



**HAL**  
open science

# ETUDE PAR SIMULATION MONTE CARLO D'ARCHITECTURES DE MOSFET ULTRACOURTS A GRILLE MULTIPLE SUR SOI

Jérôme Saint-Martin

► **To cite this version:**

Jérôme Saint-Martin. ETUDE PAR SIMULATION MONTE CARLO D'ARCHITECTURES DE MOSFET ULTRACOURTS A GRILLE MULTIPLE SUR SOI. Micro et nanotechnologies/Microélectronique. Université Paris Sud - Paris XI, 2005. Français. NNT: . tel-00011335

**HAL Id: tel-00011335**

**<https://theses.hal.science/tel-00011335>**

Submitted on 10 Jan 2006

**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.

ORSAY

N° D'ORDRE : 8107

UNIVERSITÉ DE PARIS-SUD  
U.F.R. SCIENTIFIQUE D'ORSAY

THÈSE  
présentée  
pour obtenir

**LE GRADE DE DOCTEUR EN SCIENCES  
DE L'UNIVERSITÉ PARIS XI ORSAY**

**Spécialité Electronique**

PAR

Jérôme SAINT-MARTIN

**Sujet : ETUDE PAR SIMULATION MONTE CARLO D'ARCHITECTURES  
DE MOSFET ULTRACOURTS A GRILLE MULTIPLE SUR SOI**

Soutenu le 2 décembre 2005 devant la commission d'examen

MM.	Frederic BOUILLAULT	Président
	Arnaud BOURNEL	
	Fabian M. BUFLER	
	Mireille MOUIS	
	Olivier VANBÉSIEN	Rapporteur
	Luca VARANI	Rapporteur







# Remerciements

A l'issue de cette thèse, je souhaite tout d'abord remercier Monsieur Jean Michel LOURTIOZ, directeur de l'IEF (Institut d'Electronique Fondamentale), pour m'avoir accueilli au sein de son laboratoire.

J'adresse mes plus vifs remerciements à Monsieur Olivier VANBESIEN, Professeur à l'université de Lille 1, et à Monsieur Luca VARANI, Chargé de Recherche au CNRS au CEM2 de Montpellier, pour l'intérêt qu'ils ont porté à cette étude et pour m'avoir fait l'honneur d'en être les rapporteurs.

Je suis heureux que Monsieur le professeur Frédéric BOUILLAULT du LGEP, M. Fabian BUFLER ingénieur Synopsis à Zurich et Madame Mireille MOUIS directrice de recherche au CNRS à l'IMEP me fassent l'honneur de participer au jury.

Je remercie Monsieur Simon DELEONIBUS pour m'avoir accueilli pendant quelques semaines au sein du CEA/LETI et toute son équipe pour m'avoir permis d'aborder sereinement les aspects expérimentaux.

J'adresse toute ma reconnaissance à Monsieur Philippe DOLLFUS responsable du département CMO (Composant pour la Micro et l'Opto-électronique), pour m'avoir donné les moyens de mener à bien ce travail. Ses précieux conseils m'ont été très précieux, et m'ont, entre autres, permis de m'initier aux arcanes du logiciel MONACO...

C'est avec une profonde sincérité que j'adresse toute ma reconnaissance à Monsieur Arnaud BOURNEL pour m'avoir confié cette étude. Ses grandes qualités humaines, sa disponibilité à toute épreuve et ses remarquables compétences ont rendu ce travail particulièrement enrichissant. Je voudrais aussi le remercier pour la confiance qu'il m'a accordée tout au long de ces trois années.

Je tiens aussi à remercier Madame Sylvie RETAILLEAU, ma tutrice de Monitorat, qui a grandement facilité mes premiers pas dans l'enseignement, ainsi que les différentes équipes enseignantes qui m'ont accueilli.

J'aimerais également remercier chaleureusement les autres membres permanents de l'équipe que j'ai côtoyés quotidiennement pendant ces trois années M. Christophe CHASSAT, Mme Valérie AUBRY-FORTUNA et Mme Françoise MEYER.

J'adresse toutes mes amitiés aux doctorants et jeunes docteurs avec qui j'ai partagé de très bons moments dans le bureau 113 : Hugues CAZIN, Van Nam DO, Johann SEE, Abdennour HATTAB et Florian MONSEF. J'ai eu avec eux de nombreuses discussions toujours très intéressantes et souvent fructueuses. J'ai également une pensée chaleureuse pour les exilés de la « 119 » : Audrey

VALENTIN, Damien QUERLIOZ et Karim HUET. Je ne manquerais pas d'envoyer toute ma sympathie à la joyeuse, mais très efficace, équipe des JNRDM 2005.

Enfin, tous mes remerciements à Virginie qui m'a supporté durant ces années de thèse, à ma famille et à mes amis qui m'ont aussi bien soutenu pour franchir la dernière étape de mes études !  
Du moins en sciences...

# Table des Matières

<b>Remerciements</b>	<b>5</b>
<b>Table des Matières</b>	<b>7</b>
<b>Constantes, symboles, abréviations</b>	<b>11</b>
<b>Introduction générale :</b>	<b>17</b>
<b>Chapitre I :</b>	<b>19</b>
<b>Introduction aux transistors MOS</b>	<b>19</b>
<b>1. Les transistors à effet de champ</b>	<b>20</b>
1.1 Historique : de l'essor à l'âge d'or	20
1.2 Comment ça marche ?	20
1.3 Principaux paramètres des MOSFET	24
1.4 Transistor MOS : schéma équivalent	25
1.5 Transistor et inverseur CMOS	27
<b>2. Fin du CMOS conventionnel</b>	<b>31</b>
2.1 Loi d'échelle	31
2.2 La « Roadmap »	33
2.3 Effets de canal court	34
2.4 Effet de canal court et caractéristiques électriques	39
2.5 Solutions conventionnelles	40
2.6 Effets de canal ultracourt	41
2.7 Dispositifs émergents	42
<b>3. Présentation des architectures multigrille</b>	<b>44</b>
3.1 La technologie SOI	44
3.2 Transistor MOS sur SOI	44
3.3 L'inversion volumique	45
3.4 Les architectures multigrilles	46
3.5 Comparaison actuelle des différentes architectures MOS	51
3.6 Nouvelles logiques	52
<b>4. Objectifs de l'étude</b>	<b>53</b>
<b>Chapitre II :</b>	<b>55</b>
<b>Transport balistique dans les dispositifs ultracourts</b>	<b>55</b>
<b>1. Description du transport électronique</b>	<b>56</b>
1.1 Equation de transport	56
1.2 Approximations usuelles de l'équation de Boltzmann	59
1.3 Modèles analytiques du transport fortement non stationnaire	62
<b>2. Mise en évidence du transport non stationnaire</b>	<b>71</b>
2.1 Dispositifs simulés	71
2.2 Dépassement de la vitesse de saturation stationnaire	72
2.3 Spectres en vitesse et comptage d'interactions	74
<b>3. Evolution de la balisticité</b>	<b>78</b>
3.1 Définitions de $B_{int}$ et $B_{eff}$	78
3.2 Paramètres physiques	79
3.3 Relation entre les balisticités effective et intrinsèque	81
3.4 Paramètres électriques	82

<b>4. Localisation des interactions et courant</b>	<b>85</b>
4.1 Importance des interactions en fin de canal	85
4.2 Localisation des interactions dans le canal	86
4.3 Evolution du coefficient de rétrodiffusion	87
4.4 Rétrodiffusion, hauteur de barrière et concentration en entrée de canal	90
<b>5. Mesures de balisticité</b>	<b>91</b>
5.1 Bases théoriques	91
5.2 Dispositif expérimental	92
5.3 Résultats	93
<b>Chapitre III :</b>	<b>97</b>
<b>Dimensionnement des transistors MultiGrilles</b>	<b>97</b>
<b>1. Nouvelles lois d'échelle</b>	<b>97</b>
<b>2. Présentation des dispositifs simulés</b>	<b>98</b>
<b>3. Normalisation du courant des architectures 3D</b>	<b>99</b>
<b>4. Influence de l'architecture sur le transport</b>	<b>100</b>
<b>5. Caractéristiques statiques</b>	<b>101</b>
5.1 Etat passant	101
5.2 Etat bloqué	102
5.3 Simulations et modèle analytique	103
5.4 Analyse des effets de canal court	104
5.5 Bilan et rapports $I_{on}/I_{off}$	109
<b>6. Evolution du paramètre « CV/I »</b>	<b>110</b>
<b>7. Influences des paramètres géométriques</b>	<b>111</b>
7.1 Influence de la largeur du film	111
7.2 Longueur des caissons et impédances parasites	112
<b>8. Bilan</b>	<b>115</b>
<b>Chapitre IV :</b>	<b>117</b>
<b>Prise en compte de la quantification perpendiculaire</b>	<b>117</b>
<b>1. Influence de la quantification perpendiculaire</b>	<b>118</b>
1.1 Résolution Schrödinger/Poisson 1D	118
1.2 Nouvelle répartition des porteurs par vallée équivalente	122
1.3 Modification du contrôle électrostatique	123
<b>2. Etude du transport en gaz 2Dk</b>	<b>125</b>
2.1 La méthode Monte Carlo à champ constant	125
2.2 Interactions avec le gaz 2Dk	126
2.3 Mobilités	137
<b>3. Le simulateur de dispositifs MONACO 2Dk</b>	<b>142</b>
3.1 Les simulations de dispositifs « quantiques »	143
3.2 Présentation MONACO 2Dk	146
3.3 Résultats obtenus avec MONACO 2Dk	153
<b>4. Conclusions et perspectives :</b>	<b>162</b>
<b>Conclusion générale</b>	<b>165</b>
<b>ANNEXES</b>	<b>169</b>
<b>ANNEXE 1 : Modèle balistique : cas dégénéré</b>	<b>171</b>
<b>ANNEXE 2 : Calcul du coefficient de rétrodiffusion</b>	<b>173</b>
<b>ANNEXE 3 : Approche par modes ou « mode-space »</b>	<b>175</b>
<b>ANNEXE 4 : Tirages au sort dans MONACO</b>	<b>177</b>

<i>ANNEXE 5 : Interactions avec les impuretés ionisées</i>	<u>181</u>
<i>ANNEXE 6 : Résolution de l'équation de Schrödinger</i>	<u>183</u>
<i>Références bibliographiques</i>	<u>185</u>
<i>Publications liées à ce travail</i>	<u>195</u>



# *Constantes, symboles, abréviations*

<b>Constantes</b>	<b>Significations/Valeurs</b>	<b>Unités</b>
$\epsilon_0$	Permittivité diélectrique du vide, $\epsilon_0 = 8,85.10^{-12}$ F/m	F/m
$\epsilon_{Si}$	Permittivité diélectrique du silicium, $\epsilon_{Si} = 11,8.\epsilon_0$	F/m
$\epsilon_{SiO_2}$	Permittivité diélectrique de l'oxyde de silicium, $\epsilon_{SiO_2} = 3,9.\epsilon_0$	F/m
$\hbar$	Constante de Dirac, $\hbar = 1,054 \times 10^{-34}$ kg.m <sup>2</sup> /s	kg.m <sup>2</sup> /s
$k_B$	Constante de Boltzmann, $k_B = 8,617385.10^{-5}$ eV/K	eV/K
$m_0$	Masse de l'électron, $m_0 = 9,109389.10^{-31}$ kg	kg
$m_l$	Masse longitudinale, $m_l = 0,916.m_0$ pour les électrons dans le silicium	kg
$m_t$	Masse transverse, $m_t = 0,191.m_0$ pour les électrons dans le silicium	kg
$q$	Charge élémentaire, $q = 1,602.10^{-19}$ C	C

<b>Symboles</b>	<b>Significations</b>	<b>Unités</b>
$ \Delta k $	Différence des vecteurs d'onde avant et après interaction	1/m
$\epsilon_{SC}$	Permittivité diélectrique d'un semi-conducteur	F/m
$\epsilon_{ox}$	Permittivité diélectrique de l'oxyde.	F/m
$\lambda$	Longueur onde	1/m
$\lambda_{th}$	Longueur onde thermique	1/m
$\lambda_{tot}$	Fréquence d'interaction totale	1/s
$\lambda(E)$	Fréquence d'interaction	1/s
$\mu_{eff}$	Mobilité effective	m <sup>2</sup> /Vs
$\xi_n(z)$	Fonction d'onde associée au niveau n	1/m <sup>2</sup>
$\rho$	Densité volumique du matériau	kg/m <sup>3</sup>
$\rho(z)$	Densité de charges électrique totale	C
$\tau(E)$	Temps de relaxation des moments	s
$\Phi_m, \Phi_s$	Travaux de sortie respectivement d'un métal, d'un semi-conducteur	eV
$\phi$	Potentiel électrostatique	eV
$\phi_B$	Différence entre le niveau de Fermi et le quasi niveau de Fermi en bandes plates	eV

$\phi_D$	Potentiel de diffusion de la jonction source/substrat $\phi_D = \frac{k_B T}{q} \ln \frac{N_A N_D}{n_i^2}$	eV
$\Phi_f$	potentiel d'inversion	eV
$\Psi_b$	Fonction de Bloch	
$\psi$	hauteur de barrière en entrée de canal de canal	eV
$\omega$	Pulsation d'un phonon	rad/s
$B_{\text{eff}}$	Balisticité effective : $B_{\text{eff}} = I_{\text{on}}/I_{\text{on\_bal}}$	%
$B_{\text{int}}$	Balisticité intrinsèque : pourcentage d'électrons balistiques en fin de canal	%
$C_{\text{eq}}$	Capacité surfacique équivalente de grille	F/m <sup>2</sup>
$C_G$	Capacité de grille	F
$C_{\text{ox}}$	Capacité d'oxyde par unité de surface, $C_{\text{ox}} = \epsilon_{\text{ox}}/T_{\text{ox}}$	F/m <sup>2</sup>
$C_{\text{ZCE}}$	Capacité de substrat par unité de surface associée à la zone de charge d'espace c'est-à-dire à la couche de déplétion.	F/m <sup>2</sup>
$C_{\text{inv}}$	Capacité d'inversion ou de la couche d'inversion par unité de surface	F/m <sup>2</sup>
$D_0$	Potentiel de déformation associé aux phonons intervalles d'ordre 0	J
$D_1$	Potentiel de déformation associé aux phonons intervalles d'ordre 1	J
$D_{\text{ac}}$	Potentiel de déformation associé aux phonons acoustiques	J
$D_n$	Coefficient de diffusion des électrons	m <sup>2</sup> /s
$D_p$	Coefficient de diffusion des trous	m <sup>2</sup> /s
$E$	Energie cinétique (avant interaction)	J
$E'$	Energie cinétique après interaction	J
$\vec{E}$	Champ électrique	V/m
$E_{\text{BC}}$	Energie de la bande de conduction	eV
$E_{\text{BV}}$	Energie de la bande de valence	eV
$E_c$	Energie cinétique	J
$E_{\text{eff}}$	Champ effectif : $E_{\text{eff}} = \frac{\int_{z_{\text{int}}}^{z_{\text{int}}+T_{\text{Si}}} \mathbf{n}(z) \times E_z(z) dz}{\int_{z_{\text{int}}}^{z_{\text{int}}+T_{\text{Si}}} \mathbf{n}(z) dz}$	V/m
$E_F$	Niveau de Fermi	eV
$E_m$	Energie quantifiée associée au niveau 'm'	eV
$E_n$	Energie quantifiée associée au niveau 'n'	eV
$E_x$	Champ électrique parallèle au transport	V/m
$E_z$	Champ électrique perpendiculaire au transport	V/m

$\vec{F}$	Force extérieure appliquée	N
f	Fréquence	Hz
$f(\vec{r}, \vec{p}, t)$	Fonction de distribution	
$G_d$	Conductance de drain	S ou S/m
$G_m$	Transconductance de sortie	S ou S/m
$g_v$	Nombre de dégénérescence de la vallée v	
$I(k, k')$	Fonction de recouvrement des fonctions d'enveloppe	
$I_D$	Courant de drain	A ou A/m
$I_{DSsat}$	Courant de drain de saturation	A ou A/m
$I_{off}$	Courant de drain $I_D$ à $V_{DS} = V_{DD}$ et $V_{GS} = 0$ V	A ou A/m
$I_{on}$	Courant de drain $I_D$ à $V_{DS} = V_{GS} = V_{DD}$	A ou A/m
$J(\vec{r}, t)$	Densité de courant au point $\vec{r}$ à l'instant t	A/m <sup>2</sup>
$J(E, \theta)$	Jacobien de transformation $\vec{k} \rightarrow E$ (2Dk)	
$\vec{k}$	Vecteur d'onde	m <sup>-1</sup>
$\vec{k}'$	Vecteur d'onde après interaction	m <sup>-1</sup>
$k_{x/y/z}$	Composante du vecteur d'onde dans la direction x, y ou z	
$L_{eff}$	Longueur de la zone rétrodiffusante du canal	m
$L_G$	Longueur de la grille d'un transistor	m
$L_z$	Longueur de la fenêtre ondulatoire (ou tranche Schrödinger)	m
lpm	Libre parcours moyen des porteurs	m
$M(k, k')$	Matrice de transition	SI
$M_v$	Dégénérescence du premier niveau	
$m_{x/y/z}$	Masse effective dans la direction x, y ou z	kg
$m'_{x/y/z}$	Masse effective dans la direction x, y ou z après interaction	kg
$N_A$	Concentration de dopants de type accepteur d'électrons	cm <sup>-3</sup>
$N_B$	Dopage du substrat	at./cm <sup>3</sup>
$N_D$	Concentration de dopants de type donneur d'électrons	cm <sup>-3</sup>
$N_q$	Nombre moyen de phonons d'énergie $\hbar\omega$	
$N_{inv}$	Charge surfacique d'inversion	cm <sup>-2</sup>
$n(\vec{r}, t)$	Concentration d'électrons au point $\vec{r}$ à l'instant t	cm <sup>-3</sup>
$n_{2D}$	Concentration d'électron 2Dk	cm <sup>-3</sup>

$n_i$	Concentration intrinsèque de porteurs libres	$\text{cm}^{-3}$
$P_E(\beta)$	Densité de probabilité d'avoir une interaction avec un angle de déviation de $\beta$	1/rad
$p(\vec{r}, t)$	Concentration de trous au point $\vec{r}$ à l'instant t	$\text{cm}^{-3}$
$p_{2D}$	Concentration de trous 2Dk	$\text{cm}^{-3}$
$Q_{ZCE}$	Densité surfacique de charges dans la couche de déplétion	$\text{C.m}^{-2}$
$Q_{inv}$	Densité surfacique de charges dans la couche d'inversion	$\text{C.m}^{-2}$
S	Pente sous le seuil	mV/dec
$s(\vec{k}, \vec{k}')$	Densité de probabilité qu'une interaction modifie le vecteur onde $\vec{k}$ en $\vec{k}'$	
$s(\vec{r}, \vec{p}, t)$	taux net de génération et/ou recombinaison par unité de temps	
T	Température	K
$T_{ox}$	Epaisseur d'oxyde	m
$T_{Si}$	Epaisseur de la zone active de silicium	m
$T_{ZCE}$	Epaisseur de la zone de charge d'espace dans le substrat de la capacité MOS	m
$u_k(\vec{r})$	Fonction d'enveloppe de la fonction de Bloch	SI
V	Potentiel électrostatique	eV
$V_{DS}$	Tension Drain/Source	V
$V_{DD}$	Tension d'alimentation	V
$V_{DSsat}$	Tension Drain/Source à partir de laquelle a lieu la saturation du courant $I_D$	V
$V_{FB}$	Tension de bande plate (« flat band » en anglais)	V
$V_{GS}$	Tension Grille/Source	V
$V_p$	Tension de percement volumique	V
$v_p$	Vitesse de propagation d'une onde acoustique	m/s
$v_{sat}$	Vitesse de saturation du transport stationnaire	m/s
$V_T$	Tension de seuil	V
W	Largeur du transistor	m
x	Direction selon l'axe x source-drain	m
y	Direction selon l'axe y de la profondeur de transistor	m
z	Direction selon l'axe z, direction perpendiculaire au transport	m
$z_1, z_2$	Début et fin de la fenêtre ondulatoire (ou tranche Schrödinger)	m
$z_{int}$	Position de l'interface rugueuse	m

<b>Cigle/Abréviation</b>	<b>Signification</b>
DG, DGMOS	« Double Gate MOSFET » : Transistor MOS à double grille
DIBL	« Drain Induced Barrier Lowering » : Abaissement de la barrière d'injection source/drain due à la tension de drain
ITRS	« International Technology Roadmap of Semiconductor » : feuille de route internationale des semi-conducteurs.
NMOS	Transistor à effet de champ MOS à canal N
MEB	Microscope électronique à balayage
MOS	Métal/Oxyde/Semi-conducteur
MOSFET	« MOS Field Effect Transistor » : transistor à effet de champ MOS
PMOS	Transistor à effet de champ MOS à canal P
SG, SGMOS	« Single Gate MOSFET » : Transistor MOS Simple Grille
SOI	« Silicon On Insulator » : Substrat silicium sur isolant
STI	« shallow trench isolation »
TCAD	« Technology Computer Aided Design »
TG, TGMOS	« Triple Gate MOSFET » : Transistor MOS Triple Grille
TEM	Microscopie par transmission électronique
QG	« Quadruple Gate MOSFET » : Transistor MOS Quadruple Grille
VMT	« Velocity Modulation Transistor » : Transistor à vitesse modulée
ZCE	Zone de charge d'espace



## *Introduction générale :*

Depuis l'invention du premier transistor en 1947, puis celle du circuit intégré en 1958, par Jack Kilby, les progrès de la microélectronique ont été considérables, tant au niveau de l'amélioration des performances et de l'accroissement de la complexité des circuits que de la baisse des coûts de production.

Les transistors MOS (Métal Oxyde Semi-conducteur) sur silicium utilisés en architecture CMOS (Complementary MOS : MOS Complémentaires) sont les principaux artisans de cette progression continue et dominante de façon écrasante le marché des semi-conducteurs. L'amélioration des performances de ces transistors nécessite toujours plus d'imagination de la part des concepteurs de composants. En effet, les gains en performance attendus de la miniaturisation des dispositifs sont fortement réduits à cause de phénomènes inhérents aux grilles ultrasubmicroniques actuelles (inférieures à 100 nm).

L'importance de ces effets, dits « de canal court », et la complexité des méthodes pour les contrer sont telles que le monde de la microélectronique, pourtant très frileux, est en train de s'intéresser très activement à toutes les architectures alternatives possibles aux transistors MOSFET (Metal Oxide Semiconductor Field Effect Transistor : transistor à effet de champ MOS) massifs 'traditionnels'. Ainsi, pour atteindre les prochains objectifs fixés par la feuille de route internationale des semi-conducteurs (ITRS : International Technology Roadmap of Semiconductor), l'utilisation d'architectures MOSFETs multigrille sur SOI (Silicon On Insulator : Silicium sur Isolant) est très sérieusement envisagée [COLI04]. Ces architectures sont, en effet, particulièrement prometteuses en terme de contrôle électrostatique de canal.

Dans ce contexte, cette thèse a été consacrée à l'étude des transistors MOS multigrilles par simulation particulière Monte Carlo. Les simulations particulières Monte Carlo classiques ont été réalisées avec le logiciel MONACO développé dans le département CMO (Composants pour la Microélectronique et l'Optoélectronique) de l'IEF (Institut d'Electronique Fondamentale).

Le premier chapitre de ce mémoire présente les transistors à effet de champ de type MOS, et décrit l'intérêt de leur miniaturisation pour les performances des applications logiques. La situation actuelle permettant l'émergence des architectures multigrilles est particulièrement soulignée.

Le second chapitre est entièrement dédié à l'analyse du transport électronique. En effet, les lois d'échelle qui encadrent la miniaturisation des composants reposent sur l'hypothèse d'un régime de transport quasi stationnaire. Pour les transistors ultracourts, des approches analytiques alternatives ont été proposées, afin de refléter l'influence de plus en plus grande des porteurs balistiques. La validité de ces modèles dans les dispositifs de longueur de grille décanométrique est discutée

grâce à une analyse fine du transport électronique, naturellement permise par l'algorithme Monte Carlo. Les porteurs « balistiques » sont particulièrement étudiés, afin d'estimer le gain en courant que peut apporter une telle miniaturisation.

Dans la troisième partie, les performances électriques des nouveaux dispositifs à grille multiple sont évaluées en détail. Pour optimiser le dimensionnement de ces dispositifs, nous nous sommes intéressés en particulier aux effets de canal court et aux valeurs de différentes impédances parasites qui tendent à augmenter dramatiquement à cause des lois d'échelle.

Le dernier chapitre décrit une nouvelle version très novatrice du logiciel MONACO développée durant cette thèse. Cette version « semi-quantique » permet la prise en compte des effets de quantification d'énergie qui ont lieu perpendiculairement à la direction du transport de charges dans les structures double-grille. En effet, leurs zones actives ne seront pas seulement ultracourtes mais aussi ultrafines. Les premiers résultats concernant de tels dispositifs à double grille sont présentés.

Enfin, les annexes 1 et 2 regroupent des détails de calculs relatifs au chapitres II, celles numérotées de 3 à 6 apportent des précisions vis-à-vis de la nouvelle version de MONACO qui fait l'objet du chapitre IV.

## Chapitre I :

### Introduction aux transistors MOS

Depuis l'antique boulier chinois (cf. Figure 1), en passant par l'additionneur mécanique inventé par Blaise Pascal en 1642 et la machine à recenser de Hollerith en 1890, la puissance de calcul n'a jamais cessé d'être améliorée. Les systèmes électroniques ont supplanté depuis longtemps les systèmes mécaniques dans cette course à la performance. Cette augmentation de la vitesse de calcul a naturellement provoqué une réduction du coût de calcul. La Figure 2 retrace cette fulgurante baisse des prix au cours du vingtième siècle, à travers l'augmentation colossale de la vitesse de calcul à prix constant. Elle a permis la diffusion à grande échelle de systèmes très complexes améliorant notre quotidien : ordinateurs, téléphones portable, ABS... Les tubes à vide puis les transistors, d'abord discrets puis intégrés en grande quantité sur des « puces », ont permis cette croissance sur-logarithmique.

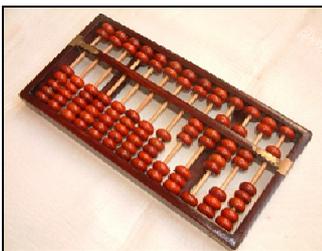


Figure 1 :  
Photographie d'un boulier  
chinois traditionnel.

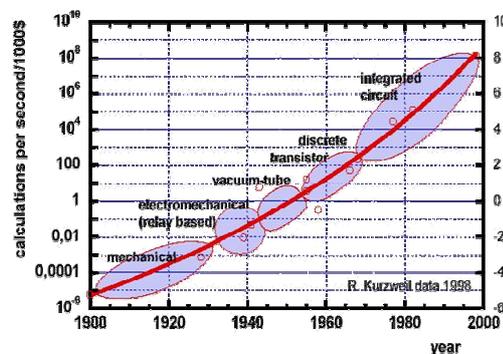


Figure 2 : Evolution de la vitesse de calcul  
à prix constant au cours du vingtième siècle  
([www.penguininputnam.com/static/packages/us/kur-zweil/excerpts/chap1/ch1note19.htm](http://www.penguininputnam.com/static/packages/us/kur-zweil/excerpts/chap1/ch1note19.htm)).

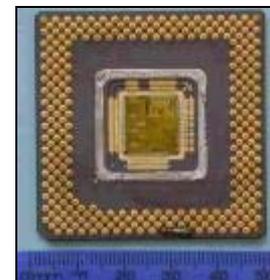


Figure 3 : Photo de  
processeur de type Pentium : puce  
( $\approx 1 \text{ cm}^2$ ) montée sur son boîtier en  
céramique.

A l'heure actuelle, les acteurs de cette accélération sont les transistors de type MOS (Métal Oxyde Semiconducteur). Ce sont les briques de base des microprocesseurs, tels le Pentium d'INTEL présenté en Figure 3, et autres microcontrôleurs qui autorisent sur une surface de quelques centimètres carrés des calculs toujours plus complexes, à des prix toujours plus faibles.

Dans ce premier chapitre, nous allons commencer par rappeler les principales caractéristiques physiques et électriques des transistors à effet de champ de type MOS. Nous dégagerons ainsi les motivations de la course effrénée à la miniaturisation des composants. Ensuite, nous détaillerons les

limites actuelles de cette miniaturisation et nous clôturerons ce chapitre en présentant les très prometteuses structures multigrilles.

## **1. Les transistors à effet de champ**

### **1.1 Historique : de l'essor à l'âge d'or**

Les concepts du transistor à effet de champ ont été brevetés par Lilienfeld et Heil en 1930. Cependant des difficultés technologiques très importantes ont retardé sa réalisation pratique. En effet, il n'apparaîtra sous sa forme moderne qu'en 1955 grâce à Ross. Cela bien après la réalisation par Shockley en 1947 du premier transistor de type bipolaire, pourtant théoriquement bien plus sophistiqué. C'est en 1960 que Kahng et Attala ont présenté le premier transistor MOS sur Silicium en utilisant une grille isolée dont le diélectrique de grille était en oxyde de silicium  $\text{SiO}_2$ . Le silicium fut un choix très judicieux. Il est l'élément le plus abondant de la croûte terrestre, après l'oxygène. De plus son oxyde est non seulement un très bon isolant électrique mais il s'est aussi révélé parfaitement adapté pour former des couches dites de passivation protégeant les circuits, accroissant remarquablement leur fiabilité.

Les transistors MOS sur silicium, plus simples et moins chers que leurs concurrents bipolaires, mais intrinsèquement moins performants à génération technologique équivalente, ont connu leur essor dans les années 70-80 grâce à la technologie CMOS inventée en 1968 qui consomme très peu d'énergie. Ainsi des systèmes à bas prix possédant une grande autonomie ont été très largement diffusés : montres à quartz, calculatrices...

Depuis, les applications en logique CMOS mais aussi les mémoires qui constituent l'autre grande application des MOSFET ont bénéficié d'une très forte et continue augmentation de leur rapidité et de leur densité d'intégration.

### **1.2 Comment ça marche ?**

#### **1.2.1 Effet de champ & MOSFET**

Un transistor est avant tout un interrupteur commandé : un signal électrique (courant ou tension) de faible puissance, dit de commande, doit contrôler un signal électrique de plus forte puissance. Cela permet deux types de fonctionnement : soit en amplification pour les applications de type analogique, soit en tout ou rien pour les applications logiques qui vont nous intéresser dans ce mémoire.

La Figure 4 présente le schéma en coupe d'un transistor MOS sur substrat massif « classique ». Ce transistor est construit sur substrat en silicium faiblement dopé, dopé P. Il possède 4 électrodes :

la source et le drain qui sont les électrodes de puissance, la grille de longueur  $L_G$  (cf. Figure 4.a) qui est l'électrode de commande, et le contact de substrat. L'isolant de grille en oxyde de silicium  $\text{SiO}_2$ , d'épaisseur notée  $T_{ox}$ , est intercalé entre l'électrode de grille et le substrat de silicium. Les électrodes de source et de drain sont en contact avec des régions très fortement dopées, d'un type de dopage inverse à celui du substrat. Dans l'exemple de la Figure 4, ils sont dopés  $N^+$  et de largeur  $W$  (cf. Figure 4.b). Ils constituent des réservoirs de charges symétriquement répartis de chaque côté de l'oxyde.

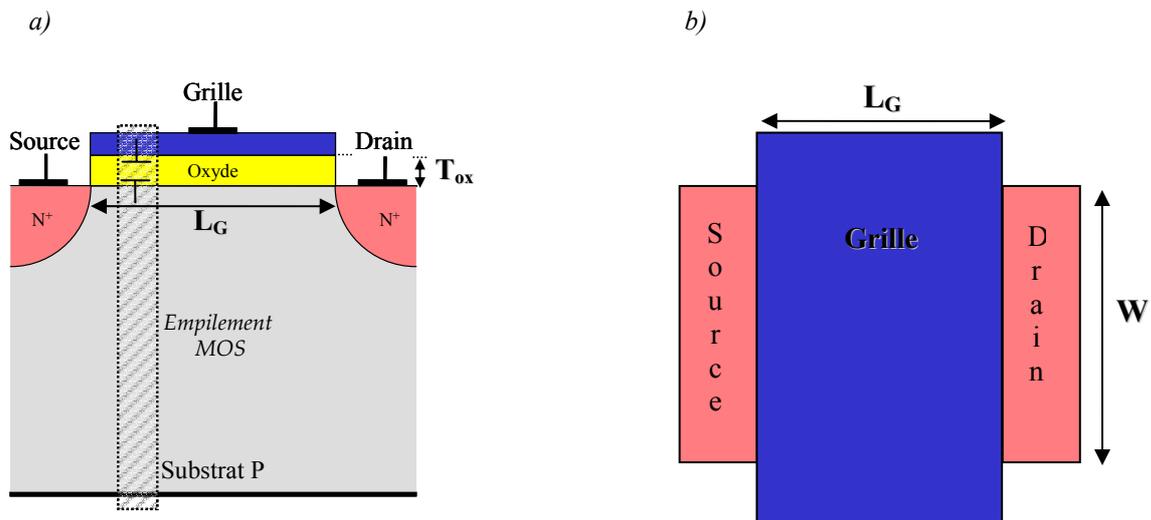


Figure 4 : Schémas d'un transistor MOSFET à canal N sur substrat P ("normally off") conventionnel. a) Schéma en coupe. b) Schéma en vue de dessus.  $L_G$  longueur de la grille,  $W$  largeur du canal et  $T_{ox}$  épaisseur d'oxyde. L'empilement Métal/Oxyde/Semi-conducteur est hachuré en a).

Le nom du transistor MOS découle de sa structure verticale : Métal/Oxyde/Semi-conducteur. Cette structure, hachurée sur le schéma de la Figure 4.a, est celle d'une capacité Conducteur/Isolant/Semi-conducteur.

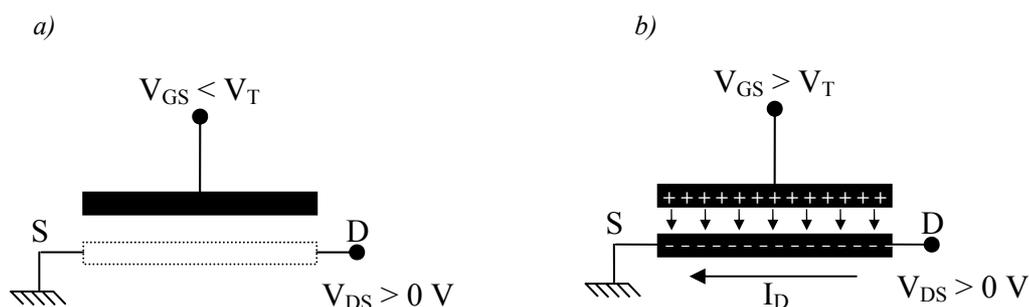


Figure 5 : Principe de fonctionnement d'un transistor MOS [GAUT03]. a) Etat bloqué. b) Etat passant.

Le principe de fonctionnement d'un tel dispositif est schématisé en Figure 5. La tension de grille crée un champ vertical qui, par l'intermédiaire de la capacité MOS, module la densité de porteurs libres à l'interface Substrat/Oxyde permettant ainsi de commander sa conductivité. Ce phénomène est appelé « effet de champ ».

Lorsque la tension de grille ne permet pas de charger « convenablement » les électrodes de la capacité MOS, aucun courant ne peut circuler entre la source et le drain quelle que soit la tension appliquée à l'électrode de drain ; c'est l'état bloqué. Comme illustré en Figure 5.a, pour le type de transistor étudié, la tension de grille  $V_{GS}$  appliquée doit rester inférieure à une tension de seuil notée  $V_T$ . Ces transistors bloqués à tension de grille nulle sont dits "normally off" (par opposition aux "normally on", tels les transistors à canal N pré implanté).

Lorsque  $V_{GS}$  devient supérieure à  $V_T$ , le nombre de porteurs libres à l'interface Substrat/Oxyde est suffisant pour qu'ils forment un « canal » conducteur (si et seulement s'ils sont du même type que ceux des réservoirs de source et drain). Si, comme illustré sur la Figure 5.b, les porteurs du canal sont des électrons, le transistor est dit à canal N et est appelé NMOS. Lorsque la tension de drain devient positive, un courant de drain  $I_D$  peut circuler dans le canal, c'est l'état passant.

### 1.2.2 Caractéristiques Courant/Tension des MOSFET « longs »

De manière plus quantitative, on obtient les trois zones de fonctionnement classiques :

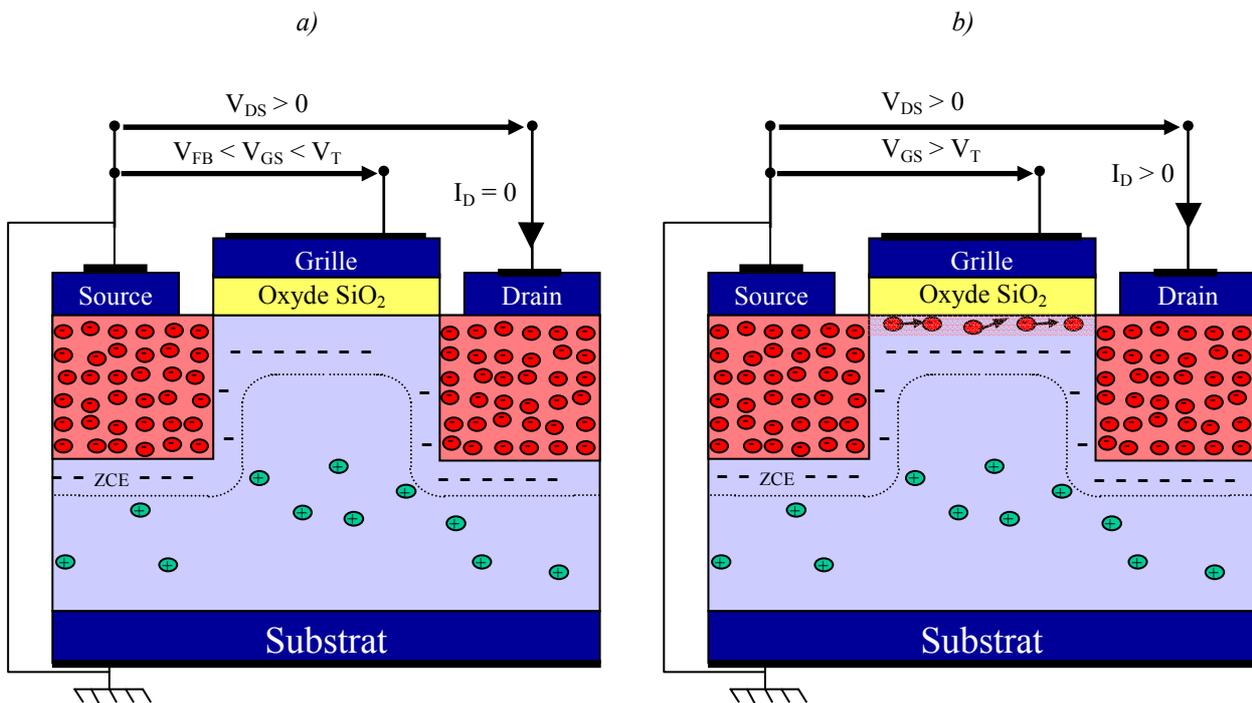


Figure 6 : Régimes de polarisation d'un transistor MOSFET à canal N sur substrat P ("normally off") à faible  $V_{DS}$  positif. En (a) désertion en trous sous la grille, en (b) formation du canal de conduction à partir des réservoirs à électrons de source et drain. Les symboles  $-$ ,  $\ominus$  et  $\oplus$  représentent respectivement les impuretés dopantes ionisées négativement, les électrons libres et les trous.

#### Zone bloquée :

Pour des tensions de grille inférieures à la tension de seuil mais supérieures à la tension de bande plate  $V_{FB}$ , la zone sous l'oxyde est désertée par les porteurs libres. La capacité MOS se trouve en régime de déplétion ; cette situation est illustrée en Figure 6.a. Dans cette zone, la

conduction du canal tend à s'annuler et le courant de drain est très majoritairement d'origine diffusive [DOLL01] :  $I_D \propto q.D_n \left| \overline{\text{grad}}(n) \right|$  avec  $n$  concentration en électrons libres et  $D_n$  coefficient de diffusion. On peut écrire [GAUT03] :

$$I_D = I_{D_0} \exp\left(\frac{V_{GS} - V_T}{S} \ln 10\right) \left[ 1 - \exp\left(-\frac{qV_{DS}}{k_B T}\right) \right] \text{ où } S \text{ est le paramètre nommé pente sous le seuil.}$$

Finalement, dans les applications pratiques basées sur l'utilisation de MOSFET longs à 300 K, le courant  $I_D$  sous le seuil est quasi nul ( $I_{D0} \approx 10^{-7} \times W/L_G$  (A)) On peut alors considérer :  $I_D \approx 0$  A pour  $V_{DS} > 0$  V et  $V_{GS} < V_T$

### Zone ohmique :

Comme illustré en Figure 6.b, pour des tensions de grille supérieures à la tension de seuil, on atteint le régime d'inversion, il y a création d'un canal formé par les porteurs libres injectés par les réservoirs de source et de drain. Pour de faibles tensions de drain  $V_{DS}$ , la vitesse  $\vec{v}$  des charges libres varie linéairement avec le champ électrique parallèle à la direction source-drain  $\vec{E}_{//}$  (variant comme  $V_{DS}/L_G$ ) :  $\vec{v} = -\mu_{\text{eff}} \times \vec{E}_{//}$ , où  $\mu_{\text{eff}}$  est appelée mobilité effective des porteurs dans le canal. Le transistor a alors un comportement équivalent à celui d'une résistance commandée par  $V_{GS}$ . En modélisant habilement le contrôle de charge dû à l'effet de champ et le courant de conduction dans le canal [MATH04], on obtient :

$$I_D = \frac{W}{L_G} \mu_{\text{eff}} C_{\text{ox}} \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \text{ pour } V_{DS} < V_{DS\text{sat}} \text{ et } V_{GS} > V_T$$

Avec  $C_{\text{ox}} = \epsilon_{\text{ox}}/T_{\text{ox}}$  capacité surfacique de la structure MOS.

Il est à noter que le terme en  $V_{DS}^2$  provient de la variation de la charge d'inversion le long de la grille sous l'effet de la tension de drain qui tend à dépeupler le canal. Quand, dans l'expression du courant  $I_D$ , le terme en  $V_{DS}^2$  n'est plus négligeable devant  $(V_{GS} - V_T) \times V_{DS}$ , on sort du régime purement ohmique, il s'agit d'une zone de transition où l'augmentation du courant avec  $V_{DS}$  croissant tend à diminuer.

### Zone source de courant :

Lorsque  $V_{DS} > V_{DS\text{sat}}$ , un phénomène de saturation du courant  $I_D$  apparaît : pincement du canal, saturation de la vitesse des porteurs... Alors  $I_D$  ne dépend plus de  $V_{DS}$  mais seulement de  $V_{GS}$ . Le transistor se comporte alors comme une source de courant commandée par  $V_{GS}$ .

Pour le cas d'une saturation par pincement,  $I_D$  devient [MATH04] :

$$I_{D\text{sat}} = \frac{W}{L_G} \mu_{\text{eff}} C_{\text{ox}} \frac{(V_{GS} - V_T)^2}{2} \text{ pour } V_{DS} > V_{GS} - V_T \text{ et } V_{GS} > V_T$$

Sur cette formule du courant de saturation, on découvre l'atout majeur du transistor MOS, celui qui a assuré son succès industriel. Le courant de saturation du transistor MOS est directement lié à ses dimensions géométriques : réduire la longueur  $L_G$  d'un transistor MOS entraîne automatiquement l'amélioration de ses performances électriques et cela sans aucun effet négatif (du moins pour les composants faiblement submicroniques...). Cette relation justifie la course à la miniaturisation des composants et explique la très forte augmentation des performances des composants MOS.

### 1.3 Principaux paramètres des MOSFET

Les trois régimes de fonctionnement (bloqué, ohmique et source de courant) décrits au paragraphe précédent sont indiqués sur les caractéristiques  $I_D(V_{DS})$  à différentes tensions de grille  $V_{GS}$  reportées en Figure 7.

L'état passant d'un transistor (zone ohmique et source de courant) MOSFET se caractérise principalement par :

- La résistance à l'état passant  $R_{on}$  : inverse de la pente de courbe  $I_D(V_{DS})$  à  $V_{GS}=V_{DD}$  et faible  $V_{DS}$  où  $V_{DD}$  est la tension d'alimentation du transistor (cf. Figure 7).
- Le courant de saturation :  $I_{on}$ , c'est-à-dire le courant  $I_D$  à  $V_{GS}=V_{DD}$  et  $V_{DS}=V_{DD}$  (cf. Figure 8).
- la transconductance  $G_m$ .  $G_m$  correspond à la pente de la courbe  $I_D(V_{GS})$  à fort  $V_{DS}$  (cf.

Figure 8) soit  $G_m = \left. \frac{\partial I_D}{\partial V_{GS}} \right|_{V_{DS}=cste}$

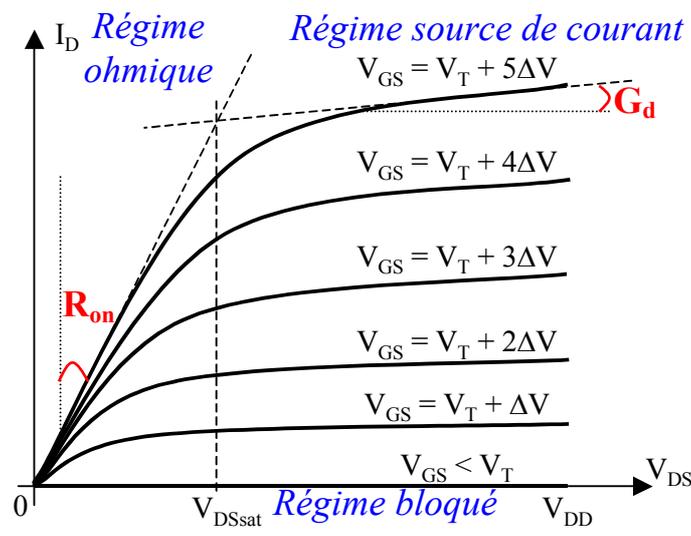


Figure 7 : Caractéristiques  $I_D(V_{DS})$  typiques à différents  $V_{GS}$  d'un transistor NMOS.

- la conductance  $G_d$  quantifie l'imperfection de la saturation. Elle est égale à la pente de la

courbe  $I_D(V_{DS})$  à  $V_{DS} > V_{DSsat}$  soit :  $G_d = \left. \frac{\partial I_{DS}}{\partial V_{DS}} \right|_{V_{GS}=cste}$  en source de courant.

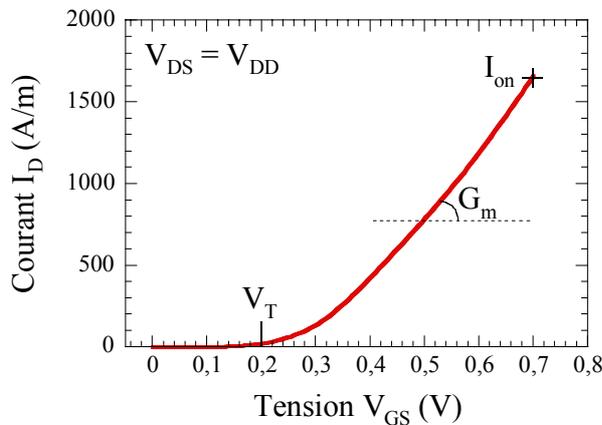


Figure 8 : Caractéristique  $I_D(V_{GS})$  à  $V_{DS} = V_{DD}$  typique d'un NMOS.  $I_{on}$ ,  $G_m$  et  $V_T$  sont indiqués.

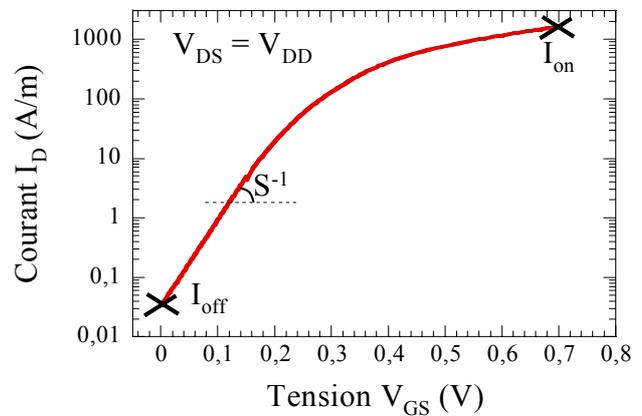


Figure 9 : Caractéristique  $\log[I_D(V_{GS})]$  à  $V_{DS} = V_{DD}$  typique d'un NMOS.  $I_{on}$ ,  $I_{off}$  et  $S$  sont indiqués.

Au-dessous du seuil, le blocage n'est pas parfait, il existe un courant faible mais non nul ( $I_D \neq 0$ ). Ce courant n'est pas un courant de conduction comme à l'état passant mais un courant diffusif, d'où la dépendance exponentielle de la caractéristique  $I_D(V_{GS})$  sous le seuil illustrée en Figure 9.

- le courant à l'état bloqué  $I_{off}$  correspond au courant  $I_D$  à  $V_{GS} = 0$  V et  $V_{DS} = V_{DD}$  (cf. Figure 9),
- de plus on définit la pente sous le seuil  $S$  comme l'inverse de la pente  $\log[I_D(V_{GS})]$  à faible  $V_{GS}$

soit  $S = \left( \left. \frac{\partial \log(I_D)}{\partial V_{GS}} \right|_{V_{DS}=cste} \right)^{-1}$ . La modélisation du courant sous le seuil [GAUT03] donne une

pente  $S = \frac{k_B T}{q} \ln(10) \left( 1 + \frac{C_{ZCE}}{C_{ox}} \right)$ , où  $k_B$  est la constante de Boltzmann,  $T$  la température,  $q$  la

charge élémentaire,  $C_{ox}$  la capacité d'oxyde et  $C_{ZCE}$  la capacité de la zone de charge d'espace ; ces capacités sont définies plus précisément dans le paragraphe suivant. Pour  $C_{ox} \gg C_{ZCE}$  on obtient la pente  $S$  idéale  $S_{idéale} = k_B \times T / q \times \ln(10)$ , c'est-à-dire égale à 60 mV par décade à 300 K.

#### 1.4 Transistor MOS : schéma équivalent

La Figure 10 présente le schéma électrique équivalent d'un transistor MOSFET intrinsèque dans le régime d'inversion ( $V_{GS} > V_T$ ). Cette modélisation inclut les différentes résistances et capacités parasites habituelles. La superposition du schéma électrique et de la coupe du transistor met en évidence les origines physiques de ces éléments parasites.

Les caissons source et drain sont des zones fortement dopées, les porteurs qui les traversent y subissent donc un grand nombre d'interactions avec les impuretés dopantes ionisées. Les résistances d'accès intrinsèques  $R_S$  et  $R_D$  de ces caissons peuvent être non négligeables si elles deviennent du même ordre, voire plus importantes, que la résistance minimale du canal comme nous le reverrons plus en détails au paragraphe 2.3.6 .

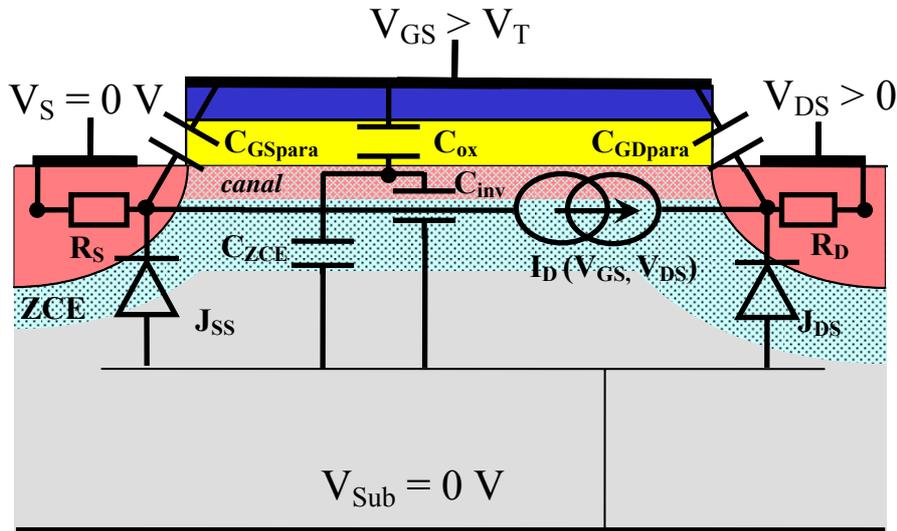


Figure 10 : Schéma électrique équivalent [VAPA87] superposé à un schéma en coupe d'un MOSFET en inversion.

On a vu précédemment que l'effet de champ crée un canal de porteurs libres par l'intermédiaire de la capacité MOS. La longueur et la largeur de l'oxyde de grille étant très grandes devant son épaisseur ( $W$  et  $L_G \gg T_{ox}$ ), les effets de bord sont négligeables en première approximation. La capacité MOS n'a donc qu'une seule dimension et on ne va considérer que des grandeurs surfaciques. La variation de charges surfaciques commandée par la grille  $\Delta Q(V_{GS})$  se répartit dans le semi-conducteur entre le canal d'inversion  $\Delta Q_{inv}$  et la zone de charge d'espace  $\Delta Q_{ZCE}$  soit :  $\Delta Q(V_{GS}) = \Delta Q_{inv} + \Delta Q_{ZCE}$ . Cette capacité MOS est modélisée par une capacité d'oxyde  $C_{ox}$  en série avec un montage en parallèle d'une capacité d'inversion  $C_{inv}$  et d'une capacité de substrat  $C_{ZCE}$ . La capacité surfacique d'oxyde  $C_{ox}$  est reliée à la structure métal de grille/isolant :  $C_{ox} = \epsilon_{ox}/T_{ox}$ . La tension à ses bornes est égale à la différence de potentiel aux limites de l'oxyde. La capacité surfacique  $C_{inv}$  est liée à la variation, sous l'effet de la tension de grille, de la charge d'inversion formée par les porteurs libres du canal  $\Delta Q_{inv}$ . En outre, on montre [VAPA87] que pour  $V_G > V_T$ , la charge surfacique d'inversion  $Q_{inv}$  est donnée par  $Q_{inv} = C_{ox} \cdot (V_G - V_T)$ . La capacité surfacique  $C_{ZCE}$  correspond à la variation de charge surfacique de la zone de charge d'espace du substrat, d'extension  $T_{ZCE}$ , commandée par la grille donc  $C_{ZCE} = \epsilon_{Si}/T_{ZCE}$  [VAPA87] avec  $\epsilon_{Si}$  permittivité diélectrique du silicium. Pour augmenter le contrôle de la grille sur le canal,  $\Delta Q_{ZCE}$  et donc la

capacité  $C_{ZCE}$  doivent rester les plus faibles possibles pour ne pas dégrader les caractéristiques du transistor, en particulier sous le seuil.

Les capacités  $C_{GSpara}$  et  $C_{GDpara}$  modélisent des phénomènes capacitifs supplémentaires qui se rajoutent, côté drain et côté source, à ceux de la structure MOS idéale. Chacune de ces capacités peut inclure différents types de couplage électrostatique comme illustré en Figure 11. Tout d'abord, il peut exister un couplage dû à des effets de bord nommé  $C_{bord}$  : des lignes de champ (à 2 dimensions) se propagent entre l'électrode de grille et les caissons dans l'espace au-dessus des caissons. Dans le cas d'un recouvrement par la grille des caissons, il y a création, en régime d'inversion du canal, d'une zone d'accumulation dans les caissons et apparition de la capacité  $C_{rec}$  associée qui peut être très pénalisante. De plus, si la grille recouvre les ZCE de jonction caisson/substrat, il y a compétition entre les ZCE commandées par la grille et la source. Cela se traduit par une modification de la capacité surfacique  $C_{ZCE}$  près des caissons.

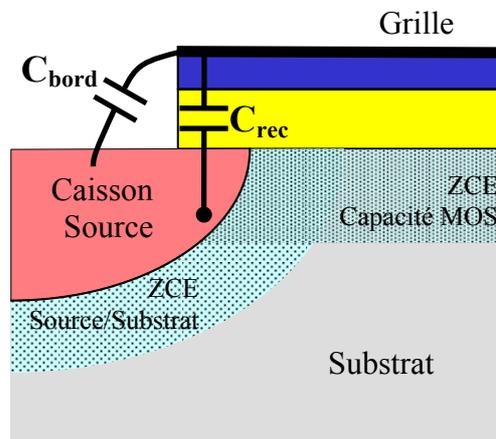


Figure 11 : Décomposition de la capacité parasite  $C_{GSpara}$  de la Figure 10 en capacité  $C_{bord}$  et capacité de recouvrement  $C_{rec}$  grille/caisson.

Enfin, on constate la présence des jonctions N+/P Source/Substrat et aussi Drain/Substrat qui induisent des comportements résistifs et capacitifs susceptibles de détériorer aussi le comportement dynamique du transistor.

Dans la suite de ce mémoire, la capacité de grille totale, prenant en compte toutes les capacités décrites ci-dessus, sera notée  $C_G$  (en F).

## 1.5 Transistor et inverseur CMOS

Les portes logiques élémentaires (NAND, NOR...) sont construites grâce à l'interconnexion de montages inverseurs. L'inverseur est la brique de base d'une technologie, ses performances conditionnent donc étroitement celles des applications plus complexes [TAUR98]. Après avoir présenté le schéma de l'inverseur CMOS, on décrira comment ses performances sont liées aux caractéristiques des transistors dont il est constitué.

### 1.5.1 Schéma de l'inverseur CMOS

Le schéma d'un inverseur CMOS est représenté en Figure 12. Il est constitué de deux transistors « normally off » complémentaires à canaux N et P, leur tension de seuil  $V_T$  étant respectivement positive et négative mais de même valeur absolue. Dans un état logique donné '1' ( $V_{out} = V_{DD}$ ) ou '0' ( $V_{out} = 0$  V), un des transistors se trouve à l'état bloqué et l'autre est passant. La capacité  $C_L$  modélise la capacité d'entrée de l'étage inverseur suivant.

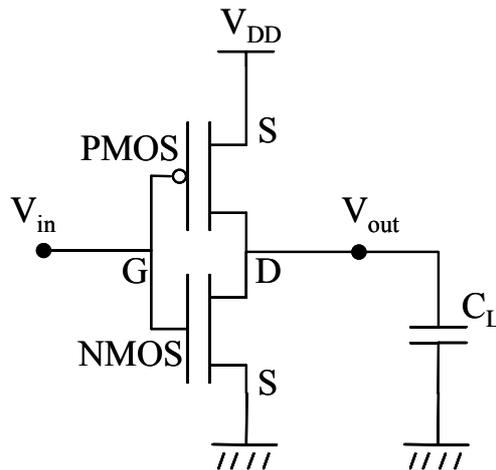


Figure 12 : Schéma électrique d'un inverseur CMOS.

### 1.5.2 Caractéristique de transfert $V_{out}(V_{in})$

Pour obtenir la caractéristique de transfert  $V_{out}(V_{in})$  complète tracée en Figure 14, on peut représenter sur un même graphe les caractéristiques  $I_D(V_{DS})$  à différents  $V_{GS}$  du transistor NMOS, en trait plein sur la Figure 13, et celle du PMOS translaturée de  $V_{DD}$  :  $V_{GS} = V_{in} - V_{DD}$  et  $V_{DS} = V_{out} - V_{DD}$ , en tirets sur la Figure 13. Les intersections de chaque courbe en tirets et en trait continu à même  $V_{in}$  forment la caractéristique statique  $V_{out}(V_{in})$ . Ainsi, pour  $V_{in}$  inférieur à  $V_T$ , le NMOS est bloqué et le PMOS passant en régime ohmique (cf. points a. de la Figure 13 et Figure 14) ; la capacité de sortie ne peut se décharger et  $V_{out}$  reste fixé à  $V_{DD}$ . Puis, quand  $V_{in}$  varie entre  $V_T$  et une valeur limite  $V_{eT}$ , le transistor NMOS est débloqué, il fonctionne en régime source de courant, alors que le PMOS tend à passer progressivement du régime ohmique au régime source de courant (cf. points b.). La tension  $V_{out}$  décroît avec l'augmentation de  $V_{in}$ . Pour  $V_{in} = V_{eT}$ , les caractéristiques des deux transistors se croisent en régime source de courant. Pour  $V_{in}$  compris entre  $V_{eT}$  et  $V_{DD} - V_T$ , le NMOS passe en régime ohmique et le PMOS reste en régime source de courant,  $V_{out}$  continue à diminuer (cf. points c.). Enfin, le PMOS se bloque pour  $V_{in}$  supérieur à  $V_{DD} - V_T$  et  $V_{out}$  atteint 0 V (cf. points d.).

La caractéristique de transfert présentée en Figure 14 est séparée en différentes zones en fonction de la valeur absolue du gain  $|dV_{out}/dV_{in}|$ . Les plateaux, c'est-à-dire les zones où le gain est inférieur à 1, sont appelés zones de décisions. Celle où le gain est supérieur à 1 est dite de basculement ou d'amplification. La zone de basculement est naturellement utilisée pour ses vertus d'amplification, et les zones de décisions pour générer des signaux logiques. L'état logique '0' correspond au premier plateau, c'est-à-dire aux tensions  $V_{in} \in [0..V_{inB}]$  et  $V_{out} \in [0..V_{outB}]$ , l'état '1' au second plateau à  $V_{in} \in [V_{inH}..V_{DD}]$  et  $V_{out} \in [V_{outH}..V_{DD}]$ . La condition indispensable pour qu'il y ait restauration du signal traversant un ensemble d'inverseurs est que la plage des tensions d'entrée définissant un état logique soit plus grande que la plage de sortie de l'état correspondant, c'est-à-dire :  $[0..V_{inB}] \gg [0..V_{outB}]$  et  $[V_{inH}..V_{DD}] \gg [V_{outH}..V_{DD}]$ . Les marges de bruit, définies comme les intervalles  $[V_{outB}..V_{inB}]$  et  $[V_{inH}..V_{outH}]$ , doivent donc être les plus grandes possible. A partir de la Figure 14 on pressent intuitivement (et les calculs le confirment) qu'une augmentation de  $V_T$  induit celle de  $V_{inB}$  et une diminution de  $V_{inH}$ . De même, une augmentation de  $V_{DD}$  améliore aussi les marges de bruit [WEST93].

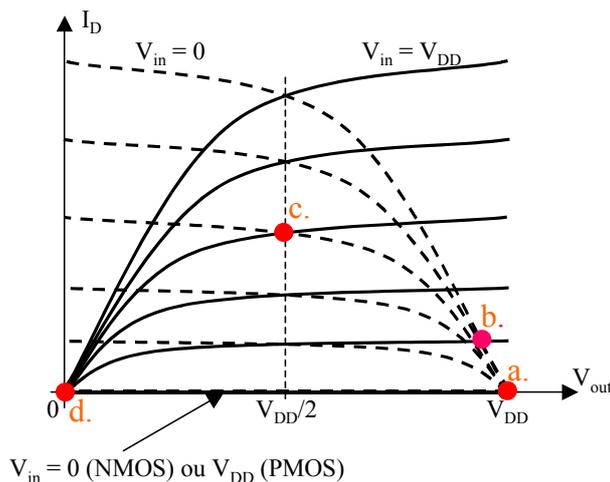


Figure 13 : Caractéristiques  $I_D(V_{DS})$  du NMOS (en traits continus) et du PMOS (en tirets) constituant un inverseur CMOS dans le plan  $(V_{out}; I_D)$ .

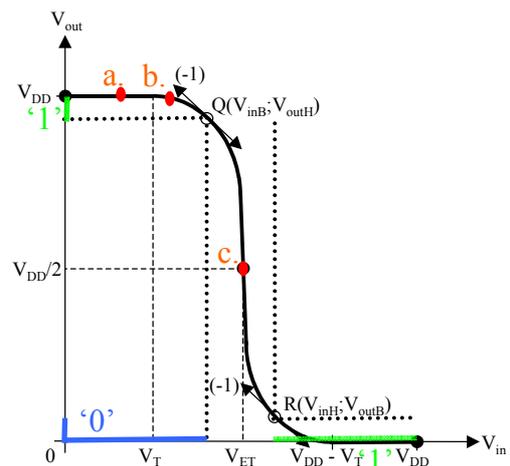


Figure 14 : Caractéristique statique  $V_{out}(V_{in})$  typique d'un inverseur CMOS. Les pentes aux points Q et R valent -1. Les états logiques '0' et '1' sont indiqués respectivement en rouge et en vert.

La tension d'entrée  $V_{eT}$  (cf. Figure 13 et Figure 14), rendue égale à  $V_{DD}/2$  grâce à "l'appariement" (réglage des tensions de seuil et largeurs des transistors) des NMOS et PMOS, est le seuil de commutation de l'inverseur. La variation de  $V_{out}$  autour de  $V_{eT}$  est d'autant plus abrupte que la conductance de drain  $G_D$  en régime source de courant est faible. Pour  $G_D$  nul, il se produit une transition verticale sur  $V_{out}$  pour  $V_{in} = V_{eT}$ , les caractéristiques  $I_D(V_{DS})$  des deux transistors se balayant en régime source de courant, avec une infinité de valeurs de  $V_{out}$  possibles pour cette seule valeur de  $V_{in}$ . On a tout intérêt à tendre vers cette limite idéale qui permet une meilleure

régénération des niveaux logiques [LAHM04]. Si la conductance  $G_D$  n'est pas nulle, une forte valeur de la transconductance  $G_m$  permet néanmoins d'avoir une transition abrupte.

Le régime sous le seuil des transistors est aussi très important pour une utilisation des transistors en logique CMOS. Si le courant à l'état bloqué  $I_{off}$  n'est pas nul, la porte consomme du courant même sans commuter entre 2 états logiques :  $P_{stat} = V_{DD} \times I_{off}$ . Cette consommation statique  $P_{stat}$  du transistor doit rester très faible, car sur une puce elle doit être multipliée par le nombre total de transistors : typiquement plusieurs millions !

### 1.5.3 Dynamique de l'inverseur

Le comportement dynamique de l'inverseur est aussi crucial pour permettre la hausse continue de la fréquence de travail des circuits. Le temps de réponse de l'inverseur  $t_d$  doit rester au moins inférieur à la période du signal d'horloge. Pour évaluer  $t_d$ , on utilise souvent le paramètre «  $CV/I$  » qui correspond à la constante de temps caractéristique de la charge (ou de décharge) de la capacité  $C_L$  de l'inverseur du schéma en Figure 12. Le courant de charge est supposé constant en première approximation et égal à  $I_{on}$ . La charge transférée est égale à  $Q = C_L \times V_{DD}$ . La capacité  $C_L$  (capacité d'entrée de l'étage inverseur suivant) la plus défavorable correspond à celle d'un transistor bloqué :  $C_L = C_G$  à  $V_{GS} = V_{DD}$  et  $V_{DS} = 0$  V. Le temps de charge est alors égal à  $C_G \times V_{DD} / I_{on}$ . En conséquence, l'augmentation du courant  $I_{on}$  (à capacité  $C_G$  constante) améliore la rapidité des circuits logiques.

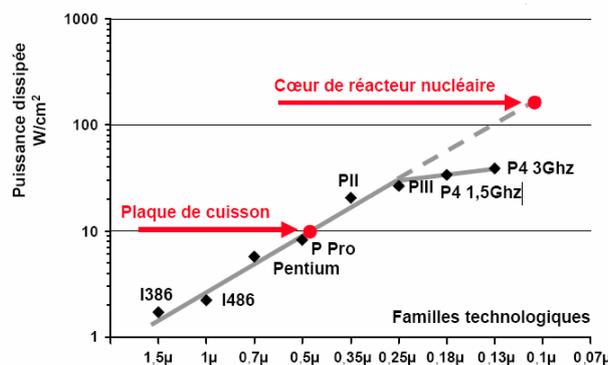


Figure 15 : Evolution de la puissance surfacique dissipée dans les processeurs d'Intel (d'après Fred Pollack, Intel).

La puissance dynamique  $P_{dyn}$  s'exprime comme le produit de la fréquence d'horloge  $f$  et de l'énergie consommée lors d'une commutation, c'est-à-dire l'énergie de charge ou de décharge de la capacité  $C_L$  (capacité équivalente de l'étage suivant), d'où  $P_{dyn} \propto f \times C_L \times V_{DD}^2$  [ARBE98]. Elle est proportionnelle au carré de la tension d'alimentation. De plus le produit  $C_L \times V_{DD}^2$  doit décroître plus vite que l'augmentation de la fréquence d'horloge des circuits pour que la puissance dynamique  $P_{dyn}$  reste constante. Il ne faut pas en effet dégrader l'autonomie des systèmes embarqués et augmenter les dissipations thermiques. En effet, comme indiqué sur la Figure 15, en extrapolant la croissance exponentielle observée dans les années 90 (des processeurs 486 aux Pentiums II), on

prévoyait d'atteindre sur les puces de technologie actuelle des puissances thermiques voisines de celles rencontrées dans le cœur d'un réacteur de centrale nucléaire ! Même si les nouvelles générations de processeurs ont permis de reculer cette échéance, elles sont encore loin de respecter la philosophie du développement durable qui prône une consommation contrôlée des ressources : la croissance reste exponentielle...

### 1.5.4 Bilan des relations entre inverseur CMOS et transistor

		$V_{DD}$	$V_T$	$I_{on}$	$G_m$	$G_d$	$I_{off}$	} Evolution des caractéristiques du transistor
		↗	↗	↗	↗	↗	↗	
<b>Statique</b>	<b>Marges de bruit</b>	+	+		+	-		} Gain sur les performances de l'inverseur
	$P_{stat}$	-	+				-	
<b>Dynamique</b>	$P_{dyn}$	-						
	$t_d \approx CV/I$		-	+	+			

Tableau 1 : Influences des caractéristiques intrinsèques des transistors MOS sur les performances de l'inverseur CMOS. [ARBE98]

Les interdépendances entre le transistor et l'inverseur CMOS sont résumées et complétées dans le Tableau 1. On constate qu'il y a des compromis souvent délicats à réaliser pour obtenir :

- $V_T$  élevé et  $G_d$  faible pour avoir de bonnes marges de bruit,
- $I_{off}$  faible pour avoir une faible consommation statique,
- $I_{on}$  et  $G_m$  élevés pour augmenter la rapidité.

Pour le  $V_T$ , le bon compromis est généralement atteint lorsque  $V_T \approx V_{DD}/3$ .

## 2. Fin du CMOS conventionnel

### 2.1 Loi d'échelle

Le monde de la microélectronique est en perpétuelle recherche d'améliorations de la rapidité et de la densité d'intégration de ses composants élémentaires. La réduction de la taille du composant, c'est-à-dire de sa longueur de grille  $L_G$ , le rend intrinsèquement plus rapide car le temps de transit des porteurs dans le canal diminue :  $t = L_G^2 / (\mu_{eff} \cdot V_{DS})$  à faible  $V_{DS}$ . Cette accélération de la réponse du transistor se répercute sur les circuits logiques dérivés comme discuté au paragraphe précédent.

Cette miniaturisation doit être effectuée sans dégrader les caractéristiques du transistor. Pour cela, il faut adapter les paramètres géométriques (épaisseur d'oxyde  $T_{ox}$ , largeur de canal  $W...$ ), électriques ( $V_{DD}...$ ), et physiques (dopages) en fonction de la longueur de grille visée. Ces

paramètres de dimensionnement suivent ce que l'on appelle les lois d'échelle. Par exemple, afin de conserver le pouvoir de contrôle de la grille identique quelle que soit la longueur  $L_G$ , il faut travailler à capacité d'oxyde de grille par unité de largeur  $C_{ox} \cdot L_G = \frac{\epsilon_{ox}}{T_{ox}} \times L_G$  constante ( $C_{ox}$  capacité surfacique et  $\epsilon_{ox}$  permittivité diélectrique du  $\text{SiO}_2$ ). Cela impose de réduire  $T_{ox}$  proportionnellement à  $L_G$ . La diminution de l'épaisseur  $T_{ox}$  doit entraîner une diminution de la tension d'alimentation  $V_{DD}$  pour que le champ électrique maximum dans l'oxyde, qui vaut  $V_{DD}/T_{ox}$  ne dépasse pas le seuil de claquage. La réduction de  $V_{DD}$  doit en outre permettre de mieux gérer le bilan énergétique d'un circuit CMOS en fonctionnement.

Les lois d'échelle idéales permettent de conserver l'équation de Poisson invariante par rapport à la miniaturisation, c'est-à-dire que les champs électriques présents dans un transistor court sont toujours identiques à ceux apparaissant dans des transistors plus longs. Un dispositif ainsi dimensionné bénéficie des bienfaits de la miniaturisation sans subir de dégradation de ses caractéristiques. On vérifie facilement que l'équation de Poisson :  $\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\epsilon_{Si}}$  ( $V$  potentiel électrostatique,  $\rho$  densité de charge) reste invariante si l'on substitue les variables par  $V' = 1/K \times V$ ,  $x' = 1/K \times x$ ,  $y' = 1/K \times y$ ,  $z' = 1/K \times z$  et  $\rho' = K \times \rho$ . On obtient ainsi les lois d'échelle détaillées au Tableau 2 : si toutes les dimensions et les tensions sont réduites d'un facteur  $1/K$  il faut augmenter les dopages des caissons et du substrat d'un facteur  $K$ .

	Transistor MOS et circuit	Facteur multiplicatif $K > 1$
Dimensionnement du transistor	Dimensions du transistor : $T_{ox}, L_G, W \dots$	$1/K$
	Tension	$1/K$
	Concentrations de dopants	$K$
Effets sur les transistors	Courant [A]	$1/K$
	Capacité [F]	$1/K$
Effets sur les circuits	Temps de retard (CV/I)	$1/K$
	Puissance dissipée par circuit (VI)	$1/K^2$
	Densité Puissance (P/Surface)	1

Tableau 2 : Lois d'échelle des transistors MOS [TAUR98]

Ces lois d'échelle renforcent le contrôle de la grille, quantifié par la capacité surfacique  $C_{ox} = \frac{\epsilon_{ox}}{T_{ox}}$ , qui est augmentée d'un facteur  $K$ . Elles induisent aussi une diminution proportionnelle au facteur  $1/K$  du courant  $I_D$  ( $I_D \propto W/L_G \times C_{ox} \times V^2$ ) et de la capacité totale de grille  $C_G = \frac{\epsilon_{ox}}{T_{ox}} \times L_G \times W$ . Ainsi, le rapport  $I_D/C_G$  reste inchangé par changement d'échelle ainsi que le courant normalisé par la largeur  $I_D/W$ . La charge surfacique d'inversion  $Q_{inv} = C_{ox} \cdot (V_G - V_T)$  restant invariante, il est à noter que malheureusement le courant sous le seuil, de nature diffusive

$(I_{diff} \propto \left| \frac{\partial Q_{inv}}{\partial x} \right| \propto \frac{Q_{inv}}{L_G})$ , augmente aussi d'un facteur K [TAUR98]. Cependant, ce courant reste suffisamment faible pour être négligé dans les dispositifs légèrement submicroniques.

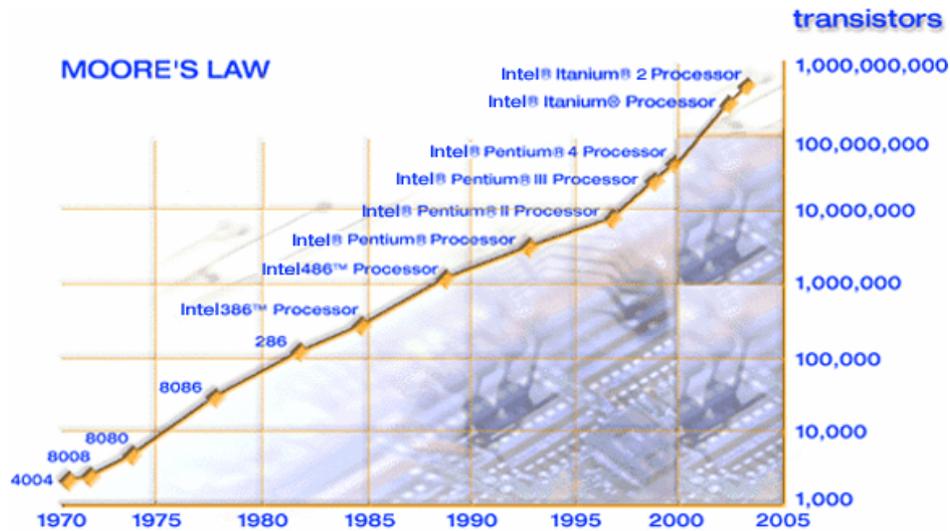


Figure 16 : Evolution du nombre de transistors dans les microprocesseurs Intel ([www.intel.com](http://www.intel.com)).

L'utilisation de transistors dimensionnés selon ces lois d'échelle a permis d'améliorer les performances des circuits CMOS sans faire aucun compromis. La densité d'intégration est augmentée d'un facteur  $K^2$  ( $\propto W \times L_G$ ), la puissance surfacique diminuée du même facteur :  $P=U \times I$  et la vitesse ( $\propto I/(C_G \times V)$ ) progressent d'un facteur K. Cela a, entre autres, permis une augmentation exponentielle du nombre de transistors par puce au cours du temps, nommée communément loi de Moore. La Figure 16 illustre cette évolution à travers les microprocesseurs d'Intel.

## 2.2 La « Roadmap »

Pour suivre la loi de Moore, l'évolution des transistors MOSFET, fers de lance de cette course à la miniaturisation, est encadrée par des feuilles de route, les « roadmaps » de l'ITRS (International Technology Roadmap for Semiconductors) qui doivent orienter les recherches. Des mises à jour de la roadmap paraissent tous les ans et sont publiées par une association internationale de laboratoires et d'industriels du secteur des semi-conducteurs. Un extrait de la feuille de route 2003 (dont les mises à jour 2004 sont mineures), concernant les circuits logiques hautes performances, est présenté dans le Tableau 3. Les paramètres  $T_{ox}$ ,  $I_{on}$ ,  $I_{off}$  et «  $CV/I$  » ont été marqués en gras.

Ce tableau indique des lois d'échelle plus complexes que celles du Tableau 2 car elles tiennent compte de nombreux paramètres technologiques et électriques supplémentaires qui vont être détaillés plus abondamment dans les sections suivantes. Cependant, on peut constater (avec émotion !) que les solutions technologiques pour atteindre des nœuds qui devraient apparaître en 2007, correspondent à des cases marquées en rouge, c'est-à-dire qu'elles ne sont toujours pas connues ! D'ailleurs, déjà pour l'année 2004, les meilleurs transistors ayant fait l'objet de

publications (cf. mise à jour 2004 de l'ITRS) n'atteignent pas les spécifications de la feuille de route 2003 :  $I_{on} = 980 \text{ A/m}$ ,  $CV/I = 1,2 \text{ ps} \dots$

Year in Production	Var.	units	2004	2005	2006	2007	2008	2010	2018
Technology Generation			hp90			hp65		hp45	
Physical Lgate (High Performance)	$L_G$	nm	37	32	28	25	22	18	7
Power Supply Voltage	$V_{DD}$	V	1.20	1.10	1.10	1.10	1.00	1.00	0.70
EOT (Equivalent Oxide Thickness)	$T_{ox}$	Å	12	11	10	9	8	7	5
Effective NMOS Current Drive	$I_{on}$	$\mu\text{A}/\mu\text{m}$	1110	1090	1170	1510	1530	1900	2190
Subthreshold Off-State Leakage Drain	$I_{off}$	$\mu\text{A}/\mu\text{m}$	0.05	0.05	0.05	0.07	0.07	0.1	0.5
Device $\tau$ Performance Scaling Target	$CV/I$	ps	1.03	0.88	0.75	0.64	0.55	0.40	0.11
Saturation Threshold Voltage	$V_T$	V	0.20	0.20	0.21	0.18	0.17	0.15	0.11
Gate Poly Depletion & Inversion Layer Thickness		Å	8	7	7	4	4	4	4
Maximum Gate Leakage Limit		$\text{A}/\text{cm}^2$	4.5E+02	5.2E+02	6.0E+02	9.3E+02	1.1E+03	1.9E+03	2.4E+04
Sub-threshold Slope	S	mV/dec	86	85	87	79	74	71	81
Effective NMOS Transconductance	$G_m$	$\mu\text{S}/\mu\text{m}$	1445	1584	1686	2145	2335	2785	4331
Effective Parasitic Rsd	$R_{s,R_D}$	$\Omega\text{-}\mu\text{m}$	180	180	171	162	153	135	60
Ideal NMOS Device Gate Capacitance	$C_G$	$\text{F}/\mu\text{m}$	6.4E-16	6.1E-16	5.7E-16	6.6E-16	6.3E-16	5.7E-16	2.7E-16
Parasitic Fringe/Overlap Capacitance		$\text{F}/\mu\text{m}$	2.4E-16	2.4E-16	2.3E-16	2.2E-16	2.0E-16	1.8E-16	8.0E-17

Tableau 3 : Besoins pour les circuits à Logique Haute Performance [ITRS 2003] (blanc : des solutions existent et sont en voie d'optimisation, jaune : des solutions sont connues, rouge : pas de solutions connues).

## 2.3 Effets de canal court

Comme on peut le constater sur l'extrait de l'ITRS du Tableau 3, la loi d'échelle indiquée au paragraphe 2.1 laissant invariants les champ électriques, n'est pas suivie pour des longueurs de grille submicroniques : par exemple  $V_{DD}/T_{ox}$  n'est pas constant (on a toujours  $T_{ox} \propto 1/K$  mais  $V_{DD} \propto \alpha/K$  avec  $K > \alpha > 1$  [BACC00]...). En effet, dans les transistors courts, l'augmentation du courant sous le seuil, mentionnée précédemment, est problématique. La contrainte sur la réduction des tensions d'alimentation  $V_{DD}$  est trop forte car des  $V_{DD}$  trop faibles ne permettent pas de garantir des marges de bruit suffisantes dans les circuits. Par conséquent, les champs électriques présents dans le transistor augmentent avec la miniaturisation et certaines caractéristiques des transistors se dégradent. Ces dégradations sont appelées « effets de canal court ».

### 2.3.1 Problèmes liés aux forts champs

La présence d'un fort champ transverse confine les électrons à l'interface Si/SiO<sub>2</sub>. Or, à cause des réflexions diffusives contre cette interface très rugueuse, le nombre de collisions subies par les porteurs augmente très sensiblement par rapport à celui inhérent au substrat massif. En conséquence, la mobilité des porteurs diminue dans le canal lorsque le champ transverse augmente.

Des modèles plus ou moins empiriques de mobilité effective tenant compte de l'influence du champ de confinement  $E_z$  ont été proposés, tels que celui-ci :

$$\mu_{\text{eff}} = \frac{\mu_0}{\left[1 + \frac{E_z}{E_0}\right]^\alpha} \text{ avec } \begin{cases} E_0 = 200 \text{ kV/cm} \\ \alpha = \text{de } 0,5 \text{ à } 2 \end{cases} \text{ et } \mu_0 \text{ mobilité faible champ.}$$

Pour bien prendre en compte cette mobilité variable avec le champ, il faut recourir à la simulation numérique à deux dimensions...

### 2.3.2 Diminution de la tension de seuil $V_T$ dans les canaux courts

Les dopages des dispositifs courts ne correspondant pas aux lois d'échelle indiquées au paragraphe 2.1, on observe alors une augmentation des épaisseurs des zones de charges d'espace relativement à la longueur de canal  $L_G$ . Cela modifie notablement les caractéristiques électriques des transistors courts.

En effet, l'injection des électrons dans la zone active du transistor est contrôlée au niveau du canal par la barrière de potentiel commandée par  $V_{GS}$  (cf. Figure 17), et au niveau du substrat par la barrière de potentiel de la jonction Source(N+)/Substrat(P). Lorsque les épaisseurs des ZCE des jonctions caissons/canal s'étendent sur toute la longueur du canal, la barrière de potentiel dans le canal court est abaissée par rapport à sa valeur dans un canal « long ». Le nombre de porteurs présents dans le canal augmente alors. La tension de seuil à partir de laquelle se produit le phénomène d'inversion est donc plus faible.

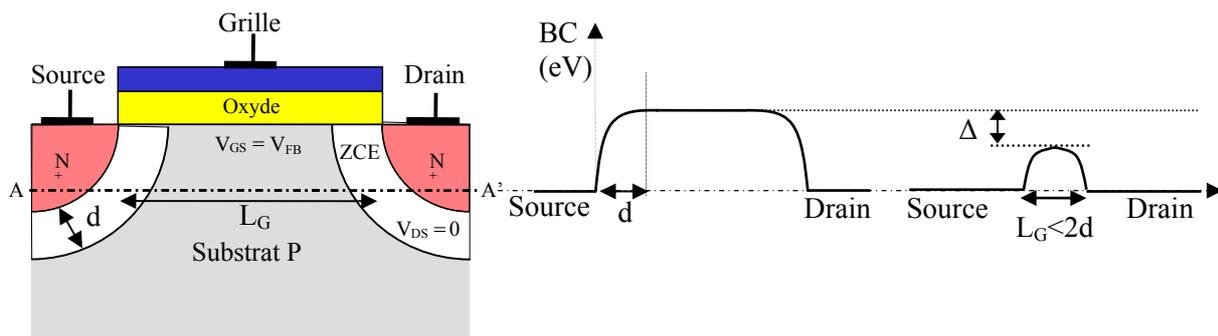


Figure 17 : Evolution de la bande de conduction dans un MOS « long » ( $L_G > 2.d$ ) et un MOS « court » ( $L_G < 2.d$ ) selon l'axe source-drain à faible  $V_{DS}$  et avec  $V_{GS}$  égal à la tension de bande plate  $V_{FB}$ . «  $d$  » épaisseur des ZCE des jonctions caissons/canal [GAUT03].

### 2.3.3 Effets de percement

De même, des phénomènes dits de percement surviennent lorsque les dimensions des zones désertées (ZCE) Source/Substrat et Drain/Substrat deviennent comparables à la longueur de la grille

$L_G$ , la distribution du potentiel dans le canal dépend alors à la fois du champ transversal (contrôlé par la tension de grille), mais aussi du champ longitudinal (contrôlé par la tension de drain). En effet, comme indiqué en Figure 18, une augmentation de la tension de drain induit un accroissement de la ZCE côté drain, ce qui provoque l'abaissement de la barrière de potentiel Source/Substrat. Cet effet est appelé DIBL : « Drain Induced Barrier Lowering ».

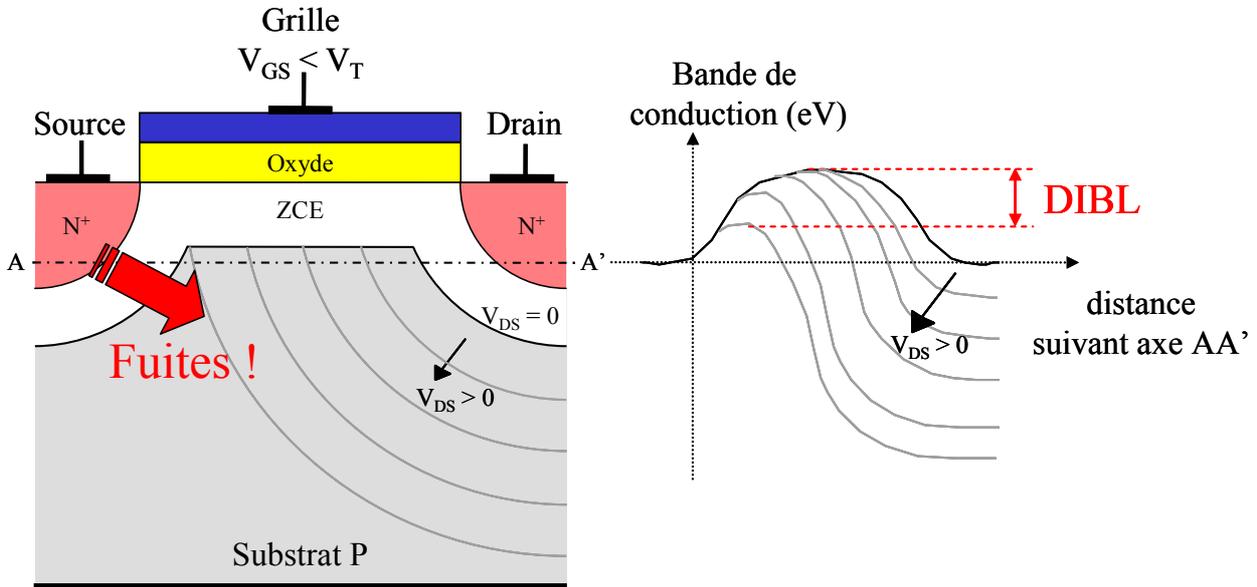


Figure 18 : Illustration des effets de perçement. La tension de drain vient modifier la barrière de potentiel qui limite l'injection des porteurs dans le canal (perçement en volume).

Le perçement (ou DIBL), comme la réduction de la longueur de grille  $L_G$ , a pour premier effet de diminuer la tension de seuil. La formule analytique suivante quantifie (après quelques « petites » approximations) la diminution de  $V_T$  en fonction de  $L_G$  et de  $V_{DS}$  [GAUT03] :

$$V_T = V_{T, L_G \rightarrow \infty} - \frac{\epsilon_{Si} \phi_d T_{ZCE}}{C_{ox} \cdot L_G^2} - \frac{\epsilon_{Si} T_{ZCE}}{C_{ox} \cdot L_G^2} V_{DS}$$

Avec

$$V_{T, L_G \rightarrow \infty} : \text{tension de seuil dans un canal long } V_{T, L_G \rightarrow \infty} = V_{FB} + 2\phi_F + \frac{qN_A T_{ZCE}}{C_{ox}},$$

$V_{FB}$  : tension de bandes plates,

$\phi_F$  : différence entre le niveau de Fermi et le quasi niveau de Fermi en bandes plates des électrons dans le substrat,

$$T_{ZCE} : T_{ZCE} = \sqrt{\frac{2\epsilon_{Si}}{q \cdot N_A}} \phi_D \quad \text{où } \phi_D \text{ le potentiel de diffusion de la jonction drain-substrat (et source-substrat) à l'équilibre : } \phi_D = \frac{k_B T}{q} \ln \frac{N_A N_D}{n_i^2}.$$

$N_D$  est le niveau de dopage des zones de contact source et drain et  $N_A$  celui du substrat.

Pour limiter les diminutions de  $V_T$  et l'effet du drain, on a donc intérêt à doper le canal pour réduire  $T_{ZCE}$  et à augmenter la capacité d'oxyde  $C_{ox}$ .

En plus de cette modification de la tension de seuil, le percement favorise l'apparition de courants de fuite comme indiqué sur la Figure 18. Sous l'effet de  $V_{DS}$ , il peut se produire soit un percement en surface qui traduit une injection supplémentaire de porteurs dans le canal à la surface de l'oxyde, soit un percement en volume favorisant l'injection parasite de porteurs dans le volume du substrat. Ces deux types de courant de fuite ont des caractéristiques différentes.

### 2.3.4 Percement en volume

Lorsque la tension de drain est supérieure à la tension de percement  $V_p$  qui permet de désorber le substrat entre les caissons, l'injection d'un courant de fuite dans le volume du substrat loin de grille est possible. Ce courant n'est pas du tout contrôlé par la tension de grille mais uniquement par la tension de drain.

### 2.3.5 Percement en surface

Les fuites par percement qui apparaissent près de la grille, à proximité de la surface de l'oxyde, sont contrôlées par la tension de drain mais aussi par la tension de la grille.

En l'absence de percement, le courant sous le seuil est indépendant de  $V_{DS}$  (cf. paragraphe 1.2.2). Ce n'est plus le cas lorsque la barrière est « percée » en surface. Soit  $\alpha$  la fraction de  $V_{DS}$  qui se répercute sur la barrière d'injection. On trouve alors que la pente  $S$  théorique est augmentée du facteur  $\alpha V_{DS}$  :

$$S = \frac{kT}{q} \ln 10 \left( 1 + \frac{C_{sub}}{C_{ox}} \right) (1 + \alpha V_{DS})$$

Cette dépendance  $\alpha$  est d'autant plus marquée que la longueur de grille est plus faible (l'influence du drain augmente) et que l'épaisseur d'oxyde est plus élevée (l'influence de la grille diminue).

## 2.3.6 Résistances séries parasites

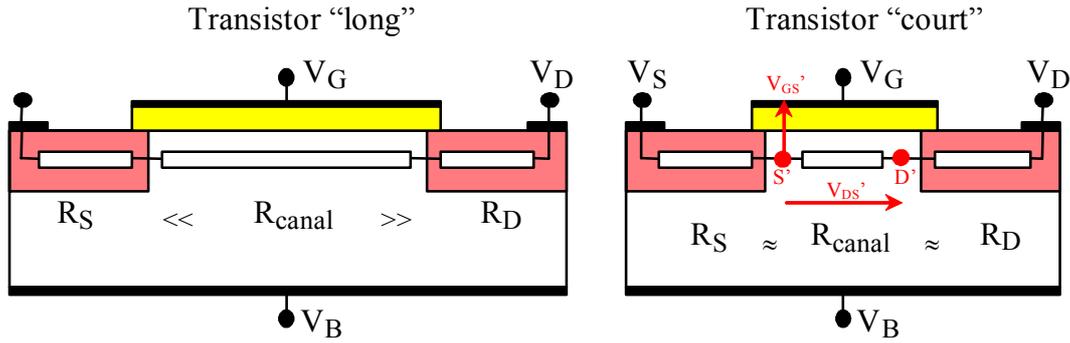


Figure 19 : Effet de la réduction de la longueur de grille sur la résistance de canal qui devient comparable aux résistances d'accès.

Lorsque la longueur du canal diminue, sa résistance ( $\propto L_G$ ) devient plus faible et éventuellement comparable à celle des caissons Source et Drain (cf. Figure 19). L'influence des résistances d'accès modifie alors fortement les caractéristiques  $I(V)$  du transistor. Les chutes de potentiel dans les caissons viennent diminuer la tension  $V_{DS}$  appliquée effectivement aux bornes du canal ainsi que la tension de grille effective  $V_{GS}$ . Cela modifie le courant  $I_{\text{on}}$  et la transconductance  $G_m$ .

En effet, il faut alors tenir compte des chutes de potentiel dans les zones d'accès. Ainsi, si  $R_S = R_D$ , les chutes de tension d'une part entre grille et source et d'autre part entre drain et source ne sont plus  $V_{GS}$  et  $V_{DS}$  mais :

$$\begin{cases} V_{GS}' = V_{GS} - R_S I_D \\ V_{DS}' = V_{DS} - 2 R_S I_D \end{cases}$$

que l'on reporte dans l'expression du courant non saturé :

$$I_D = K \left[ (V_{GS}' - V_T) V_{DS}' - \frac{V_{DS}'^2}{2} \right] \text{ avec } K = \frac{W}{L} \mu_{\text{eff}} C_{\text{ox}}$$

c'est-à-dire, en négligeant les termes en  $(R_S I_D)^2$  :

$$I_D = \frac{K}{1 + 2KR_S \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right)} \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

Dans le cas d'une saturation par pincement, le courant de saturation s'écrit sous la forme :

$$I_{D_{\text{sat}}} = \frac{K/2}{1 + K R_S (V_{GS} - V_T)} (V_{GS} - V_T)^2$$

On retrouve dans l'extrait de l'ITRS du Tableau 3, les valeurs maximales tolérées pour ces résistances d'accès en fonction de la génération technologique.

## 2.4 Effet de canal court et caractéristiques électriques

Pour résumer, les effets dits de « canal court » ont pour conséquence de dégrader le comportement des transistors submicroniques à la fois sous le seuil (pente sous le seuil, courant de fuite  $I_{off}$ , décalage de  $V_T$ ) et aussi en régime de saturation car ils augmentent la conductance  $G_d$  (le « DIBL » au dessus du seuil est appelé « DICE » : Drain Induced Conduction Enhancement). Cela peut être très préjudiciable pour les performances de l'inverseur CMOS associé, notamment pour la consommation statique  $P_{stat}$  (cf. paragraphe 1.5 ). On se propose ici de détailler comment ces phénomènes modifient les caractéristiques électriques des transistors courts.

Sur la courbe  $\log[I_D(V_{GS})]$  à faible  $V_{DS}$  de la Figure 20, on observe sous le seuil une translation vers la gauche de la caractéristique du transistor court par rapport à la courbe idéale en canal « long » traduisant la diminution de la tension de seuil  $\Delta V_T(L_G)$  décrite au paragraphe 2.3.2 . Ensuite, par effet DIBL, cette caractéristique se décale encore vers la gauche quand  $V_{DS}$  augmente :  $\Delta V_T(V_{DS})$ . De plus, les percements surfaciques induisent une augmentation de la pente sous le seuil à fort  $V_{DS}$ . Enfin à très faible  $V_{GS}$ , la contribution des fuites par percement volumique (quasi-indépendante de la tension  $V_{GS}$ ) apparaît sur la Figure 20.

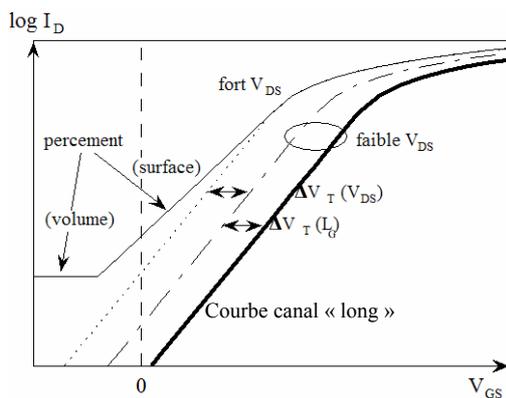


Figure 20 : Conséquences des différents effets de canal court sur la caractéristique  $I_{DS}(V_{GS})$  sous le seuil d'un MOSFET canal N.

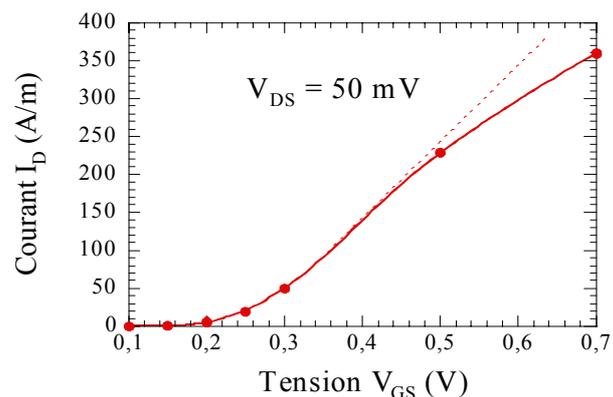


Figure 21 : Caractéristique typique  $I_{DS}(V_{GS})$  à faible  $V_{DS}$  d'un NMOS. En tirets : la caractéristique sans l'effet des résistances d'accès.

L'effet des résistances d'accès  $R_S$  et  $R_D$  se traduit à faible  $V_{DS}$  (zone ohmique) par une diminution de la transconductance  $G_m$  à fort  $V_{GS}$  visible sur la Figure 21. En effet, quand la résistance du canal  $R_{on}$  en régime ohmique est plus petite que  $R_S$  et  $R_D$ , ce sont ces dernières qui limitent le courant  $I_D = V_{DS}/(R_S + R_D)$ . Cela est moins visible dans la zone source de courant car la résistance du canal est très élevée (en  $1/G_d$ ).

## 2.5 Solutions conventionnelles

Pour contrer ces effets de canal court, différentes stratégies de miniaturisation sont mises en place au niveau des composants.

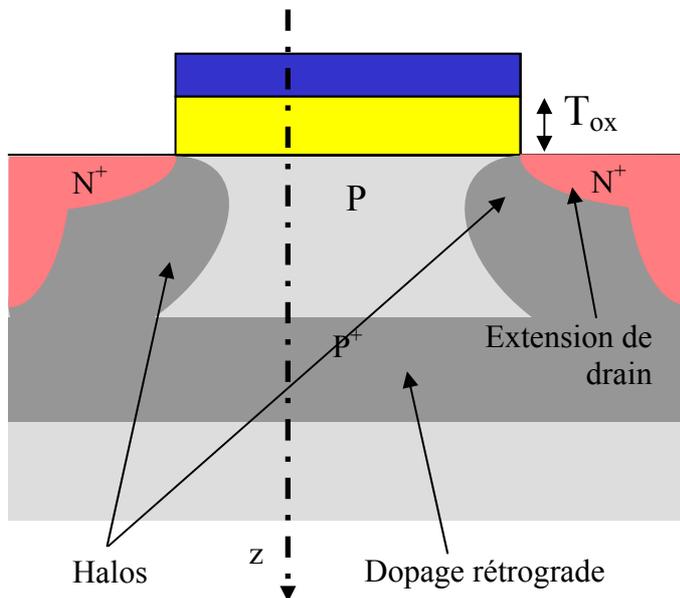


Figure 22 : Schéma de MOSFET avec dopage rétrograde (contre le percement volumique) de canal, halos autour des caissons source et drain (contre le percement surfacique) et extension de source/drain (contre le percement volumique).

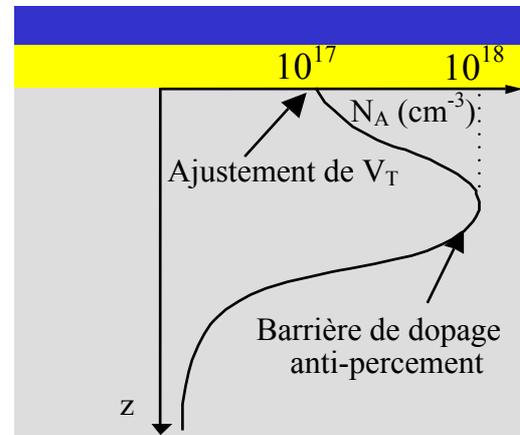


Figure 23 : Profil de dopage rétrograde.

De façon générale, la réduction de la longueur de grille d'un transistor s'accompagne de la réalisation d'un fort dopage rétrograde dans le substrat (cf. [THOM98]) sous la grille, illustré en Figure 22, afin de contrôler les effets de percement volumique. On limite ainsi l'extension des zones de charge d'espace aux endroits critiques (sous le canal) tout en permettant un réglage du  $V_T$  par un dopage plus faible du canal, ce qui est aussi bénéfique pour la mobilité des porteurs. On réalise aussi des poches, ou halos, fortement dopés, autour des caissons source et drain pour lutter contre le percement surfacique.

De plus, une ingénierie des caissons est mise en oeuvre (adaptation de la profondeur et implantations des caissons de source et de drain) afin de minimiser les résistances d'accès sans favoriser les percements. On réalise ainsi des caissons profonds loin des zones critiques pour le percement et près de ces zones on réduit leur profondeur, comme indiqué sur le schéma de la Figure 23, afin de limiter les surfaces des caissons qui peuvent être sensibles au percement et donc minimiser les courants de fuite éventuels.

## 2.6 Effets de canal ultracourt

Cependant de nouveaux problèmes apparaissent lorsque l'on atteint des longueurs de grille ultrasubmicroniques (inférieures à  $0,1 \mu\text{m}$ ) :

*Fluctuation du dopage* : dans un canal de transistor de longueur de grille ultrasubmicronique, le nombre d'atomes de dopants est suffisamment faible pour que la distribution des dopants ait une influence non négligeable sur les caractéristiques du transistor. La répartition aléatoire des atomes dopants provoque une dispersion des caractéristiques électriques des transistors réalisés sur un même substrat [DOLL04]. Par exemple pour un substrat dopé à  $10^{17} \text{at.cm}^{-3}$ , et dans une zone active de  $50 \times 50 \times 50 \text{ nm}^3$ , il n'y a que quelques dizaines d'atomes dopants. Une variation de dopants de quelques atomes dans cette zone a notamment une grande influence sur la tension de seuil  $V_T$  du transistor [ASEN98] voire sur le transport dans le canal [DOLL04].

*Courant de fuite tunnel dans la grille* : lorsque l'oxyde de grille est très mince (inférieur à 3 nm), les porteurs peuvent traverser l'isolant de grille par effet tunnel [CASS99] [LERO03]. Cela provoque une sensible augmentation du courant de fuite pour une épaisseur d'environ 1,5 nm, ce qui est très préjudiciable pour la consommation statique. On retrouve dans l'extrait de l'ITRS du Tableau 3 les valeurs maximales tolérées pour ces fuites tunnel par la grille.

*Grille très fine* : pour viser les oxydes de grille très fins (quelques Å d'épaisseur pour les dispositifs ultimes de la roadmap), il est nécessaire de contrôler leur épaisseur à la couche atomique près et cela sur des millions de transistors présents dans un circuit intégré. Cette précision semble, pour le moins, très difficile à atteindre. Dans le cas contraire, on obtient une forte variation de  $C_{ox}$  et donc de  $V_T$  [OSBU02]. Cela induit une forte dispersion des caractéristiques des composants réalisés sur un même substrat.

*Désertion de la grille* : la grille de polysilicium dopée, utilisée habituellement, est désertée en porteurs libres (cf. Figure 25) près de l'interface avec l'oxyde, comme le substrat pour  $V_{GS} > V_{FB}$ . Cette zone de charge d'espace, négligeable pour de grandes épaisseurs d'oxyde de grille, devient relativement importante en proportion pour des  $T_{ox}$  de quelques Å. Cela modifie la valeur effective de  $C_{ox}$  et influe sur les propriétés du transistor avec essentiellement un décalage du  $V_T$ . On retrouve dans l'extrait de l'ITRS du Tableau 3 les valeurs de cette épaisseur de déplétion de grille en fonction de la génération technologique.

*Effets quantiques* :

Le fort dopage de canal, nécessaire pour contrôler les effets de canal court, augmente le confinement des porteurs près de l'oxyde. Cela entraîne l'apparition d'une quantification significative de l'énergie des porteurs selon l'axe transversal (cf. Figure 24). Les fréquences

d'interaction des porteurs sont modifiées et la mobilité des porteurs s'en trouve diminuée [GAMI01] [MONS02].

Cette quantification de l'énergie provoque aussi une modification dans la répartition des porteurs qui ne se trouvent plus confinés majoritairement à l'interface Si/SiO<sub>2</sub> mais plutôt décalés d'environ 1 nm vers le substrat, comme illustré sur la Figure 25. Cela provoque aussi une modification de la capacité effective d'oxyde.

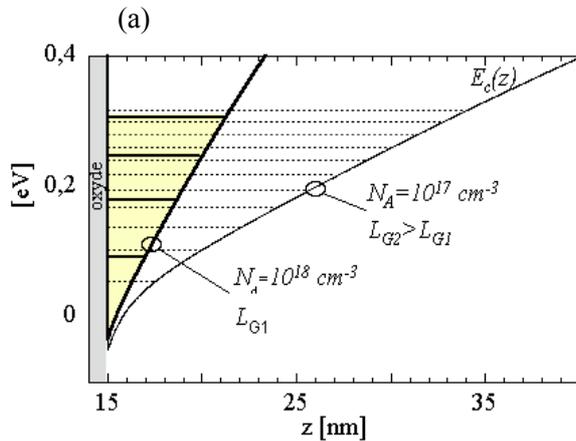


Figure 24 : Étude de la quantification en énergie selon l'axe transversal dans le canal (Gaz 2D) en fonction du dopage de canal et de la longueur de grille [MONS02].

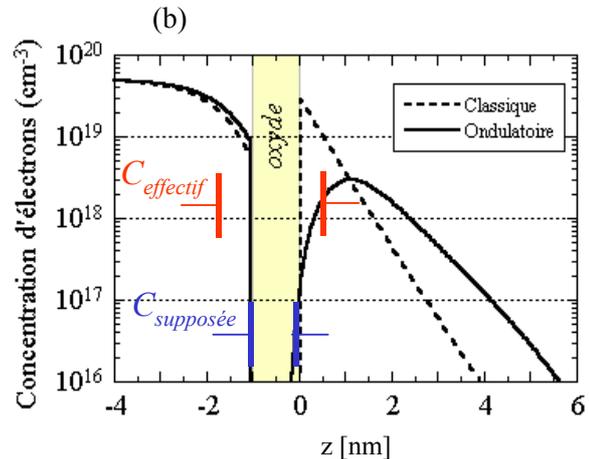


Figure 25 : Concentration des électrons dans un NMOS selon une coupe transversale au niveau du canal. Cela met en évidence le phénomène de déplétion de la grille en polysilicium et la nouvelle répartition des porteurs si l'on prend en compte la quantification selon l'axe transverse dans le canal (gaz 2D) [MONS02].

De plus, pour des longueurs de canal proches de la longueur d'onde de de Broglie ( $\lambda = 10$  nm pour des porteurs thermiques à 300 K), on ne peut plus négliger les aspects ondulatoires dans l'axe du transport. Les propriétés du transport, complètement quantiques, seront alors très différentes. C'est une limite physique incontournable (prévue vers 2010-2015 selon l'ITRS 2002) qui nécessitera de repenser complètement les dispositifs et leur fonctionnement...

On peut noter qu'à cause de ces limitations électriques, technologiques, voire physiques, les dernières technologies CMOS seront loin d'atteindre les limites théoriques ultimes [BACC00] : incertitudes quantiques, électronique à un électron... Ainsi, les portes CMOS auront toujours, malheureusement, des consommations énergétiques de plusieurs ordres de grandeur supérieures aux dispositifs idéaux permettant une commutation réversible (au sens thermodynamique, c'est-à-dire sans déperdition de chaleur) d'un état logique binaire [NEUM66].

## 2.7 Dispositifs émergents

Pour contrer tous ces effets de canal ultracourt, les solutions conventionnelles ne sont plus adaptées. En effet, pour viser des longueurs de grille de moins 25 nm, il faudrait réaliser des oxydes

de moins de 1 nm, très favorables aux fuites tunnel, et des dopages rétrogrades avec des gradients de dopants difficilement réalisables. Le pic de dopage doit atteindre  $10^{19}$  at./cm<sup>3</sup> à une profondeur de 10 nm tout en conservant un dopage de canal raisonnable d'environ  $10^{17}$  at./cm<sup>3</sup>.

Comme suggéré par l'extrait de l'ITRS du Tableau 3, le monde de la microélectronique est actuellement face à des problèmes bien plus complexes qu'auparavant. Pour poursuivre la miniaturisation, il faudra avoir recours à des alternatives au MOSFET massif. L'introduction de plusieurs innovations par nœud technologique constituera un défi majeur mais indispensable [SKOT05]. Ces innovations nécessiteront de nouvelles architectures, qui seront détaillées dans la dernière partie du chapitre, mais aussi de nouveaux matériaux.

#### *Métal de grille :*

A cause de leur faible densité d'états, les grilles en polysilicium sont sujettes aux phénomènes de déplétion décrits au paragraphe 2.6 . La zone de charge d'espace à l'interface oxyde/métal est quasi-inexistante car la densité d'états d'un métal est colossale. Le polysilicium de grille peut donc être avantageusement remplacé par un métal. En outre, pour régler la tension de seuil  $V_T$ , on peut ajuster le travail de sortie de cette grille métallique, sans devoir modifier le dopage de canal. Cela permet de s'affranchir d'un fort dopage de canal et donc d'envisager des canaux peu, voire pas dopés à plus grande mobilité.

#### *Canal à haute mobilité :*

Une augmentation de la mobilité  $\mu_{\text{eff}}$  des porteurs dans le canal induit une augmentation du courant de saturation  $I_{\text{on}}$  (cf. paragraphe 1.2.2 ) bénéfique pour la rapidité de commutation de l'inverseur. Pour augmenter la mobilité on peut soit diminuer le dopage du canal (avec une grille métallique), soit utiliser des canaux de silicium contraints [BUFL04]. On peut aussi jouer sur la direction cristallographique [YANG03].

#### *Oxyde de grille à haute permittivité diélectrique :*

Afin d'augmenter la capacité surfacique d'oxyde de grille  $C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{T_{\text{ox}}}$ , on peut réduire  $T_{\text{ox}}$ , c'est la solution traditionnelle, ou choisir un isolant avec une permittivité diélectrique  $\epsilon_{\text{ox}}$  plus importante. Cela doit permettre la réalisation de transistors fonctionnels sans exiger des oxydes ultrafins difficilement réalisables et donnant lieu à des fuites tunnel trop importantes. Malheureusement, pour l'instant, les interfaces des matériaux à haute permittivité diélectrique (« high k » en anglais) avec le silicium ne sont pas de bonne qualité et elles détériorent sensiblement la mobilité des porteurs [LUJA05]. De plus le compromis entre matériaux à forte permittivité diélectrique et à grande barrière isolante avec le silicium, pour limiter l'effet tunnel canal/grille, semble difficilement réalisable [ROBE04]. Si bien que l'interface Si/SiO<sub>2</sub> reste, à

l'heure actuelle, considérée par certains comme un « cadeau de la nature » [LUCO05] difficilement remplaçable.

### 3. Présentation des architectures multigrille

Parmi les alternatives possibles aux transistors MOS massifs 'traditionnels', les architectures multigrilles sur SOI (Silicon On Insulator : Silicium sur Isolant) sont particulièrement prometteuses particulièrement en termes de contrôle électrostatique de canal [COLI04]. Après avoir présenté la technologie SOI, différentes techniques permettant de réaliser les architectures multigrilles seront décrites et leur performances électriques actuelles comparées. Enfin, leur formidable potentiel d'innovation sera évoqué.

#### 3.1 La technologie SOI

Longtemps quasiment exclusivement utilisée pour sa bonne tenue face aux radiations [ARBE98], la technologie SOI a connu une grande diversification de ses applications aussi bien en électronique qu'en optique où elle permet la réalisation de guides optiques très performants [VIVI04]. Un substrat SOI, schématisé en Figure 26, se compose d'un empilement de trois couches. Un oxyde, dit « enterré », est intercalé entre 2 couches de silicium cristallin : la zone active et le substrat.

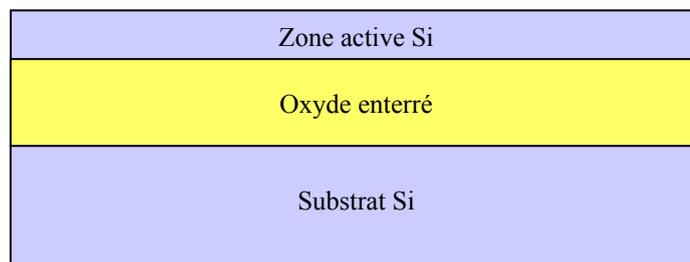


Figure 26 : Schéma d'un substrat SOI (Silicon on Insulator).

La très nette amélioration de cette technologie en qualité et en coût la rend très compétitive et ce d'autant plus qu'elle permet une bonne compatibilité avec les technologies CMOS actuelles [ALLI01]. Par exemple, la méthode révolutionnaire nommée « Smart cut », permet à l'heure actuelle de fabriquer des substrats SOI dits « Unibond » avec des couches actives de silicium ultrafines et de très bonne qualité cristalline [GAUT03]. On peut aussi citer la technologie SON (Silicon On Nothing) qui permet d'associer des zones « SOI » et « non SOI » sur une même puce [SKOT99].

#### 3.2 Transistor MOS sur SOI

Le schéma d'un transistor MOS sur SOI est présenté en Figure 27. On constate que les caissons source et drain et la zone active de silicium, tous d'épaisseur  $T_{Si}$ , reposent directement sur l'oxyde enterré ('buried oxide' en anglais) d'épaisseur  $T_{box}$ .

L'utilisation de substrats SOI permet la réalisation de transistor MOS à plus fort pouvoir bloquant et plus rapides. Dans un transistor MOS massif, le substrat de silicium est le siège d'un courant de fuite dû au percement volumique (cf. paragraphe 2.3.3 ). Dans un MOS sur SOI, en lieu et place de cette zone de percement, on trouve un isolant : l'oxyde enterré. Le SOI a donc naturellement une action bénéfique sur les effets de canal court (à condition de bien dimensionner le transistor :  $T_{\text{box}}$  faible [FENO04]...).

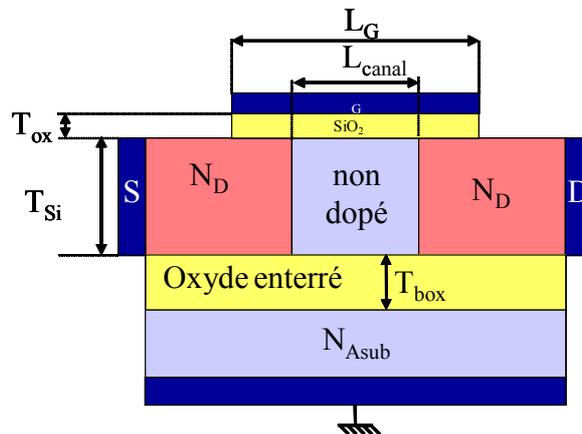


Figure 27 : Schéma d'un transistor MOS sur SOI.

De plus, le SOI réduit les capacités parasites favorisant la dynamique de l'inverseur CMOS. La diminution des surfaces des jonctions caisson/canal, par rapport à celles du transistor massif équivalent, minimise les impédances associées. En outre, si l'on utilise des zones actives non dopées, l'épaisseur de la zone de charge d'espace sous l'oxyde de grille  $T_{ZCE}$  est très grande. Cela réduit considérablement la capacité surfacique  $C_{ZCE}$  ( $=\epsilon_{\text{Si}}/T_{ZCE}$  cf. paragraphe 1.3 ). De plus, pour de faibles épaisseurs  $T_{\text{Si}}$  inférieures à  $T_{ZCE}$ , la zone active peut être complètement désertée dans la plage normale de fonctionnement. L'épaisseur de la zone de charge d'espace reste constante. Il n'y a pas de variation  $\Delta Q_{ZCE}$  de la charge de la ZCE et donc la capacité  $C_{ZCE}$  est quasi-nulle.

Malheureusement, la relativement faible profondeur des caissons source et drain dans les transistors MOS sur SOI peut accroître dramatiquement la valeur des résistances d'accès  $R_{\text{accès}}$ . En outre, ces dispositifs peuvent être le siège d'effets d'échauffement importants dus à l'importance de la résistance thermique de l'oxyde enterré [GAUT03]. Pour de forts niveaux de dissipation de puissance, ce phénomène entraîne une diminution de la mobilité effective des porteurs et donc une diminution de la conductance de drain qui peut s'avérer préjudiciable.

### 3.3 L'inversion volumique

Dès 1987, Francis Balestra et ses collaborateurs se sont intéressés aux avantages de l'inversion volumique [BALE87] sur SOI. L'utilisation d'un canal de silicium pris en sandwich entre 2 grilles face-à-face permet, sous certaines conditions, de créer une inversion dans tout le volume de silicium compris entre les caissons de source et de drain (cf. Figure 28.b). Ce fonctionnement est bien

différent de la simple mise en parallèle de transistors à inversion surfacique (cf. Figure 28.a). La répartition volumique des porteurs permet d'augmenter la section traversée par le flux de courant et donc le courant total. De plus, les porteurs moins nombreux à proximité de l'oxyde de grille sont moins perturbés par la rugosité de cette interface. Cela augmente la mobilité effective et la transconductance. De plus, l'amélioration du contrôle de charge du canal par l'ajout d'une grille diminue la pente sous le seuil. Cet exemple « historique » est à l'origine de la « lignée » des dispositifs multigrilles.

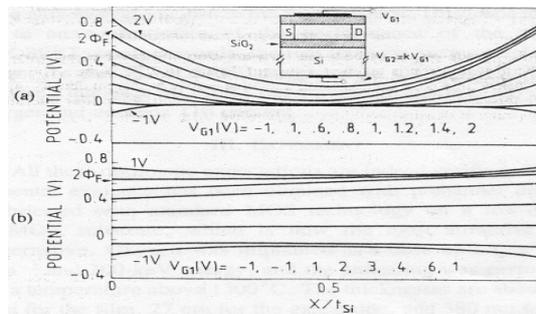


Figure 28 : Profil de potentiel à l'intérieur de la zone active pour (a) deux inversions surfaciques (dopage  $N_A = 4.10^{16} \text{ cm}^{-3}$ , épaisseur de film  $T_{si} = 300 \text{ nm}$ ) et (b) une inversion volumique ( $N_A = 3.10^{15} \text{ cm}^{-3}$ ,  $T_{si} = 100 \text{ nm}$ ) N.B.  $2\Phi_f$  correspond au potentiel d'inversion [BALE87].

### 3.4 Les architectures multigrilles

Le développement de la technologie SOI permet l'apparition de nouvelles géométries 3D de composants. Le canal de ces composants peut être contrôlé par une (c'est le transistor MOS sur SOI classique), deux, trois ou quatre grilles, comme illustré en Figure 29. Des architectures encore plus exotiques, comme par exemple le dispositif n°5 de la Figure 29 appelée transistor à « Grille en Pi », sont envisageables [PARK01].

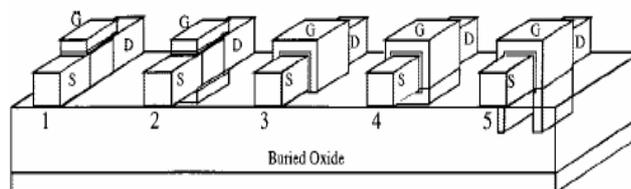


Figure 29 : MOSFET sur SOI pour différentes configurations de grille 1) Simple Grille, 2) Double grille, 3) Triple Grille, 4) Quadruple Grille et 5) une nouvelle structure : «Grille en Pi » [PARK01]

L'utilisation de plusieurs grilles permet, à épaisseur d'oxyde constante, d'améliorer le contrôle électrostatique du canal et donc de lutter efficacement contre les effets de canal court. On espère ainsi s'affranchir de la nécessité des oxydes de grille ultrafins pour réaliser les dispositifs ultimes de la feuille de route de l'ITRS. De plus, grâce au phénomène d'inversion volumique décrit précédemment, une amélioration de la mobilité effective des porteurs est attendue.

### 3.4.1 Technologies MOSFET double grille

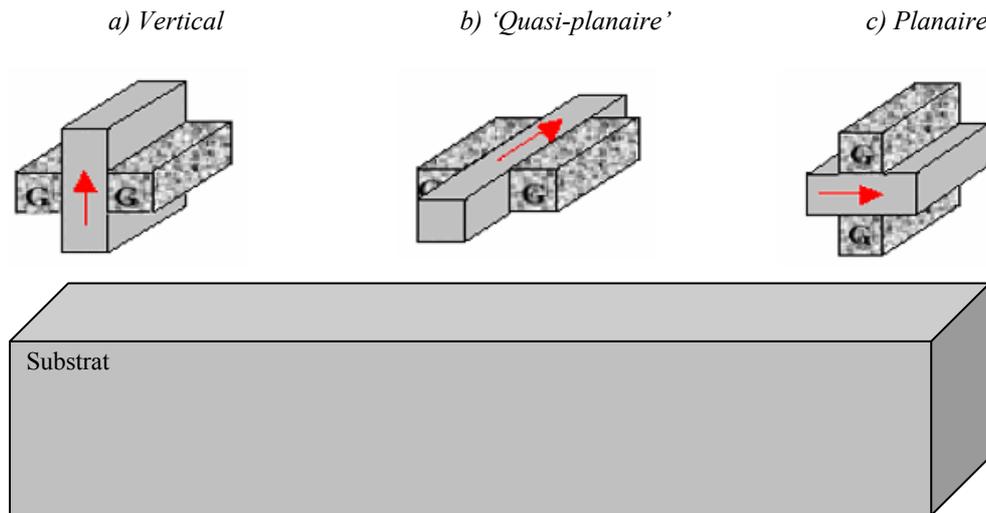


Figure 30 : Classification des transistors à double grille : a) vertical, b) 'quasi-planaire' et c) planaire. Les flèches rouges indiquent l'orientation du transport et la position des grilles la direction de l'effet de champ par rapport au plan supérieur du substrat.

La « zoologie » des transistors classe les MOSFET à double grille en fonction de la direction du transport électronique et de la direction du champ de grille par rapport au substrat. On distingue ainsi les transistors verticaux (cf. Figure 30.a), qui ont un transport électronique perpendiculaire au plan de substrat et un champ de grille parallèle, les 'quasi-planaires' (cf. Figure 30.b), où le transport électronique et le champ sont parallèles au plan de substrat et les transistors planaires (cf. Figure 30.c), qui ont une orientation inverse de celle des verticaux. Chaque « espèce » de transistors à double grille possède ses propres particularités technologiques (résolution lithographique, alignement des grilles...) qui vont influencer sur ses propriétés électriques.

#### Les transistors verticaux :

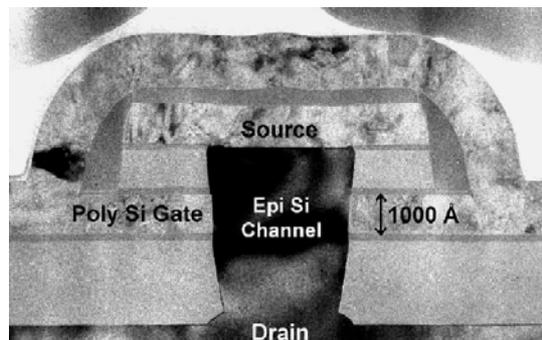


Figure 31 : Image au microscope électronique à balayage (MEB) d'un transistor vertical à grille enrobante ( $L_G = 100$  nm) [HERG02].

L'avantage des transistors verticaux, dont une image TEM (microscopie par transmission électronique) est présentée en Figure 31, est que leur longueur de grille n'est pas définie par lithographie. Cela permet de viser des longueurs de grille ultracourtes [HALL04]. De plus,

l'utilisation d'un substrat SOI n'est pas nécessaire à la fabrication de transistors à grilles multiples, grilles qui sont de plus naturellement auto-alignées.

### FinFET (transistors 'quasi-planaires') :

Les transistors à double grille 'quasi-planaires' sont nommés FinFET. Ils tirent leur nom de leur canal en forme d'aile (en anglais, 'Fin' signifie aileron).

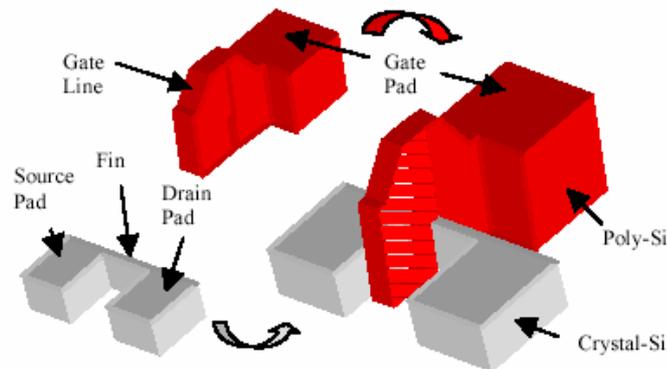


Figure 32 : Schéma d'une structure de FinFET quasi-planaire [KEDZ01].

Les étapes de réalisation de tels dispositifs sont schématisées sur la Figure 32. La zone active (en forme d'aile ou de mur) et les caissons source et drain commencent par être dessinés. L'oxyde et le contact de grille sont ensuite déposés autour de l'aile puis gravés.

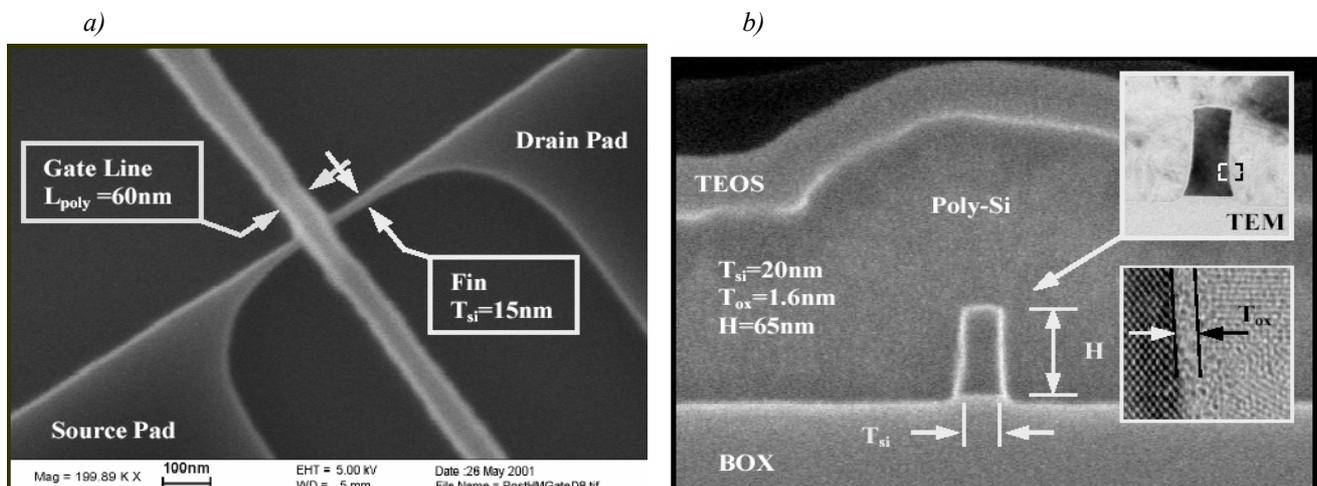


Figure 33 : Images de FinFET [KEDZ01]. a) Image au MEB, d'une vue de dessus d'un FinFET. b) Image au MEB, d'une coupe perpendiculaire au flux des porteurs d'un FinFET. L'encart montre une image en TEM de « l'aile » et de l'oxyde de grille.

Comme illustré sur les images TEM de la Figure 33, l'autoalignement des grilles d'un FinFET est naturel. De plus, des étapes technologiques supplémentaires permettent de séparer électriquement les grilles [FRIE03]. La largeur de la zone active  $W$  est déterminée par la hauteur  $H$  de l'aile, cela peut être avantageux en terme de densité d'intégration par rapport aux structures planaires. Finalement, les FinFET sont très prometteurs car ils sont simples à fabriquer si l'on

effectue la comparaison avec d'autres technologies et très compatibles avec les technologies CMOS conventionnelles [DAUG04]. Enfin, si  $T_{Si}$  est du même ordre de grandeur que  $H$  et que l'épaisseur d'oxyde est la même sur les flancs et le dessus de l'aileron, on obtient alors un transistor triple grille.

Cependant, pour prévenir les effets de canal court, ces transistors nécessitent des lithographies agressives, de résolution inférieure à  $L_G$ , pour la gravure de l'aileron qui détermine l'épaisseur de la zone active  $T_{Si}$ . La hauteur de l'aileron et la qualité des flancs de l'aileron sont aussi des paramètres délicats à contrôler [WU05]. Enfin, l'alignement des caissons source et drain, qui ne sont pas dessinés en même temps que la grille, pose problème. Pour toutes ces raisons, les FinFET semblent encore être difficilement réalisables en grande série.

*DG 'tunnel':*

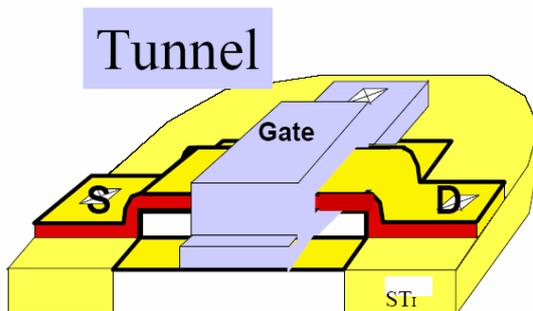


Figure 34 : Schéma d'une structure de Double grille « tunnel ».

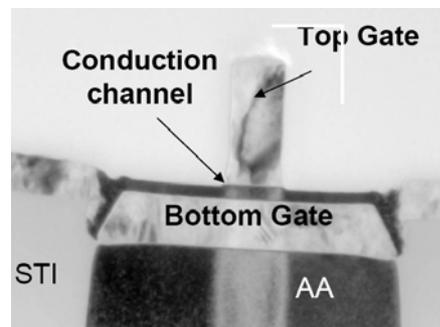


Figure 35. Image au MEB, d'une coupe perpendiculaire au flux des porteurs d'un Double grille « tunnel ». [HARR03]

Les dispositifs dits « tunnel » sont des architectures planaires sur SON réalisées sur un substrat STI (Shallow Trench Isolation). Grâce à une gravure sélective d'une couche de Silicium-Germanium (SiGe) sacrificielle, on réalise une zone active « en pont » que l'on enrobe ensuite par une grille. On obtient ainsi la structure schématisée en Figure 34 (schéma idéalisé en ce qui concerne la longueur de la grille arrière) et la coupe présentée en Figure 35 [HARR03].

Cette technologie planaire ne nécessite pas une résolution lithographique inférieure à la longueur de grille et autorise l'utilisation de zones actives très minces. Ainsi, ces dispositifs permettent de lutter efficacement contre les effets de canal et présentent des rapports  $I_{on}/I_{off}$  très intéressants [HARR03].

Cependant le non autoalignement des grilles entraîne une forte augmentation de la capacité de recouvrement  $C_{rec}$ , dégradant la vitesse des circuits. De plus, les couches d'isolation (STI) peuvent limiter la densité d'intégration [CERU04].

MOSFET à double grille à technologie planaire :

Les architectures planaires classiques sur SOI sont basées sur celles des transistors simple grille sur SOI. L'oxyde enterré est remplacé par un second empilement de grilles symétriques du premier mais électriquement indépendant. Ces dispositifs ne nécessitent pas de résolution lithographique inférieure à la longueur de grille  $L_G$  car le procédé « smart cut » permet, sans lithographie (contrairement aux FinFET), de réaliser des couches actives ultraminces ( $< 10$  nm).

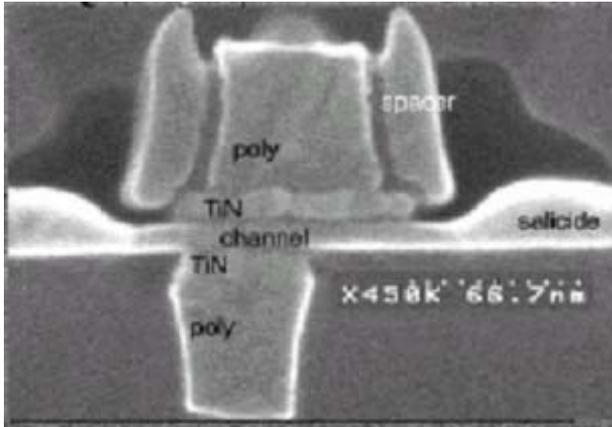


Figure 36. Image au MEB, d'une coupe de transistor à double grille non-autolignée [VINE05].

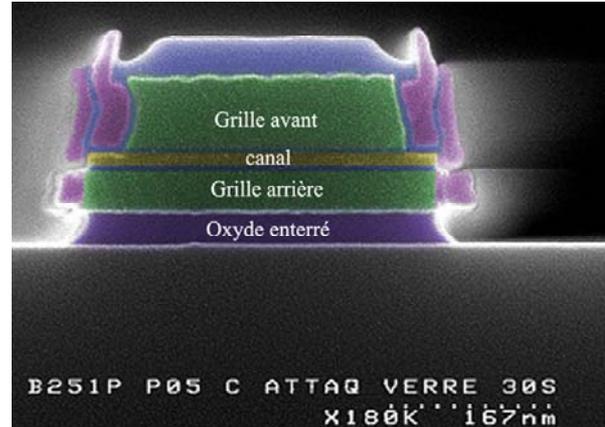


Figure 37. Image au MEB, d'une coupe de transistor à double grille autoalignée. Projet RMNT CMOS-D-ALI.

Les transistors à double grille planaires réalisés à l'heure actuelle n'ont pas leurs deux grilles parfaitement alignées. En effet, les grilles sont gravées successivement et sont alignées « optiquement ». L'imprécision sur le non alignement des grilles génère une forte dispersion des caractéristiques des composants courts [WIDI04] interdisant la fabrication en grande série. En effet, lorsque la grille inférieure recouvre un des caissons, la capacité parasite  $C_{rec}$  est accrue, dégradant le « CV/I ». A très fort désalignement, on peut même avoir un dispositif ne possédant qu'une seule grille active. Cependant, les dispositifs peu désalignés, même courts ( $L_G$  inférieur à 20 nm), sont très performants [VINE05].

Pour fabriquer les transistors ultimes, la réalisation de transistors planaires auto-alignés est indispensable mais technologiquement complexe. La Figure 37 présente la structure morphologique d'un tel transistor développée au LETI dans le cadre du projet RMNT CMOS-D-ALI auquel nous participons pour l'aspect simulation.

### 3.4.2 Technologies MOSFET Multigrille

Conventionnellement, le « nombre de grilles » indique le nombre de côtés de la zone active contrôlés efficacement par une grille. Ainsi, dans les structures à grilles enrobantes : FinFET, transistor verticaux... où l'épaisseur de la zone active tend à être du même ordre de grandeur que sa largeur ( $W \approx T_{Si}$ ), on parle de transistor triple grille si un côté de la zone active n'est pas recouvert

par une grille, ou sinon on parle de quadruple grille (grille complètement enrobante). Ces architectures nécessitent des efforts de réalisation supplémentaires par rapport aux dispositifs double grille car l'utilisation d'une lithographie très agressive est indispensable quelle que soit l'orientation du transistor.

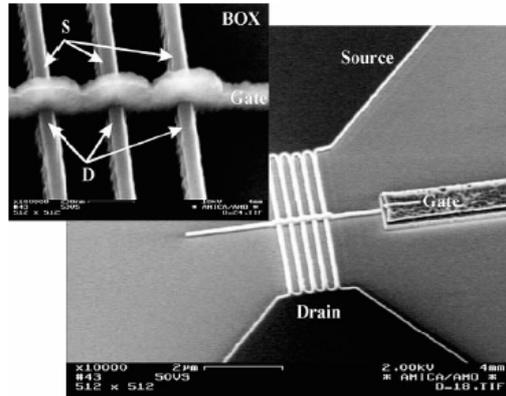


Figure 38. Image au microscope électronique à balayage, d'une structure FinFET triple grille ( $W \approx T_{Si}$ ) [LEMM04]

Les FinFET permettent de réaliser des dispositifs triple grille comme illustré sur la Figure 38 [LEMM04]. On remarquera qu'il y a alors plusieurs canaux de conduction ou « doigts » en parallèle pour augmenter le courant délivré. Les transistors quadruple grille sont réalisés à partir de structures verticales [HERG02] voire « tunnel ».

### 3.5 Comparaison actuelle des différentes architectures MOS

Le Tableau 4 résume les performances de transistors réalisés à l'heure actuelle. Le transistor massif à canal de silicium contraint d'IBM (correspondant à la zone surlignée du tableau) présenté en 2003 à la conférence IEDM reste encore compétitif pour les longueurs de grille de 45 nm. En effet, une tension d'alimentation  $V_{DD}$  de 1 V suffit au NMOS pour atteindre un courant  $I_{on}$  de plus de 1000  $\mu A/\mu m$ . La faible épaisseur  $T_{ox}$  de 1,2 nm lui permet aussi de présenter une pente sous le seuil convenable de 85 mV/dec.

Les architectures multigrilles affichent de meilleures capacités de blocage, ce qui permet, par exemple, au transistor à double grille de l'équipe du LETI [VINE05] d'atteindre des pentes sous le seuil similaires au transistor contraint de 45 nm avec une longueur de grille de 20 nm. Cependant ces dispositifs multigrilles ont encore besoin d'être optimisés pour atteindre les performances attendues par la feuille de route pour les grilles ultracourtes.

Reference	Device / Process Details	L gate	Vdd	EOT	NMOS I <sub>dsat</sub>	PMOS I <sub>dsat</sub>	NMOS I <sub>off</sub>	PMOS I <sub>off</sub>	NMOS Swing	NMOS CV/I	NMOS DIBL
		nm	V	Å	μA/μm	μA/μm	nA/μm	nA/μm	mV/dec	ps	mV/V
F.L. Yang et al. / TSMC / IEDM 2002	25 nm Omega FET	25	0.7	17 - 19	1300	550	1000	1000		0.39	
J. Kedzierski et al. / IBM / IEDM 2002	Metal-gate FinFET, undoped body, RSD, CoSi <sub>2</sub> S/D, NiSi gate	100	1.25	16	970	600	120	20	70		25
B.S. Doyle et al. / Intel / IEEE Electron Device Letters: 24 (2003) 263	Tri-gate, TSi=36 nm, WSi=55nm, no halo implants, raised source drain	60	1.3	15	521	660	70	24	68		41
V. Chan et al. / IBM / IEDM 2003	bulk transistor, 90 nm node, strain from STI and nitride	45	1.0	12	1010	400	100	100	85		100
S. Harrison / STM, LETI, Philips, L2MP / IEDM 2003	Double gate SON	70	1.2	20	1333 - 1954		1 - 283		70 - 80		30 - 70
Vinet M. et al. / LETI / IEEE EDL 2005	Metal Gate/Bonded Double gate	20	1.2	17	822-1250		0.9-1300		81-102		60-220

Tableau 4 : Comparatif des performances d'architectures multigrille actuelles ([www.clarycon.com/transistor\\_perfo.html](http://www.clarycon.com/transistor_perfo.html) et [VINE05]).

### 3.6 Nouvelles logiques

Parallèlement aux améliorations des performances CMOS classiques, les transistors multigrilles peuvent offrir de nouvelles fonctionnalités. D'ailleurs une bonne partie de leurs applications reste sûrement à être imaginée.

Les architectures double grille à grilles indépendantes sont particulièrement intéressantes. En effet, elles permettent de réaliser, sur une même puce, des transistors MOS simple grille à tension de seuil  $V_T$  variable, commandée par la tension appliquée sur la seconde grille. Cela permet, par exemple, de réaliser un comparateur à fenêtre d'hystérésis variable, dont le schéma et la caractéristique sont illustrés en Figure 39, sans nécessiter l'utilisation d'une foultitude de transistors de tailles différentes (et donc de  $V_T$  différents) [KUMA04].

Ces dispositifs permettent aussi d'envisager de nouvelles briques logiques plus compactes. Dans cet esprit, Mitra [MITR04] a proposé une nouvelle architecture de porte NOR schématisée sur la Figure 40 qui est composée de 2 transistors DGMOS en technologie NMOS. En technologie simple grille, au moins trois transistors sont nécessaires. Une nouvelle bibliothèque de circuits est ainsi à concevoir pour optimiser l'utilisation de ces nouvelles architectures.

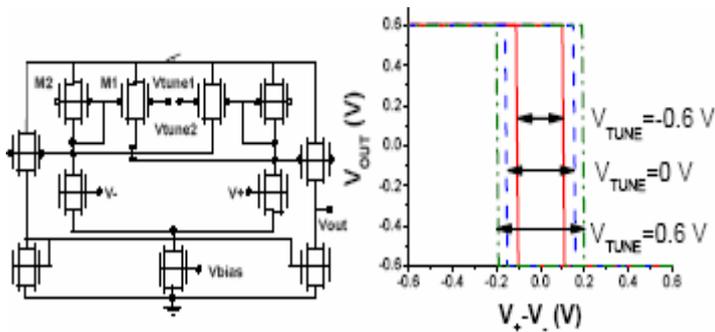


Figure 39 : Comparateur avec fenêtre d'hystérésis contrôlée par la tension  $V_{TUNE}$  [KUMA04].

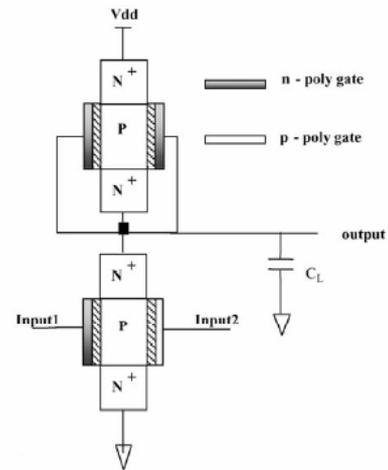


Figure 40 : Porte NOR en logique NMOS avec des DGMOS asymétriques. [MITR04].

En outre, des applications en hyper-fréquence sont envisagées. Par exemple l'utilisation des transistors à double grille comme dispositifs à VMT (Velocity Modulation Transistor) est étudiée. Une polarisation asymétrique ( $V_{GS} = V_{GS1} = -V_{GS2}$ ) des grilles de DGMOS permet de sélectionner un des deux canaux de conduction sous une des grilles. Si on a une grande différence de mobilité entre le canal de chaque grille, le courant est alors très différent en fonction du canal choisi. Modifier la polarisation des grilles permet de changer rapidement de canal de conduction s'ils sont suffisamment couplés (proches). Ainsi des simulations Monte Carlo indiquent que des DGMOS longs ( $L_G = 500$  nm) sur silicium peuvent commuter en moins d'une picoseconde [SAMP05]. Ils peuvent ainsi prétendre concurrencer les MOS ultimes en silicium voire les dispositifs en matériaux III-V.

#### 4. Objectifs de l'étude

Cette première partie a mis en évidence l'intérêt de la miniaturisation des transistors à effet de champ de type MOS en termes de performance des applications logiques mais également ses limitations. En effet, de nombreux effets néfastes (percements, fuites tunnel...) sont susceptibles de dégrader les caractéristiques des dispositifs très fortement submicroniques. Pour lutter contre ces effets de canal court, l'intérêt des nouvelles architectures multigrilles, réalisables grâce à l'amélioration des substrats SOI, a été clairement souligné. Dans ce contexte, ce travail s'appuie sur le logiciel MONACO, simulateur de transport Monte Carlo pour étudier les transistors MOS multigrilles.

Dans la seconde partie de ce mémoire, la validité du régime de mobilité dans des dispositifs ultracourts est discutée grâce à une analyse fine du transport électronique, naturellement permise par l'algorithme Monte Carlo. Les modèles reposant sur l'hypothèse de mobilité sont à la base des

lois d'échelle qui encadrent la miniaturisation des composants. Dans la troisième partie, les performances électriques des dispositifs multigrilles simulés sont évaluées en détail. Pour optimiser le dimensionnement de ces dispositifs, nous nous sommes intéressés en particulier aux effets de canal court et aux valeurs des différentes impédances parasites présentées en page 25. Enfin, la dernière partie de ce mémoire décrit une nouvelle version « semi-quantique » du logiciel MONACO développée durant cette thèse. Cette version permet la prise en compte des effets de quantification d'énergie qui ont lieu, perpendiculairement à la direction du transport de charges (effets illustrés sur la Figure 25).

## Chapitre II :

### Transport balistique dans les dispositifs ultracourts

Comme présenté dans le chapitre précédent, les modèles physiques, utilisés de nos jours par les ingénieurs en recherche et développement, permettant d'obtenir les relations courant/tension dans les transistors MOS sont basés sur l'hypothèse dite « de mobilité ». Elle suppose que les porteurs subissent un nombre d'interactions suffisamment important pour leur permettre d'être constamment en équilibre local avec le réseau cristallin, comme illustré schématiquement en Figure 41.a. Il existe alors une relation unique entre le champ électrique  $\vec{E}$  et la vitesse moyenne des porteurs  $\vec{v}$  :

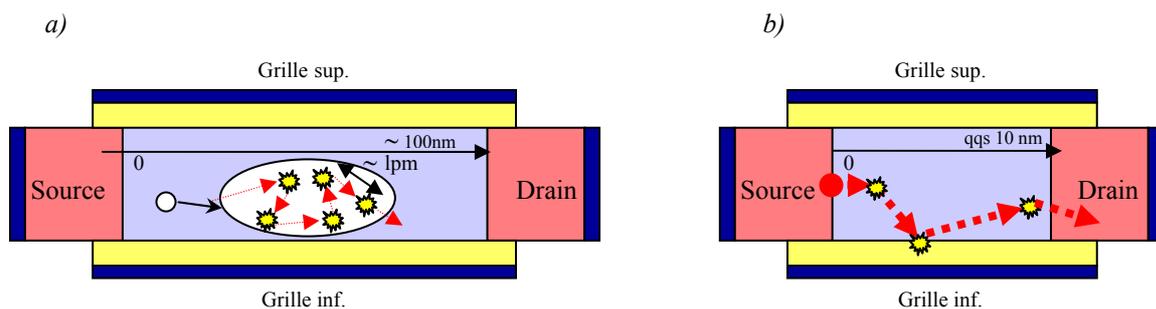
$$\vec{v} = -\mu_{\text{eff}} \times \vec{E} \quad (\mu_{\text{eff}} \text{ mobilité effective cf. page 34}).$$


Figure 41 : Schéma d'une trajectoire typique d'un porteur avec ses interactions dans a) un transistor long et b) un transistor ultracourt. *lpm* : libre parcours moyen des porteurs

Dans les transistors ultimes où la longueur de grille est fortement submicronique, la longueur du canal est de l'ordre de la distance moyenne entre deux interactions, appelée libre parcours moyen *lpm* (environ dix nanomètres dans Si à 300 K). Le nombre d'interactions subies par les porteurs lors de leur traversée d'un canal ultracourt est donc très faible, comme illustré sur le schéma de la Figure 41.b. Ce nombre peut même être nul, le porteur est alors qualifié de « balistique ». L'hypothèse d'un régime de mobilité est donc sérieusement remise en question dans les dispositifs ultracourts.

Dans le modèle classique du MOSFET basé sur le concept de mobilité, le courant varie de façon inversement proportionnelle à la longueur de grille, relation à la base de la loi d'échelle présentée dans le Tableau 2 du chapitre I. Mais, dans les dispositifs caractérisés par un transport fortement hors d'équilibre que devient cette loi d'échelle qui a suffi à assurer l'hégémonie des transistors MOS pendant plusieurs décennies ?

Ce chapitre commence par un rappel des équations fondamentales régissant le transport d'un gaz d'électrons libres dans le formalisme de Boltzmann et des méthodes numériques classiquement utilisées pour les résoudre. Les équations proposées pour décrire le transport balistique ou quasi balistique seront introduites. Ensuite, grâce aux simulations Monte Carlo classiques, nous démontrerons la présence d'un transport très fortement non stationnaire voire quasi balistique dans des dispositifs ultracourts. Nous essayerons de quantifier la part d'électrons balistiques, son influence sur les caractéristiques électriques et sa dépendance par rapport à certains paramètres physiques et électriques. Cela nous permettra de discuter les modèles analytiques proposés par quelques auteurs. Enfin, nous comparerons les résultats de simulations avec les mesures de balisticité que j'ai réalisées sur des dispositifs à double grille planaire non-autoalignée [VINE05] lors d'un séjour de trois semaines au CEA-LETI à Grenoble.

## 1. Description du transport électronique

Dans cette partie, nous présentons les équations fondamentales qui régissent le transport d'un gaz électronique classique à partir du formalisme de Boltzmann. Après avoir discuté le domaine de validité de ce dernier, nous présenterons les différentes méthodes et approximations utilisées en pratique pour les résoudre. Nous finirons cette partie en introduisant les nouveaux modèles analytiques proposés pour prendre la relève des modèles stationnaires évoqués au chapitre I pour l'étude des canaux ultracourts.

### 1.1 Equation de transport

#### 1.1.1 Fonction de distribution

L'étude classique du transport s'effectue à travers la connaissance de l'évolution temporelle de l'état des porteurs dans le dispositif. On définit la fonction de distribution  $f(\vec{r}, \vec{k}, t)$  qui représente la probabilité d'avoir à l'instant  $t$  un porteur positionné dans l'espace aux coordonnées  $\vec{r}$  et possédant un vecteur d'onde  $\vec{k}$ .

Pour une structure au repos, cette fonction se résume à la distribution de Fermi-Dirac. La connaissance de  $f(\vec{r}, \vec{k}, t)$  est importante car elle permet de déduire toutes les grandeurs caractéristiques du système telles que la concentration en porteurs  $n(\vec{r}, t)$  ou la valeur totale de toute autre grandeur transportée  $A(\vec{r}, t)$  (vitesse, énergie...):

- $n(\vec{r}, t) = \frac{1}{\Omega} \sum_{\vec{k}} f(\vec{r}, \vec{k}, t)$  où  $\Omega$  est le volume de normalisation,
- $A(\vec{r}, t) = \frac{1}{\Omega} \sum_{\vec{k}} A(\vec{k}) \cdot f(\vec{r}, \vec{k}, t)$ .

L'expression de la densité de courant est ainsi donnée par :

- $\vec{J}(\vec{r}, t) = \frac{1}{\Omega} \sum_{\vec{k}} q \cdot \vec{v}(\vec{k}) \cdot f(\vec{r}, \vec{k}, t)$  car  $\vec{J}(\vec{k}) = q \cdot \vec{v}(\vec{k})$

### 1.1.2 Equation de Boltzmann

L'étude du transport consiste à déterminer l'évolution des fonctions de distribution hors d'équilibre, c'est-à-dire à résoudre l'équation de Boltzmann.

Un bilan de variation d'une fonction  $f(x, t)$  entre  $t$  et  $t+dt$  (cf. Figure 42) donne :

$$\frac{\partial f}{\partial t} = \frac{\partial x}{\partial t} \cdot f(x, t) - \frac{\partial x}{\partial t} \cdot f(x + dx, t) + G(x, t)$$

avec  $G(x, t)$  taux net de génération par

unité de temps.

Soit 
$$\frac{\partial f}{\partial t} = -v_x \cdot \frac{\partial f(x, t)}{\partial x} + G(x, t)$$

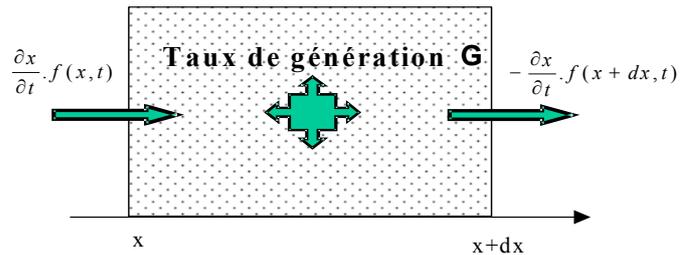


Figure 42 : Bilan 1D de la fonction  $f(x)$  entre  $x$  et  $x+dx$  durant  $dt$

En faisant de même un bilan de variation entre  $t$  et  $t+dt$  de la fonction de distribution  $f(\vec{r}, \vec{k}, t)$  selon ses 6 dimensions ( $x, y, z, k_x, k_y, k_z$ ), il vient l'Equation de Boltzmann (EB) [LUND00] :

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \vec{\nabla}_{\vec{r}} f + \frac{\vec{F}}{\hbar} \cdot \vec{\nabla}_{\vec{k}} f = \left. \frac{\partial f}{\partial t} \right|_{\text{coll}} + G(\vec{r}, \vec{k}, t)$$

Où :

$\vec{F} = \hbar \frac{\partial \vec{k}}{\partial t}$  est la force extérieure appliquée,

$\vec{v} = \frac{1}{\hbar} \vec{\nabla}_{\vec{k}} E(\vec{k})$  avec  $E(\vec{k})$  la relation de dispersion de l'énergie,

$G(\vec{r}, \vec{k}, t)$  correspond au taux net de génération/recombinaison par unité de temps,

$\left. \frac{\partial f}{\partial t} \right|_{\text{coll}}$  décrit la variation de  $f(\vec{r}, \vec{k}, t)$  due aux collisions que subissent les porteurs avec le milieu

de propagation.

A l'équilibre thermique ( $\frac{\partial}{\partial t} = 0$  dont  $\left. \frac{\partial f}{\partial t} \right|_{\text{coll}} = 0$ ), et en négligeant les phénomènes de

génération/recombinaison, l'EB se réduit à :  $\vec{v} \cdot \vec{\nabla}_{\vec{r}} f + \frac{\vec{F}}{\hbar} \cdot \vec{\nabla}_{\vec{k}} f = 0$ . Les distributions maxwelliennes

$f_{\text{équilibre}}(\vec{r}, \vec{k}, t) = e^{\frac{E(\vec{r}, \vec{k}) - \Phi_F}{k_B T}}$  où  $E(\vec{r}, \vec{k}) = E_{\text{BC}}(\vec{r}) + \frac{\hbar k^2}{2m^*}$  avec  $E_{\text{BC}}(\vec{r})$  énergie de la bande de conduction,  $m^*$  masse effective et  $\Phi_F$  niveau de Fermi, sont solutions de cette équation. On remarque alors que les distributions en vecteur d'onde  $f_{\text{équilibre}}(\vec{k})$  ou en vitesse ( $m^* \vec{v} = \hbar \vec{k}$ ) sont gaussiennes (si la bande est parabolique).

### 1.1.3 Intégrale de collision

La détermination de l'intégrale de collision issue du terme  $\left. \frac{\partial f}{\partial t} \right|_{\text{coll}}$  constitue une difficulté majeure dans la résolution de l'EB. Il s'agit de déterminer à l'instant  $t$  et à la position  $\vec{r}$  les probabilités d'interactions qui provoquent le passage d'un état initial de vecteur onde  $\vec{k}$  vers un autre état  $\vec{k}'$  ainsi que de celles permettant de passer d'un état  $\vec{k}'$  vers l'état  $\vec{k}$ .

En tenant compte de ces 2 possibilités et en sommant sur toutes les transitions possibles on peut écrire l'intégrale de collision sous la forme suivante.

$$\left. \frac{\partial f}{\partial t} \right|_{\text{coll}} = \int [s(\vec{k}', \vec{k}) \cdot f(\vec{k}') - s(\vec{k}, \vec{k}') \cdot f(\vec{k})] \cdot d\vec{k}'$$

Avec

$s(\vec{k}, \vec{k}')$  : probabilité de passer d'un état  $\vec{k}$  vers un état  $\vec{k}'$ ,

$f(\vec{k})$  : distribution en vecteur d'onde ou probabilité d'occupation de l'état  $\vec{k}$ .

Dans les simulation de type Monte Carlo la détermination de  $s(\vec{k}, \vec{k}')$  est réalisée grâce à la règle d'or de Fermi [COHE96].

### 1.1.4 Domaine de validité

Il est important de comprendre les limitations de l'équation de Boltzmann appliquée à l'étude du comportement d'un système d'électrons dans un cristal. Cette approche constitue une approximation **semi-classique** du transport de particules individuelles dont on néglige les propriétés ondulatoires : chaque électron est considéré comme un corpuscule classique (parfaitement localisé) qui obéit aux lois de la dynamique de Newton.

La mécanique quantique n'est utilisée que pour décrire la structure de bandes d'énergie et les collisions, c'est-à-dire uniquement pour traiter des phénomènes liés à des variations rapides de potentiel, dans le temps ou dans l'espace.

Ces hypothèses semi-classiques ne sont valables que lorsque les dimensions du dispositif sont très grandes devant :

- la longueur d'onde des électrons ( $\lambda = \sqrt{\frac{\hbar^2}{2m^*E}} \approx 2 \text{ nm}$  à 300 K pour  $E = 3k_B T/2$ ),
- la longueur de relaxation de phase de la fonction d'onde.

Si une de ces conditions n'est pas remplie, on ne peut plus négliger certains phénomènes quantiques :

- quantification de l'énergie des porteurs dans la direction perpendiculaire au transport,
- effet tunnel,
- phénomènes d'interférence entre porteurs en opposition de phase,
- réflexion possible sur une barrière d'énergie inférieure à l'énergie cinétique de l'électron,
- ....

## 1.2 Approximations usuelles de l'équation de Boltzmann

L'équation de Boltzmann est une équation intégral-différentielle difficile à résoudre ou, du moins, dont la résolution est mal adaptée à l'étude des dispositifs à plus d'une direction dans l'espace réel, en raison d'un coût (trop) important en ressources informatiques. C'est pourquoi des approches plus simples, basées sur des dérivations de l'EB, ont été mises en œuvre en vue de pouvoir analyser, concevoir et optimiser de manière plus souple les structures des composants.

### 1.2.1 Dérive diffusion et temps de relaxation

On simplifie généralement l'intégrale de collision en écrivant (approximation de Boltzmann) :

$$\left. \frac{\partial f}{\partial t} \right|_{\text{coll}} = -\frac{f - f_{\text{équilibre}}}{\tau} \text{ avec } \tau \text{ temps de relaxation moyen. Ainsi, en régime permanent } \left( \frac{\partial}{\partial t} = 0 \right), \text{ dans}$$

un matériau de dopage et de température uniformes (d'où  $\vec{\nabla}_r f = 0$ ), l'EB simplifiée devient :

$$\frac{\vec{F}}{\hbar} \cdot \vec{\nabla}_k f + \frac{f - f_{\text{équilibre}}}{\tau} = 0$$

Si on considère un fonctionnement sous faible champ électrique, les porteurs restent constamment dans un état proche de l'équilibre, c'est-à-dire à la température du réseau cristallin. Les fonctions de distribution en vecteur d'onde ou en vitesse sont des gaussiennes déplacées

soit  $f(\vec{k}) = f_{\text{équilibre}}(\vec{k} - \vec{k}_0)$  ou  $f(\vec{v}) = f_{\text{équilibre}}(\vec{v} - \vec{v}_0)$  avec  $\vec{k}_0$  vecteur d'onde moyen et  $\vec{v}_0$  vitesse moyenne, comme illustré par la courbe noire de la Figure 43. Grâce à ces simplifications de l'EB, on obtient [GAUT03] un courant de conduction (fonction du champ électrique) qui suit la relation bien connue dite de « dérive » :  $\vec{v} = -\mu_{\text{eff}} \vec{E}$  avec  $\mu_{\text{eff}} = \frac{q\tau}{m^*}$  mobilité effective et  $\vec{E}$  champ électrique.

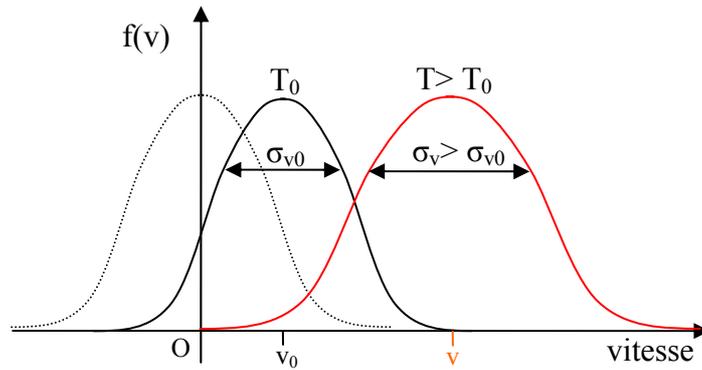


Figure 43 : Distributions en vitesse. En tirets distribution à l'équilibre sous champ électrique nul : fonction gaussienne, en noir gaussienne déplacée de vitesse moyenne  $v_0$  sous faible champ, en rouge gaussienne déplacée « élargie » qui traduit un « chauffage » des électrons sous champ plus fort.  $T$  température du gaz d'électrons.  $T_0$  température du réseau cristallin.  $\sigma_{v,0}$  écart type de la distribution à l'équilibre thermique :  $T=T_0$  ( $\sigma_{v,0}=k_B T_0/q$ )

En couplant cette relation avec les équations de diffusion, les équations de continuité et l'équation de Poisson, on obtient pour les électrons et les trous le système d'équations linéaires suivant :

$$\vec{J}_n(\vec{r}, t) = -qD_n \vec{\nabla} n(\vec{r}, t) + qn(\vec{r}, t) \mu_n \vec{E} \quad (\text{densité de courant de diffusion et de conduction d'électrons})$$

$$\vec{J}_p(\vec{r}, t) = -qD_p \vec{\nabla} p(\vec{r}, t) + qp(\vec{r}, t) \mu_p \vec{E} \quad (\text{densité de courant de diffusion et de conduction de trous})$$

$$\frac{\partial n(\vec{r}, t)}{\partial t} = G_n(\vec{r}, t) - \frac{1}{q} \vec{\nabla} \cdot \vec{J}_n(\vec{r}, t) \quad (\text{équation de conservation des électrons})$$

$$\frac{\partial p(\vec{r}, t)}{\partial t} = G_p(\vec{r}, t) - \frac{1}{q} \vec{\nabla} \cdot \vec{J}_p(\vec{r}, t) \quad (\text{équation de conservation des trous})$$

$$\vec{\nabla} \cdot (\epsilon_{sc} \vec{E}) = q(p(\vec{r}, t) - n(\vec{r}, t) + N_D + N_A) \quad (\text{équation de Poisson})$$

$$D_n = \frac{k_B T}{q} \mu_n \quad \text{et} \quad D_p = \frac{k_B T}{q} \mu_p \quad (\text{relations d'Einstein})$$

$G_n(\vec{r}, t)$  et  $G_p(\vec{r}, t)$  dépendent des modèles de génération/recombinaison choisis.

Avec  $N_D$  et  $N_A$  qui correspondent respectivement aux dopages en ions donneurs et accepteurs d'électrons.

Les 4 inconnues sont respectivement la densité de courant et la concentration pour les électrons :  $\vec{J}_n(\vec{r}, t)$  et  $n(\vec{r}, t)$ , et pour les trous :  $\vec{J}_p(\vec{r}, t)$  et  $p(\vec{r}, t)$ . Ce système dit de dérive-diffusion est assez simple à résoudre avec des méthodes numériques adaptées. Cette approche est utilisée dans la plupart des simulateurs commerciaux de dispositifs car elle ne nécessite qu'un temps de calcul réduit. Cependant elle ne permet pas de prendre en compte les phénomènes hors d'équilibre.

Pour remédier à cette lacune, on autorise, en plus du décalage des distributions gaussiennes, leur élargissement, comme illustré par la courbe rouge de la Figure 43. Cet élargissement de la distribution traduit un « chauffage » du gaz électronique par rapport à la température du réseau cristallin. En outre, le temps de relaxation des porteurs  $\tau$  peut dépendre de leur énergie. On obtient alors le système dit « d'énergie balance » avec approximation du temps de relaxation. Ce modèle plus précis pour étudier le transport légèrement hors d'équilibre reste cependant incomplet pour étudier le transport fortement hors d'équilibre où les distributions de porteurs ne sont plus gaussiennes.

### 1.2.2 Simulation Monte Carlo

Les simulations de transport particulaire de type Monte Carlo constituent l'outil idéal pour prendre en compte les effets hors d'équilibre dans les dispositifs pour la microélectronique. Elles consistent en des « simulations expérimentales » de problèmes mathématiques [NOUG93], dans lesquelles des nombres aléatoires sont utilisés pour trouver une solution qui, elle, n'est pas forcément aléatoire.

Ainsi, dans un simulateur de transport de type Monte Carlo (cf. Chapitre IV pour plus de précisions), on décompose la trajectoire des particules en un grand nombre de séquences aléatoires (quelques milliers ou dizaines de milliers suivant les cas). De nombreux tirages au sort sont effectués selon des lois de probabilité issues d'une fine modélisation physique des matériaux. Durant chaque séquence, un porteur suit une trajectoire déterministe, puis subit une collision. On simule ainsi un grand nombre de particules dont le mouvement d'ensemble donne les valeurs des caractéristiques macro- et microscopiques : courant, fonction de distribution, densité de charges...

Dans le cas de l'étude du transport d'une seule particule, l'EB se réduit à l'équation fondamentale de la dynamique. L'avantage de la méthode Monte Carlo, qui étudie la trajectoire d'une seule particule à la fois, est qu'elle permet de résoudre l'équation de Boltzmann sans aucune approximation. Les résultats sont d'autant plus précis que le nombre de particules simulées ainsi que le temps de simulation sont importants.

Cette méthode, quoique bien plus gourmande en temps de calculs, permet de s'affranchir de toutes les approximations précédentes et permet naturellement l'étude de tout type de transport, de

purement stationnaire jusqu'à purement balistique. Contrairement aux méthodes de résolution directe, elle est bien adaptée à l'étude des composants à deux, voire trois dimensions dans l'espace réel.

### 1.3 Modèles analytiques du transport fortement non stationnaire

Pour donner de nouvelles relations courant/tension dans les dispositifs ultracourts, des modèles analytiques, considérant le cas limite du transport purement balistique, ont été proposés pour faire le pendant aux relations décrivant un régime purement stationnaire (dérive/diffusion). Pour assurer la continuité entre ces deux cas limites, des relations ont été étendues au régime intermédiaire, appelé régime « quasi balistique ».

#### 1.3.1 Equation de Boltzmann balistique

L'équation de Boltzmann balistique à une dimension s'écrit :  $v_x \frac{\partial f(x, v_x)}{\partial x} - \frac{F_x}{m^*} \frac{\partial f(x, v_x)}{\partial v_x} = 0$ .

Cette équation est issue d'une vision « eulérienne » de l'équation de transport : on suit l'évolution de la fonction de distribution dans un élément de volume de l'espace des phases  $(\vec{r}, \vec{k})$ , comme décrit au paragraphe 1.1.2 .

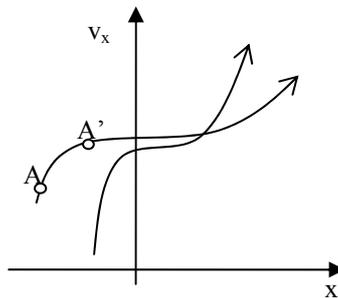


Figure 44 : Trajectoires possibles dans l'espace des phases 1D. Sans interaction, les porteurs se déplacent selon les lois de Newton sur une trajectoire unique dépendant des conditions initiales [LUND00].

Cependant, dans un matériau sans interaction, l'approche « lagrangienne » (plus « Monte Carlisle ») s'intéressant aux trajectoires des porteurs au cours du temps est plus efficace. En effet, les porteurs se trouvant à une position A' de l'espace des phases de la Figure 44 à l'instant t étaient à la position A à l'instant t-dt. Ainsi la probabilité d'occupation de l'état A' est la même que la probabilité d'occupation de l'état A un instant dt auparavant [LUND00]. On peut donc écrire en régime balistique :

$$f(x, v_x, t) = f(x - v_x \cdot dt, v_x - F/m^* \cdot dt, t - dt) \text{ (soit la dérivée totale } df(x, v_x, t)/dt = 0)$$

Appliquons cette démarche dans le canal balistique 1D d'un MOSFET. Comme schématisé en Figure 45, on considère une relation de dispersion parabolique :  $E(x, v_x) = E_{BC}(x) + \frac{1}{2} m^* v_x^2$ .

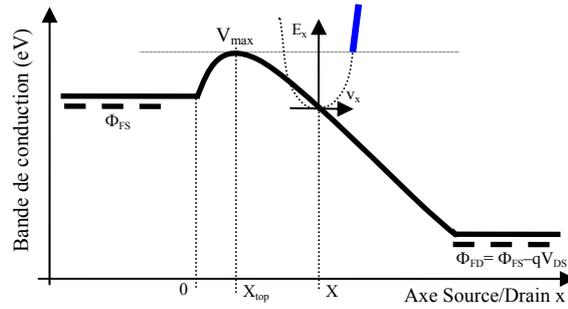


Figure 45 : Evolution typique de la bande de conduction  $E_{BC}(x)$  le long de l'axe source drain d'un transistor balistique en trait continu. Les tirets en gras symbolisent les niveaux de Fermi  $\Phi_{FS}$  et  $\Phi_{FD}$  respectivement dans les caissons de Source et de Drain. La relation de dispersion parabolique  $E_x(v_x)$  est indiquée en tirets fins. En bleu sont indiqués les états en équilibre avec la source.

Les caissons étant à l'équilibre, les fonctions de distribution  $y$  sont connues et donnent les conditions aux limites suivantes dans le cas non dégénéré (statistique de Boltzmann) :

$$f(0, v_x) = f_{\text{équilibre}}(E(x, v_x) - \Phi_{FS}) = e^{\frac{E(x, v_x) - \Phi_{FS}}{k_B \cdot T}} \quad \text{dans la source}$$

$$f(L_G, v_x) = f_{\text{équilibre}}(E(x, v_x) - \Phi_{FD}) = e^{\frac{E(x, v_x) - \Phi_{FD}}{k_B \cdot T}} \quad \text{dans le drain}$$

Dans le canal, à une position  $x > x_{\text{top}}$  quelconque après la barrière d'injection source/canal, les électrons à vitesse positive qui ont une énergie supérieure au sommet de la barrière ont tous été injectés depuis le réservoir de source. Sur la relation de dispersion  $E_x(v_x)$  à la position  $x$  de la Figure 45, ces électrons sont indiqués en gras sur la courbe de dispersion correspondante. Tous les autres électrons présents à cet endroit proviennent du réservoir de drain. Ainsi, en régime permanent, la fonction de distribution d'un porteur de vitesse  $v_x$  injecté par la source possède la même fonction de distribution qu'un porteur de la source de vitesse  $v'_x = v_x - \sqrt{2 \cdot |E_{BC}(x) - E_{BC}(0)| / m^*} \times \text{sign}(E_{BC}(x) - E_{BC}(0))$  par conservation de l'énergie totale.

En outre, seuls les porteurs de la source ayant une énergie  $E_{BC}$  supérieure à celle du sommet de la barrière  $V_{\text{max}}$  sont injectés dans le canal. Cette condition en terme de vitesse se traduit par  $v_x > v_{\text{max}} = \sqrt{2 \cdot (V_{\text{max}} - E_{BC}(x)) / m^*}$  (distribution parabolique).

On peut donc vérifier que la fonction suivante est solution pour  $x > x_{\text{top}}$  [RHEW02] :

$$f(x, v_x) = f(0, v'_x) = f_{\text{équilibre}}(E(x, v_x) - \Phi_{FS}) \quad \text{si } v_x > v_{\text{max}}$$

et, par un raisonnement similaire pour les électrons venant du drain,

$$f(x, v_x) = f(L_G, v'_x) = f_{\text{équilibre}}(E(x, v_x) - \Phi_{FD}) \quad \text{si } v_x < v_{\text{max}}.$$

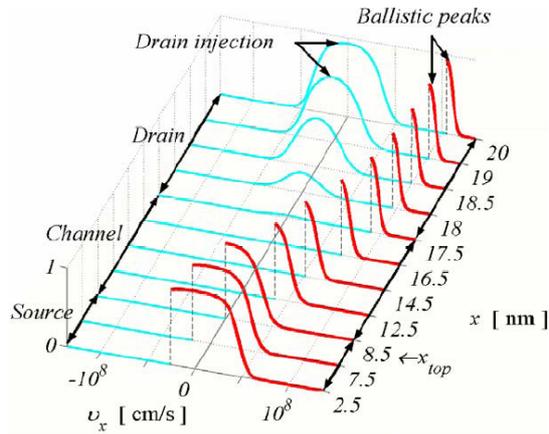


Figure 46 : Evolution des fonctions de distribution dans un canal balistique [RAHM03].

Grâce aux relations précédentes, on peut reconstituer les fonctions de distribution de la population balistique pour chaque position du canal. Ainsi construite, la Figure 46 présente l'évolution de la fonction de distribution en vitesse, en considérant la statistique de Fermi-Dirac, dans un canal balistique de porteurs ayant tous la même masse effective  $m^*$ . On observe la formation du pic d'électrons balistiques (formation décrite en détail par la suite) et la diminution du nombre de porteurs injectés par le drain à mesure que l'on s'approche du sommet de la barrière.

### 1.3.2 Modèle analytique balistique

Le paragraphe précédant a montré que les fonctions de distribution dans le canal des populations balistiques, fortement hors d'équilibre, peuvent être reconstruites par l'intermédiaire des fonctions d'occupation des porteurs dans les caissons qui sont eux à l'équilibre. Préférant utiliser les distributions à l'équilibre  $f_{\text{équilibre}}$ , les modèles analytiques du courant balistique sont généralement issus du formalisme de Landauer [NATO94][DATT95]. Dans ce formalisme, le courant de drain  $I_D$  est décomposé en une composante  $I^+$  correspondant aux porteurs injectés depuis la source vers le drain, et  $I^-$  correspondant à ceux injectés par le drain (cf. annexe page 171). On a alors :  $I_D = I^+ - I^-$ .

Comme schématisé sur la Figure 47.a, dans un canal balistique, tous les porteurs présents dans la source, à vitesse positive (direction source/drain) et possédant une énergie  $E$  supérieure à la barrière d'injection source/canal  $V_{\text{max}}$ , vont contribuer au courant  $I^+$ . De même, tous les électrons suffisamment énergétiques du drain vont contribuer à  $I^-$ .

Comme illustré sur la Figure 47.b, au sommet de la barrière, le nombre  $n_s$  d'électrons par unité de longueur dans la direction  $x$  est composé d'électrons à vitesse positive  $n_s^+$  injectés depuis la source et d'électrons à vitesse négative  $n_s^-$  injectés depuis le drain avec :  $n_s = n_s^+ + n_s^-$

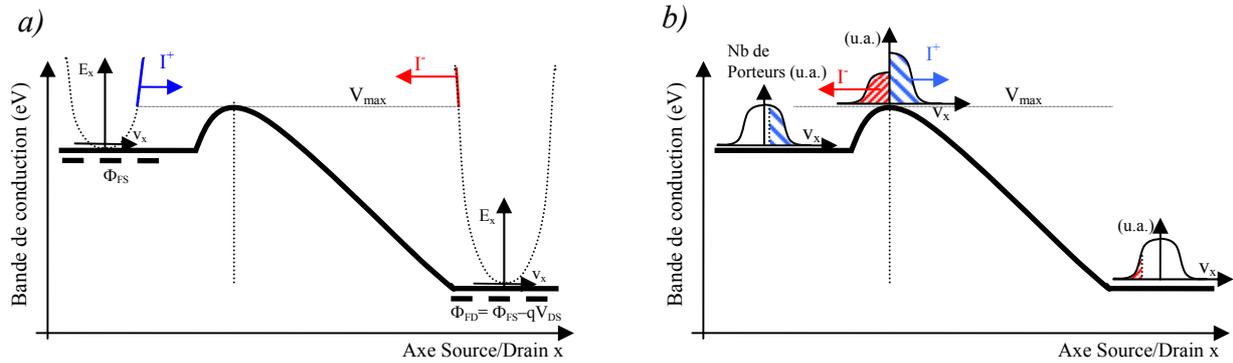


Figure 47 : Evolution typique de la bande de conduction le long de l'axe source-drain dans un transistor balistique en train continu. a) Les tirets en gras symbolisent les niveaux de Fermi  $\Phi_{FS}$  et  $\Phi_{FD}$  respectivement dans les caissons de Source et de Drain. Les relations de dispersion paraboliques  $E_x(v_x)$  sont indiquées en tirets fins. Les états de la courbe  $E_x(v_x)$  intervenant dans le calcul des courants  $\Gamma^+$  et  $\Gamma$  sont surlignés. b) Distributions en vitesse dans la source, au sommet de la barrière d'injection et dans le drain. Les composantes de chaque distribution intervenant dans le courant  $\Gamma^+$  et  $\Gamma$  sont respectivement hachurées en bleu et rouge.

Si l'on applique la loi de continuité au sommet de la barrière d'injection source/canal pour les populations d'électrons à vitesse positive et négative, il vient :

$$\Gamma^+ = q \times n_s^+ \times v_{inj}$$

$$\Gamma = q \times n_s^- \times v_{inj}$$

Avec  $v_{inj}$  vitesse d'injection moyenne des porteurs.

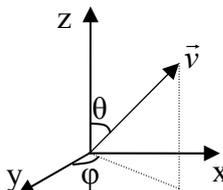
#### a) Vitesse moyenne d'injection au sommet de la barrière $v_{inj}$ :

Dans toute la suite de ce mémoire, nous allons nous restreindre à l'étude des porteurs non dégénérés car cela correspond au modèle utilisé dans nos simulations Monte Carlo. En effet, le principe d'exclusion de Pauli n'étant pas pris en compte, les populations simulées sont régies par des fonctions de distribution maxwelliennes et non par des distributions de Fermi-Dirac. Le lecteur intéressé pourra se reporter à l'annexe page 171 pour l'étude analytique du cas dégénéré en régime balistique.

Dans le cas non dégénéré (statistique de Boltzmann) la vitesse d'une population se trouve

distribuée sur une gaussienne :  $f(v) \propto \exp\left(\frac{m^* \cdot v^2}{2 \cdot k_B \cdot T}\right)$  (si bandes paraboliques ; cf. page 57)

En  $x = x_{top}$ , la vitesse moyenne des électrons à vitesse positive sur l'axe  $x$  et injectés par la source en  $x = x_{top}$  s'exprime en coordonnées sphériques comme :



$$v_{inj} = \frac{\int_0^\pi \int_0^\pi \int_0^\infty v \cdot f(v) \cdot v^2 \cdot \sin \theta \cdot d\theta \cdot d\phi \cdot dv}{\int_0^\pi \int_0^\pi \int_0^\infty f(v) \cdot v^2 \cdot \sin \theta \cdot d\theta \cdot d\phi \cdot dv} = \frac{1}{2} \frac{\int_0^\infty v \cdot f(v) \cdot v^2 \cdot dv}{\int_0^\infty f(v) \cdot v^2 \cdot dv} = \sqrt{\frac{2 \cdot k_B \cdot T}{\pi \cdot m^*}} = v_{th}$$

Avec un raisonnement similaire pour les électrons à vitesse négative, on trouve la même expression de vitesse d'injection thermique  $v_{th}$ . De plus, contrairement au cas dégénéré [NATO94],  $v_{inj}=v_{th}$  ne varie pas avec la tension de grille.

b) Charge inversion au sommet de la barrière  $n_s^-$  :

En supposant que dans un dispositif bien dimensionné (et en négligeant les effets 2D) la concentration au sommet de la barrière  $n_s$  (en électrons/m) est imposée par la tension de grille, il vient :

$q.n_s \approx W.C_{eq}(V_{GS}-V_T)$  avec  $W$  largeur du transistor,  $V_T$  la tension de seuil de transistor et  $C_{eq}$  capacité surfacique de grille équivalente (en reprenant les notations du chapitre I,  $C_{eq} \approx C_{ox}$  dans le cas d'une simple grille,  $C_{eq} \approx 2 \times C_{ox}$  dans un double grille...).

c) Calcul du rapport  $n_s^+ / n_s^-$  :

D'après la résolution de l'EB balistique, les électrons à vitesse positive et négative présents au sommet de la barrière proviennent respectivement des réservoirs source et drain. Leurs fonctions de distribution en énergie sont donc données par celles des réservoirs. Ainsi, en régime permanent, le rapport de concentration  $n_s^+ / n_s^-$  au sommet de la barrière est le même que celui des concentrations « initiales » dans les caissons :

$$\text{On obtient en effet : } \frac{n_s^-}{n_s^+} = e^{\frac{E_{\text{Drain}}(v=0) - E_{\text{Source}}(v=0)}{k_B \cdot T}} = e^{\frac{\phi_{FD} - \phi_{FS}}{k_B \cdot T}} = e^{\frac{-q \cdot V_{DS}}{k_B \cdot T}}$$

c) Caractéristique courant/tension balistique :

Finalement, les relations précédentes permettent d'écrire [LUND02b] :

$$I_D^{\text{bal}} = I^+ - I^- = v_{th} \cdot q \cdot (n_s^+ - n_s^-) = v_{th} \cdot q \cdot n_s^+ \left( 1 - \frac{n_s^-}{n_s^+} \right) = v_{th} \cdot q \cdot n_s \frac{1 - \frac{n_s^-}{n_s^+}}{1 + \frac{n_s^-}{n_s^+}} \text{ car } n_s^+ = \frac{n_s^-}{1 + \frac{n_s^-}{n_s^+}}$$

On obtient ainsi la caractéristique courant/tension balistique suivante qui ressemble à celle utilisée habituellement :

$$I_D^{\text{bal}}(V_{GS}, V_{DS}) = W \cdot v_{th} \cdot C_{eq} (V_{GS} - V_T) \frac{1 - e^{\frac{-q \cdot V_{DS}}{k_B \cdot T}}}{1 + e^{\frac{-q \cdot V_{DS}}{k_B \cdot T}}}$$

En simplifiant l'expression en fonction de la valeur de  $V_{DS}$  devant  $k_B T/q$ , on retrouve les plages de fonctionnement habituelles :

- $V_{DS} \ll k_B.T/q$ , régime linéaire : 
$$I_D^{bal}(V_{GS}, V_{DS}) = W \cdot v_{th} \cdot C_{eq} (V_{GS} - V_T) \frac{q \cdot V_{DS}}{k_B.T} \frac{1}{2}$$
- $V_{DS} \gg k_B.T/q$ , régime de saturation : 
$$I_D^{bal}(V_{GS}, V_{DS}) = W \cdot v_{th} \cdot C_{eq} (V_{GS} - V_T)$$

On remarque avec intérêt et stupeur (et tremblements...) que le courant ne dépend plus de la longueur de grille contrairement au cas stationnaire (cf. paragraphe 1.2.2 ).

### 1.3.3 Modèle quasi balistique :

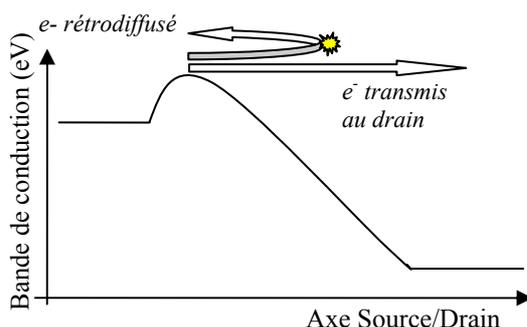


Figure 48 : Illustration de la transmission et de la rétrodiffusion des porteurs dans un canal quasi balistique à l'état passant.

Dans un canal quasi balistique, tous les électrons injectés par la source dans le canal ne sont pas transmis jusqu'au drain. Ils peuvent subir quelques interactions lors de leur traversée du canal et certains sont même rétrodiffusés dans la source, comme illustré en Figure 48. Ce type de transport est plus complexe à décrire que le régime purement balistique et certaines hypothèses plus difficiles à justifier. La relation de continuité au sommet de la barrière d'injection, en négligeant la contribution des porteurs du drain à fort  $V_{DS}$ , donne un courant de saturation :  $I_{Dsat} = q \cdot n_s \cdot v_{inj}$ .

On suppose souvent que si le transistor est bien dimensionné, l'électrostatique ne devrait pas être modifiée (en première approximation et en négligeant les effets 2D) par les interactions (du moins au sommet de la barrière). On pose alors :  $q \cdot n_s \approx W \cdot C_{eq} (V_{GS} - V_T)$ .

#### a) Coefficient de balisticité effective $B_{eff}$

L'expression de la vitesse d'injection en entrée de canal  $v_{inj}$  est modifiée par les interactions. En effet par rapport au courant du canal balistique, il faut tenir compte de la contribution des électrons qui ont subi une interaction dans le canal et qui ont été rétrodiffusés vers la source. Ainsi, même à fort  $V_{DS}$ , la population d'électrons à vitesse négative au sommet de la barrière n'est plus nulle. La vitesse d'injection est donc plus faible que dans le cas balistique  $v_{inj} = B_{eff} \times v_{th}$  avec  $0 < B_{eff} < 1$ .

En considérant au sommet de la barrière d'injection les populations d'électrons présentes, on

trouve un coefficient de "balisticité" effective  $B_{eff}$  donné par  $B_{eff} = \frac{1 - \frac{n_s^-}{n_s^+}}{1 + \frac{n_s^-}{n_s^+}} = \frac{1 - R_c}{1 + R_c}$  pour

$V_{DS} \gg k_B.T/q$  [LUND97] avec  $R_c$  coefficient de rétrodiffusion du canal.

A faible  $V_{DS}$ , il faut aussi considérer la population venant du drain dans la détermination de  $n_s^-$ . La balisticité effective s'exprime alors telle que  $B_{eff} = 1 - R_c$  pour  $V_{DS} \ll k_B.T/q$  [LUND02b].

b) Coefficient de rétrodiffusion du canal  $R_c$

Une succession d'interfaces réfléchissantes présente un coefficient de rétrodiffusion total  $R_c$  donné par [DAT95] (se reporter au calcul en annexe page 173) :

$$R_c = \frac{1}{1 + \frac{L_0}{L_{eff}}} \quad \text{avec} \quad L_{eff} : \text{longueur effective de la zone de rétrodiffusion}$$

$$\text{et} \quad L_0 = \frac{T}{(1-T)v} \quad \text{où } T \text{ est le coefficient de transmission et } v \text{ le nombre}$$

d'interfaces réfléchissantes par unité de longueur dans la zone de rétrodiffusion.

Tout d'abord, dans le cas d'un canal quasi balistique,  $L_0$  peut être assimilé au libre parcours moyen des porteurs  $l_{pm}$ . En effet, on a pour  $T$  tendant vers 1 (quasi balistique) :  $L_0(1-T)v \approx 1$  avec  $(1-T).v$  probabilité d'avoir une interaction rétrodiffusive par unité de longueur.

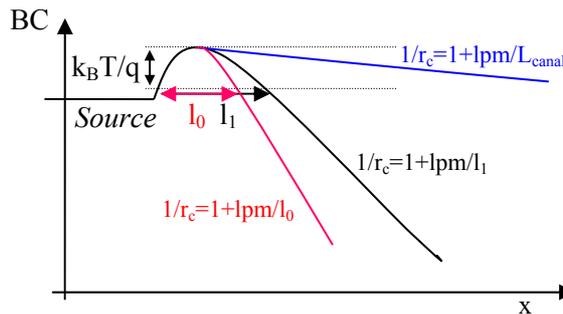


Figure 49 : Illustration de la modélisation du coefficient de rétrodiffusion dans un canal quasi balistique en fonction de la forme du potentiel après la barrière [GAUT03].

La détermination de  $L_{eff}$  est plus problématique. Dans un canal long, la longueur effective de la zone de rétrodiffusion peut être considérée égale à la longueur du canal  $L_{canal}$ . Cependant, lorsque le  $l_{pm}$  devient du même ordre de grandeur que  $L_{canal}$ , cette approximation ne semble plus être acceptable. Certains auteurs [LUND02] ont suggéré d'estimer  $L_{eff}$  par la longueur  $l_{KT}$  nécessaire à la bande de conduction pour s'abaisser d'une valeur de  $k_B.T/q$  après le sommet de la barrière, comme illustré par la Figure 49. Cette zone franchie, les électrons ont, a priori, peu de chance d'être

rétrodiffusés vers la source. Il est important de souligner que la longueur  $l_{kT}$  est fonction de  $L_{\text{canal}}$  et aussi de la polarisation de drain  $V_{DS}$ . En conjecturant que dans cette zone rétrodiffusante les porteurs sont encore proches de l'équilibre,  $L_0$  sera supposé égal en pratique au  $l_{pm}$  extrait dans un canal long.

Cependant, la détermination plus précise de cette longueur caractéristique  $L_0$  et celle du coefficient  $R_c$  font encore l'objet de nombreux travaux [FUCH05].

### c) Caractéristique courant/tension quasi balistique

Il faut noter l'importance de la balisticité effective  $B_{\text{eff}}$  qui de par sa définition  $B_{\text{eff}} = v_{\text{inj}}/v_{\text{th}}$ , exprime aussi le rapport du courant quasi balistique sur le courant balistique :  $B_{\text{eff}} = I_D / I_D^{\text{bal}}$ .

Finalement, le courant de saturation quasi balistique et la résistance à l'état passant  $R_{\text{on}}$  s'expriment de la façon suivante :

- $V_{DS} \ll k_B \cdot T/q$ ,  $R_{\text{on}} = 1 / \left. \frac{\partial I_D}{\partial V_{DS}} \right|_{V_{DS} \ll k_B T/q} = R_{\text{on}}^{\text{bal}} / B_{\text{eff}}(V_{DS})$  avec  $B_{\text{eff}}(V_{DS}) = 1 - R_c = \frac{1}{1 + \frac{L_{\text{eff}}}{l_{pm}}}$
- $V_{DS} \gg k_B \cdot T/q$ ,  $I_{D_{\text{sat}}}(V_{GS}) = B_{\text{eff}}(V_{DS}) \times I_{D_{\text{sat}}}^{\text{bal}}(V_{GS})$  avec  $B_{\text{eff}}(V_{DS}) = \frac{1 - R_c}{1 + R_c} = \frac{1}{1 + 2 \frac{L_{\text{eff}}}{l_{pm}}}$

Avec

$$L_{\text{eff}} = L_{\text{canal}} \quad \text{pour } l_{pm} \ll L_{\text{eff}}$$

$$L_{\text{eff}} = l_{kT} \quad \text{pour } l_{pm} \approx L_{\text{eff}}$$

### 1.3.4 Transition entre régime stationnaire et quasi balistique

Pour essayer de décrire la transition entre les régimes de fonctionnement quasi balistiques et stationnaires plusieurs approches ont été utilisées :

#### a) Du régime quasi balistique au régime stationnaire

Pour introduire la notion de mobilité canal long  $\mu_0$  dans les relations quasi balistiques précédentes, la relation  $v_{\text{th}} = 2 \cdot k_B T/q \times \mu_{\text{eff}}/l_{pm}$  est très pratique. Cette relation est issue de la loi de Fick  $D = v_{\text{th}} \cdot l_{pm}/2$  associée à la relation d'Einstein  $D = k_B T/q \times \mu_{\text{eff}}$  avec  $D$  coefficient de diffusion.

Cela permet d'exprimer la résistance à l'état passant  $R_{\text{on}}$  comme suit :

$$1/R_{on} = \left. \frac{\partial I_D}{\partial V_{DS}} \right|_{V_{DS} \ll k_B T/q} = \frac{W}{L_{eff}} \cdot \frac{\mu_{eff}}{1 + \frac{l_{pm}}{L_{eff}}} \cdot C_{eq} (V_{GS} - V_T)$$

Dans les cas où  $L_{eff} = L_{canal}$ , on reconnaît l'expression du modèle stationnaire pondérée (cf. chapitre I) par une fraction du rapport  $\frac{l_{pm}}{L_{eff}}$ . Si ce rapport devient important,  $R_{on}$  ne dépend plus de  $L_{eff}$  donc de  $L_{canal}$ , c'est le régime balistique.

### b) Du régime stationnaire au régime quasi balistique

Certains auteurs ont amélioré la notion de mobilité effective dans les canaux longs pour qu'elle prenne en compte des effets non stationnaires [BACC99], voire balistiques. Par analogie avec l'expression de la mobilité canal long  $\mu_0 = \frac{q \cdot l_{pm}}{m \cdot \pi \cdot v_{th}}$  et en remarquant que dans un canal balistique le libre parcours moyen des porteurs  $l_{pm}$  vaut  $L_{canal}$ , une mobilité balistique  $\mu_{bal}$  a été définie telle que :  $\mu_{bal} = \frac{q \cdot L_{canal}}{m \cdot \pi \cdot v_{th}}$  [RHEW03].

En appliquant la règle de Mathiessen pour pondérer ces mobilités, on obtient une mobilité effective  $\mu_{eff}$  telle que :  $\mu_{eff} = \frac{1}{\mu_0^{-1} + \mu_{Bal}^{-1}} = \mu_0 \frac{1}{1 + \frac{\mu_0}{\mu_{Bal}}} = \mu_0 \frac{1}{1 + \frac{l_{pm}}{L_{canal}}}$ .

On obtient ainsi deux valeurs limites :

- Si  $L_{canal} \gg l_{pm}$  alors  $\mu_{eff} = \mu_0$ . C'est le régime de mobilité canal long.
- Si  $L_{canal} \ll l_{pm}$  alors  $\mu_{eff} = \mu_{bal}$  et  $\mu_{eff}/L_{canal}$  indépendant de  $L_{canal}$ . C'est le régime balistique.

### c) Bilan

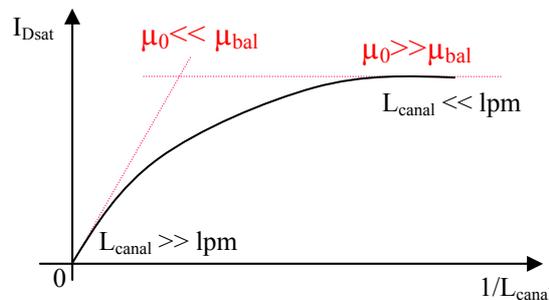


Figure 50 : Evolution du courant de saturation  $I_{Dsat}$  et de la mobilité effective  $\mu_{eff}$  en fonction de l'inverse de la longueur de canal. Dans les canaux courts ( $L_{canal} \ll l_{pm}$ ) le courant de saturation ne dépend plus de la longueur de grille [RHEW03] et  $\mu_0 \gg \mu_{bal}$ .

Des approches différentes ont mis en évidence l'importance du rapport  $\frac{l_{pm}}{L_{canal}}$  pour décrire la transition entre les régimes stationnaires et balistiques en passant par le régime quasi balistique. L'évolution du courant de saturation en fonction des longueurs de canal  $L_{canal}$  est illustrée en Figure 50. On constate ainsi que dans un canal ultracourt, le gain en courant attendu de la miniaturisation ne dépend plus directement de la longueur de grille comme dans les canaux longs (où  $I_D^{sat} \propto 1/L_{canal}$ ) mais seulement à travers le rapport  $L_{eff}(L_{canal})/l_{pm}$ . De ce fait, le gain en courant ne va faire que diminuer et s'annulera même pour les dispositifs ultimes, où le régime balistique pur sera éventuellement atteint. Le gain en courant que l'on peut espérer du transport balistique dans un canal ultracourt est quantifié par le paramètre  $1-B_{eff}$  avec  $B_{eff}$  balisticité effective qui vaut 1 en régime purement balistique.

De plus, en régime quasi balistique ou balistique la notion de vitesse d'injection des porteurs  $B_{eff} \times v_{th}$  est centrale, comme l'est la notion de mobilité dans l'étude des canaux longs. Pour augmenter  $B_{eff}$ , il faut réduire le rapport  $L_{eff}(L_{canal})/l_{pm}$ , soit en augmentant le  $l_{pm}$  en jouant sur la mobilité, soit en réduisant les dimensions géométriques. Pour augmenter la vitesse thermique  $v_{th}$ , on peut réduire la masse effective des porteurs en utilisant, par exemple, des matériaux contraints.

## 2. Mise en évidence du transport non stationnaire

### 2.1 Dispositifs simulés

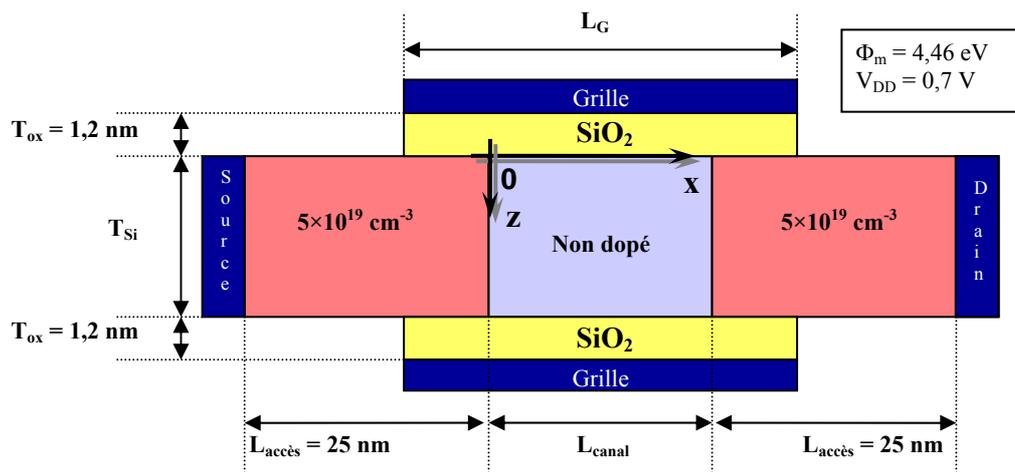


Figure 51 : Géométrie des transistors MOSFET à double grille simulés.

Nous allons étudier le transport électronique grâce à des simulations effectuées à l'aide du logiciel MONACO classique. Les dispositifs étudiés dans cette partie sont des transistors à double grille (DGMOS) dont le schéma ainsi que les dimensions géométriques répondent aux

spécifications de l'ITRS 2001 pour l'année 2007. Ils sont présentés en Figure 51. Les dispositifs ont des longueurs métallurgiques de canal  $L_{\text{canal}}$  égales respectivement à 15, 25, 50, 100 et 200 nm. Les zones actives (canal) ont une épaisseur  $T_{\text{Si}}$  pouvant valoir soit 5, soit 10 nm. Tous les oxydes de grille ont une épaisseur  $T_{\text{ox}}$  égale à 1,2 nm. Le canal a un dopage intrinsèque de  $2 \times 10^{15}$  atomes/cm<sup>-3</sup>. Les accès Source/Drain, de longueur  $L_{\text{accès}}$  valant, sauf indication contraire, 25 nm, sont dopés  $N^+$  à  $5 \times 10^{19}$  atomes/cm<sup>-3</sup>. Les jonctions PN sont supposées abruptes. Le travail de sortie du métal de grille  $\Phi_m$  est fixé à 4,46 eV pour viser une tension de seuil  $V_T$  de 0,2 V [CHAN00]. La tension d'alimentation  $V_{\text{DD}}$  est de 0,7 V.

## 2.2 Dépassement de la vitesse de saturation stationnaire

Pour déterminer la nature du transport, stationnaire ou quasi balistique, dans ces dispositifs de tailles différentes allant de 15 à 200 nm, nous nous sommes tout d'abord intéressés à l'évolution de la vitesse des porteurs et de la bande de conduction.

La Figure 52 présente l'évolution à l'état passant ( $V_{\text{GS}} = V_{\text{DS}} = 0,7$  V) de la vitesse le long de l'axe source-drain (axe x) pour trois dispositifs simulés :  $L_{\text{canal}} = 15, 50$  et 200 nm. L'axe x a pour origine le début du canal. La Figure 53 présente pour ces mêmes transistors l'évolution de la bande de conduction correspondante. On remarque un « effet 2D » illustré en Figure 53.b pour le 15 nm : l'évolution de la bande de conduction n'est pas la même à toutes les profondeurs z. Les électrons étant très majoritairement situés près de l'oxyde à fort  $V_{\text{GS}}$ , nous nous sommes concentrés sur cette profondeur. Dans le dispositif le plus long ( $L_{\text{canal}} = 200$  nm), on distingue une zone comprise entre  $x = 40$  nm et  $x = 150$  nm dans laquelle la bande de conduction et la vitesse moyenne des porteurs évoluent peu avec x. De plus, la vitesse reste bien inférieure à la vitesse de saturation stationnaire qui vaut environ  $10^7$  cm/s. Dans cette zone de faible champ électrique (inférieur à 15 kV/cm), on peut observer sur l'insert de la Figure 53.a qu'il existe un équilibre entre la vitesse  $v_x$  et le champ électrique  $E_x$  : la vitesse évolue de façon à peu près linéaire avec x. Nous sommes en présence d'un régime de mobilité. La pente moyenne de  $v_x(E_x)$  ( $320 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ ) correspond à une valeur de mobilité tout à fait « raisonnable » dans un canal de MOS.

Pour le dispositif le plus court ( $L_{\text{canal}} = 15$  nm), ce régime de mobilité n'existe manifestement pas. La vitesse s'accroît très vite et cela dès l'entrée du canal. Ainsi, dès le début du canal à  $x = 2$  nm pour  $L_{\text{canal}} = 15$  nm, la vitesse des porteurs devient supérieure à la vitesse de saturation stationnaire dans le silicium massif.

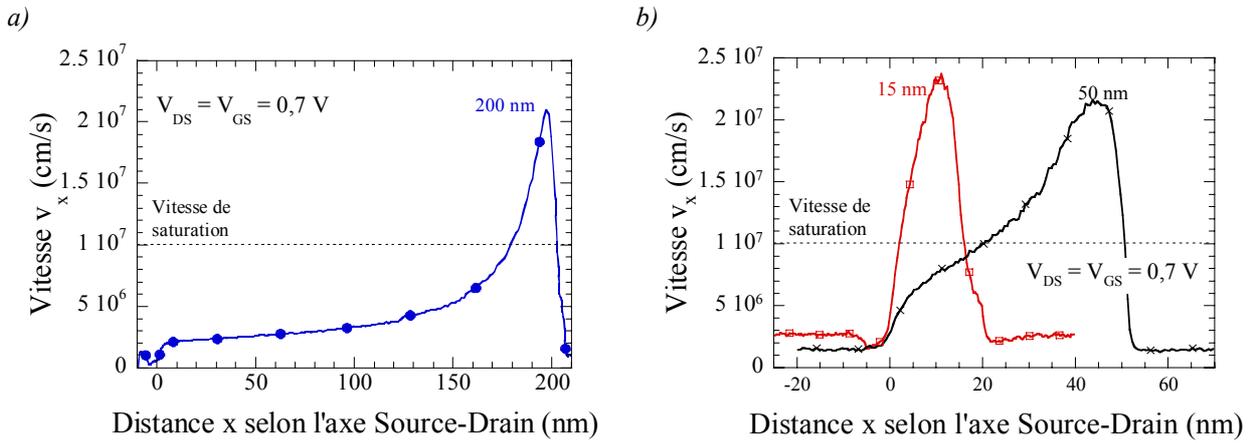


Figure 52 : Evolution de la vitesse moyenne des porteurs le long de l'axe source drain dans des structures DG-MOS à  $V_{GS} = V_{DS} = 0,7 V$ . En tirets, vitesse de saturation stationnaire dans le silicium massif. a)  $L_{canal} = 200 nm$  ;  $T_{Si} = 5 nm$ . b)  $L_{canal} = 15$  (carrés rouges) et  $50 nm$  (croix noires) ;  $T_{Si} = 10 nm$ .

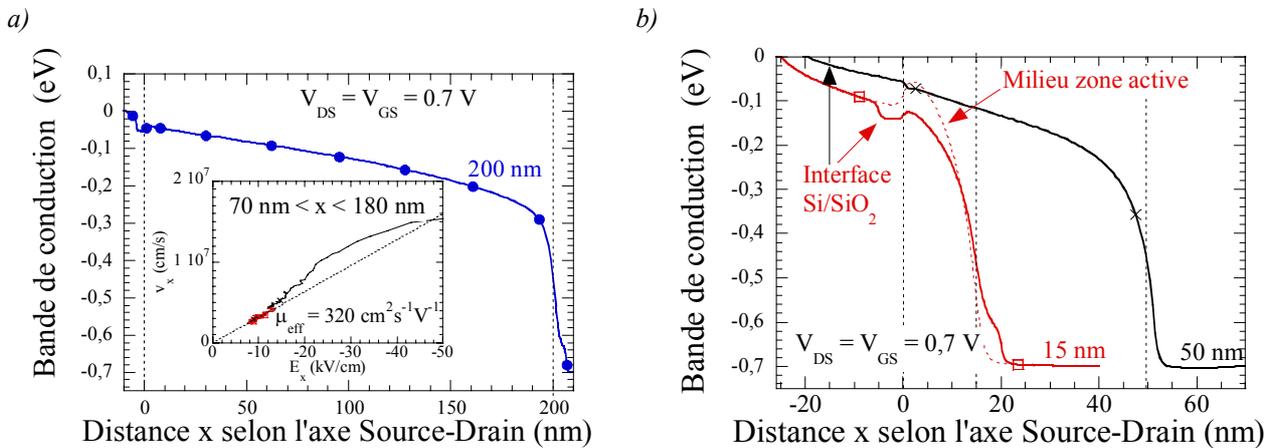


Figure 53 : Evolution en trait continu de la bande de conduction dans des structures DG-MOS à  $V_{GS} = V_{DS} = 0,7 V$  à une profondeur  $z=0,5 nm$  sous un des oxydes de grille. a)  $L_{canal} = 200 nm$  ;  $T_{Si} = 5 nm$ . Insert : évolution de la vitesse  $v_x$  des porteurs suivant  $x$  en fonction de la composante du champ électrique dans cette direction  $E_x$ . b)  $L_{canal} = 15$  (carrés rouges) et  $50 nm$  (croix noires) ;  $T_{Si} = 10 nm$ . En tirets rouges l'évolution au milieu du canal  $z = 5 nm$  pour  $L_{canal} = 15 nm$ .

L'évolution de la vitesse le long du canal de longueur  $L_{canal} = 50 nm$  illustre un régime intermédiaire. On observe un pseudo régime de mobilité entre  $x = 15$  et  $30 nm$  où le champ est relativement faible et l'évolution de la vitesse présente un léger plateau. Cependant, la vitesse est rapidement proche de la vitesse stationnaire. Cela révèle le début du régime quasi balistique, décrit

dans la partie précédente. En conséquence, le rapport  $\frac{l_{pm}}{L_{canal}}$  n'est plus négligeable pour

$L_{canal} = 50 nm$  dans un canal non dopé.

## 2.3 Spectres en vitesse et comptage d'interactions

### 2.3.1 Principe

L'étude des spectres en vitesse issus des simulations Monte Carlo permet d'examiner plus en détail le transport électronique à de telles échelles. Un spectre en vitesse présente la distribution en vitesse d'une population d'électrons. La population considérée ici est l'ensemble des électrons présents dans le volume de plusieurs mailles accolé à une surface transverse au transport, comme par exemple le volume hachuré en Figure 54.

Afin d'étudier plus finement les spectres en vitesse, on a « marqué » les électrons qui sont injectés dans la zone active depuis la source, tel qu'illustré en Figure 54 (les électrons venant du drain ne sont évidemment pas pris en compte). On peut ainsi sélectionner les populations d'électrons en fonction de leur provenance et calculer des spectres en ne comptabilisant que les électrons venant de la source. Par différence, on peut en déduire les spectres des porteurs provenant du drain.

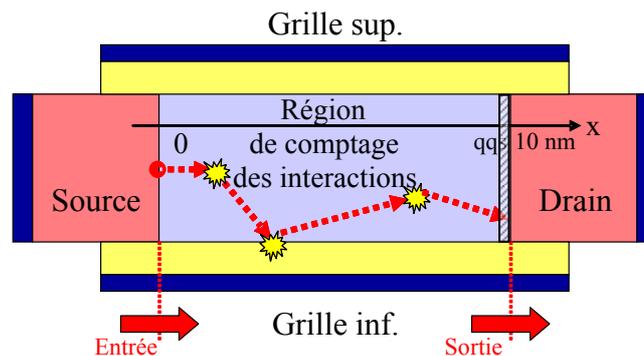


Figure 54 : Principe du marquage des électrons entre la zone d'entrée et de sortie. Ainsi le comptage des interactions subies par les porteurs est possible de même que la différenciation de leur provenance : la source ou le drain. De plus une zone de calcul de spectre en fin de canal est illustrée (zone hachurée).

De plus, par cette méthode, le comptage des interactions subies par les porteurs « marqués » entre la zone d'entrée et de sortie est possible. Nous pouvons ainsi classer les porteurs en fonction du nombre total d'interactions (phonon, rugosité...) qu'ils ont subies lors de leur traversée du canal.

### 2.3.2 Spectres de vitesse en fin de canal et pics balistique

Pour le dispositif de 15 nm, les spectres en vitesse, en fin de canal, des porteurs venant de la source et aussi du drain ont été tracés en Figure 55. Il est à noter qu'on ne considère ici que les électrons proches de l'oxyde :  $0 < z < 1$  nm.

Sur cette figure, on constate que le spectre des électrons injectés depuis le drain dans la zone active est composé de 2 quasi-gaussiennes d'écart types différents. En différenciant dans les résultats les électrons en fonction de leur vallée, nous avons pu vérifier que la population

d'électrons à masse transverse correspond bien à la gaussienne de plus grand écart type et ceux à masse longitudinale à la gaussienne de plus petit écart type. En effet dans la zone active très faiblement dopée, les électrons ne subissent pas assez d'interactions (phonons intervalles...) pour qu'ils changent de masse et que ces 2 populations se confondent au niveau de leur fonction de distribution comme dans les caissons.

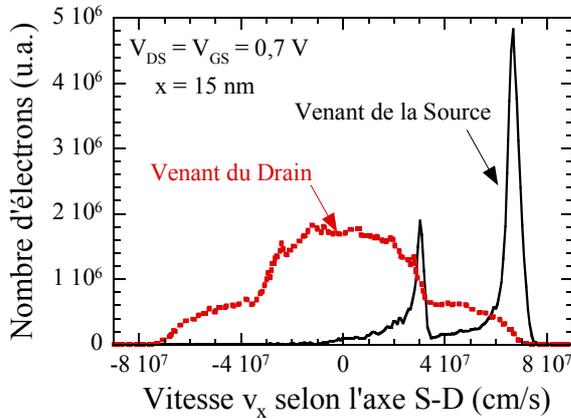


Figure 55 : Spectres en vitesse  $v_x$  calculés en fin de canal ( $x = 15 \text{ nm}$  et  $0 < z < 1 \text{ nm}$  sous l'oxyde), dans le DGMOS avec  $L_{\text{canal}} = 15 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$ . La distribution des électrons ayant été injectés dans le canal par la source est tracée en trait continu et pour ceux injectés par le drain en tirets.  $V_{\text{GS}} = V_{\text{DS}} = 0,7 \text{ V}$ .

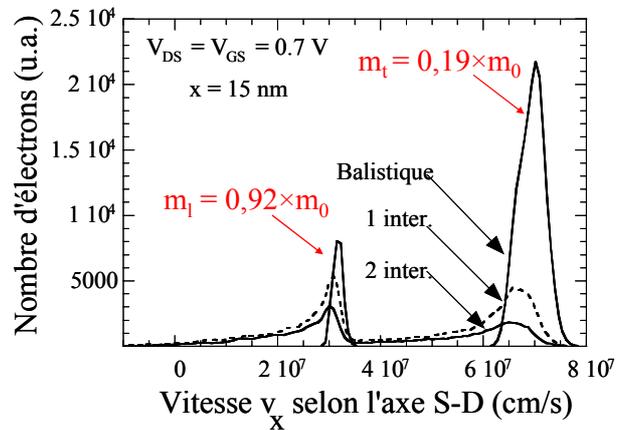


Figure 56 : Spectres en vitesse  $v_x$  calculés en fin de canal ( $x = 15 \text{ nm}$  et  $0 < z < T_{\text{Si}}$ ) dans le DGMOS avec  $L_{\text{canal}} = 15 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$  pour les électrons injectés depuis la source ayant subi 0, 1 et 2 interactions durant leur traversée du canal.  $V_{\text{GS}} = V_{\text{DS}} = 0,7 \text{ V}$ .

La forme est très différente pour la distribution en vitesse des électrons injectés depuis la source et qui ont traversé tout le canal. Elle est constituée de deux pics. La Figure 56 représente le même type de spectre mais relevé sur toute l'épaisseur du film :  $0 < z < T_{\text{Si}}$  (cela élargit les pics) et, surtout, en distinguant les électrons ayant subi 0, 1 ou 2 interactions. Elle montre que les pics très marqués sont dus aux électrons qui traversent le canal sans subir d'interactions. Le premier pic ( $v = 2,9 \times 10^7 \text{ cm/s}$ ) correspond aux électrons balistiques de masse longitudinale et le second ( $v = 6,55 \times 10^7 \text{ cm/s}$ ) à ceux de masse transverse. On retrouve en effet ces vitesses en appliquant trivialement la conservation de l'énergie dans le canal (cf. bande de conduction en Figure 53) et en supposant une vitesse d'injection nulle. A gauche des pics, vers les plus faibles vitesses, on observe une petite queue de distribution qui correspond aux électrons qui ont été ralentis par des interactions.

### 2.3.3 Formation des pics de balistique

La Figure 57 permet de suivre l'évolution des spectres en vitesse des électrons venant de la source le long du canal selon l'axe source-drain (axe  $x$ ), et donc la formation des pics de balistiques observés précédemment en fin de canal ( $x = 15 \text{ nm}$ ).

On part d'une distribution gaussienne des électrons quasi-thermiques dans la source. Puis vers  $x = 1$  nm, la courbe se déforme très fortement et il ne reste quasiment plus que des électrons à vitesse positive car on s'approche du sommet de la barrière de potentiel en entrée de canal (cf. Figure 53). La Figure 58 décompose cette distribution en fonction du nombre d'interactions subies par les porteurs. La distribution des électrons balistiques est une demi-gaussienne et les électrons à vitesse négative provenant de la source ont été rétrodiffusés par le canal.

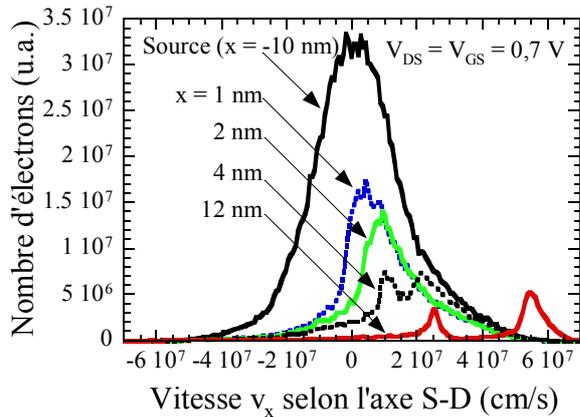


Figure 57 : Evolution des spectres en vitesse des électrons venant de la source le long du canal sous un oxyde de grille du DGMOS avec  $L_{\text{canal}} = 15$  nm et  $T_{\text{Si}} = 10$  nm à  $V_{\text{GS}} = V_{\text{DS}} = 0,7$  V

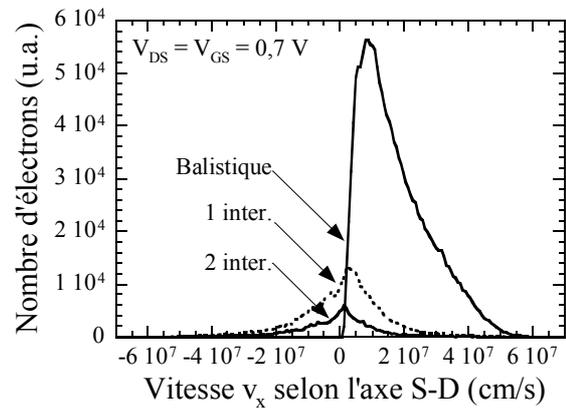


Figure 58 : Spectres en vitesse  $v_x$  calculés juste après le sommet de la barrière dans le DGMOS avec  $L_{\text{canal}} = 15$  nm et  $T_{\text{Si}} = 10$  nm pour les électrons injectés depuis la source ayant subi 0, 1 et 2 interactions durant leur traversée du canal.  $V_{\text{GS}} = V_{\text{DS}} = 0,7$  V.

Entre 1 nm et 2 nm, le pic du spectre de la Figure 57 se déplace vers les vitesses positives, la moyenne de la demi-gaussienne augmente, en clair les électrons accélèrent.

Pour  $y = 4$  nm, on distingue la population des électrons à masse longitudinale (« mamelon » de plus faible vitesse) de celle à masse transverse («mamelon» de plus de grande vitesse). Enfin à 12 nm, les pics se sont propagés et affinés (flux conservatif), chacun à sa vitesse propre.

En comparant ces spectres de vitesse avec ceux des électrons injectés depuis le drain présentés en Figure 59, on constate que le pourcentage d'électrons venant du drain (par rapport à la population totale d'électrons présents dans la tranche), important en fin de canal (72% pour  $x = 15$  nm), devient rapidement négligeable dès qu'on s'en éloigne et vaut 42%, 6% et 2% pour respectivement  $x = 12, 4$  et 1 nm.

Ces évolutions des spectres en vitesse sont semblables à celles obtenues dans un canal purement balistique déjà présentées en Figure 46. Cependant, les deux populations d'électrons à masses transverses et longitudinales présentes dans le silicium sont considérées dans notre étude. Contrairement au cas purement balistique, les pics en fin de canal ont une queue de distribution non nulle. De plus au sommet de la barrière, la population d'électrons à vitesse négative est due aux électrons injectés par la source et rétrodiffusés par le canal. Très peu proviennent du drain, comme

supposé précédemment en 1.3.3. Tout cela confirme que pour  $L_{\text{canal}} = 15 \text{ nm}$  dans un canal non dopé, le transport est quasi balistique comme décrit au paragraphe 1.3 .

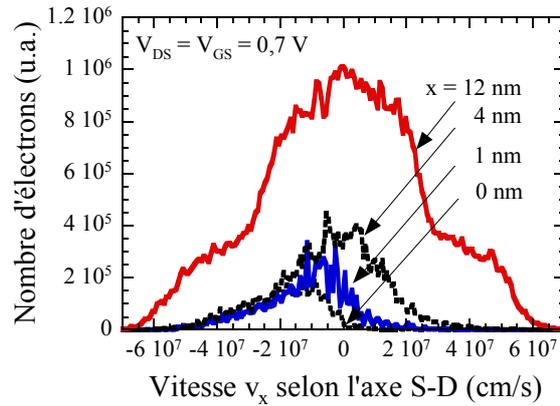


Figure 59 : Evolution des spectres en vitesse des électrons venant du drain le long du canal sous un oxyde de grille du DGMOS avec  $L_{\text{canal}} = 15 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$  à  $V_{\text{GS}} = V_{\text{DS}} = 0,7 \text{ V}$

### 2.3.4 Au niveau du drain

Sur la Figure 53.b présentant l'évolution de la bande de conduction dans le transistor de 15 nm, on peut constater que la chute de potentiel dans le drain est bien moindre que celle encaissée par la source. Cela signifie que la résistance d'accès au caisson de drain  $R_D$  est faible devant celle de la source  $R_S$ .

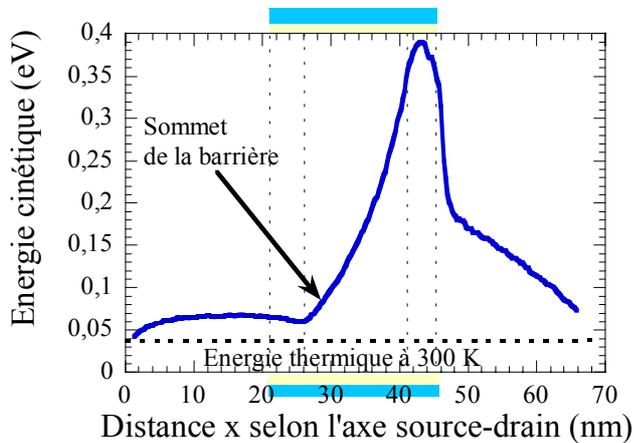


Figure 60 : Coupe longitudinale représentant l'évolution de l'énergie cinétique des porteurs, sous l'oxyde, dans la structure DGMOS avec  $L_{\text{canal}}=15 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$ .  $V_{\text{GS}}=0,7 \text{ V}$  et  $V_{\text{DS}}=0,7 \text{ V}$ .

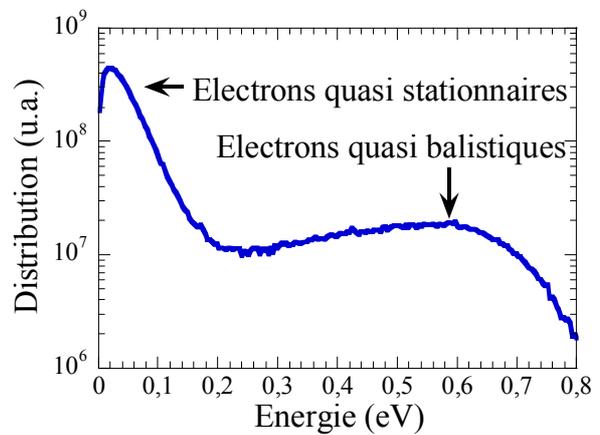


Figure 61 : Spectres en énergie cinétique au milieu du drain dans la structure DGMOS avec  $L_{\text{canal}}=15 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$ .  $V_{\text{GS}}=0,7 \text{ V}$  et  $V_{\text{DS}}=0,7 \text{ V}$ .

La Figure 60 indique l'évolution de la valeur moyenne de l'énergie cinétique des électrons le long du dispositif (axe  $x$ ). On vérifie bien que les électrons sont quasiment thermiques dans le caisson de source ( $E_{\text{thermique}} = 3kT/2 = 38,8 \text{ meV}$  à 300 K) et que cette énergie moyenne augmente fortement dans le canal. Dans le drain, par contre, les porteurs sont très « chauds ». Au milieu du

drain, les électrons ont encore une énergie moyenne de 130 meV et, même en fin de drain, ils ne sont pas complètement thermalisés.

Pour comprendre ce phénomène, nous avons tracé en Figure 61 le spectre en énergie des porteurs (calculé selon la même méthode qu'un spectre en vitesse) au milieu de la zone  $N^+$  de drain. On découvre qu'il reste dans le drain un nombre non négligeable d'électrons n'ayant subi qu'un faible nombre d'interactions à la traversée du canal (électrons quasi balistiques) et qui ont donc une grande énergie cinétique :  $0,2 < E < 0,6$  eV. Ces électrons font sensiblement augmenter la moyenne de l'énergie cinétique observée en Figure 60. De plus, une très grosse part de ces électrons participe au transport du courant dans le drain.

Dans le courant total, la proportion d'électrons qui subissent un grand nombre d'interactions (électrons quasi-stationnaires) et qui assurent le caractère résistif du drain est donc diminuée. Cela doit expliquer la faible valeur de  $R_D$  obtenue [MOUI03]. Cependant, la prise en compte de l'interaction électron/électron, plus particulièrement dans sa forme électron/plasmon [FISC91] [PALE05], permettrait peut être d'accélérer la thermalisation des électrons dans le drain.

Dans ce paragraphe, on a montré que, dans des dispositifs à canal non dopé de longueur inférieure à 50 nm, le transport est très fortement hors d'équilibre : dépassement de la vitesse de saturation stationnaire dès l'entrée du canal, apparition de pics balistiques sur les spectres en vitesse... Le transport dans des dispositifs encore plus courts semble même quasi balistique.

### 3. Evolution de la balisticité

#### 3.1 Définitions de $B_{int}$ et $B_{eff}$

Pour étudier le transport dans les dispositifs ultracourts, on utilise deux paramètres pour quantifier la « balisticité » du dispositif. Le premier paramètre appelé « balisticité effective »  $B_{eff}$  a déjà été rencontré dans la présentation des modèles analytiques quasi balistiques du paragraphe 1.3.3 . Il correspond au rapport de courants  $B_{eff} = I_{on}/I_{on\_bal}$ .  $I_{on\_bal}$  est le courant à l'état passant dans un transistor virtuel parfaitement identique mais ayant un canal balistique. Pour simuler un tel canal, aucun phénomène d'interactions n'est pris en compte dans la gestion du transport des porteurs qui suivent ainsi des trajectoires purement balistiques.

Pour décrire à un niveau plus microscopique le transport, on s'est aussi intéressé à un paramètre appelé « balisticité intrinsèque »  $B_{int}$ .  $B_{int}$  correspond au pourcentage d'électrons parfaitement balistiques en fin de canal, c'est-à-dire n'ayant subi aucune interaction lors de la traversée du canal depuis l'injection à travers la surface d'entrée à la jonction source-canal jusqu'à la collection par la surface de sortie au niveau de la jonction canal-drain. Il est important de noter que  $B_{int}$  est donc un

pourcentage en terme de flux de porteurs, c'est-à-dire de densité électronique pondérée par la vitesse.

## 3.2 Paramètres physiques

### 3.2.1 Balisticité effective et longueur de canal

On a tracé sur la Figure 62 l'évolution du courant de saturation  $I_{on}$  en fonction de la longueur de canal  $L_{canal}$  pour des DGMOS standard et balistiques. On constate tout d'abord que l'augmentation du courant en  $1/L_{canal}$  ( $W$ ,  $\mu_{eff}$ ,  $V_{DD}$  identiques) prévue par les théories stationnaires, indiquée en tirets noirs, ne donne pas de résultats satisfaisants pour ces dispositifs courts.

En accord avec les théories balistiques, le courant dans un transistor balistique n'évolue pas avec la longueur du canal. L'augmentation du courant pour  $L_{canal} < 15$  nm n'est due qu'à des courants de fuite causés par des effets de canal court dans ces dispositifs moins bien dimensionnés (comme précisé dans le chapitre suivant, le rapport  $L_{canal}/T_{Si}$  commence à être trop faible).

La balisticité effective  $B_{eff}$ , rapport des courants des canaux standard et balistiques, est tracée en Figure 63. L'évolution de  $B_{eff}$  révèle un changement de pente autour de 50 nm qui traduit la transition entre le régime stationnaire et quasi balistique.

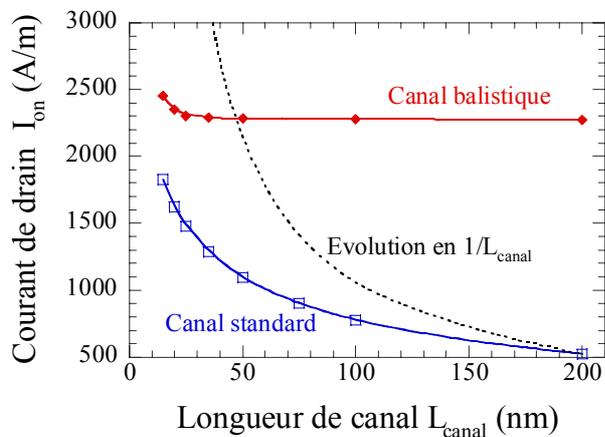


Figure 62 : Evolution du courant  $I_{on}$  en fonction de  $L_{canal} = 15, 25, 50, 100$  nm dans des DGMOS standard et balistiques avec  $T_{Si} = 5$  nm. En tirets : évolution du courant en  $1/L_{canal}$  à partir du courant  $I_{on}$  obtenu pour  $L_{canal} = 200$  nm.

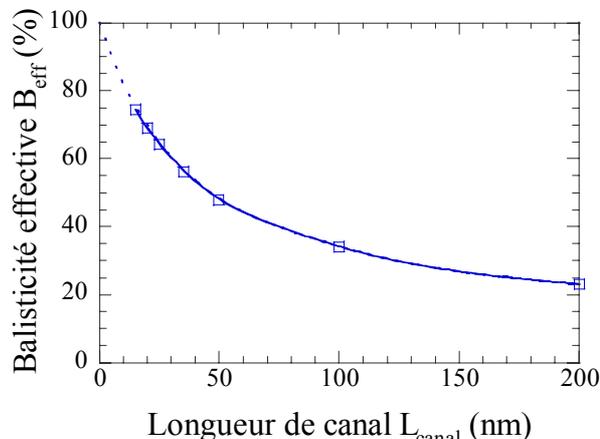


Figure 63 : Evolution de la balisticité effective  $B_{eff}$  en fonction de  $L_{canal}$  dans des structures DGMOS avec  $T_{Si} = 5$  nm.

On a  $B_{eff} = 25\%$  dans le transistor de 200 nm de long, le courant maximum possible est donc 4 fois plus grand. En outre, selon ces simulations, les dispositifs ultimes de 15 nm de long devraient flirter avec le fonctionnement idéal et délivrer 75% du courant balistique. A ces dimensions, l'objectif principal de la miniaturisation ne devrait donc plus être le gain en courant.

### 3.2.2 Balisticité intrinsèque et longueur de canal

Pour mettre clairement en évidence la transition entre les régimes de transport quasi stationnaire et quasi balistique évoquée précédemment, on a représenté en Figure 64 les spectres d'interactions pour des DGMOS de longueur de canal  $L_{\text{canal}}$  variant de 15 à 200 nm. Un spectre d'interactions représente la population d'électrons en fin de canal en fonction du nombre d'interactions  $N_{\text{int}}$  subies lors de la traversée du canal, depuis la jonction source/canal jusqu'à la jonction canal/drain.

Pour le transistor le plus long  $L_{\text{canal}} = 200$  nm, le maximum du spectre (en forme de cloche) est centré sur 26 interactions. Le pourcentage d'électrons balistiques en fin de canal est infime, inférieur à 0,1%. Cette forme est caractéristique d'un transport quasi stationnaire avec un libre parcours moyen qui vaut environ  $l_{\text{pm}} \approx 200/26 \approx 8$  nm.

Pour les transistors plus courts, la décroissance exponentielle du spectre, quand  $N_{\text{int}}$  augmente au-delà du maximum de la courbe, s'accélère. De plus, à mesure que l'on réduit les dimensions, le maximum du spectre se décale vers les nombres d'interactions plus faibles et, pour  $L_{\text{canal}} = 50$  nm, le maximum correspond à la population d'électrons balistiques. Pour  $L_{\text{canal}} = 25$  et 15 nm, on a des spectres d'interactions qui ne font que décroître exponentiellement quand  $N_{\text{int}}$  augmente, cette forme est la signature typique d'un transport quasi balistique. La transition entre ces deux régimes de transport apparaît pour des longueurs de canaux d'environ 50 nm ; le rapport  $l_{\text{pm}}/L_{\text{canal}}$ , qui apparaît de façon récurrente dans toutes les formules analytiques présentées au paragraphe 1.3.2 , vaut 0,16 et est en effet non négligeable devant l'unité.

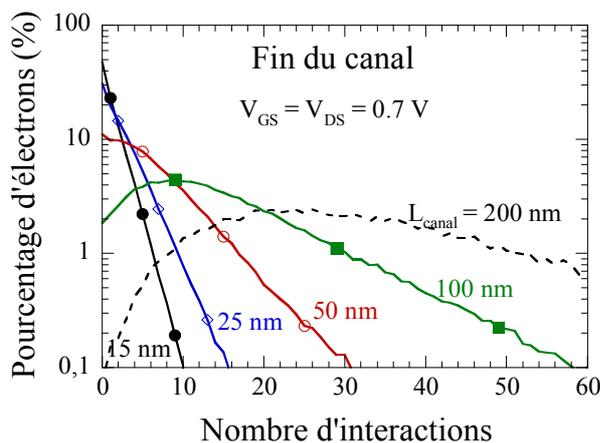


Figure 64 : Spectres d'interactions dans des DGMOS tels que  $L_{\text{canal}} = 15, 25, 50, 100$  et  $200$  nm avec  $T_{\text{Si}} = 5$  nm, soit le pourcentage de porteurs en fonction du nombre d'interactions qu'ils ont subies pendant la traversée du canal de la source jusqu'au drain.

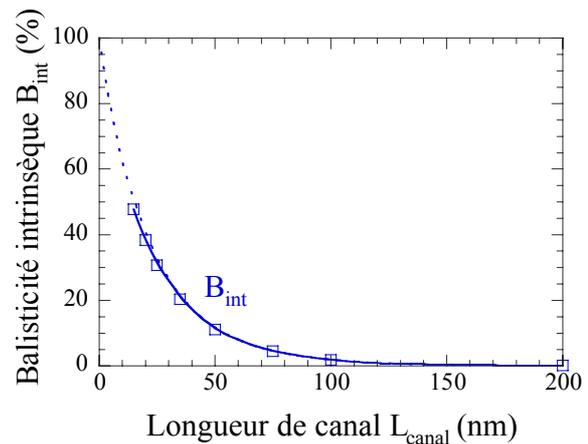


Figure 65 : Evolution de la balisticité intrinsèque  $B_{\text{int}}$  en fonction de  $L_{\text{canal}}$  dans des structures DGMOS telles que  $T_{\text{Si}} = 5$  nm.  $V_{\text{GS}} = 0,7$  V et  $V_{\text{DS}} = 0,7$  V.

Pour extraire un paramètre unique de l'étude des spectres en vitesse, nous nous sommes focalisés sur le pourcentage de la population balistique, c'est-à-dire la balisticité intrinsèque  $B_{\text{int}}$ . De même que pour  $B_{\text{eff}}(L_{\text{canal}})$ , une transition apparaît autour de 50 nm dans l'évolution  $B_{\text{int}}(L_{\text{canal}})$  tracée en

Figure 65, pour  $B_{\text{int}} \approx 10\%$ . Cette balisticité apparaît être un paramètre très pertinent pour décrire le régime de transport ce qui est en accord avec des résultats de Valérie Aubry-Fortuna dans le cas de dispositifs contraints massifs [STMA04].

### 3.3 Relation entre les balisticités effective et intrinsèque

Pour analyser la relation entre le courant et le régime de transport, l'évolution de la balisticité effective  $B_{\text{eff}}$  (et donc du courant) en fonction de la balisticité intrinsèque  $B_{\text{int}}$  a été représentée sur la Figure 66 pour chaque longueur de Double Grille simulée. Pour de faibles valeurs de balisticité ( $B_{\text{int}} < 10\%$ ), on s'aperçoit que la pente de la courbe  $B_{\text{int}}(B_{\text{eff}})$  est importante : une faible augmentation du nombre d'électrons balistiques en fin de canal induit une forte augmentation du courant de drain. Pour  $B_{\text{int}}$  croissant, cette pente diminue : l'impact de  $B_{\text{int}}$  sur  $B_{\text{eff}}$  est donc plus faible lorsque le transport est déjà sensiblement balistique ( $B_{\text{int}} > 10\%$ ). La relation entre  $B_{\text{int}}$  et  $B_{\text{eff}}$  tend alors à être linéaire. On constate, une fois de plus, que la transition entre les régimes de transport se situe autour de  $L_{\text{canal}} = 50$  nm.

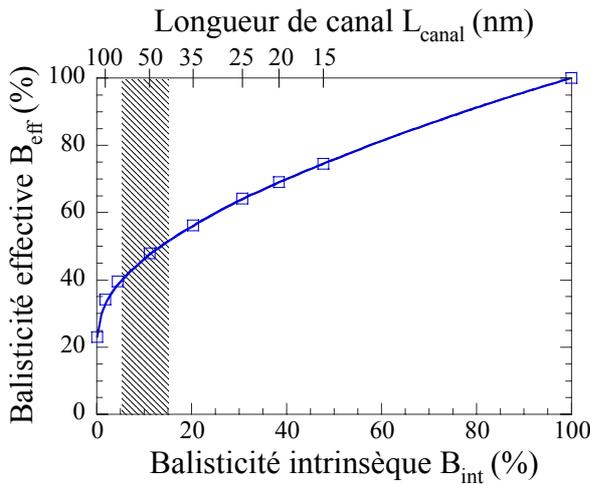


Figure 66 : Evolution de la balisticité effective  $B_{\text{eff}}$  en fonction de balisticité intrinsèque  $B_{\text{int}}$  dans des structures DG MOS avec  $L_{\text{canal}}$  comme paramètre et  $T_{\text{Si}} = 5$  nm.  $V_{\text{GS}} = 0,7$  V et  $V_{\text{DS}} = 0,7$  V.

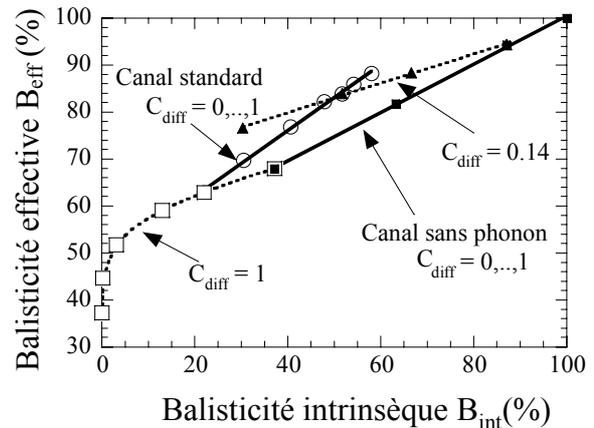


Figure 67 : Evolution de la balisticité effective  $B_{\text{eff}}$  en fonction de balisticité intrinsèque  $B_{\text{int}}$  dans une structure DG MOS avec  $L_{\text{canal}} = 15$  nm et  $T_{\text{Si}} = 10$  nm. L'intensité des interactions avec l'interface rugueuse  $C_{\text{diff}}$  ( $=0,14$  dans un canal standard) et celle avec les phonons  $K_{\text{ph}}$  ( $=1$  dans un canal standard) sont artificiellement modifiées pour que  $B_{\text{int}}$  et  $B_{\text{eff}}$  évoluent.  $V_{\text{GS}} = 0,7$  V et  $V_{\text{DS}} = 0,7$  V.

Malgré l'apparence trompeuse de la Figure 66, la courbe  $B_{\text{int}}(B_{\text{eff}})$  n'est pas universelle. En effet, l'impact d'une interaction sur le courant dépend de sa nature : isotropie, élasticité... La Figure 67 indique l'évolution des paramètres  $B_{\text{int}}$  et  $B_{\text{eff}}$  dans un DG MOS de 15 nm dans lequel on a modifié artificiellement les fréquences d'interaction utilisées dans les simulations. Les fréquences d'interaction électrons/phonons ont été multipliées par un facteur  $K_{\text{ph}}$  variant entre 0 et 20. Pour

$K_{ph} = 1$ , on retrouve les fréquences d'interaction standard. De plus, nous avons fait varier la proportion  $C_{diff}^1$  de réflexions diffusives contre les interfaces Si/SiO<sub>2</sub> rugueuses dans un intervalle compris entre 0 (interface lisse) et 1 (interface totalement diffuse).  $C_{diff} = 0,14$  correspond à la valeur standard pour un canal peu dopé. Les évolutions de  $B_{eff}(B_{int})$  à  $C_{diff}$  donné pour différentes fréquences de phonons sont tracées en tirets sur la Figure 67. Les résultats obtenus pour  $C_{diff} = 1$  (respectivement  $C_{diff} = 0,14$ ) et pour  $K_{ph} = 0, 1, 2, 5, 10$  et  $20$  (respectivement  $K_{ph} = 0, 0,05, 1, 2$ ) sont indiqués en carrés vides (respectivement en triangles pleins). Les courbes obtenues en faisant varier  $C_{diff}$  à  $K_{ph}$  constant sont indiquées en traits continus. Les cercles vides correspondent à  $K_{ph} = 1$  et  $C_{diff} = 0, 0,07, 0,14, 0,21, 0,4, 0,7$  et  $1$ , les pleins à  $K_{ph} = 0$  et  $C_{diff} = 0, 0,14, 0,5$  et  $1$ . Cette figure permet de distinguer l'effet de chaque type d'interactions (phonon ou rugosité) sur le courant et le transport.

Pour  $B_{int} > 10\%$ , les relations entre  $B_{int}$  et  $B_{eff}$  sont linéaires, mais les coefficients de linéarité dépendent du type de l'interaction modifiée. En dessous de cette limite, le transport devient plus stationnaire et les relations ne sont plus simplement linéaires. La valeur de  $B_{eff}$  à  $B_{int}$  donné n'est donc pas unique ; elle dépend de la valeur du jeu de paramètres  $C_{diff}$  et  $K_{ph}$ . En outre, sur la Figure 66 et la Figure 67, la balisticité effective  $B_{eff}$  est toujours plus grande que la balisticité intrinsèque  $B_{int}$ , par exemple  $B_{int} = 52\%$  et  $B_{eff} = 84\%$  pour le DGMOS standard ( $K_{ph} = 1$  et  $C_{diff} = 0,14$ ) de la Figure 67.

Néanmoins, ces deux paramètres sont fortement corrélés. La pente de la courbe  $B_{int}(B_{eff})$  est toujours beaucoup plus importante pour les  $B_{int}$  faibles. De plus, dans les cas simulés, à  $B_{int}$  donné,  $B_{eff}$  varie seulement sur une plage de 10%. Ainsi, l'évaluation de  $B_{eff}$  permet d'avoir une bonne description du transport électronique et vice-versa.

De plus, ces simulations indiquent que même si le transport purement balistique ( $B_{int} = 100\%$ ) n'est qu'une chimère compte tenu du grand nombre de phénomènes diffusifs présents dans les dispositifs ultracourts [FISC03], les MOSFET ultimes devraient cependant fonctionner très près de la limite balistique ( $B_{eff} > 80\%$  à 300 K).

### 3.4 Paramètres électriques

Dans un transistor de dimensions géométriques données, la balisticité peut évoluer avec la polarisation électrique du transistor, c'est-à-dire dans différents régimes de fonctionnement :

<sup>1</sup> Dans la version semi-classique de MONACO, l'effet sur le transport de la rugosité de l'interface entre le canal et l'oxyde de grille est prise en compte en tirant au sort la nature de la réflexion (spéculaire ou diffuse) subie par les porteurs contre ce type d'interface. La proportion  $C_{diff}$  de réflexions diffusives est ajustée pour retrouver la loi de mobilité universelle dans les couches d'inversion de structures MOS.

régimes de saturation, « ohmique »... On s'intéresse dans ce paragraphe à l'évolution des balisticités  $B_{\text{int}}$  et  $B_{\text{eff}}$  en fonction des tensions de drain  $V_{\text{DS}}$  et de grille  $V_{\text{GS}}$ .

### 3.4.1 Balisticité et tension de grille

La Figure 68 présente, pour un DGMOS avec  $L_{\text{canal}} = 25$  nm, l'évolution de la balisticité intrinsèque  $B_{\text{int}}$  en fonction de la tension  $V_{\text{DS}}$  pour plusieurs tensions  $V_{\text{GS}}$  : 0,3, 0,5 et 0,7 V. Quelle que soit la tension  $V_{\text{DS}}$ ,  $B_{\text{int}}$  diminue quand  $V_{\text{GS}}$  augmente. Dans ces simulations classiques (sans prise en compte du confinement quantique), l'augmentation de  $V_{\text{GS}}$  concentre les porteurs à proximité de l'interface rugueuse et accroît le nombre de réflexions diffusantes. La Figure 69 indique que la part de la rugosité dans le nombre total d'interactions subies par les porteurs traversant le canal de la source vers le drain fait plus que doubler pour  $V_{\text{GS}}$  variant de 0,3 V à 0,7 V. Cette dégradation du transport électronique sous l'effet de  $V_{\text{GS}}$  se traduit par une diminution de  $B_{\text{int}}$  et aussi de la balisticité effective  $B_{\text{eff}}$  comme illustré sur la Figure 69.

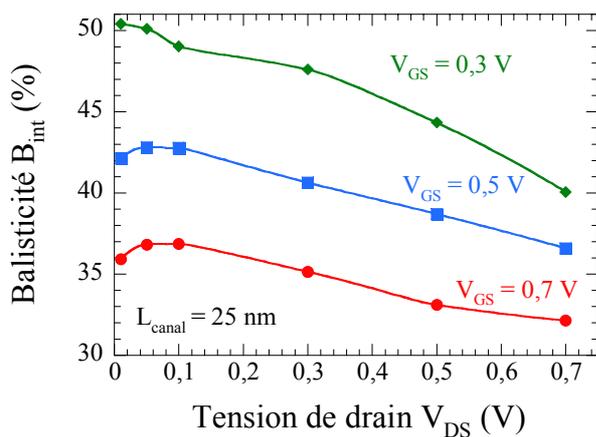


Figure 68 : Evolution de la balisticité intrinsèque  $B_{\text{int}}$  en fonction de la tension de drain  $V_{\text{DS}}$  dans un DGMOS avec  $L_{\text{canal}} = 25$  nm et  $T_{\text{Si}} = 10$  nm pour différents  $V_{\text{GS}} = 0,3, 0,5$  et  $0,7$  V.

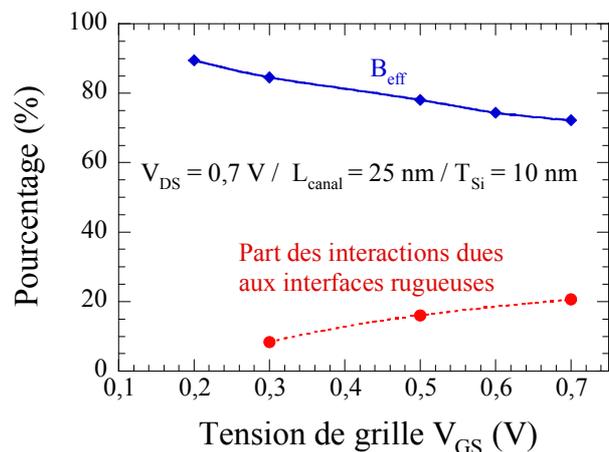


Figure 69 : Pourcentage d'interactions dues à l'interface rugueuse (ronds rouges) et balisticité effective  $B_{\text{eff}}$  (losanges bleus) en fonction de la tension de grille  $V_{\text{GS}}$  dans un DGMOS avec  $L_{\text{canal}} = 25$  nm et  $T_{\text{Si}} = 10$  nm pour  $V_{\text{DS}} = 0,7$  V.

Il faut cependant noter que cet effet néfaste de  $V_{\text{GS}}$  sur  $B_{\text{int}}$  est, en réalité, atténué par la quantification des porteurs perpendiculairement au transport qui limite l'accumulation des porteurs à l'interface (cf. chapitre IV).

### 3.4.2 Balisticité et tension de drain

L'influence de  $V_{\text{DS}}$  sur  $B_{\text{int}}$  est à la fois moins importante mais aussi moins évidente que celle de  $V_{\text{GS}}$ . Pour de très faibles  $V_{\text{DS}}$  ( $< 50$  mV) et à  $V_{\text{GS}}$  fort ( $> 0,5$  V),  $B_{\text{int}}$  est une fonction croissante de  $V_{\text{DS}}$ . Cependant pour  $V_{\text{DS}} = 10$  mV,  $B_{\text{int}}$  reste, à tous les  $V_{\text{GS}}$ , très supérieur à sa valeur lorsque

$V_{DS} = V_{DD}$ . Ainsi même dans la zone dite ohmique le transport est fortement hors d'équilibre dans ce dispositif de 25 nm.

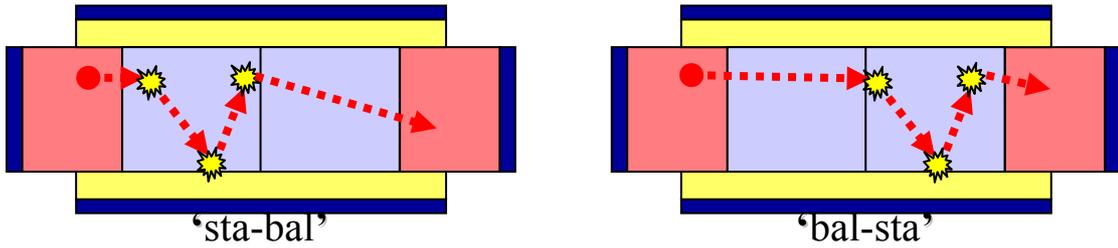


Figure 70 : DGMOS « sta-bal » et « bal-sta ». « sta » moitié du canal standard, « bal » moitié du canal balistique.

Pour des tensions  $V_{DS}$  supérieures,  $B_{int}$  diminue quand  $V_{DS}$  augmente. Pour comprendre cette évolution, nous avons simulé deux nouveaux dispositifs : « sta-bal » et « bal-sta ». Ces transistors sont en tous points identiques à celui étudié en Figure 68, si ce n'est que leurs canaux ont été séparés en deux moitiés, chacune caractérisée par des modèles d'interactions différents. Comme indiqué sur la Figure 70, tous les processus diffusifs ont été supprimés dans la première moitié du canal du dispositif « bal-sta », alors que dans la seconde moitié les fréquences d'interaction standard sont prises en compte. Dans « sta-bal », la première moitié du canal est standard alors que la seconde est balistique, c'est-à-dire sans interaction. Il faut noter que les tensions de seuil  $V_T$  des dispositifs standard, « sta-bal » et « bal-sta » sont quasiment identiques (voir plus loin sur l'insert de la Figure 73). On peut donc les comparer directement pour une même polarisation.

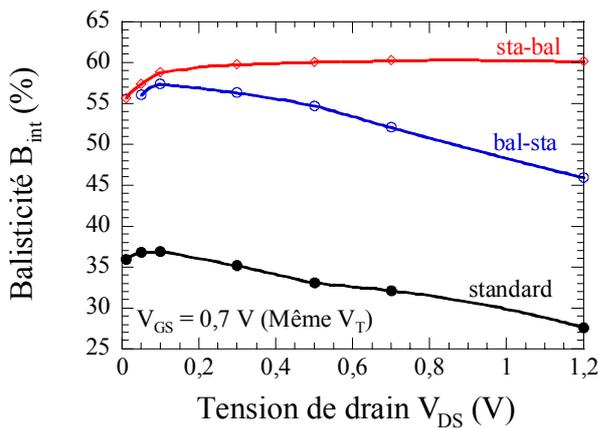


Figure 71 : Evolution de la balisticité intrinsèque  $B_{int}$  en fonction de la tension de drain  $V_{DS}$  dans des DGMOS avec  $L_{canal} = 25$  nm et  $T_{Si} = 10$  nm à  $V_{GS} = 0,7$  V. Cercles pleins noirs pour le transistor standard, losanges rouges pour « sta-bal » et cercles bleus pour « bal-sta ».

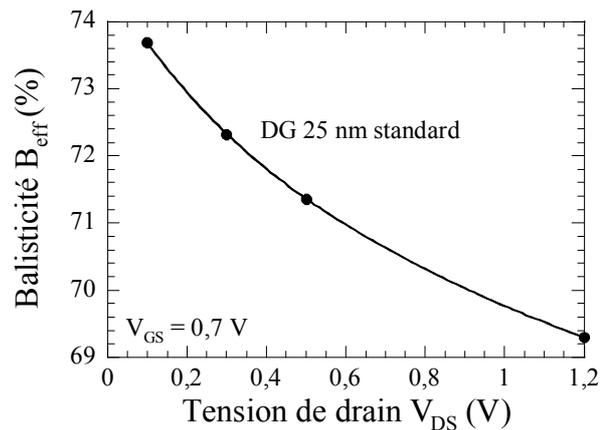


Figure 72 : Evolution de la balisticité effective  $B_{eff}$  en fonction de la tension de drain  $V_{DS}$  dans le DGMOS standard avec  $L_{canal} = 25$  nm et  $T_{Si} = 10$  nm à  $V_{GS} = 0,7$  V.

La Figure 71 compare l'évolution de  $B_{int}$  avec  $V_{DS}$  dans les dispositifs DGMOS standard, « bal-sta » et « sta-bal ». A faible  $V_{DS}$  ( $< 0,1$  V), les évolutions de  $B_{int}$  en fonction de  $V_{DS}$  dans les

deux nouveaux dispositifs sont très similaires. Pour de plus fortes valeurs de  $V_{DS}$ , seul « bal-sta » reproduit la diminution de  $B_{int}$  avec  $V_{DS}$  croissant observée sur le transistor standard. Cette baisse est donc liée aux interactions subies par les électrons chauds (chauffés par la tension de drain) en fin de canal. Cette dégradation du transport dans le canal standard induite par l'augmentation de  $V_{DS}$  engendre une diminution de la balisticité effective  $B_{eff}$  dont l'évolution en fonction de  $V_{DS}$  est tracée en Figure 72.

#### 4. Localisation des interactions et courant

Les différences de comportement entre les dispositifs de « bal-sta » et « sta-bal » observées précédemment suggèrent que la localisation des interactions (en début ou en fin de canal) a des conséquences importantes sur le transport. Dans cette partie, nous allons tout d'abord étudier précisément quel est l'effet des interactions sur le courant de drain suivant leur localisation le long du canal. Puis, nous nous intéresserons plus particulièrement au phénomène de rétrodiffusion.

##### 4.1 Importance des interactions en fin de canal

La Figure 73 permet de comparer les caractéristiques  $I_D(V_{DS})$  et en insert  $I_D(V_{GS})$  de DGMOS standard, « sta-bal », « bal-sta » et balistique ( $L_{canal} = 25$  nm), en reprenant les notations du paragraphe précédent.

L'insert permet de constater sur la courbe  $I_D(V_{GS})$  que les tensions de seuil de ces dispositifs sont quasi-identiques :  $V_T \approx 0,3$  V. De plus, les courbes en échelle semilogarithmique indiquent que les pentes sous seuil  $S$ , elles-aussi quasi-identiques, sont très correctes :  $S=70$  mV/dec. Les comportements sous le seuil ne sont donc pas dégradés par des effets de canal court.

Au début de la zone « ohmique » ( $V_{DS} < 0,1$  V), les courbes « sta-bal » et « bal-sta » sont confondues. Les résistances à l'état passant mesurées à 50 mV sont :  $R_{on\_sta} = 139 \Omega\mu m$  pour le dispositifs standard,  $R_{on\_sta-bal} = 125 \Omega\mu m$  pour « sta-bal »,  $R_{on\_bal-sta} = 119 \Omega\mu m$  pour « bal-sta », et pour le transistor balistique  $R_{on\_bal} = 104 \Omega\mu m$ .  $R_{on\_sta-bal}$  et  $R_{on\_bal-sta}$  correspondent, à moins de 2,5% d'erreur relative, à la conductivité moyenne du canal (loi d'Ohm) :  $\frac{1}{2}(R_{on\_bal} + R_{on\_sta}) = 122 \Omega\mu m$ . Même si  $R_{on\_sta-bal} > R_{on\_bal-sta}$ , la conduction du canal n'est que très faiblement influencée par la position des interactions à faible  $V_{DS}$ . Ce résultat est totalement cohérent avec les évolutions similaires de la balisticité  $B_{int}$  préalablement observées en Figure 71 à faible  $V_{DS}$  pour « sta-bal » et « bal-sta ».

A fort  $V_{DS}$ , dans le régime de saturation, bien que les  $V_{DSsat}$  soient proches, une différenciation des caractéristiques entre « sta-bal » et « bal-sta » apparaît. Leurs conductances valent

respectivement  $G_{d\_sta-bal} = 270$  S/m et  $G_{d\_bal-sta} = 340$  S/m. Dans les canaux standard et balistique les conductances sont égales respectivement à  $G_{d\_sta} = 190$  S/m et  $G_{d\_bal} = 360$  S/m. Plus le canal est balistique, en particulier dans sa première moitié, plus la conductance est grande. On en déduit qu'une haute conductivité, surtout dans le début du canal, réduit  $R_{on}$  et est donc bénéfique au courant  $I_{on}$ , mais en contre partie, cela dégrade la conductance de sortie  $G_d$ .

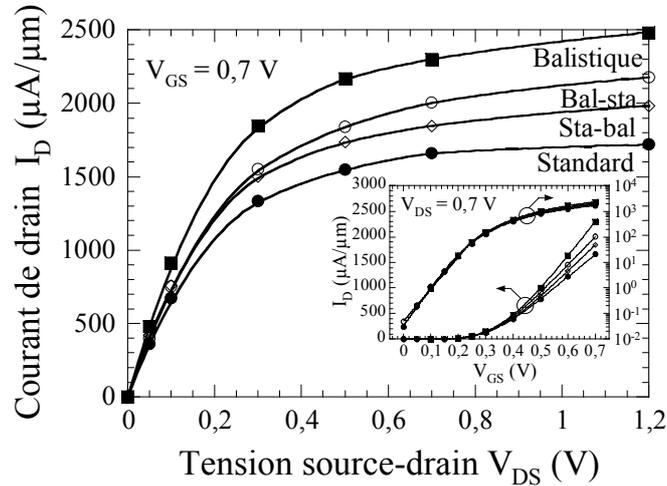


Figure 73 : Caractéristiques  $I_D(V_{DS})$  de DGMOS avec  $L_{canal} = 25$  nm à  $V_{GS} = 0,7$  V : transistors standard (cercles pleins noirs), « sta-bal » (losanges), « bal-sta » (cercles) et balistique (carrées pleines). Insert : caractéristiques  $I_D(V_{GS})$  à  $V_{DS} = 0,7$  V.

A l'état passant le courant vaut  $I_{on\_sta} = 1660$  A/m dans le transistor standard,  $I_{on\_sta-bal} = 1850$  A/m pour « sta-bal »,  $I_{on\_bal-sta} = 2000$  A/m pour « bal-sta » et  $I_{on\_bal} = 2300$  A/m dans le transistor balistique. Le courant calculé en considérant une conductance moyenne du canal donne  $I_{on\_moy} = \frac{1}{2}(I_{on\_bal} + I_{on\_sta}) = 1980$  A/m. L'erreur relative entre  $I_{on\_moy}$  et  $I_{on\_sta-bal}$  est alors supérieure à 6%, or les erreurs sur  $R_{on}$  sont plus faibles. La localisation des interactions a donc un effet non négligeable sur le courant surtout à fort  $V_{DS}$  ; la mobilité en début de canal a un impact plus important sur les performances des composants que celle en fin de canal.

Cependant, comme le laissait augurer l'étude précédente sur  $B_{int}(V_{DS})$ , les effets de la seconde partie du canal sont loin d'être négligeables. En effet  $I_{on\_sta-bal}$  est 12% plus grand que  $I_{on\_sta}$  et  $I_{on\_bal}$  est 15% plus grand que  $I_{on\_bal-sta}$ . De plus, l'écart relatif entre les grandeurs  $I_{on\_sta-bal}$  et  $I_{on\_bal-sta}$  reste faible, c'est-à-dire inférieur à 8%.

## 4.2 Localisation des interactions dans le canal

La Figure 74 présente l'histogramme du nombre d'interactions subies par les porteurs traversant le canal de la source au drain en fonction de la distance  $x$  dans un DGMOS de longueur 25 nm avec  $T_{Si} = 10$  nm. Dans le canal, le nombre total d'interactions reste à peu près constant, exception faite des deux ou trois premiers nanomètres où il est supérieur d'un facteur 1,5 environ. Cette constatation est confirmée par la décroissance relativement linéaire du pourcentage d'électrons

balistiques en fonction de la distance selon l'axe source drain indiquée sur la Figure 75. Les interactions plus nombreuses dans la zone faible champ du début du canal justifient la primauté de la première partie du canal observée précédemment et les différences de courant entre les dispositifs « sta-bal » et « bal-sta ». La relative homogénéité du nombre d'interactions dans le reste du canal explique aussi la faiblesse de cet écart de courant.

En outre, le nombre d'interactions avec l'interface rugueuse, majoritaires en début de canal, se réduit à mesure que l'on s'en éloigne. En effet, la proportion de porteurs aux interfaces se réduit à mesure qu'ils se rapprochent du drain. Les interactions avec les phonons intervallés ont une évolution opposée ; elles deviennent majoritaires dans la seconde moitié du canal où elles « thermalisent » les électrons chauds.

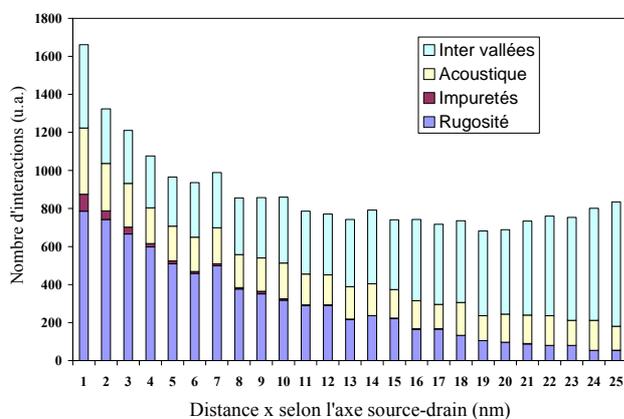


Figure 74 : Répartition du nombre d'interactions le long du canal en fonction du type d'interaction dans un DGMOS tel que  $L_{\text{canal}} = 25 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$  à  $V_{\text{DS}} = V_{\text{GS}} = 0,7 \text{ V}$ . En partant du bas de l'histogramme : en mauve interactions avec l'interface rugueuse, en marron avec les impuretés ionisées, en jaune avec les phonons acoustiques et en bleu ciel avec les phonons intervallés.

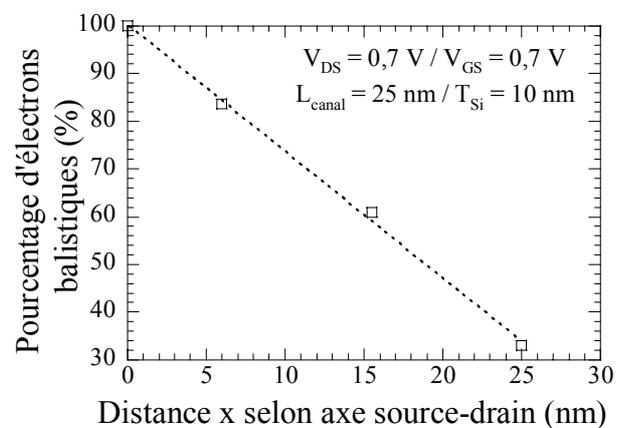


Figure 75 : Evolution du pourcentage d'électrons balistiques le long du canal dans un DGMOS tel que  $L_{\text{canal}} = 25 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$  à  $V_{\text{DS}} = V_{\text{GS}} = 0,7 \text{ V}$ .

Les modèles quasi balistiques, présentés dans cette partie en page 67, soulignent l'importance des interactions rétrodiffusantes. Or toutes les interactions du modèle numérique étant isotropes, un porteur a donc une chance sur deux d'avoir une vitesse négative après une interaction qui est alors potentiellement rétrodiffusante. Le paragraphe suivant va dénombrer les interactions qui le sont réellement, c'est-à-dire compter le nombre d'électrons qui sont effectivement réinjectés du canal vers la source.

### 4.3 Evolution du coefficient de rétrodiffusion

La théorie des flux (ou approche de Landauer présentée dans la partie 1 en page 64) à la base des modèles analytiques en régime quasi balistique considère que seules les interactions pouvant rétrodiffuser les électrons et les réinjecter dans la source ont un effet néfaste sur le courant de drain.

En conséquence, plus l'interaction a lieu à proximité du sommet de la barrière, plus elle a de chances d'être rétrodiffusante [LUND02] donc de dégrader le courant. Cette théorie prend élégamment en compte l'effet prédominant des interactions en début de canal sur la valeur du courant.

Grâce à la procédure de marquage décrite en page 74, nous avons déduit des simulations Monte Carlo l'évolution du coefficient de rétrodiffusion  $R_c$  pour étudier la pertinence de ce modèle. Comme indiqué sur le schéma de Figure 76,  $\Phi_I^+$  est le flux d'électrons injectés au travers de la surface d'entrée  $S_e$  fixée à l'abscisse  $x_{top}$  qui correspond au sommet de la barrière.  $\Phi_0^+$  est le flux d'électrons qui, injecté depuis  $S_e$ , est transmis au travers de la surface de sortie  $S_s$  positionnée en  $x$  ( $x > x_{top}$ ). Enfin,  $\Phi_1^-$  et  $\Phi_2^-$  sont les flux d'électrons rétrodiffusés respectivement dans la zone d'étude comprise entre  $S_e$  et  $S_s$  et après cette zone d'étude.  $R_c(x)$  est donc égal au rapport :  $\Phi_{11}^-(x) / \Phi_I^+$ .

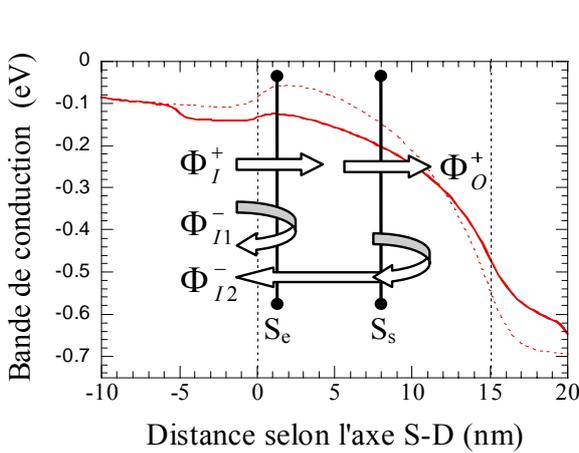


Figure 76 : Schéma des flux et évolution de la bande de conduction dans le DGMOS de 15 nm avec  $T_{Si} = 10$  nm. Trait continu : BC sous l'oxyde, tirets : BC au milieu de la zone active.  $S_e$  surface d'entrée (sommet de la barrière),  $S_s$  surface de sortie.

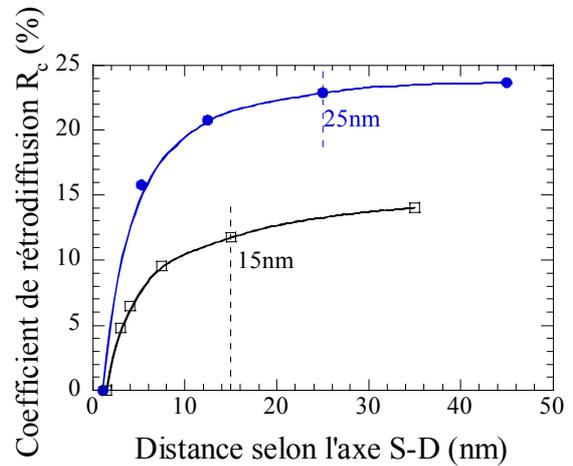


Figure 77 : Evolution du coefficient de rétrodiffusion le long de l'axe source drain dans des DGMOS tels que  $T_{Si} = 10$  nm à  $V_{DS} = V_{GS} = 0,7$  V. Carrés pour  $L_{canal} = 15$  nm et losanges pleins pour  $L_{canal} = 25$  nm.

Pour évaluer la théorie des flux, nous avons reporté dans le Tableau 5 la balisticité effective  $B_{eff}$  obtenue par l'intermédiaire de la méthode utilisée dans les modèles quasi balistiques (cf. section 1.3.3, page 67) :  $B_{eff} = (1-R_c)/(1+R_c)$  [LUND97] et par sa définition :  $B_{eff} = I_{on}/I_{on\_bal}$ . Les deux formules donnent des résultats relativement identiques : les écarts sont inférieurs à 5% pour les deux longueurs étudiées, en considérant le coefficient total «  $R_c$  max » ( $S_s$  placée dans la dernière maille du drain). Ces écarts, très faibles, peuvent être dus au déplacement selon l'axe  $x$  du sommet de la barrière en fonction de la profondeur (axe  $z$ ) qui n'ont pas été pris en compte.

Dans un second temps, nous nous sommes intéressés à la détermination du coefficient  $R_c$  : existe-t-il dans le canal des zones plus rétrodiffusantes que d'autres ? Dans le cas d'un canal à coefficient de rétrodiffusion par unité de longueur constant  $R_c(x) = (1 + l_{pm}/x)^{-1}$  d'après l'annexe page 173. Pour obtenir les valeurs de  $R_c < 25\%$  obtenues ici, il faudrait se trouver dans la partie quasi

linéaire de cette expression. Or sur la courbe  $R_c(x)$  extraite des simulations, tracée en Figure 77, on constate que ce n'est pas le cas. On en déduit (par l'absurde) que la rétrodiffusion n'est pas la même dans tout le canal. L'augmentation de  $R_c(x)$  est en effet bien plus importante en début de canal. Cependant, l'accroissement de  $R_c$  reste monotone dans tout le canal.

Pour des DGMOS avec  $L_{\text{canal}} = 15$  et  $25$  nm, le Tableau 5 indique les valeurs de  $R_c$  à différentes positions dans le canal : après la zone où la bande de conduction chute de  $k_B T/q$  à partir du sommet de la barrière, au milieu du canal, à la fin du canal et à la fin du drain ( $R_c \text{ max}$ ). La dernière ligne du tableau correspond au calcul inverse de  $R_c$  à partir de  $B_{\text{eff}}$  ( $R_c = (1 - B_{\text{eff}}) / (1 + B_{\text{eff}}) = I_{\text{on}} / I_{\text{on\_bal}}$  (cf. paragraphes précédents).

	$L_{\text{canal}} = 15 \text{ nm}$		$L_{\text{canal}} = 25 \text{ nm}$	
	$R_c$	$(1 - R_c) / (1 + R_c)$	$R_c$	$(1 - R_c) / (1 + R_c)$
$R_c(I_{kT})$	8,5% = $0,73 \times R_c(L_{\text{canal}})$	84,3%	19,7% = $0,86 \times R_c(L_{\text{canal}})$	67,0%
$R_c(L_{\text{canal}}/2)$	9,6% = $0,82 \times R_c(L_{\text{canal}})$	82,5%	20,8% = $0,91 \times R_c(L_{\text{canal}})$	65,6%
$R_c(L_{\text{canal}})$	11,7%	79%	22,9%	62,7%
$R_c \text{ max}$	14,0% = $1,20 \times R_c(L_{\text{canal}})$	75,4%	23,7% = $1,03 \times R_c(L_{\text{canal}})$	61,7%
$I_{\text{on}} / I_{\text{on\_bal}}$	14,61% = $1,04 \times R_c \text{ max}$	74,5%	24,15% = $1,02 \times R_c \text{ max}$	61,1%

Tableau 5 : Différentes estimations du coefficient de rétrodiffusion  $R_c$  et de la balisticité effective  $B_{\text{eff}}$  dans des DGMOS.

On constate quantitativement que la très grande majorité des interactions rétrodiffusantes a lieu dans la première partie du canal, voire dans la zone «  $k_B T/q$  » :  $R_c(L_{\text{canal}}/2) / R_c(L_{\text{canal}})$  et  $R_c(I_{kT}) / R_c(L_{\text{canal}})$  valent en effet respectivement 82 et 73% pour  $L_{\text{canal}} = 15$  nm, et 91 et 86% pour  $L_{\text{canal}} = 25$  nm

Cependant, la réduction du canal à la zone «  $k_B T/q$  », suggérée par M. Lundstrom [LUND97], génère une erreur non négligeable sur l'estimation de  $R_c$  dans ces dispositifs. Cette imprécision tend de plus à augmenter lorsque le canal est réduit : 27% d'erreur relative pour  $L_{\text{canal}} = 15$  nm contre 14% pour  $L_{\text{canal}} = 25$  nm. Ainsi dans un canal ultracourt la longueur du canal susceptible d'être rétrodiffusante  $L_{\text{eff}}$  (cf. notation du paragraphe 1.3.3 ) tend vers la longueur du canal  $L_{\text{canal}}$  (comme pour un canal long).

En conclusion, la théorie des flux faisant le lien en  $R_c$  et  $B_{\text{eff}}$  semble donc être pertinente. Il faut néanmoins porter une attention particulière à l'estimation de  $R_c$ . Selon nos simulations, les rétrodiffusions dans la seconde moitié du canal ne sont jamais totalement négligeables et tendent à prendre de l'importance dans les dispositifs de fin de roadmap.

#### 4.4 Rétrodiffusion, hauteur de barrière et concentration en entrée de canal

Il a été suggéré dans la référence [SVIZ03] que les électrons rétrodiffusés dans la seconde moitié du canal ont une influence indirecte sur le courant drain. Pour conserver constante la concentration d'électrons au sommet de la barrière d'injection (hypothèse forte de la page 67), les porteurs rétrodiffusés qui seraient en excès au sommet de la barrière devraient être compensés par un accroissement de la hauteur de la barrière induisant une diminution du courant.

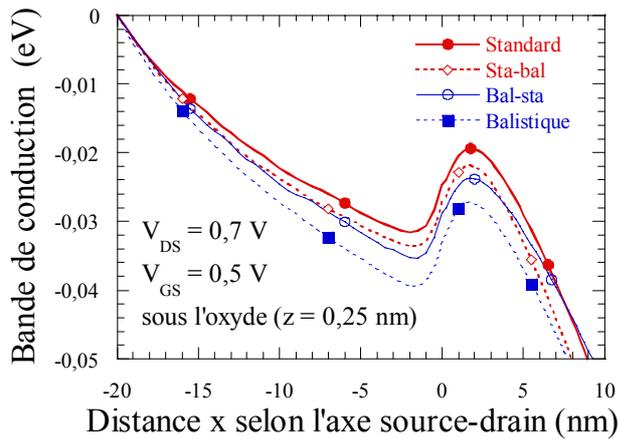


Figure 78 : Evolution de la bande de conduction sous l'oxyde de DGMOS avec  $L_{\text{canal}} = 25 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$  à  $V_{\text{DS}} = 0,7 \text{ V}$ ,  $V_{\text{GS}} = 0,5 \text{ V}$  : transistors standard (cercles pleins rouges), « sta-bal » (losanges rouges), « bal-sta » (cercles vides bleus) et balistique (carrés pleins bleus).

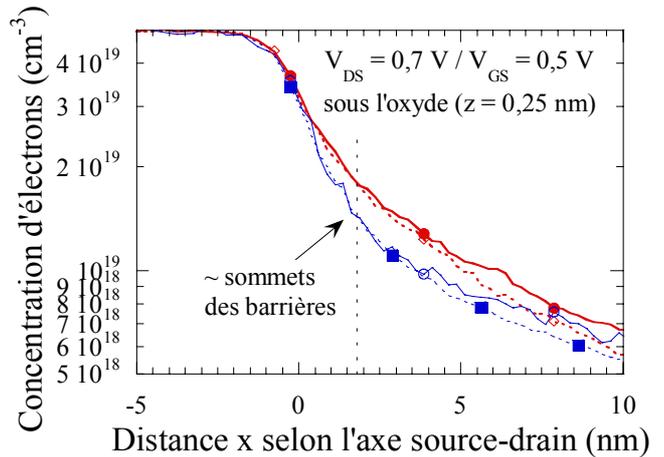


Figure 79 : Evolution de la concentration d'électrons sous l'oxyde de DGMOS avec  $L_{\text{canal}} = 25 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$  à  $V_{\text{DS}} = 0,7 \text{ V}$ ,  $V_{\text{GS}} = 0,5 \text{ V}$  : transistors standard (cercles pleins rouges), « sta-bal » (losanges rouges), « bal-sta » (cercles vides bleus) et balistique (carrés pleins bleus).

Pour étudier cette possible relation entre le sommet de la barrière et les rétrodiffusions, nous présentons l'évolution, sous l'oxyde de grille, de la bande de conduction en Figure 78 et de la concentration en électrons en Figure 79 pour les transistors, désormais bien connus, standard, « sta-bal », « bal-sta » et balistique avec  $L_{\text{canal}} = 25 \text{ nm}$ . La polarisation est  $V_{\text{DS}} = 0,7 \text{ V}$  et  $V_{\text{GS}} = 0,5 \text{ V}$  (à  $V_{\text{GS}} = 0,7 \text{ V}$ , il n'existe plus de barrière d'injection conséquente en entrée de canal...). Le courant dans chaque composant ainsi que la valeur absolue du sommet de la barrière d'injection  $E_B$  et la hauteur de cette barrière (différence entre  $E_B$  et le minimum de la bande de conduction) sont reportés dans le Tableau 6.

	Standard	“Sta-Bal”	“Bal-Sta”	Balistique
Courant (A/m)	780	837	916	1002
Sommet de la barrière $E_B$ (meV)	19,0	21,7	23,8	27,2
Hauteur de la barrière (meV)	12,2	11,7	11,6	12,0

Tableau 6 : DGMOS 25 nm avec  $T_{\text{Si}} = 10 \text{ nm}$  à  $V_{\text{DS}} = 0,7 \text{ V}$  et  $V_{\text{GS}} = 0,5 \text{ V}$

La valeur de l'énergie  $E_B$  augmente bien quand le courant décroît. Cependant, comme la source est résistive, la chute de potentiel dans les caissons est plus significative à mesure que le courant augmente. Cette chute de potentiel augmente mais la hauteur de barrière reste, elle, quasiment

indépendante du courant : 12,2 meV pour ‘standard’, 11,7 meV pour ‘sta-bal’, 11,6 meV pour ‘bal-sta’ et 12,0 meV pour ‘bal-bal’. Un changement de la valeur du courant ne peut donc être simplement expliqué par une modification de la hauteur de barrière.

Une hypothèse forte, utilisée dans les modèles analytiques, considère que la concentration électronique au sommet de la barrière est uniquement commandée par la tension de grille ( $q.n_s \approx W.C_{eq}(V_{GS} - V_T)$  avec les notations du paragraphe 1.3.2 ) dans un dispositif bien dimensionné. Or ces dispositifs de 25 nm sont bien dimensionnés et très peu sensibles aux effets de canal court (ce point est détaillé dans le chapitre suivant). Cependant près de la barrière (entre  $x = 0$  and  $x = 5$  nm) le profil de concentration est similaire pour le dispositif standard et « sta-bal », mais il est nettement plus faible pour « bal-sta » et le balistique. La concentration au sommet de la barrière semble donc plus influencée par les fréquences d’interaction locales que par les interactions dans la seconde moitié du canal. Ainsi cette concentration ne dépend pas que de l’électrostatique mais aussi du transport et donc du coefficient de rétrodiffusion  $R_c$ .

## 5. Mesures de balisticité

Pour valider les simulations numériques, des mesures expérimentales de mobilité balistique [LUSA05] et de balisticité effective  $B_{eff}$  [RUMS00] ont été proposées. Cependant à notre connaissance, une seule équipe a proposé une démarche entièrement expérimentale pour extraire  $B_{eff}$  directement à partir des courbes I-V d’un composant [CHEN02]. Après avoir présenté brièvement la méthode utilisée, nous discuterons les résultats obtenus à partir des caractéristiques électriques de dispositifs réalisés au CEA-LETI<sup>2</sup>.

### 5.1 Bases théoriques

En reprenant l’expression du courant de saturation dans un double grille quasi balistique, il vient :  $I_{D_{sat}}(V_{GS}) = W \times B_{eff} \times v_{th} \times Q_{inj}$  avec  $Q_{inj} = C_G(V_{GS} - V_T)$ ,  $v_{th} = \sqrt{\frac{2.k_B.T}{\pi.m^*}}$  et  $B_{eff} = \frac{1}{1 + 2 \frac{L_{eff}}{l_{pm}}}$

Le courant de drain d’un transistor est dépendant de la température, on peut dériver l’expression précédente [CHEN02], il vient :

$$\frac{\partial I_{D_{sat}}}{\partial T} = I_{D_{sat}} \left[ \frac{\partial B_{eff}}{\partial T} \frac{1}{B_{eff}} + \frac{\partial v_{th}}{\partial T} \frac{1}{v_{th}} + \frac{\partial Q_{inj}}{\partial T} \frac{1}{Q_{inj}} \right]$$

<sup>2</sup> Travail effectué dans une équipe composée de Simon Deleonibus, Thierry Poiroux, Maud Vinet, Jérôme Lolivier et Quentin Raffhay.

On suppose que les dépendances en température  $T$  de certains paramètres sont du type loi de puissance [CHEN04], soit :  $v_{th} \propto T^a$ ,  $l_{pm} \propto T^b$  et  $L_{eff} \propto T^c$ . D'où finalement :

$$\frac{1}{I_{D_{sat}}(T_0)} \left. \frac{\partial I_{D_{sat}}}{\partial T} \right|_{T_0} = \frac{1}{T_0} \left( \frac{2(b-c)}{2 + \frac{l_{pm}}{L_{eff}}} + a \right) - \left. \frac{\partial V_T}{\partial T} \right|_{T_0} \frac{1}{V_{GS} - V_T(T_0)}$$

En supposant, de plus, que  $\frac{l_{pm}}{L_{eff}} \propto T^4$  [CHEN02] et  $v_{th} \propto \sqrt{T}$  on a :  $a=0,5$  et  $b-c = 2$ .

Ainsi, les mesures des termes  $\alpha(T_0) = \frac{1}{I_{D_{sat}}(T_0)} \left. \frac{\partial I_{D_{sat}}}{\partial T} \right|_{T_0}$ ,  $V_T(T_0)$  et  $\eta(T_0) = \left. \frac{\partial V_T}{\partial T} \right|_{T_0}$  permettent de

déterminer le rapport  $\frac{l_{pm}}{L_{eff}}$  et donc la balisticité effective  $B_{eff}$ .

## 5.2 Dispositif expérimental

### 5.2.1 Transistors étudiés

Les mesures ont été réalisées avec des transistors à double grille planaire non autoalignée réalisés dans le cadre du projet européen NESTOR<sup>3</sup>. Une image MEB d'un tel dispositif a été donnée en Figure 36.

Les zones actives des transistors ont des épaisseurs  $T_{Si}$  d'environ 5 nm et des dopages intrinsèques de  $2 \times 10^{15} \text{ at.cm}^{-3}$ . Les caissons ont été dopés à l'Arsenic As à une dose de  $2 \text{ keV}/6 \times 10^{14} \text{ cm}^{-2}$  pour les LDD (Lightly Doped Drain) et  $13 \text{ keV}/2 \times 10^{15} \text{ cm}^{-2}$  pour les HDD (Highly Doped Drain).

Les longueurs de grille  $L_G$  variant entre 370 nm et 10 nm et les largeurs des transistors utilisés sont reportées dans le Tableau 7.

Nom	N°	$L_{mask}$ (nm)	$L_G \approx L_{mask} - 30 \text{ nm}$ (nm)	$W$ ( $\mu\text{m}$ )
LN	4	400	370	0,4
100N	4	100	70	0,4
50W	2	50	20	10
40W	4	40	10	10
40M	6	40	10	2

Tableau 7 : Dimensions des DGMOS caractérisés. N.B Le numéro du dispositif correspond au désalignement volontaire de la grille inférieure (de positif à négatif). Le N° 6 est à grille débordante.

<sup>3</sup> IST-2001-37114

## 5.2.2 Banc de mesure

Les mesures ont été effectuées sur un banc cryogénique de mesures sous pointes. Le bâti sous vide est refroidi par de l'hélium liquide. La régulation de la température s'effectue par un contrôle du débit d'hélium associé à un asservissement en température par chauffage électrique.

Les mesures I/V statiques ont été effectuées à l'aide d'un testeur HP4245.

## 5.3 Résultats

### 5.3.1 Mesures de $B_{eff}$

Les évolutions des caractéristiques  $I_D(V_{GS})$  à fort et faible  $V_{DS}$  en fonction de la température du dispositif 50W n°2 avec  $L_{canal}=20$  nm sont présentées respectivement en Figure 80 et en Figure 81.

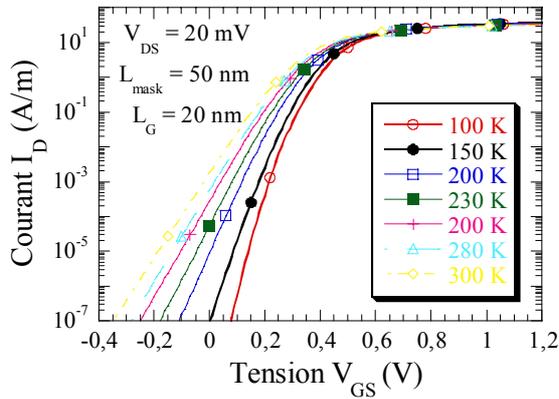


Figure 80 : Caractéristiques  $\log[I_D(V_{GS})]$  du DGMOS NESTOR 50W n°2 à  $V_{DS}=20$  mV pour différentes températures.

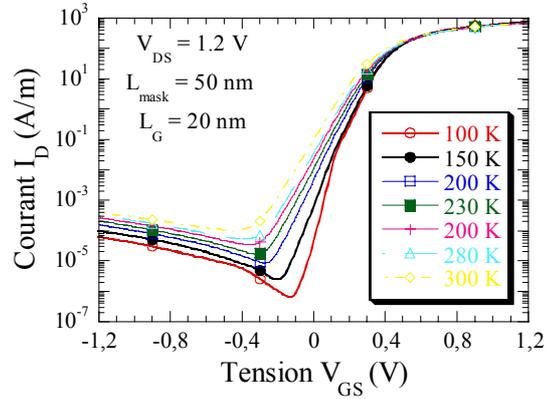


Figure 81 : Caractéristiques  $\log[I_D(V_{GS})]$  du DGMOS NESTOR 50W n°2 à  $V_{DS}=1,2$  V pour différentes températures.

L'évolution de la pente sous le seuil en fonction de la température tracée en Figure 82 indique que la température améliore le comportement sous le seuil. En outre, on peut constater, avec soulagement, que son évolution linéaire est bien cohérente avec l'expression analytique donnée au chapitre I.

La tension de seuil  $V_T$  a été évaluée par la méthode Y [GHIB88]. La température a de plus un effet sur la tension de seuil qui diminue quand la température augmente. Cette évolution quasi linéaire est reportée en Figure 83. On en déduit que  $\eta(T_0) = \left. \frac{\partial V_T}{\partial T} \right|_{T_0}$  est constant sur la plage de températures étudiée. Le décalage de  $V_T$  dû à l'effet DIBL (cf. chapitre I, page 35)  $\Delta V_T(V_{DS})$  est quasiment indépendant de la température et vaut 170 mV. Enfin, l'évolution du courant  $I_D$  est

linéaire sur une grande plage de températures ; la détermination de  $\alpha(T_0)$  est donc facilement réalisable.

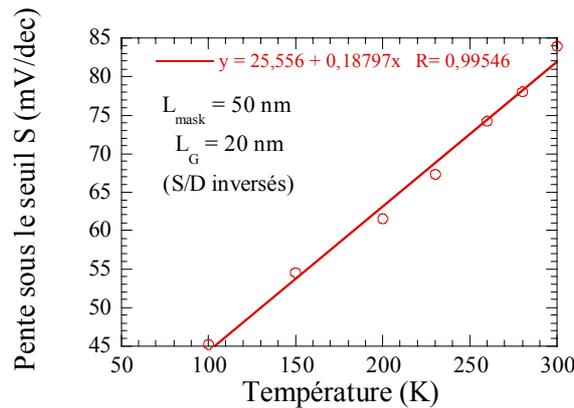


Figure 82 : Evolution de la pente sous le seuil en fonction de la température dans le DGMOS NESTOR 50W n°2.  $R=0.99546$  est le coefficient de régression associé à la droite dont l'équation est indiquée en rouge.

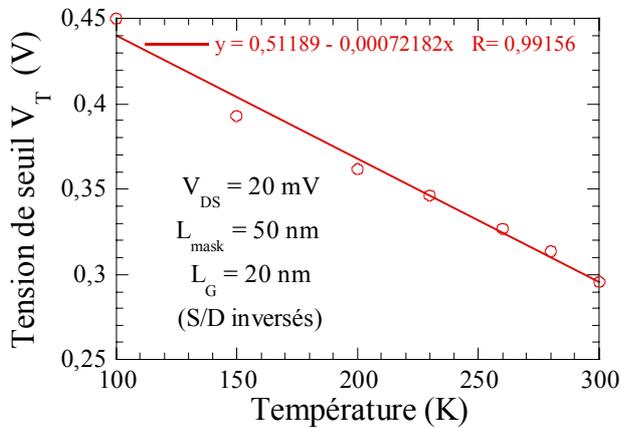


Figure 83 : Evolution de la tension de seuil en fonction de la température dans le DGMOS NESTOR 50W n°2.

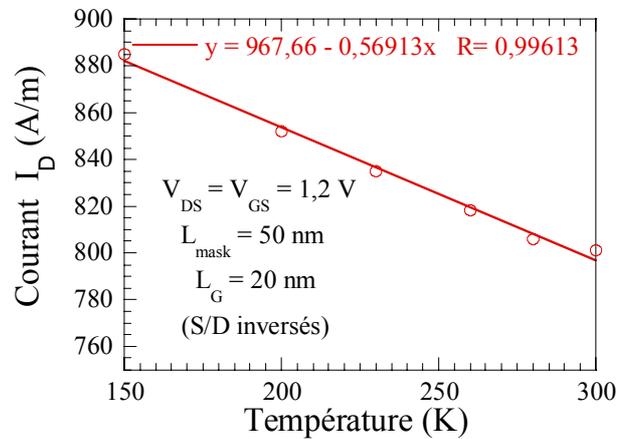


Figure 84 : Evolution du courant de drain  $I_D$  en fonction de la température dans le DGMOS NESTOR 50W n°2.

Le Tableau 8 récapitule les paramètres utiles à la détermination de  $B_{\text{eff}}$  à partir de la relation :

$$\frac{l_{\text{pm}}}{L_{\text{eff}}}(T_0) = \frac{4}{1/2 - \left( \alpha(T_0) + \frac{\eta(T_0)}{V_{\text{GS}} - R_{\text{accès}} \cdot I_D(T_0) - V_T(T_0) - \Delta V_T(V_{\text{DS}})} \right) T_0} - 2$$

Cette expression corrige les effets des différents éléments parasites telles que les résistances d'accès  $R_{\text{accès}}$  et le  $\Delta V_T(V_{\text{DS}})$ . Les paramètres  $R_{\text{accès}}$  ont été extraits par la méthode « shift and ratio » [TAUR92].

	$I_D(T_0)$ (A/m)	$\left. \frac{\partial I_{D_{\text{sat}}}}{\partial T} \right _{T_0}$ (A/m/K)	$\alpha(T_0)$ (K <sup>-1</sup> )	$V_T(T_0)$ (V)	$\eta(T_0)$ (V/K)	$\Delta V_T(V_{\text{DS}})$ (mV)	$R_{\text{accès}}$ $\Omega \mu\text{m}$	$\frac{l_{\text{pm}}}{L_{\text{eff}}}(T_0)$	$B_{\text{eff}}$ (%)
$T_0 = 300$ K	796	0,569	$-7,15 \times 10^{-4}$	0,29	$-7,22 \times 10^{-4}$	170	330	0,421	51,0

Tableau 8 : DGMOS NESTOR 50W n°2 à  $V_{\text{DS}} = 1,2$  V.

Cette méthode donne l'évolution de  $B_{\text{eff}}$  en fonction de la longueur de canal  $L_{\text{canal}}$  tracée en Figure 85. Les résultats du transistor de 10 nm ne sont pas reportés car ils sont trop sensibles aux effets de canal court. Le  $B_{\text{eff}}$  extrait grâce à cette méthode vaut entre 51% (mesure source/drain inversés) et 54% pour  $L_G = 20$  nm. Ces valeurs sont faible comparée aux résultats issus des simulations Monte Carlo de transistor à double grille de 20 nm (cf. page 79). La valeur pour le canal long est, par contre, trop importante. Les ordres de grandeur et les tendances sont cependant concordants.

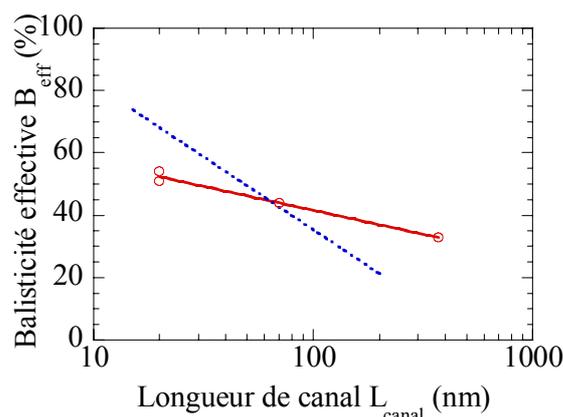


Figure 85 : Evolution de la balisticité effective  $B_{\text{eff}}$  en fonction de la longueur de canal  $L_{\text{canal}}$ . En tirets : simulation Monte Carlo. Ronds rouges : points expérimentaux.

### 5.3.2 Comparaison des méthodes et bilan

Référence	Méthode	$L_{\text{eff}}$ (nm)	$R_c$	$B_{\text{eff}}$
[RUMS00]	Simulation ( $I_{\text{on\_bal}}$ ) +Expérimental $I_{\text{on}}$	115- Massif	0,39	44%
[CHEN04]	Expérimental	68-Massif	0,40	43%
Ce travail	Simulation	25-SG SOI	0,24	62%
[LOLI05]	Simulation ( $I_{\text{on\_bal}}$ ) +Expérimental $I_{\text{on}}$	25-SG SOI	0,26	58%
[LOLI05]	Expérimental	25-SG SOI	0,27	57%
Ce travail	Simulation	20-DG NESTOR	0,18	69%
Ce travail	Expérimental	20-DG NESTOR	0,30	54%

Tableau 9 : Comparaison de résultats d'extraction de  $B_{\text{eff}}$ . La valeur de  $I_{\text{on\_bal}}$  dans [LOLI05] correspond à nos résultats de simulation Monte Carlo.

Le Tableau 9 reprend des extractions parues dans la littérature en utilisant les diverses méthodes proposées. Nos simulations Monte Carlo sont en bon accord avec les mesures pour les transistors simple grille (SG) sur SOI, l'écart relatif est de moins de 10%. Cependant on observe pour les DGMOS un écart sensible alors que les zones actives sont non dopées. Il faut souligner que les

méthodes d'extraction de balisticité sont, à l'heure actuelle, loin d'être matures. Nous avons choisi la seule méthode d'extraction, à notre connaissance, qui soit purement expérimentale. Cependant, de nombreuses questions sur le modèle analytique proposé initialement restent en suspens [CHEN04]. Les résultats présentés ici doivent donc être considérés pour leur aspect prospectif et leurs interprétations quantitatives nécessitent d'être validées.

Les simulations Monte Carlo « classiques » ont clairement mis en évidence la transition entre les régimes de transport quasi stationnaires et quasi balistiques du point de vue de l'évolution de la vitesse et de l'évolution du spectre d'interactions. Dans les transistors à double grille à zone active non dopée qui ont été simulés, cette transition apparaît pour des longueurs de canal  $L_{\text{canal}}$  d'environ 50 nm. A ces dimensions, et ce même à faible tension  $V_{\text{DS}}$ , plus de 10% des électrons traversent le canal sans subir d'interactions ( $B_{\text{int}} \geq 10\%$ ). En outre, nos simulations ont montré que plus le canal est court, plus les phénomènes de rétrodiffusion dans la seconde partie du canal prennent de l'importance contredisant ainsi une idée largement répandue. Les modèles analytiques présentés dans la littérature décrivent mal la charge d'inversion au sommet de la barrière. De plus, même si le transport purement balistique ( $B_{\text{int}} = 100\%$ ) n'est qu'une chimère compte tenu du grand nombre de phénomènes diffusifs présents dans les dispositifs ultracourts [FISC03], les simulations Monte Carlo indiquent que les MOSFET ultimes (de longueur de canal  $L_{\text{canal}} < 15$  nm) devraient cependant fonctionner très près de la limite balistique ( $B_{\text{eff}} > 80\%$ ). Notons que dans ce cas la miniaturisation ne permettra plus d'améliorer les performances des transistors. D'ailleurs, l'influence des porteurs balistiques n'est pas seulement pertinente pour l'étude du courant, mais aussi pour tous les autres aspects du dimensionnement : fiabilité, bruit [VARA97]...

## **Chapitre III :**

### ***Dimensionnement des transistors MultiGrilles***

La réduction de la taille des composants microélectroniques, dictée par la demande de performances toujours plus grande, ne cesse de confronter le transistor MOS sur silicium massif à des effets parasites, dits de canal court (cf. chapitre I), de plus en plus difficiles à contrer par des méthodes classiques. Naturellement très performantes en terme de contrôle électrostatique du canal, les architectures multigrilles, dont la réalisation a été rendue possible par l'amélioration des technologies SOI, sont présentées comme l'avenir de la logique CMOS.

L'apparition de ces composants nécessite l'établissement de nouvelles lois d'échelles pour encadrer leur miniaturisation. De nouveaux modèles de dimensionnement ont donc dû être conçus, puis vérifiés, voire affinés grâce à des simulations numériques, plus lentes mais bien plus précises. Ces étapes préalables doivent limiter le temps d'optimisation des dispositifs réels, ce qui constitue l'objectif principal du TCAD (« Technology Computer Aided Design »).

Dans ce contexte, nous avons utilisé le logiciel de simulation particulière MONACO, développé au sein de l'équipe, pour étudier les caractéristiques électriques de différents transistors multigrilles. Ainsi, après avoir brièvement décrit les grandes lignes du dimensionnement fournies par les modèles analytiques, nous discuterons leur validité grâce aux architectures simulées. Nous détaillerons ensuite certains aspects du dimensionnement concernant la maîtrise des différentes impédances parasites.

#### **1. Nouvelles lois d'échelle**

Les modèles analytiques permettant de dimensionner les dispositifs massifs ont dû être adaptés pour décrire les nouvelles architectures multigrilles sur SOI. Il faut ainsi ajuster de nouveaux facteurs de formes ou longueurs caractéristiques pour optimiser les caractéristiques électriques et assurer l'immunité vis-à-vis des effets de canal court.

Ces règles de dimensionnement sont généralement obtenues par résolution de l'équation de Poisson pour une structure 2D (SGMOS ou DGMOS) en faisant l'hypothèse d'un canal complètement déserté : 
$$\frac{\partial^2 \phi(x, z)}{\partial x^2} + \frac{\partial^2 \phi(x, z)}{\partial z^2} = \frac{qN_A}{\epsilon_{Si}}$$
 où  $N_A$  est le dopage (en ions accepteurs

d'électrons dans les transistors à canal N) de la zone active,  $\phi$  est le potentiel électrostatique et  $\epsilon_{Si}$  permittivité diélectrique du silicium.

Si l'on suppose en outre un potentiel de forme parabolique dans la direction de l'effet de champ (axe z), il vient :

$$\phi(x, z) = c_0(x) + c_1(x) \times z + c_2(x) \times z^2 \text{ [YAN92]}$$

En considérant des conditions aux limites plus ou moins raffinées, et en posant  $\phi(x) = c_0(x) - \frac{qN_A}{\epsilon_{Si}}$ , on peut réécrire l'équation de Poisson sous l'oxyde ( $z = 0$ ) sous la forme :

$$\frac{d^2\phi(x)}{dx^2} - \frac{\phi(x)}{\lambda_1^2} = 0$$

Sous cette forme, la distance caractéristique  $\lambda_1$  apparaît clairement. Ce paramètre contrôle la diffusion du champ électrique dans la direction du transport. Pour se prémunir des effets de canal court, cette distance doit être très inférieure à la longueur du canal et le facteur de forme  $L_{\text{canal}}/\lambda_1$  doit donc être dimensionné avec précaution. L'expression de  $\lambda_1$  dépend des conditions aux limites prises en compte ; différentes expressions en fonction de l'architecture sont reportées dans le Tableau 10.

	SGMOS sur SOI	DGMOS	QGMOS
$\lambda_1$	$\sqrt{\frac{\epsilon_{Si}}{\epsilon_{ox}} T_{Si} T_{ox}}$	$\sqrt{\frac{\epsilon_{Si}}{2\epsilon_{ox}} T_{Si} T_{ox}}$	$\sqrt{\frac{\epsilon_{Si}}{4\epsilon_{ox}} T_{Si} T_{ox}}$

Tableau 10 : Expression de la longueur caractéristique  $\lambda_1$  en fonction de l'architecture : Simple, Double ou Quadruple Grille [YAN92] [COLI04].

En poursuivant les calculs précédents, on obtient l'expression complète du potentiel à l'état bloqué  $\phi(x, z)$ . On peut alors en déduire les différentes grandeurs électriques : valeurs de la tension de seuil, de la pente sous le seuil... [SUZI96] [BOUR06].

## 2. Présentation des dispositifs simulés

Les transistors Simple (SG), Double (DG), Triple (TG) et Quadruple (QG) Grille étudiés correspondent aux spécifications de l'ITRS 2002 pour les composants hautes performances du nœud 65 nm, leurs schémas sont présentés en Figure 86 et Figure 87. L'épaisseur de l'oxyde de grille ( $\text{SiO}_2$ )  $T_{ox}$  est fixée à la limite haute préconisée, c'est-à-dire à 1,2 nm. Le travail de sortie de la grille métallique est de 4,46 eV pour viser une tension de seuil  $V_T$  de 0,2 V. La tension d'alimentation  $V_{DD}$  est égale à 0,7 V. Les concentrations d'impuretés dopantes valent

$N_D = 5 \times 10^{19} \text{ at./cm}^{-3}$  dans les caissons  $N^+$  source/drain et  $N_A = 2 \times 10^{15} \text{ at./cm}^{-3}$  dans le canal de dopage intrinsèque P. Les jonctions sont supposées abruptes. Dans les dispositifs Simple et Triple Grille l'épaisseur de l'oxyde enterré est égale à 25 nm et le dopage P du substrat enterré vaut  $N_A = 2 \times 10^{17} \text{ at./cm}^{-3}$ . L'électrode de substrat est à la masse. Sauf exception (dûment signalée), les dispositifs TG et QG ont une section carrée : la largeur de la zone active  $W$  est égale à son épaisseur  $T_{Si}$ .

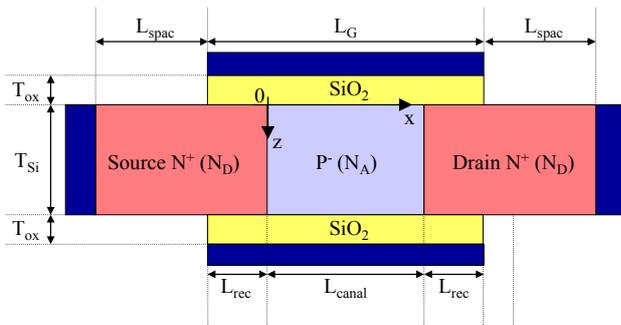


Figure 86 : Schéma d'une coupe de DGMOS.

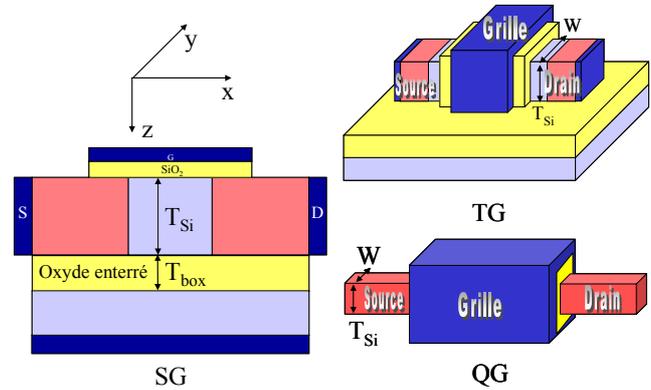


Figure 87 : Schéma 2D d'une coupe de SGMOS et schéma 3D d'un TGMOS et d'un QGMOS.

### 3. Normalisation du courant des architectures 3D

Pour simplifier les comparaisons entre les architectures 2D (simple ou double grille) de tailles différentes, on utilise le courant de drain normalisé par unité de largeur  $W$ . Pour les transistors Triple ou Quadruple Grille, la normalisation du courant par unité de largeur est plus problématique. En effet, comme indiqué sur la Figure 88, ces architectures 3D sont constituées de doigts de conduction qui peuvent être multiples pour atteindre un débit de courant suffisant pour les applications.

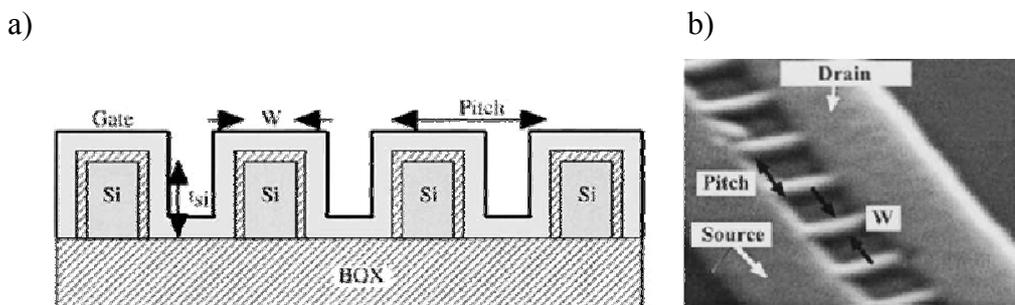


Figure 88 : TGMOS multi-doigts [COLI04]. a) Schéma en coupe et b) image MEB

Plusieurs types de normalisations peuvent être utilisées pour l'étude de ces dispositifs. Pour souligner l'effet des grilles multiples, la normalisation par le périmètre total de la grille pour un triple grille semble la plus adaptée :  $I_D(A/m) = I_D(A)/(W + 2 \times T_{Si})$ . Certains, intéressés par la densité

d'intégration, ont divisé le courant d'un doigt par la distance entre deux doigts, nommée le « pitch »  $P$  [COLI04].

Dans ce travail, pour pouvoir comparer directement les dispositifs 3D et 2D, les courants obtenus ont tous été normalisés de la même façon :  $I_D(A/m) = I_D(A)/W$ . Néanmoins, seul l'utilisation d'un paramètre indépendant de la normalisation («  $CV/I$  ») permet de s'affranchir de ces problèmes récurrents aux études des architectures multigrilles.

#### 4. Influence de l'architecture sur le transport

Cette première étude quantifie l'effet du nombre de grilles, c'est-à-dire le nombre de côtés du canal contrôlés par une grille, sur le transport ainsi que l'influence de l'épaisseur du film. Dans ce contexte, nous avons reporté en Figure 89 les spectres d'interactions à l'état passant des SG, DG, TG et QGMOS avec  $T_{Si} = 10$  nm et celui d'un QGMOS avec  $T_{Si} = 5$  nm.

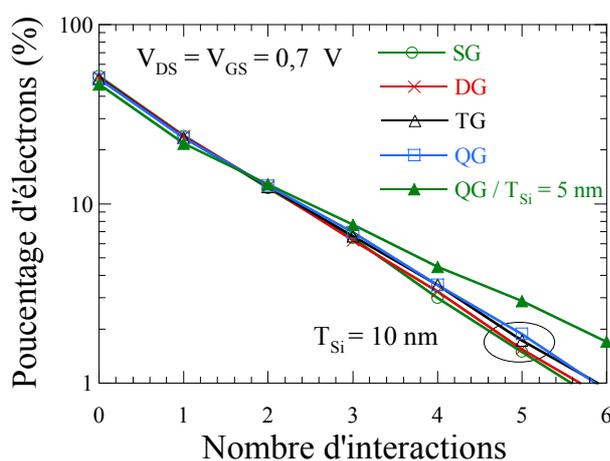


Figure 89 : Spectres d'interactions dans différentes architectures avec  $L_{canal} = 15$  nm : SG, DG, TG et QGMOS avec  $T_{Si} = 10$  nm et un QGMOS avec  $T_{Si} = 5$  nm.  $V_{GS} = V_{DS} = 0,7$  V.

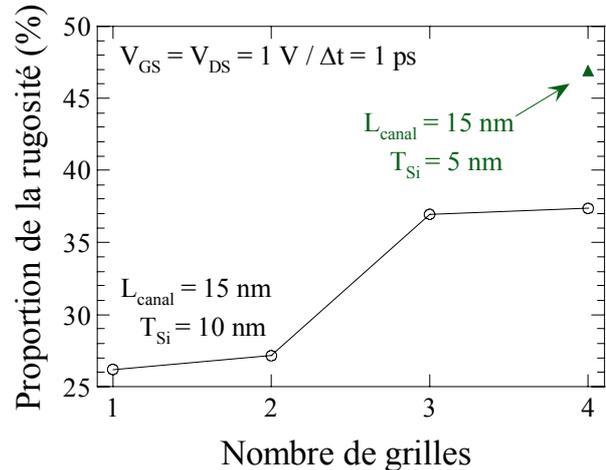


Figure 90 : Evolution de la part des interactions rugueuses en fonction du nombre de grilles dans différentes architectures avec  $L_{canal} = 15$  nm : SG, DG, TG et QGMOS avec  $T_{Si} = 10$  nm et un QGMOS avec  $T_{Si} = 5$  nm.  $V_{GS} = V_{DS} = 0,7$  V.

On constate que le transport n'est affecté que très légèrement par le nombre de grilles : pour  $T_{Si} = 10$  nm,  $B_{int}$  varie de 51,4% dans le SG à 49,8% dans le QGMOS. Toutefois, plus le nombre de grilles est élevé, plus la surface des interfaces rugueuses est importante et plus le pourcentage d'interactions dues à la rugosité parmi toutes celles subies par les porteurs présents dans le canal, indiqué en Figure 90, est élevé. En outre, on constate que les dispositifs 2D (SG et DG) ont sensiblement le même pourcentage d'interactions avec les interfaces rugueuses, de même pour les dispositifs 3D (TG et QG). En effet, en considérant l'interface avec l'oxyde enterré, il y a le même nombre de parois rugueuses dans les dispositifs 2D (2 parois rugueuses) et dans les 3D (quatre parois rugueuses).

De même, on constate que l'épaisseur du film  $T_{Si}$  a une influence mineure sur le transport :  $B_{int}$  vaut 46,4% pour le QGMOS avec  $T_{Si} = 5$  nm. Cette valeur est plus faible que pour  $T_{Si} = 10$  nm à cause de l'augmentation du nombre d'interactions avec les interfaces rugueuses.

Finalement, le transport reste très fortement balistique dans ces dispositifs ultracourts ( $L_{canal} = 15$  nm) et est très peu influencé, de ce point de vue, par l'architecture et l'épaisseur de la zone active  $T_{Si}$ .

## 5. Caractéristiques statiques

### 5.1 Etat passant

Les Figure 91 et Figure 92 représentent les caractéristiques  $I_D(V_{DS})$  à fort  $V_{GS}$  pour toutes les architectures simulées avec  $L_{canal} = 15$  nm respectivement pour  $T_{Si} = 10$  et 5 nm.

Sur la caractéristique du transistor simple grille le plus épais ( $T_{Si} = 10$  nm), les effets de canal court sont tellement importants que l'on distingue à peine la zone « ohmique » de la zone de saturation ( $G_d=880$  S/m). La conductance s'améliore avec le nombre de grilles pour  $T_{Si} = 10$  nm :  $G_d$  vaut 540 S/m pour le DG, 480 S/m pour le TG et 370 S/m pour le QG. La réduction de l'épaisseur  $T_{Si}$  a aussi un effet très bénéfique sur la conductance. Avec  $T_{Si} = 5$  nm,  $G_d$  vaut 150 S/m dans DG, mais elle reste inacceptable dans le SG.

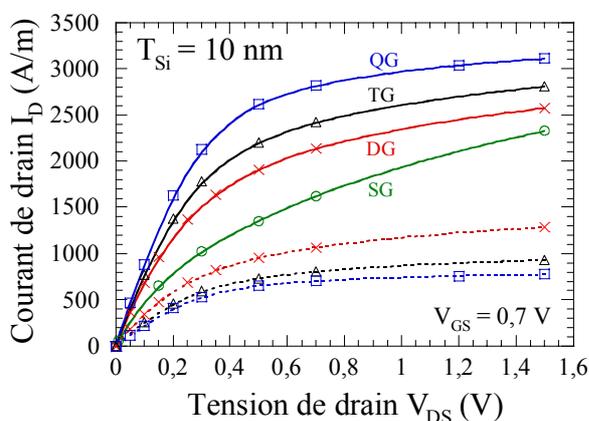


Figure 91 : Caractéristiques  $I_D(V_{DS})$  à  $V_{GS} = 0,7$  V dans différentes architectures : SG, DG, TG et QGMOS.  $L_{canal}=15$  nm et  $T_{Si} = 10$  nm. En tirets, le courant est divisé par le nombre de grilles : /2 pour le DG, /3 pour TG, /4 pour le QG.

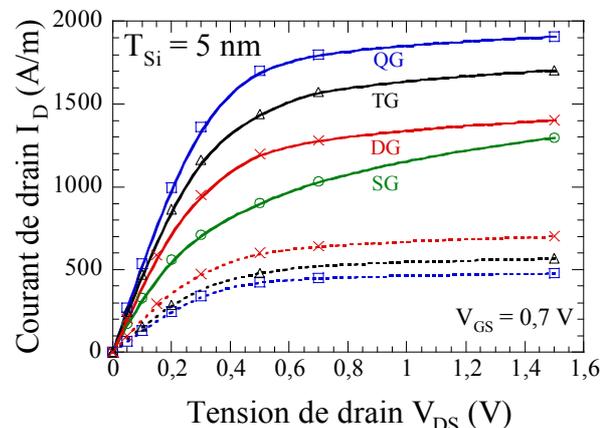


Figure 92 : Caractéristiques  $I_D(V_{DS})$  à  $V_{GS} = 0,7$  V dans différentes architectures : SG, DG, TG et QGMOS.  $L_{canal}=15$  nm et  $T_{Si} = 5$  nm. En tirets, le courant est divisé par le nombre de grilles : /2 pour le DG, /3 pour TG, /4 pour le QG.

Le courant à l'état passant  $I_{on}$  augmente quand le nombre de grilles augmente. Avec  $T_{Si} = 10$  nm,  $I_{on}$  atteint 1623 A/m pour le SG, 2138 A/m pour le DG, 2420 A/m pour le TG et 2815 A/m pour le QG. Cependant, cette augmentation n'est pas proportionnelle au nombre de grilles ; les caractéristiques normalisées par le nombre de grilles représentées en tirets sont diminuées quand le

nombre de grilles augmente. Ainsi, deux transistors à double grille en parallèle débitent plus de courant qu'un quadruple grille. Il faut noter que le courant est supérieur aux 900 A/m recommandés par l'ITRS 2002 dans tous ces dispositifs sur SOI. Néanmoins, la diminution de  $T_{Si}$  réduit de façon quasi proportionnelle le courant  $I_{on}$ , ainsi le courant du DG passe de 2420 A/m à 1280 A/m pour  $T_{Si}$  variant de 10 à 5 nm.

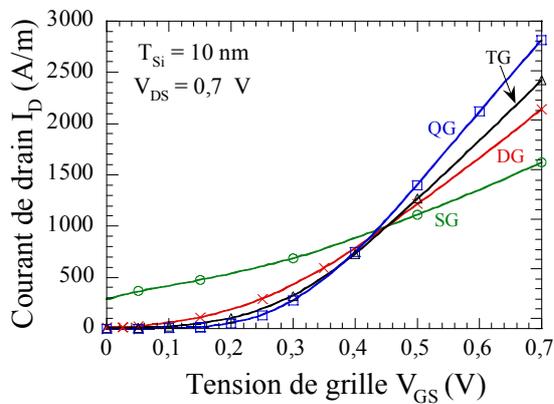


Figure 93 : Caractéristiques  $I_D(V_{GS})$  à  $V_{DS} = 0,7 V$  dans différentes architectures : SG, DG, TG et QGMOS.  $L_{canal} = 15 nm$  et  $T_{Si} = 10 nm$ .

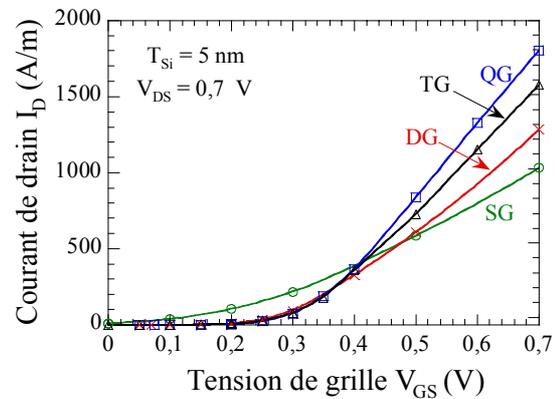


Figure 94 : Caractéristiques  $I_D(V_{GS})$  à  $V_{DS} = 0,7 V$  dans différentes architectures : SG, DG, TG et QGMOS.  $L_{canal} = 15 nm$  et  $T_{Si} = 5 nm$ .

Sur les caractéristiques  $I_D(V_{GS})$  à fort  $V_{DS}$  de la Figure 93 et de la Figure 94, nous retrouvons les mauvaises performances des SG. De plus, la transconductance  $G_m$  est, elle aussi, améliorée par le nombre de grilles. Avec  $T_{Si} = 10 nm$ ,  $G_m$  atteint 4170 S/m pour le DG, 5700 S/m pour le TG et 7070 S/m pour le QG. Cette augmentation n'est donc pas non plus proportionnelle au nombre de grilles. La diminution de l'épaisseur du dispositif réduit la transconductance : pour le DG avec  $T_{Si} = 5 nm$ ,  $G_m = 3350 S/m$  ce qui reste plus que convenable (l'ITRS préconise 1580 S/m).

## 5.2 Etat bloqué

La Figure 95 et la Figure 96 représentent les caractéristiques en échelle semi-logarithmique  $I_D(V_{GS})$  à fort  $V_{DS}$  des dispositifs avec  $L_{canal} = 15 nm$  et respectivement pour  $T_{Si} = 10$  et 5 nm.

Une fois de plus les caractéristiques des transistors simple grille ( $T_{Si} = 5$  et 10 nm) sont déplorables : leur comportement sous le seuil est très mauvais. A cause d'effets de canal court totalement incontrôlés, les pentes sous le seuil sont très supérieures à 100 mV/dec, très loin des recommandations de l'ITRS 2002 :  $S < 80 mV/dec$ .

Le comportement sous le seuil (pente sous le seuil et  $\Delta V_T(V_{DS})$ ) s'améliore avec le nombre de grilles : pour  $T_{Si} = 10 nm$ ,  $S$  vaut 110 mV/dec et  $\Delta V_T(V_{DS}) = 135 mV$  pour le DG,  $S = 96 mV/dec$  et  $\Delta V_T(V_{DS}) = 100 mV$  pour le TG et  $S = 83 mV/dec$  et  $\Delta V_T(V_{DS}) = 44 mV$  pour le QG. La réduction

de l'épaisseur  $T_{Si}$  a aussi un effet très bénéfique sur la conductance. Avec  $T_{Si} = 5$  nm,  $S$  vaut 80 mV/dec et  $\Delta V_T(V_{DS}) = 24$  mV dans le DG.

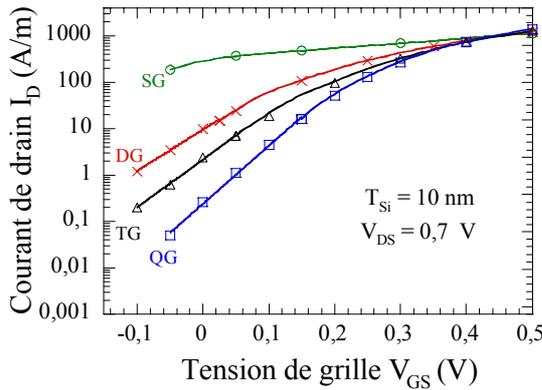


Figure 95 : Caractéristiques  $\log[I_D(V_{GS})]$  à  $V_{DS} = 0,7$  V pour différentes architectures : SG, DG, TG et QGMOS.  $L_{canal} = 15$  nm et  $T_{Si} = 10$  nm.

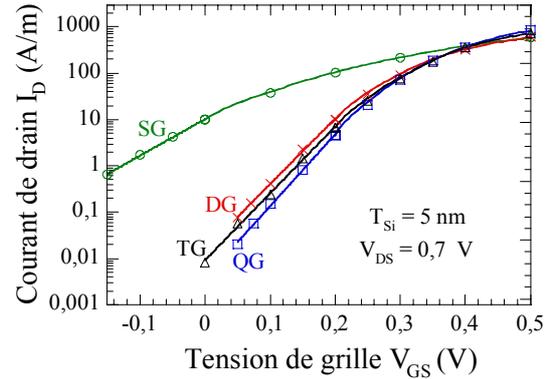


Figure 96 : Caractéristiques  $\log[I_D(V_{GS})]$  à  $V_{DS} = 0,7$  V pour différentes architectures : SG, DG, TG et QGMOS.  $L_{canal} = 15$  nm et  $T_{Si} = 5$  nm.

Donc le pouvoir bloquant des transistors est amélioré par la réduction de l'épaisseur de la zone active  $T_{Si}$ . En effet pour  $T_{Si} = 10$  nm, quatre grilles sont nécessaires pour atteindre une pente sous le seuil convenable :  $S < 80$  mV/dec, alors que pour  $T_{Si} = 5$  nm, deux sont suffisantes.

### 5.3 Simulations et modèle analytique

Le Tableau 11 reporte les calculs du facteur de forme  $L_{canal}/\lambda_1$  présentés au paragraphe 1. pour les différentes architectures simulées. Les cases surlignées correspondent aux transistors mal dimensionnés d'après les simulations précédentes. Comme mentionné dans la référence [COLI04], ce tableau confirme que, pour se prémunir des effets de canal court, il faut que  $L_{canal}/\lambda_1$  soit au moins égal à 5.

$T_{Si}$ (nm)	$L_{canal}$ (nm)	SGMOS sur SOI	DGMOS	QGMOS
		$L_{canal}/\lambda_1$	$L_{canal}/\lambda_1$	$L_{canal}/\lambda_1$
10	15	2,5	3,5	5,0
5	15	3,5	5,0	7,0

Tableau 11 : Calcul du rapport  $L_{canal}/\lambda_1$  pour divers  $L_{canal}$ ,  $T_{Si}$  avec  $T_{ox} = 1.2$  nm en fonction de l'architecture : Simple, Double ou Quadruple Grille. Les cases surlignées correspondent aux transistors mal dimensionnés.

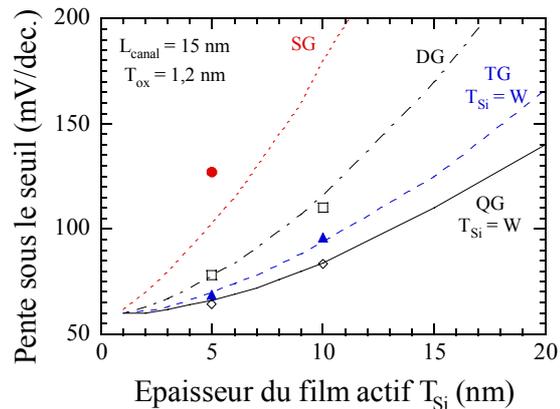


Figure 97 : Pente sous le seuil  $S$  en fonction de l'épaisseur du film du silicium  $T_{Si}$  pour différentes architectures MOS sur SOI : simple (SG), double (DG), triple grille (TG) et quadruple grille (QG).  $L_{canal} = 15$  nm. Les courbes et les symboles sont respectivement issus d'un modèle analytique [BOUR06] et des simulations Monte Carlo.

Les résultats des simulations Monte Carlo ont été confrontés au modèle analytique développé au sein de l'équipe [BOUR06]. Sur la Figure 97, les résultats des dispositifs simulés ont été superposés aux faisceaux de pentes sous le seuil en fonction de l'épaisseur  $T_{Si}$  fournis par ce modèle. On obtient une bonne cohérence pour les DG (carrés vides), TG (triangles pleins) et QG (losanges vides), ce modèle semble donc très performant pour l'étude des architectures multigrilles. Cependant il existe des différences sensibles pour les transistors simple grille (ronds pleins) où l'hypothèse de complète désertion est sujette à caution...

#### 5.4 Analyse des effets de canal court

Dans cette partie, les effets de canal court ont été étudiés plus en détail. Grâce aux simulations Monte Carlo, nous avons analysé les densités de courant pour localiser les courants de fuite et les perturbations en terme de contrôle électrostatique du canal. Cela permet d'avoir une meilleure compréhension des équations de la page 97.

##### 5.4.1 Localisation des courants de fuite

Le SGMOS avec  $L_{canal} = 15$  nm et  $T_{Si} = 10$  nm est, d'après l'analyse des caractéristiques I-V précédentes, très sensible aux effets de canal court. Les cartographies 2D de densité de courant à l'état passant ( $V_{GS} = 0,7$  V et  $V_{DS} = 0,7$  V) et à l'état bloqué ( $V_{GS} = 0$  V et  $V_{DS} = 0,7$  V), présentées en Figure 98, confirment ces soupçons. Un courant de fuite important dû à l'effet parasite du drain, du même ordre de grandeur que la densité de courant à l'état passant sous la grille, est localisé près de l'oxyde enterré. La zone la plus éloignée de la grille est aussi la zone la moins bien contrôlée ; c'est donc la plus sensible aux phénomènes de percement volumique.

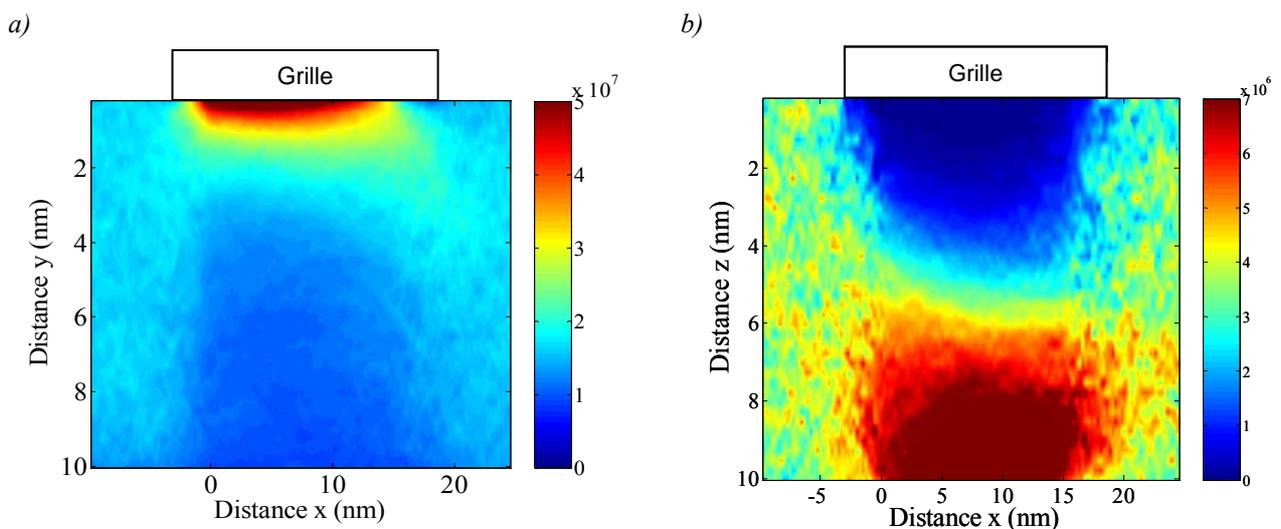


Figure 98 : Cartographies 2D de densité de courant  $J_x$  selon l'axe  $x$  en  $A/cm^2$  dans un SGMOS avec  $L_{canal} = 15$  nm et  $T_{Si} = 10$  nm. a) Etat passant :  $V_{GS} = 0,7$  V et  $V_{DS} = 0,7$  V et b) état bloqué :  $V_{GS} = 0$  V et  $V_{DS} = 0,7$  V.

L'ajout d'une grille supplémentaire pour contrôler la zone de perçement à la place de l'oxyde enterré est une solution naturelle. Dans le transistor à double grille de même dimension que le SGMOS présenté en Figure 98.a), on constate sur la Figure 99.a que la zone la moins bien contrôlée par les grilles se situe au milieu de la zone active. Bien que les fuites soient nettement plus faibles dans ce dispositif, elles demeurent importantes.

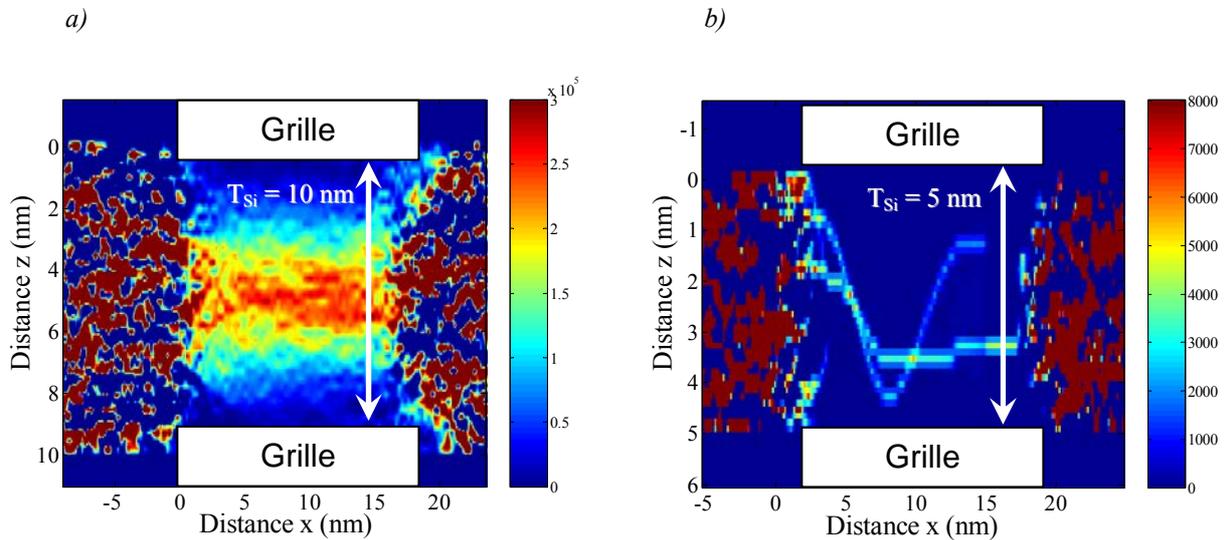


Figure 99 : Cartographie 2D de densité de courant  $J_x$  selon l'axe  $x$  en  $A/cm^2$  dans des DGMOS avec  $L_{\text{canal}} = 15 \text{ nm}$  : a)  $T_{\text{Si}} = 10 \text{ nm}$ , b)  $T_{\text{Si}} = 5 \text{ nm}$ .  $V_{\text{GS}} = 0 \text{ V}$  et  $V_{\text{DS}} = 0,7 \text{ V}$ .

Pour augmenter encore le contrôle électrostatique dans la profondeur du film, il faut rapprocher les grilles et donc diminuer l'épaisseur  $T_{\text{Si}}$ . Avec  $T_{\text{Si}} = 5 \text{ nm}$ , l'état bloqué est quasi idéal : pour un temps de simulation identique aux cas précédents, la Figure 99.b indique les trajectoires des quelques rares électrons qui ont réussi à traverser le canal ; l'aspect particulière des simulations Monte Carlo est mis en évidence.

#### 5.4.2 Etude de l'effet du « DIBL » (Drain Induced Barrier Lowering)

Les extractions précédentes de  $\Delta V_{\text{T}}(V_{\text{DS}})$  sont accessibles aux expérimentateurs, mais elles ne sont que des mesures indirectes de l'évolution de la hauteur de la barrière de potentiel en entrée de canal (effet DIBL, cf. page 35 du chapitre I). Ces analyses peuvent être réalisées grâce aux résultats de simulations, qui donnent accès aux évolutions de la bande de conduction dans le dispositif.

La Figure 100 indique l'évolution des bandes de conduction dans le dispositif à l'état bloqué ( $V_{\text{GS}} = 0 \text{ V}$ ) pour différentes polarisations de drain  $V_{\text{DS}} = 10 \text{ mV}$ ,  $50 \text{ mV}$  et  $0,7 \text{ V}$  du DGMOS avec  $L_{\text{canal}} = 15 \text{ nm}$  et  $T_{\text{Si}} = 10 \text{ nm}$ . On constate clairement l'abaissement de la barrière en fonction de la tension de drain. Les valeurs de la hauteur de barrière en fonction de  $V_{\text{DS}}$  au milieu de la zone active, mais aussi sous l'oxyde, sont reportées sur la Figure 101. Comme suggéré par l'étude précédente des cartographies en courant, l'influence du drain n'est pas la même à toutes les

profondeurs du film de silicium. La barrière en entrée de canal qui assure le blocage du transistor est bien plus haute sous l'oxyde qu'au milieu de la zone active et surtout son abaissement est bien moins important lorsque  $V_{DS}$  augmente. La région la plus éloignée des grilles, c'est-à-dire le milieu du canal, est aussi la moins bien contrôlée. Nous allons focaliser notre étude du DIBL sur cette zone qui correspond au pire cas.

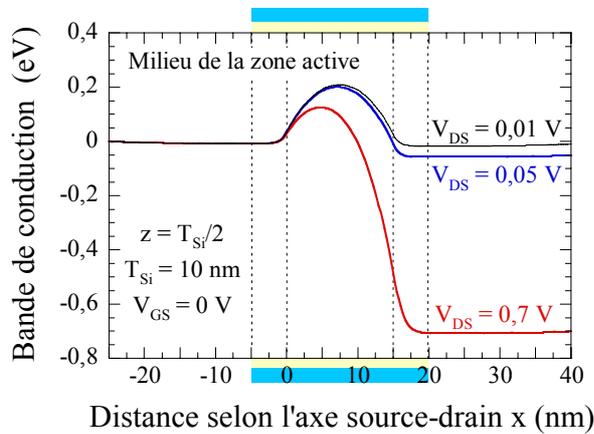


Figure 100 : Coupes longitudinales de la bande de conduction à  $V_{GS} = 0$  V au milieu de la zone active ( $z = T_{Si}/2$ ) et à différentes polarisations de drain ( $V_{DS} = 10$  mV, 50 mV et 0,7 V) du DGMOS avec  $L_{canal} = 15$  nm et  $T_{Si} = 10$  nm. Les pointillés verticaux indiquent la zone de recouvrement des caissons par les grilles.

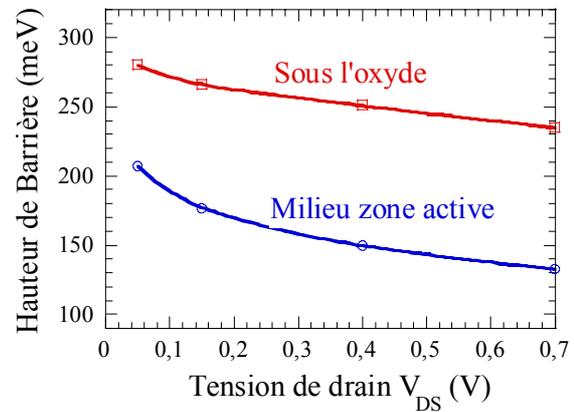


Figure 101 : Evolution de la hauteur de barrière d'injection source/canal en fonction de la tension de drain  $V_{DS}$  à  $V_{GS} = 0$  V à différentes profondeurs : sous l'oxyde ( $z = 0$  nm) et au milieu de la zone active ( $z = T_{Si}/2$ ) pour différentes polarisations de drain  $V_{DS}$  du DGMOS avec  $L_{canal} = 15$  nm et  $T_{Si} = 10$  nm.

Déterminer le DIBL pour  $V_{DS}$  variant entre 50 mV et 0,7 V donne une approximation à moins de 10% d'erreur relative près. Cette erreur nous paraît acceptable d'autant plus que l'obtention de la convergence de simulations Monte Carlo à plus faible polarisation devient délicate... Sur cette plage, le DIBL du DG « épais » ( $T_{Si} = 10$  nm) est conséquent, il vaut 85 mV soit 40% d'abaissement relativement à sa valeur à faible  $V_{DS}$ .

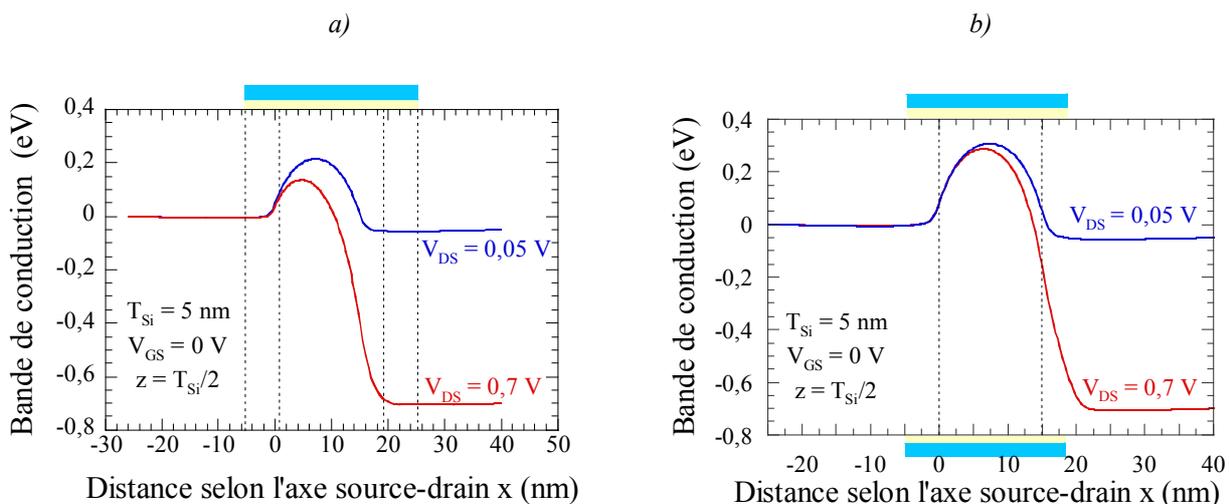


Figure 102 : Coupes longitudinales de la bande de conduction à  $V_{GS} = 0$  V à différentes profondeurs et à différentes polarisations de drain ( $V_{DS} = 50$  mV et 0,7 V) avec  $L_{canal} = 15$  nm et  $T_{Si} = 5$  nm pour différentes architectures : a) SGMOS, b) DGMOS.

La comparaison des bandes de conduction du SGMOS de la Figure 102.a et de celles du DGMOS de la Figure 102.b illustre clairement la diminution du DIBL grâce à l'ajout d'une nouvelle grille. Comme reporté dans le Tableau 10, le DIBL est encore plus faible pour le QGMOS de même dimension. Ce dernier est d'ailleurs quasiment insensible à la polarisation de drain. Le contrôle de grille est donc bien renforcé par l'ajout de nouvelles grilles.

	SG	DG	QG
<b>DIBL (meV)</b>	83	19	3
<i>écart en %</i>	39%	6%	<1%

Tableau 12 : Valeur du DIBL en fonction de l'architecture : Simple, Double, Quadruple Grille avec  $L_{canal} = 15$  nm et  $T_{Si} = 5$  nm.

Curieusement, le Tableau 13 indique que le DIBL maximal dans le TG (25 meV, soit 7,9%) est plus élevé que celui du transistor à double grille équivalent (19 meV, soit 6%). Cependant, la comparaison de ces extractions est sujette à caution car cette importante valeur du DIBL à l'interface avec le l'oxyde enterré est induite par des effets de bord 3D. Or, les DG n'étant simulés qu'en 2D, on ne connaît pas leur DIBL en bord de dispositif ( $y = 0$  et  $y = W$ ). Pour effectuer une comparaison plus pertinente, il est préférable de s'intéresser aux valeurs du DIBL extraites au niveau du plan médian entre l'oxyde enterré et la grille supérieure. Le DIBL  $y$  est égal à 15 meV (4,7%) et est donc meilleur que celui du DG.

	$z = T_{Si}$	$z = T_{Si}/2$	$z = 0$
<b>DIBL (meV)</b>	25	15	8
<i>écart en %</i>	7,9%	4,7%	2,4%

Tableau 13 : Expression du DIBL dans un Triple Grille avec  $L_{canal} = 15$  nm et  $T_{Si} = 5$  nm sur le plan médian entre les deux grilles latérales ( $y = W/2$ ) pour différentes profondeurs : à l'interface avec l'oxyde enterré ( $z = T_{Si}$ ), au milieu de la zone active ( $z = T_{Si}/2$ ) et sous l'oxyde de grille ( $z = 0$ ).

En outre, la réduction de l'épaisseur  $T_{Si}$  a un effet très bénéfique pour le DIBL. Le DIBL des DGMOS de 15 nm passe de 85 meV (soit 40% de variation relative) à 19 meV (soit 6% de variation relative) quand  $T_{Si}$  passe de 10 à 5 nm.

### 5.4.3 Contrôle de charge

Afin d'étudier plus en détail le contrôle de la grille sur le milieu de la zone active, nous avons comparé des coupes de concentrations en électrons dans le sens perpendiculaire au transport, présentées en Figure 103, de transistors SG, DG, TG et QGMOS avec  $L_{canal} = 15$  nm et  $T_{Si} = 10$  nm. Ces coupes ont été faites au milieu du canal sous faible polarisation de drain ( $V_{DS} = 50$  mV) pour n'étudier que l'effet de la tension de grille (le DIBL a été étudié précédemment...).

A faible  $V_{GS}$  (état bloqué), les coupes de concentration dans les dispositifs SG, DG voire TGMOS ont toujours une certaine concavité : la concentration varie fortement en fonction de la distance dans la profondeur du film. Les valeurs de la concentration de porteurs dans les zones éloignées des grilles, au milieu du film dans le DGMOS et près de l'oxyde enterré pour le SGMOS, sont beaucoup moins bien contrôlées par le potentiel de grille. Une faible tension de commande ( $V_{GS} = 0$  V) ne permet pas de dépeupler convenablement ces zones. Elles conservent des concentrations importantes à l'état bloqué, facilitant la génération d'un courant de fuite commandée par la tension de drain. Par contre à fort  $V_{GS}$ , la concavité et les concentrations sont très proches quel que soit le nombre de grilles ; les contrôles de charge dans tous ces dispositifs sont similaires à l'état passant.

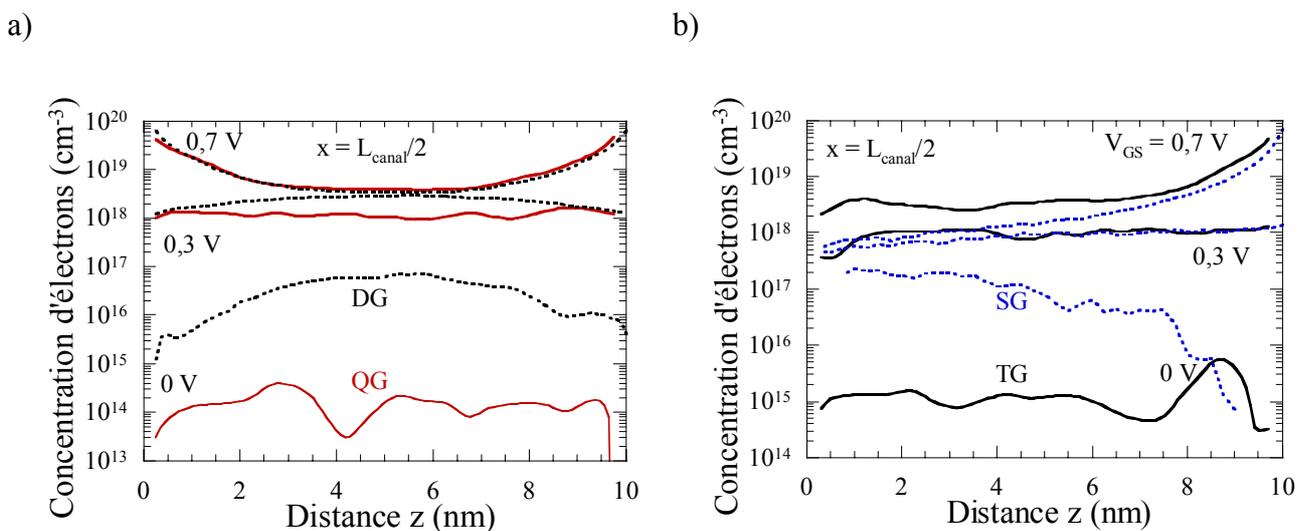


Figure 103 : Coupes transversales (axe  $z$ ) en  $x = L_{\text{canal}}/2$  de concentrations en électrons à différents  $V_{GS}$  pour différents transistors MOS avec  $L_{\text{canal}} = 15$  nm et  $T_{Si} = 10$  nm à faible  $V_{DS} = 50$  mV a) DG et QG, b) SG et TG.

Dans le QGMOS, les concentrations en électrons au milieu de la zone et sous l'oxyde sont identiques sauf à fort  $V_{GS}$ . On a donc un contrôle de grille quasi équivalent dans tout le volume de la zone active. Ces résultats sont cohérents avec l'étude du DIBL et confirment une amélioration du contrôle du canal avec le nombre de grilles. Comme indiqué en Figure 104, les dispositifs étroits ( $T_{Si} = 5$  nm) atteignent un contrôle de grille satisfaisant avec deux grilles, c'est-à-dire au moins aussi bon qu'avec quatre grilles et  $T_{Si} = 10$  nm.

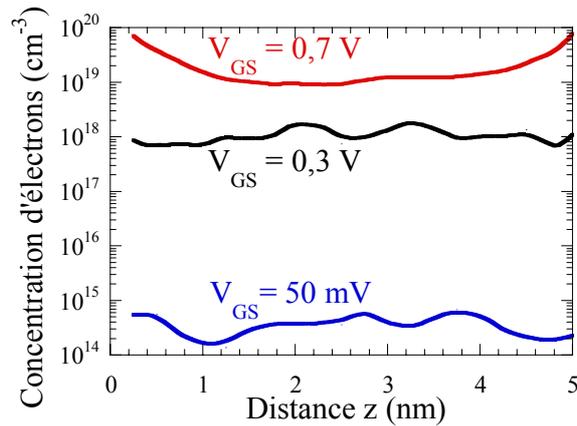


Figure 104 : Coupes transversales (axe  $z$ ) en  $x = L_{\text{canal}}/2$  de concentrations en électrons à différents  $V_{GS}$  pour un DGMOS avec  $L_{\text{canal}} = 15 \text{ nm}$  et  $T_{Si} = 5 \text{ nm}$  à faible  $V_{DS} = 50 \text{ mV}$ .

### 5.5 Bilan et rapports $I_{\text{on}}/I_{\text{off}}$

Les résultats précédents sont synthétisés en Figure 105. Tous les résultats par simulation de dispositifs simulés à simple, double, triple et quadruple grille avec  $L_{\text{canal}} = 15 \text{ nm}$  et  $T_{Si} = 5$  et  $10 \text{ nm}$  ainsi que d'autres, tels que  $L_{\text{canal}} = 25 \text{ nm}$  et  $T_{Si} = 10 \text{ nm}$ , sont représentés en fonction de leur courant à l'état passant  $I_{\text{on}}$  et à l'état bloqué  $I_{\text{off}}$ .

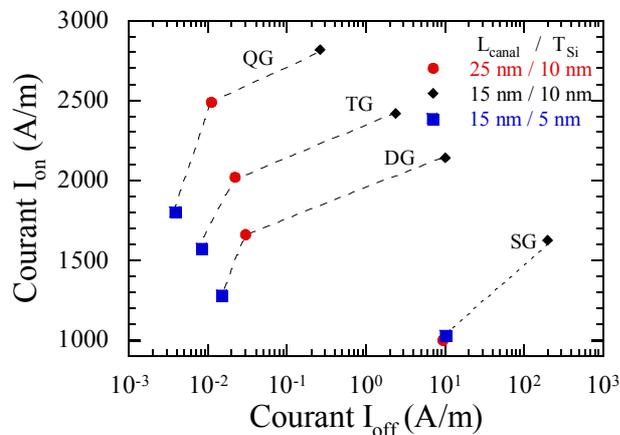


Figure 105 : Courant à l'état passant  $I_{\text{on}}$  en fonction du courant à l'état bloqué  $I_{\text{off}}$  pour différentes architectures SG, DG, TG et QGMOS :  $L_{\text{canal}} = 25 \text{ nm}$  avec  $T_{Si} = 10 \text{ nm}$  (ronds rouges),  $L_{\text{canal}} = 15 \text{ nm}$  avec  $T_{Si} = 10$  (losanges noirs) et  $5 \text{ nm}$  (carrés bleus).

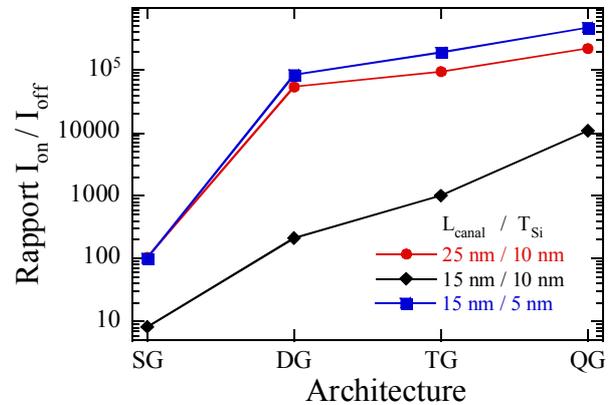


Figure 106 : Evolution du rapport  $I_{\text{on}}/I_{\text{off}}$  en fonction du nombre de grilles de l'architecture : SG, DG, TG et QGMOS :  $L_{\text{canal}} = 25 \text{ nm}$  avec  $T_{Si} = 10 \text{ nm}$  (ronds rouges),  $L_{\text{canal}} = 15 \text{ nm}$  avec  $T_{Si} = 10$  (losanges noirs) et  $5 \text{ nm}$  (carrés bleus).

L'augmentation du nombre de grilles améliore d'une part l'état passant  $I_{\text{on}}$  et d'autre part l'état bloqué des transistors. Ainsi, comme indiqué en Figure 106, le rapport  $I_{\text{on}}/I_{\text{off}}$  s'accroît quand le nombre de grilles augmente. Cependant, à partir de deux grilles, ces courbes tendent à saturer dans les dispositifs où le rapport  $L_{\text{canal}}/T_{Si}$  est égal à 2,5 ou 3. Ajouter des grilles à un transistor DG déjà bien dimensionné améliore peu les performances  $I_{\text{on}}/I_{\text{off}}$ .

La réduction de l'épaisseur de la zone active  $T_{Si}$  permet de réduire significativement les courants de fuite du transistor bloqué  $I_{off}$ . Cette réduction induit parallèlement une dégradation du courant  $I_{on}$ . Toutefois, le bilan est positif car comme l'indique la Figure 106 le rapport  $I_{on}/I_{off}$  tend à croître quand le rapport  $L_{canal}/T_{Si}$  augmente.

Il faut néanmoins souligner que les valeurs en A/m des courants  $I_{on}$  et  $I_{off}$  des transistors TG et QG dépendent de la normalisation du courant (cf. paragraphe 3. ). Les conclusions précédentes ne sont peut être pas générales ; notre étude doit donc être prolongée et s'affranchir des problèmes de normalisation.

## 6. Evolution du paramètre « CV/I »

Le paramètre « CV/I », défini au chapitre I (cf. page 30), a l'énorme avantage d'être indépendant de la normalisation utilisée : C correspond à la capacité de grille totale d'un dispositif,  $V = V_{DD}$  est la tension d'alimentation,  $I = I_{on}$  courant en ampère délivré par un transistor et la même normalisation est appliquée pour C et pour I. La Figure 107 représente les valeurs de « CV/I » en fonction du courant  $I_{off}$  pour tous les dispositifs simulés.

Pour une architecture donnée, si le rapport  $L_{canal}/T_{Si}$  augmente, cela induit une augmentation du délai « CV/I ». En effet, la diminution du courant  $I_{on}$  est plus rapide que la celle de la capacité de grille qui, elle, croit quasi linéairement avec le nombre de grilles, comme illustré en Figure 108.

Si l'on travaille à facteur de forme constant, l'augmentation du nombre de grilles génère aussi un allongement du temps de réponse. En effet, l'augmentation de la capacité de grille est quasi linéaire avec le nombre de grilles alors que celle du courant  $I_{on}$  ne l'est pas, comme on l'a vu dans les paragraphes précédents.

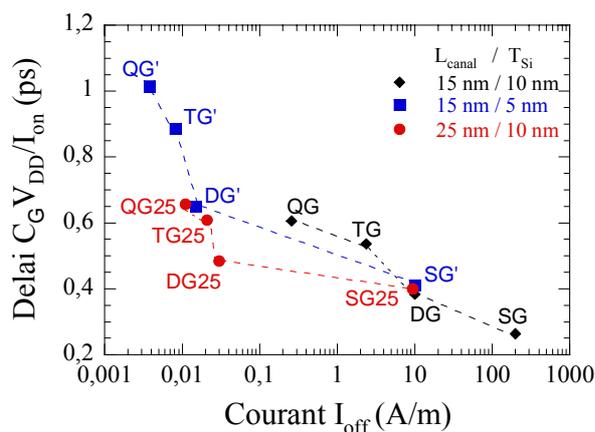


Figure 107 : Paramètre « CV/I » en fonction du courant à l'état bloqué  $I_{off}$  pour différentes architectures SG, DG, TG et QGMOS :  $L_{canal} = 25$  nm avec  $T_{Si} = 10$  nm (ronds rouges),  $L_{canal} = 15$  nm avec  $T_{Si} = 10$  (losanges noirs) et 5 nm (carrés bleus)..

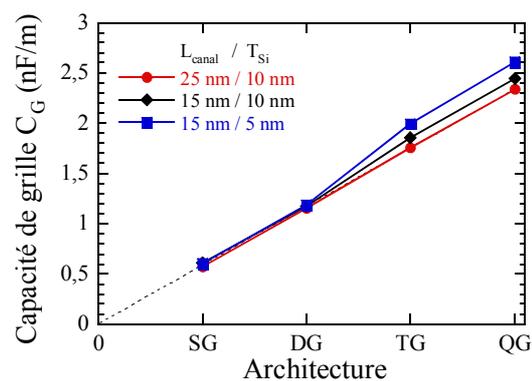


Figure 108 : Evolution de la capacité totale de grille  $C_G$  (évaluée à  $V_{GS} = V_{DD}$  et  $V_{DS} = 0$  V) en fonction du nombre de grilles de l'architecture : SG, DG, TG et QGMOS :  $L_{canal} = 25$  nm avec  $T_{Si} = 10$  nm (ronds rouges),  $L_{canal} = 15$  nm avec  $T_{Si} = 10$  (losanges noirs) et 5 nm (carrés bleus). NB : toutes les longueurs de grilles valent 25 nm !.

Un compromis entre favoriser le pouvoir bloquant des transistors ou la rapidité des applications logiques associées apparaît inéluctable. Dans ce contexte, le DGMOS ayant un rapport  $L_{\text{canal}}/T_{\text{Si}} = 2,5$  semble être un bon compromis : son courant  $I_{\text{off}}$  est convenable, inférieur à 0,1 A/m, et son délai « CV/I » (0,49 ps) reste compétitif par rapport au simple grille équivalent (0,40 ps).

## 7. Influences des paramètres géométriques

On a vu précédemment quel était le meilleur compromis entre nombre de grilles et rapport  $T_{\text{Si}}/L_{\text{canal}}$  à épaisseur d'oxyde donnée. Ce paragraphe s'intéresse au contrôle des impédances parasites et donc au dimensionnement de paramètres tels que la largeur du film  $W$  dans les architectures 3D, la longueur des zones d'accès  $L_{\text{accès}}$  et la longueur d'espaceur  $L_{\text{spac}}$  entre la grille et les prises de contact de source et drain.

### 7.1 Influence de la largeur du film

Est-il possible dans une architecture 3D de diminuer les marges sur le courant  $I_{\text{off}}$  pour accroître le courant  $I_{\text{on}}$  et qu'en conséquence la vitesse, c'est-à-dire le délai « CV/I », devienne acceptable ? Pour répondre à cette question, nous avons tracé en Figure 109 les évolutions de la pente sous le seuil  $S$  en fonction de  $W$  pour  $T_{\text{Si}} = 5$  et 10 nm. Les résultats fournis par les simulations Monte Carlo ont été comparés avec succès à ceux du modèle analytique de la référence [BOUR06]. De plus, les évolutions des paramètres  $I_{\text{off}}$  et « CV/I » dans un QGMOS en fonction de la valeur de  $W$  pour  $T_{\text{Si}} = 5$  nm sont reportées en Figure 110.

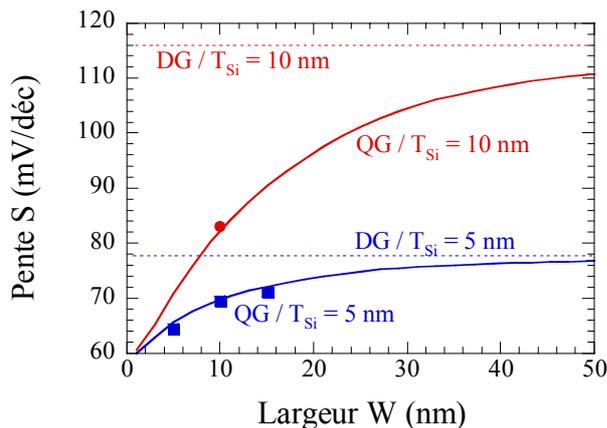


Figure 109 : Pente sous le seuil  $S$  en fonction de la largeur du transistor  $W$  pour différentes architectures MOS sur SOI avec  $L_{\text{canal}} = 15$  nm : simple (SG), double (DG), triple grille (TG) et quadruple grille (QG). Les courbes et les symboles sont respectivement issus d'un modèle analytique [BOUR06] et des simulations Monte Carlo.

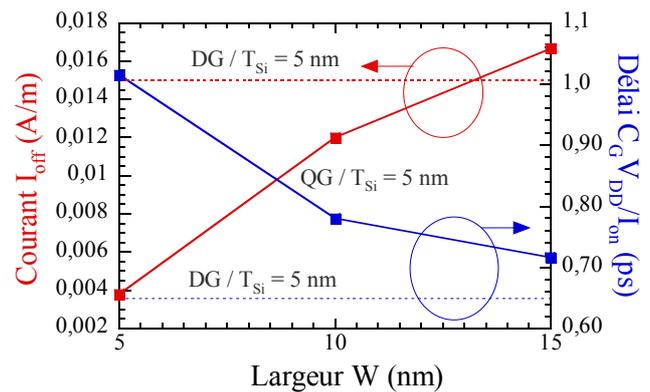


Figure 110 : Evolutions des paramètres  $I_{\text{off}}$  et « CV/I » (issus des simulations Monte Carlo) en fonction de la valeur de  $W$  dans un QGMOS avec  $T_{\text{Si}} = 5$  nm et  $L_{\text{canal}} = 15$  nm

Si  $W$  est doublé, on obtient une réduction de 20% du délai « CV/I » et les paramètres  $S$  et  $I_{off}$  restent convenables : la pente sous le seuil passe de 65 mV/dec à 69 mV/dec et le  $I_{off}$ , même multiplié par 3, reste inférieur à 12 mA/m. Pour un rapport  $W/T_{Si} = 3$ , on retrouve quasiment les caractéristiques sous le seuil d'un double grille et on perd donc l'intérêt de la structure quadruple grille.

Nous avons aussi étudié l'influence de  $W$  dans un TG. Pour  $T_{Si} = 5$  nm et  $W$  variant de 5 à 10 nm, « CV/I » décroît de 0,86 ps à 0,66 ps. Cependant  $I_{off}$  s'est dramatiquement accru ; il passe de 8 mA/m à 0,14 A/m. Il est à noter que dans un TG les meilleurs performances sont obtenues pour  $W$  inférieur à  $T_{Si}$ . Pour  $T_{Si} = 10$  nm et  $W = 5$  nm, « CV/I » et  $I_{off}$  sont respectivement égaux à 0,72 ps et 8,4 mA/m en normalisant par le périmètre de grille total :  $2 \times T_{Si} \times W$ . Cependant un DG de même épaisseur reste très compétitif : « CV/I » = 0,65 ps et  $I_{off} = 7,5$  mA/m en normalisant par le périmètre de grille total :  $2 \times T_{Si} \times W$ . En outre, si l'on utilise la normalisation habituelle par  $W$  le  $I_{off} = 42$  mA/m du TG est bien plus grand que celui du DG (15 A/m)...

## 7.2 Longueur des caissons et impédances parasites

### 7.2.1 Résistances d'accès

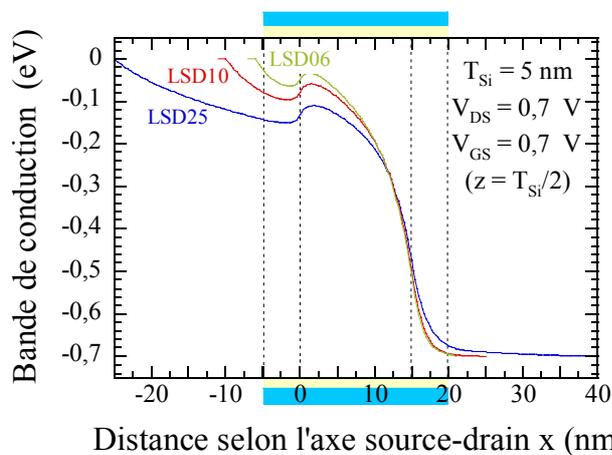


Figure 111 : Coupes longitudinales (axe  $x$ ) de la bande de conduction à  $V_{GS}=V_{DS}=0,7$  V du DGMOS avec  $L_{canal}=15$  nm,  $T_{Si}=5$  nm et  $N_D=5 \times 10^{19}$  at./cm<sup>3</sup> sous l'oxyde et au milieu de la zone active.

	Unité	DGMOS ( $L_{canal} = 15$ nm)					ITRS 2001
		$T_{Si}$	$L_{accès}$	$N_d$	$R_{accès}$		
$T_{Si}$	nm	10	5	5	5	5	
$L_{accès}$	nm	25	25	25	10	6	
$N_d$	cm <sup>-3</sup>	$5.10^{19}$	$5.10^{19}$	$2.10^{20}$	$5.10^{19}$	$5.10^{19}$	
$R_{accès}$	$\Omega \cdot \mu m$	68	112	34	54	32	140

Tableau 14 : Résistances de caissons  $R_{accès}$  extraites dans des DGMOS avec  $L_{canal}=15$  nm et  $T_{Si}=5$  nm pour différents dopages de caissons  $N_D$  et différentes longueurs de caissons  $L_{accès}$ .

Les évolutions des bandes de conceptions selon l'axe  $x$  dans des transistors à double grille amincis ( $T_{Si} = 5$  nm) pour différentes longueurs des caisson  $L_{accès}$  tracées en Figure 111 indiquent la présence d'une importante chute de potentiel dans le contact de source. On peut ainsi évaluer la résistance d'accès  $R_{accès}$  à 112  $\Omega \mu m$  (=  $\Delta V/I_{on}$  avec  $I_{on} = 1281$  A/m) sans prendre en compte les résistances de contacts alors que pour un DGMOS moins étroit ( $T_{Si} = 10$  nm) le même calcul donne  $R_{accès} = 68 \Omega \mu m$ . A titre indicatif, l'ITRS recommande une résistance totale inférieure à 140  $\Omega \mu m$ .

En conséquence, les résistances propres des caissons source et drain deviennent problématiques dans les films ultrafins. Néanmoins, dans ces dispositifs ultracourts, les caissons de drain semblent être beaucoup moins résistifs grâce aux électrons chauds (cf. chapitre II, page 77).

Comme l'indique les extractions de  $R_{\text{accès}}$  reportées dans le Tableau 14, pour réduire  $R_{\text{accès}}$  à épaisseur  $T_{\text{Si}}$  constante, on peut réduire la longueur des caissons  $L_{\text{accès}}$  et/ou augmenter le dopage des caissons  $N_D$ . Ces valeurs de  $R_{\text{accès}}$  révèlent un comportement de type ohmique dans la source. La résistance des caissons raccourcis  $L_{\text{accès}} = 6 \text{ nm}$  et dopés à  $N_D = 5 \times 10^{19} \text{ at.cm}^{-3}$  est égale à celle des caissons quasiment 4 fois plus longs  $L_{\text{accès}} = 25 \text{ nm}$  mais aussi 4 fois plus dopés  $N_D = 2 \times 10^{20} \text{ at.cm}^{-3}$ . La loi d'Ohm ( $R_{\text{accès}} = L_{\text{accès}} / (S \times N_D \times \mu)$  avec  $S = W \times T_{\text{Si}}$ ) est donc bien vérifiée.

Les effets néfastes de l'augmentation des  $R_{\text{accès}}$  sont nettement visibles sur les caractéristiques statiques des transistors simulés avec différents dopages  $N_D$  et longueurs  $L_{\text{accès}}$ . Les caractéristiques  $I_D(V_{\text{GS}})$ , tracées en Figure 112, indiquent que la tension de seuil est indépendante de  $L_{\text{accès}}$ . Cependant l'augmentation de  $L_{\text{accès}}$ , induisant celle de  $R_{\text{accès}}$ , provoque une diminution de la transconductance à fort  $V_{\text{DS}}$  (ou plutôt à fort courant quand la chute de potentiel dans les caissons est comparable à celle aux bornes du canal, cf. chapitre I, page 38). Cet accroissement a aussi pour effet de dégrader la pente à l'origine des courbes  $I_D(V_{\text{DS}})$  de la Figure 113 ainsi que le courant  $I_{\text{on}}$ .

Il est important de constater que l'augmentation du dopage  $N_D$  a permis de diminuer les résistances d'accès  $R_{\text{accès}}$ , mais en contrepartie elle favorise les effets de canal court : la diffusion des électrons des caissons dans le canal est plus importante. Cela se traduit sur la Figure 112, par la diminution du  $V_T$  (le potentiel de diffusion  $\phi_D$  des formules analytiques du chapitre I, page 35 augmente) et une dégradation du comportement sous le seuil. Sur la courbe  $I_D(V_{\text{DS}})$ , on constate une augmentation de la conductance de sortie.

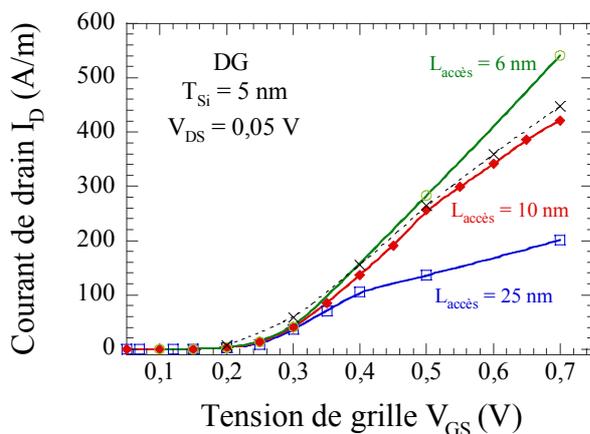


Figure 112 : Caractéristiques  $I_D(V_{\text{GS}})$  à  $V_{\text{DS}} = 50 \text{ mV}$  dans des différents DG MOS avec  $N_D = 5 \times 10^{19} \text{ at./cm}^{-3}$ ,  $L_{\text{canal}} = 15 \text{ nm}$ ,  $T_{\text{Si}} = 5 \text{ nm}$  et  $L_{\text{accès}} = 6, 10$  et  $25 \text{ nm}$ . En tirets caissons dopés à  $N_D = 2 \times 10^{20} \text{ at./cm}^{-3}$  et  $L_{\text{accès}} = 25 \text{ nm}$ .

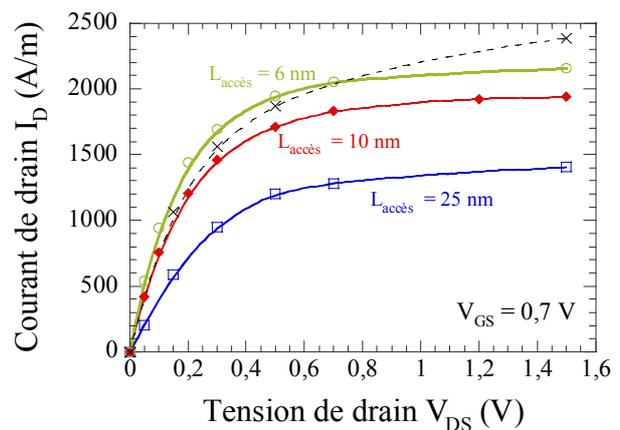


Figure 113 : Caractéristiques  $I_D(V_{\text{DS}})$  à  $V_{\text{GS}} = 0,7 \text{ V}$  dans des différents DG MOS avec  $N_D = 5 \times 10^{19} \text{ at./cm}^{-3}$ ,  $L_{\text{canal}} = 15 \text{ nm}$ ,  $T_{\text{Si}} = 5 \text{ nm}$  et  $L_{\text{accès}} = 6, 10$  et  $25 \text{ nm}$ . En tirets caissons dopés à  $N_D = 2 \times 10^{20} \text{ at./cm}^{-3}$  et  $L_{\text{accès}} = 25 \text{ nm}$ .

L'augmentation du dopage des caissons  $N_D$  à des niveaux très importants va être technologiquement délicate et, d'un point de vue électrique, favoriser les effets de canal court. Pour contrôler les résistances  $R_{\text{accès}}$ , il va donc falloir ajuster  $N_D$  (dans la mesure du possible) mais surtout  $L_{\text{accès}}$ .

### 7.2.2 Capacités de bord

Comme on vient de le voir, le dimensionnement des  $L_{\text{accès}}$  a une importance particulière dans les dispositifs ultrafins :  $L_{\text{accès}}$  doit être court pour conserver des  $R_{\text{accès}}$  modérées. Cependant la diminution de  $L_{\text{accès}}$ , c'est-à-dire de la longueur des espaceurs  $L_{\text{spac}}$  entre la grille et les prises de contact de source et drain, augmente dangereusement la capacité de bord  $C_{\text{bord}} = 2 \times \epsilon_{\text{spac}} \times T_{\text{MG}} \times L_{\text{spac}}$  (cf. Figure 11 du chapitre I, le facteur 2 tenant compte des  $C_{\text{bord}}$  du drain et de la source) avec  $T_{\text{MG}}$  épaisseur du métal de grille comme schématisé sur la Figure 114.

Le délai « CV/I » en considérant la capacité  $C_{\text{bord}}$  est tracé en Figure 115 en fonction de la longueur  $L_{\text{spac}}$  pour 2 types d'espaceurs :  $\text{SiO}_2$  et  $\text{Si}_3\text{N}_4$ . La diminution de  $L_{\text{accès}}$  pour diminuer les  $R_{\text{accès}}$  a pour effet d'augmenter les capacités de bord et donc le paramètre « CV/I ». Pour ne pas détériorer les performances dynamiques des circuits logiques, il ne faut donc pas trop réduire  $L_{\text{accès}}$ . Il existe un optimum pour le dimensionnement de  $L_{\text{accès}}$  ou plutôt une limite à ne pas dépasser : quel que soit le matériau de l'espaceur,  $L_{\text{spac}}$  doit rester supérieur à 5 nm dans ces dispositifs.

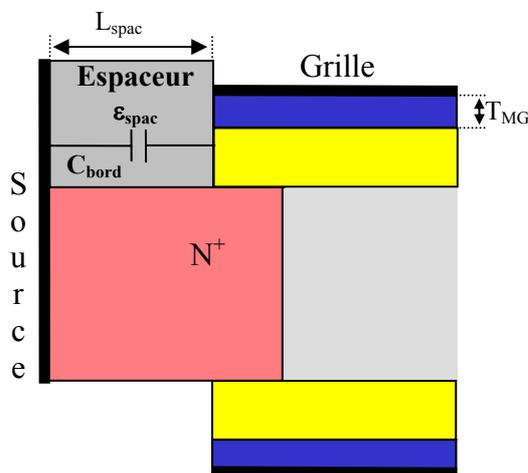


Figure 114 : Schéma de la capacité de bord  $C_{\text{bord}}$  avec  $\epsilon_{\text{spac}}$  permittivité de l'espaceur,  $L_{\text{spac}}$  longueur de l'espaceur,  $T_{\text{MG}}$  épaisseur du métal de grille.

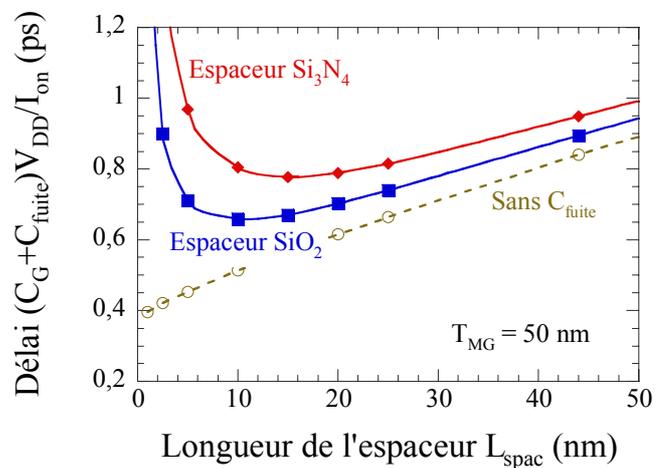


Figure 115 : Evolution du paramètre « CV/I » en fonction de la longueur de l'espaceur  $L_{\text{spac}}$  pour différents matériaux :  $\text{SiO}_2$ ,  $\text{Si}_3\text{N}_4$  comparée aux valeurs obtenues sans prise en compte de la capacité de bord  $C_{\text{bord}}$ .  $L_{\text{canal}} = 25 \text{ nm}$  avec  $T_{\text{Si}} = 5 \text{ nm}$ .

## 8. Bilan

Les architectures multigrilles améliorent clairement l'état bloqué des transistors. Néanmoins, les multigrilles sont intrinsèquement plus lents que les simple grille. Le courant  $I_{on}$  n'étant pas proportionnel au nombre de grilles, le gain en terme de pouvoir bloquant se paye en terme de rapidité des applications logiques associées, surtout pour les architectures 3D à section carrée  $T_{Si} = W$ .

Ce chapitre a permis de valider certains modèles analytiques utilisés dans le dimensionnement des architectures multigrilles. Les transistors à double grille sont apparus être de très bons compromis entre comportement sous le seuil et rapidité ; ils devraient succéder aux architectures simple grille pour les générations telles que  $L_{canal} \approx 20$  nm, prévues à l'horizon 2008.

En outre, pour contrôler les impédances parasites, ces dispositifs nécessitent une optimisation du dimensionnement des caissons ( $L_{spac}$  et  $N_D$ ) qui prennent de plus en plus d'importance à mesure que l'on tend vers les limites ultimes du MOS.

Nos simulations n'étant pas couplées à des simulations thermiques, nous n'avons pas intégré les problèmes de dissipation thermique. Des observations récentes n'ont toutefois pas reporté de dégradations sensibles. En effet, les dissipations par le dessus des composants semblent être suffisamment efficaces [LOLI05].



## ***Chapitre IV :***

### ***Prise en compte de la quantification perpendiculaire***

Le chapitre précédent a permis de mettre en évidence que les transistors à double grille, dont la réalisation a été permise par l'amélioration des technologies SOI, sont des dispositifs très prometteurs qui devraient succéder aux architectures « bulk » à l'horizon 2008. Grâce aux couplages électrostatiques entre les grilles de DGMOS suffisamment minces, tout le volume de la zone active se trouve en régime de forte inversion et participe au transport de charges. Grâce à ce régime d'inversion volumique (cf. chapitre I, page 45), les caractéristiques électriques sont très performantes :  $I_{\text{off}}$ , « CV/I »...

Pour assurer l'immunité des DGMOS vis-à-vis des effets de canal court, la réduction de la longueur de canal  $L_{\text{canal}}$  nécessite une diminution en parallèle de l'épaisseur de la zone active  $T_{\text{Si}}$ . En effet, le rapport  $L_{\text{canal}}/T_{\text{Si}}$  doit toujours rester supérieur à encore 2,5 (cf. chapitre III). L'épaisseur de canal  $T_{\text{Si}}$  des transistors ultimes de longueur  $L_{\text{canal}}$  décanométrique sera donc subdécanométrique. Ainsi, c'est essentiellement et d'abord dans cette direction, perpendiculaire au transport de charges, que les effets quantiques vont devenir conséquents ; le confinement des porteurs dans le puits de potentiel défini par les deux oxydes de grille ne sera pas négligeable. En conséquence, les porteurs de la couche d'inversion doivent être en toute rigueur considérés comme un gaz d'électrons confinés possédant seulement deux degrés de liberté dans l'espace réciproque (2Dk).

Pour expliciter les modifications du comportement électrostatique d'un composant ultrafin modélisé selon une approche « quantique » par rapport à une modélisation purement classique, nous avons mis en place et utilisé un algorithme 1D de résolution couplée des équations de Schrödinger et de Poisson. Ensuite, le transport d'un gaz électronique 2Dk soumis à un champ électrique constant a été étudié grâce à un simulateur Monte Carlo modifié dans ce but. Enfin, un simulateur de dispositifs couplé à l'équation de Poisson 2D avec prise en compte de la quantification perpendiculaire, MONACO 2Dk, a été développé au cours de cette thèse.

## 1. Influence de la quantification perpendiculaire

Pour étudier l'influence de la quantification dans les composants ultrafins, des capacités MOS à double grille ont été étudiées. Cette partie débute par une présentation de notre algorithme 1D de résolution couplée des équations de Schrödinger et de Poisson selon l'axe z, qui est la direction du confinement. Ensuite, les modifications par rapport aux modélisations purement classiques, observées grâce à ce logiciel, sont détaillées.

### 1.1 Résolution Schrödinger/Poisson 1D

#### 1.1.1 Modélisation physique du silicium

Structures de bandes :

On utilise des relations analytiques de dispersion en énergie  $E(k)$  pour décrire les extrema (vallées) de la structure de bandes du semiconducteur susceptibles d'être occupés par des porteurs.

La bande de valence du silicium, présentée en Figure 116, est modélisée de façon simple par deux vallées sphériques. Chacune est caractérisée par une masse effective  $m^*$  (trous lourds  $m^* = m_{hh} = 1,08 \times m_0$  et trous légers  $m^* = m_{lh} = 0,24 \times m_0$ , avec  $m_0$  masse des électrons libres) et éventuellement un coefficient de non-parabolicité  $\alpha$ . Les relations de dispersion associées sont de la

forme :  $E(1 + \alpha.E) = \frac{\hbar^2 k^2}{2.m^*}$ . Ne nous intéressant dans cette étude qu'aux structures de type N-MOS, cette représentation simple de la bande de valence nous suffira.

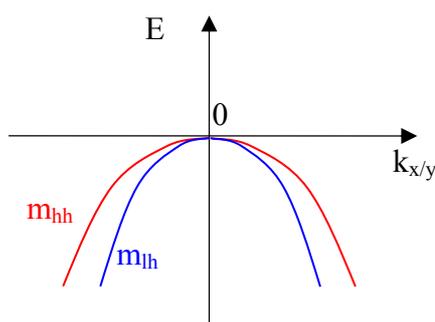


Figure 116 : Les bandes de valence analytiques du silicium. Bande de trous ( $m_{hh}$ ) en rouge et trous légers ( $m_{lh}$ ) en bleu.

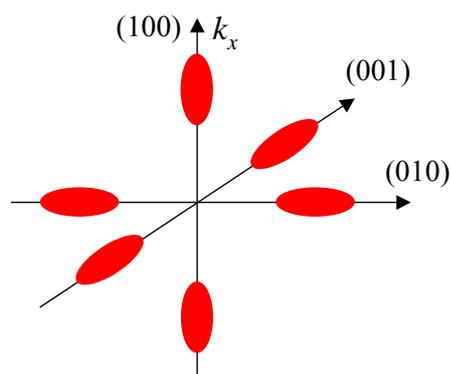


Figure 117 : Les 6 vallées ellipsoïdales (surfaces isoénergies) de la bande de conduction (décrite sous forme analytique) du silicium.

Le bas de la bande de conduction du silicium est décrit en considérant les 6 vallées ellipsoïdales  $\Delta$  centrées à 85% du bord de la zone de Brillouin (cf. Figure 117). La relation de dispersion pour un électron  $2Dk$  du niveau (appelé aussi sous-bande) numéro 'n' s'écrit :

$$E(1 + \alpha.E) = E_n + \frac{\hbar^2}{2} \left( \frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} \right)$$

où  $k_x$  et  $k_y$  sont les composantes du vecteur d'onde à partir du minimum du niveau,  $m_x$  et  $m_y$  celles des masses effectives et  $E_n$  l'énergie quantifiée (dans la direction  $z$ ) du niveau 'n'. En fonction de la vallée à laquelle appartient l'électron,  $m_x$  et  $m_y$  valent soit la masse transverse ( $m_t = 0,1905 \times m_0$ ), soit celle longitudinale  $m_l = 0,916 \times m_0$ . Le coefficient de non parabolicité est tel que  $\alpha = 0,5 \text{ eV}^{-1}$ .

Densité d'états

On a considéré des relations de dispersion paraboliques ( $\alpha = 0$ ) pour le calcul de la densité d'états 'DOS' en énergie du gaz 2Dk. On a ainsi pour chaque sous-bande :  $\text{DOS}_{2\text{Dk}}(E) = \frac{\sqrt{m_x m_y}}{\pi \hbar^2}$  avec  $m_x$  et  $m_y$  masses effectives dans les directions du transport (plan x,y).

1.1.2 *Principe de la résolution Schrödinger/Poisson 1D.*

Les différentes étapes nécessaires à la résolution de Schrödinger/Poisson 1D sont résumées sur la Figure 118.

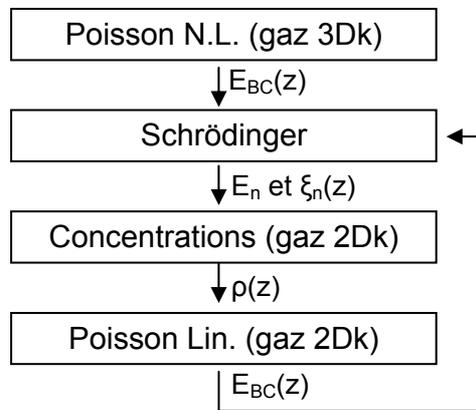


Figure 118 : Les séquences de la résolution couplée des équations de Schrödinger et de Poisson. 'N.L.' signifie que l'équation est non linéaire et 'Lin.' qu'elle est linéaire.

1) Détermination du potentiel initial

La première étape d'un tel algorithme passe par l'obtention d'un potentiel initial assez proche de la solution. Pour ce faire, on commence par résoudre une équation non linéaire dérivée de l'équation de Poisson 1D :  $\text{div}(\epsilon_{sc}(z) \times -\overrightarrow{\text{grad}}(V)(z)) = \rho(z)$  avec  $\rho(z)$  densité totale de charges,  $\epsilon_{sc}(z)$  permittivité du matériau et  $V(z)$  potentiel électrostatique.

$\rho(z)$  s'obtient en sommant les différentes contributions des charges fixes et des porteurs libres. Les ions donneurs d'électrons ont une densité de charges égale à  $+q.N_D(z)$  et les accepteurs à  $-q.N_A(z)$ . Il faut encore ajouter la densité d'électrons 3Dk (pour cette étape d'initialisation du potentiel nous considérons des porteurs non confinés) :  $-q.n(z)$ , et de trous 3Dk :  $+q.p(z)$ . Finalement, on écrit :

$$\rho(z) = q \times [N_D(z) - N_A(z) + p_{3D}(z) - n_{3D}(z)]$$

avec

$$N_D(z) = N_D \text{ (dans l'hypothèse de complète ionisation des donneurs)}$$

$$N_A(z) = N_A \text{ (dans l'hypothèse de complète ionisation des accepteurs)}$$

$$p_{3D}(z) = N_v \times \mathfrak{F}_{1/2} \left( \frac{E_{BV}(z) - E_F}{k_B T} \right)$$

$$n_{3D}(z) = N_c \times \mathfrak{F}_{1/2} \left( \frac{E_F - E_{BC}(z)}{k_B T} \right)$$

$$\text{où } N_{v/c} = 2 \left( \frac{2\pi m_{t+/e-}^* k_B T}{h^2} \right)^{3/2} \text{ et } \mathfrak{F}_{1/2}(\eta) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\varepsilon^{1/2}}{1 + e^{(\varepsilon - \eta)}} d\varepsilon$$

Dans ces expressions,  $E_F$  est le niveau de Fermi,  $E_{BC}$  et  $E_{BV}$  sont respectivement l'énergie de la bande de conduction et de valence, enfin  $m_{t+/e-}^* = \sqrt{m_x m_y}$  avec  $m_x$  et  $m_y$  masses effectives de trous ou d'électrons dans les directions du transport (plan x,y).

Les largeurs des bandes interdites ( $E_{BC} - E_{BV}$ ) des matériaux et les écarts de la bande de conduction  $\Delta E_{BC}$  et de valence  $\Delta E_{BV}$  par rapport au matériau de référence (par exemple le film de silicium) sont connus.

On peut donc réécrire l'équation de Poisson sous la forme :

$$\frac{\partial^2 V}{\partial z^2} - \frac{q}{\varepsilon_{sc}(z)} \left( N_v \times \mathfrak{F}_{1/2} \left( \frac{-V(z) + \Delta E_{BV}(z) - E_F}{k_B T} \right) - N_c \times \mathfrak{F}_{1/2} \left( \frac{E_F + V(z) + \Delta E_{BC}(z)}{k_B T} \right) + N_D - N_A \right) = 0$$

Il faut donc résoudre une équation non linéaire avec comme inconnue le potentiel  $V(z)$ . Pour cela l'équation est discrétisée par une méthode de différence finie, puis un algorithme itératif de type Newton/Raphson [NOUG93] est utilisé. Cette première résolution fournit la solution « classique » du problème : potentiel, concentration...

## 2) Résolution de l'équation de Schrödinger

Dans l'approximation de la masse effective, validée par la référence [SACC04] pour des énergies inférieures à 0,5 eV, l'équation de Schrödinger stationnaire 1D se résume à l'équation aux

valeurs propres de l'Hamiltonien 1D :  $H = -\frac{\hbar^2}{2.m_z} \frac{\partial}{\partial z^2} + E_{BC}(z)$  ( $m_z$  masse effective selon l'axe de confinement), c'est-à-dire à :  $H.\xi_n = E_n.\xi_n$

Dans cette équation,  $E_n$  et  $\xi_n$  correspondent respectivement à l'énergie et à la fonction d'onde associées au niveau numéro 'n'. En considérant les 6 vallées équivalentes du silicium (cf. Figure 115), on obtient deux masses effectives possibles dans le sens de confinement, soit  $m_z = m_l$  ou  $m_t$  ( $m_l > m_t$ ).

La normalisation de la fonction d'onde s'effectue sur l'intervalle  $[z_1, z_2]$ , appelé fenêtre ondulatoire ou tranche Schrödinger, c'est-à-dire  $\int_{z_1}^{z_2} |\xi_n(z)|^2 dz = 1$ . Hors de cet intervalle, les fonctions d'onde sont considérées comme nulles.

La recherche des valeurs et des vecteurs propres s'effectue numériquement grâce à un algorithme de type 'HQR' sur l'équation matricielle obtenue par différence finie. Des précisions sur le schéma numérique et la méthode de résolution sont fournies dans l'annexe page 183.

### 3) Détermination de la densité de charges $\rho$

La densité totale de charges  $\rho(z)$  s'obtient en sommant les différentes contributions :

$$\rho(z) = q \times [N_D(z) - N_A(z) + p_{2D}(z) - n_{2D}(z)]$$

avec (dans l'hypothèse de complète ionisation des impuretés dopantes)

$$N_D(z) = N_D$$

$$N_A(z) = N_A$$

Pour obtenir les concentrations des gaz 2Dk constitués uniquement d'électrons (les trous sont considérés comme étant classiques :  $p_{2D} = p_{3D}$ ), il faut sommer les contributions de chaque vallée 'val' (au nombre de six dans le silicium) et de chaque niveau d'énergie 'niv'. Ces contributions sont pondérées par la densité de probabilité de présence  $|\xi_i(z)|^2$  associée. Cela se traduit par les expressions suivantes :

$$n_{2D}(z) = \sum_{val} \sum_{niv} n_{val,niv} |\xi_{niv}(z)|^2 \quad \text{où } n_{val,niv} = \frac{k_B T \sqrt{m_x m_y}}{\pi \hbar^2} \text{Ln} \left( 1 + \exp \left( \frac{E_F - E_{val,niv}}{k_B T} \right) \right)$$

### 4) Résolution de l'équation de Poisson 1D (linéaire)

Une fois que la concentration  $\rho(z)$  est déterminée, il suffit de résoudre l'équation linéaire de Poisson :  $\text{div}(\epsilon_{sc}(z) \times -\overrightarrow{\text{grad}}(V)(z)) = \rho(z)$  avec  $V(z)$  comme inconnue.

Les 3 dernières séquences sont effectuées en boucle jusqu'à ce que la convergence du système soit atteinte. Cette convergence est testée en calculant la différence sur la somme des écarts sur le potentiel en chaque noeud entre chaque itération.

## 1.2 Nouvelle répartition des porteurs par vallée équivalente

Le logiciel de simulation de résolution de Schrödinger/Poisson a été utilisé sur une structure MOS, schématisée en Figure 119, dont l'épaisseur de la zone active de silicium  $T_{Si}$  vaut 5 nm. On obtient ainsi deux séries de valeurs propres, l'une associée aux vallées de masse longitudinale dans la direction de confinement et l'autre aux vallées de masse transverse. Ces résultats sont respectivement présentés en Figure 120.a et en Figure 120.b.

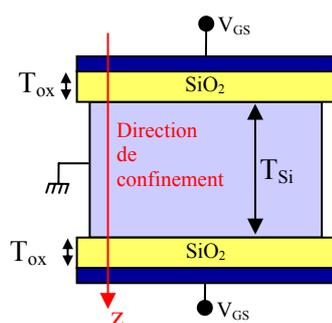


Figure 119 : Schéma d'une capacité MOS.

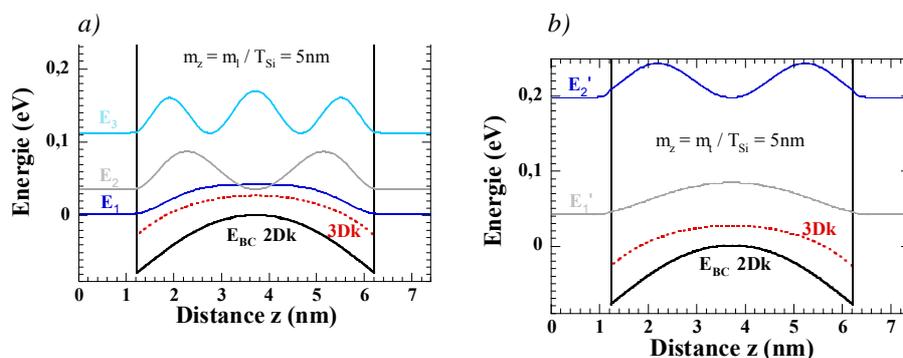


Figure 120 : Niveaux d'énergie et densités de présence pour les électrons de la capacité MOS avec  $T_{Si} = 5 \text{ nm}$ ,  $T_{ox} = 1,2 \text{ nm}$  à  $V_{GS} = 0,7 \text{ V}$ . Le niveau de Fermi est fixé à  $E_F = 0 \text{ eV}$ . a) Vallées « lourde »  $m_z = m_l$ . b) Vallées « légères »  $m_z = m_t$ . En pointillés rouges : le potentiel de confinement, en tirets : les niveaux d'énergie et en traits pleins : les densités de probabilité de présence.

A même polarisation de grille, la bande de conduction « quantique » est plus basse que la « classique ». De ce fait, le niveau de la première sous-bande  $E_1$  (cf Figure 120.a) se situe quasiment au niveau moyen de la bande de conduction classique. L'observation des probabilités de présence des électrons 2Dk révèle que plus le niveau d'énergie est élevé, plus l'électron a de chances de se rapprocher des interfaces. De plus, au sein de la zone active, le champ électrique est plus fort que dans le cas classique pour repousser les porteurs du centre du film et compenser les effets quantiques. Cependant, le champ électrique 2Dk à l'interface avec l'oxyde est plus faible que le 3Dk (l'impression visuelle sur la Figure 120 est trompeuse...), ce qui traduit que la charge dans une tranche Schrödinger est moins importante en 2Dk comme on le reverra par la suite.

Les niveaux d'énergie les plus bas sont bien évidemment les plus remplis. Dans ce puits de 5 nm d'épaisseur, le premier niveau d'énergie de chaque vallée est supérieur au maximum du fond du

puits. Par conséquent, nous ne sommes pas en présence de deux puits triangulaires couplés mais plutôt d'un puits quasi carré unique. Les fonctions d'onde associées ainsi que les probabilités de présence des électrons de ces niveaux seront donc maximales au milieu du film de silicium. Dans une capacité MOS à double grille d'épaisseur 5 nm polarisée à  $V_{GS} = 0,7$  V, l'inversion volumique est optimale.

En outre, le niveau le plus bas de toutes les vallées appartient aux vallées les plus lourdes dans la direction de confinement ( $m_z = m_l$ ). En conséquence, contrairement aux vallées équivalentes du silicium massif identiquement peuplées, les 2 vallées où  $m_z = m_l$  sont plus remplies dans un gaz d'électrons 2Dk. D'après la Figure 121, ce déséquilibre, faible dans des puits de 10 nm d'épaisseur (50% d'électrons dans les vallées où  $m_z = m_l$  au lieu de 33% dans le silicium massif), atteint son paroxysme dans les films de moins de 3 nm où seules les 2 vallées avec  $m_z = m_l$  sont peuplées.

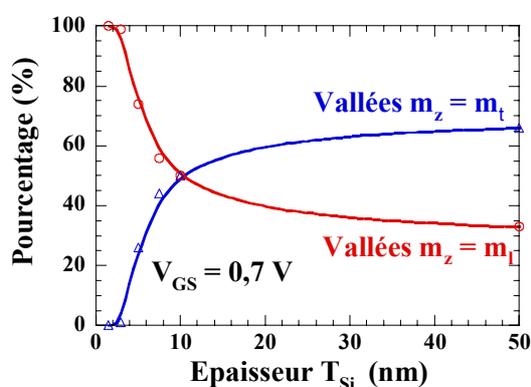


Figure 121 : Evolution de la répartition des porteurs sur les vallées équivalentes du silicium en fonction de l'épaisseur du film  $T_{Si}$  de capacités MOS avec  $T_{ox} = 1,2$  nm à  $V_{GS} = 0,7$  V. Insert : Evolution de la masse moyenne du gaz 2Dk dans le sens du transport (en pourcentage de  $m_0$ ) en fonction de  $T_{Si}$ .

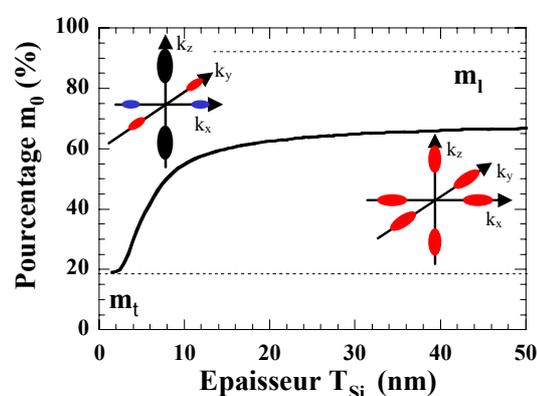


Figure 122 : Evolution de la masse moyenne du gaz 2Dk dans le sens du transport (en pourcentage de  $m_0$ ) en fonction de  $T_{Si}$ . Insert : Schématisation de la modification de la répartition des porteurs sur les 6 vallées  $\Delta$  du silicium avec (faible  $T_{Si}$ ) et sans (fort  $T_{Si}$ ) confinement selon l'axe z.

Cette modification de la répartition des porteurs sur les 6 vallées  $\Delta$  du silicium due au confinement, schématisée en Figure 122, a une influence sur la masse effective globale du gaz d'électrons. En effet, les vallées les plus remplies ( $m_z^* = m_l$ ) du film confiné, présentent dans le plan du transport (axe x et y) des masses transverses. La masse effective de conduction globale est donc réduite par le confinement ; elle tend vers la masse transverse  $m_t$  à mesure que l'on diminue l'épaisseur  $T_{Si}$ . Ce phénomène est très bénéfique pour le transport 2Dk.

### 1.3 Modification du contrôle électrostatique

Après l'étude des niveaux d'énergie et des fonctions d'onde, intéressons-nous maintenant à la modification de la répartition des porteurs entre l'étude classique et quantique.

Grâce aux simulations Schrödinger/Poisson de diverses structures MOS (cf. Figure 119), nous avons obtenu les évolutions des concentrations d'électrons présentées en Figure 123. Les résultats sont présentés pour diverses épaisseurs de film actif  $T_{Si}$  et sont comparés à ceux issus d'une modélisation purement classique, tracés en tirets (cf. étape 1 du paragraphe 1.1.2 ).

La concentration maximale « quantique » des électrons est réduite par rapport à ce que prédit un modèle entièrement classique. Elle est aussi décalée vers le milieu de la zone active. Cette influence des effets de quantification est d'autant plus importante que les grilles sont proches (faible  $T_{Si}$ ). Pour des épaisseurs  $T_{Si}$  inférieures à 5 nm, on observe même l'apparition d'un maximum unique localisé au milieu de la zone active. Grâce aux effets quantiques, l'inversion volumique est alors totale.

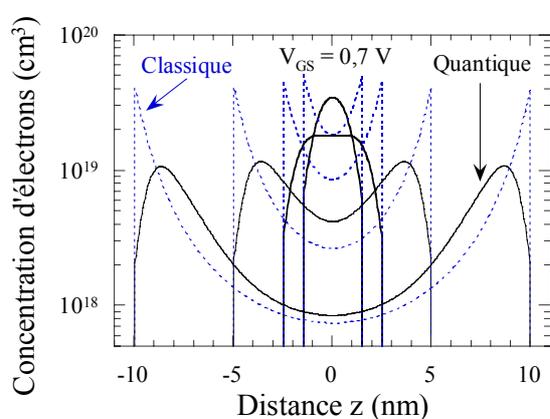


Figure 123 : Profils de concentration volumique d'électrons dans des capacités MOS à double grille de différentes épaisseurs  $T_{Si}$  : 3, 5, 10 et 20 nm à  $V_{GS} = 0,7$  V. En pointillés : calcul classique. En traits continus : calcul quantique.

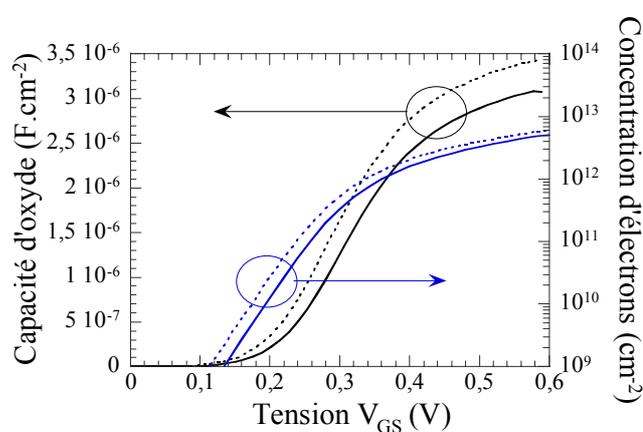


Figure 124 : Evolution de la concentration surfacique d'électrons et de la capacité de grille en fonction de la tension  $V_{GS}$  dans une structure MOS à double grille avec  $T_{Si} = 5$  nm et  $T_{ox} = 1,2$  nm. En pointillés : calcul classique. En traits continus : calcul quantique.

Ces modifications du profil de concentration entraînent celle de l'évolution de la charge d'inversion en fonction de la tension de grille, évolution tracée en Figure 124. Le confinement quantique vertical, éloignant les porteurs des électrodes de grille, réduit la capacité effective de commande dans le cas de film ultramince. Cela induit une augmentation de la tension du seuil d'inversion  $V_T$  [HARE96].

Pour résumer notre étude des capacités MOS, le confinement diminue sensiblement la masse effective moyenne du gaz d'électrons pour des dispositifs de moins de 10 nm d'épaisseur. De plus, l'inversion volumique est obtenue pour des puits dont l'épaisseur est inférieure à 5 nm environ. Tout cela semble être très favorable au transport électronique dans des films minces. Cependant, le confinement dans la direction du transport réduit le contrôle de la grille et favorise donc a priori les effets de canal court.

## 2. Etude du transport en gaz 2Dk

Pour étudier le transport d'un gaz électronique 2Dk, les résultats issus de l'algorithme Schrödinger/Poisson, concentrations et niveaux d'énergie, ont été intégrés dans un simulateur Monte Carlo de gaz d'électrons 2Dk sous champ constant.

La première partie de cette section décrira le simulateur de transport sous champ constant de type Monte Carlo. Notre modélisation des phénomènes d'interactions sera particulièrement détaillée. Enfin, nous étudierons les réponses d'un gaz 2Dk soumis à un champ électrique constant.

### 2.1 La méthode Monte Carlo à champ constant

Pour résoudre les équations de transport, la méthode Monte Carlo décompose la trajectoire d'une particule en un grand nombre de séquences (quelques milliers ou dizaines de milliers suivant les cas) [NOUG93]. Comme illustré en Figure 125.a, chaque séquence débute par un vol libre durant lequel la particule suit une trajectoire déterminée par le champ de force extérieur (loi de Newton). Ce vol libre est suivi d'une collision, ou interaction, du porteur avec son environnement. Pour générer de telles séquences aléatoires, il faut tirer au sort (cf. Figure 125.b) (i) la durée du vol libre  $t_v$ , (ii) le type d'interaction subie en fin de vol libre ainsi que (iii) son effet. Les lois de probabilité nécessaires, c'est-à-dire les différentes fréquences d'interaction possibles, sont déduites des lois de la physique.

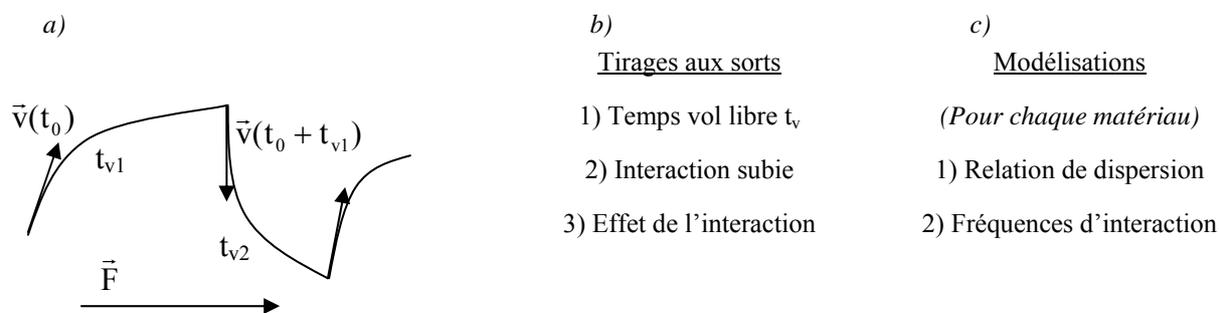


Figure 125 : Présentation de la méthode Monte Carlo. a) Une trajectoire d'un porteur dans l'espace réel.  $\vec{F}$  force extérieure appliquée.  $t_v$  temps de vol libre. b) Les différents tirages aux sorts nécessaires à la génération d'une telle trajectoire. c) Les modélisations physiques nécessaires.

Un grand nombre de particules est ainsi simulé, le mouvement d'ensemble donne les valeurs des caractéristiques macroscopiques : mobilité, énergie moyenne... Les simulations sont d'autant plus précises que le nombre de particules est important ou que le temps simulé est grand. Le choix de la représentation des relations de dispersion (ou diagrammes de bandes, cf. paragraphe 1.1.1 ) et des fréquences d'interaction constituent le cœur de la modélisation physique des simulations de transport (cf. Figure 125.c).

Nous commencerons par décrire la trajectoire durant le vol libre. Nous détaillerons ensuite les calculs des fréquences d'interaction.

### 2.1.1 Equation du mouvement

Il faut noter que dans les simulateurs à champ constant le mouvement n'est géré que dans l'espace réciproque, seuls les simulateurs de dispositifs traitent l'espace réel (cf. dernière partie). Par conséquent, le problème consiste simplement à déterminer l'évolution du vecteur d'onde  $\vec{k}(t)$  sous l'action des forces extérieures  $\vec{F}$  durant le vol libre, c'est-à-dire à résoudre l'équation fondamentale de la dynamique :

$$\frac{d\vec{k}(t)}{dt} = \frac{1}{\hbar} \frac{d\vec{p}(t)}{dt} = \frac{\vec{F}}{\hbar}$$

Si l'on considère le mouvement durant un pas sur le temps  $\delta t$ ,  $\vec{F}$  restant constant, le vecteur d'onde après mouvement  $\vec{k}(t + \delta t)$  est relié à sa valeur initiale par :  $\vec{k}(t + \delta t) = \vec{k}(t) + \frac{\delta t}{\hbar} \vec{F}$ .

## 2.2 Interactions avec le gaz 2Dk

### 2.2.1 Modélisation des fréquences d'interaction 2Dk

#### Expressions des fréquences d'interaction 2Dk

Les imperfections du cristal semiconducteur, telles que les vibrations des atomes ou la présence d'impuretés, perturbent le potentiel cristallin dans lequel les porteurs se déplacent. Cela se répercute sur leur mouvement. Ces perturbations, appelées interactions, modifient l'état du porteur qui les subit. Elles peuvent être considérées comme des événements instantanés à l'échelle des temps de vol libre, on peut donc utiliser la théorie des chocs. Soit  $\vec{k}$  et  $\vec{k}'$  les vecteurs d'onde avant et après interaction. Pour des porteurs 2Dk, ces vecteurs d'onde appartiennent au plan x,y (pas de composante selon l'axe de quantification z). La densité de probabilité de transition par unité de temps  $s(\vec{k}, \vec{k}')$ , pour un gaz 2Dk, est donnée par la règle d'or de Fermi [COHE77] :

$$s(\vec{k}, \vec{k}') = \frac{2\pi}{\hbar} |M(\vec{k}, \vec{k}')|^2 \cdot \text{DOS}_{2\text{Dk}}(\vec{k}') \cdot \delta(E - E') \quad \text{et} \quad M(\vec{k}, \vec{k}') = \langle \vec{k} | H_i | \vec{k}' \rangle = \iiint \Psi_b^*(\vec{r}, \vec{k}) \cdot H_i \cdot \Psi_b(\vec{r}, \vec{k}') d^3\vec{r}$$

où  $H_i$  est le potentiel de couplage de l'interaction,  $\text{DOS}_{2\text{Dk}}(\vec{k}')$  densité d'états dans l'espace réciproque, les  $\Psi_b$  correspondent respectivement aux fonctions d'onde de Bloch associées aux états

final et initial. Dans le cas d'un gaz 2Dk, on considère, dans l'approximation de la masse effective, des fonctions d'onde de Bloch du type :  $\Psi_b(n, \vec{r} = \vec{r}_{xy} + z.\vec{u}_z, \vec{k}_{xy}) \propto u_{\vec{k}}(\vec{r}).\xi_n(z).e^{i\vec{k}.\vec{r}_{xy}}$  où  $u_{\vec{k}}(\vec{r})$  est la fonction enveloppe de Bloch et  $\xi_n(z)$  est la fonction d'enveloppe du niveau d'énergie numéro n. E et E' sont les énergies du porteur avant et après interaction.

Finalement, on obtient :

$$s(\vec{k}, \vec{k}') = \frac{1}{2\pi\hbar} \left| \iint e^{-i(\vec{k}-\vec{k}')\cdot\vec{r}_{xy}} \times \left( \int \xi_m^*(z).\xi_n(z).H_i.dz \right) d^2\vec{r}_{xy} \right|^2 \times \delta(E-E')$$

La fréquence d'interaction est obtenue par intégration de  $s(\vec{k}, \vec{k}')$  sur tous les états finaux possibles, c'est-à-dire les valeurs possibles de  $\vec{k}'$  :

$$\lambda(\vec{k}) = \iint s(\vec{k}, \vec{k}') d^2\vec{k}' = \iint s(\vec{k}, \vec{k}') J(E', \theta'). dE'. d\theta'.$$

Dans cette expression E' est l'énergie finale (après interaction),  $\theta$  est l'angle horaire, J(E,  $\theta$ ) est le Jacobien de transformation entre un repère cartésien ( $k_x', k_y'$ ) et un repère cylindrique (E(k'),  $\theta'$ ). En gaz 2Dk, dans le cas d'une vallée ellipsoïdale non parabolique, les différentes composantes  $k_x'$  et  $k_y'$  du vecteur d'onde final  $\vec{k}'$  sont données par les expressions suivantes (cf. Figure 126) :

$$k_x' = K_{\max} \times \cos(\theta')$$

$$k_y' = (m_y/m_x)^{1/2} \times K_{\max} \times \sin(\theta')$$

avec

$$K_{\max}^2 = (k_x^2 + m_x/m_y \times k_y^2)$$

$\theta'$  ( $= \theta_0 + \beta$ ) angle entre  $\vec{k}'$  et l'axe x

Dans ce repère, le Jacobien J(E',  $\theta'$ ) a pour expression [MONS02] :

$$J(E', \theta') = \frac{1}{\hbar^2} \left[ \frac{\cos^2 \theta'}{m_x'} + \frac{\sin^2 \theta'}{m_y'} \right]^{-1} (1 + 2.\alpha.E')$$

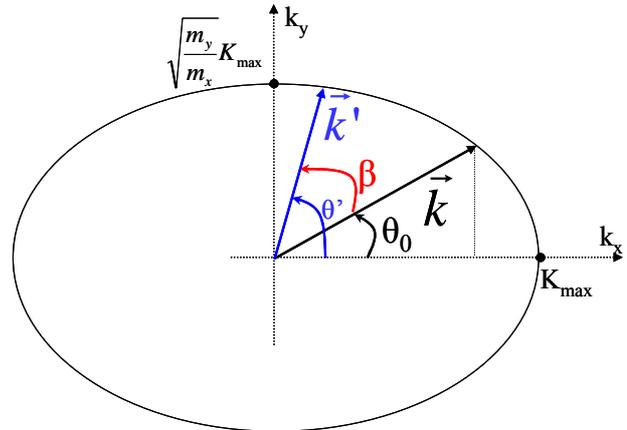


Figure 126 : Représentation du vecteur d'onde final  $\vec{k}'$  et du vecteur initial  $\vec{k}$  dans le cas d'une interaction élastique entre deux vallées elliptiques.

Finalement, les fréquences d'interaction 2Dk peuvent se mettre sous la forme :

$$\lambda_{2Dk}(\vec{k}) = \frac{(1 + 2.\alpha.E')}{2\pi\hbar^3} \int_0^{2\pi} \iint \left( \int \xi_m^*(z).\xi_n^*(z).H_i.dz \right) e^{-i(\vec{k}-\vec{k}')\cdot\vec{r}} \times d^2\vec{r} \left[ \frac{\cos^2(\beta + \theta_0)}{m_x'} + \frac{\sin^2(\beta + \theta_0)}{m_y'} \right]^{-1} d\beta$$

où  $\beta$  est l'angle entre  $\vec{k}'$  et  $\vec{k}$ ,  $m_x'$  et  $m_y'$  les masses effectives dans la vallée finale.

### Effet des interactions

Connaissant la fréquence de l'interaction, il faut ensuite s'intéresser à ses effets. De manière générale, la norme (dans le cas d'interactions inélastiques) et/ou la direction du vecteur d'onde peuvent être modifiées, comme illustré en Figure 126. Pour décrire l'effet d'une interaction, il faut connaître la valeur de la norme du vecteur d'onde final  $|\vec{k}'|$  liée à la valeur de l'énergie finale  $E'$ , et la densité de probabilité  $P_E(\beta)$  pour que l'angle de déviation du vecteur d'onde vaille  $\beta$  (cf. Figure 126).

Dans le cas d'une interaction à **déviatio**n isotrope ( $P_E(\beta)$  constant) vers une vallée **sphérique**, ou rendue isotrope par l'utilisation du temps de relaxation, on obtient (cf. annexe)  $P_E(\beta) = \frac{1}{2\pi}$ . Pour considérer une déviation isotrope dans une vallée **ellipsoïdale** ( $P_E(\beta) \neq 1/2\pi$ ), on peut se ramener au cas uniforme (vallée sphérique) en utilisant la transformation d'« Herring-Vogt » (cf. [MONS02]). C'est l'approche utilisée, par exemple, pour les interactions avec les phonons.

Pour traiter le cas des interactions à déviation anisotrope, il faut en toute rigueur effectuer un tirage au sort selon une loi de densité de probabilité non uniforme (cf. annexe). Cependant, on préfère en pratique, pour ne pas alourdir les calculs numériques, gérer des interactions à déviations isotropes. Grâce à la notion de « temps de relaxation du moment  $\tau(\vec{k})$  », on transforme les interactions élastiques **anisotropes** en **interactions fictives isotropes** [HEST84]. Cet artifice moyenne les effets d'une interaction anisotrope sur un grand nombre d'interactions. Cela fournit un temps de relaxation moyen ayant le même effet statistique qu'une interaction fictive isotrope telle que «  $\lambda(\vec{k}) = 1/\tau(\vec{k})$  ». Pour évaluer le temps de relaxation, on moyenne les interactions possibles en accordant un poids plus important aux interactions diffusantes dans des directions éloignées de la direction incidente, c'est-à-dire aux déviations  $\beta$  proches de  $\pi$  :

$$\frac{1}{\tau(\vec{k})} = \int s(\vec{k}, \vec{k}') \left(1 - \frac{\vec{k} \cdot \vec{k}'}{|\vec{k}'||\vec{k}|}\right) d^2\vec{k}' = \int s(\vec{k}, \vec{k}') \times (1 - \cos\beta) d\beta \approx \lambda_{\text{eq}}(\vec{k}) \quad (\text{avec } \beta \text{ angle entre } \vec{k}' \text{ et } \vec{k})$$

Cette approche est souvent utilisée pour traiter les interactions avec les impuretés dans les calculs de mobilité.

### Tirages au sort

Pour résumer, les grandeurs  $\lambda(\vec{k})$  et  $P_E(\beta)$  sont à la base de la procédure Monte Carlo. Ces densités de probabilité vont permettre d'effectuer les 3 types de tirages au sort nécessaires.

Soit  $r_1$ ,  $r_2$  et  $r_3$  trois variables aléatoires uniformes entre 0 et 1 dont les tirages sont indépendants.

- Le temps de vol libre vaut  $t_v = \frac{-\ln r_1}{\lambda_{tot}}$
- Le numéro 'i' de l'interaction est égal au plus petit entier 'i' tel que :  $r_2 > \frac{\sum_{n=1..i} \lambda_n}{\lambda_{tot}}$
- L'angle de déviation s'identifie à  $\beta = 2.\pi.r_3$ .

Le  $\lambda_n$  correspond à la fréquence de l'interaction numérotée 'n' et  $\lambda_{tot} = \sum_{n \in \mathbb{N}} \lambda_n$ .

Des précisions sur la méthode de tirage au sort et la détermination du temps de vol libre sont données en annexe.

## 2.2.2 Interactions avec le réseau cristallin (phonons)

Les expressions des fréquences d'interaction avec les phonons utilisées dans ce travail reprennent celles de la thèse de Florian Monsef [MONS02]. Elles correspondent aux interactions d'un gaz d'**électrons entièrement 2Dk** (pas de porteurs 3Dk étant donné le caractère ultramince des couches étudiées) dans le **silicium**. Toutes les variables utilisées sont exprimées en unité du système international (SI).

### a) Phonons acoustiques de faible énergie

#### Fréquence de l'interaction

On montre que la probabilité de transition d'un électron 2Dk d'une sous-bande 'm' vers une sous-bande 'n' par l'intermédiaire d'un phonon acoustique de faible énergie est telle que :

$$\lambda_{m,n}^{ac}(E) = \frac{k_B \cdot T \cdot D_{ac}^2}{2\hbar^3 \cdot \rho \cdot v_p^2} \sqrt{m'_x m'_y} (1 + 2 \cdot \alpha \cdot E') \int_{z_1}^{z_2} \xi_n^2(z) \xi_m^2(z) dz \quad (\text{SI})$$

Avec  $D_{ac}$  : potentiel de déformation,  $\rho$  : densité volumique du matériau,  $v_p$  : vitesse de propagation d'une onde longitudinale (phonon longitudinal) ou transverse (phonon transverse). En pratique la vitesse de propagation considérée est une vitesse moyenne entre les deux composantes. Les valeurs des différentes grandeurs sont reportées dans la référence [DOLL99]. E et E' sont respectivement les énergies cinétiques du porteur avant et après interaction.

Les transitions seront d'autant plus probables que les intégrales de recouvrement des fonctions d'enveloppe seront importantes. En conséquence, les transitions intra sous-bandes sont les plus probables.

### Effets de l'interaction

Les phonons sont de faible énergie, l'interaction est donc élastique :  $E' = E - (E_n - E_m)$  avec  $E_n$  et  $E_m$  qui sont respectivement les énergies associées aux niveaux  $n$  et  $m$  (la norme de  $\vec{k}$  est à déterminer grâce à la relation de dispersion  $k = f(E')$ ). De plus, on suppose que la déviation du vecteur d'onde lors de l'interaction est isotrope.

#### b) Phonons intervallées d'ordre 0

Les phonons considérés ici peuvent être induits par des modes de vibrations acoustiques ou optiques du cristal. Les vecteurs d'onde des phonons mis en jeu sont toujours importants et leur module est environ égal à la distance entre les minima des vallées initiales et finales. Dans le cadre de notre étude, on considère que les transitions  $2Dk/2Dk$  sont causées par des phonons  $3Dk$ . Cela est cohérent avec des résultats expérimentaux à température ambiante [SONG97]. Finalement, six phonons intervallées  $\Delta$ - $\Delta$  sont pris en compte : trois d'ordre 0 et, comme on le verra par la suite, trois d'ordre 1 [DOLL99].

### Fréquence de l'interaction

Le potentiel de déformation  $D_0$  étant supposé constant et indépendant de la valeur du vecteur d'onde du phonon, la probabilité de passage d'un électron  $2Dk$  d'énergie cinétique  $E$  d'une sous-bande 'm' vers une sous-bande 'n' par l'intermédiaire d'un phonon intervallée d'ordre 0 est telle que :

$$\lambda_{m,n}^{iv0}(E) = \frac{Z_{iv0} \cdot D_0^2}{2\hbar \cdot \rho} \frac{\left[ N_q + \frac{1}{2} + \frac{\sigma}{2} \right]}{[\hbar\omega]} \sqrt{m'_x m'_y} (1 + 2\alpha \cdot E') \int_{z_1}^{z_2} \xi_n^2(z) \xi_m^2(z) dz \quad (SI)$$

Avec

$\rho$  : densité volumique du matériaux,

$N_q$  : nombre moyen de phonons d'énergie  $\hbar\omega$  à la température  $T = 300$  K,

$m'_x$  et  $m'_y$  : masses effectives dans la vallée finale,

$E' = E - \sigma \times \hbar\omega - (E_n - E_m)$  : énergie cinétique finale après émission ( $\sigma = 1$ ) ou absorption ( $\sigma = -1$ ) de phonons.  $E_n$  et  $E_m$  sont respectivement les énergies associées aux niveaux  $n$  et  $m$ .

Le nombre de vallées d'arrivée possibles  $Z_{iv0}$  dépend du type de processus de transition (f ou cf. [MONS02]). Enfin, comme pour les phonons acoustiques intravallés, la présence de l'intégrale

des carrés des fonctions d'enveloppe pondère la probabilité de transition en fonction du recouvrement des probabilités de présence.

### Effets de l'interaction

L'interaction est inélastique :  $E' = E - \sigma \times \hbar\omega - (E_n - E_m)$  et la déviation isotrope.

c) Phonons intervallées d'ordre 1

### Fréquence de l'interaction

La probabilité de passage d'un électron 2Dk d'énergie cinétique E d'une sous-bande 'm' vers une sous-bande 'n' par l'intermédiaire d'un phonon intervallée d'ordre 1 est donnée par l'expression suivante (SI) :

$$\lambda_{m,n}^{iv1}(E) = \frac{Z_{iv1} \cdot D_1^2}{\hbar \cdot \rho} \frac{\left[ N_q + \frac{1}{2} + \frac{\sigma}{2} \right]}{[\hbar\omega]} \sqrt{m'_x m'_y} \times (1 + 2 \cdot \alpha \cdot E') \times \left[ \frac{1}{\hbar^2} \left( \gamma(E) \sqrt{m_x m_y} + \gamma(E') \sqrt{m'_x m'_y} \right) \cdot F_{mn} - \frac{G_{mn}}{2} \right]$$

Avec

$$\gamma(E) = (1 + 2 \cdot \alpha \cdot E),$$

$$F_{mn} = \int_{z_1}^{z_2} \xi_n^2(z) \xi_m^2(z) dz,$$

$$G_{mn} = \int_{z_1}^{z_2} \xi_n(z) \xi_m(z) \frac{\partial^2 [\xi_n(z) \xi_m(z)]}{\partial z^2} dz,$$

$D_1$  : potentiel de déformation,

$\rho$  : densité volumique du matériau,

$N_q$  : nombre moyen de phonons d'énergie  $\hbar\omega$  à une température  $T = 300$  K,

$Z_{iv1}$  : nombre de vallées d'arrivée possibles.

$m_x$  et  $m_y$  : masses effectives dans la vallée initiale,

$m'_x$  et  $m'_y$  : masses effectives dans la vallée finale,

$E' = E - \sigma \times \hbar\omega - (E_n - E_m)$  : énergie cinétique finale après émission ( $\sigma = 1$ ) ou absorption ( $\sigma = -1$ ) de phonons.  $E_n$  et  $E_m$  sont respectivement les énergies associées aux niveaux n et m.

Pour ce type d'interactions intervallées, les masses et énergies des vallées initiale et finale sont présentes dans l'expression de la probabilité d'interactions. De plus, une nouvelle intégrale de recouvrement  $G_{mn}$  apparaît.

### Effets de l'interaction

De même que pour les phonons intervallées d'ordre 0, l'interaction est inélastique et la déviation isotrope.

### 2.2.3 Interactions avec les impuretés ionisées

La diminution de l'épaisseur des films de silicium pour lutter contre les effets de canal court rend nécessaire l'augmentation du dopage des caissons source/drain pour limiter l'augmentation des résistances d'accès (cf. chapitre III). Les impuretés ionisées sont donc des centres de déviations inévitables dans les caissons de source et de drain des dispositifs ultrafins.

Les interactions électron-ion sont en toute rigueur anisotropes cependant, en pratique, l'utilisation d'un temps de relaxation permet de gérer une interaction équivalente qui est isotrope. Cette approximation est justifiée dans le calcul de mobilité et dans le cas des transistors que nous étudions (DGMOS à canal non dopé) par le nombre important d'interactions dans les zones dopées de source et de drain.

### Fréquence de l'interaction

Le temps de relaxation associé à la transition d'un porteur d'une sous-bande  $n$  vers une sous-bande  $m$  est donné par l'expression suivante (cf. annexe) :

$$\frac{1}{\tau_{m,n}^{\text{imp}}(\vec{k})} = \frac{q^4(1+2\alpha E')}{8\pi\hbar^3\epsilon_{sc}^2} \int_0^{2\pi} \frac{(1-\cos\beta)}{\frac{\cos^2(\beta+\theta_0)}{m'_x} + \frac{\sin^2(\beta+\theta_0)}{m'_y}} \int_{z_1}^{z_2} N_{\text{imp}}(z_0) \left| \int_{z_1}^{z_2} \frac{\xi_m(z) e^{-\Delta k(\beta,\theta_0)|z-z_0|} \xi_n(z)}{\Delta k(\theta) + \Delta k_{\text{scr}}(\theta)} dz \right|^2 dz_0 d\beta$$

avec

$$\Delta k^2(\beta, \theta_0) = |\vec{k}' - \vec{k}|^2 = (k'_x - k_x)^2 + (k'_y - k_y)^2 \quad (\beta \text{ angle de déviation et } \theta_0 \text{ angle entre } \vec{k} \text{ et l'axe } x),$$

$E, E'$  : respectivement énergie cinétique initiale et finale

$\epsilon_r$  : permittivité diélectrique du matériau,

$m_x$  et  $m_y$  : masses effectives dans la vallée,

$N_{\text{imp}}(z_0)$  : densité d'impuretés dopantes ( $\text{at./m}^3$ ) à la profondeur  $z_0$ ,

Le modèle d'écrantage employé est celui de Debye-Hückel (déjà utilisé pour les électrons 3Dk dans MONACO classique) revisité par Fetter [FETT74] dans le cadre bidimensionnel. L'inverse de la longueur d'écran vaut  $\Delta k_{scr} = \frac{q^2 \times N_{scr} n_{2D}}{2\epsilon_r k_B T} g_1\left(\frac{\Delta k \times \lambda_{th}}{4\sqrt{\pi}}\right)$  avec  $g_1(x) = \frac{1}{x} e^{-x^2} \int_0^x e^{-t^2} dt$ . La longueur

d'onde thermique est telle que :  $\lambda_{th} = \sqrt{\frac{2\pi\hbar^2}{\sqrt{m_x m_y} k_B T}}$ . Dans ces expressions  $N_{scr} \times n_{2D}$  correspond à la

fraction  $N_{scr}$  (sans dimension) de la concentration totale d'électrons  $n_{2D}$  qui contribue à l'écrantage. En première approximation, on ne considère généralement que la contribution de la première sous-bande. En pratique,  $N_{scr}$  est ajusté, comme nous le verrons plus loin, pour adapter l'effet des impuretés ionisées sur la mobilité.

### Effets de l'interaction

L'interaction est élastique  $E' = E - (E_n - E_m)$  avec  $E_n$  et  $E_m$  qui sont respectivement les énergies associées aux niveaux  $n$  et  $m$ , et isotrope grâce à l'approximation du temps de relaxation.

#### 2.2.4 Interactions avec les interfaces rugueuses

Les interfaces silicium/oxyde de grille, schématisées en Figure 127, ne sont pas lisses et l'épaisseur effective de la zone active des dispositifs multigrilles fluctue. Ces défauts d'interfaces modifient sensiblement le transport des porteurs à proximité des zones rugueuses.

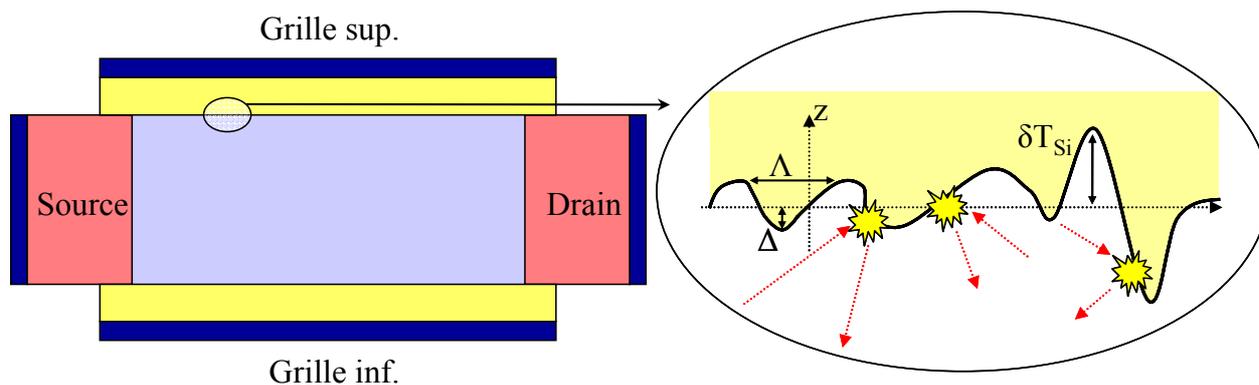


Figure 127 : Schématisation de l'interface rugueuse Si/SiO<sub>2</sub> dans un transistor à double grille et de l'interaction électron/interface rugueuse.

### Modélisation des surfaces rugueuses

Pour modéliser les interfaces rugueuses, la théorie des signaux aléatoires est très pertinente. La position de l'interface autour de sa valeur moyenne, considérée comme aléatoire, est décrite par des lois statistiques qui sont caractérisées par leur fonction d'auto-corrélation [COUL98]  $C_{\delta T_{Si}}(\Delta \vec{r}_{xy}) = E(\delta T_{Si}(\vec{r}_{xy} + \Delta \vec{r}_{xy}) \times \delta T_{Si}(\vec{r}_{xy}))$  avec  $E$  espérance mathématique,  $\vec{r}_{xy}$  coordonnées dans le

plan (x,y) de l'interface rugueuse et  $\delta T_{Si}(\vec{r}_{xy})$  écart de l'interface par rapport à sa position moyenne selon l'axe z.

Plusieurs modèles aléatoires ont été utilisés dont principalement :

- le modèle gaussien :  $C_{\delta T_{Si}}(\vec{r}_{xy}) = \Delta^2 e^{-(|\vec{r}_{xy}|/\Lambda)^2}$ ,
- le modèle exponentiel :  $C_{\delta T_{Si}}(\vec{r}_{xy}) = \Delta^2 e^{-|\vec{r}_{xy}|/\Lambda}$

avec  $\Delta$  écart type et  $\Lambda$  longueur caractéristique. Les fonctions d'auto-corrélation expérimentales  $C_{\delta T_{Si}}(\vec{r}_{xy}) = \delta T_{Si}(\vec{r}_{xy} + \Delta \vec{r}_{xy}) * \delta T_{Si}(\Delta \vec{r}_{xy})$ , convolutions des profils mesurés  $\delta T_{Si}(\vec{r}_{xy})$  (signaux déterministes), sont plus proches du modèle exponentiel [GOOD85], cependant seul le modèle gaussien fournit des expressions analytiques des fréquences d'interaction.

Notre calcul des fréquences d'interaction étant numérique, nous avons utilisé une loi exponentielle dont la densité spectrale de puissance (transformée de Fourier de la fonction d'autocorrélation) vaut :

$$\overline{C_{\delta T_{Si}}(\Delta \vec{k}_{xy})} = \iint C_{\delta T_{Si}}(\vec{r}_{xy}) \times e^{-i\Delta \vec{k}_{xy} \cdot \vec{r}_{xy}} d\vec{r}_{xy} = \frac{2\pi\Delta^2\Lambda^2}{\left(1 + |\Delta \vec{k}_{xy}|^2 \Lambda^2\right)^{3/2}}$$

Dans le cas de double grille et en supposant les rugosités de chaque interface indépendantes, les densités spectrales de chaque grille s'additionnent.

### Modélisation des interactions électron/surface rugueuse

Différentes approches ont été mises en œuvre pour modéliser les interactions avec les interfaces rugueuse. L'approche la plus simple dans le cas de particules classiques consiste à utiliser un pourcentage de réflexions diffusives non spéculaires aux interfaces rugueuses [SANG92]. Ce pourcentage est ajusté pour retrouver les courbes de mobilité universelle [VALI05]. Cette méthode empirique est simple et peu consommatrice en temps. D'autres équipes ont simulé des interfaces réellement rugueuses (obtenues par tirage au sort). Les réflexions des porteurs [ASEN01] et/ou les modifications sur les niveaux d'énergie [WANG05] sont ainsi naturellement prises en compte. Cependant ces méthodes nécessitent une résolution selon les 3 dimensions de l'espace réel (3D). En effet, en fonction de leur position et de leur forme, les irrégularités des interfaces peuvent dévier différemment les particules incidentes selon les directions x, y ou z. De plus, dans des transistors ultracourts la longueur de corrélation de la rugosité n'est pas négligeable devant la longueur de canal. Il faut donc simuler de nombreux dispositifs pour obtenir un effet moyen.

Les approches précédentes nécessitant la résolution des équations dans l'espace réel (résolution spécifique aux simulateurs de dispositifs), nous nous sommes intéressés à une approche perturbative

pour calculer les fréquences d'interaction  $\lambda(\vec{k})$  associées à ce type d'interactions. Le potentiel de déformation dépend de la variation de position de l'interface par l'intermédiaire de la densité spectrale de puissance (qui est la transformée de Fourier de la fonction d'auto-corrélation).

L'Hamiltonien de perturbation dû à la rugosité est tel que :

$$H_{SR}(\vec{r}_{xy}, z) = -q \times \delta V_{\delta T_{Si}}(\vec{r}_{xy}, z) \text{ avec } \delta V_{\delta T_{Si}}(\vec{r}_{xy}, z) \text{ potentiel de perturbation.}$$

A l'ordre 1, on suppose que le potentiel de perturbation peut s'écrire

$$\delta V_{\delta T_{Si}}(\vec{r}_{xy}, z) = V_{T_{Si} + \delta T_{Si}}(\vec{r}_{xy}, z) - V_{T_{Si}}(\vec{r}_{xy}, z) \approx \frac{\delta V_{\delta T_{Si}}(z)}{\delta T_{Si}} \times \delta T_{Si}(\vec{r}_{xy})$$

La matrice de transition s'écrit alors :

$$|M_{m,n}(\vec{k}, \vec{k}')|^2 = \left| \langle n, \vec{k}' | H_{SR} | m, \vec{k} \rangle \right|^2 = \left| \int_{z_1}^{z_2} \xi_m(z) \frac{-\delta V_{\delta T_{Si}}(z)}{\delta T_{Si}} \times \xi_n(z) \iint e^{-i\Delta\vec{k} \cdot \vec{r}_{xy}} \delta T_{Si}(\vec{r}_{xy}) d\vec{r}_{xy} dz \right|^2$$

Soit, en posant  $\Delta\vec{k} = \vec{k}' - \vec{k}$  et  $\overline{\delta T_{Si}}(\Delta\vec{k})$  pour la transformée de Fourier de  $\delta T_{Si}(\vec{r}_{xy})$  :

$$|M_{m,n}(\Delta\vec{k})|^2 = \left| \int_{z_1}^{z_2} \xi_m(z) \frac{\delta V_{\delta T_{Si}}(z)}{\delta T_{Si}} \times \xi_n(z) dz \right|^2 \times |\overline{\delta T_{Si}}(\Delta\vec{k})|^2$$

La densité de probabilité associée est :

$$s(\vec{k}, \vec{k}') = \frac{2\pi}{\hbar} |M_{m,n}(\Delta\vec{k})|^2 \frac{1}{4\pi^2} = \frac{2\pi}{\hbar} \times \left| \int_{z_1}^{z_2} \xi_m(z) \frac{\delta V_{\delta T_{Si}}(z)}{\delta T_{Si}} \times \xi_n(z) dz \right|^2 \times |\overline{\delta T_{Si}}(\Delta\vec{k})|^2 \times \frac{\sqrt{m_x m_y}}{\pi \hbar^2}$$

Dans le modèle exponentiel, la densité spectrale de puissance  $\overline{\delta T_{Si}}(\Delta\vec{k})$  ne dépend que de la norme du vecteur  $|\Delta\vec{k}| = |\vec{k}' - \vec{k}|$ . On en déduit que  $s(\vec{k}, \vec{k}')$  est fonction de l'angle de déviation  $\beta$  entre  $\vec{k}'$  et  $\vec{k}$ . La déviation est donc anisotrope, nous la caractériserons par un temps de relaxation :

$$\frac{1}{\tau_{m,n}^{\text{rug}}(\vec{k})} = \int s(|\Delta\vec{k}|(\beta)) (1 - \cos \beta) d\beta$$

Finalement,

$$\frac{1}{\tau_{m,n}^{\text{rug}}(\vec{k})} = \frac{q^2 (1 + 2\alpha \cdot E')}{2\pi \hbar^3} \int_0^{2\pi} \int_{z_1}^{z_2} \xi_m(z) \frac{\delta V_{\delta T_{Si}}(z)}{\delta T_{Si}} \xi_n(z) dz \left| \overline{\delta T_{Si}}(|\Delta\vec{k}|) \right|^2 \times \frac{(1 - \cos \beta)}{\frac{\cos^2(\beta + \theta_0)}{m'_x} + \frac{\sin^2(\beta + \theta_0)}{m'_y}} d\beta$$

Approximations de la perturbation  $\frac{\delta V_{\delta T_{Si}}(z)}{\delta T_{Si}}(z) = ?$

Une première expression de la déformation de la bande de conduction due à l'interface rugueuse dans un gaz 2Dk donnée par Goodnick [GOOD85]  $\left( \frac{\delta V_{\delta T_{Si}}}{\delta T_{Si}}(z) = E_z(z) \right)$  est dérivée de celle donnée

par Ando et Stern [ANDO82] pour un puits triangulaire (cas des dispositifs à simple grille). Il s'adapte donc mal aux dispositifs à double grille. L'approche de Gamiz [GAMI00] est plus générale mais nécessite une étape préalable de calcul (résolution couplée de Schrödinger/Poisson) du profil de potentiel  $V_{T_{Si}+\Delta T_{Si}}(z)$  pour une certaine valeur d'écart  $\delta T_{Si} = \Delta T_{Si}$ . Ensuite, les autres potentiels perturbés par un  $\delta T_{Si}$  quelconque sont évalués par interpolation linéaire. En conséquence, la

déformation a pour valeur :  $\frac{\delta V_{\delta T_{Si}}}{\delta T_{Si}}(z) = \frac{V_{T_{Si}+\Delta T_{Si}}(z) - V_{T_{Si}}(z)}{\Delta T_{Si}}$ . Cette méthode est difficilement

transposable dans les simulateurs de dispositifs, on s'est donc intéressé au calcul analytique de la déformation de potentiel [BOUR06] calculé dans un dispositif double grille polarisé symétriquement. On peut alors considérer la déformation  $B(z, \Delta k_{xy})$  suivante :

$$\frac{\delta V_{\delta T_{Si}}}{\delta T_{Si}}(z, \Delta k_{xy}) = B(z, \Delta k_{xy}) = \frac{q \times E_z(z_{int}) \times (1 - \epsilon_{Si}/\epsilon_{ox})}{\epsilon_{Si}/\epsilon_{ox} \times \text{th}(\Delta k_{xy} \times T_{ox}) \times \text{sh}(\Delta k_{xy} \times T_{Si}/2) + \text{ch}(\Delta k_{xy} \times T_{Si}/2)} \times \text{ch}(\Delta k_{xy} \times z)$$

avec  $z_{int}$  profondeur de l'interface ( $z_{int} = 0$  ou  $T_{Si}$ ). Ce modèle reste en cours de validation.

D'autres approches directement liées à l'aspect ondulatoire, s'intéressent en plus de la perturbation du potentiel électrostatique, aux modifications de la position des niveaux d'énergie  $\frac{\delta V_{\delta T_{Si}}}{\delta T_{Si}}(z) = \frac{\partial E_n}{\partial T_{Si}}$  générée directement par la fluctuation d'épaisseur du film de silicium.

Pour augmenter la vitesse de calcul, des approximations analytiques sont très utiles. Le potentiel de confinement peut être assimilé à puits triangulaire dans le cas de transistor simple grille ou à un puits carré [SAKA87] [STRO98] pour les DG étroits polarisés symétriquement. La référence [ESSE03] a montré que l'approximation analytique du puits carré infini s'avère parfaitement en accord avec les solutions exactes dans les films de moins de 5 nm d'épaisseur. En effet, dans ces puits ultrafins, le premier niveau est très éloigné de la bande de conduction, tout se passe comme s'il ne subissait que le potentiel de confinement de type puits carré. Ainsi, dans un dispositif à

double grille polarisé symétriquement ( $T_{Si} \leq 5\text{nm}$ ) :  $\frac{\delta V_{\delta T_{Si}}}{\delta T_{Si}}(z) = \frac{\partial E_n}{\partial T_{Si}} = \frac{-\pi \hbar^2 n^2}{m_z^* T_{Si}^3}$ .

### Fréquence de l'interaction

Finalement, dans un transistor à double grille polarisé symétriquement, les différentes probabilités de passage d'un électron 2Dk d'énergie cinétique  $E$  d'une sous-bande 'm' vers une

sous-bande 'n' par l'intermédiaire d'une interaction avec **deux** interfaces rugueuses supposées indépendantes sont données par les expressions suivantes (SI) :

$$\frac{1}{\tau_{m,n}^{\text{rug-dV}}(\vec{k})} = \frac{1}{\pi\hbar^3} (1 + 2\alpha \cdot E') \times \int_0^{2\pi} \int_{z_1}^{z_2} \xi_m(z) \left[ B(z, \Delta k) + \frac{-\pi\hbar^2 n^2}{m_z^* T_{Si}^3} \delta_{m,n} \right] \xi_n(z) dz \left| \frac{2\pi\Delta^2 \Lambda^2}{(1 + |\Delta k|^2 \Lambda^2)^{3/2}} \frac{(1 - \cos\beta)}{\frac{\cos^2(\beta + \theta_0)}{m'_x} + \frac{\sin^2(\beta + \theta_0)}{m'_y}} \right. d\beta$$

avec  $\Delta$  écart type et  $\Lambda$  longueur caractéristique des interfaces rugueuses,  $\delta_{m,n}$  symbole de Kronecker ( $\delta_{m,n} = 1$  si  $m = n$  sinon  $\delta_{m,n} = 0$ ) et  $\Delta k = |\vec{k}' - \vec{k}|$ .

### Effets de l'interaction

L'interaction est élastique  $E' = E - (E_n - E_m)$  avec  $E_n$  et  $E_m$  qui sont respectivement les énergies associées aux niveaux  $n$  et  $m$ , et isotrope grâce à l'approximation du temps de relaxation.

## 2.3 Mobilités

Dans cette section, nous présentons notre étude du transport d'électrons en gaz 2Dk à température ambiante ( $T = 300$  K) effectuée grâce aux simulations Monte Carlo à champ constant décrites précédemment. Les structures étudiées sont à double grille et polarisées symétriquement.

Pour caractériser les gaz 2Dk, nous avons utilisé les notions de charges d'inversion  $N_{\text{inv}}$  et de champ effectif  $E_{\text{eff}}$  (dans la direction  $z$  de confinement). La charge surfacique d'inversion dans une structure à double grille polarisée symétriquement représente la moitié de la concentration sommée

sur toute une tranche Schrödinger, soit  $N_{\text{inv}} = \int_{z_{\text{int}}}^{z_{\text{int}} + T_{Si}/2} n(z) dz$ .  $E_{\text{eff}}$  est défini comme la moyenne du

champ électrique perpendiculaire pondéré par la concentration d'électrons :

$$E_{\text{eff}} = \frac{\int_{z_{\text{int}}}^{z_{\text{int}} + T_{Si}/2} n(z) \times E_z(z) dz}{N_{\text{inv}}} \quad \text{où } E_z \text{ est la composante du champ électrique dans la direction du}$$

confinement et  $n$  la concentration d'électrons (en  $\text{cm}^{-3}$ ).

### 2.3.1 Phonons seuls

Dans ce paragraphe, seul l'effet des interactions avec le réseau cristallin (phonons) est étudié. Pour donner un premier aperçu des différences entre le transport d'un gaz 2Dk et celui 3Dk, les évolutions des fréquences d'interaction avec les phonons acoustiques de faible énergie pour les 2 types de gaz ont été tracées en Figure 128. A chaque fois que l'énergie cinétique des électrons 2Dk devient supérieure à un niveau d'énergie quantifié, la sous-bande associée fournit une nouvelle réserve d'états finaux possibles après interaction. Cela se traduit au niveau des fréquences

d'interaction par des marches d'escalier, typiques des gaz 2Dk, chaque discontinuité révélant la présence d'un niveau d'énergie.

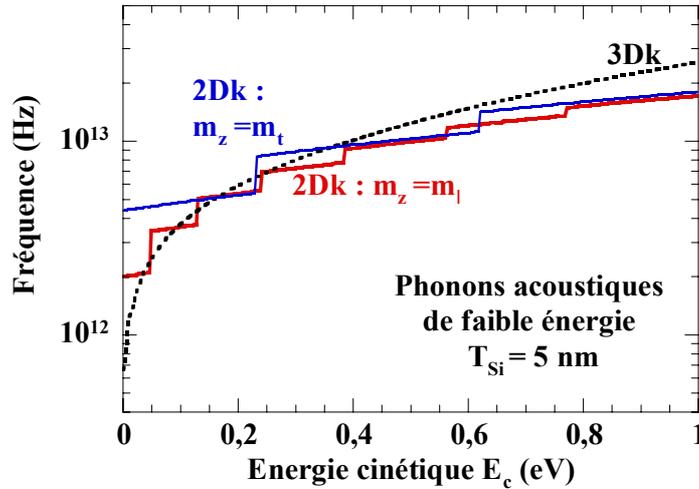


Figure 128 : Fréquences d'interaction pour les phonons acoustiques de faibles énergies. En traits continus : pour les première sous-bande d'électrons 2Dk (pour les vallées où  $m_z = m_t$  et celles où  $m_z = m_l$ ), en tirets fréquences 3Dk.

Quelle que soit la dimensionnalité du gaz (2Dk ou 3Dk), les fréquences sont du même ordre de grandeur. Cependant à faible énergie, les fréquences 2Dk sont plus importantes. En effet, la densité d'états des niveaux d'arrivée est plus importante à faible énergie dans un gaz 2Dk (constante en fonction de l'énergie) que dans un gaz 3Dk où elle tend à s'annuler à basse énergie ( $\propto \sqrt{E_c}$  où  $E_c$  est l'énergie cinétique des électrons). En outre, plus les niveaux d'énergie quantifiés sont élevés, plus les écarts d'énergie entre deux niveaux successifs (longueur des marches) s'élargissent dans un puits quasi carré. Cela conduit à des fréquences 3Dk supérieures aux 2Dk pour des énergies élevées. Dans un puits de 5 nm le croisement des fréquences 2Dk et 3Dk se produit aux environs de 0,3 eV.

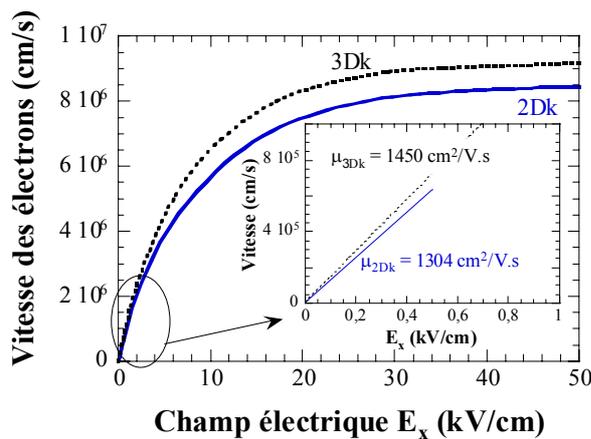


Figure 129 : Caractéristique vitesse en fonction du champ électrique accélérateur en régime de mobilité pour un gaz 2Dk d'électrons dans un film de silicium d'épaisseur  $T_{Si} = 5$  nm non dopé. Tirets : cas du silicium massif (MONACO 3Dk).

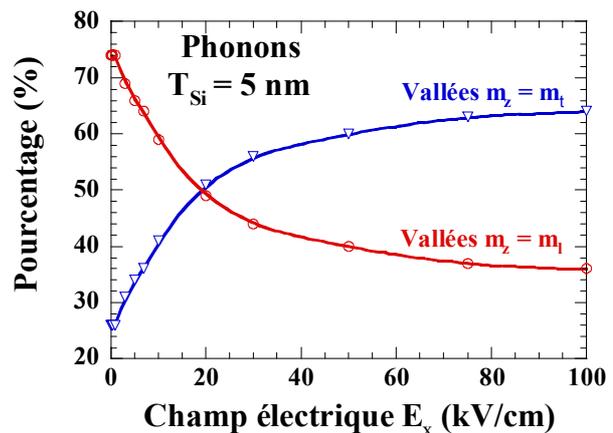


Figure 130 : Evolution de la répartition des porteurs par vallée en fonction du champ électrique pour un gaz 2Dk d'électrons dans un film de silicium d'épaisseur  $T_{Si} = 5$  nm non dopé.

Après ce petit préambule intéressons nous plus directement au transport électronique. Lorsque l'on applique un champ électrique constant sur un gaz électronique initialement au repos, après un régime transitoire durant lequel la vitesse moyenne des porteurs évolue avec le temps, le régime stationnaire est atteint. Alors, l'effet du champ électrique est équilibré par les phénomènes diffusifs (interactions) ; la vitesse moyenne des porteurs n'évolue plus en fonction du temps. En Figure 129, l'évolution de la vitesse stationnaire  $v_x$  des électrons 2Dk est tracée en fonction du champ électrique parallèle appliqué  $E_x$ . Dans un puits étroit de silicium, d'épaisseur  $T_{Si} = 5$  nm, on a sous  $V_{GS} = 0,7$  V :  $E_{eff} = 4 \times 10^5$  V/cm et  $N_{inv} = 4 \times 10^{12}$  cm<sup>-2</sup>. A titre de comparaison, la relation  $v_x(E_x)$  est indiquée en tirets pour un gaz 3Dk.

Le régime de mobilité, où la relation  $v_x(E_x)$  est linéaire, se produit pour des champs faibles, inférieurs à 1 kV/cm. L'insert de la Figure 129 indique que la mobilité 2Dk est légèrement plus faible qu'en 3Dk. En effets, les interactions 2Dk plus probables donc plus nombreuses qu'en 3Dk (cf. Figure 128) ralentissent les porteurs. Parallèlement, nous avons tracé sur la Figure 130 l'évolution de la répartition des électrons sur les différentes vallées en fonction du champ électrique. Plus le champ électrique est élevé, plus la répartition des porteurs sur les vallées tend à être uniforme comme dans le silicium massif. Cela a pour effet de d'augmenter la masse effective de conduction moyenne du gaz 2Dk et donc de ralentir les électrons 2Dk. A cause de la faible mobilité et de l'augmentation de la masse moyenne, la vitesse de saturation 2Dk ( $v_{sat2Dk} = 0,85 \cdot 10^7$  cm/s) est plus faible que celle 3Dk ( $v_{sat3Dk} = 0,92 \cdot 10^7$  cm/s). Il faut signaler que même si les valeurs des fréquences d'interaction 2Dk à haute énergie ( $E_c > 0,4$  eV, cf. Figure 128) sont plus faibles que les fréquences 3Dk, très peu d'électrons sont aussi énergétiques en régime stationnaire, même sous fort champ électrique.

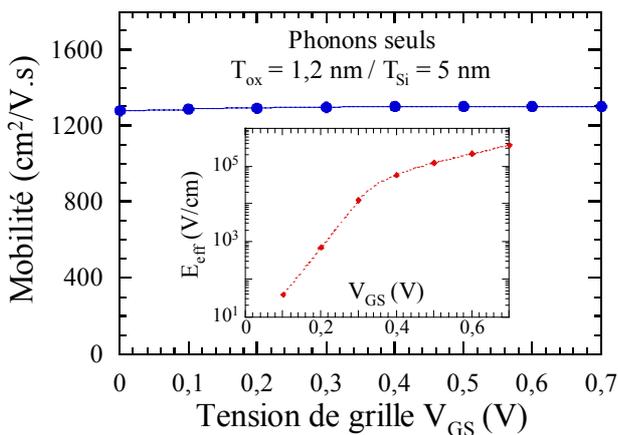


Figure 131 : Evolution de la mobilité en fonction de la tension de grille  $V_{GS}$  pour un gaz 2Dk d'électrons dans un film de silicium non dopé avec  $T_{ox} = 1,2$  nm et  $T_{Si} = 5$  nm. Seules les interactions électron 2Dk/phonon sont prises en compte. Insert : Champ effectif en fonction de  $V_{GS}$ .

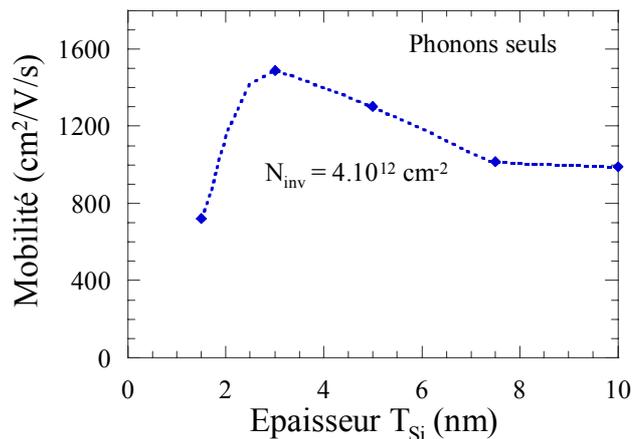


Figure 132 : Evolution de la mobilité en fonction de l'épaisseur  $T_{Si}$  du film de silicium pour un gaz 2Dk d'électrons avec  $T_{ox} = 1,2$  nm et à  $V_{GS} = 0,7$  V pour une charge d'inversion  $N_{inv} = 4 \cdot 10^{12}$  cm<sup>-2</sup>. Seules les interactions électron 2Dk/phonon sont prises en compte..

Nous avons ensuite analysé l'évolution de la mobilité en fonction du champ effectif perpendiculaire  $E_{\text{eff}}$ . Nous avons donc simulé une structure MOS à double grille d'épaisseur  $T_{\text{Si}} = 5 \text{ nm}$  avec une épaisseur d'oxyde  $T_{\text{ox}} = 1,2 \text{ nm}$  à plusieurs polarisations de grille  $V_{\text{GS}}$ . Pour  $E_{\text{eff}} < 4 \times 10^5 \text{ V/cm}$ , la Figure 131 indique clairement que dans de telles structures  $E_{\text{eff}}$  ne modifie pas la mobilité. En effet, dans des dispositifs aussi fins les niveaux d'énergie quantifiés restent supérieurs au fond du puits (cf. Figure 120) ; alors, le puits de potentiel reste toujours de forme quasi carrée. Les niveaux d'énergie et surtout les fonctions d'onde associées sont très peu modifiés. En conséquence, les fréquences d'interaction électrons/phonons sont très peu sensibles à  $E_{\text{eff}}$ , du moins tant que le puits de potentiel reste quasi carré.

Grâce au logiciel de Schrödinger/Poisson, on a précédemment observé que l'évolution de la répartition des porteurs sur les vallées induit une diminution de la masse effective moyenne d'un gaz 2Dk à mesure que l'épaisseur des puits se réduit (cf. insert de la Figure 121). Pour mettre en évidence l'effet bénéfique de ce phénomène sur le transport, la Figure 132 indique la mobilité électronique 2Dk pour des puits d'épaisseurs  $T_{\text{Si}}$  variant de 1,5 nm à 10 nm. Dans tous ces dispositifs, le champ effectif  $E_{\text{eff}}$  et la charge d'inversion  $N_{\text{inv}}$  varient peu. Comme attendu, la mobilité augmente quand  $T_{\text{Si}}$  diminue de 10 nm à 3 nm.

Pour des épaisseurs plus faibles, la mobilité baisse. En effet, plus les porteurs sont confinés, plus les interactions inter sous-bandes sont probables à cause du recouvrement des fonctions d'onde. Par exemple, les fonctions d'onde associées au puits carré infini :  $\xi_n = \sqrt{2/T_{\text{Si}}} \sin(n.\pi.x / T_{\text{Si}})$ , ont des

intégrales de recouvrement  $F_{m,n} = \int_0^{T_{\text{Si}}} |\xi_n(z)|^2 |\xi_m(z)|^2 dz$  qui évoluent proportionnellement à  $1/T_{\text{Si}}$  ( $F_{1,1} = 1,5/T_{\text{Si}} \dots$ ).

Du fait des évolutions opposées des intégrales de recouvrement et de la masse effective moyenne en fonction de  $T_{\text{Si}}$ , la mobilité maximale (due seulement aux phonons) est obtenue dans des films d'épaisseur  $T_{\text{Si}} = 3 \text{ nm}$ .

### 2.3.2 Impuretés ionisées

Tout d'abord, analysons les fréquences d'interaction avec les impuretés ionisées tracées en fonction de l'énergie cinétique  $E_c$  en Figure 133. Contrairement aux fréquences d'interaction avec les phonons, elles sont décroissantes pour  $E_c$  croissant à haute énergie. Si l'on ne prend pas en compte l'effet du gaz d'électrons qui « écrante » l'effet d'une charge ionisée isolée, les fréquences d'interaction des porteurs de faible énergie avec les impuretés ionisées tendent vers l'infini. Le paramètre  $N_{\text{scr}}$  quantifie l'écrantage pris en compte dans nos simulations (cf. 2.2.3 ). Comme on

peut le constater sur la Figure 133, plus il est important, plus les fréquences à basses énergies sont faibles.

A cause des intégrales de forme présentes dans l'expression des fréquences d'interaction, les interactions inter sous-bandes sont d'autant moins probables que les niveaux d'énergie initial et final sont différents :  $\lambda_{1,1} > \lambda_{1,2} > \lambda_{1,3} \dots$ .

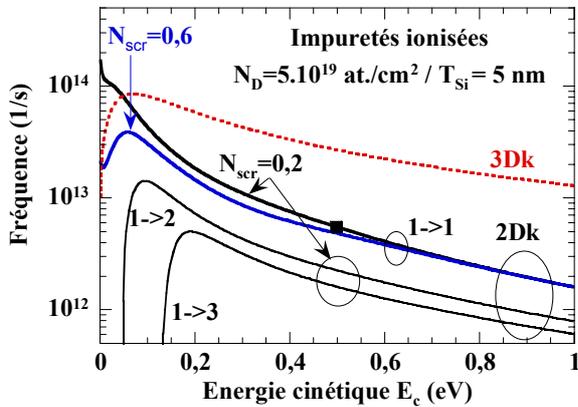


Figure 133 : Evolution des fréquences d'interaction avec les impuretés ionisées en fonction de l'énergie cinétique des porteurs dans du silicium dopé  $N_D = 5 \times 10^{19} \text{ at./cm}^3$ . En tirets : silicium massif, en traits pleins : film de silicium d'épaisseur  $T_{Si} = 5 \text{ nm}$  pour les transitions inter sous-bande (1->1) et intra sous-bande pour diverses valeurs de la constante d'écrantage  $N_{scr}$ .

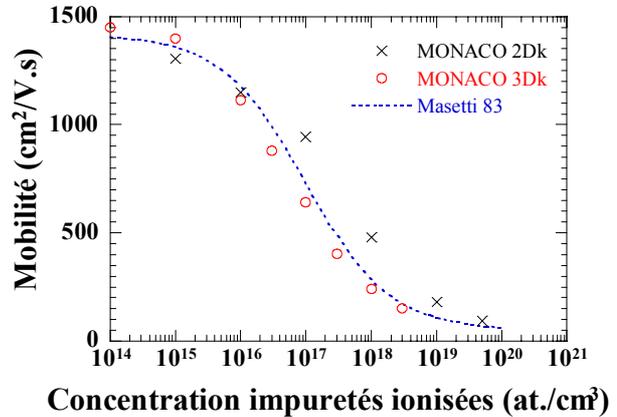


Figure 134 : Evolution de la mobilité des électrons 2Dk en fonction de la concentration d'impuretés dans un film de silicium d'épaisseur  $T_{Si} = 5 \text{ nm}$ . Les cercles rouges correspondent aux résultats MONACO 3Dk, les croix noires à ceux de MONACO 2DK. En tirets sont reportées les mesures de la référence [MASE83].

N'ayant pas à notre disposition de mesures expérimentales relatives à l'évolution de la mobilité en fonction du dopage dans un gaz 2Dk, le paramètre d'écrantage  $N_{scr}$  a été ajusté pour correspondre aux résultats 3Dk à fort dopage pour une valeur :  $N_D = 5.10^{19} \text{ at./cm}^3$  choisie arbitrairement. En considérant les interactions avec les phonons et les impuretés ionisées, les évolutions de la mobilité en fonction de la concentration d'impuretés sont tracées en Figure 134. Même avec un faible niveau d'écrantage, les interactions avec les impuretés ionisées 2Dk sont moins efficaces que les 3Dk, cela induit une mobilité 2Dk plus importante dans les zones dopées, bien que dans les zones faiblement dopées (dopage inférieur à  $10^{16} \text{ at./cm}^3$ ) elle soit plus faible.

### 2.3.3 Comparaisons et bilan

Pour décrire les interactions avec les interfaces rugueuses, nous avons choisi d'implémenter seulement le modèle intégrant les modifications des niveaux d'énergie ( $B(z, \Delta k) = 0$ ) dans l'approximation du puits carré infini qui est très pertinente dans les films ultraminces d'épaisseurs inférieures à 5 nm. Il faut noter que dans ce modèle, contrairement au cas 3Dk, le champ perpendiculaire n'a aucune influence.

La Figure 135 compare les mobilités obtenues par différents simulateurs pour différentes épaisseurs de film à des points expérimentaux. Par rapport aux autres résultats de simulation

publiés, les mobilités 2Dk que nous obtenons sont plus importantes, particulièrement celles dues aux phonons. Nos expressions des fréquences d'interaction sont identiques à celles des autres références. Cependant, les potentiels de déformation utilisés par [ESSE03] ont été modifiés par rapport à ceux précédemment ajustés dans les films épais. L'expérience ayant de nombreuses fois validé les valeurs des paramètres que nous utilisons pour les fréquences d'interaction, en 3Dk [DOLL97] comme en 2Dk [MONS04], nous attendons plus de résultats expérimentaux concernant les structures DG MOS pour les faire évoluer. De même, l'addition de nouveaux types d'interactions tels que les phonons optiques de surface [ESSE03] ou la prise en compte des impuretés distantes [FISC01] [GAMI03] est envisagée, avec cependant un succès relatif, particulièrement dans le cas des phonons de surface. Le transport d'un gaz 2Dk dans un dispositif à double grille polarisé symétriquement reste encore un sujet de recherche intense aussi bien d'un point de vue expérimental que théorique.

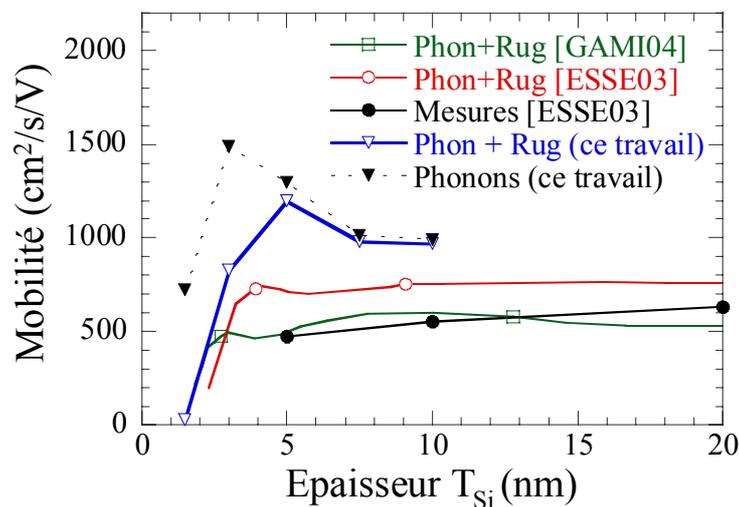


Figure 135 : Comparaison de l'évolution de la mobilité en fonction de l'épaisseur électrique en régime de mobilité pour un gaz 2Dk d'électrons dans un film de silicium non dopé (rugosité et phonons). Cercles : mesures (pleins noirs) et résultats (vides rouges) de simulations reportés à  $N_{inv} = 10^{12} \text{ cm}^{-2}$  en référence [ESSE03]. Carrés verts : résultats de simulations reportés à  $N_{inv} = 5 \times 10^{12} \text{ cm}^{-2}$  en référence [GAMI04]. Triangles bleus : nos résultats à  $N_{inv} = 4 \cdot 10^{12} \text{ cm}^{-2}$  avec prise en compte des phonons et de la rugosité.

### 3. Le simulateur de dispositifs MONACO 2Dk

Les paragraphes précédents ont montré que les phénomènes quantiques apparaissant dans les dispositifs ultracourts, qui sont aussi ultrafins selon la direction perpendiculaire au transport, ont une influence non négligeable sur l'électrostatique, le contrôle de charges, la tension de seuil... mais aussi sur le transport. L'intégration de ces phénomènes dans les simulateurs de dispositifs apparaît donc nécessaire.

Dans ce contexte, le simulateur MONACO, basé sur la méthode particulière Monte Carlo et développé au sein de l'équipe, a été enrichi et couplé à un algorithme de résolution de l'équation de Schrödinger pour prendre en compte le confinement quantique. Ainsi une résolution auto-cohérente des équations de Schrödinger (dans la direction perpendiculaire au transport), de Poisson 2Dr et de Boltzmann 2Dk a ainsi été développée.

Après une brève description des différentes approches proposées dans la littérature pour inclure les effets quantiques, nous décrirons en détail le logiciel MONACO 2Dk. Enfin, nous analyserons les premiers résultats obtenus afin de valider le simulateur.

### **3.1 Les simulations de dispositifs « quantiques »**

Pour prendre en compte les nouveaux phénomènes quantiques qui apparaissent dans les transistors, différentes solutions ont été mises en œuvre. Les premières constituent de simples corrections des modèles 3Dk existants. Dans d'autres cas, on a anticipé depuis longtemps l'arrivée de la mécanique quantique dans le monde des composants, et donc développé des formalismes totalement nouveaux.

#### **3.1.1 « *Density-gradient* »**

Les simulateurs les plus largement utilisés en modélisation physique de dispositifs microélectroniques sont basés sur les modèles de dérive-diffusion. Ces modèles décrivent le gaz de porteurs d'un point de vue macroscopique, et pour traiter les phénomènes ondulatoires un terme « quantique » est ajouté dans les équations de flux de porteurs [ANCO90]. Ce terme variant en fonction du gradient de densité de porteurs a donné le nom à la méthode : « density-gradient ». Peu de modifications sont nécessaires et les nombreuses calibrations accumulées durant de nombreuses années sur les logiciels de dérive-diffusion sont conservées. Ces simulateurs, peu consommateurs de temps de calcul, permettent de rendre compte de certains effets de la quantification dans les dispositifs fins. Avec peu de paramètres d'ajustements (1 par matériau et par type de porteur) l'éloignement des porteurs de l'interface et la modification de la tension de seuil associée sont relativement bien décrits. La prise en compte des courants tunnels (de grille ou source/drain) par une telle méthode reste encore un sujet très controversé [HORH03]...

Cependant, à cause des hypothèses simplificatrices permettant d'aboutir aux équations de dérive-diffusion, ces simulateurs traitent des porteurs, classiques ou quantiques, proches de l'équilibre. Ils ne permettent donc pas de traiter rigoureusement le transport fortement non stationnaire apparaissant dans les dispositifs ultracourts (cf. chapitre II).

### 3.1.2 Monte Carlo avec « corrections quantiques »

Dans le même esprit que les simulateurs « density-gradient », des simulateurs de type Monte Carlo, mieux à même d'étudier les dispositifs fortement hors d'équilibre, ont été modifiés pour mimer le comportement macroscopique d'un gaz 2Dk.

Dans les simulateurs Monte Carlo, les « corrections » quantiques sont effectuées au niveau du potentiel utilisé pour le calcul du champ électrique. Différents types de correction de potentiel plus ou moins empiriques ont été proposés. La correction la plus utilisée à l'heure actuelle est appelée potentiel effectif [FERR00]. Mis à part les difficultés liées à l'ajustement de certains paramètres, cette méthode s'avère relativement facile à mettre en œuvre, néanmoins elle ne peut reproduire finement le profil de porteurs issu de la résolution directe de l'équation de Schrödinger/Poisson [JAUD05]. Des approches plus sophistiquées [KATH03] corrigent le potentiel à l'aide de résolutions de l'équation de Schrödinger permettant d'obtenir de meilleurs profils de concentrations mais encore à l'aide d'un paramètre d'ajustement... On citera aussi l'approche du « potentiel quantique » qui est dérivée du formalisme des fonctions de Wigner [TSUC01]. Théoriquement plus séduisante, cette approche est difficile à appliquer d'un point de vue pratique (dérivée troisième d'un potentiel naturellement bruité...). Elle devrait cependant permettre de reproduire le courant tunnel selon l'axe source/drain.

Ces méthodes de « corrections quantiques » ont l'avantage de ne pas être beaucoup plus coûteuses en calcul que les algorithmes Monte Carlo classiques, en contre partie il faut accepter certaines approximations sur les résultats. La correction « parfaite », c'est-à-dire facile à mettre en œuvre et donnant des résultats rapides (à l'échelle des simulations Monte Carlo...) et exacts, n'existe pas encore...

### 3.1.3 Nouvelles approches quantiques

Pour décrire le comportement de composants où la mécanique quantique a une grande importance (MOSFET ultimes, diodes tunnels résonnantes, électronique moléculaire [XUE05]...) des approches purement quantiques ont été développées. Ces formalismes gèrent la conservation de la phase de la fonction d'onde des porteurs et autorisent donc la description des phénomènes ondulatoires tels que les interférences et les courants tunnels.

Tous d'abord, des résolutions directes de l'équation stationnaire de Schrödinger ont été réalisées dont les plus courantes utilisent des matrices de transfert [CASS00]. D'autres équipes résolvent l'équation de Schrödinger 3D temporelle [SORI04]; les porteurs sont alors décrits comme des paquets d'ondes. Cependant, ces méthodes relativement faciles à mettre en œuvre n'incluent pas les

effets dissipatifs qui ne sont pourtant pas négligeables, même dans les MOSFET ultimes (cf. chapitre II).

Il existe des formalismes quantiques qui modélisent aussi les phénomènes dissipatifs, c'est-à-dire essentiellement les interactions des porteurs avec le réseau (phonons). Des dispositifs qui ne sont ni à l'équilibre thermique, ni entièrement cohérents peuvent ainsi être étudiés. Parmi ces méthodes, on citera le formalisme très général des fonctions de Green hors d'équilibre [LAKE92], et celui des fonctions de Wigner [NEDJ04] où le traitement des dissipations est peut-être plus « naturel ». Ces méthodes extrêmement gourmandes en temps de calcul, sont encore peu matures et un gros travail d'optimisation et de validation reste à réaliser.

### 3.1.4 L'approche par mode

Pour étudier de manière efficace les effets quantiques dans les dispositifs ultimes, l'approche par mode ou « mode-space » [REN01] est très répandue. Elle a pour objectif de découpler l'équation de Schrödinger **2D** stationnaire :

$$-\frac{\hbar^2}{2m_x} \frac{\partial^2 \Psi(x,z)}{\partial x^2} - \frac{\hbar^2}{2m_z} \frac{\partial^2 \Psi(x,z)}{\partial z^2} - qV(x,z)\Psi(x,z) = E \times \Psi(x,z) \quad [1],$$

en **2** équations **1D** selon l'axe  $z$  et selon l'axe  $x$ . Alors, l'équation de transport (selon l'axe  $x$ ) et le calcul des modes (selon l'axe  $z$ ) sont totalement découplés.

Dans cette approche, les fonctions d'onde des électrons, a priori quelconques, sont mises sous la forme :

$$\Psi(x,z) = \sum_n \tilde{\Psi}_n(x) \times \xi_n(x,z) \quad [2]$$

avec  $\xi_n(x,z)$  fonction d'onde associée à l'état propre, ou mode, numéro 'n' d'énergie  $E_n(x)$  dans une tranche de coordonnée  $x$ . D'après ces notations,  $|\tilde{\Psi}_n(x)|^2$  représente la probabilité de trouver un électron sur le mode (état propre) 'n' de la tranche Schrödinger située à la position  $x$ .

Si les éventuels couplages entre les modes dus aux termes  $\int \xi_n^*(x_0,z) \frac{\partial \xi_m}{\partial x}(x_0,z) dz$  et  $\int \xi_n^*(x_0,z) \frac{\partial^2 \xi_m}{\partial x^2}(x_0,z) dz$  sont négligeables devant celui en  $\int \xi_n^*(x_0,z) \xi_m(x_0,z) dz$  (cf. calcul en annexe), l'équation 2D [1] est alors équivalente au système d'équations 1D suivant :

$$-\frac{\hbar^2}{2m_x} \frac{d^2}{dx^2} \tilde{\Psi}_n(x) - E_n(x) \tilde{\Psi}_n(x) = E \times \tilde{\Psi}_n(x) \quad [3]$$

$$-\frac{\hbar^2}{2m_z} \frac{\partial^2}{\partial z^2} \xi_n(x_0, z) - qV(x_0, z) \times \xi_n(x_0, z) = E_n(x_0) \times \xi_n(x_0, z), \text{ pour tout } x_0 \quad [4]$$

Dans ce système, les énergies des modes  $E_n(x)$  peuvent évoluer librement, et ces évolutions déterminent celles des fonctions d'onde dans le sens du transport (cf. équation [3]).

Finalement, l'approche par mode (équations 1D [3] et [4]) permet bien de simplifier le problème 2D (équation [1]) tant que les fonctions d'onde associées aux modes n'évoluent pas trop le long du dispositif ( $\frac{\partial \xi_m}{\partial x}(x_0, z) \ll \xi_n(x_0, z)$  et  $\frac{\partial^2 \xi_m}{\partial x^2}(x_0, z) \ll \xi_n(x_0, z)$ ).

### 3.2 Présentation MONACO 2Dk

Dans le cadre de l'approche par mode, MONACO 2Dk découple le problème selon l'axe du transport (axe x) et celui du confinement (axe z). De plus, en s'inspirant des travaux de la référence [FISC93], approfondie par les références [WILL00] et [EZAK03], la modélisation des porteurs dans MONACO 2Dk diffère selon ces directions comme illustré en Figure 136.

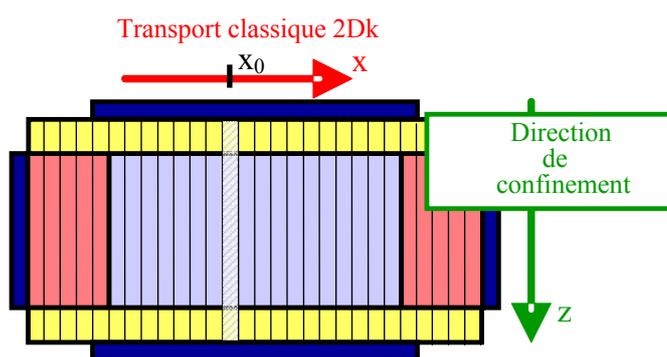


Figure 136 : Schéma du transistor à double grille. L'axe x du transport et l'axe z du confinement sont indiqués. Une tranche « Schrödinger » est hachurée.

Sur une tranche de coordonnée  $x_0$  selon l'axe x, appelée par la suite tranche Schrödinger, les fonctions d'onde des électrons  $\xi_n(x_0, z)$  selon l'axe z (direction perpendiculaire au transport) correspondent à des solutions stationnaires, ou modes, de l'équation de Schrödinger 1D résolue numériquement. L'énergie potentielle dans cette direction est quantifiée et la probabilité de présence des porteurs n'est plus ponctuelle comme dans les simulations classiques (3Dk) mais répartie sur toute une tranche. On note que tous les porteurs présents dans le dispositif ont leur énergie quantifiée dans la direction z ; il n'y a pas de cohabitation avec des porteurs 3Dk dans ces structures ultrafines.

Dans la direction du transport (axes x et y), les porteurs sont considérés de manière purement classique c'est-à-dire comme étant ponctuels, toutefois une sous-bande correspondant à leur énergie potentielle quantifiée leur est affectée. Le transport des électrons qui possèdent seulement deux

degrés de liberté (2Dk : axes x et y), est régi par l'équation classique de Boltzmann résolue par une méthode Monte Carlo. Les transitions inter sous-bandes sont assurées par les interactions entre le gaz 2Dk et son environnement.

Finalement, le problème étant considéré à géométrie 2D, c'est-à-dire invariant par translation selon l'axe y, on associe aux électrons des fonctions d'onde de la forme :  $\Psi(x,y,z)=\delta(x-x_0)\times\xi_n(x,z)\times e^{ik_y y} / \sqrt{W}$  avec  $\delta$  impulsion de Dirac,  $k_y$  composante du vecteur d'onde selon l'axe y et W largeur (selon l'axe y) du transistor. Ce découplage du problème selon les axes x et z (approche « mode-space ») est valide tant que la « forme » du puits de potentiel n'évolue par trop ( $d\xi_n(x,z)/dx \approx 0$ ) le long du dispositif. Ces conditions étant respectées dans les transistors à double grille [VENU02], la quantification perpendiculaire est ainsi traitée de manière rigoureuse. Contrairement aux approches par potentiel effectif ou « density-gradient », quelles que soient la géométrie et la polarisation, aucun paramètre d'ajustement empirique n'est utilisé ; cependant le temps de calcul en pâtit...

### 3.2.1 Présentation de la méthode Monte Carlo pour la simulation de dispositifs

Dans ce paragraphe, nous allons préciser quels sont les différences et les compléments ajoutés aux simulateurs Monte Carlo de dispositifs par rapport au simulateur à champ constant.

#### Equations du mouvement :

Le problème consiste à déterminer les évolutions de  $\vec{k}(t)$  et  $\vec{r}(t)$  sous l'action des forces extérieures lors du temps de vol libre. Dans l'espace réciproque, le problème est équivalent aux simulations sous champ constant, il nous manque seulement les évolutions dans l'espace réel.

La vitesse de groupe est donnée par  $\vec{v} = \frac{d\vec{r}}{dt} = \frac{1}{\hbar} \vec{\nabla}_{\vec{k}} E(\vec{k})$ . Dans le cas d'une vallée ellipsoïdale

$$\text{non parabolique, il vient : } \vec{v} = \begin{cases} v_x = \frac{\hbar \cdot k_x(t)}{m_x(1+2\alpha \cdot E(t))} \\ v_y = \frac{\hbar \cdot k_y(t)}{m_y(1+2\alpha \cdot E(t))} \end{cases} .$$

Finalement :  $\vec{r}(t + dt) = \vec{r}(t) + \vec{v}(t) \times dt$  avec  $dt$  pas sur le temps (entre deux résolutions de l'équation de Poisson).

#### Couplage avec l'équation de Poisson

Afin d'assurer l'auto-cohérence entre le mouvement des porteurs et la valeur du potentiel électrostatique, l'équation de Poisson est résolue numériquement à intervalles de temps réguliers.

Pour éviter que les grandeurs physiques simulées dans un schéma discrétisé ne s'éloignent des 'vraies' valeurs, les pas de discrétisations spatiales et temporelles doivent être respectivement inférieurs à la longueur de Debye  $L_D$  et au temps de relaxation diélectrique  $\tau_D$ . Les longueurs maximales des mailles en fonction du dopage et des caractéristiques du matériau considéré sont discutées et données par la référence [BRIS94].

Aux électrodes le potentiel est fixé (condition de Dirichlet). Aux contacts ohmiques, on assure dans les mailles adjacentes une densité de porteurs majoritaires constante et égale à la densité d'impuretés ionisées. Pour pallier un défaut de porteurs dans ces mailles, on y injecte à la fin d'un pas de temps  $\delta t$  le nombre convenable de porteurs dans ces mailles « ohmiques ». La composante normale de la vitesse d'un porteur injecté est tirée au sort sur une distribution maxwellienne pondérée par la vitesse pour ne pas injecter des électrons à vitesse nulle.

### Algorithme complet

Pour compléter cette description du simulateur classique, l'algorithme complet issu de la référence [BRIS94] est présenté en Figure 137. Bien évidemment, on retrouve les séquences typiques aux simulations Monte Carlo : chaque porteur subit une succession de vols libres interrompus par des interactions aux effets instantanés.

De plus, on remarque que la contribution de chaque porteur à la concentration électronique  $n$  est pondérée par le temps passé dans chaque maille  $t_m$  :  $n=n+t_m/\delta t$ . Grâce à ce calcul, la résolution de Poisson peut s'effectuer périodiquement à tous les pas sur le temps  $\delta t$

Enfin, cet algorithme dévoile certains problèmes annexes peu ou pas mentionnés jusqu'à présent : gestion du changement de maille et des diverses conditions aux limites. Pour plus de précisions sur le logiciel MONACO, nous conseillons aux lecteurs curieux les thèses ou habilitations suivantes [BRIS94], [DOLL99] et [BARR01].

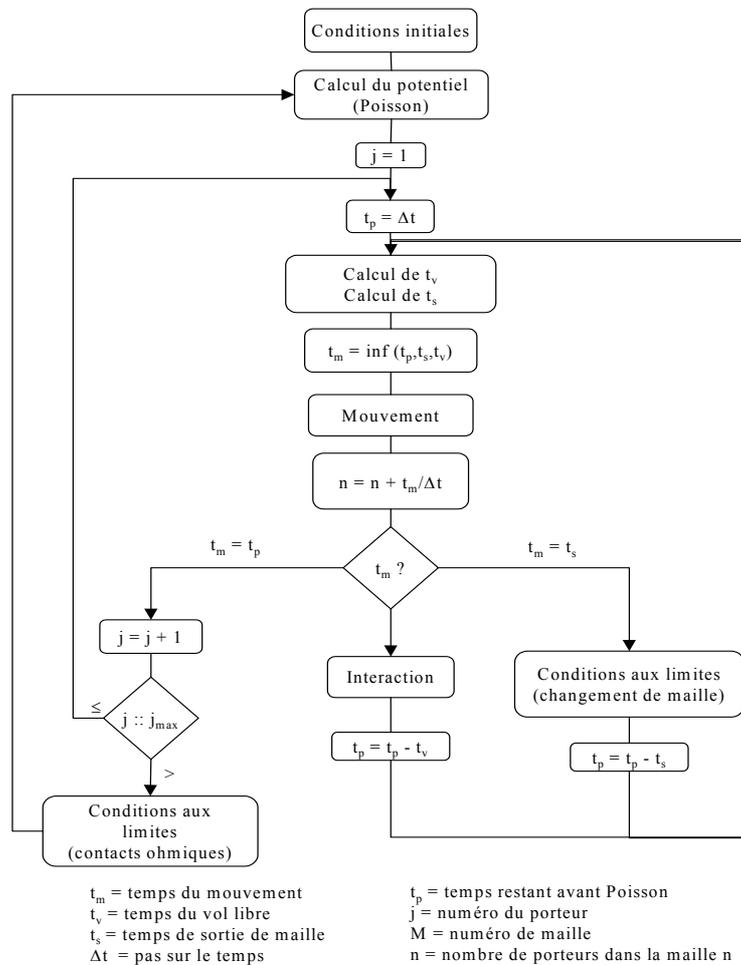


Figure 137 : Algorithme du simulateur MONACO classique [BRIS94].

### 3.2.2 Couplage de la résolution de Schrödinger

Pour inclure les effets de la quantification perpendiculaire dans le logiciel MONACO, on résout dans chaque tranche « Schrödinger » l'équation de Schrödinger 1D à partir du potentiel issu d'une résolution préalable de l'équation de Poisson. La résolution s'effectue par une méthode de différence finie associée à un algorithme classique 'HQR' pour résoudre l'équation aux valeurs propres (cf. annexe).

### 3.2.3 Champ électrique 2Dk

Une grande nouveauté concernant le traitement d'un gaz électronique 2Dk vient du fait que le champ électrique vu par les porteurs 2Dk n'est pas issu du gradient du potentiel électrostatique. En effet, le champ électrique « accélérateur » pertinent provient du gradient d'énergie potentielle des porteurs, c'est à dire des niveaux d'énergie fournis par la résolution de l'équation de Schrödinger dans la direction de confinement (axe z). On remarque que l'équation de transport (axe x) issue de l'approche par mode (cf. paragraphe 3.1.4) prend bien en compte les évolutions des niveaux

d'énergie des porteurs et non celles du potentiel électrostatique. L'évolution typique de ces niveaux est représentée en Figure 138 dans d'un DGMOS, et la Figure 139 indique, seulement pour le premier niveau, l'évolution de la composante suivant  $x$  du champ électrique associé

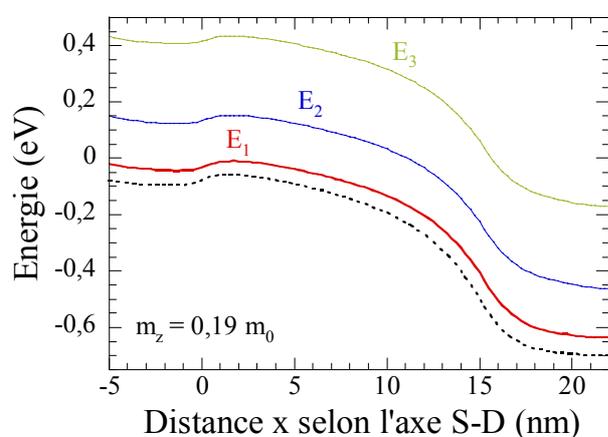


Figure 138 : Evolution typique des niveaux d'énergie des électrons selon l'axe  $x$  (source-drain) dans un DGMOS à l'état passant. En tirets : évolution de la bande de conduction moyennée sur les tranches Schrödinger (axe  $z$ ).

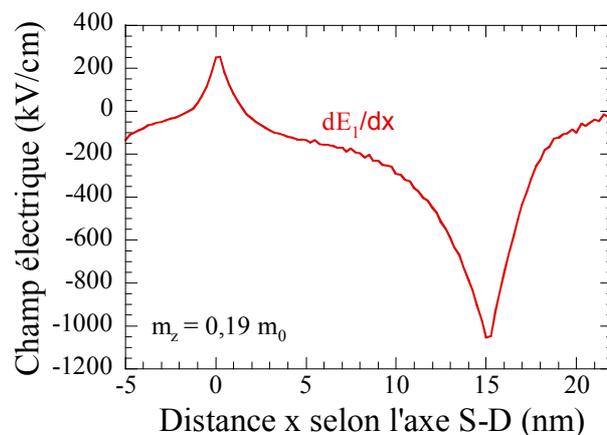


Figure 139 : Evolution typique du champ électrique ressenti par les électrons de la première sous-bande selon l'axe source-drain dans un DGMOS à l'état passant.

En 2Dk, le champ électrique accélérateur subi par les électrons dépend du niveau d'énergie auquel ils appartiennent. En effet, les niveaux peuvent a priori évoluer différemment selon qu'ils sont proches ou éloignés du sommet du puits de potentiel.

### 3.2.4 Calcul des interactions 2Dk à la volée

Les niveaux d'énergie pouvant évoluer, les fréquences d'interaction doivent être recalculées aux cours de la simulation. Le calcul des fréquences d'interaction « à la volée », c'est-à-dire à chaque fois que cela est nécessaire, supprime tous les problèmes de tabulations en énergie des fréquences d'interaction rencontrés parfois dans MONACO classique (effets de seuil...), mais le temps de calcul en pâtit encore une fois...

Pour limiter l'accroissement de la durée des simulations à cause des nombreuses intégrales numériques présentes dans les expressions des fréquences d'interaction (en particulier celles associées aux impuretés ionisées), ces fréquences sont malgré tout conservées en mémoire durant un certain nombre d'itérations. Pour suivre l'évolution des niveaux d'énergie et des fonctions d'onde, les fréquences mémorisées sont périodiquement remises à zéro. Contrairement à ce qui est fait dans les méthodes de tabulation, dans MONACO 2Dk, les fréquences d'interaction pour des niveaux d'énergie jamais atteints par les porteurs ne sont jamais ni calculées, ni a fortiori mémorisées.

Il faut noter que la place mémoire nécessaire à la sauvegarde des fréquences d'interaction  $2Dk$  est potentiellement beaucoup plus importante que dans le cas classique. En effet, un dispositif typique possède une centaine de mailles dans la direction du transport donc autant de tranches Schrödinger avec un jeu de fréquences d'interaction différentes pour chaque tranche. De plus, il y a plusieurs sous-bandes à considérer, habituellement une petite dizaine dans chaque vallée, de nombreuses transitions inter sous-bandes sont donc possibles. On peut donc facilement dépasser les tailles nécessaires au stockage des fréquences  $3Dk$  (de l'ordre du Méga octet) de trois ordres de grandeur... Pour conserver une occupation mémoire raisonnable, seules les fréquences d'interaction les plus souvent utilisées, c'est-à-dire associées aux interactions de faibles énergies (typiquement d'énergie inférieure à la tension d'alimentation  $V_{DD}$  dans des dispositifs fortement hors d'équilibre) sont conservées en mémoire.

### 3.2.5 Calcul des concentrations de porteurs $2Dk$

Pour résoudre l'équation de Poisson il faut préalablement avoir déterminé la concentration de porteurs dans les mailles du dispositif. L'algorithme de transport « habituel » Monte Carlo détermine le nombre d'électrons présents dans chaque tranche Schrödinger. Dans chaque tranche Schrödinger (axe  $z$  de confinement), les électrons sont délocalisés. La contribution d'un électron numéro 'i' appartenant à la sous-bande  $n(i)$  doit ainsi être répartie sur l'axe  $z$  en fonction de sa probabilité de présence dans la tranche Schrödinger  $x_j$  :  $|\xi_{n(i)}(x_j, z)|^2$ . Ensuite, cette contribution est pondérée par le temps qu'il a effectivement passé dans la tranche Schrödinger  $x_j$  :  $\Delta t_i(x_j)$  (et non plus dans la maille, cf. Figure 137) et normalisée par rapport à l'écart temporel  $\delta t$  entre deux résolutions de Poisson. Si le porteur n'a pas été présent dans la tranche durant  $\delta t$ ,  $\Delta t_i(x_j)$  est bien évidemment nul.

Finalement, la concentration en électrons dans la maille de coordonnées  $x_j$  et  $z_j$  est donnée par :

$$n(x_j, z_j) = \sum_i \frac{\Delta t_i(x_j)}{\delta t} \times |\xi_{n(i)}(x_j, z_j)|^2 \text{ avec } i \text{ balayant tous les électrons présents dans le dispositif.}$$

### 3.2.6 Problème de la référence de potentiel

Dans MONACO classique  $3Dk$  et MONACO  $2Dk$ , le potentiel de référence correspond au niveau maximum de la bande de conduction au contact ohmique de référence. Le calcul des positions des bandes de conduction en fonction des tensions appliquées aux diverses électrodes suppose les contacts ohmiques à l'équilibre. En conséquence, pour déterminer l'écart entre la bande

de conduction et le niveau de Fermi aux contacts ohmiques, la formule analytique classique (3Dk) suivante est appliquée :

$$E_F - E_{BC} = k_B T \times \ln(n_{3D} / N_c) \text{ où } N_{v/c} = 2 \left( \frac{2\pi \cdot m_{t/e}^* \cdot k_B T}{h^2} \right)^{3/2} \text{ avec } m_{t/e}^* = \sqrt{m_x m_y} \text{ où } m_x \text{ et } m_y$$

sont les masses effectives de trous ou d'électrons dans les directions du transport (plan x,y). De plus, les bandes de conduction ainsi calculées sont supposées uniformes sur toute la surface des contacts ohmiques.

Dans les simulations 2Dk, ces conditions aux limites ne sont plus valides car on ne peut pas utiliser de relation biunivoque entre le niveau de Fermi au contact et le haut de la bande de conduction qui d'ailleurs évolue sur la surface du contact. En l'absence de solutions analytiques satisfaisantes, une résolution préalable des équations couplées de Schrödinger/Poisson est donc nécessaire pour déterminer le profil de la bande conduction à appliquer aux mailles ohmiques, toujours supposées à l'équilibre.

### 3.2.7 Zone 2Dk.

On nomme « zone 2Dk », la région du dispositif où s'effectuent les résolutions des équations de Schrödinger 1D (les fonctions d'ondes s'annulent au-delà de cette zone). Cette zone s'étend sur toute la longueur du film de silicium (selon l'axe x) depuis le contact de source jusqu'à celui du drain. En effet, dans les dispositifs ultrafins, les porteurs présents dans la totalité des zones actives se trouvent confinés. Par conséquent, la cohabitation d'électrons 2Dk et 3Dk qui nécessiterait l'intégration des transitions entre 3Dk et 2Dk (cf. [MONS02]) n'est donc pas gérée actuellement par MONACO 2Dk. En outre, lorsqu'une géométrie particulière des accès de source et de drain induit une telle cohabitation, les caractéristiques électriques semblent peu affectées si l'on tient compte des nombreux phénomènes diffusifs possibles dans les caissons [VENU04].

Comme illustré en Figure 136, on peut choisir d'inclure l'oxyde dans la zone 2Dk de MONACO 2Dk. Par la suite, cette inclusion pourrait être utile pour déterminer le courant de fuite tunnel par la grille... Cependant pour respecter l'approche par mode ( $d\xi_n(x,z)/dx \approx 0$ ), il faut veiller à ne pas faire varier trop brutalement la taille (selon l'axe z) des fenêtres Schrödinger le long du dispositif. Il faut donc faire pénétrer les fonctions d'onde de façon identique dans l'oxyde jouxtant les caissons source-drain et dans l'oxyde de grille comme illustré en Figure 136.



Schrödinger/Poisson 1D pour une capacité MOS de mêmes épaisseurs d'oxyde et de film de silicium. Dans MONACO 2Dk le nombre de porteurs dans une tranche Schrödinger ainsi que le niveau auquel ils sont associés dépend de l'algorithme de transport.

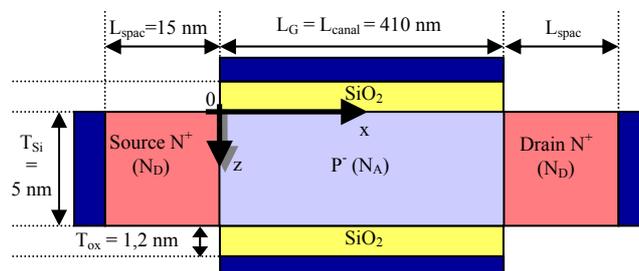


Figure 141 : Schéma en coupe du DGMOS « long ».

La Figure 142 permet de comparer les profils de concentration obtenus grâce à MONACO 2Dk (en traits continus) à ceux obtenus grâce aux résolutions Schrödinger/Poisson 1D (en tirets) à diverses polarisations de grille. Les profils très voisins indiquent que la localisation des charges issue de MONACO 2Dk est très satisfaisante. De même, l'évolution de la charge d'inversion surfacique extraite au milieu du dispositif 2D suit celle de l'algorithme 1D indiquée en tirets sur la Figure 143.

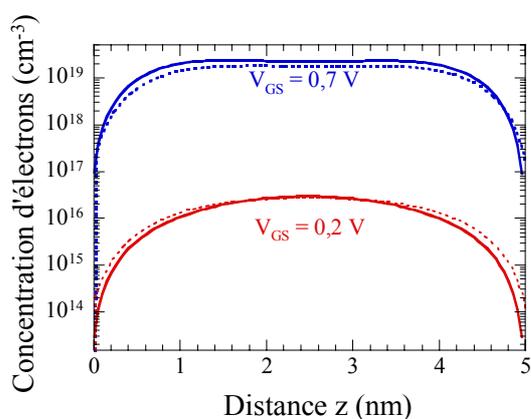


Figure 142 : Profils de concentration d'électrons dans une tranche Schrödinger pour différentes polarisations de grille. Traits continus : résultats issus de MONACO 2Dk au milieu d'un DGMOS long ( $x = L_{\text{canal}}/2$ ) avec  $L_{\text{canal}} = L_G = 460$  nm,  $T_{\text{Si}} = 5$  nm et à faible  $V_{\text{DS}} = 0,1$  V. Tirets : résultats issus de l'algorithme Schrödinger/Poisson dans une capacité MOS à double grille de même dimension.

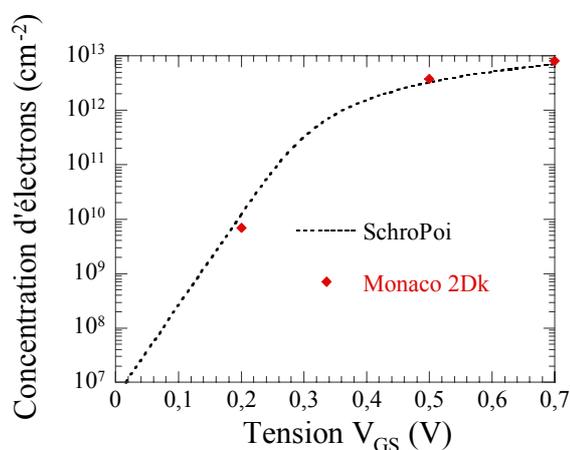


Figure 143 : Evolution de la concentration surfacique d'électrons dans une tranche perpendiculaire en fonction de la tension de grille au milieu d'un DGMOS long ( $x = L_{\text{canal}}/2$ ) avec  $L_{\text{canal}} = L_G = 460$  nm,  $T_{\text{Si}} = 5$  nm et à faible  $V_{\text{DS}} = 0,1$  V. Tirets noirs : résolutions couplées Schrödinger/Poisson, losanges rouges : MONACO 2Dk.

Ces résultats partiels sont très encourageants et montrent que MONACO 2Dk permet de résoudre de façon auto-cohérente les équations de Schrödinger et de Poisson et qu'il reproduit convenablement l'électrostatique d'un gaz 2Dk confiné. En outre, une étude plus systématique (mais nécessitant un temps de calcul d'autant plus important) est en cours.

### 3.3.2 Dispositif de référence

Nous avons simulé un transistor à double grille, dit de référence et schématisé en Figure 144, de longueur de canal  $L_{\text{canal}} = L_G = 15 \text{ nm}$ , d'épaisseur d'oxyde  $T_{\text{ox}} = 1,2 \text{ nm}$  et dont l'épaisseur du film actif de silicium vaut  $T_{\text{Si}} = 5 \text{ nm}$ . Dans ce paragraphe, nous comparons les modélisations 2Dk et 3Dk.

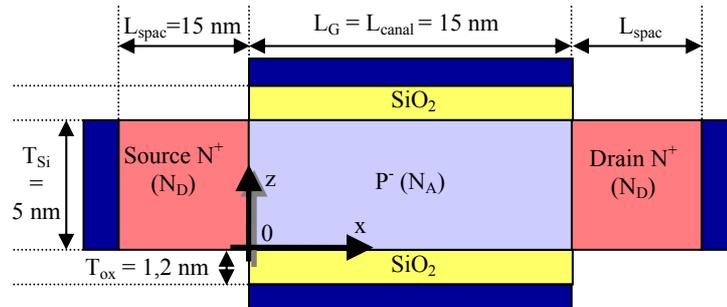


Figure 144 : Schéma en coupe du DGMOS de référence.

A l'état passant ( $V_{\text{DS}} = V_{\text{GS}} = 0,7 \text{ V}$ ), les cartographies de potentiel 3Dk et 2Dk ont été tracées respectivement en Figure 145 et en Figure 146. Les différences entre les 2 modélisations sont nettes ; les porteurs 2Dk sont repoussés des interfaces aussi bien dans le canal que dans la source et le drain. Pour compléter les cartographies, les profils 1D de concentrations au milieu du canal ( $x = L_{\text{canal}}/2$ ) dans la direction de confinement sont présentés en Figure 147. Contrairement au cas 3Dk où les inversions surfaciques se produisent sous les oxydes de grille, le maximum de la concentration 2Dk est centré sur le milieu du dispositif. Le transistor 2Dk fonctionne donc en régime d'inversion volumique.

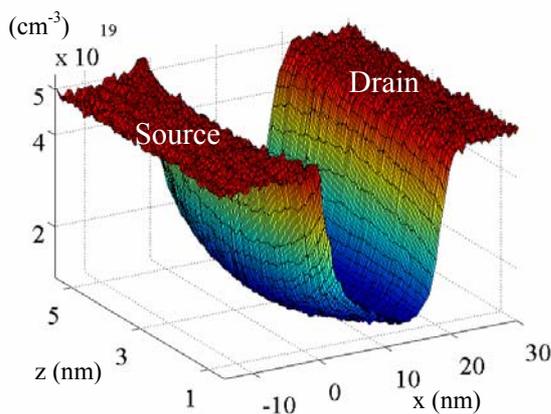


Figure 145 : Cartographie 2D de la concentration d'électrons issue de MONACO classique pour le DGMOS de référence à  $V_{\text{GS}} = V_{\text{DS}} = 0,7 \text{ V}$ .

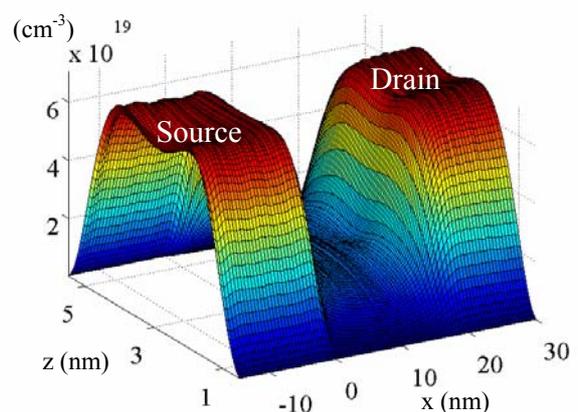


Figure 146 : Cartographie 2D de la concentration d'électrons issue de MONACO 2Dk pour le DGMOS de référence à  $V_{\text{GS}} = V_{\text{DS}} = 0,7 \text{ V}$ .

Dans la direction du transport, les évolutions de la concentration surfacique d'électrons sont indiquées en Figure 148. Les modélisations 2Dk/3Dk diffèrent peu dans les caissons de source et de

drain. En effet, la résolution de l'équation de Poisson garantit la neutralité électrique dans chaque tranche Schrödinger (direction  $z$ ).

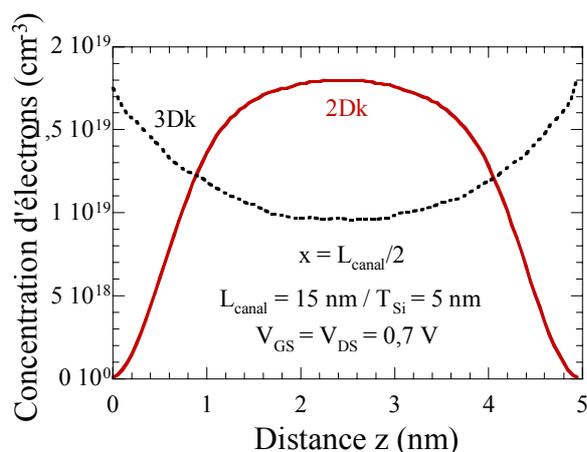


Figure 147 : Profils de concentration d'électrons selon l'axe  $z$  au milieu du canal ( $x = L_{\text{canal}}/2$  pour le DGMOS de référence à  $V_{\text{GS}} = V_{\text{DS}} = 0,7$  V. Tirets rouges : classique, Traits continus bleus : 2Dk.

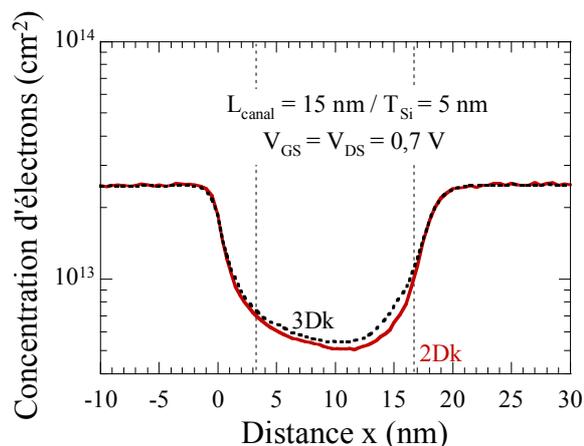


Figure 148 : Profils de concentration d'électrons selon l'axe  $x$  au milieu du canal ( $z = T_{\text{Si}}/2$ ) pour le DGMOS de référence à  $V_{\text{GS}} = V_{\text{DS}} = 0,7$  V. Tirets rouges : classique, Traits continus bleus : 2Dk.

Cependant, on observe que dans le canal la densité surfacique 2Dk est légèrement plus faible que la 3Dk. Ces résultats sont cohérents avec les évolutions de la charge totale 3Dk/2Dk en fonction de la tension de grille  $V_{\text{GS}}$  à faible  $V_{\text{DS}}$ . Dans le dispositif de référence, la charge 2Dk est toujours inférieure à la charge 3Dk, révélant une tension de seuil 2Dk plus importante qu'en 3Dk comme on le reverra lors de l'étude des caractéristiques I-V. De plus, le contrôle de charges, caractérisé par la pente de la caractéristique  $Q(V_{\text{GS}})$  (ou capacité de grille  $C_{\text{G}}$ ) reportée en Figure 149, est moins efficace en gaz 2Dk. En effet, la tension de grille est moins efficace pour peupler la zone active 2Dk mais aussi pour la dépeupler.

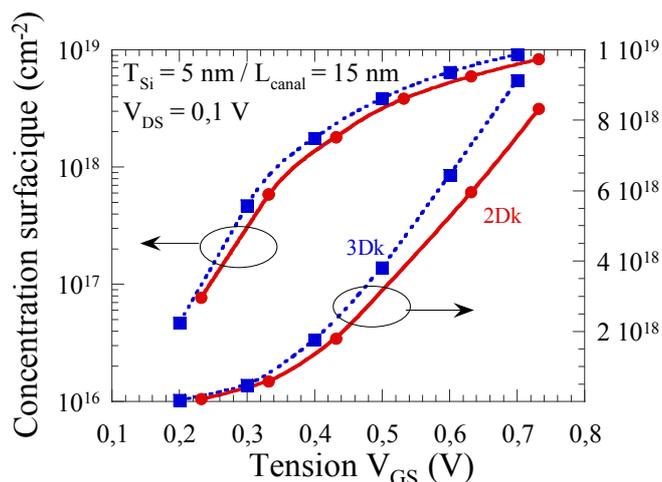


Figure 149 : Evolution de la concentration surfacique en fonction de la tension de grille  $V_{\text{GS}}$  au milieu ( $x = L_{\text{canal}}/2$ ) du DGMOS de référence à  $V_{\text{DS}} = 0,1$  V. Tirets rouges : classique, traits continus bleus : 2Dk.

Les différences entre les cartographies des bandes de conduction 3Dk et 2Dk, respectivement présentées en Figure 150 et en Figure 151, sont moins évidentes que celles précédemment

observées sur les concentrations. On note toutefois une courbure du potentiel suivant  $z$  dans les caissons de source et de drain due au confinement des porteurs dans ces régions créant une zone de charge d'espace sur les bordures. De même dans le canal, la concavité de la bande de conduction au milieu de la zone active est accentuée ; ce phénomène est plus visible sur les coupes relevées au milieu du canal et reportées sur la Figure 152.

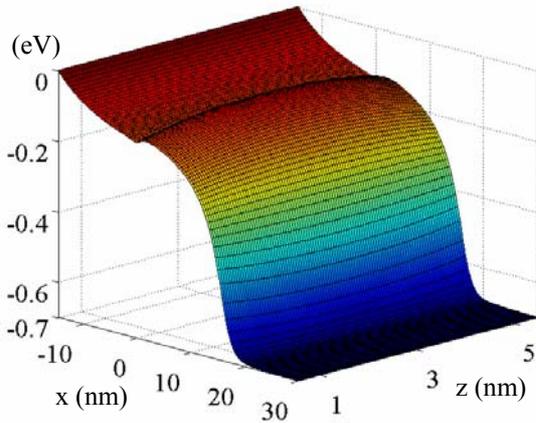


Figure 150 : Cartographie 2D de la bande de conduction issue de MONACO 3Dk pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V.

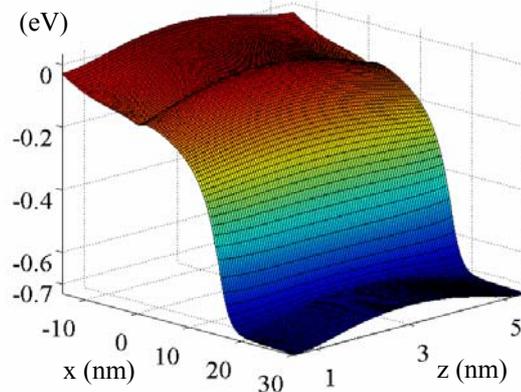


Figure 151 : Cartographie 2D de la bande de conduction issue de MONACO 2Dk pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V.

Pour comparer les évolutions des profils des bandes de conduction 3Dk et 2Dk selon l'axe source/drain tracées en Figure 153, les bandes de conduction ont été moyennées sur une tranche Schrödinger. En 2Dk, les évolutions de niveaux d'énergie le long des dispositifs qui déterminent le champ électrique accélérateur vu par les porteurs ont été superposées.

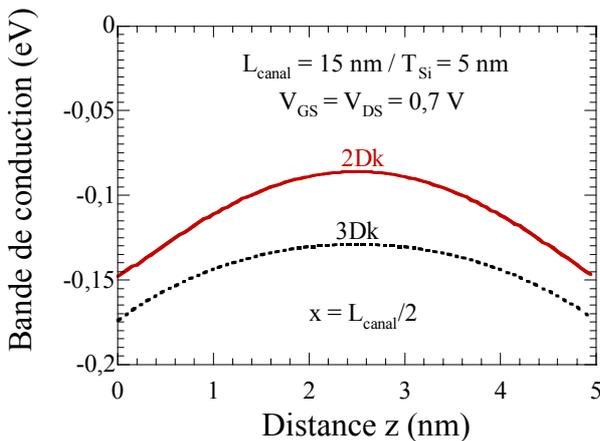


Figure 152 : Profils de la bande de conduction selon l'axe  $z$  au milieu du canal ( $x = L_{canal}/2$  pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V. Tirets rouges : classique, traits continus bleus : 2Dk.

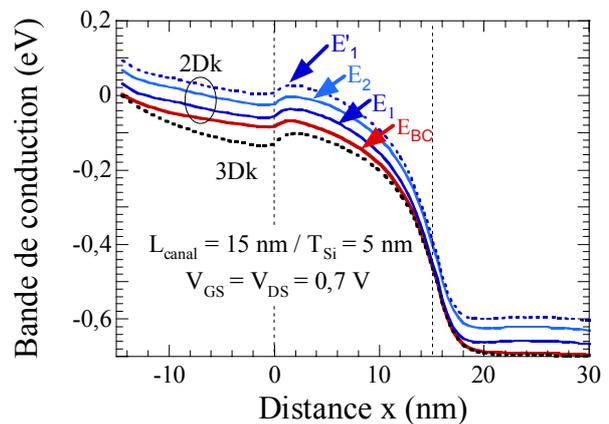


Figure 153 : Profils de la bande de conduction selon l'axe  $x$  au milieu du canal ( $z = T_{Si}/2$ ) pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V. Tirets rouges : classique, traits continus bleus : 2Dk et évolutions de quelques niveaux.

Tout d'abord, on constate que dans la source les évolutions 2Dk et 3Dk ne sont pas les mêmes : la bande conduction 2Dk est toujours au dessus de la 3Dk. Bien que dans un canal dopé à  $5.10^{19}$  at./cm<sup>3</sup> les mobilités 2Dk et 3Dk soient proches ( $\mu_{2Dk} = \mu_{3Dk} \times 70\%$ , cf. Figure 134), les

résistances d'accès sont différentes dans les sources :  $R_{\text{accès\_2Dk}} = 42 \Omega\mu\text{m} = R_{\text{accès\_3Dk}} \times 55\%$ . En effet, dans des dispositifs aussi courts les différences de potentiel encaissées par la source sont de l'ordre de 100 mV sur 15 nm, elles correspondent à des champs électriques de plusieurs dizaines de kV/cm. La mobilité faible champ est donc peu pertinente pour comparer les résistances d'accès de dispositifs ultracourts... Lors de l'étude des caractéristiques I-V, ces différences de résistances d'accès devront être prises en compte. Au niveau du drain, les chutes de potentiel restent négligeables quel que soit le type de modélisation.

En entrée de canal, la barrière d'injection 3Dk moyenne (selon l'axe z) s'avère être plus grande que les barrières 2Dk (pour chaque niveau d'énergie). Même si les barrières 3Dk sont plus basses sous l'oxyde, en 2Dk, la conduction volumique est favorisée. De plus, les évolutions des niveaux d'énergie le long du canal sont parallèles. Par conséquent, les champs électriques de chaque sous-bande, indiqués en Figure 154, sont quasiment identiques dans ce dispositif de référence.

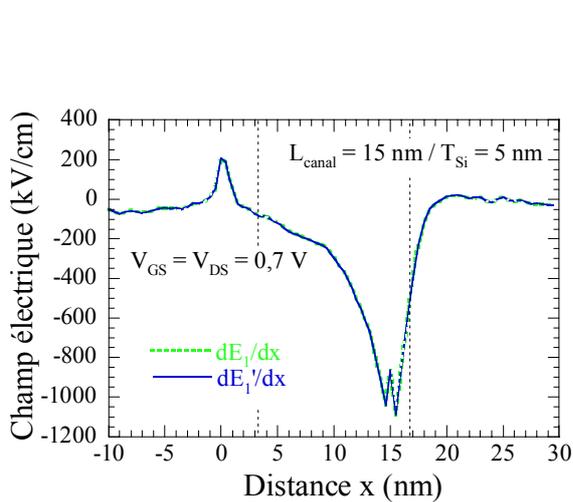


Figure 154 : Evolution des différents champs électriques 2Dk selon l'axe x pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V.  $dE_1'/dx$  et  $dE_1/dx$  correspondent aux champs des niveaux les plus bas de chaque vallée (respectivement  $m_z = m_l$  et  $m_z = m_l$ ).

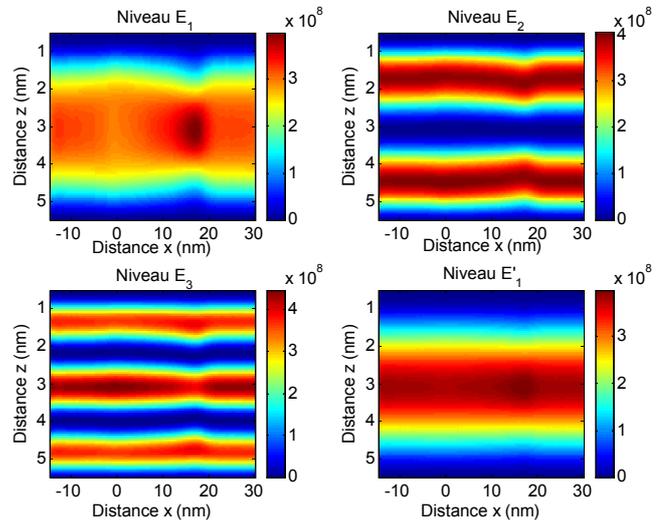


Figure 155 : Evolution des densités de présence ( $m^{-1}$ ) pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V.  $E_1$ ,  $E_2$  et  $E_3$  correspondent aux trois premiers niveaux des vallées où  $m_z = m_l$  et  $E'_1$  au premier niveau des vallées où  $m_z = m_r$ .

Pour justifier a posteriori notre découplage selon la direction du transport et celle du confinement (approche « mode-space » cf. annexe), il nous faut vérifier que les fonctions d'onde associées aux différents niveaux d'énergie (ou mode) restent à peu près constantes le long du transistor. Dans ce but, la Figure 155 présente l'évolution des probabilités de présence des électrons pour les 4 premiers niveaux d'énergie du dispositif de référence à l'état passant. On peut remarquer une légère évolution de ces probabilités dans les régions de forts champs, c'est-à-dire en entrée de canal et, surtout, en fin de canal. Cependant, ces perturbations restent faibles :

$\max\left(\frac{\partial \xi_m}{\partial x}(x_0, z), \frac{\partial^2 \xi_m}{\partial x^2}(x_0, z)\right) < \max(\xi_n(x_0, z))/100$  (cf. annexe). On en déduit que la forme du

puits de potentiel reste tout le long du dispositif de type puits carré. En conséquence, le découplage par mode se révèle être adapté à l'étude de l'architecture à double grille (de référence).

### 3.3.3 Transport :

Pour étudier le transport dans notre dispositif double grille de référence, nous nous sommes intéressés seulement à l'effet des phonons et des impuretés dans les caissons. Les effets de la rugosité ont en effet été négligés, ils n'apparaissent d'ailleurs pas être dominants dans les simulations à champ constant pour des dispositifs de 5 nm d'épaisseur (cf. Figure 135).

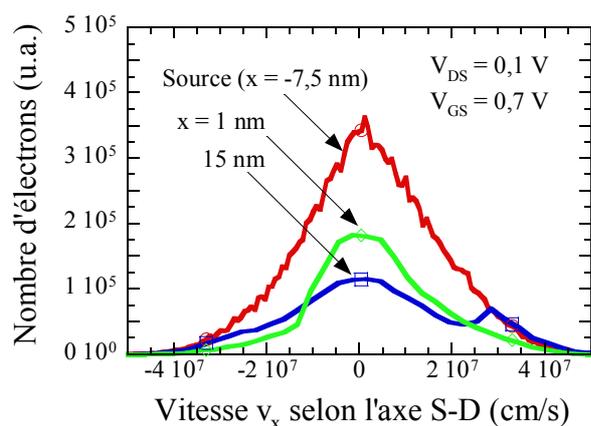


Figure 156 : Spectres en vitesse 2Dk pour le DGMOS de référence à  $V_{GS} = 0,7$  V et  $V_{DS} = 0,1$  V. En rouge (cercles) : dans la source ( $x = -10$  nm), en vert (losanges) : début du canal avant la barrière d'injection ( $x = 1$  nm) et en bleu (carrés) : fin de canal ( $x = 15$  nm).

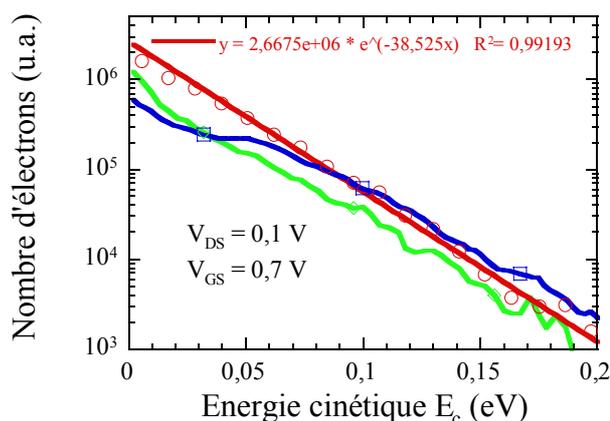


Figure 157 : Spectres en énergie 2Dk pour le DGMOS de référence à  $V_{GS} = 0,7$  V et  $V_{DS} = 0,1$  V. En rouge (cercles) : dans la source ( $x = -10$  nm), en vert (losanges) : début du canal ( $x = 1$  nm) et en bleu (carrés) : fin de canal ( $x = 15$  nm).

#### Faible $V_{DS}$

Pour étudier l'évolution des spectres dans un cas faiblement non-stationnaire nous avons faiblement polarisé le drain :  $V_{DS} = 0,1$  V. Les spectres en vitesse sont tracés en Figure 156. Les spectres en vitesse au milieu de la source sont gaussiens (du moins aussi gaussiens que dans les simulation 3Dk). Dans la zone active, on retrouve les évolutions des spectres en vitesse décrites au chapitre II : en début de canal, la barrière « filtre » les électrons à vitesse négative et en fin de canal on distingue un pic d'électrons qui correspond aux électrons de masse transverse ultra majoritaires dans un gaz confiné. Les spectres en énergie correspondants, reportés en Figure 157, restent sensiblement proches de l'équilibre sauf en fin de canal. En effet, la décroissance exponentielle du spectre dans la source correspond à une température électronique de 301,5 K ( $k_B \times T = 26,0$  meV) très proche de la température ambiante : 300 K.

#### Fort $V_{DS}$

Pour étudier un transport fortement non stationnaire, nous nous sommes placés à l'état passant :  $V_{GS} = V_{DS} = 0,7$  V. La Figure 158, présentant l'évolution de la balisticité intrinsèque (cf. chapitre

II) 2Dk et 3Dk, indique que la balisticité est très peu affectée par la prise en compte du caractère 2D du gaz électronique. Le transport 2Dk apparaît toutefois légèrement plus balistique.

Les évolutions des énergies cinétiques moyennes en 2Dk comme en 3Dk sont présentées en Figure 159. Quelle que soit la modélisation, les électrons sont un peu « chauds » dans la source. Au début du dispositif, les porteurs 2Dk sont moins énergétiques que les 3Dk car on ne prend pas en compte leur énergie dans la direction de confinement, de plus, la chute de tension est plus faible dans la source en 2Dk. De plus, dans une zone proche de l'équilibre (la source) ou d'accélération des porteurs (« chauffage » dans le canal), plus le niveau d'énergie de quantification est élevé plus l'énergie cinétique est faible. Cependant cette répartition s'inverse dans le drain qui est une zone de forte décélération (« refroidissement »). En effet, dans cette zone pour perdre de l'énergie cinétique les porteurs transfèrent vers des niveaux plus élevés grâce aux interactions.

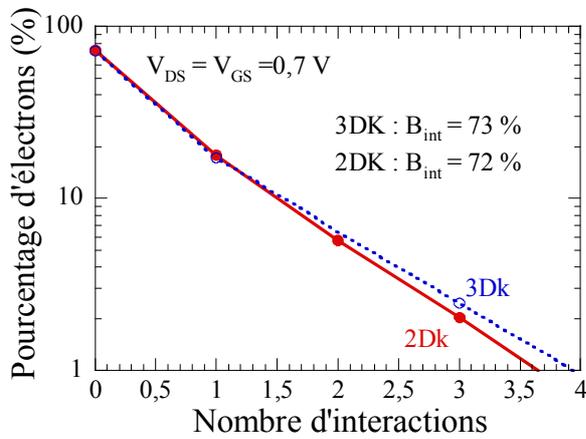


Figure 158 : Spectres d'interactions pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V. Tirets rouges : classique, traits continus bleus : 2Dk.

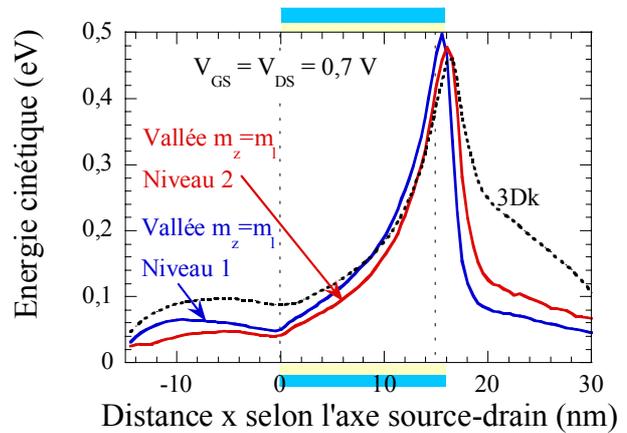


Figure 159 : Evolution de l'énergie cinétique moyenne selon l'axe  $x$  pour le DGMOS de référence à  $V_{GS} = V_{DS} = 0,7$  V. Tirets : simulation classique, traits continus : simulations 2D. Traits rouges : énergie cinétique des porteurs appartenant au niveau 1 de la vallée 1 avec  $m_z = m_b$ , traits bleus : niveau 2 vallée 1.

Ces premiers résultats indiquent que MONACO 2Dk est opérationnel. De plus, ils suggèrent que les transports 2Dk et 3Dk dans les dispositifs ultracourts d'épaisseur 5 nm sont relativement proches. Globalement, les conclusions du chapitre II ne devraient pas être modifiées par la prise en compte du gaz 2Dk. En particulier, le transport dans les dispositifs ultracourts devrait rester très fortement balistique.

### 3.3.4 Courant : Dispo 2Dk vs 3Dk

Après nous être intéressés aux aspects microscopiques, nous allons examiner les caractéristiques électriques de 2 dispositifs à double grille. Nous avons ainsi étudié le dispositif double grille de référence ( $L_{\text{canal}} = L_G = 15$  nm,  $T_{Si} = 5$  nm et  $T_{ox} = 1,2$  nm) et un double grille de mêmes dimensions mais plus fin :  $T_{Si} = 3$  nm. Seules les interactions avec les phonons et les impuretés ont été prises en

compte (pas de rugosité). Les caractéristiques  $I_D(V_{GS})$  sont tracées en Figure 160 pour le dispositif de référence et en Figure 161 pour le dispositif étroit. Les mêmes caractéristiques en échelle semi logarithmique sont indiquées respectivement en Figure 162 et en Figure 163.

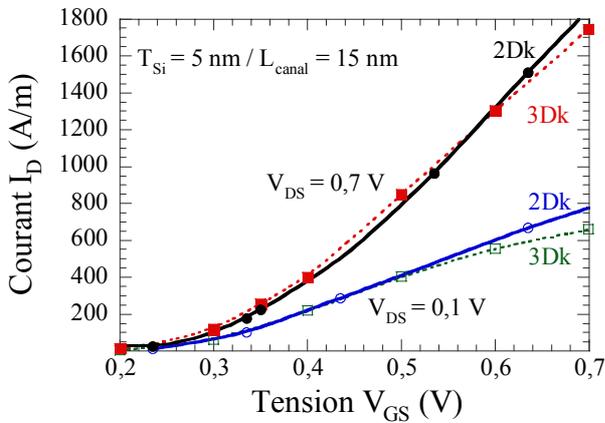


Figure 160 : Caractéristiques  $I_D(V_{GS})$  pour le DG MOS de référence à  $V_{DS} = 0,1$  V et  $0,7$  V. Tirets : classique, traits continus : 2Dk.

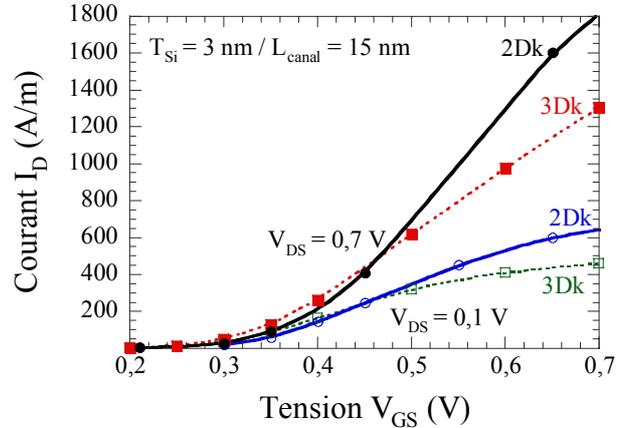


Figure 161 : Caractéristiques  $I_D(V_{GS})$  pour le DG MOS de référence à  $V_{DS} = 0,1$  V et  $0,7$  V. Tirets : classique, traits continus : 2Dk.

Tout d'abord, on constate que la tension de seuil dans le dispositif de référence ( $T_{Si} = 5$  nm) reste en première approche inchangée par la modélisation :  $V_{T\_2Dk} \approx V_{T\_3Dk} \approx 0,28$  V ( $\pm 10$  mV, méthode du maximum de la transconductance). Par contre, dans le dispositif plus étroit on observe que la tension de seuil  $V_{T\_2Dk} \approx 0,33$  V est augmentée par la quantification :  $V_{T\_3Dk} \approx 0,30$  V.

Malgré la valeur des tensions de seuil et des mobilités plus faibles, le courant 2Dk est toujours plus important que le courant 3Dk à fort  $V_{GS}$ , c'est-à-dire avec une meilleure transconductance. L'importance des résistances d'accès  $R_{accès}$  3Dk explique en partie ce phénomène. En effet, dans ces deux dispositifs, on constate que, sur les courbes à faibles  $V_{DS}$ , l'effet des résistances d'accès est très significatif sur les dispositifs 3Dk.

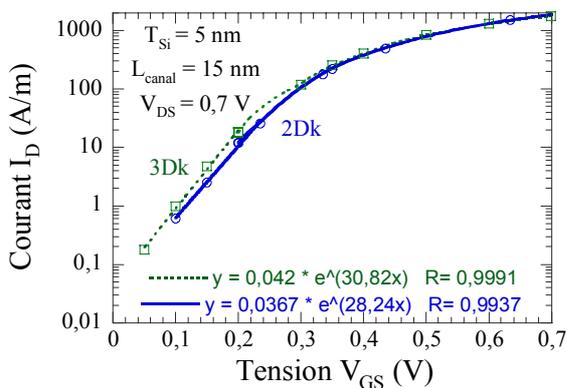


Figure 162 : Caractéristiques  $\text{Log}[I_D(V_{GS})]$  dans le DG MOS de référence ( $T_{Si} = 5$  nm) à  $V_{DS} = 0,7$  V. Tirets verts : classique, traits continus bleus : 2Dk.

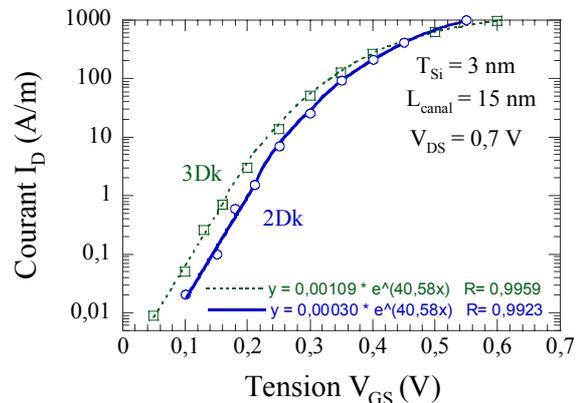


Figure 163 : Caractéristiques  $\text{Log}[I_D(V_{GS})]$  dans un DG MOS fin ( $T_{Si} = 3$  nm) à  $V_{DS} = 0,7$  V. Tirets verts : classique, traits continus bleus : 2Dk.

Le dispositif fin d'ailleurs profite au mieux de la forte mobilité observée sur la Figure 132. Il débite un courant à l'état passant équivalent à celui du dispositif de 5 nm. Cependant, dans un

dispositif plus réaliste l'effet des interfaces rugueuses devrait fortement dégrader ses performances si l'on se réfère à l'évolution de la mobilité (cf. Figure 135).

Le comportement sous le seuil s'avère légèrement dégradé par la quantification perpendiculaire dans le cas du dispositif de référence :  $S_{2DK} = 82 \text{ mV/dec} > S_{3DK} = 75 \text{ mV/dec}$ . Entre les 2 modélisations, à courant sous le seuil constant, on mesure un décalage de tension de seuil  $\Delta V_{T_{2DK/3DK}}$  inférieur à 10 mV cohérent avec la détermination effectuée au dessus du seuil. Dans le dispositif étroit, de très grand rapport  $L_{\text{canal}}/T_{\text{Si}}$ , les pentes sous le seuil 3Dk et 2Dk restent idéales :  $S_{2DK} = S_{3DK} = 60 \text{ mV/dec}$ . Dans ce dispositif très étroit, le décalage de la tension de seuil est plus important :  $\Delta V_{T_{2DK/3DK}} = 30 \text{ mV}$  (extrait à courant constant en régime sous le seuil).

D'après ces premiers résultats, le comportement des dispositifs ne semble pas profondément modifié par la prise en compte de la quantification. Cependant, une étude beaucoup plus exhaustive doit être entreprise pour vérifier l'impact de la quantification perpendiculaire sur les règles de dimensionnement des transistors multigrille issues de modélisations purement classiques.

#### **4. Conclusions et perspectives :**

La variante de MONACO 2Dk prenant en compte la quantification dans le sens perpendiculaire au transport réalisée au cours de cette thèse est totalement achevée d'un point de vue informatique. Les profils de concentrations et les décalages de tensions de seuil obtenus semblent satisfaisants, néanmoins la modélisation du transport 2Dk est encore en développement. Notre description des interactions 2Dk doit être affinée : paramètre d'écrantage, potentiel de déformation des phonons acoustiques, interaction avec les interfaces rugueuses... et surtout confrontée à des résultats expérimentaux dans des structures à double grille, actuellement encore trop peu nombreux dans la littérature.

D'après nos premiers résultats, la quantification perpendiculaire ne semble pas modifier sensiblement les performances des dispositifs étroits en termes de contrôle des effets de canal court par rapport aux modélisations purement classiques. De plus, la « balisticité » des dispositifs apparaît être très peu affectée par l'aspect spécifique des interactions en gaz 2Dk, les conclusions des études classiques du chapitre II ne semblent donc pas être contredites. Les dispositifs ultimes devraient donc fonctionner très près de la limite balistique.

De nombreuses évolutions du logiciel sont envisagées. Tout d'abord d'un point de vue numérique des optimisations de l'algorithme sont possibles afin d'améliorer la rapidité du simulateur. Limiter le calcul aux valeurs propres les plus faibles serait une bonne piste. D'ailleurs, un perfectionnement de la résolution de l'équation de Schrödinger [LEPA96] sera impératif pour

simuler des dispositifs 3D avec une résolution 2D (dans la direction perpendiculaire) de l'équation de Schrödinger. Du point de vue de la modélisation physique, des diagrammes de bandes plus réalistes pourraient être utilisés. Par exemple, la résolution de l'équation de Schrödinger pourrait utiliser une bande non parabolique [GOME05]. L'utilisation d'un diagramme de bande complet [FISC88] [BUFL03] permettrait de gérer idéalement le transport d'un gaz 2Dk de trous. Enfin, l'utilisation d'un algorithme approprié [BORO05] permettrait de gérer plus rigoureusement les effets de dégénérescence.

Afin d'apprécier l'influence des effets quantiques ainsi que l'impact réel des interactions, nous allons comparer les résultats obtenus avec MONACO 2Dk à d'autres approches ondulatoires avec ou sans interactions.

Cependant, pour des valeurs de  $L_{\text{canal}} < 10$  nm, le courant tunnel Source/Drain devient trop important [RHEW02] rendant l'approximation semi-classique dans le sens du transport inacceptable. Pour inclure ces phénomènes dans MONACO 2Dk, on envisage de superposer des courants tunnel Source/Drain et canal/grille calculés, par exemple, grâce une méthode de transmission analytique telle que WKB [MERZ61]. Mais peut-être que d'ici là, d'autres méthodes purement quantiques (fonctions de Green, de Wigner...) seront plus efficaces pour traiter le transport nanométrique...



## *Conclusion générale*

Pour continuer la course à la miniaturisation des composants, la microélectronique ne cesse, depuis ses débuts, de surmonter des obstacles considérés initialement comme infranchissables. Néanmoins, le nombre de défis à relever actuellement pour lutter contre les effets de canal court dans les prochaines générations de transistors MOS est particulièrement impressionnant. Comme il a été détaillé dans le premier chapitre de ce mémoire, il devient nécessaire de réaliser, d'une part, l'intégration de nouveaux matériaux (utilisation de matériaux contraints dans la zone active, de métaux pour électrodes de grilles et des « high-k » comme isolants de grille) et, d'autre part, d'envisager le développement d'architectures à grille multiple et à zone active ultrafine.

Dans ce contexte, le travail présenté dans ce mémoire et réalisé grâce au simulateur particulière MONACO avait trois objectifs : la description du transport fortement non-stationnaire apparaissant dans les dispositifs ultracourts de fin de roadmap, l'étude des performances électriques des nouvelles architectures à grille multiple et enfin la réalisation d'une nouvelle version du logiciel MONACO capable de prendre en compte la quantification dans la direction perpendiculaire au transport de charges.

Tout d'abord, les simulations Monte Carlo « classiques » présentées dans le chapitre II ont permis de mettre clairement en évidence la transition entre les régimes de transport quasi stationnaires et quasi balistiques. Dans les transistors à double grille et à zone active non dopée qui ont été simulés, cette transition apparaît pour des longueurs de canal  $L_{\text{canal}}$  d'environ 50 nm. A ces dimensions, plus de 10% des électrons traversent le canal sans subir d'interactions ( $B_{\text{int}} \geq 10\%$ ). Pour des transistors ultracourts ( $L_{\text{canal}} = 15$  nm) cette proportion dépasse les 50% ( $B_{\text{int}} > 50\%$ ). Ainsi le transport dans les dispositifs de fin de roadmap devrait être fortement balistique, et cela même à faible tension de drain  $V_{\text{DS}}$ . En outre, notre étude a montré que, contrairement aux idées largement répandues, les interactions subies dans la seconde partie du canal sont d'autant plus préjudiciables au courant que le canal se miniaturise. En termes de courant à l'état passant des transistors, les simulations Monte Carlo prévoient une valeur de courant à plus de 80% de la limite balistique dans les dispositifs tels que  $L_{\text{canal}} = 15$  nm ( $B_{\text{eff}} = 80\%$ ). Les mesures que j'ai effectuées au CEA-LETI fournissent des résultats légèrement moins optimistes que les simulations Monte Carlo, cependant la balisticité effective  $B_{\text{eff}}$  des transistors de fin de roadmap devrait largement dépasser les 50%. En conséquence, même si le transport purement balistique ( $B_{\text{int}} = 100\%$ ) ne peut être atteint, les MOSFET ultimes devraient délivrer des courants très proches de la limite balistique ( $B_{\text{eff}} > 50\%$ ).

Dans le chapitre III, l'étude des caractéristiques électriques de dispositifs, obtenues grâce à la version « semi-classique » de MONACO, a montré que les architectures multigrilles sur SOI permettent d'améliorer nettement le pouvoir bloquant des transistors. Malheureusement, les dispositifs à grille multiple sont intrinsèquement plus lents que les transistors à simple grille. En considérant l'ensemble de ces paramètres, les architectures MOS à double grille sont apparues être de très bons compromis entre comportement sous le seuil acceptable et rapidité. Ils devraient donc succéder aux architectures à simple grille pour les générations où  $L_{\text{canal}} \approx 20$  nm, prévues à l'horizon 2008. En outre, la géométrie des caissons doit impérativement être optimisée pour contrôler les impédances parasites ; elles prennent, en effet, de plus en plus d'importance à mesure que les dimensions des accès de source et drain (longueur et épaisseur) sont réduites pour accompagner la miniaturisation.

La variante de MONACO 2Dk prenant en compte la quantification dans le sens perpendiculaire au transport, présentée dans le dernier chapitre et réalisée au cours de cette thèse, est totalement achevée d'un point de vue informatique. Les profils de concentrations et les décalages de tensions de seuil obtenus semblent satisfaisants. Néanmoins, la modélisation du transport 2Dk reste perfectible. Le modèle d'interaction avec les interfaces rugueuses a besoin d'être affiné, mais pour ce faire des mesures détaillées et concordantes de mobilité dans des structures à double grille sont nécessaires. D'après nos premiers résultats, la quantification perpendiculaire ne semble pas dégrader sensiblement les performances sous le seuil de dispositifs dimensionnés à l'aide d'approches purement classiques. De plus, le transport dans les dispositifs ultracourts semble être peu affecté par les modifications de mobilité liées au passage d'une modélisation 3Dk à une version 2Dk. Les conclusions des chapitre II et III devraient rester valides dans le cas de gaz 2Dk. Pour confirmer ces résultats, cette variante de MONACO devra être utilisée pour comparer les performances des dispositifs de l'électronique "ultime" : dispositifs à couche de silicium contraint, MOSFET à double, triple ou quadruple grille...

Dans l'avenir, il semble que les applications nécessiteront désormais des architectures de plus en plus spécifiques : les dispositifs de faibles consommations seront multigrilles, les hautes performances plutôt simple grille... Elles demanderont de grands efforts de recherche et de développement. L'industrie se souviendra alors avec nostalgie du temps où les transistors sur silicium massif suffisaient à remplir l'essentiel des fonctions.

D'un point de vue plus conceptuel, cette miniaturisation ne peut pas se poursuivre de manière infinie, une limite physique existe. L'obtention de canaux de moins de 10 nm compromet le fonctionnement classique des dispositifs MOS. L'importance du courant tunnel source-drain n'est pas compatible avec les spécifications de courant de fuite sous le seuil nécessaire au fonctionnement des circuits logiques « traditionnels ». En outre, la réduction des dimensions à ces échelles

décananométriques ne permettra plus d'améliorer les performance des transistors :  $B_{\text{eff}} = 90\%$  pour  $L_{\text{canal}} = 12 \text{ nm}$  d'après des simulations Monte Carlo classiques (cf. [BOUR06]) ; on devra donc chercher ailleurs les motivations de nouvelles miniaturisations.

Pour faire face aux insatiables besoins en performance, une électronique alternative doit donc se développer. On pense à l'électronique moléculaire et en particulier aux transistors à nanotube de carbone qui sont les plus aboutis à l'heure actuelle [BAND05]. Des dispositifs de conceptions radicalement nouvelles tirant partie de la granularité de la charge tels que le transistor à un électron ou SET [SEE03], ou utilisant le spin [BOUR06] [ZUTI04] sont autant de candidats potentiels.



# *ANNEXES*



## ANNEXE 1 : Modèle balistique : cas dégénéré

Comme schématisé sur la Figure 164, pour calculer le courant dans un canal balistique, on utilise le formalisme de Landauer dans lequel le courant  $I_D$  dans un transistor MOS s'écrit comme la différence d'un courant  $I^+$  provenant de la source et d'un courant  $I^-$  provenant du drain :

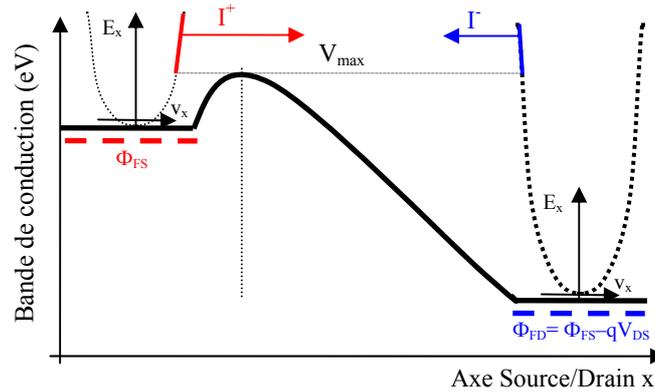
$$I_D = I^+ - I^-$$


Figure 164 : Evolution typique de la bande de conduction le long de l'axe source drain dans un transistor balistique en trait continu. Les tirets en gras symbolisent les niveaux de Fermi  $\Phi_{FS}$  et  $\Phi_{FD}$  respectivement dans les caissons de Source et de Drain. Les relations de dispersions paraboliques  $E_x(v_x)$  sont indiquées en tirets fins. Les états de la courbe  $E_x(v_x)$  intervenant dans le calcul des courants  $I^+$  et  $I^-$  sont surlignés.

En considérant les différentes vallées et les niveaux multiples, il vient :

$$I_D = q \sum_{\text{vallée}} \sum_{\text{niveaux}} \int \left[ v^+ \cdot D^+(E) \times f(\Phi_{FS}, E) (1 - f(\Phi_{FD}, E)) - v^- \cdot D^-(E) \times f(\Phi_{FD}, E) (1 - f(\Phi_{FS}, E)) \right] T(E) \cdot dE$$

Avec

$E$  : énergie des porteurs,

$v^+$ ,  $v^-$  : vitesse de groupe respectivement dans la source et le drain,

$D^+(E)$ ,  $D^-(E)$  : densité d'états respectivement dans la source et le drain,

$\Phi_{FS}$  : niveau de Fermi dans la source,

$\Phi_{FD}$  : niveau de Fermi dans le drain,

$f(\Phi_F, E)$  : probabilité d'occupation d'un état d'énergie  $E$  avec  $\Phi_F$  comme niveau de Fermi,

$T(E)$  : coefficient de transmission d'un porteur injecté dans le canal à l'énergie  $E$ .

Dans le canal d'un transistor MOS balistique, aucune interaction ne vient dégrader le coefficient de transmission  $T$  du canal.  $T$  est donc soit nul, soit unitaire. En conséquence, tous les porteurs présents dans la source à vitesse positive (direction source/drain) et possédant une énergie  $E$

supérieure à la barrière  $V_{\max}$  (selon les notations de la Figure 164) vont contribuer au courant  $I^+$ . De même, tous les électrons suffisamment énergétiques du drain vont contribuer à  $I^-$ .

Dans l'approximation WKB, le produit de la vitesse de groupe et de la densité d'état est constant [NATO94] :  $v^{+/-} \times D^{+/-}(E) = (\pi\hbar)^{-1}$ . Ensuite, en développant la formule de Landauer dans le cas d'une statistique de Fermi-Dirac et ne considérant que la première sous-bande (bandes paraboliques), K. Natori a obtenu la caractéristique courant/tension balistique suivante :

$$I_D = W \cdot I_0 \cdot [F_{1/2}(u) - F_{1/2}(u - v_d)] \text{ avec } F_{1/2}(u) = \int_0^\infty \frac{\sqrt{y}}{1 + e^{y-u}} dy$$

$$u = \Phi_{FS} - E_{\max} = \ln \left[ \sqrt{(1 + e^{v_d})^2 + 4 \cdot e^{v_d} (e^p - 1)} - (1 + e^{v_d}) \right] - \ln 2 \text{ avec } v_d = \frac{q \cdot V_{DS}}{k_B \cdot T}$$

$$\rho = \frac{2 \cdot \pi \cdot \hbar^2 \cdot C_{eq} (V_G - V_T)}{q \cdot k_B \cdot T \cdot m^* \cdot M_v}$$

$$I_0 = \frac{\sqrt{2} q (k_B \cdot T)^{3/2}}{\pi^2 \hbar^2} \cdot M_v \sqrt{m^*} .$$

avec  $m^*$  masse effective des porteurs et  $M_v$  dégénérescence du premier niveau.

Dans ce modèle le courant de saturation obtenu à forte tension de drain ( $V_{DS} \gg k_B T/q$ ) vaut :

$$I_{Dsat} = W \cdot I_0 \cdot F_{1/2}(\ln[e^p - 1]) .$$

Si on pose :  $I_{Dsat} = W \cdot v_{inj} \cdot C_{eq} (V_{GS} - V_T)$  avec  $v_{inj}$  vitesse d'injection,

dans le cas de porteurs dégénérés ou à faibles températures ( $\rho \gg 1$ ), la vitesse d'injection vaut :

$$v_{inj} = \frac{8 \cdot \hbar \cdot \sqrt{C_{eq} (V_G - V_T)}}{3 \cdot m^* \sqrt{q \cdot \pi \cdot M_v}} .$$

Dans le cas de porteurs non dégénérés ou à hautes températures ( $\rho \ll 1$ ), la vitesse d'injection vaut :

$$v_{inj} = \sqrt{\frac{2 \cdot k_B T}{\pi \cdot m^*}} = v_{th} . (\text{vitesse thermique}).$$

On remarque que dans le cas non dégénérés la vitesse d'injection  $v_{inj}$  est indépendante de la tension de grille  $V_{GS}$ .

Une extension de ce modèle aux cas multi sous-bandes a été développée plus récemment [NATO03].

## ANNEXE 2 : Calcul du coefficient de rétrodiffusion

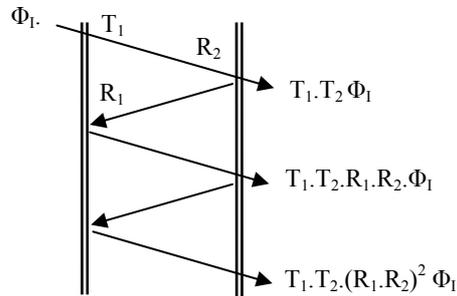


Figure 165 : Schéma de la transmission après 2 interfaces réfléchissantes

Comme suggéré par le schéma Figure 165, le coefficient de transmission total  $T(2)$  à travers une suite de deux interfaces réfléchissantes de coefficients de transmissions  $T_1$  et  $T_2$  et de réflexions  $R_1$  et  $R_2$  est tel que :  $T(2) = T_1 T_2 \times [1 + R_1 R_2 + (R_1 R_2)^2 + (R_1 R_2)^3 + \dots] = T_1 T_2 / (1 - R_1 R_2)$ .

En remarquant que  $T_1 = 1 - R_1$  et  $T_2 = 1 - R_2$ , le résultat précédant peut se mettre sous la forme additive suivante :

$$\frac{1 - T(2)}{T(2)} = \frac{1 - T_1}{T_1} + \frac{1 - T_2}{T_2}$$

On en déduit que le coefficient de transmission total  $T(N)$  d'une série de  $N$  interfaces chacune ayant le même coefficient de transmission  $T$  s'écrit :

$$\frac{1 - T(N)}{T(N)} = \sum_i \frac{1 - T_i}{T_i} = N \frac{1 - T}{T} \text{ d'où } T(N) = \frac{T}{N(1 - T) + T}$$

Finalement, une succession de  $N$  interfaces réfléchissantes présente un coefficient de rétrodiffusion total  $R$  donné par la relation suivante [DATT95] :

$$R = 1 - T(N) = \frac{L_{\text{eff}}}{L_{\text{eff}} + L_0} \text{ avec } L_{\text{eff}} \text{ longueur de la zone de réfléchissante et } L_0 = \frac{T}{(1 - T)v} \text{ où}$$

$v = N/L_{\text{eff}}$  est le nombre d'interfaces réfléchissantes par unité de longueur dans la zone de rétrodiffusion.



## ***ANNEXE 3 : Approche par modes ou « mode-space »***

### Présentation

Dans l'approche par mode ou « mode-space », on choisit de représenter les fonctions d'onde des électrons, a priori quelconques, sous la forme [REN01] :  $\Psi(x,z) = \sum_n \tilde{\Psi}_n(x) \times \xi_n(x,z)$  [1]

avec  $\xi_n(x,z)$  fonction d'onde associée à l'état propre, ou mode, numéro 'n' d'énergie  $E_n(x)$  dans une tranche de coordonnée x. On a donc par définition :

$$-\frac{\hbar^2}{2m_z} \frac{\partial^2}{\partial z^2} \xi_n(x,z) - qV(x,z) \times \xi_n(x,z) = E_n(x) \times \xi_n(x,z) \quad [2]$$

Ces fonctions forment une base orthogonale :  $\int \xi_n(x,z) \times \xi_m(x,z) dz = \delta_{m,n}$  [3] avec  $\delta_{m,n}$  symbole de Kronecker. Toute fonction  $\Psi(x,z)$  peut donc être bien est décrite par [1]. D'après ces notations,  $|\tilde{\Psi}_n(x)|^2$  représente la probabilité de trouver un électron sur le mode (état propre) 'n' de la tranche Schrödinger située à la position x.

Grâce à ce choix de fonctions d'onde, on peut, sous certaines conditions qui vont être déterminées dans la suite, découpler l'équation de Schrödinger **2D** stationnaire :

$$-\frac{\hbar^2}{2m_x} \frac{\partial^2 \Psi(x,z)}{\partial x^2} - \frac{\hbar^2}{2m_z} \frac{\partial^2 \Psi(x,z)}{\partial z^2} - qV(x,z) \Psi(x,z) = E \times \Psi(x,z), \quad [4]$$

en **2** équations **1D** selon l'axe z (équation [2]) et selon l'axe x où il faut résoudre l'équation suivante :  $-\frac{\hbar^2}{2m_x} \frac{d^2}{dx^2} \tilde{\Psi}_n(x) - E_n(x) \tilde{\Psi}_n(x) = E \times \tilde{\Psi}_n(x)$  [5]

Alors, l'équation de transport (selon l'axe x) et le calcul des modes (selon l'axe z) sont totalement découplés.

### Justification

Pour déterminer les conditions nécessaires à la réalisation du découplage des équations [2] et [5], commençons par multiplier l'équation [4] par  $\delta(x-x_0) \times \xi_n(x,z)$  et intégrons sur l'espace réel selon les axes x et z, il vient la relation [6] suivante :

$$-\iint \delta^*(x-x_0) \xi_n^*(x,z) \left[ -\frac{\hbar^2}{2m_x} \frac{\partial^2}{\partial x^2} \Psi(x,z) \right] dx dz$$

$$\begin{aligned}
& - \iint \delta^*(x - x_0) \xi_n^*(x, z) \left[ \frac{\hbar^2}{2m_z} \frac{\partial^2}{\partial z^2} - qV(x, z) \right] \Psi(x, z) dx dz \\
& = E \times \iint \delta^*(x - x_0) \xi_n^*(x, z) \Psi(x, z) dx dz
\end{aligned} \tag{6}$$

Par définition de la distribution de Dirac  $\delta(x)$  et grâce aux relations [1] et [3], le second membre de l'équation [6] peut s'écrire :

$$E \times \int \xi_n^*(x_0, z) \Psi(x_0, z) dz = E \sum_m \tilde{\Psi}(x_0) \int \xi_n^*(x_0, z) \xi_m(x_0, z) dz = E \times \tilde{\Psi}(x_0) \tag{7}$$

De même, le second terme du premier membre de l'équation [6] peut se simplifier en utilisant l'équation [2] :

$$\begin{aligned}
& \iint \delta^*(x - x_0) \xi_n^*(x, z) \left[ \frac{\hbar^2}{2m_z} \frac{\partial^2}{\partial z^2} - qV(x, z) \right] \Psi(x, z) dx dz = \int \xi_n^*(x_0, z) \left[ \frac{\hbar^2}{2m_z} \frac{\partial^2}{\partial z^2} - qV(x_0, z) \right] \Psi(x_0, z) dz \\
& = \sum_m \tilde{\Psi}_m(x_0, z) \int \xi_n^*(x_0, z) \left[ \frac{\hbar^2}{2m_z} \frac{\partial^2}{\partial z^2} - qV(x_0, z) \right] \xi_m^*(x_0, z) dz \\
& = \sum_m \tilde{\Psi}_m(x_0, z) \int \xi_n^*(x_0, z) E_m(x_0) \xi_m(x_0, z) dz = E_n(x_0) \tilde{\Psi}_n(x_0, z)
\end{aligned} \tag{8}$$

Le premier terme du premier membre de l'équation [6], peut s'écrire :

$$\begin{aligned}
& \iint \delta^*(x - x_0) \xi_n^*(x, z) \left[ \frac{\hbar^2}{2m_x} \frac{\partial^2}{\partial x^2} \Psi(x, z) \right] dx dz = \frac{\hbar^2}{2m_x} \int \xi_n^*(x_0, z) \left[ \frac{\partial^2}{\partial x^2} \sum_m \tilde{\Psi}_m(x_0) \xi_m(x_0, z) \right] dz \\
& = \frac{\hbar^2}{2m_x} \sum_m \left( \int \xi_n^*(x_0, z) \xi_m(x_0, z) dz \frac{\partial^2 \tilde{\Psi}_m(x_0)}{\partial x^2} + 2 \int \xi_n^*(x_0, z) \frac{\partial \xi_m}{\partial x}(x_0, z) dz \frac{d\tilde{\Psi}_m}{dx}(x_0) + \tilde{\Psi}_m(x_0) \int \xi_n^*(x_0, z) \frac{\partial^2 \xi_m}{\partial x^2}(x_0, z) dz \right)
\end{aligned}$$

On constate que si l'on a  $\frac{\partial \xi_n^*(x, z)}{\partial x} \approx 0$  l'équation se simplifie en :

$$\iint \delta^*(x - x_0) \xi_n^*(x, z) \left[ \frac{\hbar^2}{2m_x} \frac{\partial^2}{\partial x^2} \Psi(x, z) \right] dx dz = \frac{\hbar^2}{2m_x} \frac{d^2 \tilde{\Psi}}{dx^2}(x_0) \tag{9}$$

Dans cette approximation, les éventuels couplages entre les modes dus aux termes :

$$\int \xi_n^*(x_0, z) \frac{\partial \xi_m}{\partial x}(x_0, z) dz \text{ et } \int \xi_n^*(x_0, z) \frac{\partial^2 \xi_m}{\partial x^2}(x_0, z) dz \text{ sont négligés...}$$

Finalement si  $\frac{\partial \xi_n^*(x, z)}{\partial x} \approx 0$ , en regroupant les relations [6] à [9], on constate que l'équation [4]

est équivalente au système formé par les équations [2] et [5]. L'approche par mode permet donc de découpler le problème 2D selon les axes  $x$  et  $z$  tant que les fonctions d'onde associées aux modes

n'évoluent pas trop ( $\frac{\partial \xi_m}{\partial x}(x_0, z) \ll \xi_n(x_0, z)$  et  $\frac{\partial^2 \xi_m}{\partial x^2}(x_0, z) \ll \xi_n(x_0, z)$ ) le long du dispositif.

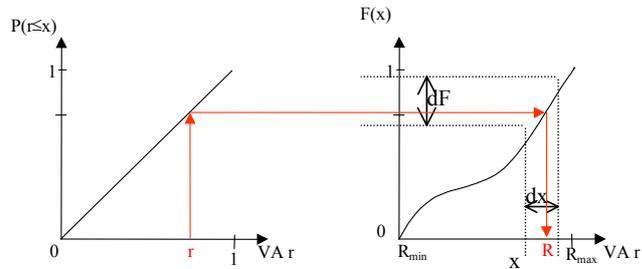
Néanmoins, les énergies des modes  $E_n(x)$  peuvent évoluer librement...

## ANNEXE 4 : Tirages au sort dans MONACO

### Principe des tirages au sort

Le problème est d'obtenir par tirage au sort la valeur de variables aléatoires (VA) qui suivent des lois de probabilité quelconque.

Par exemple partons d'une variable aléatoire  $r$  qui suit une loi de probabilité uniforme. On affecte alors à  $R$  la valeur  $R=F^{-1}(r)$  (ou  $r=F(R)\dots$ ), comme indiqué en Figure 166.



*Figure 166 : Exemple de tirage au sort d'une variable aléatoire qui a une fonction de répartition  $F(x)$  à partir d'une variable aléatoire uniforme.*

La probabilité  $dP$  que  $R$  tombe dans l'intervalle  $dx$  est égale à celle que  $r$  tombe dans l'intervalle  $dF$ , qui vaut  $dF$  ( $r$  est une VA de densité de probabilité uniforme) d'où

$$dP = dF = \frac{dF}{dx} dx$$

La VA  $R$  ainsi obtenue suit bien une loi de densité de probabilité  $f(x) = \frac{dF}{dx}$  et a  $F(x)$  pour fonction de répartition. (Par définition de  $f(x)$  :  $dP = \text{Pr ob}(x \leq R \leq x + dx) = f(x).dx$ )

L'outil informatique nous donne facilement accès à une variable aléatoire  $r$  qui suit une loi de probabilité uniforme. On en déduit alors facilement, dans les cas simples, un tirage de la VA  $R$  de fonction de répartition  $F$  :  $R=F^{-1}(r)$ .

### Calcul des fréquences d'interaction

La fréquence d'interaction est obtenue par intégration de  $s(\vec{k}, \vec{k}')$  sur toutes les états finaux possibles, c'est-à-dire les valeurs possibles de  $\vec{k}'$  :

$$\lambda(E) = \frac{1}{(2\pi)^2/A_c} \times \iint s(\vec{k}, \vec{k}') d^2\vec{k}' = \frac{A_c}{(2\pi)^2} \iint s(\vec{k}, \vec{k}') J(E', \theta') .dE' .d\theta' .$$

Dans cette expression  $E'$  est l'énergie finale (après interaction),  $\theta$  est l'angle horaire,  $J(E, \theta)$  est le Jacobien de transformation entre un repère cartésien et un repère sphérique, il est obtenu en

considérant la relation de dispersion  $E(k)$ . En gaz 2Dk, dans le cas d'une vallée ellipsoïdale non parabolique, le Jacobien  $J(E,\theta)$  a pour expression :  $J(E,\theta) = \frac{1}{\hbar^2} \left[ \frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_y} \right]^{-1} (1 + 2.\alpha.E)$ .

Dans le cas, d'une interaction faisant passer le porteur d'un état  $(E, \vec{k})$  à un état final  $(E', \vec{k}')$ . Les différentes composantes  $k'_x$  et  $k'_y$  du vecteur d'onde final  $\vec{k}'$  sont données par les expressions suivantes :

$$k'_x = K_{\max} \times \cos(\theta')$$

$$k'_y = (m_y'/m_x')^{1/2} \times K_{\max} \times \sin(\theta')$$

avec

$$K_{\max} = (k_x^2 + m_x/m_y \times k_y^2)$$

$\theta'$  angle entre l'axe x et  $\vec{k}'$ .

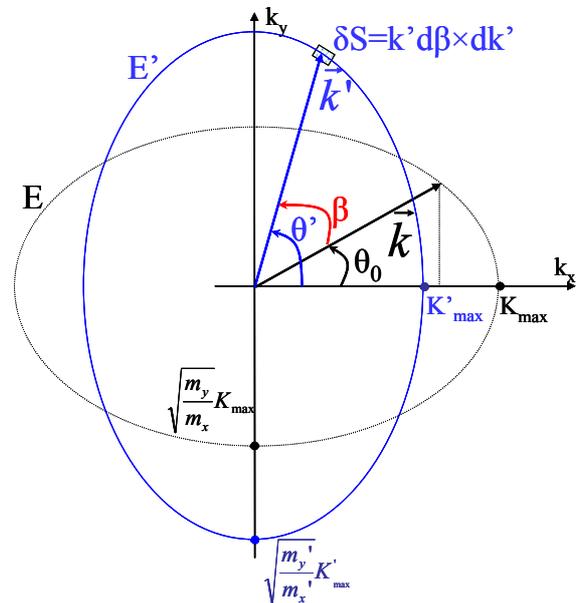


Figure 167 : Représentation du vecteur d'onde final  $\vec{k}'$  et du vecteur initial  $\vec{k}$  dans le cas d'une interaction élastique entre deux vallées elliptiques.

Finalement, les fréquences d'interaction 2Dk peuvent se mettre sous la forme :

$$\lambda_{2Dk}(\vec{k}) = \frac{1}{2\pi\hbar^3} (1 + 2.\alpha.E') \int_0^{2\pi} \left| \int \xi_m^*(z) \xi_n^*(z) H_i \cdot dz \right| e^{-i(\vec{k}-\vec{k}') \cdot \vec{r}} \times d^2\vec{r} \times \left[ \frac{\cos^2(\beta + \theta_0)}{m'_x} + \frac{\sin^2(\beta + \theta_0)}{m'_y} \right]^{-1} d\beta$$

où  $\beta$  est l'angle en  $k'$  et  $k$  et  $\theta_0$  l'angle entre  $k$  et l'axe x.

On notera que :  $\int_0^{2\pi} \left[ \frac{\cos^2 \theta}{m_x} + \frac{\sin^2 \theta}{m_y} \right]^{-1} d\theta = 2\pi \sqrt{m_x m_y}$

### Tirage au sort du type d'interactions subies par le porteur à la fin du temps de vol libre

On effectue un tirage au sort de l'interaction subie à la fin du vol libre grâce à celui d'une variable aléatoire uniforme et de la fonction de répartition de la probabilité d'interactions comme indiqué sur la Figure 168.

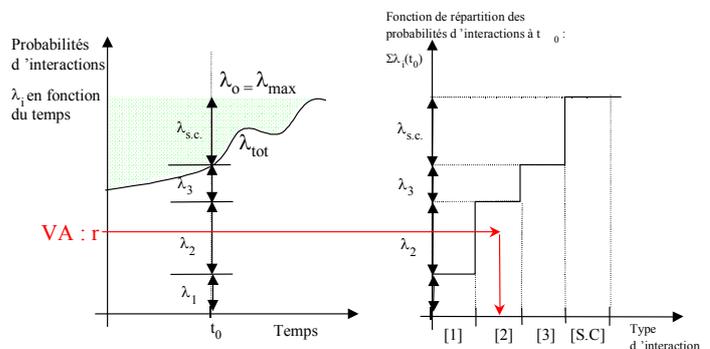


Figure 168 : Tirage au sort du type d'interactions.

## Effet des interactions

Connaissant la fréquence de l'interaction, il faut ensuite s'intéresser à ses effets. De manière générale, la norme (dans le cas d'interactions inélastiques) et/ou la direction du vecteur d'onde initial peuvent être modifiées comme illustré en Figure 167. Pour décrire l'effet d'une interaction, nous devons introduire la valeur de la modification d'intensité associée à l'interaction  $\Delta|k|$  et la densité de probabilité  $P_E(\beta)$  pour que l'angle de déviation du vecteur d'onde soit compris entre  $\beta$  et  $\beta+d\beta$ .

Dans le cas d'un électron  $2Dk$  appartenant à une vallée sphérique de vecteur  $\vec{k}$  donné, sa densité de probabilité d'avoir un vecteur final  $\vec{k}'$  de norme  $k'$  et dévié d'un angle  $\beta$  est donnée par  $s(\beta, k') = s(\vec{k}, \vec{k}')$ . Selon les notations du schéma de la Figure 167, on peut écrire :

$$P_E(\beta)d\beta = \frac{\int_0^\infty s(\beta, k') \delta S(k') dk'}{\int_0^\infty \int_0^{2\pi} s(\beta, k') \delta S(k') dk' d\beta} = \frac{\int_0^\infty s(\vec{k}, \vec{k}') k' dk' d\beta}{\int_0^\infty \int_0^{2\pi} s(\vec{k}, \vec{k}') k' dk' d\beta}$$

Dans le cas d'une interaction à effet isotrope ( $s(\beta, k')$  est indépendant de  $\beta$  et  $k'$ ), on obtient  $P_E(\beta)d\beta = \frac{1}{2\pi}$ . Pour considérer une déviation **isotrope** dans une vallée **ellipsoïdale** ( $P_E(\beta)d\beta \neq 1/2\pi$ ), on peut se ramener au cas uniforme (vallée sphérique) en utilisant la transformation d'« Herring-Vogt » (cf. [MONS02]).

## Calcul de la loi de probabilité du temps de vol libre

Soit :

$\lambda_{\text{tot}}(t).dt$  la probabilité que le porteur subisse une interaction entre  $t$  et  $t+dt$

$dq(t)$  la probabilité que le porteur ait un vol libre de durée  $t$  et qu'il subisse une interaction entre  $t$  et  $t+dt$ .

De plus  $Q(t) = \int_t^\infty \frac{dq(u)}{du} du$  est la probabilité du porteur de n'avoir aucune interaction avant  $t$  (avec

ou sans interaction plus tard). C'est la probabilité d'avoir un temps de vol libre de durée  $t$ .

Si l'on considère que les chocs et les vols libres sont indépendants on peut évaluer la probabilité  $dq(t)$  :

$P(\text{'temps de vol libre de durée } t \text{' et 'interaction entre } t \text{ et } t+dt\text{'}) = P(\text{temps de vol libre de durée}) \times P(\text{interaction entre } t \text{ et } t+dt)$

$$\text{Soit : } dq(t) = Q(t) \times \lambda_{\text{tot}}(t)$$

$$\text{En remarquant que } \int_{q(t=0)}^{q(t=\infty)} dq = 1 \text{ alors } \frac{dQ(t)}{dt} = \frac{d}{dt} \left( \int_0^t \frac{dq(u)}{du} du \right) = \frac{d}{dt} \left( 1 - \int_t^\infty \frac{dq(u)}{du} du \right) = -dq(t)$$

Donc

$$\frac{-dQ(t)/dt}{Q(t)} = \lambda_{\text{tot}}(t) \text{ et finalement } Q(t) = \exp\left[-\int_0^t \lambda_{\text{tot}}(t).dt\right]$$

Il n'est possible d'expliciter de manière analytique l'expression du temps de vol libre  $t_v$  par l'équation précédente que si la probabilité totale est constante en fonction du temps. C'est le cas par exemple si on utilise un pas sur le temps très petit :  $\delta t \ll t_v$ . Sinon on utilise la méthode du « self-scattering » qui consiste à remplacer le  $\lambda_{\text{tot}}(t)$  variable par une fréquence  $\lambda_0$  constante. Pour ce faire on introduit une interaction fictive dite de « self-scattering », sans influence sur la trajectoire du porteur quand elle est tirée au sort. On lui associe la fréquence d'interaction  $\lambda_{\text{sc}}$  qui est définie ainsi (cf. Figure 168) :  $\lambda_0 = \sum_i \lambda_i(E) + \lambda_{\text{sc}}(E)$ . En général, afin de limiter le temps de simulation, on choisit  $\lambda_0$  comme le maximum des  $\sum_i \lambda_i(E)$ .

On a alors une expression très simple du temps de vol en utilisant une variable aléatoire uniforme  $r$  :

$$t_v = -\frac{\text{Ln}(r)}{\lambda_0}$$

Remarque : On retrouve le paramètre d'un processus poissonien

## ANNEXE 5 : Interactions avec les impuretés ionisées

On suppose qu'une impureté ionisée génère un potentiel de perturbation coulombien non écranté

de la forme :  $V(r) = \frac{q^2}{4\pi\epsilon_0\epsilon_r r}$  (J) [1]

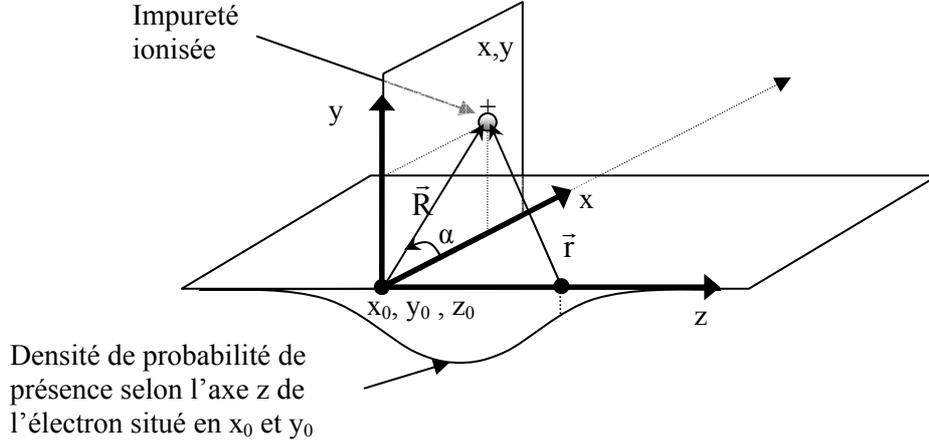


Figure 169 : Schéma de l'interaction entre un électron situé en  $x = x_0$  et  $y = y_0$  et délocalisé sur l'axe  $z$  selon la probabilité de présence indiquée le long des  $y$  négatifs. La distance  $\vec{r}$  qui sépare les porteurs est décomposé en  $z_0$  (direction de confinement) et en un vecteur  $\vec{R}$  dans le plan  $\{x,y\}$ .

Selon les notations de la Figure 169 et en considérant le potentiel de perturbation [1], la matrice de transition d'un électron dans un plan parallèle aux couches, repéré en  $z$  dans la direction de confinement peut s'écrire :

$$|M_{m,n}(\Delta\vec{k})| = \left| \langle n, \vec{k} | H_{IMP} | m, \vec{k}' \rangle \right| = \iiint \xi_m(z) \frac{q^2}{4\pi\epsilon_0\epsilon_r \sqrt{R^2 + (z-z_0)^2}} \xi_n(z) e^{-i\Delta\vec{k} \cdot \vec{R}} d\vec{R} dz \quad \text{avec}$$

$$\Delta k^2(\beta, \theta_0) = |\vec{k}' - \vec{k}|^2 \quad (\beta \text{ angle de déviation entre } \vec{k} \text{ et } \vec{k}', \text{ et } \theta_0 \text{ angle entre } \vec{k} \text{ et l'axe } x).$$

En utilisant les coordonnées cylindriques  $(R, \alpha)$  dans le plan  $(x,y)$  on peut écrire :

$$|M_{m,n}(\Delta\vec{k})| = \frac{q^2}{4\pi\epsilon_0\epsilon_r} \int \xi_m(z) \xi_n(z) \int \frac{R \times \int e^{-i\Delta k R \cos \alpha} d\alpha}{\sqrt{R^2 + (z-z_0)^2}} dR dz$$

En se rappelant la définition de l'intégrale de Bessel d'ordre 0, il vient :

$$|M_{m,n}(\Delta\vec{k})| = \frac{q^2}{4\pi\epsilon_0\epsilon_r} \int \xi_m(z) \xi_n(z) \int \frac{R \times 2\pi J_0(\Delta k R)}{\sqrt{R^2 + (z-z_0)^2}} dR dz$$

D'après l'expression des primitives reportées dans la référence [GRAD80] p. 682 équation ETII 7(4), on peut simplifier la matrice et la mettre sous la forme suivante :

$$\left| M_{m,n}(\Delta\vec{k}) \right|^2 = \left| \frac{q^2}{4\pi\epsilon_0\epsilon_r} \int \xi_m(z)\xi_n(z) \frac{2\pi e^{\Delta k(\beta)|z-z_0|}}{\Delta k(\beta)} dz \right|^2$$

On constate que la matrice de transition dépend de  $\Delta k$ , l'interaction est donc anisotrope.

La densité de probabilité qu'un électron ait son vecteur d'onde modifié de  $\vec{k}$  vers  $\vec{k}'$  à cause d'une impureté ionisée localisée en  $z_0$  vaut :

$$s(\vec{k}, \vec{k}') = \frac{2\pi}{\hbar} \left| \frac{q^2}{4\pi\epsilon_0\epsilon_r} \int \xi_m(z)\xi_n(z) \frac{2\pi e^{\Delta k|z-z_0|}}{\Delta k} dz \right|^2 \times \frac{\delta(E - E')}{4\pi^2}$$

En considérant une répartition uniforme des impuretés selon la direction de confinement et en supposant leurs effets indépendants, la densité de probabilité avec l'ensemble des atomes dopants est la somme de toutes les contributions soit :  $s_{\text{imp}}(\vec{k}, \vec{k}') = \int_{z_1}^{z_2} N_{\text{imp}}(z_0) \times s(\vec{k}, \vec{k}') \times dz_0$ .

Enfin, en prenant en compte les phénomènes d'écrantage, la fréquence d'interaction d'un électron de vecteur d'onde  $\vec{k}$  avec les impuretés ionisées peut se mettre sous la forme :

$$\frac{1}{\lambda_{m,n}^{\text{imp}}(\vec{k})} = \frac{q^4(1+2\alpha.E')}{8\pi\hbar^3\epsilon_{\text{sc}}^2} \int_0^{2\pi} \frac{1}{\cos^2(\beta + \theta_0) + \frac{m'_x}{m'_y} \sin^2(\beta + \theta_0)} \int_{z_1}^{z_2} N_{\text{imp}}(z_0) \left| \int_{z_1}^{z_2} \frac{\xi_m(z) e^{-\Delta k(\beta, \theta_0)|z-z_0|} \xi_n(z)}{\Delta k(\theta) + \Delta k_{\text{scr}}(\theta)} dz \right|^2 dz_0 d\beta$$

Avec  $\Delta k^2(\beta, \theta_0) = |\vec{k}' - \vec{k}|^2$  ( $\beta$  angle de déviation et  $\theta_0$  angle entre  $\vec{k}$  et l'axe x),  $E, E'$  : énergies cinétiques respectivement initiales et finales,  $\epsilon_{\text{sc}}$  : permittivité diélectrique du matériau,  $m'_x$  et  $m'_y$  : masses effectives dans la vallée,  $N_{\text{imp}}(z_0)$  : densité d'impuretés dopantes (at./m<sup>-3</sup>) à la profondeur  $z_0$ ,  $\Delta k_{\text{scr}}$  est l'inverse de la longueur d'écran (modèle de Debye-Hückel revisité par Fetter [MONS02]).

Pour conclure, on peut signaler que la gestion des interactions avec les impuretés ionisées pose des problèmes de mise en œuvre informatique à cause de leur anisotropie mais aussi des problèmes plus théoriques. En effet la longueur de corrélation et de temps d'interaction non négligeable devant le temps de vol libre rend la validité de l'approche par la règle d'or de Fermi discutable [HRUS02]

....

## ANNEXE 6 : Résolution de l'équation de Schrödinger

L'équation de Schrödinger 1D s'écrit :  $H(\psi_n(z)) = -\frac{\hbar^2}{2} \frac{\partial}{\partial z} \left[ \frac{1}{m^*(z)} \frac{\partial}{\partial z} \cdot + V(z) \right] \cdot \psi_n(z) = E_n \cdot \psi_n(z)$

On peut l'écrire sous la forme :

$$-\frac{\hbar^2}{2} \left\{ \frac{\partial}{\partial z} \left( \frac{1}{m^*(z)} \right) \frac{\partial}{\partial z} (\psi_n(z)) + \frac{1}{m^*(z)} \frac{\partial^2}{\partial z^2} (\psi_n(z)) \right\} + V(z) \cdot \psi_n(z) = E_n \cdot \psi_n(z)$$

En discrétisation sur la base des impulsions de Dirac  $\{\delta(z-z^i)\}$  avec  $\delta z^{i+1} = z^{i+1} - z^i$  ( $i \in [1..N]$ ,  $z^1$  et  $z^N$  limite de la fenêtre Schrödinger où  $\delta z^{N+1} = \delta z^N$ ). En approximant les opérateurs par les méthodes classiques de différence finie pour la dérivée première et seconde suivantes :

$$\frac{\partial}{\partial z} \left( \frac{1}{m^*(z)} \right) \approx \frac{1}{(\delta z^{i+1} + \delta z^i)} \left( \frac{1}{m_z^{i+1}} - \frac{1}{m_z^{i-1}} \right) \text{ et } \frac{\partial^2}{\partial z^2} (\psi_n(z)) \approx \frac{\frac{\Psi_{i+1} - \Psi_i}{\delta z^{i+1}} - \frac{\Psi_i - \Psi_{i-1}}{\delta z^i}}{\frac{\delta z^{i+1} + \delta z^i}{2}},$$

on obtient le schéma numérique suivant :

$$H \cdot \Psi = A_i \left( \frac{2}{\delta z^{i+1} \cdot m_z^i} + B_i \right) \cdot \Psi_n^{i+1} + D_i \cdot \Psi_n^i + A_i \left( \frac{2}{\delta z^i \cdot m_z^i} - B_i \right) \cdot \Psi_n^{i-1} \quad \text{avec} \quad A_i = -\frac{\hbar^2}{2} \frac{1}{(\delta z^{i+1} + \delta z^i)},$$

$$B_i = \frac{1}{(\delta z^{i+1} + \delta z^i)} \left( \frac{1}{m_z^{i+1}} - \frac{1}{m_z^{i-1}} \right) \text{ et } D_i = \frac{\hbar^2}{m_z^i (\delta z^{i+1} + \delta z^i)} \left( \frac{1}{\delta z^{i+1}} - \frac{1}{\delta z^i} \right) + V(z_i)$$

En considérant des conditions aux limites nulles :  $\Psi_n^{N+1} = \Psi_n^0 = 0$ , on peut écrire ce schéma sous la forme matricielle suivante :

$$H \cdot \Psi = \begin{pmatrix} D_1 & C_1 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & A_i \left( \frac{2}{\delta z^{i+1} \cdot m_z^i} + B_i \right) & D_i & A_i \left( \frac{2}{\delta z^i \cdot m_z^i} - B_i \right) & 0 \\ 0 & 0 & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & A_N & D_N \end{pmatrix} \begin{pmatrix} \vdots \\ \Psi_n^{i-1} \\ \Psi_n^i \\ \Psi_n^{i+1} \\ \vdots \end{pmatrix} = E_n \begin{pmatrix} \vdots \\ \Psi_n^{i-1} \\ \Psi_n^i \\ \Psi_n^{i+1} \\ \vdots \end{pmatrix}$$

Les niveaux d'énergies  $E_n$  (valeurs propres) et les fonctions d'ondes associées  $\Psi_n$  (vecteurs propres) de la matrice réelle  $H$  (tridiagonale mais malheureusement potentiellement asymétrique à cause des masses et des pas variables) sont obtenus par méthode HQR [PRES03].



# *Références bibliographiques*

- [ALLI01] **Frédéric ALLIBERT** et al. « From SOI materials to innovative devices », *Solid-State Electronics*, vol 45, p. 559, February 2001.
- [ANCO90] **M. G. ANCONA** « Macroscopic description of quantum-mechanical tunneling », *Phys. Rev. B*, vol. 42, p. 1222–1233, 1990.
- [ANDO82] **T. ANDO et al.** « Electronic properties of two-dimensional systems », *Rev. Mod. Phys.*, Vol. 54, n°2, p. 437-621, 1982.
- [ARBE98] **Marie-Emma ARBEY-RAZATOVO** « Filière CMOS 0.1µm sur substrat SOI : étude du temps de propagation de l'inverseur par simulation particulière Monte Carlo », Thèse de doctorat, Université Paris-Sud, 1998.
- [ASEN98] **Asen ASENOV**, « Random dopant-induced threshold voltage lowering and fluctuations in sub-0.1-µm MOSFET » *IEEE Trans. Electron Devices*, vol. 45, p. 2505–2513, 1998.
- [ASEN01] **Asen ASENOV**, « Discrete impurities and quantum potentials in MOSFET modeling » *Workshop on Quantum Transport in Semiconductors*, Maratea, Italie, 17-22 June 2001.
- [BACC99] **Giorgio BACCARANI et Susanna REGGIANI**, « A compact Double-Gate MOSFET Model Comprising Quantum-Mechanical and Nonstatic Effects » *IEEE Trans. Electron Devices*, vol. 46, p. 1656–1666, 1999.
- [BACC00] **Giorgio BACCARANI et Susanna REGGIANI**, « Performance limits of CMOS technology and perspectives of quantum devices », *C.R. Acad. Sci. Paris*, t. 1, Série IV, p. 843-873, 2000.
- [BALE87] **Francis BALESTRA et al.** « Double-Gate Silicon-on-Insulator Transistor with Volume Inversion: A new Device with Greatly Enhanced Performance », *IEEE Electron Device Letters*, vol. EDL-8, No. 9, 1987
- [BAND05] **P. R. BANDARU, C. DARAIIO, S. JIN et A. M. RAO** « Novel electrical switching behaviour and logic in carbon nanotube Y-junctions », *Nature Materials* Vol. 4, p. 663-666, Sept 2005.
- [BARR01] **Sylvain BARRAUD** « Effet d'une répartition discrète et aléatoire des impuretés dans le canal des MOSFET sub-100 nm : Etude théorique par simulation Monte Carlo », Thèse de doctorat, Université Paris-Sud, 2001.
- [BARR02] **Sylvain BARRAUD** et al. « Short-range and long-range Coulomb interactions for 3D Monte Carlo device simulation with discrete impurity distribution », *Solid-State Electronics*, Vol. 46, p. 1061, 2002.
- [BORO05] **Piotr BOROWIK, Leszek ADAMOWICZ** « Improved algorithm for Monte Carlo studies of electron transport in degenerate semiconductors », *Physica B*, Vol. 365, p. 235–239, 2005.

- [BOUR06] **Arnaud BOURNEL** « Composants pour la spintronique, la microélectronique ultime et la nanoélectronique », Habilitation à Diriger les Recherches, Université Paris-Sud, 2006 (?!).
- [BRIS94] **Christophe BRISSET** « Etude théoriques par simulation Monte-Carlo 3D de la sensibilité aux irradiations des technologies CMOS/SOI », Thèse de doctorat, Université Paris-Sud, 1994.
- [BUFL03] **Fabian M BUFLER** « Full-Band Monte Carlo Simulation of Nanoscale Strained-Silicon MOSFETs », Series in Microelectronics, Vol. 140, 2003.
- [BUFL04] **Fabian M BUFLER, A SCHENK and W FICHTNER** « Strained-Si single-gate versus unstrained-Si double-gate MOSFETs », Semicond. Sci. Technol. 19, p. S122–S124, 2004.
- [CASS99] **Eric CASSAN et al.** « Study of direct tunneling through ultrathin gate oxide of field effect transistors using Monte Carlo simulation », Journal of Applied Physics, Vol. 86, No. 7, p. 3804, 1999.
- [CASS00] **Eric CASSAN** « Etude théorique du courant tunnel de grille dans les transistors MOSFET sub-0,1  $\mu\text{m}$  », Thèse de doctorat, Université Paris-Sud, 2000.
- [CERU04] **Robin CERUTTI et al.** « Gate All Around (GAA) adapté au design de circuits », Actes des JNRDM, p. 120, 2004.
- [COHE96] **Claude COHEN-TANNOUJJI et al.**, « Mécanique quantique », Hermann, Tome 2, 1996.
- [COLI04] **Jean Pierre COLINGE** « Multiple-gate SOI MOSFETs », Solid State Electron., vol. 48/6, p. 897–905, 2004.
- [COUL98] **Frédéric de COULON** « Théorie et traitement des signaux », Presses polytechniques et universitaires romanes, Traité d'électricité, vol. 6, 1998.
- [CHAN00] **Leland CHANG et al.** « Gate Length Scaling and Threshold Voltage Control of double-Gate MOSFETS », IEDM, 2000.
- [CHAU02] **Robert CHAU et al.** « Advanced Depleted-Substrate Transistors: single-gate, Double-gate and Tri-gate », ICSSDM, 2002.
- [CHEN02] **M-J CHEN et al.** « Temperature Dependent Channel Backscattering Coefficients in Nanoscale MOSFETs », IEDM Tech. Dig., p. 39–40, 2002.
- [CHEN04] **M-J CHEN et al.** « Separation of Channel Backscattering Coefficients in Nanoscale MOSFETs », IEEE Trans. Electron Devices vol. 51, No. 9, p. 1409-1415, 2004.
- [DATT95] **Supriyo DATTA** « Electronic Transport in Mesoscopic Systems », Cambridge University Press, 1995.
- [DAUG04] **F. DAUGE et al.** « Coupling effects and channels separation in FinFETs », Solid-State Electronics, vol. 48, p. 535-542, 2004.

- [DOLL97] **Philippe DOLLFUS** « Si/Si<sub>1-x</sub>Ge<sub>x</sub> heterostructures : Electron transport and field-effect transistor operation using Monte Carlo simulation », Journal of Applied Physics, Vol. 82, No. 8, p. 3911, 1997.
- [DOLL99] **Philippe DOLLFUS** « Etudes théoriques de structures pour l'électronique rapide et contribution au développement d'un simulateur particulaire Monte Carlo », Habilitation à Diriger les Recherches, Université Paris-Sud, 1999.
- [DOLL01] **Philippe DOLLFUS** et **Sylvie GALDIN-RETAILLEAU** « Physique des dispositifs à Semiconducteurs », Cours de DEA, mars 2001.
- [DOLL04] **Philippe DOLLFUS et al.** « Effect of discrete impurities on electron transport in ultra-short MOSFET using 3D Monte Carlo simulation », IEEE Trans. Electron Devices vol. 51, No. 5, p. 749-756, 2004.
- [ESSE03] **David ESSENI et al.** « Physically Based Modeling of Low Field Electron Mobility in Ultrathin Single- and Double-Gate SOI n-MOSFETs », IEEE Trans. Electron Devices vol. 50, No. 12, p. 2445, 2003.
- [EZAK03] **Tatsuya EZAKI et al.** « Self-Consistent Quantum Mechanical Monte Carlo MOSFET Device Simulation », Journal of Computational Electronics, vol. 2, p. 97-103, 2003.
- [FENO04] **C. FENOUILLET-BERANGER et al.** « Requirements for ultra-thin-film devices and new materials for the CMOS roadmap », Solid-State Electronics, vol. 48, p. 961-967, 2004.
- [FERR00] **D. K. FERRY et al.** « Quantum effects in MOSFETs: Use of an effective potential in 3D Monte Carlo simulation of ultra-short channel devices », IEDM Tech. Digest, p.287-290, 2000.
- [FETT74] **A. L. FETTER** « Electrodynamics and thermodynamics of a classical electron surface layer », Phys. Rev. B, vol. 10, n°9, p. 3739-3745, 1974.
- [FISC88] **M.V. FISCHETTI et S.E. LAUX** « Monte carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects », Phys. Rev. B, vol. 38, No. 14 p. 9721-9745, 1988.
- [FISC91] **M.V. FISCHETTI** « Effect of the electron-plasmon interaction on the electron mobility in silicon », Phys. Rev. B, vol. 44, p. 5527-5534, 1991.
- [FISC93] **M.V. FISCHETTI et S.E. LAUX** « Monte Carlo study of electron transport in silicon inversion layers », Phys. Rev. B, vol. 48, p. 2244-2274, 1993.
- [FISC01] **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, n°2, p. 1205, 2001.
- [FISC03] **M.V. FISCHETTI.** « Scaling MOSFETs to the Limit: A Physicists's Perspective », Journal of Computational Electronics, Vol. 2, N° 2-4, p. 73-79, 2003.

- [FRIE03] **David M. FRIED et al.** « Improved Independent Gate N-Type FinFET Fabrication and Characterization », IEEE Electron Devices Let., vol. 24, No. 9, p. 592-594, 2003.
- [FUCH05] **Emmanuel FUCHS al.** « A New Backscattering model for Nano-MOSFET Compact Modeling », IEEE Trans. Electron Devices vol. 52, No.12, p. 2280-2289, 2005.
- [GALD92] **Sylvie GALDIN**, « Etude du transistor bipolaire à double hétérojonction Si/SiGe /Si par simulation Monte Carlo », Thèse de doctorat, Université Paris-Sud, 1992.
- [GAMI01] **F. GAMIZ et al.** « Role of surface-roughness scattering in double gate silicon-on-insulator inversion layers », Journal of Applied Physics, Vol. 89, n°3, p. 1764, 2001.
- [GAMI03] **F. GAMIZ et al.** « Monte Carlo simulation of remote-Coulomb-scattering-limited mobility in metal–oxide–semiconductor transistors », Journal of Applied Physics, Vol. 42, No. 19, p. 3251, 2003.
- [GAMI04] **F. GAMIZ et al.** « Double gate silicon on insulator transistors. A Monte Carlo study », Solid-State Electronics, Vol. 48, p. 937-945, 2004.
- [GAUT03] **Jacques GAUTIER et al.** « Physique des dispositifs pour circuits intégrés silicium », Hermès, 2003.
- [GHIB88] **G. GHIBAUDO et al.**, « New method for the extraction of MOSFET parameters », Electronics Letters, Vol. 24, n°9 , p. 543-545, 1988.
- [GOME05] **F. M. GOMEZ-CAMPOS et al.** « A solution of the effective-mass Schrödinger equation in general isotropic and nonparabolic bands for the study of two-dimensional carrier gases », Journal of Applied Physics, Vol. 98, p. 033717, 2005.
- [GOOD85] **S. M GOODNICK. et al.** « Surface roughness at the Si(100)-SiO<sub>2</sub> interface », Phys. Rev. B, vol. 32, n°12,p. 8171, 1985.
- [GRAD80] **I. S. GRADSHTEYN, I. M. RYZHIK.** « Table of Integrals, Series, and Products », Academic Press, 1980.
- [HALL04] **S. HALL et al.**,« Recent developments in deca-nanometer vertical MOSFETs », Microelectronics Engineering, vol. 72, p. 230-235, 2004.
- [HARE96] **S.A. HARELAND et al.**, « A computationally efficient model for inversion layer quantization effects in deep submicron N-channel MOSFET's », IEEE Trans. Electron Devices, vol. 43, p. 90-96, 1996.
- [HARR03] **S. HARRISON et al.** « Highly performant double gate MOSFET realized with SON process », IEDM Tech. Dig., p. 449–452, 2003.
- [HEST84] **Patrice HESTO** « Simulation Monte Carlo du transport non stationnaire dans les dispositifs submicroniques : importance du phénomène balistique dans GaAs à 77 K », Thèse d'état de l'Université Paris-Sud, Orsay, 1984.

- [HERG02] **J.M. HERGENROTHER et al.** « The vertical replacement-gate (VRG) MOSFET », Solid-State Electronics, vol. 46, p. 939-950, 2002.
- [HORH03] **Tim HORH et al.**, « On Density-Gradient Modeling of Tunneling through Insulators », IEICE Trans. Electron Devices, vol. E86-C, No. 3, p. 379, 2003.
- [HRUS02] **M. HRUSKA and B. SPIVAK** « Conductivity of the classical two-dimensional electron gas », PHYS. REV. B, vol. 65, n°3, p. 033315-1, 2002.
- [JAUD05] **Marie-Anne JAUD et al.** « Quantum-corrected Monte-Carlo simulation of double-Gate nMOSFET », 15th Workshop on the Modeling and Simulation of Electron Devices (MSED 2005), Pise, Italie, 4-5 juillet, soumis à Journal of Computational Electronics, 2005.
- [KATH03] **Gulzar A. KATHAWALA, Mohamed MOHAMED et Umberto RAVAIOLI** « Comparison of Double-Gate MOSFETs and FinFETs with Monte Carlo Simulation », Journal of Computational Electronics, Vol. 2, N° 2-4, p. 85-89, 2003.
- [KEDZ01] **Jakub KEDZIERSKI et al.** « High-performance symmetric-gate and CMOS-compatible VT asymmetric-gate FinFET devices », Proc. IEDM, p. 437, 2001.
- [KUMA04] **E. KUMAR et al.** « Low voltage and performance tunable CMOS circuit design using independently driven double gate MOSFETs », SOI Conf., p. 119-121, 2004.
- [LAKE92] **R. K. LAKE and S. DATTA**, « The Non-equilibrium Green's Function Method Applied to Double Barrier Resonant Tunneling Diodes », Phys. Rev. B, vol. 45, p. 6670-6685, 1992.
- [LAHM04] **Marcel LAHMANI et al.** « Les nanosciences. Nanotechnologies et nanophysique », Belin, 2004.
- [LEMM04] **M.C. LEMME** « Subthreshold behavior of triple-gate MOSFETs », Solid-State Electronics, vol. 48, p. 529-534, 2004.
- [LEPA96] **S. LEPAUL, F. BOUILLAULT et A. DE LUSTRAC** « Simulation of electron states in quantum wires with mixed finite elements », COMPEL: Int J for Computation and Maths. in Electrical and Electronic Eng., Vol. 15, No. 3, p. 58-69(12), 1996.
- [LERO03] **M. LE ROY, E. LHEURETTE, O. VANBÉSIEN et D. LIPPENS** « Wave-mechanical calculations of leakage current through stacked dielectrics for nanotransistor metal-oxide-semiconductor design », Journal of Applied Physics, Vol. 93, No. 5, p. 2966, 2003.
- [LOLI05] **Jérôme LOLIVIER** « Etude, réalisation et caractérisation de transistors silicium sur isolant complètement désertés de longueur de grille inférieure à 25 nm », Thèse de doctorat, Institut National Polytechnique de Grenoble, 2005.
- [LUCO05] **G. LUCOVSKY, J.C. PHILLIPS** « Defects and defect relaxation at internal interfaces between high- $k$  transition metal and rare earth dielectrics and interfacial native oxides in metal oxide semiconductor (MOS) structures », Thin Solid Films, vol. 486, p. 200-204, 2005.

- [LUJA05] **G.S. LUJAN et al.** « Modelling mobility degradation due to remote Coulomb scattering from dielectric charges and its impact on MOS performance », *Microelectronics Reliability*, vol. 45, p. 794-797, 2005.
- [LUND97] **Mark LUNDSTROM** « Elementary Scattering Theory of the Si MOSFET », *IEEE Electron Device Letters*, vol. 18, p. 361-363, 1997.
- [LUND00] **Mark LUNDSTROM** « Fundamentals of carrier transport », Second edition, Cambridge University Press, 2000.
- [LUND02] **Mark LUNDSTROM et Zhibin REN** « Essential Physics of Carrier Transport in Nanoscale MOSFETs », *IEEE Trans. Electron Devices* vol. 49, No. 1, p. 133-141, 2002.
- [LUND02b] **Mark LUNDSTROM et J.H. RHEW** « A Landauer Approach to Nanoscale MOSFETs », *J. of Comp. Elec.*, Vol. 1, No. 4, p. 481-489, 2002.
- [LUSA05] **J. ŁUSAKOWSKI et al.** « Ballistic and pocket limitations of mobility in nanometer Si metal-oxide semiconductor field-effect transistors », *Applied Physics Letters* 87, 053507, 2005.
- [MASE83] **G. MASETTI et al.** « Modeling of carrier mobility against carrier concentration in Arsenic-Phosphorus, and Boron-Doped Silicon », *IEEE Trans. Electron Devices* vol. 30, No. 7, p. 764-769, 1983.
- [MATH04] **Henry MATHIEU** « Physique des semiconducteurs et des composants électroniques », Dunod, 2004
- [MERZ61] **E MERZBACHER** « Quantum Mechanics », John Wiley and Sons, 1961
- [MITR04] **S. MITRA et al.** « Double gate (DG)-SOI ratioed logic with symmetric DG load- a novel approach for sub 50 nm low-voltage/low-power circuit design », *Solid-State Electronics*, vol. 48, p. 1727-1732, 2004.
- [MONS02] **Florian MONSEF** « Etude par simulation Monte Carlo des effets de quantification sur le transport dans les structures à effet de champ », Thèse de doctorat, Université Paris-Sud, 2002.
- [MONS04] **Florian MONSEF et al.** « Electron transport in Si/SiGe modulation-doped heterostructures using Monte Carlo simulation », *Journal of Applied Physics*, Vol. 95, No. 7, p. 3587, 2004.
- [MOUI01] **M. MOUIS, A. PONCET**, ESSDERC (European Solid-State Device Research Conference), Nuremberg (Allemagne), 11-13 septembre 2001, p. 211-214
- [MOUI03] **Mireille MOUIS et Sylvain BARRAUD** « Velocity distribution of electrons along channel of channel of nanoscale MOS transistors », *Proc. ESSDERC*, p. 147-150, 2003
- [NATO94] **K. NATORI** « Ballistic metal-oxide-semiconductor field effect transistor », *Journal of Applied Physics*, Vol. 76, No. 8, p. 4879, 1994.
- [NATO03] **K. NATORI** « Multi-Subband Effects on Performance Limit of Nanoscale MOSFETs », *Journal of Applied Physics*, Vol. 42, No. 4B, p. 2063, 2003.

- [NEDJ04] **M. NEDJALKOV et al** « Unified particle approach to Wigner-Boltzmann transport in small semiconductor devices », Phys. Rev. B, vol. 70, 115319, 2004.
- [NEUM66]. **J. VON NEUMAN** « Theory of Self-Reproducing Automata », Univ. Illinois Press, Urbana, p. 66, 1966.
- [NOUG93] **J.P. NOUGIER**, « Méthodes de calcul numérique », Masson, 1993
- [OSBU02] **C. M. OSBURN et al.** « Vertically scaled MOSFET gate stacks and junctions: How far are we likely to go ? », IBM J. Res. & Dev., Vol. 46, No 2/3, p. 299, March/May 2002.
- [PALE05] **P. PALESTRI et al.** « An improved semi-classical Monte-Carlo approach for nano-scale MOSFET simulation », Solid-State Electronics, vol. 49, p. 727-732, 2005.
- [PARK01] **Jong-Tae PARK et al.** « Pi-Gate SOI MOSFET » IEEE Electron Device Letters, vol. 22, No. 8, 2001.
- [PRES03] **W. H. PRESS et al.** « Numerical Recipes in Fortran 77: The art of Scientific Computing », Second edition, Cambridge University Press, 2003.
- [RAHM03] **Anisur RAHMAN et al.**, « Theory of ballistic nanotransistors », IEEE Trans. Electron Devices, vol. 50, p. 1853-1864, 2003.
- [REN01] **Zhibin REN** « Nanoscale MOSFETs: Physics, simulation and design », Thèse de doctorat, Purdue University, 2001.
- [RHEW02] **Jung-Hoon RHEW et al.** « A numerical study of ballistic transport in a nanoscale MOSFET », Solid-State Electronics, vol. 46 , p. 1899-1906, 2002.
- [RHEW03] **Jung-Hoon RHEW** « Physics and simulation of quasi-ballistic transport in nanoscale transistors », Thèse de doctorat, Purdue University, 2003.
- [ROBE04] **J. ROBERSON, P.W. PEACOCK** « Atomic structure, band offsets, growth and defects at high-*K* oxide: Si interfaces », Microelectronics Engineering, vol. 72, p. 112-120, 2004.
- [RUMS00] **David RUMSEY, Zhibin REN et Mark LUNDSTROM** « Electrical Characterization of MOSFET's in Terms of Backscattering Parameters », Presented at SRC Techcon, 2000
- [SACC04] **Fabio SACCONI et al.** « Full-band approaches to the electronic properties of nanometer-scale MOS structures », Solid-State Electronics, vol. 48, p. 575-580, 2004.
- [SAKA87] **H. SAKAKI et al.** « Interface roughness scattering in GaAs/AlAs quantum wells », Applied Physics Letters 51, Vol. 23, p. 1934-1936, 1987.
- [SAMP05] **C. SAMPREDO et al.** « Monte Carlo simulation of double gate silicon on insulator devices operated as velocity modulation transistors », Applied Physics Letters 86, 202115, 2005.

- [SANG92] **E. SANGIORGI and M. R. PINTO**, "A semi-empirical model of surface scattering for Monte Carlo simulation of silicon n-MOSFET", IEEE Trans. Electron Devices, vol. 39, p. 356-361, 1992.
- [SEE03] **Johann SEE** « Théorie du blocage de Coulomb appliquée aux nanostructures semi conductrices : modélisation des dispositifs à nanocristaux de silicium », Thèse de doctorat, Université Paris-Sud, 2003.
- [SHAO05] **Xue SHAO et Zhiping YU** « Nanoscale FinFET simulation: A quasi-3D quantum mechanical model using NEGF », Solid-State Electronics, vol. 49, p. 1435-1445, 2005.
- [SKOT99] **Thomas SKOTNIKI et al.** « Heavily doped and extremely shallow junctions on insulator by SONCTION (SilicON Cut-off junction) process », IEDM Tech. Dig., p. 513-516, 1999.
- [SKOT05] **Thomas SKOTNIKI** « The end of CMOS scaling », IEEE circuits and design magazine, Jan/Feb 2005.
- [SONG97] **S.H. SONG et al.** « Energy relaxation of two-dimensional carriers in strained Ge/Si<sub>0.4</sub>Ge<sub>0.6</sub> and Si/Si<sub>0.7</sub>Ge<sub>0.3</sub> quantum wells: Evidence for two-dimensional acoustic phonons », Applied Physics Letters 70, p. 3422-3424, 1997.
- [SORI04] **Antonio SORIANO et al.** « Analysis of the finite difference time domain technique to solve the Schrödinger equation for quantum devices », Journal of Applied Physics, Vol. 95, N°12, p. 8011-8018, 2004.
- [STMA04] **J. SAINT MARTIN et al.** « Influence of Ballistic Effects in Ultra-Small MOSFETs », Journal of Computational Electronics, Vol. 3, N° 3-4, p. 207-210, 2004.
- [STRO98] **M. A. STROSCIO, V MITRIN et V. A. KOHELAP** « Quantum Heterostructures », Cambridge University Press, 1998.
- [SUZI96] **K. SUZUKI, Y. TOSAKA, T. SUGII** « Analytical threshold voltage model for short channel double-gate SOI MOSFET's », IEEE Trans. Electron. Dev., vol. 43, N°7, p. 1166-1168, 1996.
- [SVIZ03] **A. SVIZHENKO and M. P. ANANTRAM** « Role of scattering in nanotransistors », IEEE Trans. Electron. Dev., vol. 50, p. 1459-1466, 2003.
- [TAUR92]. **Yuan TAUR et al.** « 'A new "shift and ratio" method for MOSFETs channel length extraction », IEEE Electron Devices Let., vol. 13, No. 5, p. 267-269, 1992.
- [TAUR98] **Yuan TAUR et Tak H. NING** « Fundamentals of modern VLSI devices », Cambridge University Press, 1998.
- [THOM98] **THOMPSON Scott et al.** « MOS Scaling: Transistor Challenges for the 21<sup>st</sup> Century », Intel Technology Journal Q 398, 1998
- [TSUC01] **B. TSUCHIYA et U. RAVAIOLI** « Particle Monte Carlo simulation of quantum phenomena in semiconductor nanostructures », Journal of Applied Physics, Vol. 89, p. 4023-4029, 2001.

- [VARA97] **Luca VARANI** « Noise Phenomena in Semiconductor Transport », phys. stat. sol. (b) Vol. 204, p. 442, 1997.
- [VALI05] **Valérie AUBRY-FORTUNA, P. DOLLFUS, S. GALDIN-RETAILLEAU** « Electron effective mobility in strained-Si/Si<sub>1-x</sub>Ge<sub>x</sub> MOS devices using Monte Carlo simulation », Solid-State Electronics, vol. 49, p. 1320-1329, 2005.
- [VAPA87] **André VAPAILLE et René CASTAGNE** « Dispositif et circuits intégrés semiconducteurs », Dunod, 1987
- [VENU02] **R. VENUGOPAL, Z. REN, S. DATTA, M. S. LUNDSTROM, and D. JOVANOVIC** « Simulating Quantum Transport in Nanoscale MOSFETs: Real vs. Mode Space Approaches » J. Appl. Physics, Vol. 92, p. 3730-3739, 2002.
- [VENU04] **R. VENUGOPAL et al.** « Simulating Quantum Transport in Nanoscale MOSFETs: Real vs. Mode Space Approaches » J. Appl. Physics, Vol. 95, No. 1, p. 292-305, 2002.
- [VINE05] **M. VINET et al.** « Bonded Planar Double-Metal-Gate NMOS transistors down to 10 nm », IEEE Electron Devices Let., vol. 26, No. 5, p. 317-319, 2005.
- [VIVI04] **Laurent VIVIEN et al.** « Experimental demonstration of a low loss optical H-tree distribution using silicon-on-insulator microwaveguides », Applied Physics Letters, vol. 85, p. 701-703, 2004.
- [WANG05] **Jing WANG et Eric POLIZZI** « Theoretical investigation of surface roughness scattering in silicon nanowire transistors », Applied Physics Letters, vol. 87, p. 043101, 2005.
- [WEST93] **Neil H.E. WESTE et Kamran ESHRAGHIAN** « Principles of CMOS VLSI design », Second Edition, Addison-Wesley, 1993.
- [WIDI04] **Julie WIDIEZ et al.** « Experimental gate misalignment analysis on double gate SOI MOSFETs », SOI Conf. ,p. 185-186, 2004.
- [WILL00] **S. C. WILLIAMS et al.** « Ensemble Monte Carlo Study of Channel Quantization in a 25-nm n-MOSFET », IEEE Trans. Electron Devices vol. 47, No. 10, p. 1864, 2000.
- [WU05] **X. WU et al.** « Impacts of Nonrectangular Fin Cross Section on the Electrical Characteristics of FinFET », IEEE Trans. Electron Devices vol. 52, No. 1, p. 63-68, 2005.
- [XUE05] **Y. XUE et M. A. RATNER** « Molecular Electronics: From Physics to Computing », cond-mat/0508477, à paraître dans Natural Computing.
- [YAN92] **Ran-Hong YAN et al.** « Scaling the Si MOSFET: From Bulk to SOI to BULK », IEEE Transactions on Electron Devices, Vol. 39, No. 7, p. 1704, 1992.
- [YANG03] **M. YANG et al.** « High performance CMOS fabricated on hybrid substrate with different crystal orientations », IEDM Tech. Dig., p. 453, 2003.
- [ZUTI04] **I. ZUTIC, J. FABIAN, S. DAS SARMA** « Spintronics: Fundamentals and applications », Review of Modern Physics, Vol. 76, No. 2, p. 323-410, 2004.



# *Publications liées à ce travail*

## **1) Publications dans des revues avec comité de lecture**

Saint-Martin J., Bournel A., Dollfus P.  
*Comparison of multiple-gate MOSFET architectures using Monte Carlo simulation*  
Accepté sous réserve de modifications mineures à Solid State Electronics

Saint Martin J., Bournel A., Dollfus P.  
*On the ballistic transport in nanometer-scaled DG MOSFETs*  
IEEE Transaction on Electron Devices 51 (2004) 1148-1155

## **2) Conférences nationales et internationales**

### **Avec actes**

Saint-Martin J., Bournel A., Dollfus P.  
*Comparison of multiple-gate MOSFET architectures using Monte Carlo simulation*  
(Oral) 6th European Conference on ULtimate Integration of Silicon (ULIS 2005), Bologne, Italie, 7-8 avril 2005, Proceedings ISBN 88-900847-0-7, p. 61-64 (2005)

Saint Martin J., Aubry-Fortuna V., Bournel A., Dollfus P., Galdin S., Chassat C.  
*Influence of ballistic effects in ultra-small MOSFETs*  
(Oral) International Workshop on Computational Electronics (IWCE-10), West Lafayette, USA, 24-27 octobre 2004, Journal of Computational Electronics 3 (2004) 207-210

### **Sans actes**

Saint Martin J.  
*Prise en compte de la quantification perpendiculaire au transport dans les transistors MOSFET à double grille par simulations Monte Carlo*  
(Poster) 8<sup>èmes</sup> Journées Nationales du Réseau Doctoral en Microélectronique (JNRDM 2005), Paris, 10-12 mai 2005

Saint Martin J., Bournel A., Dollfus P.  
*Etude du transport quasi-balistique dans les transistors MOSFET à double-grille ultracourts*  
(Poster) Xe Journées Nationales Microélectronique Optoélectronique, La Grande Motte, 8-11 juin 2004

Dollfus P., Saint Martin J., Bournel A.  
*Les composants balistiques*  
(Oral) Journées Nationales Nanoélectronique, conférence invitée, Aussois, 10-12 mai 2004

Bournel A., Saint Martin J., Dollfus P., Grémion E.  
*Etude d'architectures de MOSFET ultra-courts à grille multiple*  
(Oral) Journées Nationales Nanoélectronique, Aussois, 10-12 mai 2004

Saint Martin J., Bournel A., Dollfus P.  
*Transport balistique dans les transistors MOSFET à double-grille ultracourts*  
(Poster) Journées Nationales Nanoélectronique, Aussois, 10-12 mai 2004

Saint Martin J.  
*Etude par simulation Monte Carlo de MOSFETs ultra-courts à grille multiple sur SOI*  
(Oral) Journées Nationales du Réseau Doctoral en Microélectronique, Marseille, 4-6 mai 2004

### **3°) Séminaires**

Saint Martin J.  
*Etude par simulation Monte Carlo d'architectures de MOSFET ultra courts à grille multiple sur SOI*  
Séminaire doctorant IEF, 1er décembre 2004