open(ca,'>caricad.out') or die "non apre caricad.out: $?";
        print ca $i;
        close(ca);
        print "----\n";
        print "----\n";
        print "SIMULA CON Vg=$vg e N = $i elettroni\n";
        print "----\n";
        print "----\n";
        system('struttura')==0 or die "errore struttura: $?";
        system('prnano')==0 or die "errore prnano: $?";
#prnano viene eseguito due volte se step è 0.5 V per garantire la convergenza
        system('prnano')==0 or die "errore prnano: $?";
        open(ag,'<gii.dat') or die "non apre gii.dat: $?";
        @ah[$i]=<ag>;
        open(ae,'<rii.dat') or die "non apre rii.dat: $?";
        @af[$i]=<ae>;
        if ($i==1) {
            $aa=@ah[$i-1]/@af[$i];
            print as "$aa \n";
            print abc @ah[$i-1];
            print ac @af[$i];
        }
    }
}
```

```
        if ($aa<=0.001)
        { $alert=1;}
        }
    if ($i>1) {
        $bb=@ah[$i-1]/@af[$i];
        if ($bb < 0.001)
        { $alert=1;}
        $cc= $bb / $aa;
        if ($cc < 0.001 && $bb<0.001)
        { $delta=1;}
        print as "$bb \n";
        print abc @ah[$i-1];
        print ac @af[$i];
        if ($alert==1 && $i>=3)
        { $delta=1;}
        if ($i==12)
        { $delta=1;}
        $aa=$bb;
        }
        $i+=1;
        $dot=$i*(-4.3e24);
    close(ag);
    close(ae);
#ora sto definendo il file contenente i valori g/r=ai per un dato valore di vg da 0 a dot #elettroni
    }
    print abc @ah[$i-1];
    close(abc);
    close(ac);
    close(as);
    close(o);
    copy("aitot.out", "ai$vg.out");
    copy("generation.out", "generation$vg.out");
    copy("recombination.out", "recombination$vg.out");
#ciclo per la definizione di resoconto1.out che mi dice per ogni tensione
# quanti elettroni stanno nel dot
        $jj=$i-1;
        print u $vg;
        print u ' ';
        print u "$jj\n";
#genero resoconto per poterlo dare in ingresso al file probabilita che ha bisogno
#di aprirlo per associarlo ad una unita ,ma il file e' già in uso nel perl
#il modulo copy non funziona(speriamo bene)
    open(ab,'>resoconto.out') or die "non apre resoconto.out:$?";
        print ab $vg;
        print ab ' ';
        print ab $i-1;
    close(ab);
    open(ao,'>elettroni.out') or die "non apre elettroni.out: $?";
    print ao $i-1;
    close(ao);
    copy("elettroni.out", "elettroni$vg.out");
    system('probabilita')==0 or die "errore probabilita: $?";
    open(a,'<ennemedio.out') or die "non apre ennemedio.out :$?";
#ora mi accingo a definire il file ennem che contiene per ogni tensione il valore
#della stessa seguita da quello di ennemedio
```

```
copy("pzero.out","pzero$vg.out") || warn "Could not copy files: $!";
copy("probtot.out","probtot$vg.out") ||warn "Could not copy files: $!";
while ($gg=<a>) {
    print x $vg;
    print x $gg;
}
open(ee,'<deltav.out') or die "non apre deltav.out :$?";
while ($ff=<ee>) {
    print dd "$vg $ff";
#    print dd $ff;
}
close(a);
close(ee);
#fine del ciclo
$vg += $step;
}
close(x);
close(u);
close(dd);
```

APPENDICE A

Teoria delle perturbazioni

Il fatto che il coefficiente di trasmissione di una barriera tunnel sia molto piccolo, come si vede dalla equazione 1.4.46, suggerisce che l'uso della teoria delle perturbazioni per il tunneling è molto appropriata. Parleremo adesso della teoria delle perturbazioni e inizieremo dalle perturbazioni indipendenti dal tempo.

A.1 Perturbazioni indipendenti dal tempo

Pochi problemi consentono una soluzione esatta in meccanica quantistica, ma in molti casi c'è un termine di hamiltoniana H' piccolo rispetto ad un H_0 , tale che il problema con l'hamiltoniana non perturbata H_0 sia esattamente risolubile. In questi casi si usa la teoria delle perturbazioni: il metodo consiste nello sviluppare sia la funzione d'onda che l'energia in serie di potenze di H' .

Il problema da risolvere è quindi:

$$H\Psi = (H_0 + H')\Psi = E\Psi \quad (\text{A.1.1})$$

Per ipotesi è nota la soluzione del problema non perturbato

$$H_0 \Psi_n^{(0)} = E_n^{(0)} \Psi_n^{(0)} \quad (\text{A.1.2})$$

dove abbiamo indicato con l'apice (0) i termini non perturbati e con il pedice n l' n -esimo livello energetico.

Supporremo che i livelli non perturbati $E_n^{(0)}$ siano discreti e non degeneri.

Una funzione d'onda qualsiasi può essere sviluppata in termini di $\{\Psi_m^{(0)}\}$:

$$\Psi = \sum_m c_m \Psi_m^{(0)}, \text{ oppure in notazione di Dirac } |\Psi\rangle = \sum_m c_m |m\rangle$$

ovviamente i ket $|m\rangle$ si riferiscono agli autostati non perturbati.

Sostituendo quest'ultima espressione nella equazione per l'hamiltoniana, si ha:

$$\sum_m c_m (E_m^{(0)} + H') |m\rangle = \sum_m c_m E |m\rangle \quad (\text{A.1.3})$$

Proiettiamo i due membri sul k-esimo stato (prodotto scalare con il bra $\langle k|$):

$$c_k E_k^{(0)} + \sum_m c_m H'_{km} = E c_k \quad (\text{A.1.4})$$

che è ancora esatta.

Vogliamo studiare la correzione all'energia e alla funzione d'onda di uno stato particolare

(ad esempio l'n-esimo); sviluppiamo ora:

$$\begin{aligned} E &= E^{(0)} + E^{(1)} + E^{(2)} + \dots \\ c_m &= c_m^{(0)} + c_m^{(1)} + c_m^{(2)} + \dots \end{aligned} \quad (\text{A.1.5})$$

dove $E^{(k)}, c_m^{(k)}$ sono proporzionali a $(H')^k$; sostituendo nella (A.1.4) si ha:

$$(c_k^{(0)} + c_k^{(1)} + c_k^{(2)} + \dots) E_k^{(0)} + \sum_m (c_m^{(0)} + c_m^{(1)} + c_m^{(2)} + \dots) H'_{km} = (E^{(0)} + E^{(1)} + E^{(2)} + \dots) (c_k^{(0)} + c_k^{(1)} + c_k^{(2)} + \dots)$$

All'ordine zero troviamo:

$$c_k^{(0)} (E_k^{(0)} - E^{(0)}) = 0 \quad (\text{A.1.6})$$

e poiché vogliamo studiare le correzioni all'n-esimo stato avremo:

$$c_k^{(0)} = \begin{cases} 1 & \text{se } k = n \\ 0 & \text{se } k \neq n \end{cases} \quad (\text{A.1.7})$$

Allora troviamo:

$$E^{(0)} = E_n^{(0)} \quad (\text{A.1.8})$$

Al primo ordine, invece, troviamo:

$$c_k^{(1)} E_k^{(0)} + \sum_m c_m^{(0)} H'_{km} = E_n^{(0)} c_k^{(1)} + E^{(1)} c_k^{(0)} \quad (\text{A.1.9})$$

che, per $k = n$, dà:

$$E^{(1)} = H'_{nn} = \langle n | H' | n \rangle \quad (\text{A.1.10})$$

Abbiamo trovato un risultato importante: lo spostamento dell'energia dell' n -esimo livello dovuto alla perturbazione è l'elemento diagonale n -esimo della matrice associata all'operatore H' .

Per $k \neq n$ si ha:

$$c_k^{(1)} (E_n^{(0)} - E_k^{(0)}) = H'_{kn}, \text{ da cui } c_k^{(1)} = \frac{H'_{kn}}{E_n^{(0)} - E_k^{(0)}}.$$

Il coefficiente $c_n^{(1)}$ rimane indeterminato; possiamo trovarlo dalla condizione di normalizzazione della funzione d'onda, cioè dalla condizione che la funzione d'onda al primo ordine sia normalizzata.

$$\Psi_n^{(0)} + \Psi_n^{(1)} = (1 + c_n^{(1)}) \Psi_n^{(0)} + \sum_{k \neq n} c_k^{(1)} \Psi_k^{(0)} \quad (\text{A.1.11})$$

Otteniamo quindi:

$$\left| 1 + c_n^{(1)} \right|^2 + \sum_{k \neq n} \left| c_k^{(1)} \right|^2 = 1 \quad (\text{A.1.12})$$

che implica $c_n^{(1)} = 0$.

Ricapitolando abbiamo trovato, al primo ordine della perturbazione, l'energia:

$$E_n = E_n^{(0)} + E_n^{(1)} \text{ con } E_n^{(1)} = H'_{nn} \text{ e la funzione d'onda } \Psi_n = \Psi_n^{(0)} + \sum'_k \frac{H'_{kn}}{E_n^{(0)} - E_k^{(0)}} \Psi_k^{(0)}.$$

L'apice al simbolo di sommatoria significa che manca il termine per $k = n$ nella sommatoria stessa.

Dall'ultima equazione segue anche la condizione di applicabilità del metodo:

$$|H'_{kn}| \ll |E_n^{(0)} - E_k^{(0)}| \quad (\text{A.1.13})$$

Procediamo adesso con il secondo ordine:

$$c_k^{(2)} E_k^{(0)} + \sum_m c_m^{(1)} H'_{km} = E^{(2)} c_k^{(0)} + E^{(1)} c_k^{(1)} + E^{(0)} c_k^{(2)} \quad (\text{A.1.14})$$

Per $k = n$ si trova la correzione al secondo ordine all'energia:

$$E^{(2)} = \sum'_m c_m^{(1)} H'_{km} = \sum'_m \frac{|H'_{mn}|^2}{E_n^{(0)} - E_m^{(0)}} \quad (\text{A.1.15})$$

La correzione alla funzione d'onda si calcola per $k \neq n$:

$$\begin{aligned} c_k^{(2)} &= -\frac{H'_{nn} H'_{kn}}{(E_n^{(0)} - E_k^{(0)})^2} - \sum'_m \frac{H'_{mn} H'_{km}}{(E_n^{(0)} - E_m^{(0)})(E_k^{(0)} - E_n^{(0)})} = \\ &= -\frac{H'_{nn} H'_{kn}}{\hbar^2 \omega_{nk}^2} + \sum'_m \frac{H'_{mn} H'_{km}}{\hbar^2 \omega_{nk} \omega_{nm}} \end{aligned} \quad (\text{A.1.16})$$

$$\text{con } \omega_{nk} = \frac{E_n^{(0)} - E_k^{(0)}}{\hbar}.$$

Il termine $c_n^{(2)}$ v'è calcolato dalla condizione di normalizzazione; la norma di

$(1 + c_n^{(2)})|n\rangle + \sum'_k (c_k^{(1)} + c_k^{(2)})|k\rangle$ deve essere 1, allora:

$$c_n^{(2)} = -\frac{1}{2} \sum'_k |c_k^{(1)}|^2 = -\frac{1}{2} \sum'_k \frac{|H'_{kn}|^2}{\hbar^2 \omega_{nk}^2} \quad (\text{A.1.17})$$

Dopo aver visto le correzioni all'energia e alla funzione d'onda per il caso di non degenerazione, dobbiamo affrontare il caso più generale di degenerazione.

A.2 Teoria delle perturbazioni con degenerazione

Se l'hamiltoniana non perturbata H_0 possiede autovalori degeneri non si può usare la procedura precedente. Prima avevamo trovato $c_k^{(1)} = \frac{H'_{kn}}{E_n^{(0)} - E_k^{(0)}}$, ma ora può esistere uno

stato $k \neq n$ tale che $E_k^{(0)} = E_n^{(0)}$; dovremo cambiare metodo di calcolo per le perturbazioni.

Siano allora $\Psi_{n1}^{(0)}, \Psi_{n2}^{(0)}, \Psi_{n3}^{(0)}, \dots, \Psi_{ns}^{(0)}$ gli stati corrispondenti ad un livello con $E = E_n^{(0)}$ dell'hamiltoniana H_0 . L'insieme di questi stati $\{\Psi_{nr}\}$ formano un insieme ortonormale nel sottospazio appartenente all'autovalore $E_n^{(0)}$.

Per semplicità indicheremo tali stati con $\Psi_1^{(0)}, \Psi_2^{(0)}, \dots, \Psi_s^{(0)}$ omettendo l'indice n.

All'ordine zero in H' qualsiasi altra scelta di insieme ortonormale $\Psi_i^{(0)} = \sum_j c_{ij} \Psi_j^{(0)}$ è buona.

Al primo ordine in H' , come abbiamo detto sopra, troveremo:

$$c_k^{(1)} = \frac{H'_{kn}}{E_n^{(0)} - E_k^{(0)}} = \infty \quad (\text{A.2.1})$$

L'idea è quella di usufruire della libertà di base all'ordine zero, per "preparare gli stati" in modo che l'effetto della perturbazione sia effettivamente piccolo; una buona base deve ancora essere tale che H' sia diagonale.

Riportiamo la formula esatta trovata prima:

$$c_k E_k^{(0)} + \sum_m c_m H'_{km} = E c_k \quad (\text{A.2.2})$$

e la consideriamo per gli stati $k = n_1, n_2, \dots, n_s$ appartenenti al livello degenerare n .

Ponendo $E = E^{(0)} + E^{(1)}$, $c_k = c_k^{(0)} + c_k^{(1)}$ nella formula sopra si ha, al primo ordine:

$$E^{(1)} c_i^{(0)} = \sum_j H'_{ij} c_j^{(0)} \quad (\text{A.2.3})$$

Tale equazione la possiamo scrivere in forma matriciale:

$$\left(H' - E^{(1)} I \right) \underline{c} = 0 \quad (\text{A.2.4})$$

Tale equazione è detta equazione secolare, cioè un'equazione agli autovalori della matrice (nel sottospazio $s \times s$) H' .

In altre parole, la correzione al primo ordine dell'energia è data da uno degli s autovalori di H' , determinato da:

$$\det \left| H' - E^{(1)} I \right| = 0 \quad (\text{A.2.5})$$

e gli autovettori corrispondenti,

$$c^{(r)} = \begin{pmatrix} c_1^{(r)} \\ \cdot \\ \cdot \\ \cdot \\ c_s^{(r)} \end{pmatrix} \quad \text{per } r = 1, 2, \dots, s \quad (\text{A.2.6})$$

definiscono s funzioni d'onda all'ordine zero in H' .

Se s autovalori di H' sono uno diverso dall'altro, la degenerazione è eliminata al primo ordine, altrimenti sarà solo parzialmente eliminata.

Il caso che si incontra frequentemente è quello di un livello non perturbato doppiamente degenerare. L'equazione secolare diventa:

$$\det \begin{vmatrix} H'_{11} - \lambda & H'_{12} \\ H'_{21} & H'_{22} - \lambda \end{vmatrix} = 0 \quad (\text{A.2.7})$$

le cui soluzioni sono

$$E^{(1)} = \lambda = \frac{1}{2} \left\{ H'_{11} + H'_{22} \pm \sqrt{(H'_{11} - H'_{22})^2 + 4|H'_{12}|^2} \right\} \quad (\text{A.2.8})$$

A.3 Perturbazioni dipendenti dal tempo: transizioni quantistiche.

L'interazione con campi esterni esplicitamente dipendenti dal tempo è qualitativamente diversa da quella di prima. Adesso scriveremo che:

$$H = H_0 + H'(t) \quad (\text{A.3.1})$$

L'energia di un sistema sottoposto a tali potenziali esterni non si conserva e questo fa sì che siano possibili transizioni quantistiche tra i diversi livelli.

Sia $\Psi_k^{(0)}(q, t)$ la funzione d'onda non perturbata di uno stato stazionario con il fattore $e^{-i\frac{E_k t}{\hbar}}$ incluso. L'equazione di Schrödinger generale è:

$$i\hbar \frac{\partial \Psi}{\partial t} = (H_0 + H'(t)) \Psi \quad (\text{A.3.2})$$

Sviluppando la funzione d'onda in termini di stati stazionari non perturbati

$$\Psi = \sum_k a_k(t) \Psi_k^{(0)} \quad (\text{A.3.3})$$

dove $\Psi_k^{(0)}$ soddisfa l'equazione

$$i\hbar \frac{\partial \Psi_k^{(0)}}{\partial t} = H_0 \Psi_k^{(0)} \quad (\text{A.3.4})$$

si ha

$$i\hbar \sum_k \frac{da_k}{dt} \Psi_k^{(0)} = \sum_k a_k H'(t) \Psi_k^{(0)} \quad (\text{A.3.5})$$

Moltiplicando questa equazione per $\Psi_m^{(0)*}$ e integrando su q , abbiamo:

$$i\hbar \frac{da_m}{dt} = \sum_k a_k H'_{mk}(t) \quad (\text{A.3.6})$$

dove $H'_{mk}(t) = \int \Psi_m^{(0)*} H'(t) \Psi_k^{(0)} dq = V_{mk} e^{i\omega_{mk}t}$, con $\omega_{mk} = \frac{E_m - E_k}{\hbar}$ e $V_{mk} = \langle m | H'(t) | k \rangle$.

Come stato non perturbato prendiamo $\Psi = \Psi_n^{(0)}$, cioè $a_n^{(0)} = 1$, $a_k^{(0)} = 0$ se $k \neq n$, e poniamo $a_m(t) = a_m^{(0)} + a_m^{(1)}(t)$.

Al primo ordine abbiamo

$$i\hbar \frac{da_{mn}^{(1)}(t)}{dt} = \sum_k a_k^{(0)} H'_{mk}(t) = H'_{mn}(t) \quad (\text{A.3.7})$$

nel primo termine abbiamo ripristinato l'indice n per indicare lo stato prima di "accendere" la perturbazione.

Integrando l'equazione sopra si ha:

$$a_{mn}^{(1)}(t) = -\frac{i}{\hbar} \int_{t_0}^t H'_{mn}(t) dt = -\frac{i}{\hbar} \int_{t_0}^t V_{mn} e^{i\omega_{mn}t} dt \quad (\text{A.3.8})$$

dove $V_{mn} = \langle m | H'(t) | n \rangle$ dipende dal tempo per la dipendenza esplicita da esso dell'hamiltoniana.

Prenderemo adesso in esame alcuni casi particolari per poter ricavare alcune espressioni delle approssimazioni introdotte.

Consideriamo dapprima perturbazioni che durano per un intervallo finito di tempo; se la perturbazione è assente sia per $t \rightarrow -\infty$ che per $t \rightarrow +\infty$, ha senso studiare la probabilità totale di transizione tra uno stato iniziale i e uno finale f . L'ampiezza di transizione, al primo ordine di H' , è dato dalla formula vista sopra:

$$a_{fi}^{(1)}(\infty) = -\frac{i}{\hbar} \int_{-\infty}^{+\infty} V_{fi} e^{i\omega_{fi}t} dt \quad (\text{A.3.9})$$

allora la probabilità di trovare il sistema nello stato f sarà:

$$P_{i \rightarrow f} = \left| a_{fi}^{(1)}(\infty) \right|^2 = \frac{1}{\hbar^2} \left| \int_{-\infty}^{+\infty} V_{fi} e^{i\omega_{fi}t} dt \right|^2 \quad (\text{A.3.10})$$

E' interessante studiare il caso delle perturbazioni che (essendo assenti a $t = -\infty$) tende ad un operatore costante a $t \rightarrow +\infty$. L'integrale è però divergente; la ragione di ciò è semplice; integriamo per parti e otteniamo

$$a_{fi}^{(1)} = -\frac{V_{fi}(\infty) e^{i\omega_{fi}t}}{\hbar\omega_{fi}} + \int_{-\infty}^t \frac{\partial V_{fi}}{\partial t} \frac{e^{i\omega_{fi}t}}{\hbar\omega_{fi}} dt \quad (\text{A.3.11})$$

Il primo termine è la correzione alla funzione d'onda (che avevamo trovato come

$\frac{H'_{kn}}{E_n^{(0)} - E_k^{(0)}}$) ma visto che il sistema a $t = \infty$ è in uno stato diverso da quello iniziale, tale

correzione è da aspettarsi. La probabilità di transizione è allora data solo da

$$P_{i \rightarrow f} = \frac{1}{\hbar^2 \omega_{fi}^2} \left| \int_{-\infty}^{+\infty} \frac{\partial V_{fi}}{\partial t} e^{i\omega_{fi}t} dt \right|^2 \quad (\text{A.3.12})$$

Dall'espressione precedente risulta che se la perturbazione varia poco col tempo allora la

transizione non avrà luogo; infatti $P_{i \rightarrow f} \approx 0$ se $\left| \frac{\partial V_{fi}}{\partial t} \right| \frac{1}{|V_{fi}|} \ll \omega_{fi}$.

Vediamo adesso il caso in cui l'hamiltoniana vari istantaneamente (per $t = t_0$) allora il fattore esponenziale si può portare fuori dall'integrale visto che $\frac{\partial V_{fi}}{\partial t}$ è non nullo solo per $t = t_0$.

$$a_{fi} \approx \frac{e^{i\omega_{fi}t_0}}{\hbar\omega_{fi}} \int_{-\infty}^t \frac{\partial V_{fi}}{\partial t} dt = \frac{e^{i\omega_{fi}t_0}}{\hbar\omega_{fi}} \cdot V_{fi} \quad (\text{A.3.13})$$

Otteniamo allora l'espressione della probabilità di transizione:

$$P_{i \rightarrow f} \approx |a_{fi}|^2 = \frac{|V_{fi}|^2}{\hbar^2 \omega_{fi}^2} \quad (\text{A.3.14})$$

Come ultimo caso particolare consideriamo quello di perturbazioni periodiche, che è di grande importanza.

$$H'(t) = \begin{cases} Fe^{-i\omega t} + h.c. & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases} \quad (\text{A.3.15})$$

dove con *h.c.* abbiamo indicato il coniugato hermitiano.

Andiamo a calcolare il quadrato dell'ampiezza, che è dato da

$$|a_{fi}|^2 = |F_{fi}| \frac{4 \text{sen}^2 \left[\frac{1}{2} (\omega_{fi} - \omega) t \right]}{\hbar^2 (\omega_{fi} - \omega)^2} \quad (\text{A.3.16})$$

Per t grande $\frac{\text{sen}^2 \left(\frac{1}{2} \alpha t \right)}{\alpha^2}$ tende a $\delta(\alpha)$, la cui costante moltiplicativa è data da

$$\int_{-\infty}^{+\infty} \frac{\text{sen}^2 z}{z^2} dz = \pi.$$

Allora $\frac{\text{sen}^2 \left(\frac{1}{2} \alpha t \right)}{\alpha^2} \rightarrow \frac{\pi t}{2} \delta(\alpha)$ da cui troviamo:

$$|a_{fi}|^2 = |F_{fi}|^2 \frac{2\pi}{\hbar^2} t \delta(\omega_{fi} - \omega) = |F_{fi}|^2 \frac{2\pi}{\hbar} t \delta(E_f - E_i - \omega \hbar) \quad (\text{A.3.17})$$

Questa espressione diverge col tempo, si definisce allora una probabilità di transizione $i \rightarrow f$ per un intervallo unitario di tempo, data da

$$\omega_{fi} = |F_{fi}|^2 \frac{2\pi}{\hbar} \delta(E_f - E_i - \omega \hbar) \quad (\text{A.3.18})$$

La funzione δ rappresenta la conservazione dell'energia; se lo stato finale f fa parte di uno spettro continuo, la probabilità di transizione va moltiplicata per il numero di stati quantistici tra E_f e $E_f + dE_f$, ossia $dv_f = \rho(E_f) dE_f$ dove $\rho(E_f)$ è la densità di stati.

Se integriamo la ω_{fi} moltiplicata per dv_f troviamo:

$$\omega_{fi} = |F_{fi}|^2 \frac{2\pi}{\hbar} \rho(E_f) \Big|_{E_f = E_i^{(0)} + \hbar\omega} \quad (\text{A.3.19})$$

detta regola d'oro di Fermi.

APPENDICE B

Discretizzazione delle equazioni del modello.

Vogliamo proporre, in questo paragrafo, il metodo numerico utilizzato dal programma di simulazione per la soluzione delle equazioni di Poisson e Schrödinger. Data una equazione differenziale alle derivate parziali, si può trovare una soluzione analitica solo in pochi casi particolari, introducendo pesanti approssimazioni. L'approccio che si usa per la simulazione è quello di trasformare l'equazione in un sistema di equazioni algebriche, operando una discretizzazione.

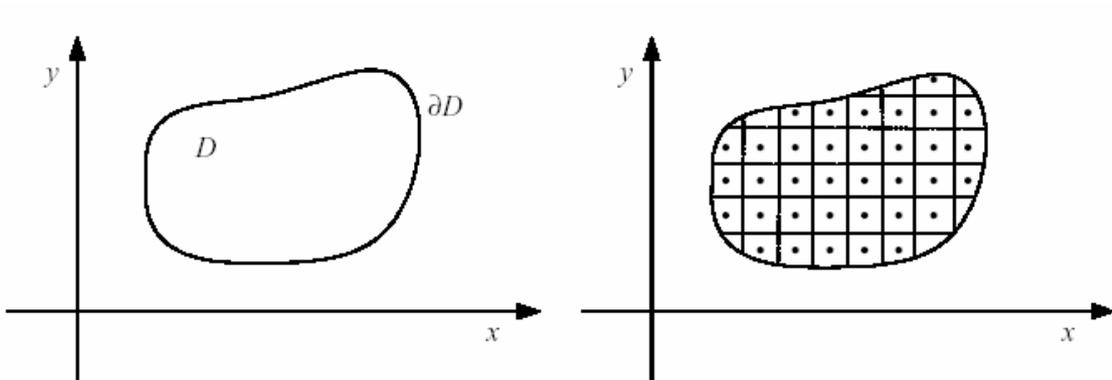


Figura B.1: regione di definizione della soluzione e sua discretizzazione.

Supponiamo che $u(x, y)$, soluzione della nostra equazione differenziale, sia definita nel dominio D di figura B.1; possiamo pensare di cercare la soluzione solamente in alcuni punti (x_i, y_i) del dominio D : l'insieme di tali punti viene definito griglia.

Il dominio D viene poi suddiviso in tante sottoregioni, nelle quali si suppone che la soluzione sia costante e pari al valore della stessa nel punto corrispondente della griglia; così facendo le incognite del sistema diventano i valori della funzione in punti della griglia, che sono in numero finito, quindi abbiamo trasformato l'equazione in un sistema algebrico.

Il metodo che useremo per discretizzare le equazioni del nostro modello è quello delle differenze finite, del quale riporteremo i tratti principali, limitandoci al caso monodimensionale.

Supponiamo che $u(x)$ sia la soluzione della nostra equazione differenziale e sviluppiamola in serie di Taylor considerando un passo costante Δx :

$$u(x + \Delta x) = u(x) + \frac{\partial u(x)}{\partial x} \Delta x + \frac{1}{2!} \frac{\partial^2 u(x)}{\partial x^2} (\Delta x)^2 + \frac{1}{3!} \frac{\partial^3 u(x)}{\partial x^3} (\Delta x)^3 + O(\Delta x)^3 \quad (\text{B.1})$$

$$u(x - \Delta x) = u(x) - \frac{\partial u(x)}{\partial x} \Delta x + \frac{1}{2!} \frac{\partial^2 u(x)}{\partial x^2} (\Delta x)^2 - \frac{1}{3!} \frac{\partial^3 u(x)}{\partial x^3} (\Delta x)^3 + O(\Delta x)^3 \quad (\text{B.2})$$

Dalla prima si ha:

$$\frac{\partial u(x)}{\partial x} = \frac{u(x + \Delta x) - u(x)}{\Delta x} - O(\Delta x) \quad \text{differenze in avanti} \quad (\text{B.3})$$

e dalla seconda si ottiene:

$$\frac{\partial u(x)}{\partial x} = \frac{u(x) - u(x - \Delta x)}{\Delta x} - O(\Delta x) \quad \text{differenze indietro} \quad (\text{B.4})$$

Facendo la sottrazione fra la seconda e la prima si ottiene invece:

$$\frac{\partial u(x)}{\partial x} = \frac{u(x + \Delta x) - u(x - \Delta x)}{2\Delta x} + O(\Delta x)^2 \quad \text{differenze centrali} \quad (\text{B.5})$$

che permette di approssimare la derivata prima con un'accuratezza maggiore. Se si prendono in considerazione anche termini del tipo $u(x + 2\Delta x)$, $u(x - 2\Delta x)$, $u(x + 3\Delta x)$,

$u(x - 3\Delta x)$, etc. possiamo semplificare anche le derivate di ordine superiore al primo; così facendo si ottiene però un sistema numerico troppo complicato, conviene invece infittire la griglia. La derivata seconda si può approssimare sommando la prima e la seconda equazione:

$$\frac{\partial^2 u(x)}{\partial x^2} = \frac{u(x + \Delta x) - 2u(x) + u(x - \Delta x)}{\Delta x^2} + O(\Delta x)^2 \quad (\text{B.6})$$

Se si usa una griglia non uniforme i risultati precedenti si modificano nel modo seguente:

$$h_i = x_{i+1} - x_i \quad \text{passo} \quad (\text{B.7})$$

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} = \frac{u_{i+1} - u_i}{h_i} \quad \text{differenze in avanti} \quad (\text{B.8})$$

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} = \frac{u_i - u_{i-1}}{h_{i-1}} \quad \text{differenze indietro} \quad (\text{B.9})$$

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} = \frac{u_{i+1} - u_{i-1}}{h_i - h_{i-1}} \quad \text{differenze centrali} \quad (\text{B.10})$$

Per il calcolo della derivata seconda bisogna considerare i punti di mezzo della griglia, detti punti fittizi:

$$\begin{aligned} \left. \frac{\partial^2 u}{\partial x^2} \right|_{x=x_i} &= \frac{\left. \frac{\partial u}{\partial x} \right|_{i+1/2} - \left. \frac{\partial u}{\partial x} \right|_{i-1/2}}{x_{i+1/2} - x_{i-1/2}} = \frac{\frac{u_{i+1} - u_i}{h_i} - \frac{u_i - u_{i-1}}{h_{i-1}}}{\frac{h_i - h_{i-1}}{2}} = \\ &= u_{i+1} \left[\frac{2}{h_i (h_i + h_{i-1})} \right] - u_i \left[\frac{2}{h_i (h_i + h_{i-1})} \left(\frac{1}{h_i} + \frac{1}{h_{i-1}} \right) \right] + u_{i-1} \left[\frac{2}{h_{i-1} (h_i + h_{i-1})} \right] \end{aligned} \quad (\text{B.11})$$

Possiamo adesso passare all'equazione di Poisson discretizzata, che fornisce:

$$\begin{aligned}
 \varphi_1 &= \varphi_A \\
 \varepsilon_{i+1/2} h_i \varphi_{i-1} - (\varepsilon_{i+1/2} h_{i-1} + \varepsilon_{i-1/2} h_i) \varphi_i + \varepsilon_{i+1/2} h_{i-1} \varphi_{i+1} &= -\frac{h_i h_{i-1} (h_i + h_{i-1})}{2} \rho_i \\
 \varphi_N &= \varphi_B
 \end{aligned} \tag{B.12}$$

con $i = 2, \dots, N-1$ e dove ε indica il prodotto $\varepsilon_0 \varepsilon_r$. Si ottiene così un sistema tridiagonale di $N-2$ equazioni in $N-2$ incognite (i punti interni della griglia) che il programma risolve mediante la routine *tridag*. In caso di equazione bidimensionale o tridimensionale si ottiene rispettivamente una matrice pentadiagonale o eptadiagonale: in tali casi non esiste una forma chiusa di risoluzione ma dobbiamo ricorrere a metodi iterativi.

In modo analogo a quanto fatto con l'equazione di Poisson otteniamo la discretizzazione dell'equazione di Schrödinger, imponendo l'annullamento della funzione d'onda agli estremi del dominio.

$$\left\{ \begin{array}{l} \Psi_1 = 0 \\ -\frac{\hbar^2}{h_i + h_{i-1}} \left(\frac{\Psi_{i+1} - \Psi_i}{m_{i+1/2} h_i} - \frac{\Psi_i - \Psi_{i-1}}{m_{i-1/2} h_{i-1}} \right) + E_{C_i} \Psi_i = E \Psi_i \\ \Psi_N = 0 \end{array} \right. \tag{B.13}$$

con $i = 2, \dots, N-1$. Anche in questo caso abbiamo un sistema tridiagonale che, se moltiplichiamo entrambi i membri per $(h_i + h_{i-1})$, diventa simmetrico. Possiamo scriverlo in forma matriciale, nella forma:

$$A\Psi = ED\Psi \tag{B.14}$$

con A matrice tridiagonale simmetrica e D matrice diagonale.

Riscriviamola come:

$$AD^{-\frac{1}{2}} D^{\frac{1}{2}} \Psi = ED^{\frac{1}{2}} D^{\frac{1}{2}} \Psi \tag{B.15}$$

moltiplichiamo a sinistra per $D^{-\frac{1}{2}}$ e otteniamo:

$$D^{-\frac{1}{2}}AD^{-\frac{1}{2}}D^{\frac{1}{2}}\Psi = ED^{\frac{1}{2}}\Psi \quad (\text{B.16})$$

Definiamo $A' = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ che è una matrice tridiagonale simmetrica e $\Psi' = D^{\frac{1}{2}}\Psi$ si ha:

$$A'\Psi' = E\Psi' \quad (\text{B.17})$$

dalla quale si ricavano gli autovalori, mentre da

$$\Psi = D^{-\frac{1}{2}}\Psi' \quad (\text{B.18})$$

si ricavano gli autovettori.

Le equazioni vengono risolte con un metodo iterativo autoconsistente, i cui passi sono indicati nel diagramma di flusso B.2. Si parte da una stima iniziale del potenziale $\varphi^{(0)}$ (ossia φ al passo zero) dalla quale si valutano $E_c^{(0)}$ e $E_v^{(0)}$ tramite la 1.2.23.

Il passo successivo è il calcolo della concentrazione di donatori e di accettori ionizzati, utilizzando le 1.2.24 e 1.2.25, e la concentrazione di portatori calcolate con le 1.2.20 e 1.2.22 (trattazione semiclassica) o con le 1.3.9 e 1.3.10 (trattazione quantistica). In quest'ultimo caso vengono risolte le equazioni di Schrödinger 1.3.7 e 1.3.8 per gli elettroni e una equazione del tutto equivalente, ma con massa efficace m_n , per le lacune. A questo punto è nota $\rho_i^{(0)}$ e l'equazione di Poisson diventa lineare; risolta quest'ultima si confronta la soluzione ottenuta con quella ottenuta al passo precedente ($\varphi^{(k-1)}$ se siamo al passo k -esimo) e, se la norma euclidea del vettore differenza tra le due soluzioni consecutive è minore di una certa tolleranza prefissata, si esce dal ciclo prendendo come soluzione quella dell'ultimo passo, altrimenti si incrementa k finché non si arriva alla tolleranza voluta. In questo modo però il metodo converge difficilmente perché la densità di carica è molto sensibile a piccole variazioni del potenziale. Si usa allora il metodo del

sottorilassamento che consiste nel prendere come potenziale al passo k -esimo una media pesata tra la soluzione ottenuta e il potenziale al passo precedente:

$$\varphi^{(k)} = \alpha\varphi^* + (1-\alpha)\varphi(k-1) \quad \text{con} \quad 0 < \alpha \leq 1 \quad (\text{B.19})$$

Il coefficiente α dipende dalla struttura che stiamo analizzando; per strutture più complesse α deve essere abbastanza piccolo quindi il metodo converge più lentamente, ma siamo certi della convergenza ad una soluzione.

Il valore della tolleranza sulla norma euclidea, che determina l'uscita dal ciclo, e il valore del coefficiente di sottorilassamento sono strettamente collegati visto che, se si diminuisce α per garantire la convergenza, bisogna diminuire anche la tolleranza perché se quest'ultima fosse troppo grande il sistema potrebbe anche convergere, ma non è detto che converga ad una soluzione.

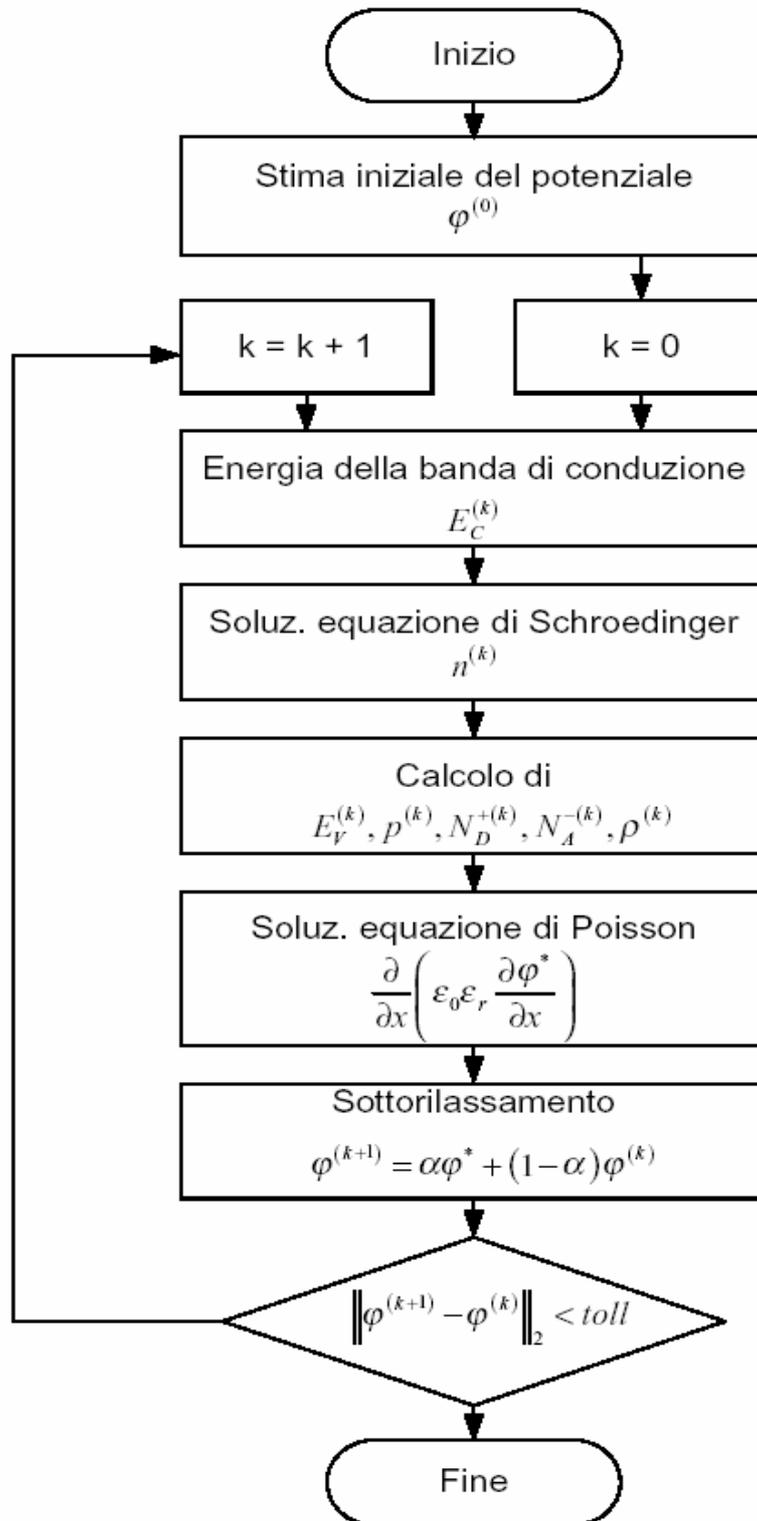


Figura B.2: diagramma di flusso per la soluzione autoconsistente delle equazioni di Schrödinger e di Poisson.

Conclusioni

Siamo riusciti, tramite confronto tra risultati sperimentali e simulazioni, ad estrarre i parametri dei materiali utilizzati nei gate stack alternativi. Le curve sperimentali vengono riprodotte fedelmente dal codice di simulazione e questo ci permette di fare anche analisi qualitative su altri fenomeni che si presentano, oltre naturalmente a quelle quantitative.

Abbiamo infatti rilevato che l'ossido di afnio ha una maggiore quantità di trappole interfacciali che comportano problemi di stabilità della tensione di soglia a causa di fenomeni quali il Fermi level pinning. I silicati, al contrario, soffrono molto meno di questi fenomeni, pur pagando, rispetto all'ossido di afnio, una costante dielettrica nettamente inferiore. L'altra conclusione qualitativa che possiamo trarre riguarda le interfacce ai lati dello stack; la densità di trappole interfacciali è risultata molto minore nei silicati rispetto all'ossido, anche per l'introduzione dello strato interfacciale di *SiON*. Non ci sono, o sono trascurabili, correnti di perdita per basse tensioni negative, comunque inferiori rispetto a quelle che si avrebbero utilizzando il classico ossido di silicio. Sembrano inoltre assenti fenomeni conduttivi come TAT (Trap Assisted Tunneling).

I valori estratti dalle nostre simulazioni concordano, per la maggioranza dei parametri, con gli importanti lavori che si trovano in letteratura; in alcuni casi differiscono da essi, ma la forte variabilità degli stessi in dipendenza del processo di crescita utilizzato consente una "varianza" abbastanza alta sui valori. Delle tre strutture analizzate, sia con ossido di afnio che con silicato di afnio, due si sono dimostrate molto bene in accordo con i valori ottenuti dalle simulazioni, mentre per una struttura (quella con 4 nm di silicato e 1 nm di *SiON*) i

Conclusioni

risultati delle simulazioni di caratteristiche I-V non sono stati molto soddisfacenti. Le motivazioni di questa discordanza potrebbero essere moltissime, a partire dalla scarsa conoscenza dei parametri di processo utilizzati. Il fatto che l'altra struttura con silicato di afnio abbia dato subito dei risultati buoni ci fa credere che una più approfondita conoscenza dei parametri e delle procedure di misura possa risolvere questo problema.

Sull'altro argomento di questa tesi, ovvero le memorie a nanocristalli, dobbiamo necessariamente lasciare qualche dubbio in più. Le simulazioni effettuate sulle due strutture presentate nel capitolo 3 hanno portato a dei risultati che sono in disaccordo con le rilevazioni sperimentali che si trovano in letteratura. In particolare il valore della tensione di gate per cui si ha il massimo della tensione di soglia (o, il che è equivalente, del numero medio di elettroni nel dot) nelle curve sperimentali, risulta superiore a quello ottenuto dalle simulazioni. Non c'è nessuna variazione significativa neanche cambiando i parametri fisici e geometrici della struttura. Abbiamo allora dovuto ipotizzare che il livello energetico in cui sono confinati gli elettroni si trovi nel gap del silicio e non dove il calcolo quantistico prevede. Questi livelli energetici profondi sono associati a trappole interfacciali che possono crearsi all'interfaccia nanocristallo ossido di silicio durante la crescita dei nanocristalli e che sono capaci di intrappolare l'elettrone. Con questa ipotesi le curve simulate si spostano verso le curve misurate, senza però riprodurle perfettamente. Un possibile motivo della discordanza dei massimi può essere individuato nel fatto che le curve sperimentali non rappresentino una condizione di regime; al passare del tempo i massimi si spostano infatti per tensioni sempre più basse. Se gli elettroni, come mostrano le nostre simulazioni, fossero confinati sulla superficie del nanocristallo, lo sforzo

Conclusioni

tecnologico dovrebbe essere rivolto alla ricerca di processi che minimizzino i difetti superficiali.

Bibliografia

- [1] Giovanni Ghione: Dispositivi per la microelettronica, Mc Graw – Hill, 1998
- [2] E. Burstein, S. Lundquist: Tunneling Phenomena in solids, Plenum Press, New York, 1969.
- [3] J. Cai, C.T. Sah: “Gate tunneling current in ultrathin oxide metal-oxide-semiconductor transistors”, Journal of Applied Physics, Volume 89, Number 4, 15 February 2001, pp. 2272-2285.
- [4] G. Iannaccone, B. Pellegrini: “Unified approach to electron transport in double-barrier structures”, Physical Review B, Volume 52, Number 24, 15 December 1995-II, pp. 17 406-17 412.
- [5] G. Iannaccone, B. Pellegrini: “A compact formula for the density of states in a quantum well”
- [6] C.T. Sah: Fundamentals of Solid State Electronics, World Scientific, 1991
- [7] E.H. Nicollian, J.R. Brews: MOS (Metal Oxide Semiconductor) Physics and Tecnology, Bell Laboratories, Murray Hill, New Jersey, 1982
- [8] C.C. Hobbs, L.R.C. Fonseca, A. Knizhnik, V. Dhandapani, S.B. Samavedam, W.J. Taylor, J.M. Grant, L.G. Dip, D.H. Triyoso, R.I. Hegde, D.C. Gilmer, R. Garcia, D. Roan, M.L. Lovejoy, R.S. Rai, E.A. Hebert, H.H. Tseng, S.G.H. Anderson, B.E. White, P.J. Tobin: “Fermi level pinning at the Polysilicon/Metal Oxide interface – Part I”, IEEE Transactions on Electron Devices, Volume 51, Number 6, June 2004, pp. 971-977.

- [9] C.C. Hobbs, L.R.C. Fonseca, A. Knizhnik, V. Dhandapani, S.B. Samavedam, W.J. Taylor, J.M. Grant, L.G. Dip, D.H. Triyoso, R.I. Hegde, D.C. Gilmer, R. Garcia, D. Roan, M.L. Lovejoy, R.S. Rai, E.A. Hebert, H.H. Tseng, S.G.H. Anderson, B.E. White, P.J. Tobin: “Fermi level pinning at the Polysilicon/Metal Oxide interface – Part II”, IEEE Electron Device Letters, Volume 51, Number 6, June 2004, pp. 978-984.
- [10] J. Robertson: “Band offsets of wide-band-gap oxides and implications for future electronic devices” J. Vac. Sci. Technol. B, Microelectron. Process. Phenom., vol. 18, no. 3, pp. 1785–1791, 2000.
- [11] R. Garg, N.A. Chowdhury, M. Bhaskaran, P.K. Swain and D. Misra: “Electrical Characteristic of Thermally Evaporated HfO_2 ”, Journal of the Electrochemical Society, 2004
- [12] W.J. Zhu, T.P. Ma, T. Tamagawa, J. Kim, Y. Di: “Current transport in Metal/Hafnium-Oxide/Silicon structure”, IEEE Transactions on Electron Devices, Volume 23, Number 2, February 2002, pp. 97-99.
- [13] M. Togo, K. Watanabe, T. Yamamoto, N. Ikarashi, T. Tatsumi, H. Ono and T. Mogami: “Electrical Properties of 1.5-nm SiON Gate-Dielectric Using Radical Oxygen and Radical Nitrogen”, IEEE Transactions on Electron Devices, Vol. 49, NO. 11, NOVEMBER 2002.
- [14] Crupi Iannaccone Ammendola
- [15] G. Iannaccone, F. Crupi, B. Neri, S. Lombardo: “Theory and experiment of suppressed shot noise in Stress Induced Leakage Currents”, IEEE Transactions on Electron Devices, Volume 50, Number 5, May 2003, pp. 1363-1369.
- [16] Crupi Corso Ammendola
- [17] G. Ammendola, M. Vulpio, M. Bileci, N. Nastasi, C. Gerardi, G. Renna, I. Crupi, G. Nicotra and S. Lombardo: “Nanocrystal metal-oxide-semiconductor memories obtained by chemical vapor deposition of Si nanocrystal”, J. Vac. Sci. Technol., October 2002.

- [18] Jan De Blauwe: “Nanocrystal Nonvolatile Memory Devices”, IEEE Transactions on Nanotechnology, VOL. 1, NO. 1, MARCH 2002.

Bibliografia Capitolo 1

- R. Clerc, A. Spinelli, G. Ghibuado, G. Pananakakis: “Theory of direct tunneling current in metal-oxide-semiconductor structures”, Journal of Applied Physics, Volume 91, Number 3, 1 February 2002, pp. 1400-1409.
- G. Iannaccone, M. Macucci, B. Pellegrini: “Shot noise in resonant-tunneling structures”, Physical Review B, Volume 55, Number 7, 15 February 1997-I, pp. 4539-4549.

Bibliografia Capitolo 2

- F. Crupi, R. Degraeve, A. Kerber, D:H: Kwak, G. Groeseneken: “Correlation between Stress Induced Leakage Current (SILC) and the $H_F O_2$ bulk trap density in a $SiO_2 / H_F O_2$ stack”
- H. Hu, C. Zhu, Y.F. Lu, M.F. Li, B.J. Cho, W.K. Choi: “A high performance MIM capacitor using $H_F O_2$ dielectrics”, IEEE Electron Device Letters, Volume 23, Number 9, September 2002, pp. 514-516.
- E. Spitale, D. Corso, I. Crupi, G. Nicotra, S. Lombardo, D. Deleruyelle, M. Gely, N. Buffet, B. De Salvo, C. Gerardi: “Effect of high-k materials in the control dielectric stack of nanocrystal memories”
- S. Takagi, N. Yasuda, A. Toriumi: “Experimental evidence of inelastic tunneling in Stress Induced Leakage Current”, ”, IEEE Transactions on Electron Devices, Volume 46, Number 2, February 1999, pp. 335-341.

- S. Takagi, N. Yasuda, A. Toriumi: “A new I-V model for Stress Induced Leakage Current including inelastic tunneling”, IEEE Transactions on Electron Devices, Volume 46, Number 2, February 1999, pp. 348-354.

Bibliografia Capitolo 3

- B. De Salvo, G. Ghibaudo, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, B. Guillaumot : “Experimental and theoretical investigation of nano-crystal and nitride-trap memory devices”, IEEE Transactions on Electron Devices, Volume 48, Number 8, August 2001, pp. 1789-1799.