



**Étude des mécanismes de dégradation de la mobilité sur
les architectures FDSOI pour les noeuds technologiques
avancés (**
Sebastien Guarnay

► **To cite this version:**

Sebastien Guarnay. Étude des mécanismes de dégradation de la mobilité sur les architectures FDSOI pour les noeuds technologiques avancés (

HAL Id: tel-01219287

<https://tel.archives-ouvertes.fr/tel-01219287>

Submitted on 22 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Comprendre le monde,
construire l'avenir®



UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : STITS
Laboratoire de Simulation et de Modélisation

DISCIPLINE PHYSIQUE

THÈSE DE DOCTORAT

soutenue le 21/04/2015

par

Sébastien GUARNAY

Étude des mécanismes de dégradation de la
mobilité sur les architectures FDSOI pour les
nœuds technologiques avancés (<20nm)

Directeur de thèse :

Arnaud BOURNEL

Professeur (Université Paris-Sud - IEF)

Composition du jury :

Président du jury :

Abdelkader SOUIFI

Professeur (INSA Lyon)

Rapporteurs :

Mireille MOUIS

Directrice de Recherche CNRS (IMEP-LAHC, Grenoble)

Marc BESCOND

Chargé de recherche CNRS (IM2NP, Marseille)

Examineur :

François TRIOZON

Ingénieur de Recherche (CEA-Leti, Grenoble)

Membre invité :

Sébastien MARTINIE

Ingénieur de Recherche (CEA-Leti, Grenoble)

Sommaire

Notations	5
Abréviations.....	7
Index	8
Table des figures.....	9
Introduction générale.....	14
Chapitre I. Vers les transistors ultimes et leurs limites	16
1. Introduction au transistor	17
1.1. Description du transistor MOSFET	17
2. Caractéristiques courant-tension	21
2.1. Les différents régimes	21
2.2. Un modèle simple.....	22
2.3. Perte de contrôle électrostatique	23
3. Évolution des MOSFET.....	25
3.1. Lois de changement d'échelle	25
3.2. Architectures ultimes de transistors	26
4. Réduction de mobilité à faible longueur de grille	31
4.1. Mise en évidence expérimentale	31
4.2. Modélisation des effets de dégradation	33
4.3. Mécanisme d'interaction avec des défauts neutres	34
4.4. Influence des défauts	34
4.5. Régime quasi-balistique	36
4.6. Effets de résistance d'accès et contraintes	38
5. Conclusion du chapitre	38
Chapitre II. Modélisation du transport.....	39
1. Structure de bandes	40
1.1. Principe de calcul de structure de bandes	40
1.2. Modèle de masse effective	43
1.3. Confinement quantique	44
2. Équation de transport de Boltzmann	47
2.1. Hypothèse semi-classique	47

2.2.	Équation de Boltzmann 2Dk	48
2.3.	Calcul des taux de transition	49
2.4.	Interactions considérées	50
2.5.	Grandeurs physiques obtenues.....	51
3.	Modélisation Kubo-Greenwood	51
3.1.	Linéarisation de l'équation de Boltzmann et expression de la mobilité	52
3.2.	Calcul des temps de collision.....	53
4.	Simulation Monte-Carlo 2Dk.....	54
4.1.	Résolution de l'équation de Boltzmann 2Dk	55
4.2.	Résolution de l'équation de Schrödinger 1D.....	57
4.3.	Résolution de l'équation de Poisson 2D.....	58
5.	Simulation TCAD	59
5.1.	Approximations sur la fonction de distribution	60
5.2.	Modèles classiques.....	61
5.3.	Modèle avec corrections quantiques	64
6.	Conclusion du chapitre	64
Chapitre III. Méthodes d'extraction de la mobilité		66
1.	Théorie sur les méthodes d'extraction.....	67
1.1.	Méthode de Hamer	67
1.2.	Méthode de la fonction Y.....	68
1.3.	Extraction de la résistance d'accès.....	71
1.4.	Méthode split-CV.....	73
2.	Extraction sur les simulations Monte-Carlo	74
2.1.	Extraction sur le courant seul.....	74
2.2.	Extraction sur le courant et la charge.....	75
3.	Extraction sur données expérimentales.....	77
3.1.	Choix de la méthode.....	77
3.2.	Conclusion du chapitre	78
Chapitre IV. Analyse de la mobilité à canal court par simulation Monte-Carlo		80
1.	Grandeurs microscopiques extraites.....	81
1.1.	Énergie potentielle	81
1.2.	Fonctions d'onde	82
1.3.	Transport non stationnaire.....	83
2.	Mobilités extraites sur les simulations.....	85

2.1. Extraction sur les simulations.....	85
2.2. Mobilité extraite en fonction de la longueur	86
3. Modélisation de la dégradation de mobilité.....	87
3.1. Loi de Matthiessen	87
3.2. Mobilité à canal long	89
3.3. Résistances balistique et d'accès.....	89
3.4. Autre évaluation de la résistance d'accès	93
4. Comparaison avec les mobilités expérimentales	97
4.1. Analyse du transport par les méthodes classiques	97
5. Bilan sur le transport à canal court	101
Conclusion générale	103
Annexes	104
1. Calcul des taux de transition	104
1.1. Interaction avec les phonons	104
1.2. Interactions coulombiennes avec les impuretés ionisées.....	105
1.3. Rugosité d'interface avec l'oxyde.....	105
2. Calcul de la résistance balistique.....	107
Références bibliographiques	109

Notations

Constante	Signification	Valeur	Unité
\hbar	constante de Planck réduite (ou de Dirac)	$1,054.10^{-34}$	eV.s
k_B	constante de Boltzmann	$8,617.10^{-5}$ ou $1,38.10^{-23}$	eV/K J
q	charge élémentaire de l'électron	$1,602.10^{-19}$	C
m_0	masse de l'électron	$9,109.10^{-31}$	kg
m_l	masse longitudinale	$0,916 m_0$	kg
m_t	masse transverse	$0,191 m_0$	kg
ϵ_0	permittivité diélectrique du vide	$8,85.10^{-12}$	F/m
ϵ_{Si}	permittivité diélectrique du silicium	$3,9 \epsilon_0$	F/m
ϵ_{SiO_2}	permittivité diélectrique de l'oxyde de silicium	$11,8 \epsilon_0$	F/m

Grandeur	Signification	Valeur	Unité
B_{eff}	taux de balisticité effectif : rapport du courant sur le courant d'un même transistor dont le canal serait balistique.	$I_{ON}/I_{ON\ bal}$	%
B_{int}	taux de balisticité intrinsèque : proportion de porteurs balistiques, i.e. ne subissant aucune interaction dans le canal		%
C_{ox}	capacité d'oxyde	ϵ_{ox}/t_{ox}	F/m ²
DIBL	diminution de la tension de seuil par la tension de drain dans les canaux courts (<i>Drain Induced Barrier Lowering</i>)	$\frac{V_{t\ lin} - V_{t\ sat}}{V_{dd} - V_{d\ lin}}$	mV/V ou mV
E_c	énergie de la bande de conduction		eV
E_f	énergie niveau de Fermi		eV
E_v	énergie de la bande de valence		eV
ϕ_m	tension correspondant au travail de sortie $e\phi_m$ du métal (énergie pour libérer un électron du niveau de Fermi au niveau du vide)		V
g_m	transconductance de drain (dynamique)	$\partial I_d / \partial V_{gs}$	S ou S/m
g_d	conductance de drain (dynamique)	$\partial I_d / \partial V_{ds}$	S ou S/m
I_d	courant de drain		A ou A/m
I_{OFF}	courant de drain à l'état bloqué $V_{ds} = V_{dd}$ et $V_{gs} = 0$ V		A ou A/m
I_{ON}	courant de drain à l'état passant $V_{ds} = V_{gs} = V_{dd}$		A ou A/m
L_c	longueur de canal		nm
L_{eff}	longueur effective de canal L_c ôtée des longueurs de zones de charge d'espace (ZCE)		nm
L_g	longueur de grille		nm
m^*	masse effective		kg
μ_{eff}	mobilité effective		cm ² /Vs
μ_{cs}	composante de la mobilité due aux interactions avec les impuretés ionisées (<i>Coulomb scattering</i>)		cm ² /Vs
μ_{ph}	composante de la mobilité due aux phonons		cm ² /Vs
μ_{sr}	composante de la mobilité due à la rugosité de surface (<i>surface roughness</i>)		cm ² /Vs
n_i	concentration volumique de porteurs libres (charge d'inversion)		cm ⁻³

Notations

N_i	concentration surfacique de porteurs libres (charge d'inversion)	cm^{-2}
N_N	concentration de défauts neutres	cm^{-3}
t_{box}	épaisseur du BOX	nm
t_{Si}	épaisseur du canal de silicium	nm
t_{ox}	épaisseur d'oxyde	nm
V_{dd}	tension d'alimentation	V
V_{ds}	tension de drain par rapport à la source	V
$V_{ds \text{ lin}}$	tension de drain en régime linéaire	V
$V_{ds \text{ sat}}$	tension de drain en régime de saturation	V
V_{gs}	tension de grille par rapport à la source	V
V_t	tension de seuil (<i>threshold voltage</i>)	V
R	coefficient de rétrodiffusion	-
R_{sd}	résistances d'accès (source et drain) parasites	$\Omega \cdot \mu\text{m}$
S	pente sous le seuil (<i>subthreshold slope</i>)	$\frac{\partial V_{gs}}{\partial (\log I_{ds})}$ mV/dec
T	température	K
v_{th}	vitesse thermique	$\sqrt{\frac{8k_B T}{\pi m^*}}$ m/s
W	largeur du transistor	nm
χ	tension liée à l'affinité électronique $e\chi$ du semi-conducteur (eV)	V
χ_i	tension liée à l'affinité électronique $e\chi_i$ de l'oxyde (eV)	V

Abréviations

BEOL	<i>Back End of Line</i> : mise en place des contacts, interconnexions une fois le transistor créé.
BOX	<i>Buried OXide</i> : oxyde mince enterré sous le canal de silicium pour l'isoler du substrat.
BTI	<i>Biased Temperature Instability</i>
CESL	<i>Contact Etch Stop Layer</i> : couche d'arrêt de gravure.
DIBL	<i>Drain Induced Barrier Lowering</i> : réduction de la tension de seuil par la tension de drain dans les canaux courts. $DIBL = (V_{t\ lin} - V_{t\ sat}) / (V_{dd} - V_{d\ lin})$.
DICE	<i>Drain Induced Conductivity Enhancement</i> : augmentation de la conductance du drain à la saturation dans les canaux courts.
Dit	<i>Densité d'états d'interface</i> : pièges aux interfaces.
DIVF	<i>Drain-Induced Vertical Field effect</i> : concerne plutôt le bulk.
DIVSB	<i>Drain-Induced Virtual Substrate Biasing</i> .
DLTS	<i>Deep Level Transient Spectroscopy</i> .
EOR	<i>End Of Range</i> .
EOT	<i>Equivalent Oxide Thickness</i> : épaisseur d'oxyde équivalente, excluant la zone de capacité quantique.
ETB	Équation de Transport de Boltzmann : équation régissant la fonction de distribution.
GIDL	<i>Gate Induced Drain Leakage</i> .
HCI	<i>Hot Carrier Injection</i> .
LDD	<i>Lightly Doped Drain</i> .
MPI	<i>Message Passing Interface</i> .
PID	<i>Process Induced Strain</i> .
RBB	<i>Reverse Body Bias</i> .
RPA	<i>Random Phase Approximation</i> .
RTA	<i>Relaxation Time Approximation</i> : approximation du temps de relaxation.
SCE	<i>Short Channel Effects</i> : effets de canal court.
SEM	<i>Scanning Electron Microscopy</i> : microscope électronique à balayage (MEB).
SMT	<i>Stress Memory Technique</i> : technique permettant d'appliquer une contrainte en tension.
SPER	<i>Solid Phase Epitaxy Regrowth</i> .
STI	<i>Shallow Trench Isolation</i> : tranchées d'isolation peu profondes.
TED	<i>Transient Enhanced Diffusion</i> : effet d'élargissement inattendu des profils d'implantation.
TEM	<i>Transmission Electron Microscopy</i> : microscope électronique en transmission.
TSV	<i>Through Silicon Via</i> : interconnexions à travers le silicium pour l'intégration 3D.
ZCE	<i>Zone de Charge d'Espace</i> : zone du canal désertée en porteurs majoritaires mais pourvue de charges fixes issues de la diffusion depuis les accès, i.e. de la pénétration des dopants sous l'oxyde.

Index

affinité électronique, 17
Boltzmann, 48
canal court, 22
déplétion, 18
DIBL, 23
double grille, 28
FDSOI, 25
FinFET, 28
halo
 HDD, 26
 LDD, 26
impuretés ionisées, 104
Kubo-Greenwood, 48
masse effective, 42
Monte-Carlo, 53
MOSFET, 16
phonons, 103
rugosité de surface, 104
scaling (lois de changement d'échelle), 24
simulation
 Monte-Carlo, 53
split-CV, 72
tension de seuil, 18
travail de sortie, 17
trigate, 28

Table des figures

Figure 1 : schéma simplifié d'un transistor à canal N (bulk)	17
Figure 2 : diagramme de bandes d'énergie dans la direction normale à l'interface oxyde/semiconducteur en régime d'accumulation	18
Figure 3 : tension de bandes plates.....	19
Figure 4 : inversion faible (désertion ou déplétion).....	19
Figure 5 : inversion forte	20
Figure 6 : courant de drain en fonction de la tension de drain appliquée, à différentes tensions de grille. Identification des régimes de fonctionnement.	21
Figure 7 : courant de drain en fonction de la tension de grille appliquée, à différentes tensions de drain. Identification des régimes de fonctionnement.	22
Figure 8 : courant de drain à l'échelle logarithmique en fonction de la tension de grille pour une tension V_{ds} donnée. Représentation de la pente sous le seuil et des courants I_{ON} et I_{OFF}	22
Figure 9 : effet canal court SCE et de DIBL sur la barrière de potentiel. Simulations à $t_{Si} = 4$ nm et $V_g = 0,2$ V. $L_g = 30$ nm et $V_{ds} = 50$ mV en ligne continue, $L_g = 10$ nm et $V_{ds} = 50$ mV en tirets longs, $L_g = 10$ nm et $V_{ds} = 0,8$ V en tirets courts.....	24
Figure 10 : effets de canal court et décalages de courbe associés. $V_{t\ lin} = V_{gs}$ à $I_{ds} = I_t$ (courant de seuil de l'ordre de $1\ \mu A/\mu m$) et $V_{ds} = V_{ds\ lin}$, $V_{t\ sat} = V_{gs}$ à $I_{ds} = I_t$ et $V_{ds} = V_{ds\ lin}$, $I_{ON} = I_{ds}$ à $V_{gs} = V_{ds} = V_{dd}$, $I_{OFF} = I_{ds}$ à $V_{gs} = 0$ V et $V_{ds} = V_{dd}$, $S = \partial V_{ds} / \partial \log I_{ds}$ pente sous le seuil, $DIBL = V_{t\ lin} - V_{t\ sat} / V_{dd} - V_{ds\ lin}$ [ESSENI11]	24
Figure 11 : schéma simplifié d'un transistor FDSOI et notations utilisées	27
Figure 12 : coupe TEM d'un FDSOI [MORVAN12].....	27
Figure 13 : fabrication de SOI par technologie Smart Cut™ de SOITEC.....	28
Figure 14 : transistor double grille.	29
Figure 15 : schéma du FinFET.	30
Figure 16 : coupe TEM d'un FinFET.	30
Figure 17 : nanofil <i>gate all around</i>	30
Figure 18 : coupe TEM d'un nanofil [BIDAL09]	31
Figure 19 : courbes de décroissance de mobilité μ_0 avec L_{eff} [BIDAL09].....	32
Figure 20 : dégradation de mobilité effective (i. e. dépendant de la charge en abscisse) avec la réduction de L_g [RAMOS06].....	33
Figure 21 : évolution typique de la mobilité en fonction de L_{eff} . Insert : représentation schématique des dépendances aux mécanismes d'interaction en fonction du champ effectif E_{eff} . ..	34
Figure 22 : localisation proche des interfaces canal/source et canal/drain. Augmentation de la densité moyenne globale de défauts avec la réduction de la longueur de canal.....	36
Figure 23 : régime balistique	36
Figure 24 : régime quasi-balistique	36
Figure 25 : régime diffusif	36
Figure 26 : maille cubique de silicium. Structure « diamant » composée d'un réseau cfc en bleu, et d'un autre décalé de $a/4, a/4, a/4$ représenté en vert. Paramètre de maille : $a = 0,357$ nm à température ambiante.....	41

Figure 27 : première zone de Brillouin du réseau réciproque du silicium, présentant les principaux points de haute symétrie Γ , X, L et les droites caractéristiques associées Δ et Λ .	41
Figure 28 : structure de bandes du silicium. La vallée Δ est visible en bas de bande de conduction. Quant aux trous, les deux bandes les plus hautes de la bande de valence sont représentées (trous lourds et trous légers).	42
Figure 29 : ellipsoïdes isoénergies autour des six vallées Δ de la bande de conduction du silicium et situés le long des directions principales du réseau réciproque.	42
Figure 30 : diagramme de bandes du silicium.	42
Figure 31 : coupe d'énergie potentielle selon y . $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.	46
Figure 32 : coupe de concentration en électrons selon y . $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.	46
Figure 33 : coupe de la fonction d'onde en y dans le silicium. Vallée Δy , premier niveau. $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.	46
Figure 34 : coupe de la fonction d'onde en y dans le silicium. Vallée Δy , deuxième niveau. $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.	47
Figure 35 : coupe de la fonction d'onde en y dans le silicium. Vallée Δx , premier niveau. $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.	47
Figure 36 : fonction de distribution de Fermi-Dirac à l'équilibre f_{n0} en fonction de l'énergie autour du niveau de Fermi, à différentes températures.	48
Figure 37 : bilan de variation de la fonction de distribution pendant dt sur un intervalle de longueur dx permettant d'obtenir l'équation de Boltzmann.	49
Figure 38 : image TEM de l'interface Si-SiO ₂ [ZHAO09].	50
Figure 39 : paramètres géométriques de rugosité de surface.	50
Figure 40 : géométrie du FDSOI dans la résolution Monte-Carlo, supposée invariante par translation selon l'axe z représenté en pointillés. Une tranche de résolution de l'équation de Schrödinger est représentée hachurée en vert selon l'axe y . Les axes des principales équations sont représentés au-dessus.	54
Figure 41 : illustration de la simulation Monte-Carlo à champ constant sur la trajectoire d'un électron. E est le champ électrique, v la vitesse de la particule simulée et chaque δt_i est un temps de vol libre entre deux interactions.	55
Figure 42 : cycle des séquences de la simulation Monte-Carlo.	55
Figure 43 : résolution de l'équation couplée Poisson-Schrödinger 1D [SAINT-MA06].	58
Figure 44 : algorithme détaillé de la simulation Monte-Carlo [JAUD06].	59
Figure 45 : fonction de distribution à faible champ sous forme de maxwelliennes déplacées selon l'axe de vecteur d'onde.	61
Figure 46 : potentiel dans un FDSOI 14 nm comprenant un empilement de TiN sur poly-Si avec un contact Schottky. $L = 20$ nm, $N_{poly} = 1018$ cm ⁻³ , $V_g = 1$ V et $V_{ds} = 50$ mV.	63
Figure 47 : ajustement des paramètres μ_0 et Vt sur la fonction Y dans l'exemple d'une simulation Monte-Carlo. Plage d'extraction : de 0,6 à 0,85 V. $L = 80$ nm et $V_{ds} = 50$ mV. En rouge, les points de mesure. En vert, la fonction affine ajustée.	69
Figure 48 : ajustement de paramètres θ_1 et θ_2 sur la fonction θ_{eff} . Plage d'extraction : de 0,6 à 0,85 V. $L = 80$ nm et $V_{ds} = 50$ mV.	69
Figure 49 : courant obtenu après les ajustements de paramètres précédents μ_0 et Vt sur la fonction Y , et θ_1 et θ_2 sur la fonction θ_{eff} . Plage d'extraction : de 0,6 à 0,85 V. $L = 80$ nm et $V_{ds} = 50$ mV.	70
Figure 50 : extraction de la mobilité par fonction Y améliorée.	71

Figure 51 : exemple de courbe $R_{tot}L$ (en bleu) avec ajustement de la droite (en rouge) pour l'extraction du paramètre $R_{sd} + R_{bal}$ à l'origine. $V_{gs} = 0,8 V$	72
Figure 52 : $R_{sd} + R_{bal}$ en fonction de N_{inv} (en rouge) et écart-type (en noir). Chaque valeur est obtenue par méthode $R_{tot}L$ à N_{inv} fixé.	72
Figure 53 : courant et ajustement par fonction Y de $\mu_0, \theta_1, \theta_2$ et V_t . Plage d'extraction : V_g allant de 0,55 à 0,8 V. $L = 80$ nm et $V_{ds} = 50$ mV.	74
Figure 54 : courant et ajustement par fonction Y de $\mu_0, \theta_1, \theta_2$ et V_t . Plage d'extraction : V_g allant de 0,55 à 0,9 V. $L = 80$ nm et $V_{ds} = 50$ mV.	74
Figure 55 : courant et ajustement par la méthode de Hamer de $\mu_0, \theta_1, \theta_2$ et V_t . $L = 80$ nm et $V_{ds} = 50$ mV.....	75
Figure 56 : concentration surfacique par sous-bande. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm. Représentation des zones d'extraction : canal complet en noir, zone réduite en tirets bleus.	76
Figure 57 : Charge en fonction de la tension de grille V_g intégrée sur les concentrations surfaciques précédentes. $L = 60$ nm. En trait noir, extraction sur le canal complet. En tirets bleus, extraction sur une zone réduite.	76
Figure 58 : mobilité effective extraite sur le courant et la charge. $L = 60$ nm.....	77
Figure 59 : courbes de capacités d'oxyde C_{ox} et avec suppression de l'offset	78
Figure 60 : charge d'inversion extraite par split-CV.....	78
Figure 61 : énergie potentielle dans le silicium. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.	81
Figure 62 : coupe d'énergie potentielle au milieu du canal, dans le BOX, le canal et l'oxyde. $L = 30$ nm, $t_{box} = 4$ nm, $t_{Si} = 4$ nm, $t_{ox} = 10$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV.....	81
Figure 63 : Évolution suivant l'axe source-drain (x) des niveaux quantifiés des 4 sous-bandes les plus basses. $L = 30$ nm, $t_{box} = 4$ nm, $t_{Si} = 4$ nm, $t_{ox} = 10$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV.	82
Figure 64 : densité de probabilité dans le silicium. Vallée Δy , premier niveau. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.	82
Figure 65 : densité de probabilité dans le silicium. Vallée Δy , deuxième niveau. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.	83
Figure 66 : densité de probabilité dans le silicium. vallée Δx , premier niveau. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.	83
Figure 67 : vitesse le long du canal des porteurs des différentes sous-bandes. $L = 30$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV.	84
Figure 68 : vitesse le long du canal des porteurs des différentes sous-bandes. $L = 30$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 0,8 V$	84
Figure 69 : répartition des électrons selon leur nombre d'interactions subies dans le canal. $L = 30$ nm, $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV.	84
Figure 70 : taux de balisticité intrinsèque en fonction de la longueur de canal. $V_{gs} = 0,8 V$ et $V_{ds} = 50$ mV.....	85
Figure 71 : $I_d(V_g)$ avec interactions dans le canal. Échelle logarithmique. $L = 10, 20, 30$ et 100 nm. $V_{ds} = 50$ mV.....	86
Figure 72 : mobilité effective extraite des simulations à canal diffusif (avec interactions dans le canal). Ligne continue : simulation Monte-Carlo. Pointillés : modèle analytique.....	86

Figure 73 : dégradation de la mobilité avec la réduction de la longueur de canal à différentes charges d'inversion : $N_{inv} = 5 \cdot 10^{12}$, $7,5 \cdot 10^{12}$ et 10^{13} cm^{-2} . Chaque point est obtenu par interpolation linéaire sur les courbes précédentes en faisant varier L à N_{inv} fixé.....	87
Figure 74 : ajustement des paramètres $KR_{bal} + R_0$ et μ_{long} sur la mobilité en fonction de la longueur de canal. Exemple à $N_{inv} = 5 \cdot 10^{12} \text{ cm}^{-2}$	88
Figure 75 : mobilité canal long. Ligne noire : extraction d'après des simulations à différentes longueur de canal, en ajustant les paramètres de la formule donnée en insert. En rouge, l'écart-type. Ligne bleue et carrés : mobilité extraite de simulations à champ constant.....	88
Figure 76 : facteur de mobilité $KR_{bal} + R_0$ de résistance balistique et d'accès, en fonction de la charge d'inversion. Ligne continue : extraction sur les simulations Monte-Carlo des paramètres en insert. Ligne pointillée : modèle analytique.....	89
Figure 77 : extraction de la résistance d'accès par différence entre les simulations à canal balistique (source et drain diffusifs) et la résistance balistique analytique. R_0 s'approche de $35 \Omega \cdot \mu\text{m}$ à fort N_{inv}	90
Figure 78 : résistance balistique issue du présent modèle en fonction de la charge d'inversion (ligne continue), comparée à la résistance balistique du modèle de Shur (tirets).....	91
Figure 79 : énergie des trois premières sous-bandes des vallées Δy (y étant la direction de confinement). $L = 100 \text{ nm}$, $V_g = 0,2 \text{ V}$ et $V_{ds} = 50 \text{ mV}$. Les bandes d'énergie sont plates dans les simulations à canal balistique en l'absence d'interactions permettant l'équilibre avec la tension de drain imposée.....	92
Figure 80 : comparaison des courants des simulations diffusives et balistiques. Ordonnée à droite, échelle linéaire. À gauche, échelle logarithmique. $L = 30 \text{ nm}$	92
Figure 81 : mobilité effective extraite des simulations à canal balistique (sans interaction dans le canal). Ligne continue : simulation Monte-Carlo. Pointillés : modèle analytique.....	92
Figure 82 : énergie de Fermi le long du canal. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$. Les pics correspondent aux jonctions considérées comme abruptes entre canal et zones dopées de source et drain.....	93
Figure 83 : résistance cumulée depuis la source. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$	94
Figure 84 : résistance cumulée depuis la source pour différentes longueurs de canal. $L = 10, 20, 40, 60$ et 80 nm , $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$	94
Figure 85 : résistance linéique. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$	95
Figure 86 : énergie de Fermi pour un canal balistique. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$	95
Figure 87 : résistance linéique avec un canal balistique. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$	96
Figure 88 : résistance de source cumulée pour différentes valeurs de tensions. $L = 30 \text{ nm}$, $V_{gs} = 0,2, 0,4, 0,6$ et $0,8 \text{ V}$. $V_{ds} = 50 \text{ mV}$	96
Figure 89 : géométrie des dispositifs étudiés et polarisations appliquées pour l'extraction de la mobilité. (a) Tension de seuil en régime linéaire et saturé. (b) Courant I_{OFF} à l'état bloqué en fonction du courant I_{ON} à l'état passant.....	97
Figure 90 : extraction de la mobilité à champ faible μ_0 par la méthode de la fonction Y	98
Figure 91 : extraction de la résistance d'accès R_0 par la pente de la courbe $\theta_1 \beta_0$	98
Figure 92 : faisceau de courbes de capacités totales $C_{gs} V_g$ à différentes longueurs de grille ($L = 240, 160, 80, 40$ et 20 nm).....	99

Table des figures

Figure 93 : mobilité extraite par split-CV en fonction de la charge d'inversion pour différentes longueurs de grille ($L = 240, 160, 80$ et 40 nm).....	99
Figure 94 : mobilité extraite expérimentalement et ajustement des paramètres $\mu_{bal} = K_{bal} \cdot L$ et μ_{long} dans le cas particulier $N_{inv} = 0,96 \cdot 10^{12} \text{ cm}^{-2}$	100
Figure 95 : mobilité extraite expérimentalement dans les cas particuliers $N_{inv} = 2,08 \cdot 10^{13} \text{ cm}^{-2}$ et $N_{inv} = 4,4 \cdot 10^{13} \text{ cm}^{-2}$	100
Figure 96 : facteur de mobilité balistique $KR_{bal} + R_0$ de résistance balistique et d'accès extrait des résultats expérimentaux, en fonction de N_{inv}	100
Figure 97 : résistance d'accès extraite des résultats expérimentaux par notre modèle, comparée à la résistance d'accès extraite par fonction Y (indépendante de N_{inv}).	101
Figure 98 : composante de mobilité canal long μ_{long} extraite des dispositifs expérimentaux, en fonction de N_{inv}	101

Introduction générale

Pour améliorer les performances des circuits intégrés et de leurs applications tout en maintenant des coûts de fabrication raisonnables, il est nécessaire de réduire les dimensions des composants. Depuis des décennies, cette réduction d'échelle s'est déroulée à un rythme énoncé dès les débuts de la microélectronique par la loi de Moore, qui s'est avérée prédictive.

La feuille de route rédigée par l'ITRS, *International Technology Roadmap for Semiconductors*, accompagne les professionnels de la microélectronique en déterminant chaque année les obstacles à franchir pour les quinze années suivantes et encourage les solutions innovantes, se voulant ainsi un document de référence. Composée d'organisations issues des principales régions productrices de transistors dans le monde, les États-Unis, l'Asie (le Japon, Taiwan et la Corée du Sud), et l'Europe, cette association a été créée en 1999 par des professionnels de la microélectronique réunis pour faire face à une nécessité grandissante d'investissements.

Après une décennie où l'électronique a été marquée par le développement du Web social, elle évolue actuellement vers l'*Internet of Things*, l'extension d'internet à des dispositifs de plus en plus interactifs. Cet Internet des objets dont font partie l'e-santé (ou télésanté) et la domotique requiert également l'émergence des *big data*, i. e. de la gestion d'une grande quantité de données.

Pour augmenter les performances et réduire les coûts de production des transistors CMOS, l'industrie de la micro-électronique cherche depuis ses débuts à réduire l'échelle de ces composants. Ainsi, conformément à une formulation de la loi de Moore, empirique, le nombre de transistors dans les microprocesseurs Intel a doublé tous les deux ans depuis le premier microprocesseur Intel 4004 inventé en 1971.

Malheureusement, la réduction d'échelle des composants se heurte à plusieurs mécanismes qui obligent à utiliser de nouvelles technologies : fuite de grille, effet tunnel, perte de contrôle de la grille sur la charge dans le canal due à la tension de drain (effet de DIBL ou *drain-induced barrier lowering*), confinement quantique, etc. Suivre la loi de Moore devient de plus en plus complexe pour les nouveaux composants.

L'un de ces effets néfastes est la réduction forte de mobilité des porteurs de charge dans le canal du transistor quand sa longueur décroît en deçà de 100 nm. Si cette tendance a été observée expérimentalement sur de nombreux types de composants depuis une dizaine d'années, son interprétation a suscité beaucoup d'interrogations si ce n'est de controverses. Il s'agit d'un paramètre clé pour les performances des nouveaux transistors, ainsi qu'un élément de base pour leur simulation au niveau des logiciels d'aide à la conception technologique des dispositifs à semiconducteurs (*technology computer-aided design* ou TCAD) ou des circuits. Dans cette thèse, le code MONACO de simulation Monte-Carlo de l'Institut d'Électronique Fondamentale (IEF) à l'Université Paris-Sud (Orsay) est utilisé pour décrire précisément le transport dans le canal des transistors planaires à film de silicium complètement déserté sur isolant (*fully depleted silicon on insulator* ou FDSOI) et pour modéliser la dégradation de mobilité extraite lorsque la longueur du canal diminue. La notion de mobilité est complexe et sera longuement discutée.

Introduction générale

Les principaux mécanismes de dégradation du transport des électrons et trous dans le transistor étudiés sont leurs interactions avec des défauts de la structure du composant. L'arrivée au maximum de performance correspond au régime balistique, régime idéal dans lequel aucune collision n'est subie par les porteurs de charge dans le canal. La caractérisation électrique permet d'explorer le fonctionnement des dispositifs aux échelles ultracourtes, ouvrant la possibilité de vérifier si on tend ou non vers ce régime ultime de transport. Ses résultats doivent toutefois être interprétés avec précaution.

Dans le premier chapitre, les principes du transistor seront rappelés, avec leurs modes de fonctionnement. Les différentes architectures seront définies et les techniques de fabrication seront évoquées, en insistant sur le cas du transistor planaire FDSOI qui fait l'objet de ce manuscrit. Puis les effets concernant la limitation de la performance, à savoir la dégradation de mobilité, les interactions et le transport balistique, étudiés en détail par la suite, seront introduits.

Dans le deuxième chapitre, les méthodologies utilisées pour modéliser le transport seront expliquées et comparées. Pour cela, l'équation de transport de Boltzmann (ETB) sera détaillée, ainsi que les interactions prises en compte, à savoir les phonons, les impuretés ionisées et la rugosité. Ensuite, la méthode de simulation TCAD, permettant d'obtenir des résultats relativement rapidement, et surtout la méthode Monte-Carlo, plus précise, seront décrites. Enfin la modélisation Kubo-Greenwood de la mobilité sera présentée.

Dans le troisième chapitre, les méthodes d'extraction de paramètres de performance que sont la mobilité, mais aussi la tension de seuil, les résistances d'accès, etc. seront présentées et comparées. Les problématiques théoriques et pratiques d'extraction par la méthode de Hamer et la fonction Y , d'évaluation de la charge par la méthode dite de split-CV, seront illustrées d'exemples.

Enfin, la méthodologie choisie sera utilisée dans le quatrième chapitre d'analyse des résultats de simulations Monte-Carlo obtenus. D'autres grandeurs plus directes telles que la charge dans le canal, l'énergie et la vitesse des porteurs, seront d'abord analysées. Un modèle de comportement de la mobilité à canal court, prenant en compte un terme de résistance « balistique », sera décrit en détails. Puis cette méthodologie d'extraction de mobilité sera appliquée à des données expérimentales.

Chapitre I. Vers les transistors ultimes et leurs limites

1. Introduction au transistor

Après le tube électronique (ou tube à vide), inventé en 1904 par le Britannique John Ambrose Fleming, la découverte de l'effet transistor (abréviation de *transfer resistor*¹) peut être attribuée à Julius Edgar Lilienfeld, physicien austro-hongrois qui a breveté cette invention en 1925, avant que le groupe de William Bradford Shockley n'en présente une version opérationnelle en 1947.

Ce dispositif, plus fiable que son prédécesseur qui n'est plus guère utilisé que dans certaines applications de forte puissance telles que le four à micro-ondes, a deux modes de fonctionnements : celui d'interrupteur et celui d'amplificateur.

1.1. Description du transistor MOSFET

Dans le cas d'un nMOSFET (transistor à effet de champ à canal N basé sur une capacité métal-oxyde-semiconducteur ou *metal oxide semiconductor field effect transistor*) réalisé sur substrat massif (ou *bulk*) tel que schématisé sur la figure 1, le courant d'intérêt traverse un canal dopé P, situé entre les zones d'accès constituées de la source S et du drain D dopés N, et est contrôlé par la tension appliquée sur la grille.

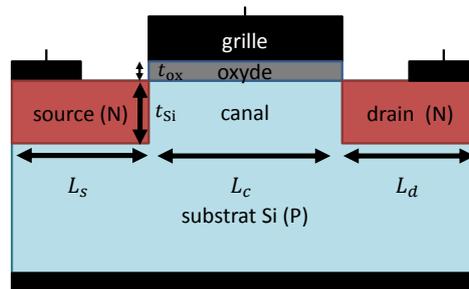


Figure 1 : schéma simplifié d'un transistor à canal N (bulk)

Plusieurs paramètres contribuent à la performance d'un transistor : parmi ceux-ci figurent la longueur physique du canal, sa longueur électrique (déduite de la précédente en excluant les queues de distribution de dopants des accès), la vitesse d'injection des porteurs dans le canal, la vitesse de saturation des porteurs, la mobilité. Ils sont déterminants pour la valeur du courant à l'état passant I_{ON} qui au niveau des circuits numériques conditionne la fréquence d'horloge. Ces paramètres sont liés entre eux par la physique du transport, régie par l'équation de Boltzmann qui sera détaillée plus loin, en incluant les effets quantiques de structure de bandes, de confinement au centre du canal et le régime de transport balistique (idéal).

D'autres paramètres dégradent les performances d'un MOSFET : les capacités ou résistances parasites. Ces effets sont d'autant plus importants que la longueur de canal est faible, car alors les diverses zones du transistor sont proches et peuvent interagir, induisant des charges ou des courants indésirables. Seul l'effet des résistances d'accès sera réellement étudié dans le présent manuscrit.

¹ Selon la Fondation Nobel ayant récompensé Shockley, Bardeen et Brattain en 1956. À moins que ce

Ces paramètres sont étroitement liés à l'architecture et à la géométrie du transistor ainsi qu'à la physique du matériau. La présence de défauts, les interactions avec les phonons, la rugosité de l'interface avec l'oxyde de grille, les interactions coulombiennes éventuellement à distance sont des facteurs bien connus de dégradation du transport de charges dans le canal.

1.1.1. Contrôle de charge par la grille

Pour générer un courant dans l'axe source-canal-drain, il faut disposer de charges dans le canal en appliquant un champ électrique dans l'axe perpendiculaire grille-oxyde-canal.

La charge dans le canal découle directement de la modification des extrema de bandes d'énergie par les polarisations. Selon la valeur de la tension de grille appliquée, quatre régimes de fonctionnement de la capacité MOS peuvent être atteints. Dans un nMOSFET, plus cette tension augmente, plus les électrons peuplent la bande de conduction dans le canal et pourront être accélérés de la source jusqu'au drain par la tension appliquée entre ses bornes. Ils sont en premier lieu déterminés par rapport à la tension $V_g = V_{FB}$ dite de bandes plates (ou « *flat band* ») c'est-à-dire telle que les extrema de bandes du Si sont constants dans la direction perpendiculaire à la capacité MOS jusqu'à l'interface avec l'oxyde.

1.1.1.i. Régime d'accumulation : $V_g < V_{FB}$

Dans ce régime représenté sur la figure 2, le transistor est bloqué car la tension de grille faible ou négative attire des trous qui s'accumulent dans la zone de canal, principalement à la surface de contact avec l'oxyde, et empêchent le passage des électrons de la source vers le drain même lorsque la tension de drain V_{ds} est positive.

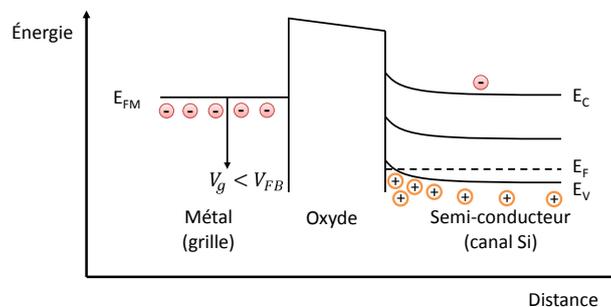


Figure 2 : diagramme de bandes d'énergie dans la direction normale à l'interface oxyde/semiconducteur en régime d'accumulation

1.1.1.ii. Régime de bandes plates : $V_g = V_{FB}$

Dans le régime de bandes plates (figure 3), la tension de grille annule la courbure de bandes du semiconducteur due à la différence de travail de sortie avec la grille et d'affinité électronique. La valeur V_{FB} de V_g dépend de ces deux derniers paramètres ainsi que du gap et du dopage du semiconducteur.

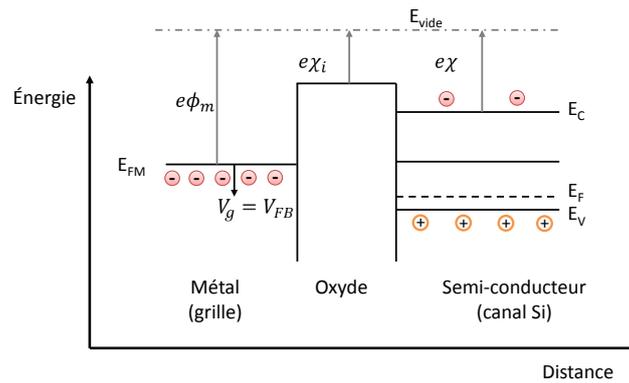


Figure 3 : tension de bandes plates²

1.1.1.iii. Régime de désertion : $V_{FB} < V_g < V_t$

Dans ce régime, il n'y a ni accumulation de trous dans la bande de valence ni d'électrons dans la bande de conduction, du moins en grande quantité, cf. figure 4. La zone du canal est désertée par les charges mobiles, d'où le terme de zone de déplétion ou de désertion.

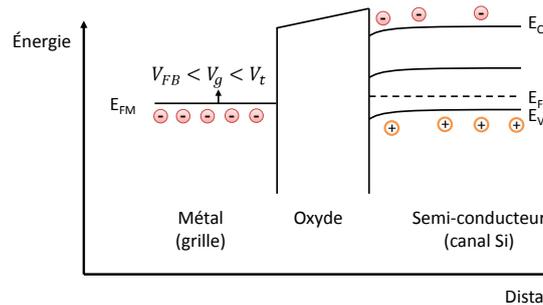


Figure 4 : inversion faible (désertion ou déplétion)

À la tension de seuil V_t (*threshold voltage*), la courbure de la bande de conduction à l'interface oxyde/Si est telle que le niveau de Fermi intrinsèque (environ le milieu de la bande interdite) passe clairement en dessous du niveau de Fermi, avec un écart identique à celui existant dans l'autre sens dans le fond du substrat. À partir de cette tension le canal d'inversion est considéré comme créé, puisque la densité d'électrons dans cette zone quand V_g croît devient plus grande que celle en trous dans le bulk.

1.1.1.iv. Régime d'inversion : $V_g > V_t$

En régime d'inversion, les électrons sont nombreux à la surface du canal à l'interface avec l'oxyde (cf. figure 5) et forment un canal N de conduction entre la source et le drain (dopés N).

² cf. <http://easytp.cnam.fr/algani/>

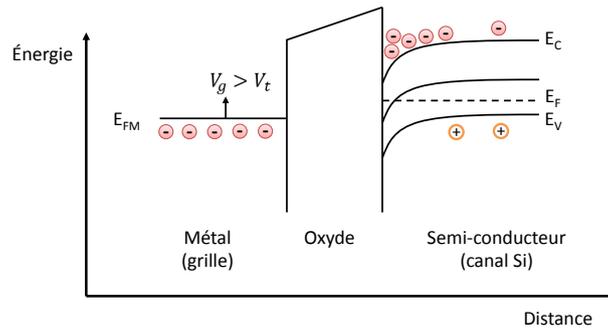


Figure 5 : inversion forte

Ce canal de conduction peut être mis à profit en imposant également une tension entre le drain du transistor et la source. L'effet de champ de la capacité MOS permet alors de moduler le passage de courant du transistor, d'où l'appellation de MOSFET (*field-effect transistor*). Le courant de drain fonctionne en régime linéaire par rapport à la tension V_{ds} jusqu'au pincement du canal près du drain à une tension au-delà de laquelle le régime saturé est atteint.

1.1.2. Autres types de FET

Il existe une multitude d'autres types de FET dont il ne sera plus question par la suite. On peut notamment citer les catégories suivantes :

- JFET (*junction gate field-effect transistor*) : architecture la plus simple utilisant une zone dopée sous la grille, créant une jonction p-n avec le substrat, au lieu d'un oxyde comme pour le MOSFET.
- MESFET (*metal semiconductor field-effect transistor*) : similaire au JFET, il présente une jonction Schottky à la place de la jonction p-n.
- HEMT (*high electron mobility transistor*) ou HFET (*heterostructure FET*) : il s'agit d'une évolution des MESFET dans laquelle une couche supérieure à grand gap (AlGaAs par exemple, ou InAlAs) dopée N est entièrement désertée par la grille Schottky. Les porteurs s'accumulent alors dans un canal formé dans une zone de petit gap (GaAs, InGaAs...) intentionnellement non dopée, située juste en dessous et formant une hétérojonction avec la précédente. Les électrons sont confinés à l'interface dans un puits quantique quasi triangulaire et la tension de grille module leur densité. L'éloignement des impuretés ionisées et la qualité de l'hétérointerface entre les deux semiconducteurs favorisent la mobilité électronique.
- TFT (*thin-film transistor*) : c'est un transistor dans lequel le canal est une couche mince déposée et non plus une partie du substrat, qui peut être en verre comme dans le cas des écrans à cristaux liquides LCD.
 - OFET (*organic field-effect transistor*) ou OTFT : il désigne un TFT dont le substrat est cette fois-ci organique tel le PMMA³ (polyméthacrylate de méthyle), ce qui peut lui conférer des propriétés de flexibilité (et donc de facilité d'impression), de transparence, et de coût, au détriment d'une faible performance.

³ plus connu sous le nom commercial de Plexiglas®.

- GFET (*graphene field-effect transistor*) : ce transistor vise à exploiter la forte mobilité du graphène. Malgré certains effets d'annonce comme la démonstration par IBM en 2011 d'un mélangeur radiofréquence intégré sur graphène [LIN11], cette technologie reste en l'état éloignée de la production industrielle (difficulté de production à grande échelle, ingénierie nécessaire pour rendre le graphène semiconducteur).
- TFET (*tunnel field-effect transistor*) : l'effet tunnel entre la bande de valence dans la source et la bande de conduction dans le canal est modulé par les tensions V_{gs} et V_{ds} imposées. Les dopages des accès sont de types opposés (jonction p-i-n) contrairement aux transistors usuels. L'effet tunnel a pour intérêt de présenter une transition nette entre le régime passant et le régime bloqué. La pente sous le seuil est améliorée. Beaucoup d'efforts sont encore à porter sur cette architecture pour lever les obstacles techniques et atteindre les performances théoriques escomptées.
- SET (*single electron transistor*) : le transistor dit « à un électron » utilise l'effet tunnel entre la source et le canal qui est réduit à un îlot de dimension nanométrique, puis entre le canal et le drain. Les accès ont cette fois le même dopage (jonction n-i-n par exemple). L'état OFF est permis par le blocage de Coulomb qui consiste en la répulsion d'un électron de la source par un électron de même énergie dans l'îlot. La faiblesse de sa taille permet en effet d'atteindre une capacité de chargement depuis la source suffisamment petite pour que l'énergie associée au stockage d'un électron soit bien plus grande que celle d'agitation thermique.

2. Caractéristiques courant-tension

Dans un circuit, les caractéristiques du transistor sont mesurées sur les courbes de courant de drain I_d en fonction des tensions entre ses diverses bornes.

2.1. Les différents régimes

Le premier faisceau de courbes, $I_d(V_{ds})$ à différents V_{gs} , permet de distinguer deux régimes du transistor à l'état passant : linéaire et saturé.

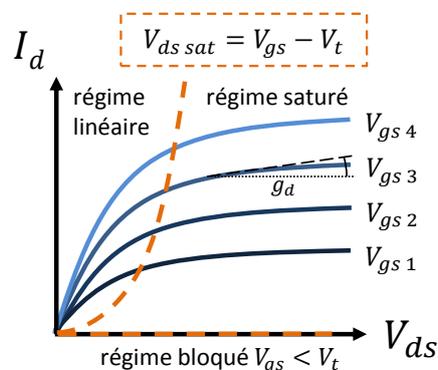


Figure 6 : courant de drain en fonction de la tension de drain appliquée pour un MOSFET à canal long, à différentes tensions de grille. Identification des régimes de fonctionnement.

En effet, sur chacune des courbes de la figure 6, à faible V_{ds} , dans une première région le courant I_d augmente presque linéairement avec la tension V_{ds} . C'est le régime linéaire ou ohmique.

À fort V_{ds} , le courant I_d atteint un plateau. En réalité, le courant n'atteint pas un maximum. Au contraire, il présente une légère pente dont la valeur correspond à une conductance de drain (dynamique) $g_d = \frac{\partial I_d}{\partial V_{ds}}$ (à V_{gs} donné) indésirable. C'est pourquoi les courbes ne sont pas exactement compatibles avec l'équation (4) du modèle développé dans la sous-section suivante.

Sur la figure 7 sont représentées des courbes $I_d(V_{gs})$ à différents V_{ds} . Elles permettent de visualiser d'une autre manière l'apparition de la saturation du courant à fort V_{ds} qui se manifeste par le rapprochement des courbes pour i croissant. Tout comme pour le faisceau de courbes précédent, il faut donc se rappeler que c'est à fort V_{gs} que se situe le régime linéaire.

C'est dans ce régime qu'est définie la transconductance $g_m = \frac{\partial I_d}{\partial V_{gs}}$ (à V_{ds} donné) qui est une caractéristique principale des MOSFET⁴.

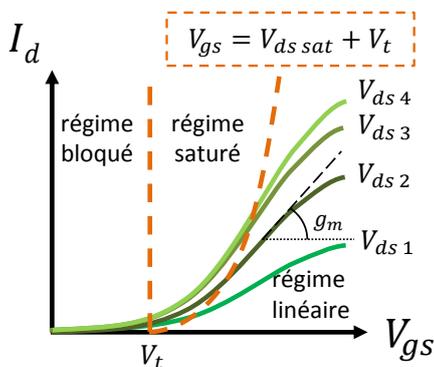


Figure 7 : courant de drain en fonction de la tension de grille appliquée pour un MOSFET à canal long, à différentes tensions de drain. Identification des régimes de fonctionnement.

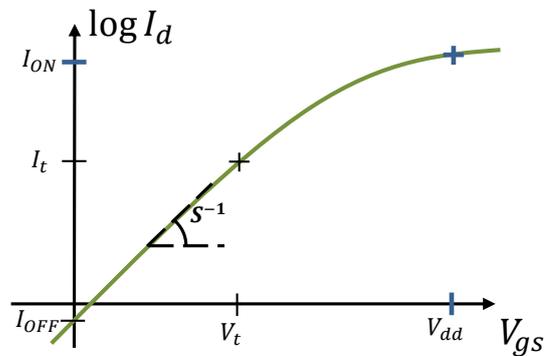


Figure 8 : courant de drain à l'échelle logarithmique en fonction de la tension de grille pour une tension V_{ds} donnée. Représentation de la pente sous le seuil et des courants I_{ON} et I_{OFF} .

Ce comportement du courant peut être modélisé simplement par quelques équations qui permettent de visualiser les principaux paramètres en jeu dans le transport, et d'introduire des effets indésirables plus complexes à décrire et à prédire, les effets de canal court.

2.2. Un modèle simple

Pour décrire simplement le fonctionnement du transistor dans tous ses régimes de fonctionnement, les équations de conservation de la charge sont simplifiées selon plusieurs hypothèses.

En régime sous le seuil $V_{gs} < V_t$, le transistor conduit peu de courant :

$$I_d = I_{d0} \exp\left(\frac{V_{gs} - V_t}{S} \ln 10\right) \left[1 - \exp\left(-\frac{qV_{ds}}{k_B T}\right)\right] \quad (1)$$

⁴ Dans g_m , l'indice m évoque la contribution « mutuelle » entre le courant de drain et la tension de grille.

Dans cette équation, le terme de pente sous le seuil S (*subthreshold slope*) est l'inverse de la pente de la courbe $I_d(V_{gs})$ tracée sur une échelle semi-logarithmique, voir figure 8. C'est donc un abus de langage et c'est pourquoi il est également appelé *subthreshold swing* (littéralement « balancement » sous le seuil).

Elle peut être estimée par l'équation suivante [SAINT-MA05] :

$$S = \frac{k_B T}{q} \ln(10) \left(1 + \frac{C_{ZCE}}{C_{ox}} \right) \quad (2)$$

Courant en régime linéaire $V_{ds} < V_{gs} - V_t$:

$$I_d = \frac{W}{L_{eff}} \mu_{eff} C_{ox} \left(V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} \quad (3)$$

Courant en régime saturé pour un transistor long $V_{ds} > V_{gs} - V_t$:

$$I_d = \frac{W}{L_{eff}} \mu_{eff} C_{ox} \frac{(V_{gs} - V_t)^2}{2m} \quad (4)$$

où m est le coefficient de l'effet de body lié à la concentration de dopant et à l'épaisseur d'oxyde.

Quel que soit le régime de fonctionnement, le courant de drain est fortement lié au facteur de gain $\beta = \frac{W}{L_{eff}} \mu_{eff} C_{ox}$. Autrement dit, pour améliorer le courant de sortie du transistor, il faut augmenter la largeur W , la mobilité effective μ_{eff} , la capacité de l'oxyde de grille C_{ox} (qui est en réalité une capacité surfacique), ou diminuer la longueur de canal effective L_{eff} . Malheureusement, certains de ces effets sont contradictoires (en diminuant L_{eff} , la mobilité μ_{eff} diminue également) et il est nécessaire de trouver un compromis. De plus, certains effets parasites apparaissent, dont la perte de contrôle du canal par la grille.

2.3. Perte de contrôle électrostatique

Les équations précédentes ne sauraient rendre compte parfaitement du courant mesuré sur un transistor ni même déduit d'une simulation de transport, sans y ajouter au moins les effets de canal court qui induisent un décalage de ces courbes cf. figure 10.

2.3.1. Effet de canal court

Le premier effet de canal court (SCE ou *short channel effect*) généralement cité est dû à la présence des zones de charges d'espace (ZCE) aux jonctions entre canal et accès, source ou drain. Les champs régnant dans ces zones bloquent *a priori* la diffusion des porteurs. Mais si elles sont trop proches, les effets des champs s'opposent, ils peuvent s'annihiler ce qui induit un courant de fuite échappant au contrôle de la grille. Cet effet de canal court peut se traduire par un abaissement de la tension de seuil, ou V_t *roll-off* [YU97].

En d'autres termes, la barrière de potentiel que la tension de grille est censée contrôler est plus faible qu'à plus grande longueur de grille, comme dans l'exemple représenté en figure 9.

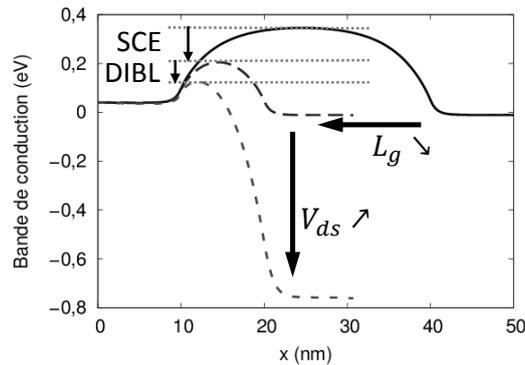


Figure 9 : effet canal court SCE et de DIBL sur la barrière de potentiel. Simulations Monte Carlo d'un transistor FDSOI à canal d'épaisseur $t_{Si} = 4$ nm et à $V_g = 0,2$ V. $L_g = 30$ nm et $V_{ds} = 50$ mV en ligne continue, $L_g = 10$ nm et $V_{ds} = 50$ mV en tirets longs, $L_g = 10$ nm et $V_{ds} = 0,8$ V en tirets courts.

2.3.2. DIBL et DICE

L'augmentation de la tension de drain peut se combiner à l'effet de canal court, précédemment évoqué et contribue d'autant plus à l'abaissement de barrière. C'est le DIBL (pour *drain induced barrier lowering*), soit la diminution du maximum de la bande de conduction lorsque V_{ds} augmente, comme dans l'exemple représenté en figure 9.

Au-dessus du seuil, cet abaissement de barrière se cumule à l'effet de la tension de grille et est nommé DICE (pour *drain induced conductivity enhancement*), accroissant ainsi la conductivité de drain g_d . La mauvaise saturation du transistor qui en découle est préjudiciable aux marges de bruit de circuits intégrés numériques à base de MOSFET (technologie CMOS pour *Complementary MOS*).

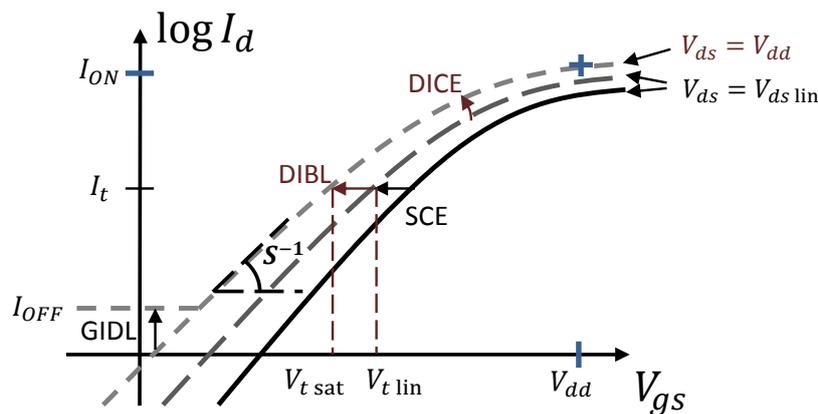


Figure 10 : effets de canal court et décalages de courbe associés. $V_{t \text{ lin}} = V_{gs}$ à $I_{ds} = I_t$ (courant de seuil de l'ordre de $1 \mu\text{A}/\mu\text{m}$) et $V_{ds} = V_{ds \text{ lin}}$, $V_{t \text{ sat}} = V_{gs}$ à $I_{ds} = I_t$ et $V_{ds} = V_{ds \text{ lin}}$, $I_{ON} = I_{ds}$ à $V_{gs} = V_{ds} = V_{dd}$, $I_{OFF} = I_{ds}$ à $V_{gs} = 0$ V et $V_{ds} = V_{dd}$, $S = \partial V_{ds} / \partial (\log I_d)$ pente sous le seuil, $\text{DIBL} = (V_{t \text{ lin}} - V_{t \text{ sat}}) / (V_{dd} - V_{ds \text{ lin}})$ [ESSENI11]

Afin de lutter contre la perte de contrôle électrostatique sur la charge dans le canal, des lois d'échelle ont été proposées comme la réduction de l'épaisseur d'oxyde (augmentation de la capacité de grille) ou l'augmentation de la concentration en dopants dans le canal (diminution des tailles des ZCE aux jonctions) dans le même rythme que la loi de Moore.

Comme nous allons le voir, ces solutions ont toutefois atteint leurs limites depuis quelques années, notamment du fait d'effets de variabilité d'ordre « atomistique » (épaisseur de la couche d'oxyde, nombre de dopants dans le canal) qui rendent indispensables des modifications dans la nature des matériaux employés ou la topologie du transistor.

3. Évolution des MOSFET

Pour favoriser un ou plusieurs des paramètres évoqués précédemment, il est nécessaire de modifier la géométrie du transistor par rapport à la géométrie classique *bulk*, dans laquelle les charges sont réparties en volume à l'intérieur du canal, en trouvant un compromis entre sophistication et simplicité de fabrication.

3.1. Lois de changement d'échelle

Lorsque les dimensions sont réduites d'un facteur de réduction d'échelle $\alpha > 1$, plusieurs autres grandeurs diminuent selon des facteurs résumés sur le tableau 1 [ESSENI11], [WONG99]. Ces lois sont les règles de *scaling* :

Paramètre	Valeur	Notation générale	Champ constant	Tension constante	Scénario mixte
dimensionnement du transistor	dimensions	L_g, t_{Si}, t_{ox}, W	$1/\alpha$	$1/\alpha$	$1/\alpha$
	concentration de dopants	n	α	α^2	α^2/λ
caractéristiques du transistor	tension	V	$1/\alpha$	1	$1/\lambda$
	champ	E	1	α	α/λ
	courant	I	$1/\alpha$	α	α/λ^2
	capacité	C	$1/\alpha$	$1/\alpha$	$1/\alpha$
caractéristique du circuit	résistance d'interconnexion	R	α	α	α
	délai ($C_{ox}V_{dd}/I_{ON}$)	τ	$1/\alpha$	$1/\alpha^2$	λ/α^2
	délai d'interconnexion	τ_i	1	1	1
	puissance dissipée ($V_{ds}I_d$)	P	$1/\alpha^2$	α	α/λ^3
	énergie dissipée par cycle	$P\tau$	$1/\alpha^3$	$1/\alpha$	$1/\lambda^2\alpha$
	densité de puissance	$P/(t_{Si}W)$	1	α^3	α^3/λ^3

Tableau 1 : lois de changement d'échelle. α : facteur géométrique de réduction d'échelle, λ : facteur de tension de réduction d'échelle ($1 < \lambda < \alpha$).

C_{eff} et I_{eff} sont respectivement la capacité et le courant effectifs du transistor durant un cycle de fonctionnement en tant qu'interrupteur. Dans le cas d'un scénario mixte, la réduction des tensions est de l'ordre d'un facteur λ où $1 < \lambda < \alpha$.

Alors que la mobilité des électrons est censée être stable, celle-ci décroît comme nous le reverrons plus loin lors de la réduction de la longueur de canal.

3.2. Architectures ultimes de transistors

Les lois d'échelle précédentes ont fait le succès de l'industrie CMOS jusqu'en 2000. La réduction envisagée par la suite de la longueur de grille en dessous de la limite d'une trentaine de nm pose cependant des difficultés insurmontables si on continue à utiliser le couple Si/SiO₂ pour l'empilement de grille. D'après le *scaling* habituel, la couche de silice doit alors compter moins de trois couches atomiques en épaisseur ce qui est inenvisageable à l'échelle d'une production en grande masse. L'introduction – difficile – d'isolants à plus haute permittivité diélectrique comme l'oxyde d'hafnium à la fin des années 2000 a permis de relâcher la contrainte de dimensionnement de l'épaisseur d'oxyde de grille mais n'a fait que retarder l'échéance pour l'architecture *bulk* des MOSFET, également mise à mal par les effets importants de variabilité dus à la réduction importante du nombre de dopants dans le canal. Si l'ingénierie de contraintes a été bénéfique pour trouver d'autres sources de gains en performance à longueur de grille fixée, l'utilisation de structures alternatives de transistor est aujourd'hui plus que nécessaire.

Généralement, pour améliorer le contrôle électrostatique et ainsi ne pas avoir à employer des tensions d'utilisation élevées, la réduction des épaisseurs est souvent utilisée. Elle permet de favoriser le contrôle du courant par la grille en la situant à proximité du canal. Comme nous allons le voir dans le cas des différentes alternatives au *bulk* utilisées ou étudiées, il peut y avoir une ou plusieurs réductions de dimension.

3.2.1. Transistor planaire SOI

Une couche d'oxyde enterrée (ou BOX pour *buried oxide*) dans le substrat permet de réduire l'épaisseur du canal d'un transistor par rapport au cas du bulk et d'obtenir un transistor planaire SOI (*silicon on insulator*), le canal Si se retrouvant sur l'isolant qu'est le BOX. Cette architecture a pour avantage d'améliorer le contrôle électrostatique de grille, mais sa fabrication implique plus d'étapes technologiques. Les transistors SOI sont développés depuis longtemps et peuvent être divisés en plusieurs catégories :

- PDSOI (*partially depleted silicon on insulator*) : le canal a une épaisseur d'au moins plusieurs dizaines de nanomètres. L'épaisseur de BOX peut être de plusieurs centaines de nanomètres.
- FDSOI (*fully depleted*) : de même que pour le PDSOI, le canal est une couche mince, mais cette fois-ci d'une épaisseur de l'ordre de la dizaine de nanomètres à quelques nanomètres. La finesse du canal permet de mieux contrôler le potentiel dans cette zone non polarisée par une électrode. On lutte ainsi contre le *floating body effect*, responsable par exemple d'un décalage de la tension de seuil avec effet mémoire⁵, ou d'un changement de pente dans les caractéristiques de courant $I_d(V_{ds})$ (effet *kink* [JUNGEMAN12]).

Le BOX est aussi fin, de l'ordre de quelques dizaines de nanomètres, de façon à ce que le champ électrique induit dans l'oxyde enterré par V_{ds} reste perpendiculaire aux interfaces et ne produise pas d'effet DIBL par déviation des lignes de champ vers la

⁵ www.soiconsortium.org

source. C'est pourquoi le FDSOI peut avoir pour autre dénomination UTBB (*ultra-thin body and BOX*) SOI. Le terme d'ETSOI (*extremely thin silicon on insulator*) est aussi utilisé dans le cas des films les plus minces.

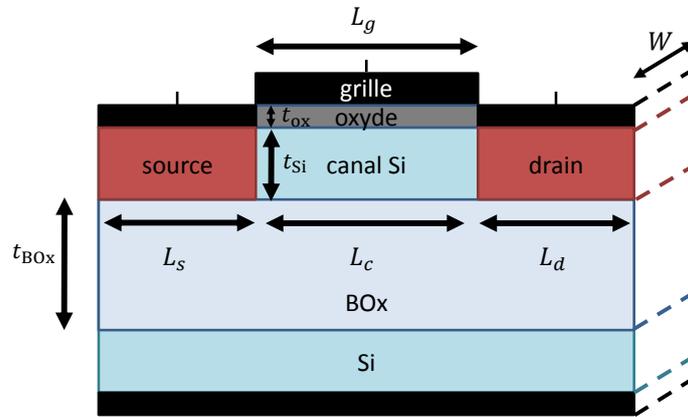


Figure 11 : schéma simplifié d'un transistor FDSOI et notations utilisées

Le schéma simplifié de la figure 11 présente la géométrie d'un transistor FDSOI dont le manuscrit fait l'étude. Les notations seront donc utilisées par la suite. La géométrie du transistor simulé en Monte-Carlo ne différera de celle-ci que par l'absence du substrat silicium sur la face arrière, opposée à la grille.

Il ne présente pas les éventuels halos et poches fortement (HDD) et légèrement (LDD) dopés parfois utilisés pour limiter les effets de canal court (décalage de V_t). Un exemple de vue en coupe effectuée par microscopie électronique à transmission (TEM) de FDSOI est présenté sur la figure 12.

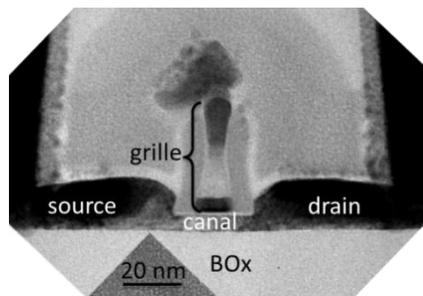


Figure 12 : coupe TEM d'un FDSOI [MORVAN12]

Les électrons sont confinés selon le plan perpendiculaire au canal et donc à l'axe du transport.

Les étapes de fabrication dans l'exemple d'un FDSOI avec $L_g = 32$ nm sont rappelées sur le tableau 2 [FENOUILLO8] :

Étape de fabrication	Description de l'étape
isolation STI (<i>shallow trench isolation</i>)	tranchées d'isolation peu profondes entre deux transistors
implantation de la face arrière optionnelle : P+ type for NMOS N+ type for PMOS	dopage du substrat (en porteurs majoritaires)
dépôt d'une couche de HfZrO ₂ (2,5 nm)	dépôt d'oxyde
dépôt d'ALD TiN (10 nm)	dépôt de métal
dépôt de polysilicium (80 nm)	formation de la grille
pré-dopage de la grille	dopage avant formation
dessin de la grille	gravure de la grille
formation de l'espaceur (offset)	
épitaxie sélective de Si (10 nm)	
implantation de LDD	dopage léger entre les accès et le canal
formation de l'espaceur (d-shape)	
implantation de source et drain	dopage
recuit haute température (RTA spike 1080°C)	activation de dopants lors de la recristallisation
siliciuration NiPtSi	
contrainte par CESL (<i>contact etch stop layer</i>)	couche d'arrêt de gravure
backend standard	création des contacts, réalisation des interconnexions entre transistors

Tableau 2 : étapes de fabrication

Pour créer une architecture SOI, une couche de silicium contraint sur isolant peut être fabriquée puis découpée en utilisant la technologie Smart Cut™ pour garder l'épaisseur de silicium nécessaire, comme illustré sur la figure 13.

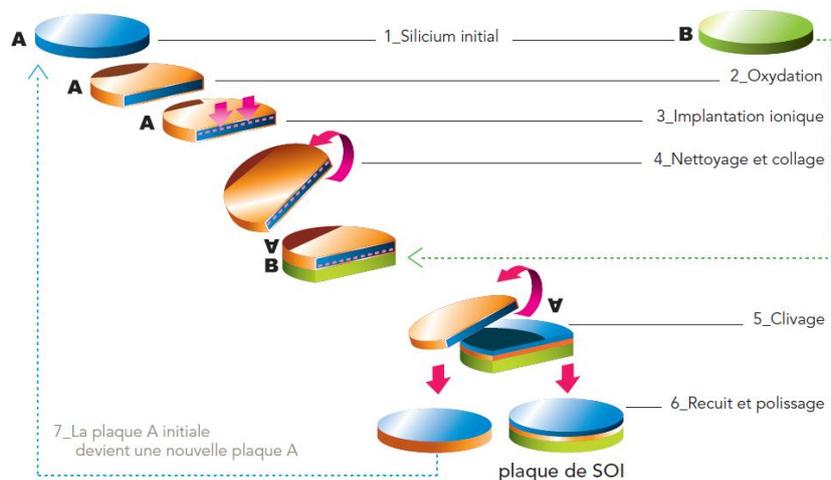


Figure 13 : fabrication de SOI par technologie Smart Cut™ de SOITEC⁶

À partir d'une plaque initiale de silicium, une oxydation est effectuée, suivie d'une implantation ionique qui se retrouve principalement en surface. La plaque obtenue est alors retournée après nettoyage puis collée sur une autre pour que la surface implantée soit en contact avec la plaque de silicium, pour générer une contrainte. Ensuite, le clivage permet de ne conserver que la partie nécessaire de silicium, puis le recuit et le polissage permettent le nettoyage de celle-ci.

⁶ www.soitec.com

Grâce au procédé Smart Cut™, SOITEC propose des substrats SOI avec des couches actives de très bonne qualité, potentiellement contraintes, à des épaisseurs de moins de 10 nm.

L'architecture FDSOI à canal extrêmement mince a été choisie par STMicroelectronics pour les nœuds technologiques en deçà de 28 nm. Comme décrit dans la synthèse annuelle de l'Observatoire des Micro et Nanotechnologies, cela a permis la démonstration en 2013 de processeurs ST Ericsson à double cœur permettant soit d'atteindre à 3 GHz d'horloge un fonctionnement deux fois plus rapide qu'une puce à quatre cœurs sur Si massif, soit un faible niveau de consommation à 1 GHz et sous 0,63 V de tension d'alimentation. En 2014, Samsung puis le fondeur taïwanais TSMC ont passé des accords avec STMicroelectronics pour le développement de cette technologie.

3.2.2. Transistor double grille

Afin d'améliorer encore le contrôle par la grille de la formation du canal, il est possible de prendre la couche active du transistor entre deux empilements oxyde/métal, cf. figure 14. Cette structure à double grille peut posséder un canal de largeur ajustable comme l'architecture bulk planeaire, avec une orientation dans le plan du substrat ou verticale. La réalisation de grilles alignées est néanmoins un problème complexe [VINET09] qui limite le développement de cette solution.

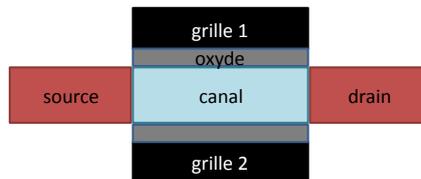


Figure 14 : transistor double grille.

3.2.3. Transistors FinFET et trigate

La fabrication d'architectures à grille multiple, c'est-à-dire avec plusieurs flancs du canal contrôlés par une grille, peut sembler plus facile dans le cas des FinFET, schématisés sur la figure 15. Il s'agit de graver le Si en latéral, sous forme de fil ou « d'aileron » (fin en anglais), puis de réaliser un empilement de grille sur 3 côtés. Comme sur l'exemple de la figure 16, l'oxyde peut être plus épais sur le dessus du fin que sur les flancs et le FinFET est alors un transistor à double-grille non planeaire. Si l'empilement de grille est identique sur les 3 côtés du canal, il s'agit d'un dispositif à triple grille ou *trigate*, tel que celui introduit par Intel en mai 2011 pour les technologies dites 22 nm, ou par d'autres fondeurs depuis.

L'introduction de ces FinFET non planeaires constitue néanmoins un changement fort au niveau de la conception des circuits. L'ensemble des règles de dessin doit être repensé. L'augmentation de la largeur du transistor passe par la mise en parallèle d'un plus grand nombre de fils, entre lesquels il n'y a pas de Si actif. La qualité de la gravure du Fin constitue en outre l'étape clé pour cette technologie. Ce dispositif reste en concurrence avec le FDSOI à canal ultra-mince pour les prochains nœuds technologiques.

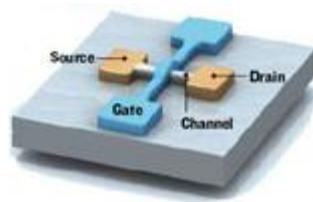


Figure 15 : schéma du FinFET⁷.

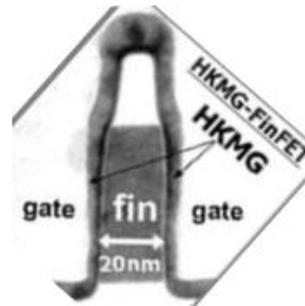


Figure 16 : coupe TEM d'un FinFET⁸.

3.2.4. Transistor à nanofils *gate all around*

La solution ultime pour poursuivre la loi de Moore semble être celle des transistors à nanofils complètement enrobés par une grille (*gate all around* ou GAA), comme représenté sur la figure 17.

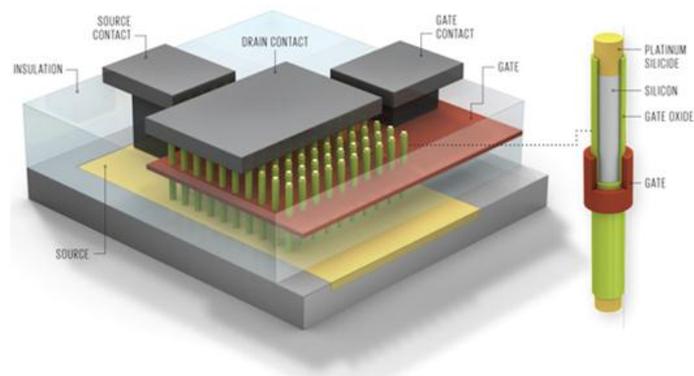


Figure 17 : nanofil *gate all around*⁹

Une coupe du nanofil *gate all around* est donnée sur la figure 18.

⁷ www.spectrum.ieee.org

⁸ World of Electronics, 2011.

⁹ Illustration : Emily Cooper

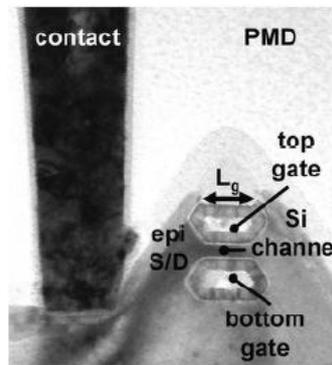


Figure 18 : coupe TEM d'un transistor à nanofil [BIDAL09]

Entre le trigate et le nanofil GAA se trouvent le Ω -gate [JAHAN05] ou le Π -gate [PARK01] qui sont des compromis entre ces deux types d'architectures. Grâce à une forme en Ω ou Π de la grille, le fond du canal y est contrôlé par effet de bord.

L'utilisation de ce type de dispositifs innovants au niveau circuit reste très prospective.

4. Réduction de mobilité à faible longueur de grille

La mobilité est une notion qui peut se déduire de la résolution approchée de l'équation de Boltzmann, ou d'approches plus simples décrivant l'état d'équilibre entre un porteur de charge mobile et les collisions qu'il subit dans son environnement cristallin. Comme illustré par le modèle analytique de base des MOSFET, elle est indispensable pour la simulation de transistors par une approche compacte au niveau circuit. La mesure de la mobilité est également largement employée pour qualifier une technologie. Sa définition devient néanmoins complexe dans le cas d'un canal dont la longueur devient du même ordre de grandeur que le libre parcours moyen des porteurs.

L'augmentation des performances des transistors MOSFET par la diminution de la longueur de canal se heurte à une dégradation de la mobilité, du moins apparente. Celle-ci est expliquée par plusieurs facteurs, liés à une limitation physique qui n'est pas visible de façon quantitative par les méthodes de caractérisation. Les principaux facteurs possibles de dégradation additionnelle sur la mobilité seront présentés ci-après.

4.1. Mise en évidence expérimentale

Comme illustré par les résultats de la figure 19, la mobilité μ_0 mesurée dans les dispositifs avancés ultracourts est réduite avec la diminution de la longueur de canal en dessous de $L_{eff} = 100$ nm environ. Bien que cette décroissance apparaisse en réalité déjà au-dessus de cette valeur, elle peut être décrite en première approximation par un plateau pour L_{eff} supérieur et par une décroissance rapide vers une valeur nulle lorsque L_{eff} tend vers zéro.

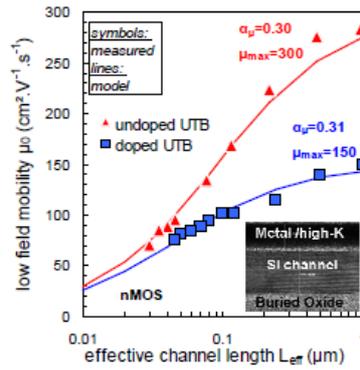


Figure 19 : courbes de décroissance de mobilité μ_0 avec L_{eff} dans des transistors MOSFET à canal d'électrons à canal ultrarmince [BIDAL09].

Cet effet a été observé quelle que soit la géométrie de transistor choisie, principalement à partir d'une longueur de canal L_{eff} inférieure à une centaine de nanomètres : dans les MOSFET *bulk* [CROS06], FDSOI [PHAM-NGU08], *gate all around* [CROS06], FinFET [RAMOS06].

De même, elle est constatée indépendamment des matériaux constituant le transistor :

- une grille en Si et un oxyde *high-κ* ou SiO_2 [PHAM-NGU08]. Ce constat est issu de l'étude d'un même MOSFET de type FDSOI par comparaison de la mobilité extraite sur la face avant présentant une interface *Si/high-κ* avec celle extraite sur la face arrière (Si/SiO_2). Même si les mobilités sont différentes, elles diminuent entre 1 μm et 40 nm. Par ailleurs cette diminution est également observée à faible température.
- une grille en métal et un oxyde *high-κ* [BIDAL09]. Cet effet n'est donc pas lié à une fuite de courant dans la grille ou à une interaction spécifique à l'interface entre la grille et l'oxyde, telle la rugosité de surface ou des charges piégées.
- un NMOS ou un PMOS [ANDRIEU05, CROS06, WEBER10]. Cette diminution concerne donc les électrons et les trous puisqu'elle a été vérifiée sur des empilements de grilles identiques dans les deux cas, avec des matériaux variés.
- un canal contraint ou non [ANDRIEU05, RAMOS06]. L'amélioration de la mobilité par la contrainte n'empêche pas sa diminution inéluctable lorsque L_g décroît.
- un canal fortement dopé ou non [CROS06, PHAM-NGU08]. C'est ce que tend à montrer une étude de nMOSFET de type *bulk* dont les implantations des extensions de source et de drain ont été effectuées à différentes énergies. Le dopage du canal n'est pas responsable de cette décroissance de mobilité. À l'inverse, de même que pour le canal contraint, l'utilisation d'un canal non dopé n'est pas suffisante pour empêcher la diminution de la mobilité pour les transistors à canaux plus courts.

Enfin, elle apparaît quelle que soit la méthode d'extraction utilisée pour la mesurer [CASSÉ08] : la méthode par fonction Y [GHIBAUDO09], le split-CV [RAMOS06] ou la magnétorésistance [CHASANT06]. Il ne s'agit donc pas d'un artefact de mesure.

La figure 20 précise que cette dégradation de mobilité concerne même toute la courbe de mobilité effective μ_{eff} dépendant de la charge d'inversion dans le canal.

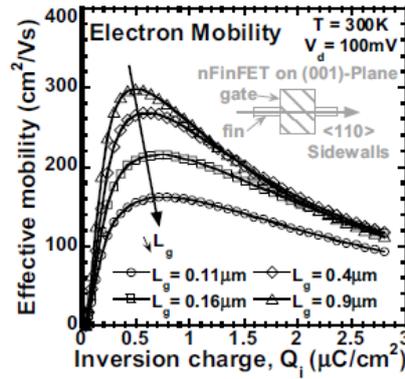


Figure 20 : dégradation de mobilité effective (i. e. dépendant de la charge en abscisse) avec la réduction de L_g [RAMOS06].

Plusieurs effets ont été proposés pour expliquer cette tendance. Les raisons principalement avancées sont décrites et résumées ci-après.

4.2. Modélisation des effets de dégradation

Le courant dans un MOSFET est limité par les interactions dans le canal entre les électrons et leur environnement, décrites globalement par un paramètre macroscopique : la mobilité μ qui représente la capacité des électrons à se mouvoir dans la couche d'inversion. Les autres facteurs limitatifs sont dépendants du type d'architecture ou du matériau utilisé comme les fuites de grille ou encore les fuites de jonction.

Usuellement la mobilité est en première approximation décrite par une loi de Matthiessen (à champ électrique source-drain faible). Cette loi suppose des mécanismes indépendants pour pouvoir sommer les probabilités d'occurrence des collisions avec les centres coulombiens (cb), la rugosité de surface (sr) et les phonons (ph) :

$$\frac{1}{\mu_{\text{long}}} = \frac{1}{\mu_{\text{cb}}} + \frac{1}{\mu_{\text{sr}}} + \frac{1}{\mu_{\text{ph}}} \quad (5)$$

Comme nous avons pu le démontrer à travers les nombreuses références bibliographiques du paragraphe précédent, expérimentalement, il a été démontré que cette mobilité chutait de façon inattendue avec la réduction de longueur du canal. Comme décrit schématiquement sur la figure 21, cette dégradation limite fortement les performances des transistors sub 100 nm. Il devient essentiel de comprendre et de modéliser cette chute de mobilité. Plusieurs explications sont fournies pour décrire cette forte diminution, dont la présence de défauts neutres et le régime quasi-balistique. La mobilité à champ faible est alors décrite de la façon suivante :

$$\frac{1}{\mu_{\text{eff}}} = \frac{1}{\mu_{\text{long}}} + \frac{1}{\mu_{\text{add}}} \quad (6)$$

Le problème réside dans la description physique de ce dernier terme : μ_{add} . D'autres hypothèses ont été explorées et peuvent aussi expliquer cette réduction de mobilité comme la théorie des plasmons, une mauvaise maîtrise des profils de diffusion, les contraintes mécaniques

dues au procédé de fabrication ou encore l'erreur induite par les méthodes d'extraction de cette mobilité.

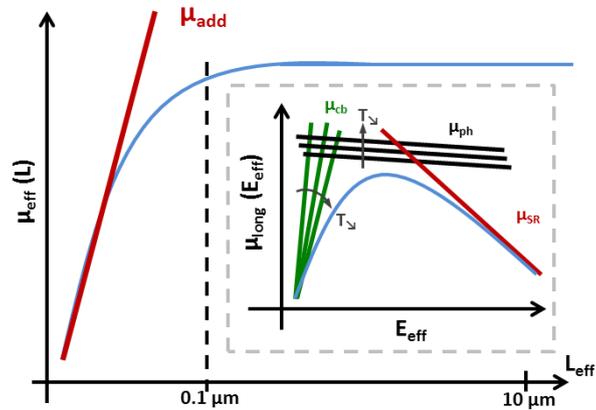


Figure 21 : évolution typique de la mobilité en fonction de L_{eff} . Insert : représentation schématique des dépendances aux mécanismes d'interaction en fonction du champ effectif E_{eff} .

4.3. Mécanisme d'interaction avec des défauts neutres

4.4. Influence des défauts

Des défauts spécifiques à l'interface entre zones dopées de source et drain et canal, dont l'effet est invisible sur de grandes longueurs, sont suspectés, en particulier des défauts neutres dus à l'implantation des dopants de source et drain [CROS06, DUPRÉ07], dont il existe une multitude de types, et dont la localisation est mal connue [GHIBAUDO09, BIDAL09]. Quand la longueur de canal diminue, l'influence des défauts proches du profil de jonction et générés par l'implantation des dopants risque de devenir critique, comme illustré par la figure 22.

Par ailleurs, en faisant des essais sur un matériau simple, il a été montré [PHAM-NGU09] que l'interface avant avec l'oxyde de grille présentait une dégradation de la mobilité plus importante avec la réduction de L_{eff} que celle contre l'oxyde enterré (BOX), laissant entendre que les défauts en face arrière se recombinaient avec le BOX.

Néanmoins, les défauts neutres ont une caractéristique spécifique : leur impact doit *a priori* être indépendant de la température. Comme on le voit sur l'insert de la figure 21, les mesures de mobilité en fonction du champ effectif transverse et de la température peuvent aider à discriminer les phénomènes de dégradation de la mobilité. Le tableau 3 résume la dépendance des mécanismes d'interaction en fonction de la température.

Mobilité	Interaction	Dépendance en température
μ_{cb}	coulombienne	T
μ_{sr}	rugosité de surface	—
μ_{ph}	phonons	$1/T^{1\sim 1.5}$
μ_N	défauts neutres	—

Tableau 3 : dépendance des mécanismes d'interaction en fonction de la température.

Du point de vue de la modélisation, les défauts neutres et chargés sont traités dans la littérature de façon assez similaire d'un point de vue théorique [KIRÉEV75, ERGINSOY50]. Le potentiel dû à un défaut neutre vu par un porteur de charge $\pm e$ vaut :

$$V(r) = \pm \frac{Ze^2}{\epsilon r} \underbrace{e^{-k_0 r}}_{\text{écranage}} \quad (7)$$

où ϵ est la permittivité diélectrique du matériau considéré. De façon surprenante, les défauts neutres sont traités comme des défauts avec une charge $\pm Ze$ artificielle. La différence avec les défauts chargés étant leur caractère plus local, modélisé par le facteur k_0 de décroissance exponentielle plus important devant k la norme du vecteur d'onde de la charge incidente, que dans le cas des défauts chargés.

On peut en déduire la section efficace différentielle de diffusion $\sigma(\theta, \varphi)$ caractérisant la probabilité qu'une particule se trouve après collision à l'intérieur d'un angle solide $d\Omega$ déterminé par l'angle θ de déviation et φ l'angle de rotation autour de la direction initiale du transport. En observant son expression calculée en première approximation, on constate qu'elle est isotrope :

$$\sigma(\theta) = 4 \left(\frac{Ze^2 m^*}{\epsilon \hbar^2 k_0^2} \right)^2 \quad (8)$$

contrairement au cas des impuretés ionisées¹⁰. ϵ désignant la permittivité diélectrique du matériau.

On peut déduire de (7) le temps de relaxation

$$\tau_N = 1/N_N v \sigma_c \quad (9)$$

où

$$\sigma_c = \int_0^{2\pi} \int_0^\pi \sigma(\theta) (1 - \cos \theta) \sin \theta \, d\theta \, d\varphi \quad (10)$$

est la section efficace de conduction σ_c , qui est l'intégrale de σ pondérée par $(1 - \cos \theta)$, terme lié à la déviation par rapport au sens de transport.

Enfin on en déduit la mobilité dépendant des défauts neutres :

$$\mu_N = \frac{e \tau_N}{m^*} = \frac{m^* e^3}{20 \epsilon \hbar^3 N_N} \quad (11)$$

¹⁰ $\sigma(\theta) = 4 \left(\frac{Ze^2 m^*}{\epsilon \hbar^2} \right)^2 \frac{1}{k_0^2 + 4k^2 \sin^2 \frac{\theta}{2}}$ pour les impuretés ionisées.

En réalité, plusieurs hypothèses ont été utilisées ici pour permettre le calcul de cette mobilité. En effet, la diffusion devient plus anisotrope pour des porteurs de forte énergie, à la manière des interactions avec les défauts chargés. De plus, à plus longue distance, l'effet d'un potentiel de polarisation prend le relais sur le potentiel local, bien qu'il soit généralement négligé.

L'hypothèse principale proposée est celle d'une localisation des défauts proche des jonctions [CROS06], dus aux implantations de dopants, ce qui expliquerait la dépendance de la dégradation avec la longueur de canal (comme illustré sur la figure 22).

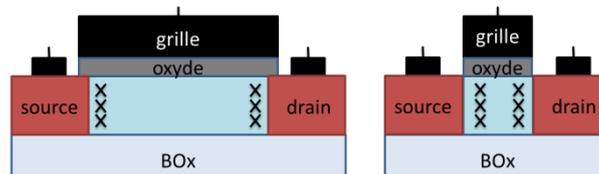


Figure 22 : localisation proche des interfaces canal/source et canal/drain. Augmentation de la densité moyenne globale de défauts avec la réduction de la longueur de canal.

4.5. Régime quasi-balistique

Dans des canaux de longueur qui se rapproche du libre parcours moyen des électrons, le transport balistique doit prendre une part *a priori* de plus en plus importante [SAINT-MA04]. V. K. Arora [ARORA11] a suggéré que le régime balistique est à moitié atteint dès 50 nm, en d'autres termes que 50 % des porteurs injectés par la source ne subiront pas une seule interaction à la traversée du canal de la source jusqu'au drain.

En effet, plusieurs régimes de transport peuvent être distingués selon le nombre d'interactions : le régime balistique, idéal, est un régime dans lequel un électron ne subit aucune interaction, comme simplement illustré sur la figure 23. Dans le régime quasi-balistique (figure 24), l'électron subit seulement quelques interactions (moins d'une dizaine) entre lesquelles il est accéléré par le champ (vol libre). Lorsqu'il y a beaucoup d'interactions, le régime est diffusif (figure 25). Le libre parcours moyen des électrons est faible et l'électron est donc moins accéléré par le champ.



Figure 23 : régime balistique

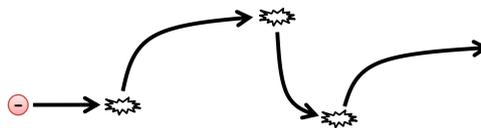


Figure 24 : régime quasi-balistique

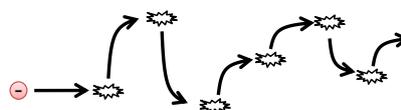


Figure 25 : régime diffusif

Lorsque la longueur de canal diminue, le faible nombre d'interactions peut rendre le modèle de dérive-diffusion inadéquat, il n'est alors possible que de décrire une mobilité apparente, prenant en considération le régime quasi-balistique.

M. S. Shur [SHUR02] a été l'un des premiers à introduire une notion de mobilité balistique qui repose sur les travaux de Lundstrom [LUNDSTRO00] fondé sur la méthode des flux. L'expression de la mobilité devient alors :

$$\frac{1}{\mu_{\text{eff}}} = \frac{1}{\mu_{\text{long}}} + \frac{1}{\mu_{\text{bal}}} \quad (12)$$

où μ_{long} est la mobilité à canal long et μ_{bal} est la mobilité balistique apparente définie par :

$$\mu_{\text{bal}} = \frac{2eL}{\pi m^* v_{\text{th}}} \quad (13)$$

où v_{th} désigne la vitesse thermique et m^* la masse effective des porteurs.

Ainsi, pour des longueurs de canal grandes devant le libre parcours moyen des porteurs, la valeur de la mobilité balistique est trop élevée pour influencer sur la mobilité effective. À l'inverse, lorsque la longueur diminue, ce terme devient prédominant dans le calcul, ce qui a pour effet de faire décroître la mobilité totale.

Il est possible de redémontrer simplement cette expression. L'expression du courant en régime linéaire dans le cas dérive-diffusion est :

$$I_{dd} = \frac{W}{L} \mu e N_{\text{inv}} V_{ds} \quad (14)$$

et par la méthode des flux¹¹ :

$$I_{QB} = \frac{W}{L} \mu \frac{e v_{\text{th}}}{2k_B T} (1 - R) V_{ds} \quad (15)$$

Avec $R = L_{\text{eff}} / (L_{\text{eff}} + \lambda)$ le coefficient de rétrodiffusion du canal en régime linéaire. En égalisant ces deux équations basées sur les hypothèses de l'équation de transport de Boltzmann semi-classique, il est possible de déterminer l'expression de la mobilité de Shur.

$$\frac{1}{\mu_{\text{eff}}} = \frac{2k_B T}{e v_{\text{th}}} \left(\frac{1}{L_{\text{eff}}} + \frac{1}{\lambda} \right) \quad (16)$$

En posant les quantités suivantes nous retrouvons l'expression (12) :

$$\mu_{\text{bal}} = \frac{2eL_{\text{eff}}}{\pi m^* v_{\text{th}}} \text{ et } \mu_{\text{long}} = \frac{2e\lambda}{\pi m^* v_{\text{th}}} \quad (17)$$

Huet *et al.* [HUET07] ont utilisé la simulation particulière Monte-Carlo pour reproduire théoriquement l'extraction de la mobilité dans des canaux ultracourts. Ils ont identifié une tendance sur l'évolution de la mobilité en fonction de la longueur de grille qui se rapproche bien de celle prédite par Shur. Néanmoins, la valeur de μ_{bal} à utiliser est inférieure à celle de la formule (13) et le

¹¹ I_{QB} dénote le courant de drain en régime quasi-balistique.

modèle de transport mis en œuvre ne prenait pas en compte les effets de quantification dans les canaux ultra-minces considérés. Les travaux de Zilli *et al.* [ZILLI07] reposant sur un simulateur Monte-Carlo plus sophistiqué et concernant un nombre de types de structures plus important n'ont pas permis d'expliquer totalement la dégradation observée par des effets quasi-balistiques.

4.6. Effets de résistance d'accès et contraintes

Une erreur systématique dans l'interprétation des résultats expérimentaux [ZILLI07], liée notamment au problème de la prise en compte des résistances d'accès au canal [HUET09, RIDEAU14], de même que les contraintes, pourraient également jouer un rôle dans la dégradation de la mobilité.

5. Conclusion du chapitre

Si les solutions sont nombreuses pour améliorer la mobilité, par modification de la géométrie, des matériaux, l'utilisation de contraintes, ou encore le choix de l'orientation cristalline la plus favorable, aucune d'entre elles ne peut empêcher la diminution inéluctable de la mobilité à longueur de grille décroissant en deçà de 100 nm.

Il faut donc déterminer un modèle physique pour rendre compte précisément de cette dégradation de mobilité, pour pouvoir la prédire et optimiser les solutions pour y remédier. La physique relativement simple implémentée en TCAD et la modélisation compacte ne permettent pas de modéliser cette dépendance de la mobilité avec la longueur.

Nous allons mener par la suite une étude théorique sur la base de simulations Monte-Carlo de dispositifs FDSOI, analysés par les méthodes utilisées pour extraire la mobilité de données expérimentales, afin de proposer un modèle simple prédisant la réduction de la mobilité dans les transistors ultracourts.

Chapitre II. Modélisation du transport

Les modèles pour simuler la physique des composants à semiconducteurs reposent en premier lieu sur une description de leurs états électroniques. Nous débuterons donc ce chapitre par une description des bandes d'énergie dans le silicium, en particulier l'approche par la masse effective largement employée dans Monaco. Le remplissage de ces états dans le cas d'un gaz d'électrons 2D sera ensuite abordé. L'essentiel du chapitre sera consacré aux équations régissant le transport des charges dans les semiconducteurs et à leur résolution dans différents cadres de simplification.

1. Structure de bandes

1.1. Principe de calcul de structure de bandes

Pour déterminer la structure de bandes, il faut déterminer les énergies des fonctions d'onde solutions de l'équation de Schrödinger stationnaire dans un potentiel périodique $U_c(\mathbf{r})$ dû au réseau cristallin. Celle-ci s'écrit dans le cas d'un électron seul :

$$H\psi = \left[-\frac{\hbar^2}{2m_0} \Delta + U_c(\mathbf{r}) \right] \psi = E\psi \quad (18)$$

où H désigne l'opérateur hamiltonien du système, ψ est la fonction d'onde qui est une fonction propre et E est l'énergie correspondant à la valeur propre associée. \hbar est la constante de Planck réduite et m_0 la masse de l'électron libre. Δ est l'opérateur laplacien.

Alors que dans le vide la résolution de ce système conduit à une superposition d'ondes planes de la forme $e^{i\mathbf{k}\cdot\mathbf{r}}$, dans un cristal les solutions peuvent s'écrire comme des ondes de Bloch :

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \quad (19)$$

où n est l'indice de bande, \mathbf{k} le vecteur d'onde, \mathbf{r} la position, et $u_{n,\mathbf{k}}(\mathbf{r})$ est une fonction ayant la périodicité du cristal, autrement dit :

$$u_{n,\mathbf{k}}(\mathbf{r} + \mathbf{r}_0) = u_{n,\mathbf{k}}(\mathbf{r}) \quad (20)$$

pour tout \mathbf{r}_0 choisi parmi les trois vecteurs de base de la maille cristalline élémentaire.

Les énergies de ces ondes sont notées $E_n(\mathbf{k})$.

Pour représenter la relation de dispersion de l'énergie $E_n(\mathbf{k})$, il faut déjà déterminer les directions privilégiées des vecteurs d'onde \mathbf{k} en revenant tout d'abord à la description du réseau cristallin de silicium.

Celui-ci possède une structure cristalline de type « diamant », obtenue à partir d'un réseau cubique à faces centrées (cfc) et de son décalage d'un quart selon une grande diagonale $\left(\frac{a}{4}, \frac{a}{4}, \frac{a}{4}\right)$ de la maille cubique élémentaire, visible sur la figure 26. Il possède donc huit atomes par maille cubique : quatre pour le réseau de base (les atomes situés à la frontière avec d'autres mailles étant également représentés) et autant pour le réseau décalé.

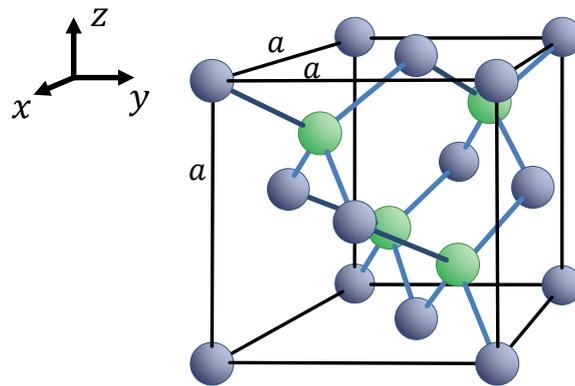


Figure 26 : maille cubique de silicium. Structure « diamant » composée d'un réseau cfc en bleu, et d'un autre décalé de $(\frac{a}{4}, \frac{a}{4}, \frac{a}{4})$ représenté en vert. Paramètre de maille : $a = 0,543$ nm à température ambiante.

Le réseau réciproque du réseau cfc est un réseau cubique centré (cc). La structure de bandes peut être restreinte à la première zone de Brillouin, représentée sur la figure 27.

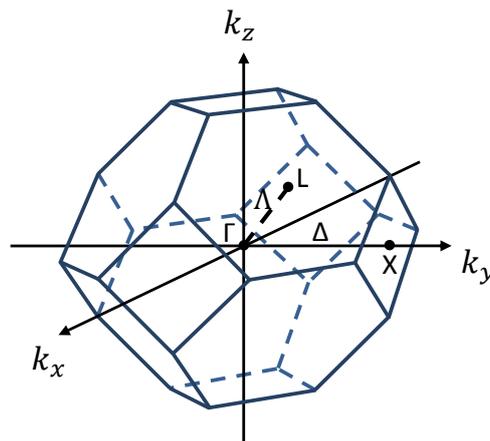


Figure 27 : première zone de Brillouin du réseau réciproque du silicium, présentant les principaux points de haute symétrie Γ , X, L et les droites caractéristiques associées Δ et Λ .

La structure de bandes est représentée pour quelques directions particulières de la zone de Brillouin. La figure 28 schématise les relations de dispersion $E_n(k)$ des bandes de valence et de conduction dans la direction (ΓX), désignée par la droite Δ .

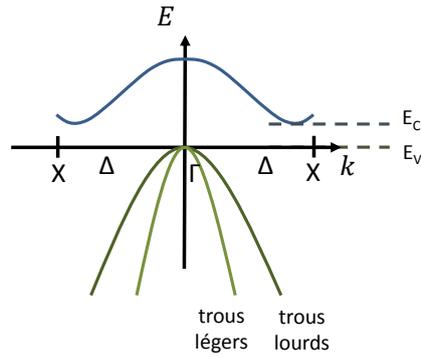


Figure 28 : structure de bandes du silicium. La vallée Δ est visible en bas de bande de conduction. Quant aux trous, les deux bandes les plus hautes de la bande de valence sont représentées (trous légers et trous lourds).

Cette direction est importante pour la bande de conduction du silicium car les minima d'énergie sont situés sur les six droites Δ à 85% du segment $[\Gamma X]$. Ces six minima sont appelés vallées Δ , autour desquelles six ellipsoïdes isoénergies sont représentées sur la figure 29.

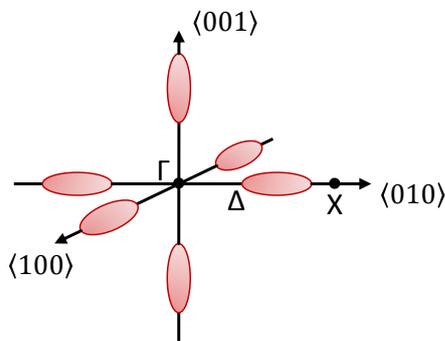


Figure 29 : ellipsoïdes isoénergies autour des six vallées Δ de la bande de conduction du silicium et situés le long des directions principales du réseau réciproque.

Une représentation des bandes plus complète et selon d'autres axes de vecteur d'onde est donnée en figure 30.

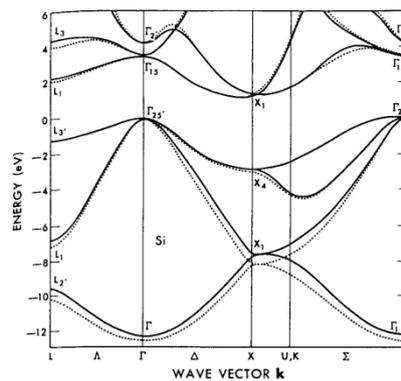


Figure 30 : diagramme de bandes du silicium¹²

¹² D'après M. L. Cohen et J. R. Chelikowsky: "Electronic Structure and Optical Properties of Semiconductors" Solid-State Sciences 75, Springer-Verlag 1988

À partir de l'analyse de la structure de bandes, il est possible de déduire la vitesse de groupe des porteurs qui sera utilisée dans le calcul du transport, selon la relation :

$$v_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E_n(\mathbf{k}) \quad (21)$$

Cette vitesse de groupe intervient donc dans le calcul de courant et par là même dans le calcul de la mobilité.

Pour modéliser le plus simplement possible la structure de bandes, le modèle de masse effective peut être utilisé pour les porteurs dont l'énergie se trouve au voisinage d'un minimum de vallée.

1.2. Modèle de masse effective

1.2.1. Modèle parabolique

Au voisinage d'un extremum d'énergie \mathbf{k}_0 , la relation de dispersion $E(\mathbf{k})$ est souvent approchée par une forme quadratique, les dérivées partielles premières étant nulles en ce point. Le tenseur de masse effective $\left(\frac{1}{m^*}\right)$ est alors défini à partir des dérivées secondes de l'énergie par rapport au vecteur d'onde :

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \partial k_j} \quad (22)$$

L'énergie autour de \mathbf{k}_0 est alors approchée par son développement de Taylor :

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2}{2} \left(\frac{1}{m^*}\right)_{ij} (\mathbf{k} - \mathbf{k}_0)_i (\mathbf{k} - \mathbf{k}_0)_j \quad (23)$$

Un autre avantage de l'approximation de la masse effective est la simplification de l'équation de Schrödinger dans un potentiel extérieur V :

$$\left[-\frac{\hbar^2}{2m_0} \Delta + U_c(\mathbf{r}) + V(\mathbf{r}) \right] \psi = E\psi \quad (24)$$

La fonction d'onde peut être approchée par l'onde de Bloch en \mathbf{k}_0 modulée par la fonction enveloppe $\varphi(\mathbf{r})$:

$$\psi(\mathbf{r}) = \varphi(\mathbf{r}) u_{n,\mathbf{k}_0}(\mathbf{r}) e^{i\mathbf{k}_0 \cdot \mathbf{r}} \quad (25)$$

La fonction enveloppe vérifie une équation de Schrödinger simplifiée :

$$-\frac{\hbar^2}{2} \left(\frac{1}{m^*}\right)_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} + V(\mathbf{r})\varphi = E'\varphi \quad (26)$$

Ce résultat provient de la théorie $\mathbf{k} \cdot \mathbf{p}$ restreinte à une seule bande. L'effet du potentiel cristallin est pris en compte dans le tenseur de masse effective.

En se plaçant dorénavant dans un système d'axes dans lequel le tenseur est diagonal, celui-ci s'écrit pour les deux vallées Δ_x du silicium, selon les axes de la maille cubique :

$$\left(\frac{1}{m^*}\right) = \begin{bmatrix} \frac{1}{m_l} & 0 & 0 \\ 0 & \frac{1}{m_t} & 0 \\ 0 & 0 & \frac{1}{m_t} \end{bmatrix} \quad (27)$$

L'équation de Schrödinger pour la fonction enveloppe peut alors s'écrire plus simplement :

$$-\frac{\hbar^2}{2m_l} \frac{\partial^2 \varphi}{\partial x^2} - \frac{\hbar^2}{2m_t} \frac{\partial^2 \varphi}{\partial y^2} - \frac{\hbar^2}{2m_t} \frac{\partial^2 \varphi}{\partial z^2} + V(\mathbf{r})\varphi = E'\varphi \quad (28)$$

Les équations pour les deux vallées Δ_y et les deux vallées Δ_z se déduisent de celle-ci par symétrie.

Cette formule est analogue au cas de l'électron libre mais avec des masses effectives $m_l = 0,92m_0$ et $m_t = 0,19m_0$ inférieures à la masse de l'électron libre m_0 grâce à l'effet du potentiel cristallin.

1.2.2. Correction des bandes non-paraboliques

Pour étendre la gamme d'énergie sur laquelle l'approximation de la masse effective est valable, un coefficient α de non parabolicité peut être introduit.

$$E(1 + \alpha E) = \frac{\hbar^2 k^2}{2m^*} \quad (29)$$

pour une bande à masse isotrope centrée au point Γ de la zone de Brillouin. Dans le cas du silicium, l'hypothèse $\alpha \simeq \frac{1}{2}(eV)^{-1}$ est généralement utilisée.

1.3. Confinement quantique

1.3.1. Confinement quantique

Le confinement quantique apparaît lorsque l'épaisseur du canal est aussi faible que les longueurs d'ondes des électrons. Ceux-ci sont alors confinés dans un puits de potentiel étroit, (cf. figure 31) et sont décrits par des fonctions enveloppes correspondant à des états propres associés à un nombre restreint de valeurs d'énergies discrètes, contrairement au cas classique.

Par ailleurs, ces fonctions enveloppes présentent une densité maximale décalée de l'interface avec l'oxyde. La fonction enveloppe étant très faible dans l'oxyde en valeur absolue, elle n'atteint pas son maximum directement à l'interface. La zone où la charge est faible à proximité de l'interface avec l'oxyde est alors dénommée *dark space*. Cela est d'autant plus vrai que le nombre de sous-bandes peuplées de façon significative est réduit. Ce phénomène est un peu comparable au filtrage d'un signal carré ne préservant qu'une somme de quelques harmoniques sinusoïdales s'annulant aux limites de leur intervalle et dont le maximum est alors décalé.

Dans le cadre du confinement quantique, les électrons ne sont donc plus représentés comme un gaz 3Dk évoluant librement dans les trois dimensions du transport, mais comme un gaz 2Dk dont le transport est modélisé dans les deux dimensions du plan du canal et dont le confinement se produit selon la dimension de l'épaisseur du canal. La fonction enveloppe de la fonction d'onde s'écrit alors comme le produit d'une fonction enveloppe selon la direction de confinement y par une onde plane dans le plan (xz) du canal.

$$\varphi(\mathbf{r}) = \xi(y)e^{i\mathbf{k}\cdot\mathbf{r}} \quad (30)$$

À présent, le vecteur position \mathbf{r} et le vecteur d'onde \mathbf{k} sont inclus dans le plan du confinement.

L'équation de Schrödinger pour la fonction enveloppe devient alors unidimensionnelle :

$$-\frac{\hbar^2}{2m_z} \frac{\partial^2 \xi}{\partial y^2} + V(y)\xi = E'\xi \quad (31)$$

Les solutions obtenues forment un ensemble discret de fonctions ξ_n d'énergie E_n . Les fonctions enveloppes complètes de vecteur d'onde \mathbf{k} s'écrivent :

$$\varphi_{n,\mathbf{k}}(\mathbf{r}) = \xi_n(y)e^{i\mathbf{k}\cdot\mathbf{r}} \quad (32)$$

et leur énergie vaut :

$$E_n(\mathbf{k}) = E_n + \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_z^2}{2m_z} \quad (33)$$

1.3.2. Concentration totale

La concentration en électrons est donnée par la superposition des densités de probabilités issues des fonctions d'onde des différentes vallées et sous-bandes :

$$n(z) = \frac{\sqrt{m_x m_y} kT}{\pi \hbar^2} \sum_{v,i} \zeta_n(y) \mathcal{F}_0 \left(\frac{E_F - E_n^v}{k_B T} \right) \quad (34)$$

où \mathcal{F}_0 est l'intégrale de Fermi-Dirac d'ordre 0.

Une coupe d'énergie potentielle selon y dans le canal, issue d'une simulation Monte-Carlo, est représentée sur la figure 31.

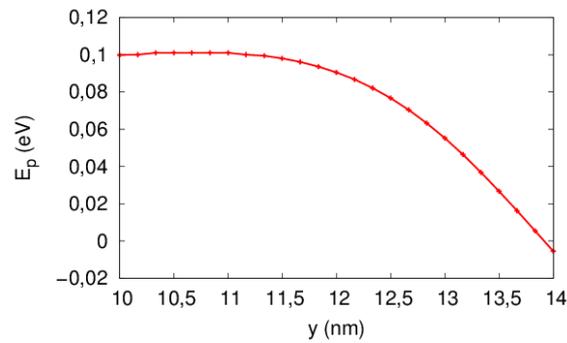


Figure 31 : coupe d'énergie potentielle selon l'axe y normal au plan de l'interface entre le canal et l'oxyde de grille (située à $y = 14$ nm. L'interface canal/BOX est à $y = 10$ nm). $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

La dépendance en y due à la polarisation et au confinement quantique est visible sur la coupe de concentration en électrons donnée sur la figure 32.

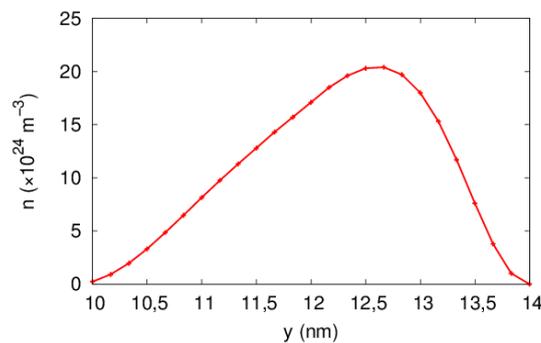


Figure 32 : coupe de concentration en électrons selon y . $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

La densité de probabilité de présence sur la première sous-bande est présentée sur la figure 33. Étant la plus peuplée, elle est la composante principale dont résulte la concentration en électrons vue sur la figure précédente : comme celle-ci, elle tend à s'annuler à proximité de l'oxyde de grille et du BOX. De plus, dans le canal, la densité de probabilité de présence est plus importante du côté de la grille ($y = 10$ nm) que du BOX ($y = 14$ nm) dans cet exemple à forte polarisation de grille V_g .

Voici les coupes en y des fonctions d'ondes précédentes.

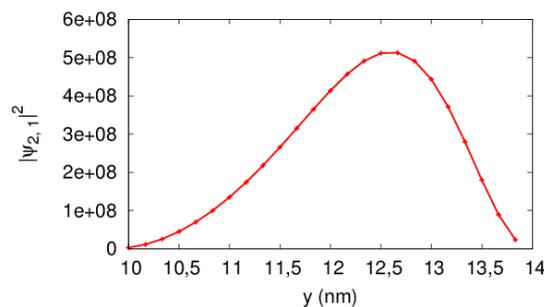


Figure 33 : coupe de la fonction d'onde en y dans le silicium. Vallée Δ_y , premier niveau. $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

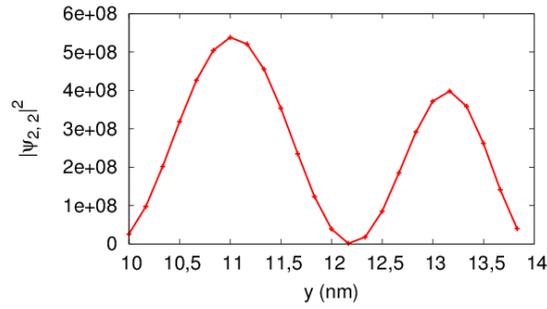


Figure 34 : coupe de la fonction d'onde en y dans le silicium. Vallée Δ_y , deuxième niveau. $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

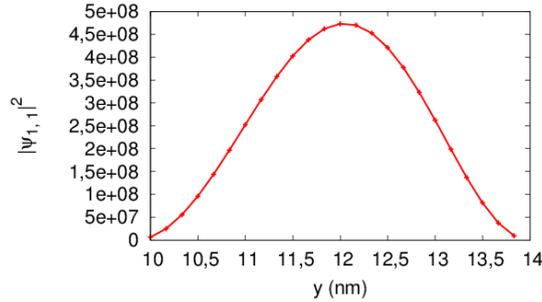


Figure 35 : coupe de la fonction d'onde en y dans le silicium. Vallée Δ_x , premier niveau. $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

2. Équation de transport de Boltzmann

C'est une équation de transport semi-classique dans laquelle la trajectoire des particules est décrite de manière classique et leur interaction (collisions) de manière quantique, comme le sont leurs bandes d'énergie.

2.1. Hypothèse semi-classique

L'hypothèse semi-classique utilisée est valide dans le cas où les dimensions du dispositif sont assez grandes par rapport à la longueur d'onde des électrons $\lambda = \sqrt{\frac{\hbar^2}{2m^*E}}$ (2 nm à 300 K pour une énergie $E = \frac{3k_B T}{2}$) et à la longueur de relaxation de phase des fonctions d'ondes.

La vitesse de groupe est donnée par la relation $\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k})$, dans laquelle $E(\mathbf{k})$ est l'énergie donnée par la relation de dispersion. En présence d'une force \mathbf{F} appliquée à l'électron, due à un champ électrique, l'équation du mouvement semi-classique est :

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F} \quad (35)$$

Si une ou plusieurs dimensions du système sont petites, des phénomènes de transport quantiques apparaissent :

- quantification de l'énergie dans le plan perpendiculaire au transport, allant de pair avec l'effet de confinement quantique éloignant les électrons de l'interface avec l'oxyde.
- effet tunnel induisant un courant de fuite entre différentes zones du dispositif, que ce soit la source, le canal, le drain ou la grille.
- réflexion sur une barrière d'énergie inférieure à l'énergie cinétique de l'électron.

De plus, les collisions entre des porteurs ou entre un porteur et son environnement constituent des interférences qui doivent également être traitées de manière quantique.

2.2. Équation de Boltzmann 2Dk

Un système comprenant un grand nombre de porteurs peut être décrit par une fonction de distribution $f(\mathbf{r}, \mathbf{k}, t)$, qui représente la probabilité de présence d'un porteur ayant un vecteur d'onde \mathbf{k} , à la position \mathbf{r} et à l'instant t .

Dans un modèle 2Dk, les particules se déplacent dans un plan défini par les deux directions du transport et sont soumises au confinement quantique dans la troisième. f est alors la somme des fonctions de distribution $f_n(\mathbf{r}, \mathbf{k}, t)$ par sous-bande n :

$$f(\mathbf{r}, \mathbf{k}, t) = \sum_n f_n(\mathbf{r}, \mathbf{k}, t) \quad (36)$$

À l'équilibre, elle est donnée par la fonction de Fermi-Dirac :

$$f_n^0(\mathbf{k}) = \frac{1}{1 + e^{\frac{E_n(\mathbf{r}, \mathbf{k}) - E_F}{k_B T}}} \quad (37)$$

Une représentation de cette fonction de distribution à l'équilibre est donnée en fonction de l'énergie sur la figure 36 pour différentes températures.

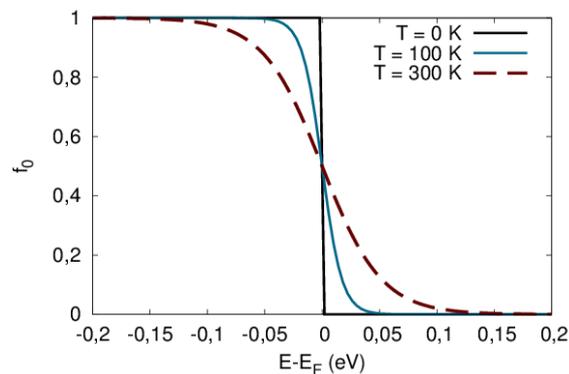


Figure 36 : fonction de distribution de Fermi-Dirac à l'équilibre f_n^0 en fonction de l'énergie autour du niveau de Fermi, à différentes températures.

Le niveau de Fermi définit l'énergie en-dessous de laquelle se trouvent la majorité des électrons.

L'équation de transport de Boltzmann (ETB) donne l'évolution temporelle et spatiale de la fonction de distribution. Pour simplifier la démonstration, seule la dimension de l'espace réel x est

prise en compte, selon le schéma de la figure 37, G_n étant le taux de génération/recombinaison par unité de temps.

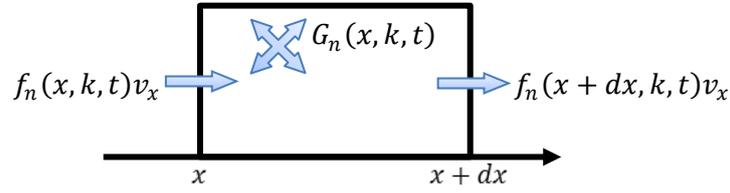


Figure 37 : bilan de variation de la fonction de distribution pendant dt sur un intervalle de longueur dx permettant d'obtenir l'équation de Boltzmann.

Dans ce cadre, un bilan d'évolution de la fonction de distribution est effectué sur un intervalle de temps dt , une zone de longueur dx et un intervalle de vecteurs d'onde dk sur lequel la force extérieure F fait entrer ou sortir des porteurs :

$$\begin{aligned} & (f_n(x, k, t + dt) - f_n(x, t)) dx dk \\ &= f_n(x, k, t) v_x dk dt - f_n(x + dx, k, t) v_x dk dt + f_n(x, k, t) \frac{F}{\hbar} dt dx \\ & - f_n(x, k + dk, t) \frac{F}{\hbar} dt dx + G_n(x, k, t) dx dk dt \end{aligned} \quad (38)$$

En divisant les termes de l'équation par $dx dk dt$, il vient :

$$\frac{\partial f_n}{\partial t} = - \frac{\partial f_n(x, k, t)}{\partial x} v_x - \frac{\partial f_n(x, k, t)}{\partial x} \frac{F}{\hbar} + G_n(x, k, t) \quad (39)$$

En 2Dk, en faisant un bilan sur les deux dimensions du transport, l'équation s'écrit :

$$\frac{\partial f_n}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f_n + \frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f_n = \left. \frac{\partial f_n}{\partial t} \right|_{\text{coll}} + G_n(\mathbf{r}, \mathbf{k}, t) \quad (40)$$

où \mathbf{r} et \mathbf{k} sont restreints au plan du transport, $\left. \frac{\partial f_n}{\partial t} \right|_{\text{coll}}$ est l'intégrale de collision due aux interactions entre les porteurs et leur milieu de propagation et $G_n(\mathbf{r}, \mathbf{k}, t)$ est le taux de génération/recombinaison par unité de temps.

L'évaluation de l'intégrale de collision $\left. \frac{\partial f_n(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right|_{\text{coll}}$ est un obstacle à la résolution de l'ETB. Celle-ci est déterminée par les probabilités d'interactions faisant passer de l'état \mathbf{k} à un état \mathbf{k}' ou inversement, évaluée de différentes manières selon le type de simulations, les modélisations TCAD, Monte-Carlo et Kubo-Greenwood étant détaillées dans les dernières sections de ce chapitre.

2.3. Calcul des taux de transition

Les taux de transition $S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'}$ intervenant dans le calcul du taux d'interaction sont le plus souvent évalués par la règle d'or de Fermi. Dans le cas d'une interaction élastique décrite par une perturbation V_p du hamiltonien :

$$S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'} = \frac{2\pi}{\hbar} |\langle \varphi_{\mathbf{k}'} | V_p | \varphi_{\mathbf{k}} \rangle|^2 \delta(E_{n'}(\mathbf{k}') - E_n(\mathbf{k})) \quad (41)$$

Ce type d'interaction correspond à une perturbation indépendante du temps.

Dans le cas d'une interaction inélastique, telle que l'interaction avec un phonon optique ou acoustique de grand vecteur d'onde, cette règle d'or s'écrit :

$$S_{n,k \rightarrow n',k'} = \frac{2\pi}{\hbar} |\langle \varphi_{k'} | V_{e-ph} | \varphi_k \rangle|^2 \delta(E_{n'}(\mathbf{k}') - E_n(\mathbf{k}) \pm \hbar\omega) \quad (42)$$

le signe devant l'énergie du phonon $\hbar\omega$ étant positif dans le cas de l'absorption de phonon par le porteur de charge et négatif dans le cas de l'émission.

2.4. Interactions considérées

Les interactions modélisées sont celles avec les phonons acoustiques de faible vecteur d'onde (collisions intravallées considérées comme élastiques), les phonons intervallées, les impuretés ionisées et la rugosité à l'interface entre Si et oxyde de grille, illustrée par l'image en microscopie électronique à transmission (TEM) de la figure 38.

La fréquence d'interaction $\lambda_{n \rightarrow n'}(\mathbf{k})$ d'un porteur de vecteur d'onde \mathbf{k} de la sous-bande n vers la sous-bande n' est donnée dans le cas générique par l'intégration des taux de transition vus précédemment sur tous les états finaux \mathbf{k}' possibles, soit dans le cas présent d'un gaz d'électrons 2Dk :

$$\lambda_{n \rightarrow n'}(\mathbf{k}) = \int S_{n,k \rightarrow n',k'} d\mathbf{k}' = \iint S_{n,k \rightarrow n',k'} J(E', \theta') dE' d\theta' \quad (43)$$

où $S_{n,k \rightarrow n',k'}$ est le taux de transition entre l'état n, k et l'état n', k' et $J(E', \theta')$ est le jacobien de passage entre le repère cartésien (k'_x, k'_z) et le repère cylindrique $(E(k'), \theta')$.

Les paramètres d'interaction listés dans le tableau 4 sont utilisés par la suite pour décrire les interactions détaillées en annexe 1. Les phonons considérés correspondent aux vibrations des atomes dans le volume de Si. Les valeurs sont essentiellement celles généralement implémentées dans le code Monte-Carlo MONACO de l'IEF [DOLLFUS97, AUBRY-FO05, SAINT-MA06, MONSEF04]. La rugosité d'interface peut être vue comme un ensemble de défauts de surface, caractérisés par leur amplitude Δ et longueur caractéristique Λ comme schématisé sur la figure 39. Les valeurs considérées sont celles souvent choisies dans le cas de l'interface Si-SiO₂.

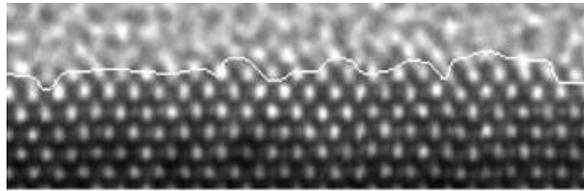


Figure 38 : image TEM de l'interface Si-SiO₂ [ZHAO09]

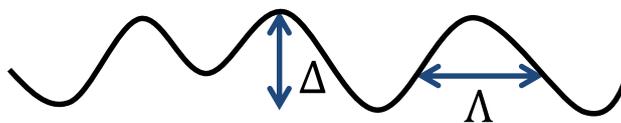


Figure 39 : paramètres géométriques de rugosité de surface

Paramètre		Notation	Valeur	Unité
potentiel de déformation de phonon acoustique		D_{ac}	6,6	eV
potentiel de déformation de phonon intervallée	ordre 0	D_0	$3,4 \cdot 10^8$	eV/cm
	ordre 1	D_1	3	eV
énergie du phonon d'ordre 0	type g	$\hbar\omega$	63,2	meV
	type f		46,3 – 59,1	
énergie du phonon d'ordre 1	type g		11,4 – 18,8	
	type f		21,9	
vitesse du son dans le silicium ¹³		v_p	6607	m/s
taux d'écrantage des impuretés ionisées		N_{scr}	0,2	-
longueur de corrélation (rugosité d'interface avec l'oxyde)		Λ	2,12	nm
hauteur moyenne quadratique (rugosité)		Δ	0,5	

Tableau 4 : paramètres du modèle d'interaction

2.5. Grandeurs physiques obtenues

À partir de la fonction de distribution évaluée par cette équation, plusieurs grandeurs liées au transport peuvent être exprimées, dont la densité surfacique de charge N et la densité surfacique de courant j :

$$N(\mathbf{r}, t) = \langle N \rangle = \frac{1}{\Omega} \int f(\mathbf{r}, \mathbf{k}, t) \text{DOS}(\mathbf{k}) d^2k \quad (44)$$

$$\mathbf{v}(\mathbf{r}, t) = \langle \mathbf{v} \rangle = \frac{1}{\Omega} \int \mathbf{v}(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \text{DOS}(\mathbf{k}) d^2k \quad (45)$$

$$j(\mathbf{r}, t) = \langle j \rangle = \frac{1}{\Omega} \int j(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \text{DOS}(\mathbf{k}) d^2k \quad (46)$$

$$E(\mathbf{r}, t) = \langle E \rangle = \frac{1}{\Omega} \int E(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \text{DOS}(\mathbf{k}) d^2k \quad (47)$$

Ω étant une surface de normalisation et $\text{DOS}(\mathbf{k}) d^2k$ le nombre d'états dans la surface Ω dont le vecteur d'onde est dans une « surface » d^2k autour de \mathbf{k} .

3. Modélisation Kubo-Greenwood

Cette approche permet d'obtenir rapidement la mobilité dans un matériau dans l'hypothèse de champ faible et homogène [KUBO57, GREENWOOD58]. Le calcul est beaucoup plus rapide que les méthodes de résolution de type Monte-Carlo puisque contrairement à celles-ci le mouvement des

¹³ Vitesse moyennée géométriquement entre la vitesse longitudinale et la vitesse transverse

porteurs le long du canal n'est pas simulé. Elle consiste en une résolution de l'équation de Boltzmann en régime linéaire.

3.1. Linéarisation de l'équation de Boltzmann et expression de la mobilité

En régime permanent et en supposant que le système est homogène, c'est-à-dire que la distribution en \mathbf{k} (autrement dit en vitesses) des porteurs est indépendante de la position, l'équation de Boltzmann devient :

$$\frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f_n = \left. \frac{\partial f_n}{\partial t} \right|_{\text{coll}} \quad (48)$$

où f_n ne dépend plus que de \mathbf{k} , et pas de \mathbf{r} et t . Cette équation reste valable en régime non linéaire, mais est difficile à résoudre. Nous nous plaçons maintenant en régime linéaire. En utilisant une approche perturbative, la fonction de distribution f_n est alors exprimée comme la somme de son état d'équilibre f_0 et de sa variation g :

$$f_n(\mathbf{k}) = f_n^0(\mathbf{k}) + g_n(\mathbf{k}) \quad (49)$$

On suppose que le terme de collision peut aussi être linéarisé et qu'il est proportionnel à $g_n(\mathbf{k})$.

$$\left[\frac{\partial f_n(\mathbf{k})}{\partial t} \right]_{\text{coll}} = -\frac{g_n(\mathbf{k})}{\tau_n(\mathbf{k})} \quad (50)$$

Notons que la seule hypothèse faite ici est celle de linéarisation. Elle sera démontrée dans la section suivante. Le fait d'écrire le facteur de proportionnalité sous la forme $1/\tau_n(\mathbf{k})$ n'implique pas que l'on fasse l'approximation du temps de relaxation. Injectée dans l'ETB, cette linéarisation donne :

$$\frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f_n = -\frac{g_n(\mathbf{k})}{\tau(\mathbf{k})} \quad (51)$$

Dans le membre de gauche, la fonction de distribution peut être approchée par son développement linéaire en fonction du champ :

$$\nabla_{\mathbf{k}} f_n = \frac{\partial f_n^0}{\partial E} \nabla_{\mathbf{k}} E_n(\mathbf{k}) \quad (52)$$

En reconnaissant dans le gradient de l'énergie l'expression de la vitesse de groupe vue précédemment, l'équation peut être transformée en :

$$\frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f_n = -eE(\mathbf{k})v(\mathbf{k}) \frac{\partial f_n^0(\mathbf{k})}{\partial E} \quad (53)$$

Finalement la variation de la fonction de distribution peut s'écrire comme suit :

$$g_n(\mathbf{r}, \mathbf{k}, t) = -e\tau(\mathbf{k})v(\mathbf{k})E(\mathbf{k}) \frac{\partial f_n^0(\mathbf{k})}{\partial E} \quad (54)$$

Où \mathbf{E} est le champ électrique. Cette forme peut être injectée dans l'expression du courant, qui vaut dans le cas d'un transistor planaire, pour une boîte 2Dk de dimensions L et W :

$$I = -\frac{1}{L} g_s e \frac{LW}{(2\pi)^2} \int v(\mathbf{k})(f_n(\mathbf{k}) - f_n^0(\mathbf{k})) d\mathbf{k} \quad (55)$$

Le courant peut alors s'écrire :

$$I = \frac{1}{L} g_s e \frac{LW}{(2\pi)^2} \int v^2(\mathbf{k}) e E \tau(\mathbf{k}) \left(-\frac{df_n^0}{dE} \right) d\mathbf{k} \quad (56)$$

Or la mobilité a pour expression dans le cas général d'un transistor planaire :

$$\mu = \frac{\sigma}{en_{2D}} = \frac{I}{Wn_{2D}eE} \quad (57)$$

L'expression de la mobilité par Kubo-Greenwood peut aisément être déduite :

$$\mu_{KG} = \frac{eg_s}{n_{2D}(2\pi)^2} \int v_x^2(\mathbf{k}) \tau(\mathbf{k}) \left(-\frac{df_n^0}{dE} \right) d\mathbf{k} \quad (58)$$

Pour obtenir des valeurs de cette mobilité, il faut déterminer les temps de collision $\tau_n(\mathbf{k})$.

3.2. Calcul des temps de collision

Exprimé à l'aide des taux de transition, le terme de collision s'écrit sous la forme :

$$\left[\frac{\partial f_n(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right]_{\text{coll}} = - \sum_{n', \mathbf{k}'} S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'} f_n(\mathbf{r}, \mathbf{k}, t) [1 - f_{n'}(\mathbf{r}, \mathbf{k}', t)] + \sum_{n', \mathbf{k}'} S_{n', \mathbf{k}' \rightarrow n, \mathbf{k}} f_{n'}(\mathbf{r}, \mathbf{k}', t) [1 - f_n(\mathbf{r}, \mathbf{k}, t)] \quad (59)$$

où f_n est la fonction de distribution de la bande d'indice n et $S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'}$ est le taux de transition entre l'état $|n, \mathbf{k}\rangle$ et l'état $|n', \mathbf{k}'\rangle$, dont l'évaluation sera donnée dans la sous-section suivante.

Puis pour simplifier le terme de collision, l'équation de la balance détaillée peut être utilisée :

$$S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'} f_n^0(\mathbf{r}, \mathbf{k}, t) [1 - f_{n'}^0(\mathbf{r}, \mathbf{k}', t)] = S_{n', \mathbf{k}' \rightarrow n, \mathbf{k}} f_{n'}^0(\mathbf{r}, \mathbf{k}', t) [1 - f_n^0(\mathbf{r}, \mathbf{k}, t)] \quad (60)$$

Après linéarisation de f_n (éq. 49), et en ne conservant que les termes du premier ordre, le terme de collision devient :

$$\left[\frac{\partial f_n(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right]_{\text{coll}} = - \sum_{n', \mathbf{k}'} S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'} \frac{1 - f_n^0(\mathbf{r}, \mathbf{k}', t)}{1 - f_n^0(\mathbf{r}, \mathbf{k}, t)} \left(g_n(\mathbf{r}, \mathbf{k}, t) - g_n(\mathbf{r}, \mathbf{k}', t) \frac{f_n^0(\mathbf{r}, \mathbf{k}', t)}{f_n^0(\mathbf{r}, \mathbf{k}, t)} \frac{1 - f_n^0(\mathbf{r}, \mathbf{k}, t)}{1 - f_n^0(\mathbf{r}, \mathbf{k}', t)} \right) \quad (61)$$

Finalement, en utilisant l'expression (54) de g_n , un système d'équations implicites pour les temps de collision peut en être déduit :

$$\frac{1}{\tau_n(\mathbf{k})} = \sum_{n', \mathbf{k}'} S_{n, \mathbf{k} \rightarrow n', \mathbf{k}'} \frac{1 - f_n^0(\mathbf{r}, \mathbf{k}', t)}{1 - f_n^0(\mathbf{r}, \mathbf{k}, t)} \left(1 - \frac{\tau_{n'}(\mathbf{k}') v_{n'}(\mathbf{k}')}{\tau_n(\mathbf{k}) v_n(\mathbf{k})} \right) \quad (62)$$

Il est possible de résoudre ce système d'équations de façon itérative. Mais dans la plupart des études Kubo-Greenwood le système est transformé sous une forme explicite par une approximation dans laquelle le facteur de droite est remplacé par le terme angulaire $\cos \theta$:

$$\frac{1}{\tau_n(\mathbf{k})} = \sum_{n',k'} \left[S_{n,k \rightarrow n',k'} \frac{1 - f_n^0(\mathbf{r}, \mathbf{k}', t)}{1 - f_n^0(\mathbf{r}, \mathbf{k}, t)} \cos \theta \right] \quad (63)$$

où θ est l'angle entre \mathbf{k} et \mathbf{k}' .

La méthode de Kubo-Greenwood n'a pas été utilisée dans cette thèse mais les équations (50) et (63) sont importantes pour comprendre les approximations utilisées en Monte-Carlo et en TCAD.

4. Simulation Monte-Carlo 2Dk

Plus exacte que les différents modèles utilisés habituellement en TCAD, la modélisation Monte-Carlo consiste en une résolution statistique précise de l'équation de transport de Boltzmann (ETB) semi-classique, n'utilisant pas d'hypothèse d'approximation du temps de relaxation pour décrire les interactions. La simulation se faisant au niveau des particules simulées aléatoirement, le régime quasi-balistique et même le transport balistique (idéal) peuvent naturellement être pris en compte.

Cette résolution de l'ETB dans les directions du transport 2Dk, selon les axes x et z , se fait de manière couplée avec l'équation de Poisson 2D, selon les directions x et y , de même qu'avec l'équation de Schrödinger 1D le long du confinement perpendiculaire, selon l'axe y . La géométrie simulée et les directions de résolution de ces différentes équations sont présentées sur la figure suivante :

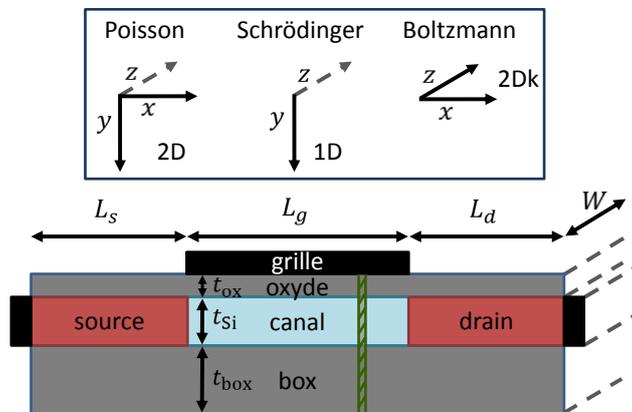


Figure 40 : géométrie du FDSOI dans la résolution Monte-Carlo, supposée invariante par translation selon l'axe z représenté en pointillés. Une tranche de résolution de l'équation de Schrödinger est représentée hachurée en vert selon l'axe y . Les axes des principales équations sont représentés au-dessus.

La notation 2Dk pour l'équation de Boltzmann qualifie les dimensions du transport, puisque ce sont les vecteurs d'onde \mathbf{k} qui évoluent selon celles-ci. La dépendance selon l'axe z (dans le sens de la largeur du transistor) est prise en considération. En effet, malgré l'invariance du système selon l'axe z , la composante k_z du vecteur d'onde intervient dans le calcul des temps de relaxation, contrairement aux autres équations (Poisson et Schrödinger).

Dans chaque tranche de dispositif dans laquelle est résolue l'équation de Schrödinger, la fonction d'onde est supposée indépendante de x , sauf à un facteur de phase $e^{ik_x x}$ près. Finalement, ces trois équations sont couplées dans le code Monte-Carlo 2Dk.

4.1. Résolution de l'équation de Boltzmann 2Dk

Pour résoudre l'ETB, la méthode Monte-Carlo est une décomposition de la trajectoire de chaque particule en plusieurs étapes :

- une séquence de vol libre, dont la durée est choisie aléatoirement en fonction de la fréquence d'apparition d'une interaction.
- une interaction, choisie elle aussi aléatoirement parmi les différents types d'interaction, tout comme son effet sur la particule, en tenant compte de leurs fréquences d'apparition respectives.

La trajectoire modélisée dans le cas simple d'un champ constant est illustrée sur la figure 41 suivante :

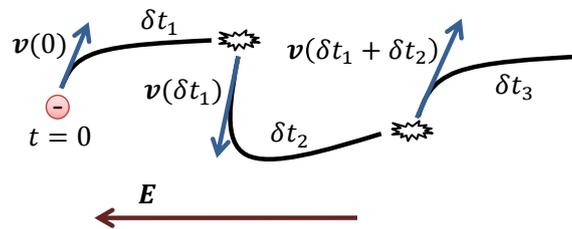


Figure 41 : illustration de la simulation Monte-Carlo à champ constant sur la trajectoire d'un électron. E est le champ électrique, v la vitesse de la particule simulée et chaque δt_i est un temps de vol libre entre deux interactions.

L'algorithme de simulation se divise en plusieurs étapes dont des tirages au sort successifs. Il est résumé sur le schéma de la figure 42 : pour simuler l'effet des différentes probabilités d'interaction, un tirage au sort est tout d'abord effectué selon le temps moyen d'occurrence de l'ensemble des interactions pour déterminer le temps de vol libre. Puis un nombre est choisi aléatoirement sur un intervalle décomposé en intervalles ayant pour longueurs les probabilités des différentes interactions afin de sélectionner celle subie par le porteur en fin de vol libre. Un autre nombre est tiré au sort pour obtenir l'effet de l'interaction. L'algorithme est itéré jusqu'à la fin d'un pas sur le temps pour une particule simulée donnée.

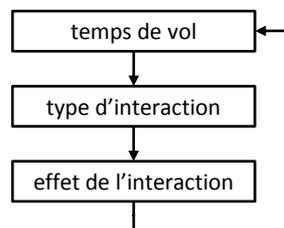


Figure 42 : cycle des séquences de la simulation Monte-Carlo

Le modèle MONACO suppose une forme ellipsoïdale des surfaces isoénergies, à l'image de la figure théorique vue précédemment (figure 29). Un coefficient de non-parabolicité est introduit.

4.1.1. Modélisation du vol libre

Lors de la séquence de vol libre, le mouvement de la particule sous l'influence du champ électrique peut simplement être décrit par l'équation de Newton :

$$\frac{d\mathbf{k}}{dt} = \frac{1}{\hbar} \frac{d\mathbf{p}}{dt} = \frac{-e\mathbf{E}}{\hbar} \quad (64)$$

Dans le cas d'un champ constant, cette équation peut être intégrée sur un intervalle de temps δt pour obtenir :

$$\mathbf{k}(t_0 + \delta t) = \mathbf{k}(t_0) - \frac{e\mathbf{E}}{\hbar} \delta t \quad (65)$$

Le champ électrique dérive du potentiel calculé régulièrement pour être à jour, par résolution de l'équation de Poisson. Le potentiel dépend à son tour de la répartition des particules, la simulation est en fait réalisée de manière autocohérente.

Les coordonnées de la particule sont :

$$\mathbf{r}(t_0 + \delta t) = \mathbf{r}(t_0) + \mathbf{v}(t_0) \delta t \quad (66)$$

où la vitesse de groupe est donnée par $\mathbf{v}(t_0) = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}(t))$, soit dans le cas d'une vallée ellipsoïdale avec correction de non parabolicité :

$$\mathbf{v}(t_0) = \begin{pmatrix} \frac{\hbar k_x}{m_x(1 + 2\alpha E(t))} \\ \frac{\hbar k_z}{m_z(1 + 2\alpha E(t))} \end{pmatrix} \quad (67)$$

4.1.2. Modélisation des interactions

Plusieurs mécanismes de diffusion sont considérés, chacun ayant un taux d'interaction moyen différent, inverse du temps d'interaction $\tau_{n,i}(p)$, leur somme permettant de définir un taux d'interaction total :

$$\frac{1}{\tau_n(p)} = \sum_{i=1}^{N_{\text{int}}} \frac{1}{\tau_{n,i}(p)} \quad (68)$$

où N_{int} est le nombre de types d'interaction. En l'occurrence, trois interactions sont prises en compte : les interactions avec les phonons, les impuretés ionisées et la rugosité de surface.

Pour simplifier les calculs, on fait une approximation isotrope pour les taux de transition $S_{n,\mathbf{k} \rightarrow n',\mathbf{k}'}$: à chaque collision, l'angle du vecteur d'onde final \mathbf{k}' est choisi de manière aléatoire avec une probabilité isotrope. Mais on définit la fréquence d'interaction totale $\tau_n(\mathbf{k})$ d'après la formule (63) en prenant bien en compte l'anisotropie de $S_{n,\mathbf{k} \rightarrow n',\mathbf{k}'}$. De ce fait, la mobilité à champ faible obtenue en Monte-Carlo avec des collisions isotropes est la même que celle obtenue en Kubo-Greenwood sans l'approximation isotrope. Les fréquences de collision utilisées en Monte-Carlo sont ainsi bien calibrées.

Le principe d'exclusion de Pauli interdit les transitions d'un état quantique vers un autre si ce dernier est déjà occupé. Sa limitation de l'effet des interactions a été introduite dans MONACO via

une pondération des fréquences d'interaction par un terme multiplicatif en $1 - f_{MC}(E')$, où $f_{MC}(E')$ est la fonction d'occupation de l'état d'énergie E' d'arrivée [CAROSELLO8]. Cette fonction est régulièrement mise à jour au cours des simulations. En pratique, la prise en compte du principe de Pauli a une incidence très faible sur les résultats présentés dans ce mémoire.

4.2. Résolution de l'équation de Schrödinger 1D

Il est nécessaire de prendre en compte la quantification perpendiculaire pour un canal planaire mince FDSOI ayant quelques nanomètres d'épaisseur t_{Si} , de l'ordre de la longueur de Broglie, de même qu'il est nécessaire de recourir à d'autres méthodes de résolution pour des canaux de longueur inférieure à cette limite.

Lorsque le potentiel varie peu avec l'axe principal du transport x , l'approche par mode (ou *mode space*) [REN01, SAINT-MA06] peut être utilisée. On note $\psi(x, y, z)$ la fonction enveloppe 3D. L'invariance selon z permet de la décomposer sous la forme :

$$\psi(x, y, z) = \frac{\varphi(x, y)e^{ik_z z}}{\sqrt{W}} \quad (69)$$

où W est la largeur du transistor (selon l'axe z).

L'approche *mode space* consiste à découpler l'équation de Schrödinger 2D stationnaire :

$$-\frac{\hbar^2}{2m_x} \frac{\partial^2 \varphi(x, y)}{\partial x^2} - \frac{\hbar^2}{2m_y} \frac{\partial^2 \varphi(x, y)}{\partial y^2} + V(\mathbf{r})\varphi(x, y) = E\varphi(x, y) \quad (70)$$

en deux équations 1D : une équation régissant le transport selon l'axe x et une équation selon l'axe y pour la détermination des modes des fonctions d'ondes électroniques.

À la position x , la fonction d'onde peut alors s'écrire sous la forme :

$$\varphi(x, y) = \sum_n \chi_n(x) \xi_n(x, y) \quad (71)$$

où $\xi_n(x, y)$ est l'état propre (ou mode) n d'énergie $E_n(x)$ et $|\chi_n(x)|^2$ est la probabilité de présence d'un électron à l'état propre n en x . Si les termes de couplage entre modes $\frac{d\chi_n(x)}{dx} \int \xi_n^*(x, y) \frac{\partial \xi_m}{\partial x}(x, y) dy$ et $\int \xi_n^*(x, y) \frac{\partial^2 \xi_m}{\partial x^2}(x, y) dy$ sont négligeables devant $\frac{d^2 \chi_n(x)}{dx^2} \int \xi_n^*(x, y) \xi_m(x, y) dy$, l'équation de Schrödinger 2D devient équivalente au système d'équations 1D suivant :

$$\begin{cases} -\frac{\hbar^2}{2m_x} \frac{d^2 \chi_n(x)}{dx^2} - E_n(x) \chi_n(x) = E \chi_n(x) \\ -\frac{\hbar^2}{2m_y} \frac{\partial^2 \xi_n(x, y)}{\partial y^2} - eV(x, y) \xi_n(x, y) = E \xi_n(x, y) \end{cases} \quad (72)$$

L'approche par mode est adaptée dans le code MONACO 2Dk : l'équation de Schrödinger 1D selon y est utilisée alors que celle selon x est remplacée par l'équation classique de Boltzmann, exposée précédemment, régissant le mouvement des porteurs considérés comme ponctuels dans le plan (xz) du transport. L'hypothèse de validité est respectée si les états propres évoluent peu dans la direction du transport, autrement dit si $\left| \frac{\partial \xi_m}{\partial x}(x, y) \frac{d\chi_n(x)}{dx} \right| \ll \left| \xi_n(x, y) \frac{d^2 \chi_n(x)}{dx^2} \right|$ et $\left| \frac{\partial^2 \xi_m}{\partial x^2}(x, y) \right| \ll \left| \xi_n(x, y) \frac{d^2 \chi_n(x)}{dx^2} \right|$ pour tous niveaux n et m , ce qui est le cas dans les transistors FDSOI entre autres, pour une tension drain-source assez faible [VENUGOPA02].

À la position x_0 , la fonction d'onde associée à un électron peut s'écrire sous la forme :

$$\psi(y, z) = \frac{\xi_n(y) e^{ik_z z}}{\sqrt{W}} \quad (73)$$

où W est la largeur du transistor (selon l'axe z).

Pour simplifier la résolution des fonctions enveloppes, elles sont considérées comme nulles dans les oxydes. C'est une bonne approximation qui a été vérifiée en pratique.

4.3. Résolution de l'équation de Poisson 2D

Le schéma de résolution de l'équation couplée Poisson-Schrödinger 1D est celui présenté en Figure 43 :

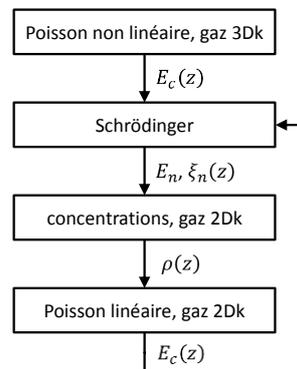


Figure 43 : résolution de l'équation couplée Poisson-Schrödinger 1D [SAINT-MA06].

Plus précisément, l'algorithme du principe de la simulation Monte-Carlo est détaillé en figure 44 :

Deux types de simulation existent : la simulation d'Ensemble Monte-Carlo, où les particules sont suivies en parallèle, comme dans MONACO, et l'approche par flux incident où les charges sont injectées depuis un contact et suivies individuellement pendant leur parcours dans le composant.

Cette méthode semi-classique a pour intérêt de prendre en compte des interactions, régies par les équations de la mécanique quantique, et le confinement quantique. Contrairement à l'expérience, elle permet d'étudier des cas en ayant une connaissance précise de la répartition de charge, mais elle a pour inconvénients la présence d'un bruit à faible charge et un besoin de temps de simulation et d'espace mémoire importants.

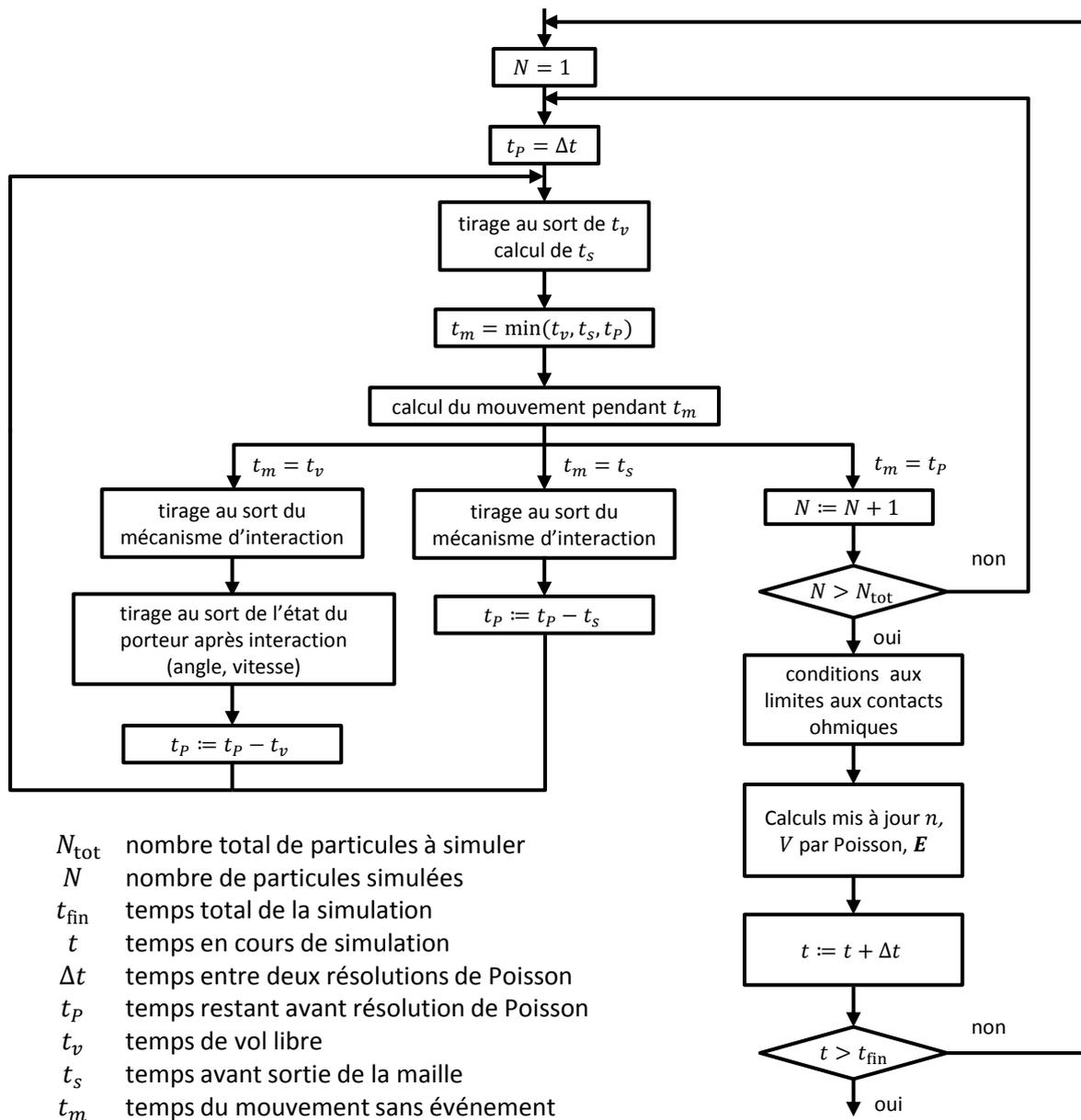


Figure 44 : algorithme détaillé de la simulation Monte-Carlo [JAUD06]

5. Simulation TCAD

La TCAD (*technology computer aided design*) est très utilisée dans le monde de la microélectronique car elle permet de coupler deux types de simulations fondamentales pour évaluer les nouveaux concepts de transistor. D'une part il est possible de réaliser en simulation le procédé de fabrication et d'autre part, à l'aide des modèles de transport associés, de faire la simulation électrique de cette même architecture. Seule la partie de la TCAD relative à la description des modèles de transport dans le transistor est considérée ici.

Les approches TCAD se fondent toutes sur la méthode des moments issue de l'intégration de transport de Boltzmann. Nous nous plaçons donc dans le cadre de la physique semi-classique où seul le terme d'interaction est décrit par la physique quantique [KIRÉEV75, SAINT-MA05]. Le point de départ des modélisations présentées dans de ce paragraphe et de la méthode des moments est alors l'équation de transport de Boltzmann semi-classique :

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f = \left. \frac{\partial f}{\partial t} \right|_{\text{coll}} \quad (74)$$

La difficulté réside dans l'expression de ce terme de collision et des hypothèses physiques sous-jacentes.

5.1. Approximations sur la fonction de distribution

5.1.1. Approximation du temps de relaxation

Pour surmonter l'obstacle de l'évaluation de l'intégrale de collision, celle-ci est remplacée par sa valeur donnée en supposant un temps de relaxation moyen τ correspondant au temps caractéristique de vol libre entre deux interactions. Il vient donc :

$$\left. \frac{\partial f}{\partial t} \right|_{\text{coll}} = -\frac{f - f_0}{\tau} \quad (75)$$

En régime stationnaire, pour un matériau uniformément dopé, à température constante et supposant le terme de génération/recombinaison nul, l'ETB devient :

$$\frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f + \frac{f - f_0}{\tau} = 0 \quad (76)$$

5.1.2. Fonction de distribution maxwellienne déplacée

À champ faible, les porteurs sont supposés proches de l'équilibre $f(\mathbf{k}) = f_0(\mathbf{k} - \mathbf{k}_0)$ où \mathbf{k}_0 est le vecteur d'onde moyen. Dans l'approximation de la masse effective m^* obtenue dans l'hypothèse de relation de dispersion sous forme quadratique, les fonctions de distribution sont des maxwelliennes (fonctions gaussiennes) décalées de \mathbf{k}_0 (figure 45) et dont la largeur augmente avec la température :

$$f_n^0(\mathbf{k}) = e^{\frac{E_n(\mathbf{r}, \mathbf{k}) - E_F}{k_B T}} \quad (77)$$

avec $E_n(\mathbf{r}, \mathbf{k}) = E_c(\mathbf{r}) + \frac{\hbar^2 \mathbf{k}^2}{2m^*}$, où E_c est la bande de conduction et E_F le niveau de Fermi. Cette solution à l'équilibre est proche de la fonction de Fermi-Dirac donnée précédemment.

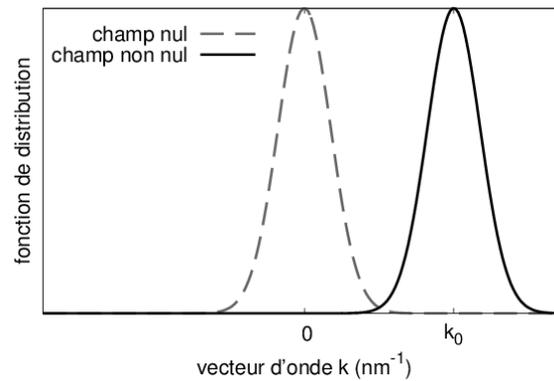


Figure 45 : fonction de distribution à faible champ sous forme de maxwelliennes déplacées selon l'axe de vecteur d'onde.

La vitesse moyenne obtenue par la résolution de l'ETB ainsi simplifiée varie alors linéairement avec le champ électrique appliqué ce qui donne lieu à une loi de mobilité effective, décrite dans le modèle de dérive-diffusion ci-après.

5.2. Modèles classiques

5.2.1. Modèle de dérive-diffusion

Le modèle de dérive-diffusion décrit le courant de porteurs, électrons et trous, traversant le canal du transistor, en utilisant les approximations du temps de relaxation et de la fonction de distribution en maxwellienne déplacée. Le modèle dérive-diffusion représente les deux premiers moments correspondant à l'intégration simple suivant la vitesse de groupe de l'équation de transport de Boltzmann présentée précédemment. Pour résoudre le système à $n+1$ inconnues pour n équations alors obtenu, il faut trouver une relation de fermeture [GRITSCHO2] fondée sur des considérations physiques.

Ce modèle est la résultante de deux composantes principales :

- la dérive, correspondant à la vitesse imposée par le champ électrique et qui dépend linéairement de la charge et de la mobilité des porteurs.

$$\mathbf{v} = \mu_{\text{eff}} \mathbf{E} \quad (78)$$

où

$$\mu_{\text{eff}} = \frac{e\tau}{m^*} \quad (79)$$

- la diffusion, au cours de laquelle les porteurs se dispersent ce qui tend à homogénéiser leur concentration, où ils se déplacent donc globalement des fortes concentrations vers les concentrations inférieures selon l'opposé du gradient, proportionnellement au coefficient de diffusion déduit de la relation d'Einstein.

De plus, les équations de conservation des porteurs, l'équation de Poisson et les relations d'Einstein sont aussi utilisées, ce qui donne finalement le modèle suivant :

$$\left\{ \begin{array}{ll}
 \mathbf{J}_n(\mathbf{r}, t) = en(\mathbf{r}, t)\mu_n\mathbf{E} + eD_n\nabla n(\mathbf{r}, t) & \text{dérive et diffusion des électrons} \\
 \mathbf{J}_p(\mathbf{r}, t) = ep(\mathbf{r}, t)\mu_p\mathbf{E} - eD_p\nabla p(\mathbf{r}, t) & \text{dérive et diffusion des trous} \\
 \frac{\partial n}{\partial t} = G_n(\mathbf{r}, t) + \frac{1}{e}\nabla \cdot \mathbf{J}_n(\mathbf{r}, t) & \text{conservation des électrons} \\
 \frac{\partial p}{\partial t} = G_p(\mathbf{r}, t) + \frac{1}{e}\nabla \cdot \mathbf{J}_p(\mathbf{r}, t) & \text{conservation des trous} \\
 \nabla(\varepsilon_{sc}\mathbf{E}) = e(p(\mathbf{r}, t) - n(\mathbf{r}, t)) + N_D + N_A & \text{équation de Poisson}
 \end{array} \right. \quad (80)$$

où les constantes de diffusion et les mobilités sont données par :

$$\left\{ \begin{array}{ll}
 D_n = \frac{k_B T}{e} \mu_n & \text{relation d'Einstein pour les électrons} \\
 D_p = \frac{k_B T}{e} \mu_p & \text{relation d'Einstein pour les trous} \\
 \mu_n = \frac{e\tau_m}{m^*} & \text{mobilité des électrons} \\
 \mu_p = \frac{e\tau_{m,h}}{m_h^*} & \text{mobilité des trous}
 \end{array} \right. \quad (81)$$

Dans ces équations, \mathbf{J}_n et \mathbf{J}_p sont les densités de courant d'électrons et de trous, n et p les densités volumiques de charge, μ_n et μ_p les mobilités, \mathbf{E} le champ électrique, D_n et D_p les coefficients de diffusion, N_D et N_A les concentrations de dopage en donneurs et accepteurs d'électrons, G_n et G_p les termes de génération/recombinaison. τ_m est le temps de relaxation en moment.

Les courants et charges sont des solutions déterminées rapidement de ce modèle simple, d'où son utilisation courante dans les logiciels commerciaux. La figure 46 illustre les résultats que l'on peut ainsi obtenir. En revanche, ce modèle ne prend pas en compte les phénomènes hors équilibre, car l'énergie est supposée implicitement constante pour aboutir à la relation d'Einstein et égale à :

$$w = \frac{3}{2}k_B T \quad (82)$$

Cette relation la n+1^e équation permettant de trouver une solution au système.

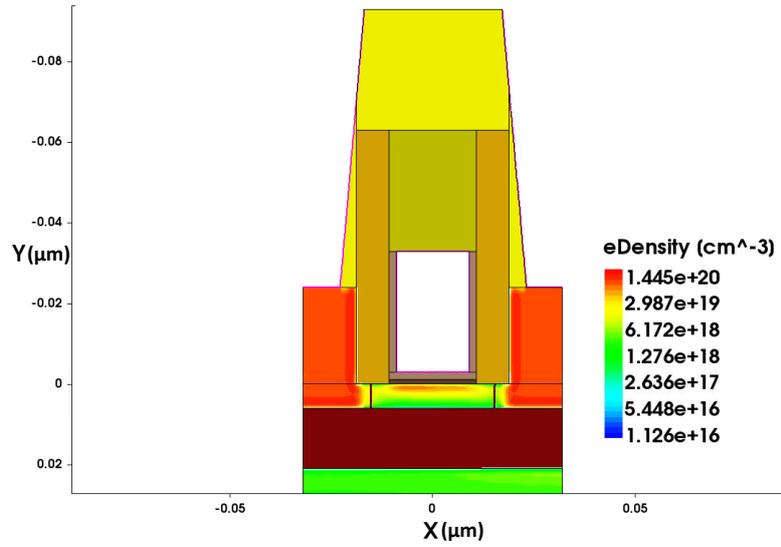


Figure 46 : potentiel dans un FDSOI 14 nm comprenant un empilement de TiN sur poly-Si avec un contact Schottky.
 $L = 20 \text{ nm}$, $N_{\text{poly}} = 10^{18} \text{ cm}^{-3}$, $V_g = 1 \text{ V}$ et $V_{ds} = 50 \text{ mV}$.

5.2.2. Modèle hydrodynamique

Ce modèle utilise l'intégration simple, l'intégration suivant la vitesse de groupe, l'intégration suivant l'énergie de l'ETB et il décrit les phénomènes non stationnaires. Pour cela des termes de relaxation supplémentaires seront définis tels que la relaxation en énergie et en moment. Donc dans ce modèle, la température électronique n'est plus supposée égale à celle du réseau. La relation de fermeture sur l'énergie w est remplacée par une relation sur le flux énergétique S . Il se divise en une partie de dérive liée au courant, et une partie de diffusion liée au flux d'énergie thermique Q :

$$nS = -\frac{5k_B T_n}{2e} J_n + nQ \quad (83)$$

où T_n est la température des électrons, supposée différente de la température du réseau T_L .

La composante de flux d'énergie thermique Q est modélisée par la loi de Fourier dans laquelle la conductivité thermique κ_n est évaluée par la loi de Wiedermann-Franz [BLOTEKJA70].

Les dépendances des fonctions avec l'espace et le temps n'étant dorénavant plus notées, le modèle hydrodynamique se résume alors pour les électrons à :

$$\left\{ \begin{array}{ll} J_n = en\mu_n E + \mu_n k_B \nabla(nT_n) - \tau_m \frac{\partial J_n}{\partial t} & \text{dérive et diffusion des électrons} \\ \frac{\partial n}{\partial t} = G_n + \frac{1}{e} \nabla \cdot J_n & \text{conservation des électrons} \\ \nabla(nS) = -\frac{3k_B}{2} \frac{T_n - T_L}{\tau_\varepsilon} + \frac{3k_B}{2} \frac{\partial(nT_n)}{\partial t} + E \cdot J_n & \text{relation de fermeture sur le flux énergétique} \end{array} \right. \quad (84)$$

où τ_m est le temps de relaxation en moment et τ_ε le temps de relaxation en énergie.

5.2.3. Modèle « énergie balance »

Ce modèle proche du modèle hydrodynamique suppose une relation entre le temps de relaxation et l'énergie d'après la méthode de Stratton [APANOVIC94]. Il diffère uniquement en ce qui concerne l'expression du courant, et le système d'équations pour les électrons est dans ce cas :

$$\left\{ \begin{array}{ll} J_n = en\mu_n E + \mu_n k_B [T_n \nabla(n) + (1 + v_n)n \nabla(T_n)] & \text{dérive et diffusion des électrons} \\ \frac{\partial n}{\partial t} = G_n + \frac{1}{e} \nabla \cdot J_n & \text{conservation des électrons} \\ \nabla(nS) = -\frac{3k_B T_n - T_L}{2} \frac{1}{\tau_\varepsilon} + \frac{3k_B}{2} \frac{\partial(nT_n)}{\partial t} + E \cdot J_n & \text{relation de fermeture sur le flux énergétique} \end{array} \right. \quad (85)$$

où v_n est un terme correctif décrivant la variation de la mobilité avec la température :

$$v_n = \frac{\partial \ln \mu_n}{\partial \ln T_n} = \frac{T_n}{\mu_n} \frac{\partial \mu_n}{\partial T_n} \quad (86)$$

5.3. Modèle avec corrections quantiques

Pour prendre en compte facilement l'effet de la quantification lorsque les dimensions du MOS sont réduites, il est possible de garder une approche analogue aux modèles classiques précédents en remplaçant seulement le potentiel par un autre issu d'une approche quantique.

Le modèle *density gradient* consiste à ajouter au potentiel classique un terme de correction quantique, différent pour les électrons et les trous, avant de l'utiliser dans les équations de transport. Le potentiel résultant vaut alors :

$$V_n = V_{\text{class}} + V_{\text{quant } n} \quad (87)$$

$$V_p = V_{\text{class}} + V_{\text{quant } p} \quad (88)$$

où

$$V_{\text{quant } n} = \alpha_n \frac{\Delta(\sqrt{n})}{\sqrt{n}} \quad (89)$$

$$V_{\text{quant } p} = \alpha_p \frac{\Delta(\sqrt{n})}{\sqrt{n}} \quad (90)$$

Pour évaluer les paramètres α , une approche simple [WETTSTEI01] parmi d'autres consiste à appliquer la méthode des moments à la méthode de Wigner :

$$\alpha_n = -\frac{\gamma \hbar^2}{6m_{\text{DOS}}} \quad (91)$$

Une seule vallée peut donc être prise en compte par cette approche.

6. Conclusion du chapitre

Dans ce chapitre, nous avons exposé les différentes approches modélisant le transport électronique, à savoir les approches de type stochastique comme la simulation Monte-Carlo ou continue comme celle utilisant le modèle dévié de la méthode des moments (modèle de dérive-diffusion, hydrodynamique...). De son côté l'approche Kubo-Greenwood résout uniquement

l'équation de transport de Boltzmann en régime linéaire. Elle permet simplement d'étudier le transport à faible champ longitudinal. Nous avons aussi présenté les différentes façons de modéliser le confinement quantique, que ce soit à partir d'une approche par résolution des équations de Poisson-Schrödinger, ou par la modèle *density gradient* dans le cas des approches de type dérive-diffusion.

Les approches de type dérive-diffusion, largement utilisées dans les simulateurs TCAD, permettent de bien prendre en compte de façon comportementale les effets du transport électronique. L'approche hydrodynamique est une amélioration permettant de prendre en compte une pseudo-survitesse dans le canal (par exemple pour reproduire les effets d'avalanche). Néanmoins les approches de type TCAD se fondent sur la résolution de la méthode des moments et seront dans tous les cas limités intrinsèquement par l'approximation du temps de relaxation qui introduit un nombre important d'hypothèses sur les quantités macroscopiques (vitesse moyenne, énergie moyenne...). Il est alors extrêmement difficile pour ces approches d'être prédictives sans un gros travail de calibration sur les résultats expérimentaux ou des approches plus fondamentales.

L'objectif de cette thèse, comme expliqué précédemment, est l'étude fine du transport électronique pour les longueurs ultimes. Il nous faut conjointement étudier l'impact de phénomènes d'interactions dues aux champs transverse et longitudinal. Or actuellement seules les approches de type Monte-Carlo ou NEGF (*non equilibrium Green functions*) décrivent correctement les différents termes d'interaction en introduisant l'effet du champ transverse. De plus, de par sa nature stochastique, cette approche est particulièrement intéressante pour les cas ayant peu d'interactions [SAINT-MA05]. Si nous prenons l'exemple concret du transistor MOS par une approche de type dérive-diffusion, la notion de temps de relaxation est erronée dans le cas où les interactions sont faibles. Certes nous pouvons reproduire correctement les effets mis en jeu, mais cette démarche ne peut pas être prédictive. Dans le cas d'une approche de type Monte-Carlo et en supposant que les interactions mises en jeu sont bien décrites, nous pouvons étudier l'impact de variation de mobilité du canal et de l'effet de saturation des porteurs (en supposant que le temps de simulation n'est pas une limitation). Nous utiliserons pleinement la possibilité de cette approche tout au long du chapitre IV.

Chapitre III. Méthodes d'extraction de la mobilité

Dans le chapitre précédent nous avons présenté les fondamentaux de la théorie du transport. Nous allons maintenant présenter les différentes méthodes d'extraction permettant d'obtenir la mobilité à champ faible et les caractéristiques de transport telles que la résistance d'accès et la mobilité en fonction de la charge d'inversion. Pour chacune de ces méthodes, nous expliciterons les hypothèses de base qui permettent de construire le modèle théorique, puis nous expliciterons en détail la technique d'extraction. Nous ferons aussi une analyse critique de l'utilisation de ces méthodes.

1. Théorie sur les méthodes d'extraction

Pour pouvoir extraire des paramètres, nous avons besoin *a fortiori* d'un modèle simple décrivant la physique du dispositif qui sera caractérisé. Malheureusement la plupart du temps il est impossible d'avoir une bonne description physique avec une extraction de paramètre efficace. Les méthodes d'extraction sont nombreuses et diffèrent tout d'abord par leur modèle de transport. Le modèle en courant peut utiliser une approche de type dérive-diffusion, balistique ou quasi-balistique et aura donc une grande influence sur le résultat [BARRAL08]. Néanmoins nous exposerons ici uniquement les principales méthodes qui seront utilisées dans le chapitre IV.

Les méthodes d'extraction reposent sur l'hypothèse d'une expression du courant dépendant d'un terme de mobilité considéré comme paramètre suffisant pour décrire le transport dans le dispositif. L'utilisation de ce seul paramètre est contestée pour plusieurs raisons, dont le fait qu'elle ne prend pas naturellement en compte les effets non stationnaires dont va dépendre le courant. Les approches seront théoriquement plus simples que celles développées dans le chapitre précédent mais sont plus à même d'être utilisées sur des données expérimentales.

1.1. Méthode de Hamer

Comme expliqué dans les chapitres précédents, pour calculer la mobilité qui est inversement liée à la résistivité du canal due aux différents termes d'interaction, il faut se placer dans le régime ohmique (à faible tension de drain). Le canal du transistor se comporte alors comme une résistance quasi bidimensionnelle contrôlée par la tension. La méthode de Hamer est une méthode d'extraction présupposant l'expression du courant la plus simple en régime ohmique [HAMER86] :

$$I_d(V_{gs}) = \frac{WC_{ox}}{L_{eff}} \mu_{eff}(V_{gs} - V_t) V_{ds} \quad (92)$$

où la mobilité effective peut contenir différents ordres de termes de dégradation θ , de telle sorte que :

$$\mu_{eff}(V_{gs}) = \frac{\mu_0}{1 + \theta_{eff}(V_{gs} - V_t)} \quad (93)$$

La fonction θ_{eff} peut elle-même dépendre de V_{gs} selon le modèle de dégradation choisi.

Au second ordre, la fonction θ_{eff} vaut $\theta_1 + \theta_2(V_{gs} - V_t)$ où θ_1 et θ_2 sont les paramètres de dégradation de la mobilité.

$$I_d(V_{gs}) = \frac{\mu_0 C_{ox} W}{L_{eff}} \frac{(V_{gs} - V_t) V_{ds}}{1 + \theta_1(V_{gs} - V_t) + \theta_2(V_{gs} - V_t)^2} \quad (94)$$

En pratique, la dégradation de mobilité peut souvent être limitée au terme de premier ordre θ_1 . La signification empirique des termes θ_1 et θ_2 est directement liée aux mécanismes d'interaction considérés : le premier est couramment considéré comme celui représentant les interactions avec les

phonons et l'impact de la résistance d'accès, alors que le second est lié aux interactions de type rugosité de surface [ROMANJEK04].

1.2. Méthode de la fonction Y

La méthode par fonction Y est censée être plus précise que la méthode de Hamer, grâce à la réduction du problème d'ajustement de paramètres, décomposé en plusieurs extractions plus simples et surtout plus fiables [GHIBAUDO88].

1.2.1. Principe de la méthode par fonction Y

Comme pour la méthode de Hamer, la fonction Y se fonde sur le courant en régime ohmique de type dérive-diffusion. Néanmoins celle-ci introduit astucieusement une fonction qui permet d'isoler les termes de transport. La fonction Y est définie par le rapport du courant sur la racine carrée de la transconductance :

$$Y(V_{gs}) = \frac{I_d(V_{gs})}{\sqrt{g_m(V_{gs})}} \quad (95)$$

Si l'expression en courant présuppose un unique terme de dégradation ($\theta_2 = 0$), nous obtenons :

$$Y(V_{gs}) = \sqrt{\frac{C_{ox}W}{L_{eff}}} \mu_0 V_{ds} (V_{gs} - V_t) \quad (96)$$

Expérimentalement c'est un puissant moyen pour extraire à l'aide de la pente de cette courbe la valeur de la mobilité à champ faible, en effectuant seulement une opération sur le courant et sa dérivée. Il nous reste néanmoins à déterminer le terme de dégradation, qui est censé être constant, à partir de l'équation suivante :

$$\theta(V_{gs}) = \left(\frac{I_d}{(V_{gs} - V_t)g_m} - 1 \right) \frac{1}{V_{gs} - V_t} \quad (97)$$

Même si cette méthode peut paraître simpliste, elle permet de trouver la valeur de la résistance d'accès en extrayant la pente de θ en fonction de β_0 .

1.2.2. Utilisation de la méthode par fonction Y simple

Nous présentons ici une utilisation de la fonction Y pour extraire les paramètres de transport. La détermination de la zone où Y est une fonction affine de V_{gs} peut être délicate. La figure 47 fournit un exemple d'extraction de μ_0 et V_t réalisée sur les points de la fonction Y obtenus par simulation Monte-Carlo.

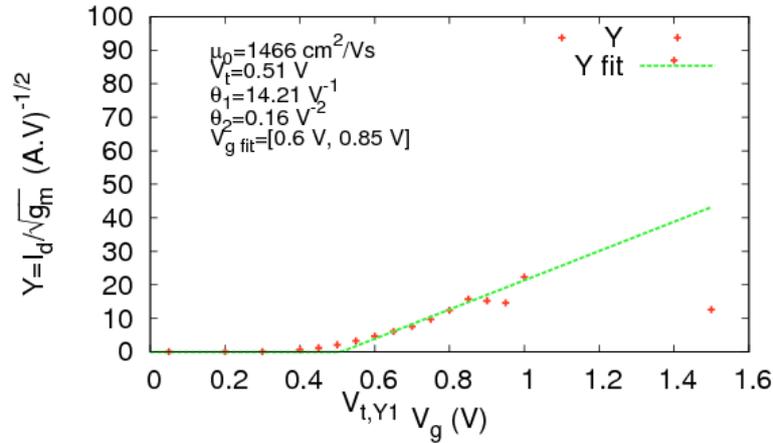


Figure 47 : ajustement des paramètres μ_0 et V_t sur la fonction Y dans l'exemple d'une simulation Monte-Carlo. Plage d'extraction : de 0,6 à 0,85 V. $L = 80$ nm et $V_{ds} = 50$ mV. En rouge, les points de mesure. En vert, la fonction affine ajustée.

Pour vérifier la validité de l'hypothèse de θ_2 faible, il est possible d'extraire les paramètres θ_1 et θ_2 , toujours par un ajustement, cette fois-ci sur la fonction θ_{eff} définie ci-dessous :

$$\theta_{\text{eff}} = \frac{\mu_0 W C_{\text{ox}} V_{ds}}{L I_d} - \frac{1}{V_{gs} - V_t} \quad (98)$$

En effet, en injectant l'expression du courant dans celle-ci, cette fonction peut s'écrire :

$$\theta_{\text{eff}} = \theta_1 + \theta_2 (V_{gs} - V_t) \quad (99)$$

Sur la figure 48, une telle extraction des paramètres θ_1 et θ_2 est représentée. Le paramètre θ_1 correspond à la valeur de la fonction au seuil et θ_2 à la pente.

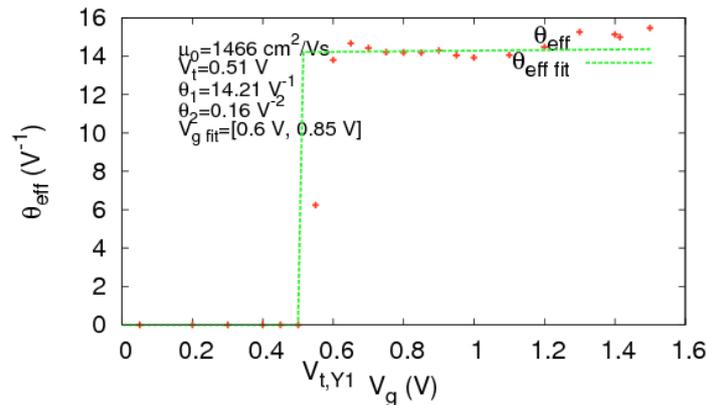


Figure 48 : ajustement de paramètres θ_1 et θ_2 sur la fonction θ_{eff} . Plage d'extraction : de 0,6 à 0,85 V. $L = 80$ nm et $V_{ds} = 50$ mV.

En première approximation, la fonction θ_{eff} se réduit bien à la constante θ_1 et le paramètre θ_2 est négligeable. Il est censé être positif dans le modèle simple décrit ici.

Finalement, ces deux extractions de paramètres, μ_0 et V_t sur la fonction Y , puis θ_1 et θ_2 sur la fonction θ_{eff} , permettent de revenir à une expression du courant obtenu par ces ajustements. La

figure 49 permet de comparer les variations du courant avec V_{gs} ainsi obtenues aux résultats de simulation Monte-Carlo.

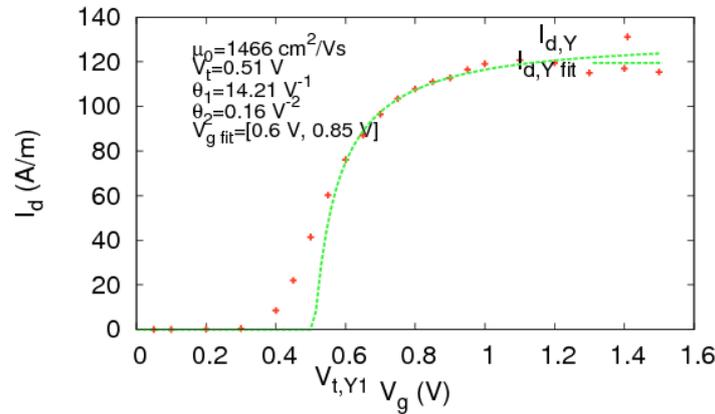


Figure 49 : courant obtenu après les ajustements de paramètres précédents μ_0 et V_t sur la fonction Y , et θ_1 et θ_2 sur la fonction θ_{eff} . Plage d'extraction : de 0,6 à 0,85 V. $L = 80 \text{ nm}$ et $V_{ds} = 50 \text{ mV}$.

Les courants simulés et extraits se superposent bien dans la région au-dessus du seuil et à V_{gs} pas trop grand. Ce critère n'est en revanche pas suffisant pour valider l'extraction car une bonne superposition peut être obtenue par compensation des différents paramètres, elle est alors trompeuse.

1.2.3. Principe de la méthode par fonction Y améliorée

Supposons maintenant que le modèle de courant défini par une expression de la dégradation de la mobilité au deuxième ordre est valide. Pour extraire précisément la mobilité μ_0 , il est préférable de ne pas l'extraire en même temps que le paramètre de dégradation θ_1 , pour éviter le phénomène de compensation entre ces deux paramètres. Dans la méthode par fonction Y , la fonction de courant sur laquelle l'extraction est réalisée est remplacée par une autre fonction, indépendante du paramètre θ_1 . On vérifie facilement qu'au premier ordre, cette fonction est indépendante de θ_1 :

$$Y(V_{gs}) = \sqrt{\frac{C_{\text{ox}}W}{L_{\text{eff}}} \frac{\mu_0}{1 - \theta_2(V_{gs} - V_t)^2} V_{ds}(V_{gs} - V_t)} \quad (100)$$

Le fait d'avoir un ordre de dégradation supplémentaire permet de mieux calibrer la courbe de courant sans pour autant avoir de compensation de paramètres. Donc pour prendre en compte le terme de dégradation au deuxième ordre, la méthode de la fonction Y doit être corrigée.

$$Y_{\text{new}} = Y \sqrt{1 - \theta_2(V_{gs} - V_t)^2} \quad (101)$$

La même série d'extractions peut être effectuée sur cette nouvelle fonction Y qui est en principe plus proche d'une fonction affine que la précédente. En supposant que le paramètre θ_2 est représentatif des résistances d'accès, la procédure d'extraction par fonction Y améliorée peut se résumer au schéma de la figure 50.

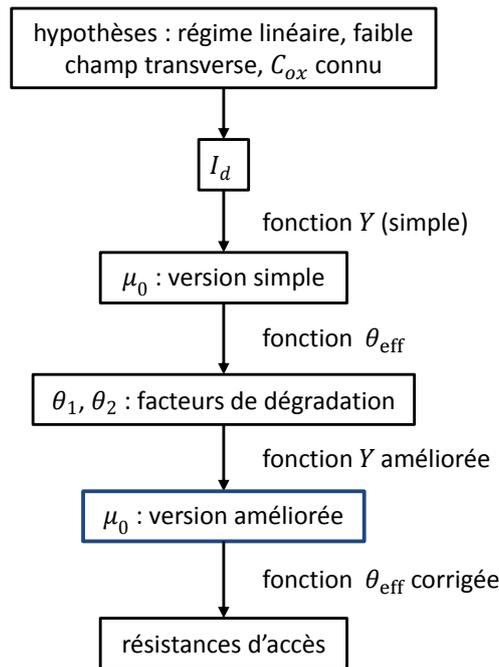


Figure 50 : extraction de la mobilité par fonction Y améliorée

En pratique, cette procédure d'extraction modifie légèrement les valeurs de mobilité μ_0 extraites sur les valeurs expérimentales pour lesquelles $\theta_2(V_{gs} - V_t)$ n'est pas négligeable devant θ_1 sur les plages d'extraction en V_{gs} considérées.

Sur les autres expériences, cette méthode ne s'avère pas très utile, de même pour les simulations Monte-Carlo, où de plus ce terme est très bruité.

1.2.4. Utilisation de la méthode par fonction Y améliorée

Cette méthode présente quelques inconvénients. Elle présuppose de connaître parfaitement la largeur et la longueur du dispositif (ce qui est un problème récurrent des méthodes d'extraction). Elle demande aussi de très bien connaître la capacité C_{ox} , mais cela peut se faire grâce à des structures de test dédiées avec de grandes surfaces, couplées à des simulations numériques de type Poisson-Schrödinger.

Enfin, le plus gros inconvénient de cette approche et qui a déjà été évoqué est sa forte dépendance à l'intervalle d'extraction. Cela sera problématique pour les dispositifs ayant une inversion modérée très élargie, c'est-à-dire les plus petits ou ceux marqués par des effets coulombiens importants (charges piégées dans l'oxyde...).

1.3. Extraction de la résistance d'accès

Différentes méthodes [FLEURY09] existent : la méthode classique $R_{tot}(L)$ et la méthode améliorée en $R_{tot}(\beta_0)$ censée s'affranchir de l'indétermination de la longueur effective L_{eff} .

Dans le modèle prenant en compte la résistance d'accès R_{sd} , la chute de potentiel V_{ds} est ôtée de la tension $R_{sd}I_d$, ce qui donne :

$$I_d(V_{gs}) = \frac{\mu_0 C_{ox} W}{L_{eff}} \frac{(V_{gs} - V_t)(V_{ds} - R_{sd}I_d)}{1 + \theta_1(V_{gs} - V_t) + \theta_2(V_{gs} - V_t)^2} \quad (102)$$

1.3.1. Méthode $R_{tot}(L)$

Cette méthode repose sur une évaluation simple de la résistance totale des dispositifs pour différentes longueurs de canal dans le but de décorrélérer les contributions de résistance d'accès R_0 de la résistance de canal R_{canal} :

$$R_{tot}(L) = \frac{V_{ds}}{I_d} = R_0 + R_{canal}(L) \quad (103)$$

L'équation étudiée est :

$$R_{tot} = \frac{V_{ds}}{I_d} = \frac{L_{eff}}{\mu_0 C_{ox} W} \frac{1 + \theta_1(V_{gs} - V_t) + \theta_2(V_{gs} - V_t)^2}{(V_{gs} - V_t)} + R_0 \quad (104)$$

Un exemple d'une telle extraction est donné sur la figure 51.

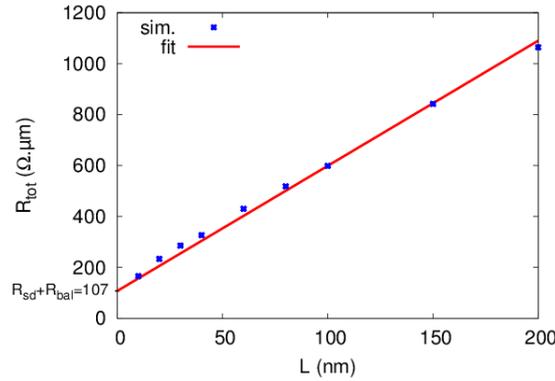


Figure 51 : exemple de courbe $R_{tot}(L)$ (en bleu) avec ajustement de la droite (en rouge) pour l'extraction du paramètre $R_{sd} + R_{bal}$ à l'origine. $V_{gs} = 0,8$ V.

Dans cet exemple, on obtient $R_{sd} + R_{bal} = 107 \Omega \cdot \mu\text{m}$. En réalisant la même extraction pour différentes valeurs de N_{inv} , on obtient la fonction $R_{sd}(N_{inv}) + R_{bal}(N_{inv})$ sur la figure 52.

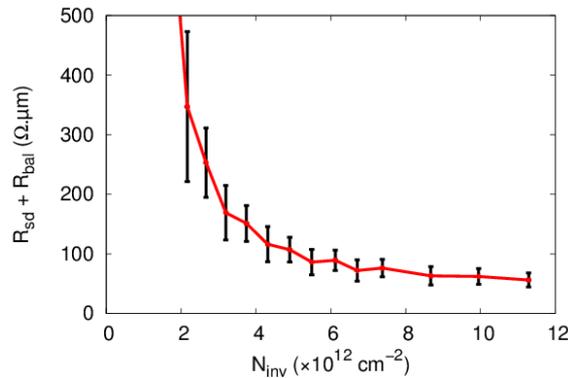


Figure 52 : $R_{sd} + R_{bal}$ en fonction de N_{inv} (en rouge) et écart-type (en noir). Chaque valeur est obtenue par méthode $R_{tot}(L)$ à N_{inv} fixé.

Pour simplifier cette extraction, nous avons ici supposé que N_{inv} ne dépend que de V_{gs} et non pas de L . Par exemple, $V_{gs} = 0,8 \text{ V}$ correspond à $N_{inv} = 4,9.10^{12} \text{ cm}^{-2}$.

Nous présenterons une méthode qui présente des similitudes dans le chapitre suivant et qui à la différence de celle-ci utilise une extraction plus précise de la mobilité et donc de la résistance d'accès.

1.3.2. Méthode $R_{tot}(\beta_0)$

L'équation étudiée est :

$$R_{tot} = \frac{V_{ds}}{I_d} = \frac{1}{\beta_0} \frac{1 + \theta_1(V_{gs} - V_t) + \theta_2(V_{gs} - V_t)^2}{(V_{gs} - V_t)} + R_{sd} \quad (105)$$

1.4. Méthode split-CV

Les méthodes d'extraction précédentes se fondaient uniquement sur la mesure du courant. La méthode dite « split-CV » demande préalablement d'avoir une mesure de la capacité en plus de la mesure du courant. Cela nous permettra d'avoir directement la valeur de charge d'inversion et donc celle de mobilité en fonction de cette dernière à l'aide de l'équation classique :

$$\mu = \frac{L}{W} \left(\frac{I_d}{Q_{inv} V_{ds}} \right) \quad (106)$$

où Q_{inv} est la charge d'inversion directement calculée en fonction de la capacité mesurée :

$$Q_{inv} = \frac{1}{WL} \int_0^{V_g} C_{gs}(V_g) dV_g \quad (107)$$

Cette méthode est particulièrement précise et permet d'avoir une image claire des termes d'interaction relatifs aux phonons et à la rugosité de surface.

Néanmoins cette méthode souffre des mêmes inconvénients que la fonction Y , elle présuppose de connaître parfaitement la largeur et la longueur du dispositif. Elle est aussi fortement perturbée par le ou les termes d'interaction coulombienne. Elle reste cependant très efficace pour obtenir la mobilité dans un canal long et constitue une méthode standard en analyse des lots pour l'optimisation du transport électronique.

2. Extraction sur les simulations Monte-Carlo

2.1. Extraction sur le courant seul

Pour réaliser des extractions rapides de mobilité uniquement sur la courbe de courant, et non sur celle de la charge, la méthode de Hamer semble la plus adaptée car elle est plus robuste. Il est vrai que les simulations Monte-Carlo sont fortement bruitées, ce qui rend plus difficile l'utilisation de méthodes faisant appel à la dérivée du courant comme la fonction Y .

De plus, elle est moins dépendante que la méthode par fonction Y des points erronés dus à la présence de fluctuations statistiques inhérentes à la méthode Monte-Carlo ou au choix d'un intervalle d'extraction incluant des valeurs de tension trop proches du seuil en-dessous duquel le modèle de dégradation de mobilité au premier ordre n'est plus valide. Cela a été vérifié en pratique : les extractions par fonction Y avec quelques points supplémentaires (une vingtaine en général pris en compte) changent notablement la valeur de la mobilité extraite, avec une imprécision de l'ordre du dixième, comme on peut le voir en comparant les figures 53 et 54.

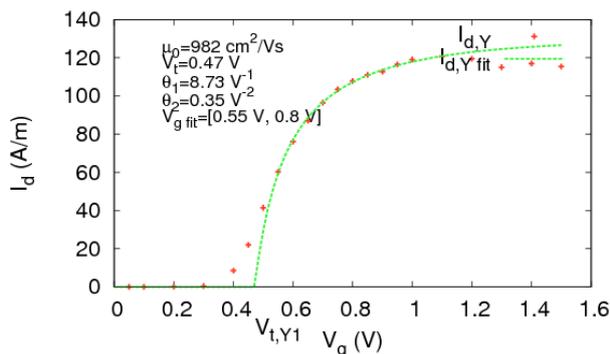


Figure 53 : courant et ajustement par fonction Y de μ_0 , θ_1 , θ_2 et V_t . Plage d'extraction : V_g allant de 0,55 à 0,8 V. $L = 80$ nm et $V_{ds} = 50$ mV.

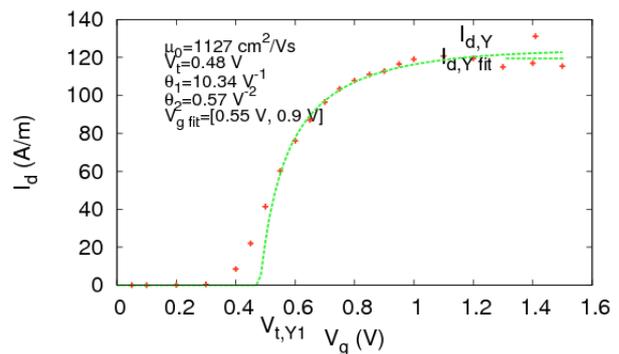


Figure 54 : courant et ajustement par fonction Y de μ_0 , θ_1 , θ_2 et V_t . Plage d'extraction : V_g allant de 0,55 à 0,9 V. $L = 80$ nm et $V_{ds} = 50$ mV.

Dans le premier exemple, la mobilité extraite vaut $\mu_0 = 982$ cm²/Vs, alors que dans le deuxième exemple, $\mu_0 = 1127$ cm²/Vs. Les paramètres θ_1 et θ_2 sont tout autant variables. Cette grande différence s'explique pourtant par la modification de la plage d'extraction de V_g allant de 0,55 à 0,8 V dans le premier cas à V_g allant de 0,55 à 0,9 V dans le deuxième cas. Or il n'existe pas de règle unique de détermination d'une zone linéaire de la fonction Y , qui est laissée à l'appréciation de chacun.

Visiblement, il est difficile d'évaluer la pertinence d'une plage d'extraction. Comme elle ne suppose pas de linéarité d'une fonction annexe comme la méthode par fonction Y , la méthode de Hamer est utilisée sur une plage d'extraction plus large en pratique, comme on peut le voir sur la figure 55.

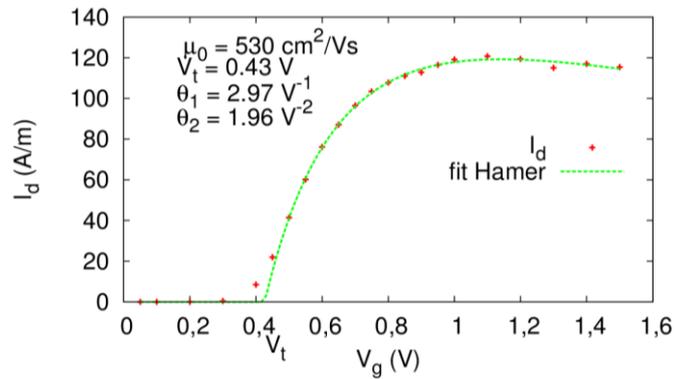


Figure 55 : courant et ajustement par la méthode de Hamer de μ_0 , θ_1 , θ_2 et V_t . $L = 80 \text{ nm}$ et $V_{ds} = 50 \text{ mV}$.

La plage d'extraction comprenant aussi l'inversion faible, la mobilité extraite est nettement plus faible que précédemment, de l'ordre de la moitié sur cet exemple choisi pour mettre en évidence ce phénomène, tout comme l'est le terme principal de dégradation, θ_1 .

La différence entre les deux méthodes d'extraction ne tient pas tant aux formules calculées qu'au fait que la méthode Y amène naturellement à considérer une hypothèse plus forte que la méthode de Hamer, à savoir sa linéarité sur une certaine zone, sur laquelle l'extraction doit alors être restreinte.

2.2. Extraction sur le courant et la charge

Les simulations Monte-Carlo permettant d'avoir accès à toutes les données physiques du problème de transport modélisé, il est possible d'extraire la mobilité en utilisant directement la charge simulée plutôt que de supposer un modèle de charge comme c'est le cas dans la méthode par split-CV.

La concentration en électrons dans le canal non dopé est logiquement inférieure de plusieurs ordres de grandeurs à celle dans les accès source et drain. À forte polarisation de grille (et faible polarisation de drain), la concentration selon la direction y perpendiculaire au transport est plus importante du côté de l'oxyde sous la grille.

Sur une coupe suivant l'axe x dans la longueur du transistor, la figure 56 présente la concentration surfacique par sous-bande. La bande d'énergie la plus basse est naturellement bien plus occupée que les autres et ce quelle que soit la zone de silicium considérée. La deuxième sous-bande de la vallée Δ_y et la première sous-bande d'une vallée Δ_x présentent des profils de concentration analogues et doivent être prises en compte pour analyser précisément la concentration totale. Les sous-bandes suivantes apparaissent clairement comme négligeables.

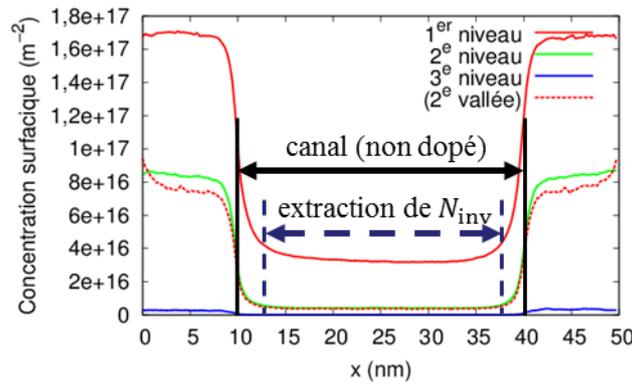


Figure 56 : concentration surfacique par sous-bande. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm. Représentation des zones d'extraction : canal complet en noir, zone réduite en tirets bleus.

La charge intégrée sur le canal peut alors être extraite et tracée en fonction de la tension de grille V_g , sur la figure 57. L'intervalle d'intégration peut être réduit pour diminuer l'influence des accès source et drain fortement dopés.

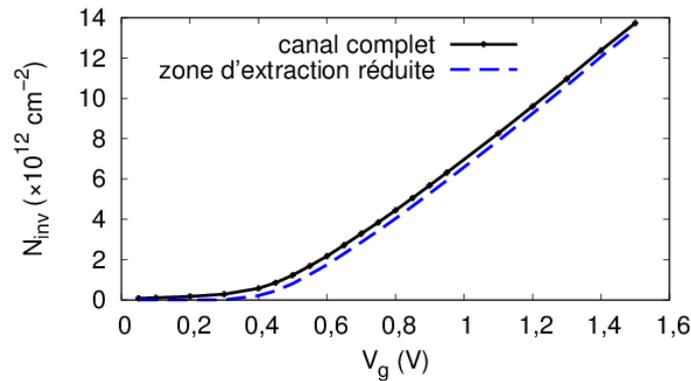


Figure 57 : Charge en fonction de la tension de grille V_g intégrée sur les concentrations surfaciques précédentes. $L = 60$ nm. En trait noir, extraction sur le canal complet. En tirets bleus, extraction sur une zone réduite.

La zone d'extraction a une influence minimale sur la courbe de charge en fonction de V_g en valeur absolue. Cependant, en proportion, elle a une influence à faible charge N_{inv} (obtenue à faible V_g).

Un intérêt des simulations Monte-Carlo est donc une extraction précise de la charge, qui permet alors d'en déduire la mobilité effective de manière directe, selon la loi suivante :

$$\mu_{\text{eff}} = \frac{L}{eW} \frac{I_d}{N_{\text{inv}} V_d} \quad (108)$$

Par conséquent, en réduisant la zone d'extraction de la charge, la diminution de la mobilité extraite quand N_{inv} diminue peut être atténuée.

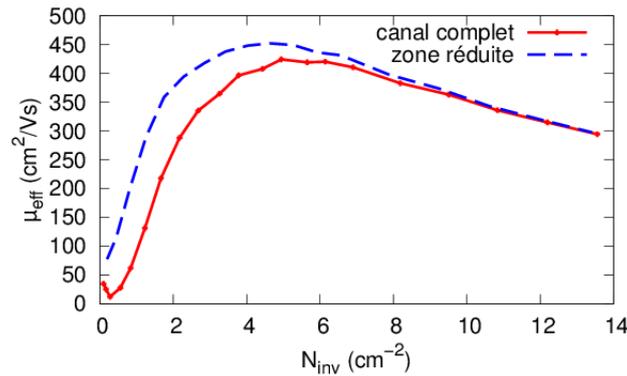


Figure 58 : mobilité effective extraite sur le courant et la charge. $L = 60$ nm.

De plus, cette amélioration de la forme de la courbe de mobilité, en cohérence avec les tendances observées par la méthode de calcul de Kubo-Greenwood, ne nécessite pas d'hypothèse de comportement idéal de la courbe de charge qui suivrait une loi affine déterminée par une capacité effective C_{ox} , comme c'est le cas des méthodes précédentes d'extraction sur le courant seul, de Hamer et par fonction Y :

$$N_{inv} = C_{ox}(V_g - V_t) \quad (109)$$

La méthode d'extraction sur le courant et la charge est donc plus précise. Il resterait à déterminer précisément l'intervalle d'extraction réduite la plus légitime, mais ce n'est pas le propos ici puisque nous ne nous intéressons pas spécifiquement aux valeurs de charge d'inversion les plus faibles.

3. Extraction sur données expérimentales

La fonction Y est adaptée à la caractérisation expérimentale car elle permet de choisir visuellement une partie linéaire. Cependant, cette méthode présente une sensibilité à la plage d'extraction.

3.1. Choix de la méthode

Pour réaliser des extractions de mobilité uniquement sur des courbes de courant obtenues sur des dispositifs réels, et non sur leurs courbes de charge, la méthode par fonction Y est cette fois-ci plus adaptée. Si l'extraction de la capacité le permet, la mobilité peut aussi être extraite par split-CV.

Les données expérimentales étudiées dans ce mémoire proviennent de notre partenariat avec IBM à Albany, elles concernent des FDSOI d'épaisseur de film de 6 nm et de BOX de 20 nm. Nous avons accès à un grand nombre de résultats expérimentaux qui seront analysés dans le chapitre suivant.

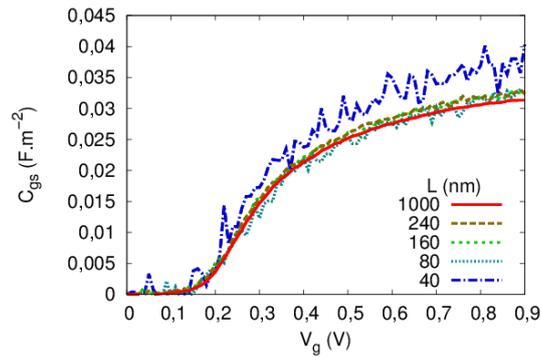


Figure 59 : courbes de capacités d'oxyde C_{ox} avec suppression de l'offset dans des FDSOI fabriqués à Albany.

La figure 59 représente la capacité d'oxyde $C_{ox}(V_g)$. Elle permet d'en déduire la charge d'inversion, cf. figure 60.

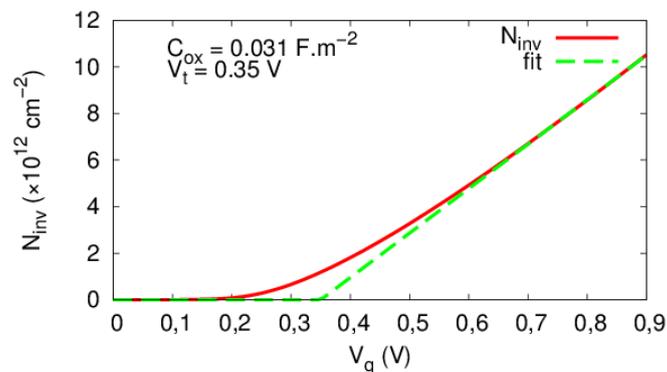


Figure 60 : charge d'inversion extraite par split-CV sur les dispositifs d'Albany.

Lorsque les effets de canal court affectent la charge, il peut être intéressant de ne garder que l'information de la capacité d'oxyde extraite à fort V_{gs} . Cependant, sur les résultats présentés ici, la charge d'inversion a pu être exploitée sans effectuer une telle procédure.

En revanche, la capacité a été extraite à fort W , ce qui estompe de manière indésirable les effets de canal court. La charge extraite a donc été corrigée par le décalage de V_t du courant, en supposant que ce décalage est identique pour la charge, ce qui affecte la mobilité autour du seuil.

3.2. Conclusion du chapitre

Dans cette partie, nous avons détaillé l'ensemble des méthodes d'extraction permettant de connaître au mieux la mobilité que ce soit à champ faible ou en fonction de la charge d'inversion. Il a été démontré que la méthode de Hamer était la plus pertinente pour les approches Monte-Carlo du fait du bruit inhérent à ce type de simulation.

Nous avons pu éprouver ces méthodes par l'extraction de la mobilité sur des simulations Monte-Carlo et sur des résultats expérimentaux. La plus grande difficulté réside dans le choix de la plage d'extraction, particulièrement pour la fonction Y .

Pour déterminer l'intervalle de tension V_{gs} sur lequel est effectuée l'extraction, il faut choisir une zone où la fonction Y présente une forme linéaire, ce qui est un critère subjectif pour lequel il n'y a pas de consensus. Néanmoins, une procédure d'extraction est déterminée en considérant que le régime de courant linéaire est atteint à partir de la valeur de tension pour laquelle le maximum de gain en courant est obtenu, à laquelle est ajoutée la tension correspondant à l'énergie thermique.

Par conséquent, comme il a été vu sur les données issues des simulations, le choix de la plage d'extraction modifie quantitativement les valeurs de mobilités extraites. Mais les tendances relatives sont préservées. Finalement, ce sont elles qui seront comparées dans la prochaine partie, de même que les extractions obtenues par méthode split-CV.

Chapitre IV. Analyse de la mobilité à canal court par simulation Monte-Carlo

Nous allons désormais appliquer les méthodes décrites dans le chapitre précédent pour extraire de caractéristiques de transistors simulés par MONACO l'évolution de la mobilité en fonction de la longueur de grille. L'analyse des résultats va nous conduire à proposer un modèle simple du type de celui proposé par Shur pour reproduire l'évolution de $\mu_{\text{eff}}(L)$. Un bon accord est en effet obtenu en calculant plus précisément la résistance d'un canal balistique et en prenant en compte des résistances d'accès réalistes au dispositif. Enfin ce modèle sera appliqué sur des données expérimentales provenant du partenariat avec IBM et ses limites seront discutées.

1. Grandeurs microscopiques extraites

L'extraction de mobilité requiert une bonne connaissance préalable des grandeurs microscopiques extraites directement. Le problème de l'évaluation de la charge dans le canal a déjà été discuté dans le chapitre précédent. Le taux de porteurs balistiques, permet par ailleurs d'identifier le régime de transport, diffusif ou (quasi-)ballistique. Nous reviendrons dans un premier temps sur l'énergie potentielle et les fonctions enveloppes qui illustrent le confinement quantique dans les couches ultra-minces considérées.

1.1. Énergie potentielle

La prise en compte du confinement quantique engendre une diminution de l'énergie près de l'oxyde et du BOX, que ce soit dans le canal ou dans les accès, ce qui peut être vu, dans l'exemple d'une forte polarisation de grille et de faible polarisation de drain, sur la courbe d'énergie potentielle dans le canal et les accès, source et drain.

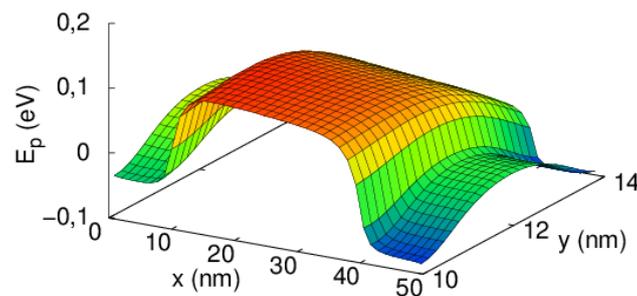


Figure 61 : énergie potentielle dans le silicium. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.

Cette courbure de l'énergie potentielle est plus facile à visualiser sur une coupe selon le plan transverse au transport dans le canal. Dans l'épaisseur des oxydes, l'énergie potentielle varie linéairement avec la distance y puisqu'aucune charge qui pourrait modifier le champ électrique n'est supposée présente dans les simulations effectuées.

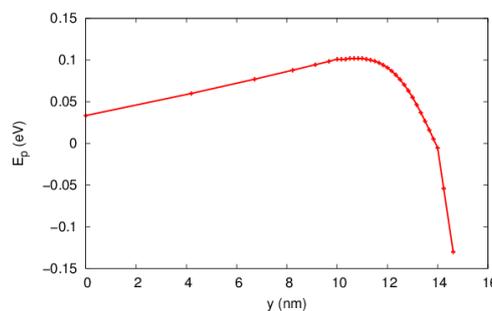


Figure 62 : coupe d'énergie potentielle au milieu du canal, dans le BOX, le canal et l'oxyde. $L = 30$ nm, $t_{box} = 4$ nm, $t_{Si} = 4$ nm, $t_{ox} = 10$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

L'abaissement de la barrière de potentiel par la polarisation de grille permet de créer le canal traversé par les charges contribuant au courant. Ces charges sont déterminées par la superposition des probabilités de présence déduites des fonctions d'onde dans le canal.

Pour ce qui est des variations suivant l'axe source-drain, les variations suivant x des niveaux d'énergie quantifiés pour les sous-bandes les plus basses sont reportées sur la figure 63 pour le même dispositif que précédemment.

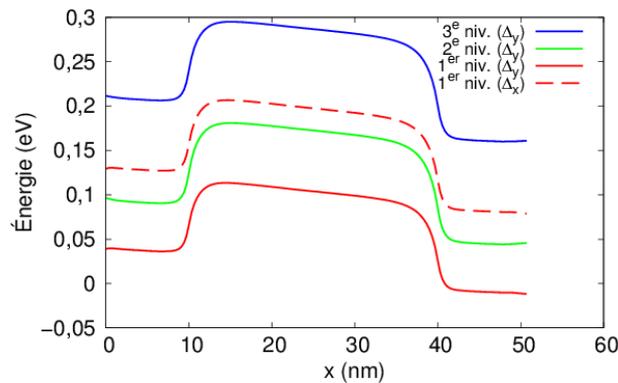


Figure 63 : Évolution suivant l'axe source-drain (x) des niveaux quantifiés des 4 sous-bandes les plus basses. $L = 30$ nm, $t_{\text{box}} = 4$ nm, $t_{\text{Si}} = 4$ nm, $t_{\text{ox}} = 10$ nm, $V_{g_s} = 0,8$ V et $V_{d_s} = 50$ mV.

On remarque que le champ accélérateur est quasi uniforme dans la zone du canal et est de l'ordre de $10 \text{ kV} \cdot \text{cm}^{-1}$. Le champ est également non nul dans les zones fortement dopées d'accès au canal.

1.2. Fonctions d'onde

Les fonctions d'onde sont résolues par l'équation de Schrödinger couplée aux équations de Poisson et de Boltzmann. La densité de probabilité, définie par le module de la fonction d'onde au carré, est représentée sur la figure 64 dans le cas du premier niveau d'énergie de la vallée Δ_y .

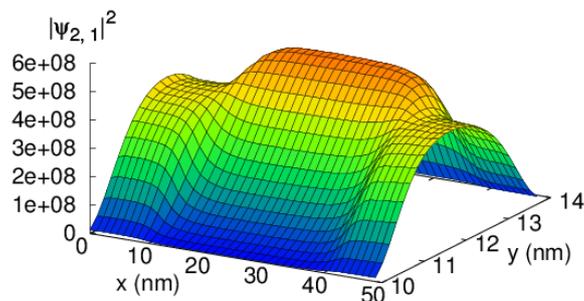


Figure 64 : densité de probabilité dans le silicium. Vallée Δ_y , premier niveau. $L = 30$ nm, $t_{\text{Si}} = 4$ nm, $V_{g_s} = 0,8$ V et $V_{d_s} = 50$ mV. $L_s = L_d = 10$ nm.

Cette fonction d'onde est par hypothèse nulle aux limites que sont ici les interfaces avec l'oxyde d'un côté et le BOX de l'autre, elle est donc variable selon y , plus que ne l'est la concentration en électrons. Les autres sous-bandes, telle que celle représentée en figure 65, peuvent être vues comme des harmoniques donnant une résultante plus uniforme correspondant à une concentration en électrons variant peu avec y .

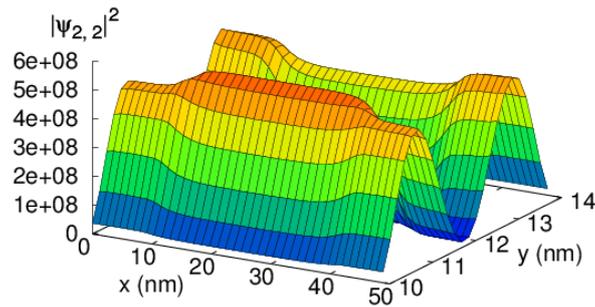


Figure 65 : densité de probabilité dans le silicium. Vallée Δ_y , deuxième niveau. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.

Sur la figure 66, la fonction d'onde de la vallée Δ_x , moins peuplée, présente un profil standard.

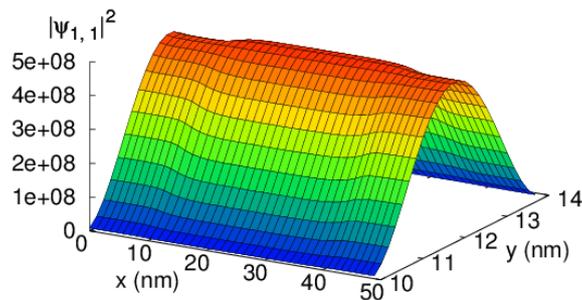


Figure 66 : densité de probabilité dans le silicium. vallée Δ_x , premier niveau. $L = 30$ nm, $t_{Si} = 4$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV. $L_s = L_d = 10$ nm.

Des coupes selon y dans le canal ont été données au Chapitre II. 1.3, p. 44 et suivantes.

Finalement, lorsque les bandes d'énergie supérieures sont occupées, comme c'est le cas pour des épaisseurs de silicium supérieures à quelques dizaines de nanomètres, les densités de probabilité correspondantes conduisent à une répartition de la charge plus uniforme selon y , pour laquelle une résolution classique sans prise en compte du confinement devient pertinente, ce qui n'est pas le cas pour la géométrie étudiée ici.

Ainsi est calculée la densité d'électrons qui sera utilisée par la suite avec le courant pour l'extraction de mobilité.

1.3. Transport non stationnaire

Pour comprendre les courbes de mobilité et leurs liens avec celles obtenues dans le cas du canal balistique, il peut être utile d'étudier le transport à travers d'autres grandeurs physiques, telles que la vitesse, le taux de « balisticité » déduit de la répartition statistique des porteurs en fonction du nombre d'interactions subies à travers le canal.

1.3.1. Vitesse le long du canal

Les figures 67 et 68 décrivent l'évolution de la vitesse moyenne des électrons le long du canal à, respectivement, faible et forte tension drain-source, pour les sous-bandes les plus peuplées. À fort V_{ds} , le régime de mobilité n'est clairement plus valide, puisque pour les niveaux les plus remplis la vitesse dépasse de beaucoup la valeur atteinte en régime de saturation v_{sat} dans le silicium massif, qui est de l'ordre de 10^5 ms^{-1} . À faible V_{ds} , la vitesse reste inférieure à v_{sat} , le caractère non stationnaire du transport n'est pas directement visible.

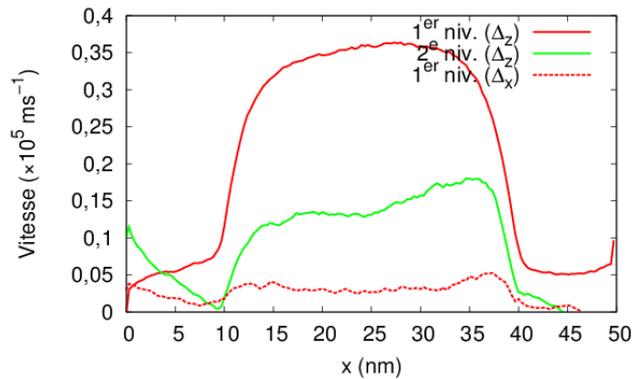


Figure 67 : vitesse le long du canal des porteurs des différentes sous-bandes. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$.

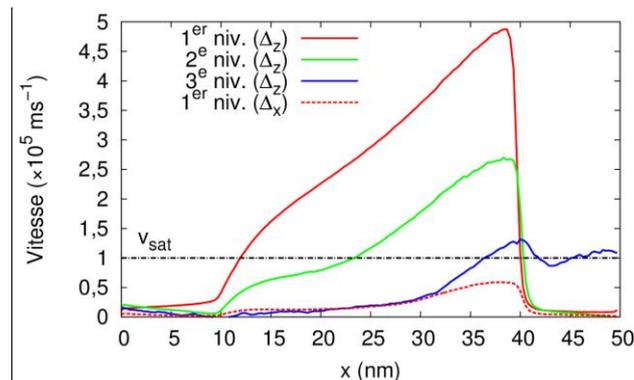


Figure 68 : vitesse le long du canal des porteurs des différentes sous-bandes. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 0,8 \text{ V}$.

La figure 69 présente la répartition des électrons ayant traversé le canal de la zone dopée de source jusqu'à celle de drain en fonction du nombre d'interactions subies.

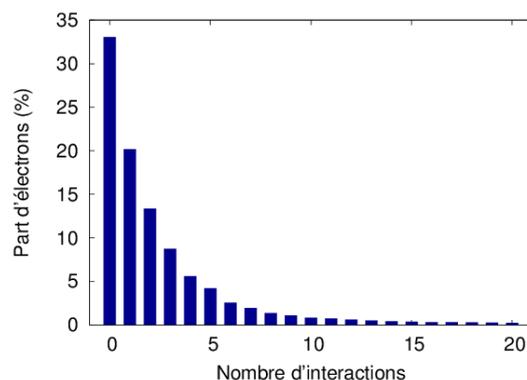


Figure 69 : répartition des électrons selon leur nombre d'interactions subies dans le canal. $L = 30 \text{ nm}$, $V_{gs} = 0,8 \text{ V}$ et $V_{ds} = 50 \text{ mV}$.

Pour cette longueur de grille de 30 nm, les électrons balistiques sont une catégorie majoritaire par comparaison à chaque catégorie à nombre d'interactions subies donné, le taux de balisticité intrinsèque (pourcentage d'électrons balistiques parmi ceux traversant le canal de la l'interface avec la source jusqu'à celle avec le drain) étant de $B_{\text{int}} = 33\%$. L'évolution de cette grandeur avec la longueur de canal est représentée sur la figure 70.

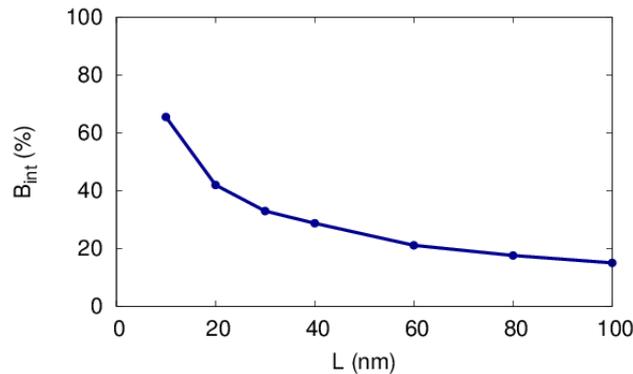


Figure 70 : taux de balisticité intrinsèque en fonction de la longueur de canal. $V_{gs} = 0,8\text{ V}$ et $V_{ds} = 50\text{ mV}$.

Le taux de balisticité en sortie d'électrons balistiques traversant le canal augmente bien continûment avec la réduction de la longueur de grille. Il ne dépasse 50% que pour L_g proche de 10 nm. Si l'on prend en compte les électrons subissant 1 à 3 interactions à la traversée du canal, la nature du transport est toutefois assez radicalement modifiée par la présence de ces porteurs quasi-balistiques quand L_g diminue en deçà de 50 nm [SAINT-MA04].

2. Mobilités extraites sur les simulations

2.1. Extraction sur les simulations

En simulant des dispositifs de longueur L allant de 10 à 100 nm, les charges et les courants directement obtenus permettent d'en déduire grâce aux méthodes décrites dans le chapitre précédent différentes courbes de mobilité pour différentes longueurs.

Tout d'abord, l'augmentation attendue du courant lorsque L diminue est visible par simulation comme le montre la figure 71.

Mais cette augmentation de courant lorsque L diminue n'est pas aussi grande qu'espérée, en effet la mobilité effective diminue, et ce quelle que soit la charge d'inversion, comme en témoigne la figure 72.

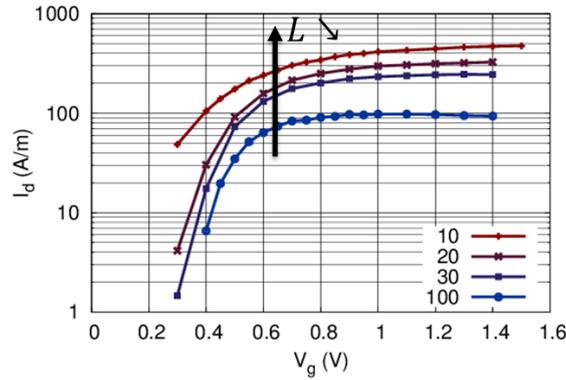


Figure 71 : $I_d(V_g)$ avec interactions dans le canal. Échelle logarithmique. $L = 10, 20, 30$ et 100 nm. $V_{ds} = 50$ mV.

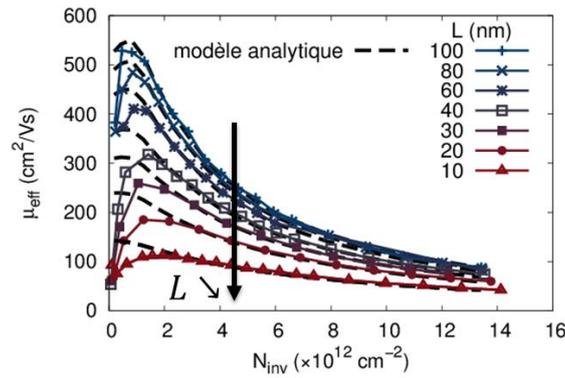


Figure 72 : mobilité effective extraite des simulations à canal diffusif (avec interactions dans le canal). Ligne continue : simulation Monte-Carlo. Pointillés : modèle analytique complet de l'équation 119.

2.2. Mobilité extraite en fonction de la longueur

La mobilité peut être extraite pour différentes valeurs de charge d'inversion, comme sur les trois courbes de la figure 73.

Le pourcentage de dégradation de mobilité est simplement la variation relative de mobilité entre celle obtenue pour le canal de plus grande longueur (L_{long}) et celle obtenue pour le canal de plus faible longueur (L_{court}) :

$$\Delta\mu = \frac{\mu_{eff}(L_{long}) - \mu_{eff}(L_{court})}{\mu_{eff}(L_{long})} \quad (110)$$

En l'occurrence, $L_{court} = 10$ nm et $L_{long} = 200$ nm.

Finalement, pour tout N_{inv} , le pourcentage de dégradation obtenu est $\Delta\mu \simeq 60$ %.

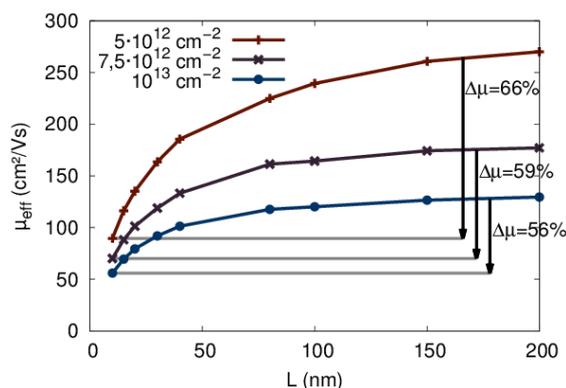


Figure 73 : dégradation de la mobilité avec la réduction de la longueur de canal à différentes charges d'inversion : $N_{\text{inv}} = 5 \cdot 10^{12}$, $7,5 \cdot 10^{12}$ et 10^{13} cm^{-2} . Chaque point est obtenu par interpolation linéaire sur les courbes précédentes en faisant varier L à N_{inv} fixé.

Cette dégradation de mobilité avec la réduction de L est décrite par le modèle [GUARNAY14] proposé dans la sous-section suivante traitant des contributions de la résistance balistique, de la résistance d'accès et de la mobilité à canal long.

3. Modélisation de la dégradation de mobilité

3.1. Loi de Matthiessen

Nous avons analysé les variations de la mobilité avec L par le modèle suivant à différentes charges d'inversion N_{inv} en ajustant les paramètres d'une loi de Matthiessen :

$$\frac{1}{\mu_{\text{eff}}(L)} = \frac{1}{K_{R_{\text{bal}}+R_0} L} + \frac{1}{\mu_{\text{long}}} \quad (111)$$

avec pour paramètres d'ajustement μ_{long} (composante diffusive de la mobilité) et le facteur $K_{R_{\text{bal}}+R_0}$ qui inclut comme nous le verrons non seulement la contribution de la résistance balistique du canal mais aussi celle des résistances d'accès.

Ces deux composantes de la mobilité peuvent être visualisées sur la figure 74.

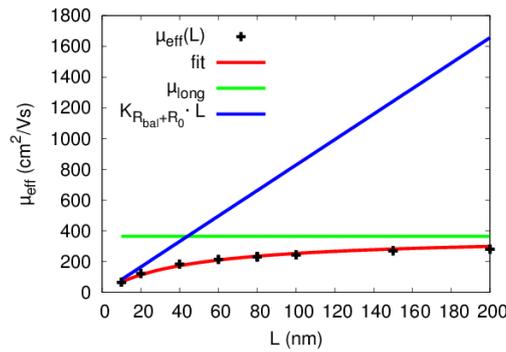


Figure 74 : ajustement des paramètres $K_{R_{bal}+R_0}$ et μ_{long} sur la mobilité en fonction de la longueur de canal. Exemple à $N_{inv} = 5 \cdot 10^{12} \text{ cm}^{-2}$.

Cette figure indique que ces deux paramètres suffisent pour décrire parfaitement la dégradation de mobilité avec la réduction de la longueur de canal L .

D'un côté, la mobilité canal long μ_{long} tenant compte de la diffusion dans le canal peut être extraite comme une extrapolation de la mobilité lorsque L tend vers l'infini, ou par des simulations à champ constant.

De l'autre, il est possible d'étudier la composante $K_{R_{bal}+R_0}$ par un modèle descriptif analytique ou par des simulations à canal balistique.

Ces deux composantes, μ_{long} et $K_{R_{bal}+R_0}$, sont représentées en fonction de la charge d'inversion N_{inv} , sur les deux figures suivantes. Les valeurs obtenues par ajustement de la mobilité extraite sont représentées en même temps que la mobilité déduite de simulations Monte-Carlo à champ constant (figure 75) et du résultat du modèle analytique de résistance balistique (prenant en compte des résistances d'accès) qui sera présenté en section 3.3. On constate un très bon accord entre μ_{long} et la mobilité en champ constant d'une part, $K_{R_{bal}+R_0}$ et les résultats du modèle analytique d'autre part.

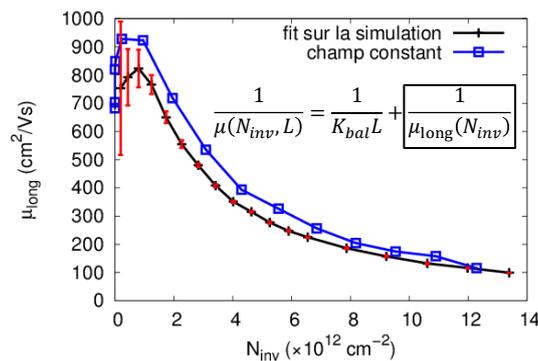


Figure 75 : mobilité canal long. Ligne noire : extraction d'après des simulations à différentes longueur de canal, en ajustant les paramètres de la formule donnée en insert. En rouge, l'écart-type. Ligne bleue et carrés : mobilité extraite de simulations à champ constant.

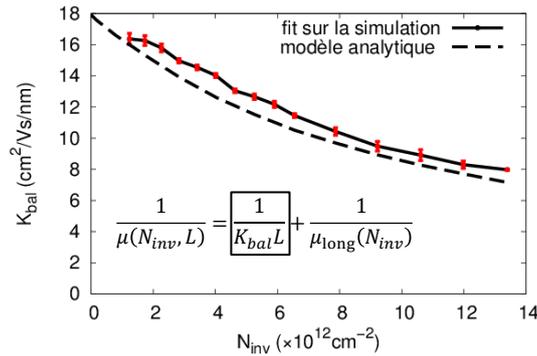


Figure 76 : facteur de mobilité $K_{R_{\text{bal}}+R_0}$ de résistance balistique et d'accès, en fonction de la charge d'inversion. Ligne continue : extraction sur les simulations Monte-Carlo des paramètres en insert. Ligne pointillée : modèle analytique pour $K_{R_{\text{bal}}+R_0}$ (équation 112).

3.2. Mobilité à canal long

Afin d'évaluer l'extraction de la mobilité μ_{long} pour L tendant vers l'infini, nous avons procédé à des simulations Monte-Carlo, celle des simulations à champ électrique constant. Il est uniforme selon l'axe du transport x et faible (500 V.cm^{-1}) alors que selon l'axe y nous utilisons le champ de confinement déduit de la résolution 1D des équations de Schrödinger et de Poisson (grâce au code AQUILA¹⁴).

Suivant x , les conditions aux limites sont périodiques : tout électron collecté à la fin du canal simulé est réinjecté dans le même état en début de canal, afin de considérer une structure infiniment longue. La mobilité est déduite de la valeur stationnaire atteinte par la vitesse moyenne des électrons simulés (divisée par le champ selon x).

La mobilité déduite des simulations à canal long se superpose donc bien à la composante μ_{long} déduite des extractions de mobilité sur des dispositifs courts. De plus, elle a des valeurs proches de celles de la mobilité universelle dans un canal de MOSFET en Si [TAKAGI94], non représentée ici pour ne pas surcharger la figure.

3.3. Résistances balistique et d'accès

L'autre composante principale de l'évolution de la mobilité en fonction de la longueur L , à savoir $K_{R_{\text{bal}}+R_0}$, extraite comme la composante linéaire en L , peut également être modélisée analytiquement.

$$K_{R_{\text{bal}}+R_0} = \frac{1}{eN_{\text{inv}}W(R_0 + R_{\text{bal}})} \quad (112)$$

Cette composante principale dépend donc à son tour de deux termes, la résistance d'accès R_0 et la résistance balistique R_{bal} dont les contributions s'additionnent.

¹⁴ www.mathworks.com/matlabcentral/fileexchange/3344-2d-schroedinger-poisson-solver-aquila

D'une part, la résistance d'accès R_0 est simplement supposée indépendante de la polarisation de la grille V_g , ce qui est censé être une bonne approximation pour des simulations sur des géométries simples comme c'est le cas ici, mais qui mériterait une meilleure description dans le cas d'une extraction expérimentale [RIDEAU14].

Cette résistance de source et drain est estimée par différence entre les simulations à canal balistique (source et drain diffusifs) et la résistance balistique analytique. Plus précisément, la résistance d'accès vaut :

$$R_0(N_{inv}) = \frac{V_{ds}}{I_d(N_{inv})} - R_{bal. anal.}(N_{inv}) \quad (113)$$

En l'occurrence, $R_0 = 35 \Omega \cdot \mu m$ d'après l'extraction de la figure 77, en considérant les fortes valeurs de N_{inv} .

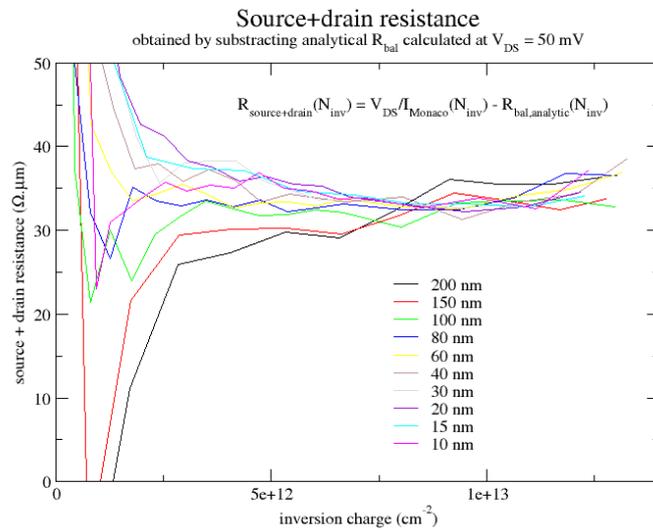


Figure 77 : extraction de la résistance d'accès par différence entre les simulations à canal balistique (source et drain diffusifs) et la résistance balistique analytique. R_0 s'approche de $35 \Omega \cdot \mu m$ à fort N_{inv} .

D'autre part, la résistance balistique R_{bal} est calculée dans l'hypothèse de transport purement balistique en analysant le courant provenant de la source et du drain, obtenu par intégration des fonctions de distribution f_s et f_d sur tous les vecteurs d'onde \mathbf{k} prenant leurs valeurs dans deux dimensions de l'espace réciproque :

$$I_{sd} = \frac{-2eL}{(2\pi)^2} \int_{v_x > 0} (f_s(\mathbf{k}) - f_d(\mathbf{k})) v_x(\mathbf{k}) d\mathbf{k} \quad (114)$$

À faible V_{ds} , la différence entre les fonctions de distribution peut être simplifiée par un développement selon l'énergie :

$$f_s - f_d = -\frac{df}{dE} eV_{ds} \quad (115)$$

Cela permet d'intégrer le courant en fonction de l'énergie, faisant alors apparaître les intégrales de Fermi-Dirac d'ordre $\frac{1}{2}$, comme décrit en annexe 2.

Finalement, la résistance balistique vaut :

$$\frac{1}{WR_{\text{bal}}} = \sum_{\nu,n} \underbrace{\frac{e\sqrt{\pi m_y^\nu k_B T}}{\sqrt{2}\pi^2 \hbar^2}}_{\mu_{\text{bal}}} \times \underbrace{\left[\mathcal{F}_{1/2} \left(\frac{E_F - E_n^\nu}{k_B T} \right) - \mathcal{F}_{1/2} \left(\frac{E_F - eV_{ds} - E_n^\nu}{k_B T} \right) \right]}_{\text{Fermi-Dirac}} \frac{k_B T}{V_{ds}} \quad (116)$$

où W est la longueur de canal. Pour chaque vallée ν , m_y^ν est la masse effective transverse et E_n^ν est l'énergie de la sous-bande n , obtenue par simulation. $\mathcal{F}_{1/2}$ est l'intégrale de Fermi-Dirac d'ordre $\frac{1}{2}$:

$$\mathcal{F}_j(\eta_F) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{\xi^j d\xi}{1 + e^{\xi - \eta_F}} \quad (117)$$

La résistance balistique est calculée en équilibrant les flux de courants balistiques et en utilisant la statistique de Fermi-Dirac. Le facteur μ_{bal} de l'éq. (116) correspond à la mobilité balistique apparente originelle calculée par M. S. Shur [SHURO2]. Il est précisé par le facteur comprenant les intégrales de Fermi-Dirac déjà connu [ZILLIO7], qui implique une dégradation lorsque V_{gs} augmente.

La résistance balistique du présent modèle est donc supérieure à la résistance balistique du modèle de Shur d'environ $50 \Omega \cdot \mu\text{m}$ quelle que soit la charge d'inversion, d'après la figure 78.

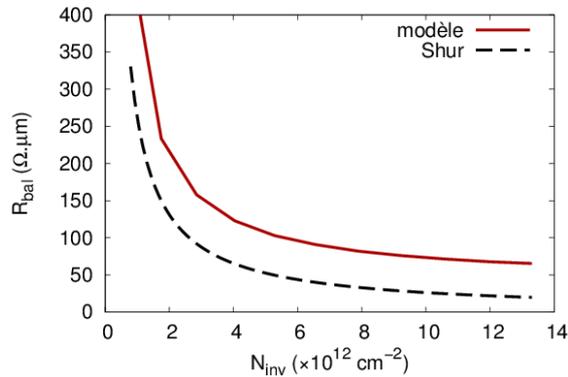


Figure 78 : résistance balistique issue du présent modèle (équation 116) en fonction de la charge d'inversion (ligne continue), comparée à la résistance balistique du modèle de Shur (tirets).

Ces valeurs sont intéressantes à connaître puisqu'elles sont assez universelles. En effet, elles ne varient pas selon la longueur de canal et peu selon son épaisseur.

En termes de mobilité, ce modèle se traduit par la formule suivante de mobilité apparente balistique [ZILLIO7], qui est donc plus précise que celle de Shur, mais que nous n'utiliserons pas en pratique par la suite :

$$\mu_{\text{bal}} = \frac{eL}{\sqrt{2\pi m_x k_B T}} \frac{\mathcal{F}_{-1/2} \left(\frac{E_F - E_n}{k_B T} \right)}{\mathcal{F}_0 \left(\frac{E_F - E_n}{k_B T} \right)} \quad (118)$$

Ce modèle de transport balistique peut être confronté à des simulations Monte-Carlo à canal balistique mais en conservant la diffusion dans les accès, source et drain. La comparaison des niveaux

d'énergie et des courants avec les simulations à canal « diffusif » sont données sur les deux figures 79 et 80 :

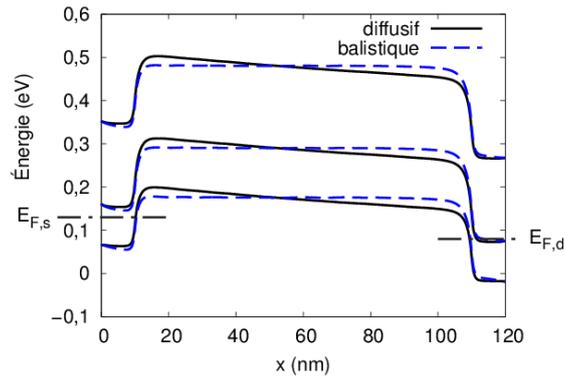


Figure 79 : énergie des trois premières sous-bandes des vallées Δ_y (y étant la direction de confinement). $L = 100$ nm, $V_g = 0,2$ V et $V_{ds} = 50$ mV. Les bandes d'énergie sont plates dans les simulations à canal balistique en l'absence d'interactions permettant l'équilibre avec la tension de drain imposée.

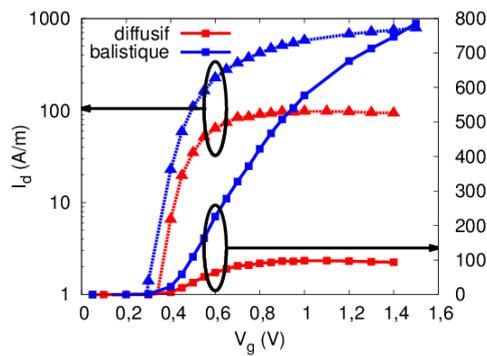


Figure 80 : comparaison des courants des simulations diffusives et balistiques. Ordonnée à droite, échelle linéaire. À gauche, échelle logarithmique. $L = 30$ nm.

La mobilité apparente extraite de la même manière que pour le canal diffusif est alors donnée sur la figure 81 et se superpose bien au modèle analytique proposé, à l'exception des faibles charges d'inversion pour les raisons évoquées précédemment au Chapitre III. 2.2, p. 75.

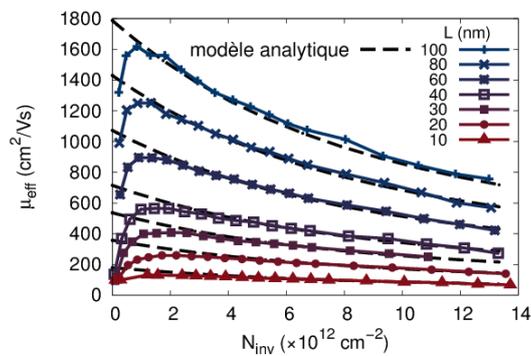


Figure 81 : mobilité effective extraite des simulations à canal balistique (sans interaction dans le canal). Ligne continue : simulation Monte-Carlo. Pointillés : modèle analytique.

Cette mobilité extraite présente bien la tendance de diminution prévue par le modèle lorsque N_{inv} augmente. Finalement, la formule de la mobilité effective des simulations diffusives est obtenue en ajoutant un terme μ_{long} de mobilité à canal long.

$$\frac{1}{\mu_{eff}} = \frac{eN_{inv}W(R_{bal} + R_0)}{L} + \frac{1}{\mu_{long}(N_{inv})} \quad (119)$$

Ce modèle décrit parfaitement la dégradation de mobilité effective présente sur la figure 72, aux faibles charges d'inversion près pour lesquelles l'extraction de la charge d'inversion est plus difficile, de même que la détermination de la zone d'extraction, comme expliqué précédemment.

3.4. Autre évaluation de la résistance d'accès

Nous nous sommes efforcés d'évaluer la résistance d'accès au canal sur les résultats de simulation Monte-Carlo, en étudiant les variations d'un niveau de Fermi extrait de la densité de porteurs.

3.4.1.i. Détermination de l'énergie de Fermi

La densité électronique dans le cas d'un semiconducteur 2D est donnée par [KIM08] :

$$N_{inv} = \sum_{v,n} \frac{m_{eff}^v k_B T}{\pi \hbar^2} \mathcal{F}_0 \left(\frac{E_F - E_n^v}{k_B T} \right) \quad (120)$$

où m_{eff}^v est la masse effective de la vallée v .

Réciproquement, en connaissant la charge d'inversion, il est possible d'ajuster l'énergie de Fermi permettant de retrouver cette charge. Cet ajustement de niveau de Fermi pour chaque position, correspondant à une charge d'inversion particulière, donne la figure suivante :

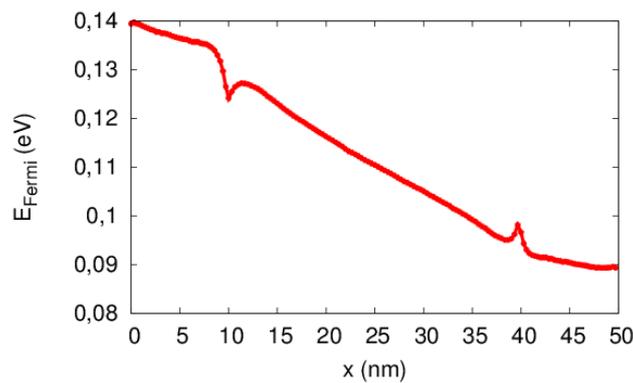


Figure 82 : énergie de Fermi le long du canal. $L = 30$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV. Les pics correspondent aux jonctions considérées comme abruptes entre canal et zones dopées de source et drain.

3.4.1.ii. Calcul de la résistance d'accès

La chute de niveau de Fermi entre deux positions, divisée par le courant, donne alors une valeur de résistance.

Plus précisément, il est intéressant de connaître la résistance cumulée depuis la source, présentée sur la figure 83 et définie par la fonction suivante :

$$R_s(x) = \frac{E_F(0) - E_F(x)}{I_d} \quad (121)$$

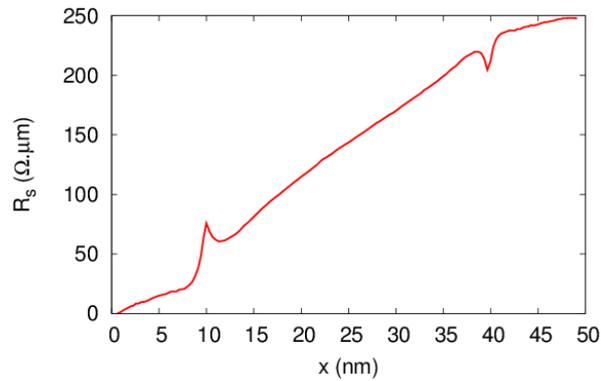


Figure 83 : résistance cumulée depuis la source. $L = 30$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

En effet, la résistance de la source dépend de sa zone effective de définition. Pour la visualiser, les courbes de résistance cumulée superposées pour différentes longueurs de canal donnent une indication :

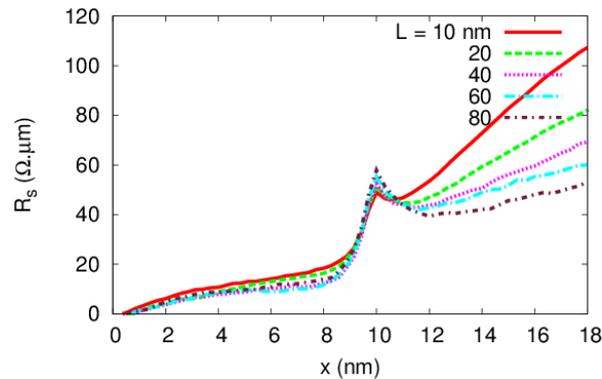


Figure 84 : résistance cumulée depuis la source pour différentes longueurs de canal. $L = 10, 20, 40, 60$ et 80 nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

Les courbes se superposent jusqu'à 1 nm avant le pic. La résistance d'accès au canal par la zone dopée de source apparaît ainsi de l'ordre de $20 \Omega \cdot \mu\text{m}$. Au niveau du drain, la résistance d'accès est légèrement plus faible ($15 \Omega \cdot \mu\text{m}$). En sommant ces deux contributions, la résistance d'accès totale obtenue correspond à celle évaluée précédemment ($35 \Omega \cdot \mu\text{m}$, cf. figure 77) par différence entre la simulation du canal balistique et la résistance balistique analytique.

Il est de même possible de déterminer une résistance linéique pour une étude locale des simulations.

$$\rho(x) = -\frac{1}{I_d} \frac{dE_F(x)}{dx} \quad (122)$$

En tant que dérivée de la résistance cumulée, elle accentue les pics à l'interface avec la source et le drain. Par ailleurs, c'est une grandeur locale qui ne sera donc pas comparée à des mesures expérimentales.

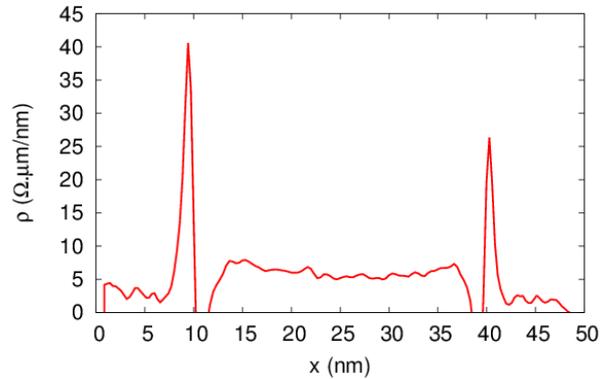


Figure 85 : résistance linéique. $L = 30$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

3.4.1.iii. Résistance cumulée pour un canal balistique

Dans le cas d'un canal balistique, les grandeurs précédentes peuvent être extraites de manière analogue. Le niveau de Fermi présente cette fois-ci un palier, du moins une chute faible, dans le canal.

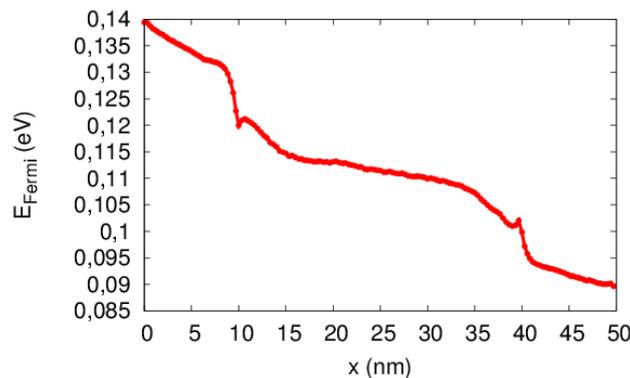


Figure 86 : énergie de Fermi pour un canal balistique. $L = 30$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

En contrepartie de cette faible chute dans le canal, l'énergie de Fermi calculée dans ce cas présente très clairement une chute à l'interface entre le canal et la source et entre le canal et le drain, plus encore que dans le cas des simulations diffusives précédentes, même en faisant abstraction du pic dû aux jonctions abruptes.

Cette chute à l'interface correspond à une résistance qui est à prendre en compte dans le calcul de la résistance d'accès en considérant une zone de définition de la source comprenant également de l'ordre de 1 nm du canal, en cohérence avec l'analyse de l'énergie de Fermi dans le cas des simulations diffusives.

De plus, cette chute se poursuit avant le palier ce qui se mesure plus directement d'après la résistance linéique représentée sur la figure 87.

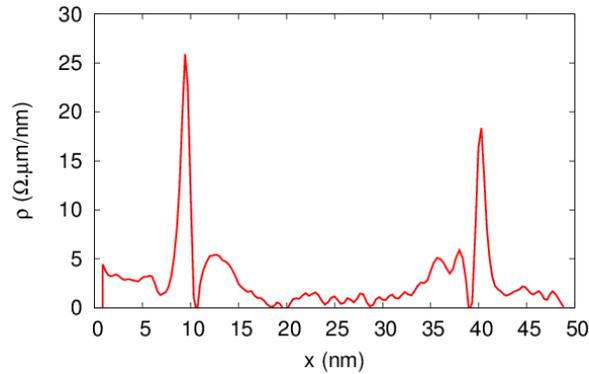


Figure 87 : résistance linéique avec un canal balistique. $L = 30$ nm, $V_{gs} = 0,8$ V et $V_{ds} = 50$ mV.

Cette résistance linéique est naturellement quasi-nulle au centre du canal. En d'autres termes, la résistance balistique se situe à proximité des jonctions, plus précisément à quelques nanomètres de celles-ci. Cela la rend d'autant plus difficile à distinguer de la résistance d'accès à l'interface.

En revanche, le modèle de résistance d'accès présente des limites, celle-ci dépendant non seulement de la zone de définition de la source et du drain, mais également de la longueur, du type de canal considéré, et des polarisations d'après la figure 88.

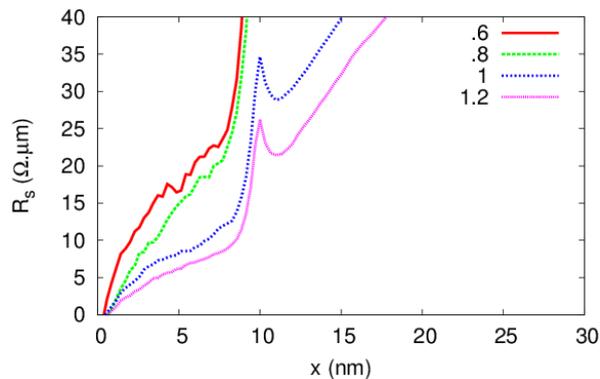


Figure 88 : résistance de source cumulée pour différentes valeurs de tensions. $L = 30$ nm, $V_{gs} = 0,2, 0,4, 0,6$ et $0,8$ V. $V_{ds} = 50$ mV.

Par ailleurs, le concept d'énergie de Fermi est valide dans l'hypothèse de porteurs à l'équilibre, ce qui n'est pas le cas notamment au niveau des jonctions dans lesquelles se situent les pics. Mais il permet de déduire une valeur de résistance d'accès qui coïncide avec celle obtenue par les méthodes présentées auparavant.

4. Comparaison avec les mobilités expérimentales

4.1. Analyse du transport par les méthodes classiques

4.1.1. Présentation des données expérimentales

Ce paragraphe s'appuiera sur des données expérimentales provenant du partenariat CEA-Leti avec IBM à Albany. Ce lot constitue une étude préparatoire pour l'optimisation de la technologie FDSOI et nous permettra de confronter notre modèle présenté précédemment à des données expérimentales. Celles-ci sont décrites sommairement par la figure 89, et concernent une technologie FDSOI à BOX épais (145 nm) avec un film de silicium (aucun *booster* de mobilité n'est ajouté) de 6 nm et une épaisseur équivalente d'oxyde (EOT) de 0,74 nm réalisée à partir d'un empilement de SiO₂-HfO₂.

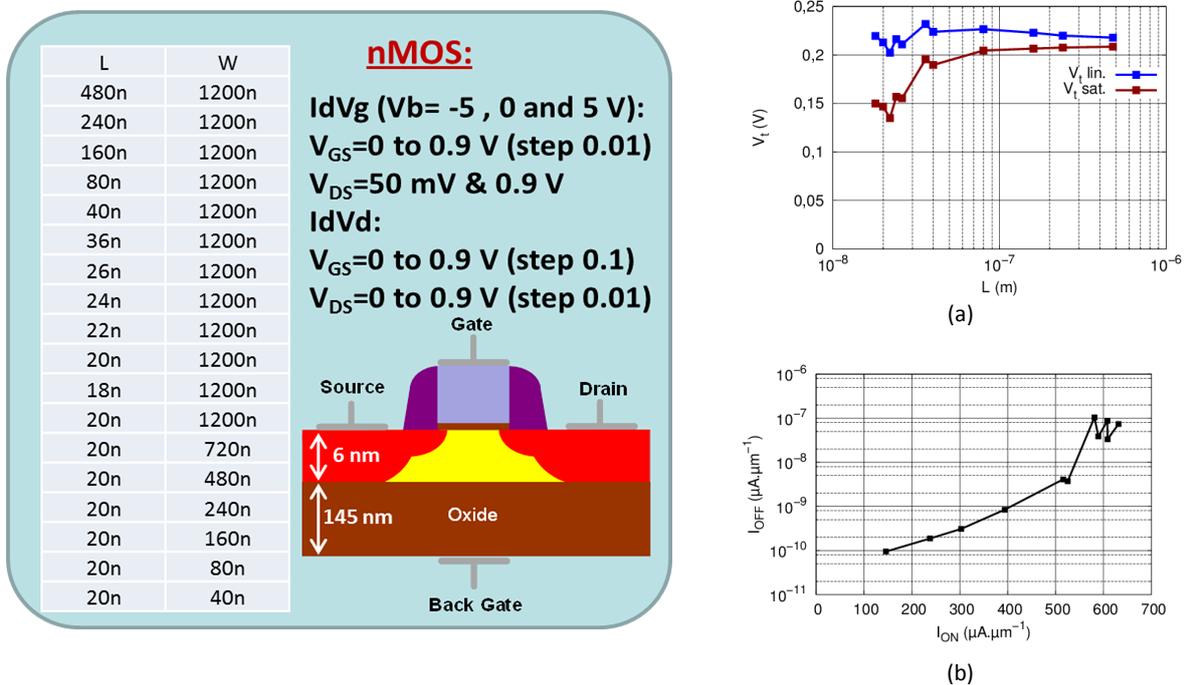


Figure 89 : géométrie des dispositifs étudiés expérimentalement et polarisations appliquées pour l'extraction de la mobilité. (a) Tension de seuil en régime linéaire et saturé. (b) Courant I_{OFF} à l'état bloqué en fonction du courant I_{ON} à l'état passant.

4.1.2. Analyse des données par fonction Y

Nous avons dans un premier temps extrait la mobilité effective par la méthode de la fonction Y présentée dans le chapitre III. Les mobilités extraites des données expérimentales, sont plus bruitées que celles issues de simulations Monte-Carlo, comme on peut le constater sur la figure 90. Cette courbe montre une mobilité à champ faible canal long de l'ordre de $200 \text{ cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$, valeur qui semble légèrement faible par rapport à la littérature [FLEURY09]. Néanmoins nous verrons plus tard lors de l'extraction par split-CV que cette conclusion n'est en rien un gage de mauvaise performance des transistors les plus courts.

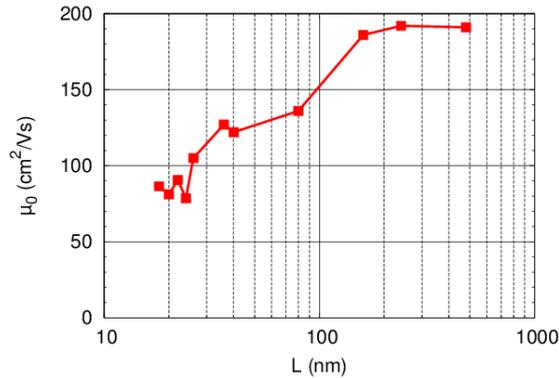


Figure 90 : extraction de la mobilité à champ faible μ_0 par la méthode de la fonction Y

Grâce à cette extraction avec la fonction Y , il nous est possible d'évaluer la résistance d'accès par régression linéaire sur la courbe $\theta_1(\beta_0)$ (illustrée sur la figure 91). Cette méthode nous donne une résistance d'accès de $260 \Omega \cdot \mu\text{m}$, qui reste dans les spécifications des technologies actuelles au meilleur de l'état de l'art [FLEURY09].

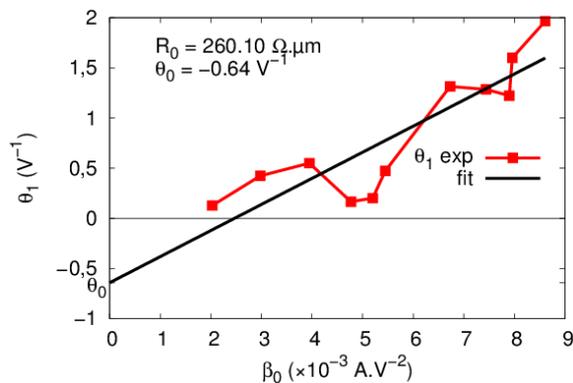


Figure 91 : extraction de la résistance d'accès R_0 par la pente de la courbe $\theta_1(\beta_0)$

4.1.3. Extraction de la mobilité par split-CV

Avec la série de données présentée dans le paragraphe précédent, un nombre réduit de mesures capacitives nous a été fourni, nous permettant d'extraire par la méthode split-CV présentée dans le chapitre III la mobilité pour plusieurs longueurs de canal.

La figure 92 représente le tracé des différentes capacités disponibles. Malgré le bruit de mesure de ces données, elles sont exploitables, sauf pour la longueur la plus courte ($L = 20 \text{ nm}$) qui

présente un niveau de capacité parasite trop important pour tirer des conclusions fiables par cette méthode.

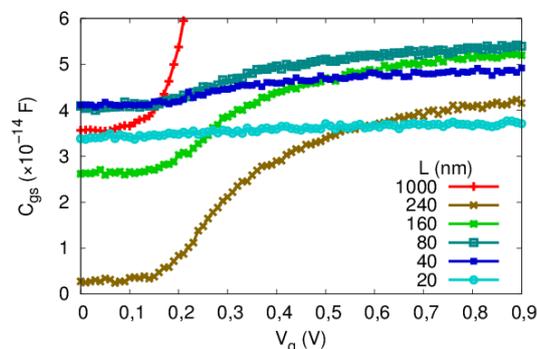


Figure 92 : faisceau de courbes de capacités totales $C_{gs}(V_g)$ à différentes longueurs de grille ($L = 240, 160, 80, 40$ et 20 nm).

La mobilité extraite par split-CV, comme expliqué dans le chapitre III, est donnée sur la figure 93. Nous remarquons que les performances des transistors sont assez bonnes. Il est vrai que la mobilité à champ faible est en dessous de la littérature, mais la performance globale de la mobilité en fonction de l'augmentation de la charge d'inversion est intéressante même pour des dispositifs à faible longueur de grille sans booster de mobilité [FLEURY09].

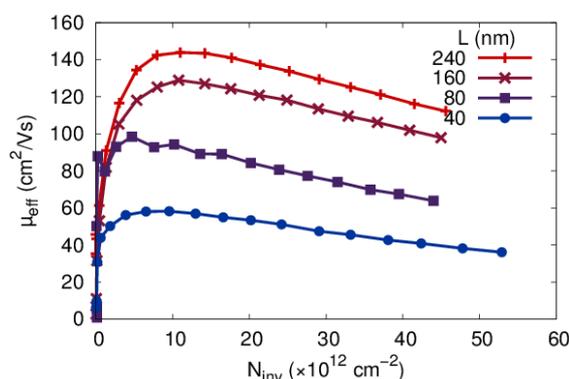


Figure 93 : mobilité extraite par split-CV en fonction de la charge d'inversion pour différentes longueurs de grille ($L = 240, 160, 80$ et 40 nm).

4.1.4. Mobilité effective expérimentale vs modèle

Nous avons ajusté les paramètres du modèle développé précédemment pour reproduire ces variations. Les paramètres $K_{R_{bal}+R_0}$ et μ_{long} sont extraits sur $\mu_{eff}(L)$ à différents N_{inv} (cf. figures 94 et 95). Les courbes sont lisses puisque les mobilités $\mu_{eff}(N_{inv})$ ont été extraites par la méthode Y et que les contributions de dégradation selon L sont d'autant plus précises qu'il y a de longueurs testées.

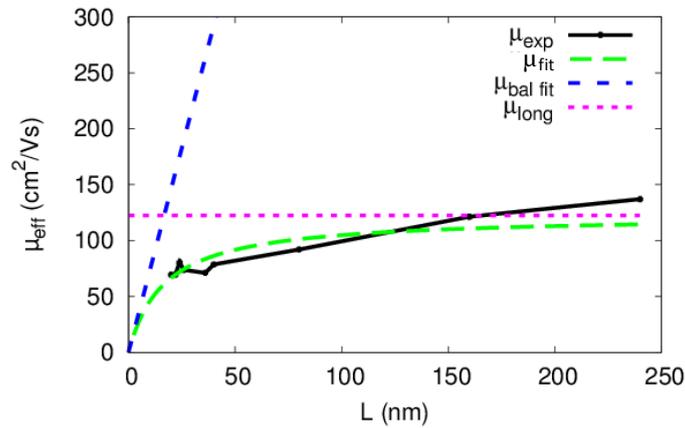


Figure 94 : mobilité extraite expérimentalement et ajustement des paramètres $\mu_{\text{bal}} = K_{\text{bal}} \cdot L$ et μ_{long} dans le cas particulier $N_{\text{inv}} = 0,96 \cdot 10^{12} \text{ cm}^{-2}$.

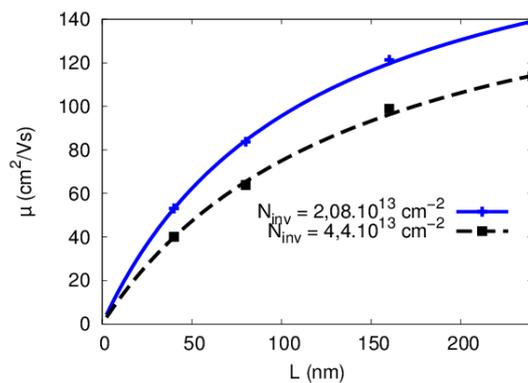


Figure 95 : mobilité extraite expérimentalement dans les cas particuliers $N_{\text{inv}} = 2,08 \cdot 10^{13} \text{ cm}^{-2}$ et $N_{\text{inv}} = 4,4 \cdot 10^{13} \text{ cm}^{-2}$

4.1.5. Contributions $K_{R_{\text{bal}}+R_0}$ et μ_{long}

Les paramètres $K_{R_{\text{bal}}+R_0}$ et μ_{long} peuvent ainsi être représentés en fonction de la charge d'inversion, respectivement sur les figures 96 et 98.

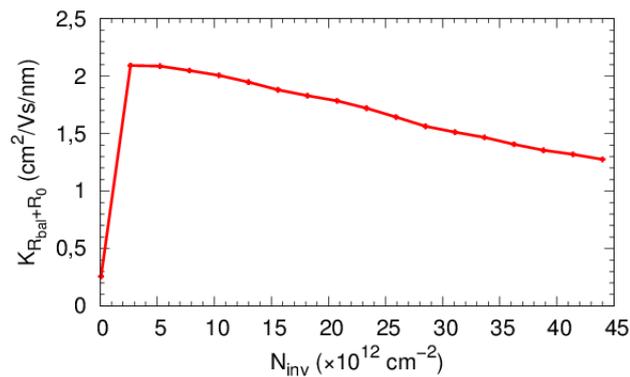


Figure 96 : facteur de mobilité balistique $K_{R_{\text{bal}}+R_0}$ de résistance balistique et d'accès extrait des résultats expérimentaux, en fonction de N_{inv} .

La figure 97 montre la résistance (en $\Omega \cdot \mu\text{m}$) totale $R_0 + R_{\text{bal}} = 1/eN_{\text{inv}}K_{\text{bal}}$ ainsi que la résistance balistique théorique R_{bal} (calculée par résolution de Poisson-Schrödinger sur un film de 6 nm d'épaisseur). On observe que le terme de résistance d'accès R_0 domine sur R_{bal} . De plus, sa

valeur moyenne est compatible avec l'extraction de la résistance (indépendante de N_{inv}) par fonction Y .

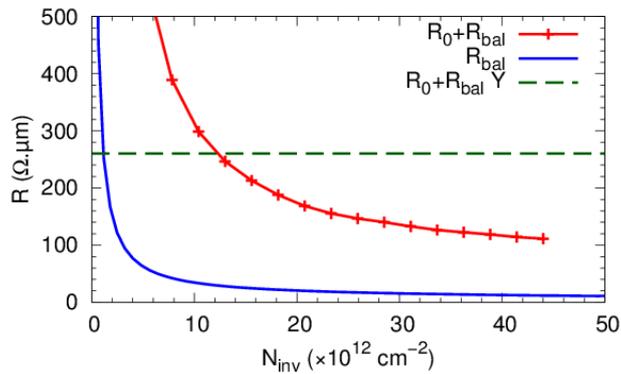


Figure 97 : résistance d'accès extraite des résultats expérimentaux par notre modèle, comparée à la résistance d'accès extraite par fonction Y (indépendante de N_{inv}).

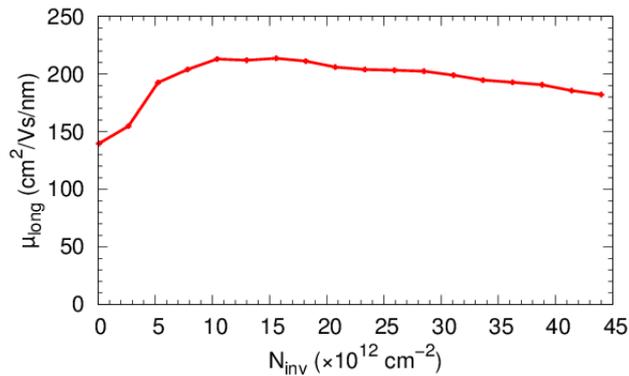


Figure 98 : composante de mobilité canal long μ_{long} extraite des dispositifs expérimentaux, en fonction de N_{inv} .

5. Bilan sur le transport à canal court

Nous avons dans ce chapitre utilisé la modélisation Monte-Carlo ainsi que construit une modélisation analytique nous permettant d'extraire les principaux paramètres du transport électronique dans les dispositifs ultimes. Comme expliqué dans les chapitres précédents, le choix de différentes méthodes d'extraction couplées à la simulation Monte-Carlo semble être le meilleur compromis pour analyser finement la mobilité.

Notre modèle prend en compte la notion de résistance balistique qui est généralement sous-estimée par l'utilisation du modèle de Shur. Néanmoins, la dépendance à la charge d'inversion de la résistance d'accès reste la contribution principale de la dégradation de mobilité avec la réduction de la longueur du canal.

Il faudrait donc poursuivre l'étude de cette résistance d'accès pour connaître :

- les étapes de procédé de fabrication responsables de la diffusion latérales des dopants et ainsi mieux identifier les solutions à mettre en œuvre pour contrôler cette résistance,

- l'effet de la contrainte et les valeurs les plus favorables pour diminuer la résistance,
- la zone de définition de cette résistance, à l'interface entre les accès et le canal, notamment par rapport à la longueur effective de canal,
- la validité de la procédure d'extraction de paramètres par rapport à celle obtenue par les méthodes existantes. Il faudrait prendre en compte la dépendance de cette résistance avec la charge d'inversion pour améliorer le modèle proposé.
- l'importance de la résistance balistique, qui devrait devenir non négligeable pour des technologies plus optimisées (en termes de zones d'accès).

Nous avons aussi éprouvé notre approche sur des données expérimentales permettant d'avoir une idée de la robustesse des méthodes d'extraction mises en jeu. Cette étude a aussi permis d'évaluer, sur des données expérimentales, le poids de la résistance balistique par rapport à la résistance d'accès. C'est un résultat qui peut être intéressant pour le *benchmark* d'architectures puisqu'il représente une limite physique. De plus nous voyons clairement la décroissance de la résistance en fonction de l'accroissement de la charge d'inversion ; un résultat attendu puisque dans les zones moyennement dopées (sous l'espaceur), l'augmentation progressive de la charge due au contrôle de la grille favorise le transport sans pour autant que cela soit une région où les interactions entre porteurs sont dominantes.

Conclusion générale

Nous avons proposé une évaluation des principaux mécanismes de dégradation de la mobilité des transistors FDSOI avec la longueur de canal, pour aider à mieux comprendre le transport et à mesurer sa qualité afin d'améliorer les performances des futurs nœuds technologiques. Pour ce faire, nous avons étudié par simulation Monte-Carlo (simulateur MONACO de l'IEF) différents mécanismes de dégradation de la mobilité extraite des caractéristiques électriques : les interactions de la mobilité extraite des caractéristiques électriques, le transport quasi-balistique et les résistances d'accès.

D'une part, la simulation a permis de décorrélérer les effets des résistances d'accès et balistique. En outre, une évaluation locale de la charge, à faible inversion, a pu être effectuée (toujours pour des jonctions abruptes), ce qui a permis d'améliorer l'évaluation de la mobilité dans ce régime de transport.

D'autre part, la confrontation avec des données expérimentales a permis à la fois de valider le modèle théorique et de proposer une méthode d'extraction de la mobilité à canal long et de la somme des résistances d'accès et balistique.

Il reste des efforts à fournir pour améliorer l'évaluation de ces différents termes intervenant dans l'expression de la mobilité et pour mesurer localement l'effet des phénomènes de dégradation, toujours difficiles à découpler et à prédire. Pour étudier les MOSFET à canal court et améliorer les procédés de fabrication, il faudra prendre en compte les phénomènes de transport non stationnaire et quasi-balistique et améliorer la compréhension physique des résistances d'accès.

Annexes

1. Calcul des taux de transition

Les taux de transition sont obtenus en intégrant la probabilité de transition sur les états finaux possibles. Dans le cas 2Dk, ils s'écrivent :

$$\lambda(\mathbf{k}) = \iint S(\mathbf{k}, \mathbf{k}') J(E', \theta') dE' d\theta' \quad (123)$$

où le premier terme de l'intégrale $S(\mathbf{k}, \mathbf{k}')$ est calculé d'après la règle d'or de Fermi :

$$S(\mathbf{k}, \mathbf{k}') = \frac{1}{2\pi\hbar} \left| \iint e^{-i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{r}_{xz}} M'(\mathbf{k}, \mathbf{k}') \right|^2 \delta(E - E') \quad (124)$$

dans laquelle $M'(\mathbf{k}, \mathbf{k}')$ représente la matrice de transition :

$$M'(\mathbf{k}, \mathbf{k}') = \int \xi_m^*(\mathbf{y}) \xi_n(\mathbf{y}) H_i d\mathbf{y} \quad (125)$$

Le deuxième terme de l'intégrale double du taux de transition $\lambda(\mathbf{k})$ est le Jacobien :

$$J(E', \theta') = \frac{1 + 2\alpha E'}{\hbar^2} \left[\frac{\cos^2 \theta'}{m'_x} + \frac{\sin^2 \theta'}{m'_z} \right]^{-1} \quad (126)$$

1.1. Interaction avec les phonons

Les phonons correspondent aux vibrations du réseau cristallin.

1.1.1. Phonons acoustiques intravallées

Le taux de transition des phonons acoustiques intravallées vaut :

$$\lambda_{m,n}^{\text{ac}}(E) = \frac{k_B T D_{\text{ac}}^2}{2\hbar^3 \rho v_p^2} \sqrt{m'_x m'_z} \gamma(E') F_{mn} \quad (127)$$

1.1.2. Phonons intervallées

Les phonons intervallées sont de deux types.

1.1.2.i. Phonons d'ordre 0

Les phonons d'ordre 0 sont modélisés par l'équation :

$$\lambda_{m,n}^{\text{iv}0}(\mathbf{k}) = \frac{Z_{\text{iv}0} D_0^2 N_q + 1/2 \pm 1/2}{2\hbar \rho} \frac{1}{\hbar \omega} \sqrt{m'_x m'_z} \gamma(E') F_{mn} \quad (128)$$

où la notation abrégée $N_q + 1/2 \pm 1/2$ signifie $N_q + 1$ dans le cas de l'émission et N_q dans le cas de l'absorption.

1.1.2.ii. Phonons d'ordre 1

Les phonons d'ordre 0 sont modélisés par l'équation :

$$\lambda_{m,n}^{iv1}(\mathbf{k}) = \frac{Z_{iv1} D_1^2 N_q + \delta_{ab}}{\hbar \rho} \frac{1}{\hbar \omega} \sqrt{m'_x m'_z} \gamma(E') \left[\frac{1}{\hbar^2} \left(\gamma(E) \sqrt{m_x m_z} + \gamma(E') \sqrt{m'_x m'_z} \right) F_{mn} - \frac{G_{mn}}{2} \right] \quad (129)$$

dans laquelle l'intégrale de recouvrement vaut :

$$G_{mn} = \int_{y_1}^{y_2} \xi_m(y) \xi_n(y) \frac{\partial^2 \xi_m(y) \xi_n(y)}{\partial y^2} dy \quad (130)$$

1.2. Interactions coulombiennes avec les impuretés ionisées

Les impuretés ionisées sont modélisées par l'équation :

$$\frac{1}{\tau_{m,n}^{\text{imp}}(\mathbf{k})} = \frac{q^4 \gamma(E')}{8\pi \hbar^3 \varepsilon_{\text{Si}}^2} \int_0^{2\pi} \Theta(\beta) \int_{y_1}^{y_2} N_{\text{imp}}(y_0) \left| \int_{y_1}^{y_2} \frac{\xi_n(y) e^{-\Delta k(\beta, \theta_0) |y-y_0|} \xi_m(y)}{\Delta k(\theta) + \Delta k_{\text{scr}}(\theta)} dy \right|^2 dy_0 d\beta \quad (131)$$

où

$$\Theta(\beta) = \frac{1 - \cos \beta}{\frac{\cos^2(\beta + \theta_0)}{m'_x} + \frac{\sin^2(\beta + \theta_0)}{m'_z}} \quad (132)$$

Le modèle d'écrantage de Debye-Hückel revisité [FETTER74] a été utilisé :

$$\Delta k_{\text{scr}} = \frac{e^2 N_{\text{scr}} n_{2d}}{2\varepsilon_r k_B T} g_1 \left(\frac{\Delta k \lambda_{\text{th}}}{4\sqrt{\pi}} \right) \quad (133)$$

En revanche, les interactions coulombiennes à distance [FISCHETT01, GÁMIZ03] ne sont pas modélisées de manière spécifique.

1.3. Rugosité d'interface avec l'oxyde

Le taux d'interaction avec la rugosité d'interface est donné par :

$$\frac{1}{\tau_{m,n}^{\text{rug}}(\mathbf{k})} = \frac{\gamma(E')}{2\pi \hbar^3} \int_0^{2\pi} \Theta(\beta) |M_{m,n}^{\text{rug}}|^2 d\beta \quad (134)$$

où

$$|M_{m,n}^{\text{rug}}|^2 = \left| \int_{y_1}^{y_2} \xi_m(y) \left[B(y, \Delta k) - \frac{\pi^2 \hbar^2 n^2}{m_z^* T_{\text{Si}}^3} \delta_{m,n} \right] \xi_n(y) dy \right|^2 S(|\Delta k|) \quad (135)$$

La densité spectrale de rugosité $S(|\Delta k|)$ du modèle exponentiel [GOODNICK85] employé de manière usuelle [BUFLER10, GÁMIZ99] et ici s'élève à :

$$S(|\Delta k|) = \frac{\pi \Delta^2 \Lambda^2}{\left(1 + \frac{|\Delta k|^2 \Lambda^2}{2}\right)^{3/2}} \quad (136)$$

et la fonction d'autocorrélation vaut :

$$C(\mathbf{r}_{xz}) = \Delta^2 e^{-|\mathbf{r}_{xz}|/\Lambda} \quad (137)$$

L'autre modèle principalement utilisé est le modèle gaussien :

$$S(|\Delta k|) = \pi \Delta^2 \Lambda^2 \exp\left[-\frac{(\Lambda q)^2}{4}\right] \quad (138)$$

avec sa fonction d'autocorrélation :

$$C(\mathbf{r}_{xz}) = \Delta^2 e^{-(|\mathbf{r}_{xz}|/\Lambda)^2} \quad (139)$$

D'autres modèles plus complexes [GÁMIZ99] tendent à rendre compte avec précision de la rugosité de surface, en incluant notamment l'interaction avec la rugosité de l'oxyde enterré, et de l'écrantage des interactions qui y sont liées.

2. Calcul de la résistance balistique

La résistance balistique R_{bal} est calculée dans l'hypothèse de transport purement balistique en analysant le courant provenant de la source et du drain, obtenu par intégration des fonctions de distribution f_s et f_d sur tous les vecteurs d'onde \mathbf{k} prenant leurs valeurs dans les deux dimensions de la zone de Brillouin 2Dk.

Pour un canal planaire de dimensions $L \times L$, le courant venant de la source vaut :

$$-\frac{2e}{L} \frac{L^2}{(2\pi)^2} \int_{v_x > 0} d\mathbf{k} f_s(\mathbf{k}) v_x(\mathbf{k}) \quad (140)$$

Par symétrie $\mathbf{k} \leftrightarrow -\mathbf{k}$, le courant balistique total vaut donc :

$$I_{sd} = \frac{-2eL}{(2\pi)^2} \int_{v_x > 0} d\mathbf{k} (f_s(\mathbf{k}) - f_d(\mathbf{k})) v_x(\mathbf{k}) \quad (141)$$

Or, à faible V_{ds} , la différence entre les fonctions de distribution peut être simplifiée par un développement selon l'énergie :

$$f_s - f_d = -\frac{df}{dE} eV_{ds} \quad (142)$$

La résistance balistique est alors donnée par la relation :

$$\frac{1}{R_{\text{bal}}} = \frac{I_{ds}}{V_{ds}} = \frac{2e^2 L}{(2\pi)^2} \int_{v_x > 0} d\mathbf{k} \left(-\frac{df}{dE} \right) v_x(\mathbf{k}) \quad (143)$$

Cette expression peut être intégrée numériquement dans la zone de Brillouin 2Dk.

Elle peut aussi être calculée analytiquement dans l'hypothèse d'une sous-bande parabolique d'énergie E_n et de masses effectives m_x et m_z . Il est alors possible d'exploiter le remplacement de $d\mathbf{k}$ par $kdkd\theta$. En posant $E_n = 0$ pour simplifier la démonstration :

$$E = \frac{\hbar^2 k^2}{2} \left(\frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_z} \right) \quad (144)$$

Les expressions suivantes en découlent :

$$k = \frac{\sqrt{2E}}{\hbar} \left(\frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_z} \right)^{-\frac{1}{2}} \quad (145)$$

$$dk = \frac{dE}{\hbar \sqrt{2E}} \left(\frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_z} \right)^{-\frac{1}{2}} \quad (146)$$

$$v_x = \frac{\hbar k}{m_x} \cos \theta \quad (147)$$

Ces remplacements permettent d'intégrer le courant en fonction de l'énergie :

$$\frac{1}{R_{\text{bal}}} = \frac{2e^2L}{(2\pi)^2} \int_{\theta=-\pi/2}^{\pi/2} \int_{E=0}^{\infty} \frac{\sqrt{2E}}{\hbar^2 m_x} \left(-\frac{df}{dE} \right) \frac{\cos \theta}{\left(\frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_z} \right)^{3/2}} d\theta dE \quad (148)$$

$$= \frac{2\sqrt{2}e^2L}{(2\pi)^2 \hbar^2 m_x} \int_{\theta=-\pi/2}^{\pi/2} \frac{\cos \theta}{\left(\frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_z} \right)^{3/2}} d\theta \int_{E=0}^{\infty} \sqrt{E} \left(-\frac{df}{dE} \right) dE \quad (149)$$

Or l'intégrale sur l'énergie fait apparaître les intégrales de Fermi-Dirac d'ordre $\frac{1}{2}$:

$$\frac{1}{2} \int_{E=0}^{\infty} \frac{dE}{\sqrt{E}} f(E) = \frac{\sqrt{\pi kT}}{2} \mathcal{F}_{-1/2} \left(\frac{E_F - E_n}{kT} \right) \quad (150)$$

De plus, l'intégrale sur θ est connue, ce qui simplifie l'expression de la résistance en :

$$\frac{1}{R_{\text{bal}}} = \frac{2\sqrt{2}e^2L}{(2\pi)^2 \hbar^2 m_x} 2m_x \sqrt{m_z} \left[\frac{\sqrt{\pi kT}}{2} \mathcal{F}_{-1/2} \left(\frac{E_F - E_n}{kT} \right) \right] \quad (151)$$

$$= \frac{\sqrt{\pi m_y kT} e^2 L}{\sqrt{2} \pi^2 \hbar^2} \mathcal{F}_{-1/2} \left(\frac{E_F - E_n}{kT} \right) \quad (152)$$

Finalement, la résistance balistique vaut :

$$\frac{1}{WR_{\text{bal}}} = \sum_{\nu, n} \underbrace{\frac{e \sqrt{\pi m_y^\nu kT}}{\sqrt{2} \pi^2 \hbar^2}}_{\mu_{\text{bal}}} \times \underbrace{\left[\mathcal{F}_{1/2} \left(\frac{E_F - E_n^\nu}{kT} \right) - \mathcal{F}_{1/2} \left(\frac{E_F - eV_{ds} - E_n^\nu}{kT} \right) \right]}_{\text{Fermi-Dirac}} \frac{kT}{V_{ds}} \quad (153)$$

où W est la longueur de canal. Pour chaque vallée ν , m_z^ν est la masse effective transverse et E_n^ν est l'énergie de la sous-bande n , obtenue par simulation. $\mathcal{F}_{1/2}$ est l'intégrale de Fermi-Dirac d'ordre $\frac{1}{2}$.

L'énergie de Fermi utilise l'intégrale de Fermi-Dirac d'ordre 0 :

$$\mathcal{F}_0 \left(\frac{E_F - E_n^\nu}{k_B T} \right) = \ln \left(1 + e^{\frac{E_F - E_n^\nu}{k_B T}} \right) \quad (154)$$

Références

bibliographiques

- [ANDRIEU05] F. Andrieu et al., « Experimental and comparative investigation of low and high field transport in substrate- and process-induced strained nanoscaled MOSFETs », *Symposium on VLSI Technology. Digest of Technical Papers*, pp. 176-177, juin 2005.
- [APANOVIC94] Y. Apanovich, E. Lyumkis, B. Polsky, A. Shur et P. Blakey, « Steady-state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models », *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, pp. 702-711, juin 1994.
- [ARORA11] V. K. Arora, M. S. Zainal Abidin, S. Tembhurne, M. A. Riyadi, « Concentration dependence of drift and magnetoresistance ballistic mobility in a scaled-down metal-oxide semiconductor field-effect transistor », *Applied Physics Letters*, vol. 111, no. 5, p. 063106, mars 2011.
- [AUBRY-FO05] V. Aubry-Fortuna, P. Dollfus et S. Galdin-Retailleau, « Electron effective mobility in strained-Si/Si_{1-x}Ge_x MOS devices using Monte Carlo simulation », *Solid State Electronics*, vol. 49, no. 8, août 2005.
- [BARRAL08] V. Barral, « Étude, simulation et caractérisation du transport quasi-balistique dans les dispositifs nanométriques », Université de Provence, Thèse de doctorat 2008.
- [BIDAL09] G. Bidal, D. Fleury, G. Ghibaudo, F. Boeuf et T. Skotnicki, « Guidelines for MOSFET Device Optimization accounting for L-dependent Mobility Degradation », *Silicon Nanoelectronics Workshop*, 2009.
- [BLOTEKJA70] K. Blotekjaer, « Transport equations for electrons in two-valley semiconductors », *Electron Devices, IEEE Transactions on*, vol. 17, pp. 38-47, jan. 1970.
- [BUFLER10] F. M. Bufler et al., « Comparison of semiclassical transport formulations including quantum corrections for advanced devices with High-K gate stacks », *14th International Workshop on Computational Electronics*, pp. 1-4, oct. 2010.
- [CAROSELL08] F. Carosella, J. Saint-Martin, A. Bournel, S. Galdin-Retailleau, P. Dollfus, « Monte Carlo study of 2D electron gas transport including Pauli exclusion principle in highly doped silicon », *physica status solidi (c)*, vol. 5, no. 1, pp. 98-101, jan. 2008.
- [CASSÉ08] M. Cassé et al., « Mobility of strained and unstrained short channel FD-SOI MOSFETs: New insight by magnetoresistance », *IEEE Symposium on VLSI*

- Technology*, pp. 170-171, juin 2008.
- [CHAISANT06] M. W. Chaisantikulwat et al., « Differential magnetoresistance technique for mobility extraction », *Solid-State Electronics*, vol. 50, no. 4, pp. 637-643, avril 2006.
- [CROS06] A. Cros et al., « Unexpected mobility degradation for very short devices: a new challenge for CMOS scaling », *International Electron Devices Meeting*, 2006.
- [DOLLFUS97] P. Dollfus, « Si/Si_{1-x}Ge_x heterostructures: Electron transport and field-effect transistor operation using Monte Carlo simulation », *Journal of Applied Physics*, vol. 82, no. 8, pp. 3911-3916, oct. 1997.
- [DUPRÉ07] C. Dupré et al., « Carrier mobility degradation due to high dose implantation in ultrathin unstrained and strained silicon-on-insulator films », *Journal of Applied Physics*, vol. 102, no. 10, pp. 104505-104505-8, nov. 2007.
- [ERGINSOY50] C. Erginsoy, « Neutral Impurity Scattering in Semiconductors », *Physical Review*, vol. 79, pp. 1013-1014, 1950.
- [ESSENI11] D. Esseni, P. Palestri et L. Selmi, « Nanoscale MOS Transistors ». : Cambridge University Press, 2011.
- [FENOUILLO8] C. Fenouillet-Béranger et al., « FDSOI devices with thin BOX and ground plane integration for 32nm node and below », *38th European Solid-State Device Research Conference*, pp. 206-209, sept. 2008.
- [FETTER74] A. L. Fetter, « Electrodynamics and thermodynamics of a classical electron surface layer », *Phys. Rev. B*, vol. 10, no. 9, pp. 3739-3745, nov. 1974.
- [FISCHETT01] M. V. Fischetti et S. E. Laux, « Long-range Coulomb interactions in small Si devices. Part I: Performance and Reliability », *Journal of Applied Physics*, vol. 89, no. 2, pp. 1205-1231, 2001.
- [FLEURY09] D. Fleury, A. Cros, G. Bidal, J. Rosa et G. Ghibaudo, « A New Technique to Extract the Source/Drain Series Resistance of MOSFETs », *IEEE Electron Device Letters*, vol. 30, no. 9, pp. 975-977, sept. 2009.
- [GÁMIZ03] F. Gámiz, A. Godoy, J. B. Roldan, J. E. Carceller et P. Cartujo, « Effect of polysilicon depletion charge on electron mobility in ultrathin oxide MOSFETs », *Semiconductor Science and Technology*, vol. 18, no. 11, p. 927, 2003.
- [GÁMIZ99] F. Gámiz, J. B. Roldan, J. A. Lopez-Villanueva, P. Cartujo-Cassinello et J. E. Carceller, « Surface roughness at the Si-SiO₂ interfaces in fully depleted silicon-on-insulator inversion layers », *Journal of Applied Physics*, vol. 86, no. 12, pp. 6854-6863, 1999.
- [GHIBAUDO09] G. Ghibaudo et al., « Electrical transport characterization of nano CMOS devices with ultra-thin silicon film », *International Workshop on Junction Technology*, pp. 58-

63, juin 2009.

- [GHIBAUDO88] G. Ghibaudo, « New method for the extraction of MOSFET parameters », *Electronics Letters*, vol. 24, no. 9, pp. 543-545, avril 1988.
- [GOODNICK85] S. M. Goodnick et al., « Surface roughness scattering at the Si(100)-SiO₂ interface », *Phys. Rev. B*, vol. 32, no. 12, p. 8171, déc. 1985.
- [GREENWOOD58] D. A. Greenwood, « The Boltzmann Equation in the Theory of Electrical Conduction in Metals », *Proceedings of the Physical Society*, vol. 71, p. 585, avril 1958.
- [GRITSCH02] M. Gritsch, « Numerical Modeling of Silicon-on-Insulator MOSFETs », Technische Universität Wien, PhD thesis 2002.
- [GUARNAY14] S. Guarnay, F. Triozon, S. Martinie, Y.-M. Niquet et A. Bournel, « Monte Carlo study of effective mobility in short channel FDSOI MOSFETs », *IEEE International Conference on Simulation of Semiconductor Processes and Devices*, pp. 105-108, sept. 2014.
- [HAMER86] M. F. Hamer, « First-order parameter extraction on enhancement silicon MOS transistors », *IEE Proceedings I Solid-State and Elec. Dev.*, vol. 133, no. 2, pp. 49-54, avril 1986.
- [HUET07] K. Huet et al., « Monte Carlo Study of Apparent Mobility Reduction in Nano-MOSFETs », *European Solid State Device Research Conference*, no. 37, pp. 382-385, 2007.
- [HUET09] K. Huet, J. Saint-Martin, A. Bournel, D. Querlioz et P. Dollfus, « Effect of access resistance on apparent mobility reduction in nano-MOSFET », *10th International Conference on Ultimate Integration of Silicon*, pp. 35-38, mars 2009.
- [JAHAN05] C. Jahan et al., « QFETs transistors with TiN metal gate and HfO₂ down to 10nm », *VLSI Technology*, pp. 112-113, juin 2005.
- [JAUD06] Marie-Anne Jaud, « Effective Potential approach for the Monte-Carlo simulation of electronic transport with quantization effects in MOSFETs devices », Université Paris Sud - XI, Orsay, Thèse de doctorat 2006.
- [JUNGEMAN12] C. Jungemann, A.-T. Pham, S.-M. Hong et B. Meinerzhagen, « Deterministic simulation of 3D and quasi-2D electron and hole systems in SiGe devices », *Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, pp. 318-321, sept. 2012.
- [KIM08] R. Kim et M. Lundstrom, « Notes on Fermi-Dirac Integrals », *ArXiv e-prints*, nov. 2008.

- [KIRÉEV75] P. Kiréev, « La Physique des Semiconducteurs », *Mir*. Moscou, 1975.
- [KUBO57] R. Kubo, « Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems », *Journal of the Physical Society of Japan*, vol. 12, no. 6, pp. 570-586, juin 1957.
- [LIN11] Y.-M. Lin et al., « Wafer-Scale Graphene Integrated Circuit », *Science*, vol. 332, no. 6035, pp. 1294-1297, juin 2011.
- [LUNDSTRO00] M. Lundstrom, « Fundamentals of Carrier Transport », *Second Edition - Cambridge University Press.*, 2000.
- [MONSEF04] F. Monsef, P. Dollfus, S. Galdin-Retailleau, H.-J. Herzog et T. Hackbarth, « Electron transport in Si/SiGe modulation-doped heterostructures using Monte Carlo simulation », *Journal of Applied Physics*, vol. 95, no. 7, pp. 3587-3593, mars 2004.
- [MORVAN12] S. Morvan et al., « Efficiency of mechanical stressors in Planar FDSOI n and p MOSFETs down to 14nm gate length », *Symposium on VLSI Technology*, pp. 111 - 112, juin 2012.
- [PARK01] J.-T. Park, J. Colinge et C. H. Diaz, « Pi-Gate SOI MOSFET », *IEEE Electron Device Letters*, vol. 22, no. 8, pp. 405-406, août 2001.
- [PHAM-NGU09] L. Pham-Nguyen, C. Fenouillet-Beranger, A. Vandooren et T. Skotnicki, « In Situ Comparison of Si/High- κ and Si/SiO₂ Channel Properties in SOI MOSFETs », vol. 30, no. 10, pp. 1075-1077, oct. 2009.
- [PHAM-NGU08] L. Pham-Nguyen et al., « Direct comparison of Si/High-K and Si/SiO₂ channels in advanced FD SOI MOSFETs », *IEEE International SOI Conference*, pp. 25-26, oct. 2008.
- [RAMOS06] J. Ramos et al., « Effective Mobility Extraction Based on a Split RF C-V Method for Short-Channel FinFETs », *Proceeding of the 36th European Solid-State Device Research Conference*, pp. 363 -366, sept. 2006.
- [REN01] Zhibin Ren, « Nanoscale MOSFETs: Physics, Simulation and Design », Purdue University, PhD thesis 2001.
- [RIDEAU14] D. Rideau et al., « Experimental and theoretical investigation of the ‘apparent’ mobility degradation in Bulk and UTBB-FDSOI devices: a focus on the near-spacer-region resistance. », *IEEE International Conference on Simulation of Semiconductor Processes and Devices*, pp. 101-104, sept. 2014.
- [ROMANJEK04] K. Romanjek, « Caractérisation et modélisation des transistors CMOS des technologies 50nm et en deçà », Institut Polytechnique de Grenoble, Thèse de doctorat 2004.

- [SAINT-MA04] J. Saint-Martin, A. Bournel et P. Dollfus, « On the ballistic transport in nanometer-scaled DG MOSFETs », *IEEE Transactions on Electron Devices*, vol. 51, no. 7, pp. 1148-1155, juil. 2004.
- [SAINT-MA06] J. Saint-Martin, A. Bournel, F. Monsef, C. Chassat et P. Dollfus, « Multi sub-band Monte Carlo simulation of an ultra-thin double gate MOSFET with 2D electron gas », *Semiconductor Science and Technology*, vol. 21, no. 4, p. L29, avril 2006.
- [SAINT-MA05] J. Saint-Martin, « Étude par simulation Monte Carlo d'architectures de MOSFET ultracourts à grille multiple sur SOI », Université Paris XI, Orsay, Thèse de doctorat 2005.
- [SHUR02] M. S. Shur, « Low ballistic mobility in submicron HEMTs », *IEEE Electron Device Letters*, vol. 23, no. 9, pp. 511-513, sept. 2002.
- [TAKAGI94] S. Takagi, A. Toriumi, M. Iwase et H. Tango, « On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration », *IEEE Transactions on Electron Devices*, vol. 41, no. 12, pp. 2357-2362, déc. 1994.
- [VENUGOPA02] R. Venugopal, Z. Ren, S. Datta, M.S. Lundstrom et D. Jovanovic, « Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches », *Journal of Applied Physics*, vol. 92, no. 7, pp. 3730-3739, oct. 2002.
- [VINET09] M. Vinet et al., « Self-Aligned Planar Double-Gate MOSFETs by Bonding for 22-nm Node, With Metal Gates, High- κ Dielectrics, and Metallic Source/Drain », *IEEE Electron Device Letters*, vol. 30, no. 7, pp. 748-750, juil. 2009.
- [WEBER10] O. Weber et al., Work-function Engineering in Gate First Technology for Multi-VT Dual-Gate FDSOI CMOS on UTBOX, 2010.
- [WETTSTEI01] A. Wettstein, A. Schenk et W. Fichtner, « Quantum device-simulation with the density-gradient model on unstructured grids », *IEEE Transactions on Electron Devices*, vol. 48, no. 2, pp. 279-284, fév. 2001.
- [WONG99] H.-S. P. Wong, D. J. Frank, P. M. Solomon, C. H. Wann et J. J. Welser, « Nanoscale CMOS », *Proceedings of the IEEE*, vol. 87, no. 4, pp. 537-570, avril 1999.
- [YU97] B. Yu, C. H. Wann, E. D. Nowak, K. Noda et C. Hu, « Short-channel effect improved by lateral channel-engineering in deep-submicrometer MOSFETs », *Electron Devices, IEEE Transactions on*, vol. 44, no. 4, pp. 627-634, avril 1997.
- [ZHAO09] Y. Zhao, M. Takenaka et S. Takagi, « On Surface Roughness Scattering-Limited Mobilities of Electrons and Holes in Biaxially Tensile-Strained Si MOSFETs », *IEEE Electron Device Letters*, vol. 30, no. 9, pp. 987-989, sept. 2009.
- [ZILLI07] M. Zilli, D. Esseni, P. Palestri et L. Selmi, « On the Apparent Mobility in Nanometric

n-MOSFETs », *IEEE Elec. Dev. Lett.*, vol. 28, no. 11, pp. 1036-1039, Nov. 2007.