Alma Mater Studiorum Università di Bologna

Corso di Dottorato in

Tecnologie dell'Informazione

Tesi di Dottorato in

Modelling and characterization of decananometric electronic devices

Scritta da

Simone Eminente

Relatore

Prof. Enrico Sangiorgi

Coordinatore

Prof. Riccardo Rovatti

Settore Scientifico-Disciplinare: ING-INF/01 Elettronica

Anni Accademici: 2003/04 - 2004/05 - 2005/06

List of Symbols

Indices

avg	average value
ON	on bias condition, with scattering
BL	on bias condition, without scattering
$+$	Source-to-drain direction
$-$	Drain-to-source direction
G, D, S, B, SUB	Gate, drain, source, bulk, substrate
x	direction parallel to the Si–SiO_2 interface
inj	evaluated at the injection point
$1, 2$	gate stack 1 or 2 in multi-gate devices

Symbols

$A_{i_{short}}$	Current gain with short circuit as load
$A_{v_{open}}$	Voltage gain with open circuit as load
a_D	Carrier flux from drain to channel
a_S	Carrier flux from source to channel
C_{ox}	Gate oxide capacitance, fF
δ	Coefficient of SOI scaling function $L_G - t_{si}$
E	Carrier energy, eV
$\mathcal{E}$	Electric field, V/cm
$\mathcal{E}_{EFF}$	Effective field, V/cm
ϵ_{ox}	Oxide Dielectric permittivity, 3.45×10^{-13} F/cm
ϵ_{si}	Silicon Dielectric permittivity, 1.04×10^{-12} F/cm
$\mathcal{F}$	Boltzmann transport equation (BTE)
$\mathcal{F}[f]$ or $\widetilde{f}$	F-transform of function f
F_T	Transition frequency, Hz
F_{3dB}	Cutoff frequency of the voltage gain, Hz
f	Frequency, Hz
Γ	Scattering rate, s^{-1}
g_m	Transconductance, Ω^-1
$g_d s$	Drain conductance, Ω^-1
$\widetilde{\mathcal{H}}$	Fourier tranfer function
I	Current, A/μm
I_{OFF}	Device off-current, A/μm
J_n	Electron current density, A/cm^2
k	Boltzmann's constant, 8.617×10^{-5} JK^{-1}
κ	Scaling factor

λ	Mean free path, nm
L_G	Gate length, nm
L_{KT}	Length of the kT-layer, nm
L_{SCATT}	Decay length of backscattering contribution, nm
L_{SPAC}	Spacer length for S/D contact diffusion, nm
m^*	Carrier effective mass, kg
μ_{EFF}	Inversion layer effective mobility, $\mathrm{cm}^2/\mathrm{Vs}$
N_a and N_d	p-type and n-type Substrate doping, cm^{-3}
n and p	Electron and hole concentration, cm^{-3}
$\vec{p}$	Momentum vector
Q_d	Depletion charge, cm^{-2}
Q_i	Inversion charge concentration, cm^{-2}
q	Electron charge, 1.6×10^{-19} C
$\vec{r}$	Position vector
r or r_C	Backscattering coefficient
τ	Average time between two scattering events, s
τ_{DC}	DC channel transit time, s
τ_{DS}	Source-drain delay time, s
τ_{gm}	Transconductance delay time, s
T	Temperature, K
T^*_{SI}	Critical silicon fin thickness, nm
t	Time variable, s
t_{box}	Back gate oxide thickness, nm
t_C	Transmission coefficient from channel to drain
t_{ox}	Gate oxide thickness, nm
t_{SI}	Silicon fin thickness, nm
V_{DD}	Supply voltage, V
V_{TH} or V_T	Threshold voltage, V
v	Carrier velocity, cm/s
v_g	Carrier group velocity, cm/s
v_{inj}	Injection velocity, cm/s
v_{sat}	Inversion layer saturation velocity, cm/s
v_T	Thermal velocity, cm/s
W	Gate width, μm
x	Horizontal axis, parallel to the Si–SiO_2 interface
x_{inj}	Injection point
y	Vertical axis, normal to the Si–SiO_2 interface
Ψ	Electric potential, V
Ψ_F	Fermi potential, V
Ψ_S	Surface potential, V
ω	radian frequency, rad Hz

Contents

Part I

Introduction

Introduzione e Riassunto della Tesi

I circuiti integrati (ICs) sono stati tra i protagonisti dello sviluppo tecnologico degli ultimi decenni. Il loro successo è in particolare dovuto alla continua crescita della densità di integrazione che, come previsto dalla famosa legge di Moore [1], ha permesso di realizzare circuiti sempre più complessi. I dispositivi MOSFET moderni hanno una lunghezza di gate L_G inferiore ai 100 nanometri, e i processi produttivi più recenti realizzano transistori con lunghezze pari a 30-40 nm.
Negli ultimi anni la riduzione ("scaling") della lunghezza di gate ha portato alla luce una serie di problemi dovuti al fatto che le caratteristiche elettriche del MOSFET non scalano idealmente con L_G. Questi effetti di non-idealità obbligano i progettisti a cercare compromessi e nuove soluzioni per mantenere il progressivo miglioramento delle prestazioni nei nodi tecnologici successivi. Uno tra i problemi di scaling più studiati è legato alla corrente di drain, che cresce meno del previsto quando le dimensioni del transistor vengono ridotte. La causa principale risiede nella limitata mobilità elettronica dei portatori e nelle resistenze serie, il cui impatto diviene via via più rilevante con la progressiva riduzione della lunghezza di gate.
Possiamo distinguere due tipologie di soluzione ai problemi di non idealità: introduzione di materiali innovativi per realizzare il canale e le regioni di source e drain e introduzione di architetture innovative. Nella prima categoria troviamo, ad esempio, lo strain per aumentare la mobilità ed eterogiunzioni tra il canale e il source/drain. Nella seconda vi è l'introduzione della tecnologia SOI per sostituire l'attuale architettura di tipo Bulk.

In questa tesi abbiamo studiato alcune proprietà fisiche dei dispositivi MOSFET rappresentativi dei prossimi nodi tecnologici. Lo scopo finale è duplice. Da un lato vogliamo studiare lo scaling delle prestazioni dei MOSFET, e quindi i suoi limiti. Dall'altro, allo scopo di avere una più accurata modellizzazione dei dispositivi a canale corto, utilizziamo metodi di simulazione differenti rispetto a quelli tradizionali. In questo modo possiamo trattare, nel miglior modo possibile, gli effetti di non-idealità che più ci interessano. Ove possibile mostriamo i benefici di avere architetture di dispositivo non convenzionali, con un particolare riguardo ai vantaggi dei transistori di tipo SOI rispetto ai Bulk.
Di seguito sono brevemente descritti i principali argomenti trattati.

Analisi del moto quasi-ballistico

Nei dispositivi a canale ultra-corto, i portatori subiscono un numero

limitato di perturbazioni del moto (eventi di "scattering") e il concetto di mobilità diventa di difficile definizione. Inoltre il trasporto nel canale è influenzato da condizioni di forte scostamento rispetto all'equilibrio e questa caratteristica non è sempre adeguatamente considerata quando si simulano tali dispositivi. La simulazione Monte Carlo costituisce un tipo di approccio molto utile per studiare condizioni molto lontane dall'equilibrio. Per questo motivo un simulatore Monte Carlo è stato usato per studiare le proprietà del trasporto nei MOSFET per i futuri nodi tecnologici. Il simulatore utilizzato è auto-consistente e comprende i più accurati modelli di scattering ed una correzione quantistica per il calcolo della carica di inversione. I risultati mostrano che lo scattering ha ancora un effetto rilevante sui dispositivi fino a lunghezze di gate pari ad almeno 14 nm. Pertanto non possiamo ancora parlare di un trasporto completamente ballistico per i prossimi nodi tecnologici. Tuttavia abbiamo verificato che lo scaling delle dimensioni sta progressivamente avvicinando il trasporto al limite ballistico, in particolare negli SOI con body non drogato dove lo scattering da impurezze ionizzate e da rugosità all'interfaccia silicio-ossido é trascurabile.

Analisi delle prestazioni RF

Lo stesso simulatore Monte Carlo è stato adattato per includere l'analisi delle prestazioni in frequenza nei MOSFET. Anche i moderni MOSFET per applicazioni RF presentano un trasporto molto lontano dall'equilibrio, cosa che potrebbe diminuire l'accuratezza degli approcci drift-diffusion normalmente utilizzati per la simulazione AC. Abbiamo considerato diversi parametri di prestazioni, tra cui la transconduttanza g_m, la frequenza di transizione F_T e il guadagno di tensione a circuito aperto A_{Vopen}. I nostri risultati hanno dimostrato che le prestazioni dei dispositivi MOSFET di tipo Bulk ed SOI Single-Gate continuano a migliorare con lo scaling di L_G. I dispositivi SOI con substrato non drogato hanno un certo vantaggio grazie all'elevata mobilità elettronica che garantisce una migliore transconduttanza. Abbiamo anche confrontato la simulazione Monte Carlo con l'approccio drift-diffusion e abbiamo verificato la maggiore accuratezza del primo metodo nel descrivere le prestazioni RF previste per i dispositivi MOSFET dei prossimi nodi tecnologici.

Analisi dell'accoppiamento capacitivo negli SOI Multi-Gate

Nell'ultima parte della tesi sono riportati i risultati di una analisi delle proprietà dell'accoppiamento capacitivo ("coupling") tra i contatti di gate nei MOSFET SOI completamente svuotati (Fully-Depleted). In questi dis-

positivi le caratteristiche di un'interfaccia silicio-ossido sono collegate alle caratteristiche dell'interfaccia opposta attraverso il film sottile di silicio (che costituisce anche il body del transistor). Questa mutua interazione é molto utilizzata per scopi di caratterizzazione elettrica. Nei dispositivi avanzati essa devia dal normale comportamento ed evidenzia caratteristiche complesse. Abbiamo considerato due casi differenti. Nel primo caso abbiamo studiato le caratteristiche dell'accoppiamento capacitivo negli SOI a gate singolo con film di silicio più sottili di 10 nm: questi spessori saranno richiesti per L_G minore di 25 nm. Prove sperimentali hanno mostrato che l'accoppiamento appare rafforzato nei film ultra-sottili. A partire da questi risultati, abbiamo studiato, tramite simulazioni, le nuove proprietà del'accoppiamento in termini di potenziale elettrico e densità di carica interne al film di silicio. Nel secondo caso abbiamo analizzato, attraverso caratterizzazione elettrica, le caratteristiche dell'accoppiamento capacitivo nei MOSFET SOI non-planari con contatti multipli di gate. In questi dispositivi, la geometria tridimensionale e la presenza di più contatti di gate rende molto complesso il coupling tra le varie interfacce. Abbiamo visto che solo alcune di esse sono accoppiate tra loro in modo non trascurabile e che questa caratteristica dipende fortemente dalle dimensioni del dispositivo.

Résumé et présentation génerale de Mémoire

La vitesse d'intégration des circuits CMOS a été maintenue jusqu'à maintenant, en suivant la loi célèbre de Moore qui stipule que la densité des transistors dans une puce double chaque 24 mois [1]. Même si beaucoup de chercheurs avaient prédié, dans le passé, que cette progression pourrait s'arrêter bientôt, la tendance n'a pas encore rencontré des problèmes majeurs. Les dispositifs MOSFET modernes ont des longueurs de grille plus petites que 100 nanométres, donc nous pouvons dire qu'il s'agit d'un régime decananometrique. Beaucoup de nouvelles questions surviennent avec la miniaturisation ("scaling") et les ingénieurs préparent des solutions différentes. Par exemple la conduction de canal est affectée par de très petites mobilités du porteur et par le fortes résistances en série. Parmi les solutions proposées que nous pouvons mettre en valeur: amélioration de la mobilité par contrainte, l'hétérojunction au bords du canal, transport quasi balistique. La dernière solution exige que le transport du canal approche la limite balistique pour les longueurs de grille les plus courtes. Le transport balistique est accompli quand les porteurs dans le canal ne souffrent aucun événement de collision ("scattering") qui dérange leur mouvement. D'autres ingénieurs ont préféré vaincre l'architecture Bulk classique et ont proposé des technologies différents comme le SOI MOSFET Fully-Depleted et le Double-Grille SOI MOSFETs qui, potentiellement, double le courant. Dans cette thèse nous discutons quelques aspects physiques qui est très importants dans les MOSFET avec longueur de grille ultracourte. Notre travail est divisé en deux parties.

La première partie adresse les propriétés du transport du porteur dans le canal des MOSFETs avancés pour les applications numériques et analogiques. L'approche sera entièrement basée sur des simulations. Les premiers deux chapitres sont une introduction à notre analyse. Le premier discute le scaling de la technologie MOSFET et deux sujets principaux: modéliser le transport quasi balistique et les avantages des nouvelles architectures MOSFET. Le deuxième chapitre décrit le deux méthodes de la simulation les plus importantes pour les dispositifs avancés, en particulier, il présentera l'approche Monte Carlo, qui est la base de nos recherches dans les chapitres suivants. Dans les chapitres III et IV, nous essaierons de répondre à une question commune: est-ce que le moderne Technologie MOS s'approche à la limite balistique? Pour donner une réponse nous utiliserons le modèle Monte Carlo pour étudier les propriétés du transport dans les dispositifs à canal court pour les applications numériques plus performantes. Dans le chapitre V l'approche Monte Carlo serait utilisée avec succès pour étudier les performances AC dans

les dispositifs avancés pour les applications RF. En particulier nous montrerons les avantages de la simulation Monte-Carlo par rapport à la méthode "drift-diffusion".

La deuxiéme partie du manuscrit discute le couplage capacitif entre les contacts de grille dans les SOI MOSFET Fully-Depleted et dans les MOSFETs Multi-Grilles. Ceux-ci sont les candidats les plus prometteurs pour remplacer l'architecture Bulk du Noeud de la Technologie 45nm et au-delà. Un grand effort est consacré à leur processus, modelisation et caractérisation. L'effect de couplage est défini comme la dépendance du voltage de seuil du canal avant V_{TH} en function du voltage et des propriétés de la porte arrière. Cet effect est utilisé communément pour la caractérisation des dispositifs SOI. Le Chapitre VI est consacré à la simulation du couplage dans les SOI MOSFET avec épaisseur ultra mince. En fait, quand l'épaisseur du silicium t_{SI} est de l'ordre de 10 nanomètres, l'effect de couplage change et s'amplifie. Cet effet a été étudié par simulations et nous avons donné une explication en termes de potentiel électrique dans la couche mince de silicium. Enfin, nous caractériserons, dans le chapitre VII, les effects de couplage dans le transistors Multiple MOSFETs (MIGFET). Ces dispositifs non-planar, qui sont très semblables aux FinFET, ont un couplage capacitif complexe entre les trois contacts de grille que nous pouvons utiliser pour former un canal. Notre but sera de séparer les différents types de couplage et comprendre lesquels sont les plus considérables.

Analyse du Transport Balistique

La modélisation des dispositifs avec longueur de grille L_G plus courte que 25 nm doit traiter, comme projeté par la ITRS [3], la quantisation de la charge d'inversion et le transport non-local, très loin d'équilibre. La méthode Monte Carlo (MC) est regardée comme un outil très prometteur pour simuler le transport des porteurs dans les dispositifs nanometriques, au moins jusqu'à $L_G = 10$ nm.

Bandit est un simulateur Monte-Carlo Full-Band pour le gaz 3D d'électron avec des corrections au potentiel électrostatique pour inclure l'effet de quantification d'energie sur la distribution spatiale de la charge d'inversion [35]. Bandit inclut les principaux mécanismes: (i) collision avec le phonon, (ii) un modèle pour la rugosité de la surface (SR) basé sur la moyenne du champ effectif éprouvé par les porteurs (ce modéle a été proposés et validés dans [33]), (iii) scattering par les ions du dopants, (iv) scattering par interaction Électron-Plasmon à l'intérieur des régions lourdement dopées.

Pour étudier le transport balistique, nous avons utilisé le modèle présenté en [13, 14, 15, 16]. Ce modèle:

1. décrit le courant comme un flux de porteurs et des coefficients de transmission et réflexion. En particulier le modèle distingue un flux positif, de la source au drain, et un flux négatif, du drain à la source. La somme des deux flux est le courant total du dispositif;

2. reconnaît que la plus grande contribution au courant négatif est due aux collisions, qui reorientent les porteurs vers la source, ("back-scattering") dans le kt-layer. Cette région est définie comme la zone de canal où le profil de potentiel diminue de kT/q par rapport au pic de potentiel au point d'injection x_{inj} entre la source et le canal;

3. définit un coefficient de backscattering r qui tient compte des porteurs qui, après avoir été injectés de la source dans le canal, reviennent à la source à cause d'un événement de scattering. En general r dépend de la mobilité et de la longueur de la region kT;

4. avec r, il calcule une expression du courant I_{ON} du dispositif et montre que le rapport entre la courant avec et sans scattering est égal à $(1 - r)/(1 + r)$.

Une analyse approfondie a été faite sur un dispositif réaliste: une structure du dispositif est définie en suivant la spécification pour les noeud technologique 45 nm de la Haute Performance de l'édition 2003 de l'ITRS. La simulation a été faite avec et sans scattering dans le canal. La courant balistique I_{BL} est environ 50% supérieur á celui qui inclut le scattering I_{ON}. La Figure 1 montre les courants simulés. Donc l'effet du scattering reste encore considérable dans les dispositifs avec longueur de grille très court. En d'autres termes les dispositifs des prochains noeuds de la technologie ne fonctionneront pas dans le régime idéal du transport balistique.
Le résultat a été interprété en termes de flux de courant qui va de la source au drain (I^+) et du drain à la source (I^-). Le courant total est $I = I^+ - I^-$. Nous définissons le point d'injection x_{inj} au pic du profil de potentiel entre la source et le canal. Quand le transport est balistique il n'y a aucun flux I^- au le point d'injection. Nous avons vérifié que dans les dispositifs très courts (jusqu'à $L_G = 14$ nm) il y a encore un flux de courant négatif en x_{inj} dû au backscattering.
La Fig. 2 montre la somme des événements de scattering le long du canal et la contribution du backscattering au courant I^- au point d'injection. Cette

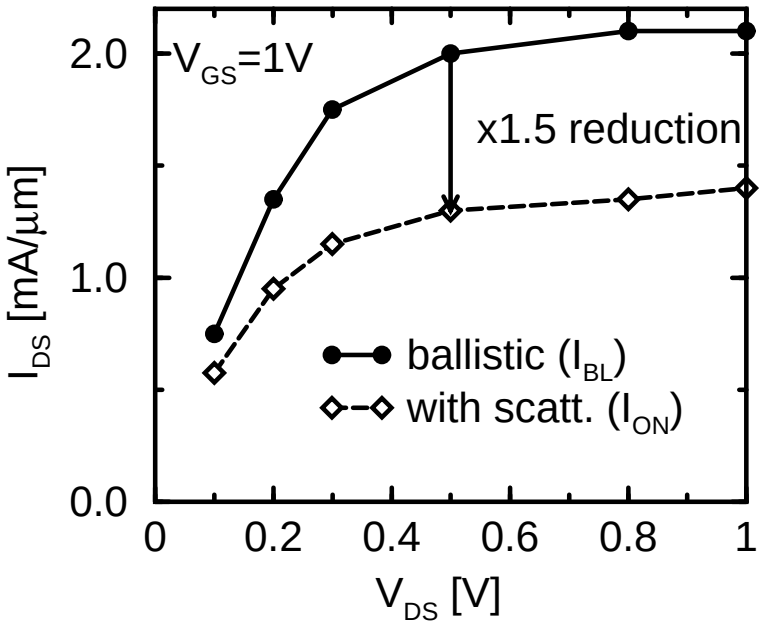

Figure 1: Courants I_{ON} et I_{BL} simulés pour un transistor Double-Gate SOI avec $L_G = 25$ nm.

contribution déchoit de manière exponentielle avec une longueur que nous avons montrée égale à la longueur de la region kT. Notons que l'effect du scattering sur le courant I_{ON} n'est pas simplement proportionnel au nombre d'événements de back-scattering. En effet le scattering change le profil de potentiel le long du canal. A cause de cet effect la longeur de la region kT ne peut pas être calculée avec des simulateurs balistiques simples; ce calcul demande une simulation auto-consistante comme dans le cas MC.

Le modèle présenté en [13] prédit aussi que le rapport entre le courant avec scattering et le courant ballistique $BR = I_{ON}/I_{BL}$ est égal à $(1-r)/(1+r)$, où r est le coefficient de backscattering. Ce modèle ne prévoit pas de formule rigoureuse de r mais seulement une possible expression déduite en utilisant l'idée de mobilité et la longueur de la region kT. Cependant cette approche ne peut pas être utilisée pour évaluer I_{ON} directement à partir de quelque expression approximée de I_{BL} car:

1. seulement une simulation auto-consistante peut calculer r;
2. la définition de mobilité dans les dispositifs ultra-courts n'est pas simple.

Après avoir défini r comme le rapport I^-/I^+, calculé au point d'injection,

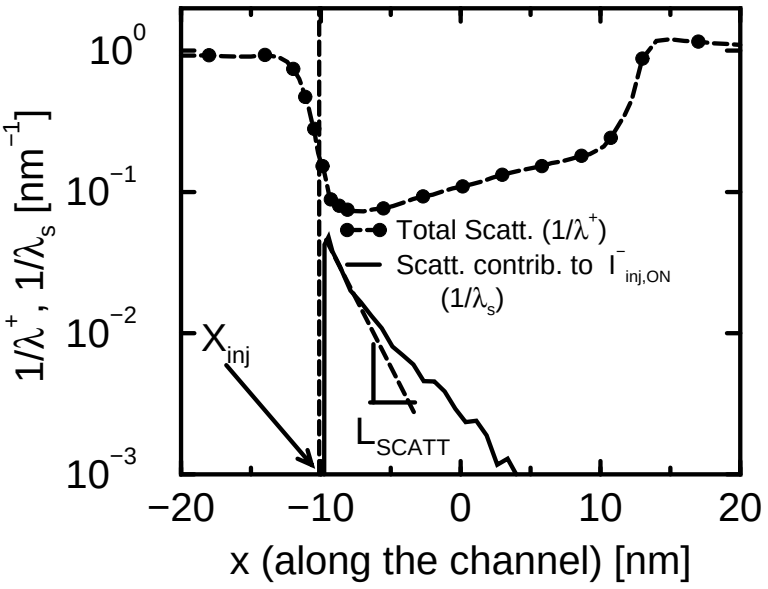

Figure 2: Nombre total d'événements de collision électronique dans le canal et d'événements de backscattering qui contribuent au courant au point d'injection. Dans cette figure $L_G = 19$ nm.

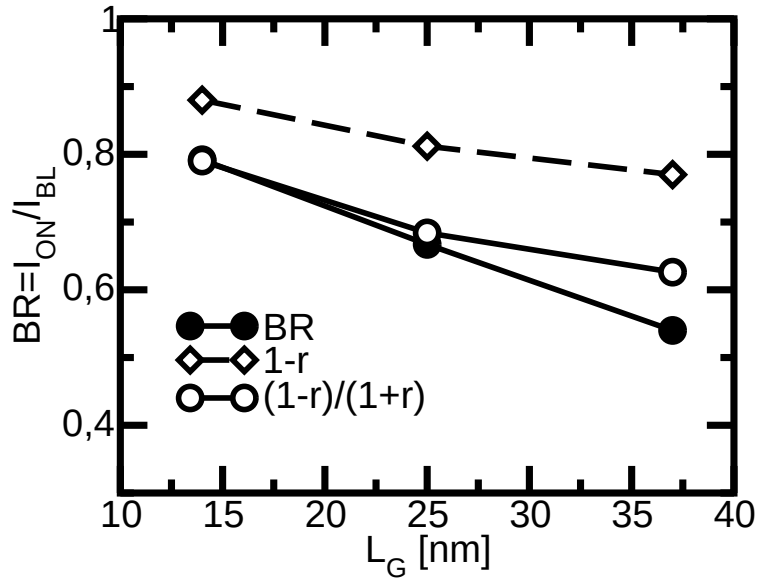

Figure 3: Comparaison entre le taux de ballisticité $BR = I_{ON}/I_{BL}$ et le coefficient $(1-r)/(1+r)$, ou r est le rapport I^-/I^+, calculé au point d'injection.

nos simulations ont confirmé que $BR = (1 - r)/(1 + r)$, comme est illusté dans la Fig. 3.

Effet de la miniaturisation sur le Transport Balistique

Le même méthode Monte Carlo a été appliquée à une étude systématique du transport quasi balistique dans les transistors MOSFET Bulk et Double-Grille (DG) SOI envisagés selon l'ITRS 2003. Notre but est comprendre comment le transport quasi-balistique change avec la miniaturisation des dispositifs et, en particulier, si la diminution de la longueur de grille L_G nous permet de nous rapprocher du cas balistique pur. Nous avons pris en considération les noeuds technologiques 130, 90, 65 et 45 nm. Pour chaque noeud nous avons défini et simulé des dispositifs avec différentes longeurs L_G. Nous avons montré comment l'effet DIBL (Drain Induced Barrier Lowering) et les autres effets de canal court (SCE) augmentent avec la diminution de L_G. Pour chaque dispositif, le rapport $BR = I_{ON}/I_{BL}$ a été calculé, en simulant le courant avec ou sans scattering.

Selon le modèle [13, 14, 15, 16], il est possible de se rapprocher de la limite balistique avec l'augmentation, dans la région de canal voisine à x_{inj}, du champ électrique parallèle au mouvement des porteurs. Nous pouvons "mesurer" l'augmentation du champ comme l'accroissement des effets SCE, et ensuite, dans notre cas, du DIBL. L'analyse n'a pas montré une dépendance univoque entre BR et le DIBL: on observe, en général, que le transport dans le canal est plus balistique si le DIBL grandit. Telle dépendance change selon le noeud technologique considéré.
Le résultat est sensiblement modifié si nous analysons BR en fonction de la longueur de grille L_G. On observe que:

- Dans le cas SOI Double-Grille, le rapport BR reste invariant entre dispositifs avec la même longueur de grille mais de différents noeuds. Ce résultat nous a mené à une courbe "universelle" montrée en 4. Ce n'est pas un résultat physique mais un produit utile des régles de scaling du roadmap.

- Nous avons interprété la courbe BR par l'analyse de la region kT: la longueur de cette région est une fonction "universelle" de L_G pendant que le mobilitè est environ la même entre tous les dispositif SOI;

- la mobilitè dans les DG SOI est due seulement aux collisions avec les phonons et moins à la rugosité de surface(SR) ou aux impuretés. Le

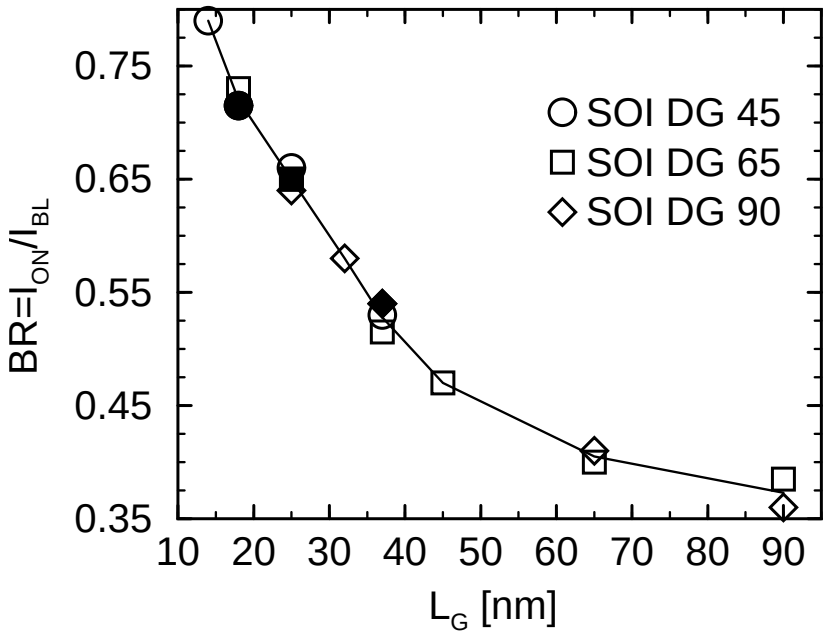

Figure 4: Rapport de ballisticité BR calculé dans le cas de transistors Double-Grille SOI avec différentes longeurs, issus de générations technologique successives.

contrôle des SCE est assuré en réduisant l'épaisseur du film lorsque L_G est réduit. On n'utilise pas de concentrations fortes de dopant: tous nos dispositifs Double-Grille ont un substrat sans dopant intentionnel;

- Dans le MOSFET Bulk, les rapports BR sont plus bas et le transport est plus éloigné du cas balistique. Nous avons vérifié que cette aggravation est due au scattering par rugosité de surface. Il s'agit d'une consequence indirecte du au haut niveau de dopage dans le substrat, nécessaire pour garder le SCE à un niveau acceptable, mais qui conduit à l'augmentation du champ électrique vertical.

Les Figures 4 et 5 montrent, respectivement, les courbes de $BR(L_G)$ dans le cas DG SOI et Bulk pour tous les noeuds technologiques. Donc, nos résultats montrent comment le canal devient plus balistique avec la diminution de L_G dans le cas DG SOI, pendant que les transistors Bulk sont désavantagés par les hauts niveaux de dopant nécessaires pour contenir le DIBL.

RF Performance

Le but de ce chapitre est d'étudier les performance RF dans les dispositifs MOSFET par l'analyse Monte Carlo. Notre simulateur Bandit a été modifié

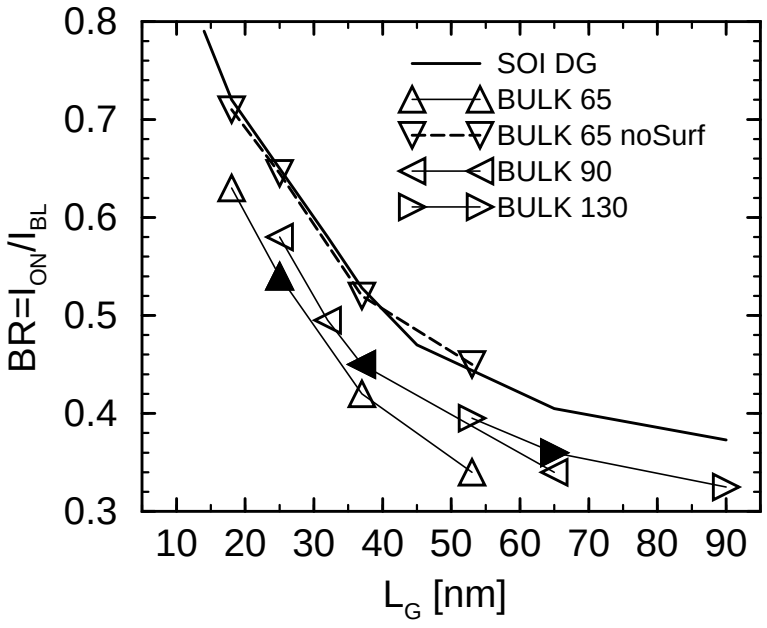

Figure 5: Rapport de ballisticité BR calculé dans le cas de transistors Bulk et SOI. On observe que, sans l'effet de collisions par les rugosités de surface, le rapport BR des MOSFET Bulk devient égal à celui SOI.

pour inclure l'analyse temporelle et ensuite calculer

- le retard de propagation du signal dans le canal;
- la matrice des admittances Y.

La description de l'adaptation du simulateur MC à l'analyse AC est présentée dans le chapitre.

Les dispositifs de test sont représentatifs de la technologie RF pour les années 2006, 2007 et 2008 (L_G=53, 45 et 37 nm). Le dispositifs MOSFET considérés sont soit de type Bulk, avec un fort dopage dans le substrat pour reduire les SCE, soit de type SOI Single-Grille (SG), avec film ultra-mince (Ultra-Thin Body ou UTB) sans dopant. Les résultats confirment les bonnes propriétés de miniaturisation pour les paramètres AC principaux, fréquence F_T et bande passante. Fig. 6 montre les résultats pour F_T. En particulier, nous avons montré que l'approche quasi-statique, est adéquate pour l'analyse du délai de la propagation du signal et pour le calcul de la fréquence de transition (Fig. 7). Cette méthodologie permet une analyse distribuée qui décrit la formation du retard le long du canal et indique que le délai est en rapport direct avec la vitesse moyenne du porteur le long de la région de canal

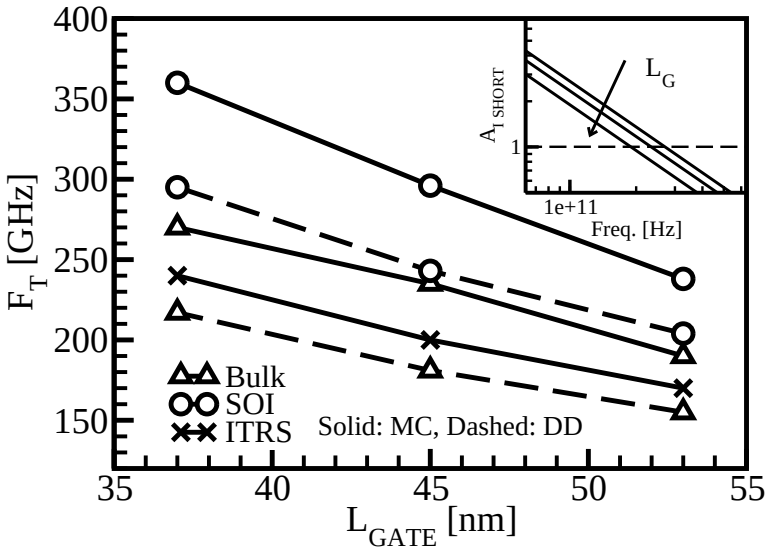

Figure 6: Fréquence de transition obtenue par les simulations MC de tous les dispositifs (Bulk et SOI de longeurs variabiles). Les résultats drift-diffusion sont inclus (lignes pointillées).

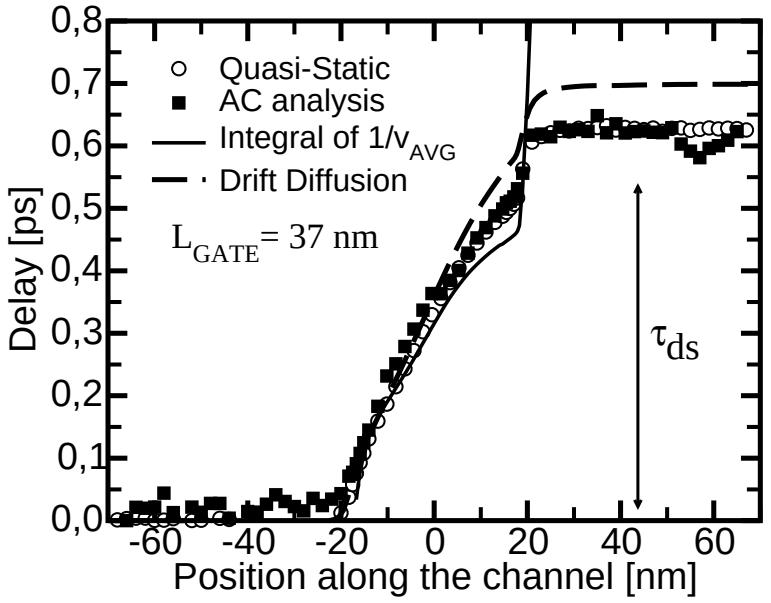

Figure 7: Comparaison entre le retard des signals obtenu avec une analyse en fréquence et le retard quasi-statique. Le cas montré ici est un transistor Bulk $L_G = 37$ nm.

intrinsèque. Le bon accord entre l'approche quasi-statique et l'analyse en fréquence est une bonne nouvelle du point de vue méthodologique parce que le cas quasi-statique demande seulement des simulations DC très rapides. Nos résultats indiquent que les transistors UTB-SG SOI MOSFETs peuvent fournir des valeurs considérablement meilleurs pour la transconductance et F_T, grâce à la plus grande vitesse du porteur le long de le canal, et à la plus grande bande passante du gain en tension, dû à la capacité du drain réduite . En revanche, ils se caractérisent pour une plus grande conductance de drain à cause des effets SCE légèrement amplifiés. En effet, pour réduire les effets de canal court ultérieurement, on devrait utiliser des épaisseurs de film de silicium $t_{SI} < 5$ nm, avec certaines problèmes concernant également le processus technologique et les résistances en série.

Ce chapitre présente une comparaison entre la simulation Monte-Carlo et l'analyse en fréquence dans le cas de simulateurs drift-diffusion (DD). Les modèles de mobilité implémentés dans le cas drift-diffusion sous-estiment la mobilité, et ensuite la transconductance. Nous avons observé une différence considérable entre les résultats DD et MC, spécialement dans le cas SOI (Fig. 6). Par conséquent les retards de propagation dans le canal, qui résultent des ces deux type des calcul, peuvent être très diffèrents.

Effets de couplage dans les transistors SOI MOSFET avancé

Dans la deuxième partie du manuscrit, nous avons étudié les effects de couplage capacitif non-conventionnel qui est présent dans les dispositifs SOI extrêmement minces. Le couplage est le lien qui se forme entre les deux interfaces silicium-oxyde à travers la couche mince de silicium. Le résultat le plus visible du couplage est la dépendance de la tension de seuil V_{TH} d'une grille de la tension V_{G2} appliquée sur l'autre grille. Le couplage capacitif entre les grilles est un effet très utilisé pour la caractérisation des propriétés des interfaces silicium-oxyde et du film mince [66].

Les règles de scaling des dispositifs SOI FD prévoient que l'épaisseur de la couche de Si t_{SI} diminue continuellement avec L_G. La variation du V_{TH} par V_{G2}, selon le principal modèle analytique du couplage [74], devient donc plus forte dans les noeuds avancés. Il est observé que, dans le cas $t_{SI} < 10$ nm, le couplage change également de comportement [67, 68]. La possibilité de polariser une interface en inversion et l'autre en accumulation est en outre physiquement discutable à cause de l'attraction entre les charges positives et

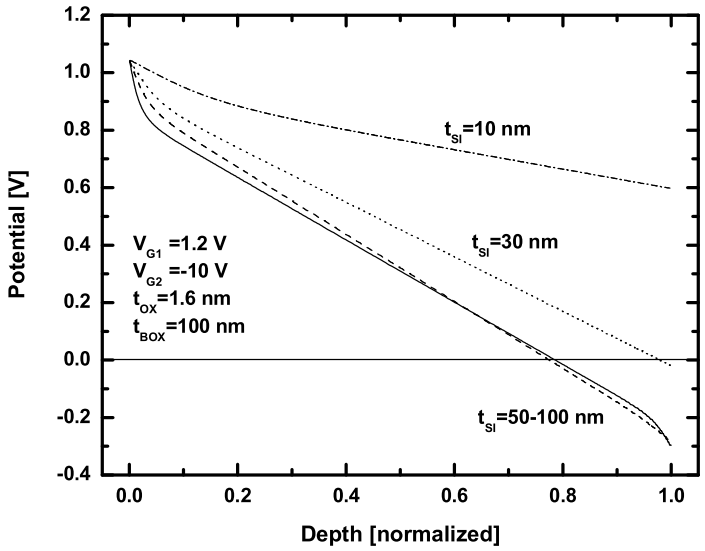

Figure 8: Profil de potentiel électrique entre les deux interfaces silicium-oxyde, pour épaisseurs différentes de film t_{SI}. On observe que la charge d'accumulation à l'interface arrière disparaît dans la couche ultra-mince (10 nm).

négatives. Nous avons vérifié, par simulations, que, dans certaines conditions, il n'est pas possible d'imposer des types de charge différente sur les deux interfaces: un seul type de charge, positive ou négative, peut être présent dans le film ultra-mince.

Une épaisseur critique a été définie analytiquement, comme la valeur de t_{SI} pour laquelle, quand l'interface avant est en inversion, il n'y a pas plus de charge d'accumulation à l'interface opposée. La Figure 8 montre ce résultat en termes de potentiel électrique. Comme conséquence, il devient impossible de caractériser un canal indépendamment du canal opposée, en masquant les défauts de l'interface opposés par accumulation, comme s'etait le cas dans les couches SOI plus épaisses. Dans ce cas, la courbe de $V_{TH}(V_{G2})$ change et la saturation de la tension de seuil, pour V_{G2} très negatif (accumulation), n'est plus observée car l'autre interface n'est jamais en accumulation. Pour t_{SI} très minces nous avons vérifié la présence du super couplage: le profil de potentiel dans le film ultra-mince devient presque plat (Fig. 9). Il est contrôlé par l'un ou l'autre des contacts de grille. Le transconductance ne refléte plus la mobilité du canal de surface car elle intègre le profil de la mobilité à travers le film.

Les concepts de super-couplage et d'épaisseur critique sont tout à fait

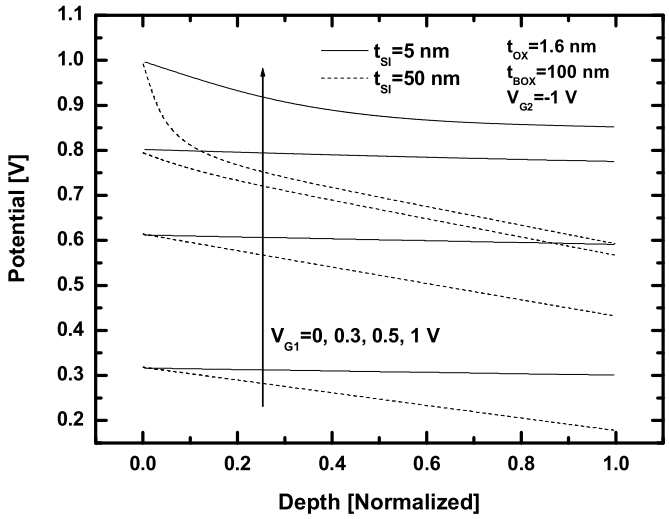

Figure 9: Profils de potentiel pour épaisseurs t_{SI} différentes et tensions de grille avant variables. Dans le cas ultra-mince, le profil de potentiel est pratiquement plat (super-couplage).

inédits. Nous pensons qu'ils seront pleinement utilisés lorsque les transistors SOI atteinderont une épaisseur sous-critique.

Couplage dans le MIGFET

Autres effets de couplage non-conventionnel ont été caractérisés sur des dispositifs MOS non-planaires. Les MIGFETs sont une variation des FinFETs où les deux contacts de grille latéraux sont séparés par un processus technologique spécial [84]. Dans les FinFET il y a un seul type de couplage capacitif: entre le contact de substrat et les deux contacts de grille latéraux. Nous avons appelé ce type d'effet *couplage vertical*. On observe que la variation de la tension de seuil en function de la tension de substrat V_{SUB} est relativement faible dans les fins étroites, car le potentiel à l'interface silicium-oxyde enterré contrôle une région très petite du fin.

Dans les MIGFETs, au-delà du couplage vertical, il existe aussi un *couplage latéral* entre les deux contacts de grille séparés: ce type de couplage est mono-dimensionnel comme dans les SOI MOSFET Single-Grille. En général, la tension de seuil à une interface est contrôlée par la tension appliquée à l'interface opposée V_{G2} et par la tension du substrat V_{SUB}. À travers la car-

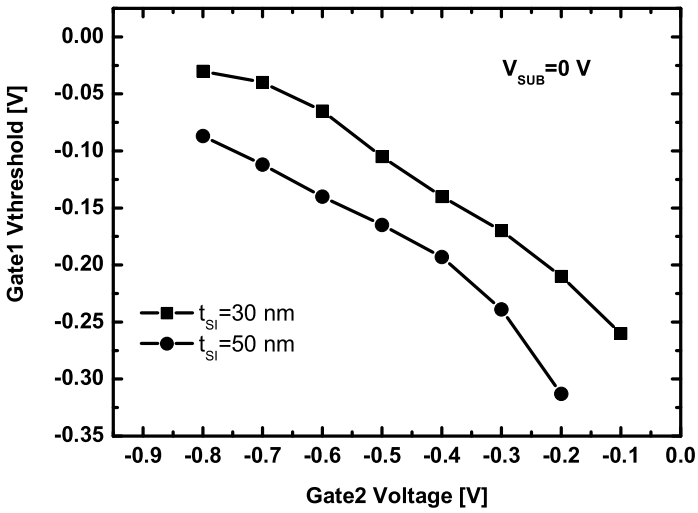

Figure 10: Valeurs de la tension de seuil V_{T1}, dans un MIGFET, en fonction de V_{G2}, extraites pour différentes t_{SI}.

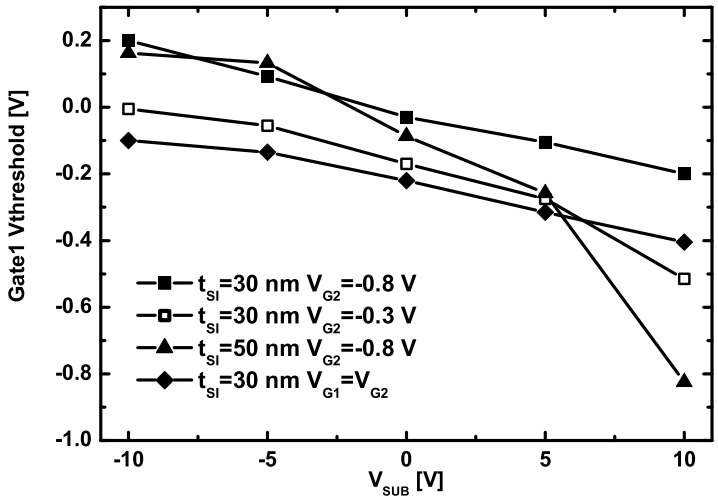

Figure 11: Tension de seuil en fonction de la tension du substrat pour dispositifs MIGFET avec differénts t_{SI} et conditions de polarisation.

actérisation électrique des MIGFETs avec différentes épaisseurs du fin t_{SI}, nous avons observé que le couplage latéral entre les deux interfaces laterales est l'effet dominant dans les MIGFET. Il est très fort dans les dispositifs avec du petit t_{SI} (Fig. 10). Pour faire une comparaison nous avons aussi considéré le cas où le dispositif est utilisé comme un FinFET, c'est-à-dire avec la même tension sur les deux contacts de grille latéraux $V_{G1} = V_{G2}$. Dans ce cas le couplage latéral est absent et celui vertical est moindre. Nous avons vu que, si t_{SI} est assez mince, un MIGFET avec $V_{G1} \neq V_{G2}$, peut avoir le même faible couplage vertical que dans le cas double-grille.

Dans le cas de fins plus larges, le contact de substrat a un effet considérable parce qu'il contrôle une part ample du film: dans ce cas le couplage vertical ne doit plus être négligé. La Figure 11 montre les résultats de la comparaison du couplage vertical dans les différents dispositifs.

Introduction

Integrated circuits (ICs) have driven the technological development during the last forty years: modern ICs are present in every electronic system going from the common personal computer to the communication systems, from medical equipment to automotive industry, etc. The main actor of this improvement has been the MOS transistor, due to its unmatched scalability. The digital ICs are completely built with MOSFETs and the old technologies, like the bipolar one, have been almost completely replaced. In the last years the MOSFET have been also common in other applications, like power or RF circuits, for which other devices were the "only" choice. The actual trend is to integrate all types of required applications into a single chip (system-on-chip, SoC). This will require to overcome the problems related to the coexistence of the various technologies that are currently integrated in separate chips. It is easy to predict that the MOSFET will become more and more popular also in the non-digital world.

The progress of device integration has procedeed for more than forty years, following the well-known Moore's law which states that the transistor density doubles every 24 months [1]. Modern, high-performance MOS devices feature gate lengths much shorter than 100 nanometers. Many, new issues are arising right now and the engineers are preparing many different solutions to face them. For example the channel current is degraded due to very low carrier mobilities and high series resistances. Among the solutions proposed to overcome these problems, we can highlight mobility improvement by strain and heterojunctions at the channel edges. Engineers also expect to approach the ballistic transport regime which will compensate the mobility degradation in gate lengths shorter than 20 nm. The ballistic transport is achieved when the carriers in the channel do not suffer scattering events that perturbe their movement. Another solution suggests to overcome the classical Bulk architecture and proposes different technologies like the Fully-Depleted SOI MOSFET and the Double-Gate SOI MOSFETs which, potentially, doubles the on-current.

In this thesis we discuss physical issues which are very important in MOSFET with ultra-short channel length. The work is divided into two parts.

The first part addresses the carrier transport properties in the channel of advanced MOSFETs for digital and analog applications. The approach will be entirely simulative. The first two chapters are an introduction to our analysis. The first one discusses the main issues affecting the scaling of the

CMOS technology and in particular two main arguments: modelling of the quasi-ballistic transport and advantages of the new MOSFET architectures. The second chapter describes the most important simulation methods for advanced devices with particular regard to the Monte Carlo approach.
In chapters III and IV we use the Monte Carlo approach to study how close to the ballistic limit are modern CMOS technologies.
In chapter V the Monte Carlo approach is used successfully to study the AC performance in advanced devices designed for RF applications.

The second part of this thesis discusses the capacitive coupling between the gate contacts in Fully-Depleted SOI MOSFET and in Multi-Gate MOSFETs. These devices are promising candidates to replace the Bulk architecture starting from the 45 nm Technology Node. A great effort in terms of process, modelling and characterization is necessary. The coupling effect is defined as the dependence of the front-gate threshold voltage V_{TH} on the bias and on the properties of the back gate. This effect is absent in Bulk devices and is commonly used for characterization purposes in the SOI case.
Chapter VI is dedicated to the simulation of gate coupling in SOI MOSFET with ultra-thin body. In fact, when the film thickness t_{SI} is only few nanometers, the coupling effect changes and gets reinforced: in this regime we choose to use the term *supercoupling*.
In Chapther VII, we will report about the characterization of the coupling effects in Multiple Independent Gate MOSFETs (MIGFET). These non-planar devices, which are very similar to FinFET devices, exhibit a complex capacitive coupling between the three gate contacts that we can use to form a channel.

Part II

Modeling of decananometric MOS Devices

Chapter 1

The short-channel MOSFET: device modelling and scaling

The aim of this chapter is to review the physics of the MOSFET and to introduce concepts that are used throughout the thesis. For space reasons, we limit our review to those scaling issues and transport properties which are essential for the comprehension of our work (section I). Then, in section II, we review the quasi-ballistic model for the modelling of the on-current in short-channel MOSFET: this model will be extensively used in our work. In the last section we show the main properties of SOI devices, which are important candidates for the ultimate MOS scaling in the coming years.

1.1 Channel transport in long-channel and ultra-short MOSFETs

1.1.1 General current-voltage characteristic

The n-channel MOSFET, whose typical structure is shown in Fig. 1.1, features a relevant drain current when biased above-threshold condition ($V_{GS} \geq V_{TH}$) and with drain voltage $V_{DS} > 0$ V. The most common current-voltage expressions for an nMOS are the well-known formulas:

$$I_{DS} = \frac{W}{L_G} C_{OX} \mu_{EFF} \left((V_{GS} - V_{TH}) V_{DS} - \frac{1}{2} V_{DS}^2 \right) \quad \text{linear region} \tag{1.1}$$

$$I_{DS} = \frac{W}{2L_G} C_{OX} \mu_{EFF} (V_{GS} - V_{TH})^2 \quad \text{saturation region} \tag{1.2}$$

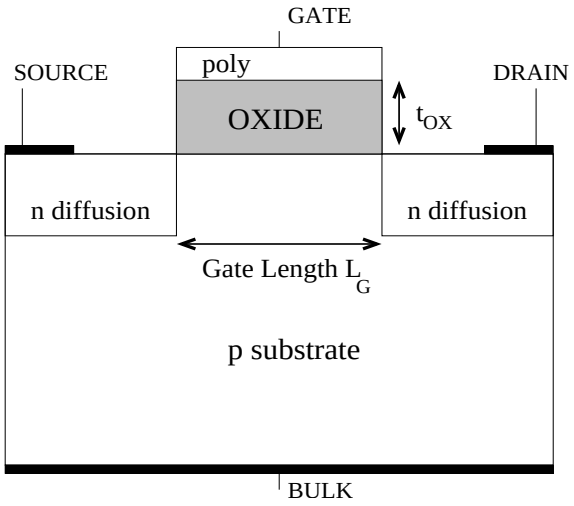

Figure 1.1: Schematic structure of a standard Bulk MOSFET. The bulk contact is also called substrate contact.

where V_{TH} is the threshold voltage that can be expressed as:

$$V_{TH} = V_{FB} + 2\Psi_F + \frac{\sqrt{4\epsilon_{si} q N_a \Psi_F}}{C_{OX}} + \frac{\sqrt{2\epsilon_{si} q N_a}}{C_{OX}} (\sqrt{V_{SB} + 2\Psi_F} - \sqrt{2\Psi_F}) \quad (1.3)$$

The main parameters shown in the above expressions are: L_G is the channel length, $C_{OX} = \epsilon_{OX}/t_{ox}$ is the oxide capacitance, μ_{EFF} is the effective mobility, N_a is the substrate doping concentration.

These simple expressions can be obtained from hand-calculation under many approximations. The most important one is the *Gradual Channel Approximation* [2], which assumes that the electric field in the direction normal to the silicon/oxide interface (from now it will be called simply the normal electric field) is larger than the one parallel to the transport direction (the parallel electric field). The GCA approximation will be useful in section 1.2, where we will discuss the quasi-ballistic transport. The GCA is reasonable at low drain-source voltage V_{DS}, while it fails at large drain bias and the concept of pinch-off is introduced to explain the saturation of the current.

The main characteristic of the MOSFET is its extreme scalability. Even if simple, the above equations are useful to evaluate the dependence of the current on the different geometrical parameters and applied voltages. In particular we observe that I_{DS} increases as the gate length L_G decreases. However the scaling of the MOSFET device is not a straightforward task, in particular in the short-channel regime where many second-order effects tend to become more and more relevant.

1.1.2 Theory of the MOSFET scaling

The scaling theory identifies the design criteria that must be adopted when we modify the device parameters in order to improve the device performance and increase the density of integration. Two strategies can be adopted: i) constant-voltage scaling, for which all device dimensions are scaled-down by the same factor $\kappa > 1$, but the voltages are kept constant; ii) constant-field scaling [3], for which both dimensions and applied voltages are scaled down by the same factor κ. The first solution allows to increase the on-current by the same factor, then greatly increasing the device performance. This improvement comes at the cost of drain-source punch-through, large threshold sensitivity to channel length and drain voltage (short-channel effects, often referred as SCE) and hot carrier related reliability problems. The constant-field scaling keeps the above issues under control since the internal average electric fields remain constant. The current scales as $1/\kappa$, thus it is reduced, but the circuit delay time CV/I and the power dissipation per circuit scale-down, respectively, as $1/\kappa$ and $1/\kappa^2$. This set of rules suffers for limited flexibility and has not been rigorously applied in the modern semiconductor technology.
In fact, engineers prefer to scale-down the device dimensions and the applied voltages at the same time *but* with two different factors. This scaling rule is called generalized scaling theory, well described in [4]. The aim of the generalized theory is to preserve the field pattern within the scaled device so that punch-through and the other detrimental effects listed above are expected to remain essentially constant, in spite of the increase in the electric field strength. Moreover a relevant flexibility is possible since the two scaling factors can be varied independently in order to find the best tradeoff among the different device characteristics.

The scaling of modern, short-channel devices is very difficult due to the increasing importance of non-scalable parameters, like the noise and the junction built-in potentials, and second-order effects, like the leakage currents. The subsection 1.1.4 will describe some common issues in short channel MOSFETs. Engineers have to face many tradeoffs by scaling the different device parameters with different rules. When necessary, the device architecture is modified (for example introducing different materials) in order to maintain the scaling trends.

In order to spread information and elaborate shared visions of the future trends, the International Technology Roadmap for Semiconductor (ITRS) [5] is published every two years and updated every year. The ITRS is a reference

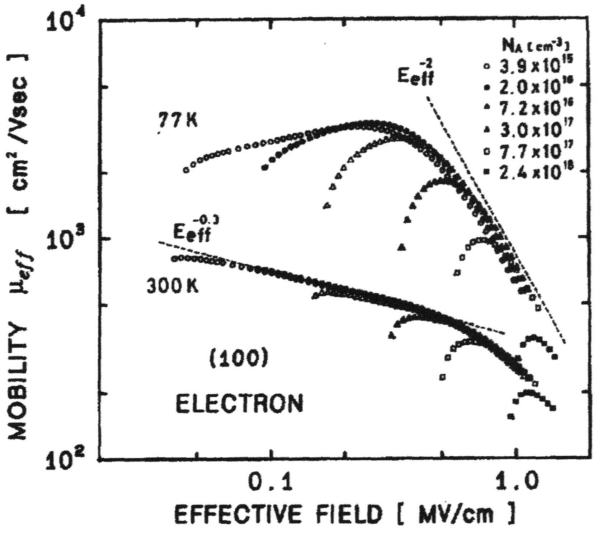

Figure 1.2: Example of channel mobility versus the effective field. An universal behavior can be observed in all plotted mobility curves. The figure also emphasizes the impact of the doping concentration on μ_{EFF}. The degradation due to the surface roughness is clearly visible at large $\mathcal{E}_{EFF}$, where the device operates. Picture taken from [7].

for the microelectronic industry since it contains all the scaling rules for the next technology nodes, in all field of application (digital, analog, memories, etc.). Moreover it also resumes all the scaling issues and presents all the known technological solution. We will often use the roadmap in this thesis as a guideline for the definition of our template devices.

1.1.3 Carrier mobility in the MOSFET inversion layer

Carrier mobility is of utmost importance as it limits the device performance and a short discussion about follows.

The carrier mobility is defined as the ratio between the carrier velocity and the electric field. In a three-dimensional silicon lattice the velocity-field relation is linear at low electric fields, than saturates at a value v_{sat} for fields approaching $10^4\,\mathrm{V/cm}$. Thus the low-field mobility is constant and it is $\approx 1400\,\mathrm{cm^2/Vs}$ for electrons and $\approx 500\,\mathrm{cm^2/Vs}$ for holes in undoped silicon. Inside the thin inversion layer which forms the channel region the carrier mobility μ_{EFF} is much lower than in a three-dimensional lattice. The channel mobility is commonly analyzed by following the well-known approach

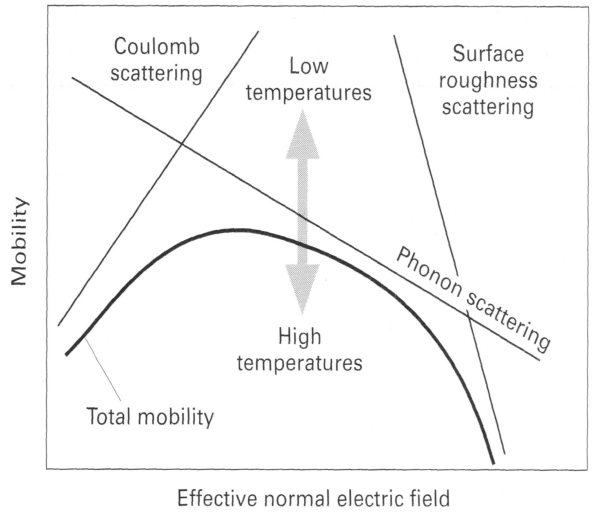

Figure 1.3: Carrier mobility in the channel region as a function of the effective field $\mathcal{E}_{EFF}$. The dependence on the three main scattering sources and on temperature is shown. Picture taken from [7].

from Sabnis and Clemens [6] and extended by Takagi [7], stemming from the observation that μ_{EFF} is an universal function of the *effective vertical field* $\mathcal{E}_{EFF}$. The effective field for electrons is defined as

$$\mathcal{E}_{EFF} = \frac{1}{\epsilon_{si}}\left(|Q_d| + \frac{|Q_i|}{2}\right) \tag{1.4}$$

and depends on the inversion charge density Q_i and on the depletion charge Q_d. Examples of mobility curves in nMOS devices are plotted in Fig. 1.2. This picture is very useful for a fast comparison of μ_{EFF} in different MOSFETs.

From a microscopic point of view it is possible to demonstrate that mobility is proportional to the average time between two perturbations of the carrier motion. These perturbations, called *scattering events*, are caused by interactions with the semiconductor lattice. There are many sources of scattering and we now list the most important ones:

- phonon scattering, caused by interactions with lattice vibrations. It is the most important source of perturbations, except at low temperatures T;
- impurity scattering, which is caused by the electrostatic interaction

between the carriers and the ionized atoms of the dopant. Its relative impact increases with the doping level and at low T;

- surface roughness scattering, which is caused by the roughness of the Si–SiO_2 interface.

The impact of the different scattering mechanisms is described by Fig. 1.3 which shows the typical dependence of the channel mobility on the effective field and on three types of scattering previously described. In particular we notice that the surface roughness is very detrimental at large effective fields, required by MOSFETs with large substrate dopant concentration in order to operate in strong inversion.

This mobility degradation can be clearly observed in Fig. 1.2 and is a major issue which limits the performance of modern MOS transistors.

1.1.4 The short-channel MOSFETs

Short-channel MOSFETs differ in many aspects from long-channel devices. In this section we briefly describe concepts necessary for the discussion of this work: short-channel effects, velocity overshoot, quantization of the inversion charge.

Short-channel effects. The Short-Channel Effects (SCEs) consist in the loss of control of the gate on the potential profile along the channel, due to the reduction of the gate length. The main macroscopic effect is the decrease of the threshold voltage while L_G is scaled-down: designers use to calculate the V_{TH} roll-off in order to evaluate the impact of SCEs on the scaled devices.

We can explain the short-channel effect starting from this observation: the 2D potential profile inside the channel is given by the superposition of the normal electric field (controlled by V_{GS}) and the parallel electric field. The latter contribution is given by V_{DS} and by the effect of the junctions between the source/drain and the channel. While L_G decreases, the parallel contribution increases and the gradual-channel approximation collapses: at this point the drain junction controls a non-negligible part of the channel. The parallel field raises the current and this result is equivalent to a decrease in the threshold voltage.

Short channels effects are strongly enhanced when the drain voltage is close to the supply voltage. A different way to evaluate the impact of SCE

on a MOS transistor is the calculation of the Drain-Induced Barrier-Lowering (DIBL), which is the difference of the threshold voltage at low V_{DS} and the one at large V_{DS} [8]. When we apply a large drain voltage, the potential barrier between the source and the drain is lowered and the leakage current is increased. This is equivalent to a reduction of the V_{TH} at large V_{DS}. Long channel MOSFETs have DIBL$\sim$ 10 mV/V (thus there is a variation $\Delta V_{TH} =$ 10 mV for V_{DS} increasing by 1 V) while DIBL$>$ 150 mV/V is not acceptable.

Short-channel effects can be kept under control in many ways: the most common one is to raise the substrate doping concentration N_a. Unfortunately a large doping level decreases the mobility μ_{EFF} of the device, as we have seen in section 1.1.3, thus degrading the on-current. Instead of a single diffusion, constant over the entire substrate, it is also possible to design complex doping profiles like:

- retrograde profile doping: it uses a lower dopant level in the channel region and a larger one in the rest of the substrate. Its design is rather difficult.

- halos (also called pockets) at the source (or drain) end of the channel: they are small regions were the substrate doping concentration is increased. In this way the largest portion of the channel has a limited doping level.

Velocity overshoot. The velocity-electric field relationship is linear only at low fields. When the parallel electric field is large, the velocity saturates at a value $v_{sat} \approx 1 \times 10^7$ cm/s for electrons and holes in silicon [9]. The velocity saturation is common in short devices, due to the large parallel electric fields. In the case L_G is very short, the spatial variation of the potential can be very rapid and a fraction of the carriers may acquire energy much higher than thermal energy. These carriers are not in thermal equilibrium with the silicon lattice and are generally referred as *hot carriers*. Moreover their velocity can exceed the v_{sat} in a confined spatial region: this phenomenon is called *velocity overshoot*. The presence of velocity overshoot is a condition for which transport is highly out of equilibrium. This effect is often neglected in device modeling by assuming a velocity saturation approach that does not allow $v > v_{sat}$. Even if the velocity overshoot is present, the MOSFET current remains essentially controlled by the source end of the channel and we will see in the next section an approach to analyze this transport condition.

Quantization of the inversion charge. In the MOSFET inversion layer, carriers are confined in a potential well close to the silicon surface. This well

is formed by the bent silicon conduction band and the silicon-oxide barrier. In modern devices, the scaling rules dictate very thin t_{ox} and large N_a, thus increasing the electric field normal to the interface. This field is related to the bending of conduction band, which is very rapid, and causes a strong confinement of the minority carriers. From quantum mechanics, we know that the energy of confined carriers can assume only discrete levels, corresponding to the *subbands* of the electron gas. Each subband is associated to a discrete value of energy for motion perpendicular to the interface, and to a continuum of energies for motion in the parallel plane. The carriers now form a 2-D gas [10], instead of the 3-D gas of the silicon bulk. Because of the discretization of energy, the number of carriers in the inversion layer, at fixed V_{GS}, is lower than in the case without quantization (also called *classical* case). This effect is equivalent to an increase of V_{TH} with respect to the classical case. An exhaustive discussion about quantization is not necessary for our work: for further reading, starting from [10], the bibliography is very vast.

1.2 Theory of the Quasi-Ballistic MOSFET

In MOS device with $L \leq 100$ nm, biased at large drain voltages, the carrier transport is strongly out of equilibrium and velocity overshoot takes place. Carriers cross the channel in a very short time and the number of scattering events in these ultra-small devices is small. The concept of mobility starts to collapse, since it is defined by the average time between two interactions that perturb the motion of the carrier. In the case where no interaction happens, carriers travel along the channel in the so called *ballistic* regime, which was studied by Natori [11][12]. Transport in *real* ultra-short devices is *quasi-ballistic*, which means that only few scattering events happen. The expected trend of MOS scaling is to approach the ballistic limit with the decrease of the channel length [5].

A well-known model of the quasi-ballistic transport in MOSFETs has been developed by Lundstrom [13][14][15][16] and we will often refer to its results in order to explain our work. The main characteristics of this approach are now described.

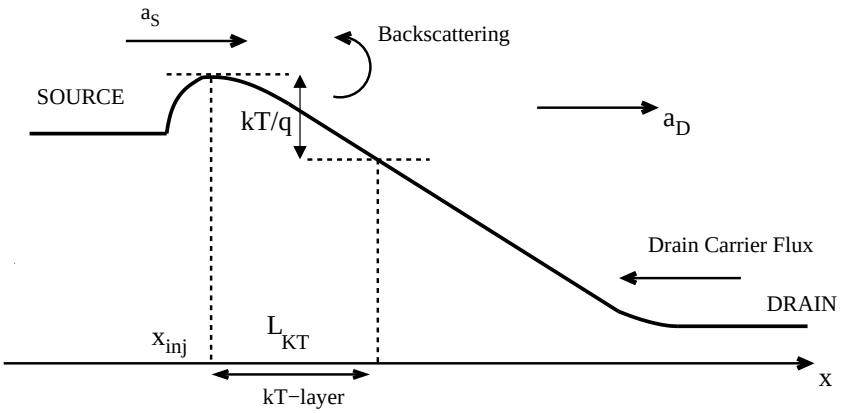

Figure 1.4: Potential profile of a quasi-ballistic MOSFET. The fluxes a_S and a_D coming from the source are equal only in the absence of scattering. a_S is the flux of carriers injected at the source-end of the channel. a_D is the flux of carriers which have not been backscattered. The drain flux is neglected for large drain voltages.

1.2.1 Carrier fluxes

In the Lundstrom's model the current is described by carrier fluxes moving through the channel. The model does not use the concept of mobility. On the contrary it defines *transmission* and *reflection* coefficients to study the transport along the channel. Source and drain are reservoirs of carriers at thermal energy which are injected into the channel. The total current I_{DS} is the difference between the carrier flux going to the drain (positive flux) and the one towards the source (negative flux)[13]. Figure 1.4 depicts this approach.

Low drain voltage. At low V_{DS} both the positive and the negative fluxes are important and the carriers are close to equilibrium. If no scattering is present inside the channel, every carrier entering from one side reaches the opposite channel-edge without any perturbation. If the channel is not ballistic, transmission and reflection coefficients must be defined to account for scattering. From the potential profile (refer again to Fig. 1.4) along the device, it is possible to recognize the top of the potential barrier at the source-end side of the channel, which is called *injection point* x_{inj}. At the injection point, the parallel field is zero and the Gradual Channel Approximation can be applied[15]. Thus, in this point, the inversion charge density is entirely controlled by the oxide capacitance and the gate voltage, as in a MOS capacitor. If we observe the distribution function at the injection point, we will find a symmetrical profile, similar to the equilibrium distribution function,

as shown in the left part of Fig. 1.5. The positive half of the distribution function is due to carriers injected from the source to the channel (carriers with velocity $v > 0$), while the negative half is due to carriers injected from the drain (with velocity $v < 0$) which have reached the source-end. In near-equilibrium condition the average velocity v_{avg} of the carriers at the top of the source potential barrier is nearly zero. On the contrary the average velocity of the *injected* carriers is *not* zero and it is equal to the *injection velocity* v_{inj}, which is the average velocity of the carriers with positive values of v (right half of the distribution function in Fig. 1.5).

For non-degenerate electrons in silicon at T=300 K, v_{inj} is approximately 1.2×10^7 cm/s. In the degenerate case v_{inj} is larger and depends on the inversion charge density.

Large drain voltage. At large V_{DS} the drain-to-source flux can be neglected because no carrier can overcome the large potential barrier from drain to the source. Any negative contribution to the total current is provided only by backscattering, which redirects a small portion of the injected carriers to the source. Therefore the negative half of the distribution function can either disappear or present a reduced tail. On the contrary the positive half increases. Indeed, it is important to remember that the top of the injection point features a 1D electrostatic control, dependent only of the gate voltage and not of the parallel field. The inversion charge at this point, is then given by $C_{OX}(V_{GS} - V_{TH})$, a well-known result of the theory of the MOS capacitor. From the point of view of the carrier distribution, the inversion charge at $x = x_{inj}$ is given by carriers moving with both positive and negative velocities. The total charge is calculated by integrating both halves of the distribution function. If there is not any negative flux, as in the case of a full-ballistic channel, the density of carriers with $v > 0$ is increased to maintain the same inversion charge [15]. This electrostatic effect is shown in the right part of Fig. 1.5.

1.2.2 Current under quasi-ballistic condition

We consider only the case of a large drain bias, when quasi-ballistic transport occurs. The flux of carriers transmitted across the source barrier into the channel is marked a_S. A fraction t_C of a_S reaches the drain while a fraction $r_C = 1 - t_C$ is *backscattered* from the channel and reenters the source (as

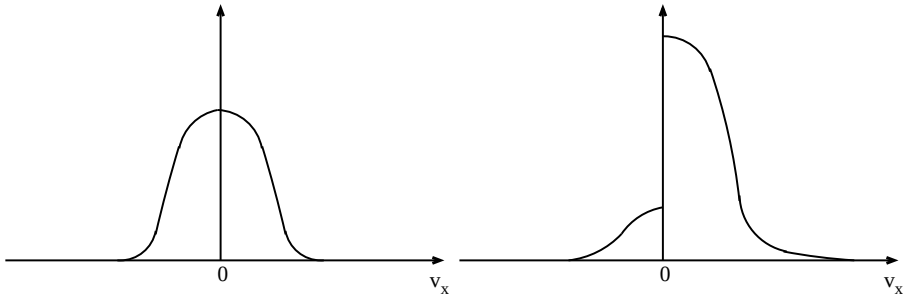

Figure 1.5: Distribution function of carriers at the top of the source-to-channel barrier. Left plot is the quasi-equilibrium distribution, valid for low V_{DS}. The inversion charge is found from the integration over all values of x. Right plot is the function shape for which the negative half is given by backscattered carriers. In order to maintain the same inversion charge as in the left-side figure, the positive half is increased. If the channel is ideally ballistic, the negative tail disappears. The injection velocity v_{inj} is calculated as the average velocity of carriers in the positive half only.

shown in Fig. 1.4). These backscattered particles have velocity $v < 0$ and they give a negative contribution to I_{DS}. The flux reaching the drain is then

$$a_D = t_C a_S \tag{1.5}$$

The inversion charge density at the injection point x_{inj} is [14]

$$n(x_{inj}, y) = \frac{a_S + r_C a_S}{v_{inj}} \tag{1.6}$$

where y is the axis normal to the interface. The flux entering the drain is then

$$a_D = n(x_{inj}, y) v_{inj} \frac{t_C}{1 + r_C} \tag{1.7}$$

The current I_{DS} is found by integrating eq. (1.7) over y. We now remind that the inversion charge Q_i at x_{inj} is

$$Q_i = \int_0^{y_{max}} n(x_{inj}, y) dy = \frac{C_{OX}}{q}(V_{GS} - V_{TH}) \tag{1.8}$$

where y_{max} is the border of the depletion region. The combination of eq. (1.7) and (1.8) provides the drain current I_{DS}

$$I_{DS} = W C_{OX} v_{inj} \frac{1 - r_C}{1 + r_C}(V_{GS} - V_{TH}) = \frac{1 - r_C}{1 + r_C} I_{DS,BL} \tag{1.9}$$

where W is the width of the MOSFET and $I_{DS,BL}$ is the current in the full-ballistic case. The average carrier velocity at $x = x_{inj}$ is $v_{avg} = v_{inj}(1 -$

$r_C)/(1 + r_C)$. The injection velocity v_{inj} is equal to the thermal velocity v_T, calculated by Natori [12]. In the non-degenerate gas $v_T = 1.2 \times 10^7$cm/s. In the case $r_C = 0$, the transport is full-ballistic and the average velocity equals the injection velocity because the distribution function at x_{inj} has only the positive half. In order to use eq. (1.9), a viable expression of r_C must be found. The Lundstrom's model suggests to evaluate the backscattering coefficient as [14]

$$r_C = \frac{L_{KT}}{L_{KT} + \lambda} \tag{1.10}$$

where λ is the mean free path of the carriers near the source-end of the channel and L_{KT} is the distance over which the potential drops by kT/q from the top of the source-channel barrier (refer again to Fig. 1.4). This distance is called *critical distance* or *kT-layer*. The model states that if an electron is injected from the source and do not encounter any scattering event in the kT-layer, it cannot be backscattered to the source in any other position inside the channel.
The expression (1.10) is very useful because it relates the low field mobility (through λ) and the lateral electric field near x_{inj} (through L_{KT}). If the mobility increases, λ increases and the backscattering coefficient is reduced, approaching the ballistic limit [16]. This result is also obtained by raising the lateral electric field near the injection point but this effect can be a result of strong short-channel effects and it is not desired. Even if the model treats the quasi-ballistic transport, the concept of low field mobility remains useful to evaluate the distance from the fully-ballistic device.

1.3 SOI devices for the decananometric length: SOI Single-Gate and Double-Gate MOSFETs

The Silicon-On-Insulator technology is a viable alternative to the mainstream Bulk architecture for the next technology nodes. In a SOI wafer the silicon active region, where devices are patterned, is separated from the substrate by a thick oxide region. A MOS transistor designed on a SOI wafer has a cross-section like the one in Fig. 1.6. The main differences with the standard MOS of Fig. 1.1 are:

- the vertical isolation protects the active region from many parasitic effects like radiation-induced photo-currents and latch-up;

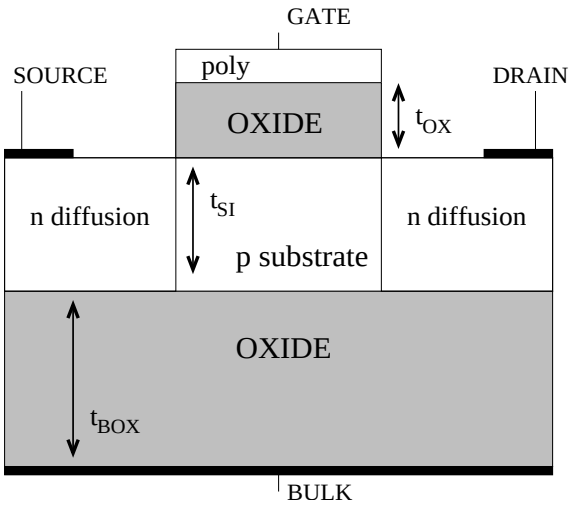

Figure 1.6: Schematic structure of a Single-Gate SOI MOSFET. A thick back-oxide region separates the active area from the bulk contact, which, for SOI devices only, is often called back-gate contact.

- the isolation reduces the parasitic capacitances and leakage currents of the pn junctions.
- the lateral inter–device isolation in the SOI case does not need trenches or well formation;

Starting from an SOI wafer, many different types of MOS transistor can be obtained. The first classification is between the Single-Gate SOI MOSFET, like the one in Fig. 1.6, and the Multi-Gate SOI MOSFET, shown in Fig. 1.7. While the former has found a commercial application, the latter is still confined in the research area and will be described in a specific subsection.

The Single-Gate SOI MOSFETs can be further classified into two categories:

- *Partially-Depleted (PD)*: if the active silicon body is thick enough to contain completely the depletion region in strong inversion;
- *Fully-Depleted (FD)*: if the fin thickness t_{SI} is so thin that the depletion region touches the bottom of the silicon body. Also the depths of the source and drain diffusions are limited by t_{SI}.

When the SOI process was presented [17], its higher cost limited its diffusion to the applications where radiation hardness was a main problem. With the continuous technological development the IC designers were founding increasing issues in parasitic control and power consumptions. At the same

time the research on SOI technology has increased exponentially and SOI has become an interesting alternative to the standard devices because of the reduced parasitic effects.
At the moment all the integrated circuits on SOI wafers use the Partially-Depleted structure because the technological process to create a "thick" body is much easier. The design rules of a PD-SOI MOS transistor are not very different from the bulk case and the scaling rules are almost the same, for example the doping concentration of the body must be increased while decreasing the gate length L_G to control the SCE.

On the contrary the FD-SOI MOSFET is quite different from the bulk transistor and its behavior will be described in the next section.

1.3.1 The Fully-Depleted SOI MOSFET

The term Fully-Depleted means that the body thickness t_{SI} is smaller than the depletion layer width in inversion. The FD concept is applied not only to Single-Gate but also to Multiple Gates devices as it will be shown in the next section. The parameter t_{SI} is critical to determine the behavior of this type of MOSFET.

The main issues about the use of an ultra thin silicon body are

1. because the depletion layer width is inversely proportional to $\sqrt{N_a}$, a large doping concentration cannot be used without losing the fully-depleted behavior. Unfortunately, N_a must increase with scaling or the short-channel effects will become relevant;

2. the last term of Eq. (1.3) is proportional to the depletion charge density. Due to the small t_{SI}, this charge concentration is negligible and the threshold voltage is low, unless N_a is further increased;

3. if we use a very large N_a to satisfy points 1 and 2, the mobility will be severely degraded, as seen in section 1.1.3. Moreover the device becomes sensible to impurity fluctuations.

In order to avoid all this problems, FD Single-Gate MOSFETs often have an almost undoped body ($N_a \sim 10^{15}\text{cm}^{-3}$). The threshold voltage can be set to the appropriate value by changing V_{FB}. This requires metal gate with midgap workfunction instead of the standard polysilicon gate contacts. The integration of metal gates in the MOS process is currently under research [19].

The penetration of the drain electric field, and then the SCE, is detrimental in short-channel devices with undoped body, but can be limited if the t_{SI} thickness is chosen following some special scaling rule. Many $L_G - t_{SI}$ relations have been proposed [18], [20]. Following [20], the body thickness should be chosen to have $6\delta > L_G > 8\delta$, where δ is

$$\delta = \sqrt{\frac{\epsilon_{si} t_{SI} t_{ox}}{\epsilon_{ox}}} \tag{1.11}$$

It can be shown that, for gate length shorter than 50 nanometers, a body thickness smaller than 10 nm is required. The creation of such an ultra small t_{SI} is still a major technological problem because of the difficult reproducibility of silicon fins with the same small thickness.

On the other hand the FD Single-Gate MOSFET is attractive for many reasons:

1. if the active area is kept undoped, the mobility is very large. Unfortunately the use of thin bodies enhances the source/drain series resistances, then compensating the advantage of large μ_{EFF};

2. the back-oxide thickness t_{box} is an important parameter to set the subthreshold behavior. It is important to have $t_{box} \gg t_{si,ox}$ or the subthreshold slope can be too large [21]. It can be seen that this requirement enhances self-heating in the device [22];

3. coupling effects between the front gate and the back gate arise [23]. This argument will be developed in Part II.

The Fully-Depleted operating mode has found an interesting application in the case of Multi-Gate devices because the presence of more gate contacts relaxes the $L_G - t_{SI}$ relation and permits the use of thicker bodies.

1.3.2 The Double-Gate SOI MOSFET

The advancement in the MOS technology has allowed the creation of devices with multiple gate contacts, like the Double-Gate MOSFET or the Trigate [24]. All these devices take advantage of the fully-depleted behavior. Thanks to the combination of thin t_{SI} and multiple gate contacts, the control of the body by the gate voltage is greatly enhanced and the short-channel effects are reduced in comparison with single-gate devices. The Double-Gate SOI MOSFET is the most common example of multi-gate transistor and the research about it has increased in the last years.

Depending on the technology, the double-gate structure can be designed in many different ways. The FinFET type [25], which is plotted in Fig. 1.7, is the most promising one even if its technological process is rather difficult due to the non-planar shape of the device.
Because the single-gate SOI and the Double-Gate SOI will be addressed many times in the rest of this work, we will sometimes refer to the former with SG and to the latter with DG.

Like the Single-Gate device, the silicon thickness t_{SI} is a dominant parameter in the Double-Gate MOSFETs. Anyway the relation $L_G - t_{SI}$ is more relaxed than the SG case. For example, following again [20], we can rewrite the expression (1.11) as

$$\delta = \sqrt{\frac{\epsilon_{si} t_{SI} t_{ox}}{2\epsilon_{ox}}} \tag{1.12}$$

Comparing (1.12) with (1.11), it can be demonstrated that, for the same gate length and oxide thickness, the DG transistor can have a larger t_{SI} than the SG case. This is positive from the point of view of the technological process and guarantees a better scalability. Moreover the thicker body improves the series resistances that are detrimental in the short-channel SG devices.
The immunity to SCE can be degraded by the penetration of field lines from the drain to the BOX. These fields enhance a parasitic current in the bottom part of the film, without any control from the gate bias. We can usually ignore this issue if the the vertical dimension H_{fin} of the silicon fin is: $H_{fin} \gg t_{SI}$. In this way the parasitic current is concentrated in a negligible portion of the the entire fin.

From the point of view of the current drive, the double-gate transistors can exhibit an I_{DS} and transconductance about twice the values of a SG transistor with the same dimensions, because of the two channels in parallel. The DG transistor also has some peculiar mobility enhancements:

- *Volume Inversion* is an increase in the mobility when the channel is in moderate inversion [26]. The name comes from the fact that the inversion charge is not concentrated at the two Si/SiO_2 interfaces but is distributed over the entire fin width. This a well-known advantage of the double-gate configuration;

- higher surface roughness limited mobility [27]. In fact, for fixed inversion charge, the DG case features a smaller effective field than the SG case.

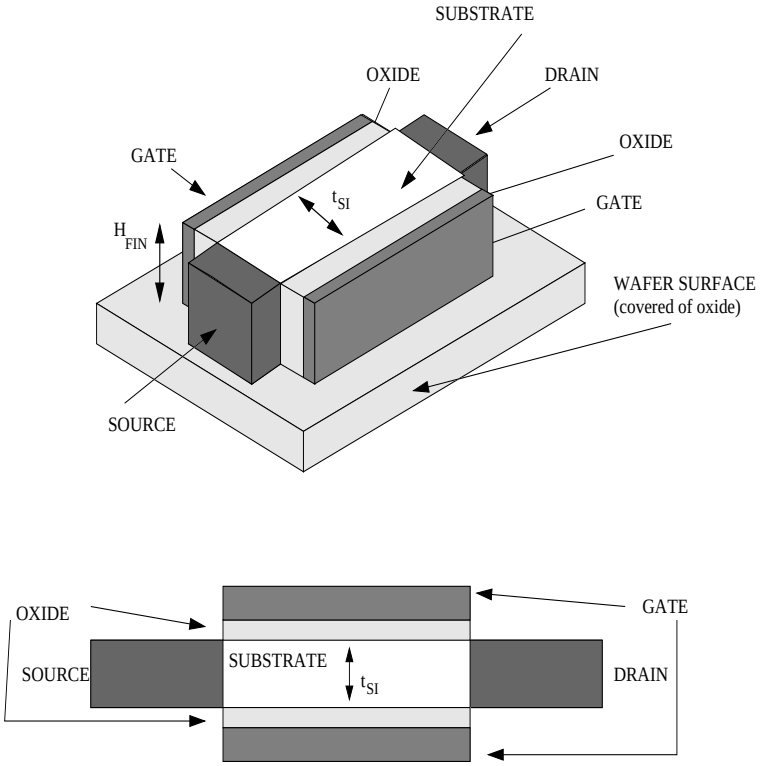

Figure 1.7: Top: 3D schematic structure of a Double-Gate SOI MOSFET realized with the FinFET technology. Bottom: simplified 2D section of the same device, where the back oxide is not considered. In both figures, the bulk contact (or substrate contact) under the back oxide is not shown.

The FinFET device has still many technological problems to solve, mainly related to its non-planar process and layout. The vertical dimension plays the role of the transistor width W. Because it is not possible to build thin fins taller than about 100 nm, many FinFETs must be connected in parallel to obtain a relevant current. While the research is advancing, the DG technology is more and more optimized. Many solutions under study for advanced SG devices, like metal gates, are going to be used in the DG case too.
According to the last ITRS predictions [5], the Double-Gate technology will probably enter production in 2012.

Chapter 2

Simulation techniques for the deca-nanometric MOSFET

The simulation of a silicon device is not an easy task to perform, expecially in the short-channel regime in which parasitic effects must be accounted for. Here we will not review all the possible simulation approaches but we will concentrate only on the two methods adopted in this work: the drift-diffusion model and the Monte Carlo method. First, the Boltzmann transport equation will be introduced. This is a master equation for the simulation of a silicon transistor. Since this equation is very difficult to solve, we will derive the famous drift-diffusion model, which is the basis of most of commercial simulation tools. Finally we will describe the Monte-Carlo approach, which will be the main simulation method used in the following chapters.

2.1 The Boltzmann Transport equation and its approximate solution

In order to evaluate the behavior of the carriers inside a semiconductor device, we need to compute the distribution function $\mathcal{F}(\vec{r}, \vec{p}, t)$. This function, which depends on the carrier position, carrier momentum and time, describes the probability to find an electron/hole in a certain position $\vec{r}$ with momentum $\vec{p}$ at the instant t. The function $\mathcal{F}$ is the solution of the *Boltzmann Transport Equation* (BTE)

$$\frac{\partial \mathcal{F}}{\partial t} = -\nabla_{\vec{r}} \cdot \left(\frac{d\vec{r}}{dt} \mathcal{F} \right) - \nabla_{\vec{p}} \cdot \left(\frac{d\vec{p}}{dt} \mathcal{F} \right) + \left(\frac{\partial \mathcal{F}}{\partial t} \right)_C \tag{2.1}$$

The Boltzmann equation represents a charge balance inside an elementary volume in the $(\vec{r}, \vec{p})$ space. The first and the second term in the right-hand side are the net flux of $\mathcal{F}$ in the $\vec{r}$ and $\vec{p}$ space, respectively. The third term describes collisions due to scattering events (already introduced in sec. 1.1.3) that modify the carrier motion. The collision term can be expressed as:

$$\left(\frac{\partial \mathcal{F}}{\partial t}\right)_C = \int_{\vec{p}} \left[\mathcal{S}\left(\vec{r}, \vec{p}', \vec{p}\right) \mathcal{F}\left(\vec{r}, \vec{p}', t\right) - \mathcal{S}\left(\vec{r}, \vec{p}, \vec{p}'\right) \mathcal{F}\left(\vec{r}, \vec{p}, t\right)\right] d\vec{p}' \tag{2.2}$$

where $\mathcal{S}(\vec{r}, \vec{p}, \vec{p}')$ is the probability of the collision event.

The Boltzmann transport equation is valid within the *semi-classical* approach, which assumes a classical description of the particle while the scattering form is calculated by quantum mechanics. The closed-form solution of the BTE is particularly difficult for the case of simple device geometries, mainly due to the form of the collision term.

There are two main methods for the solution of eq. (2.1)

1. Approximated methods, where a set of simpler equations is derived from BTE and then solved;

2. Direct methods, which need complex numerical calculations.

The first approach includes the well-known "moments method" which is described next. The second one includes the statistical Monte Carlo approach.

2.1.1 The moments method

The moments method is based on:

- Reduction of the number of dimensions of the unknown variables. For this purpose the function $\mathcal{F}$ is replaced by its statistical moments.

- Strong approximation of the collision term which is described by a single parameter τ. This parameter represents the characteristic time needed by the system to return to equilibrium. Once τ is defined, the last term of the BTE is written as

$$\left(\frac{\partial \mathcal{F}}{\partial t}\right)_C = \frac{\mathcal{F}_{eq} - \mathcal{F}(\vec{r}, \vec{p}, t)}{\tau} \tag{2.3}$$

  with $\mathcal{F}_{eq}$ is the distribution function in the equilibrium condition.

The dependence of the momentum $\vec{p}$ is eliminated by evaluating the statistical moments of the distribution function up to a given order.

Zero-order moment $n(\vec{r},t)$ is the number of carrier in a volume $d\vec{r}$ at a certain instant t. It is given by

$$\int_{\vec{p}} \mathcal{F}(\vec{r},\vec{p},t)d\vec{p} = n(\vec{r},t) \tag{2.4}$$

First-order moment $< \vec{v} >$ is the average velocity of the carrier population obtained by averaging the group velocity according to

$$\frac{\int_{\vec{p}} \vec{v}_G \mathcal{F}(\vec{r},\vec{p},t)d^3p}{\int_{\vec{p}} \mathcal{F}(\vec{r},\vec{p},t)d^3p} =< \vec{v} > \tag{2.5}$$

Second-order moment $< \vec{v}^2 >$ is the mean squared velocity, related to the kinetic energy of the carriers, defined as

$$\frac{\int_{\vec{p}} v_G^2 \mathcal{F}(\vec{r},\vec{p},t)d^3p}{\int_{\vec{p}} \mathcal{F}(\vec{r},\vec{p},t)d^3p} =< v^2 > \tag{2.6}$$

Higher-order moments can be obtained but the most common methods for solving the BTE consider only low-order functions. In particular the *drift-diffusion* model uses only the moments above and we will discuss it in the next section.

2.1.2 The Drift-Diffusion model

The drift-diffusion model (DD) is the most popular approximated method to solve the Boltzmann transport equation. It is widely diffused in commercial device simulation tools.
In modern devices the quantum effects and many non-local effects (like hot-carriers and velocity overshoot) are difficult to track with the approximated methods. However the drift-diffusion model can be calibrated to improve its accuracy in the short-channel regime. This characteristic, together with its robustness and efficiency, explains the success of the DD model.

The drift-diffusion model involves three variables: electron concentration, hole concentration and electric potential. The mathematical system to solve includes five equations which are briefly described in the following paragraphs, without including the mathematical passages for deriving them from the BTE.

The Poisson equation

It is the simplest equation of the drift-diffusion model and is the master equation of any electrostatic problem. For a silicon volume with both a donor dopant (concentration N_d) and acceptor dopant (concentration N_a), the equation is expressed as

$$\nabla^2 \Psi = -\frac{\rho}{\epsilon} = -\frac{q}{\epsilon}(N_d - N_a + p - n) \tag{2.7}$$

where ϕ is the electric potential, p is the hole concentration and n is the electron concentration.

Charge continuity equations

The charge continuity equation for electrons is

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \vec{J_n} + \mathcal{G} - \mathcal{R} \tag{2.8}$$

where $\vec{J_n} = -qnv_{avg}$ is the electron current density, v_{avg} is the average velocity. The last term, $\mathcal{G} - \mathcal{R}$, substitutes the collision term of the BTE, and is the difference between the generation and recombination function. These functions represent the electron-hole pairs that are generated and recombinated in the volume unit and time unit.
In the case of holes, the charge continuity equation is

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \vec{J_p} + \mathcal{G} - \mathcal{R} \tag{2.9}$$

Current density equations

Equations (2.8) and (2.9) require two additional constitutive relations for the density currents $\vec{J_n}$ and $\vec{J_p}$. Here we consider a simple steady-state 1D case, where the density currents can be written as

$$\frac{q\mathcal{E}_x}{m^*}\frac{d\mathcal{F}}{dx} + v_x\frac{d\mathcal{F}}{dx} = \frac{\mathcal{F}_{eq} - \mathcal{F}(v_x, x)}{\tau} \tag{2.10}$$

where $\mathcal{E}_x$ is the electric field along the x axis and the the generation-recombination term is expressed using approximation (2.3). The current density can be expressed as

$$J(x) = q\int v_x \mathcal{F}(v_x, x)dv_x = q\frac{q\tau}{m^*}\mathcal{E}_x n(x) - q\tau\frac{dn}{dx} < v^2 > \tag{2.11}$$

In eq. (2.11), the second-order moment is evident in the last term. This quadratic dependence on v_x is simplified by approximating the average kinetic energy with the average thermal energy. This approximation assumes that carrier temperature is in equilibrium with the silicon lattice, and no hot-carriers effects are present. Thus, for the 1D carrier gas, we obtain

$$\frac{1}{2}m^* < v^2 >= \frac{1}{2}k_B T \quad \Rightarrow \quad < v^2 >= \frac{k_B T}{m^*} \tag{2.12}$$

Then we introduce the concept of mobility

$$\mu = \frac{q\tau}{m^*} \tag{2.13}$$

and the *diffusion coefficient* by using the Einstein's relation

$$D = \frac{\mu k_B T_o}{q} \tag{2.14}$$

The final expressions of the electron/hole current densities are

$$J_n = qn(x)\mu_n \mathcal{E}(x) + qD_n \frac{dn}{dx} \tag{2.15}$$

$$J_p = qp(x)\mu_p \mathcal{E}(x) - qD_n \frac{dn}{dx} \tag{2.16}$$

The drift-diffusion model is composed of equations (2.7), (2.8), (2.9), (2.15) and (2.16). This set of equations must be solved over the entire MOSFET to obtain the potential ϕ and the charge densities n and p.

2.1.3 An example of commercial simulation tool: *Dessis*

Dessis is a multidimensional, mixed-mode device and circuit simulator for one-, two-, and three-dimensional semiconductor devices [28]. It is part of the large TCAD tool suite for simulation in the field of semiconductor technology. It incorporates advanced physical models and robust numerical methods for the simulation of semiconductor devices with different levels of accuracy. DESSIS computes the electrical behavior of a semiconductor device by solving a set of physical device equations that describes the carrier distribution and conduction mechanisms.
A real semiconductor device, such as a transistor, is represented in the simulator as a "virtual" device whose physical properties are discretized onto a non-uniform grid (or mesh) of nodes. In order to simulate, for example, a MOSFET it is necessary to create a numerical representation and gridding

as more accurate as possible. The second critical point is the choice of the physical models to use during the calculations: for this purpose DESSIS has a very large choice of models.
DESSIS implements the drift-diffusion equations as the standard method for the simulation of a semiconductor device. Even if other sets of equations, like the hydrodynamic one, are available, we prefer the DD model for its robustness and reliability. All the simulations that we have performed with DESSIS, include other physical models to improve the accuracy:

- the electron/hole mobility is described by the model from [29], which has been developed at the University of Bologna. All the main scattering mechanisms, which limit the mobility in a short-channel MOSFET, are included in this model.
- the Density Gradient Model [30] is used for describing the quantization of the inversion charge.
- the saturation of the carrier mobility at large electric fields is taken into account by the Canali model [31].

In this thesis, DESSIS has been mainly used to:

- create the initial condition of our Monte-Carlo simulations (see sec. 2.2.3);
- simulate the subthreshold current I_{OFF}, which is an important parameter for the design of our template devices;
- make a comparison between the results from the drift-diffusion approach and the result from the Monte-Carlo approach.

Further details about the DD simulations performed in this work, in particular about the device structures to analyze, are described in each related section.

2.2 The Monte Carlo simulation of silicon devices

In the previous section, we introduced the Boltzmann Transport Equation for calculating the electron (hole) distribution function. The BTE is difficult to solve and a solution is usually found by using the approximated methods. An

alternative way is based on direct methods, like the Monte Carlo approach, based on statistical calculations. Historically the MC simulators were used to study high-field effects in MOS transistors. Now this simulation approach is still used in short-channel devices because it can describe the transport within the channel more accurately than the drift-diffusion model. Quasi-ballistic transport, complex scattering models and carrier energy distribution can be properly simulated with the MC method.
In this section we discuss the Monte-Carlo (MC) simulation applied to silicon devices. It can be shown that the Monte Carlo method is a rigorous solution of the Boltzmann equation, but we will not demonstrate it. On the contrary we will dedicate more space to describe in a simple way how a MC simulator works. The argument is developed here in a way similar to [32], due to its simplicity. A rigorous treatment can be found in [33].

2.2.1 Basic knowledge of a Monte Carlo Simulation

Simply speaking, the Monte Carlo method simulates the motion of all particles inside a lattice, under the effect of an electric field. The carrier is subject to a force given by the presence of an electric field

$$\vec{F} = \frac{d\vec{p}}{dt} = (-q)\vec{\mathcal{E}} \tag{2.17}$$

The force $\vec{F}$ accelerates the carrier, thus increasing its energy E and momentum $\vec{p}$. The evolution of the movement is calculated over the time t. At the same time, also the position $\vec{r}$ of the electron changes with t. The movement is stopped at a certain instant by a perturbation (also called scattering event or collision). The time between two perturbations defines the duration t' of the *free-flight*, during which eq. (2.17) is integrated for evaluating the variation of $\vec{p}$. The duration of a scattering event is usually negligible with respect to the flight duration, so it is considered istantaneus.
At the end of a free-flight, the position and momentum of the particle are updated. For example, a carrier, that has moved in a three-dimensional lattice under the influence of a constant electric field directed along the z axis, has its position and momentum changed according to

$$p_x(t') = p_x(0) \tag{2.18}$$
$$p_y(t') = p_y(0) \tag{2.19}$$
$$p_z(t') = p_z(0) + (-q)\mathcal{E}_z t \tag{2.20}$$

$$x(t') = x(0) + \frac{p_x(0)}{m^*}t \tag{2.21}$$

$$y(t') = y(0) + \frac{p_y(0)}{m^*}t \tag{2.22}$$

$$z(t') = z(0) + \left(\frac{E(t) - E(0)}{(-q)\mathcal{E}_z}\right) \tag{2.23}$$

In eq. (2.21), (2.22), (2.23) we have used a parabolic relation between the energy E and the momentum $\vec{p}$

$$E(t) = \frac{p^2(t)}{2m^*} \tag{2.24}$$

where m^* is the effective mass of the electron. The E–p relation, called dispersion relation, takes into account the energy bandstructure of the silicon lattice and, because of its complexity, it is often approximated by an analytical expression. The parabolic expression (2.24) is commonly used for this purpose but a MC code can include more complicated models, not only analytical but also based on look-up tables.
The determination of the free-flight t' is a very difficult task because it depends on the scattering rate Γ which is the frequency of collisions: the higher the frequency, the shorter the time $\tau = 1/\Gamma$ between two collisions, which is the duration of the free flight. The rate Γ is a function of the scattering probability $\mathcal{S}$, which always depends on the final energy $E(t')$.

If more sources of perturbation, like phonons, impurities, etc, are involved, the total collision rate is

$$\Gamma(E) = \sum_{i=1}^{k} \Gamma_i(E) = \sum_{i=1}^{k} \frac{1}{\tau_i(E)} \tag{2.25}$$

where the sum is done over all sources of scattering. In order to overcome the difficult calculation of t', a constant Γ scheme is adapted. For a given constant rate $\Gamma = \Gamma_0$, it can be shown that the duration of the free-flight is given by

$$t' = -\frac{1}{\Gamma_0}\ln(r_c), \tag{2.26}$$

where r_c is a casual number uniformly distributed between 0 and 1.
Once t' is found from eq. (2.26), the simulation of the particle free-flight can proceed as follows:

1. the movement of the particle is evaluated from equation (2.17) while the final momentum and position are updated using expressions similar to (2.18)–(2.23);

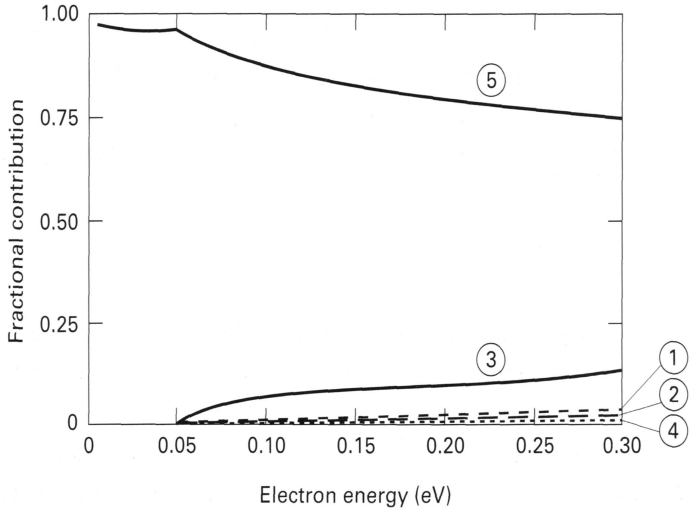

Figure 2.1: Example of contribution, normalized over Γ_0, to the total scattering rate from different scattering sources as a function of energy. The dependence on energy varies from one scattering mechanism to another one. In the figure, curves labeled from 1 to 4 are various real scattering mechanisms, the curve labeled 5 is self-scattering. Picture taken from [32].

2. we need to find which source of scattering has caused the end of the free flight. From the knowledge of $E(t')$, we can calculate $\Gamma_i(E(t')) = 1/\tau_i(E(t'))$ for each type of scattering, as shown in Fig. 2.1. The total scattering rate, given by eq. (2.25), must be lower than Γ_0;

3. the rates Γ_i are normalized by Γ_0. This will result in a scale of probability from 0 to 1 where each type of collision has a certain probability to happen. Fig. 2.2 reports an example of probability scale;

4. because the total rate Γ is less than Γ_0, a certain range of the probability scale does not correspond to any type of collision. This is treated as a new scattering mechanism, denoted as *self-scattering*;

5. a random number is mapped on the probability scale and identifies the collision event that stopped the flight. If a real collision has happened, the state $(E, \vec{p})$ of the particle is updated following the model of the right scattering source. If self-scattering has happened, the state of the particle remains the same as at the end of point 1.

The cycle is repeated over time and the state of the particle is recorded to create the statistics of energy, velocity, etc. The simulation can be stopped

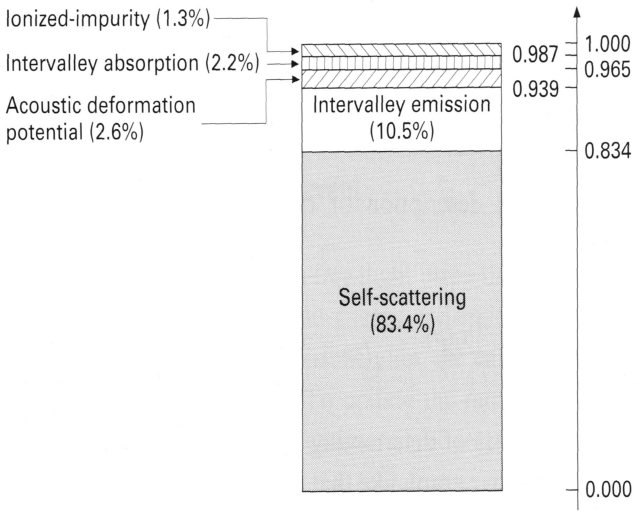

Figure 2.2: Probability scale, from 0 to 1, of scattering events at a fixed energy, which is equal to the final energy $E(t')$ at the end of the free-flight. The figure is built on the values from Fig. 2.1. A random number defines the scattering mechanism which has stopped the flight. Picture taken from [32].

after a reasonable number of steps or after the convergence has been achieved.

2.2.2 Ensemble Monte Carlo

In the previous section we have seen an example of cyclic algorithm to simulate one particle in a semiconductor lattice under the influence of an electric field. In a device many particles are present and must be simulated in the proper way to describe the complete behavior of the device. In general there are two types of Monte-Carlo simulation for semiconductor devices: *ensemble* and *incident flux* approach. While the former calculates the trajectories of all particles at the same time, the latter simulates one particle for a certain time, builds up the statistic and then considers another particle. The first approach is the most popular and we discuss only it.

In the ensemble MC method, the two dimensional MOSFET is divided into cells by a numerical grid. Each cell is populated with carriers with a given charge *weight*, that is the quantity of charge associated to each particle. In fact, because the number of particles in a real devices is very large, only a smaller number can be handled by the MC code. This number must be

representative of the entire carrier population. Each simulated carrier has not an elementary charge q, like electrons or holes, but has a charge $Q = kq$, where k can be different from one particle to the other. The new *superparticle* represents the charge of many carriers. Together with k, it is necessary to set an initial energy and momentum for each created particle. This assignment can be done by randomly choosing a sample from a Maxwellian or Fermi-Dirac distribution. As a final step the carriers must be distributed over the grid in a "smart" way. The choice of the carrier distribution and of the initial electrostatic potentials is usually based on a known solution, which is the initial condition of the MC simulation.
At this point, the simulation can start: the momentum, energy and position of all particles are traced by the techniques discussed in sec. 2.2.1. Poisson's equation must be solved after each simulation step to update the electric field. At any time during the simulation the average carrier density, velocity, energy versus position can be computed by averaging over the particles within each cell. The whole process is repeated until numerical convergence is achieved.

This type of simulation approach has many problems, like the treatment of the boundary conditions and the rules for creating the grid. The number of particles involved in a single simulation can be very large and limits the computational efficiency. In particular the choice of the scattering events and the calculation of the final state for all particle after each flight can add a lot of simulation time. This last issue is the drawback of the possibility to include very complex models of the scattering events. This is an important characteristic of the MC approach, that does not reduce the entire analysis of the collisions to a single mobility value, like the drift-diffusion model. This observation is a key point to understand why the Monte Carlo is useful to study the transport in MOSFETs with very short channel length. In particular the quasi-ballistic transport regime cannot be accounted properly in a drift-diffusion simulation, while MC can handle particles that experience few collision events within the channel.
Another key advantage is the inclusion of realistic models for the dispersion relation $E - \vec{p}$. Because this characteristic permits a proper treatment of carriers with large energies and velocities, the analysis of transport highly out-of-equilibrium is one of the best field of application of the MC method.

2.2.3 *Bandit*: an example of Monte Carlo tool

The results from MC simulations, reported in chapters III, IV and V, have been obtained using a MC code called *Bandit*, which is presented in this sec-

tion.
Bandit is a Full-Band self-consistent Monte-Carlo simulator for the 3D electron gas with corrections to the electrostatic potential in order to include the effect of carrier quantization on the spatial distribution of the inversion charge [34]. Quantum-mechanical corrections are introduced by the effective potential approach proposed in [35] and its implementation in the MC code is described in [34]. The effective potential is defined as:

$$V_{eff}(x,y) = \int\int V(x',y')G(x'-x,y'-y)dx'dy' \tag{2.27}$$

where the potential energy profile $V = -q\Psi + \chi$, including both the electrostatic potential (Ψ) and the electron affinity (χ), is smoothed by a Gaussian function $G(\xi,\zeta)$; the standard deviation of the Gaussian is chosen in order to reproduce the inversion charge density of a coupled Schrödinger-Poisson solver even in devices with very thin gate oxides and over a wide range of voltages. The initial solution of the MC analisys, is calculated from the solution of a DESSIS drift-diffusion simulation of the same device structure (see sec 2.1.3).

Besides the phonon scattering the MC code includes a model for ionized impurities scattering, which follows the usual 3DEG formalism. Electron-plasmon scattering inside the heavily doped regions is also included. This scattering mechanism plays an important role because it thermalizes the particles in the source and drain regions. The carrier-plasmon interaction is a very strong inelastic scattering and has to be included when simulating quasi-ballistic transport, since the amount of back-scattered carriers depends on the balance between elastic and inelastic scattering.
Finally a model for the surface roughness (SR) scattering is necessary, since the mobility of MOSFETs in the "on state" is limited by this scattering mechanism. In MC simulators, surface roughness scattering is usually modeled with a specular-diffusive reflection of the particles hitting the Si/SiO_2 interface, and the percentage of diffused particles is adjusted to fit experimental data [36]. However, the effective potential repels the carriers from the surface and almost none of them can reach the interface; therefore the specular-diffusive approach cannot be used. As a consequence, surface roughness must be included as an additional scattering mechanism.
Bandit features an original approach to adapt the SR scattering model for a 2D electron gas to the full-band 3D electron gas MC corrected by the effective potential. The scattering rate is calculated from the effective field $\mathcal{E}_{EFF}$ since, as documented in [7], the experimental mobility is an unique function of $\mathcal{E}_{EFF}$. The details of this model can be found in [37].

This tool suffers limitations as it does not account for the 2D-gas subband structure and for its effects on the phonons and surface-roughness scattering rate. Nonetheless, since the deformation potentials for acoustic phonons and the parameters for SR-scattering have been adjusted in order to fit the mobility curves of both bulk and SOI devices [7, 27, 37], reasonable accuracy in terms of terminal-currents can be expected. Furthermore, scattering mechanisms that assume an increasingly important role as t_{si} is scaled below 10 nm such as surface optical phonons and the effects of body-thickness fluctuations [38] are not included. For this reason, the simulated current may be overestimated for ultra-thin body SOI MOSFETs. In [39] the results obtained by the MC simulator adopted in this work have been compared with those of a MC simulator for a 2D confined electron gas that explicitly accounts for the effects of quantization on the dispersion relation and on the scattering rates [40]. The results of this comparison confirm that the simulation approach adopted in this work provides terminal currents in good agreement with the more accurate simulator for the 2D electron gas, at least for the devices of interest in this work.

The Bandit code implements a RF analysis for the calculation of: the drain-source delays, the Y matrix and the small-signal parameters [41]. This last feature will be described in details in chapter 5.

Chapter 3

Study of Quasi-Ballistic Transport in decananometric MOSFET

The accurate modeling of devices with gate length L_G shorter than 25 nm must account for both quantization and far from equilibrium, non-local transport. Many different transport models have been used to simulate the current in this scaling regime. The approach to the ballistic limit has suggested the use of full-ballistic simulators, that neglects scattering, but for actual devices this approach may only provide an upper estimate of the drain current [12, 42]. The self-consistent Monte Carlo method to solve the BTE is regarded as a very promising tool to simulate carrier transport in nano-scale devices, at least down to $L_G = 10$ nm.

In this chapter we apply the MC method to study the properties of electron transport in advanced MOSFET. The simulation results point out the strong influence of the scattering on the on-current even in ultra-scaled devices. We analyze in detail the flux of back-scattered carriers; the role of scattering in different parts of the device is clarified and the Monte-Carlo results are compared to the simple models for quasi-ballistic transport presented in sec. 1.2.

3.1 Study of scattering effects inside the channel

In this section MC simulations are performed to: i) evaluate the contribution of scattering to the on-current of advanced MOSFETs; ii) understand the

effect of scattering mechanisms in the channel. In the next section we will see that also scattering in the drain region has its own importance when analyzing the ballistic transport.
The original results and discussion are available in [46].

3.1.1 Description of the devices under study

We consider a Double-Gate SOI with gate length equal to 25 nm. The device structure is defined following the specification for the 45 nm High Performance (HP) node from the 2003 edition of the ITRS.
The device design requires many simulations for evaluating the leakage current I_{OFF}, the treshold voltage and the DIBL. We have used the drift-diffusion model of the commercial simulator Dessis to this purpose.
The gate stack is composed of a metal gate on a thin (EOT=0.7 nm) nitride oxide layer with dielectric constant $\epsilon_{ox} = 7$. Since the body is undoped, the gate metal contact has a midgap workfunction to set the correct threshold voltage. We want to design test devices with low short channel effects and good electrostatic behavior. We enforce the electrostatic integrity by setting $DIBL = 110\,mV/V$ for the device with the nominal gate length of this technology node, that, in the case of 45 nm HP technology node, is $L_G = 18$ nm. Thus the device under test, with $L_G = 25$ nm, has much lower SCE. The constrain about the DIBL fixes the body thickness $t_{SI} = 10$ nm. We have also verified that the leakage current I_{OFF} is lower than the maximum value predicted by the roadmap.
The Source and Drain doping profiles feature the Light-Doped Drain (LDD) diffusion. The profiles of both the contact and the extended diffusion are approximated by Gaussian functions and their main characteristics (junction depth, lateral abruptness, length of the spacer defining the mask regions) are selected according to the roadmap.

3.1.2 Results about scattering mechanisms

Fig. 3.1 shows the simulated output characteristic of the Double Gate SOI MOSFET with $L_G = 25$ nm. The ballistic current I_{BL} is calculated by turning off scattering inside the channel and inside a portion of the drain much larger than the mean free path, while the scattering mechanisms remain active near the source and drain contacts, in order to keep the carriers thermalized near the contacts. The quantum correction by effective potential is active in both the scattering and the ballistic simulations. The important role played by scattering in this ultra-short device is demonstrated by the large reduction

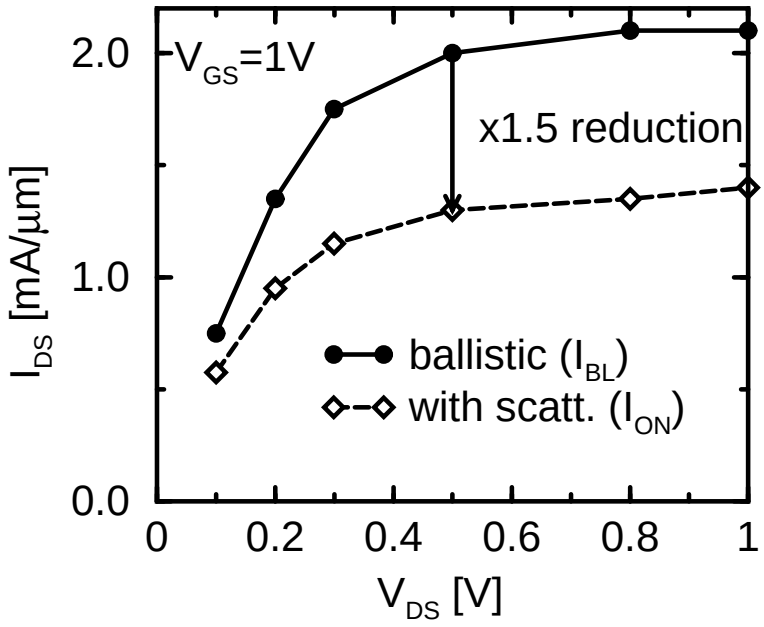

Figure 3.1: Output characteristics of the Double-Gate SOI transistor with $L_G = 25$ nm. Dashed line: scattering activated in the whole device. Solid line: ballistic simulation. The device is designed according to the 45nm Technology node of the ITRS roadmap (2003 edition)

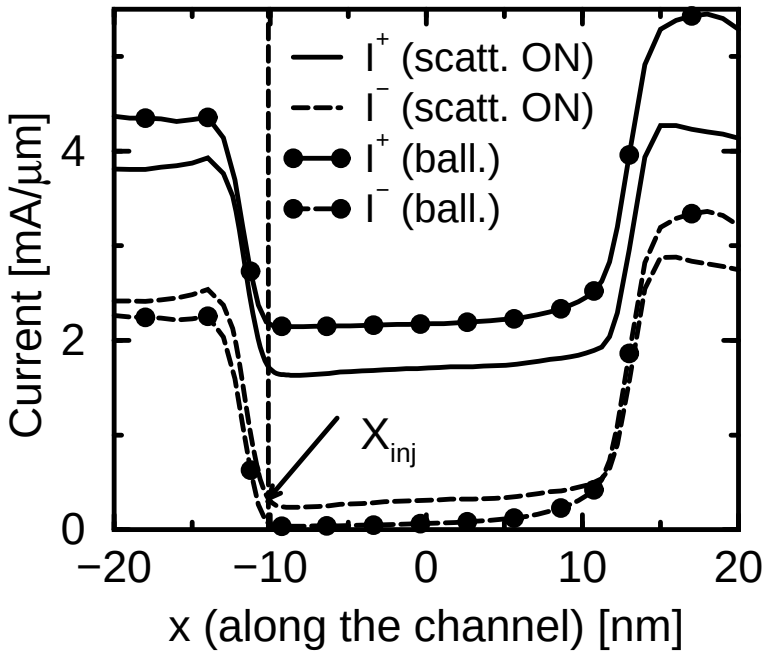

Figure 3.2: I^+ (solid lines) and I^- (dashed lines) profiles comparing the case with (only lines) and without (lines plus circles) scattering. Same device as in Fig. 3.1. $V_{GS} = V_{DS} = 1V$. The virtual source is highlighted by the vertical line.

(about 50%) of the on-current due to scattering. This result is in good quantitative agreement with the analysis presented in [47], where I_{BL} in a silicon MOSFET with L_G=23 nm was found to be 1.47 times larger than I_{ON}.

As seen in section 1.2, a powerful approach to understand quasi-ballistic transport is to separate the carrier fluxes moving with group velocity ($\vec{v_g}$) oriented from source to drain ($v_{gx} > 0$) or from drain to source ($v_{gx} < 0$). For each of these fluxes we are able to compute the corresponding current, electron concentration and average velocity.
From this point we will analyze the bias point $V_{GS} = V_{DS} = V_{DD} = 1$ V, which is the bias for the maximum currents. This is also the condition where the quasi-ballistic effects are more interesting because of the high electric field along the channel.
The absolute values per unit width of the currents I^+ and I^- along the channel for simulations with and without scattering are reported in Fig. 3.2. The qualitative shapes of the I^+ and I^- profiles are the same in both simulations with and without scattering. However, the absolute values of these currents are different, in particular inside the channel, and near the virtual source x_{inj}. The position of the injection point, which corresponds to the maximum of the potential energy, is found to be the same with and without scattering. We define:

- $I^+_{inj} = I^+(x_{inj})$, that is the current injected into the channel from the source,

- $I^-_{inj} = I^-(x_{inj})$, that is the current back-scattered to the source.

We will use the symbols $I^+_{inj,ON}$ and $I^-_{inj,ON}$ when referring to simulations with scattering, while the symbols $I^+_{inj,BL}$ and $I^-_{inj,BL}$ will refer to ballistic simulations.

By inspection of Fig. 3.2 we see that, as expected, $I^-_{inj,BL} \simeq 0$. Furthermore, $I^+_{inj,BL}$ and $I^+_{inj,ON}$ are not equal. This latter important fact will be analyzed in the next section. Here we investigate further the spatial regions that mostly contribute to $I^-_{inj,ON}$. To this purpose, Fig. 3.3 plots the total number of scattering events per unit length and time suffered by the electrons moving with positive group velocity. The details about how this curve is generated can be found in [46]. $1/\lambda^+$ is a sort of inverse mean free path for the electrons moving from source to drain. In the channel $1/\lambda^+ \approx 0.1\, nm^{-1}$, meaning that, for this device and this bias point, forward moving carriers suffer approximately one scattering event every 10 nm.

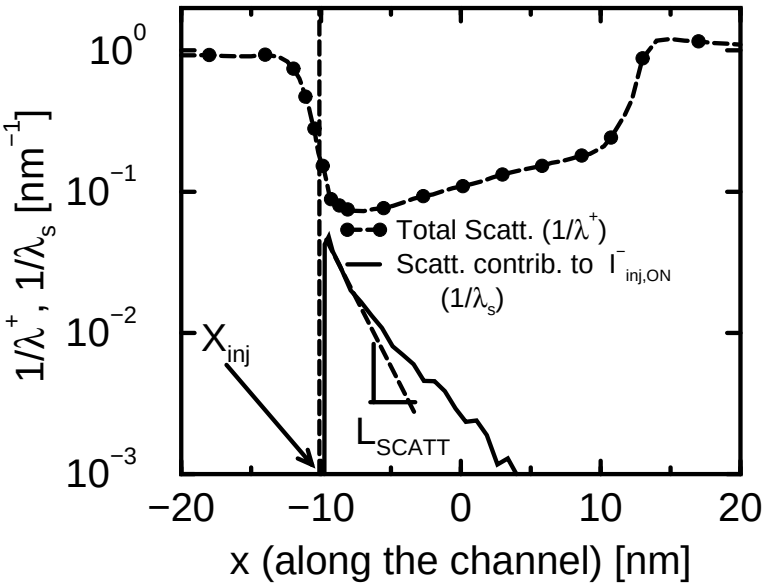

Figure 3.3: Dashed line with symbols: number of scattering events per unit length and unit time (that is a sort of inverse mean free path), suffered by the electrons moving from source to drain. Solid line: number of scattering events contributing to $I^-_{inj,ON}$ per unit length and unit time. Both are normalized to $I^+_{inj,ON}/q$. L_{SCATT} is the decay length of the contribution to $I^-_{inj,ON}$. Same device as in Fig. 3.1. $V_{GS} = V_{DS} = 1$ V.

In the same figure we also report the number of scattering events per unit length ($1/\lambda_S$) that contribute to the current $I^-_{inj,ON}$ as a function of the position along the channel. This quantity is evaluated by counting the number of particles crossing the virtual source with $v_{gx} < 0$ that were back-scattered by a collision in a spatial interval Δx placed around the position x. This number is then divided by Δx, by the simulation time, and by $I^+_{inj,ON}/q$. This contribution, which is obviously zero for the x position before the virtual source, is peaked at x_{inj} and decays rapidly for increasing x. The decay length near the virtual source is denoted as L_{SCATT}, which is of the order of few nanometers. The comparison between the total number of scattering events and the number of scattering events contributing to $I^-_{inj,ON}$ points out that only a small part of the total scattering events taking place inside the channel controls how close to its ballistic limit the device operates. In particular the dominant contribution to backscattering is given by collisions close to the peak of the potential barrier, in agreement with [14].

After the separation of the positive and negative carrier fluxes, it is also helpful to analyze the carrier velocity in the device. To this purpose, Fig. 3.4 shows the average drift velocity v_{avg} along the channel for the same device

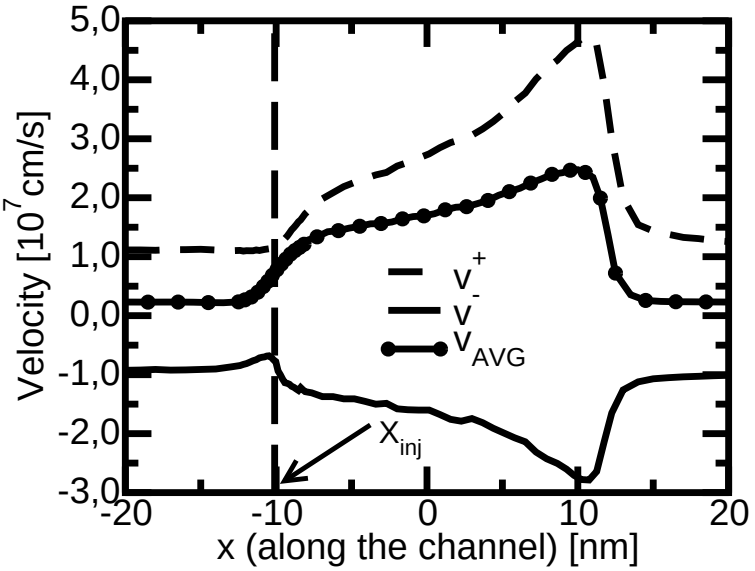

Figure 3.4: Average drift velocity (v_{avg}) compared with the average velocities v^+ and v^- of the electrons with group velocity $v_{gx} > 0$ and $v_{gx} < 0$, respectively. Simulation with scattering active in the whole device.

and bias as in the previous figure. This is the average velocity of all carriers at each position x along the device, regardless of their vertical position. Inside v_{avg}, the information about carrier moving with positive group velocity ($v_{gx} > 0$) and with negative group velocity ($v_{gx} < 0$) are mixed together. In order to separate and compare the two contributions, we report in the same figure the velocities v^+ and v^-, which are the average velocity of the carriers with, respectively, $v_{gx} > 0$ and $v_{gx} < 0$. We see that the drift velocity is low in the source and drain, since the electric field in these regions is low, whereas both v^+ and $|v^-|$ are close to the thermal velocity $v_T \simeq 10^7$ cm/s. The drift velocity at the virtual source is lower than v_T, while $v^+ \simeq v_T$, as assumed by the present models for ballistic and quasi-ballistic transport [12, 14]. On the other hand, $|v^-|$ is slightly lower than v_T.
We also see that v^+ grows very rapidly along the channel and largely exceeds $|v^-|$. As a result the velocity overshoot at the drain end of the channel is relevant.

The picture coming out from our study, in terms of backscattering current and velocity at the injection point, is similar to the picture of the quasi-ballistic model in section 1.2. In the next sections we continue the comparison between our analisys and the result of that model.
Till now the analysis has considered a high drain voltage, $V_{DS} \gg kT/q$. At low drain voltage a relevant fraction of the carrier injected from the drain

can reach the virtual source, also in the case of ballistic simulations, so that $I^{-}_{inj,BL} > 0$. Moreover, since the electric field along the channel is low, the current is controlled by the whole channel and the lenght L_{SCATT} becomes comparable to L_G. The conclusions drawn in this section and in the following ones are thus valid at high drain bias and we will not consider any other operating condition.

3.1.3 Comparison with analytical models for quasi-ballistic MOSFETs

In this section we apply the Monte Carlo simulations to investigate the validity of the assumptions at the basis of the analytical model for quasi-ballistic MOSFETs proposed in [14, 15], that we have already presented in section 1.2. In particular we would like to check its two main features: the first one is the expression for the ballistic ratio

$$BR = \frac{I_{ON}}{I_{BL}} = \frac{1-r}{1+r} \tag{3.1}$$

while the second one is the statement that only backscattering events within a kT-length from the virtual source contribute to the negative carrier flux.

Validation of the model for the Ballistic Ratio

In order to verify the validity of the first point, Fig. 3.5 compares the *ballisticity ratio* BR=I_{ON}/I_{BL} with the term $(1-r)/(1+r)$ in the same $L_G = 25$ nm DG MOSFET of the previous section and in other two similar devices with different gate length. The value of r has been evaluated as $I^{-}_{inj,ON}/I^{+}_{inj,ON}$, which is the ratio between the negative flux (given by backscattered carriers) and the positive flux. The figure shows that the term $(1-r)/(1+r)$ reproduces fairly well the BR.

In Fig. 3.5 we have also reported the term $1-r$, that would be the ballistic ration if $I_{inj,BL}$=$I^{+}_{inj,ON}$, since $I^{-}_{inj,BL} = 0$. In fact, under this assumption:

$$\frac{I_{ON}}{I_{BL}} = \frac{I^{+}_{inj,ON} - I^{-}_{inj,ON}}{I^{+}_{inj,ON}} = 1 - r. \tag{3.2}$$

From the figure we see that $1-r$ term does not approximate the ballistic ratio and is very different from $(1-r)/(1+r)$. In fact, we have already seen in

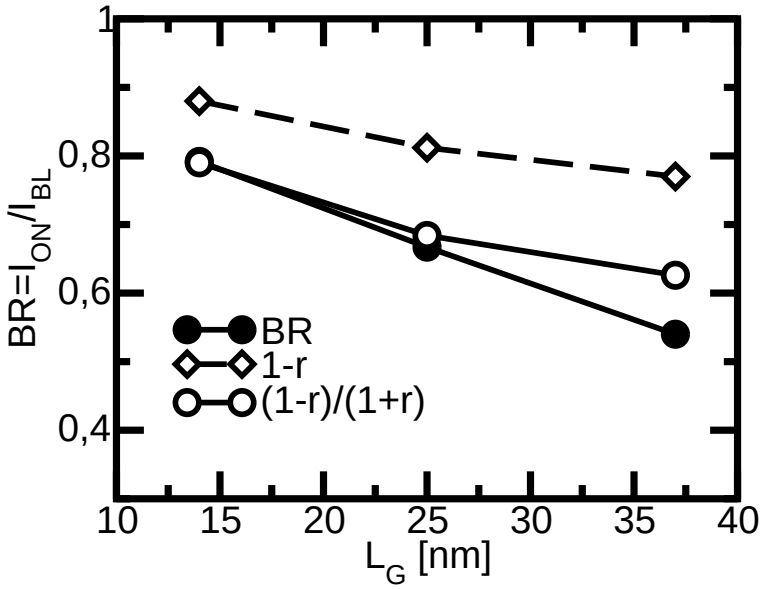

Figure 3.5: Filled symbols: ballistic ratio BR between the drain current I_{ON} and its ballistic counterpart I_{BL}. Open symbols: terms $1 - r$ (diamonds) and $(1 - r)/(1 + r)$ (circles). $r = I^{-}_{inj,ON}/I^{+}_{inj,ON}$. $V_{GS} = V_{DS} = 1$ V.

Fig. 3.2 that $I^{+}_{inj,BL} > I^{+}_{inj,ON}$. An explanation for this is provided by considering the assumptions of the quasi-ballistic model under study [15]: a) at the virtual source $v^{+} = |v^{-}| = v_T$; b) the inversion charge at the virtual source is controlled only by the gate capacitance and voltage, and it is therefore the same with and without scattering, meaning that without collisions there would be more carriers moving with $v_{gx} > 0$, and thus higher $I^{+}_{inj,BL}$.

As it can be seen in Fig. 3.4, the first assumption is reasonably verified in our simulations for v^{+}, while $|v^{-}| \approx 0.7\, v_T$. In order to check the validity of the second assumption, we plot in Fig. 3.6 the inversion charge concentration at the virtual source for the case with scattering, separated into the contributions Q^{+}_{i} (due to carriers with positive group velocity) and Q^{-}_{i} (carriers with negative group velocity). Also shown is the inversion charge concentration in the ballistic case $Q_{i,BL}$, calculated at the virtual source. We remind that the virtual source remains in the same position x_{inj} with and without scattering. We see that:

$$Q^{+}_{i} + Q^{-}_{i} \approx Q_{i,BL}, \tag{3.3}$$

that verifies the above assumption. This is probably due to the fact that the template DG-SOI devices considered here are fairly well scaled and free of significant short channel effects. If the electrostatic behavior at the injection point is dependent on the SCE, the interpretation of the results, following

the model of sec. 1.2, would not be possible.

Scattering near the injection point

We consider now the other feature of the analytical model, i.e. the fact that the scattering events contributing to $I^{-}_{inj,ON}$ take place within the kT-length L_{KT}. We have already seen in Fig. 3.3 that $1/\lambda_S$ (the number of scattering events contributing to $I^{-}_{inj,ON}$) is maximum at the virtual source and then decays rapidly. From that figure it was possible to extract the decay length L_{SCATT}. On the contrary L_{KT} is simply evaluated by observing the potential profile along the device. Fig. 3.7 compare the decay length L_{SCATT} with L_{KT} in our template devices with different gate lengths. As expected L_{SCATT} is very close to the kT-length, meaning that, in the proximity of the virtual source, the probability for a back-scattered electron to overcome the potential barrier is related to the average lateral electric field in the kT-layer: $F_S \simeq (kT/q)/L_{KT}$. This result supports the simplified picture proposed in [14, 15].

3.2 Accurate methodology for the quasi-ballistic simulation

It is important to notice that, although the results shown so far essentially validate the model proposed in section 1.2, this approach cannot be used to post-process full-ballistic simulations in order to account for the effect of scattering.

In fact, one would be tempted to use Eq. (3.1) to correct the current I_{BL} obtained with a ballistic tool, but the model for r requires to know L_{KT}. For example Eq. (1.10) suggests to use an appropriate mean free path that can be tentatively calculated from the low-field mobility. However, since the scattering inside the channel implies higher inversion charge in the channel, the potential profile in ballistic simulations is different than in the case with scattering, as shown in Fig. 3.8. As a result, L_{KT} calculated assuming ballistic transport differs from that calculated including scattering, as demonstrated in Fig. 3.9. Thus, the kT-length extracted from a ballistic tool cannot be used to calculate the back-scattering coefficient in a reliable way. Only self-consistent simulations with scattering are adequate for this purpose and the Monte Carlo approach seems to be one of most effective ways to evaluate the quasi-ballistic transport.

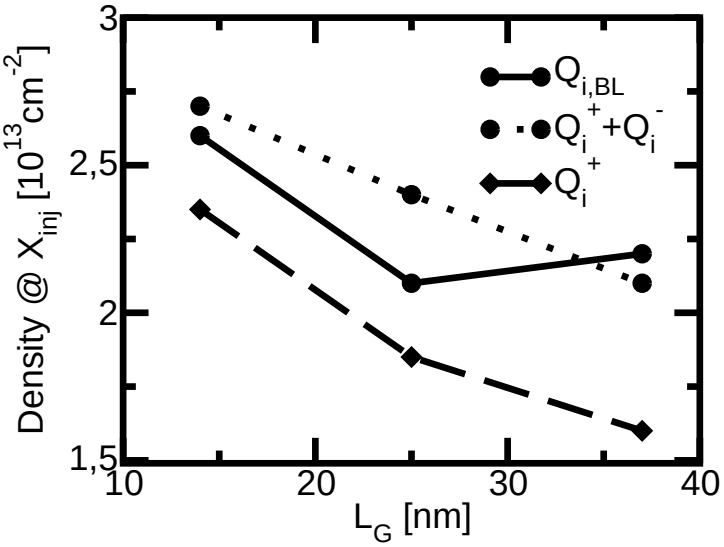

Figure 3.6: Inversion charge concentration at the virtual source for devices with different gate length. Q_i^+ and Q_i^- refer to electron with group velocity $v_{gx} > 0$ and $v_{gx} < 0$ respectively, in the simulations with scattering. $Q_{i,BL}$ refers to ballistic simulations. $V_{GS} = V_{DS} = 1$ V.

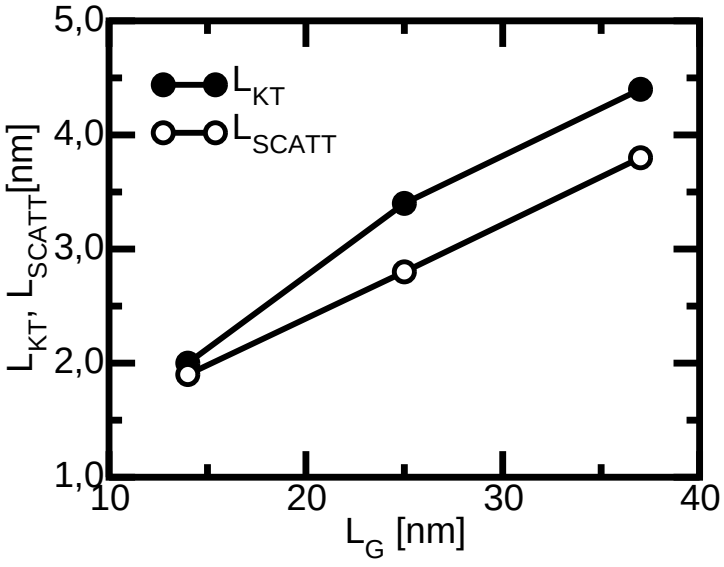

Figure 3.7: Comparison between the decay length L_{SCATT} (see solid line in Fig. 3.3) and the length of the kT-layer L_{KT}. $V_{GS} = V_{DS} = 1$ V.

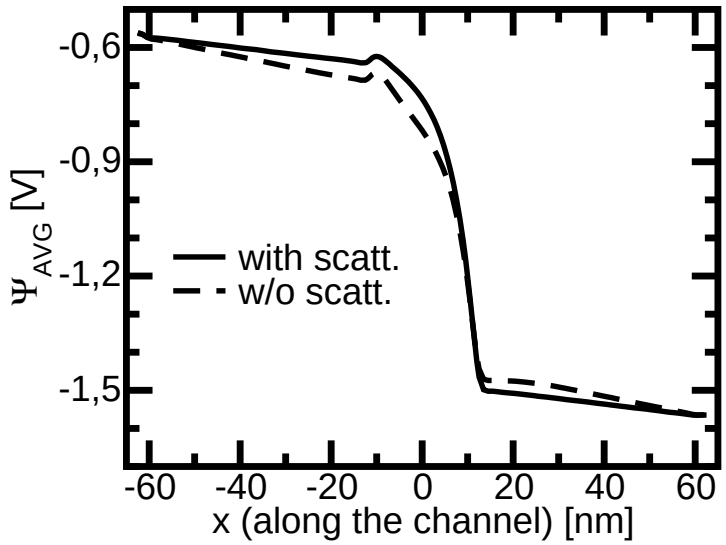

Figure 3.8: Potential energy profile along the template 25 nm DG-SOI MOSFET, averaged in the silicon film: $\Psi_{AVG}(x) = -(1/Q_i(x)) \int_0^{T_{si}} \Psi(x,y)n(x,y)dy$. The simulation with scattering (solid line) is compared with the case without scattering (dashed line). At $x = x_{inl}$ the difference is about 40 mV. $V_{GS} = V_{DS} = 1$ V.

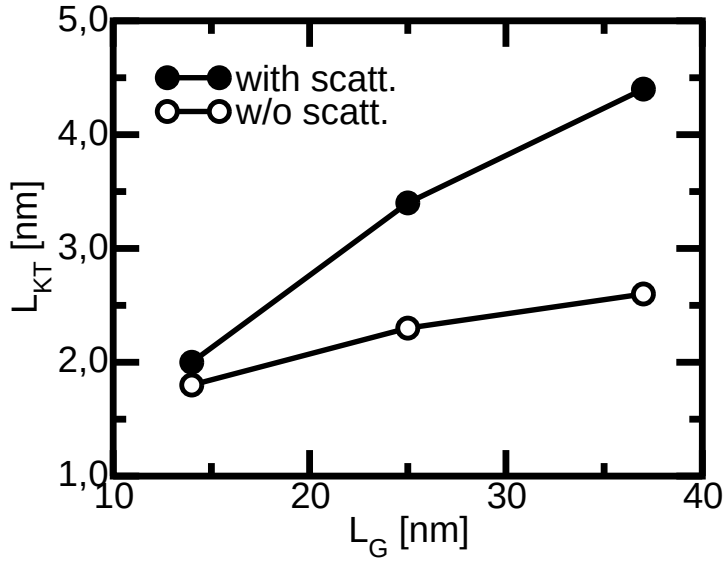

Figure 3.9: Extension of the kT-layer (L_{KT}) as extracted from simulations with (filled circles) and without (empty circles) scattering. Same devices as in Fig. 3.5.

We could also ask if simple simulators implementing the equations (1.9) and (1.10) and simple 1D electrostatic at the virtual source could be enough for the evaluation of the quasi-ballistic current. The evaluation of model parameters, such as the position and the charge at the virtual source, as well as the extension of the KT-layer and the carrier mean-free-path λ, is not trivial. For example, in the DG SOI devices considered in this work, the virtual source is in the overlap region, so that simple formulas based on the oxide capacitance and gate overdrive ($V_G - V_{TH}$) cannot be used to compute the inversion charge. The self-consistent Monte Carlo simulation has demonstrated to be a tool of primary importance for the simulation of the quasi-ballistic transport in short-channel MOSFETs.

3.3 Effect of scattering in the drain region

It has been suggested that in ultra-short devices back-scattering at the drain could have a relevant impact on the drain current [48, 49]. However, we have already seen in Fig. 3.3 that the contribution to $I^{-}_{inj,ON}$ due to scattering at the drain is negligible.

The case of a shorter device (L_G = 14 nm) is shown in Fig. 3.10: the contribution from back-scattering (filled squares) still confirms that scattering at the drain has no influence on $I^{-}_{inj,ON}$. In the same figure we plot the contribution from backscattering events in the case of ballistic channel with and without plasmon scattering at the drain. When the scattering inside the channel is turned off (as done in [48, 49]), the contribution to $I^{-}_{inj,ON}$ due to back-scattering at the drain becomes much higher (filled circles), and even larger if plasmon scattering in the drain is switched off (open circles). This is because particles injected into a ballistic channel enter the drain with an energy equal to, or higher than, the source barrier. Therefore, if back-scattered by elastic collisions, they have a high probability to travel all the way back to the source. Plasmons, which are a very inelastic scattering source, reduce the back-scattered flux, because they subtract a significant energy to the electrons entering the drain and therefore they efficiently thermalize the hot electrons coming from the channel. On the other hand, if inelastic scattering in the channel (e.g. optical phonon scattering) reduces the carrier energy, then carriers entering the drain have not enough energy to go back to the source if back-scattered. This result points out that it is methodologically incorrect to estimate the impact of scattering at the drain on I_{ON} by switching off scattering in the channel.

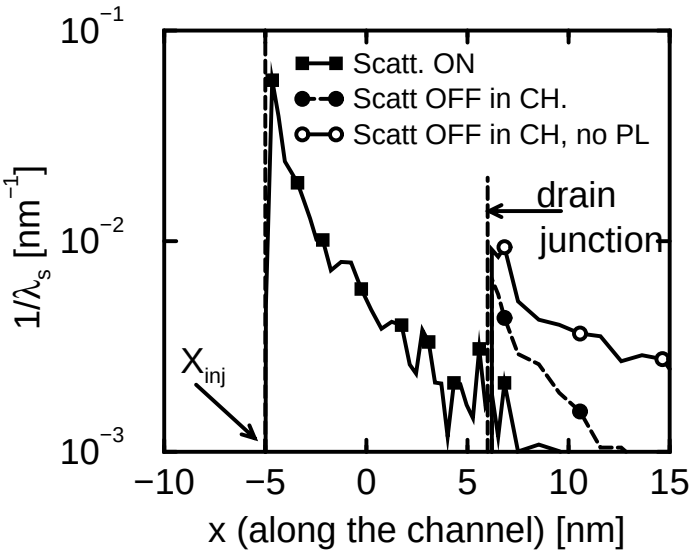

Figure 3.10: Contribution to $I^{-}_{inj,ON}$ given by the scattering events at a given position along the channel. Filled squares: the scattering active in the whole device. Filled circles: no scattering inside the channel. Open circles: no scattering inside the channel and plasmon scattering is turned off inside the drain. DG SOI MOSFET with $L_G = 14$ nm. $V_{GS} = V_{DS} = 1$ V.

Obviously these considerations hold at relatively high drain voltages.

Since all the simulations reported in this work are self-consistent and our MC simulator features an electron-plasmon interaction as a scattering mechanism, it could be argued that we are double-counting the effect of the e-pl interaction [50]. However, this is not the case, because e-pl interaction can be taken into account by the self-consistency only if the grid spacing, the time-step and the number of particles in each grid element are properly selected [50, 51]. This was not the case for the parameters of our simulations, and the demonstration is found in [46].

Chapter 4

Study of the ballisticity in nano-MOSFET along the roadmap

In the previous chapter we presented a detailed Monte Carlo analysis of the role of scattering in the channel and in the drain of deca-nanometric MOSFETs. In particular we compared our results to the analytical model for quasi-ballistic transport presented in section 1.2. This model provides useful physical insight because it links the on-current to the scattering events taking place in the kT-layer. Nonetheless, the informations needed for quantitative predictive analysis (location of x_{inj}, the values of L_{KT} and λ) would require a 2-D self-consistent simulation. On the contrary, self-consistent Monte Carlo simulation represents the ideal tool for this analysis.

In this chapter, the methodology and the simulation tool of section 3.1 are applied to a systematic study of Bulk and DG MOSFETs designed according to the ITRS 2003 in order to understand to which extent ballistic transport is going to affect devices with channel lengths down to 14 nm. The result of this chapter has been presented in [52].

4.1 Scaling of Bulk and Double-Gate MOSFETs

In this study we consider MOS devices representative of several technology nodes (TN) and two technological options. In particular (between parenthesis the nominal gate length for each TN is indicated):

- three nodes for the Bulk technology: 130 nm (65 nm), 90 nm (37 nm) and 65 nm (25 nm);
- three nodes for the Double-Gate SOI technology: 90 nm (37 nm), 65 nm (25 nm) and 45 nm (18 nm).

We follow the ITRS 2003 roadmap to design the devices with nominal gate length. Within each TN we scale L_G without changing any other parameter. We consider a large number of devices, which permits us to simulate: (i) the effect of scaling on the ballisticity *within each TN*, (ii) the effect of moving from a technology node to another one. Further details about the device design are described in the following sections.

4.1.1 Description of the device structures

The design of the simulated devices follows the same general rules already used in 3.1.1. In particular we use the drift-diffusion simulator Dessis to check the electrostatic behavior and the leakage currents. All device parameters have been chosen to comply the ITRS 2003 High-Performance specifications. A simple sketch of the bulk and DG SOI structure is found in Fig. 4.1 and 4.2 respectively. All the parameters like oxide thickness, supply voltage, etc are listed in Table 4.1.

The gate stack is realized with nitride oxide instead of SiO_2 (dielectric constant $\epsilon_{ox} = 7$) in order to limit the gate leakage current. N-poly and metal gates are used, for bulk and DG SOI MOSFETs, respectively. The metal gates have midgap workfunction (4.6 eV) to set the correct threshold voltage. The polysilicon contact is supposed to be ideal. The doping profiles in the channel of the bulk MOSFETs are tailored in order to keep short channel effects (SCE) and drain leakage current (I_{OFF}) below the limits set by the ITRS for each TN. We set $DIBL = 110$ mV/V for the device with the nominal gate length of each technology node to limit short channel effects. The source and drain regions are formed by a relatively deep contact region and a thinner extension superimposed to a heavily-doped halo. S/D extensions and halos are aligned to the gate electrode edges; S/D contact regions are separated from the gate edge by a lateral spacer. The profiles of dopant concentrations are approximated by gaussian functions and their main characteristics are listed in Table 4.1.

The average substrate doping level, resulting from the ITRS and our DIBL specifications, are rather large ($N_{a,avg} > 10^{18}\,\mathrm{cm}^{-3}$) which let us predict a

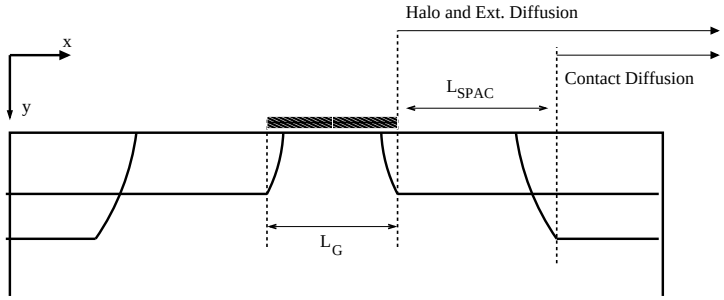

Figure 4.1: Sketch of the simulated Bulk MOSFETs. Drain and Source contacts are at the two edges. Details of this device for all technology nodes can be found in Table 4.1

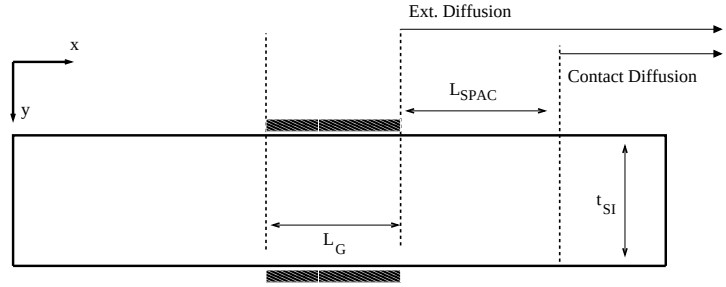

Figure 4.2: Sketch of the simulated Double-Gate SOI MOSFETs. Details of this device for all technology nodes can be found in Table 4.1

Technology node		**130 nm**	**90 nm**	**65 nm**	**45 nm**
	Supply Voltage V_{DD} (V)	**1.2**	**1.2**	**1.1**	**1.0**
	Nominal Gate Length (nm)	**65**	**37**	**25**	**18**
	Rel. Diel. Const. ε_{ox}	**7**			
	L_{SPAC} (nm)	**70**	**40**	**27**	**19**
	t_{ox}/EOT (nm)	**2.87/1.6**	**2.15/1.2**	**1.62/0.9**	**1.26/0.7**
	S/D Ext.: Peak Conc. [cm^{-3}]	10^{20}			
	S/D Ext.: Vert. Std. dev / Lat. Std. dev [nm]	**15/4.5**	**7.5/1.88**	**5.6/1.18**	**2.5/0.33**
	Bulk-Halo: Peak Conc. [cm^{-3}]	**-**	**6$\cdot$ 10^{18}**	**6$\cdot$ 10^{18}**	**-**
	Bulk-Halo: Vert. Std. dev / Lat. Std. dev [nm]	**-**	**8/2.8**	**6/2.1**	**-**
SOI DG	**t_{SI} (nm)**		**17**	**12**	**10**
	DIBL (mV/V)		**91**	**97**	**112**
	I_{OFF} (nA/μm)		**4**	**7**	**10**
	Gate Workfunc.[eV]	**4.6**			
BULK	**N_a (10^{18}cm^{-3})**	**2.5**	**2 + halo**	**3 + halo**	
	DIBL (mV/V)	**105**	**108**	**120**	
	I_{OFF} (nA/μm)	**2.5**	**30**	**60**	
	Gate Workfunc.[eV]	**4.05**			

Table 4.1: Main technological parameters of the simulated devices. For each technology node the nominal gate length for high-performance transistors is reported. The results of the DIBL and I_{OFF} from drift-diffusion simulations are reported.

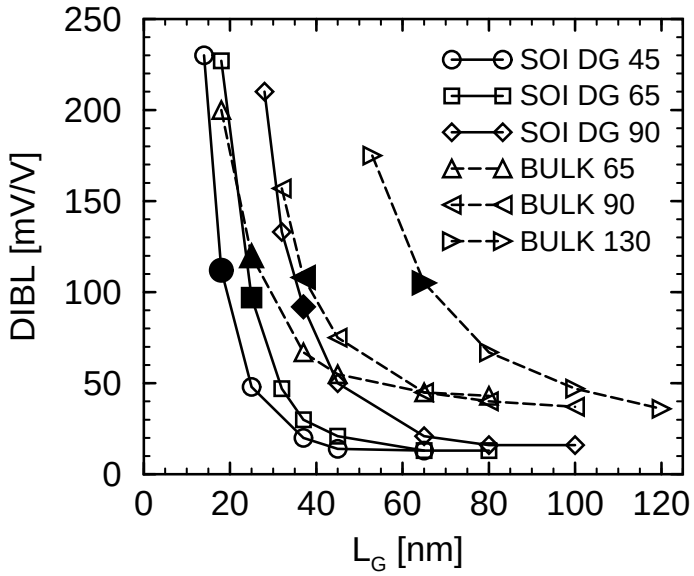

Figure 4.3: DIBL vs. L_G for Bulk and DG SOI MOSFETs from Drift-Diffusion, simulations; filled symbols denote devices with the nominal gate length of each TN.

relatively low current due to the degraded electron mobility. For the DG SOI transistors, we assume lightly-doped silicon film ($N_a = 10^{15}\,\mathrm{cm}^{-3}$) and no halos, while the specification on the DIBL and drain leakage are met by scaling the thickness of the silicon layer t_{SI}. The donor concentration profiles of the source and drain regions for the 90 nm and 65 nm TNs are the same adopted for the corresponding bulk MOSFETS. In the case of the 45 nm technology node, the profiles were obtained by aggressive scaling of the doping-profile parameters.

4.1.2 General performance of the simulated devices

Before studying the on current we check the electrostatic integrity of all transistors, evaluated by drift-diffusion density-gradient simulations. The results are illustrated in Fig. 4.3: the DIBL of the High-Performance transistors with nominal L_G is approximately 110 mV/V for all the considered technology nodes. Scaling down the gate length increases the DIBL as expected.

Fig. 4.4 reports the simulated I_{ON} and I_{BL} for transistors with the nominal L_G-values corresponding to the different TNs. The ballistic current is calculated by switching off all scattering mechanisms inside the channel region. In the case of the DG MOSFETs, currents are divided by a factor of 2 in order to account for a single inversion channel, therefore leading to a fair compari-

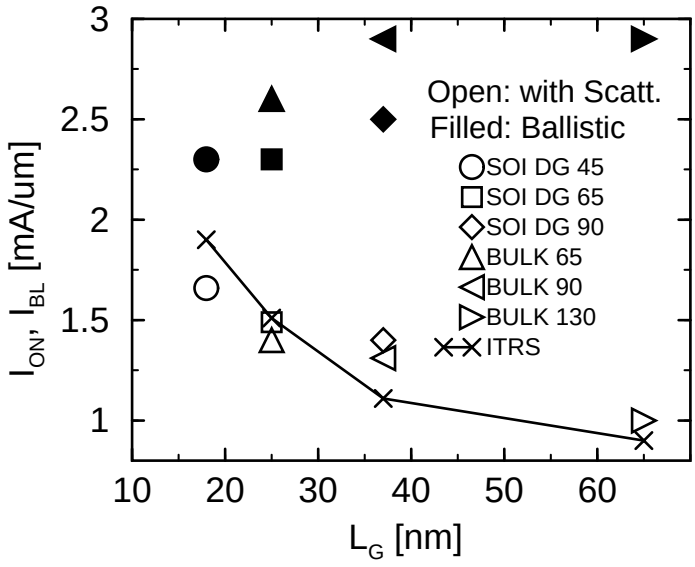

Figure 4.4: Drain current vs. the nominal gate length at $V_{GS} = V_{DS} = V_{DD}$ for the devices of each TN. Filled symbols: ballistic current I_{BL}; Open symbols: I_{ON}. In the case of DG MOSFETs, currents are divided by a factor of 2 in order to account for a single inversion channel.

son with single-gate bulk MOSFETs. As it can be seen, I_{BL} is slightly larger in the bulk MOSFETs, due to the larger source/drain parasitic resistances of the SOI devices. For a given architecture, I_{BL} is almost independent of the TN and it decreases slightly with scaling (maximum variation is within 10%), mainly due to the non-ideal scaling of the source/drain series resistances. We verified that these almost-constant I_{BL} values stem from constant values of the inversion density N_{INV} and of the injection velocity at the virtual source x_{inj} as predicted in [14, 15]. This is a result from the combination of V_{DD} and EOT values required by the ITRS (see table 4.1). Fig. 4.4 also shows that the I_{ON} values, differently from I_{BL}, increase along the Roadmap.

According to [14, 15] the ballistic upper limit of the current for a well-scaled MOSFET is essentially controlled by the 1-D electrostatics of the MOS system at the virtual source and by injection velocity. As a consequence, the ballistic current is expected to be essentially independent of channel length within a given TN, unless substantial drain-induced barrier lowering affects the inversion charge at the virtual source.
This point is confirmed by Fig. 4.5, reporting the dependence of the ballistic upper limit on L_G for a fixed TN, where the ballistic current increases significantly only when L_G is scaled to values corresponding to unacceptable short-channel effects.

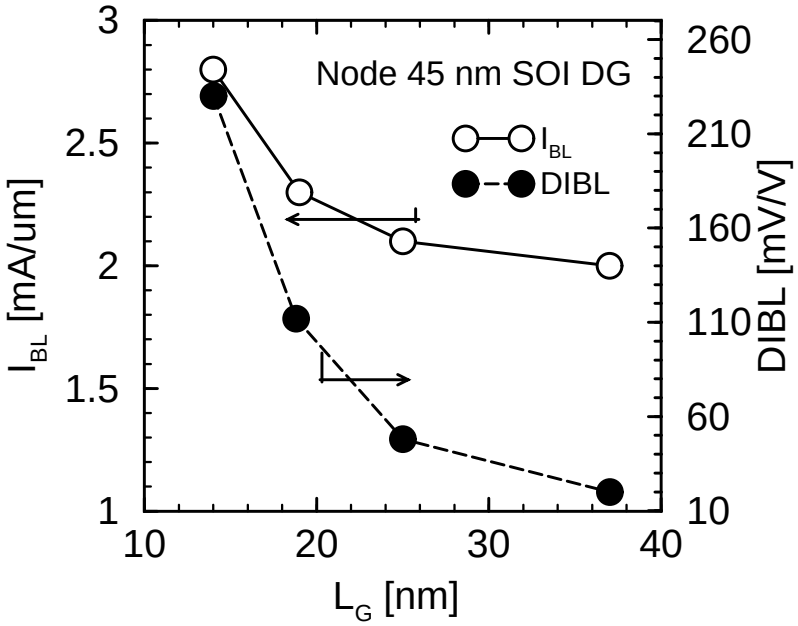

Figure 4.5: Ballistic current and DIBL vs. gate length at $V_{GS} = V_{DS} = V_{DD}$ for devices of the 45 nm technology node.

4.2 Scaling of the quasi-ballistic transport

In the previous section we have evaluated separately the scaling of I_{ON} and I_{BL} along the roadmap. From Fig. 4.4 we recognize that the ballistic ratio $BR = I_{ON}/I_{BL}$ has increased moving from the oldest 130 nm TN to the advanced 45 nm node. This result is in agreement with the ITRS specification which requires the on-current to approach the ballistic limit with the scaling of L_G [5]. We now perform an extensive study of the scaling of the ballistic ratio along the roadmap and we will analyze its dependence on the device architecture. In order to investigate the correlation of BR to L_G , we will use the quasi-ballistic model described in 1.2 and validated in 3.1. In particular we will extract the distance L_{KT} that it takes for a potential drop of $[KT/q]$ from the virtual source towards the channel: this characteristic length, together with a suitably defined carrier mean free path λ, is sufficient for a simple analysis of our results.

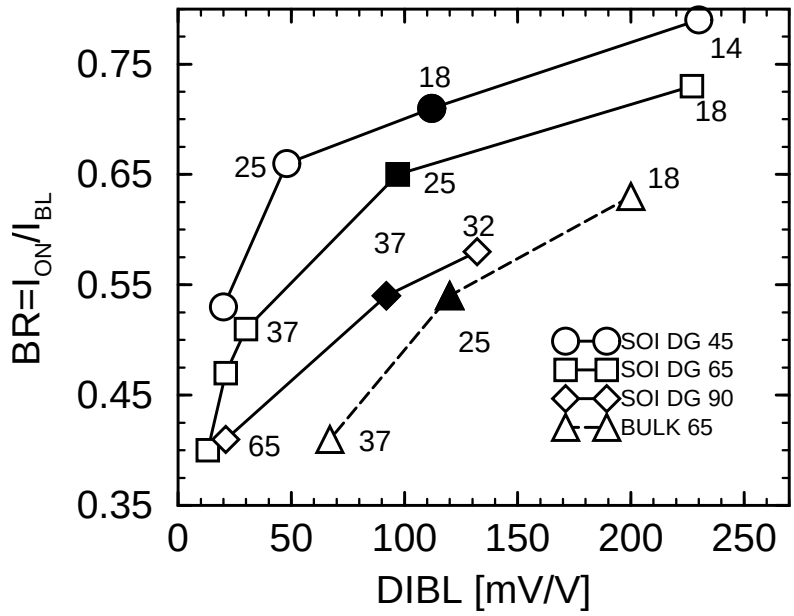

Figure 4.6: Ballisticity-ratio $BR = I_{ON}/I_{BL}$ at $V_{GS} = V_{DS} = V_{DD}$ vs. DIBL for all simulated devices; numbers close to each symbol report the corresponding L_G. Filled symbols represent devices with the nominal gate length.

4.2.1 Scaling of the Ballistic Ratio

The ballistic ratio $BR = I_{ON}/I_{BL}$ quantifies how close a device operates to its ballistic limit. According to [14], [15], scaling the gate length increases the BR, due to the increase of longitudinal field in the channel. For this reason, a clear correlation may be expected between the ballistic ratio and DIBL. Fig. 4.6 reports the ballistic ratio versus DIBL for devices belonging to different TNs. For a given TN, it is possible to increase the BR by scaling the gate length, thus increasing the longitudinal field in the channel region close the virtual source, but, when L_G is scaled beyond the nominal value for each TN, the improvement of BR comes at the cost of a significantly larger DIBL. A comparison among different TNs points out that the BR at given DIBL is improved by scaling and that the UTB-DG MOSFET are closer to the ballistic limit than bulk MOSFETs of the same node.

Another correlation is between the ballistic ratio and the gate length because we have already seen from Fig. 4.4 that BR is increasing with the scaling of L_G. Fig. 4.7 reports the BR as a function of gate length for Double-Gate SOI MOSFETs. It is interesting to observe that the BR values lay essentially on the same curve for all TNs. In this sense, Fig. 4.7 suggests that the scaling

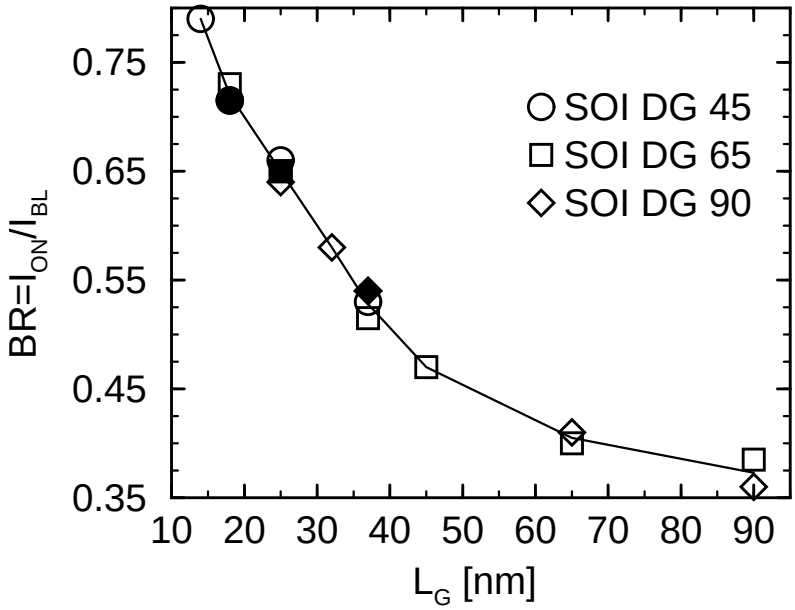

Figure 4.7: Ballisticity-ratio of fig. 4.6 plotted versus gate length. The continuous line is drawn as an eye-guide that emphasizes the correlation between BR and L_G.

strategy set by the Roadmap implies that BR is an unique function of L_G. Furthermore, BR tends to a fairly constant value close to 0.4 at long gate length, whereas it remarkably increases for $L_G < 50$ nm reaching 0.8 for the shortest device.

In order to investigate the correlation of BR to L_G in the DG case, we analyze the length L_{KT} of the kT-layer. The value of L_{KT} is calculated for the same $V_{GS} = V_{DS} = V_{DD}$ bias condition used to evaluate I_{ON} and BR. Fig. 4.8 demonstrates that, *for the scaling strategy of the Roadmap*, L_{KT} is uniquely related to the channel length L_G. One of the main results of the model presented in section 1.2, is eq. 1.10 which states that r, and then BR, depends on the ratio between L_{KT} and the electron mean free path λ. From our simulation we are able to extract a value of λ following the same procedure used for the result in Fig. 3.3. In section 3.1.2 we calculated an electron mean free path nearly equal to 10 nm. We have verified that λ at the virtual source, and then the scattering rates, does not change significantly with L_G. Consequently the ratio $[\lambda/L_{KT}]$ becomes an unique function of the gate length, which qualitatively justifies the BR to L_G correlation of Fig. 4.7. It should be emphasized that such a correlation should not be considered an universal curve in a physical sense but it results from the combination of EOT and V_{DD} values given by the Roadmap.

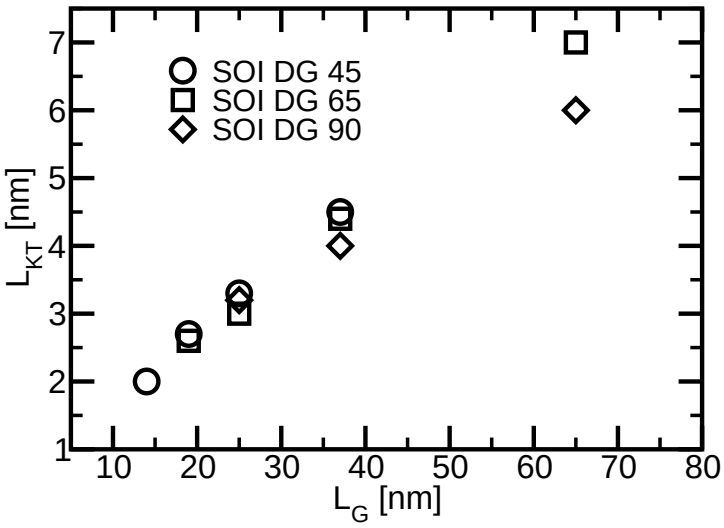

Figure 4.8: kT-layer length versus gate length for $V_{GS} = V_{DS} = V_{DD}$.

4.2.2 The Bulk case

The correlation between BR and L_G, presented in the previous section, breaks in the case of bulk MOSFETs. Fig. 4.9 shows the comparison between the BR curve from Double-Gate MOSFETs and the ratios calculated in the Bulk cases from different TNs. In particular we observe a degradation of the ballistic ratio, for fixed L_G, while moving to most advanced technology nodes. From the analysis of 4.2.1, we interpret these results as an effect of the change of channel doping concentration that affects the mean free path $\lambda_{\mathrm{m}fp}$ through ionized impurity scattering and surface-roughness scattering. The use of large doping concentrations is necessary to counteract SCE and the increase in N_a is detailed in Table 4.1. The main role here is played by surface roughness because increasing doping concentration leads to larger vertical field, hence, enhanced surface roughness scattering. This observation is proven by the fact that when this scattering mechanism is switched off (triangle-down in Fig. 4.9), the scaling trend of DG SOI devices is recovered. On the other hand, the role of impurity scattering is suppressed at large inversion charge concentration, due to the screening effects.

The larger BR obtained in the UTB-DG SOI case, compared to the Bulk counterpart at given gate length, is due to a higher average velocity at the virtual source, as shown in Fig. 4.10. In fact, in agreement with [14, 15], the

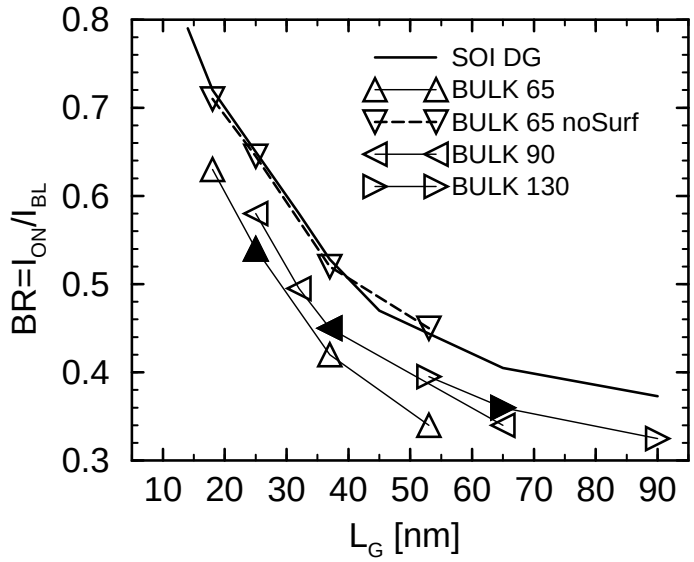

Figure 4.9: Ballisticity-ratio versus gate length for Bulk MOSFETs belonging to all considered TN's. Filled symbols denote devices with the nominal gate length. The solid line represents the results for DG MOSFETs.

velocity at the injection point is $v_d(x_{inj}) \approx BR \cdot v_T$. As the surface roughness scattering is turned off in the bulk device (crosses in Fig. 4.10) we can observe that:

1. the velocity profile of the UTB-DG-SOI device is exactly recovered at the source-end of the channel;

2. at the drain end of the channel, as the inversion layer broadens away from the Si-dielectric interface, the velocity profile merges with the one of the bulk MOSFET simulated including surface roughness.

The results reported in the inset of Fig. 4.10 confirm that surface scattering redirects many injected carriers back to the source and then plays an important role in limiting the I_{ON} of bulk MOSFETs.

4.2.3 Scaling of the transit time

We now analyze the quasi-ballistic transport from a different point of view: we use the concept of carrier transit time inside the channel. Before introducing this parameter we briefly comment the shape of the velocity profile inside the channel region. Fig.4.11 illustrates the electron average drift velocity $v_d(x)$ calculated at each section of the channel as the average of the

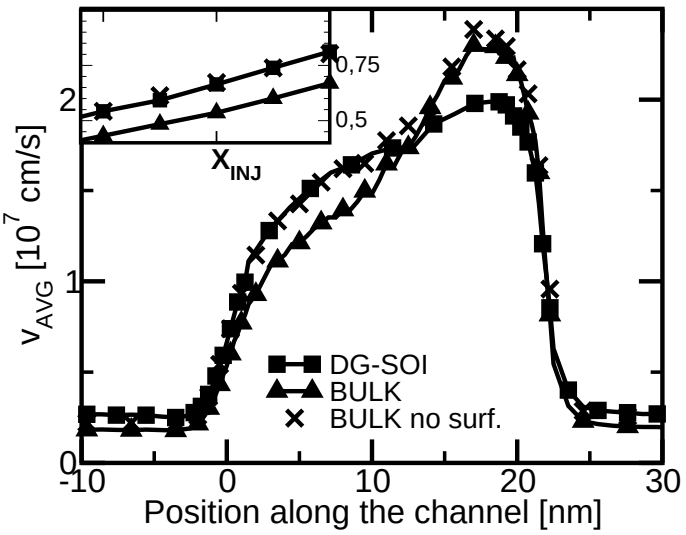

Figure 4.10: Average drift velocity along the channel for DG SOI and Bulk MOSFETs belonging to the 65 nm TN with $L_G = 25$ nm biased at $V_{GS} = V_{DS} = V_{DD}$. Crosses: bulk MOSFET with surface roughness turned-off. The origin of the x axis corresponds to the virtual source. The inset provides a magnification around the virtual source.

velocity x-component weighted by electron concentration in the direction normal to the silicon-oxide interface. The ballistic and non-ballistic cases are considered for different DG devices of the 45 nm technology node and it is shown that scattering significantly affects the average velocity within the whole channel region. Looking at the non-ballistic case the increase of the electric field, due to the scaling of L_G, leads to shorter L_{KT}, and therefore, to larger average velocity at the virtual source (see the inset of Fig. 4.11). In fact a smaller kT-layer corresponds to a reduced number of back-scattering events that contributes to the carrier flux with negative velocity. In agreement with the models proposed in [14], [15], in the ballistic limit approximation, the reduction of the gate length does not lead to an increase of velocity at the virtual source.
The average electron velocity determines the *transit time*, defined as:

$$\tau_{\mathrm{DC}} = \int_{x_{inj}}^{x_{end}} \frac{1}{v_{\mathrm{avg}}(x)} dx, \tag{4.1}$$

where x_{end} is the position near the drain at which the drift velocity becomes $v_{avg}(x_{end}) = v_T/2$ (v_T is the thermal velocity). The region within the channel where the average velocity is smaller gives the larger contribution to τ_{DC}: from Fig. 4.11 this region extends immediately beyond the injection point. In a ballistic device the minimum average velocity is found at x_{inj} and is equal to the thermal velocity: in this case the transit time is minimized. A

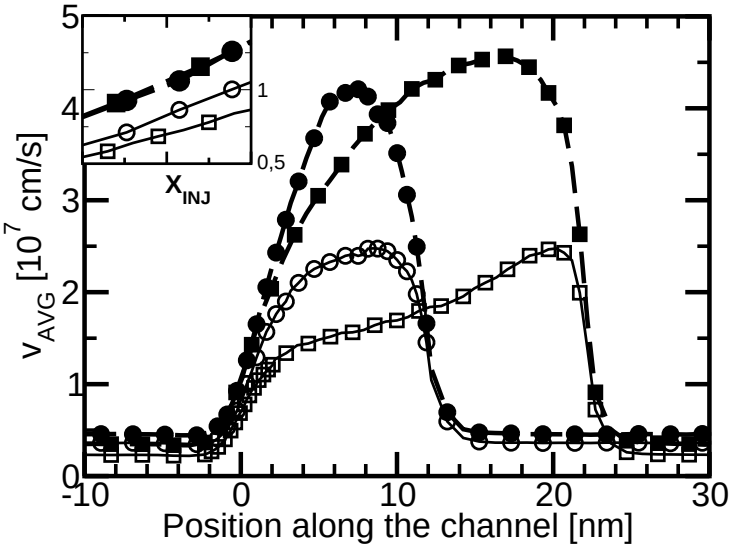

Figure 4.11: Average drift velocity along the channel for 14 nm and 25 nm L_G DG MOSFETs designed according to the 45 nm TN, biased at $V_{GS} = V_{DS} = V_{DD}$. Open symbols: simulation with scattering; filled symbols: ballistic case. The origin of the x axis corresponds to x_{inj}. The inset provides a magnified view at the virtual source.

non-ballistic device has smaller carrier velocity in the same region and τ_{DC} will be larger.
We are also able to relate the transit time and the drain current in the case of ballistic and non-ballistic simulation. According to the Charge Control Model, τ_{DC} is related to the transistor current (either I_{ON} or I_{BL}) as:

$$I_{ON} = \frac{Q_i}{\tau_{DC}} \quad ; \quad I_{BL} = \frac{Q_i}{\tau_{DC}^{BL}}, \tag{4.2}$$

where Q_i is the inversion charge in the channel of the device.
Fig. 4.12 reports τ_{DC} calculated for DG MOSFETs of the different technology nodes. Both the ballistic τ_{DC}^{BL} and the non-ballistic τ_{DC} values decrease by reducing the gate length. Note that the ballistic transit time decreases with the shrinking of L_G, even if I_{BL} stays almost constant (see Fig. 4.4), due to the reduction of the inversion charge Q_i with scaling-down (see equation 4.2). The results of Fig. 4.12 indicate that scattering plays a role even in the dynamic performance of short devices by increasing the transit time compared to the ballistic-limit case.

A strong correlation between the ratio τ_{DC}^{BL}/τ_{DC} and $BR = I_{ON}/I_{BL}$, is present too. Fig. 4.13 shows that the ratios τ_{DC}^{BL}/τ_{DC}, plotted for many L_G of different TNs, lay on the same curve of the ballistic ratios, shown

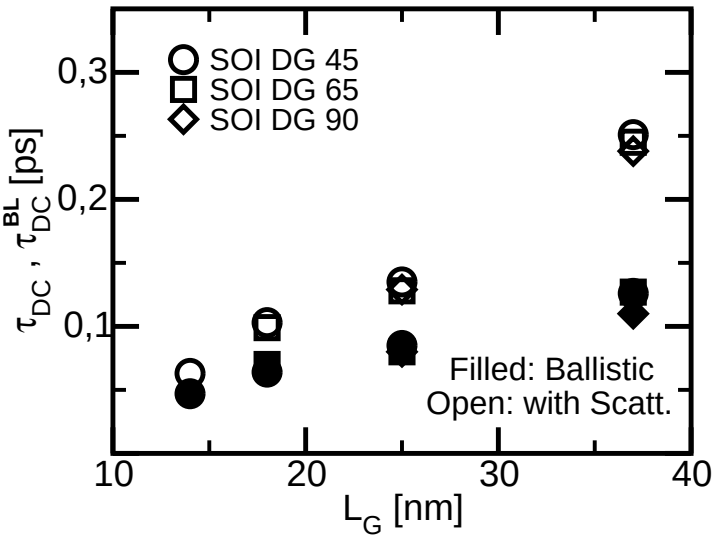

Figure 4.12: Average electrons transit time τ_{DC} versus L_G for DG MOSFETs biased at $V_{GS} = V_{DS} = V_{DD}$. Open symbols: simulations with scattering; filled symbols: ballistic case.

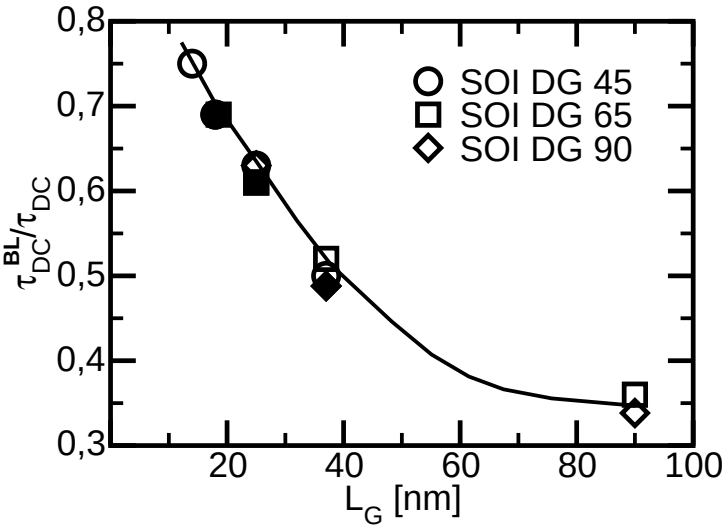

Figure 4.13: Ratio of the non-ballistic to ballistic average transit time from fig. 4.12 plotted vs. L_G. Filled symbols: devices with nominal gate length, solid line is the same curve of fig. 4.7.

in Fig. 4.7. We conclude that the ratio of the transit times is the same, unique function of the gate length as the ratio BR between the currents. The results reported in Fig. 4.13, together with eq. 4.2 implies that Q_i has to be the same in the ballistic and non-ballistic case. We numerically verified that this is approximately the case. As discussed in [53], τ_{DC} can be related to the small signal delay and therefore may provide important information about the performance of the device at RF. From our results we expect an improvement in terms of signal delay in advanced RF MOSFETs as the relevance of quasi-ballistic transport increases with the scaling of the gate length.

Chapter 5

RF Analysis of deca-nanometer bulk and SOI MOSFETs using the Time-Dependent Monte Carlo Approach

The continuous progress of the CMOS-IC technology has allowed the adoption of the MOSFET in the field of RF application [54] in view of the opportunity to realize system on chip for telecommunications adopting the same basic technology for digital circuits, baseband analog circuits and RF front-end. Thanks to technology scaling, the speed of the CMOS RF circuits has been increasing by about one order of magnitude every ten years. As a consequence, while nowadays MOSFETs are commonly adopted in the 0.8 - 2GHz range (GSM and Bluetooth), very good performance has already been demonstrated in the 20-40 Ghz range [55]. More recently, a CMOS receiver for the unlicensed band around 60 GHz has been realized [56].

As recognized in the Modeling and Simulation section of the ITRS [5], the accurate modeling of advanced MOSFETs operating at radio frequencies is a challenging objective, since quantization and non-equilibrium transport effects have to be properly taken into account. In the previous chapters we have demonstrated that the Monte Carlo method is one of the most appropriate techniques for the simulations of devices working in the quasi-ballistic transport regime. The same approach is applied here to the investigation of the RF performance of Bulk and Single Gate SOI MOSFETs designed according to the 2005 ITRS Roadmap. Section I will introduce the general concept about the evaluation of the RF performance of a MOSFET. Section II will review the Monte Carlo simulation procedure for AC simulation, in-

cluded in the MC code Bandit. Section III is dedicated to the analysis of the RF performance of advanced bulk and SOI MOSFETs: the signal-delay build-up along the channel, the scaling of the small signal parameters and of some RF figures of merit will be discussed in details. A useful comparison between the results of drift-diffusion and Monte Carlo simulations is also included in order to assess the advantages of the MC approach.

5.1 General analysis of the RF Performance of the MOS device

The evaluation of the RF performance of the MOSFET can be performed in different ways.In this section we discuss how to evaluate the RF performance of a MOSFET device, starting from the 2-port description through the admittance matrix Y. Most of these methods are included in many simulation tools. The next section will be dedicated to the implementation of the methods for RF analysis in the Monte Carlo simulation.

5.1.1 Description through Admittance matrix Y

The small-signal AC behavior of the MOSFET can be described through the 2-port admittance matrix that provides the following relationship among the F-transforms of the AC currents and voltages

$$\widetilde{I}_1 = \widetilde{Y}_{11}\widetilde{V}_1 + \widetilde{Y}_{12}\widetilde{I}_2$$

$$\widetilde{I}_2 = \widetilde{Y}_{21}\widetilde{V}_1 + \widetilde{Y}_{22}\widetilde{I}_2$$

by using the admittance parameters of the Y matrix

$$\mathbf{Y} = \begin{pmatrix} \widetilde{Y}_{11} & \widetilde{Y}_{12} \\ \widetilde{Y}_{21} & \widetilde{Y}_{22} \end{pmatrix}$$

Each $\widetilde{Y}_{ij}$ is, in general, a complex value $\widetilde{Y}_{ij} = \widetilde{G}_{ij} + j\omega\widetilde{C}_{ij}$ whose real part is a conductance and imaginary part is proportional to a capacitance. The mathematical definition of any matrix element is

$$\widetilde{Y}_{ij} = \left.\frac{\widetilde{I}_i}{\widetilde{V}_j}\right|_{\widetilde{V}_k=0 \quad k\neq j} \tag{5.1}$$

The above expression says that each matrix element is obtained as a ratio of the F-transforms of the current response at a i contact, and of the voltage

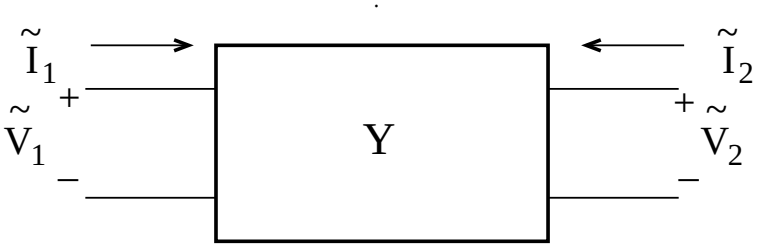

Figure 5.1: Typical 2-port configuration with Y matrix description.

variation at j contact. All contacts except the j one are grounded for signals. Even if very general, the 2-port analysis does not give many informations about the internal conduction and capacitive effects of the device. The small-signal description by *differential parameters* can be used for this purpose.

5.1.2 MOS Small-signal description by differential parameters

The description through a lumped-elements linear equivalent circuit is obtained from the current-voltage and charge-voltage relationships derived from the physics-based description of the device. The device frequency behavior is modeled through an equivalent circuit whose elements are the small signal parameters. It is less general than the Y-matrix description. More sophisticated approximations usually generate complex circuits. A simple equivalent circuit for the MOSFET is shown in Fig. 5.2 for the common-source configuration. All the parameters are calculated by differentiating the current-voltage and charge-voltage relationship around a bias point and keeping only the first term of the Taylor's series. The model includes all the intrinsic effects of the device while other extrinsic effects like series resistances, gate resistance, inductances, etc, are not considered here.
The lumped-elements of Fig. 5.2 are briefly described here:

Gate Transconductance $g_{m0} = \partial i_{DS}/\partial v_{GS}$ is a measure of the current driving of the device. It increases by scaling the device to smaller channel length.

Transconductance delay τ_{g_m} takes into account the delay of the the current response.

Drain Conductance $g_{ds} = \partial i_{DS}/\partial v_{DS}$ is related to the channel modulation if the device is biased in saturation. It is very large in short-channel MOSFETs due to short-channel effects.

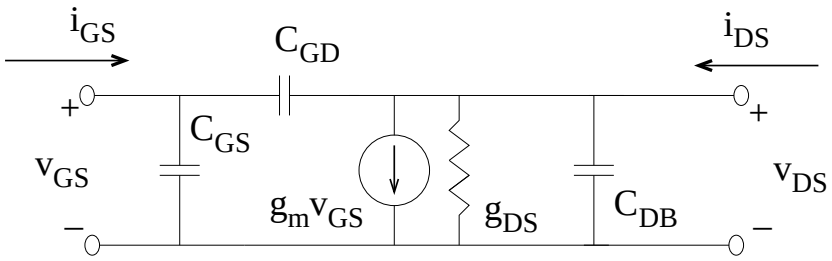

Figure 5.2: MOS small-signal model with differential parameters, where $g_m = g_{m0}e^{-j\omega\tau_{gm}}$.

Gate-source C_{GS} and gate-drain C_{GD} capacitances come from the charge-voltage relationships. The sum of these parameters gives the total gate differential capacitance C_{GG}. With the decrease of the transistor dimensions, these capacitances should decrease, thus increasing the device speed and the maximum operational frequency.

Drain bulk Capacitance C_{DB} takes into account the junction capacitance between the drain region and the substrate. Because it is not an intrinsic parameter, it would be useful to eliminate it. The SOI Fully-depleted MOSFET (see sec. 1.3.1), because of its thin body, has a negligible C_{DB}.

The Y matrix elements can be written as a function of the differential parameter by simple network analysis:

$$\widetilde{Y}_{11} = j\omega(C_{GS} + C_{GD}) \tag{5.2}$$
$$\widetilde{Y}_{12} = -j\omega C_{GD} \tag{5.3}$$
$$\widetilde{Y}_{21} = g_{m0}e^{-j\omega\tau_{gm}} - j\omega C_{GD} \tag{5.4}$$
$$\widetilde{Y}_{22} = g_{ds} + j\omega(C_{DB} + C_{GD}) \tag{5.5}$$

The differential parameter of the AC equivalent circuit can be expressed as a function of the admittance parameters

$$C_{GD} = -\frac{Im[\widetilde{Y}_{12}]}{\omega} \tag{5.6}$$
$$C_{GS} = \frac{Im[\widetilde{Y}_{12}] + Im[\widetilde{Y}_{11}]}{\omega} \tag{5.7}$$
$$C_{DB} = \frac{Im[\widetilde{Y}_{12}] + Im[\widetilde{Y}_{22}]}{\omega} \tag{5.8}$$
$$g_{ds} = Re[\widetilde{Y}_{22}] \tag{5.9}$$
$$g_{m0} = |\widetilde{Y}_{21} - \widetilde{Y}_{12}| \tag{5.10}$$

$$\tau_{gm} = -\frac{1}{\omega}\arctan\left(\frac{Im[\widetilde{Y}_{21}] - Im[\widetilde{Y}_{12}]}{Re[\widetilde{Y}_{21}]}\right) \tag{5.11}$$

where $\omega = 2\pi f$ is the radian frequency. The network shown in Fig. 5.2 is not valid at high frequency ($f > 100$ GHz) where many extrinsic effects are important and more complex models are necessary.

5.1.3 The drain-source delay τ_{ds}

The propagation delay between the source and the drain is a very important parameter because it limits the maximum operational frequency of the device. It is written as τ_{ds} and must not be confused with the transconductance delay τ_{gm} which is included in the small signal equivalent circuit.
The source-drain delay can be evaluated through two different approaches: the first and simplest one is the *Quasi-Static* approach, the second consists of a real frequency domain analysis. These methods are now briefly described.

The Quasi-Static Method. This approach assumes that the charge density inside the device responds immediately to any voltage variation at a contact, without memory effects [57]. It can demonstrated starting from the charge-continuity equation (eq. (2.8)). Here we do not report the demonstration but only the final result in the case of a nMOS: the following relation among the delay, the electron concentration and the current density can be found

$$\frac{d\tau_{ds}}{dx} = -q\frac{\Delta n(x)}{\Delta J} \tag{5.12}$$

where $\Delta n(x)$ and ΔJ are the variations of, respectively, the electron concentration and the current density due to the applied signal. The intrinsic delay $\tau_{ds}(x)$ is calculated by integrating eq. (5.12) along the channel

$$\tau_{ds}(x) = -q\int_0^x \frac{\Delta n(x)}{\Delta J}dx \tag{5.13}$$

If the integration is done along the entire device length, we obtain the total delay τ_{ds}. We notice that eq. (5.13) is independent of frequency.
It is important to remember that the quasi-static approximation is not valid at high frequency and in long-channel MOSFETs, where the charge density varies with non-negligible time constants.

Frequency-analysis method. It is possible to define the Fourier transfer function $\widetilde{\mathcal{H}}(\omega, x)$ that describes the propagation of the signal along the device. If we consider only small values of ω, this transfer function can be expanded

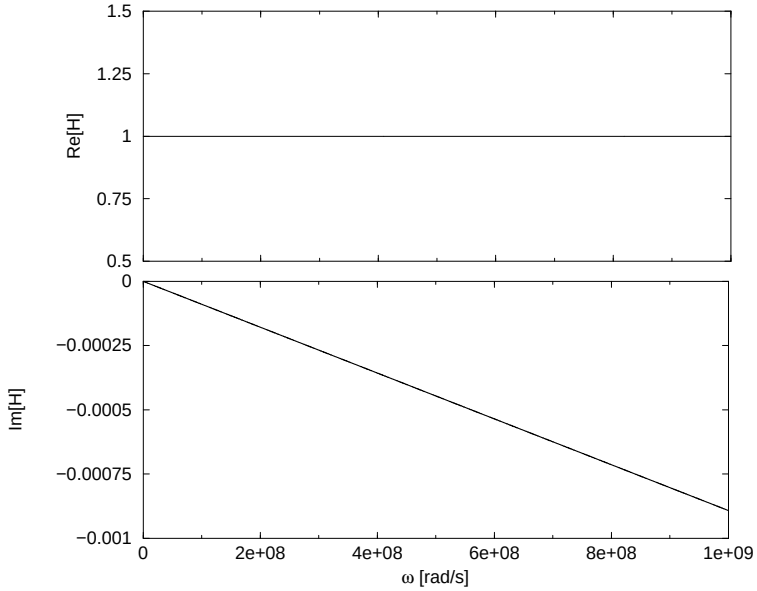

Figure 5.3: Example of real and imaginary part of $\mathcal{H}(\omega, x)$. For higher frequencies the curves deviate from the behavior given by (5.14).

in a Taylor series where only the first order term is kept. It is possible to demonstrate that

$$\widetilde{\mathcal{H}}(\omega, x) = \frac{\widetilde{J}_n(\omega, x)}{\widetilde{J}_n(\omega, 0)} \approx 1 - j\omega\tau_{ds}(x) \tag{5.14}$$

where $x = 0$ is the point inside the device which is used as reference, and $\widetilde{J}_n(\omega, x)$ is the Fourier-transformation of the current density. Figure 5.3 shows an example of $\widetilde{\mathcal{H}}(\omega, x)$ calculation. The information about the delay is within the phase of $\widetilde{\mathcal{H}}(\omega, x)$ and can be simply extracted from the imaginary part of the transfer function as

$$\tau_{ds}(x) = -\frac{Im[\widetilde{\mathcal{H}}]}{\omega} \tag{5.15}$$

This approach is valid for small frequencies. In general eq. (5.14) holds for $f < 100$ GHz.

5.1.4 Figures of Merit of the AC performance of the MOS

From the small-signal network introduced in sec. 5.1.2 we can easily evaluate the performance of a generic MOS device. For example the higher the

transconductance, the faster the transistor will be. Unfortunately the small-signal equivalent circuit assumes many approximations and can differ depending on the case under analysis. More general Figures of Merit (FOM) have been defined in order to quickly compare the AC performance among different MOSFETs. Here we briefly describe the most commonly used.

Transition Frequency F_T. If a MOSFET is in common-source configuration, we can evaluate easily a current gain with a short circuit as load. From the Y matrix description, this gain is simply

$$A_{i_{short}} = |\widetilde{Y}_{21}/\widetilde{Y}_{11}| \tag{5.16}$$

and is inversely proportional to the frequency f, if f is small. The transition frequency is evaluated from an extrapolation of the current gain at high frequency (in practice we neglect the deviation from the $1/f$ dependence) and then calculating the frequency where the gain is unitary. This frequency is called *Transition Frequency* and is indicated with F_T. From the small-signal equivalent circuit of Fig. (5.2), we can find this expression

$$F_T = \frac{g_{m0}}{2\pi(C_{GS} + C_{GD})} \tag{5.17}$$

which highlights the role played by transconductance and by the total capacitance at the input (gate) port.
The transition frequency, because obtained from an extrapolation procedure, is much higher than the real maximum operation frequency of the device but is considered a very useful figure of merit and is the first parameter to check when two devices are compared.

Voltage Gain and Cutoff Frequency F_{3dB}. We consider again the common-source configuration and we evaluate the voltage gain with an open circuit as load. The most general expression is again given by the admittance parameters and is

$$A_{V_{open}} = |\widetilde{Y}_{21}/\widetilde{Y}_{22}| \tag{5.18}$$

The frequency dependence of the voltage gain has the same shape as a low-pass filter, so a -3db corner frequency F_{3dB} can be calculated from the Bode diagram. From the small-signal equivalent circuit we can calculate the gain value at zero frequency as g_{m0}/g_{ds} while the -3dB corner frequency is

$$F_{3dB} = \frac{g_{ds}}{(2\pi C_{DD})} \tag{5.19}$$

Historically, in the RF circuits, the voltage gain and F_{3dB} was not considered an important figure of merit, but, because of SCE, the MOSFET has a drain transconductance that increases with scaling and the gain can degrade seriously in advanced devices. This issue explains why it is important to consider also the voltage gain diagram.

Maximum Oscillation Frequency F_{max}. The maximum oscillation frequency is defined as the frequency corresponding to the unity maximum available power gain (MAG). The MAG is calculated under power-matching conditions at both the input and output ports. The obtained F_{max} accounts for many type of losses like substrate distributed resistance or the gate resistance. Together with F_T, is the most popular figure of merit for RF MOSFETs. Due to the some limitation of our simulations, a correct evaluation of this frequency is not possible and it is not the case to further describe this, anyway important, figure of merit.

5.2 AC simulation in the Monte Carlo approach

In sec. 5.1 the AC characteristics of the MOS transistor have been discussed in general terms. In this section we briefly review the implementation of the above methods for RF analysis in a Monte Carlo simulator. The MC method is an adequate way to study non-equilibrium phenomena in the steady-state condition, but the AC analysis can be included without much effort. Following [58, 59, 60], the AC behavior of FETs is simulated by Fourier decomposition of the response to small voltage steps applied to the gate and the drain terminals, obtained by time-dependent MC simulations. The device is first simulated at the bias point. When this part of the simulation has reached convergence, a bias step is applied to either the gate or the drain terminal. The time-dependent current at each contact is obtained from a statistical estimator based on the extended Ramo-Shockley theorem [61], which is described in the last subsection.

The details about the implementation of the RF analysis in a MC code can be found in [41] and [62].

5.2.1 Drain-Source delay

The drain to source delay τ_{ds} is the propagation delay of the carrier through the channel. In sec. 5.1.3 we presented two methods to calculate it: the quasi-static one and the frequency-analysis. The quasi-static method is easy to perform because it requires only two separates DC simulations. Since the delay is calculated as a response to one or more voltage steps at different device contacts, we only need to perform a preliminary simulation at the starting bias point and a second one with the final voltage values at the corresponding contacts. Once the current and the inversion density have been extracted, eq. (5.13) is used to obtain τ_{ds}.
The frequency analysis method is based on eq. (5.14). Using that approximation of the transfer function, valid for small frequencies, the drain-source delay τ_{ds} is extracted from the imaginary part as expressed by eq. (5.15). The transfer function in the case of electron is calculated as

$$\widetilde{\mathcal{H}}(\omega, x) = \frac{\mathcal{F}[J_n(t,x) - J_{n0}(x)]}{\mathcal{F}[J_n(t,0) - J_{n0}(0)]} \tag{5.20}$$

which is the ratio between the Fourier transform of the current density response in a point x and the transform of the current density response at a reference point $x = 0$. The term $J_{n0}(x)$ is the value of the current density before the application of the rectangular pulse to a device contact. Eq. (5.20) is true for a rectangular pulse only; if the signal is a voltage step the expression must be changed into

$$\widetilde{\mathcal{H}}(\omega, x) = \frac{\mathcal{F}[J_n(t,x) - J_{n\infty}(x)] + \mathcal{F}[J_{n\infty}(x) - J_{n0}(x)]}{\mathcal{F}[J_n(t,0) - J_{n\infty}(0)] + \mathcal{F}[J_{n\infty}(0) - J_{n0}(0)]} \tag{5.21}$$

where $J_{n\infty}(x)$ is the value of the current density when the transient, due the voltage step, has finished. Once we have calculated the transform of the currents, the calculation of $\widetilde{\mathcal{H}}$ is quite straightforward. Here we present the final result, separating the real and imaginary part.

$$\begin{aligned}
\widetilde{\mathcal{H}}(x,\omega) &= \frac{(\Delta J_n(x) - \omega Im[\widetilde{J}_n(x)])(\Delta J_n(0) - \omega Im[\widetilde{J}_n(0)]) + \omega^2 Re[\widetilde{J}_n(x)] Re[\widetilde{J}_n(0)]}{(\Delta J_n(0) - \omega Im[\widetilde{J}_n(0)])^2 + \omega^2 (Re[\widetilde{J}_n(0)])^2} + \\
&+ j\frac{\omega Re[\widetilde{J}_n(x)][\Delta J_n(0) - \omega Im[\widetilde{J}_n(0)]) - \omega Re[\widetilde{J}_n(0)](\Delta J_n(x) - \omega Im[\widetilde{J}_n(x)])}{(\Delta J_n(0) - \omega Im[\widetilde{J}_n(0)])^2 + \omega^2 (Re[\widetilde{J}_n(0)])^2} = \\
&= Re[\widetilde{\mathcal{H}}(x,\omega)] + jIm[\widetilde{\mathcal{H}}(x,\omega)].
\end{aligned} \tag{5.22}$$

where $\Delta J_n(x) = J_{n\infty}(x) - J_n(x)$ and $\widetilde{J}_n(x)$ is the F-transform of the current density.

The imaginary part in eq. (5.22) is then used to calculate the delay τ_{ds}. From the point of view of numerical calculation, in order to use the eq. (5.22), we only need to keep track of the current density $J_n(x)$ at each section x of the device during all the simulation time.

5.2.2 Admittance Matrix

Each Y matrix element is calculated as

$$\widetilde{Y}_{ij} = \frac{\mathcal{F}[I_i(t) - I_i(0)]}{\mathcal{F}[V_j(t) - V_j(0)]} \tag{5.23}$$

where $I_i(t)$ and $V_j(t)$ are the sum of a dc bias ($I_i(0)$ and $V_j(0)$ respectively) and a perturbation. The expression above is valid for a rectangular pulse. In the case of a voltage step $V_j(t) = V_j(0) + \Delta V_j u(t)$, the final current is $I_i(\infty) \neq I_i(0)$ and thus

$$\widetilde{Y}_{ij} = \frac{\mathcal{F}[I_i(t) - I_i(0) + I_i(\infty) - I_i(\infty)]}{\mathcal{F}[\Delta V_j]} \tag{5.24}$$

We do not show the mathematical passages after eq. (5.24). It can be demonstrated that the matrix element is calculated as

$$\begin{aligned}
\widetilde{Y}_{ij} &= \frac{I_i(\infty) - I_i(0)}{\Delta V_j} + \frac{j\omega}{\Delta V_j}\int_0^T [i_i(t) - I_i(\infty)]e^{-j\omega t}dt = \\
&= \frac{I_i(\infty) - I_i(0)}{\Delta V_j} + \frac{\omega}{\Delta V_j}\int_0^T [I_i(t) - I_i(\infty)]\sin(\omega t)dt + \\
&+ \frac{j\omega}{\Delta V_j}\int_0^T [I_i(t) - I_i(\infty)]\cos(\omega t)dt = \\
&= Re[\widetilde{Y}_{ij}] + jIm[\widetilde{Y}_{ij}]
\end{aligned} \tag{5.25}$$

The real and imaginary part of any $\widetilde{Y}_{ij}$ can be easily evaluated once we track the current at the i contact over the simulation time. The time T after the voltage perturbation must be large enough to include the entire current response. From the admittance matrix it is possible to extract the small signal equivalent circuit following the expressions (5.6), and the figures of merit (equations (5.16) and (5.18)).

5.2.3 The Ramo-Shockley theorem

The DC current obtained from a Monte Carlo simulation is usually evaluated by averaging the current at all device sections. If the simulation statistics

have converged, the value of $I(x)$ is rather constant along the x position. The variation among the sections is reduced by simulating more particles for a longer time.
The definition of current related to a specific contact is not straightforward in the MC case but the *Ramo-Shockley* gives an easy way to calculate it [61]. Following this theorem, the instantaneous current at a contact i is given by

$$i_i(t) = I_i'(t) + I_i''(t) \tag{5.26}$$

where $I_i'(t)$ is a contribution given by all particles under study and $I_i''(t)$ is a contribution from the capacitive coupling among the contacts.
If the entire system contains N particles, the $I_i'(t)$ is given by

$$I_i'(t) = -\sum_{j=1}^{N} Q_j v_j(t) \cdot \nabla f_i \tag{5.27}$$

where Q_j is the charge of the j superparticle, v_j is its velocity and f_i is a spatial potential profile found as follows: the voltage at the j contact is set to unity and all the other contacts are grounded, then the Laplace equation $\nabla \cdot (\epsilon \nabla f_i) = 0$ is solved. From eq. (5.27) we see that the voltage biases at the contacts do not influence directly the current, but indirectly through the velocities of the particles.
The second term $I_i''(t)$ is an impulsive contribution, which is relevant only in the step immediately after the application of the voltage impulse. Its expression is

$$I_i''(t) = C_{ij} \frac{\Delta V_j}{\Delta t} \tag{5.28}$$

where ΔV_j is the step amplitude at the j contact and Δt is the duration of the first simulation step. After this time $I_i''(t)$ drops to zero. C_{ij} account the capacitive coupling between the i and the j contacts.
With the implementation of eqs. (5.27) and (5.28), the MC simulator can calculate the currents at all contacts and the evaluation of the admittance matrix is possible.

5.3 RF Analysis of deca-nanometer bulk and SOI MOSFETs using the Time-Dependent Monte Carlo Approach

In this section, our Monte-Carlo simulator, already introduced in sec. 2.2.3 and upgraded with the time-dependent analysis presented in sec. 5.2, is applied to the investigation of the RF performance of Bulk and Single Gate

SOI MOSFETs designed according to the prescriptions of the 2005 ITRS Roadmap. The first subsection is, as usual, dedicated to the description of the analyzed devices for AC RF simulation. The following subsections discuss, in order: the signal-delay along the channel, the scaling of the small signal parameters, the scaling of some common RF FOM. Finally we present a comparison between the results of AC drift-diffusion simulation from Dessis and the result from the Monte Carlo Bandit AC simulations. The results of this section have been presented in [64].

5.3.1 Device design and description of the Monte Carlo approach

We have considered n-MOSFETs representative of realistic low standby power transistors (LSTP) for analog/mixed-signal applications (AMS) in the years 2006, 2007 and 2008 [5]. The simulated devices are both Bulk MOSFETs (whose sketch is shown in Fig. 5.4) and Ultra-Thin Body (UTB) Single-Gate (SG) SOI MOSFETs. Their main characteristics are summarized in Table 5.1. In the bulk case the substrate doping profile has been tailored, through the use of halos, in order to meet the Roadmap specification on drain OFF-current and threshold voltage. Drain and source feature a double-diffusion with a low-doped shallow region and a highly-doped region separated by a spacer to limit short channel effects. The gate electrode is a highly doped n-poly with metallic behavior (no poly-depletion effect). Silicon dioxide is assumed for the gate dielectric.

The UTB-SG SOI MOSFETs has been analyzed in order to discuss their specific features, concerning applicability to analog and RF circuits. The simulated transistors feature a thin almost-undoped substrate with the same source/drain doping profiles assumed for the corresponding bulk MOSFET. A metal gate electrode is assumed, with workfunction tailored for each device in order to set the same threshold voltage obtained for the corresponding bulk MOSFET. Short channel effects (SCE) and off-current are controlled by adopting a thin body whose thickness is scaled with the channel length.
Unless differently stated, all the results presented in the following sections have been calculated for a bias condition defined by: V_{GS}=V_{DS}=0.9 V.
The amplitude of the voltage steps applied at the gate and drain contacts is of critical importance in the MC approach. In fact if the voltage step is too small, the evaluation of the associated variations of gate and drain currents is degraded by the statistical noise affecting Monte Carlo results.

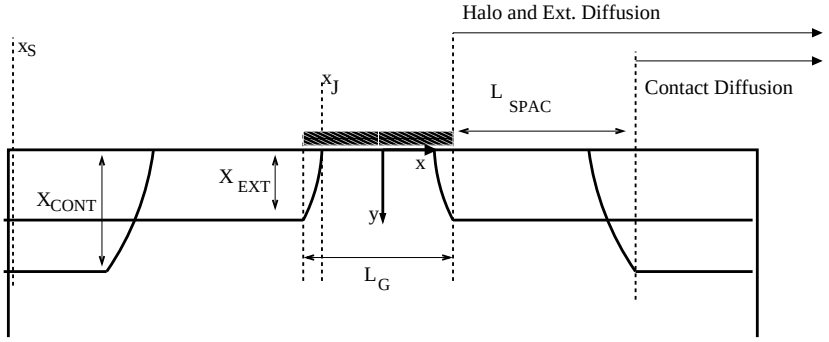

Figure 5.4: Sketch of the simulated bulk MOSFETs. x_J: position of the source junction; x_S: lateral boundary at the source end.

	Bulk			**UTB-SG SOI**		
L_G	**53 nm**	**45 nm**	**37 nm**	**53 nm**	**45 nm**	**37 nm**
V_{DD} **(V)**	**1.2**	**1.2**	**1.1**	**1.2**	**1.2**	**1.1**
t_{OX} **(nm)**	**2**	**1.9**	**1.6**	**2**	**1.9**	**1.6**
N_a **(cm^{-3})**	$2 \cdot 10^{18}$	$2.2 \cdot 10^{18}$	$2.3 \cdot 10^{18}$	$1 \cdot 10^{15}$		
I_{OFF} **(pA/μm)**	**5**	**7**	**9**	**0.3**	**0.7**	**1**
$V_{TH,lin}$ **(V)**	**0.63**	**0.63**	**0.62**	**0.65**	**0.66**	**0.65**
DIBL (mV/V)	**100**	**98**	**105**	**110**	**112**	**117**
S/D X_{EXT} **(nm)**	**11**	**10**	**8**	**8**	**7.2**	**6.7**
S/D X_{CONT} **(nm)**	**28**	**26**	**25**	**8**	**7.2**	**6.7**
L_{SPAC} **(nm)**	**27**	**25**	**22**	**27**	**25**	**22**

Table 5.1: Main characteristics of the simulated devices. Gate dielectric is SiO_2. Bulk MOSFETs include halos. In the UTB-SG SOI, $t_{SI} = X_{EXT} = X_{CONT}$ is set in order to keep I_{OFF} below the ITRS LSTP limits with a low-doped silicon. $V_{TH,lin}$ is the threshold voltage in linear region.

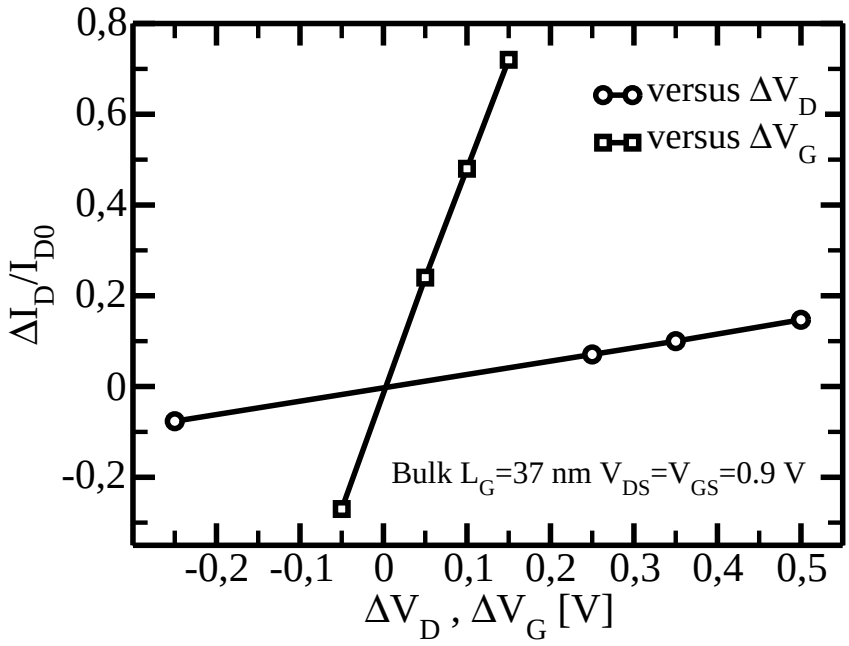

Figure 5.5: Drain current variation normalized to the bias current $\Delta I_D/I_{D0}$ versus the amplitude of the gate step and of the drain step. The initial bias is $V_{GS} = V_{DS} = 0.9$ V.

On the other hand, a too large step produces a non-linear response. Fig 5.5 shows the dependence of drain current asymptotic variation ΔI_D induced by voltage steps applied at the drain and gate terminals of the shortest bulk device, starting from the nominal bias point. The variation of drain current associated to ΔV_D is quite small and linearly related to ΔV_D for positive steps as large as 0.5 V. In this work we adopt positive ΔV_D in the range 0.3 - 0.5 V, depending on the gate length (the smallest step corresponding to the shortest device). As, for ΔV_G, due to the larger sensitivity of drain current to gate voltage, we adopt positive steps in the range 0.05 - 0.1 V, increasing with gate length. In order to reduce the impact of statistical noise, the waveforms for drain and gate currents are evaluated as an average of the results of five simulations that differ only for the sequence of random-numbers adopted in the Monte Carlo algorithm.

5.3.2 Signal-propagation delay along the MOSFET's channel

The AC signal propagation delay along the channel is evaluated by applying a small voltage step ΔV_G to the gate terminal and evaluating the ratio, given by eq. 5.22. This ratio is calculated from the Fourier-transforms of the responses

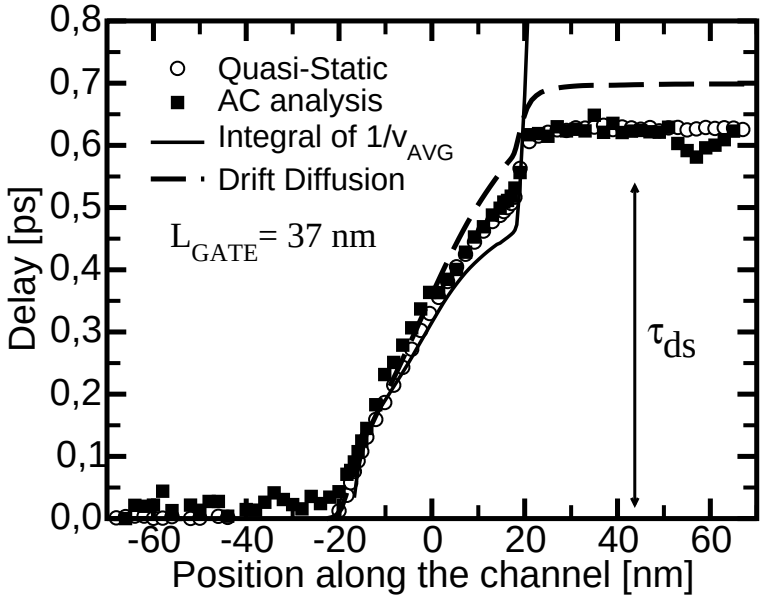

Figure 5.6: Delay along the channel for the 37 nm bulk MOSFET biased at $V_{GS} = V_{DS} = 0.9$ V. The origin of x-axis corresponds to the center of the channel; open circles: Monte Carlo QS analysis; squares: Monte Carlo AC analysis; dashed line: drift-diffusion QS analysis; solid line: $\int_{x_j}^{x} v_{avg}^{-1}(x')dx'$ applied to Monte Carlo DC simulation for the nominal bias. The d

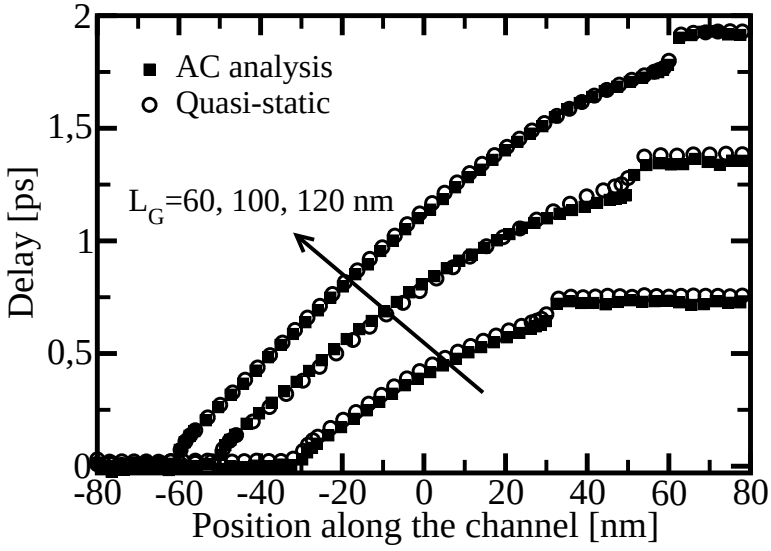

Figure 5.7: Delay along the channel for UTB SG SOI MOSFET with same characteristic as the 37 nm template device but with longer L_G. Open circles: MC-QS analysis; squares: MC-AC analysis.

of the current $I(x,t)$ at the generic abscissa x along the channel, and close to the source contact (x=x_S, see Fig. 5.4). As discussed in 5.1.3, at relatively low radian frequency ω the ratio of the transfer function can be approximated by its first-order expansion as $1 - j\omega\tau_{ds}(x)$ and the position-dependent delay is calculated according to:

$$\tau_{ds}(x) = -Im\{H(x,\omega)\}/\omega. \tag{5.29}$$

The approximation in eq. 5.29 provides accurate frequency-independent results for $\tau_{ds}(x)$ up to ω values larger than the operational frequencies of practical interest for our devices. Fig. 5.6 report the calculated signal propagation delay along the channel of the bulk MOSFETs with L_G=37 nm. The total delay from source to drain is referred as τ_{ds} and includes the charge build-up due to the electrons moving within the channel and due to the charging of the overlap regions. The relationship between τ_{ds} and the delay due to the channel alone, and the relationship with the delay time between the gate voltage step and the drain current are analyzed in detail in [63].

The propagation delay provided by this AC analysis can be compared to the result of a quasi-static (QS) approach. The QS delay is given by eq. (5.13) in order to use that expression we need the variation of inversion charge surface-density and current along the channel due to ΔV_G. These variations are calculated from two DC simulations performed for the nominal bias and for gate voltage incremented by the gate voltage step ΔV_G. Our aim is to investigate the validity of the QS approximation that is particularly useful, as it requires only two DC simulations instead of the five long transient analysis followed by waveform averaging and post-processing required by the AC analysis.
Fig. 5.6 demonstrates remarkable agreement between the QS-approximation and the result of AC analysis. The agreement between QS and AC analysis is further confirmed by Fig. 5.7 reporting the results obtained for UTB SG SOI MOSFETs featuring the same structure of the nominal 37 nm UTB-SOI MOSFET (Table 5.1) but with gate length increased up to 120 nm. These results confirms the validity of the quasi-static approximation even for relatively long-channel MOSFETs, allowing the estimation of the cut-off frequency $F_T \approx 1/(2\pi\tau_{QSds})$ and gate capacitance $C_{GG} \approx g_m\tau_{QSds}$, where τ_{QSds} represents the total signal delay from source to drain.
In addition, in Fig. 5.6 we show that, within the intrinsic channel region, the channel delay is well-approximated by:

$$\tau_{ds}(x) \approx \int_{xj}^{x} v_{avg}^{-1}(x')dx', \tag{5.30}$$

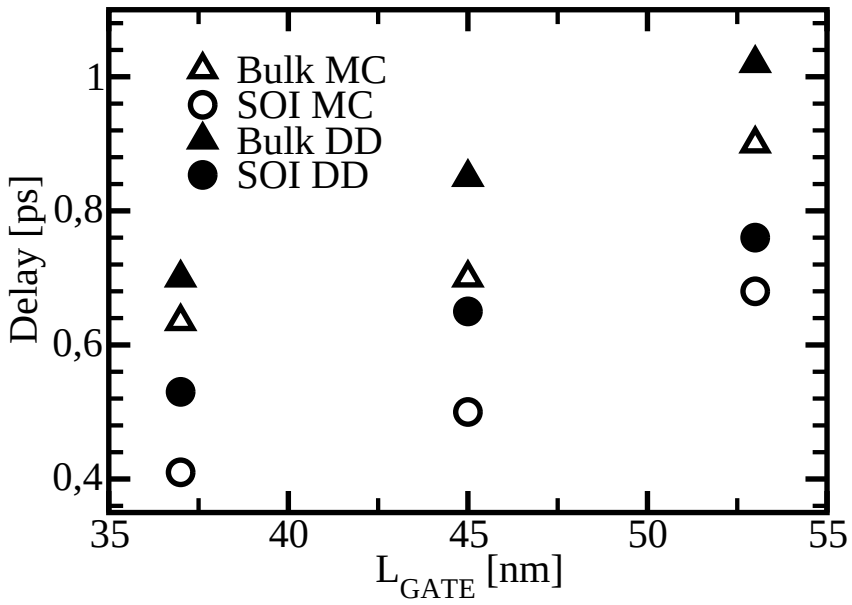

Figure 5.8: QS-delay as a function of gate length for the bulk and UTB-SG SOI MOSFETs calculated by Monte Carlo (open symbols) and drift-diffusion (filled symbols). The drift-diffusion results are discussed in section 5.3.5. Bias: $V_{GS} = V_{DS} = 0.9$ V.

$v_{avg}(x)$ being the DC electron velocity, x_j the position of the source-to-channel junction. This approximation, as might be demonstrated by a distributed analysis involving the continuity equation, holds within the QS approach under the assumption that the voltage step V_G does not affect the velocity $v_{avg}(x)$ inside the channel. The latter approximation loses validity in the source/drain quasi-neutral regions since the response of an high doped region to an external perturbation decays according to the dielectric relaxation time:

$$\tau_D = \frac{\epsilon_S}{qN_{SD}\mu} \tag{5.31}$$

where ϵ_S is the permittivity of silicon, N_{SD} the doping level in the source/drain regions and μ the mobility [65]. In our devices $\tau_D \approx 0.5$ fs which results in a negligible propagation delay in the source/drain regions. Therefore the integration in eq. 5.30 must not include the quasi-neutral regions otherwise their contribution, due to an average velocity $v_{avg} \rightarrow 0$, can greatly overestimate $\tau_{ds}(x)$. The effect of an incorrect integration is visible in Fig. 5.6 where the integral diverges inside the drain quasi-neutral region. The approximated expression for signal delay provided by equation 5.30 is particularly relevant as it shows that, while the DC on-current is related to the injection velocity at the virtual source only [15], the whole average-velocity profile along the channel is relevant and must be accurately evaluated for the calculation of

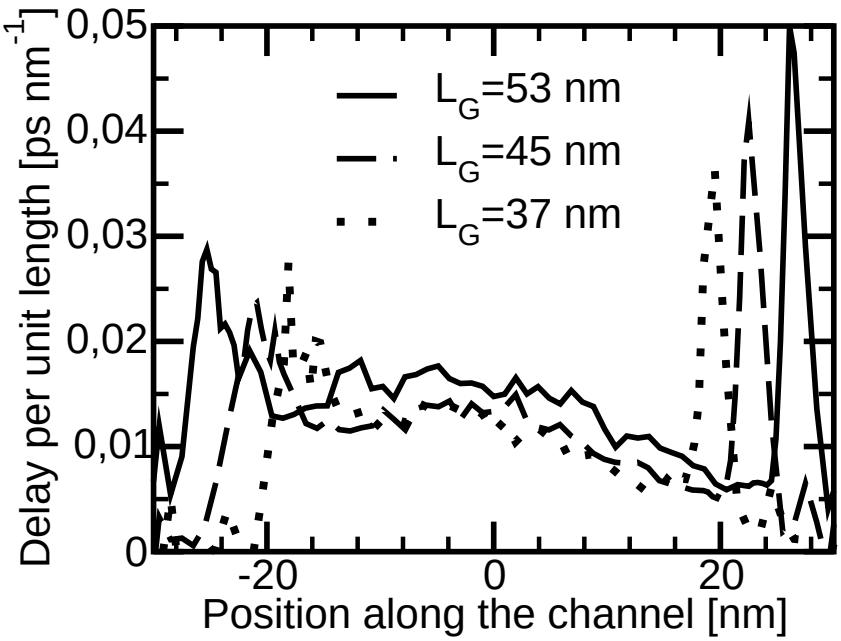

Figure 5.9: QS-delay per unit-length ($-q\Delta N_{inv}/\Delta I$) vs. position for the three Bulk MOSFETs; x=0: center of the channel. The largest delay per unit length occurs in the drain overlap regions. Delay within the intrinsic channel decreases for decreasing L_G as v_{avg} increases.

the signal propagation delay (and related AC figures of merit such as F_T).

Figure 5.8 compares the total source-drain signal delays evaluated for the bulk and UTB-SOI devices. τ_{QSds} is reduced as L_G is scaled down; furthermore, the bulk devices present systematically larger delays as a consequence of a lower average drift velocity along the channel due to a larger effect of surface-roughness scattering, caused by the stronger confinement of inversion charge towards the silicon-dielectric interface (as discussed in 4.2.2 and [52]). Figure. 5.9 reports the quasi-static delay per unit-length as a function of position along the channel for the three bulk devices; the integral from source to drain of this quantity provides the QS total source-drain delay τ_{QSds}. The largest contribution to τ_{QSds} comes from the intrinsic-channel region, while the drain and source overlap regions provide the largest specific delay per unit length, due to the large capacitive coupling to the gate electrode.

5.3.3 Frequency-domain analysis

In this section the RF performance of MOSFETs is analyzed by calculating the frequency-dependent admittance parameters of the device connected in

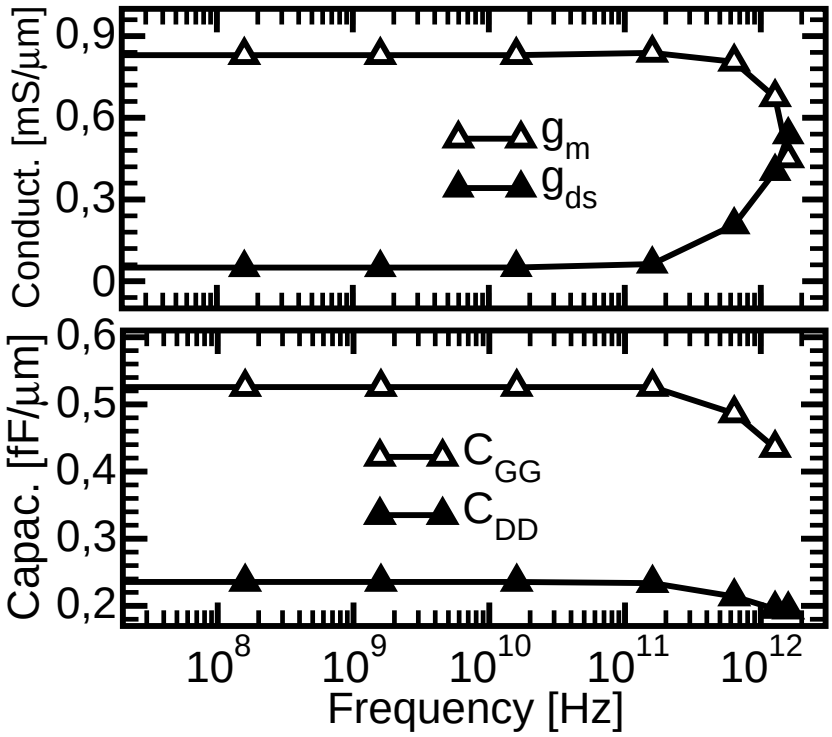

Figure 5.10: Frequency dependence of the AC-evaluated small signal parameters for the L_G=37 nm Bulk MOSFET biased at $V_{GS} = V_{DS} = 0.9$ V.

common source-bulk configuration. The time-dependent Monte Carlo simulation evaluates the AC admittance matrix following the equations presented in 5.2.2.

Figure 5.10 reports the frequency dependence of some of the small-signal parameters of the $L_G = 37$ nm bulk MOSFET: the parameters are independent of frequency up to $\omega \approx 10^{11}$ rad/s.
Figures 5.11 and 5.12, report the calculated transconductance $g_m = Re[\widetilde{Y}_{21}]$ and gate capacitance C_{GG} extracted from the admittance matrix Y obtained from Monte Carlo AC analysis, as a function of gate length. Both parameters present the expected dependence on L_G with increasing g_m and decreasing C_{GG} as L_G is scaled down. Notice that the UTB-SG SOI MOSFETs feature larger transconductances and lower gate capacitances compared to their bulk counterparts at given L_G. These effects can be related to a lower parallel electric field in the SOI MOSFETs, due to the absence of depletion charge in these undoped-channel devices. As a consequence, in the case of the UTB-SG SOI MOSFETs the centroid of inversion charge is more displaced from the silicon-dielectric interface, leading to reduced scattering with the rough

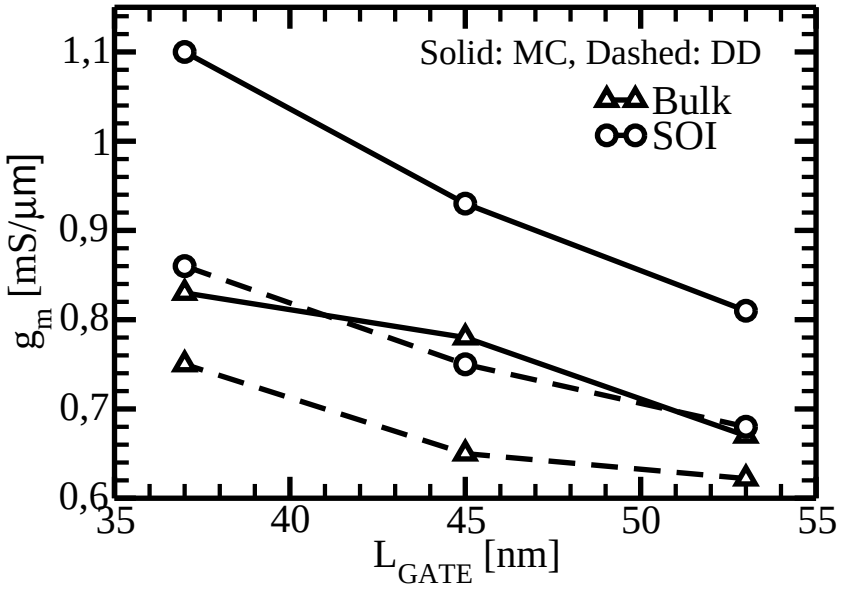

Figure 5.11: Transconductance from Monte Carlo and drift-diffusion AC simulations, as a function of gate length for bulk and UTB-SG SOI MOSFETs. Bias: $V_{GS} = V_{DS} = 0.9$ V.

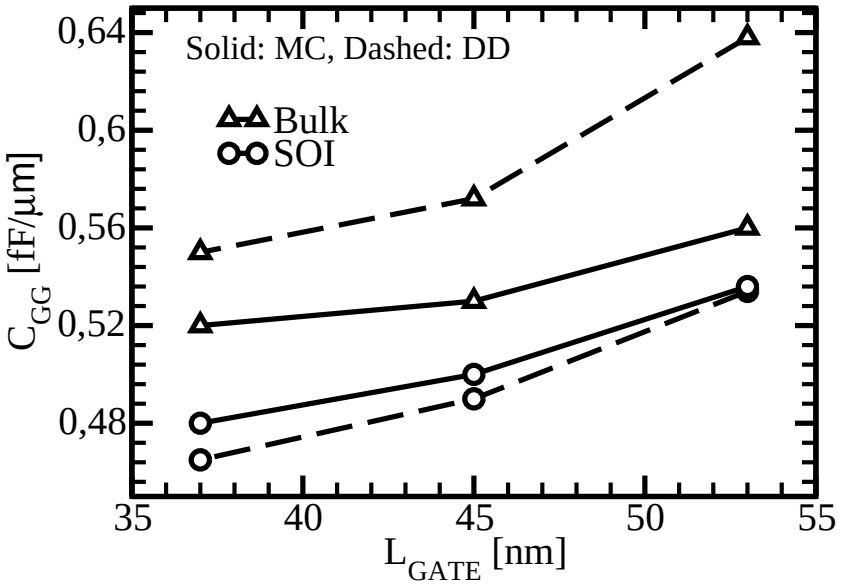

Figure 5.12: Gate capacitance $C_{GG} = Im[\widetilde{Y}_{11}]$ from AC Monte Carlo simulations, as a function of gate length for bulk and UTB-SG SOI MOSFETs. Bias: $V_{GS} = V_{DS} = 0.9$ V.

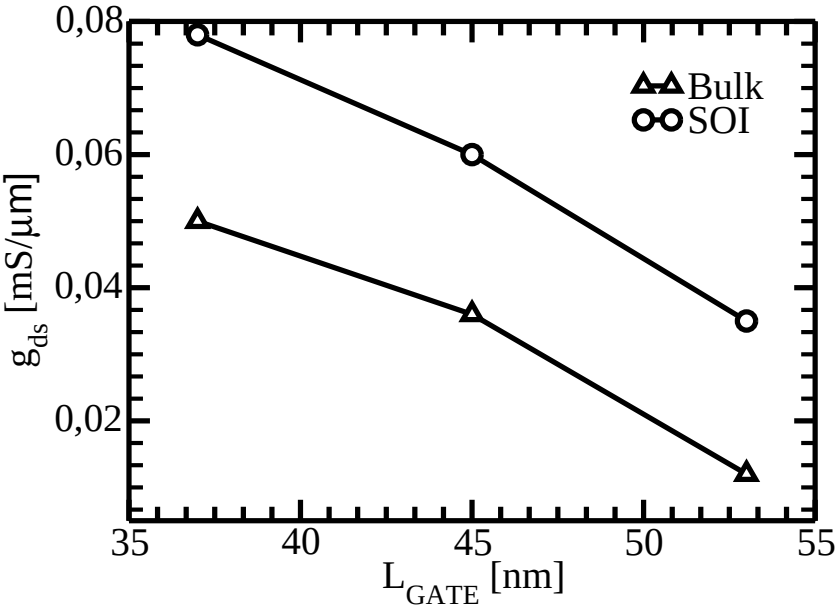

Figure 5.13: Scaling of AC Monte Carlo evaluated g_ds=Re[$\widetilde{Y}_{22}$] for Bulk and UTB-SG SOI MOSFETs. Bias: $V_{GS} = V_{DS} = 0.9$ V.

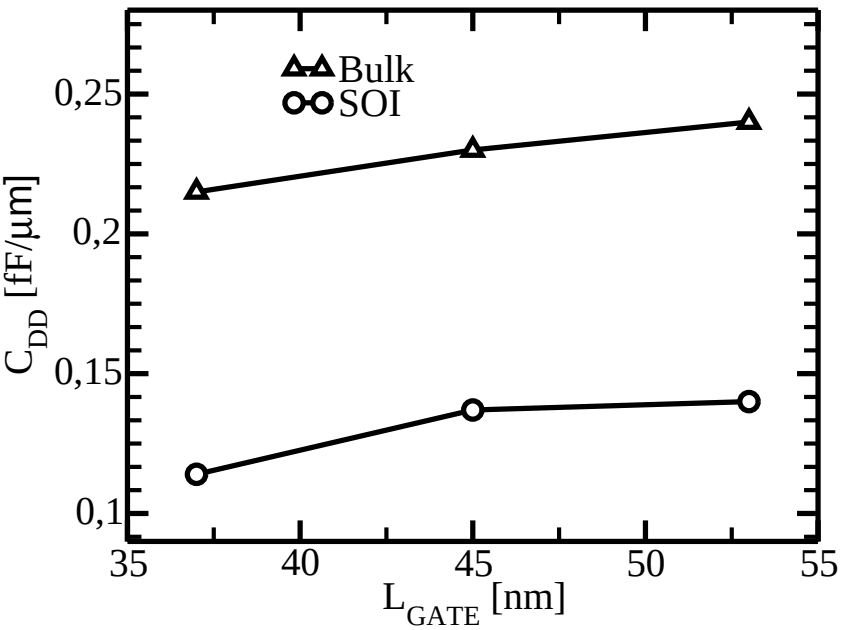

Figure 5.14: Scaling of AC Monte Carlo evaluated $C_{DD} = Im[\widetilde{Y}_{22}]$ for Bulk and UTB-SG SOI MOSFETs. Bias: $V_{GS} = V_{DS} = 0.9$ V.

interface, and attenuated gate capacitance.

Figures 5.13 and 5.14 report the drain capacitance C_{DD} and the drain conductance $g_{ds} = Re[\widetilde{Y}_{22}]$ extracted from the admittance parameters obtained by MC simulations. Notice that while g_{ds} significantly increases as L_G is scaled down, C_{DD}, that is related to the junction capacitance and to gate-drain overlap capacitance, is only slightly dependent on L_G. We may also observe that the UTB-SG SOI MOSFETs feature larger output conductance, due to slightly larger SCE, and significantly lower C_{DD} values due to the suppression of the large junction parasitic capacitance affecting the bulk MOSFETs.

5.3.4 RF figures of merit

In this subsection some relevant figures of merit (FOM) for analog and RF operation of MOSFETs are analyzed.
The cut-off frequency F_T has been defined in sec. 5.1.4 as the frequency for which the open circuit current gain $A_{ishort} = |\widetilde{Y}_{21}|/|\widetilde{Y}_{11}|$ of a common-source/bulk MOSFET drops to 0dB. On the basis of AC linear analysis F_T can be also expressed as a function of the small-signal parameters as: $F_T = g_m/(2\pi C_{GG})$. The F_T values estimated from AC MC simulations are reported in Fig 5.15 as a function of gate length. In the inset of the same figure we highlight the scaling of A_{ishort} at frequency values close to F_T. As expected, scaling the gate length causes an increase of the cut-off frequency for both bulk and SOI devices. The simulated F_T values for bulk devices are larger than the requirements by the ITRS 2005 but we have to mention that our simulations do only partially include the effects of parasitic source resistance and do not account for poly-depletion effects, both effects producing a degradation of the transconductance. The UTB-SG SOI MOSFETs feature larger transition frequencies thanks to larger transconductance and lower gate capacitance (Figs. 5.11, 5.12).
In section 5.1.4 we introduced another relevant FOM for the RF MOSFET: the 3dB bandwidth (F_{3dB}) of the open-circuit voltage gain A_{Vopen}=$|\widetilde{Y}_{21}/\widetilde{Y}_{22}|$. Figure 5.16 reports the MC-calculated F_{3dB} for bulk and UTB-SG SOI devices. A_{Vopen} features a low-frequency gain given by g_m/g_{ds} and a 3dB upper frequency that can be expressed as: $F_{3dB} = g_{ds}/(2\pi C_{DD})$. The results of MC simulation confirm that scaling degrades the low-frequency gain and increases the bandwidth due to the increase of g_{ds}. It should be noticed that the also low-frequency gain is overestimated by our simulations, because of the non-proper treatment of the source/drain resistances and poly-depletion effects. The UTB-SG SOI devices provide larger bandwidth compared to the bulk

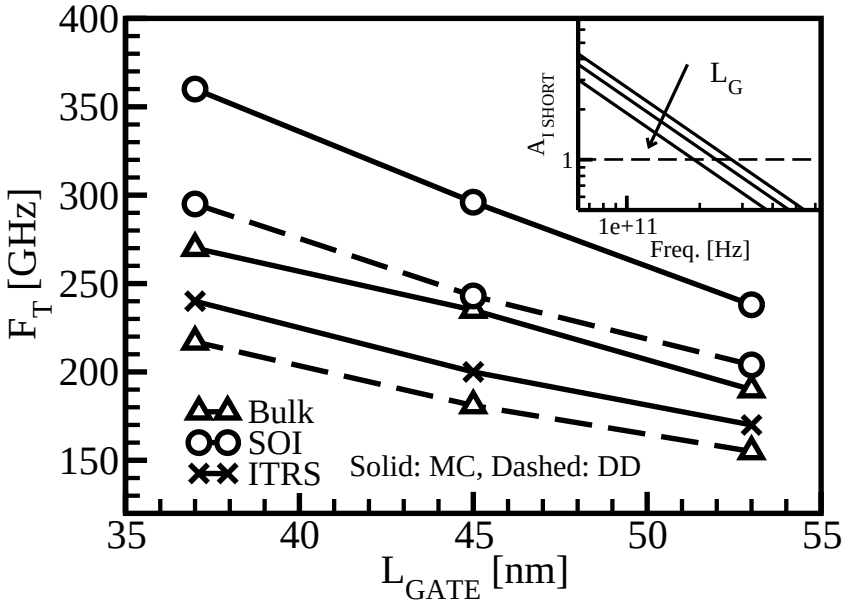

Figure 5.15: Scaling of the cut-off frequency F_T from AC drift-diffusion and Monte Carlo simulations. F_T is extracted from the short-circuit current-gain $A_{Ishort} = |\widetilde{Y}_{21}/\widetilde{Y}_{11}|$. The inset reports A_{Ishort} vs. frequency close to F_T.

counterparts, thanks to the lower C_{DD} (Fig. 5.14) and larger g_{ds} (Fig. 5.13).

5.3.5 Comparison of Monte Carlo and drift-diffusion AC analysis

Simple and efficient drift diffusion simulations have been carried out in order to perform a comparison between different transport models. In this subsection these results are compared with the result from Monte-Carlo simulations. As in the previous chapters, we have adopted the commercial program Dessis , presented in section 2.1.3, that features: (i) charge quantization through the density gradient approach, (ii) physical models for transport in bulk silicon which are consistent with the MC models, as they provide essentially the same mobility as a function of doping and longitudinal field.

Figure 5.6 compares the quasi-static signal delay along the channel calculated from drift-diffusion DC simulations using the QS approach (Eq. (5.13)) to the results of Monte Carlo QS and AC analysis. As might be expected the drift-diffusion predicts a larger delay mainly due to the lower average velocity along the channel compared to Monte Carlo. Figure 5.8 confirms that the total source/drain signal delay is significantly overestimated by drift-diffusion

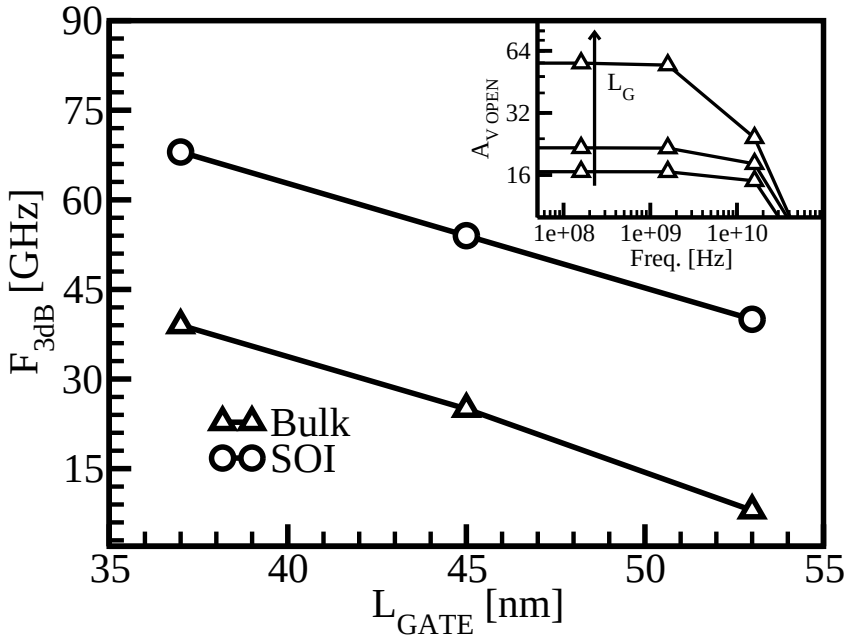

Figure 5.16: Scaling of AC-evaluated -3dB frequency (F_{3dB}) of the open-circuit voltage gain A_{Vopen} for the bulk and UTB-SG SOI MOSFETs. The inset reports A_{Vopen} vs. frequency for the Bulk case. The low-frequency gain is g_m/g_{ds}.

for all the devices considered in this study.

Figure 5.11 shows that drift-diffusion predicts lower values for the device transconductance due to the failure of this collision-limited transport model based on the concept of mobility, in the limit of very short devices. Consistently with a reduced transconductance and a lower average carrier velocity along the channel, in Figure 5.15 the drift-diffusion AC simulation provides a significantly lower F_T values compared to the Monte Carlo analysis, the discrepancy being larger in the case of the UTB-SG SOI MOSFETs.

Part III

Modeling and Characterization of Coupling in Multi-Gate MOS Devices

Chapter 6

Coupling in Ultra-Thin Body SOI MOSFET

In Part I we introduced the SOI technology and described its main characteristics. In particular we divided the SOI family in two categories: Partially Depleted (PD) and Fully Depleted (FD) devices (refer to section 1.3). The SOI devices, realized in the actual IC production, belong to the first category. Their scaling rules are similar to the Bulk case since they need a complex design of the doping profiles and a polysilicon gate contact to control the threshold voltage and the short-channel effects. On the contrary the Fully Depleted design permits to remove the body doping, thus enhancing the channel mobility. This improvement can be achieved only at the cost of using very thin body t_{SI}, a solution that raises many technological issues.

In this chapter we discuss another phenomenon which is strictly related to the use of thin bodies: the capacitive coupling between the gate contact and the substrate contact. In the first section we will present a general introduction to the coupling effect: the discussion will range from modelling to characterization of coupling. This mutual influence has a large importance in the field of device characterization and many measurement techniques are based on it [66]. In ultra thin body device however, the behavior of the coupling effect is changing and, for ultra thin bodies, is reinforced leading to the "super-coupling" effect [67, 68] . This unusual effect is investigated in the second section by detailed simulations and analytical modelling of the potential and electron/hole concentrations. Furthermore we will discuss some scaling issues related to the ultra thin t_{SI} needed for the most advanced technology nodes. The results discussed in this section have been presented in [69] and [70].

6.1 Modeling and application of the coupling effect

When the SOI film is Fully Depleted the electrical properties of MOSFETs are influenced by the charge coupling between the front and back gates. The term "front gate" in a Single gate (SG) SOI MOSFET is related to the gate stack with the thinnest gate oxide. On the contrary the term "back gate" is related to the gate stack with the thick back-oxide (BOX) region that separates the substrate from the silicon active area. Thus the Bulk contact (see Fig. 1.6) is called back-gate contact.
In Double-Gate SOI MOSFETs the above definition cannot be used, since the oxide thicknesses are equal and the two gate contacts are, in general, tied to the same V_G, hence preventing any coupling effect.
The main effect of the coupling is the dependence of the front-gate threshold voltage V_{TH} on the bias and properties of the back gate: in the next subsection we review the classical model which treats this V_{TH} variation. We also show many characterization results that can be interpreted using the concept of capacitive coupling.

6.1.1 Lim and Fossum model

Many different models describe the threshold voltage in Fully-Depleted SOI MOSFETs [71, 72, 73] but the Lim and Fossum model is the most popular since it yields closed-form expressions of V_{TH} under all possible charge condition at the back interface [74]. The model is one-dimensional and is based on the charge-sheet approximation which supposes the entire body depleted and the front (back) charge concentrated at the front (back) interface [75]. We refer to the nMOS case only.
The Lim and Fossum model considers three possible cases (we use the label 1 to refer to the front gate and the label 2 for the back gate):

Accumulated back surface. The accumulation charge concentration at the back interface increases exponentially with the surface potential Ψ_{S2} that is supposed to be virtually pinned at zero. This condition is possible for

$$V_{G2} < V_{G2}^{A} = V_{FB2} - \frac{C_{SI}}{C_{BOX}} 2\Psi_F + \frac{qN_a t_{SI}}{C_{BOX}} \tag{6.1}$$

where $C_{BOX} = \epsilon_{ox}/t_{box}$ is the back-oxide capacitance, $C_{SI} = \epsilon_{si}/t_{SI}$ is the film capacitance, V_{FB2} is the back flat-band voltage, N_a is the doping concentration and Ψ_F is the Fermi potential. In this case the threshold voltage

V_{T1} is independent of V_{G2} and is equal to

$$V_{T1} = V_{T1}^A = V_{FB1} + \Big(1 + \frac{C_{SI}}{C_{OX}}\Big)2\Psi_F + \frac{qN_a t_{SI}}{C_{OX}} \tag{6.2}$$

where $C_{OX} = \epsilon_{ox}/t_{ox}$ is the front-oxide capacitance and V_{FB1} is the front flat-band voltage.

Inverted back surface. In this case Ψ_{S2} is supposed to be virtually pinned at $2\Psi_F$. The back-gate bias is then

$$V_{G2} > V_{G2}^I = V_{FB2} + 2\Psi_F + \frac{qN_a t_{SI}}{C_{BOX}} \tag{6.3}$$

The front-gate threshold voltage is again independent of V_{G2} and is equal to

$$V_{T1} = V_{T1}^I = V_{FB2} + 2\Psi_F + \frac{qN_a t_{SI}}{C_{OX}} \tag{6.4}$$

The difference between the plateau is $V_{T1}^A - V_{T1}^I = (C_{SI}/C_{OX}2\Psi_F)$ and gives the thickness t_{SI} of the silicon fin. This technique is often used for characterization purposes. Notice that a thinner t_{SI} causes a larger variation of the threshold voltage V_{TH}: this is an obvious result of the stronger influence of the charge population at one interface on the opposite one.

Depleted Back Surface. When the back surface is depleted, Ψ_{S2} is strongly dependent on V_{G2} and its value ranges from zero to $2\Psi_F$. For $V_{G2}^A < V_{G2} < V_{G2}^I$, the front-gate threshold voltage varies linearly with the back-gate bias and has the following expression

$$V_{T1} = V_{T1}^I - \frac{C_{SI}C_{BOX}}{C_{OX}(C_{SI} + C_{BOX})}(V_{G2} - V_{T2}^A) \tag{6.5}$$

Thus as V_{G2} varies within the interval $[V_{G2}^A, V_{G2}^I]$, V_{T1} decreases linearly from V_{T1}^A to V_{T1}^I. From the slope of the linear expression 6.1.1 we can extract information about t_{box} or, again, t_{SI}.

Fig. 6.1 shows the $V_{T1}(V_{G2})$ curve given by the above equations. The Lim and Fossum model, even if very popular, has some issues:

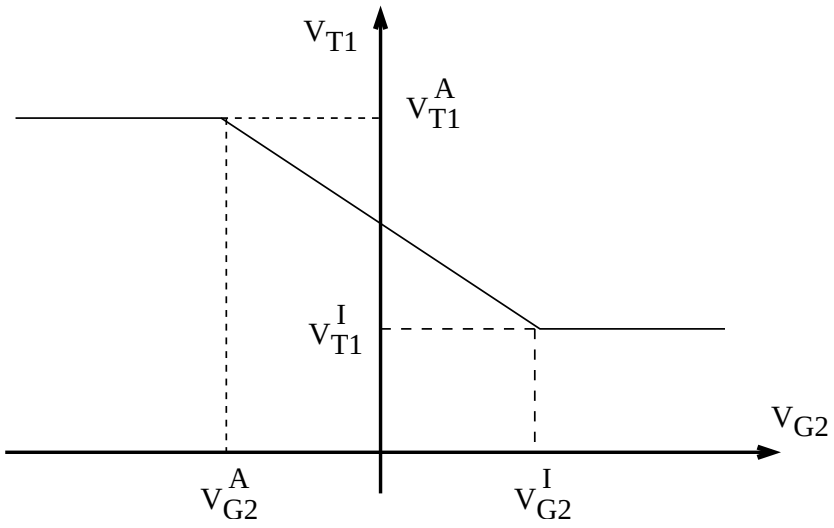

Figure 6.1: Dependence, calculated by the Lim and Fossum model, of V_{TH} on the back-gate bias for a FD SOI MOSFET.

1. the surface potential in accumulation or inversion is not exactly 0 or $2\Psi_F$, in particular in the case of undoped films with t_{SI}below 50 nm. Simple 1D Poisson simulations can demonstrate this behavior. Thus the above expressions cannot be used to describe the $V_{T1}(V_{G2})$ curve in ultra-thin body devices.

2. the model does not predict what happens for $t_{SI} \rightarrow 0$.

This issues will be discussed in the next chapter where we study the charge properties in ultra-thin body SOI MOSFETs.

6.1.2 Impact of coupling on the characterization of the MOSFET

The V_{TH} variation observed experimentally does not feature the same discontinuous shape of Fig. 6.1 since the transition between two charge conditions is not abrupt as assumed in the model. Figs. 6.2 and 6.3 show two examples of characterization results: both $V_{T1}(V_{G2})$ and its reciprocal curve $V_{T2}(V_{G1})$ are plotted in each figure. In Fig. 6.2 we notice that the thinner the film, the wider the V_{TH} variation range. From Fig. 6.3 it is clear that the two reciprocal curves have different slopes in the linear region. The latter result stems from eq. (6.1.1) if we calculate the slope of the two curves with $t_{ox} << t_{box}$.

Fig. 6.4 shows an example of measured drain current and transconduc-

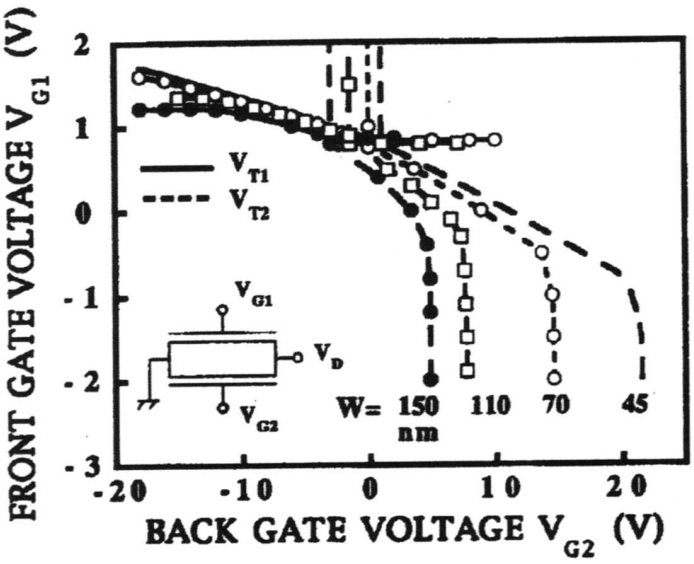

Figure 6.2: Example of variation of both front (V_{T1}) and back (V_{T2}) threshold voltages with opposite gate voltage in SOI MOSFET's for four different silicon thicknesses. Here W means t_{SI}. From Faynot *et al.* [76].

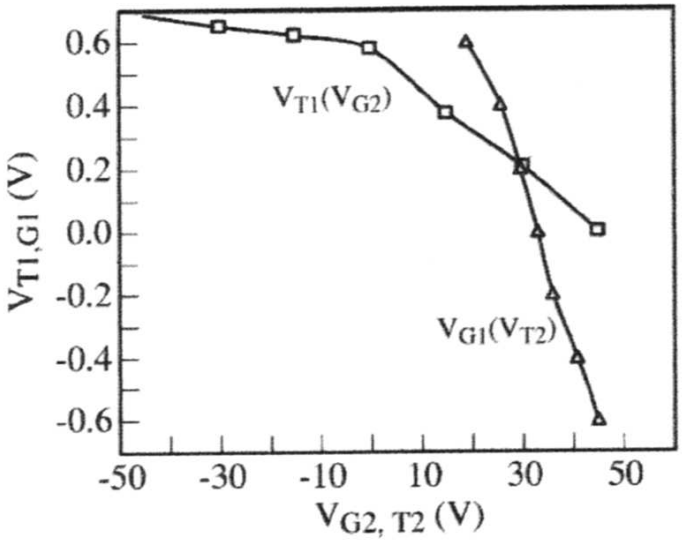

Figure 6.3: Example of variation of both front (V_{T1}) and back (V_{T2}) threshold voltages with opposite gate voltage in a SOI MOSFET with $t_{SI} = 47$ nm. From Ohata *et al.* [77].

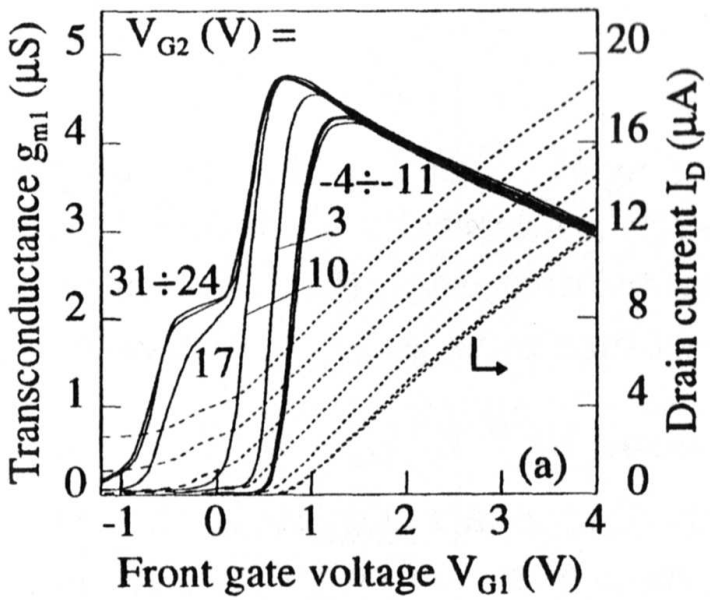

Figure 6.4: Example of the drain current and transconductance curves versus front-gate voltage in a FD SOI MOSFET ($L_G = 20$ μm, $V_D = 50$ mV ,$t_{ox} = 17$ nm, $t_{box} = 380$ nm, $t_{SI} = 80$ nm). From [66]

tance in a fully-depleted SOI MOSFET. The coupling effect not only affects the front-gate threshold voltage but also allows the activation of the back-channel. When the back-interface is accumulated (-4;-11 V in Fig. 6.4), the front channel transconductance is nearly constant because V_{T1} is independent of the back-gate bias. For larger V_{G2}, the front threshold voltage starts to decrease: the current shifts left and the electric field at the front interface decreases. The last effect increases the channel mobility and explains the variation of transconductance peak. When the back-interface is inverted ($V_{G2} \geq 17$ V in Fig. 6.4), we measure the current flowing in both the front and the back channels. Even if $V_{G1} < V_{T1}$, we observe a non-zero current, due to the back channel, and a characteristic tail in the transconductance curve. The tail presents a plateau (see again figure 6.4) which is often used to evaluate the mobility at the back interface [66].

The formation of an accumulation layer at the back interface is a common method to isolate the front-channel from the back-interface properties. In fact defects, interface traps, oxide charges induced by radiation or hot carrier injection, when concentrated at the back interface, may have a signif-

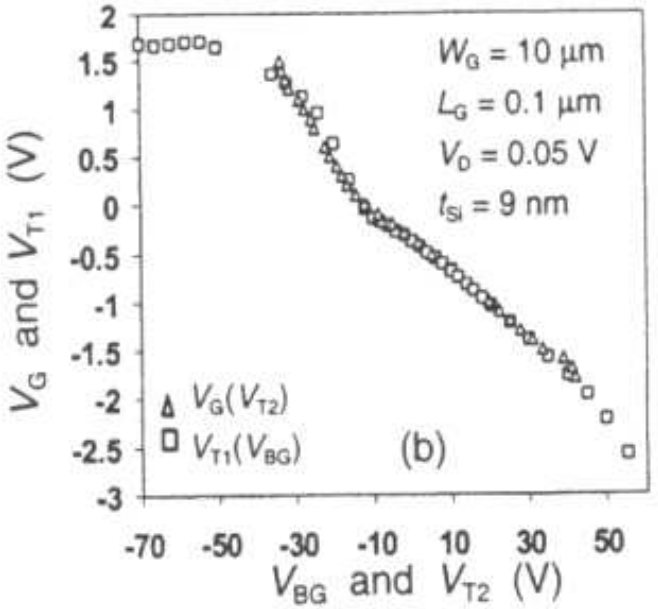

Figure 6.5: Reciprocal threshold voltage curves $V_{T1}(V_{G2})$ and $V_{T2}(V_{G1})$ for an ultra-thin FD SOI MOSFET with $t_{SI} = 9$ nm. From Aydin *et al.* [68].

icant effect on the front-channel properties. The accumulation layer at the back-gate de-couples the two interfaces and permits a correct characterization of the front channel. The opposite case is possible too: the back-channel properties can be measured independently of the front-gate characteristics.

6.2 Special charge properties and coupling effects in UTB SOI MOSFETs

For the 45 nm node and beyond, the silicon body thickness t_{SI} should be scaled below 10 nm in order to suppress the short-channel effects. In such ultra thin bodies, the co-existence of an inversion layer at the front channel and an accumulation layer at the back channel seems unphysical, making the viability of traditional analytical models and characterization procedures questionable. Moreover strange deviations from the standard coupling effects have been observed: Fig. 6.5 reports experimental data from [68] for $t_{SI} = 9$ nm :

- the saturation of V_{T1} and V_{T2} is achieved for extremely large back-gate bias, larger than the values predicted from the model of [74]. These

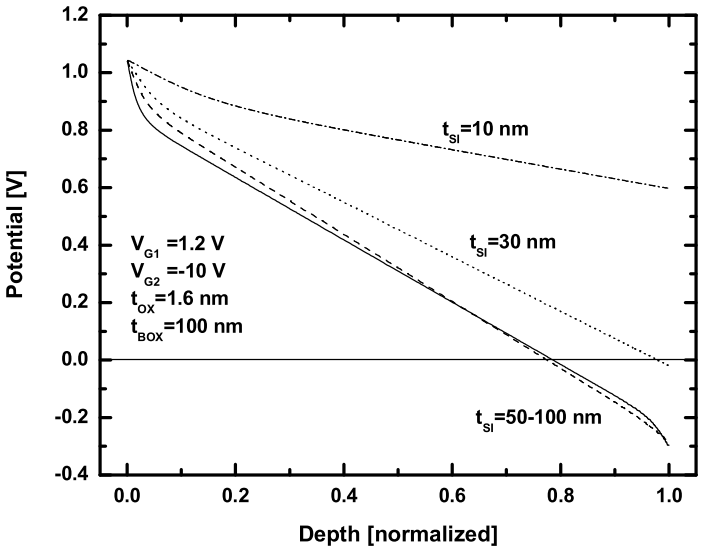

Figure 6.6: Inter-gate potential profile for different body thicknesses t_{SI}. The front interface is at $x/t_{SI} = 0$, and the back interface at $x/t_{SI} = 1$. $V_{G1} = 1.2$ V and $V_{G2} = -10$ V.

values may be dangerous for the front(back) oxide reliability;

- the two reciprocal curves $V_{T1}(V_{G2})$ and $V_{T2}(V_{G1})$ are superimposed even if $t_{ox} << t_{box}$.

The purpose of this section is to demonstrate that the possibility to accumulate charges at the back interface can indeed disappear in ultra thin body devices. Moreover, the electrostatic coupling between the front and the back gates is enhanced leading to interesting super-coupling effects.

6.2.1 Concept of "critical thickness"

In order to study the impact of the silicon body thickness on the possibility to accumulate the back interface, a generic n-channel MOS SOI device with undoped body is considered. The front oxide thickness t_{ox} is set to $t_{ox} = 1.6$ nm and the buried oxide (BOX) thickness t_{box} is much thicker. Front and back gates are supposed to be perfect mid-gap metallic contacts for simplicity. Our simulations are performed by numerically solving the coupled Poisson and Schrödinger equations [78]. The device is biased in strong inversion at the front gate ($V_{G1} = 1.2$ V) and in accumulation at the back gate ($V_{G2} = -10$ V), reproducing standard experimental conditions. Results are plotted in Figure 6.6, showing the electrostatic potential profile in the transverse

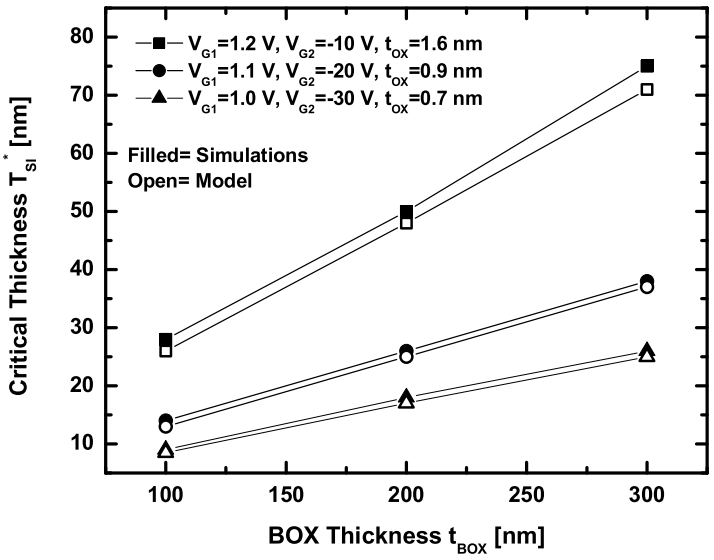

Figure 6.7: Values of the critical thickness T^*_{SI} as a function of buried oxide thickness t_{box}, for various back-gate voltages V_{G2}. The front interface is always biased in strong inversion. The values calculated with the analytical model (Eqs. 6.9–6.11) are compared to self-consistent Poisson-Schrödinger simulations.

direction. For $t_{SI} = 100$ nm (solid line), the front interface potential Ψ_{S1} is largely positive, meaning that strong inversion is achieved, while the back-interface potential Ψ_{S2} is negative, corresponding to an accumulation regime. For all the body thicknesses considered here, the front gate surface potential Ψ_{S1} remains constant, fixed to its strong inversion value. However, for very thin bodies, the back-surface potential Ψ_{S2} starts to increase by coupling becoming equal to zero for a particular t_{SI} value (referred to as the critical t_{SI} and noted T^*_{SI}) and then even positive. Of course, when Ψ_{S2} reaches zero, no accumulation charge is maintained at the back surface. For $t_{SI} < T^*_{SI}$, the Ψ_{S2} value increases further and the back interface is driven into inversion *even if* V_{G2} *is still negative*. This means that not only volume inversion regime [26] occurs but it does reach the back interface, being essentially controlled by the front-gate bias.

Our conclusion is the following: accumulation regime at the back interface can not be achieved in structures with body thinner than T^*_{SI}. This critical value depends on the device structure (especially t_{ox} and t_{box}), but also on the applied bias V_{G1} and V_{G2}. The values of the critical body thickness T^*_{SI}, extracted from numerical simulations, are plotted on Figure 6.7 as a function of the BOX thickness t_{box} and back-gate bias V_{G2}. For constant V_{G1} and t_{ox}, the critical thickness decreases if $|V_{G2}|$ increases or t_{box} decreases, in other

words, when rising the electric field at the back interface.

6.2.2 Analytical modeling of T^*_{SI}

In order to evaluate the critical thickness T^*_{SI} without carrying time-consuming Poisson-Schrödinger simulations, we develop an analytical model, using the following set of approximations:

1. (i) the front interface is biased in strong inversion ($V_{G1} > V_{T1}$),
2. (ii) the inversion (respectively accumulation) charge is exclusively concentrated at the front (respectively back) interface, so that the entire body is depleted (charge sheet approximation),
3. (iii) quantum effects are negligible.

The 1D solution of the Poisson equation, using the approximation (ii), leads to:

$$\Psi(x) = \Psi_{S1} - \mathcal{E}_{S1}x + \frac{qN_a x^2}{2\epsilon_{SI}} \tag{6.6}$$

$$\mathcal{E}(x) = \mathcal{E}_{S1} - \frac{qN_a x}{\epsilon_{SI}} \tag{6.7}$$

where Ψ is the electrostatic potential, x is the vertical position in the body, $\mathcal{E}_{S1}$ is the electric field at the front interface and N_a is the doping concentration. Because of the undoped body the N_a terms will be neglected in the following equations. At the back interface the boundary conditions are $\Psi(t_{SI}) = \Psi_{S2}$ and $\mathcal{E}(t_{SI}) = \mathcal{E}_{S2}$, with

$$\mathcal{E}_{S2} = \frac{C_{BOX}}{\epsilon_{SI}}(\Psi_{S2} + V_{FB2} - V_{G2}) \tag{6.8}$$

where V_{FB2} is the back flat-band voltage, and C_{box} the back-gate oxide capacitance. Using equations (6.6), (6.7), (6.8) and setting $\Psi_{S2} = 0$ (no accumulation charge at the back interface), we obtain:

$$T^*_{SI} = \frac{\epsilon_{SI}\Psi_{S1}}{C_{BOX}(V_{FB2} - V_{G2})} \tag{6.9}$$

To determine a simple, yet efficient expression of Ψ_{S1}, we use the definition of the potential at inversion proposed by Tsividis [79]: the front-interface potential at threshold can be found by solving $C_{INV1} = C_{OX}$, where C_{OX} is

the front-gate oxide capacitance and C_{INV1} is the front-gate inversion charge capacitance, that can be expressed as:

$$C_{INV1} = -\epsilon_{SI} \frac{d}{d\Psi}\bigg|_{\Psi=\Psi_{S1}} \left[\mathcal{E}_{S2} - \sqrt{\mathcal{E}_{S2}^2 + 2\frac{n_i^2 kT}{N_a \epsilon_{SI}} e^{\frac{q\Psi}{kT}}} \right] \tag{6.10}$$

The resulting expression of Ψ_{S1} is thus:

$$\Psi_{S1} = \frac{kT}{q} \ln \frac{N_a C_{OX} C_{BOX} (V_{FB2} - V_{G2})}{q\epsilon_{SI} n_i^2} \tag{6.11}$$

This result can be replaced in Eq. (6.9) in order to find T_{SI}^* versus V_{G2}, t_{ox} and t_{box}.

$$T_{SI}^* = \frac{kT}{q} \frac{\epsilon_{SI}}{C_{BOX}(V_{FB2} - V_{G2})} \ln \left[\frac{N_a C_{OX} C_{BOX} (V_{FB2} - V_{G2})}{q\epsilon_{SI} n_i^2} \right] \tag{6.12}$$

The critical thickness exhibits a quasi-linear variation with t_{box} as shown in Figure 6.7. Note that the obtained T_{SI}^* is not depending on V_{G1}, which is a consequence of approximation (i). Finally, despite that approximations (ii) and (iii) are crude yet usual assumptions when dealing with ultra-thin body devices ($t_{SI} < 10$ nm), the accuracy of the analytical model is satisfactory. This is demonstrated in Fig. 6.7 by the comparison between our model with rigorous numerical simulated data resulting from the self-consistent solving of the Poisson-Schrödinger equations.

6.2.3 Electric fields and control of the body

In order to address properly the coupling mechanism in thin-body SOI MOSFETs, we need to replace the conventional concept of "gate controlling the surface potential" by a more appropriate formulation like "the primary gate which controls the body". The latter definition is justified and clarified by studying the distribution of the transverse electric field in the device. It can be observed in Figure 6.6 that the potential profile becomes very linear in ultra-thin body devices: this means that the electric field is roughly constant, being controlled by one of the two gates, called 'primary gate'. The opposite gate plays a 'secondary' role (modulation effect). Which gate is prevailing depends on the bias (V_{G1}, V_{G2}). In the device with $t_{SI} = 10$ nm of Figure 6.6, the field in the middle of the device is linked to the (volume) inversion charge density imposed by the front gate.

Figure 6.8 presents a more detailed study, showing the field at the front interface $\mathcal{E}_{S1}$, at the back interface $\mathcal{E}_{S2}$ and in the middle of the film E_{mid}

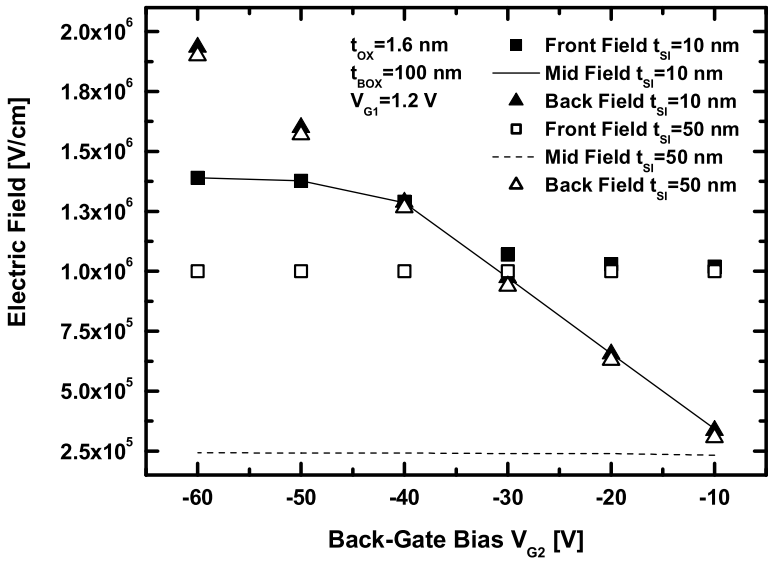

Figure 6.8: Electric field at the front-interface (squares), back-interface (triangles) and in the middle of the body (solid line) versus back-gate voltage, for two devices with $t_{SI} = 50$ and 10 nm.

for two different film thicknesses. The fields are plotted for inversion at the front interface as a function of the back-gate bias ($V_{G2} < 0$, accumulation). In the case of a thick body (50 nm, open symbols), $\mathcal{E}_{mid}$ is lower than $\mathcal{E}_{S1}$ and $\mathcal{E}_{S2}$, whose values are imposed respectively by the strong inversion charge at the front interface and by the accumulation charge at the back interface. These charge densities are independently controlled by the respective gate and de-coupling is achieved. For example, $\mathcal{E}_{S1}$ and $\mathcal{E}_{mid}$ do not depend on V_{G2} whereas $\mathcal{E}_{S2}$ increases linearly with $|V_{G2}|$.

The picture is different in ultra-thin body ($t_{SI} = 10$ nm, closed symbols in Fig. 6.8). The fields in the middle body and at the back interface take the same value: since there is no accumulation charge, V_{G2} acts as a modulation parameter. It can be shown that the transverse field distribution is linked only to the electron charge density. Therefore, V_{G1} is the primary gate and V_{G2} is the secondary gate. When V_{G2} is enough negative, $\mathcal{E}_{S2}$ equals the front field and hole accumulation arises. Complete de-coupling is mathematically achievable for even more negative V_{G2}: however, this case is not realistic because the electric field within the BOX (that is about 3 times the field in silicon) becomes too high leading to oxide damage.

It is clear that in UTB SOI the front gate is the primary gate since

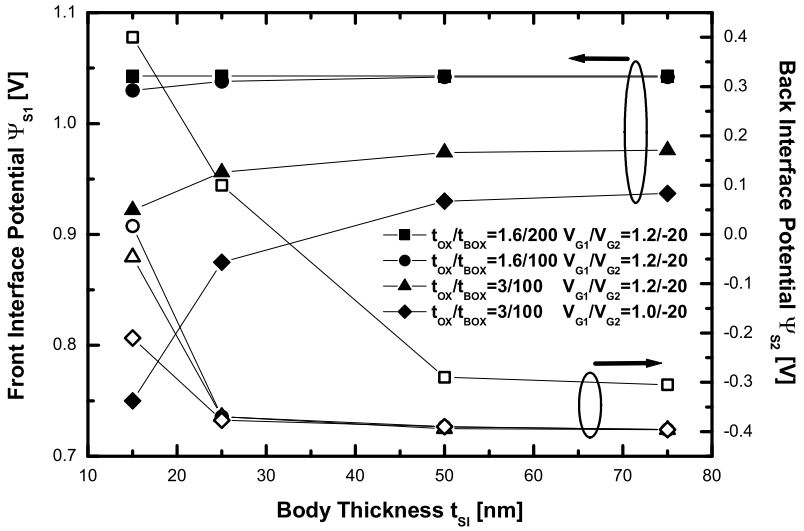

Figure 6.9: Front and back surface potentials Ψ_{S1} and Ψ_{S2} versus t_{SI} for different device parameters. Squares indicate the case of stronger front electric field and volume inversion. Diamonds show the case of stronger back field leading to volume accumulation.

$t_{ox} << t_{box}$. Nevertheless, depending on the device parameters, the back field can influence the inversion condition at the front interface. In Figure 6.9 the front and back surface potential are plotted versus body thickness, for various t_{ox}/t_{box} ratios and conditions of biasing. Squares indicate a device like in Figure 6.6: Ψ_{S2} moves from accumulation to inversion while thinning the body, whereas Ψ_{S1} does not change because V_{G1} is the primary gate. For a relatively thick front-gate oxide (3 nm), biased at $V_{G1} = 1$ V, when reducing t_{SI}, the front interface can move from inversion into depletion (diamond symbols) under the strong influence of the back-gate bias. There are conditions (low V_{G1}, high V_{G2}, thin BOX) where the front-gate action is further reduced, so that *volume accumulation*, instead of *volume inversion*, may be achieved. In such a case, V_{G2} becomes the primary gate.

Our conclusion is that in UTB MOSFETs only one type of charge, positive or negative, can be sustained for a given set of device parameters and bias.

6.2.4 Super Coupling effects in FD SOI MOSFETs with subcritical thickness ($t_{SI} < T^*_{SI}$)

In this section, the electrostatic coupling of the two gates is investigated in more detail for $t_{SI} < T^*_{SI}$. The first point consists in answering the following question: since accumulation charge cannot be forced at the back interface, how the modulation of V_{TH1} with V_{G2} is affected?
Figure 6.10 compares the threshold voltage V_{T1} versus V_{G2} curves for two different devices, one with a t_{SI} smaller than T^*_{SI} (T^*_{SI}= 15 nm for $V_{G2} = -20$ V), and the second much thicker. It is seen that the variation of the threshold voltage with the back-gate bias is not only present when $t_{SI} < T^*_{SI}$ but is even stronger. The slope of $V_{T1}(V_{G2})$ curve (filled squares) may exceed the conventional value, t_{ox}/t_{box} [74], for very thin t_{SI} due to volume inversion. Moreover, the saturation observed in the thicker (for $V_{G2} < -10$ V, filled triangles), is no longer present. This means that the interface coupling effects are enhanced in films with subcritical thickness. This trend is in agreement with previously published experimental data [67, 68]: no saturation of coupling was observed in 7 nm films because the back-channel accumulation could not be achieved. In order to de-couple again the two interfaces a very strong back-gate field must be imposed, as seen in the previous paragraph. Unfortunately the needed value of V_{G2} may be too high and could not be used for characterization purposes.

Another consequence of super-coupling was noted when drawing the reciprocal curves, $V_{T1}(V_{G2})$ and $V_{T2}(V_{G1})$, on the same graph. In thick devices, the two curves are different, and the intercept point represents the unique couple of front- and back-gate threshold voltages ($V_{TH1}; V_{TH2}$) which enables the two channels to reach inversion simultaneously [80].

$$V_{G1} - V_{T1} = \frac{t_{ox}}{t_{box}}(V_{G2} - V_{T2}) \tag{6.13}$$

In sub 10-nm-thick films, when the film capacitance exceeds the gate oxide capacitance, the two curves tend to coincide [67]. This implies that an arbitrary back-gate bias V_{G2} is promoted as threshold voltage V_{T2} as soon as the front gate is biased at threshold. This behavior is a result of the analysis in Figure 6.8: the 'secondary' gate does not control a positive charge density near its interface but acts as a modulation parameter of the electron charge density in the whole body. When $V_{G1} = V_{T1}$, the threshold has been reached from the point of view of both gate contacts and thus $V_{G2} = V_{T2}$. Our simulations of Figure 6.10 fully confirm these trends which are important for device operation in double-gate mode.

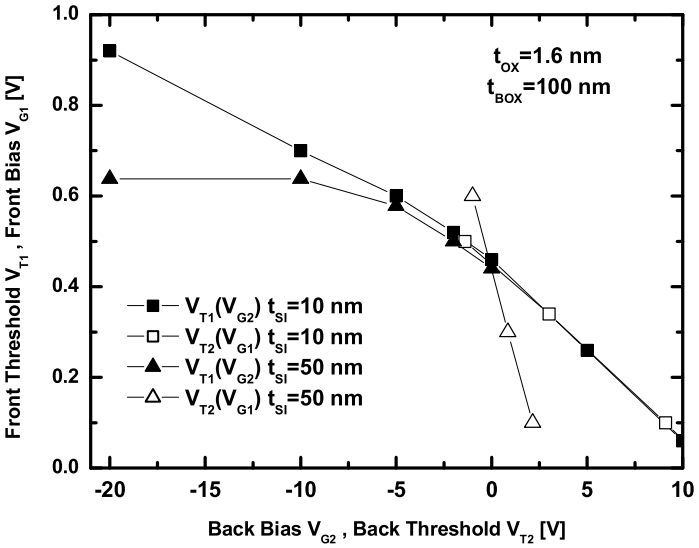

Figure 6.10: Front threshold voltage V_{T1} versus back-gate bias V_{G2} and reciprocal curve $V_{T2}(V_{G1})$ for two SOI MOSFETs with $t_{SI} > T^*_{SI}$ and $t_{SI} < T^*_{SI}$.

The enhancement of the coupling between the two gates is clarified by examining the surface potential variations. Figure 6.11 shows the front and back surface potentials, Ψ_{S1} and Ψ_{S2}, as a function of the front-gate voltage V_{G1}, for the devices of Figure 6.10 and a negative back-gate voltage $V_{G2} = -10$ V. For $t_{SI} > T^*_{SI}$, Ψ_{S2} value is perfectly controlled by the strong accumulation condition: it remains negative and independent of V_{G1}. For $t_{SI} < T^*_{SI}$ however, Ψ_{S2} is no longer constant and keeps a negative value only when the front interface is in deep subthreshold regime. The variation of Ψ_{S2} with V_{G1} parallels that of Ψ_{S1}. The difference between the two parallel curves tends to vanish when the back-gate voltage is less negative or the film is thinner. Reciprocally, the proximity of the back interface causes a lowering of the front-channel inversion charge. For a fixed V_{G1} bias, the ultra-thin device exhibits a lower inversion charge than for $t_{SI} > T^*_{SI}$. This pure electrostatic effect (not related to charge quantization [81]) is equivalent to an increase in threshold voltage, even if no back-gate accumulation is achieved near V_{TH1}. The result is somehow counter intuitive because standard MOS equations indicate that V_{T1} should be lower as the depletion charge decreases in thinner films.

Figure 6.12 shows typical in-depth potential profiles when the device is

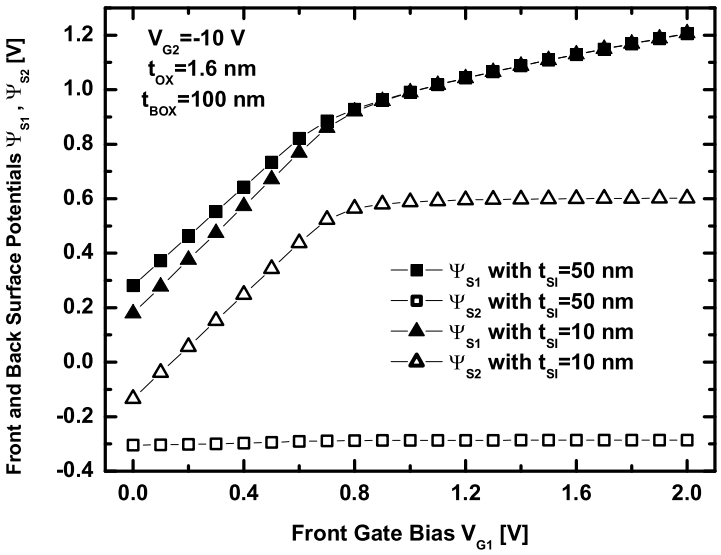

Figure 6.11: Surface potentials Ψ_{S1} (filled symbols) and Ψ_{S2} (empty symbols) versus V_{G1} for a fixed back-gate voltage $V_{G2} = -10$ V (same devices as in Fig. 6.8

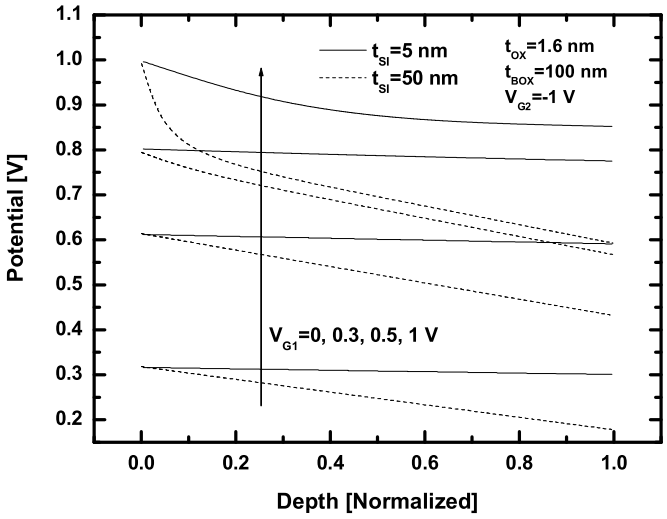

Figure 6.12: Potential profiles for two devices of different t_{SI}, and for various front-gate voltages V_{G1}. In the case of a very thin body (5 nm), the profile is almost flat and rigidly shifts up when V_{G1} increases.

driven by the front gate, while the back gate is weakly biased. The back-surface potential follows the Ψ_{S1} variation and the imbalance $(\Psi_{S1} - \Psi_{S2})$ decreases with film thickness. In ultra-thin films, the super-coupling leads to a nearly flat potential profile. Either surface potential is equally controlled by its own gate or by the opposite gate. This explains why the coupling curves in Figure 6.10 became superposed. The film behaves as a rigid quasi-rectangular well: when the potential at one interface is modified by the gate, the potential of the entire film follows. The notions of front and back channels or front and back mobilities become obsolete and need to be reconsidered by accounting for volume inversion or accumulation concepts.

Chapter 7

Coupling effects in Multiple Independent Gate MOSFETs

Planar CMOS technology has revolutionized the electronics industry over the last few decades. The rapid scaling is now reaching its limit and has forced the industry to look to novel device architectures beyond 45 nm technology node. New device architectures using multiple sides of the semiconductor not just the planar surface offer a path to overcome the performance limit. In section 1.3 we have introduced the FinFET architecture as a feasible way to produce Double-Gate MOSFET on SOI wafers. Recently a modification of the FinFET structure has been promoted: the Multiple Independent Gate FET (MIGFET) [84], is a FinFET where the two lateral gates are independent and can be controlled separately. In the FinFET device the gate coupling is not realized between the two gate contacts, because they are tied together, but between the gate contacts and the substrate contact. In the MIGFET we can distinguish the lateral coupling between the two lateral gates and the vertical coupling between the lateral gate contacts and the substrate contacts. The purpose of this work is to analyze MIGFETs with different fin widths in order to investigate these various types of coupling and to provide useful information for future simulations, modelling and device optimization. The results of this section have been presented in [85].

7.1 Characterization of gate coupling in Multi-Gate MOSFETs

The characterization of the coupling effect is essentially based on the extraction of the V_{TH} as a function of the back-gate bias. In fact the V_{TH} extraction from measurements has been extensively studied and the Lim and Fossum

model is a good method to interpret the results, except in the UTB regime as discussed in the previous chapter. In addition to the threshold voltage many other parameters are usually characterized like: subthreshold slope, transconductance, mobility. In this chapter we will discuss concepts necessary to understand the characterization of V_{TH} and of the coupling effects.

7.1.1 State-of-the-art characterization of V_{TH}

The correct characterization of the threshold voltage of the MOSFET has a critical importance when we want to evaluate the coupling effects. At this point, it is necessary to review the efficiency of the various characterization techniques that are frequently used to determine the V_{TH} [66]:

$I_{DS}(V_{GS})$ extrapolation. Once the $I_{DS}(V_{GS})$ curve has been measured for the ohmnic region of operation, the tangent is drawn at the point where the gate voltage corresponds to the maximum transconductance. The intercept of the tangent with the horizontal axis (zero current) gives the extrapolated threshold voltage. The accuracy of this technique may be improved by repeating the measurement for several V_D values (usually less than ≈ 50 mV) and re-extrapolating to $V_D = 0$ V.
This method is very simple and fast, and thus very popular. Unfortunately its accuracy is seriously degraded when large series resistances are present, like in the Fully-Depleted SOI MOSFETs. Moreover it cannot be applied when the current flows in the back channel because the observed I_D contains both the front and the back-channel contributions. Thus the extrapolation of the current may be highly inaccurate.

Constant-Current V_{TH}. The threshold voltage is defined as the gate voltage allowing a certain amount of current. This definition avoids any extrapolation error and may monitor very small V_{TH} fluctuations or shifts. Unfortunately there is no rules that defines the correct I_{DS} value that is needed to evaluate V_{TH}. Moreover, if a back-channel is activated, a current larger than the reference value flows in the device even if the front interface is still depleted. Like the previous method, this V_{TH}-extraction technique is difficult to apply in the FD SOI case.

Peak of d^2I_D/dV_G^2. The peak of the second-order derivative of the $I_{DS}(V_{GS})$ curve is another useful way to calculate V_{TH}. This method has been validated

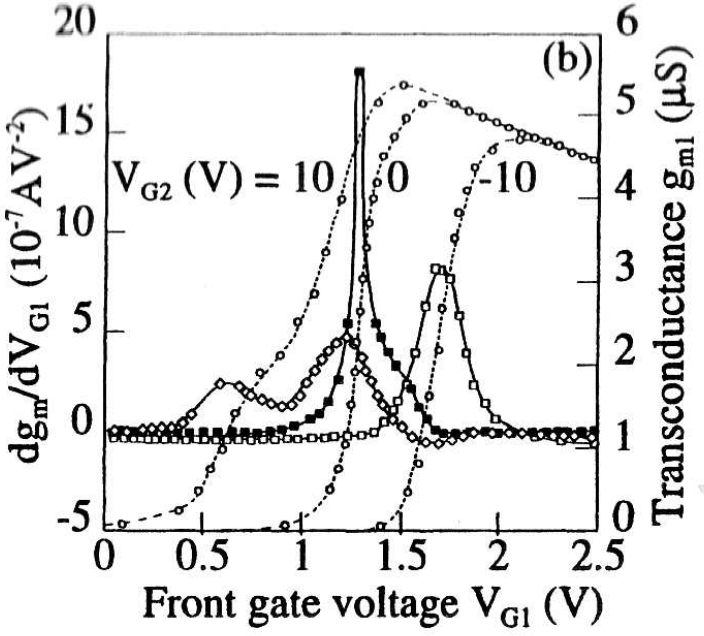

Figure 7.1: Front-channel threshold voltage evaluation from the second-order current derivative for various back-gate biases. From [66].

by numerically calculating the derivatives of the inversion charge as a function of the surface potential [82]. This method is fairly insensitive to series resistances and back-gate bias. Fig. 7.1 illustrates the technique. The peak position defines the front-gate V_{TH}. When the back interface is activated, two peaks are observed the larger one for the front-channel V_{TH}, the smaller one for the back-channel activation. The main issue of this characterization technique is the requirement of $I_{DS}(V_{GS})$ curves with negligible noise: each small variation is increased by the double derivation and spurious peaks can appear.

Extrapolation from $I_D/\sqrt{g_m}$. This technique is very powerful and is able to extract not only V_{TH} but other device parameters [83]. The simplest version of this method is based on the classic $I_{DS}(V_{GS})$ curve given by equation (1.1), valid for the ohmnic region. The curve $I_D/\sqrt{g_m}$ is then

$$\frac{I_D}{\sqrt{g_m}} = \sqrt{\frac{C_{OX}\, W\, V_D \mu_{EFF}}{L}}(V_G - V_{TH}) \tag{7.1}$$

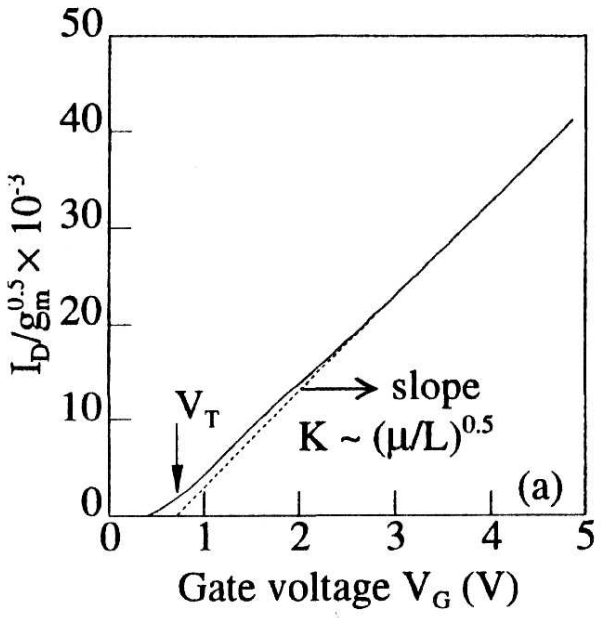

Figure 7.2: Threshold voltage extraction using the $I_{DS}/\sqrt{g_m}$. From [66].

Fig. 7.2 shows an example of eq. (7.1): it appears as a straight line that intercepts the horizontal axis at $V_G = V_{TH}$. The mobility is determined by the slope $K = C_{OX} W V_D \mu_{EFF}/L$ of the curve if the gate length is known. Otherwise the experiment is repeated for various gate length and the function K^{-2} versus L is linearized: thus the slope is equal to μ_{EFF}. The entire procedure is rather insensitive to the series resistances.

In this work we have decided to use the third method to extract the V_{TH} of our devices, since the possibility to observe different threshold voltages (corresponding to different peaks of the curve $d^2 I_D/dV_G^2$) is very useful when evaluating the coupling effects.

7.1.2 Lateral and Vertical coupling in Non-Planar MOSFETs

In section 6.1.1 we have studied the coupling effect in a simple one-dimensional picture. In this picture, the device is modeled as a stack of different regions and materials: front-gate contact, front oxide, silicon, back oxide, back-gate contact. In a Fully-Depleted Single-Gate SOI MOSFET the cross-sectional

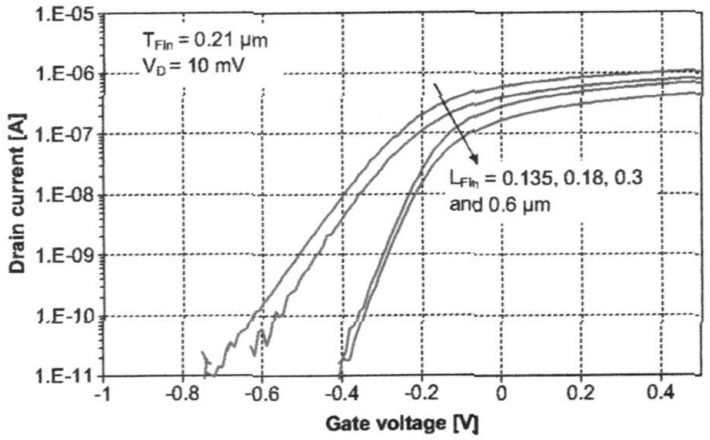

Figure 7.3: Drain current versus gate voltage in n-channel with different channel length (here called L_{fin}) and fixed fin thickness (here called T_{fin}) FinFET. From [86].

area has a structure similar to the above device stack and, in fact, the 1D model described in 6.1.1 works rather well for that type of device.
Now we show how to extend the analysis of the coupling effect by introducing a second dimension.

Vertical coupling. The top part of Fig. 7.4 shows the cross-sectional area of a typical FinFET device: the same V_G is applied to both gate contacts and the double-Gate configuration is thus realized. In this architecture the capacitive coupling between the two gates is not possible. Fig. 7.3 shows the drain current I_D versus V_G for different bias V_{SUB} at the substrate contact: we again observe a V_{TH} shift similar to the case of Fig. 6.4. The result suggests that FinFET devices still present some type of coupling effect but between the substrate contact and the two channels [86]. This type of coupling, that we will use to call *vertical coupling*, is two-dimensional because of the device geometry (arrows in the top part of Fig. 7.4). It has been demonstrated through characterization and simulation that the vertical coupling depends on the fin width t_{SI}: if t_{SI} is enough thin, the observed V_{TH} shift is small.

The MIGFET device. The MIGFET is a new type of FinFET where the two lateral gates are no more tied at the same voltage V_G. The fabrication process of MIGFET devices has been presented in [84]. SOI silicon wafers were patterned into thin fins of about 100 nm height and, using an innovative process, two independent gate regions were formed on the two sides of the fin: the separation between the two gate contacts is achieved through a top

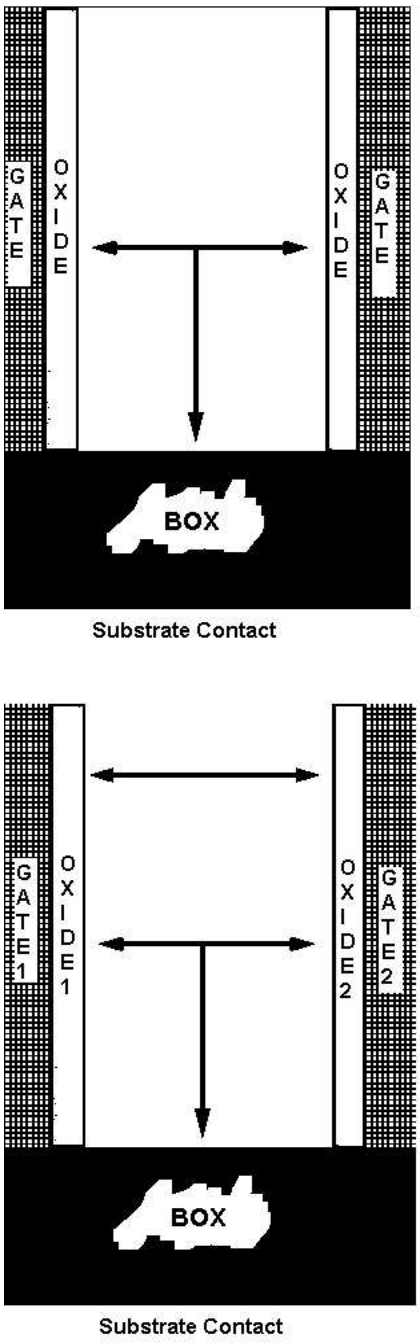

Figure 7.4: Simplified cross-section of typical FinFET (top) and MIGFET (bottom) devices. Arrows highlight: (i) the vertical coupling between the bottom interface and the two lateral channels; (ii) the lateral coupling between the lateral gates in the MIGFET case. In the MIGFET case we observe the multiple capacitive coupling between all the three gate contacts.

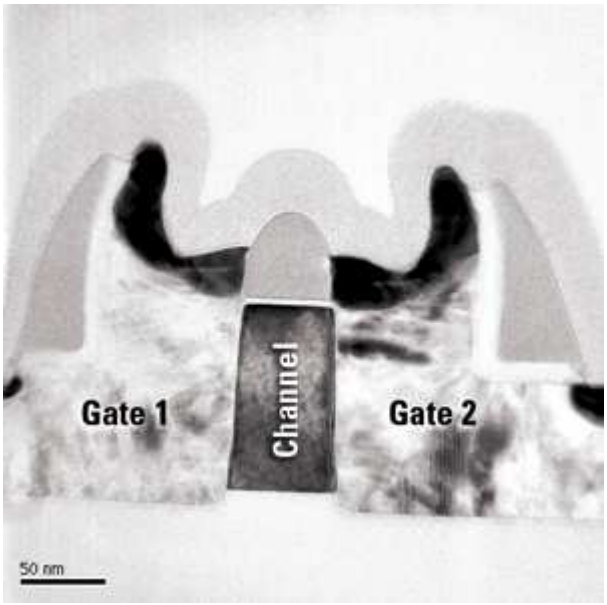

Figure 7.5: TEM figure of the cross-section of a MIGFET device. From [84].

spacer.
The bottom part of Figs. 7.4 and 7.5 show, respectively, a simplified cross-section of the MIGFET and a TEM picture of it. The capacitive coupling between the lateral gates $G1$ and $G2$ is similar to the case analyzed in chapter 6. We notice that

- the gate coupling between $G1$ and $G2$ is roughly one-dimensional, thus the Lim and Fossum model can be used to describe the V_{TH} shift. We call *Lateral Coupling* this effect;
- the substrate bias V_{SUB} is still able to influence the device body like in the FinFET case, but, in this case, it will influence each channel (controlled by $G1$ or $G2$) differently;
- it is still possible to apply $V_{G1} = V_{G2}$ to recover the FinFET behavior. In this case the Lateral coupling is neglected.

7.2 Lateral and Vertical coupling in MIGFETs

In these section we present our study, through characterization results, of the lateral and vertical coupling in MIGFET devices. The majority of our work is based on V_{TH} measurements but we will also discuss the variation of

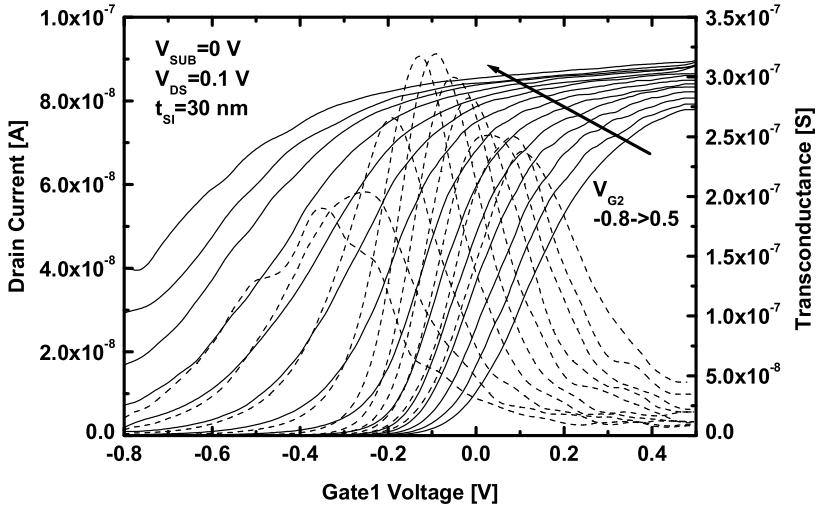

Figure 7.6: Drain current and transconductance versus gate1 voltage for different gate 2 bias, $t_{SI} = 30$ nm.

transconductance and subthreshold slope as a function of the voltage bias at the opposite gate and at the substrate. The MIGETs provided by Freescale are n-channel devices with gate length of about 80 nm. On both fin edges the gate stack is composed of a 2.4 nm oxide layer, thermally grown, and polysilicon gate contacts. We have considered devices with starting fin widths of 100 and 150 nm. After the fin trimming the real widths, measured through TEM, are respectively 30 and 50 nm. All MIGFETs considered in this work are n-channel with undoped body. The BOX thickness is about 100 nm while the fin height is about 80-100 nm.

7.2.1 Current Measurements and Lateral Coupling Effects

We consider the lateral coupling between the two main gates $G1$ and $G2$ with grounded substrate: Figure 7.6 shows the I_{DS} vs. V_{G1} curves measured for $V_{DS} = 0.1$ V, with the opposite gate voltage V_{G2} varying from -0.8 to 0.5 V. The device has a fin width t_{SI} equal to 30 nm. The same measurements have been repeated on the other gate and we have obtained similar I_{DS} vs. V_{G2}: this means that the device structure is rather symmetrical.
Because of the capacitive coupling the threshold voltage decreases while the

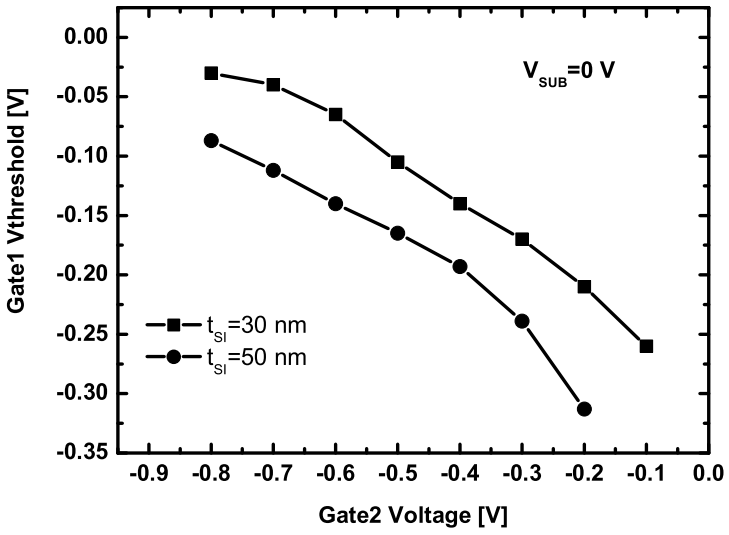

Figure 7.7: Extracted V_{TH}, for different fin thicknesses, versus voltage at the opposite gate.

opposite gate is moving from accumulation to inversion and the I_{DS} vs. V_{G1} curve shifts to left. Also the transconductance g_m changes with the variation of V_{G2}: the degradation of the g_m peak, when the opposite gate is accumulated, is a well known result of the inter-gate higher effective field. The degradation of g_m, when the opposite channel is in inversion, is due to a different mechanism: because of the presence of another channel, more current flows through the series resistances and the effective V_{DS} voltage across the device is lower [66].

In order to understand how strong the coupling effects are, we must extract the threshold voltage V_{TH}. All the threshold voltages have been evaluated as the voltages where d^2I_{DS}/dV_{G1}^2 is maximum. Figure 7.7 shows the threshold voltages extracted from Fig. 7.6 and for a MIGFET with $t_{SI} = 50$ nm under the same measurement conditions. The values of V_{G2} have been restricted to $V_{G2} < 0$ V because, when the opposite gate is inverted, we cannot distinguish the contribution of each channel and the extracted value of V_{TH} is not reliable.
Now let's consider the V_{T1} vs. V_{G2} plot for the thinner fins: the curve is quasi-linear as predicted, and the slope is about -0.35, meaning that a lateral coupling exists. We now remind the key result of the Lim and Fossum's

theory for Fully-Depleted SOI [74]:

$$\Delta V_{T1} = -\frac{C_{SI} \cdot C_{OX2}}{C_{OX1} \cdot (C_{SI} + C_{OX2})} \cdot \Delta V_{G2} \tag{7.2}$$

where C_{OX1} and C_{OX2} are the gate capacitances and C_{SI} is the fin fully-depleted capacitance. In our devices with $C_{OX1} = C_{OX2}$, Eq. (7.2) predicts a less steep slope, about -0.2. Anyway this theory is based on a simple 1D model while here the structure is 2-dimensional; other works have already evaluated through simulation that multiple-gate devices need different and more complicated models [88][89]. In particular, the bottom part of the fin should be more coupled with the back-gate, degrading the total lateral coupling coefficient. But if the fin is thin enough we cannot neglect the fringing fields between the lateral gate and the BOX: these fields screen the back-gate coupling and enhances the lateral coupling.

The lateral coupling should decrease with a larger fin thickness and indeed the slope of the curve is slightly reduced to about -0.29 when $t_{SI} = 50$ nm. The decrease of all the threshold voltages as compared to the thinner film is an expected result of the degraded lateral gate control over the fin.

7.2.2 Vertical coupling and comparison with double-gate mode

We now investigate the effect of the back-gate bias in MIGFETs: in particular we separate the variation $V_{TH}(V_{SUB})$ of the lateral gates (vertical coupling) and the variation of the lateral coupling between the two main gates versus V_{SUB}. The variation of the on-current and transconductance for $V_{SUB} = 10$ V and $V_{SUB} = -10$ V is shown, respectively, in Figs. 7.8 and 7.9, for the case $t_{SI} = 30$ nm. If we look at the case of zero substrate bias (Fig. 7.6), we notice that all threshold voltages have shifted due to the effect of V_{SUB}. To make comparison with a particular case, we have done measurements on MIGFETs with $V_{G1} = V_{G2}$, as Double-Gate devices: Figure 7.10 shows the measured I_{DS} and g_m under this condition. It is again visible a shift of the threshold voltage due to the coupling with the back-gate voltage, even if the variation is modest compared to the lateral coupling effect. Recent simulations have demonstrated that the coupling between the lateral channels and the bottom channel is weak in Double-Gate MOSFETs with thin fins [86].

The transconductance is higher than in Fig. 7.6, due to the volume inversion and higher mobility [26], but still it is lowered when the bottom interface

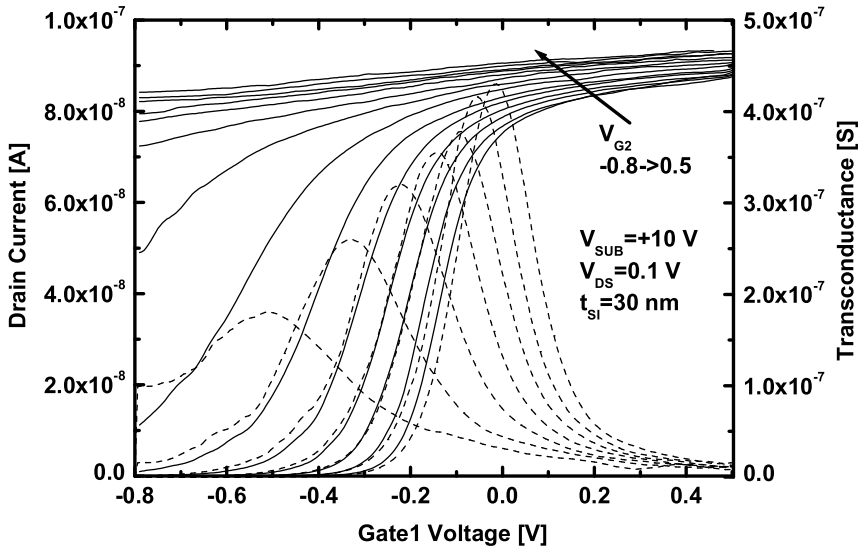

Figure 7.8: Drain current and transconductance versus gate1 voltage for different gate 2 bias, $t_{SI} = 30$ nm and $V_{SUB} = 10$ V.

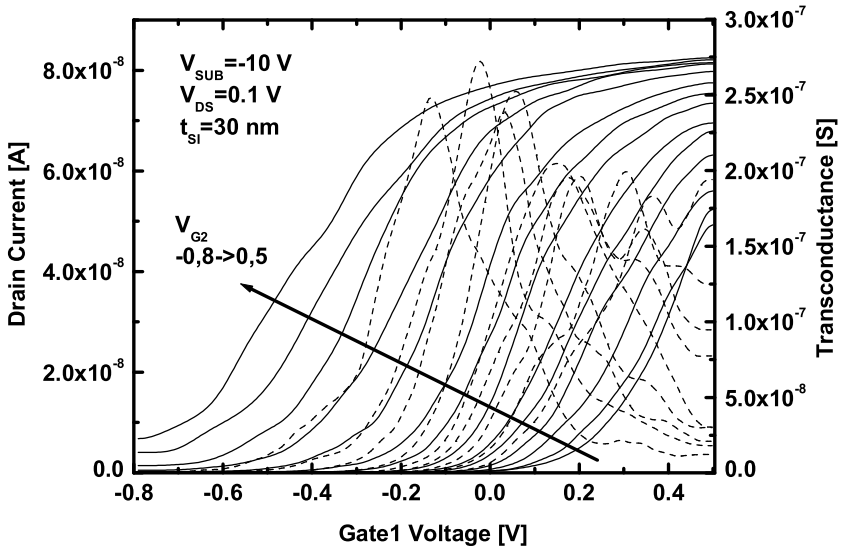

Figure 7.9: Drain current and transconductance versus gate1 voltage for different gate 2 bias, $t_{SI} = 30$ nm and $V_{SUB} = -10$ V.

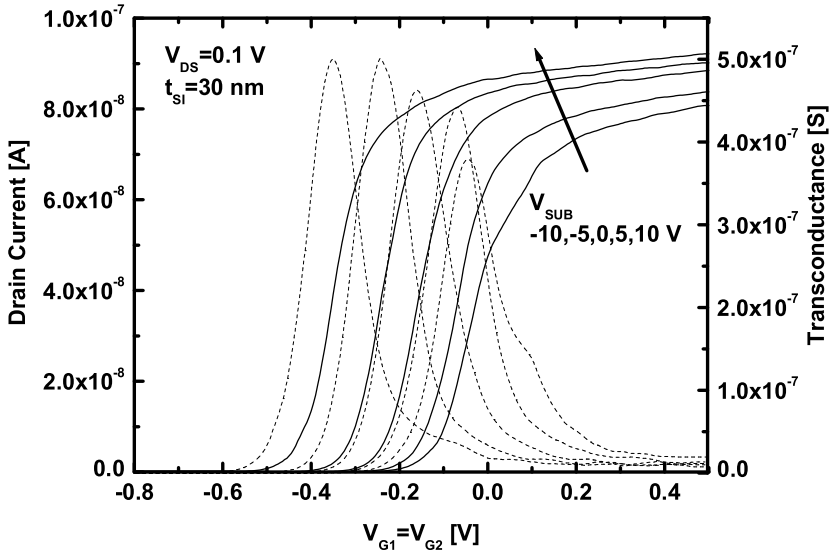

Figure 7.10: Drain current and transconductance in double-gate mode for different back-gate bias, $t_{SI} = 30$ nm.

is accumulated. For DG mode, the g_m peak is large and does not change much with V_{SUB}.

Figure 7.11 shows the variation of V_{T1} with back-gate voltage V_{SUB} varying from -10 to +10 V: the overall variation of the threshold is quite weak for $t_{SI} = 30$ nm and it is quite insensitive to the bias at the opposite gate. Moreover it is equal to the variation with $V_{G1} = V_{G2}$, when the back-coupling is minimum. Instead if the fin is thicker the back-gate has a larger influence on the V_{T1}: when $V_{SUB} = +10$ V the back-channel is inverted and V_{T1} drops rapidly; when $V_{SUB} = -5$ V the potential in the bottom part of the fin is influenced by back channel accumulation and the threshold voltage becomes insensitive of the back-gate bias. These results confirm that the vertical coupling between the lateral channels and the back-gate is weak if the fin is thin enough.

On the purpose to understand the variation of the lateral coupling with V_{SUB}, V_{T1} vs. V_{G2} curves are shown in Figure 7.12. Accumulating the back-gate does not change significantly the slope of the $V_{TH}(V_{G2})$ curve and this effect does not agree with the result in section 7.2.1 where we found a slope steeper than the predicted value. In that case we interpreted the result sug-

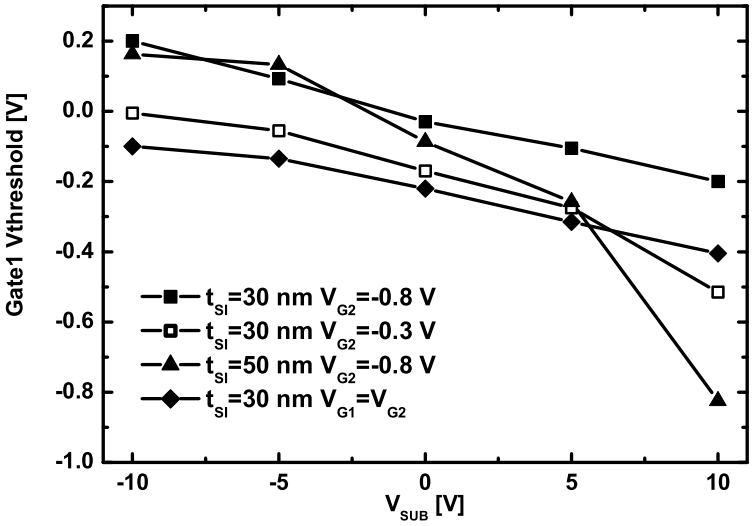

Figure 7.11: Threshold voltages versus back-gate bias for devices with two fin widths and under different measurements conditions.

gesting the presence of fringing fields at the BOX, but these fields should be less effective when the back-gate is in accumulation. Two possible explanations may be suggested now:

1. the real fin width of the device under test is not 30 nm but it is slightly smaller. This is quite possible since the technological process has some margin of error. A smaller t_{SI} could decrease the coupling between the body and the bottom interface, thus reducing the dependence of the lateral coupling from V_{SUB};

2. the top spacer, used to separate the two gates, is subject to fringing fields too and its effect remains for any back-gate voltage.

Three-dimensional simulations have been carried out to interpret our results and it seems that the first option is the most probable [85]. When $V_{SUB} = +10$ V the curve seems steeper: this result is not due to coupling but to the difficulty of extracting the V_{TH} when more than one channel is open: so the calculated value rapidly decreases if V_{G2} increases. The final result is that the back-gate bias does not seem to affect in a significant way the lateral coupling.
Figure 7.13 shows V_{T1} vs. V_{G2} curves for a fin width $t_{SI} = 50$ nm. When the bottom interface is inverted the substrate gate control is more efficient: the

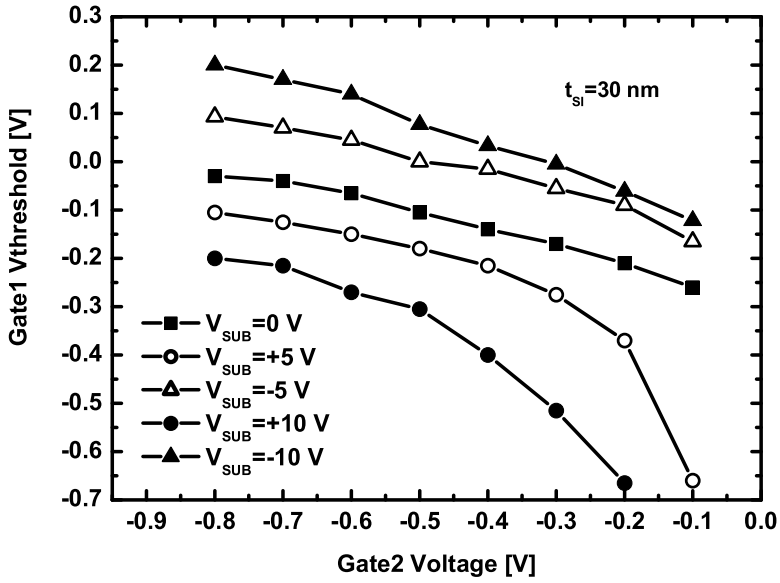

Figure 7.12: Threshold voltages of gate1 versus gate2 with different back-gate bias, $t_{SI} = 30$ nm.

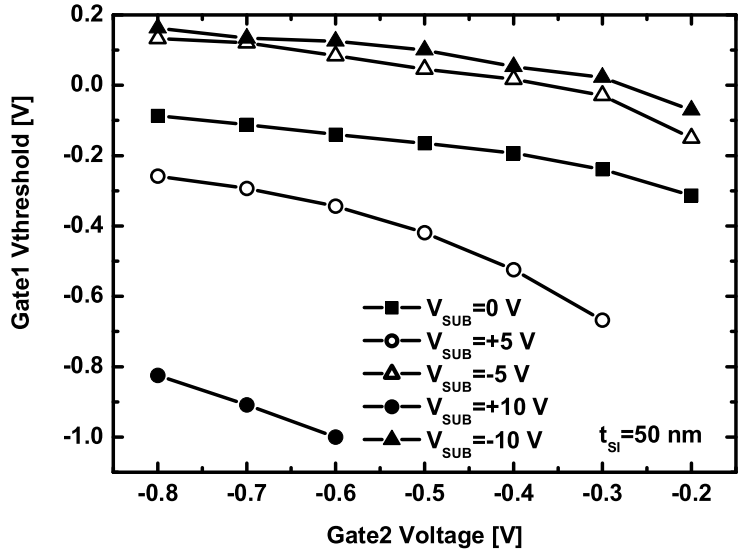

Figure 7.13: Threshold voltages of gate1 versus gate2 with different back-gate bias, $t_{SI} = 50$ nm.

curve seems steeper but again this result is due to the difficulty of extracting the threshold voltage. When the back-gate is accumulated the slope of the curve does not change and makes evident the effect of the fixed back-gate potential: even if V_{SUB} is becoming more negative the V_{th} does not change anymore.

After this analysis we can conclude that the lateral coupling is the most effective method to modify the threshold voltage of the device. The effect of the back-gate bias is quite modest, if the fin width is very thin, both in single-gate and in double-gate mode. Other device parameters have been studied through characterization: here we present the variation of the subthreshold slope and transconductance as a function of V_{G2} and V_{SUB}.

7.2.3 Other parameters extracted from measurements

Subthreshold slope. Figure 7.14 plots the subthreshold slope S for various measurement conditions. The minimum value of S for the thinner fin is about 80 mV/dec which is larger than the experimental and simulation results for the most advanced Double-Gate and Fully-Depleted SOI. The values of S is more degraded if accumulating the opposite channel: this is an obvious result of the coupling between the two channels [66]. The subthreshold slopes are obviously worse in the $t_{SI} = 50$ nm case due to the reduced control of the fin by the gate. If driving the MIGFET in double-gate mode we get the ideal 60 mV/dec swing. This ideal value is insensitive of V_{SUB} except if $V_{SUB} = -10$ V: in this case we find $S = 75$ mV/dec and this slight degradation is again due to the capacitive coupling with the back-gate in accumulation.

Transconductance. Fig. 7.15 plots the peak values of g_m vs. V_{G2} for different V_{SUB}. As already seen in Fig 7.6 the transconductance has a maximum value, reached when the opposite gate is depleted.
Back-gate accumulation modifies the inter-gate field and the lateral coupling, in particular at the bottom of the fin. The mobility is reduced by 20% and the peak is shifted to the right because of the larger threshold voltage. In contrast when the back gate is inverted, volume inversion is amplified. The mobility enhancement can reach 40% (Fig. 7.15) but this gain is lost for small V_{G2}. This is due to the lower V_{TH} of the opposite gate when $V_{SUB} = +10$ V: if V_{G2} increases, also the current flowing trough the series resistance increase and g_m decreases.

We can compare the transconductances obtained from this measurement with the ones plotted in Fig. 7.10. If we drive the MIGFET with $V_{G1} = V_{G2}$ we achieve a maximum g_m that is always larger than in Fig. 7.15 but it is never twice as obtained in the most advanced Double-Gate devices. This results may be due to a fin width not enough thin to cancel the influence of the back-gate potential and to get maximum benefit from volume inversion.

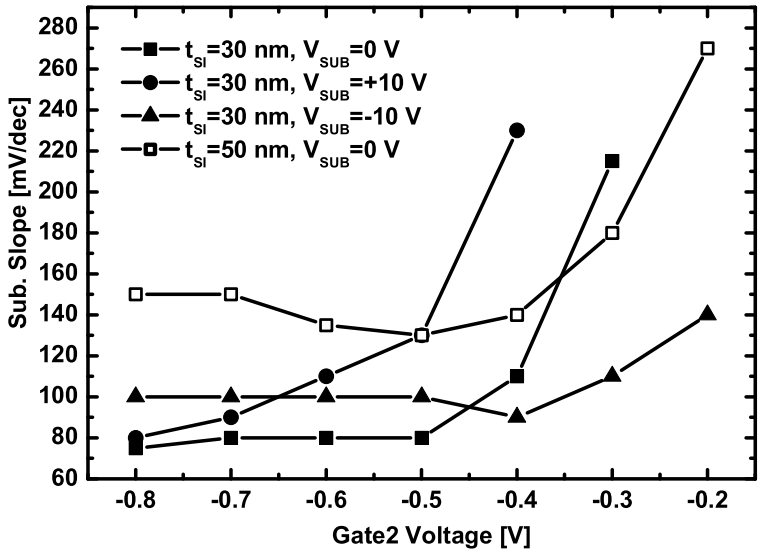

Figure 7.14: Subthreshold slopes versus gate2 with different back-gate bias and t_{SI}.

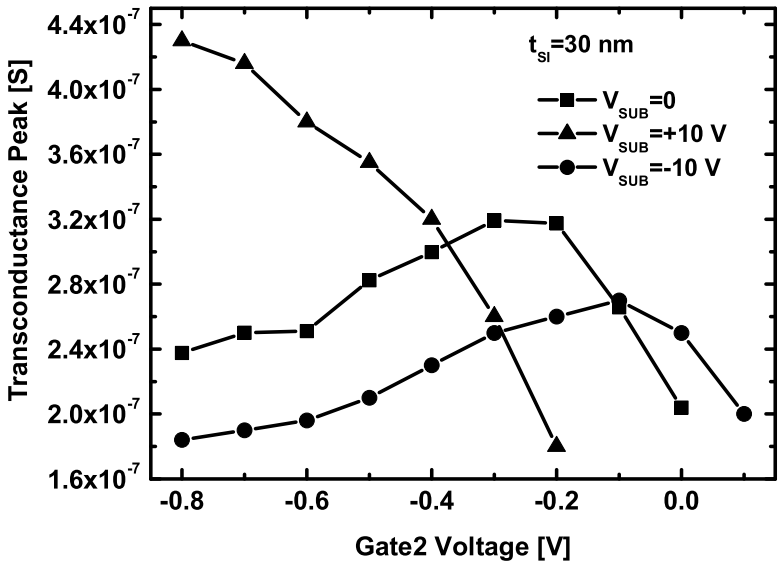

Figure 7.15: Maximum values of the transconductances for different back-gate bias, $t_{SI} = 30$ nm.

Part IV

Conclusions

In this thesis we have discussed physical issues which are important for MOSFET devices with decananometric gate length. Our research has aimed to provide useful results about the transport properties and electrostatic behavior of MOSFETs for the next technology nodes and for various types of applications.

In the first part we have presented a simulation study of nano-scale devices by using the Monte Carlo approach. Our MC model reproduces the quantization of the inversion charge as well as the universal mobility curves of bulk and SOI MOSFETs. Accurate MC simulations demonstrate that scattering still controls the ON current of decananometric devices, the main role being played by scattering events near the source barrier. Thus devices for the next technology nodes will not be in the full ballistic transport regime. The effect of scattering on the on-current I_{ON} is not simply proportional to the number of back-scattering events, since scattering changes the potential profile along the channel. We have validated the model proposed in [13]-[16] but we have seen that the evaluation of some model-parameter, like the kT-layer length, is not an easy task. For these reasons, it is not possible to use ballistic simulation tools to estimate the quasi-ballistic current. We have then demonstrated that the self-consistent Monte Carlo approach, used in this work, is a valid choice to study transport in advanced devices.

The same analysis of the transport properties has been extended to a large set of devices designed for many future technology nodes. We pointed out, by using Monte Carlo simulations, that, despite the fact that transport will be affected by scattering for L_{G} down to at least 14 nm, a significant increase of the degree of ballisticity is expected for the future TNs. This improvement follows a very tight correlation to L_{G}. Both I_{ON} and the transit time τ_{DC} will benefit of the increased ballisticity. In this sense, quasi-ballistic transport will contribute to boost the performance of scaled devices of technology generations below approximately 50 nm. An interpretation of the results is provided based on the length of the kT-layer. Improvements related to reduced scattering, are more evident in the case of low-doping UTB-DG SOI devices, than in bulk ones. In the latter case, our results point out that ballistic effects are limited by a larger impact of surface roughness effects, consequent to the high channel doping that is needed to keep short-channel effects at an acceptable level. In the case of UTB-DG MOSFETs, SOI phonons are dominant and there is no way to reduce their impact, other than changing the material properties or the operating temperature.

The same Monte Carlo code, extended for the time-dependent analysis,

has been used to simulate the AC performance of BULK MOSFETs for analog mixed-signal and RF applications with L_G down to 37 nm. The results confirm good scaling properties for the main AC parameters, F_T and 3dB bandwidth. In particular results point out that the quasi-static approach, requiring only efficient DC simulations, is adequate for the analysis of the signal-propagation delay and transition frequency. This methodology allows for a distributed analysis that describes how the delay builds-up along the channel and indicates that the delay is directly related to the carrier average velocity along the whole intrinsic channel region.
Our results indicate that UTB-SG SOI MOSFETs can provide significantly larger transconductance and F_T, thanks to the larger carrier velocity along the channel, and larger voltage-gain bandwidth, due to the reduced drain capacitance. On the other hand, they feature larger output conductance due to slightly larger SCE.

In the second part we have studied non-conventional coupling effects that occur in ultra-thin SOI devices. The co-existence of a front inversion charge and a back accumulation charge cannot be accommodated in films with sub-critical thickness. A critical film thickness has been analytically defined. As a consequence, it becomes impossible to characterize one channel independently of the opposite channel, by screening the opposite interface defects via accumulation. Super-coupling is responsible for a flat potential profile that can be equally controlled from either gate. These effects are critical for the operation of single-gate SOI MOSFETs because 'as-measured' front-channel properties include contributions from the front interface, back interface, and BOX (especially if damaged by radiations) which are not easy to isolate. The transconductance does no longer reflect the front-channel mobility as it integrates the mobility profile across the film.

Other non-conventional coupling effect have been evaluated in non-planar devices featuring independent gate contacts. We have separated the lateral coupling between the lateral gates and the vertical coupling between the lateral gates and the back-gate. The lateral coupling permits to modify the threshold voltage and can be used for characterization purposes or circuit design. The subthreshold characteristics and the transconductance are always degraded with respect to the double-gate case. The effect of the substrate bias is quite modest, if the fin thickness is very thin, both in Single-Gate and in Double-Gate mode.

Bibliography

[1] G.Moore "Cramming more components onto integrated circuits", *Electronics*, 1965, vol. 38, no. 8.

[2] H.C. Pao and C.T.Sah "Effects of diffusion current on characteristics of metal-oxide(insulator)-semiconductor transistors", *Solid-State Electron*, 1966, 9, p. 927.

[3] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous and A. R. Leblanc "Design of ion-implanted MOSFET's with very small physical dimensions", *IEEE Journal of Solid-State Circuits*, 1974, vol. 9, no. 5, p. 256.

[4] G. Baccarani, M. Wordeman and R. H. Dennard "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design", *IEEE Transactions on Electron Devices*, 1984, vol. 31, no. 4, p. 452.

[5] International Technology Roadmap for Semiconductors, http://www.itrs.net/

[6] A.G. Sabnis, J.T. Clemens "Characterization of the electron mobility in the inverted ¡100¿ Si surface", *IEDM Technical Digest*, 1979, p. 18.

[7] S. Takagi, M. Iwase and A. Toriumi "On the universality of inversion-layer mobility in n- and p- channel MOSFETs", *IEDM Technical Digest*, 1988, p. 398.

[8] R.R. Troutman "VLSI limitations for drain-induced barrier lowering", *IEEE Transactions on Electron Devices*, 1979, ED-26, p. 461.

[9] R.W. Coen and R.S. Muller "Velocity of surface carriers in inversion layers on silicon", *Solid-State Electron*, 1980, 23, p. 35.

[10] F. Stern and W.E. Howard "Properties of semiconductor surface inversion layers in the electric quantum limit", *Physical Review*, 1967, 163, p. 816.

[11] K. Natori "Current-voltage characteristics of silicon on insulator metal oxide semiconductor field effect transistors in ballistic mode", *Japanese Journal of Applied Physics*, 1994, 33, p. 554.

[12] K. Natori "Ballistic metal-oxide-semiconductor field effect transistor", *Journal of Applied Physics*, 1994, 76(8), p. 4879.

[13] M. Lundstrom "Scattering theory of the short channel MOSFET", *IEDM Technical Digest*, 1996, p. 387.

[14] M. Lundstrom "Elementary scattering theory of the Si MOSFET", *IEEE Electron Device Letters*, 1997, vol. 18, no. 7, p. 361.

[15] M. Lundstrom and Z. Ren "Essential physics of carrier transport in nanoscale MOSFETs", *IEEE Transactions on Electron Devices*, 2002, vol. 49, no. 1, p. 133.

[16] M. Lundstrom "On the mobility versus drain current relation for a nanoscale MOSFET", *IEEE Electron Device Letters*, 2001, vol. 22, no. 6, p. 293.

[17] M. Watanabe and A. Tooi "Formation of SiO_2 films by oxygen-ion bombardment", 1966, vol. 5, p. 737.

[18] R.–H. Yan, A. Ouzmard and K.F. Lee "Scaling the Si MOSFET: from bulk to SOI to bulk", *IEEE Transactions on Electron Devices*, 1992, vol. 39, no. 7, p. 1704.

[19] Y. Taur "An Analytical solution to a Double-Gate MOSFET with undoped body", *IEEE Electron Device Letters*, 2000, vol. 21, no. 5, p. 245.

[20] J.P. Colinge "Multiple-gate SOI MOSFETs", *Solid-State Electronics*, 2004, vol. 48, no. 6. p. 897.

[21] B. Mazhari, S. Cristoloveanu, D.E. Ioannou and A.L. Caviglia "Properties of ultra-thin wafer-bonded silicon-on-insulator MOSFET's", *IEEE Transactions on Electron Devices*, 1991, vol. 38, no. 6, p. 1289.

[22] L.T. Su *et al.* "Measurement and modeling of self-heating effects in SOI nMOSFETs", *IEDM Technical Digest*, 1992, p.357

[23] H-K Lim, J.G. Fossum "Threshold voltage of thin-film silicon-on-insulator (SOI) MOSFETs", *IEEE Transactions on Electron Devices*, 1983, vol. 30, no. 10, p. 1244.

[24] R. Chau *et al.* "Advanced depleted-substrate transistors: single-gate, double-gate and tri-gate", *SSDM Technical Digest*, 2002, p. 68.

[25] D. Hisamoto *et al.* "A Folded-channel MOSFET for deep-sub-tenth micron era", *IEDM Technical Digest*, 1998, p. 1032.

[26] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini and T. Elewa "Double-gate silicon on insulator transitor with volume inversion: a new device with greatly enhanced performance", *IEEE Electron Device Letters*, 1987, vol. 8, p. 410.

[27] D.Esseni, M.Mastrapasqua, GK Celler, C.Fiegna, L.Selmi, E.Sangiorgi, "An Experimental study of mobility enhancement in ultrathin SOI transistors operated in double-gate mode", *IEEE Trans. on Electron Devices*, 2003, vol.50, p. 802.

[28] TCAD User Manual, www.synopsys.com .

[29] S. Reggiani, M. Valdinoci, L. Colalongo, G. Baccarani "A unified analytical model for bulk and surface mobility in Si n- and p-channel MOSFETs", *ESSDERC Proceedings*, p. 1999.

[30] M. G. Ancona and G. J. Iafrate "Quantum correction to the equation of state of an electron gas in a semiconductor" *Phys. Rev B*, 1989, vol. 39, no. 13, p. 9536.

[31] C. Canali, G. Majni, R. Minder, and G. Ottaviani "Electron and hole drift velocity measurements in Silicon and their empirical relation to electric field and temperature" *IEEE Trans. on Electron Devices*, 1975, vol. ED-22, p. 1045.

[32] M. Lundstrom, Fundamentals of carrier transport, Second Edition, Cambridge University Press, 2000.

[33] C. Jacoboni and L. Reggiani "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials", *Rev. Mod. Phys.*, 1983, vol. 55, p. 645.

[34] P. Palestri, D. Esseni, A. Abramo, R. Clerc, L. Selmi "Carrier quantization in SOI MOSFETs using an effective potential based Monte-Carlo tool", *Solid State Electronics*, 2005, vol. 49, no. 5, p. 727.

[35] DK.Ferry, R.Akis, D.Vasileska, "Quantum effects in MOSFETs: use of an effective potential in 3D Monte Carlo simulation of ultra-short channel devices", *IEDM Technical Digest*, 2000, p. 287.

[36] E. Sangiorgi and M.R. Pinto "A semi-empirical model of surface roughness scattering for Monte Carlo simulation of silicon N-MOSFETs", *IEEE Trans. on Electron Devices*, 1992, vol. 39, p. 356.

[37] P. Palestri, S. Eminente, D. Esseni,C. Fiegna, E. Sangiorgi, L. Selmi "An improved semi-classical Monte-Carlo approach for nano-scale MOSFET simulation", *Solid State Electronics*, 2005, vol. 49, no. 5, p. 727.

[38] D. Esseni, A. Abramo, L. Selmi, E. Sangiorgi "Physically based modeling of low field electron mobility in ultrathin Single- and Double-Gate SOI n-MOSFETs", *IEEE Trans. on Electron Devices*, 2003, vol.50, p. 2445.

[39] P. Riolino M. Braccioli, L. Lucci, D. Esseni, C. Fiegna, P. Palestri, and L. Selmi, "Monte-Carlo simulation of decananometric Double-Gate SOI devices: multi-subband vs. semiclassical with quantum corrections", *ESSDERC 2006 Proceedings*

[40] L.Lucci, P.Palestri, D.Esseni, L.Selmi "Multi-subband Monte Carlo modeling of nano-MOSFETs with strong vertical quantization and electron gas degeneration", *IEDM Technical Digest*, 2005, p. 631.

[41] N. Barin, P. Palestri, C. Fiegna "Monte-Carlo analysis of signal propagation delay and AC performance of decananometric bulk and double-gate MOSFETs", *ULIS 2006 Proc.*, 2006, p.85.

[42] D. Munteanu and J. L. Autran "Two-dimensional modeling of quantum ballistic transport in ultimate Double-Gate SOI Devices", *Solid State Electronics*, 2003, vol. 47, p. 1219.

[43] S. Datta, Electronic transport in mesoscopic systems, Cambridge University Press, 1998.

[44] S. Datta, "The non-equilibrium Green's function (NEGF) formalism: an elementary introduction", *IEDM Technical Digest*, 2002, p. 703.

[45] A. Svizhenko and M.P. Anantram "Role of scattering in nanotransistors", *IEEE Trans. on Electron Devices*, 2003, vol. 50, p. 1459.

[46] P. Palestri, D. Esseni, S. Eminente, C. Fiegna, E. Sangiorgi and L. Selmi "Understanding quasi-ballistic transport in nano-MOSFETs part I: scattering in the channel and in the drain", *IEEE Trans. on Electron Devices*, 2005, vol. 52, no. 12, p. 2727.

[47] C. Jungemann, N. Subba, J.-S. Goo, C. Riccobene, Q. Xiang and B. Meinerzhagen "Investigation of strained Si/SiGe devices by MC simulation", *Solid State Electronics* ,2004, vol. 48, p. 1417.

[48] A. Svizhenko and M.P. Anantram "Role of scattering in nanotransistors", *IEEE Trans. on Electron Devices*, 2003, vol. 50, no. 6, p. 1459.

[49] M.Mouis and S. Barraud "Velocity distribution of electrons along the channel of nanoscale MOS transistors", *ESSDERC 2003 Proceedings*, 2003, p. 147.

[50] M. V. Fischetti and S. E. Laux "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", *Physical Review B*, 1988, vol. 38, no. 14, p. 9721.

[51] M.V. Fischetti and S.E. Laux "Long-range coulomb interactions in small Si devices. Part I: performance and reliability", *Japanese Journal of Applied Physics*, 2001, vol. 89, no. 2, p. 1205–1231.

[52] S. Eminente, D. Esseni, P. Palestri, C. Fiegna, L. Selmi, E. Sangiorgi "Understanding quasi-ballistics transport in nano-MOSFETS: Part II technology scaling along the ITRS ",*I*EEE Trans. on Electron Devices , 2005, Vol. 52, no. 12, p. 2736.

[53] C.Fiegna "Physics-based analysis of the RF performance of small geometry MOSFETs: methodology and application to the evaluation of the effects of scaling", *IEDM Technical Digest*, 1999, p. 543.

[54] F. Schwierz "RE Transistors: performance trends versus ITRS targets", *I*CCDCS Proc.,2006, p. 195.

[55] F. Ellinger "26-42 GHz SOI CMOS low noise amplifier", *I*EEE J. Solid-State Circuits, 2004, Vol. 39, N. 3, p. 522.

[56] B. Razavi "A 60-GHz CMOS receiver front-end", *I*EEE J. Solid-State Circuits, 2006, Vol. 41, N. 1, pp. 17.

[57] J.J.H. van den Biesen, "A simple regional analysis of transit times in bipolar transistors", *S*olid State Electr., 1986, Vol. 29, no. 5, p. 529.

[58] S. Babiker , A. Asenov, N. Cameron, S. P. Beaumont and J. R. Barker, "Complete Monte Carlo RF analysis of "real" short-channel compound FET's", *IEEE Trans. on Electron Devices*, 1998, vol. 45, no. 8, p. 1644.

[59] R. Rengel, J. Mateos, D. Pardo, T. González, and M. J. Martín, "A Monte Carlo analysis of dynamic and noise performance of submicron MOSFETs at RF and microwave frequencies", *S*emicond. Sci. Technol., 2001, vol. 16, p. 939.

[60] R. Rengel et al., "Numerical and experimental study of a 0.25 m fully depleted silicon-on-insulator MOSFET: static and dynamic radio-frequency behavior", *Semicond. Sci. Technol.*, 2002, vol. 17, p. 1149.

[61] H. Kim, H.S. Min, T.W. Tang and Y.J. Park "An extended proof of the Ramo-Shockley theorem", *Solid-State Electronics*, 1992, vol. 34, no. 11, p. 1251.

[62] N. Barin, P. Palestri, C. Fiegna "Monte-Carlo analysis of signal propagation delay and AC performance of decananometric Bulk and Double-Gate MOSFETs", accepted for publication in *Solid State Electronics*.

[63] N. Barin, P. Palestri, C. Fiegna "Monte-Carlo analysis of signal propagation delay and AC performance of decananometric bulk and double-gate MOSFETs", *U*LIS 2006 Tech. Dig., 2006, pp.85-88.

[64] S. Eminente, N. Barin, P. Palestri, C. Fiegna, E. Sangiorgi "Monte Carlo simulation of deca-nanometer MOSFETs for analog/mixed-signal and RF applications", *I*EDM Technical Digest, 2006.

[65] J.E. Carrol, Rate equations in semiconductor electronics, Cambridge University Press, 1985.

[66] S.Cristoloveanu and S.S. Li, Electrical characterization of Silicon-On-Insulator materials and devices, pp. 240-260, Kluwer 1995.

[67] J. Pretet, A. Ohata, F. Dieudonne, F. Allibert, N. Bresson, T. Matsumoto, T. Poiroux, J. Jomaah, S. Cristoloveanu, "Scaling issues for advanced SOI devices : gate oxide tunneling, thin buried oxide, and ultra-thin films", *7th International Symposium Silicon Nitride and Silicon Dioxide Thin Insulating Films*, Paris, France 2003.

[68] C.Aydin, *et al.* "Double Gate coupling and quantum tunneling in ultrathin SOI MOSFETs" in Future Trends in Microelectronics, Wiley & Sons, 2004.

[69] S. Eminente, S. Cristoloveanu, R. Clerc, A. Ohata and G. Ghibaudo "Charge properties and coupling effects in ultra-thin body fully-depleted SOI MOSFETS", *Eurosoi Conf. Proc.*, 2006.

[70] S. Eminente, S. Cristoloveanu, R. Clerc, A. Ohata and G. Ghibaudo "Ultra-thin fully-depleted SOI MOSFETs: special charge properties and coupling effects", accepted for publication in *Solid State Electronics*.

[71] E.R. Worley "Theory of the fully depleted SOS/MOS transistor", *Solid State Electronics*, 1980, vol. 33, p. 1907.

[72] E. Sano, R. Kasai, K. Ohwada and H. Ariyoshi, "A two-dimensional analysis for MOSFET's fabricated on buried S_iO_2 layer", *IEEE Trans. on Electron Devices*, 1980, vol. 27, p. 2043.

[73] P.W. Barth " Dielectric isolation technology for bipolar and MOS integrated circuits", Technical Report Stanford Electronics Laboratories, Stanford University, 1980.

[74] H.-K. Lim and J.G. Fossum "Threshold voltage of thin-film Silicon-on-Insulator (SOI) MOSFET's", *IEEE Trans. on Electron Devices*, 1983, vol. 30, no. 10, p. 1244.

[75] J.R. Brews "Physics of the MOS transistor" in Silicon Integrated Circuits, editor D. Kahng, Academic Press 1981.

[76] O. Faynot, S. Cristoloveanu, A. J. Auberton-Hervé and C. Raynaud "Performance and potential of ultrathin accumulation-mode SIMOX MOSFET's", *IEEE Trans. on Electron Devices*, 1995, vol. 42, no. 4, p.913.

[77] A. Ohata, S. Cristoloveanu, A. Vandooren, M. Cassé and F. Daugé "Coupling effect between the front and back interfaces in thin SOI MOSFETs", *Solid State Electronics*, 2005, vol. 40, p. 245.

[78] Schred Simulation Tool, www.nanohub.org.

[79] Y. Tsividis, Operation and Modeling of the MOS Transistor, McGraw-Hill, 1987, pp. 70-71.

[80] A. Ohata, J. Pretet, S. Cristoloveanu, A. Zaslavsky, "Correct biasing rules for virtual DG mode operation in SOI-MOSFETs" *IEEE Trans. on Electron Devices*, 2005, vol. 52, no. 1, pp. 124.

[81] Y. Omura, T. Ishiyama, M. Shoji et K. Izumi, "Quantum mechanical transport characteristics in ultimately miniaturized MOSFET's/SIMOX" in SOI Technology and Devices VII, Electrochem. Soc., Pennington, p. 199-211, 1996.

[82] A.M. Ionescu, S. Cristoloveanu and A. Rusu "Derivative methods for threshold voltage extraction", *16th ann. Semiconductor Conf. Proc.*, 1993, p. 197.

[83] G. Ghibaudo "New method for the extraction of MOSFET parameters", *Electronic Letters*, 1988, Vol. 24, p.543.

[84] L. Mathew et al. , "CMOS vertical multiple independent gate field effect Transistor (MIGFET)", *IEEE International SOI Conf.Proc.*, 2004.

[85] S. Eminente, Kyoung-Il Na, S. Cristoloveanu, L. Mathew and A. Vandooren "Lateral and vertical coupling effects in MIGFETs" *2005 International SOI Conference Proc.*, 2005.

[86] F. Daugé, J. Pretet, S. Cristoloveanu, A. Vandooren, L. Mathew, J. Jomaah and B.-Y. Nguyen "Coupling effects and channels separation in FinFETs", *Solid State Electronics*, 2004, Vol. 48, no. 4, p. 535.

[87] R. Ritzenthaler, S. Cristoloveanu, O. Faynot, C. Jahan, A. Kuriyama, L. Brevard and S. Deleonibus "Lateral coupling and immunity to substrate effect in ΩFET devices", *Solid State Electronics*, 2006, Vol. 50, no. 4, p. 558.

[88] J. Frei *et al.* "Body effect in Tri- and Pi- Gate SOI MOSFETs", *IEEE Electron Device Letters*,2004, vol. 25, p. 813.

[89] J. P. Colinge, J.W. Park and W. Xiong, "Threshold voltage and subthreshold slope of multiple-gaae SOI MOSFETs"; *IEEE Electron Device Letters*, 2003, vol. 24, p. 515.