

# POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in Ingegneria Fisica



**POLITECNICO**  
MILANO 1863

## **Studio della variabilità atomistica nei dispositivi nanowire tramite simulazioni 3D**

Relatore: Prof. Alessandro SOTTOCORNOLA SPINELLI  
Correlatore: Prof. Christian MONZIO COMPAGNONI

Tesi di Laurea Magistrale di  
Luca MASCETTI  
Matr. N. 819177

Anno Accademico 2014-2015



# Indice

Elenco delle figure	iii
Elenco delle tabelle	xii
Sommario	xiii
<b>1 Variabilità nei dispositivi decananometrici</b>	<b>2</b>
1.1 Lo scaling dei transistori MOS	2
1.1.1 L'effetto di canale corto (SCE)	3
1.1.2 Le regole di scaling	4
1.1.3 I dispositivi multigate	6
1.2 Il nanowire	8
1.2.1 L'inversione di volume	8
1.2.2 Scaling dei dispositivi nanowire	10
1.3 Il caso delle memorie 3D NAND	13
1.3.1 Funzionamento di una memoria Flash NAND	13
1.3.2 Le Flash NAND 3D e la cella 'macaroni'	14
1.4 La variabilità atomistica	16
1.4.1 Random Dopant Fluctuations (RDF)	17
1.4.2 Random Telegraph Noise (RTN)	19
<b>2 Simulazioni numeriche 3D</b>	<b>27</b>
2.1 Il modello drift-diffusion (DD)	27
2.2 Il modello density-gradient (DG)	29
2.2.1 Il confinamento quantistico nei MOSFET di tipo bulk	30
2.2.2 La correzione quantistica density-gradient (DG)	31
2.2.3 Il caso GAA	33
2.3 Il drogaggio atomistico	36
2.3.1 Implementazione nel modello DD	36
2.3.2 Posizionamento casuale dei droganti	40

---

<b>3</b>	<b>Analisi della variabilità atomistica nella struttura GAA con canale drogato</b>	<b>43</b>
3.1	Variabilità nella struttura planare . . . . .	43
3.2	Variabilità nella struttura GAA . . . . .	47
3.2.1	Geometria del dispositivo simulato . . . . .	47
3.2.2	Studio delle dipendenze RDF . . . . .	50
3.2.3	Studio delle dipendenze RTN . . . . .	56
<b>4</b>	<b>Nanowire a canale non drogato</b>	<b>64</b>
4.1	Analisi delle fluttuazioni RDF . . . . .	64
4.2	Analisi delle code RTN . . . . .	75
	<b>Conclusioni</b>	<b>82</b>
	<b>Bibliografia</b>	<b>83</b>

# Elenco delle figure

1.1	Andamento del nodo tecnologico nel tempo per quanto riguarda la realizzazione di CPU [1]. . . . .	2
1.2	Energia della banda di conduzione all'interfaccia Si-SiO <sub>2</sub> per due transistor con le stesse caratteristiche e diversa lunghezza di canale. (a) Caso con $L = 100$ nm. (b) Caso con $L = 30$ nm. Nel caso (b) è visibile l'abbassamento della barriera per effetto del DIBL e lo spostamento del suo massimo verso il source. . . . .	4
1.3	Transcaratteristica per i due transistor di Figura 1.2 all'aumentare di $V_{ds}$ . Per $V_{ds}$ alte il caso $L = 100$ nm ha la stessa $V_t$ che per $V_{ds}$ basse. Il transistor con $L = 30$ nm ha $V_t$ più bassa di quello con $L = 100$ nm (roll-off della tensione di soglia). Inoltre aumentando $V_{ds}$ la tensione di soglia del transistor con $L = 30$ nm diminuisce ulteriormente (DIBL). . . . .	5
1.4	Alcuni esempi di transistori multigate. (a) double-gate. (b) pi-gate. (c) omega-gate. (d) gate-all-around. Per migliorare le caratteristiche elettrostatiche del dispositivo viene ricoperta dal gate una parte sempre maggiore del canale (da [3].)	7
1.5	Soluzioni $\psi_s$ e $\psi_0$ per un nanowire con $R = 10nm$ e $t_{ox} = 2nm$ . Le due soluzioni seguono $V_g$ con lo stesso andamento fino a che $\psi_0$ rimane vincolato mentre $\psi_s$ continua ad aumentare con $V_g$ ma con una dipendenza più debole. Tale effetto è chiamato inversione di volume, e la tensione di gate a cui avviene lo stacco tra i due potenziali viene considerata la $V_t$ del transistor. . . . .	10
1.6	Energia della banda di conduzione di un nanowire lungo $r$ , per diverse tensioni di gate $V_g$ . Sotto soglia la banda è piatta e trasla rigidamente con $V_g$ , dando origine all'inversione di volume. . . . .	11

1.7	Analisi tramite simulazioni numeriche (simboli) e previsioni fornite da un modello analitico (linee continue) di dispositivi nanowire in regime di canale corto. (a) Roll-off di $V_t$ per una lunghezza del canale all'aumentare del raggio del transistor. L'effetto è più pronunciato all'aumentare di $R$ poiché si riduce l'influenza del gate sulla parte centrale del corpo di silicio. (b) Andamento della pendenza di sottosoglia al ridursi di $L$ per diversi raggi del transistor. L'aumento di STS al diminuire di $L$ è più forte per $R$ grandi, come nel caso del roll-off di $V_t$ (da [11]).	12
1.8	Previsioni per il raggiungimento dei futuri nodi tecnologici per chip che utilizzano transistori FinFET [1]. . . . .	12
1.9	Transcaratteristica di una cella Flash negli stati "cancellata" (nero) e "programmata" (rosso). Viene anche indicata la tensione di lettura $V_{read}$ . . . . .	14
1.10	Vista prospettica dell'array 3D PBiCS di Toshiba. Sono presenti la BL e la SL in direzione orizzontale e le WL che si diramano perpendicolarmente ad esse, come nel caso planare. In aggiunta sono presenti più strati verticali di WL, che estendono l'integrazione nella terza dimensione (da [13]). . . . .	15
1.11	(a) Cella di una memoria Flash 3D. La geometria è quella GAA con il gate che circonda il canale. Il corpo è in polisilicio e presenta una cavità riempita di dielettrico. Tale struttura viene detta 'macaroni'. Nelle 3D NAND il FG viene sostituito da una pila di strati dielettrici, $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$ (SONOS). L'immagazzinamento della carica avviene per intrappolamento degli elettroni nei difetti del $\text{Si}_3\text{N}_4$ , che rimangono dunque fissi riducendo i fenomeni di accoppiamento capacitivo tra celle adiacenti. (b) Transcaratteristica di due celle con lo stesso diametro e la stessa lunghezza di canale, una con il corpo pieno e l'altra con struttura macaroni. Si nota come la struttura macaroni migliori le caratteristiche elettrostatiche, avendo una $V_t$ minore, una STS maggiore e una corrente di conduzione più elevata (da [15]). . . . .	16
1.12	Energia della banda di conduzione per una sezione del transistor in direzione perpendicolare alla lunghezza del canale. Sul grafico sono mostrate anche le superfici di livello della densità di corrente lungo la sezione. La corrente si concentra nei percorsi più favorevoli, ovvero attraverso le valli del profilo di potenziale, dando origine alla percolazione (da [17]). . . . .	18

1.13	Distribuzione di potenziale all'interfaccia Si-SiO <sub>2</sub> per due n-MOS con lo stesso numero di accettori nella regione attiva ma differenti a livello microscopico. (a) La disposizione più o meno uniforme degli atomi lungo una linea perpendicolare al canale forma una barriera che riduce il passaggio di elettroni e alza $V_t$ . (b) Gli accettori si concentrano prevalentemente in un'area del canale lasciando scoperta una zona. In questa zona gli elettroni vedranno una barriera di potenziale minore e passeranno più facilmente dal source al drain attraverso tale percorso, abbassando $V_t$ (da [16]). . . . .	19
1.14	Andamento temporale della corrente di drain in funzione del tempo per una data densione di gate. Si possono vedere chiaramente i due livelli discreti di corrente associati agli stati con trappola piena ( $V_t$ alta) e trappola vuota ( $V_t$ bassa) (da [20]).	20
1.15	Variazione percentuale della corrente di drain ( $\frac{\Delta I_d}{I_d}$ ) in funzione della corrente di drain stessa ( $I_d$ ) in seguito alla presenza di una trappola all'interfaccia Si-SiO <sub>2</sub> che genera RTN. I punti rappresentano i dati sperimentali mentre la linea tratteggiata mostra l'andamento previsto dalla 1.4.2. I valori misurati sono maggiori di quasi un ordine di grandezza rispetto alle previsioni. Questo andamento rappresenta un esempio di <i>Giant RTN</i> (da [23]). . . . .	22
1.16	Distribuzioni cumulative F e 1-F del $\Delta V_t$ indotto da RTN per diverse tecnologie di memorie Flash NOR. Al diminuire delle dimensioni della cella aumenta la probabilità che si verifichino valori di Giant RTN. La pendenza della coda esponenziale della distribuzione è $\lambda$ (mV/decade) (da [24]). . . . .	23
1.17	Confronto tra le distribuzioni cumulative 1-F predette dai diversi modelli. (a) Distribuzione cumulativa dei casi 3D con doping continuo (linea tratteggiata) e 3D con doping atomistico (linea continua). Il caso con doping continuo ha una distribuzione che presenta due picchi e che non oltrepassa un certo valore di $\Delta V_t$ , a dimostrazione dell'assenza di percorsi percolativi (da [25]). (b) Grafico in scala semi-logaritmica dei tre modelli per l'RTN: 1D, 3D con doping continuo e 3D con droganti discreti. Il modello atomistico è l'unico che riproduce la coda esponenziale della distribuzione dovuta ai fenomeni di percolazione. (da [28]). . . . .	24
1.18	Esempio di profilo della densità di corrente per il transistor mostrato nel riquadro, calcolata tramite simulatore 3D. (a) Transistor con trappola vuota. Ai bordi dell'area attiva ci sono due percorsi preferenziali per la corrente, mentre verso il centro si trovano alcuni accettori che bloccano il passaggio. (b) Transistor con trappola occupata. La trappola si trova nei pressi di uno dei due filamenti di corrente e lo annulla in gran parte. Ne deriva un aumento di $V_t$ di tipo Giant RTN (da [27]). . . . .	25

2.1	<p>Rappresentazione schematica del fenomeno del confinamento quantistico in un MOSFET di tip bulk. (a) Il piegamento delle bande nei pressi dell'interfaccia Si-SiO<sub>2</sub> crea una buca di potenziale per gli elettroni dello strato di inversione. L'energia si divide in livelli discreti permessi, che aumentano verso la superficie della buca. (b) Densità di elettroni nello strato di inversione per una simulazione classica (in rosso), in cui la densità degli stati è continua, e per una simulazione quantistica (in blu), in cui la densità degli stati è discreta. La concentrazione calcolata secondo la meccanica quantistica è nulla all'interfaccia, poiché le autofunzioni si azzerano in tale punto, e presenta il picco spostato dall'interfaccia di circa 1 nm. . . . .</p>	30
2.2	<p>Confronto tra le previsioni di una simulazione DG (linea tratteggiata) e una quantistica (simboli) della densità di elettroni nello strato di inversione di un MOSFET di tipo bulk, al variare della tensione di gate. La soluzione DG riproduce con accuratezza la soluzione quantistica, prevedendo il picco della concentrazione di portatori spostato rispetto all'interfaccia Si-SiO<sub>2</sub>. L'andamento differente tra le due soluzioni lontano dall'interfaccia è dovuto in realtà ad un'approssimazione della simulazione quantistica che per semplificare i calcoli considerava solamente i primi 80 livelli discreti di energia (da [30]). . . . .</p>	32
2.3	<p>Densità di elettroni sottosoglia in funzione della posizione radiale per un nanowire con <math>R = 6</math> nm e <math>V_g = 0.3</math> V. Sono messe a confronto la soluzione classica (in rosso) e la soluzione quantistica (in blu). Le soluzioni sono state simulate con il codice utilizzato in [31]. . . . .</p>	34
2.4	<p>Densità di elettroni in funzione della posizione radiale per un nanowire con <math>R = 6</math> nm. <math>V_g = 0.7</math> V e il transistor è sopra soglia. La soluzione classica (in rosso) non risente del confinamento ed assume il suo massimo all'interfaccia. La soluzione quantistica (in blu) invece è nulla all'interfaccia ed assume la stessa forma del picco che si ha nel caso planare. . . . .</p>	34
2.5	<p>Confronto tra la densità di elettroni lungo la coordinata radiale simulata con la risoluzione dell'equazione di Poisson accoppiata con l'equazione di Schrödinger (in blu) o con la correzione DG implementata nel simulatore 3D (in rosso). (a) <math>V_g = 0.3</math> V, il transistor è sottosoglia e gli elettroni sono confinati al centro del volume di silicio. (b) <math>V_g = 0.7</math> V, il transistor è in inversione forte e gli elettroni sono confinati nello strato all'interfaccia Si-SiO<sub>2</sub>. In entrambi i casi il confinamento quantistico è riprodotto dal modello DG, che segue con qualche imprecisione la soluzione quantistica. . . . .</p>	35

2.6	Confronto tra la banda di conduzione lungo la coordinata radiale simulata dalla soluzione quantistica (in blu) o con la correzione DG (in rosso), per le stesse $V_g$ di Figura 2.5. (a) $V_g = 0.3$ V, il transistor è sottosoglia e le bande sono piatte in tutto il silicio. (b) $V_g = 0.7$ V, il transistor è in inversione forte e le bande sono piegate all'interfaccia. . . . .	35
2.7	Rappresentazione schematica della configurazione dei droganti sulla mesh (in alto) e della densità di droganti lungo la coordinata spaziale corrispondente. (a) Caso in cui il dispositivo è di grandi dimensioni. (b) Dispositivo atomistico (da [32]). . . . .	37
2.8	Rappresentazione schematica del potenziale elettrostatico causato da alcuni accettori ionizzati. In linea continua è mostrato il potenziale totale. In linea a tratti viene evidenziata la parte a lungo raggio, che è quella che viene tenuta in conto dal modello DD (da [32]). . . . .	38
2.9	Distribuzioni del potenziale elettrostatico di una carica fissa positiva in un cubo di materiale semiconduttore simulate con il modello DD (linee continue) e con il modello DG (linee a tratti), al variare della spaziatura tra i nodi della mesh. Il potenziale simulato con il modello DG è praticamente indipendente dalla spaziatura della griglia (da [33]). . . . .	39
2.10	Rappresentazione schematica di un transistor completamente atomistico. Sono indicati soltanto gli atomi droganti, ovvero gli accettori nel substrato (in blu) e i donori dei profili di source e drain (in rosso). La visione è completamente diversa rispetto a quella classica del modello DD, che ha profili di drogaggio continui e una giunzione metallurgica tra il substrato e il source o il drain. . . . .	40
2.11	Rappresentazione schematica della zona di discretizzazione utilizzata nella simulazione atomistica di un transistor di tipo bulk. Rispetto al caso illustrato in Figura 2.10 ora la zona in cui sono presenti gli atomi di drogante è ristretta ad un'area che coincide in prima approssimazione con la zona svuotata. Nel resto del dispositivo viene mantenuto un profilo di drogaggio costante. . . . .	42
3.1	Struttura utilizzata per le simulazioni statistiche del MOSFET planare. Source e drain hanno un profilo di drogaggio gaussiano con concentrazione di picco $N_d = 10^{20}$ cm <sup>-3</sup> . Nel substrato il drogaggio è uniforme con concentrazione $N_a = 5 \times 10^{18}$ cm <sup>-3</sup> . . . . .	44
3.2	Esempio del profilo atomistico implementato in uno dei 100 MOS planari simulati con il metodo Monte Carlo. Come zona di discretizzazione è stata presa una stima della zona svuotata. . . . .	45

3.3	Distribuzione delle tensioni di soglia dei 100 transistor planari con la geometria di Figura 3.1 e profilo di drogaggio atomistico simulato con il metodo Monte Carlo (in rosso). È stato sovrapposto il relativo fit gaussiano (in blu). . . . .	46
3.4	Distribuzione cumulativa delle ampiezze di RTN simulate su un campione di transistor atomistici con la geometria di Figura 3.1 (in rosso). Si vede che dall'andamento gaussiano della distribuzione si stacca una coda esponenziale con pendenza $\lambda$ , di cui è stato calcolato un fit lineare (in blu). . . . .	47
3.5	Struttura GAA implementata nel simulatore per l'analisi atomistica del nanowire. Il profilo di drogaggio di source e drain e la concentrazione di accettori nel substrato sono le stesse di Figura 3.1. . . . .	48
3.6	Transcaratteristica del transistor planare di Figura 3.1 (in rosso) e del nanowire di Figura 3.5 (in blu), a $V_{ds} = 50$ mV. Le stesse curve sono rappresentate sia in scala logaritmica che lineare. Nel nanowire, come previsto, la pendenza di sottosoglia è minore poiché il gate che circonda il canale riduce maggiormente l'effetto di canale corto. . . . .	49
3.7	Esempio del profilo di drogaggio atomistico di un nanowire utilizzato per le simulazioni di RDF e RTN. Il punto di stacco tra il profilo continuo e quello discreto è l'inizio del gate. . . . .	50
3.8	Confronto tra la percolazione nella struttura GAA e nella struttura di tipo bulk. La grandezza rappresentata in scala colore è il modulo di $\vec{J}_n$ . Sulle sezioni è tracciato il contorno degli accettori presenti. (a) Sezione del nanowire perpendicolare al diametro. È evidenziata con una linea a tratti il punto in cui è stata effettuata la sezione (b). (b) Sezione perpendicolare alla direzione del canale nel punto Z. (c) Sezione del dispositivo planare in direzione parallela al canale, ad una profondità in cui lo strato di inversione è nell'intorno del suo massimo. Si può notare come gli accettori discreti ostacolino il passaggio di corrente (toni dall'azzurro al verde). . . . .	52
3.9	Distribuzione dei valori di $V_t$ ottenuti dalle simulazioni Monte Carlo di 100 nanowire per diversi valori del doping nel canale (simboli). Sono sovrapposti i fit gaussiani corrispondenti (linee). . . . .	54
3.10	Valori della dispersione di $V_t$ dovuta a RDF in funzione dei parametri del nanowire (simboli) e andamento della stessa nel transistor planare dall'equazione (1.10) (linea continua). (a) Dipendenza di $\sigma_{V_t}$ dalla concentrazione di accettori nel canale del nanowire e relativo andamento previsto dalla (1.10). (b) Dipendenza di $\sigma_{V_t}$ dal raggio del nanowire e andamento previsto dalla (1.10) per $W$ . . . . .	55

3.11	Distribuzione cumulativa inversa delle ampiezze di RTN per 100 nanowire con $R = 4$ nm e $N_a = 5 \times 10^{18}$ cm <sup>-3</sup> (in verde). Viene riportata anche la distribuzione del caso planare di Figura 3.4 (in rosso) per un confronto. Sulle code delle curve è stato tracciato il fit della pendenza $\lambda$ . Le fluttuazioni RTN del caso GAA sono molto minori rispetto al caso bulk. . . . .	57
3.12	Confronto tra la cumulativa inversa di $\Delta V_t$ ottenuta simulando 100 nanowire con profilo di drogaggio continuo su tutto il dispositivo (in nero) e quella ottenuta da 100 simulazioni con drogaggio atomistico (in verde), per il caso $R = 4$ nm e $N_a = 5 \times 10^{18}$ cm <sup>-3</sup> . Lo stacco di una coda esponenziale è presente soltanto sulla distribuzione atomistica. . . . .	58
3.13	Distribuzioni di $\Delta V_t$ dovuto a RTN per tre campioni statistici di nanowire con diversa concentrazione di droganti nel canale. Sovrapposto ad ogni distribuzione è presente il fit della coda esponenziale a pendenza $\lambda$ . Le fluttuazioni aumentano all'aumentare degli ioni di drogante nel canale. . . . .	59
3.14	Sezioni di un nanowire atomistico simulato per lo studio dell'RTN preso dal campione $R = 8$ nm, $N_a = 2.65 \times 10^{19}$ cm <sup>-3</sup> . A sinistra ((a) e (c)) è riportato il caso senza carica all'interfaccia, mentre a destra((b) e (d)) viene mostrato l'effetto della trappola. In alto è mostrata la sezione perpendicolare al canale, all'altezza della trappola (indicata da un quadrato rosso in (b)). In basso è mostrata la sezione parallela a $L$ , passante per il centro del transistor. I droganti atomistici sono indicati con una linea che ne ricalca l'area occupata. In questo caso il loro numero è sufficientemente elevato da creare un percorso percolativo vicino all'interfaccia. La trappola casuale posizionata nei pressi di questo lo riduce e causa un'ampiezza $\Delta V_t$ grande. . . . .	60
3.15	Distribuzioni cumulative di $\Delta V_t$ per tre campioni statistici di nanowire al variare del raggio. Sovrapposto ad ogni distribuzione è presente il fit della coda esponenziale a pendenza $\lambda$ . Le fluttuazioni aumentano al diminuire di $R$ poiché i percorsi percolativi possibili diminuiscono. . . . .	61
3.16	Pendenza della coda RTN della distribuzione di $\Delta V_t$ al variare dei parametri del nanowire (simboli) e andamento previsto dall'equazione (1.10) per un dispositivo planare (linea continua). (a) Dipendenza di $\lambda$ dalla concentrazione di accettori nel canale del nanowire. (b) Dipendenza di $\sigma_{V_t}$ dal raggio del nanowire e andamento previsto dalla (1.10) per $W$ . In linea a tratti sono indicati gli andamenti seguiti dalle code simulate, che si discostano da quelli trovati per i transistor di tipo bulk. . . . .	62

4.1	Esempio del profilo di drogaggio atomistico di un nanowire con il canale non drogato, utilizzato per le simulazioni di RDF. In questo caso gli atomi discreti sono soltanto i donori del profilo gaussiano di source e drain. La zona di discretizzazione è stata aumentata e si estende anche fuori dal gate. . . . .	65
4.2	Distribuzione cumulativa delle tensioni di soglia per un campione di 100 nanowire atomistici al variare della dimensione della zona di discretizzazione. All'asse delle ordinate è stata applicata una trasformazione per facilitare il riconoscimento di distribuzioni di probabilità normali. Dal grafico si nota che è indifferente scegliere una delle tre $L_d$ per l'analisi dell'RDF. . . . .	66
4.3	Rappresentazione della $V_t$ estratta dai profili simulati, in funzione del numero totale di donori discreti presenti in questi. È riportato anche un fit lineare dell'andamento per una stima della correlazione tramite il coefficiente $R^2$ . . . . .	68
4.4	Andamento della tensione di soglia in funzione del numero di donori discreti presenti fuori dal gate del profilo simulato corrispondente. Il coefficiente $R^2$ per la stima della correlazione è minore che in Figura 4.3. . . . .	68
4.5	Grafico di $V_t$ in funzione del numero di donori presenti nel canale. Sono riportati anche i valori ottenuti per i profili con $L_d = 4$ nm e $L_d = 2$ nm per verificare che la dispersione di $V_t$ non sia dovuta al maggiore numero di atomi nella zona di discretizzazione. . . . .	69
4.6	Differenti profili di drogaggio implementati nello studio dell'RDF per il nanowire con canale non drogato. Sull'asse $x$ è indicata la posizione lungo l'asse passante per il centro del corpo cilindrico, partendo dal source del transistor. Le due linee a tratti indicano rispettivamente l'inizio e la fine del canale. . . . .	70
4.7	Distribuzioni cumulative di $V_t$ estratte da un campione di 100 nanowire atomistici per ciascuno dei due profili di drogaggio di Figura 4.6. Il profilo che ha la maggiore concentrazione di donori sotto il gate presenta un valore medio di $V_t$ più basso e una dispersione maggiore della stessa. . . . .	71
4.8	Rappresentazione dei valori di $V_t$ estratti dalle simulazioni di 100 transistor atomistici con il profilo di drogaggio $N_d(x_j) = 10^{18}$ cm <sup>-3</sup> , in funzione di diversi parametri. (a) $V_t$ in funzione del numero totale di droganti nel transistor. (b) $V_t$ in funzione del numero di droganti fuori dal gate. (c) $V_t$ in funzione del numero di droganti sotto il gate. Per ogni grafico sono riportati il fit lineare e il relativo valore di $R^2$ . . . . .	72
4.9	Valore assoluto della densità di corrente sottosoglia a $V_g$ costante su una sezione lungo l'asse del nanowire per due profili con posizionamento deterministico di un drogante. (a) Donore a $z = 0$ nm e a $r = 0$ nm. (b) Donore a $z = 0$ nm e a $r = 3.9$ nm. La differenza di $V_t$ tra i due casi è di 35 mV. . . . .	73

4.10	Superfici di livello dei valori di $V_t$ estratti da profili deterministici in cui è stata fatta variare la posizione radiale e la profondità all'interno del canale dell'unico donore atomistico presente. . . . .	74
4.11	Distribuzione cumulativa dei $\Delta V_t$ simulati su un campione di 100 transistor atomistici con il profilo di drogaggio $N_d(x_j) = 10^{18} \text{ cm}^{-3}$ per source e drain (in rosso). Viene anche riportata la cumulativa ottenuta dalle simulazioni di 100 trappole Monte Carlo sullo stesso profilo, ma con doping continuo (in nero). (a) Nanowire con $R = 6 \text{ nm}$ . (b) Nanowire con $R = 4 \text{ nm}$ . Le distribuzioni atomistiche coincidono praticamente con quelle continue, dunque l'eventuale presenza di percorsi percolativi è debole. . . . .	76
4.12	Rappresentazione schematica dell'aumento artificiale della statistica di una distribuzione. Tramite un algoritmo vengono simulati soltanto gli eventi rari per il fenomeno in esame, in modo da poter osservare la coda della cumulativa, che in genere si trova a probabilità di qualche ppm. La parte degli eventi comuni, che viene tenuta in conto correttamente pur non essendo simulata, è rappresentata in linea a tratti. . . . .	78
4.13	Valore assoluto della densità di corrente sottosoglia a $V_g$ costante su una sezione lungo l'asse del nanowire per un profilo atomistico in cui sono stati posizionati due donori sotto il gate rispettivamente a $z = 13.5 \text{ nm}$ , $r = 3.9 \text{ nm}$ e a $z = 17.5 \text{ nm}$ , $r = 3.9 \text{ nm}$ . (a) Profilo senza trappola. (b) Profilo con una trappola posizionata a $z = 13.5 \text{ nm}$ , $r = 4 \text{ nm}$ . Il valore di $\Delta V_t$ in questo caso è $> 20 \text{ mV}$ . . . . .	80
4.14	Distribuzione cumulativa dei $\Delta V_t$ estratti dal campione di 100 nanowire con $R = 4 \text{ nm}$ e profilo di drogaggio $N_d(x_j) = 10^{18} \text{ cm}^{-3}$ (in alto) e distribuzione cumulativa della coda simulata con l'aumento statistico (in basso). La coda presenta una pendenza di $8 \text{ mV/decade}$ , ma la sua probabilità di accadimento è del tutto trascurabile. . . . .	81

# Elenco delle tabelle

3.1 Matrice delle possibili combinazioni dei valori di  $R$  ed  $N_a$  che sono stati scelti per l'analisi della dipendenza delle fluttuazioni RDF dallo scaling. Il simbolo X indica quelle che sono state analizzate in modo statistico. Per ognuna di queste sono stati generati 100 profili Monte Carlo di drogaggio atomistico e si sono simulate le curve  $I - V$  per estrarre la tensione di soglia. . . . . 53

# Sommario

Nei primi anni '70 ha avuto inizio la produzione su larga scala di circuiti integrati basati sui transistori MOS. Da allora, al fine di ottenere una maggiore densità di integrazione e un aumento delle prestazioni dei componenti, le dimensioni di questi dispositivi vengono ridotte ogni anno secondo una legge esponenziale. Tuttavia, lo scaling della tecnologia convenzionale ha ormai raggiunto il suo limite fisico e, per poter proseguire nella miniaturizzazione dei dispositivi, è stato necessario ideare nuove strutture, dette multi-gate, che permettono un maggiore controllo elettrostatico del gate sul canale tramite geometrie innovative. Tra questi dispositivi, il nanowire è il principale candidato per l'avanzamento verso lunghezze di canale subdecananometriche.

Così alleviato il problema dello scaling, altri ne sorgono: quando la dimensione caratteristica di un transistor scende al di sotto dei 100 nm, la natura discreta della materia non può più essere trascurata. Le diverse configurazioni degli atomi a livello microscopico definiscono le caratteristiche elettriche di dispositivi identici dal punto di vista macroscopico. Questo effetto causa delle fluttuazioni statistiche della tensione di soglia che determinano una fonte di variabilità, atomistica, nel dispositivo. Le manifestazioni principali di questo fenomeno, le *random dopant fluctuations* (RDF) e il *random telegraph noise* (RTN), sono legate alla discretizzazione dei droganti nella zona del canale, che genera dei percorsi percolativi della corrente in regime di sottosoglia. Nel progettare il raggiungimento di un futuro nodo tecnologico diventa fondamentale conoscere se le fluttuazioni della tensione di soglia possano causare un degrado inaccettabile delle prestazioni.

Il presente lavoro di tesi si propone di investigare la variabilità atomistica nei dispositivi nanowire. Tramite simulazioni numeriche 3D saranno studiate le variazioni statistiche della tensione di soglia dovute alla posizione casuale degli atomi droganti, sia per quanto riguarda l'RDF che per l'RTN. Sono analizzate le dipendenze dal drogaggio di canale e dal raggio, dove si osserva un andamento differente rispetto a quello riscontrato nelle strutture planari.

Grazie al controllo elettrostatico superiore esercitato dal gate dei nanowire, è possibile rimuovere il drogaggio di accettori nel canale, in modo da ridurre le fluttuazioni della tensione di soglia. Nel seguito sarà analizzato questo diverso tipo di dispositivo, discu-

tendone l'efficacia nel ridurre le fluttuazioni RDF. Verrà infine presentato un metodo di *enhancement* statistico ideato per evidenziare le code RTN di questi transistor.

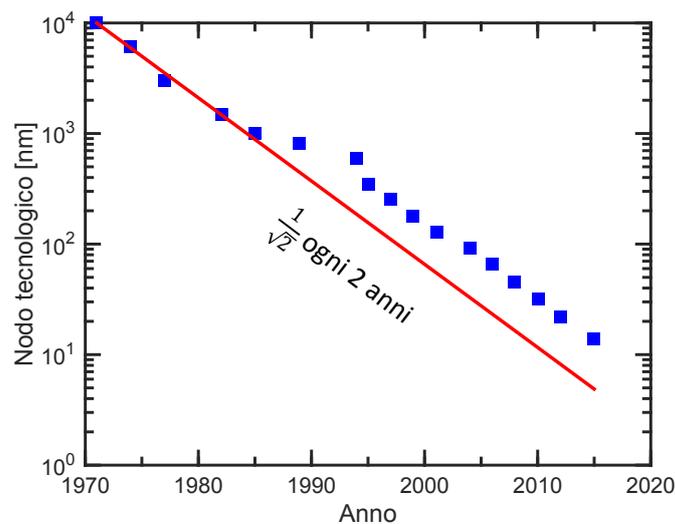


# Capitolo 1

## Variabilità nei dispositivi decananometrici

### 1.1 Lo scaling dei transistori MOS

A partire dagli anni '70 il numero di transistor integrati su un chip è raddoppiato ogni due anni, come aveva previsto Moore nel 1975. Questo significa che l'area occupata dal singolo dispositivo si dimezza e quindi la dimensione caratteristica minima si riduce di un fattore  $\sqrt{2}$  (Figura 1.1).



**Figura 1.1:** Andamento del nodo tecnologico nel tempo per quanto riguarda la realizzazione di CPU [1].

Scalare le dimensioni dei transistor non vuol dire soltanto aumentare i componenti presenti su un chip e diminuire così i costi di produzione, ma anche aumentare la velocità di risposta accorciando i tempi di transito e ridurre la potenza dissipata.

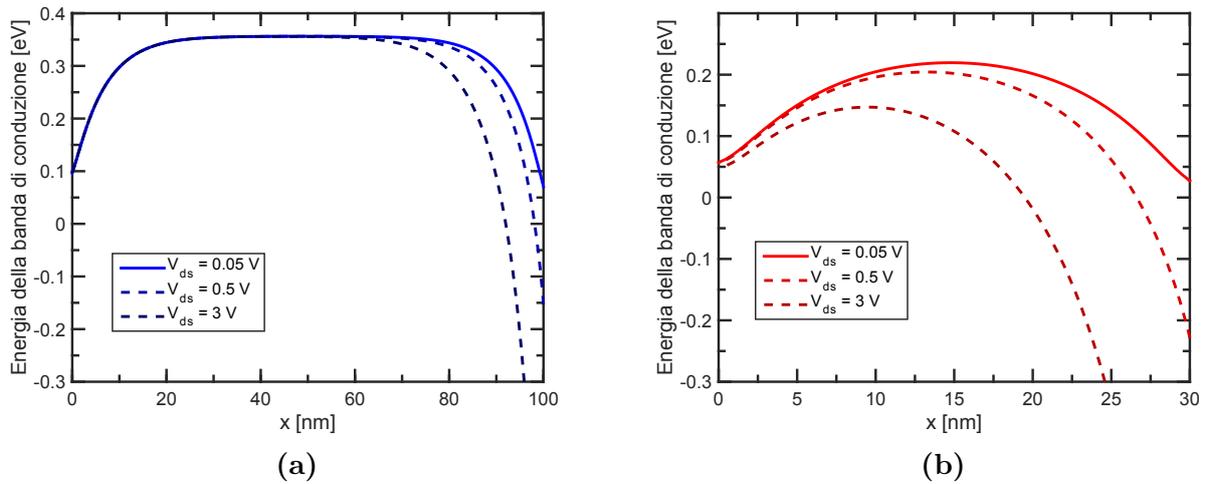
Le grandi aziende che fabbricano circuiti integrati hanno sostenuto negli anni un percorso di innovazione costante che produceva sviluppi tecnologici in tempi brevissimi e che ha portato i dispositivi a semiconduttore ad avere un ruolo fondamentale nella vita di oggi. Ogni nuovo passo di questa crescita è stato compiuto a seguito di enormi sforzi nel miglioramento dei processi produttivi e nella progettazione dei dispositivi.

### 1.1.1 L'effetto di canale corto (SCE)

Ridurre le dimensioni caratteristiche di un transistor comporta la nascita di molteplici problemi. Prendendo come esempio un n-MOS, se accorciamo la lunghezza ( $L$ ) del canale diminuisce la distanza tra source e drain. Quando tale distanza arriva a essere confrontabile con la larghezza della zona svuotata nel substrato, il potenziale di source e drain comincia ad esercitare un forte effetto sul piegamento delle bande lungo gran parte del dispositivo e il transistor è detto essere in regime di canale corto [2]. In Figura 1.2 si può vedere un confronto tra due transistori, con la stessa larghezza del gate ( $W$ ), la stessa concentrazione di accettori nel substrato ( $N_a$ ), lo stesso spessore dell'ossido di gate ( $t_{ox}$ ) ma con  $L = 100$  nm per quello di Figura 1.2a e  $L = 30$  nm per quello di Figura 1.2b. In condizioni di *off* del transistor il substrato di tipo p forma una barriera di potenziale tra source e drain. Solo un numero limitato di elettroni ha energia sufficiente per essere iniettato verso il drain (corrente di sottosoglia). Se siamo nel caso di un transistor a canale lungo la barriera di potenziale è piatta lungo tutto il canale e i campi elettrici di source e drain hanno un impatto solo sugli estremi del canale (Figura 1.2a). Il potenziale della barriera quindi è controllato principalmente dal gate. Se invece la lunghezza del canale è molto ridotta i campi di source e drain penetrano verso il centro del canale e abbassano la barriera di potenziale (Figura 1.2b). Ciò causa un aumento della corrente di sottosoglia e quindi una diminuzione della tensione di soglia ( $V_t$ ). Lo spostamento della  $V_t$  del transistor verso tensioni minori a causa delle condizioni di canale corto è detto *roll-off* della tensione di soglia.

Se viene applicata un'alta tensione di drain ( $V_{ds}$ ) ad un n-MOS in regime di canale corto l'altezza della barriera diminuisce ulteriormente e il suo punto di massimo trasla verso il source. Questo effetto è detto *drain-induced-barrier-lowering* (DIBL). In un transistor a canale corto quindi la corrente di sottosoglia aumenta all'aumentare della tensione di drain. Nel transistor a canale corto di Figura 1.2b si può vedere l'effetto del DIBL.

Se la lunghezza del canale diminuisce ancora il gate inizia a perdere il controllo del



**Figura 1.2:** Energia della banda di conduzione all'interfaccia Si-SiO<sub>2</sub> per due transistor con le stesse caratteristiche e diversa lunghezza di canale. (a) Caso con  $L = 100$  nm. (b) Caso con  $L = 30$  nm. Nel caso (b) è visibile l'abbassamento della barriera per effetto del DIBL e lo spostamento del suo massimo verso il source.

potenziale della barriera, che viene guidato prevalentemente dal drain, e la pendenza di sottosoglia (STS) della transcaratteristica degrada.

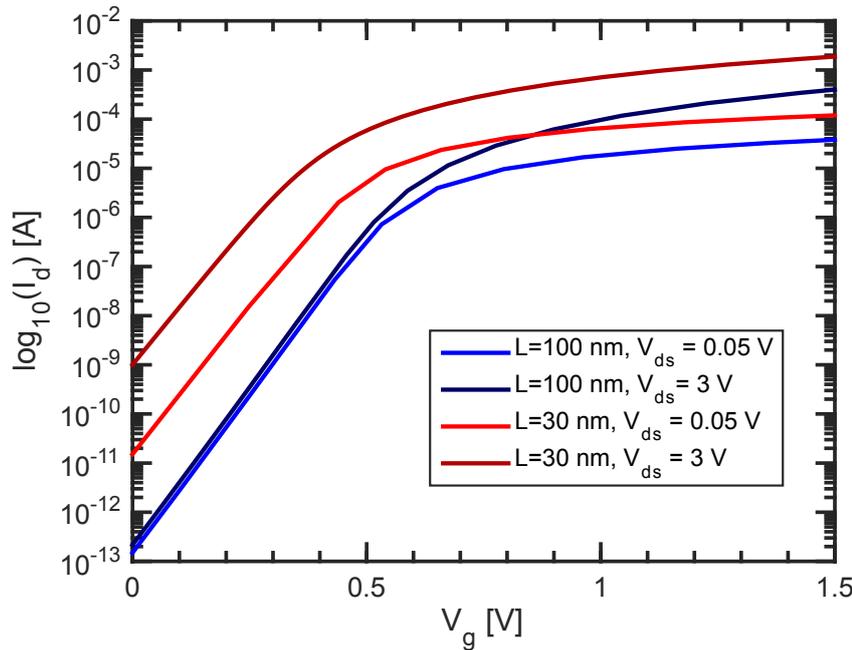
La diminuzione complessiva della tensione di soglia di un transistor in seguito al ridursi della sua lunghezza di canale è detta effetto di canale corto (SCE). Un esempio è mostrato in Figura 1.3, dove si possono vedere le caratteristiche I-V dei transistor di Figura 1.2 per due differenti tensioni di drain. Si osserva come all'aumentare di  $V_{ds}$  per il dispositivo a canale corto sorgano gli effetti che abbiamo discusso.

In un chip dove sono presenti transistori con diverse lunghezze di canale a causa delle tolleranze di processo, l'SCE porta i transistor ad avere tensioni di soglia diverse. Se la dispersione di  $V_t$  assume valori elevati nel circuito integrato viene introdotta una fonte di variabilità che ne limita il corretto funzionamento.

### 1.1.2 Le regole di scaling

Quando si vogliono scalare le dimensioni dei transistor bisogna tenere in conto degli effetti discussi, altrimenti si rischia di degradare le prestazioni del dispositivo invece che migliorarle.

Il DIBL come abbiamo detto è causato dalla progressiva perdita di controllo del gate sul canale e dalla penetrazione dei campi elettrici di source e drain nel canale. Per ridurre tale effetto si può ridurre lo spessore  $t_{ox}$  in modo da avvicinare il gate al canale ed aumentare il suo controllo sullo strato di inversione. Inoltre aumentando la concentrazione  $N_a$  si



**Figura 1.3:** Transcaratteristica per i due transistor di Figura 1.2 all'aumentare di  $V_{ds}$ . Per  $V_{ds}$  alte il caso  $L = 100$  nm ha la stessa  $V_t$  che per  $V_{ds}$  basse. Il transistor con  $L = 30$  nm ha  $V_t$  più bassa di quello con  $L = 100$  nm (roll-off della tensione di soglia). Inoltre aumentando  $V_{ds}$  la tensione di soglia del transistor con  $L = 30$  nm diminuisce ulteriormente (DIBL).

ottiene una transizione delle bande più brusca ai lati di source e drain, limitando la loro penetrazione all'interno del canale.

La saturazione delle velocità è causata da un aumento dei campi elettrici in direzione orizzontale, dunque per evitare tale effetto si possono diminuire le tensioni applicate dello stesso fattore di cui viene ridotta  $L$ .

Tuttavia questi provvedimenti influiscono anche su altre caratteristiche del dispositivo e sono quindi necessarie delle linee guida che regolino lo scaling, definendo quali effetti negativi sono in grado di ridurre e quali benefici apportano ai dispositivi realizzati.

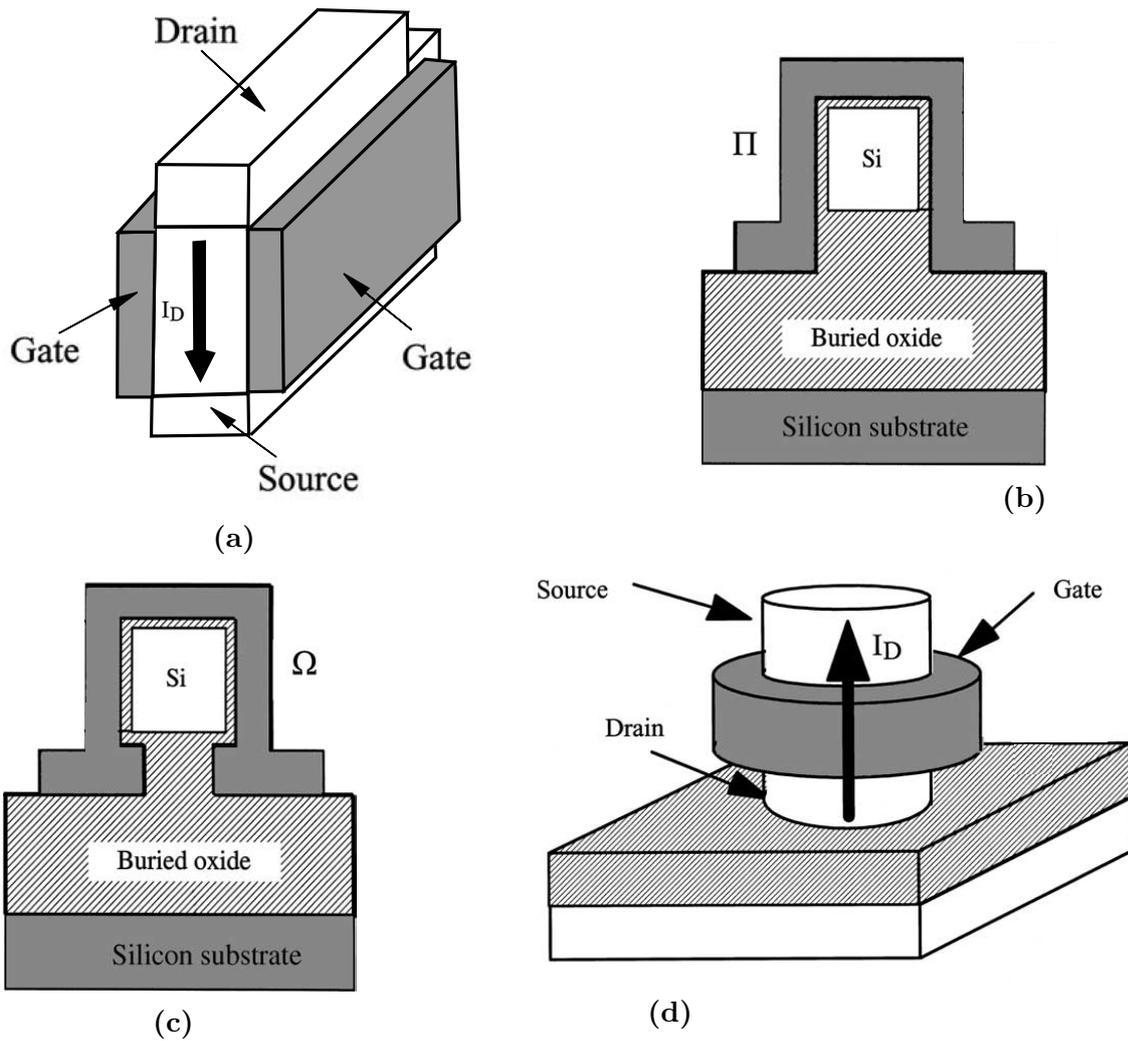
Una tipologia di regole per lo scaling è quella definita da Dennard nel 1974 che prende il nome di *constant-field*. Queste regole si propongono di mantenere lo stesso rapporto tra i campi elettrici verticali e quelli orizzontali, in modo da evitare l'SCE. Quindi se le dimensioni orizzontali ( $L$ ) vengono scalate di un fattore  $k$ , saranno scalati della stessa quantità  $k$  anche  $t_{ox}$  e  $W$ . Inoltre, per evitare la saturazione delle velocità, i campi elettrici ( $V_g, V_{ds}$ ) vengono ugualmente ridotti di  $k$ . Tali regole apportano dei benefici in termini del tempo di risposta, che diminuisce di un fattore  $k$ , e in termini di potenza dissipata, che diminuisce di un fattore  $k^2$ . Però le regole constant-field non riducono efficacemente gli effetti di canale corto in quanto lo spessore della zona svuotata ( $W_d$ ) scala di un fattore  $< k$ . A causa di questo motivo la  $V_t$  non riesce ad essere scalata in proporzione

alle tensioni di alimentazione dei circuiti.

Un'altra tipologia di regole di scaling è quella proposta da Baccarani nel 1984 chiamata *generalized*. Nella versione *constant-voltage* queste regole riducono dello stesso fattore  $k$  le dimensioni ma mantengono inalterate le tensioni applicate ai transistor. Questo permette di non avere incompatibilità con le precedenti generazioni di dispositivi e consente di scalare  $W_d$  di un fattore  $k$ , riducendo così efficacemente gli effetti di canale corto. Tuttavia tali regole di scaling hanno lo svantaggio di aumentare la potenza dissipata.

### 1.1.3 I dispositivi multigate

Dati i vantaggi e gli svantaggi che comportano le regole constant-field e constant-voltage, il raggiungimento dei diversi nodi tecnologici nel corso degli anni ha sfruttato aspetti di entrambi i tipi di regole. Ad ogni modo, una volta raggiunta una certa lunghezza di canale, non si è potuto più procedere oltre con le soluzioni tradizionali. Infatti al di sotto di certi spessori  $t_{ox}$  i campi elettrici sull'ossido di gate sono così elevati da causare tunneling quantistico di elettroni dal canale al gate e da degradarne la qualità. In aggiunta i livelli di  $N_a$  da impiantare per ridurre la zona svuotata sono arrivati a quantità che non consentirebbero più un corretto funzionamento del dispositivo. Perciò sono state pensate soluzioni innovative che migliorassero le prestazioni dei transistor e consentissero di proseguire nella riduzione delle loro dimensioni. L'idea portante è stata quella di rendere più robusto il transistor nei confronti degli effetti di canale corto aumentando l'area di influenza del gate sul canale. Sono state realizzate le cosiddette strutture *multigate* [3]. Il primo dispositivo multigate realizzato fu il *double-gate* MOS (DG MOS) nel 1989 (Figura 1.4a). Uno strato sottile e completamente svuotato di silicio è racchiuso tra due strati di dielettrico e due elettrodi di gate. Il controllo dell'elettrostatica del canale può essere ulteriormente migliorato se il gate si distribuisce su tre lati del canale o su tutti e quattro. Strutture del primo tipo sono i tri-gate MOSFET, di cui le varianti pi-gate MOSFET e omega-gate MOSFET sono mostrate in Figura 1.4b e Figura 1.4c. Dispositivi con tutto il canale circondato dal gate sono detti gate-all-around (GAA), o anche nanowire (NW) e hanno sezione quadrata o circolare (Figura 1.4d). Le strutture a multigate sono caratterizzate da un corpo di silicio che ha uno spessore ( $t_{Si}$ ) molto stretto e vengono anche dette FinFET. Alcune di queste strutture sono già in commercio. Ad esempio il nodo tecnologico precedente, che aveva una dimensione caratteristica di 22 nm, e quello attuale da 14 nm sono realizzati con un'architettura tri-gate [4]. Per confrontare la scalabilità dei transistor multigate può essere ripreso il concetto di lunghezza caratteristica  $\lambda$ , pensato nell'ambito del MOS planare. Questa grandezza è definita come la lunghezza caratteristica del decadimento esponenziale dei campi laterali. Se vogliamo limitare il roll-off di  $V_t$



**Figura 1.4:** Alcuni esempi di transistori multigate. (a) double-gate. (b) pi-gate. (c) omega-gate. (d) gate-all-around. Per migliorare le caratteristiche elettrostatiche del dispositivo viene ricoperta dal gate una parte sempre maggiore del canale (da [3].)

e il DIBL, la lunghezza minima ( $L_{min}$ ) accettabile dovrà essere compresa tra  $3\lambda$  e  $5\lambda$  [5]. Per un transistor planare [2]:

$$\lambda = \frac{(W_d + 3t_{ox})}{\pi}. \quad (1.1)$$

Per un dispositivo DG [5]:

$$\lambda_{DG} = \frac{(t_{Si} + 2\frac{\epsilon_{Si}}{\epsilon_{ox}}t_{ox})}{\pi}. \quad (1.2)$$

Mentre per un dispositivo GAA a sezione circolare [5]:

$$\lambda_{Cil} = \frac{(D_{Si} + 2\frac{\epsilon_{Si}}{\epsilon_{ox}}t_{ox})}{4.810}. \quad (1.3)$$

Essendo  $\lambda_{DG} \simeq 1.53 \lambda_{Cil}$  si riconosce che i dispositivi GAA sono i più efficienti nel ridurre gli effetti di canale corto con lo scaling e possiedono le caratteristiche elettrostatiche migliori.

## 1.2 Il nanowire

È stato mostrato come i MOSFET di tipo nanowire siano i più indicati per raggiungere lunghezze di canale decananometriche e sub-decananometriche. Numerosi studi vengono compiuti ogni anno per testarne l'affidabilità e le proprietà fisiche, e alcuni di questi dispositivi con lunghezze di canale nel range decananometrico sono già stati fabbricati [6], [7].

Vogliamo ora analizzare nel dettaglio la struttura NW (Figura 1.4d). Un sottile corpo di silicio a simmetria cilindrica, simile ad un filo, è circondato dall'ossido di gate e dal gate. La parte di silicio che si trova sotto il gate è il canale del transistor, mentre le parti restanti subiscono impiantazione di droganti e fungono da source e da drain. Il canale è flottante e, come in tutti i dispositivi FinFet,  $t_{Si}$  molto stretto. Nei nanowire, come si diceva, il gate che circonda tutto il canale riduce al minimo gli effetti di canale corto.

### 1.2.1 L'inversione di volume

I transistor GAA a simmetria cilindrica sono caratterizzati dall'inversione di volume, un fenomeno che si verifica nelle strutture FinFet con corpo flottante e che consente di raggiungere valori di STS ideali di 60 mV/decade. L'inversione di volume fu scoperta nel 1987 [8] e caratterizzata per i dispositivi DG. È stato formulato un modello analitico di questo effetto che ne mette in luce le ragioni fisiche [9]. Sulla stessa impronta sono stati ricavati dei modelli analitici per la struttura GAA da diversi autori (tra cui [10]).

Consideriamo il caso in cui nella regione del canale non ci sia carica fissa poiché non sono presenti atomi droganti. Prendiamo un nanowire che abbia  $L \gg R$ , dove  $R$  è il raggio del corpo di silicio. Vale la *gradual-channel-approximation* come nel caso del MOSFET planare [2], ovvero la variazione del campo elettrico lungo la direzione del canale è trascurabile rispetto alla variazione lungo la coordinata radiale. In queste condizioni l'equazione di Poisson in coordinate cilindriche può essere scritta come:

$$\frac{d^2\psi}{dr^2} + \frac{1}{r} \frac{d\psi}{dr} = \frac{qn_i}{\varepsilon_{Si}} e^{\frac{q\psi}{kT}}, \quad (1.4)$$

dove  $\psi$  è il potenziale elettrostatico,  $n_i$  è la concentrazione di portatori nel silicio intrinseco,  $q$  è la carica dell'elettrone,  $K$  è la costante di Boltzmann,  $T$  è la temperatura e  $\varepsilon_{Si}$  è la costante dielettrica del silicio. L'unica densità di carica che compare nell'equazione è quella degli elettroni, poiché non c'è carica fissa e assumiamo  $\frac{q\psi}{KT} \gg 1$  così che la densità di lacune è trascurabile. L'equazione (1.4) può essere risolta con le condizioni al contorno:

$$\left. \frac{d\psi}{dr} \right|_{r=0} = 0, \quad (1.5)$$

$$\varepsilon_{Si} \left. \frac{d\psi}{dr} \right|_{r=R} = C_{ox}(V_g - \phi_{ms} - \psi_s), \quad (1.6)$$

dove  $\phi_{ms}$  è la differenza tra la funzione lavoro del silicio e la funzione lavoro del gate,  $C_{ox}$  è la capacità dell'ossido di gate e  $\psi_s$  è il potenziale elettrostatico all'interfaccia Si-SiO<sub>2</sub>. La (1.5) rappresenta la continuità della componente normale del vettore spostamento dielettrico ( $D$ ) all'interfaccia Si-SiO<sub>2</sub> mentre la (1.6) nasce dalla simmetria radiale del problema. La soluzione della (1.4) è:

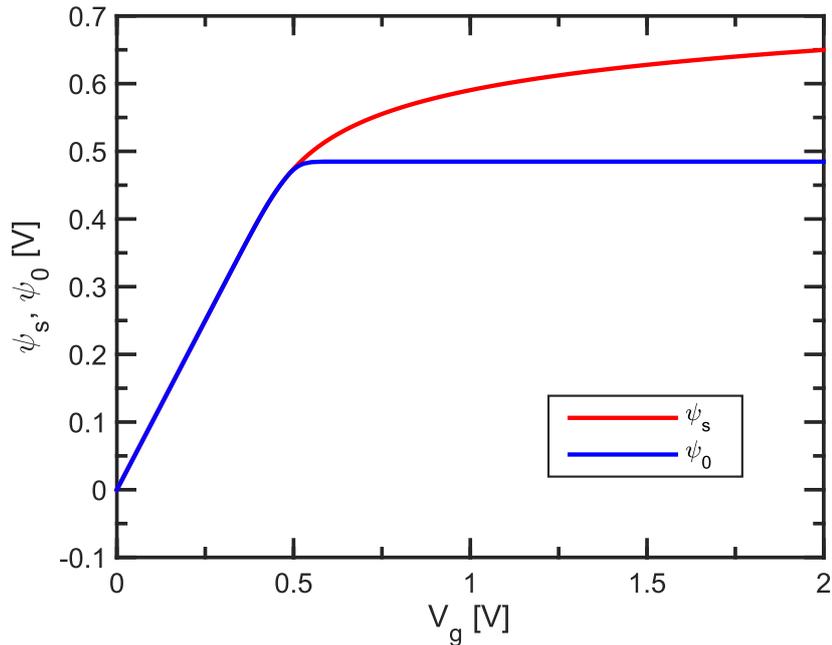
$$\psi = \psi_0 - 2 \frac{KT}{q} \ln \left( 1 - \frac{q^2 n_i}{8KT \varepsilon_{Si}} e^{\frac{q\psi}{KT}} \right), \quad (1.7)$$

dove  $\psi_0$  è il potenziale elettrostatico nel punto  $r = 0$ . Dall'equazione (1.7) si vede che  $\psi_0$  è limitato:

$$\psi_0 < \frac{KT}{q} \ln \left( \frac{8KT \varepsilon_{Si}}{R^2 q^2 n_i} \right), \quad (1.8)$$

mentre  $\psi_s$  non lo è. Se si traccia l'andamento di  $\psi_0$  e  $\psi_s$  in funzione di  $V_g$  (Figura 1.5) si nota come  $\psi_0$  e  $\psi_s$  sono uguali fino ad una certa tensione di gate, oltre la quale  $\psi_0$  rimane vincolato al suo valore di massimo mentre  $\psi_s$  continua ad aumentare. Ciò significa che fino ad un certo valore di  $V_g$ , che dipenderà esclusivamente dalla differenza ( $V_g - \phi_{ms}$ ), la carica di inversione (secondo termine del membro di destra della (1.4)) è molto piccola e le bande restano piatte seguendo tutte insieme  $V_g$ . Questa è dunque l'inversione di volume, in cui la carica non è confinata in una zona svuotata ma è compresa in tutto lo spessore del silicio. Dato che le bande traslano rigidamente seguendo la tensione di gate, la concentrazione di elettroni sarà uniforme in tutto il volume del nanowire. Se non c'è caduta di tensione nel silicio la densità di portatori sottosoglia aumenta esponenzialmente con  $V_g$  e quindi la pendenza di sottosoglia è quella ideale di 60 mV/decade.

Una volta che la carica di inversione diventa significativa schizzerà il potenziale del gate rispetto al centro del canale, così il potenziale  $\psi_s$  continua ad aumentare mentre  $\psi_0$  rimane costante. Si passa quindi in condizioni di inversione forte, come avviene nel MOSFET planare. La  $V_g$  a cui avviene lo stacco di  $\psi_s$  da  $\psi_0$  sarà allora la  $V_t$  del nanowire.



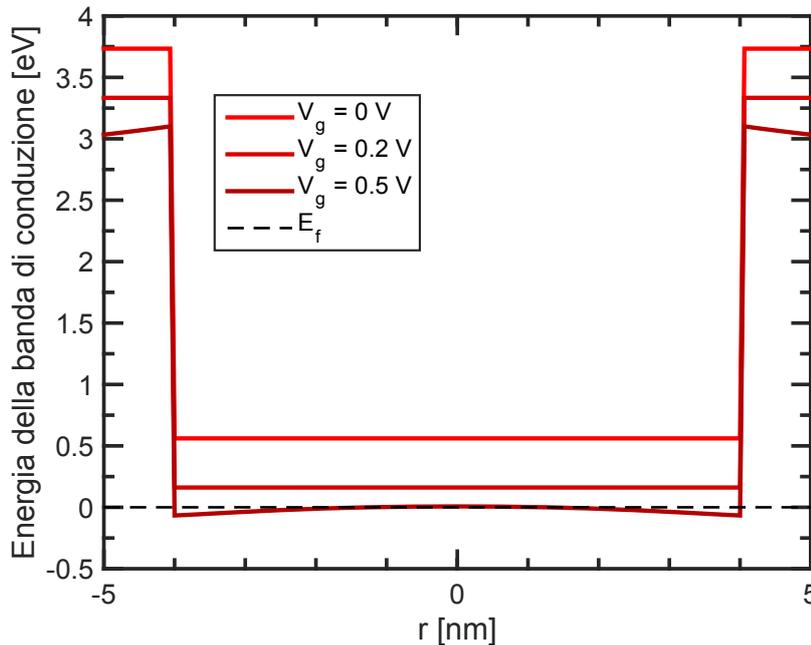
**Figura 1.5:** Soluzioni  $\psi_s$  e  $\psi_0$  per un nanowire con  $R = 10nm$  e  $t_{ox} = 2nm$ . Le due soluzioni seguono  $V_g$  con lo stesso andamento fino a che  $\psi_0$  rimane vincolato mentre  $\psi_s$  continua ad aumentare con  $V_g$  ma con una dipendenza più debole. Tale effetto è chiamato inversione di volume, e la tensione di gate a cui avviene lo stacco tra i due potenziali viene considerata la  $V_t$  del transistor.

In Figura 1.6 si vede la banda di conduzione (BC) di un nanowire per differenti tensioni di gate. Per  $V_g = 0$  V e  $V_g = 0.2$  V il transistor è sotto soglia e si osserva come la BC trasli rigidamente, rimanendo piatta lungo  $r$  con caduta di tensione nulla sull'ossido. Per  $V_g = 0.5$  V invece il transistor è sopra soglia, le bande sono piegate ed è presente un'elevata densità di carica all'interfaccia che dà origine a  $\psi_s > \psi_0$ .

### 1.2.2 Scaling dei dispositivi nanowire

È stato detto che con il ridursi della lunghezza di canale le prestazioni dei nanowire vengono degradate meno rispetto a quelle degli altri transistor (sezione 1.1.3). Per effettuare delle previsioni sui comportamenti di questi dispositivi in seguito allo scaling nel range decananometrico sarà necessario conoscere come si manifesta l'SCE e quali sono le sue dipendenze dai parametri del nanowire.

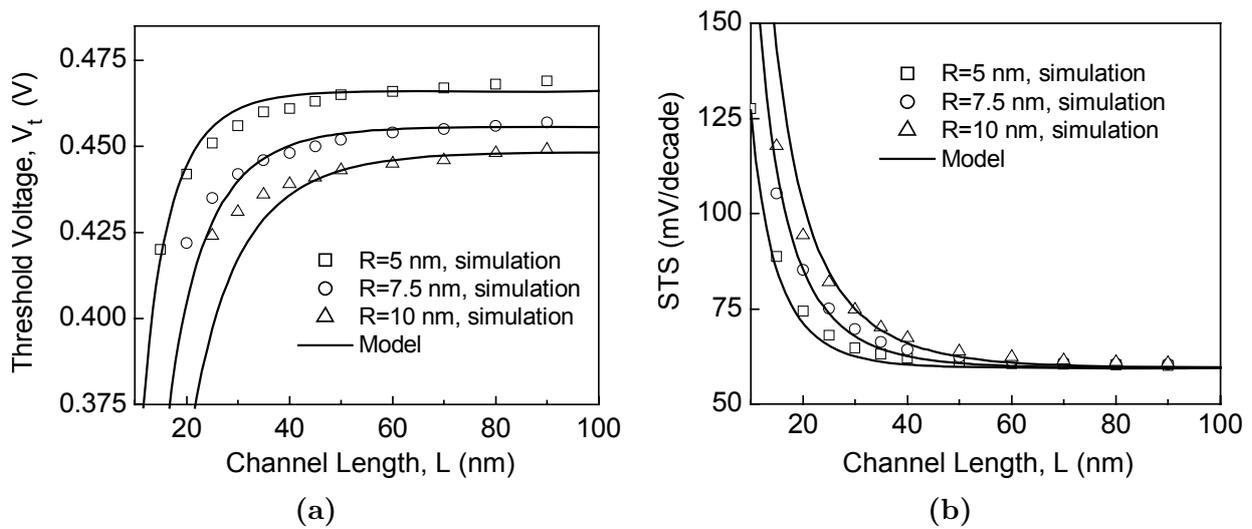
In [11] viene indagato il regime di canale corto dei transistor GAA a simmetria cilindrica tramite la formulazione di un modello analitico e con simulazioni numeriche 3D. In Figura 1.7 sono mostrati i valori di tensione di soglia e di STS in un nanowire con uno spessore  $t_{ox}$  di 2 nm al variare della lunghezza del canale e del raggio. In Figura 1.7a si vede come, scalando  $L$ ,  $V_t$  non rimane costante al valore che ha per  $L$  grandi ma



**Figura 1.6:** Energia della banda di conduzione di un nanowire lungo  $r$ , per diverse tensioni di gate  $V_g$ . Sotto soglia la banda è piatta e trasla rigidamente con  $V_g$ , dando origine all'inversione di volume.

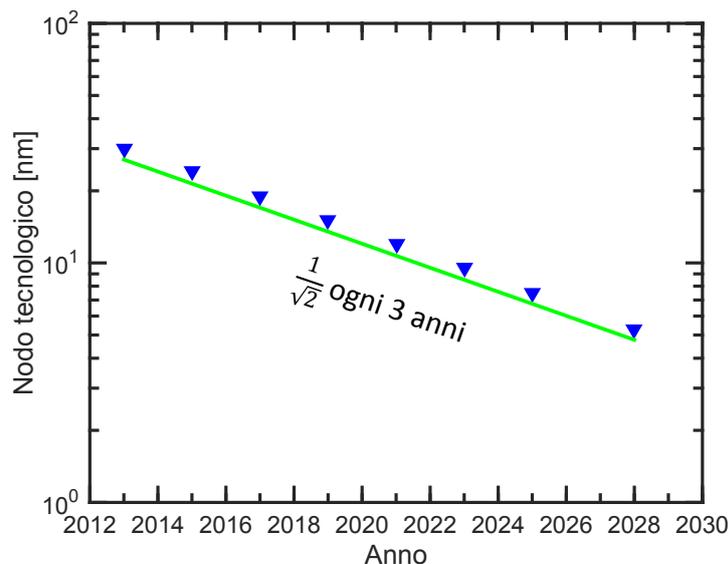
comincia a diminuire. Se  $L$  si riduce, la tensione di soglia diminuisce poiché i campi di source e drain penetrano nel canale, analogamente a quanto accade per i MOS di tipo bulk (sezione 1.1.1). Tale effetto, come si vede dalla figura, ha anche una dipendenza dal raggio del nanowire. Se  $R$  aumenta, il roll-off di  $V_t$  si manifesta a lunghezze di canale maggiori. Infatti al crescere di  $R$  il gate viene allontanato dalla parte centrale del corpo cilindrico. In questa zona il gate esercita una minore influenza elettrostatica e la barriera di potenziale che vedono i portatori viene abbassata dai campi di source e drain che penetrano nel canale. Dato che in una struttura GAA sottosoglia si verifica l'inversione di volume, il peso che hanno i portatori al centro del cilindro sulla conduzione è lo stesso che hanno quelli verso la superficie. Se i primi vedono una barriera di potenziale minore allora nel dispositivo scorrerà una corrente maggiore sottosoglia e  $V_t$  diminuisce.

In Figura 1.7b si vede l'effetto che ha sulla pendenza di sottosoglia la riduzione della lunghezza di canale del nanowire. Se il source e il drain del transistor penetrano nel canale, il potenziale dei punti più lontani dalla superficie è controllato più dal source e dal drain che dal gate. In queste condizioni, variando la tensione di gate, non si ha più il controllo completo di tutta la carica di inversione, ma soltanto le bande dei punti più vicini alla superficie traslano rigidamente con essa. Il valore di STS non è più quello ideale di 60 mV/decade. Per lo stesso motivo discusso per il roll-off di  $V_t$ , l'aumento di STS avviene a  $L$  maggiori se aumenta  $R$ .



**Figura 1.7:** Analisi tramite simulazioni numeriche (simboli) e previsioni fornite da un modello analitico (linee continue) di dispositivi nanowire in regime di canale corto. (a) Roll-off di  $V_t$  per una lunghezza del canale all'aumentare del raggio del transistor. L'effetto è più pronunciato all'aumentare di  $R$  poiché si riduce l'influenza del gate sulla parte centrale del corpo di silicio. (b) Andamento della pendenza di sottosoglia al ridursi di  $L$  per diversi raggi del transistor. L'aumento di STS al diminuire di  $L$  è più forte per  $R$  grandi, come nel caso del roll-off di  $V_t$  (da [11]).

L'importanza che i dispositivi FinFET ricopriranno nel processo di scaling che avverrà nei prossimi anni è stata riconosciuta dall'ITRS, che li ha inseriti nella sua ultima versione [1]. In Figura 1.8 sono mostrate le previsioni degli anni in cui verrà raggiunto un determinato nodo tecnologico per i chip che impiegano tali transistor. Si può notare



**Figura 1.8:** Previsions per il raggiungimento dei futuri nodi tecnologici per chip che utilizzano transistori FinFET [1].

come il trend di scaling sia leggermente inferiore a quello previsto dalla legge di Moore, diminuendo di un fattore  $\sqrt{2}$  ogni 3 anni.

## 1.3 Il caso delle memorie 3D NAND

Le memorie Flash NAND sono la tipologia di memorie non volatili più diffusa e con un mercato da 20 miliardi di dollari l'anno [12]. Anche per questi dispositivi è fondamentale aumentare la densità di celle presenti sul chip per poter aumentare la capacità di immagazzinare dati e abbattere il costo per bit.

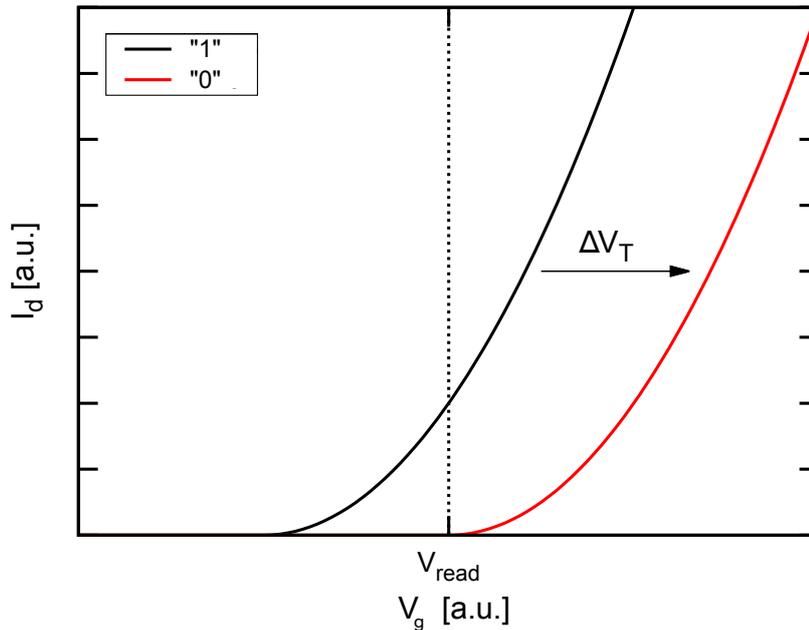
### 1.3.1 Funzionamento di una memoria Flash NAND

La cella di una memoria Flash è identica a un transistor con l'eccezione di uno strato di materiale, posto tra l'ossido di gate e una serie di strati di materiali isolanti che lo separano dal gate, che serve a raccogliere portatori provenienti dal canale iniettati per effetto tunnel. Nelle memorie Flash planari questo strato è costituito da polisilicio drogato ad altissime concentrazioni ed è equivalente ad un gate intermedio tra il canale e il gate. Viene chiamato *floating-gate* (FG). La carica immagazzinata ( $Q$ ) induce tramite  $C_{ox}$  un'identica densità di carica, ma con segno opposto, nel gate. Questa carica indotta cambierà la  $V_t$  della cella di una quantità:

$$V_t = V_{t_0} - \frac{Q}{C_{ox}}, \quad (1.9)$$

dove  $V_{t_0}$  è la tensione di soglia della cella neutra. Il cambiamento di soglia  $\Delta V_t$  produce una traslazione rigida della transcaratteristica, come mostrato in Figura 1.9. Se ci posizioniamo ad una tensione intermedia tra i due stati, detta  $V_{read}$ , e applichiamo una tensione  $V_d$  al drain della cella, avremo in uscita un valore della corrente  $I_d$  alto oppure basso a seconda che vi sia o meno carica nel FG. Possiamo allora identificare questi due valori come due diversi livelli logici e memorizzare quindi un'informazione binaria nella cella. La cella nello stato a  $V_t$  bassa è detta "cancellata" e il livello logico corrispondente è "1". La cella nello stato a  $V_t$  alta è detta "programmata" e il livello logico corrispondente è "0".

Nella struttura NAND le celle vengono collegate in serie lungo le *bit-line* (BL) con il primo transistor della stringa collegato all'alimentazione  $V_{dd}$  e il source dell'ultimo transistor collegato alla *source-line* (SL) per la lettura della corrente. I gate delle celle che occupano la stessa posizione nelle diverse BL vengono invece collegati lungo le *word-line* (WL) in direzione perpendicolare alle BL. Ogni cella è così indirizzabile singolarmente e



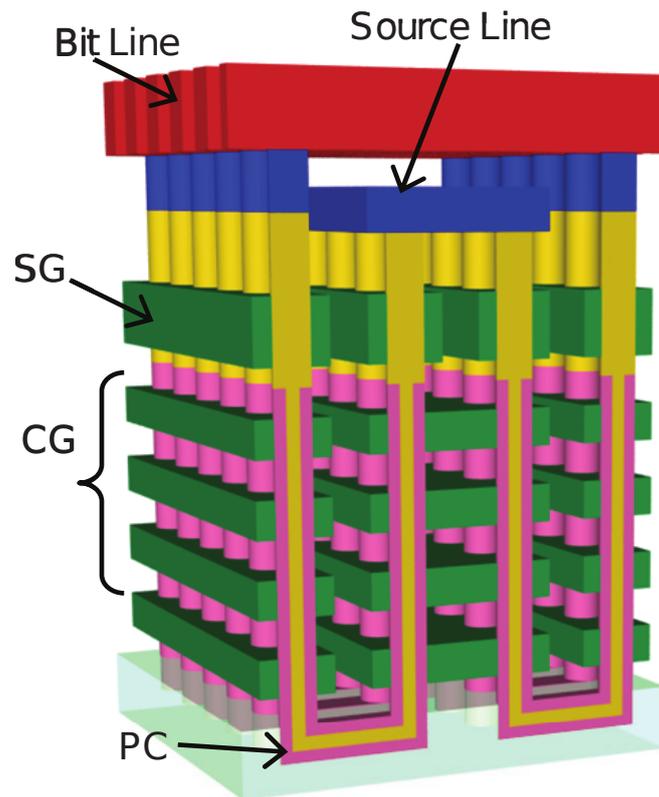
**Figura 1.9:** Transcaratteristica di una cella Flash negli stati "cancellata" (nero) e "programmata" (rosso). Viene anche indicata la tensione di lettura  $V_{read}$ .

il suo dato può venir letto portando a  $V_{dd}$  la BL corrispondente, imponendo una tensione  $V_{read}$  all'WL della cella in questione e alzando le altre celle della BL ad una tensione  $V_g \gg V_{read}$  in modo che facciano da pass-transistor.

### 1.3.2 Le Flash NAND 3D e la cella 'macaroni'

Una memoria Flash potrà contenere tanti più dati quante più celle possiede, e quindi si comprende come l'obiettivo principale per i produttori di memorie Flash in tutti questi anni sia stato di aumentare sempre più il numero di celle integrate in un dispositivo, tramite uno scaling aggressivo delle dimensioni caratteristiche dei transistor a FG. Una volta raggiunta la minima lunghezza di canale consentita dalla struttura planare (sezione 1.1.3) anche nel caso delle memorie Flash sono state introdotte nuove geometrie della cella, per poter mantenere il trend di scaling imposto dal mercato.

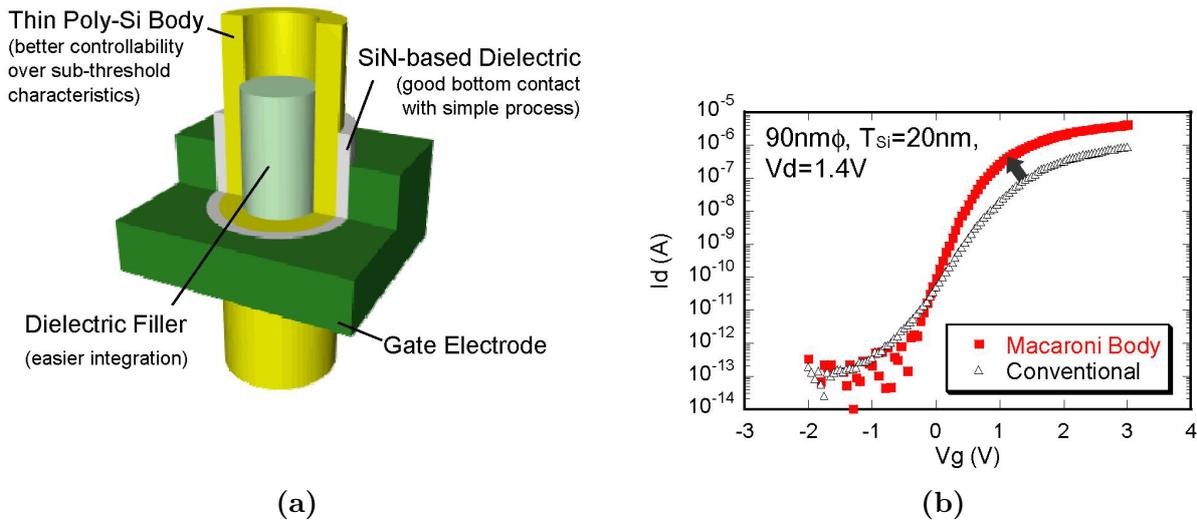
Le nuove memorie Flash hanno una cella GAA a simmetria cilindrica, che consente di estendere l'integrazione delle celle anche in verticale, formando in questo modo degli array tridimensionali (3D NAND). La dimensione del canale al momento rimane invariata e i nodi tecnologici successivi vengono raggiunti aumentando gli strati verticali di celle. Due tipologie di 3D NAND messe in commercio già da qualche anno sono le strutture P-BiCS (Pipe-shaped Bit Cost Scalable) di Toshiba (Figura 1.10) [13] e TCAT (Terabit Cell Array Transistor) di Samsung [14].



**Figura 1.10:** Vista prospettica dell'array 3D PBiCS di Toshiba. Sono presenti la BL e la SL in direzione orizzontale e le WL che si diramano perpendicolarmente ad esse, come nel caso planare. In aggiunta sono presenti più strati verticali di WL, che estendono l'integrazione nella terza dimensione (da [13]).

La cella (Figura 1.11a) è molto simile alla struttura nanowire (Figura 1.4d), ma a differenza di quest'ultima il corpo del canale è cavo e viene riempito con del dielettrico per facilitarne l'integrazione. La ragione di questa modifica sta nel fatto che il canale di una cella 3D NAND è in polisilicio. Infatti nel flusso di processo che viene usato per realizzare tali array si parte da una serie di strati alternati di isolante e materiale conduttivo che costituirà le WL, si effettuano dei fori, vengono depositati i materiali dello strato di dielettrici tra il gate e il canale e infine viene depositato il silicio. Non avendo una struttura cristallina come base, il silicio depositato avrà una struttura policristallina. Nel polisilicio c'è un'alta densità di trappole lungo le frontiere dei grani e queste catturano elettroni che agiscono da centri di scattering degradando la conduzione. Per cercare di ridurre l'effetto di queste trappole viene diminuito lo spessore di silicio realizzando un canale cavo. Questa struttura viene chiamata *macaroni*.

Avere uno spessore del canale minore comporta anche un migliore controllo su di esso da parte del gate e quindi caratteristiche elettrostatiche superiori. In Figura 1.11b sono messe a confronto le transcaratteristiche di due celle, una con il canale completamente



**Figura 1.11:** (a) Cella di una memoria Flash 3D. La geometria è quella GAA con il gate che circonda il canale. Il corpo è in polisilicio e presenta una cavità riempita di dielettrico. Tale struttura viene detta ‘macaroni’. Nelle 3D NAND il FG viene sostituito da una pila di strati dielettrici,  $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2$  (SONOS). L’immagazzinamento della carica avviene per intrappolamento degli elettroni nei difetti del  $\text{Si}_3\text{N}_4$ , che rimangono dunque fissi riducendo i fenomeni di accoppiamento capacitivo tra celle adiacenti. (b) Transcaratteristica di due celle con lo stesso diametro e la stessa lunghezza di canale, una con il corpo pieno e l’altra con struttura macaroni. Si nota come la struttura macaroni migliori le caratteristiche elettrostatiche, avendo una  $V_t$  minore, una STS maggiore e una corrente di conduzione più elevata (da [15]).

riempito e l’altra con il corpo macaroni. Nel caso macaroni si può osservare quanto detto, una  $V_t$  minore dovuta alla diminuzione del numero di trappole, una maggiore corrente  $I_d$  soprasoglia e una pendenza di sottosoglia più alta.

## 1.4 La variabilità atomistica

Abbiamo visto come gli effetti di canale corto rappresentino una grossa fonte di variabilità per i transistor che vogliono raggiungere dimensioni sempre più scalate (sezione 1.1.1). Tuttavia anche riducendo tali effetti le cause di variabilità nei transistor non vengono eliminate. Infatti quando si raggiungono dimensioni inferiori ai 100 nm la discretizzazione della materia comincia ad avere un peso decisivo. Una visione continua, cioè in cui tutti i dispositivi in media sono uguali, non è più possibile e diventa rilevante la singola realizzazione. Ad esempio, in seguito alle tolleranze di processo, il gate può variare le sue dimensioni da dispositivo a dispositivo a causa di *Line Edge Roughness*, oppure lo strato di ossido di gate può avere spessori diversi data l’incertezza di realizzazione intrinseca di

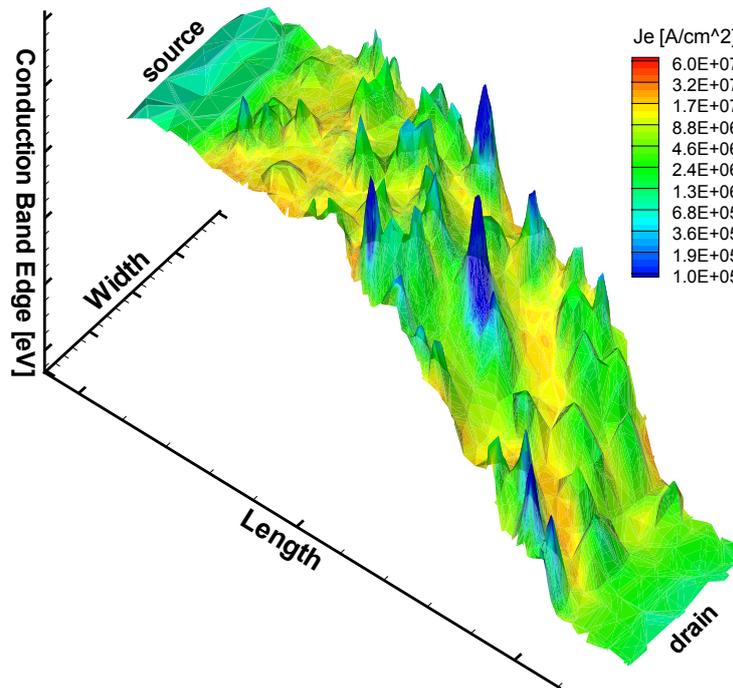
qualche strato atomico, sulle poche decine di cui arrivano a essere composti nei transistor moderni.

Queste differenze da un componente all'altro del chip portano ad avere diverse tensioni di soglia e quindi a variabilità nel funzionamento. Le fluttuazioni dei parametri spingono ad un'analisi statistica dei dispositivi realizzati, che comprenderà un valore medio della grandezza in questione e una distribuzione di probabilità per i casi che si discostano da questo. La principale fonte di variabilità che è stata individuata per i transistori è legata alla discretizzazione degli atomi droganti e comprende i fenomeni di *Random Dopant Fluctuations* (RDF) e *Random Telegraph Noise* (RTN).

### 1.4.1 Random Dopant Fluctuations (RDF)

Quando si realizza un'impiantazione di drogante, il numero di ioni che penetrano nell'area del singolo transistor non è controllabile in modo esatto. In seguito gli ioni diffondono nel materiale e la loro posizione varia in modo casuale. Se i dispositivi hanno dimensioni ridotte il numero di atomi che sono al loro interno è piccolo e quindi ogni drogante avrà un peso determinante sull'elettrostatica del transistor. Un atomo accettore ionizzato (se prendiamo un n-MOS) ha un effetto elettrostatico repulsivo sugli elettroni e dunque crea un picco di potenziale che allontana i portatori nelle sue vicinanze. Questo effetto domina il flusso di corrente soprattutto in condizioni di inversione debole, dove gli accettori ionizzati non sono ancora schermati dagli elettroni dello strato di inversione. Se le dimensioni del canale del dispositivo sono grandi, il numero di droganti contenuti al suo interno è alto e il loro effetto si media risultando in una barriera di potenziale uniforme tra source e drain per gli elettroni in condizioni di sottosoglia. Quando invece il dispositivo è piccolo, un atomo in una certa posizione alza la barriera di potenziale in quella zona e determina un cambiamento di  $V_t$  a livello globale ([16]). Diverse configurazioni di droganti porteranno a valori differenti di tensione di soglia.

Le fluttuazioni del potenziale superficiale prodotte dalle posizioni casuali degli atomi spingono la corrente a fluire attraverso le valli del profilo, generando quella che viene detta percolazione della corrente. In Figura 1.12 sono raffigurate l'energia della banda di conduzione per una certa distribuzione casuale di droganti e la densità di corrente che attraversa l'area di dispositivo in oggetto. Si possono vedere i picchi di potenziale e le valli che questi generano, dove la corrente si addensa in filamenti poiché si crea un passaggio preferenziale. L'RDF consiste nella dispersione della tensione di soglia a causa della variazione casuale della posizione dei droganti di un transistor. Il fenomeno è stato molto studiato nelle geometrie planari. Simulazioni statistiche 3D hanno portato alla formulazione di una legge empirica per la dipendenza dallo scaling della deviazione

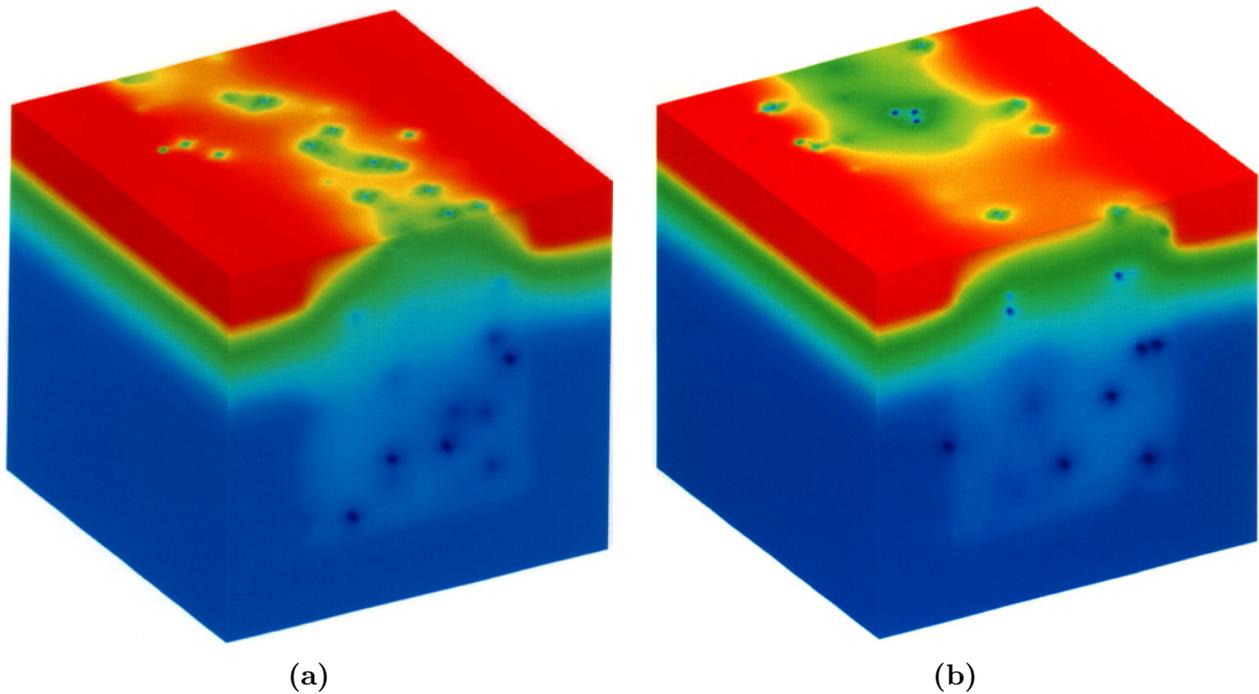


**Figura 1.12:** Energia della banda di conduzione per una sezione del transistor in direzione perpendicolare alla lunghezza del canale. Sul grafico sono mostrate anche le superfici di livello della densità di corrente lungo la sezione. La corrente si concentra nei percorsi più favorevoli, ovvero attraverso le valli del profilo di potenziale, dando origine alla percolazione (da [17]).

standard della distribuzione delle tensioni di soglia di un MOSFET [18]:

$$\sigma_{V_t} = 3.19 \times 10^{-8} \frac{t_{ox} N_A^{0.4}}{\sqrt{W_{eff} L_{eff}}} [V] \quad (1.10)$$

Questa formula ha una dipendenza più forte dalla concentrazione di droganti rispetto a formule analitiche che si ottengono considerando le variazioni statistiche del numero medio dei droganti nella regione svuotata [2] o del loro numero medio e della loro posizione [19]. Ciò testimonia l'importanza dei singoli atomi rispetto al loro comportamento medio. Per esempio se un dispositivo ha degli accettori allineati più o meno uniformemente lungo la dimensione della larghezza, tutti nelle vicinanze di un determinato punto del canale, questi formeranno una barriera di potenziale che innalzerà la tensione di soglia. Al contrario, se in una certa zona del dispositivo i droganti sono in profondità rispetto alla superficie creeranno un percorso percolativo per gli elettroni e in questo caso  $V_t$  diminuirà [16]. La situazione appena descritta è mostrata in Figura 1.13. La (1.10) mostra che le fluttuazioni della tensione di soglia diminuiscono se si riduce lo spessore dell'ossido. Infatti, come visto in sezione 1.1.2, se lo strato di ossido diventa più sottile il gate viene avvicinato maggiormente al canale e ha un migliore controllo su di esso. Sempre osservando la (1.10) notiamo che la dispersione dei valori di  $V_t$  aumenta al diminuire di  $W$  e  $L$ . Questo è



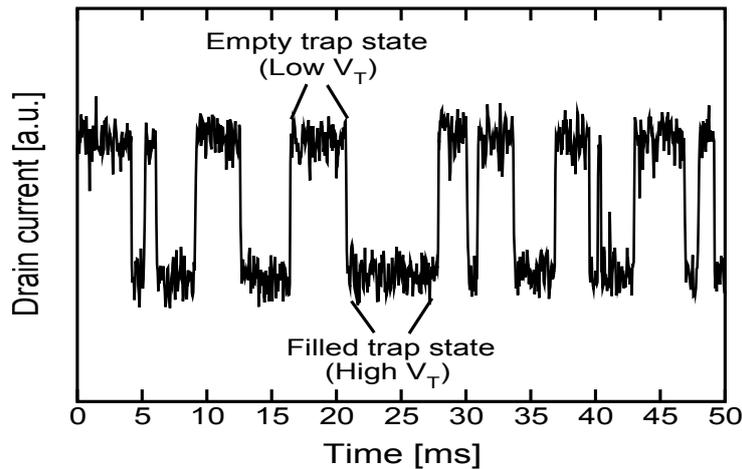
**Figura 1.13:** Distribuzione di potenziale all'interfaccia Si-SiO<sub>2</sub> per due n-MOS con lo stesso numero di accettori nella regione attiva ma differenti a livello microscopico. (a) La disposizione più o meno uniforme degli atomi lungo una linea perpendicolare al canale forma una barriera che riduce il passaggio di elettroni e alza  $V_t$ . (b) Gli accettori si concentrano prevalentemente in un'area del canale lasciando scoperta una zona. In questa zona gli elettroni vedranno una barriera di potenziale minore e passeranno più facilmente dal source al drain attraverso tale percorso, abbassando  $V_t$  (da [16]).

riconducibile al fatto che se il dispositivo diventa più piccolo aumenta l'influenza dei singoli atomi, poiché ce ne sono meno sotto il gate, e un picco di potenziale o una valle che consente il passaggio della corrente hanno un effetto più forte sulla fluttuazione di  $V_t$ . Infine la dipendenza dal doping è spiegabile col fatto che aumentando il numero di droganti presenti si aggiungono nuove possibilità di disposizione atomistiche che disperderanno ulteriormente la tensione di soglia.

### 1.4.2 Random Telegraph Noise (RTN)

L'RTN nei transistor è una fonte di rumore che presenta un andamento temporale simile a quello mostrato in Figura 1.14.

La causa di questo rumore è attribuita ai siti trappola che si formano all'interfaccia Si-SiO<sub>2</sub> o nel volume dell'ossido di gate. I difetti dovuti all'interruzione della periodicità del reticolo cristallino generano degli stati discreti all'interno del Band-Gap che consentono ad un elettrone di avere un'energia altrimenti proibita. La probabilità che questi stati vengano occupati dipende dalla posizione del livello di Fermi del materiale rispetto ad essi.



**Figura 1.14:** Andamento temporale della corrente di drain in funzione del tempo per una data densità di gate. Si possono vedere chiaramente i due livelli discreti di corrente associati agli stati con trappola piena ( $V_t$  alta) e trappola vuota ( $V_t$  bassa) (da [20]).

Gli elettroni del silicio che sono vicini all'interfaccia con l'ossido possono effettuare transizioni tra la banda di conduzione e gli stati all'interno del Band-Gap. Un portatore che occupa un tale stato viene sottratto alla banda di conduzione, come se venisse catturato, e non partecipa più alla corrente. Gli stati all'interfaccia sono perciò chiamati trappole. Il processo di cattura ed emissione di un elettrone da parte di una trappola avviene con due costanti di tempo caratteristiche  $\tau_c$  e  $\tau_e$ , diverse per ogni stato di interfaccia.

Se la trappola è occupata, nel transistor scorrerà una corrente minore rispetto al caso in cui la trappola sia vuota. Inoltre l'elettrone intrappolato agisce come centro di scattering che riduce la mobilità dei portatori che scorrono nelle vicinanze, causando un'ulteriore diminuzione della corrente. Per ottenere la stessa corrente dovremo alzare la tensione del gate e quindi l'effetto è equivalente ad un aumento della tensione di soglia. I due stati sono ben visibili in Figura 1.14 dove al livello di corrente alto, dunque  $V_t$  bassa, corrisponde il caso di trappola vuota mentre al livello di corrente basso, dunque  $V_t$  alta, corrisponde il caso di trappola occupata.

L'RTN rappresenta quindi una fonte di variabilità per i transistor integrati in un chip in quanto la  $V_t$  di un dispositivo non sarà sempre la stessa ma assumerà dei valori all'interno di un determinato intervallo, distribuiti secondo una certa probabilità. Le ampiezze di RTN vengono misurate in termini di  $\Delta V_t$ , definito come la differenza tra la tensione di soglia del dispositivo con l'elettrone intrappolato e la tensione di soglia con la trappola vuota. Determinare l'ampiezza di queste fluttuazioni è di fondamentale importanza per poter fare previsioni sull'affidabilità del circuito.

Il primo modello che è stato formulato considerava la variazione percentuale della corrente nel MOSFET dovuta al cambiamento del numero di portatori e alla varia-

ne di eventi di scattering che, come accennato sopra, l'intrappolamento di un elettrone comporta. Possiamo scrivere [21]:

$$\frac{\Delta I_d}{I_d} = \frac{1}{WL} \left( \frac{1}{N} \frac{\partial N}{\partial N_t} \pm \frac{1}{\mu} \frac{\partial \mu}{\partial N_t} \right), \quad (1.11)$$

dove  $N_t$  è il numero di portatori intrappolati per unità di area e  $N$  è il numero di portatori nel canale per unità di area. Nei dispositivi molto scalati di solito è presente al massimo una trappola, cioè  $\frac{\partial N}{\partial N_t} = -1$ . Sostituendo nella (1.11) si ottiene una variazione percentuale di corrente pari a:

$$\frac{\Delta I_d}{I_d} = \frac{1}{WL} \left( \frac{1}{N} \pm \frac{1}{\mu} \frac{\partial \mu}{\partial N_t} \right). \quad (1.12)$$

La (1.4.2) può poi essere ricondotta alla variazione di tensione di soglia  $\Delta V_t$  tramite :

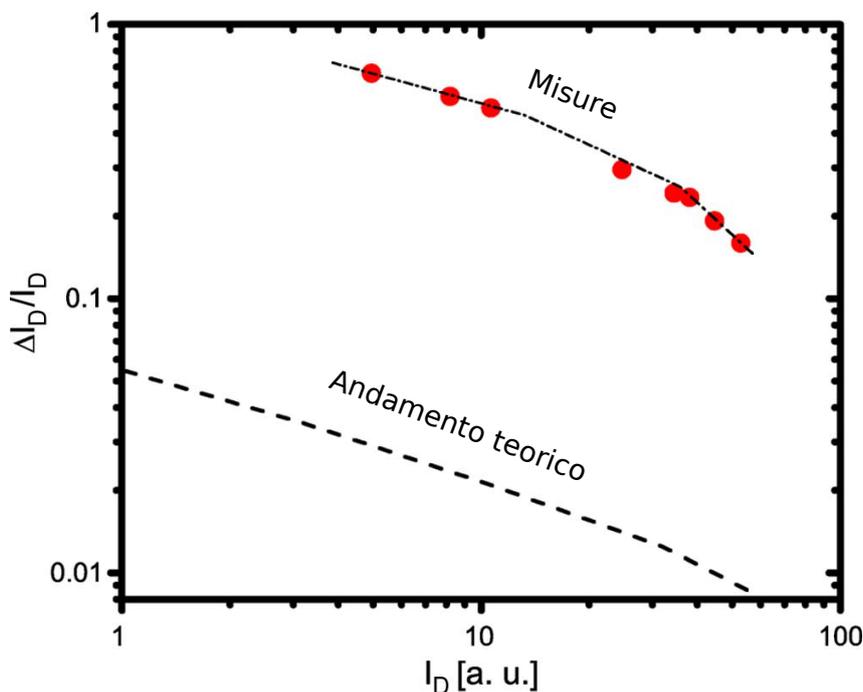
$$\Delta V_t = \frac{\Delta I_d}{g_m}, \quad (1.13)$$

dove  $g_m$  è la transconduttanza del transistor in regime lineare. Le fluttuazioni di mobilità dovute all'aggiunta di un singolo centro di scattering sono trascurabili rispetto alla fluttuazione del numero di portatori [22]. Tutta la variazione della tensione di soglia è quindi da associare alla fluttuazione del numero di elettroni di conduzione. Tuttavia i valori di RTN misurati sperimentalmente arrivano a superare quelli previsti fino ad un ordine di grandezza in più [22], [23], tanto da essere chiamati *Giant RTN*. Un esempio di questo fenomeno è mostrato in Figura 1.15, dove il valore percentuale della variazione di corrente misurato su una cella Flash reale è confrontato con l'andamento previsto dalla teoria. All'aumentare della corrente l'effetto dell'RTN diminuisce poiché quando il dispositivo passa sopra soglia e lo strato di elettroni all'interfaccia schermo l'effetto della trappola.

La distribuzione di probabilità cumulativa dei valori di  $\Delta V_t$  misurati tra due letture a tempi differenti di alcune celle di memorie Flash diverse è raffigurata in Figura 1.16. Ogni curva è caratterizzata da un andamento esponenziale della probabilità di avere valori di  $\Delta V_t$  elevati:

$$p(\Delta V_t) = \frac{1}{\sigma} e^{-\frac{|\Delta V_t|}{\sigma}}. \quad (1.14)$$

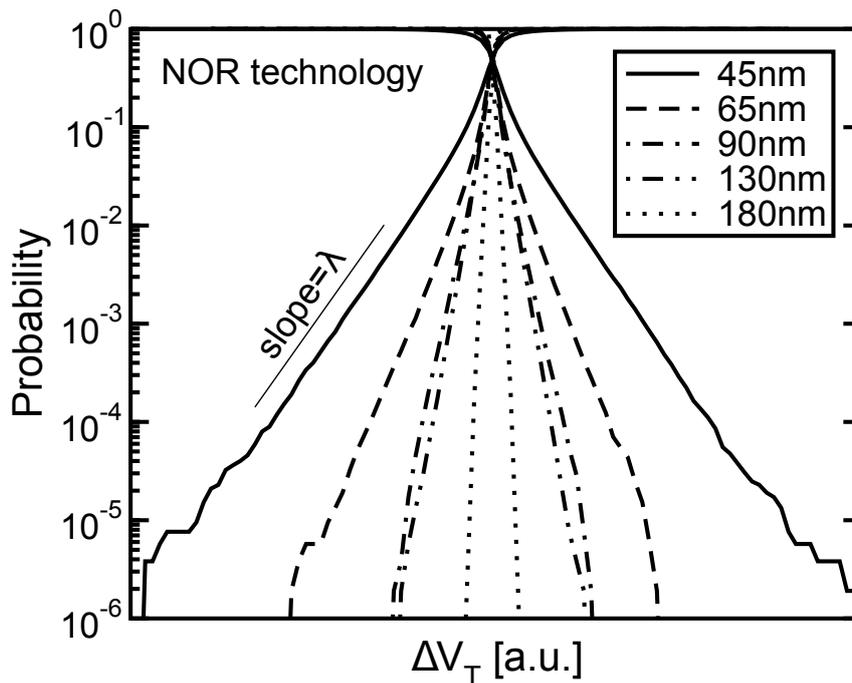
Su un grafico semi-logaritmico ne risulta una pendenza  $\lambda$  (mV/decade) per la coda della cumulativa. La pendenza  $\lambda$  cresce al diminuire delle dimensioni dei transistor, per i quali aumenta dunque la probabilità di avere Giant RTN. Anche questo fenomeno è allora da ricondursi alla discretizzazione della materia e rappresenta un serio problema per dispositivi molto scalati. Per dispositivi con un gran numero di transistor integrati, cioè prolungando la coda  $\lambda$ , si possono avere  $\Delta V_t$  elevati.



**Figura 1.15:** Variazione percentuale della corrente di drain ( $\frac{\Delta I_d}{I_d}$ ) in funzione della corrente di drain stessa ( $I_d$ ) in seguito alla presenza di una trappola all'interfaccia Si-SiO<sub>2</sub> che genera RTN. I punti rappresentano i dati sperimentali mentre la linea tratteggiata mostra l'andamento previsto dalla 1.4.2. I valori misurati sono maggiori di quasi un ordine di grandezza rispetto alle previsioni. Questo andamento rappresenta un esempio di *Giant RTN* (da [23]).

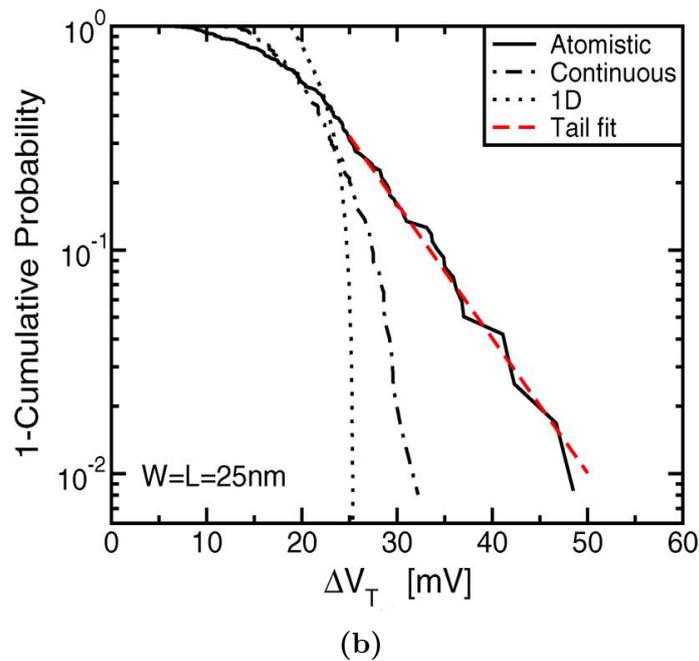
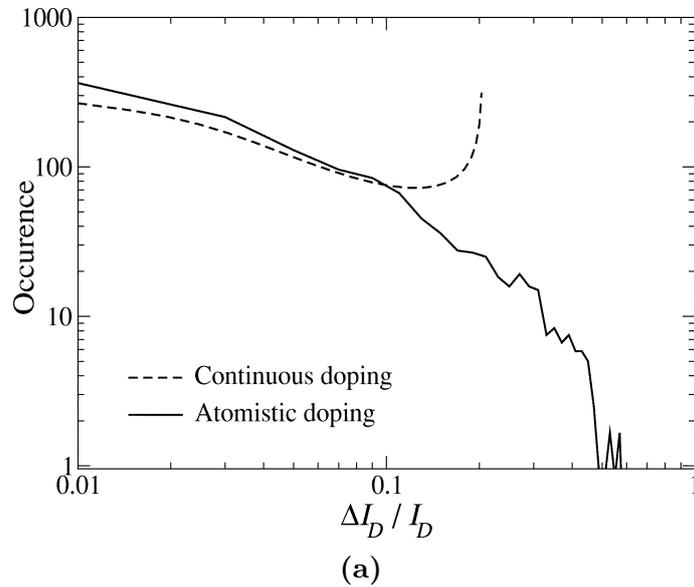
Il fenomeno dell'RTN nei dispositivi decananometrici è stato ampiamente studiato sia per i transistor ([25]) che per le memorie Flash ([26], [27]).

Se si considera che la trappola è una quantità discreta e localizzata nel canale, una visione che considera il suo effetto sulle grandezze medie del dispositivo risulta non essere corretta. In [25] viene studiato tramite simulazioni numeriche 3D l'effetto della posizione della trappola sulle ampiezze di RTN registrate in un MOSFET di tipo bulk con una concentrazione continua di accettori nel canale. Le trappole interfacciali che sono localizzate vicino al source e al drain causano delle variazioni di corrente molto piccole. L'effetto maggiore lo hanno invece le trappole che si trovano nei pressi del centro del canale. Tale andamento è intuibile se si pensa che nel centro del canale il potenziale superficiale ha il suo massimo e la concentrazione di elettroni sottosoglia è molto minore che vicino al source e al drain. Quindi la variazione in termini di numero di portatori che causa un elettrone intrappolato in mezzo al canale è molto maggiore della variazione che causerebbe se fosse intrappolato presso il source o il drain. I risultati di queste simulazioni evidenziano che sono possibili più valori per le ampiezze di RTN, che saranno quindi distribuiti secondo una certa probabilità. Tuttavia non vengono raggiunti i valori di Giant RTN registrati nei dispositivi reali. Infatti la distribuzione cumulativa del caso con doping continuo non



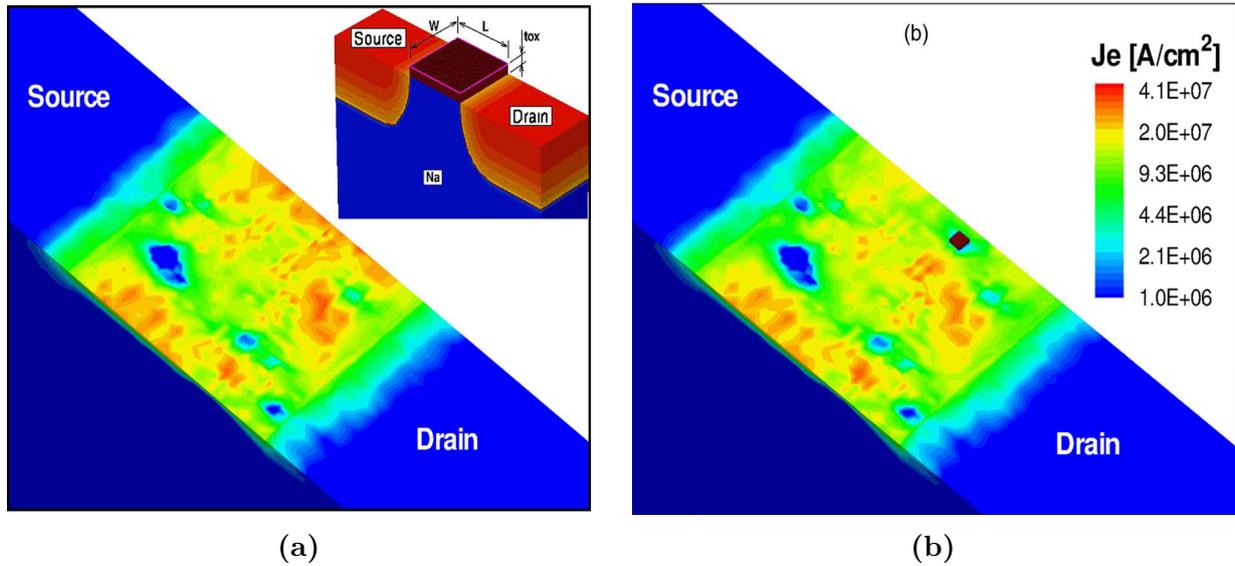
**Figura 1.16:** Distribuzioni cumulative  $F$  e  $1-F$  del  $\Delta V_t$  indotto da RTN per diverse tecnologie di memorie Flash NOR. Al diminuire delle dimensioni della cella aumenta la probabilità che si verifichino valori di Giant RTN. La pendenza della coda esponenziale della distribuzione è  $\lambda$  (mV/decade) (da [24]).

presenta una coda esponenziale ma possiede invece due picchi, uno per valori bassi dello shift e uno per valori alti (Figura 1.17a). Questo andamento è conseguenza dell'effetto che produce una trappola a seconda che sia vicino a source e drain del dispositivo oppure al centro del canale. Il picco a valori bassi di  $\frac{\Delta I_d}{I_d}$  è dovuto agli elettroni intrappolati vicino al source e al drain, quello a valori alti è dato da trappole nei pressi della zona centrale del canale. La distribuzione delle simulazioni con il doping continuo non presenta una coda esponenziale ma satura ad un valore massimo perché non ci sono altri casi possibili, oltre a quello di trappola centrale, che possano causare uno shift alto. Mueller e Schulz già nel 1996 intuirono che per avere variazioni percentuali della corrente di drain così grandi la densità di corrente lungo il canale non poteva essere uniforme [22]. Se la corrente si addensa principalmente lungo alcuni percorsi, un elettrone intrappolato nei pressi di uno di questi limita in modo molto più forte la corrente che se questa scorresse uniformemente lungo il canale. Le ragioni del Giant RTN sono dunque i percorsi percolativi della corrente, che diventano sempre più rilevanti col ridursi delle dimensioni del dispositivo, in quanto diminuendo il numero di quelli possibili ogni singolo percorso assume un'importanza maggiore. La causa della percolazione venne attribuita in principio alle cariche fisse presenti nell'ossido. Tuttavia nei dispositivi decananometrici queste cariche sono quasi del tutto assenti, mentre i valori anomali di RTN si verificano lo stesso.



**Figura 1.17:** Confronto tra le distribuzioni cumulative 1-F predette dai diversi modelli. (a) Distribuzione cumulativa dei casi 3D con doping continuo (linea tratteggiata) e 3D con doping atomistico (linea continua). Il caso con doping continuo ha una distribuzione che presenta due picchi e che non oltrepassa un certo valore di  $\Delta V_t$ , a dimostrazione dell'assenza di percorsi percolativi (da [25]). (b) Grafico in scala semi-logaritmica dei tre modelli per l'RTN: 1D, 3D con doping continuo e 3D con droganti discreti. Il modello atomistico è l'unico che riproduce la coda esponenziale della distribuzione dovuta ai fenomeni di percolazione. (da [28]).

La spiegazione dell'RTN osservato nei dispositivi attuali si può trovare allora nella percolazione indotta dai droganti casuali, di cui è stato discusso in sezione 1.4.1. Con



**Figura 1.18:** Esempio di profilo della densità di corrente per il transistor mostrato nel riquadro, calcolata tramite simulatore 3D. (a) Transistor con trappola vuota. Ai bordi dell'area attiva ci sono due percorsi preferenziali per la corrente, mentre verso il centro si trovano alcuni accettori che bloccano il passaggio. (b) Transistor con trappola occupata. La trappola si trova nei pressi di uno dei due filamenti di corrente e lo annulla in gran parte. Ne deriva un aumento di  $V_t$  di tipo Giant RTN (da [27]).

lo scaling dei dispositivi le posizioni dei singoli atomi assumono un'importanza decisiva e generano dei percorsi preferenziali per la corrente nei punti in cui creano delle valli nel profilo del potenziale superficiale. In Figura 1.18 si può vedere che una trappola posizionata sopra uno di questi percorsi lo elimina quasi totalmente. Il profilo di corrente che ne risulta, a pari tensione di gate, è molto più debole di quello del transistor senza trappola e per avere la stessa corrente sarà necessaria una  $V_g$  maggiore. Questo determina un aumento della soglia e quindi uno shift  $\Delta V_t$  tra due condizioni diverse dello stesso dispositivo. Si può vedere infine in Figura 1.17b la densità di probabilità cumulativa in un grafico semi-logaritmico dei tre diversi modelli che descrivono il fenomeno di RTN: 1D, 3D con doping continuo e 3D con drogaggio atomistico. La distribuzione del caso 1D è molto più stretta delle altre due e l'unica fonte di variabilità è la posizione della trappola lungo lo spessore dell'ossido. L'effetto dell'elettrostatica 3D aumenta la dispersione della distribuzione con doping continuo, senza avere però la coda esponenziale caratteristica dei fenomeni di percolazione. La distribuzione 3D con drogaggio discreto è l'unica che riproduce in maniera corretta i dati sperimentali di Figura 1.16.

In sezione 1.4.1 si è visto che la natura del fenomeno di RDF è la stessa di quello di RTN, ovvero il formarsi di filamenti di corrente dovuti alla configurazione che assumono gli accettori discreti presenti sotto il gate. Tuttavia mentre la distribuzione di probabilità delle ampiezze del primo è una normale, quella del secondo è esponenziale. La ragione di

ciò è che in realtà la cumulativa delle ampiezze di RTN, tramite la differenza  $\Delta V_t$  tra  $V_t$  di due distribuzioni gaussiane diverse, mette in relazione a due a due dei valori correlati, mentre la cumulativa dell'RDF analizza i valori di  $V_t$  nel loro insieme.

In [27] viene fornita una formula che analizza le dipendenze della pendenza  $\lambda$  della coda, in maniera analoga a quanto fatto in [18] per la deviazione standard della distribuzione di  $V_t$  causata da RDF:

$$\lambda = K \frac{t_{ox}^\alpha \sqrt{N_a}}{W \sqrt{L}}, \quad (1.15)$$

con  $\alpha$  leggermente minore di 1. I parametri da cui dipende  $\lambda$  sono gli stessi visti per l'RDF e anche gli andamenti di tale dipendenza sono molto simili. Il motivo è che entrambi i fenomeni sono causati dalla percolazione della corrente. Se aumenta il peso che i filamenti hanno nella conduzione, le fluttuazioni della tensione di soglia sono maggiori. Allo stesso tempo aumentano anche le ampiezze di RTN, poiché l'impatto di una trappola su uno dei percorsi di corrente diventa più forte. Rispetto alla (1.10) si nota una dipendenza maggiore dalla larghezza dell'area attiva  $W$  e dal doping di canale  $N_a$ . Se riduciamo  $W$ , i percorsi percolativi possibili per la corrente diminuiscono e quindi l'annullamento di uno di essi ha un peso più incisivo rispetto al caso di un dispositivo più largo in cui ci sono maggiori possibilità di passaggio. Anche l'aumento del doping, che accentua le disuniformità nel profilo di potenziale superficiale, riduce il numero dei filamenti che si creano e intensifica l'intensità di quelli presenti. Quindi, come accade diminuendo  $W$ , anche aumentando  $N_a$  l'impatto che una trappola ha su  $V_t$  diventa più forte.

# Capitolo 2

## Simulazioni numeriche 3D

### 2.1 Il modello drift-diffusion (DD)

In un semiconduttore il trasporto dei portatori è governato da due differenti meccanismi, la deriva (*drift*) e la diffusione (*diffusion*).

Quando viene applicato un campo elettrico  $\vec{\mathcal{E}}$  ad un mezzo conduttore, i portatori acquistano una velocità di drift  $\vec{v}_d$  che si impone sul loro moto di agitazione termica in direzione casuale. La velocità  $\vec{v}_d$  assume un valore finito poiché gli elettroni e le lacune ad ogni collisione con il reticolo cristallino perdono l'accelerazione acquisita. Se il valore del campo  $\vec{\mathcal{E}}$  non è troppo elevato si può scrivere:

$$\vec{v}_d = \mu \vec{\mathcal{E}}, \quad (2.1)$$

dove  $\mu$  è la mobilità ( $\frac{\text{cm}^2}{\text{V}\cdot\text{s}}$ ) dello specifico tipo di portatore. Una concentrazione di elettroni  $n$  ( $\text{cm}^{-3}$ ) che si muove con velocità  $\vec{v}_d$  darà origine ad una densità di corrente:

$$\vec{J}_{n,drift} = qn\mu_n\vec{\mathcal{E}}, \quad (2.2)$$

e allo stesso modo per la corrente di drift delle lacune  $\vec{J}_{p,drift}$ .

Se la concentrazione di portatori all'interno del materiale non è uniforme, questi diffonderanno a causa del gradiente di concentrazione generando una densità di corrente che può essere scritta come:

$$\vec{J}_{n,diff} = qD_n\nabla n, \quad (2.3)$$

dove  $D_n$  è il coefficiente di diffusione ( $\frac{\text{cm}^2}{\text{s}}$ ) per gli elettroni. La densità di corrente di

diffusione per le lacune  $\vec{J}_{p,diff}$  si scrive come la (2.3) ma con segno cambiato poiché la corrente scorre nel verso opposto del gradiente dei portatori. Le correnti totali di elettroni  $\vec{J}_n$  e di lacune  $\vec{J}_p$  saranno quindi la somma dei due contributi di drift e di diffusione.  $\vec{J}_n$  e  $\vec{J}_p$  soddisfano le rispettive equazioni di continuità:

$$q \frac{\partial n}{\partial t} = \text{div}(\vec{J}_n) + q(G - R), \quad (2.4)$$

$$q \frac{\partial p}{\partial t} = -\text{div}(\vec{J}_p) + q(G - R), \quad (2.5)$$

dove  $G$  e  $R$  sono i tassi di generazione e ricombinazione per unità di volume ( $\text{cm}^{-3} \text{s}^{-1}$ ).

Nel semiconduttore i portatori liberi e gli atomi di drogante ionizzati determinano il potenziale elettrostatico, secondo l'equazione di Poisson:

$$\text{div}(\nabla\varphi) = -\frac{1}{\varepsilon_{Si}}(p - n + N_d - N_a), \quad (2.6)$$

dove  $p$ ,  $n$ ,  $N_d$ ,  $N_a$  sono le concentrazioni rispettivamente di lacune, elettroni, atomi donori e atomi accettori.

Se si vogliono determinare le proprietà elettriche di un transistor devono essere risolte le precedenti equazioni, che rappresentano un sistema accoppiato:

$$\begin{cases} \text{div}(\nabla\varphi) = -\frac{1}{\varepsilon_{Si}}(p - n + N_d - N_a) & (2.7a) \\ q \frac{\partial n}{\partial t} = \text{div}(qn\mu_n\vec{\mathcal{E}} + qD_n\nabla n) + q(G - R) & (2.7b) \\ q \frac{\partial p}{\partial t} = -\text{div}(qn\mu_p\vec{\mathcal{E}} - qD_p\nabla p) + q(G - R). & (2.7c) \end{cases}$$

Il sistema (2.7) è il sistema *drift-diffusion* (DD).

Le (2.7), con opportune condizioni al contorno, sono risolvibili analiticamente soltanto in 1D e con alcune approssimazioni. Se si vogliono implementare modelli fisici complessi o analizzare geometrie 3D, per risolvere il sistema DD ci si deve avvalere delle simulazioni numeriche. Il caso dei droganti discreti, in cui le diverse posizioni degli ioni fanno variare il profilo del potenziale superficiale e il loro effetto elettrostatico si estende in tutte e tre le direzioni dello spazio, rappresenta appunto una di queste situazioni.

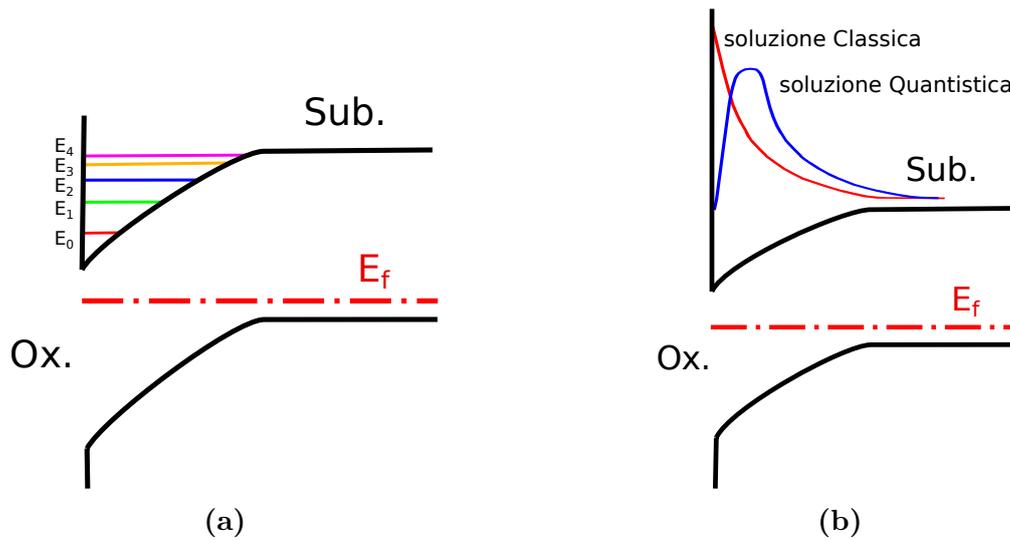
In un simulatore le equazioni differenziali (2.7) vengono trasformate in un sistema di equazioni algebriche tramite la costruzione di un dominio discreto (griglia, *mesh*) che viene mappato sulla geometria che si vuole analizzare. Le soluzioni sono quindi soluzioni approssimate e definite a tratti tra i singoli punti del dominio. Si deduce facilmente che per approssimare soluzioni con gradienti elevati o strutture complesse sia meglio avere

una griglia molto fitta. Tuttavia più sono gli elementi di cui è costituita la griglia e maggiore è la dimensione delle matrici che deve risolvere il calcolatore, aumentando lo sforzo computazionale e allungando i tempi della simulazione. La griglia non può quindi essere ridotta all'infinito. Il sistema (2.7) è costituito da tre equazioni accoppiate. La soluzione finale delle variabili principali può essere ricavata direttamente, linearizzando il sistema, oppure viene calcolata disaccoppiando le equazioni e risolvendo iterativamente un'equazione alla volta. In entrambi i casi la soluzione temporanea viene aggiornata finché non vengono soddisfatti dei criteri prestabiliti sull'errore numerico che si accetta di trascurare. Il primo metodo è più lento del secondo, poiché richiede un maggiore sforzo computazionale, però la sua convergenza è migliore. Il software di simulazione agli elementi finiti che è stato usato in questo lavoro è *Sentaurus Device* di Synopsys [29].

Il modello DD viene usato fin dagli anni '80 per la simulazione dei dispositivi a semiconduttore, sia a livello accademico che industriale. La velocità di calcolo rispetto ad altri metodi lo rende preferibile. Tuttavia le quantità che il modello tratta non sono le singole particelle ma, tramite un'ipotesi di omogeneizzazione, vengono considerate le concentrazioni numeriche delle variabili che costituiscono il sistema. Il punto di vista con cui viene analizzato il dispositivo è dunque quello delle grandezze medie. Quando le dimensioni dei transistor arrivano a qualche decina di nanometri la discretizzazione della materia incomincia ad essere rilevante (sezione 1.4) e i dispositivi non possono più essere considerati tutti uguali come invece avviene nel modello DD. Inoltre le equazioni del sistema DD sono derivate dalla fisica classica. Quando gli elettroni sono confinati in spazi così ridotti come quelli dei dispositivi ultrascalati, nascono fenomeni quantistici che le equazioni classiche (2.7) non sono in grado di descrivere. Queste problematiche saranno oggetto delle prossime due sezioni.

## 2.2 Il modello density-gradient (DG)

Alcuni parametri dei MOSFET, come ad esempio lo spessore dell'ossido e la lunghezza del canale, hanno raggiunto la scala delle distanze quantomeccaniche. Non può più essere trascurata la natura ondulatoria dei portatori, che genera fenomeni di interferenza tra le particelle o di evanescenza delle loro funzioni d'onda. Gli effetti più evidenti di questo comportamento nei MOSFET sono il tunnelling e il confinamento quantistico.



**Figura 2.1:** Rappresentazione schematica del fenomeno del confinamento quantistico in un MOSFET di tipo bulk. (a) Il piegamento delle bande nei pressi dell'interfaccia Si-SiO<sub>2</sub> crea una buca di potenziale per gli elettroni dello strato di inversione. L'energia si divide in livelli discreti permessi, che aumentano verso la superficie della buca. (b) Densità di elettroni nello strato di inversione per una simulazione classica (in rosso), in cui la densità degli stati è continua, e per una simulazione quantistica (in blu), in cui la densità degli stati è discreta. La concentrazione calcolata secondo la meccanica quantistica è nulla all'interfaccia, poiché le autofunzioni si azzerano in tale punto, e presenta il picco spostato dall'interfaccia di circa 1 nm.

### 2.2.1 Il confinamento quantistico nei MOSFET di tipo bulk

Consideriamo il caso di una buca di potenziale in 1D, ovvero una particella che si trova libera di muoversi in uno spazio racchiuso tra due barriere di potenziale infinite. Se la distanza tra le due barriere è dell'ordine della lunghezza d'onda della particella, per analizzare il problema non si possono più usare le leggi della fisica classica ma devono essere sostituite dalle equazioni della meccanica quantistica. Si risolve l'equazione di Schrödinger e si ricavano le autofunzioni  $\psi_i$  e gli autovalori  $E_i$  del problema. Come risultato si trova che la particella può occupare solo alcuni livelli di energia discreti all'interno della buca, distanziati tra loro da un intervallo di energie proibite. Le autofunzioni associate ai diversi autovalori si annullano tutte ai bordi della barriera, dove non può mai trovarsi la particella. In una situazione reale le barriere di potenziale della buca non sono mai infinite. In questo caso lo spettro dell'energia continua ad essere discreto all'interno della buca ma la spaziatura dei livelli diminuisce progredendo verso l'alto e si evolve in uno spettro continuo. Se nella buca ci sono un certo numero di elettroni la loro densità all'equilibrio si calcola riempiendo i livelli in accordo con la distribuzione di Fermi-Dirac.

Nello strato di inversione dei MOSFET di tipo bulk si verifica una situazione analoga (Figura 2.1a). Andando dal substrato all'interfaccia Si-SiO<sub>2</sub> le bande si piegano verso il basso all'aumentare della tensione di gate. Gli elettroni che si trovano nello strato

di inversione vedono quindi una buca di potenziale delimitata da un lato dalla barriera energetica dell'ossido e dall'altra dal livello di energia che la banda di conduzione possiede nel bulk. Secondo quanto detto sopra, in questa regione la densità degli stati non sarà più continua ma assumerà dei livelli discreti come accade nel caso della buca di potenziale 1D. Le autofunzioni degli stati elettronici si annulleranno all'interfaccia, dove è presente la barriera. Il caso quantistico è dunque completamente diverso da quello classico, in cui non ci sono livelli energetici proibiti e la densità di elettroni aumenta esponenzialmente assumendo il suo valore massimo proprio all'interfaccia tra silicio e ossido.

Un confronto tra il caso classico e quello quantistico è mostrato in Figura 2.1b. Come si vede, il profilo della concentrazione di elettroni del caso quantistico (in blu) è nullo a contatto con l'ossido ed ha un picco spostato di circa 1 nm dall'interfaccia. Questa configurazione dei portatori provoca un aumento dello spessore effettivo dell'ossido di gate, riducendo il controllo del gate sul canale.

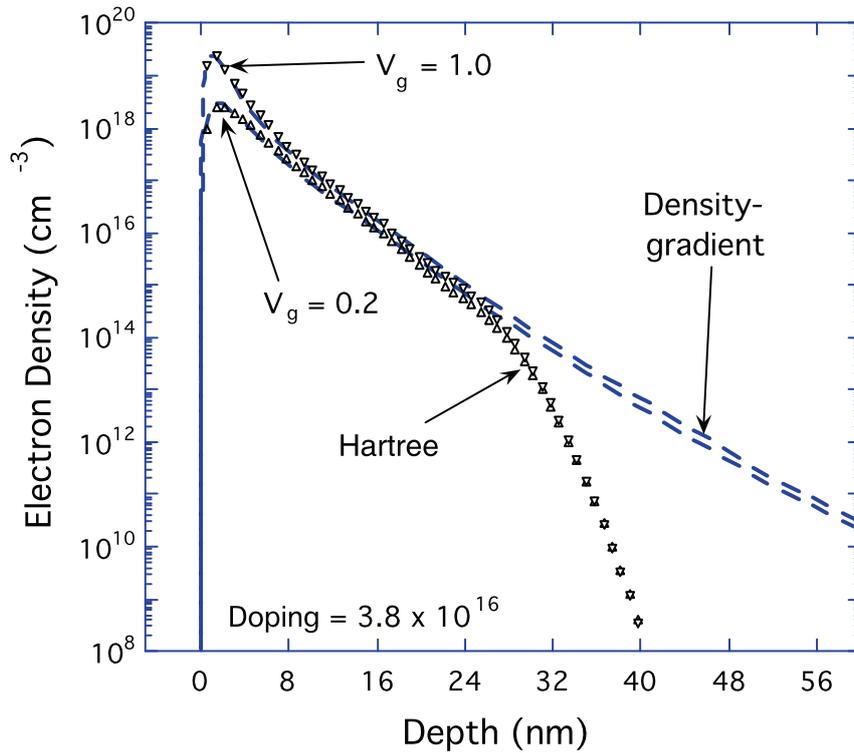
### 2.2.2 La correzione quantistica density-gradient (DG)

Una simulazione quantistica che preveda la risoluzione accoppiata delle equazioni di Schrödinger e di Poisson richiede tempi di calcolo molto più lunghi di quelli necessari alla risoluzione del sistema DD. Inoltre la risoluzione del sistema DD, consolidata ormai da tempo e implementata in numerosi simulatori commerciali, è computazionalmente efficiente ed ha robuste proprietà di convergenza. Tuttavia, con le dimensioni a cui oggi arrivano i dispositivi di cui si vuole analizzare il comportamento, non è pensabile continuare ad utilizzare dei risultati della meccanica classica.

Per rimediare a questo problema è stata proposta la teoria *density-gradient* (DG). La teoria DG, allo stesso modo della teoria DD, considera un semiconduttore come costituito da tre quantità continue interagenti tra loro: una densità di elettroni, una densità di lacune e un reticolo cristallino continuo [30]. La differenza tra il modello DG dal modello DD è che, nel primo caso, l'equazione di stato degli elettroni o delle lacune non dipende soltanto dalla loro densità ma anche dal loro gradiente di densità. Nell'equazione classica che determina la densità dei portatori (per esempio degli elettroni) viene introdotta una quantità  $\Lambda_n$ , simile ad un potenziale:

$$n = N_C \cdot F_{\frac{1}{2}} \left( \frac{E_{f,n} - E_c - \Lambda_n}{KT} \right), \quad (2.8)$$

dove  $F_{\frac{1}{2}}(\cdot)$  è l'integrale di Fermi di ordine  $\frac{1}{2}$ .  $\Lambda_n$  viene chiamato *potenziale quantistico*. Per le lacune vale un'equazione analoga, con un potenziale quantistico  $\Lambda_p$ .  $\Lambda_n$  nella (2.8)



**Figura 2.2:** Confronto tra le previsioni di una simulazione DG (linea tratteggiata) e una quantistica (simboli) della densità di elettroni nello strato di inversione di un MOSFET di tipo bulk, al variare della tensione di gate. La soluzione DG riproduce con accuratezza la soluzione quantistica, prevedendo il picco della concentrazione di portatori spostato rispetto all'interfaccia Si-SiO<sub>2</sub>. L'andamento differente tra le due soluzioni lontano dall'interfaccia è dovuto in realtà ad un'approssimazione della simulazione quantistica che per semplificare i calcoli considerava solamente i primi 80 livelli discreti di energia (da [30]).

è dato da un'equazione differenziale:

$$\Lambda_n = -\frac{\gamma \hbar}{6m_n^*} \frac{\nabla^2 \sqrt{n}}{\sqrt{n}}, \quad (2.9)$$

dove  $m_n^*$  è la massa efficace degli elettroni in banda di conduzione e  $\gamma$  è un parametro utilizzato per il fit della soluzione DG rispetto alla soluzione quantistica. La forma funzionale di  $\Lambda_n$  deriva da uno sviluppo in serie troncato dell'equazione di Schrödinger, e come essa è basata sulla continuità delle funzioni d'onda. Come si vede dalla (2.9), il potenziale quantistico smussa il profilo dei portatori riducendo la loro derivata seconda, e così evita che si creino discontinuità.

La modifica dipendente dal gradiente di densità che viene introdotta nella teoria DG riesce a riprodurre con un'accuratezza sorprendente i risultati delle simulazioni quantistiche. Se ne può vedere un esempio in Figura 2.2 dove sono confrontate le soluzioni dello strato di inversione in un MOSFET di tipo bulk ottenute con una simulazione DG e con

una simulazione quantistica, per due diverse tensioni di gate. La soluzione DG ricalca quasi perfettamente l'andamento che risulta dall'applicazione delle leggi quantistiche, con il picco della concentrazione che si sposta rispetto all'interfaccia e modifica la sua forma a seconda della profondità della buca. L'andamento differente che le due curve hanno lontano dall'interfaccia è in realtà dovuto ad un'approssimazione della simulazione quantistica, che per semplificare i calcoli considerava solo i primi 80 livelli discreti.

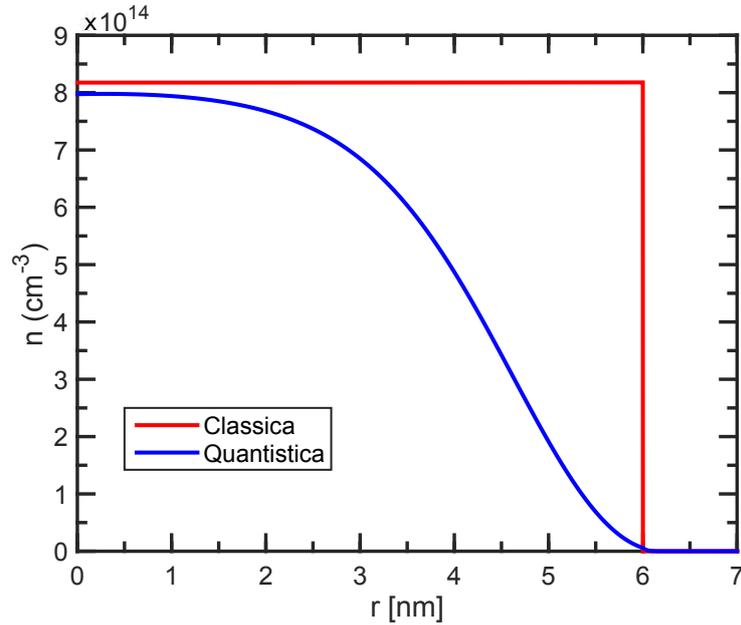
### 2.2.3 Il caso GAA

In un nanowire lo spessore di silicio circondato dal gate si aggira intorno ai 10 nm. Lungo la coordinata radiale si ha dunque una buca di potenziale, allo stesso modo di quanto visto per lo strato di inversione della struttura bulk. Anche in un dispositivo di questo tipo, dunque, gli elettroni subiranno il confinamento quantistico. Tenere in conto questi fenomeni è fondamentale per una corretta indagine elettrostatica in condizioni di sottosoglia.

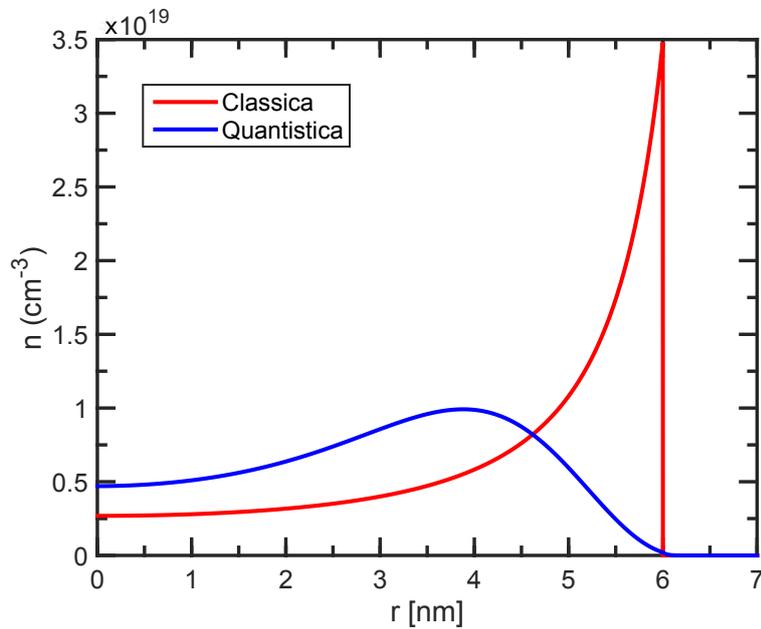
Come discusso in sezione 1.2.1, l'equazione di Poisson prevede che sottosoglia le bande siano piatte e seguano  $V_g$  su tutto il volume. In tali condizioni l'equazione di stato classica per gli elettroni stabilisce una concentrazione di elettroni costante in tutto il silicio. Se invece si osserva la situazione dal punto di vista della meccanica quantistica, questo non può accadere poiché le autofunzioni dei livelli energetici permessi si annullano tutte all'interfaccia. In Figura 2.3 si può vedere un confronto tra la soluzione classica sottosoglia e la soluzione quantistica. Per le simulazioni è stato sfruttato il codice agli elementi finiti utilizzato in [31]. Si nota come la soluzione quantistica (curva blu) sia confinata verso il centro del volume del silicio e si annulli all'interfaccia.

Se il transistor è sopra soglia, ai bordi del gate si forma lo strato di elettroni e le bande incominciano a piegarsi. All'interfaccia si crea così la stessa situazione del caso planare, con una buca ancora più stretta. In Figura 2.4 sono messe a confronto la soluzione classica e quella quantistica per lo stesso transistor di Figura 2.3, con  $V_g = 0.7$  V, ovvero ampiamente sopra soglia. A causa del confinamento dei portatori nello strato di inversione, la soluzione quantistica presenta un picco allargato e lontano dall'interfaccia, con una forma simile a quella che assume nel transistor di tipo bulk. La soluzione classica invece non risente della quantizzazione ed assume il suo massimo a contatto con l'ossido.

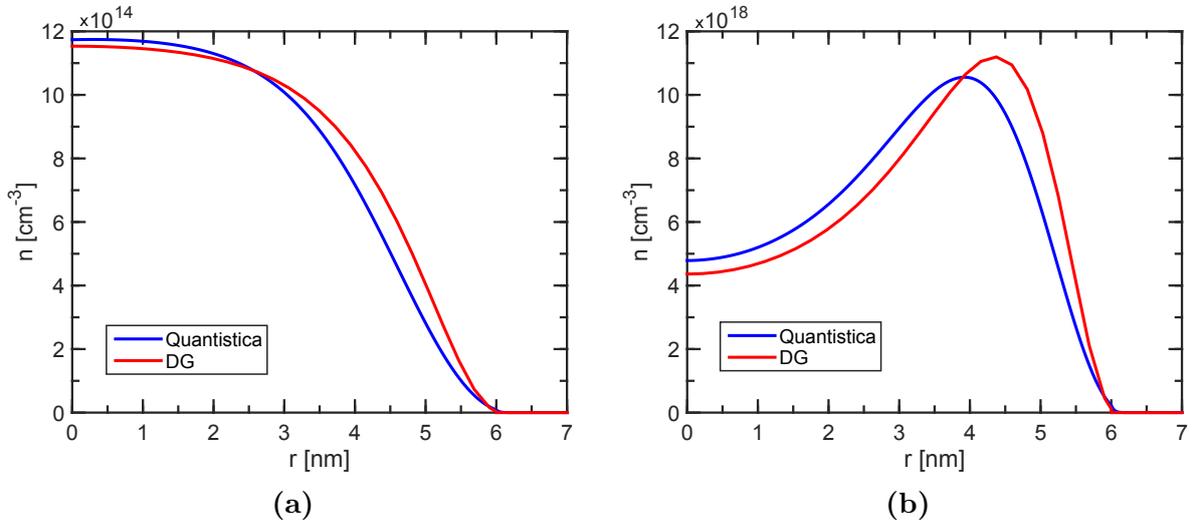
Come accennato (sezione 2.2.2), il simulatore implementa le correzioni quantistiche al modello DD tramite le equazioni (2.8) e (2.9). Per verificare che il confinamento quantistico venga reso correttamente nei dispositivi utilizzati si è simulata una rampa della tensione di gate, a  $V_{ds} = 0$  mV, per un nanowire con  $R = 6$  nm. La lunghezza del canale è di 100 nm, in modo che l'effetto di canale corto non influisse sul risultato. È stato



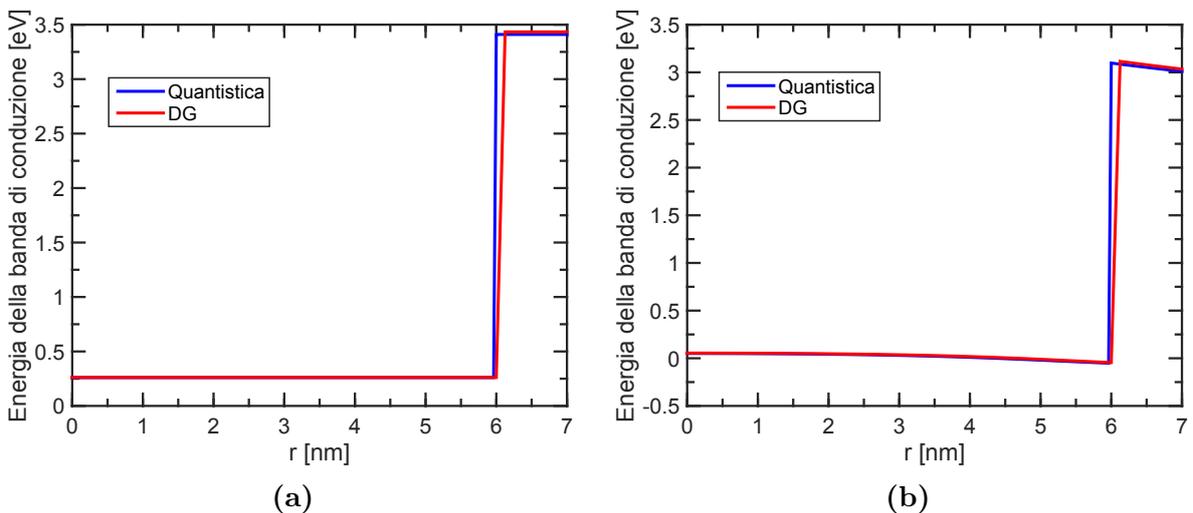
**Figura 2.3:** Densità di elettroni sottosoglia in funzione della posizione radiale per un nanowire con  $R = 6$  nm e  $V_g = 0.3$  V. Sono messe a confronto la soluzione classica (in rosso) e la soluzione quantistica (in blu). Le soluzioni sono state simulate con il codice utilizzato in [31].



**Figura 2.4:** Densità di elettroni in funzione della posizione radiale per un nanowire con  $R = 6$  nm.  $V_g = 0.7$  V e il transistor è sopra soglia. La soluzione classica (in rosso) non risente del confinamento ed assume il suo massimo all'interfaccia. La soluzione quantistica (in blu) invece è nulla all'interfaccia ed assume la stessa forma del picco che si ha nel caso planare.



**Figura 2.5:** Confronto tra la densità di elettroni lungo la coordinata radiale simulata con la risoluzione dell'equazione di Poisson accoppiata con l'equazione di Schrödinger (in blu) o con la correzione DG implementata nel simulatore 3D (in rosso). (a)  $V_g = 0.3$  V, il transistor è sottosoglia e gli elettroni sono confinati al centro del volume di silicio. (b)  $V_g = 0.7$  V, il transistor è in inversione forte e gli elettroni sono confinati nello strato all'interfaccia Si-SiO<sub>2</sub>. In entrambi i casi il confinamento quantistico è riprodotto dal modello DG, che segue con qualche imprecisione la soluzione quantistica.



**Figura 2.6:** Confronto tra la banda di conduzione lungo la coordinata radiale simulata dalla soluzione quantistica (in blu) o con la correzione DG (in rosso), per le stesse  $V_g$  di Figura 2.5. (a)  $V_g = 0.3$  V, il transistor è sottosoglia e le bande sono piatte in tutto il silicio. (b)  $V_g = 0.7$  V, il transistor è in inversione forte e le bande sono piegate all'interfaccia.

estratto il profilo radiale della concentrazione di elettroni per  $V_g = 0.3$  V e  $V_g = 0.7$  V, in modo da avere la stessa situazione di Figura 2.3 e Figura 2.4. I risultati sono messi a confronto in Figura 2.5. Sia sottosoglia (2.5a) che sopra soglia (2.5b), la soluzione DG riproduce con buona approssimazione l'effetto del confinamento quantistico. Tuttavia le due soluzioni si discostano nell'andamento in prossimità dell'ossido e nel valore di picco assunto. Le condizioni iniziali sono le stesse per entrambi i modelli, come viene mostrato in Figura 2.6, dove è riportata l'energia della banda di conduzione simulata nei due casi. Dunque lo scostamento è dovuto all'approssimazione delle equazioni quantistiche operata dal modello DG. Nei punti dove c'è un forte gradiente di concentrazione infatti, la soluzione DG è imprecisa. La differenza tra le due curve può essere ridotta tramite un fit della soluzione quantistica, su cui viene tarato il parametro  $\gamma$  dell'equazione (2.9). Per quanto riguarda il nostro lavoro, dove viene analizzato il dispositivo sottosoglia, è stato considerato sufficiente che fosse riprodotto l'andamento generale, poiché la concentrazione di elettroni in questo regime aumenta esponenzialmente con  $V_g$  e dunque errori piccoli sono in prima approssimazione trascurabili.

## 2.3 Il drogaggio atomistico

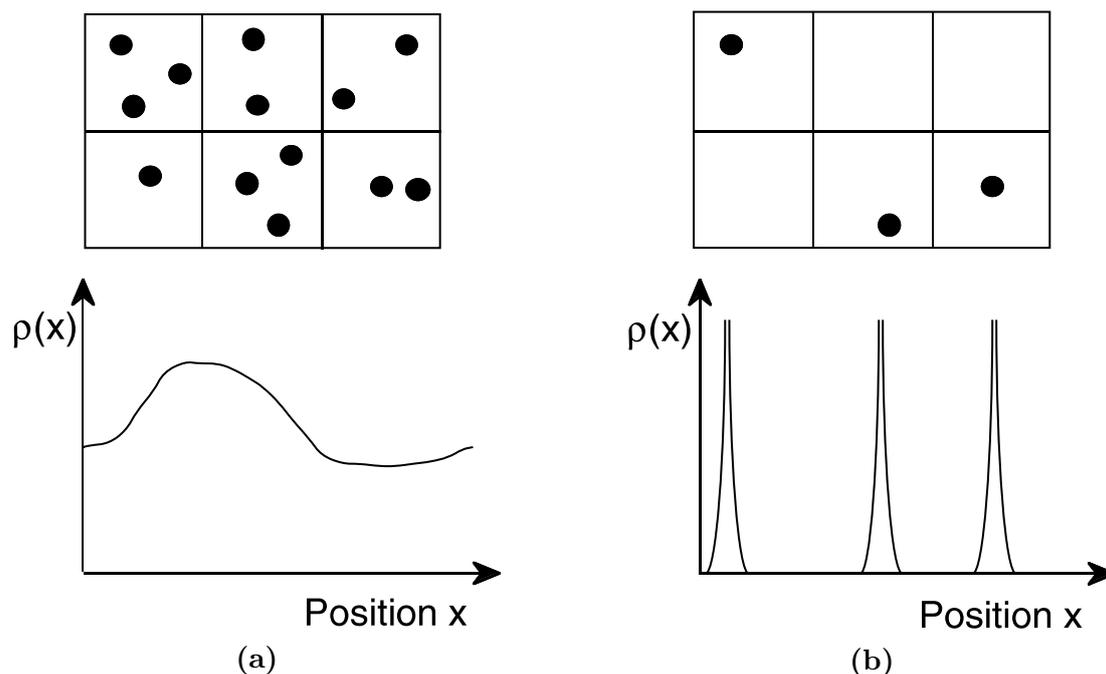
### 2.3.1 Implementazione nel modello DD

In sezione 2.1 è stato detto che il modello DD descrive le grandezze medie di un dispositivo ma non è pensato per un'analisi delle sue caratteristiche microscopiche. Tuttavia i modelli completamente atomistici, in cui viene considerata la fisica di ogni singola particella, sono molto più complessi e richiedono un notevole sforzo computazionale. Soprattutto in ambito industriale si preferisce continuare ad utilizzare il sistema DD.

Dato che poter simulare i singoli ioni di drogante è diventato fondamentale per prevedere la variabilità nei dispositivi ultrascalati (sezione 1.4), si cerca di implementarli nell'ambito delle simulazioni DD con uno stratagemma.

Uno ione ha carica  $+q$  se è un donore,  $-q$  se invece è un accettore. Esso occupa una certa area di spazio, al di fuori della quale esercita un influsso elettrostatico pari a  $\pm q$ . Circa lo stesso effetto lo si può ottenere se si considera una zona di spazio in cui è presente una densità di carica che integrata sul volume da essa occupato risulta pari a  $\pm q$ . Si riescono così a modellizzare dei singoli atomi di drogante come delle aree in cui è presente una densità di ioni, mentre all'esterno la densità del doping è nulla.

Per evitare errori di discretizzazione, il volume che viene scelto come occupato dal singolo atomo è quello corrispondente ad un nodo di mesh. Infatti se invece di tale scelta



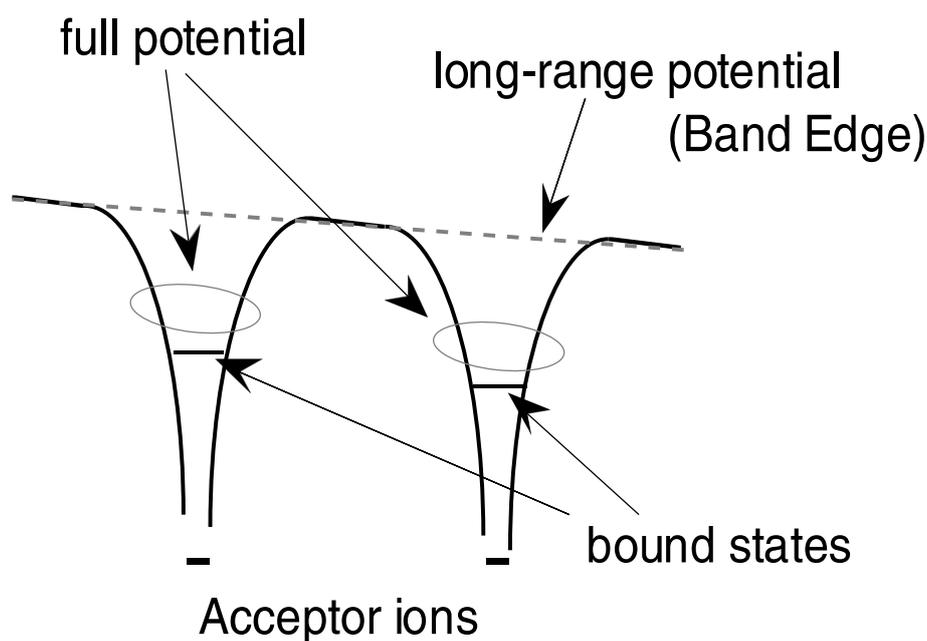
**Figura 2.7:** Rappresentazione schematica della configurazione dei droganti sulla mesh (in alto) e della densità di droganti lungo la cordinata spaziale corrispondente. (a) Caso in cui il dispositivo è di grandi dimensioni. (b) Dispositivo atomico (da [32]).

si prendessero delle dimensioni arbitrarie per simulare un atomo discreto, si rischierebbe che la densità relativa a tale volume venga mappata su uno spazio più grande a causa delle dimensioni finite della griglia, falsando l'effetto elettrostatico del drogante. Il simulatore presenta la possibilità di posizionare direttamente i droganti discreti. La loro posizione verrà fissata sul nodo di mesh più vicino alla coordinata fornita e sarà calcolata automaticamente la densità del doping da associare al volume attorno al nodo:

$$N_{d,a} = \frac{1}{V}.$$

Se nel nodo cadono più atomi droganti la concentrazione associata verrà moltiplicata per il numero di essi.

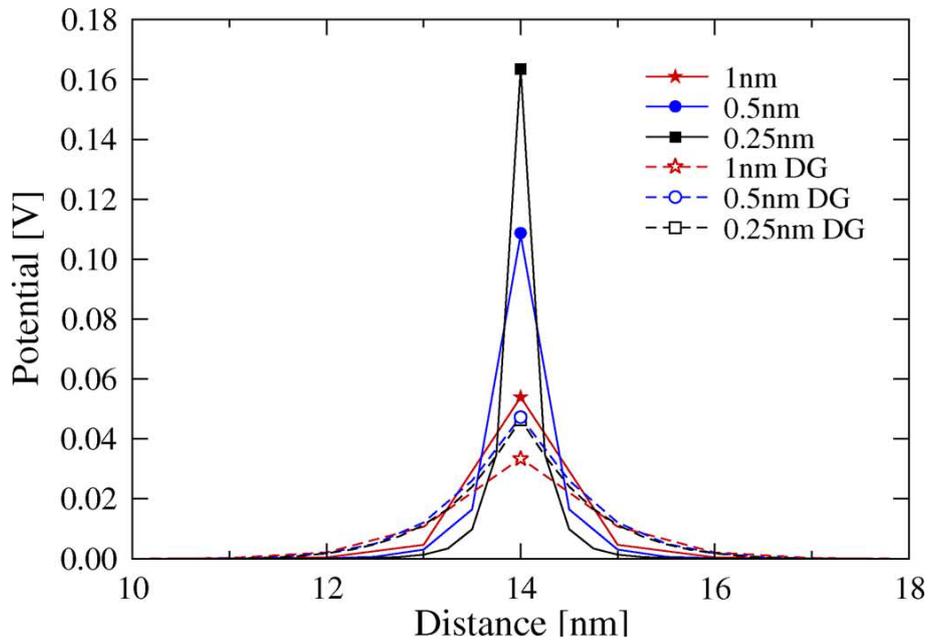
Questo approccio, che permette di rappresentare la diversità di dispositivi che dal punto di vista macroscopico sarebbero considerati uguali, presenta dei problemi che derivano dalla natura delle equazioni DD. Quando si vogliono implementare degli atomi discreti come descritto sopra, le situazioni che possono presentarsi sono riassunte in Figura 2.7. Quando le dimensioni del transistor sono grandi, per ogni nodo di mesh è presente almeno un drogante e la densità del doping varia gradualmente. Questo scenario è rappresentato in Figura 2.7a, dove nel riquadro in alto si vede una tipica configurazione di droganti rappresentata sulla griglia e in basso la densità di carica corrispondente a questa dispo-



**Figura 2.8:** Rappresentazione schematica del potenziale elettrostatico causato da alcuni accettori ionizzati. In linea continua è mostrato il potenziale totale. In linea a tratti viene evidenziata la parte a lungo raggio, che è quella che viene tenuta in conto dal modello DD (da [32]).

sizione. Nei dispositivi atomistici invece, molti dei nodi di mesh non presentano alcun drogante. Così la densità di carica sui punti della griglia si comporta come delle funzioni  $\delta$  di Dirac (Figura 2.7b) [32].

Se si pensa al potenziale coulombiano di uno ione, questo può essere suddiviso in due parti, come illustrato in Figura (2.8). La prima, che ha effetto a lungo raggio, è rappresentata con una linea tratteggiata e determina un cambiamento elettrostatico graduale nelle aree circostanti. La seconda, che ha effetto a corto raggio, varia bruscamente e presenta delle singolarità nelle posizioni degli atomi (che in questo caso sono accettori). Il modello DD considera come potenziale elettrostatico il termine a lungo raggio, che determina un piegamento moderato delle bande. Il termine a corto raggio viene tenuto in conto implicitamente nello scattering dei portatori, che determina la condizione di quasi-equilibrio in cui si assume che si trovi il sistema. Nel caso dei dispositivi atomistici, la distribuzione di carica generata dai droganti discreti (Figura 2.7b) è equivalente al termine a corto raggio del potenziale coulombiano. Così questo termine viene erroneamente contato due volte nel modello. Inoltre il suo andamento a  $\delta$  causerà un accumulo di portatori di segno opposto nelle immediate vicinanze dello ione, che scherma la carica del drogante e ne riduce l'effetto elettrostatico. Nella realtà ciò non si verifica, poiché i portatori nei pressi di una buca di potenziale sono confinati quantisticamente. Lo stato di ground ha un'energia molto vicina alla superficie della buca e, a temperatura ambiente, molti portatori non sono vincolati. Per questo motivo il potenziale che ha effetto sui portatori è soltanto il



**Figura 2.9:** Distribuzioni del potenziale elettrostatico di una carica fissa positiva in un cubo di materiale semiconduttore simulate con il modello DD (linee continue) e con il modello DG (linee a tratti), al variare della spaziatura tra i nodi della mesh. Il potenziale simulato con il modello DG è praticamente indipendente dalla spaziatura della griglia (da [33]).

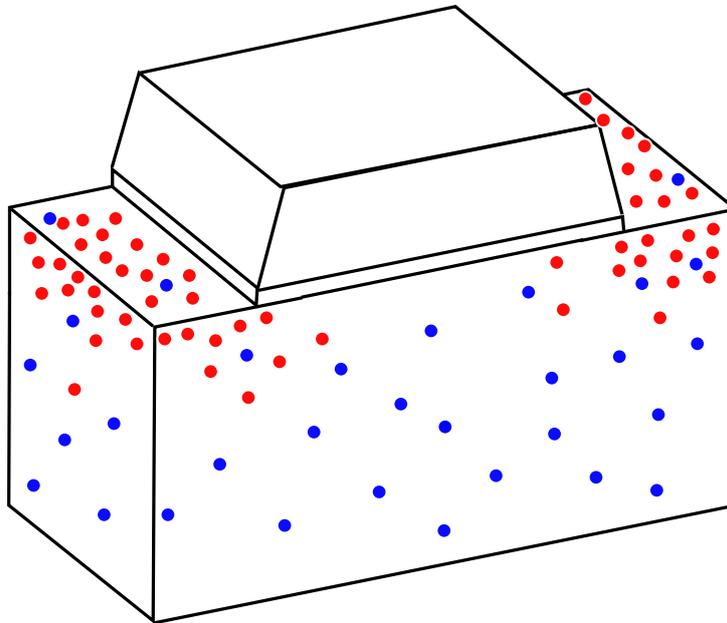
termine a lungo raggio, che è quello previsto dalla teoria DD.

Come è stato detto (2.1), per aumentare la precisione dell'approssimazione numerica delle grandezze del dispositivo che si vuole simulare la mesh sulla struttura deve essere infittita, soprattutto nelle zone dove i gradienti sono elevati. Su transistor delle dimensioni di qualche decina di nanometri la spaziatura tra i nodi arriva ad essere di qualche Ångström. In queste condizioni i droganti discreti implementati con il metodo della densità di carica mappata sul Voronoï box formano delle buche di potenziale che attraggono le cariche mobili, causando una schermatura della loro carica. Al diminuire dello spazio tra i nodi di griglia l'effetto si farà sempre più marcato. Per evitare questo artefatto è necessario applicare delle correzioni quantistiche al modello DD, in modo che l'effetto attrattivo del potenziale a  $\delta$  sia mitigato dal confinamento quantistico. In Figura (2.9) vengono mostrati i risultati delle simulazioni del potenziale elettrostatico esercitato da uno ione donore al cambiare della spaziatura tra i nodi di mesh, sia utilizzando il modello DD che utilizzando il modello DG. Con una mesh più rifinita viene aumentata la risoluzione del potenziale attorno alla posizione dell'atomo. Per le simulazioni DD il picco del potenziale aumenta quando i nodi della griglia si infittiscono e questo causa una maggiore densità di portatori che viene attratta nelle sue vicinanze e che ne scherma la carica. Nelle simulazioni DG invece il potenziale è praticamente indipendente dalla spaziatura della mesh. Inoltre la profondità della buca di potenziale DG è in buon accordo con le simulazioni

quantistiche dello stesso caso [33]. Utilizzare le correzioni DG consente quindi di eliminare il problema dovuto alla variazione brusca del potenziale elettrostatico in prossimità di ioni localizzati e permette di simulare correttamente l'effetto dei droganti atomistici all'interno del modello DD.

### 2.3.2 Posizionamento casuale dei droganti

Gli atomi di drogante vengono introdotti in un transistor tramite impiantazione ionica. Successivamente subiscono diffusione a causa dei processi termici che vengono effettuati per riparare i danni della matrice cristallina e per attivare i droganti. Una rappresentazione schematica di come appare un transistor dal punto di vista atomistico è riportata in Figura 2.10. Per semplicità sono mostrati soltanto gli ioni di drogante, in blu gli accettori



**Figura 2.10:** Rappresentazione schematica di un transistor completamente atomistico. Sono indicati soltanto gli atomi droganti, ovvero gli accettori nel substrato (in blu) e i donori dei profili di source e drain (in rosso). La visione è completamente diversa rispetto a quella classica del modello DD, che ha profili di drogaggio continui e una giunzione metallurgica tra il substrato e il source o il drain.

e in rosso i donori. In un caso simile non ha più senso nemmeno parlare della giunzione metallurgica tra source e drain e il substrato.

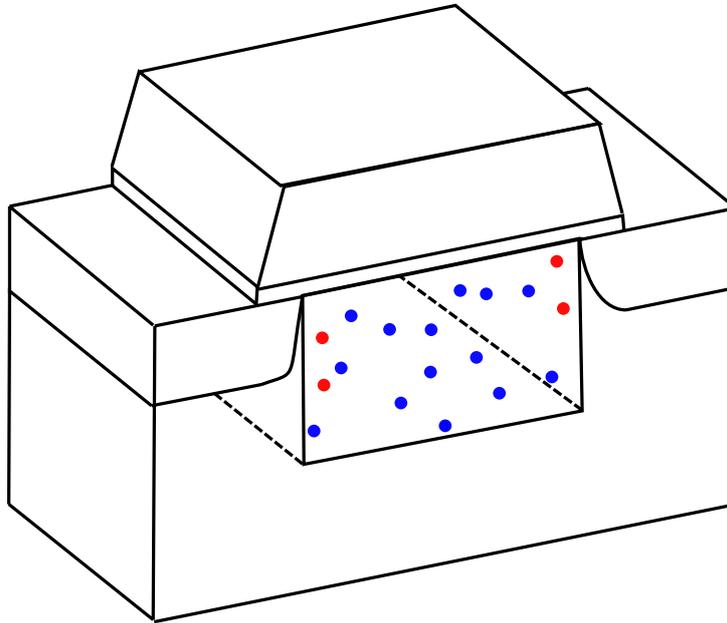
Come abbiamo visto (sezione 1.4.1), un profilo di drogaggio continuo rappresenta una grandezza media del dispositivo. Ad essere mediata è la posizione che i droganti occupano dopo il flusso di processo su un campione di transistor. Il risultato di questa operazione è un profilo continuo che avrà una determinata concentrazione di atomi, espressa in  $\text{cm}^{-3}$ , in funzione della posizione all'interno del transistor. Se si tratta del drogaggio del substrato,

gli ioni sono distribuiti con probabilità uniforme poiché hanno subito molti processi termici e la diffusione li ha posizionati casualmente su tutto il substrato. La concentrazione  $N_a$  sarà dunque costante in questa zona. Se invece si considera il drogaggio di source e drain, gli ioni donori che formano queste due zone vengono impiantati con precisione e la loro posizione è determinata dall'energia con cui sono accelerati verso il silicio e dalla loro specie atomica. Saranno localizzati ad una certa profondità nel substrato e dispersi relativamente attorno a tale punto. Un processo simile può essere ben descritto da una distribuzione normale, con una certa deviazione standard attorno alla media che dipenderà sia dal processo di impiantazione, sia dai processi termici che avvengono successivamente.

Quando si deve riportare una media fatta su molti dispositivi diversi tra loro al caso di un singolo transistor è necessario determinare il numero di droganti che saranno contenuti al suo interno e la posizione occupata da ciascuno di essi. Come discusso sopra, un certo profilo di drogaggio in una zona del transistor indica la concentrazione di atomi per unità di volume in ogni punto. Se dunque si integra la concentrazione sul volume totale in cui essa è presente si ottiene il numero medio di atomi ( $N$ ) all'interno di questo. Il numero reale di droganti in un singolo dispositivo si può assumere come distribuito secondo una statistica di Poisson con valore medio pari a  $N$ . Tale numero fluttuerà dunque con deviazione standard  $\sqrt{N}$ . A questo punto per simulare il drogaggio di un transistor reale bisogna stabilire la posizione di ogni singolo ione. Per fare questo la strada più semplice è sfruttare il profilo di concentrazione continuo, che rispecchia il comportamento medio degli atomi, come distribuzione di probabilità per il punto del dispositivo che questi andranno ad occupare. Nel substrato la concentrazione è costante e quindi la posizione di ogni drogante sarà una variabile casuale estratta da una densità di probabilità uniforme. Nelle zone di source e drain la posizione degli atomi nella direzione del canale viene estratta casualmente da una distribuzione normale, mentre rimane uniforme lungo gli altri due lati.

Negli studi condotti sulle fluttuazioni causate dal drogaggio atomistico (sezione 1.4.1) viene messo in evidenza come i maggiori responsabili dei cambiamenti della tensione di soglia siano gli ioni nella zona sotto il gate, dove la barriera di potenziale è più alta, mentre i droganti di source e drain hanno un'influenza minore. Inoltre non tutti gli ioni nella zona sotto il gate hanno lo stesso peso. Poiché lo strato di inversione si crea all'interfaccia con l'ossido, gli ioni più lontani da questo avranno un effetto elettrostatico più debole o nullo sulla conduzione sotto soglia. A fronte di questo, per alleggerire il calcolo si sceglie di solito di discretizzare soltanto la zona del dispositivo che sarà maggiormente responsabile dei cambiamenti dell'elettrostatica. Nelle altre zone il profilo di drogaggio viene lasciato continuo. Generalmente l'area del transistor dove viene implementato il drogaggio atomistico coincide con un intorno della zona svuotata [16], e può comprendere

o meno una percentuale dello spazio occupato da source e drain. In Figura 2.11 viene mostrata una possibile scelta per la zona di discretizzazione. Nelle zone senza droganti è



**Figura 2.11:** Rappresentazione schematica della zona di discretizzazione utilizzata nella simulazione atomistica di un transistor di tipo bulk. Rispetto al caso illustrato in Figura 2.10 ora la zona in cui sono presenti gli atomi di drogante è ristretta ad un'area che coincide in prima approssimazione con la zona svuotata. Nel resto del dispositivo viene mantenuto un profilo di drogaggio costante.

previsto un profilo di drogaggio continuo che per semplicità non è stato riportato.

# Capitolo 3

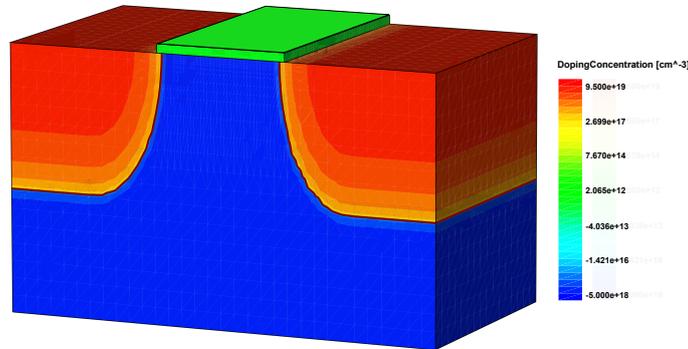
## Analisi della variabilità atomistica nella struttura GAA con canale drogato

### 3.1 Variabilità nella struttura planare

Nel nostro lavoro è stata studiata la variabilità intrinseca dei dispositivi GAA, dovuta alla discretizzazione del drogaggio. Le analisi effettuate riguardano lo studio delle fluttuazioni dovute sia all'RDF che all'RTN, in modo analogo a quanto è stato presentato in sezione 1.4 per i transistor di tipo bulk. Per poter confrontare tali risultati con quelli del nanowire, come prima cosa si è riprodotta una struttura planare da utilizzare come termine di paragone.

Il MOSFET di tipo bulk utilizzato per il confronto con la struttura GAA è mostrato in Figura 3.1. Come lunghezza del canale è stata scelta  $L = 15$  nm, per situarsi nell'attuale panorama dei dispositivi ultrascaled, dove il nodo tecnologico ha raggiunto i 14 nm ([4]). Per ridurre l'effetto di canale corto il substrato deve essere drogato con una concentrazione di accettori piuttosto elevata. Abbiamo usato  $N_a = 5 \times 10^{18}$  cm<sup>-3</sup>. Il profilo di drogaggio di source e drain è gaussiano, con una concentrazione di picco  $N_d = 10^{20}$  cm<sup>-3</sup>. Gli altri parametri sono  $W = 28$  nm e  $t_{ox} = 1$  nm.

Nella simulazione di un transistor viene calcolata la curva  $I - V$  risolvendo il sistema DD accoppiato con le equazioni DG, per una rampa di valori della tensione di gate. La tensione di soglia, che rappresenta il parametro di interesse per lo studio della variabilità



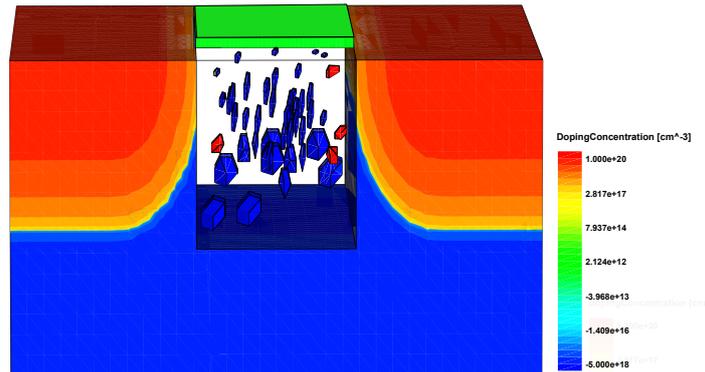
**Figura 3.1:** Struttura utilizzata per le simulazioni statistiche del MOSFET planare. Source e drain hanno un profilo di drogaggio gaussiano con concentrazione di picco  $N_d = 10^{20} \text{ cm}^{-3}$ . Nel substrato il drogaggio è uniforme con concentrazione  $N_a = 5 \times 10^{18} \text{ cm}^{-3}$ .

atomistica, viene definita come la  $V_g$  alla quale  $I_d = 1 \text{ nA}$ . Per un valore simile della corrente di drain il dispositivo è sicuramente sottosoglia. Questo permette di simulare correttamente l'elettrostatica anche per dispositivi con lunghezze di canale come quelle in esame, dove diventano determinanti il trasporto balistico e di non-equilibrio, che il modello DD non consente di simulare. Sottosoglia la corrente varia esponenzialmente con le fluttuazioni del potenziale superficiale e gli errori nella sua stima sono trascurabili. Sia nell'analisi dell'RDF che in quella dell'RTN si guarda la variazione di  $V_t$ , dunque il fatto che il valore di  $I_d$  simulato non coincida con quello esatto non rappresenta un problema, poiché l'errore è costante in tutte le simulazioni.

Si è visto in sezione 1.4 che le fluttuazioni di RDF e di RTN sono un fenomeno statistico, cioè i loro valori sono caratterizzati da una specifica densità di probabilità. Nel caso dell'RDF la distribuzione è una normale, mentre quella dell'RTN è una normale con una coda esponenziale, come osservato in sezione 1.4.2. Per ottenere queste distribuzioni è necessario avere un campione di transistor. Il campione è stato creato simulando 100 dispositivi in cui il profilo di drogaggio è stato estratto con il metodo Monte Carlo. Il metodo Monte Carlo è un metodo statistico non parametrico in cui vengono generati dei numeri casuali tra loro incorrelati che seguono la distribuzione di probabilità con cui si suppone si verifichi il fenomeno che si vuole studiare. Queste realizzazioni che il metodo calcola, con il peso della loro probabilità di evenienza, si propongono di indagare tutto lo spazio del fenomeno in esame. Nel nostro caso un algoritmo genera le posizioni casuali dei droganti discreti di ogni profilo. Queste vengono poi passate al simulatore che le implementa nella griglia di ciascun dispositivo. I transistor risultano così tutti diversi tra

loro a livello microscopico, pur avendo la stessa geometria.

I droganti discreti sono stati implementati solamente nella zona svuotata, seguendo la procedura descritta in sezione 2.3.2. Nel resto del dispositivo si è mantenuto un drogaggio continuo. Un esempio del risultato di questa procedura è mostrato in Figura 3.2. Si può

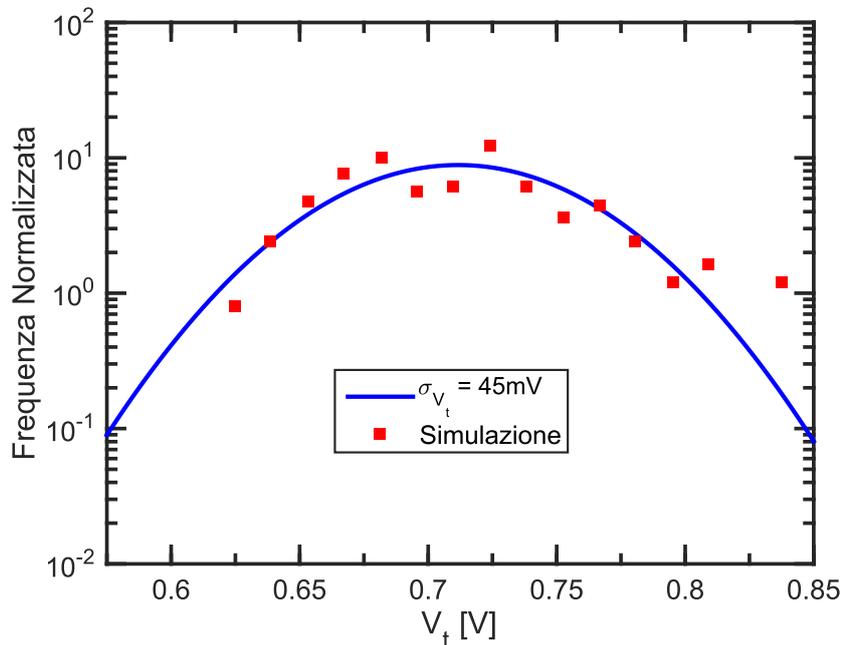


**Figura 3.2:** Esempio del profilo atomistico implementato in uno dei 100 MOS planari simulati con il metodo Monte Carlo. Come zona di discretizzazione è stata presa una stima della zona svuotata.

notare che il volume occupato dai droganti diminuisce avvicinandosi al gate. Infatti la mesh non è uniforme ma è stata infittita nella zona del canale, dove avvengono i fenomeni di interesse. Come mostrato in sezione 2.3.1, la concentrazione di doping ai nodi corrispondenti viene regolata di conseguenza dal software che genera la mesh. Si è visto che grandi concentrazioni in volumi molto piccoli possono portare a un accumulo anomalo di portatori di segno opposto in prossimità del nodo a cui è posizionato il drogante, la cui densità di carica ha la forma di un  $\delta$  di Dirac. I portatori schermano l'effetto elettrostatico dello ione, che così non viene simulato correttamente. Se viene implementato il confinamento quantistico utilizzando la correzione DG, come è stato discusso nella stessa sezione, il problema non si presenta e gli atomi sono tutti equivalenti, indipendentemente dalla concentrazione di carica presente nel volume occupato.

Per quanto riguarda l'RDF, il parametro più significativo è  $\sigma_{V_t}$ , la deviazione standard della gaussiana secondo cui si distribuiscono le ampiezze delle oscillazioni. È stato simulato il campione Monte Carlo di 100 MOS planari e per ogni dispositivo è stata estratta la tensione di soglia. Infine sui valori di  $V_t$  si è eseguito un fit gaussiano. Il risultato è

mostrato in Figura 3.3, da cui si trova una  $\sigma_{V_t}$  di 45 mV. Tale risultato è in perfetto



**Figura 3.3:** Distribuzione delle tensioni di soglia dei 100 transistor planari con la geometria di Figura 3.1 e profilo di drogaggio atomistico simulato con il metodo Monte Carlo (in rosso). È stato sovrapposto il relativo fit gaussiano (in blu).

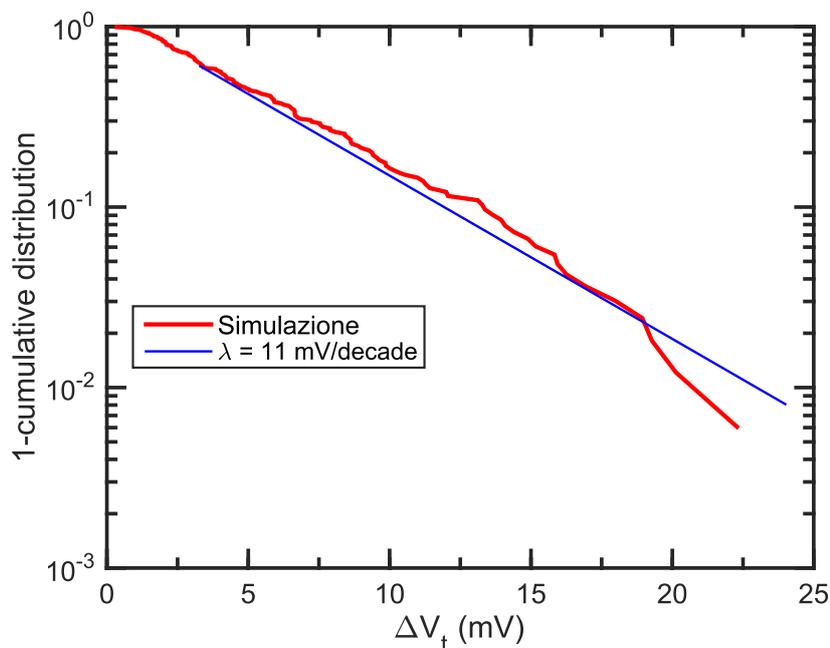
accordo con la formula (1.10) che, sostituendo i parametri del nostro dispositivo, prevede una  $\sigma_{V_t}$  di 46.8 mV.

Per stimare le variazioni della tensione di soglia dovute all'RTN sono stati presi gli stessi 100 profili atomistici del campione utilizzato nell'analisi di RDF e su ciascuno è stata posizionata casualmente una carica negativa all'interfaccia tra canale e ossido, distribuita secondo una densità di probabilità uniforme sull'area. La carica  $-q$  viene associata al nodo di mesh più vicino ed è rappresentata con una densità di carica equivalente ( $\text{cm}^{-3}$ ) distribuita sul volume del nodo in questione, esattamente come accade per l'implementazione dei singoli ioni di drogante. Tale carica negativa rappresenta l'effetto di un elettrone che viene intrappolato in uno stato di difetto. Come spiegato in precedenza (sezione 1.4.2), se l'elettrone è catturato nei pressi di un percorso percolativo causato dalla particolare configurazione di droganti nella zona svuotata del transistor darà origine ad un'ampiezza  $\Delta V_t$  caratterizzata da una probabilità di accadimento esponenziale. Gli elettroni nel volume dell'ossido sono stati trascurati e si è tenuto conto soltanto degli elettroni catturati all'interfaccia poiché sono quelli che generano le ampiezze  $\Delta V_t$  maggiori.

L'analisi di RTN che è stata compiuta è un'analisi elettrostatica, ovvero il valore di  $\Delta V_t$  è stato ricavato come la differenza tra la  $V_t$  del transistor senza carica e la  $V_t$  dello stesso con la carica. Tale ampiezza è la stessa che si osserverebbe in funzione del tempo,

ad esempio in Figura 1.14, e che in seguito all'utilizzo può rendere difettosi dei transistor considerati in principio buoni.

È stata calcolata la distribuzione cumulativa dell'insieme dei 100  $\Delta V_t$  e la si è rappresentata su un grafico semilogaritmico. Il risultato è visibile in Figura 3.4. L'andamento



**Figura 3.4:** Distribuzione cumulativa delle ampiezze di RTN simulate su un campione di transistor atomistici con la geometria di Figura 3.1 (in rosso). Si vede che dall'andamento gaussiano della distribuzione si stacca una coda esponenziale con pendenza  $\lambda$ , di cui è stato calcolato un fit lineare (in blu).

della distribuzione è quello previsto dai diversi lavori sull'argomento (sezione 1.4.2), con una coda esponenziale  $\lambda = 11$  mV/decade che si stacca dall'andamento gaussiano. I valori di probabilità vanno oltre  $10^{-2}$  poiché sono stati aggiunti altri transistor simulati al campione al fine di aumentare la statistica.

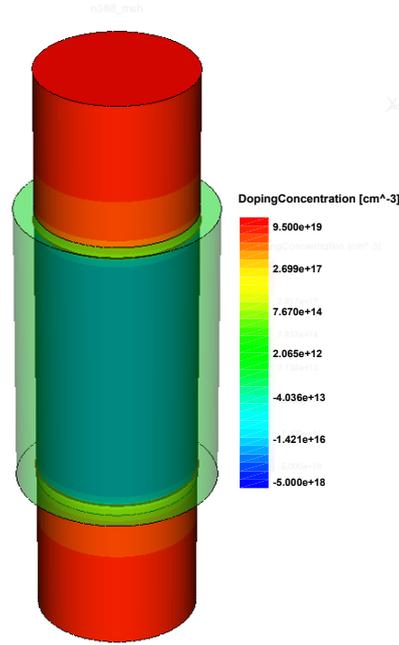
## 3.2 Variabilità nella struttura GAA

### 3.2.1 Geometria del dispositivo simulato

Le modalità con cui sono stati analizzati l'RDF e l'RTN per la struttura GAA sono le stesse del caso planare.

Il primo passo è definire la struttura nella quale verranno implementati i profili di drogaggio atomistici generati col metodo Monte Carlo. La geometria che è stata utilizzata

è riportata in Figura 3.5. La lunghezza del canale è la stessa del MOS planare della



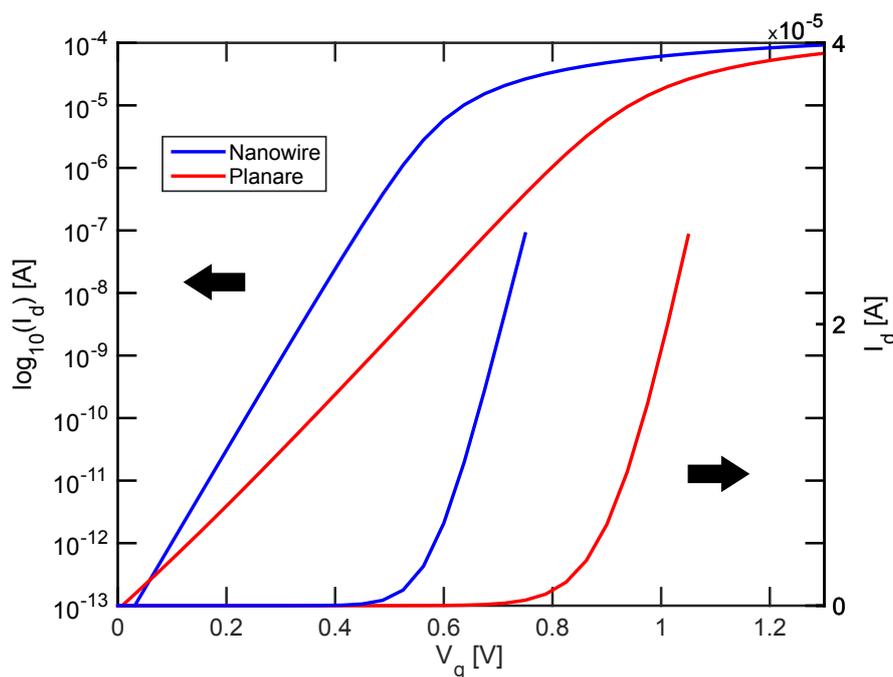
**Figura 3.5:** Struttura GAA implementata nel simulatore per l'analisi atomistica del nanowire. Il profilo di drogaggio di source e drain e la concentrazione di accettori nel substrato sono le stesse di Figura 3.1.

sezione 3.1, cioè 15 nm. La scelta del raggio ha dovuto tenere conto sia dell'effetto di canale corto, che impone valori di  $R$  molto minori di  $L$  (sezione 1.2.2), che degli effetti quantistici in uno strato di silicio della larghezza di  $2R$ , dove, se  $R$  diventa molto piccolo, le equazioni DG potrebbero non essere più sufficienti per una descrizione accurata della distribuzione dei portatori. È stato scelto così  $R = 4$  nm. Per gli altri parametri i valori sono  $t_{ox} = 1$  nm,  $N_d = 10^{20}$  cm $^{-3}$  e  $N_a = 5 \times 10^{18}$  cm $^{-3}$ . Per poter confrontare la struttura GAA con la struttura planare si è deciso di uguagliare le capacità dell'ossido dei due transistor:

$$C_{ox,bulk} = \frac{\varepsilon_{ox}}{t_{ox}} \cdot WL, \quad (3.1)$$

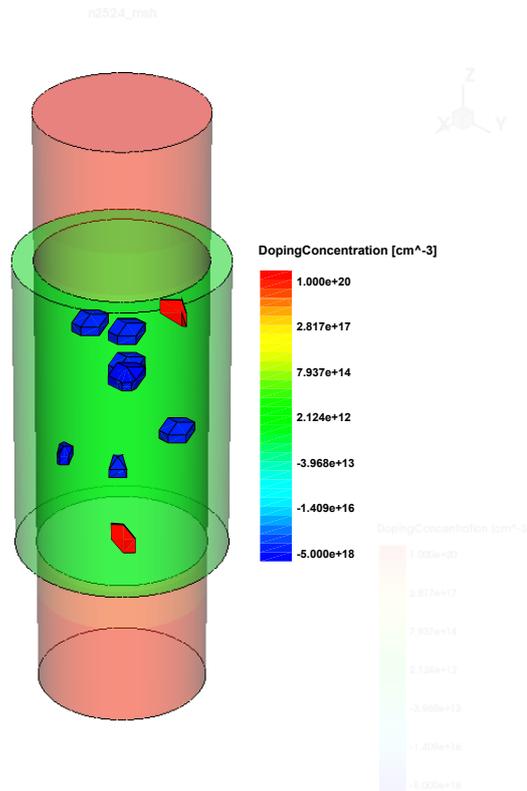
$$C_{ox,GAA} = \frac{\varepsilon_{ox} 2\pi L}{\ln\left(1 + \frac{t_{ox}}{R}\right)}, \quad (3.2)$$

dove  $C_{ox,bulk}$  è la capacità dell'ossido nella la struttura planare mentre  $C_{ox,GAA}$  è quella dell'ossido nella struttura GAA. Con tale criterio, che consente di avere la stessa transconduttanza in condizioni di *on* per le due strutture, è stata scelta la  $W$  del transistor planare.



**Figura 3.6:** Transcaratteristica del transistor planare di Figura 3.1 (in rosso) e del nanowire di Figura 3.5 (in blu), a  $V_{ds} = 50$  mV. Le stesse curve sono rappresentate sia in scala logaritmica che lineare. Nel nanowire, come previsto, la pendenza di sottosoglia è minore poiché il gate che circonda il canale riduce maggiormente l'effetto di canale corto.

Nella sezione 1.2 è stato detto che i dispositivi GAA sono più efficaci nel ridurre l'effetto di canale corto rispetto alla struttura planare. Le curve  $I - V$  dei due diversi transistor con un profilo di drogaggio continuo su tutto il dispositivo sono riportate in Figura 3.6. Si può osservare come il nanowire abbia delle caratteristiche elettrostatiche superiori rispetto al MOS planare. Infatti la pendenza di sottosoglia è minore poiché il gate che circonda tutto il volume del silicio controlla più efficacemente il canale, riducendo il DIBL e il roll-off della tensione di soglia. Inoltre l'inversione di volume contribuisce ad avvicinare il comportamento della pendenza di sottosoglia a quello ideale di 60 mV/decade. I valori di STS sono circa 110 mV/decade per il transistor planare e 75 mV/decade per il nanowire, segno che comunque la lunghezza del canale è talmente ridotta che i campi di source e drain penetrano all'interno del canale. Infine il nanowire presenta una corrente di sottosoglia maggiore del planare nonostante l'effetto di canale corto sia minore. Ciò si deve al fatto che gli elettroni di conduzione non sono limitati all'area sotto il gate come accade nel MOS di tipo bulk, ma sono distribuiti uniformemente su tutto il volume e quindi la sezione attraverso cui scorrono è maggiore. Guardando la curva in scala lineare si può notare che uguagliando la (3.2.1) e la (3.2.1) si è ottenuta la stessa transconduttanza di on per i due dispositivi.



**Figura 3.7:** Esempio del profilo di drogaggio atomistico di un nanowire utilizzato per le simulazioni di RDF e RTN. Il punto di stacco tra il profilo continuo e quello discreto è l’inizio del gate.

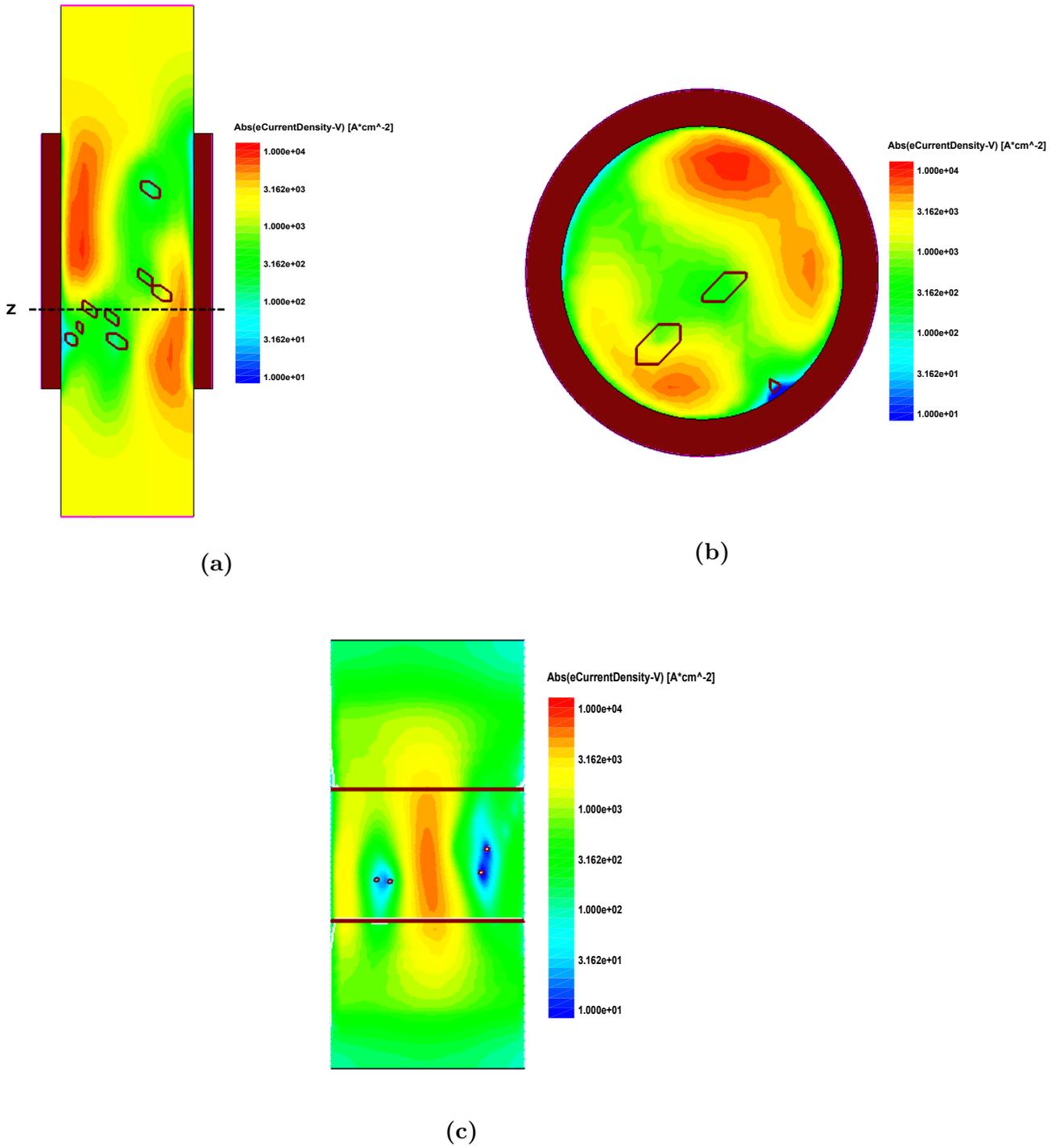
### 3.2.2 Studio delle dipendenze RDF

Per l’analisi delle fluttuazioni RDF sono stati generati 100 profili di drogaggio Monte Carlo per la struttura di Figura 3.5, allo stesso modo che nel caso del transistor planare (sezione 3.1). Un esempio di profilo atomistico è mostrato in Figura 3.7. Come fatto per la struttura di tipo bulk, si è scelto di rendere discreta la zona sotto il gate, lasciando un drogaggio continuo nel source e nel drain. Nel canale sono presenti gli accettori discreti dovuti al profilo uniforme con concentrazione  $N_a$  e anche alcuni donori dovuti alla discretizzazione delle code del profilo gaussiano di source e drain. Gli atomi verso l’ossido di gate occupano un volume più piccolo poiché la mesh diventa più fitta avvicinandosi al gate lungo la coordinata radiale.

In sezione 1.2.1 è stato mostrato che nei nanowire si verifica il fenomeno dell’inversione di volume. Sottosoglia le bande sono piatte e gli elettroni di inversione sono distribuiti con la stessa concentrazione in tutto il canale. In questa situazione, se in un punto del nanowire è presente un accettore che alza la barriera di potenziale, questo non bloccherà l’intero percorso di corrente lungo la sua direzione ma solamente in un’area attorno a sé,

poiché gli elettroni hanno a disposizione tutto il volume circostante per la conduzione. In un MOS planare, invece, gli elettroni decrescono esponenzialmente allontanandosi dall'interfaccia con l'ossido di gate, e quindi su tutta la zona svuotata l'area efficace per la conduzione è soltanto lo strato vicino al gate. Se un accettore si trova in prossimità di questo strato, il suo impatto sulla corrente sarà più forte che nel caso del nanowire, poiché gli elettroni possono aggirare l'ostacolo soltanto lateralmente. Tuttavia, proprio perché la corrente sottosoglia è distribuita uniformemente su tutto il silicio, nei dispositivi GAA gli ioni nel volume hanno lo stesso peso di quelli vicino all'interfaccia nel determinare i percorsi percolativi, mentre nel MOS planare gli accettori nella zona svuotata lontano dal gate hanno un'influenza molto minore sulla corrente. Uno ione che alza la barriera di potenziale al centro del volume di un nanowire genererà un percorso percolativo verso l'esterno del canale, esattamente come uno ione vicino all'interfaccia crea un filamento verso l'interno. Un esempio delle dipendenze appena discusse viene mostrato in Figura 3.8, dove è riportata la densità di corrente di elettroni in valore assoluto su alcune sezioni prese dalla simulazione di un profilo atomistico per ognuna delle due strutture. Sono stati scelti due profili che avessero una configurazione di droganti distribuita circa uniformemente in direzione perpendicolare al canale, ad una certa altezza lungo  $L$ . In generale si può vedere come gli accettori, di cui sono tracciati i contorni, blocchino il percorso della corrente nella zona in cui si trovano. Nel caso del MOS planare (Figura 3.8c) viene determinata una barriera per gli elettroni che si estende lungo tutto il canale. La corrente si stringe in un unico filamento e passa attraverso un piccolo spazio lasciato tra gli ioni. Per il nanowire (Figura 3.8a) la situazione sembra la stessa, con alcuni ioni che alzano la barriera di potenziale in prossimità della coordinata  $Z$  segnata sul grafico. Tuttavia una sezione perpendicolare a  $Z$ , indicata con una linea a tratti e riportata in Figura 3.8b, mostra che la corrente viene bloccata solo nell'area occupata dagli accettori, verso il centro, mentre le zone laterali sono accessibili e la corrente vi scorre normalmente. Sempre considerando la sezione di Figura 3.8b, si nota che la presenza degli accettori nel centro del canale determina la concentrazione della corrente in due filamenti lungo l'interfaccia, e dunque questi ioni sono efficaci nel determinare percorsi percolativi.

I due comportamenti appena discussi, che distinguono la percolazione nel nanowire da quella nel MOS planare, hanno un effetto opposto sulla variabilità di  $V_t$ . L'aumento dei percorsi percolativi che si possono creare a causa dell'inversione di volume tende a ridurre le fluttuazioni della tensione di soglia, poiché l'intensità dei singoli filamenti diminuisce. La presenza di un numero maggiore di atomi che influenzano la percolazione, a parità di concentrazione di drogaggio rispetto al transistor di tipo bulk, porta a delle fluttuazioni di  $V_t$  maggiori, dato che l'intensità dei filamenti che si creano è più elevata a causa di una riduzione del numero di questi. Dalla distribuzione delle  $V_t$  delle 100 simulazioni



**Figura 3.8:** Confronto tra la percolazione nella struttura GAA e nella struttura di tipo bulk. La grandezza rappresentata in scala colore è il modulo di  $\vec{J}_n$ . Sulle sezioni è tracciato il contorno degli accettori presenti. (a) Sezione del nanowire perpendicolare al diametro. È evidenziata con una linea a tratti il punto in cui è stata effettuata la sezione (b). (b) Sezione perpendicolare alla direzione del canale nel punto Z. (c) Sezione del dispositivo planare in direzione parallela al canale, ad una profondità in cui lo strato di inversione è nell'intorno del suo massimo. Si può notare come gli accettori discreti ostacolano il passaggio di corrente (toni dall'azzurro al verde).

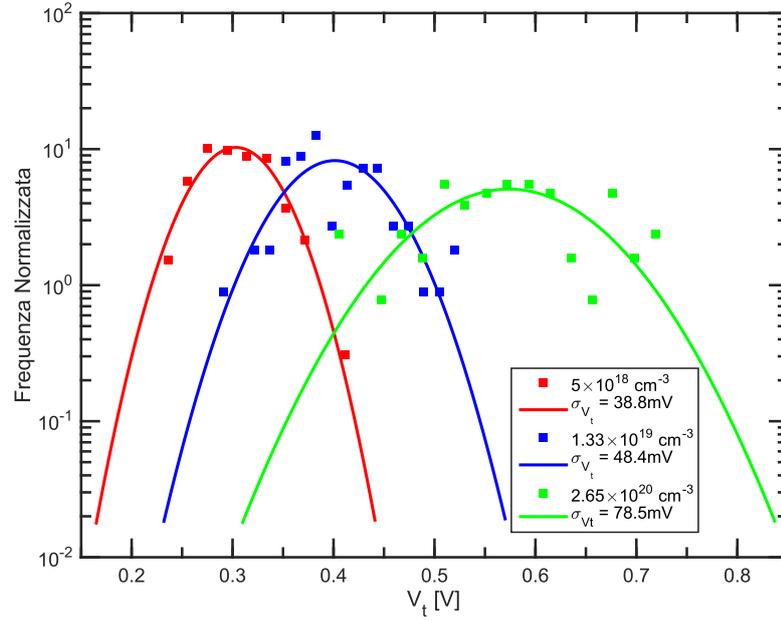
	$N_a = 5 \times 10^{18} \text{ cm}^{-3}$	$N_a = 1.33 \times 10^{19} \text{ cm}^{-3}$	$N_a = 1.33 \times 10^{19} \text{ cm}^{-3}$
$R = 4 \text{ nm}$	X	X	X
$R = 6 \text{ nm}$			X
$R = 8 \text{ nm}$			X

**Tabella 3.1:** Matrice delle possibili combinazioni dei valori di  $R$  ed  $N_a$  che sono stati scelti per l'analisi della dipendenza delle fluttuazioni RDF dallo scaling. Il simbolo X indica quelle che sono state analizzate in modo statistico. Per ognuna di queste sono stati generati 100 profili Monte Carlo di drogaggio atomistico e si sono simulate le curve  $I - V$  per estrarre la tensione di soglia.

atomistiche è stata trovata una  $\sigma_{V_t}$  di 38.8 mV. I dati simulati e la gaussiana che li interpola sono mostrati in Figura 3.9 (in rosso). La dispersione di  $V_t$  è minore che per la struttura planare analoga, dove  $\sigma_{V_t} = 45 \text{ mV}$ . Tuttavia il valore trovato non si discosta molto da quello della struttura di tipo bulk, segno che nessuno dei due effetti opposti considerati per il nanowire è prevalente sull'altro.

Si è voluto analizzare la dipendenza dai parametri di scaling della dispersione di  $V_t$  dovuta all'RDF, in analogia con quanto fatto negli studi sui dispositivi planari (sezione 1.4.1). A questo scopo sono state eseguite delle simulazioni Monte Carlo, sempre con un campione di 100 transistor, al variare della concentrazione di accettori nel canale e del raggio del nanowire. Mentre viene fatto variare un parametro, tutti gli altri sono mantenuti costanti. In Tabella 3.1 viene riportata la matrice delle possibili combinazioni dei parametri, dove sono mostrate quelle che sono state investigate con le simulazioni statistiche. Per la concentrazione di accettori, oltre a  $N_a = 5 \times 10^{18} \text{ cm}^{-3}$ , sono state scelte  $N_a = 1.33 \times 10^{19} \text{ cm}^{-3}$  e  $N_a = 2.65 \times 10^{19} \text{ cm}^{-3}$  perché davano rispettivamente un valore medio per la Poissoniana di 10 e 20 atomi sotto il gate. In seguito a quanto discusso prima sulla percolazione nei nanowire, se il numero di atomi nel canale aumenta e le dimensioni del volume rimangono inalterate, la barriera che gli ioni possono formare per il passaggio della corrente è maggiore e i filamenti che si creano sono di meno e più intensi. Dunque il loro impatto sulle fluttuazioni di  $V_t$  sarà più grande. In Figura 3.9 sono riportati i valori delle simulazioni per le tre diverse  $N_a$  e le relative gaussiane che li interpolano. Si nota che  $\sigma_{V_t}$  aumenta all'aumentare di  $N_a$ , coerentemente con la dipendenza attesa.

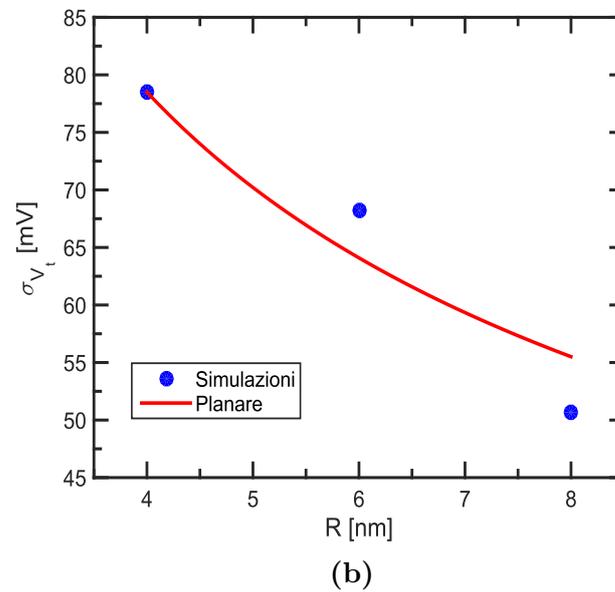
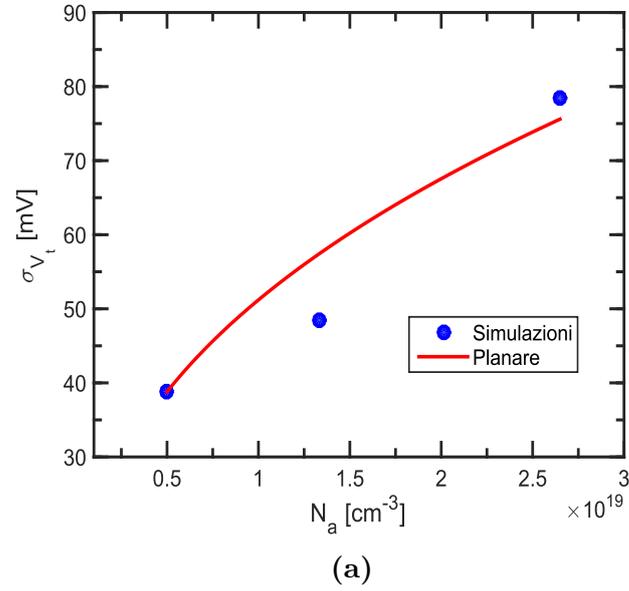
Successivamente la concentrazione di accettori nel substrato è stata fissata e si è cambiato il raggio del transistor. In questa situazione si ha sia un aumento del numero di atomi nel canale, che un aumento del volume del dispositivo. Il primo effetto, come è stato osservato in precedenza, porta ad un aumento della variabilità di  $V_t$ , poiché si formano dei filamenti più intensi e delle configurazioni di droganti che hanno un impatto maggiore



**Figura 3.9:** Distribuzione dei valori di  $V_t$  ottenuti dalle simulazioni Monte Carlo di 100 nanowire per diversi valori del doping nel canale (simboli). Sono sovrapposti i fit gaussiani corrispondenti (linee).

sulla barriera di potenziale. L'aumento del volume di silicio invece porta ad avere delle fluttuazioni di  $V_t$  ridotte, poiché lo spazio per il passaggio della corrente aumenta e si forma un numero maggiore di percorsi percolativi, con un peso relativo di ognuno di essi che è minore. I valori del raggio del nanowire che sono stati esplorati sono  $R = 4$  nm,  $R = 6$  nm ed  $R = 8$  nm. Le deviazioni standard corrispondenti sono  $\sigma_{V_t} = 78.5$ ,  $68.2$  e  $50.7$  mV. La dispersione della tensione di soglia diminuisce all'aumentare del raggio. In questo caso sembra che l'effetto dell'aumento di volume sia prevalente sull'effetto provocato dal maggiore numero di accettori sotto il gate. Nel transistor di tipo bulk l'analogo di  $R$  è  $W$ . In tale struttura, come visto in sezione 1.4.1, l'andamento di  $\sigma_{V_t}$  è decrescente con l'aumentare di  $W$  poiché diventano maggiori le possibilità di passaggio per la corrente. L'effetto è prevalente su quello dovuto all'aumento degli ioni nella zona svuotata poiché, dato che nel MOS planare gli elettroni sono concentrati principalmente in uno strato vicino all'interfaccia Si-SiO<sub>2</sub>, gli accettori efficaci nel generare una percolazione sono soltanto quelli nei pressi dell'ossido, e il loro aumento percentuale sul totale è ridotto.

Per fare un confronto tra le dipendenze dell'RDF nella struttura GAA e nella struttura planare, in Figura 3.10 sono riportate le dispersioni della distribuzione di  $V_t$  in funzione della concentrazione di accettori nel canale (Figura 3.10a) e in funzione del raggio del nanowire (Figura 3.10b), insieme all'andamento di  $\sigma_{V_t}$  per il caso planare, dato dalla (1.10). Come detto in precedenza, in Figura 3.10b l'andamento tracciato è quello per  $W$ , che è



**Figura 3.10:** Valori della dispersione di  $V_t$  dovuta a RDF in funzione dei parametri del nanowire (simboli) e andamento della stessa nel transistor planare dall'equazione (1.10) (linea continua). (a) Dipendenza di  $\sigma_{V_t}$  dalla concentrazione di accettori nel canale del nanowire e relativo andamento previsto dalla (1.10). (b) Dipendenza di  $\sigma_{V_t}$  dal raggio del nanowire e andamento previsto dalla (1.10) per  $W$ .

l'analogo di  $R$  nel caso planare. I punti simulati non hanno lo stesso andamento previsto dalla (1.10) (curva rossa). In particolare, in Figura 3.10a si ha che il punto intermedio ha un valore di dispersione molto vicino al primo punto, mentre il valore di  $\sigma V_t$  corrispondente a  $N_a = 2.65 \times 10^{19} \text{ cm}^{-3}$  cresce in modo brusco. Questo comportamento porta a pensare che una delle due caratteristiche viste per la percolazione nel nanowire prevalga sull'altra a seconda della quantità di droganti che ci sono sotto il gate. Al di sotto di un certo numero di atomi, la forza degli ioni non è sufficiente a creare dei percorsi percolativi intensi e i filamenti si disperdono nel volume. Al di sopra di tale numero invece gli accettori sono efficaci nel creare dei filamenti intensi e in questo caso, poiché nel nanowire tutti gli atomi nel volume sono determinanti per la percolazione, si avrà una dispersione maggiore che nel caso planare.

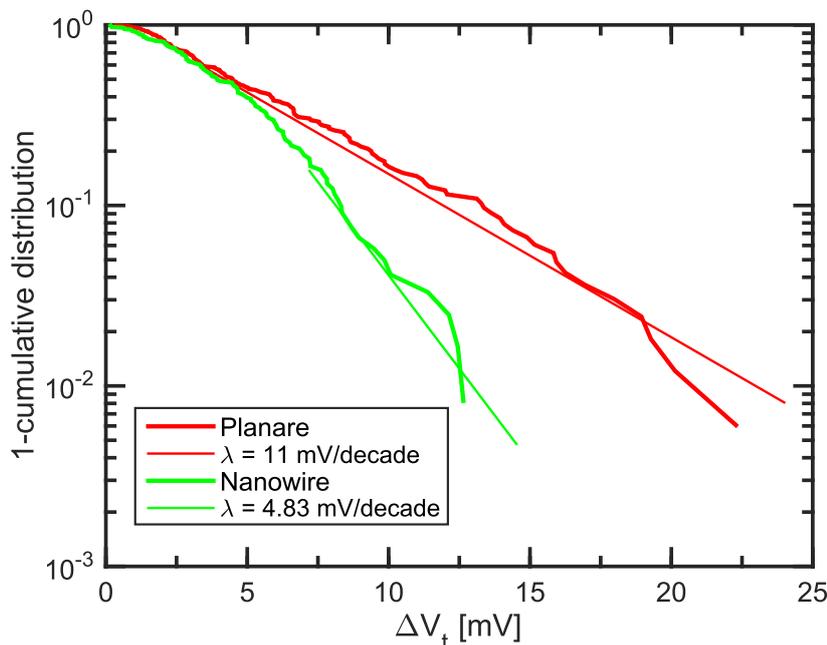
Come risultato di questa analisi di RDF condotta sulla struttura GAA si è trovato che l'inversione di volume determina delle proprietà deverse per il fenomeno della percolazione in questi dispositivi rispetto alle strutture planari. Nel nanowire ci sono due effetti, influenti in maniera opposta su  $V_t$ , che determinano la percolazione, ovvero la conduzione nell'intero volume e il peso identico che ogni drogante nel canale ha in confronto agli altri nel determinare i filamenti di corrente. Il fatto che queste due caratteristiche coesistano rende difficile prevedere quale delle due prevalga e l'andamento con i parametri di scaling non è facilmente determinabile.

### 3.2.3 Studio delle dipendenze RTN

Si vogliono confrontare ora le caratteristiche dell'RTN nel nanowire con quelle osservate nel transistor di tipo bulk per indagare le differenze tra i due casi e conoscere le dipendenze rispetto ai principali parametri coinvolti nello scaling.

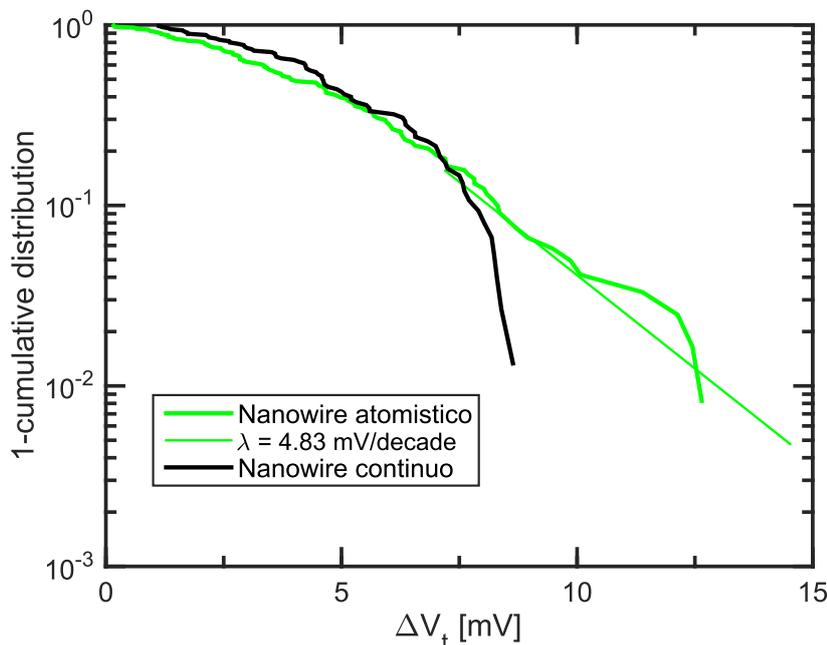
Per analizzare le fluttuazioni RTN della struttura GAA sono stati simulati gli stessi profili utilizzati per l'RDF, a cui è stata aggiunta una carica all'interfaccia generata casualmente con distribuzione di probabilità uniforme, esattamente come è stato fatto per il MOS planare (sezione 3.1). In Figura 3.11 è riportata la distribuzione cumulativa dei  $\Delta V_t$  estratti dal campione di 100 nanowire con  $R = 4 \text{ nm}$  e  $N_a = 5 \times 10^{18} \text{ cm}^{-3}$ . La curva è confrontata con la distribuzione del caso planare di Figura 3.4, che ha gli stessi parametri. Ciò che si osserva immediatamente è che le fluttuazioni di RTN del nanowire sono decisamente minori. Rispetto alla distribuzione del planare la coda della cumulativa GAA si stacca a probabilità più piccole, ma soprattutto la pendenza  $\lambda$  è molto inferiore.

Per confermare che quella del nanowire sia effettivamente una coda esponenziale dovuta al formarsi di percorsi percolativi è stato simulato anche un campione di 100 nanowire con doping continuo su tutto il dispositivo, dove è stata posizionata una trap-



**Figura 3.11:** Distribuzione cumulativa inversa delle ampiezze di RTN per 100 nanowire con  $R = 4$  nm e  $N_a = 5 \times 10^{18}$  cm $^{-3}$  (in verde). Viene riportata anche la distribuzione del caso planare di Figura 3.4 (in rosso) per un confronto. Sulle code delle curve è stato tracciato il fit della pendenza  $\lambda$ . Le fluttuazioni RTN del caso GAA sono molto minori rispetto al caso bulk.

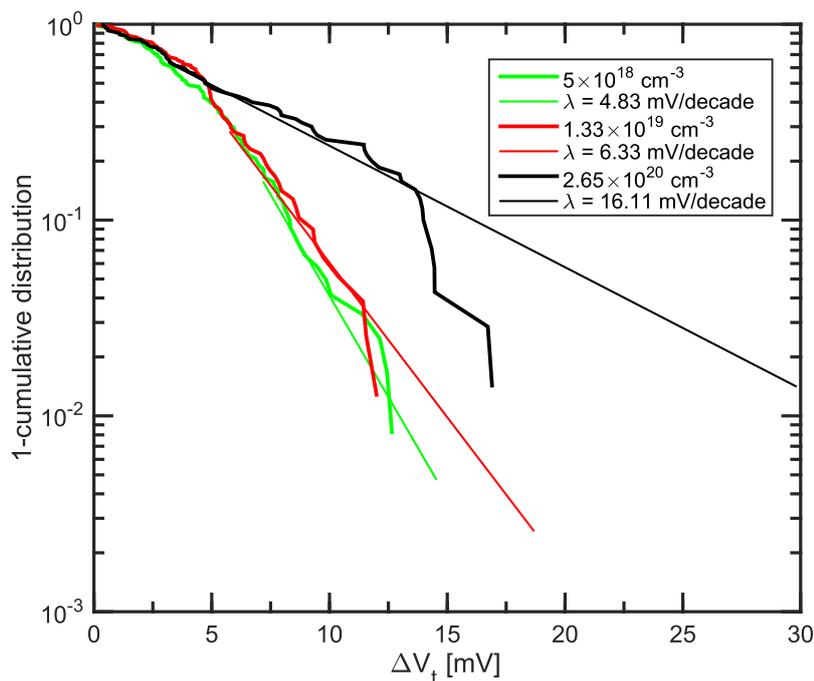
pola casuale con probabilità uniforme all'interfaccia silicio-ossido in modo da ricavare la distribuzione continua di  $\Delta V_t$ . Come visto in sezione 1.4.2, la cumulativa continua non presenta code esponenziali poiché non si generano percorsi preferenziali per la corrente dovuti agli atomi di drogante discreti. Le due distribuzioni sono messe a confronto in Figura 3.12. Si può notare che le curve hanno lo stesso andamento gaussiano fino a probabilità del 10%, oltre le quali la cumulativa continua raggiunge il valore massimo di  $\Delta V_t$ , mentre quella atomistica prosegue con una coda a pendenza  $\lambda$ . Nel caso atomistico, dunque, la natura dell'RTN è la stessa del planare, cioè sono presenti fluttuazioni anomale dovute a trappole che si trovano in prossimità di percorsi in cui la corrente si concentra a causa della presenza dei droganti discreti. Tuttavia l'effetto che una trappola ha su questi è minore che nel caso planare, e se ne vuole studiare la causa. La percolazione nella struttura GAA è determinata dall'inversione di volume, come è stato discusso per la variabilità causata da RDF (sezione 3.2.2). Ricordando quanto è stato osservato a proposito della Figura 3.10a, se il numero di atomi sotto il gate non è abbastanza elevato, i percorsi percolativi possibili per la corrente sono di più in un nanowire che in un transistor planare, poiché gli elettroni hanno a disposizione l'intero volume del silicio per la conduzione sottosoglia. In una simile situazione un filamento di corrente che viene bloccato da una trappola avrà un peso minore sul numero di filamenti totale, e dunque anche



**Figura 3.12:** Confronto tra la cumulativa inversa di  $\Delta V_t$  ottenuta simulando 100 nanowire con profilo di drogaggio continuo su tutto il dispositivo (in nero) e quella ottenuta da 100 simulazioni con drogaggio atomistico (in verde), per il caso  $R = 4$  nm e  $N_a = 5 \times 10^{18}$  cm $^{-3}$ . Lo stacco di una coda esponenziale è presente soltanto sulla distribuzione atomistica.

sulla variazione di  $V_t$ . In un MOS di tipo bulk invece i percorsi della corrente passano esclusivamente in un intorno dell'interfaccia Si-SiO $_2$ , dunque rimuoverne uno (ad esempio quello visibile in Figura 3.8c) ha un'influenza molto maggiore sulla corrente totale. In aggiunta, per la geometria cilindrica che possiede il gate, una trappola all'interfaccia è efficace nel bloccare la percolazione solamente degli elettroni che passano vicino alla superficie, mentre non influisce sui percorsi che si muovono all'interno. Ciò non succede in un MOS planare, dove, come già detto, gli elettroni che vanno da source a drain sono concentrati prevalentemente vicino all'interfaccia. Per questi motivi, le stesse ampiezze di RTN del planare con  $N_a = 5 \times 10^{18}$  cm $^{-3}$  avvengono a probabilità molto più basse nel nanowire con la stessa concentrazione di accettori.

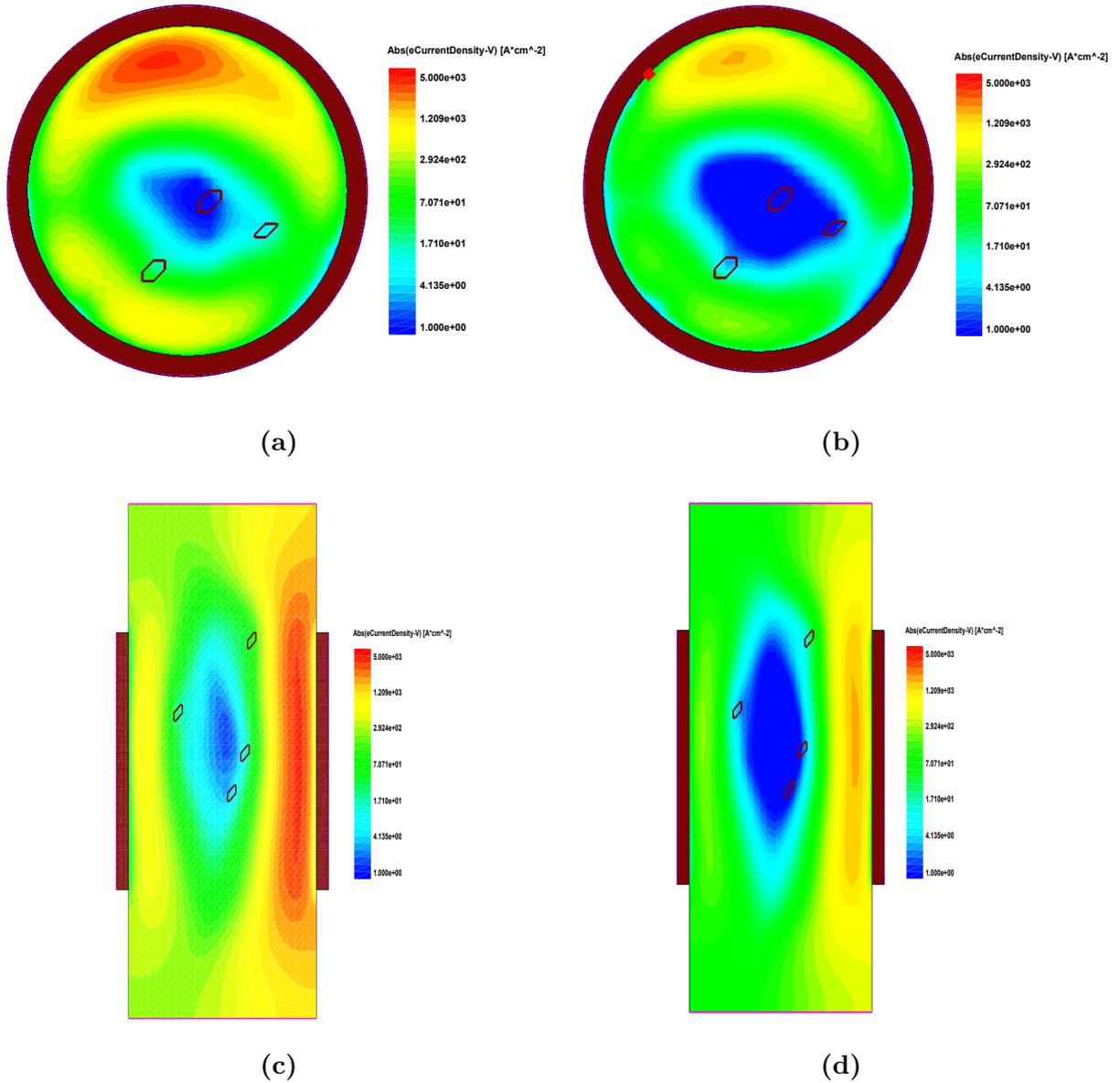
Anche per le fluttuazioni RTN si sono volute analizzare le dipendenze dai parametri coinvolti nello scaling. I casi simulati sono gli stessi di Tabella 3.1. In Figura 3.13 sono mostrate le distribuzioni cumulative di  $\Delta V_t$  al variare della concentrazione di accettori nel canale. Aumentando gli ioni presenti sotto il gate viene ridotto il numero dei filamenti in cui passa la corrente. Inoltre gli elettroni sono più concentrati all'interno di questi percorsi percolativi. L'insieme dei due effetti aumenta le possibilità di avere  $\Delta V_t$  elevati. Dalla figura si nota che le code delle cumulative  $N_a = 5 \times 10^{18}$  cm $^{-3}$  e  $N_a = 1.33 \times 10^{19}$  cm $^{-3}$  sono vicine tra loro, mentre quella del caso  $N_a = 2.65 \times 10^{19}$  cm $^{-3}$  ha una pendenza decisamente



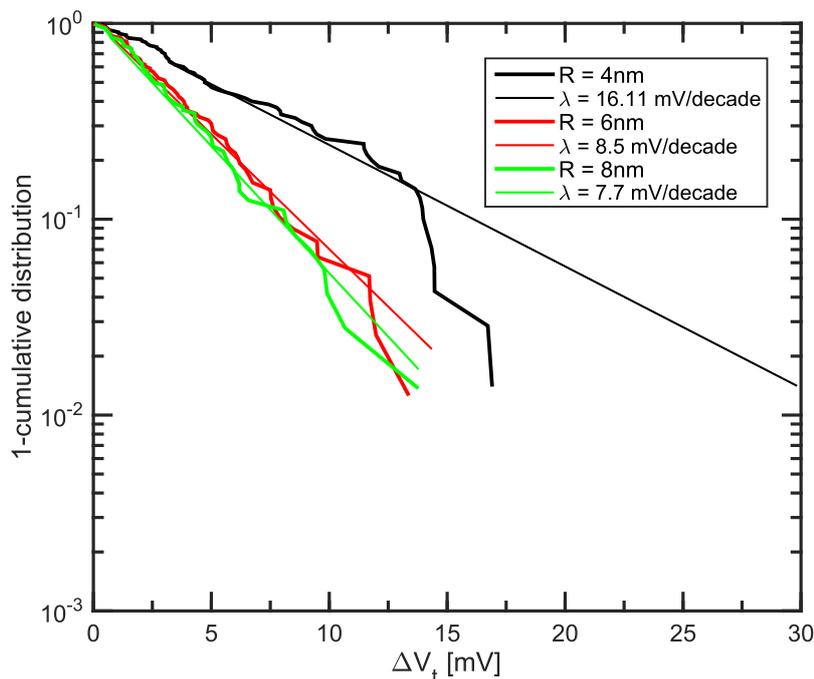
**Figura 3.13:** Distribuzioni di  $\Delta V_t$  dovuto a RTN per tre campioni statistici di nanowire con diversa concentrazione di droganti nel canale. Sovrapposto ad ogni distribuzione è presente il fit della coda esponenziale a pendenza  $\lambda$ . Le fluttuazioni aumentano all'aumentare degli ioni di drogante nel canale.

maggiore e che si stacca prima dalla gaussiana. Questo comportamento è la conseguenza di quanto già osservato in sezione 3.2.2, ovvero che la forza dei percorsi percolativi si intensifica a partire da un certo numero di atomi nel canale, che viene raggiunto solo dalla distribuzione  $N_a = 2.65 \times 10^{19} \text{ cm}^{-3}$ . Un esempio di quello che accade è riportato in Figura 3.14. Sono rappresentate le sezioni verticali e radiali di uno dei transistor di raggio 8 nm, con  $N_a = 2.65 \times 10^{19} \text{ cm}^{-3}$ . Il numero di accettori sotto il gate in questo caso è abbastanza grande da permettere che si formi un numero ridotto di percorsi percolativi. Se la configurazione casuale dei droganti è tale da spingere la corrente verso l'esterno del transistor, una trappola nelle vicinanze può bloccare il filamento e causare una fluttuazione  $\Delta V_t$  elevata.

La Figura 3.15 mostra le code delle cumulative di  $\Delta V_t$  all'aumentare di  $R$ . Anche in questo caso sembra esserci una saturazione di  $\lambda$ . Al crescere del raggio, il numero di atomi sotto il gate aumenta. Tuttavia aumenta anche il volume del dispositivo e quindi la probabilità di avere percorsi percolativi all'interno del canale. Osservando la figura si può concludere che nei casi  $R = 6 \text{ nm}$  e  $R = 8 \text{ nm}$  questa probabilità fa sì che il numero di filamenti che passano vicino alla superficie del silicio sia piccolo, e quindi l'effetto di una trappola all'interfaccia è debole, con una pendenza delle code quasi identica. Se



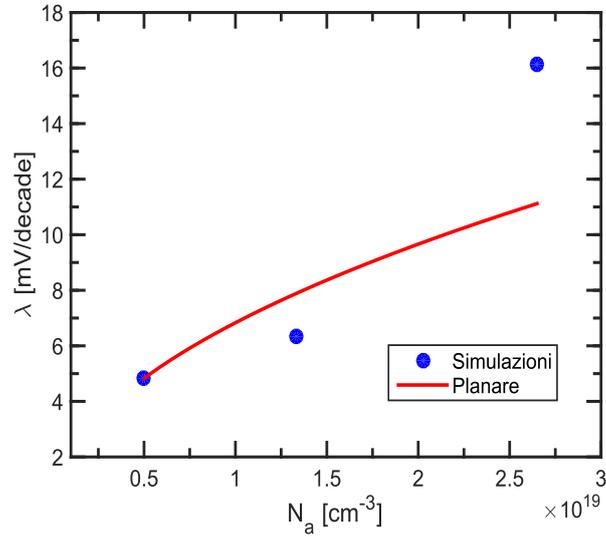
**Figura 3.14:** Sezioni di un nanowire atomistico simulato per lo studio dell'RTN preso dal campione  $R = 8$  nm,  $N_a = 2.65 \times 10^{19}$  cm<sup>-3</sup>. A sinistra ((a) e (c)) è riportato il caso senza carica all'interfaccia, mentre a destra((b) e (d)) viene mostrato l'effetto della trappola. In alto è mostrata la sezione perpendicolare al canale, all'altezza della trappola (indicata da un quadrato rosso in (b)). In basso è mostrata la sezione parallela a  $L$ , passante per il centro del transistor. I droganti atomistici sono indicati con una linea che ne ricalca l'area occupata. In questo caso il loro numero è sufficientemente elevato da creare un percorso percolativo vicino all'interfaccia. La trappola casuale posizionata nei pressi di questo lo riduce e causa un'ampiezza  $\Delta V_t$  grande.



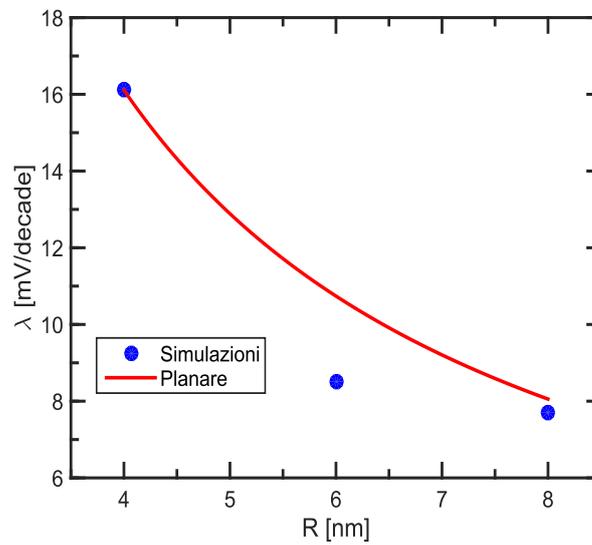
**Figura 3.15:** Distribuzioni cumulative di  $\Delta V_t$  per tre campioni statistici di nanowire al variare del raggio. Sovrapposto ad ogni distribuzione è presente il fit della coda esponenziale a pendenza  $\lambda$ . Le fluttuazioni aumentano al diminuire di  $R$  poiché i percorsi percolativi possibili diminuiscono.

consideriamo invece la struttura con  $R = 4$  nm,  $\lambda$  è molto maggiore, segno che l'effetto del numero totale di ioni nel canale è determinante per la percolazione rispetto al volume a disposizione per la conduzione degli elettroni ed è più probabile che dei filamenti di corrente intensi passino nei pressi dell'interfaccia.

Per completare il confronto con la struttura planare sono stati analizzati gli andamenti della pendenza delle code in funzione dei parametri interessati maggiormente nello scaling, come fatto in sezione 3.2.2 per l'RDF. In Figura 3.16 sono riportate le  $\lambda$  simulate al variare di  $R$  ed  $N_a$  e la legge di potenza prevista dall'equazione 1.4.2 per dispositivi di tipo bulk. Allo stesso modo di quanto visto nel caso dell'RDF, le code di RTN del nanowire non seguono la dipendenza che è stata trovata per le strutture planari. Come è stato osservato per Figura 3.13 e Figura 3.15, in generale si ha una saturazione della pendenza per bassi valori di  $N_a$  o alti valori di  $R$ . La dipendenza di  $\lambda$  è invece più forte in dispositivi che hanno raggi piccoli o concentrazioni di accettori elevate. Una dipendenza più forte da  $N_a$  si comprende tenendo presente che in un dispositivo GAA tutti gli ioni che si trovano nel volume sono determinanti nell'aumentare la concentrazione di corrente nei percorsi percolativi che si creano. Più un filamento di corrente è intenso, maggiore è l'effetto che una trappola repulsiva nelle sue vicinanze causa sulla tensione di soglia. Nel MOS planare



(a)



(b)

**Figura 3.16:** Pendenza della coda RTN della distribuzione di  $\Delta V_t$  al variare dei parametri del nanowire (simboli) e andamento previsto dall'equazione (1.10) per un dispositivo planare (linea continua). (a) Dipendenza di  $\lambda$  dalla concentrazione di accettori nel canale del nanowire. (b) Dipendenza di  $\sigma_{V_t}$  dal raggio del nanowire e andamento previsto dalla (1.10) per  $W$ . In linea a tratti sono indicati gli andamenti seguiti dalle code simulate, che si discostano da quelli trovati per i transistor di tipo bulk.

invece, gli accettori efficaci per la percolazione sono soltanto quelli nei pressi dello strato di inversione. Un aumento del numero totale di atomi nella zona svuotata porta ad un aumento relativo degli ioni che determinano i percorsi preferenziali per gli elettroni, e quindi l'effetto dell'aumento di  $N_a$  sarà minore rispetto a quello di un nanowire. Per quanto riguarda la dipendenza dal raggio per valori piccoli di  $R$ , questa può essere dovuta al fatto che, come accade riducendo  $W$  nel planare, vengono ridotti i percorsi possibili per la corrente. Nel caso di un transistor GAA, poiché la diminuzione è relativa a tutto il volume, aumenta la probabilità di avere filamenti che passano vicino all'interfaccia, e dunque l'effetto sull'ampiezza  $\Delta V_t$  è maggiore che per un dispositivo planare.

Quello che è stato osservato per l'RTN nei transistor GAA è in generale una maggiore robustezza alle fluttuazioni anomale della tensione di soglia rispetto alle strutture planari. Tuttavia, se i raggi di queste strutture diventano troppo scalati o il rapporto tra ioni sotto il gate e volume del dispositivo supera una certa soglia, le prestazioni del nanowire degradano rapidamente rispetto a quelle di un planare.

# Capitolo 4

## Nanowire a canale non drogato

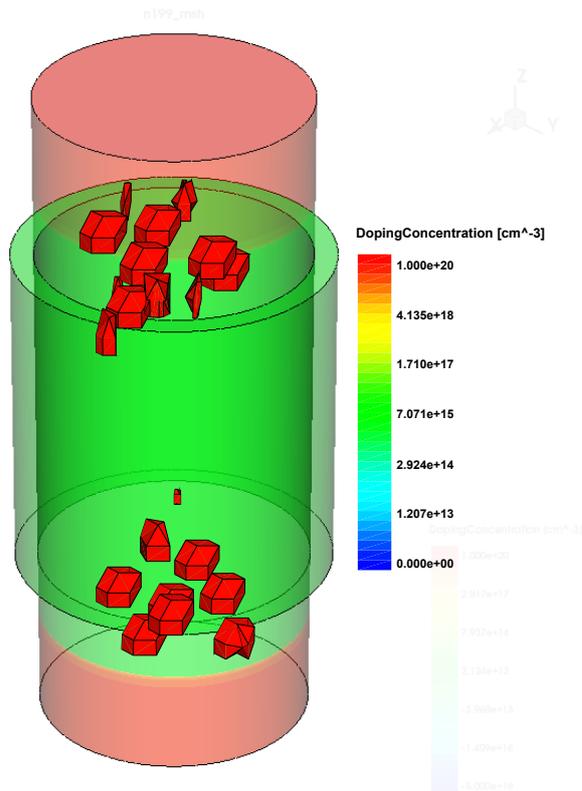
Nei precedenti capitoli è stato detto che la struttura GAA gode di proprietà elettrostatiche che la rendono più resistente all'effetto di canale corto rispetto ad altre geometrie. Tuttavia gli ioni discreti presenti nel canale sono responsabili di una dispersione della tensione di soglia di questi dispositivi, a causa dei fenomeni di RDF e RTN. Per ridurre le fluttuazioni di  $V_t$  dovute alla variabilità atomistica si può quindi pensare di realizzare i nanowire senza il drogaggio di accettori nella zona sotto il gate, poiché il roll-off di  $V_t$  e il DIBL sono già ridotti efficacemente da tali strutture. Le fluttuazioni però non vengono eliminate del tutto, poiché nel canale sono comunque presenti i donori che hanno diffuso dai profili di source e drain. Nei dispositivi a canale drogato visti nel capitolo 3 gli ioni discreti nella zona svuotata provenienti dal source e dal drain avevano un impatto trascurabile. Nel caso in esame, essendo gli unici atomi presenti nel canale, la loro influenza sulla barriera di potenziale che vedono gli elettroni sottosoglia determina le caratteristiche di RDF e RTN del dispositivo. Per indagare la variabilità di una simile struttura è quindi necessario rendere atomistici i profili gaussiani di source e drain e studiare tramite simulazioni Monte Carlo un campione di transistor.

### 4.1 Analisi delle fluttuazioni RDF

La struttura sulla quale è stato simulato il drogaggio atomistico, rispetto a quella di Figura 3.5, ha  $N_a = 0 \text{ cm}^{-3}$  e  $R = 6 \text{ nm}$ .

Per l'analisi delle fluttuazioni RDF nel nanowire sono stati simulati con il metodo Monte Carlo i profili di drogaggio di source e drain. I donori sono distribuiti con densità

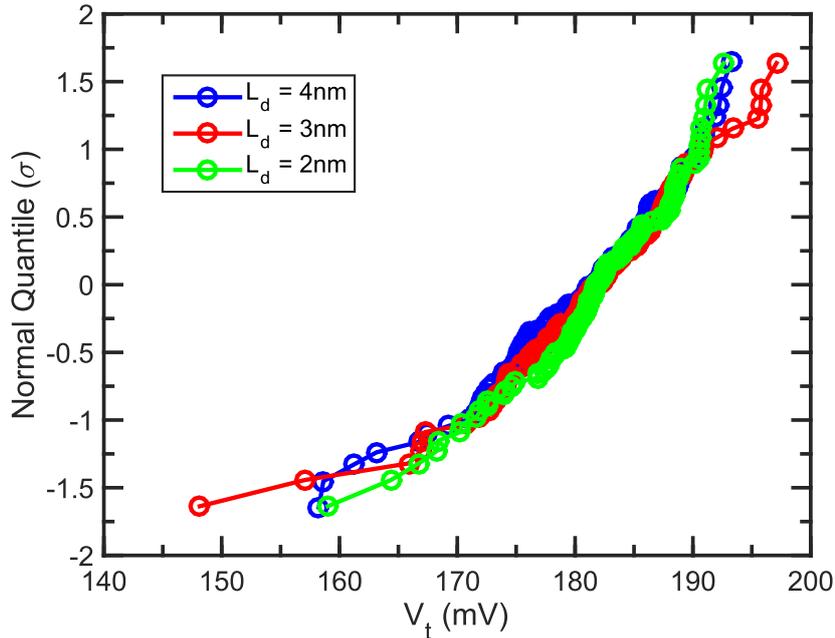
di probabilità uniforme lungo la coordinata radiale, mentre la probabilità di avere degli ioni decresce con andamento gaussiano allontanandosi da source e drain verso il centro del canale. Per tale motivo, i droganti si concentrano maggiormente all'esterno del gate. È necessario allora tenere in conto anche dell'effetto della discretizzazione di questa zona nell'analisi delle variazioni di  $V_t$ . Il profilo di drogaggio continuo non si estende più fino ai bordi del gate, ma viene interrotto ad una certa altezza dall'inizio del source e del drain. Un esempio di configurazione degli atomi implementati è mostrato in Figura 4.1. Si osserva che la maggior parte dei donori discreti si trova all'esterno del canale. È stata



**Figura 4.1:** Esempio del profilo di drogaggio atomistico di un nanowire con il canale non drogato, utilizzato per le simulazioni di RDF. In questo caso gli atomi discreti sono soltanto i donori del profilo gaussiano di source e drain. La zona di discretizzazione è stata aumentata e si estende anche fuori dal gate.

presa una lunghezza di discretizzazione ( $L_d$ ) di 3 nm, cioè la zona in cui il profilo di drogaggio viene trattato come atomistico si estende per 3 nm oltre il gate, da entrambi i lati.

I donori hanno un effetto elettrostatico opposto rispetto a quello visto per gli accettori, ovvero abbassano la barriera di potenziale e quindi attraggono gli elettroni. Ci si aspetta allora che la corrente si concentri in filamenti nei pressi delle posizioni occupate da questi ioni, e non nelle valli che si creano dove non ci sono accettori.



**Figura 4.2:** Distribuzione cumulativa delle tensioni di soglia per un campione di 100 nanowire atomistici al variare della dimensione della zona di discretizzazione. All'asse delle ordinate è stata applicata una trasformazione per facilitare il riconoscimento di distribuzioni di probabilità normali. Dal grafico si nota che è indifferente scegliere una delle tre  $L_d$  per l'analisi dell'RDF.

Per un'analisi delle fluttuazioni della tensione di soglia dei nanowire senza doping di canale bisogna verificare l'effetto che zone di discretizzazione diverse hanno sulla variabilità di  $V_t$ . Infatti, se  $L_d$  aumenta, il numero di atomi discreti presenti nel transistor è maggiore e non è chiaro a priori quale possa essere il loro impatto sulla conduzione. A tale scopo sono state effettuate delle simulazioni Monte Carlo di 100 nanowire atomistici per tre diverse lunghezze di discretizzazione:  $L_d = 2$  nm,  $L_d = 3$  nm e  $L_d = 4$  nm, al fine di valutare le rispettive distribuzioni cumulative. I risultati sono riportati in Figura 4.2. Sull'asse delle ordinate è stata applicata la trasformazione:

$$y = \text{erf}^{-1}(2 \cdot \Phi(x) - 1), \quad (4.1)$$

dove  $\Phi(x)$  è la distribuzione cumulativa della variabile che viene riportata sull'asse delle ascisse e  $\text{erf}^{-1}$  è l'inversa della funzione degli errori:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad (4.2)$$

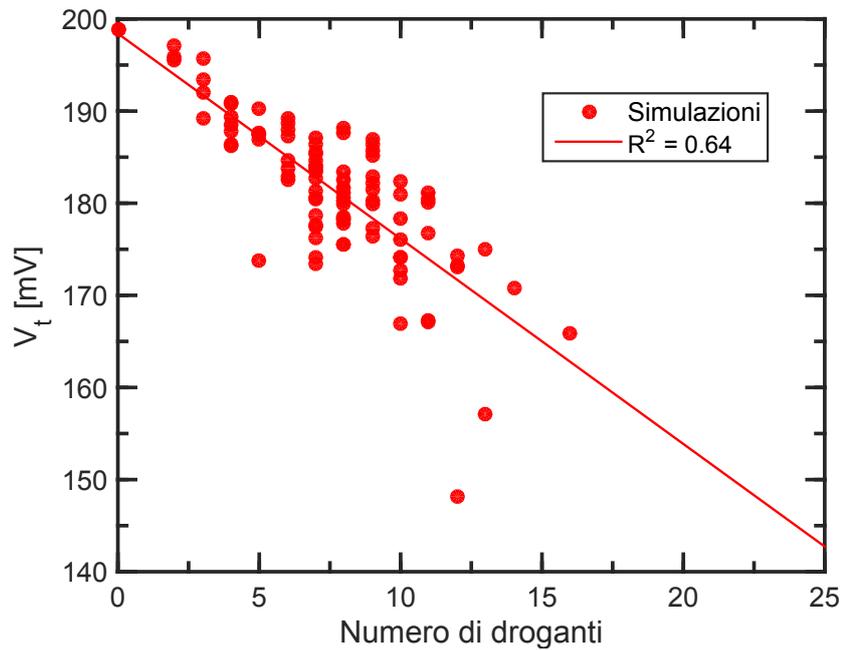
definita come due volte l'integrale della distribuzione gaussiana con media 0 e varianza  $\frac{1}{2}$ . Questa trasformazione facilita la verifica della normalità di una distribuzione poiché, nel caso la variabile  $x$  abbia distribuzione di probabilità gaussiana, sul grafico questa sarà

rappresentata da una retta. Più la pendenza della retta è alta, minore è la dispersione dei valori e viceversa. Quello che si può notare dal grafico è la sostanziale indipendenza di  $\sigma_{V_t}$  dalle tre  $L_d$  considerate, dato che le gaussiane hanno circa la stessa dispersione. Gli ioni presenti nelle zone molto esterne al gate del transistor hanno dunque poca influenza su  $V_t$ . Nella parte alta del grafico si può osservare che le distribuzioni  $L_d = 2$  nm e  $L_d = 3$  nm raggiungono un valore asintotico. Tale valore corrisponde alla configurazione in cui non ci sono droganti nella zona di discretizzazione. La tensione di soglia del transistor in questo caso è la massima possibile, poiché, come è stato detto prima, i donori discreti agevolano la conduzione abbassando la barriera di potenziale e quindi diminuiscono  $V_t$ . La distribuzione  $L_d = 4$  nm invece non raggiunge il suo valore asintotico poiché il numero medio di atomi in un profilo è tale da rendere molto inferiore ad 1 profilo su 100 la probabilità di non avere alcun donore.

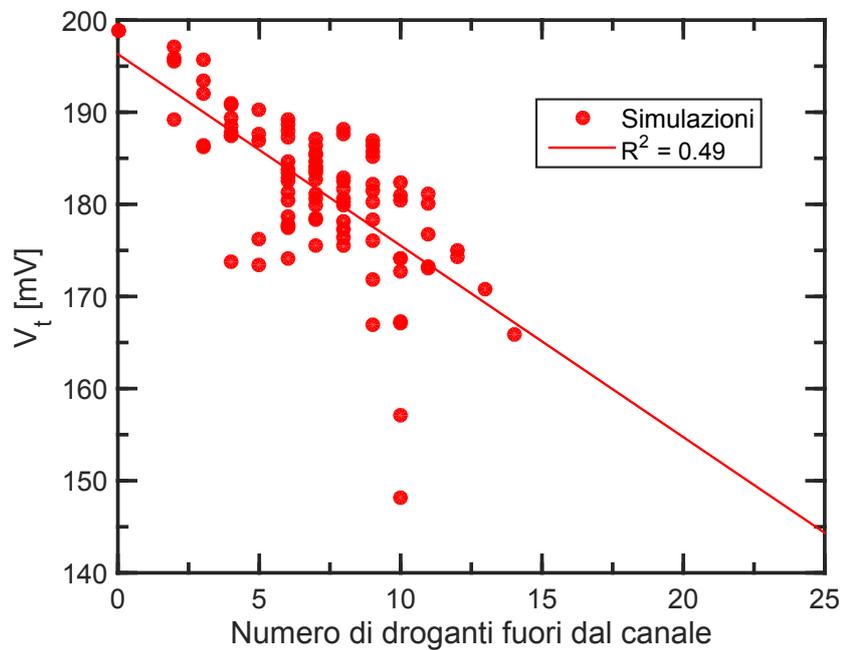
Dato che la scelta di una delle lunghezze di discretizzazione tra quelle viste non influenza la dispersione di  $V_t$ , come  $L_d$  per le analisi di RDF e RTN si è scelta quella intermedia di 3 nm. In questo caso risulta  $\sigma_{V_t} = 8.1$  mV. Il valore trovato è molto inferiore rispetto a quanto trovato per il nanowire con il canale drogato. La minore presenza di ioni sotto il gate, dovuta al profilo gaussiano di drogaggio, è quindi efficace nel ridurre la variabilità atomistica.

Poiché in questo tipo di nanowire le proprietà macroscopiche del transistor vengono determinate dalla posizione di pochi atomi, è necessario studiare approfonditamente come la tensione di soglia dipenda da questi. Per prima cosa si è verificata la dipendenza di  $V_t$  dal numero di droganti discreti. In Figura 4.3 è riportato un grafico in cui sono rappresentati i valori di  $V_t$  dei 100 nanowire con  $R = 4$  nm e  $L_d = 3$  nm simulati in precedenza, in funzione del numero di donori presenti nei corrispondenti profili di drogaggio. Dal grafico sembra esserci una correlazione tra la tensione di soglia e il numero totale di droganti nel transistor. L'andamento è decrescente poiché, come osservato prima, i donori abbassano la barriera di potenziale facilitando il passaggio degli elettroni dal source al drain e dunque una maggior presenza di questi porterà ad una ulteriore diminuzione di  $V_t$ . Per comprendere meglio questa dipendenza è stato fatto un fit lineare che è riportato sul grafico insieme al valore del coefficiente di correlazione  $R^2$ . Come si osserva, la correlazione tra il valore di  $V_t$  e il numero di droganti totali non è completa. Infatti vi è una certa dispersione tra i valori di  $V_t$  relativi a un numero fissato di atomi, e si verificano casi in cui un profilo con un numero maggiore di atomi rispetto ad un altro ha lo stesso una  $V_t$  più alta. Con l'intento di approfondire la ragione di questa variabilità è stata analizzata la dipendenza dei valori di  $V_t$  dal numero di atomi discreti presenti fuori dal gate e sotto il gate.

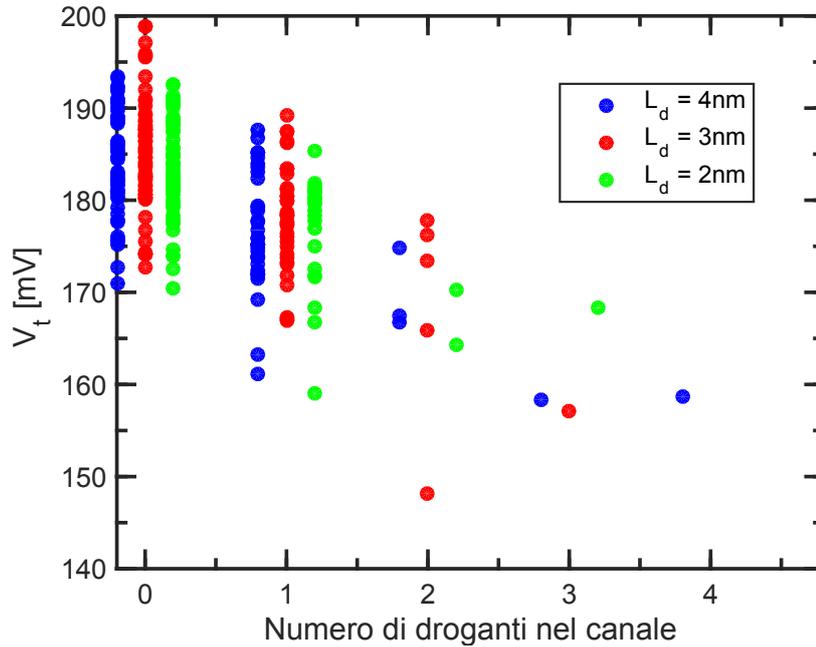
Da quello che è stato detto in precedenza, ci si aspetta che gli ioni fuori dal gate abbiano



**Figura 4.3:** Rappresentazione della  $V_t$  estratta dai profili simulati, in funzione del numero totale di donori discreti presenti in questi. È riportato anche un fit lineare dell'andamento per una stima della correlazione tramite il coefficiente  $R^2$ .



**Figura 4.4:** Andamento della tensione di soglia in funzione del numero di donori discreti presenti fuori dal gate del profilo simulato corrispondente. Il coefficiente  $R^2$  per la stima della correlazione è minore che in Figura 4.3.

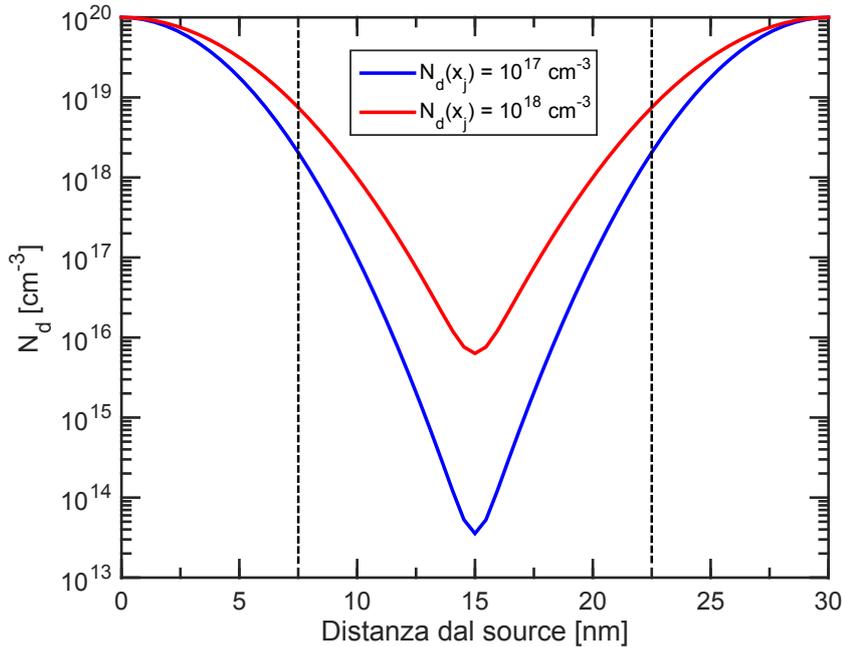


**Figura 4.5:** Grafico di  $V_t$  in funzione del numero di donori presenti nel canale. Sono riportati anche i valori ottenuti per i profili con  $L_d = 4$  nm e  $L_d = 2$  nm per verificare che la dispersione di  $V_t$  non sia dovuta al maggiore numero di atomi nella zona di discretizzazione.

un'influenza limitata sulla dispersione di  $V_t$ . In Figura 4.4 è mostrato l'andamento dei valori di  $V_t$  in funzione del numero di donori che occupano una posizione esterna al canale del transistor. In effetti la correlazione è minore rispetto al caso di Figura 4.3 in cui  $V_t$  è rappresentata in relazione al numero totale di droganti.

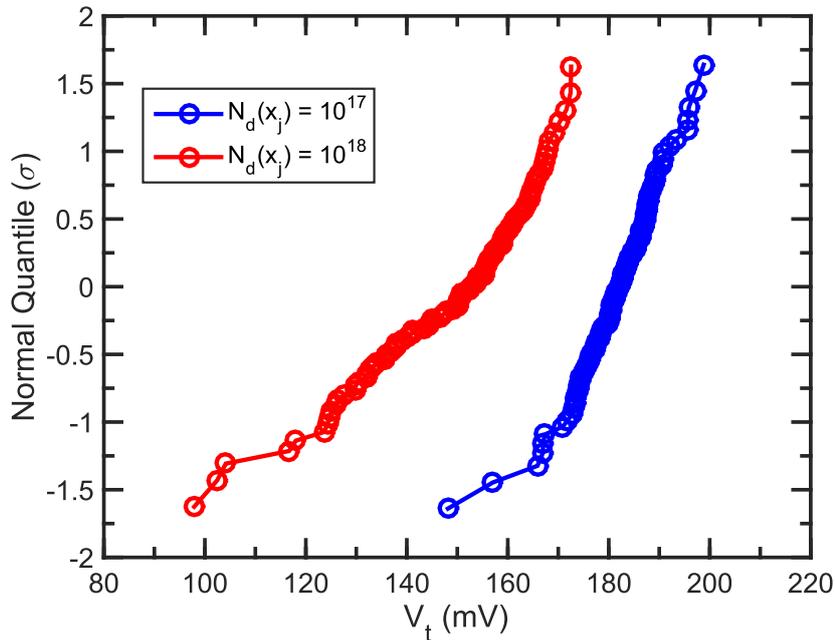
Se si considera invece l'andamento di  $V_t$  con il numero di donori presenti nel canale, ci si attende che la correlazione aumenti poiché la barriera di potenziale per gli elettroni che sottosoglia passano dal source al drain è più alta sotto il gate. In Figura 4.5 è rappresentato (in rosso) lo stesso tipo di grafico di Figura 4.3 e Figura 4.4, questa volta con il numero di donori sotto il gate sull'asse delle ascisse. Si nota come in realtà la correlazione non aumenti. La  $V_t$  infatti è molto dispersa per un determinato numero di donori sotto il gate. Inoltre si osserva una dispersione di  $V_t$  anche per il caso in cui non ci sia nessun atomo nel canale, segno che anche i droganti fuori dal gate hanno un certo peso sulle fluttuazioni di  $V_t$ . Sullo stesso grafico sono stati riportati gli andamenti delle altre due lunghezze di discretizzazione, per avere una conferma che la variabilità evidenziata per la  $V_t$  non risentisse anche dell'influenza degli ioni più lontani dal gate. La dispersione presenta la stessa ampiezza per tutte le tre  $L_d$  e dunque si può concludere che tra tutti gli atomi esterni sono soltanto quelli prossimi al gate ad esercitare un'influenza sulle fluttuazioni di  $V_t$ .

Un motivo per cui la correlazione di  $V_t$  con il numero di atomi sotto il gate è debole po-



**Figura 4.6:** Differenti profili di drogaggio implementati nello studio dell'RDF per il nanowire con canale non drogato. Sull'asse  $x$  è indicata la posizione lungo l'asse passante per il centro del corpo cilindrico, partendo dal source del transistor. Le due linee a tratti indicano rispettivamente l'inizio e la fine del canale.

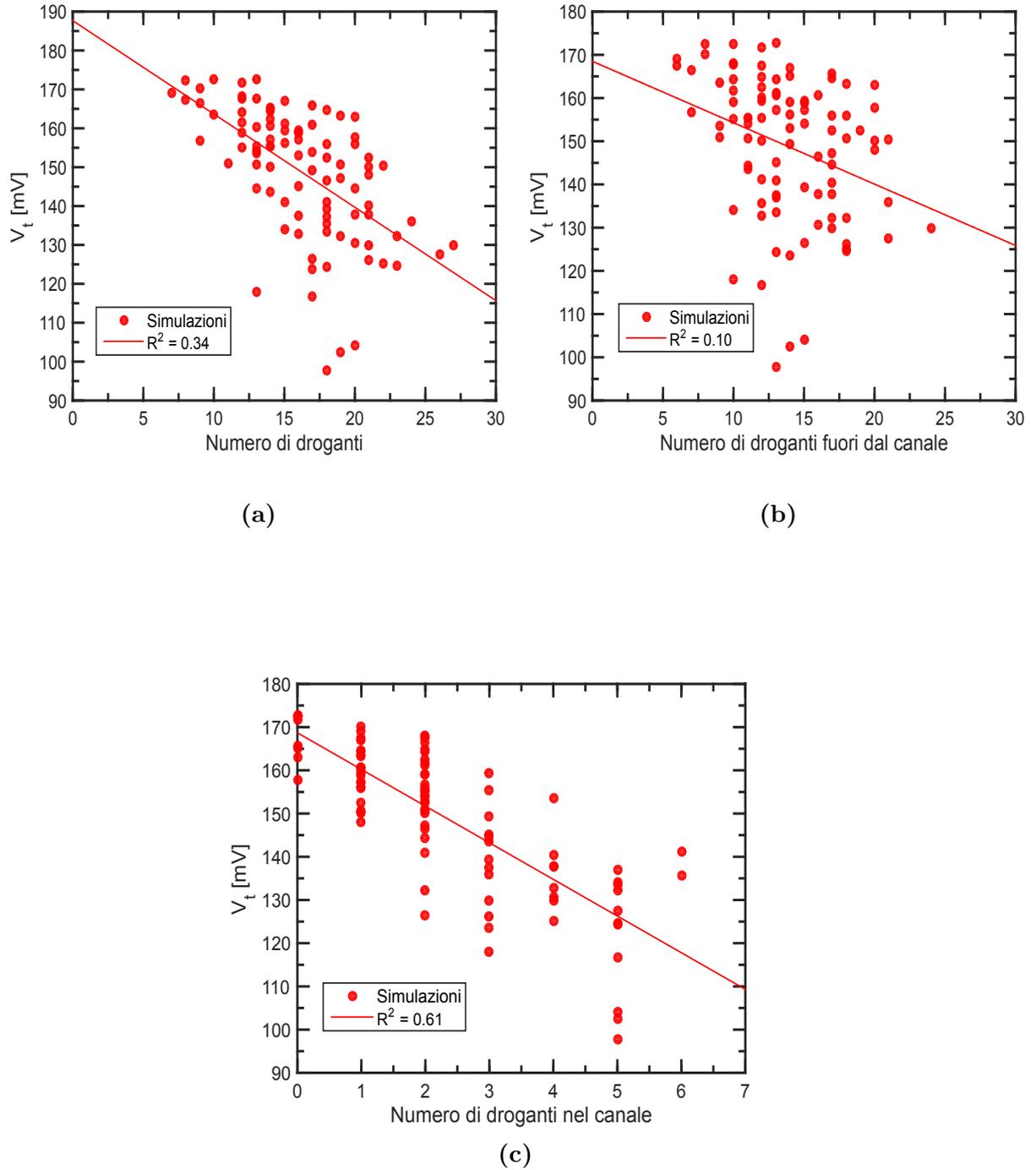
trebbe essere che il numero di questi sia in effetti troppo piccolo e la statistica sui punti a più atomi è insufficiente per trarre delle conclusioni. Al fine di ampliare il campo dell'analisi è stato introdotto un diverso profilo di drogaggio di source e drain. La concentrazione di picco è rimasta invariata,  $N_d = 10^{20} \text{ cm}^{-3}$ , mentre sono state estese le code variando la concentrazione alla profondità  $x_j = 10 \text{ nm}$ , calcolata a partire dall'inizio di source e drain andando verso il centro del canale. Per il profilo precedente  $N_d(x_j) = 10^{17} \text{ cm}^{-3}$ , mentre per quello nuovo  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$ . In Figura 4.6 è rappresentata la concentrazione di donori nei due profili in funzione della posizione lungo l'asse del nanowire a partire dal source. Come si vede, nel nuovo profilo (in rosso) aumenta il numero medio di donori sotto il gate. Se ci sono più atomi nel canale il valore medio di  $V_t$  si abbasserà poiché c'è un numero maggiore di zone in cui la corrente vede una barriera che impedisce il suo passaggio più bassa. Inoltre la dispersione dei valori della tensione di soglia dovrà aumentare poiché le possibili configurazioni per gli atomi discreti sono maggiori, esattamente come succede aumentando la concentrazione di accettori per i dispositivi con canale drogato. In Figura 4.7 è riportato il confronto tra la cumulativa di  $V_t$  del profilo  $N_d(x_j) = 10^{17} \text{ cm}^{-3}$  (in blu) e quella del profilo  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$  (in rosso). La gaussiana delle tensioni di soglia del profilo  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$  ha l'andamento atteso, con una  $\sigma_{V_t}$  di 16.7 mV, il doppio rispetto al profilo precedente.



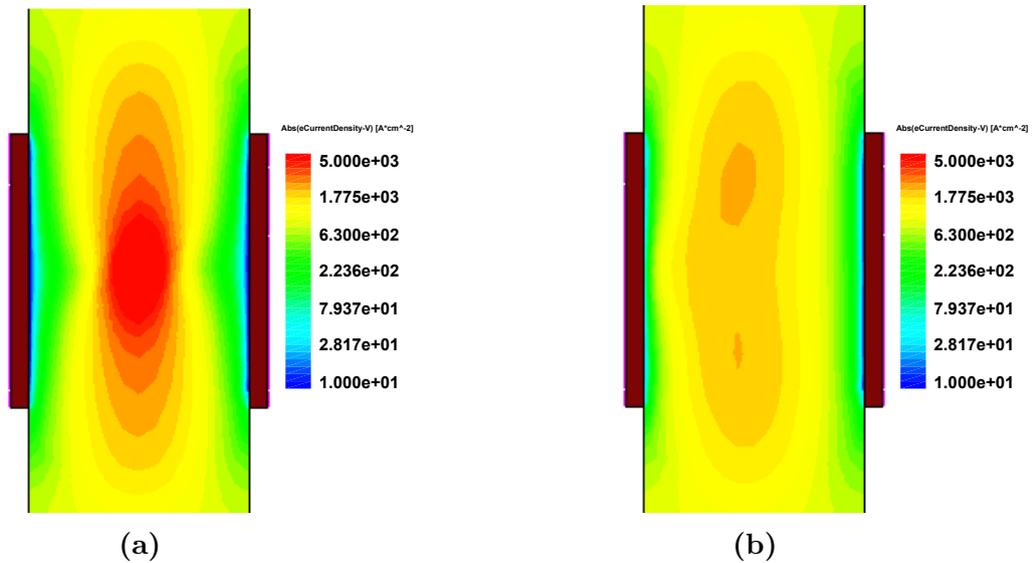
**Figura 4.7:** Distribuzioni cumulative di  $V_t$  estratte da un campione di 100 nanowire atomistici per ciascuno dei due profili di drogaggio di Figura 4.6. Il profilo che ha la maggiore concentrazione di donori sotto il gate presenta un valore medio di  $V_t$  più basso e una dispersione maggiore della stessa.

Per i transistor simulati con il nuovo profilo sono stati ripetuti gli stessi test di correlazione eseguiti in precedenza. In Figura 4.8 sono mostrati gli andamenti trovati. In Figura 4.8a si osserva che in questo caso la correlazione tra  $V_t$  e il numero totale di atomi sotto il gate diventa molto più debole che per il profilo  $N_d(x_j) = 10^{17} \text{ cm}^{-3}$  (Figura 4.3). Se guardiamo la dipendenza della tensione di soglia dal numero di donori fuori dal canale in Figura 4.8b, si nota come  $V_t$  sia diventata praticamente indipendente da questi. Invece, la correlazione tra la tensione di soglia e quanti ioni sotto il gate possiede un certo profilo migliora e l'andamento inversamente proporzionale comincia ad essere evidente (Figura 4.8c). Dunque, con l'aumento dei donori presenti nel canale dovuto all'estensione delle code gaussiane di source e drain, l'effetto di questi prevale nel determinare la tensione di soglia del dispositivo. L'elevata dispersione di  $V_t$  in funzione di quanti donori si hanno fuori dal gate è indice che il loro peso è diventato trascurabile. Tuttavia la correlazione di  $V_t$  con i droganti sotto il gate continua a non essere completa, poiché  $V_t$  varia di qualche decina di mV per un numero fissato di atomi. Si può notare che è sempre presente una fluttuazione anche per il caso di 0 atomi sotto il gate, ma è minore in confronto alla dispersione dei casi con degli atomi nel canale.

Come si è visto,  $V_t$  non dipende soltanto dal numero di droganti sotto il gate. Le altre variabili che potrebbero influire su  $V_t$  sono la posizione radiale ( $r$ ) a cui si trova un donore



**Figura 4.8:** Rappresentazione dei valori di  $V_t$  estratti dalle simulazioni di 100 transistor atomistici con il profilo di drogaggio  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$ , in funzione di diversi parametri. (a)  $V_t$  in funzione del numero totale di droganti nel transistor. (b)  $V_t$  in funzione del numero di droganti fuori dal gate. (c)  $V_t$  in funzione del numero di droganti sotto il gate. Per ogni grafico sono riportati il fit lineare e il relativo valore di  $R^2$ .



**Figura 4.9:** Valore assoluto della densità di corrente sottosoglia a  $V_g$  costante su una sezione lungo l'asse del nanowire per due profili con posizionamento deterministico di un drogante. (a) Donore a  $z = 0$  nm e a  $r = 0$  nm. (b) Donore a  $z = 0$  nm e a  $r = 3.9$  nm. La differenza di  $V_t$  tra i due casi è di 35 mV.

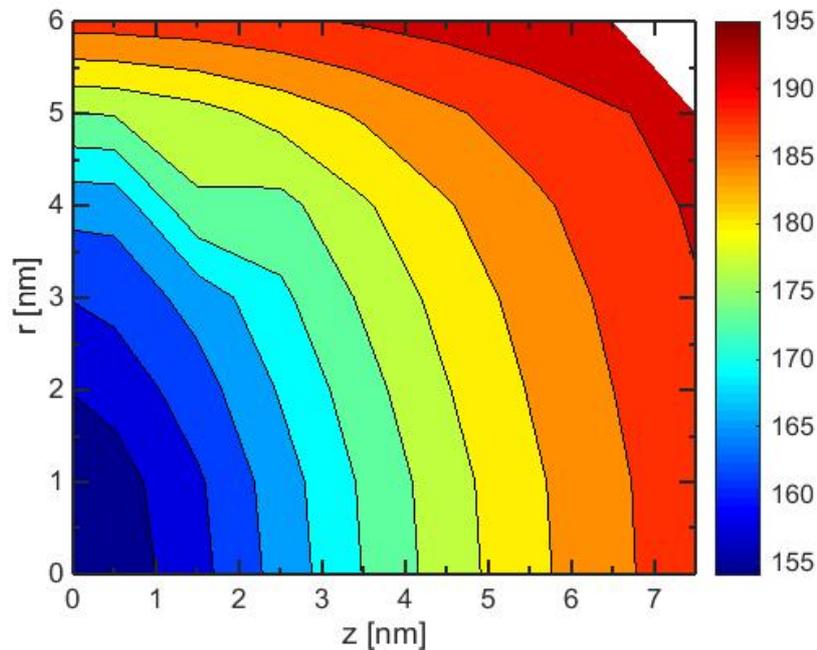
e la sua distanza dal centro del canale lungo l'asse del nanowire ( $z$ ).

Maggiore è la profondità a cui è situato un donore verso il centro del canale, più questo è vicino al massimo della barriera di potenziale che vedono gli elettroni dal source al drain. Il suo effetto sulla diminuzione di  $V_t$  sarà quindi più forte al diminuire di  $z$ .

Come è stato mostrato in sezione 2.2.3, a causa del confinamento quantistico la concentrazione di elettroni sottosoglia è nulla vicino all'interfaccia Si-SiO<sub>2</sub> e assume il suo massimo nella zona centrale del volume di silicio. Un donore che si trovi vicino all'interfaccia abbasserà la barriera di potenziale in un'area dove la concentrazione di elettroni è molto piccola. Viceversa un donore che si trova al centro del volume di silicio abbassa la barriera di potenziale nel punto in cui c'è una gran quantità di elettroni, favorendone il passaggio da source a drain e riducendo in modo maggiore la tensione di soglia. Un esempio di tale dipendenza da  $r$  è mostrato in Figura 4.9. In questo caso è messo a confronto il profilo della densità di corrente a  $V_g$  fissata su una sezione parallela al canale per due profili atomistici, scelti in modo deterministico, in cui è presente un donore a  $z = 0$  nm e  $r = 0$  nm (Figura 4.9a) nel primo caso e a  $z = 0$  nm e  $r = 3.9$  nm (Figura 4.9b) nel secondo. Come si osserva, se l'atomo è in posizione  $r = 0$  nm vi è un filamento passante per il centro del volume in cui si concentra un'alta densità di corrente. Questo abbassa di molto la  $V_t$  del transistor. Se si considera invece l'altro profilo, si nota come vi sia della corrente nell'area vicino al gate sul lato sinistro, mentre sul lato opposto questa è molto minore. Nella zona del canale la densità di corrente è circa uniforme e in questo

caso il donore influenza meno la tensione di soglia. La differenza di  $V_t$  tra i due profili è di 35 mV.

Per indagare quale sia il rispettivo peso delle dipendenze di  $V_t$  da  $r$  e da  $z$  si è pensato di procedere allo stesso modo di quanto appena fatto per verificare il diverso impatto di  $r$ , anche per la profondità  $z$ . Sono stati simulati dei profili che avessero un solo atomo donore, posizionato in modo deterministico a  $r$  e  $z$  tali da creare delle coppie di valori che esplorassero tutta la zona del canale. Le  $V_t$  estratte sono mostrate in Figura 4.10, dove sono rappresentate tramite superfici di livello in funzione di  $r$  e  $z$ . Dal grafico si può notare



**Figura 4.10:** Superfici di livello dei valori di  $V_t$  estratti da profili deterministici in cui è stata fatta variare la posizione radiale e la profondità all'interno del canale dell'unico donore atomistico presente.

che le dipendenze di  $V_t$  corrispondono a quelle attese.  $V_t$  aumenta allontanando lo ione dal centro del canale e diminuisce se la coordinata radiale si riduce. Come si può vedere, non c'è una dipendenza prevalente da uno dei due parametri. L'unica differenza che si rileva è che il peso delle diverse  $z$  è circa lo stesso su  $V_t$ , mentre  $r$  piccoli sono più determinanti nel variare la tensione di soglia rispetto a  $r$  grandi. Per posizioni  $z$  e  $r$  lontane dal centro del canale viene quasi raggiunto il valore asintotico di Figura 4.2, poiché il donore è in una posizione tale che il suo apporto alla conduzione diventa praticamente nullo. Invece non ci sono valori di  $V_t$  coincidenti con i valori più bassi delle distribuzioni di Figura 4.7,

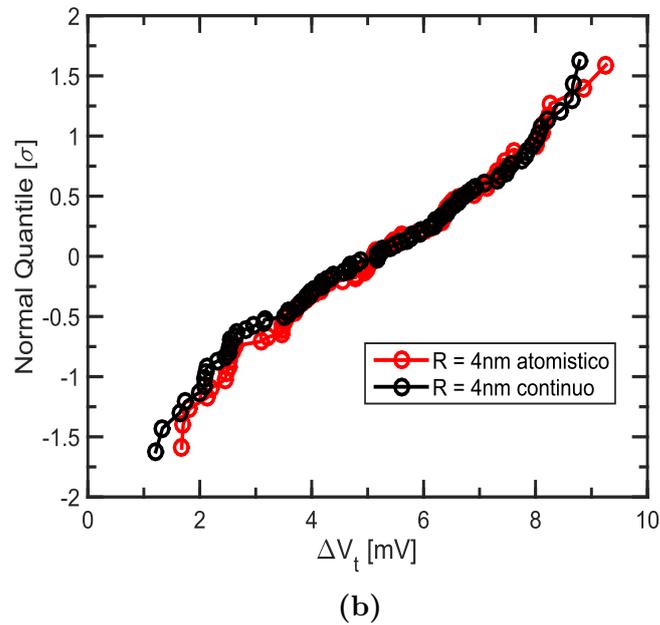
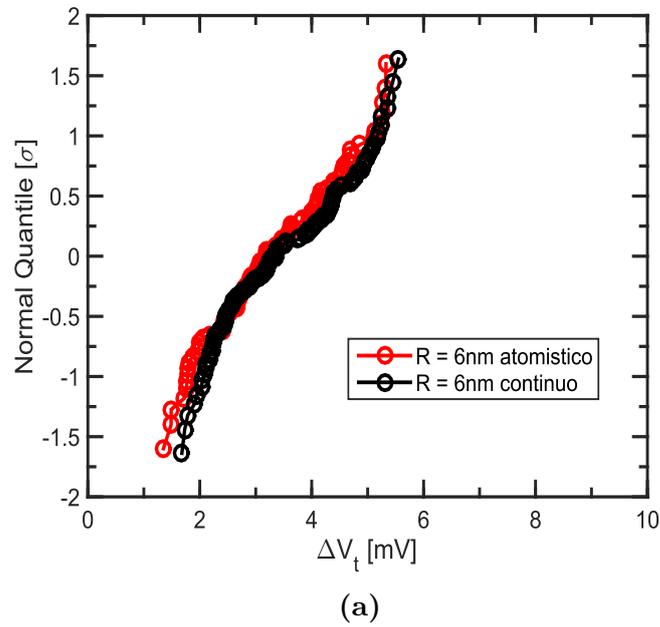
a conferma del fatto che l'aumento del numero di atomi sotto il gate, in particolare per il caso  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$ , assume maggiore importanza nella determinazione di  $V_t$ .

L'analisi di RDF ha evidenziato i tre parametri principali che influenzano  $V_t$  in un nanowire a canale non drogato, ovvero il numero di donori di source e drain sotto il gate, la profondità a cui questi sono situati lungo l'asse del canale e la posizione radiale a cui si trovano. Non vi è una dipendenza fondamentale da una di queste variabili, ma concorrono tutte e tre alla dispersione della tensione di soglia. Le fluttuazioni di  $V_t$  sono minori rispetto al caso di un nanowire a canale drogato, pur essendo ancora presenti a causa del drogaggio di source e drain. Un modo per ridurre ulteriormente la dispersione della tensione di soglia in simili dispositivi è ottenere un controllo più preciso sulla diffusione dei droganti in seguito all'impiantazione.

## 4.2 Analisi delle code RTN

In sezione 4.1 si è visto che l'effetto degli atomi donori è quello di abbassare la barriera di potenziale, e quindi la corrente si concentra nelle zone vicine alla posizione occupata dai droganti. I percorsi percolativi allora saranno attesi nei punti in cui si trovano dei donori e se il numero di questi aumenta i filamenti di corrente si intensificheranno, esattamente come avviene al crescere del numero di accettori nei dispositivi a canale drogato. In tale situazione, una trappola all'interfaccia ridurrà l'intensità dei percorsi che si trovano nelle sue vicinanze, ma lascerà inalterati quelli all'interno del volume. Da queste considerazioni segue che gli ioni che avranno un effetto sulle ampiezze di RTN saranno quelli più vicini al gate. Come si è visto in sezione 4.1, il profilo gaussiano  $N_d(x_j) = 10^{17} \text{ cm}^{-3}$  di source e drain possiede pochi ioni nel canale. Allora per lo studio dell'RTN si è deciso di utilizzare l'altro profilo, che ha delle code più estese. È stata aggiunta una trappola con probabilità uniforme all'interfaccia dei 100 profili simulati per l'RDF ed è stata estratta la nuova  $V_t$ . La distribuzione cumulativa dei  $\Delta V_t$  è visibile in Figura 4.11a (curva rossa). L'andamento è esattamente quello di una gaussiana, senza che vi siano code esponenziali. Per fare un confronto sono stati simulati anche 100 transistor con profilo di drogaggio continuo e trappola casuale all'interfaccia, che sono riportati sullo stesso grafico (in nero). Le due distribuzioni sono quasi coincidenti. Questo porta a concludere che in realtà non vi sono percorsi percolativi abbastanza intensi nei 100 transistor simulati.

Seguendo quanto trovato nel capitolo 3 per i dispositivi a canale drogato, abbiamo provato a diminuire il raggio del nanowire in modo da ridurre il numero dei percorsi percolativi che si possono formare e dunque aumentare l'intensità di quelli presenti. La distribuzione di 100 nanowire atomistici con  $R = 4 \text{ nm}$  è mostrata in Figura 4.11b (in rosso). Come prima viene riportata anche la cumulativa di 100 nanowire con profilo



**Figura 4.11:** Distribuzione cumulativa dei  $\Delta V_t$  simulati su un campione di 100 transistor atomistici con il profilo di drogaggio  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$  per source e drain (in rosso). Viene anche riportata la cumulativa ottenuta dalle simulazioni di 100 trappole Monte Carlo sullo stesso profilo, ma con doping continuo (in nero). (a) Nanowire con  $R = 6$  nm. (b) Nanowire con  $R = 4$  nm. Le distribuzioni atomistiche coincidono praticamente con quelle continue, dunque l'eventuale presenza di percorsi percolativi è debole.

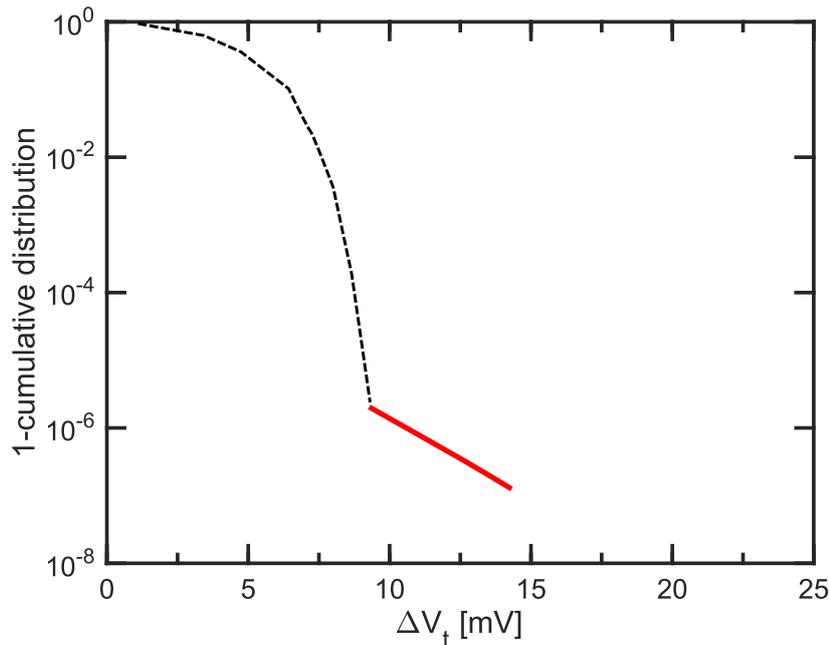
di doping continuo e trappola casuale (in nero). Di nuovo, la distribuzione atomistica coincide in pratica con quella continua. La dispersione dei valori di  $\Delta V_t$  è maggiore del caso  $R = 6$  nm poiché la capacità dell'ossido è minore per la struttura con  $R = 4$  nm.

L'assenza di un forte effetto percolativo causato dai donori discreti di source e drain potrebbe avere diverse ragioni. In primo luogo potrebbero essere ancora troppo pochi gli ioni presenti sotto il gate per essere sufficienti a intensificare la concentrazione della corrente nei percorsi. Un altro fattore potrebbe essere la profondità non elevata a cui giungono i donori che diffondono all'interno del canale. Infatti, come si era visto in sezione 1.4.2, l'effetto percolativo è più forte verso il centro del canale e minore verso le zone di source e drain, dunque l'effetto della trappola su filamenti in queste zone potrebbe essere ridotto. Infine, c'è la possibilità che il diverso apporto che i donori hanno sulla percolazione rispetto agli accettori porti ad una minore intensità dei filamenti che si creano.

Si è interessati a conoscere se esista effettivamente una coda RTN e quale sia il suo andamento poiché, pur avendo una bassa probabilità di accadimento, potrebbe essere determinante per la variabilità in un chip, dove i transistor sono milioni o miliardi. Il problema che si pone è generale, ovvero riuscire ad osservare gli eventi rari di una distribuzione, che sono di solito qualche parte per milione (ppm). L'approccio Monte Carlo in questi casi non è praticabile, poiché la simulazione di un campione così grande di dispositivi richiederebbe un tempo irragionevole. Se però il metodo Monte Carlo potesse essere focalizzato soltanto sulla generazione degli eventi rari della distribuzione, questi verrebbero simulati direttamente, rendendo possibile l'osservazione della coda. La massima attenzione deve essere prestata affinché sia mantenuta la statistica degli eventi rari simulati, ovvero la probabilità che viene loro associata deve essere la stessa che si otterrebbe simulando l'intera distribuzione. Osservando che l'algoritmo di tipo Monte Carlo che genera il campione di per sé è veloce, ed è la simulazione ad essere lenta, per arrivare ai punti della coda si può procedere nel seguente modo:

1. Si studiano i parametri che determinano il fenomeno in esame, con l'intento di conoscere quale sia il confine tra la regione dell'andamento normale della distribuzione e la regione dei casi che invece appartengono alla coda.
2. Vengono generati i punti del campione con metodo Monte Carlo.
3. In base all'analisi svolta al punto 1, gli eventi generati al punto 2 vengono distinti tra quelli che si prevede appartengano alla categoria degli eventi comuni e quelli che invece sono ritenuti potenzialmente rari.
4. Questi ultimi vengono simulati.

Il tempo necessario alla simulazione di una distribuzione costituita da un numero molto grande di eventi viene artificialmente ridotto a valori accessibili, poiché i punti 1, 2 e 3 sono veloci, mentre il punto 4 è più lento, ma si mantiene entro i tempi comuni delle simulazioni. In Figura 4.12 è rappresentato schematicamente il risultato che si ottiene con il procedimento che è stato esaminato. L'effetto è equivalente ad un aumento della



**Figura 4.12:** Rappresentazione schematica dell'aumento artificiale della statistica di una distribuzione. Tramite un algoritmo vengono simulati soltanto gli eventi rari per il fenomeno in esame, in modo da poter osservare la coda della cumulativa, che in genere si trova a probabilità di qualche ppm. La parte degli eventi comuni, che viene tenuta in conto correttamente pur non essendo simulata, è rappresentata in linea a tratti.

statistica della distribuzione, che viene tenuta in conto correttamente senza essere simulata (linea a tratti). I punti che invece vengono simulati sono soltanto quelli appartenenti alla coda (in rosso), che si trovano a probabilità molto piccole.

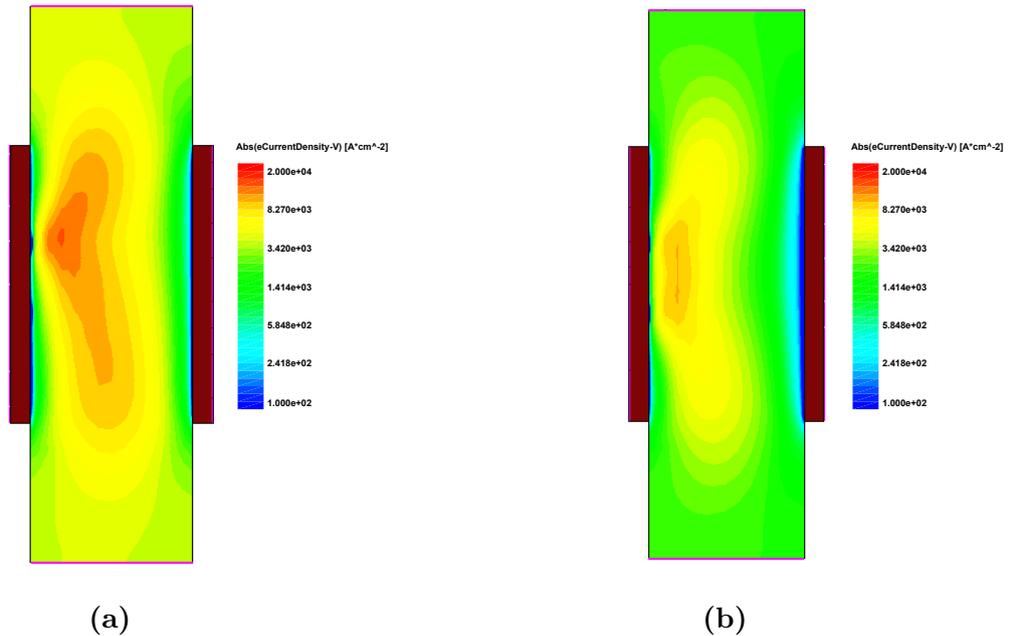
Il metodo di aumento statistico che è stato proposto può essere utilizzato per investigare la coda della distribuzione cumulativa del nanowire con il canale non drogato avente  $R = 4$  nm e profilo di drogaggio  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$  per source e drain. Infatti il campione contiene un numero insufficiente di transistor per poterne trovare alcuni che generino dei  $\Delta V_t$  dovuti a Giant RTN. Seguendo il punto 1 del procedimento appena mostrato, tramite prove deterministiche sono stati individuati dei criteri secondo cui una certa configurazione dei droganti del profilo, combinata con una determinata posizione della trappola, possa generare un valore di  $\Delta V_t$  più alto di quelli presenti nella distribuzione di Figura 4.11b. La scelta più ovvia è quella di considerare un percorso percolativo

che passi nelle vicinanze della trappola. Per simulare la corretta probabilità dei punti sulla coda della distribuzione è necessario che il confine tra la regione degli eventi comuni e quelli rari, per quanto riguarda i  $\Delta V_t$  di RTN generati dai profili di drogaggio casuali che vengono estratti dal metodo Monte Carlo, sia identificato con buona precisione. A tale scopo si è scelto un valore di  $\Delta V_t$  tra quelli anomali che facesse da limite inferiore per il criterio di selezione. In questo modo si è sicuri che tutti i valori di  $\Delta V_t$  maggiori di questo apparterranno alla coda della distribuzione.

Se consideriamo il caso di un solo donore sotto il gate, questo avrà un effetto sulla corrente sottosoglia tanto maggiore quanto più la sua profondità si avvicina al centro del canale poiché, come già accennato, la barriera ha il suo punto di massimo in questa zona. Perché una trappola all'interfaccia ossido-silicio riduca la concentrazione di corrente provocata dalla presenza dello ione, e dunque provochi un  $\Delta V_t$  elevato, è necessario che il percorso percolativo venga generato vicino ad essa. Quindi la posizione radiale dello ione deve essere scelta verso l'esterno del canale, in modo da far affluire la corrente in quest'area. La coordinata più vicina all'interfaccia è  $r = 3.9$  nm, a causa della discretizzazione del dominio. Simulando un profilo in cui il donore è posizionato in modo esatto a  $z = 0$  nm e  $r = 3.9$  nm, in assenza e in presenza della carica all'interfaccia, si ottiene un  $\Delta V_t$  di circa 20 mV. Tale valore è quello massimo ottenibile avendo un solo atomo sotto il gate, e abbiamo scelto di utilizzarlo come limite inferiore per la ricerca di profili atomistici che generino valori sulla coda della distribuzione.

Tramite prove con atomi posizionati deterministicamente si è riusciti ad individuare una serie di tolleranze sulle posizioni dei droganti e della trappola che consentono di trovare dei profili in grado di generare un  $\Delta V_t > 20$  mV. Un esempio è riportato in Figura 4.13 per il caso di due atomi nel canale. Gli ioni sono stati posizionati verso l'interfaccia con l'ossido, alla stessa coordinata radiale e anche in un intorno della stessa coordinata angolare ( $\varphi$ ), in modo da creare un percorso percolativo più intenso.

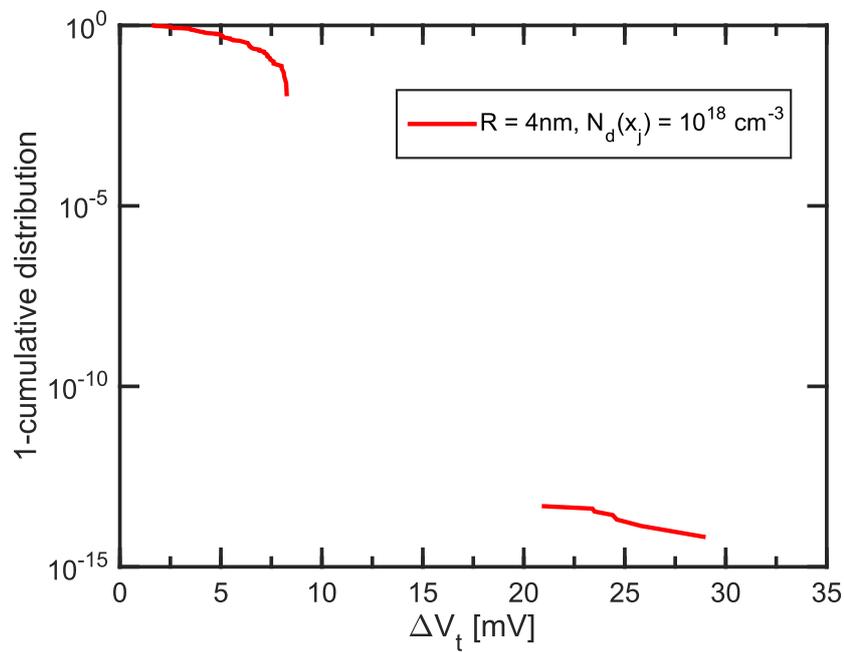
L'algoritmo utilizzato per trovare gli eventi rari della distribuzione genera in successione dei profili atomistici con metodo Monte Carlo. Per ognuno di questi le posizioni degli ioni sotto il gate e della trappola all'interfaccia vengono confrontate con le tolleranze imposte, derivate dalla precedente analisi. Per quanto riguarda la posizione  $z$  degli ioni si è cercato di prenderla il più lontano possibile dal centro del canale, entro i limiti necessari ad avere un  $\Delta V_t$  alto, poiché la probabilità di trovare un donore decresce esponenzialmente all'aumentare della sua distanza da source e drain. Essendo poi gli atomi dei profili che consideriamo in numero sempre maggiore di uno, la probabilità condizionata di avere  $n$  atomi alla posizione  $z$  che occupano sarà il prodotto delle rispettive probabilità, poiché gli eventi sono indipendenti. Inoltre la probabilità di avere più di un atomo sotto il gate dipende anche dal valore della cumulativa della distribuzione di Poisson,



**Figura 4.13:** Valore assoluto della densità di corrente sottosoglia a  $V_g$  costante su una sezione lungo l'asse del nanowire per un profilo atomistico in cui sono stati posizionati due donori sotto il gate rispettivamente a  $z = 13.5$  nm,  $r = 3.9$  nm e a  $z = 17.5$  nm,  $r = 3.9$  nm. (a) Profilo senza trappola. (b) Profilo con una trappola posizionata a  $z = 13.5$  nm,  $r = 4$  nm. Il valore di  $\Delta V_t$  in questo caso è  $> 20$  mV .

$P(n \geq 2) = \sum_{i=2}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!}$ . Le tolleranze su  $r$  e  $\varphi$ , rispettivamente  $\Delta r$  e  $\Delta\varphi$ , determinano una probabilità  $\frac{\Delta r}{R}$  e  $\frac{\Delta\varphi}{2\pi}$  di trovare uno ione di un profilo che soddisfi i requisiti.  $\Delta r$  impone che gli ioni siano vicini al gate, ma non è particolarmente stringente. Molto più restrittivo è invece  $\Delta\varphi$ , poiché gli ioni devono essere quasi allineati lungo la direzione parallela al canale per poter generare un valore di  $\Delta V_t$  elevato. Infine la trappola responsabile della fluttuazione RTN potrà trovarsi solamente all'interno di una certa area  $\Delta A_t$  nei pressi della posizione di uno degli ioni che soddisfano il requisito su  $z$ . La probabilità che una trappola casuale venga generata entro questo limite sarà  $\frac{\Delta A_t}{2\pi RL}$ . Considerando le tolleranze necessarie perché si creino dei percorsi percolativi in queste strutture, ci si attende che la coda si manifesti a probabilità molto basse. In Figura 4.14 sono riportati i  $\Delta V_t$  ottenuti simulando i profili selezionati dall'algoritmo. La pendenza della coda che si osserva è di 8 mV/decade, tuttavia le probabilità di evenienza sono talmente basse che la variabilità indotta da questa può essere tranquillamente trascurata.

Dall'analisi di RTN risulta che i nanowire a canale non drogato, in cui gli unici ioni presenti sotto il gate sono i donori provenienti dalle impiantazioni di source e drain, sono effettivamente efficaci nel ridurre le fluttuazioni  $\Delta V_t$ . Inoltre le ampiezze anomale di RTN sono praticamente assenti.



**Figura 4.14:** Distribuzione cumulativa dei  $\Delta V_t$  estratti dal campione di 100 nanowire con  $R = 4 \text{ nm}$  e profilo di drogaggio  $N_d(x_j) = 10^{18} \text{ cm}^{-3}$  (in alto) e distribuzione cumulativa della coda simulata con l'aumento statistico (in basso). La coda presenta una pendenza di 8 mV/decade, ma la sua probabilità di accadimento è del tutto trascurabile.

# Conclusioni

In questo lavoro di tesi è stata analizzata la variabilità atomistica nei dispositivi nanowire. Tramite simulazioni numeriche 3D di transistori in cui il profilo di drogaggio è reso atomistico estraendo le posizioni casuali degli ioni con il metodo Monte Carlo, sono state ricavate le distribuzioni di probabilità sia per il fenomeno di RDF che per il fenomeno di RTN. Il confronto dei risultati ottenuti con quelli già noti da studi dello stesso tipo effettuati sui transistor di tipo bulk ha messo in luce i diversi aspetti che caratterizzano la percolazione all'interno di queste strutture. In particolare è stato evidenziato come l'inversione di volume, assente nelle strutture planari, sia responsabile di due diversi effetti sulle fluttuazioni della tensione di soglia, che ne influenzano l'andamento in maniera opposta. È stato trovato che il rapporto tra il numero di atomi nel canale del dispositivo e il volume dello stesso determina quale sia il comportamento prevalente. Se il volume del nanowire è molto grande, oppure la concentrazione di accettori non è elevata, le possibilità percolative per la corrente sono maggiori che nel caso planare, e non si creano dei filamenti intensi. Per quanto riguarda l'RDF, questo porta ad avere una diminuzione della dispersione di  $V_t$  al crescere di  $R$  e al diminuire di  $N_a$ , più accentuata rispetto al caso planare. L'RTN presenta una saturazione della pendenza della coda  $\lambda$  al variare degli stessi parametri. Quando invece il numero di accettori sotto il gate supera un certo numero, che varia con le dimensioni del dispositivo, i possibili percorsi preferenziali della corrente diventano pochi, e dunque molto intensi, poiché con l'inversione di volume gli atomi più interni hanno lo stesso peso di quelli all'interfaccia nel determinare la percolazione. In questo caso sia  $\sigma_{V_t}$  sia  $\lambda$  presentano una dipendenza dalla diminuzione di  $R$  e dall'aumento di  $N_a$  che è più forte rispetto a quella trovata nei dispositivi di tipo bulk. Ulteriori studi sono necessari sia per aumentare la statistica dei dispositivi simulati, sia per esplorare nuovi valori dei parametri coinvolti nello scaling.

Successivamente è stato presentato un tipo alternativo di nanowire, con il canale non drogato, che potenzialmente è in grado di ridurre le fluttuazioni della tensione di soglia dovute ai droganti discreti. In tale dispositivo i donori che diffondono sotto il gate dai profili di source e drain diventano fondamentali nel determinare la variabilità atomistica. Nell'analisi di RDF sono state studiate le dipendenze della dispersione di  $V_t$  dalla profon-

dità sotto il gate che raggiunge lo ione, dalla coordinata radiale che questo occupa e infine dal numero totale di atomi presenti nel canale. La dipendenza dai donori discreti nelle zone del source e del drain ha mostrato di essere rilevante solamente per profili aventi delle code poco diffuse.  $\sigma_{V_t}$  risulta effettivamente minore rispetto a quanto trovato nei nanowire con il canale drogato. L'analisi statistica delle distribuzioni RTN ha mostrato l'assenza di percorsi percolativi nel campione simulato, anche riducendo le dimensioni del dispositivo. In seguito è stato ideato un algoritmo di aumento statistico artificiale che consentisse di osservare la coda della distribuzione.  $\lambda$ , per questo tipo di transistor, è a probabilità talmente basse che può essere ritenuta del tutto trascurabile.

# Bibliografia

- [1] “International Technology Roadmap for Semiconductors.” <http://www.itrs.net>.
- [2] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge University Press, second ed., 2008.
- [3] J. P. Colinge, “Multiple-gate SOI MOSFETs,” *Solid-State Electronics*, vol. 48, pp. 897 – 905, 2004.
- [4] Intel, “Tecnologia Intel a 14 nm.” <http://www.intel.it/content/silicon-innovations/intel-14nm-technology>, 2015.
- [5] S. H. Oh, D. Monroe, and J. M. Hergenrother, “Analytic description of short-channel effects in fully-depleted double-gate and cylindrical, surrounding-gate MOSFETs,” *Electron Device Letters*, vol. 21, pp. 445 – 447, 2000.
- [6] S. D. Suk, S.-Y. Lee, S.-M. Kim, and E.-J. Yoon, “High performance 5nm radius twin silicon nanowire MOSFET (TSNWFET) : fabrication on bulk si wafer, characteristics, and reliability,” *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, vol. 21, pp. 717 – 720, 2005.
- [7] J. Zhuge, R. Wang, R. Huang, and Y. Tian, “Investigation of low-frequency noise in silicon nanowire MOSFETs,” *Electron Device Letters, IEEE*, vol. 30, pp. 57 – 60, 2009.
- [8] F. Balestra, S. Cristoloveanu, M. Benachir, and J. Brini, “double-gate silicon-on-insulator transistor with volume inversion: a new device with greatly enhanced performance,” *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pp. 449 – 452, 1987.
- [9] Y. Taur, “An analytical solution to a double-gate MOSFET with undoped body,” *Electron Device Letters*, vol. 21, pp. 245 – 247, 2000.

- 
- [10] G. X. Hu and T. A. Tang, “Some physical properties of a surrounding-gate MOSFET with undoped body,” *Journal of the Korean Physical Society*, vol. 49, pp. 642 – 645, 2006.
- [11] A. Tsormpatzoglou, D. H. Tassis, C. A. Dimitriadis, G. Ghibaudo, G. Pananakakis, and R. Clerc, “A compact drain current model of short-channel cylindrical gate-all-around MOSFETs,” *Semiconductor Science and Technology*, vol. 24, pp. 1 – 7, 2009.
- [12] R. Shirota, “Developments in 3D-NAND Flash technology,” in *Advances in Non-volatile Memory and Storage Technology*, ch. 2, Eslevier, 2014.
- [13] R. Katsumata, M. Kito, Y. Fukuzumi, M. Kido, H. Tanaka, and Y. Komori, “Pipe-shaped BiCS Flash memory with 16 stacked layers and multi-level-cell operation for ultra high density storage devices,” *VLSI Technology, 2009 Symposium on*, pp. 136 – 137, 2009.
- [14] J. Jang, H.-S. Kim, W. Cho, H. Cho, J. Kim, and S. I. Shim, “Vertical cell array using TCAT (Terabit Cell Array Transistor) Technology) for ultra high density NAND Flash memory,” *VLSI Technology, 2009 Symposium on*, pp. 192 – 193, 2009.
- [15] Y. Fukuzumi, R. Katsumata, M. Kito, M. Kido, and H. Tanaka, “Optimal integration and characteristics of vertical array devices for ultra-high density, bit-cost scalable flash memory,” *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pp. 449 – 452, 2007.
- [16] A. Asenov, “Random dopant induced threshold voltage lowering and fluctuations in sub 50 nm MOSFETs: a statistical 3D ‘atomistic’ simulation study,” *Nanotechnology*, vol. 45, pp. 153 – 158, 1998.
- [17] A. Ghetti, M. Bonanomi, C. M. Compagnoni, A. Spinelli, A. Lacaita, and A. Visconti, “Physical modeling of single-trap RTS statistical distribution in Flash memories,” *Reliability Physics Symposium, 2008. IRPS 2008. IEEE International*, pp. 610 – 615, 2008.
- [18] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, “Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs,” *Electron Devices, IEEE Transactions on*, vol. 50, pp. 1837 – 1852, 2003.
- [19] P. A. Stolk, F. P. Widdershoven, and D. B. M. Klaassen, “Modeling statistical dopant fluctuations in MOS transistors,” *Electron Devices, IEEE Transactions on*, vol. 45, pp. 1960 – 1971, 1998.

- 
- [20] R. Gusmeroli, C. M. Compagnoni, A. Riva, A. S. Spinelli, A. L. Lacaita, M. Bonanomi, and A. Visconti, “Defects spectroscopy in SiO<sub>2</sub> by statistical random telegraph noise analysis,” *Electron Devices Meeting, 2006. IEDM 2006. International*, pp. 1 – 4, 2006.
- [21] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, “Random telegraph noise of deep-submicrometer MOSFETs,” *Electron Device Letters*, vol. 11, pp. 90 – 92, 1990.
- [22] H. H. Mueller and M. Schulz, “Conductance modulation of submicrometer metal-oxide-semiconductor field-effect transistors by single-electron trapping,” *Journal of Applied Physics*, vol. 79, pp. 4178 – 4186, 1996.
- [23] P. Fantini, A. Ghetti, A. Marinoni, G. Ghidini, A. Visconti, and A. Marmiroli, “Giant random telegraph signals in nanoscale floating-gate devices,” *Electron Device Letters*, vol. 28, pp. 1114 – 1116, 2007.
- [24] A. Ghetti, C. M. Compagnoni, F. Biancardi, A. L. Lacaita, and S. Beltrami, “Scaling trends for random telegraph noise in deca-nanometer Flash memories,” *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, pp. 1–4, 2008.
- [25] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, “RTS amplitudes in decanometer MOSFETs: 3-D simulation study,” *Electron Devices, IEEE Transactions on*, vol. 50, pp. 839 – 845, 2003.
- [26] C. M. Compagnoni, R. Gusmeroli, A. S. Spinelli, A. L. Lacaita, M. Bonanomi, and A. Visconti, “Statistical model for random telegraph noise in Flash memories,” *Electron Devices, IEEE Transactions on*, vol. 55, pp. 388 – 395, 2008.
- [27] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, “Comprehensive analysis of random telegraph noise instability and its scaling in deca-nanometer Flash memories,” *Electron Devices, IEEE Transactions on*, vol. 56, pp. 1746 – 1752, 2009.
- [28] A. Mauri, C. M. Compagnoni, S. M. Amoroso, A. Maconi, A. Ghetti, A. S. Spinelli, and A. L. Lacaita, “Comprehensive investigation of statistical effects in nitride memories Part I: physics-based modeling,” *Electron Devices, IEEE Transactions on*, vol. 57, pp. 2116 – 2123, 2010.
- [29] Synopsys, “Process and Device Simulation Tools, Sentaurus Device.” <http://www.synopsys.com/Tools/TCAD/DeviceSimulation/Pages/SentaurusDevice.aspx>.
- [30] M. G. Ancona, “Density-gradient theory: a macroscopic approach to quantum confinement and tunneling in semiconductor devices,” *Journal of Computational Electronics*, vol. 10, pp. 65 – 97, 2011.

- [31] A. S. Spinelli, C. M. Compagnoni, A. Maconi, S. M. Amoroso, and A. L. Lacaita, “Quantum-mechanical charge distribution in cylindrical gate-all-around MOS devices,” *Electron Devices, IEEE Transactions on*, vol. 59, pp. 1837 – 1843, 2012.
- [32] N. Sano, K. Matsusawa, M. Mukai, and N. Nakayama, “On discrete random dopant modeling in drift-diffusion simulations: physical meaning of ‘atomistic’ dopants,” *Microelectronics Reliability*, vol. 82, pp. 189 – 199, 2002.
- [33] G. Roy, A. R. Brown, F. Adamu-Lema, S. Roy, and A. Asenov, “Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-MOSFETs,” *Electron Devices, IEEE Transactions on*, vol. 53, pp. 3063 – 3070, 2006.