Contribution à l'étude expérimentale du transport dans les transistors de dimensions déca-nanométriques des technologies CMOS sub-45nm

Dominique Fleury

▶ **To cite this version:**

Dominique Fleury. Contribution à l'étude expérimentale du transport dans les transistors de dimensions déca-nanométriques des technologies CMOS sub-45nm. Micro et nanotechnologies/Microélectronique. Institut National Polytechnique de Grenoble - INPG, 2009. Français. NNT : . tel-00461948

HAL Id: tel-00461948

https://theses.hal.science/tel-00461948

Submitted on 8 Mar 2010

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.

INSTITUT POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

T H E S E

pour obtenir le grade de

DOCTEUR DE L'Institut polytechnique de Grenoble

Spécialité : *Micro et Nano Electronique*

préparée à l'Institut de Microélectronique, Electromagnétisme et Photonique et le Laboratoire d'Hyperfréquences et de Caractérisation (IMEP-LAHC) et à STMicroelectronics Crolles

dans le cadre de **l'Ecole Doctorale** *« Electronique, Electrotechnique, Automatique et Traitement du Signal »*

présentée et soutenue publiquement

par

Dominique FLEURY

le 02 Décembre 2009

CONTRIBUTION A L'ETUDE EXPERIMENTALE DU TRANSPORT DANS LES TRANSISTORS DE DIMENSIONS DECA-NANOMETRIQUES DES TECHNOLOGIES CMOS SUB 45NM

DIRECTEUR DE THESE : M. Gérard GHIBAUDO, Directeur de Recherches CNRS
CO-ENCADRANTS DE THESE : M. Antoine CROS et M. Krunoslav ROMANJEK Dr. STMicroelectronics

JURY

M. Abdelkader SOUIFI, Prof. INSA Lyon	, Président
M. Cor CLAEYS, Prof. IMEC Leuven	, Rapporteur
M. Jean-Luc AUTRAN, Prof. Université de Provence	, Rapporteur
M. Gérard GHIBAUDO, DR CNRS Grenoble	, Directeur de thèse
M. Antoine CROS, Dr. STMicroelectronics Crolles	, Co-Encadrant
M. Arnaud BOURNEL, Prof. IEF Orsay	, Examinateur
M. Thomas SKOTNICKI, DR STMicroelectronics Crolles	, Invité
M. Mikaël CASSE, Dr. CEA-LETI Grenoble	, Invité

Table des matières

Table des matières 3

Table des figures 7

Liste des tableaux 13

Remerciements 15

Introduction générale 17

I Le transistor MOS ... en bref 21
- I.1 Architecture et principe de fonctionnement du transistors MOS . . . 23
  - I.1.1 La barrière de potentiel source-drain . . . 24
  - I.1.2 Définition analytique de la tension de seuil . . . 25
  - I.1.3 Les régimes de conduction du transistor MOS . . . 25
- I.2 Les limites à la miniaturisation de la technologie CMOS . . . 27
  - I.2.1 La perte de l'intégrité électrostatique . . . 27
  - I.2.2 Influence de la miniaturisation sur le transport : des performances amoindries 29
  - I.2.3 La puissance consommée et le délai de commutation . . . 29
- I.3 Les solutions pour une poursuite de la miniaturisation . . . 31
  - I.3.1 La maîtrise des courants de fuite grâce à l'empilement métal/*high-*κ . . . 31
  - I.3.2 Des contraintes mécaniques pour *doper* le transport . . . 32
  - I.3.3 Les architectures à film de conduction mince . . . 33

II Étude théorique du transport dans le canal du transistor MOS 37
- II.1 Du transport dérive-diffusion à la balisticité : le rôle des mécanismes de collision 39
  - II.1.1 Le concept de mobilité . . . 40
  - II.1.2 Hypothèses et validité du modèle de Drude . . . 40
  - II.1.3 Le libre parcours moyen : la clé pour comprendre les mécanismes de transport 40
  - II.1.4 Notion de continuité du courant . . . 42
  - II.1.5 Description du mécanisme d'injection à la source . . . 43
- II.2 Le transport stationnaire . . . 45
  - II.2.1 Notion de mobilité effective et de champ effectif . . . 46
  - II.2.2 Interactions avec les centres de diffusion chargés . . . 49
  - II.2.3 Interactions avec les vibrations du réseau : les phonons . . . 51
  - II.2.4 L'influence de la rugosité de surface sur la mobilité . . . 56
  - II.2.5 Les mécanismes additionnels de réduction de la mobilité . . . 57
  - II.2.6 Influence de la température et du champ sur la mobilité . . . 60

II.3 Le transport non-stationnaire . . . 61
II.3.1 Les phénomènes de survitesse : une théorie vite oubliée . . . 61
II.3.2 Le modèle de transport quasi-balistique : généralités et approximations . 63
II.3.3 La notion de distance critique (kT-*layer*) et de rétrodiffusion . . . 67
II.4 Vers l'unification des mécanismes de transport . . . 69
II.4.1 Expression du courant de drain dans l'approche dérive-diffusion . . . 69
II.4.2 Le concept de mobilité apparente . . . 69
II.4.3 L'unification des théories de transport . . . 70
II.5 Bilan et discussions . . . 74

III Méthodologies d'extraction des paramètres électriques sur des transistors MOS 75
III.1 Les outils nécessaires à la caractérisation électrique en milieu industriel . . . 78
III.2 Le cœur de l'extraction : un modèle de courant adapté au comportement du transistor MOS moderne en régime linéaire et en inversion forte . . . 78
III.2.1 Influence et correction de la fuite de grille . . . 80
III.2.2 Prise en compte de la résistance série dans le modèle . . . 81
III.2.3 Discussion : les limites de l'extraction en inversion forte . . . 83
III.3 Les méthodes d'extraction de la tension de seuil ($V_{\rm th}$) . . . 84
III.3.1 L'extraction de $V_{\rm th}$ par mesures capacitives . . . 84
III.3.2 La méthode du courant constant . . . 86
III.3.3 La méthode du V_{th} extrapolé depuis le maximum de transconductance . . 86
III.3.4 Les méthodes utilisant deux dérivations successives du courant . . . 87
III.3.5 Proposition d'une nouvelle méthode : le V_{th} extrapolé en inversion faible . 89
III.4 L'extraction de tous les paramètres du modèle V_{th}, β, Θ_1 et Θ_2 . . . 91
III.4.1 Les techniques utilisant la fonction Y : des valeurs sûres . . . 91
III.4.2 L'état de l'art des améliorations de la fonction Y pour les transistors courts ($L_{eff} < 100$nm) . . . 94
III.4.3 Proposition d'une amélioration de la fonction Y : la méthode de la fonction ξ 97
III.4.4 Nouvelle méthodologie d'extraction utilisant les régimes de forte et de faible inversion . . . 101
III.5 Robustesse des méthodes d'extraction . . . 102
III.5.1 Immunité des résultats vis-à-vis des variations de R_{sd} . . . 102
III.5.2 Immunité des résultats aux variations de paramètres d'entrée . . . 106
III.6 L'extraction des résistances séries . . . 107
III.6.1 État de l'art : la méthode $\Theta(\beta)$ . . . 108
III.6.2 Proposition d'une nouvelle méthode d'extraction : $R_{tot}(1/\beta)$ . . . 110
III.6.3 "Le Graal" : vers l'extraction de R_{sd} sur un seul transistor ? . . . 113
III.7 Discussion et perspectives . . . 115

IV La modélisation et la caractérisation de la capacité MOS : innovations et nouvelles méthodologies 117
IV.1 La modélisation analytique de la réponse capacitive des transistors MOS . . . 119
IV.1.1 Introduction : la mesure de la réponse capacitive . . . 119
IV.1.2 Construction d'un modèle analytique simple (transistor long) . . . 121
IV.1.3 Prise en compte des capacités parasites de la structure MOS . . . 123

IV.2 Modélisation numérique 1D pour le calcul de la réponse capacitive 130
IV.2.1 Expression analytique de la charge d'espace 131
IV.2.2 Expression analytique de la densité surfacique de charges 132
IV.2.3 Décomposition de la charge d'espace en contributions élémentaires 133
IV.2.4 Décomposition de la charge totale en contributions élémentaires 133
IV.2.5 Calcul de la capacité dynamique d'une distribution de charges 135
IV.2.6 Calcul numérique et distribution de Fermi-Dirac 136
IV.2.7 Prise en compte des effets quantiques dans la modélisation 137
IV.3 Utilisation des mesures capacitives pour l'extraction de la charge et de la mobilité 142
IV.3.1 L'extraction expérimentale de la charge 142
IV.3.2 Cas des transistors courts : influence des capacités parasites 143
IV.3.3 Extraction de la mobilité par la mesure *split*-CV 144
IV.4 L'extraction de la longueur effective (L_{eff}) 146
IV.4.1 Etat de l'art des méthodes d'extraction 146
IV.4.2 Définition précise de la longueur effective de canal 146
IV.4.3 Proposition d'une nouvelle méthodologie d'extraction 147
IV.4.4 Automatisation de l'extraction 150
IV.4.5 Influence de la fuite de grille sur l'extraction 152
IV.4.6 La mesure de L_{eff} : applications pratiques 153
IV.5 Conclusion et perspectives 158

V Étude expérimentale du transport dans le transistor MOS 159
V.1 Évaluation des propriétés de transport en régime linéaire 161
V.1.1 L'extraction des paramètres essentiels : L_{eff}, R_{sd}, V_{th} 161
V.1.2 Extraction de la mobilité en régime linéaire 163
V.1.3 Evaluation expérimentale de la réduction de mobilité 164
V.1.4 Proposition d'un nouveau modèle de dégradation de mobilité avec L_{eff} . 165
V.1.5 L'extraction de coefficient de rétrodiffusion en régime linéaire 166
V.1.6 Influence de la balisticité dans la dégradation de mobilité 166
V.1.7 Bilan sur les propriétés de transport à bas champ 167
V.2 A la recherche de la balisticité 168
V.2.1 Un enthousiasme dans le domaine de la simulation 170
V.2.2 Les méthodes expérimentales pour évaluer la balisticité du transport . . . 170
V.2.3 Les points faibles des méthodes existantes 175
V.2.4 Proposition d'une nouvelle méthodologie d'extraction 176
V.2.5 Protocole expérimental suivi dans la thèse 178
V.3 Évaluation des propriétés de transport en régime de saturation 182
V.3.1 Extraction de la vitesse effective 183
V.3.2 Extraction de ν_{lim} et détermination de sa nature 185
V.3.3 Analyse de nos résultats en s'appuyant sur un modèle de survitesse 188
V.4 Bilan et discussions 191

Conclusion et perspectives 193

A Approximation de l'intégrale de Fermi-Dirac : cas général 197

B Rappel sur la structure crystalline du silicium 199

C Expression analytique de la kT-layer dans l'approximation dérive diffusion, inversion forte **201**

Bibliographie **203**

Publications de l'auteur **219**

Table des figures

1 Transistor amplificateur « point-contact »construit en 1947 par W. Shockley, J. Bardeen et W. Brattain (AT&T *Bell Laboratories*). Le prix Nobel de Physique leur sera décerné en 1956. 17
2 A gauche : Premier circuit intégré assemblé grâce au procédé *planar*. Commercialisé publiquement en 1961 par la société Fairchild Semiconductor, il est composé de 2 transistors inter-connectés sur une tranche de silicium (fonction logique *flip-flop*). A droite : processeur Prescott 1M commercialisé en février 2004 par la société Intel. Il et composé d'environ 125 millions de transistors et utilise le nœud 90nm qui marque un tournant dans la miniaturisation de la technologie CMOS. 18
3 Vue en coupe d'un transistor pMOS de la technologie 45nm d'Intel [Mistry07]. Son architecture inclue une grille métallique, un oxyde de type *high-*κ et des extensions S/D en SiGe qui transfèrent une contrainte compressive au canal (pour améliorer sa conductivité). 19

I.1 Principe de fonctionnement du transistor à effet champ dans les applications numériques. 23
I.2 Architecture d'un transistor nMOS sur silicium massif, d'après [Skotnicki03]. 23
I.3 Modification de la barrière de potentiel source-drain en fonction de la polarisation de grille V_{gs}. (a) régime d'accumulation, (b) régime de bandes plates (hypothèse $V_{fb} = 0$), (c) régime d'inversion forte. 25
I.4 Illustration des trois régimes de conduction qui caractérisent le fonctionnement d'un transistor nMOS 26
I.5 Illustration de l'abaissement de la barrière de potentiel source-drain induit par les effets de canaux courts (partage de charge et DIBL) 28
I.6 Évolution de la puissance dynamique (P_{on}) et de la puissance statique (P_{off}) en fonction du nœud technologique [Rochette08]. Si la puissance dynamique reste presque constante, la puissance statique augmente dramatiquement avec la miniaturisation. 30
I.7 Influence d'une contrainte uniaxiale sur le comportement des électrons et des trous, d'après [Skotnicki08]. 33
I.8 Architecture d'un transistor FDSOI. 34
I.9 Vue en coupe, prise au microscope électronique à transmission (TEM), d'un transistor d'architecture SON. 34

II.1 Illustration de la trajectoire des électrons subissant des collisions dans un cristal (a) sans champ électrique, (b) sous l'application d'un champ électrique, d'après [Grove67]. 39
II.2 Schéma simplifié du transistor MOS en régime de conduction : il est composé de deux réservoirs de porteurs reliés par un canal. L'injection est le mécanisme par lequel les porteurs quittent la source pour entrer dans le canal. Ce phénomène se produit au point de maximum de barrière de potentiel coté source appelée "source virtuelle". 41
II.3 Évolution des performances et des mécanismes de transport mis en jeu en fonction de la valeur du libre parcours moyen. L'état de l'art, pour des transistors fabriqués actuellement sur silicium, se situe proche de l'abscisse $\lambda_\mu \sim L$, correspondant à des mécanismes de transport à fort caractère non-local ou quasi-balistiques. 42
II.4 Évolution de la vitesse thermique en fonction de la dégénérescence. Pour un semiconducteur très dégénéré, ν_{th} tend vers la vitesse de Fermi qui ne dépend plus de la température, d'après [Lundstrom02a] 44
II.5 Illustration de la manière dont les collisions élastiques et inélastiques affectent les trajectoires et les énergies d'un groupe d'électrons injecté à $t = 0$ avec un moment incident, d'après [Lundstrom00]. 45
II.6 Variation de la mobilité effective en fonction du champ effectif pour différentes valeurs de dopage pour les électrons (à gauche) et les trous (à droite). La mobilité des deux types de porteurs suit une loi universelle (indépendante du dopage) à fort champ électrique [Takagi94a] lorsque η est ajusté correctement. 48
II.7 Principaux mécanismes de réduction de la mobilité dans les transistors MOS avec un empilement de grille de type Si/SiO_2. 49

II.8 Illustration du phénomène d'écrantage dune charge positive fixe par des porteurs chargés négativement. . . . 49
II.9 Évolution de la contribution de la mobilité limitée par les interactions avec les impuretés du canal en fonction du dopage en surface [Takagi94a]. . . . 50
II.10 Mobilité expérimentale pour des dispositifs possédant des empilements de grille avancés et standard, d'après [Zhu04]. . . . 51
II.11 Relation de dispersion des phonons dans le silicium le long des axes de symétrie, d'après [Yu05] . . 52
II.12 Illustration du déplacement des plans atomiques caractéristique du mode de vibration longitudinal, d'après [Kittel98]. . . . 54
II.13 Évolution de la vitesse de saturation en fonction de la température dans le silicium, d'après [Jacoboni77]. . . . 55
II.14 Evolution de la mobilité effective des porteurs en fonction du champ effectif et de la température, d'après [Takagi94a]. . . . 58
II.15 Illustration des différents mécanismes qui peuvent dégrader la mobilité dans un transistor ayant un empilement de grille *high-*κ-métal, d'après [Weber05]. . . . 58
II.16 Localisation présumée des défauts neutres résultant d'une guérison incomplète des défauts d'implantation, pour un transistor avec et sans poches de surdopage, d'après [Cros06b]. . . . 60
II.17 Vitesse moyenne et énergie cinétique des électrons injectés dans un morceau de silicium sous un fort champ électrique, d'après [Lundstrom06b] . . . 62
II.18 Corrélation entre la vitesse de dérive effective ν_{eff} (extraite par différentes méthodes) et le DIBL, d'après [Lochtefeld01]. . . . 63
II.19 Illustration du mécanisme d'injection à la source, d'après [Tsuchiya06]. Une fraction des porteurs qui franchissent le maximum de barrière de potentiel peuvent être rétrodiffusé avec une vitesse ν_{back} différente de ν_{inj}. . . . 64

III.1 A gauche : correction du courant de grille en utilisant la relation (III.7) sur un transistor à oxyde de grille mince ($T_{ox} \simeq 12$Å) de grande surface ($W \times L = 10\mu m^2$). A droite : illustration du mécanisme de partage de courant de grille, d'après [Romanjek04a]. . . . 81
III.2 Variation de la résistance série R_{sd} et de ses composantes en fonction de la tension de grille V_{gs}. La résistance de recouvrement (R_{ov}) est la contribution majeure qui provoque la variation de R_{sd} avec V_{gs}, d'après [Kim02b] . . . 82
III.3 Modélisation de la caractéristique $I_d(V_{gs})$ d'un transistor nMOS en régime linéaire, dans la zone d'inversion forte. Plusieurs combinaisons de paramètres (V_{th}, β, Θ_i) permettent de reproduire les variations du courant avec $R^2 \geq 99,6\%$. . . . 83
III.4 Extraction de la tension de seuil par mesure capacitive, **(a)** en utilisant la méthode de la dérivée première de la capacité grille-substrat dC_{gb}/dV_{gs} [Lau99, Lau01] et **(b)** en utilisant la méthode de la dérivée première de la capacité grille-canal dC_{gc}/dV_{gs}. . . . 85
III.5 Extraction de la tension de seuil depuis la courbe $Q_{inv}(V_{gs})$. La valeur extraite par ce biais est légèrement plus élevée que celle issue de l'extraction depuis la courbe $I_d(V_{gs})$ associée. . . . 86
III.6 Illustration de l'influence des résistances séries sur l'extraction de la tension de seuil en utilisant la méthode du V_{th} extrapolé depuis l'inversion forte. . . . 87
III.7 Extraction de la tension de seuil par la méthode de la dérivée seconde. L'amplification du bruit de mesure liée aux deux dérivations successives ainsi que le pas de mesure limitent la précision de cette technique. . . . 88
III.8 Extraction de la tension de seuil par la méthode de la dérivée seconde du logarithme du courant [Aoyama95]. Comme pour la technique décrite au paragraphe III.3.4.a, l'amplification du bruit de mesure liée aux deux dérivations successives ainsi que le pas de mesure (en V_{gs}) limitent fortement la précision sur V_{th}. . . . 89
III.9 Extraction des paramètres A et Γ sous le seuil, en régime de faible inversion. . . . 90
III.10 Par définition, la fonction Y est insensible aux variations de $R_{sd}(V_{gs})$ au premier ordre. . . . 92
III.11 Extraction de V_{th} depuis une régression sur la fonction Y [Ghibaudo88]. Le choix de la plage d'extraction où l'influence de Θ_2 peut être négligée conditionne la valeur du résultat. La régression linéaire en traits pointillés présente l'extraction sur la plage calibrée pour cette technologie. Les extractions présentées par des traits pointillés présentent d'autres choix possibles, soulignant l'incertitude sur la valeur de V_{th}, d'après [Fleury08a]. . . . 92

III.12 Illustration des deux plages de tensions communes à toutes les méthodologies d'extraction, qui définissent la zone dans laquelle la fonction Y peut s'appliquer ($V_{gs} \in [V_{gs,min1}, V_{gs,max1}]$) et celle où le paramètre Θ_2 peut être extrait ($V_{gs} \in [V_{gs,min2}, V_{gs,max2}]$). 93
III.13 Algorithme d'extraction itératif proposé par Mourrain dans [Mourrain00]. 94
III.14 Evolution de l'erreur sur les paramètres du modèle en fonction du nombre d'itérations n, lors de l'application de la technique décrite dans [Mourrain00], donnée par l'algorithme de la figure III.13. Transistor nMOS issu de la technologie 45nm ($L = 40nm$, $W = 1\mu$m). 95
III.15 Modification des courbes $Y(V_{gs})$ et $\Theta_{eff}(V_{gt})$ avant et après 5 itérations successives. 96
III.16 Augmentation de l'écart-type sur la valeur de V_{th} résultant du procédé d'optimisation dans la méthode Y-Φ [Tanaka07]. Un facteur deux sur l'étalement de la distribution de V_{th} résulte de la procédure de linéarisation dans le cas de la méthode Y-Φ. 96
III.17 Extraction à partir de la fonction de Faynot-McLarty (F_{Mc}). L'amplification du bruit de mesure en forte inversion dégrade la précision sur les paramètres V_{th} et β. 97
III.18 Evolution de l'erreur ε_n en fonction du nombre d'itérations n, pour différents choix de valeur initiale $V_{\mathrm{th}0}$. L'algorithme permet une convergence quasi-absolue quand l'erreur sur la valeur initiale est comprise entre -80mV et 100mV. ε_n devient de l'ordre de la précision de calcul après 7 itérations. . 99
III.19 Détermination de ΔV_{th} pour permettre la concordance avec la méthodologie d'extraction par partie décrite dans [Cros05]. 101
III.20 Extraction des paramètres du modèle grâce à une régression non-linéaire (de type Levenberg-Marquardt) dans la zone d'inversion forte. L'utilisation des propriétés sous le seuil fournit une relation du type $V_{th} = f(\beta)$ qui permet supprimer un degré de liberté dans la résolution. L'extraction est ainsi stabilisée par rapport à ce qui a été vu au §III.2.3. 102
III.21 Dérive de l'extraction de V_{th} lors de l'ajout d'une composante de résistance série ΔR_{sd}, pour les différentes méthodes d'extraction évoquées dans ce chapitre. 103
III.22 Evolution de l'erreur statistique sur V_{th} lors de l'ajout d'une composante de résistances séries ΔR_{sd}. 104
III.23 Dérive du maximum de transconductance et de la tension de grille associée lorsque R_{sd} augmente. 105
III.24 Dérive de l'extraction de V_{th} lors de l'ajout d'une composante de résistance série ΔR_{sd}, pour les méthodes de type fonction Y, suivant l'origine du V_{th} utilisé pour définir les plages d'extraction. . 106
III.25 Dérive des valeurs V_{th} et β issues des méthodes de type fonction Y, en fonction du décalage de $V_{gs,min1}$, utilisé dans la définition de la plage d'extraction. L'avantage de la nouvelle méthode "inversion faible + forte", qui ne nécessite pas de plage d'extraction, ressort clairement. 107
III.26 Incertitude sur l'extraction de la résistance série avec la méthode $R_{tot}(L)$ lorsque les variation de mobilité ne sont pas compensées. La valeur correcte est estimée à environ $100\Omega \cdot \mu$m pour cette technologie. 108
III.27 Extraction de $R_{sd}(V_{\mathrm{gs}})$ pour les nMOS et pMOS de la technologie 45nm. Les longueurs de grille vont de 35nm à 1μm, la largeur étant fixée à $W = 1\mu$m. 109
III.28 L'incertitude sur l'extraction du paramètre $\Theta_{1,eff}$ se répercute sur la valeur extraite de R_{sd}. . . . 110
III.29 Extraction de R_{sd} sur des transistors nMOS sur silicium massif en utilisant la nouvelle technique, pour $V_{gt} = 0.4$V. La valeur de R_{sd} est lue au point d'intersection tandis que la pente permet d'extraire l'amplitude de la réduction de mobilité pour cette polarisation. La distribution de l'erreur par rapport à la régression linéaire et les limites hautes et basses $\pm 2\sigma$ sont indiquées en insertion. . 111
III.30 Extraction de R_{sd} par la méthode $R_{tot}(1/\beta)$ en appliquant volontairement un décalage de R_{sd} sur la courbe $I_d(V_{gs})$. Les valeurs extraites reproduisent bien les variations appliquées, confirmant la précision relative de la nouvelle méthode d'extraction. 112
III.31 Extraction de la résistance série en fonction du champ effectif, d'après [Lin07] 114

IV.1 Différents schémas de câblage possibles pour mesurer la réponse capacitive des transistors MOS. . 120
IV.2 Schéma équivalent de la capacité C_{gb} en régime d'inversion. La formation d'une couche d'inversion reliée à la masse vient écranter la réponse des charges de désertion. 120
IV.3 Comparaison entre modèle analytique (IV.6) et la mesure directe de la réponse capacitive $C_{bg}(V_{gs})$. Le modèle donne un bon agrément avec la mesure, en accord avec [Lau99] et permettant de confirmer sa validité pour notre technologie CMOS 45nm. 122
IV.4 Validation du modèle analytique (IV.8) grâce à des simulations numériques. La fonction d'écrantage F_{shield} donne un bon accord avec le comportement $C_{gb}(V_{gs})$ et permet de reproduire fidèlement l'écrantage de la capacité de désertion en inversion. 123
IV.5 Schéma illustrant les différentes contributions à la capacité parasite C_{par}, inhérentes à l'architecture du transistor MOS, d'après [Prégaldiny02, Khakifirooz08, Wei09]. 124

IV.6 Diagramme de bande d'une coupe prise dans la largeur de la zone de recouvrement d'un transistor nMOS en régime d'inversion. L'application d'une tension positive sur la grille provoque la désertion des extensions S/D et augmente la valeur de $C_{ov}(V_{gs})$. 126

IV.7 Mesure $C_{gc}(V_{gs})$ effectuée sur une structure sans recouvrement afin d'observer le comportement de $C_{if}(V_{gs})$ dans la gamme de tensions $V_{gs} < V_{th}$. Le transistor est issu de la technologie 45nm LSTP ($L_{mask} = 45$nm, $W = 1\mu$m). Lorsque la couche d'accumulation fait son apparition dans le canal ($V_{gs} \leq V_{fb}$) la capacité C_{if} est écrantée, phénomène analogue à l'écrantage de C_{dep} observé sur la mesure C_{gb}. La valeur maximale de $C_{if}(V_{gs})$ est observée dans la gamme $V_{fb} \leq V_{gs} \leq V_{th}$ qui présente un plateau caractéristique. 127

IV.8 Simulation d'une courbe $C_{gc}(V_{gs})$ d'un transistor nMOS de longueur nominal de la technologie 45nm (sans recouvrement), pour différentes valeurs de α. 128

IV.9 Schéma équivalent d'un MOSFET au cours d'une mesure C_{gc} (la capacité de bord interne n'est pas représentée). En haut à gauche : évolution de la capacité de jonction $C_{junc}(V_{gs})$ en fonction du niveau de dopage substrat N_{sub}, le dopage des extensions N_{SDE} étant fixé à $5 \times 10^{26}/\text{m}^3$. . . . 129

IV.10 Diagramme d'énergie de la structure Silicium-Oxyde-Silicium considérée dans notre étude. 131

IV.11 Evolution de Q_{inv}, Q_{dep} et du rapport $\partial Q_{dep}/\partial Q_{inv}$ en fonction de V_{gs} pour un transistor simulé de 17.5Å d'épaisseur d'oxyde. En inversion forte, la variation de Q_{dep} faible devant celle de Q_{inv} et peut être négligée dans les calculs. L'erreur résultant de la prise en compte ou non de $\partial Q_{dep}/\partial Q_{inv}$ dans le calcul est illustré en insertion. 143

IV.12 Les charges d'inversion (Q_{inv}) et de désertion (Q_{dep}) sont définient par l'intégration respectives de la mesure capacitive $C_{gc}(V_{gs})$ et $C_{gb}(V_{gs})$. 144

IV.13 La mobilité effective extraite par la fonction Y et la technique *split*-CV fournissent des résultats en ligne en forte inversion. 145

IV.14 Illustration des différentes longueurs qui peuvent être définies dans l'architecture du transistor MOS. 147

IV.15 Courbes $C_{gc}(V_{gs})$ effectuées pour différentes longueurs de transistors ($W = 1\mu$m). 148

IV.16 Courbes $C_{gc}(V_{gs})$ issues de mesures sur des transistors nMOS de la technologie 65nm ($t_{ox} = 18.5$Å, $W = 10\mu$m). Le niveau de capacité parasite a été volontairement extrapolé dans la gamme de tensions en inversion afin d'illustrer nos explications (ligne discontinue). Le changement de pente indiqué est caractéristique de l'écrantage de la capacité C_{if}. L'évolution de la capacité de recourvrement en fonction de V_{gs} a été simulé numériquement pour différentes valeurs de dopages, en insertion de la figure. 149

IV.17 Comparaison entre les valeurs de ΔL extraites avec la méthode $\Delta L(L_{mask})$ et celles extraites grâce à la méthode du ΔL constant. L'extraction a été effectuée sur des transistors nMOS précurseurs à la technologie 45nm [Dumont07, Pouydebasque05] ($t_{ox} = 12$Å, $W = 1\mu$m). Dans ce résultat, la valeur de la capacité en inversion à été mesurée pour $V_{gt} = 850$mV. 150

IV.18 Structure de test matricées, utilisées pour mesurer L_{eff} de manière automatique en utilisant une matrice de connexion. La structure présentée ici est composées de 1980 transistors de dimensions $L_{mask} = 0.15\mu$m et $W_{mask} = 0.38\mu$m, fournissant une surface équivalente de $113\mu\text{m}^2$. 151

IV.19 Illustration de la procédure d'extraction automatique mise en place. Quelques points de mesure ($n \leq 10$) permettent d'extraire la tension de seuil qui est ensuite réutilisée pour extraire la valeur de la capacité en inversion corrigée des contributions parasites : C_{int}. Les valeurs ΔV_{inj} et ΔV_{par} sont étalonnées préalablement pour la technologie. 152

IV.20 Influence de la fuite de grille sur la mesure $C_{gc}(V_{gs})$ d'un transistor pMOS. L'apparition d'un courant au travers de l'oxyde impose ici de considérer une conductance ($1/R_p$) non nulle dans le schéma équivalent de l'appareil de mesure (HP4284). 153

IV.21 Comparaison entre la valeur de la réduction de la longueur de canal $L_{mask} - L_{eff}$ obtenue par mesures capacitive et la valeur $L_{mask} - L_{gate}$ mesurée grâce à des techniques de microscopie électronique à balayage. Les deux longueurs L_{eff} et L_{gate} présentent des variations similaires et la longueur de recouvrement $L_{ov}(L_{mask})$ est quasi-constante ($L_{ov} = 17 \pm 3$nm), en accord avec le fait que l'architecture de jonction reste la même quelque soit la longueur du dispositif. 154

IV.22 Evolution de la durée de vie lors de dégradation par porteurs chauds en fonction du rapport I_b/I_d (proportionnel au taux d'ionisation par impact) pour deux procédés de fabrication différents "A" et "B", entre lesquels L_{eff} est susceptible d'avoir changé. La dégradation de la durée de vie pour $L_{mask} = 70$nm peut être attribuée à un modification de la longueur effective d'environ 4nm, pour que les courbes $I_b/I_d(L_{mask})$ des deux procédés concordent. L'extraction de L_{eff} par mesure capacitive donne $L^A_{eff} - L^B_{eff} = (3.5 \pm 1)$nm qui valide la valeur extrapolée depuis les courbes $I_b/I_d(L_{mask})$. 155

IV.23 Variation de la longueur effective due à la diffusion des dopants lorsque la température du recuit est changée dans le procédé de fabrication. Dispositifs issus de la technologie CMOS 45nm LSTP. . 156

IV.24 Evolution de la longueur effective des transistors pour différentes doses de poche. Cette analyse semble faire apparaitre un compromis entre les performances en courant (mobilité) et la fiabilité des dispositifs courts. 157

V.1 A gauche : extraction de la réduction de longueur de canal en fonction de L_{mask} pour les deux types de transistors. A droite : évolution L_{eff} en fonction de la température. La précision de la mesure est de $\pm$1.7nm dans un intervalle de confiance de 95% (cf. histogramme en insertion). . . . 162

V.2 A gauche : extraction de la valeur R_{sd} pour différentes tensions V_{gt}. La précision de la méthode ne permet pas d'extraire une dépendance en température et une valeur constante, mesurée sur la gamme 210K–350K est considérée. A droite : évolution $R_{sd}(V_{gT})$ et $V_{th}(L_{eff})$ pour ces dispositifs. 162

V.3 Evolution de la mobilité à bas champ (μ_0) en fonction de la longueur effective de canal pour deux températures : 230K et 410K. Le modèle "linéaire" de dégradation de mobilité issue de [Bidal09a] a été ajusté sur les points expérimentaux. 164

V.4 A gauche : extraction de la dépendance en longueur de la composante neutre de mobilité : μ_N semble suivre une loi exponentielle. A droite : évolution $\mu_N(L_{eff})$ obtenue sur nos dispositifs. μ_N ne dépend pas de la température, justifiant de la validité des hypothèses utilisées pour faire ce calcul. 165

V.5 Modélisation de la dégradation de mobilité à deux températures en utilisant le modèle "exponentiel" et le modèle "linéaire". 166

V.6 Représentation de la mobilité apparente (mesurée expérimentalement) et de sa valeur corrigée de l'influence de la balisticité. Il a été supposé $\nu_{inj} \simeq 1.2 \times 10^7$cm/s. Ce résultat confirme le rôle mineur que joue la balisticité dans la dégradation de mobilité observée pour les courtes longueurs de grille, en accord avec [Cros06b, Zilli07a]. 167

V.7 Evolution du nombre de publications sur le thème de la balisticité dans le transistor MOS au cours de ces dix dernières années, dans les journaux et conférences les plus réputées. 169

V.8 Evolution du courant I_{on} et de la fréquence de coupure f_T en fonction de la longueur de grille des transistor, dans le cas d'un transport balistique et d'un transport dérive-diffusion d'après [Palestri05, Eminente07]. 169

V.9 Illustration de la méthode utilisée dans [Lochtefeld01] pour calculer la charge à la source virtuelle des transistors courts, à partir de la mesure $C_{gc}(V_{gs})$ effectuée sur un transistor long (pour lequel l'influence des capacités parasites peut être négligée). 171

V.10 A gauche : Evolution de la mobilité effective et du libre parcours moyen en fonction de L_{eff} pour $V_{gt} = 1$V. L'insertion montre une forte corrélation entre μ_{eff} et λ_μ ($R^2 \simeq 1$), permettant de justifier la cohérence de nos calculs. A droite : probabilité de passage balistique $\mathcal{P}_{bal}$ calculée grâce à la formule décrite dans [Leo94], en supposant $\nu_{inj} = (1.8 \pm 0.3) \times 10^7$cm/s. 175

V.11 Représentation schématique du calcul permettant d'aboutir à la relation (V.14) à partir des équations présentées au chapitre II. 176

V.12 Protocole d'extraction permettant la mesure de ν_{lim}. 177

V.13 Il est difficile d'identifier le mécanisme limitant car ν_{inj} ou ν_{sat} ont des valeurs proches. Ces mesures ne sont pas corrigées de la valeur des résistances séries qui contribue à la saturation des courants dans les deux régimes, à fort V_{gs}. 179

V.14 Evolution de ν_{sat} et de ν_{inj} en fonction de la température dans la gamme (200K – 400K). Les dépendances de ν_{sat} et ν_{inj} sont opposées et peuvent permettre l'identification du mécanisme responsable de la saturation de la vitesse. Remarque : les valeurs de ν_{sat} sont issues de [Jacoboni77] et mesurées dans le silicium massif, celles de ν_{inj} sont issues de simulations Monte-Carlo [Zilli07b]. 179

V.15 A gauche : évolution du courant I_{on} en fonction de L_{eff}. La saturation de courant semble avoir une dépendance en température en contradiction avec l'hypothèse de la quasi-balisticité. A droite : évolution du rapport des mobilités apparentes à 233K et 413K en fonction de la longueur de grille. La saturation pour les dispositifs courts semble en accord avec l'influence croissante d'une composante de la mobilité indépendante de la température, comme observée dans [Cros06b]. 181
V.16 Influence de la mobilité sur le courant en régime linéaire et de saturation. La perte d'influence en régime de saturation est caractéristique d'un phénomène de saturation de vitesse. En suivant l'approche de Lundstrom [Lundstrom01], il est possible d'extraire un taux de balisticité d'environ 40% sur ces dispositifs, en accord avec les résultats de la littérature. 183
V.17 Extraction de la vitesse de transport effective ν_{eff}en régime linéaire et saturé, d'après la méthode de [Lochtefeld01]. 184
V.18 Extraction de la vitesse de transport effective ν_{eff} en régime linéaire et saturé, d'après la méthode de [Lochtefeld01]. 185
V.19 La vitesse limite des porteurs à la source montre une augmentation de plus de 40% sur les dispositifs courts (en comparaison des longs). La dépendance en température, signature du mécanisme physique sous-jacent, est sensiblement la même quelque soit L_{eff}. 186
V.20 Extraction de la vitesse de transport effective ν_{eff} en régime linéaire et saturé, d'après la méthode de [Lochtefeld01]. 186
V.21 Corrélation entre l'évolution de ν_{lim} et la valeur du DIBL (intégrité électrostatique). 188
V.22 A gauche : évolution du gradient du champ en fonction du champ au niveau de la source. L'évolution du potentiel dans les canaux courts permet au phénomène de survitesse d'être efficace dès la source. A droite : utilisation du modèle de survitesse introduit dans [Ge01] pour reproduire la tendance expérimentale $\nu_{lim}(L_{eff})$. 189
V.23 A gauche : le modèle de survitesse est ajusté sur les points expérimentaux $\nu_{lim}(L_{eff})$ pour les deux températures extrêmes de cette étude. Un bon accord est trouvé lorsque le temps de relaxation en énergie est fixé à $\tau_w = 0.35$ps. A droite : évolution de la vitesse de saturation stationnaire (extrapolée depuis le modèle de survitesse) en fonction de la température. Les points extraits présentent un bon accord (valeurs et pente) avec la dépendance en température trouvée dans [Jacoboni77], compte tenu des incertitudes sur nos valeurs. 190
V.24 Neurone de rat interfaçant un dispositif électronique de type transistor MOS, expérience réalisée au Max Planck *Institute for Biochemistry* [Voelker05]. Cette application ouvre la voie à de nouvelles possibilités de calcul (non-déterminisme) sans pour autant interdire une co-intégration avec un bloc logique "traditionnel" : **l'avenir ?** 195

B.1 Structure cristalline du silicium montrant les liaisons tétraédriques avec les proches voisins (gauche) et position atomique des atomes dans la maille en projection sur une face (droite) [Kittel98] 199

Liste des tableaux

II.1 Valeurs de λ_{opt} issues de la littérature 55
II.2 Probabilité de passage balistique d'un porteur dans un transistor MOS, en considérant que seule l'interaction avec les phonons optiques limite le transport à fort champ latéral. 56
II.3 Dépendance en température et en champ des différentes contributions à la mobilité bas champ. . . 60

III.1 extraction de R_{sd} à $V_{gs} = 1.1V$ pour des dispositifs sur silicium massif et FDSOI 112

IV.1 Coefficients de Tchebyshev permettant d'approximer la fonction $F_{1/2}(\eta), \forall \eta \in \mathbb{R}$ grâce à la relation (IV.51). D'après [Werner63]. 136

B.1 Caractéristiques crystallines du silicium 200

Remerciements

"La science a-t-elle promis le bonheur ? Je ne le crois pas. Elle a promis la vérité, et la question est de savoir si l'on fera jamais du bonheur avec de la vérité."
Emile Zola

BIEN que les remerciements ne constituent pas la partie la plus conséquente du manuscrit de thèse, leur rédaction n'en est pas pour autant triviale. Elle nécessite d'avoir un recul suffisant sur l'accomplissement de la thèse, et plus largement sur une période de travail intense qui va bien au delà des trois dernières années.

Le monde de la microélectronique paraît bien rude et sauvage. J'y ai vu des entreprises achetées et négociées par des investisseurs sans âmes, telles de vulgaires marchandises embarquées dans un système boursier fragilisé, en pleine tourmente.
J'ai vu les hommes et leur savoir se disperser, évincés par une poignée de charlatans au service de quelques actionnaires : *divide et impera*[1].
Enfin, j'ai pu me confronter au comportement des quelques pseudo-scientifiques, qui servent leur propres intérêts sans jamais contribuer à promouvoir le savoir. Je les remercie pour m'avoir montré l'exemple à ne pas suivre !

Mais dans ce monde de *brutes*, au cours de ce "périple doctoral", j'ai heureusement rencontré de *vrais* esprits scientifiques qui m'ont aidé et soutenu. Je les remercie pour m'avoir guidé sur les voies du savoir.

Je tiens tout d'abord à remercier mes parents qui m'ont toujours soutenu et supporté, même dans les plus mauvais moments. Sans eux tout cela n'aurait jamais été possible. Merci également à ma chérie pour m'avoir accompagné depuis le début de cette aventure.

Je remercie STMicroelectronics et le laboratoire IMEP-LAHC pour m'avoir accueilli et permis de travailler sur ce sujet de recherche passionnant. En particulier, merci à Emmanuel Vincent de m'avoir intégré comme un membre à part entière de son équipe pendant ces trois années.

Je remercie chaleureusement Gérard Ghibaudo, mon directeur de thèse qui, grâce à ses connaissances et son ouverture d'esprit démesurées, m'a aguillé et propulsé sur des pistes de recherche prometteuses.

1. "Diviser pour mieux régner", proverbe romain utilisé par les Habsbourg

Je remercie Antoine Cros et Krunoslav Romanjek, mes encadrants, pour l'amicalité et le soutien qu'ils m'ont apportées au cours de ces années. Ensemble, nous avons formé une vraie équipe : le MOS-club !

J'adresse toute ma gratitude à David Roy, perfectionniste d'une motivation débordante qui m'a généreusement fait découvrir son métier lors de mes premiers débuts dans le monde industriel. J'aurais toujours du regret à savoir que finalement, après tant d'années, nous ne travaillerons pas ensemble.

Merci également à Perceval et Grégory, collègues et surtout proches amis, pour toutes ces discussions et ces bons moments de rigolades (au Japon, à Grenoble ou simplement au café !).

Merci à Myriam Vialle pour sa bonne humeur permanente qui m'a apporté tant de joie dans mon travail quotidien.

Enfin, à toutes les autres personnes que j'ai côtoyées ces dernières années et dont l'énumération des noms serait bien trop longue : je ne vous oublie pas et vous remercie sans modération.

Merci à tous !

Introduction générale

"The definition of "Moore's Law" has come to refer to almost anything related to the semiconductor industry that when plotted on semi-log paper approximates a straight line."
Gordon Earle Moore

L'invention du transistor par W. Shockley, J. Bardeen et W. Brattain au AT&T *Bell Laboratories* en décembre 1947 fût certainement la plus grande découverte du XXe siècle, dans le domaine de l'électronique (cf. figure 1). Le transistor a en effet permis un développement fulgurant des moyens de communications et s'impose aujourd'hui comme un élément fondamental dans la conception de presque tous les appareils électroniques. Son grand succès vient en particulier d'une capacité à être miniaturisé, intégré dans des circuits complexes et produit en masse grâce à des procédés automatisés pour atteindre des coûts unitaires extrêmement bas. La microélectronique représente aujourd'hui un secteur clé dans l'économie mondiale : un chiffre d'affaire d'environ 265 milliards de dollars et 3 millions d'emplois. Qui aurait pu croire, à l'époque, qu'il deviendrait le principal moteur de l'innovation technologique et industrielle ?

Fig. 1 – Transistor amplificateur « point-contact »construit en 1947 par W. Shockley, J. Bardeen et W. Brattain (AT&T *Bell Laboratories*). Le prix Nobel de Physique leur sera décerné en 1956.

La technologie CMOS[2] est à la base de l'électronique numérique et tient une place privilégiée dans cet essor. Elle offre en effet des possibilités d'intégration et de réduction d'échelle inégalées (qui semblent sans limite), tout en gardant une alimentation raisonnable (contrairement aux

2. *Complementary Metal-Oxide-Semiconductor*

technologies nMOS et pMOS). Ainsi, depuis 1975 jusqu'aujourd'hui, les observations de G.E. Moore sont vérifiées : la densité d'intégration et les fréquences de fonctionnement des circuits doublent tous les 18 à 24 mois [Moore95]... **mais à quel prix ?**

Aujourd'hui, l'évolution de la technologie CMOS semble en effet vivre une mutation : les gains de performances sont plus faibles que la prédiction de Moore d'un nœud à l'autre et les courants de fuite augmentent inévitablement. Des solutions existent, elles sont plus complexes technologiquement et nécessitent la mise en place de moyens de recherche coûteux (centre de recherches, unité de production) pour être viables. Ainsi, la poursuite de la miniaturisation au delà du nœud 90nm (ex : processeur Intel Pentium 4 Prescott, cf. figure 2) s'accompagne d'une explosion des coûts de conception et de production liés aux avancées technologiques (la construction d'une unité de fabrication en 2002 coûtait déjà près de 2 milliard de dollars).
Une révolution est également en marche concernant l'utilisation de l'électronique : il ne s'agit plus seulement de concevoir des circuits plus puissants, mais également d'abaisser la consommation pour les applications portatives et d'accroitre le nombre de fonctions embarquées. Par exemple, la popularisation des technologies sans-fil haute fréquence (la téléphonie mobile) introduit de lourdes contraintes sur la puissance de consommation et la fréquence de fonctionnement des circuits. La conception des technologies CMOS à très faible consommation (même en veille, *Low STandby Power* : LSTP) est un défit technologique pour l'avenir que plusieurs industries tentent de relever (STMicroelectronics en fait partie).

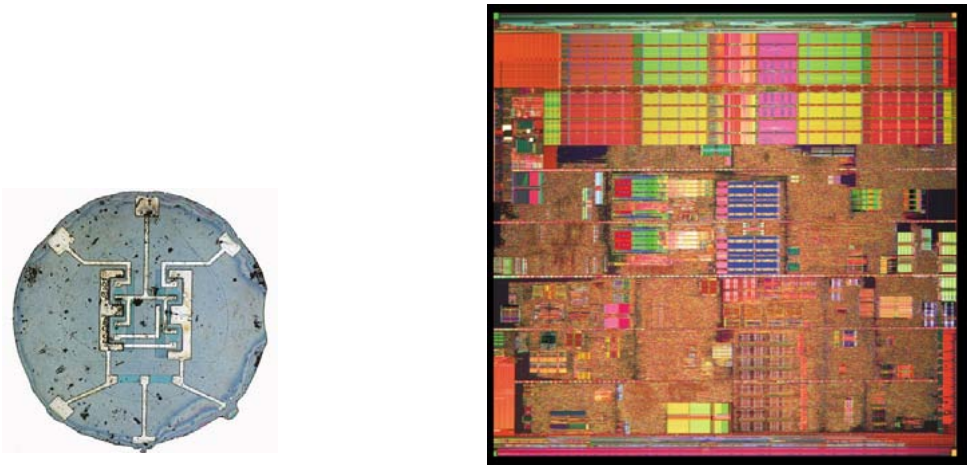


Fig. 2 – A gauche : Premier circuit intégré assemblé grâce au procédé *planar*. Commercialisé publiquement en 1961 par la société Fairchild Semiconductor, il est composé de 2 transistors interconnectés sur une tranche de silicium (fonction logique *flip-flop*). A droite : processeur Prescott 1M commercialisé en février 2004 par la société Intel. Il et composé d'environ 125 millions de transistors et utilise le nœud 90nm qui marque un tournant dans la miniaturisation de la technologie CMOS.

La vue en coupe d'un transistor *p*MOS issu de la technologie 45nm d'Intel est représenté sur la figure 3 [Mistry07]. Son architecture intègre des éléments nouveaux telle qu'une grille métallique, un diélectrique de type *high-*κ et des source/drain SiGe qui permettent de contraindre mécaniquement le canal (pour améliorer sa conductivité). Cette architecture illustre bien le contexte dans lequel s'intègre notre étude, et plus généralement, l'analyse des mécanismes de transport dans les transistors "modernes". Il s'agit de comprendre l'influence de plusieurs méca-

nismes mal connus qui influent de manière complexe sur la qualité de transport dans le canal. Le guidage des choix technologiques présents et futures dépend directement de notre capacité à comprendre cette physique.

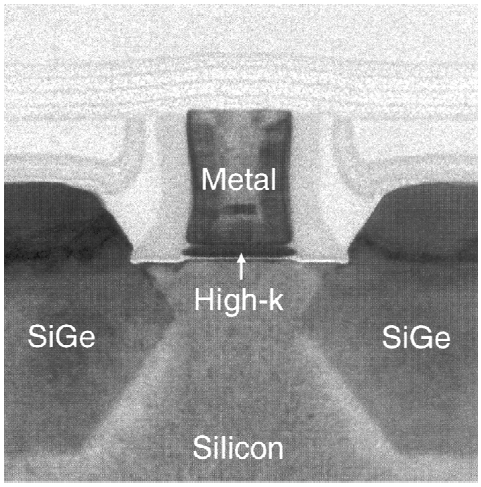


Fig. 3 – Vue en coupe d'un transistor *p*MOS de la technologie 45nm d'Intel [Mistry07]. Son architecture inclue une grille métallique, un oxyde de type *high-*κ et des extensions S/D en SiGe qui transfèrent une contrainte compressive au canal (pour améliorer sa conductivité).

Le travail présenté dans ce manuscrit a été réalisé au sein de l'équipe de Caractérisation et Fiabilité qui fait partie du Support au développement des plateformes technologiques de ST-Microelectronics, Crolles. Dans ce contexte industriel, la caractérisation électrique permet une meilleure compréhension de la physique des composants (ici le transistor MOS) afin d'anticiper les problématiques futures et de guider les technologues dans leurs choix, pour concevoir des technologies performantes et pérennes. Ce manuscrit apporte quelques éléments de réponse pour éclairer la compréhension physique des mécanismes de transport et les capacités parasites caractéristiques des transistors modernes.

Dans un premier chapitre (I) intitulé "Le transistor MOS ... en bref", nous introduirons les principes de fonctionnement du transistor MOS et les difficultés technologiques majeures rencontrées au cours de son évolution. Ce chapitre a pour but de donner les bases nécessaires au lecteur pour qu'il puisse aborder sereinement les chapitres suivants.

Une description détaillée des mécanismes de transport des porteurs dans le transistor MOS sera abordée au chapitre II. Cette partie permettra de mieux comprendre les mécanismes physique et leur rôle vis-à-vis des performances. La problématique du transport balistique sera détaillée dans cette partie.

Le troisième chapitre (III) dressera le panel des méthodologies d'extraction qui ont été utilisées dans ce travail de thèse. Elles permettent d'extraire les paramètres qui contiennent l'information physique nécessaire à la compréhension des dispositifs. Nous détaillerons en particulier les innovations qui ont été apportées au domaine de la caractérisation électrique au cours de cette thèse.

Dans le chapitre IV, nous proposons d'analyser la capacité MOS et de comprendre les contri-

butions parasites inhérentes à la structure. Nous verrons en particulier comment une meilleure compréhension de ces capacités a permis d'aboutir à la construction d'une nouvelle méthodologie d'extraction, adaptée à la mesure de la longueur de canal sur les transistors courts.

Finalement le cinquième chapitre V portera sur l'étude expérimentale du transport dans les dispositifs d'aujourd'hui. Cette analyse requiert des méthodes d'extractions et des techniques de mesures avancées, justifiant de l'utilité des chapitres précédents. Grâce à une étude solide reposant sur une nouvelle méthodologie d'extraction, nous tenterons d'élucider le mécanisme de transport majoritaire qui régit les performances de nos dispositifs. L'hypothèse du transport balistique (ou quasi-balistique) sera discutée et évaluée dans ce chapitre.

Chapitre I

Le transistor MOS ... en bref

Sommaire

I.1 Architecture et principe de fonctionnement du transistors MOS . . 23
I.1.1 La barrière de potentiel source-drain 24
I.1.2 Définition analytique de la tension de seuil 25
I.1.3 Les régimes de conduction du transistor MOS 25
I.2 Les limites à la miniaturisation de la technologie CMOS 27
I.2.1 La perte de l'intégrité électrostatique 27
I.2.2 Influence de la miniaturisation sur le transport : des performances amoindries 29
I.2.3 La puissance consommée et le délai de commutation 29
I.3 Les solutions pour une poursuite de la miniaturisation 31
I.3.1 La maîtrise des courants de fuite grâce à l'empilement métal/*high-*κ . . 31
I.3.2 Des contraintes mécaniques pour *doper* le transport 32
I.3.3 Les architectures à film de conduction mince 33

RÉSUME — Ce chapitre a pour but d'introduire les concepts de base pour mieux comprendre le fonctionnement du transistor MOS et les limitations qui résultent de la miniaturisation. Ils permettront au lecteur d'aborder le travail réalisé au cours de cette thèse, traité dans les chapitres II à V.

⋆ Dans une première partie (§I.1), nous détaillerons l'architecture du transistor MOS et expliqueront brièvement sont principe de fonctionnement. Nous aborderons les règles de miniaturisation historique qui régissaient encore l'évolution de la technologie CMOS il y a quelques années.

⋆ La seconde partie (§I.2) exposera les limitations et les problèmes qui apparaissent avec la miniaturisation des dispositifs. Nous expliquerons les raisons pour lesquelles, les règles historiques sont corrompues et ne s'appliquent plus aujourd'hui.

⋆ Finalement, dans une troisième partie (§I.3), nous exposerons quelques unes des solutions qui ont été utilisées jusqu'à présent pour repousser les limites de la miniaturisation et permettent aujourd'hui à la technologie CMOS de perdurer.

I.1 Architecture et principe de fonctionnement du transistors MOS

Dans une application logique, le transistor est utilisé comme un interrupteur : il doit pouvoir commuter entre un état passant (débit de courant *on*) et bloqué (pas de courant *off*) en fonction de la polarisation de grille. En générale, les transistors sont alimentés par une tension unique V_{dd} et les états *on* et *off* sont caractérisés respectivement par les configurations $V_{gs} = V_{dd}$ et $V_{gs} = 0\text{V}$ ($V_{ds} = V_{dd}$)[1], cf. figure I.1.

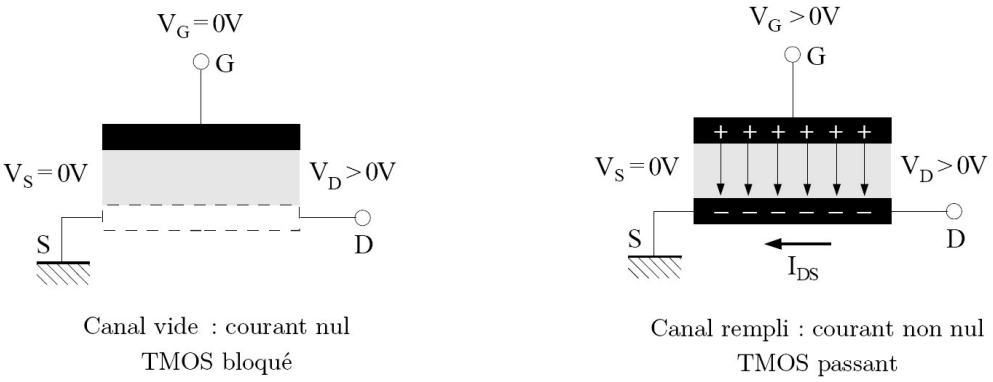


Fig. I.1 – Principe de fonctionnement du transistor à effet champ dans les applications numériques.

La figure I.2 présente l'architecture d'un transistor MOS sur silicium massif de type n (le canal en inversion est peuplé d'électrons). L'architecture classique est constituée de quatre terminaux qui permettent d'analyser le comportement électrique du transistor : la grille (V_g), la source (V_s), le drain (V_d) et le contact du substrat (V_b).

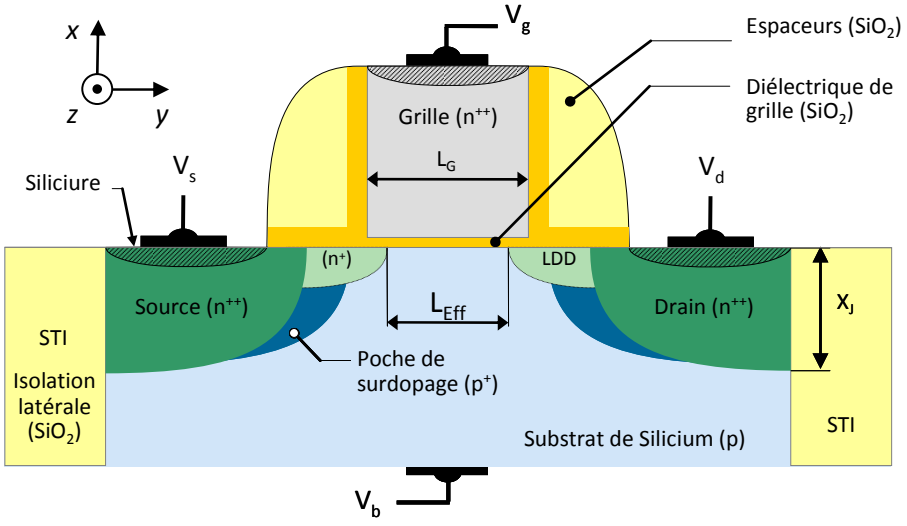


Fig. I.2 – Architecture d'un transistor nMOS sur silicium massif, d'après [Skotnicki03].

La structure du transistor MOS est constituée des éléments suivants :
- une électrode de grille qui permet de contrôler l'état du canal et le passage du courant,
- un canal qui permet le passage des porteurs entre la source et le drain,

1. Dans toute nos notations, la source sera considérée comme montre référence de potentiel : $V_s = 0$

– un diélectrique permettant d'isoler la grille du canal pour permettre un couplage grille-canal capacitif,
– les réservoirs de porteurs source et drain qui fournissent les porteurs minoritaires et permettent la formation d'une couche d'inversion.

Le principe du transistor MOS consiste à moduler la densité de porteurs du canal en jouant sur la polarisation de grille (V_{gs}). Celle-ci permet, par le biais du champ électrique ainsi créé, d'attirer une quantité de porteurs depuis les réservoirs à la surface du semiconducteur. Une couche conductrice se forme alors à ce niveau et permet la libre circulation d'un courant (I_{ds}), sous réserve de l'application d'un champ électrique (i.e. d'une tension V_{ds}). Cette couche conductrice est appelée "couche d'inversion" ou parfois "canal d'inversion". Pour plus d'informations, le lecteur pourra se référer à [Skotnicki03].

I.1.1 La barrière de potentiel source-drain

Le dopage de types opposés des réservoirs et du substrat forme deux jonctions dont la barrière de potentiel intrinsèque (*built-in potential* dans [Sze81]) empêche la libre circulation des porteurs. La tension de grille V_{gs} module la hauteur de cette barrière et quatre régimes caractéristiques de la structure MOS peuvent ainsi être définis (cf. figure I.3) :

– **le régime d'accumulation** ($V_{gs} < V_{fb}$) : les porteurs majoritaires du volume sont attirés en surface qui est localement accumulée
– **le régime de bandes plates** ($V_{gs} = V_{fb}$) : les concentrations de porteurs majoritaires et minoritaires sont constantes dans toute la profondeur du silicium,
– **le régime de désertion et d'inversion** ($V_{gs} > V_{fb}$) : les porteurs majoritaires sont repoussés de la surface, laissant les impuretés ionisées non compensées. Le terme "inversion" est utilisé lorsque la concentration de porteurs minoritaires est supérieure à celle des porteurs majoritaires en surface (condition respectée pour $V_{gs} > V_{th}$).

Dans notre exemple, V_{fb} caractérise la tension de grille qu'il est nécessaire d'appliquer sur la grille pour obtenir la condition de bandes plates telle qu'illustrée sur la figure I.3, (b). Elle résulte des différences de travaux de sortie entre l'électrode de grille et le canal et est fortement influencée par la présence de défauts électriquement actifs dans l'oxyde ou à l'interface oxyde-canal, cf. [Sze81].

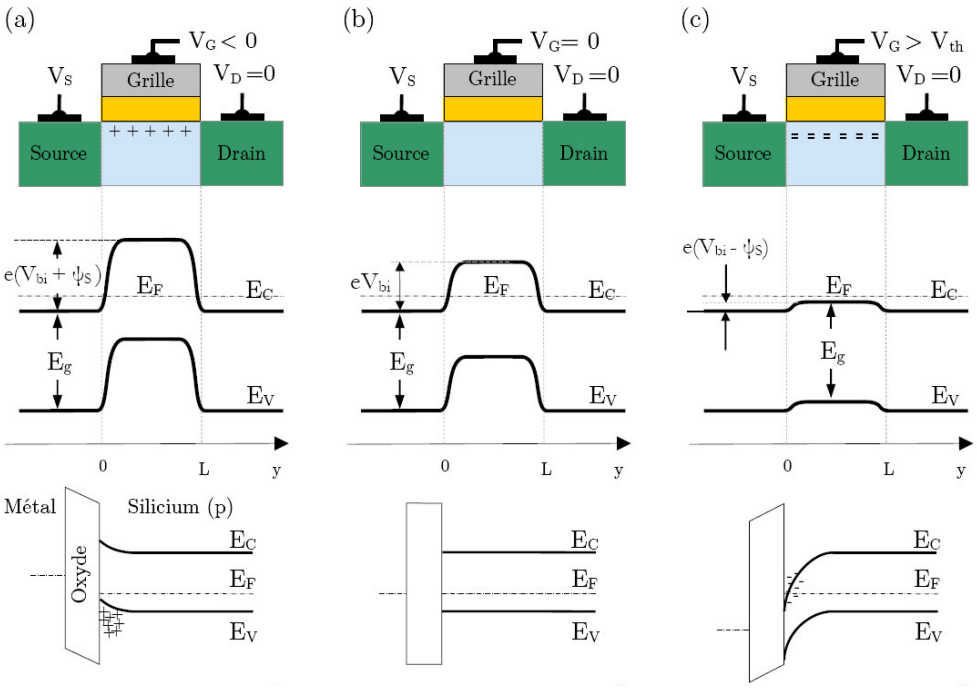


Fig. I.3 – Modification de la barrière de potentiel source-drain en fonction de la polarisation de grille V_{gs}. (a) régime d'accumulation, (b) régime de bandes plates (hypothèse $V_{fb} = 0$), (c) régime d'inversion forte.

I.1.2 Définition analytique de la tension de seuil

Lorsque V_{gs} est suffisamment élevée, la surface du semiconducteur est amenée en régime d'inversion : un canal d'inversion est créé entre la source et le drain. On parle d'inversion forte lorsque la population de porteurs est localement inversée. La tension à appliquer pour amener le dispositif dans ce régime s'appelle la tension de seuil V_{th} (pour *threshold voltage*). Son expression analytique est donnée par [Sze81, Wong05] en (I.1), où C_{eff} représente la capacité effective de l'oxyde de grille. N_D la concentration volumique en impuretés dans le canal (dopage).

$$V_{th} = V_{fb} + 2\phi_F + \frac{\sqrt{4\varepsilon_{Si} q N_D \phi_F}}{C_{eff}} \tag{I.1}$$

La variable ϕ_F quantifie l'écart d'énergie entre les niveaux de Fermi intrinsèque et extrinsèque.

I.1.3 Les régimes de conduction du transistor MOS

Le courant délivré par le transistor est gouverné par trois facteurs :
- la charge dans le canal (qui dépend de la tension de grille V_{gs}),
- la tension de drain V_{ds} qui permet le déplacement des porteurs de la source au drain,
- la mobilité des porteurs (μ_{eff}) qui quantifie la vitesse acquise par les porteurs sous l'action d'un champ électrique. Cette notion sera détaillée au chapitre II.

En fonction des polarisations V_{gs} et V_{ds} il est donc possible de définir deux régimes de conduction distincts : **le régime linéaire (ou ohmique) et le régime de saturation**. Le régime linéaire se caractérise par une dépendance *linéaire* du courant en fonction de V_{ds} tandis que le régime de saturation caractérise un état dans lequel le courant n'évolue plus avec V_{ds}, car le

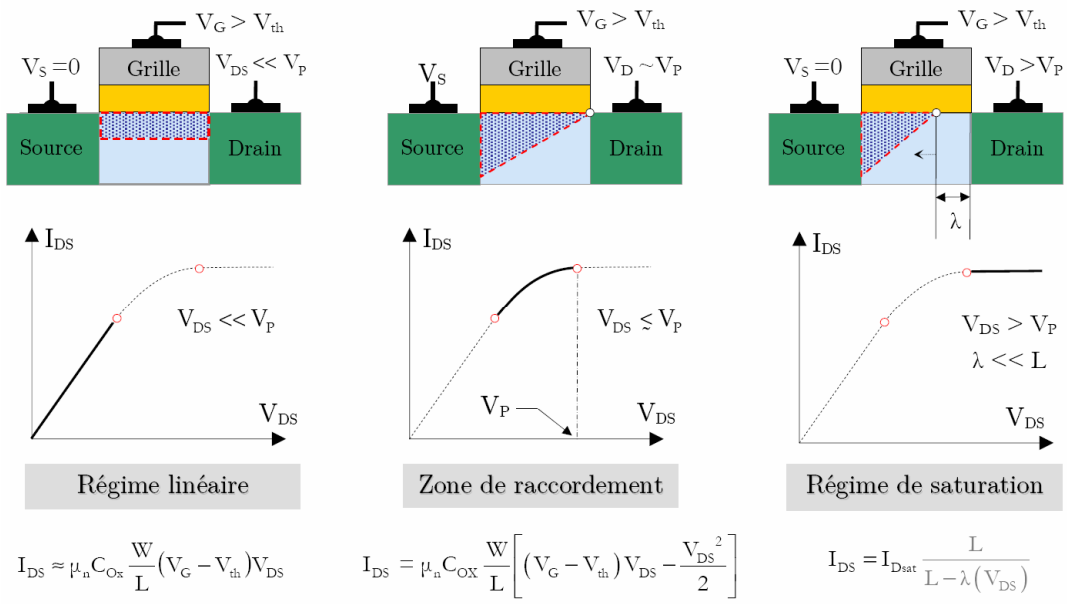


Fig. I.4 – Illustration des trois régimes de conduction qui caractérisent le fonctionnement d'un transistor nMOS

canal est dit *pincé* $V_p \simeq V_{gs} - V_{th}$[2]. Entre ces deux régimes, le courant délivré dépend à la fois de V_{ds} et de V_{gs} : c'est la zone de raccordement.

Au chapitre III, nous verrons que les propriétés du régime linéaire permettent d'aboutir à des expressions du courant relativement simples. Ce point est particulièrement intéressant pour extraire les paramètres électriques du transistor et analyser son comportement électrique.

2. Nous verrons que lorsque le transistors est soumis aux effets de canaux courts, V_{th} dépend de V_{ds} (effet de DIBL) et provoque une augmentation du courant au delà de V_p.

I.2 Les limites à la miniaturisation de la technologie CMOS

Au cours des dernières années, la technologie CMOS a montré un capacité à la miniaturisation sans précédent et qui surpasse celle des autres technologies. Aujourd'hui, la littérature reporte des dispositifs fonctionnels dont la longueur effective de canal est inférieure à 4nm [Wakabayashi03, Suk09]. Cela nous amène à la question suivante : **Quelles sont les limites à la réduction d'échelle pour la technologie CMOS ?**

Les règles du *Happy Scaling* proposée par Robert Dennard [Dennard74] rendent compte de la réduction d'échelle historique sur les différents modules technologiques du MOSFET. Révisées par Baccarani en 1984 [Baccarani84], elles prédisent une stagnation de la puissance et une augmentation de la fréquence de fonctionnement lorsque les dimensions géométriques sont réduites. Il est également prévu un gain en puissance par grille considérable lorsque la tension d'alimentation diminue.

Les paragraphes suivants vont exposer les limites à cette règle de réduction d'échelle qui ne peut plus s'appliquer au transistors de longueur de grille déca-nanométrique.

I.2.1 La perte de l'intégrité électrostatique

La réduction de la longueur de canal entraîne l'émergence des "effets de canaux courts" (ou SCE pour *Short Channel Effects*). Ils traduisent une perte de contrôle électrostatique de la grille et sont le résultat de l'influence de zones de désertions non contrôlées dans le canal. Sans aucune polarisation, le rapprochement des zones de charges d'espace qui caractérisent les jonctions de part et d'autres du canal provoquent un abaissement de V_{th}, car une partie du canal est déjà en désertion. C'est le phénomène de partage de charge, cf. [Sze81].

L'application d'une tension V_{ds} sur le drain peut abaisser d'avantage la barrière source-drain et le contrôle de la charge devient dépendant de la grille **et du drain** : c'est l'effet de DIBL (pour *Drain Induced Barrier Lowering*).

Ces deux phénomènes sont représentés schématiquement sur la figure I.5. Ils constituent une des limitations la plus problématique, qui empêchent ultimement une miniaturisation très poussée des technologies CMOS sur silicium massif.

La notion d'intégrité électrostatique (EI) introduite par T. Skotnicki permet de quantifier l'influence des mécanismes de partage de grille et DIBL sur le comportement électrique des transistors [Skotnicki03]. Son expression analytique est donnée par (I.2) et fait intervenir les dimensions caractéristiques de l'architecture du MOSFET :

- La profondeur de jonction X_J,
- l'épaisseur de l'oxyde t_{ox},
- la longueur effective de canal L_{eff},
- la profondeur de désertion T_{dep}.

$$EI = \left[1 + \left(\frac{X_J}{L_{eff}}\right)^2\right] \cdot \frac{t_{ox}}{L_{eff}} \cdot \frac{T_{dep}}{L_{eff}} \tag{I.2}$$

Le meilleur moyen de conserver une bon contrôle électrostatique tout en miniaturisant les dispositifs semble nécessiter la réduction de T_{dep} (architecture à canal de silicium fin) ou la

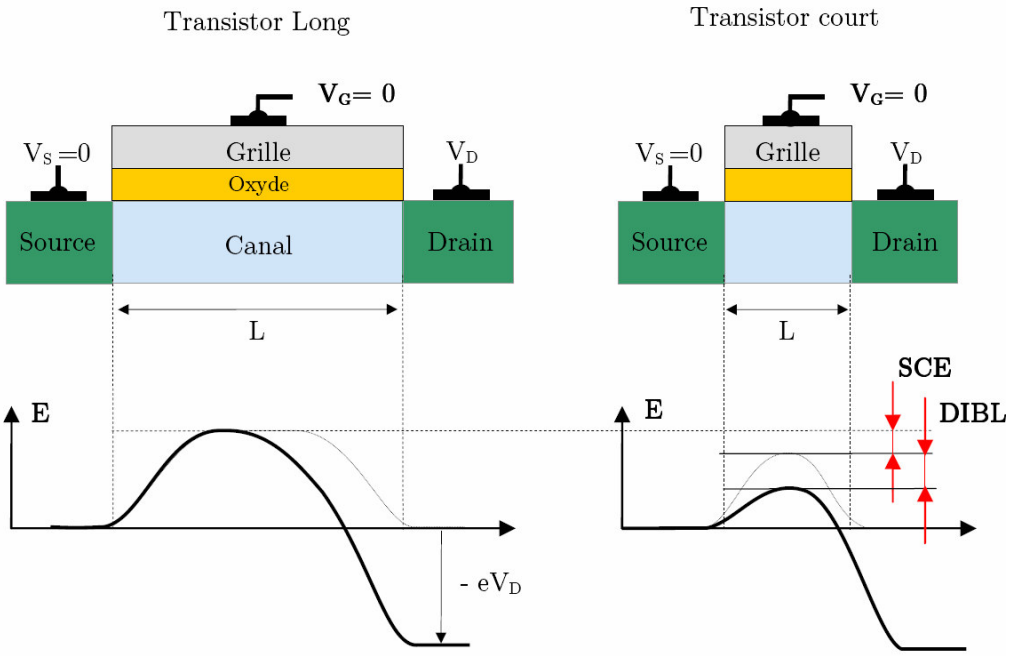


Fig. I.5 – Illustration de l'abaissement de la barrière de potentiel source-drain induit par les effets de canaux courts (partage de charge et DIBL)

conception de jonctions S/D peu profondes (cf. [Dumont07]).

I.2.2 Influence de la miniaturisation sur le transport : des performances amoindries

La réduction d'échelle a une influence directe sur la mobilité des porteurs dans le canal. Les points suivants sont particulièrement critiques et forment une réelle limitation vis-à-vis de la miniaturisation :

- l'augmentation du dopage canal, nécessaire pour contrer les effets de canaux courts, entraîne une dégradation de la mobilité [Takagi94a, Cao99],
- l'apparition de mécanismes additionnels qui provoquent une dégradation de la mobilité sur les canaux courts [Cros06b]. Ce point sera détaillé dans le chapitre II.

La mobilité a une influence directe sur le courant débité par le transistor [Lochtefeld01]. **Cela signifie donc qu'il ne suffit plus de réduire la longueur effective de canal pour augmenter le courant débité par le transistor (comme il est suggéré dans [Dennard74, Baccarani84]) : il faut également trouver des solutions qui permettent de limiter la dégradation de mobilité qu'introduit la réduction d'échelle !**

> Le concept de mobilité utilisé jusqu'à présent peut perdre sa signification physique pour les canaux très courts, dans lesquels le mécanisme de transport balistique (sans interaction dans le canal) peut devenir influent. La théorie du transport balistique (et quasi-balistique) sera détaillée au chapitre II. Une étude expérimentale des propriétés de transport dans les transistors à canaux courts sera menée au chapitre IV.

I.2.3 La puissance consommée et le délai de commutation

Le développement soutenu des technologies portatives (téléphonie, informatique) et l'augmentation fulgurante du prix du lithium, utilisé pour concevoir les batteries à forte densité d'énergie, oriente la conception technologique vers des circuits rapides qui consomment peu. C'est la problématique majeure qui gouverne la conception des technologies CMOS qui seront analysées dans ce manuscrit.

L'augmentation des performances des transistors est essentielle pour la miniaturisation des technologies CMOS. Les paramètres à optimiser pour une technologie LSTP (Low STandby Power) sont principalement les suivants :

- la puissance dissipée par le dispositifs en régime de fonctionnement à l'état "on" : P_{on},
- la puissance consommée passivement (à l'état "off") : $P_{off} = V_{dd} \times I_{off}$,
- la fréquence de fonctionnement maximale des circuits.

I.2.3.a Le délai de commutation

Le délai de commutation moyen d'un inverseur CMOS peut être approximé *grossièrement* par la relation (I.3). V_{dd} est la tension d'alimentation, I_{eff} le courant moyen en régime de fonctionnement (introduit dans [Na02]) et C_{eff} la capacité effective, caractéristique du couplage de la grille-canal (les capacités parasites de la structure sont incluses dans cette valeur).

$$\tau \simeq \frac{C_{eff} \times V_{dd}}{I_{eff}} \tag{I.3}$$

Cette relation révèle simplement les limites de la miniaturisation vis-à-vis du délai de commutation des inverseurs (en lien avec la fréquence de fonctionnement maximale des circuits).

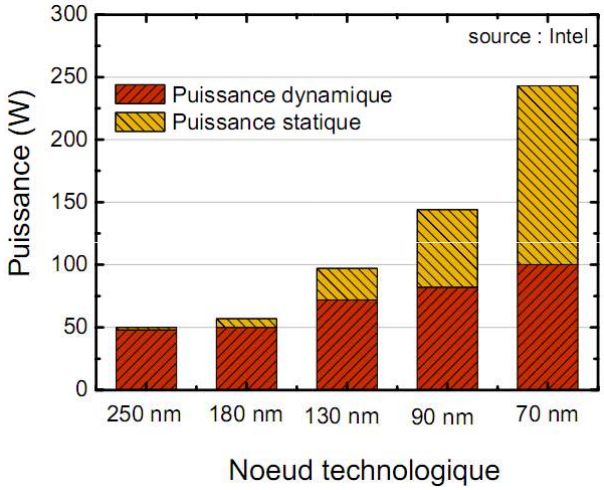


Fig. I.6 – Évolution de la puissance dynamique (P_{on}) et de la puissance statique (P_{off}) en fonction du nœud technologique [Rochette08]. Si la puissance dynamique reste presque constante, la puissance statique augmente dramatiquement avec la miniaturisation.

- le courant I_{eff} doit être maximisé (en réduisant la longueur de canal L_{eff}),
- l'influence croissante des capacités parasites[3] de la structure augmentent le délai (i.e. τ augmente lorsque L_{eff} est réduite) [Khakifirooz08],
- la tension V_{dd} doit être réduite.

I.2.3.b La puissance consommée

L'optimisation de la consommation active (P_{on}) et passive (P_{off}) est un véritable défi à relever pour les technologies d'aujourd'hui et de demain. N'oublions pas que la communauté scientifique était déjà confrontée à un problème similaire dans les années 70 (technologie bipolaire) et dans les années 80 (nMOS). A l'échelle du circuit, les solutions de *design* permettent de minimiser la consommation passive grâce à la mise en veille (coupure d'alimentation) des bloques non utilisés. La puissance P_{off} est néanmoins fondamentalement limitée à l'échelle du transistor : fuites de jonctions, fuites à travers l'oxyde de grille et fuites sous le seuil qui s'amplifient avec les effets de canaux courts [Dumont07]. Cette contribution, qui était négligeable pour les technologies jusqu'en 1990, est aujourd'hui supérieure à la dissipation dynamique P_{on} (cf. figure I.6).

Parallèlement à cela, la puissance P_{on} ne peut être réduite sans abaisser V_{dd}. La difficulté qui apparait lors de la miniaturisation de la technologie CMOS sur silicium massif est donc de taille : il est difficile d'envisager la conservation des performances si P_{on} se dégrade (à cause de la mobilité) et que P_{off} augmente (effets de canaux courts, fuites de jonction, fuites de grille).

3. Les capacités parasites qui interviennent dans l'architecture MOS seront détaillées dans le chapitre III.

I.3 Les solutions pour une poursuite de la miniaturisation

Nous venons de voir les différents phénomènes qui compromettent les règles *historiques* de la miniaturisation proposées dans [Dennard74, Baccarani84]. Il existe néanmoins des solutions pour repousser ces limites et permettre une réduction des dimensions tout en augmentant les performances des transistors.

I.3.1 La maîtrise des courants de fuite grâce à l'empilement métal/*high-*κ

L'introduction d'un empilement de grille de type *high-*κ-métal a été industrialisée par Intel dès 2007, dans le processeur Penryn [Bohr07]. Les diélectriques *high-*κ (souvent à base d'Halfnium) permettent, grâce une permittivité diélectrique plus élevée que dans le SiO_2, de relaxer les contraintes sur l'amincissement de l'oxyde de grille requis par les règles de miniaturisation. Il est ainsi possible de conserver des performances élevées (un fort couplage capacitif grille-canal C_{eff}) tout en conservant une épaisseur d'oxyde suffisamment grande pour limiter les courants de fuite à travers l'oxyde.

L'intégration des oxydes *high-*κ avec une grille en silicium polycristallin pose des problèmes quasi-insurmontables (*Fermi Level Pinning* [Hobbs03]) et il est nécessaire d'introduire une grille métallique dans l'architecture (la plupart du temps Titane, Carbure de Tantale ou en Tungstène). Cette dernière ne permet pas seulement de résoudre le problème d'intégration du diélectrique mais supprime également l'effet de désertion de grille qui augmente l'épaisseur apparente d'oxyde (perte de couplage capacitif, augmentation du délai de commutation).

Dans la littérature, la valeur de CET (pour *Capacitance Equivalent oxide Thickness*) est couramment utilisée. Elle renvoie l'épaisseur physique de SiO_2 (permittivité $\varepsilon_{SiO_2} = 3.9$) qui aurait été nécessaire pour obtenir la même capacité C_{eff} qu'avec la diélectrique *high-*κ. Le calcul du CET est donné par $\varepsilon_{SiO_2}/C_{eff}$[4].

Si l'intégration des empilements de type *high-*κ-métal semble repousser les difficultés liées à la miniaturisation (augmentation de I_{eff}), elle s'accompagne de nombreux problèmes :

- la fiabilité des dispositifs est dégradée par rapport à un empilement Si/SiO_2 [Rafik08],
- les techniques de dépôt sont plus complexes et plus coûteuses pour un diélectrique *high-*κ (ex : *Atomic Layer Deposition*) que le procédé d'oxydation thermique utilisé dans le cas du SiO_2,
- les performances sont généralement dégradées par l'empilement de grille (l'effet sur la mobilité sera approfondi au chapitre II),
- jusqu'à présent, l'intégration des diélectriques *high-*κ nécessite une couche d'oxyde piédestal SiO_2 pour limiter la dégradation de la mobilité (*Remote Coulomb Scattering*) et assurer un arrangement amorphe. Cet oxyde piédestal limite l'augmentation de C_{eff}[5],
- l'introduction d'un oxyde *high-*κ doit être associé à une grille métallique pour éviter l'effet de *Fermi Level Pinning* [Hobbs03, Samavedam03], d'où une complexification du procédé de fabrication (en particulier pour assurer les types n et p).

4. La notion de CET englobe tous les effets parasites qui peuvent augmenter l'épaisseur apparente d'oxyde : désertion de grille, effets quantiques, etc.

5. Certaines études sont en cours pour permettre de déposer le diélectrique *high-*κ à même le Si [Huang09].

– l'ajustement de la tension de seuil nécessite des techniques ingénieuses et non triviales [Cartier09]

L'intégration de l'empilement de grille *high-κ*-métal est donc un réel défi technologique que seul Intel avait su surmonter en 2007. Aujourd'hui cette solution semble indispensable pour la miniaturisation de la filière CMOS et est adoptée par la plupart des industries de la microélectronique.

I.3.2 Des contraintes mécaniques pour *doper* le transport

La contrainte mécanique permet de déformer la maille cristalline du silicium[6]. Cette déformation s'accompagne d'un changement de la courbure de bande qui peut conduire à une modification dans le peuplement des différentes vallées (donc à un changement de masse effective) [Skotnicki08][7]. La figure I.7 illustre l'influence d'une contrainte uniaxiale <110> sur le comportement des électrons et des trous d'une surface (100). Une contrainte tensive :

– améliore la mobilité des électrons en augmentant la population de la vallée transverse ,
– diminue la mobilité des trous en augmentant la population de trous lourds.

Au contraire, une contrainte compressive :

– améliore la mobilité des trous en augmentant la proportion de trous légers
– diminue la mobilité des électrons en augmentant la proportion d'électron dans la bande transverse.

Une augmentation des performances des nMOS et des pMOS, pour une orientation <110>/(100) nécessite donc l'utilisation de deux techniques de contraintes (tensive et compressive). La plupart des dispositifs utilisés dans cette étude (chapitres III, IV et V) bénéficient des techniques CESL (*Contact Etch Stop Layer*) et SMT (*Stress Memory Technique*) qui permettent d'appliquer une contrainte tensive (dépendante de L) favorable à la mobilité des électrons. Précisons que sur nos dispositifs, le substrat est tourné à 45° afin de ne pas dégrader la mobilité des trous.

La contrainte a ses propres limites, fixées par la zone d'élasticité du matériau. La limite de déformation plastique définit la contrainte maximale qu'il est possible d'appliquer sans faire apparaître de dislocations (relaxation de l'excès de contrainte par le matériau). Des gains en courant importants peuvent néanmoins être obtenus. En particulier, les techniques de dépôt de grille damascène *gate last* couplées aux techniques de contraintes via des extensions SiGe permettent d'atteindre des mobilités records [Mayuzumi09a, Mayuzumi09b].

6. Il est vivement conseillé de consulter les références [Skotnicki08, Bidal09b, Weber05] pour plus d'informations sur les contraintes mécaniques.
7. La modification de la structure de bande peut également influencer le temps de relaxation des porteurs.

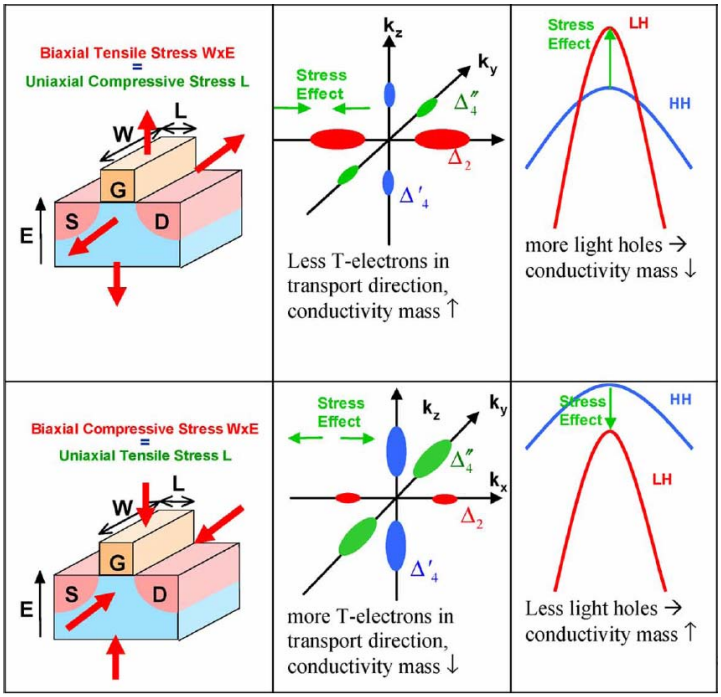


Fig. I.7 – Influence d'une contrainte uniaxiale sur le comportement des électrons et des trous, d'après [Skotnicki08].

I.3.3 Les architectures à film de conduction mince

L'architecture du transistor SOI (pour *Silicon On Insulator*) se différentie de l'architecture sur silicium massif par la présence d'un couche d'oxyde enterrée (BOX : *Buried OXide*) qui isole les transistors du substrat de silicium. La couche de conduction est ainsi limitée (T_{Si}) et, lorsque $T_{Si} < T_{dep}$, est totalement désertée (on parle alors de transistor FDSOI, pour *Fully Depleted* SOI). L'architecture typique d'un transistor FDSOI est présentée sur la figure I.8. Les principaux avantages du FDSOI devant les architectures sur silicium massif sont :

- un contrôle électrostatique accru grâce à une profondeur de jonctions limitée à l'épaisseur du film de silicium,
- un canal non dopé, qui permet potentiellement une mobilité des porteurs accrue,
- un gain de conductivité par la réduction de T_{dep} (limitée par T_{Si}),
- une meilleure isolation des dispositifs qui peut permettre une densité d'intégration plus forte.

L'épaisseur de l'oxyde enterré joue un rôle significatif sur l'amplitude du phénomène d'auto-échauffement du dispositif (dû à la faible conductivité thermique du BOX) [Lee95]. Ce phénomène modifie certains paramètres électriques et conduit à une dégradation globale des performances du transistor :

- la tension de seuil s'abaisse (augmentation de l'énergie thermique des porteurs),
- la mobilité se dégrade (plus d'interactions avec les phonons),
- la pente sous le seuil s'élève et par ce fait le courant de fuite (I_{off}) augmente.

Il est également à noter que la propagation des lignes de champ au sein du BOX perturbe la forme du potentiel entre la source et le drain et modifie la tension de seuil du dispositif. Cet effet encourage à utiliser un oxyde enterré mince de manière à limiter ce phénomène. L'emploi d'une faible épaisseur d'oxyde enterré nécessite toutefois de doper fortement le substrat afin de

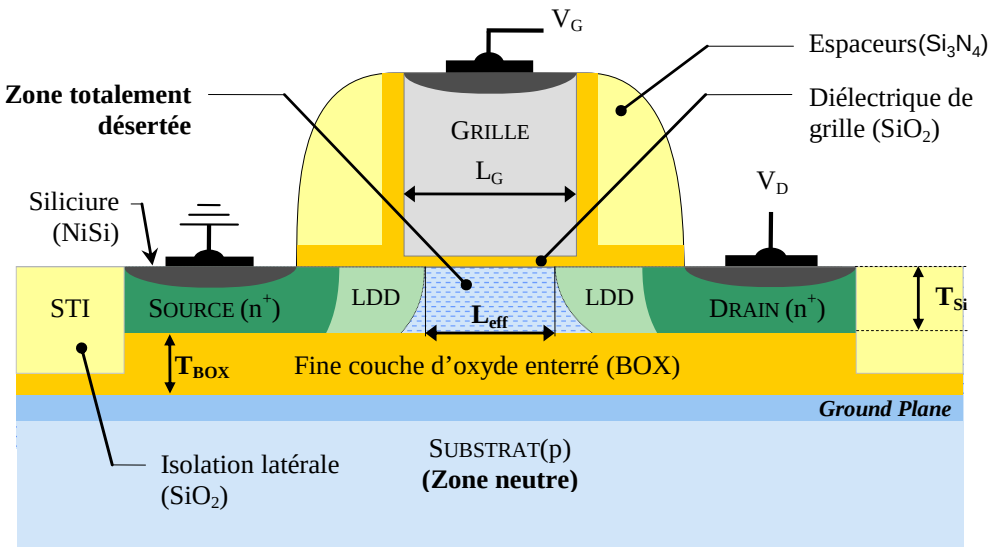


Fig. I.8 – Architecture d'un transistor FDSOI.

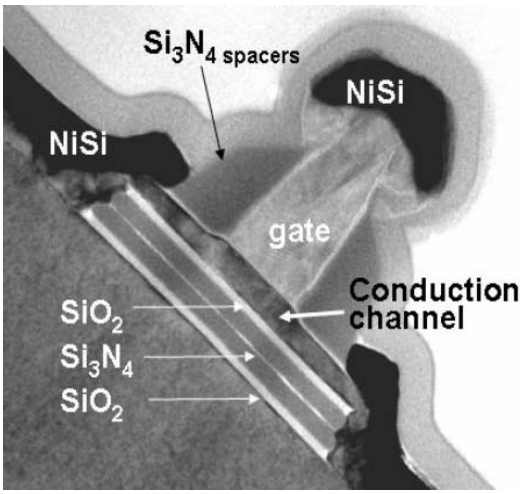


Fig. I.9 – Vue en coupe, prise au microscope électronique à transmission (TEM), d'un transistor d'architecture SON.

limiter la profondeur de la zone de désertion créée sous le BOX. Cette zone de désertion vient s'ajouter à celle classiquement observée dans le canal et dégrade de la même façon le contrôle des effets de canal court. On parle d'effet *Ground Plane* [Celler03].

I.3.3.a Du silicium sur rien ?

L'architecture *Silicon on Nothing* (SON) fût historiquement introduite par Jurczak et al. [Jurczak99]. Elle est similaire à l'architecture SOI mais permet de localiser le BOX uniquement sous le canal de conduction, laissant ainsi les jonctions source et drain en contact avec le substrat (cf. figure I.9). Cette technique permet de conserver les avantages procurés par l'architecture SOI (film de conduction mince) tout en limitant les effets parasites tels que l'auto-échauffement. Elle favorise également une bonne co-intégration avec des dispositifs sur silicium massif, contrairement au solutions FDSOI utilisant un substrat SOI.

Pour plus d'informations sur l'architecture SON, le lecteur pourra se référer au travail de thèse de G. Bidal dans [Bidal09b].

I.3.3.b Les difficultés liées aux films minces

Si les architectures de types FDSOI à film mince **et BOX mince** semblent fournir une solution à la perte d'intégrité électrostatique liée à la miniaturisation, elle soulève plusieurs problèmes qui limitent les performances globales des transistors :

- l'utilisation de films minces entraîne l'augmentation des résistances d'accès. L'élaboration de source et drain totalement siliciurés pose problème car la consommation de Si imposée par la formation du siliciure peut conduire à une rupture du canal ou à l'augmentation des résistances d'accès (cas du SOI uniquement). Deux solutions sont alors envisagées : **(a)** concevoir des architectures de type SON dans lesquelles la source et le drain sont en contact avec le substrat et où la siliciuration peut s'étendre en profondeur, sans risquer une rupture du canal, **(b)** l'élaboration une architecture où la source et le drain sont surélevés (i.e. au même niveau que l'empilement de grille) ;
- la finesse du film de conduction (et donc la profondeur de désertion) abaisse la tension de seuil qui devient difficile à ajuster uniquement grâce au dopage du canal. L'utilisation de grilles métalliques mid-gap ou totalement siliciurées est inévitable ;
- les films de conduction fins ($T_{Si} < 20$nm) sont associés à une dégradation de mobilité [Uchida02].

Chapitre II

Étude théorique du transport dans le canal du transistor MOS

Sommaire

II.1 Du transport dérive-diffusion à la balisticité : le rôle des mécanismes de collision 39
II.1.1 Le concept de mobilité 40
II.1.2 Hypothèses et validité du modèle de Drude 40
II.1.3 Le libre parcours moyen : la clé pour comprendre les mécanismes de transport 40
II.1.4 Notion de continuité du courant 42
II.1.5 Description du mécanisme d'injection à la source 43
II.2 Le transport stationnaire 45
II.2.1 Notion de mobilité effective et de champ effectif 46
II.2.2 Interactions avec les centres de diffusion chargés 49
II.2.3 Interactions avec les vibrations du réseau : les phonons 51
II.2.4 L'influence de la rugosité de surface sur la mobilité 56
II.2.5 Les mécanismes additionnels de réduction de la mobilité 57
II.2.6 Influence de la température et du champ sur la mobilité 60
II.3 Le transport non-stationnaire 61
II.3.1 Les phénomènes de survitesse : une théorie vite oubliée 61
II.3.2 Le modèle de transport quasi-balistique : généralités et approximations 63
II.3.3 La notion de distance critique (kT-*layer*) et de rétrodiffusion 67
II.4 Vers l'unification des mécanismes de transport 69
II.4.1 Expression du courant de drain dans l'approche dérive-diffusion 69
II.4.2 Le concept de mobilité apparente 69
II.4.3 L'unification des théories de transport 70
II.5 Bilan et discussions 74

RÉSUME — La qualité du transport dans les transistors MOS est l'un des paramètres clés pour la réduction d'échelle des technologies CMOS. Son amélioration permet en effet l'augmentation du courant de saturation et donc la réduction de la tension d'alimentation des circuits V_{dd} pour minimiser puissance consommée tout en maintenant des performances élevées. Ce chapitre pose les bases théoriques qui nous permettent de comprendre et d'analyser le déplacement des porteurs libres dans le canal du transistors.

⋆ Dans la première partie (§II.1), nous détaillerons le rôle des mécanismes de collision et introduirons la manière dont leur compréhension a historiquement permis d'étendre la loi d'Ohm au transistors MOS, via le modèle de Drüde. Nous y décrirons également le concept de libre parcours moyen qui constitue le lien entre un régime de transport régi par de multiples collisions et une approche balistique. Nous traiterons également le concept d'injection à la source qui décrit la manière dont les porteurs sont injectés des réservoirs vers le canal pour participer au transport.

⋆ La seconde partie (§II.2) nous permettra de décrire les différents mécanismes de collision qui interviennent dans le transport, sous une approche de transport stationnaire.

⋆ Nous introduirons ensuite les phénomènes de non-stationnarité du transport qui caractérisent les transistors courts et aborderons d'un œil critique les théories de la balisticité publiées à ce jour (§II.3).

⋆ La dernière partie de ce chapitre consistera finalement à introduire notre **modèle de transport unifié**, dans lequel les théories historiques et balistiques sont réunies de manière à fournir un modèle réaliste quelque soit la qualité du transport (§II.4).

⋆ Suite à nos discussions, nous conclurons sur les voies les plus prometteuses pour améliorer la mobilité du transistors MOS (§II.5).

II.1 Du transport dérive-diffusion à la balisticité : le rôle des mécanismes de collision

Dès les années 1900 les travaux de Paul Karl Ludwig Drude montrèrent la nécessité d'introduire le concept de collisions pour rendre compte de la loi d'Ohm locale dans les métaux et les semiconducteurs (II.1) [Drude00]. Dans cette expression, j, n, q, $m^\star$ et τ_{col} représentent respectivement la densité de courant, la densité d'électrons, la charge de l'électron, sa masse et la durée moyenne entre deux collisions successives (le temps de relaxation). Le paramètre σ est la conductivité du matériau, il permet d'exprimer la densité de courant délivré en fonction du champ électrique latéral $\mathcal{E}_{\|}$.

$$\overrightarrow{j} = \sigma \cdot \overrightarrow{\mathcal{E}_y}, \quad \sigma = \left(\frac{n \cdot q^2 \cdot \tau_{col}}{m^\star} \right) \tag{II.1}$$

La prise en compte des collisions dans le traitement du transport revient à considérer que les trajectoires des porteurs en présence d'un champ électrique ne sont pas monotones (approche de Newton) mais interrompues de nombreuses fois par des centres de diffusion qui modifient aléatoirement son orientation. En l'absence de champ électrique, l'énergie thermique est la seule source d'énergie qui est à l'origine du mouvement des porteurs. Le déplacement de groupe n'a alors pas de directionalité et une vitesse *de groupe* statistiquement nulle [Grove67] : $\nu_g = \overrightarrow{0}$. L'application d'un champ électrique ajoute une composante de vitesse non nulle au groupe de porteurs : on parle de vitesse de dérive ν_d (cf. Fig.II.1).

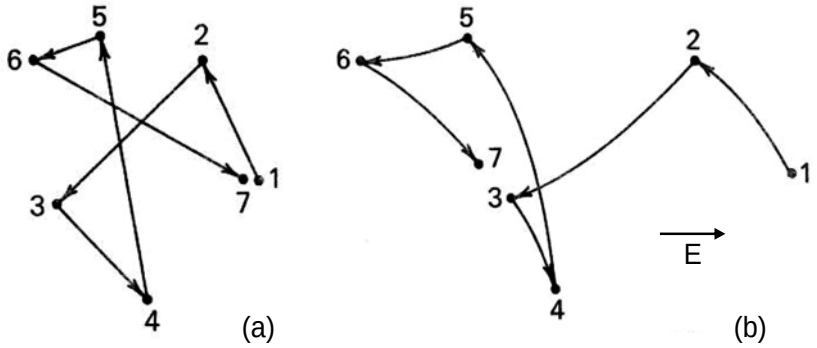


Fig. II.1 – Illustration de la trajectoire des électrons subissant des collisions dans un cristal (a) sans champ électrique, (b) sous l'application d'un champ électrique, d'après [Grove67].

II.1.1 Le concept de mobilité

La probabilité $\mathcal{P}_{col}$ qu'une collision se produise entre l'instant t et $t + dt$ est donnée par la relation $\mathcal{P}_{col} = dt/\tau_{col}$ [Ashcroft76]. Statistiquement, l'influence des interactions peut être prise en compte grâce au concept de mobilité (μ). Ce paramètre permet d'établir un lien simple entre le champ électrique appliqué ($\mathcal{E}_{\|}$) et la vitesse de dérive des porteurs (II.2). Le libre parcours moyen λ_μ caractérise la distance (moyenne) que les porteurs parcourent librement entre deux collisions successives. Il est donc un indicateur fiable de l'importance des collisions dans le déplacement des porteurs et permet d'évaluer la qualité de transport dans un matériau donné.

$$\nu_d = \mu \cdot \mathcal{E}_{\|}, \quad \mu \stackrel{\text{def}}{=} \frac{q \cdot \tau_{col}}{m^\star} \tag{II.2}$$

Cette relation correspond à un développement limité à champ faible. Elle n'est valable que si les collisions sont équiprobables en énergie, i.e. si la probabilité de collision ne dépend pas de l'énergie (et donc du champ électrique).

II.1.2 Hypothèses et validité du modèle de Drude

Le modèle de Drude se base sur des concepts de mécanique classique[1] pour décrire la conduction des électrons dans le matériau. Par exemple, il ne permet pas de prédire la dépendance en température de la conductivité (ou de la mobilité) qui doit alors être ajoutée dans le temps de relaxation [Ashcroft76]. Les hypothèses réalisées dans le modèle de Drude sont les suivantes :

- **les collisions sont de nature élastique :** elles modifient le moment des électrons de manière aléatoire et laissent leur énergie inchangée,
- **les électrons sont supposés indépendants :** les effets d'écrantages sont négligés,
- le temps de relaxation τ_{col} est indépendant de la configuration de l'électron (en particulier son énergie) au moment de la collision.

Nous verrons dans les prochaines sections qu'un modèle de transport prédictif et précis ne peut pas reposer sur de telles hypothèses et doit être étendu, en particulier à la prise en compte des collisions de nature inélastique. Ces dernières permettent en effet de rendre compte du phénomène de saturation de vitesse qui caractérise le transport à fort champ électrique [Grove67]. Ce point sera étudié en détail au §II.2.3.

II.1.3 Le libre parcours moyen : la clé pour comprendre les mécanismes de transport

Schématiquement, un transistor peut être considéré comme un ensemble constitué de deux réservoirs de porteurs (la source et le drain) reliés par un canal qui permet la circulation des charges libres (cf. figure II.2). Le courant de drain (I_d) que le dispositif peut débiter est alors fonction de deux éléments :

- la capacité de la source à injecter des porteurs dans le canal : elle dépend de la vitesse d'injection (ν_{inj}) et de la charge à la source (Q_s),
- la qualité du transport dans le canal : elle est évaluée par la mobilité μ (cf. II.1).

Le synopsis de la figure II.3 présente l'évolution des performances et des mécanismes de transport mis en jeu en fonction de la valeur du libre parcours moyen des porteurs.

1. Pour des vitesses de dérives supérieures à la vitesse thermique (10^7cm/s), la longueur d'onde des électrons devient supérieure au paramètre de maille du cristal : ils ne peuvent donc plus prétendre à un comportement corpusculaire [Wolf71]

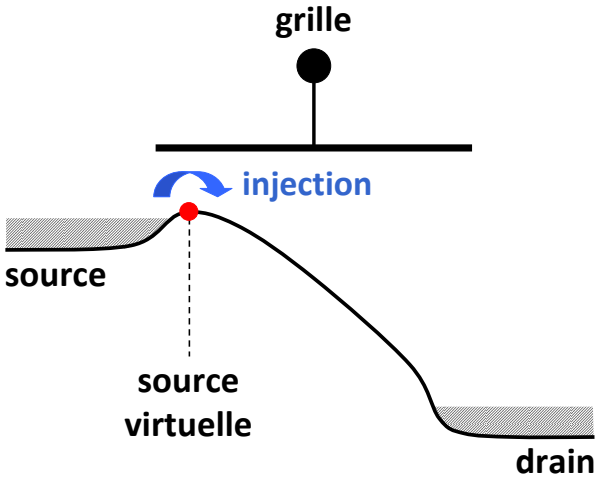


Fig. II.2 – Schéma simplifié du transistor MOS en régime de conduction : il est composé de deux réservoirs de porteurs reliés par un canal. L'injection est le mécanisme par lequel les porteurs quittent la source pour entrer dans le canal. Ce phénomène se produit au point de maximum de barrière de potentiel coté source appelée "source virtuelle".

Considérons en premier lieu un dispositif de longueur suffisamment courte et dont le canal offre des propriétés de transport suffisamment bonnes pour que les collisions avec les centres de diffusion soient rares et que la relation suivante soit respectée : $L_{eff} \ll \lambda_\mu$. Il paraît alors pertinent de considérer qu'une grande proportion de porteurs ne subit pas de collision entre la source et le drain : leur trajectoire est de type balistique. Le courant débité par le transistor est alors limité par le produit de la vitesse d'injection par la charge d'inversion à la source virtuelle : $I_{lim} = \nu_{inj} \times Q_s$ [Natori94].

Considérons maintenant le cas opposé d'un transistor dont le canal offrant des propriétés de conduction dégradées par la forte influence des collisions sur la trajectoire des porteurs : $L_{eff} \gg \lambda_\mu$. Dans ce cas, il est peu probable d'observer un transport de type balistique et la capacité d'injection de la source est peu influente car ce sont les collisions (et non le flux d'injection maximal à la source) qui limitent la valeur du courant. Un tel mécanisme de transport se modélise par l'approche dérive-diffusion (II.3) qui considère qu'un grand nombre de collisions entravent la trajectoire des porteurs entre la source et le drain. Dans cette expression, n_{inv} est la densité volumique de porteurs libres et D, le coefficient de diffusion relié au terme de mobilité par la relation d'Einstein ($D = \mu \cdot k_B T/q$) [Sze81].

$$j = q \cdot (n_{inv} \cdot \mu \cdot \mathcal{E}_y + D \cdot \nabla n_{inv}) \tag{II.3}$$

Au delà d'un certain champ critique $E_{\|c} \sim 2 \times 10^5$V/cm, la relaxation de l'énergie cinétique des porteurs par émission de phonons optiques devient fréquente et la relation linéaire entre la ν_d et $E_\|$ (II.2) n'est plus respectée. Cette déviation à la loi d'Ohm a pour effet de plafonner la vitesse de groupe : c'est le phénomène de vitesse de saturation [Shockley51, Grove67, Jacoboni77, Sodini84]. La relation d'Einstein ainsi que (II.3) ne sont alors plus valables car μ ne peut plus être considérée comme une constante, mais comme une fonction de $E_\|$ [Jacoboni77] (ce point sera abordé en détail dans la partie II.2.3.c).

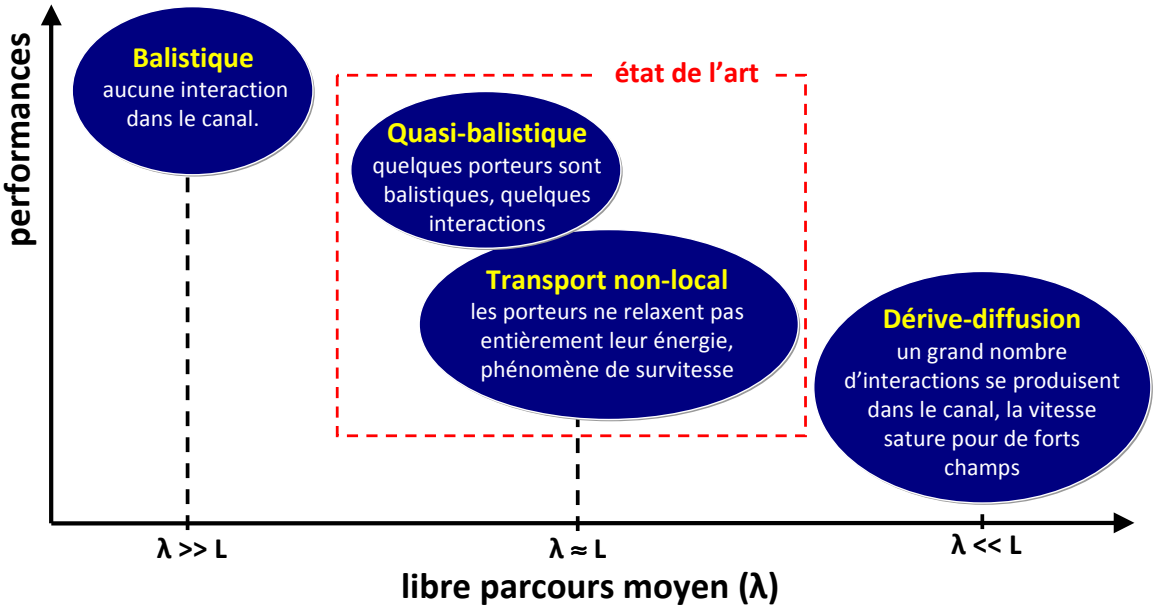


Fig. II.3 – Évolution des performances et des mécanismes de transport mis en jeu en fonction de la valeur du libre parcours moyen. L'état de l'art, pour des transistors fabriqués actuellement sur silicium, se situe proche de l'abscisse $\lambda_\mu \sim L$, correspondant à des mécanismes de transport à fort caractère non-local ou quasi-balistiques.

Que se passe-t-il entre ces deux régimes de transport ? C'est-à-dire lorsque la longueur effective et le libre parcours moyen sont du même ordre de grandeur ($L_{eff} \sim \lambda_\mu$). Le régime de transport peut être qualifié de quasi-balistique et modélisé grâce aux théories développées dans [Lundstrom97, Lundstrom02a, Tsuchiya06], qui sont des adaptations directes du modèle du transistor balistique introduit par Natori [Natori94]. Elles supposent que l'injection à la source virtuelle reste le phénomène limitant le transport mais que l'injection perd de l'efficacité car certains porteurs peuvent être rétrodiffusés à la source plutôt que d'être injectés directement dans le canal (cette théorie sera détaillée dans la section II.3.3).
Il est également possible de modéliser le transport grâce à une approche que nous appellerons "quasi dérive-diffusion" qui prend en compte le phénomène de survitesse des porteurs [Chou85, Ge01] (cette théorie sera détaillée dans la section II.3.1). Si les deux approches (quasi-balistique et survitesse) semblent similaires, elles mettent en jeu des mécanismes physiques différents. Ceux-ci doivent être identifiés pour une bonne compréhension du transistor court, dans le but de guider l'optimisation des performances de la technologie. Une méthode d'identification accompagnée de résultats expérimentaux sera détaillée dans le chapitre IV.

II.1.4 Notion de continuité du courant

Le courant $I(x)$ est le produit de la vitesse par la charge, il est continu en tout point du canal pour assurer l'auto-consistance du transport (II.4). Il est donc important de noter que les collisions qui se produisent après la source influent sur la vitesse d'injection à la source de manière à respecter cette continuité dans le transistor. Ce phénomène est couramment observé dans les simulations Monte-Carlo [Saint-Martin04, Lundstrom02a]

$$I(x) = \nu(x) \times Q(x) = \text{cste} \tag{II.4}$$

Dans les prochaines sections, nous approfondirons la compréhension du mécanisme d'injec-

tion à la source et du rôle des collisions dans le canal d'un transistor court. Nous évoquerons l'ambiguïté qui réside dans l'identification de la vitesse limitante dans le transport du fait que ν_{inj} et ν_{sat} ont des valeurs proches dans le silicium (environ 10^7cm/s).

II.1.5 Description du mécanisme d'injection à la source

L'architecture MOS est caractérisée par une barrière de potentiel située entre la source et le drain. Sa hauteur est modulée par la tension de grille V_{gs} et elle régule le passage des porteurs dans le canal : c'est le principe même de fonctionnement du MOSFET décrit au chapitre I. Dans les réservoirs (métalliques), le champ électrique est nul, par conséquent l'énergie thermique est le seul vecteur de mouvement pour les porteurs contenus dans la source et le drain (ils ont une trajectoire erratique et leur vitesse de groupe est nulle). En inversion forte ($V_{gs} > V_{th}$), la barrière de potentiel dans le canal est abaissée et les porteurs ayant une énergie suffisante peuvent sortir des réservoirs et accéder au canal. Puisque leur vitesse résulte de l'énergie thermique, la vitesse maximale à la source est fixée par la vitesse thermique unidirectionnelle [Lundstrom02a], notée ν_{th}. L'expression de ν_{th} est donnée en (II.5), dans l'hypothèse d'une seule sous-bande. Elle est une fonction de la dégénérescence du semiconducteur, de la masse effective $m^\star$ et de la température du réseau T_L [Lundstrom02a, Natori94, Assad00].

$$\nu_{th} = \sqrt{\frac{2k_BT_L}{\pi m_t^\star}} \left\{ \frac{\mathcal{F}_{1/2}(\eta)}{\ln\left(1+e^{\eta}\right)} \right\}, \quad \eta = \frac{E_F - E_c}{k_BT} \tag{II.5}$$

Cette expression peut être étendue au cas général de plusieurs sous-bandes et plusieurs vallées, elle devient alors simplement (II.6), où $n_{i,j}$ représente la densité de porteurs de la i-ème vallée de la j-ème sous-bande.

$$\langle \nu_{th} \rangle = \frac{\displaystyle\sum_{i,j} n_{i,j} \cdot \sqrt{\frac{2k_BT_L}{\pi m_t^\star}} \left\{ \frac{\mathcal{F}_{1/2}(\eta_{i,j})}{\ln\left(1+e^{\eta}\right)} \right\}}{\displaystyle\sum_{i,j} n_{i,j}}, \quad \eta_{i,j} = \frac{E_F - E_{i,j}}{k_BT} \tag{II.6}$$

La relation (II.5) met clairement en évidence l'augmentation de ν_{th} avec la dégénérescence. Cette augmentation est plafonnée par la vitesse de Fermi ν_F qui définit la vitesse maximale des électrons dans un métal (cf. II.4).

$$\lim_{\eta\to\infty}(\nu_{th}) = \left(\frac{4}{3\pi}\right)\nu_F \tag{II.7}$$

Dans un tel cas, la densité d'électrons devient constante et s'approxime par la relation (II.8).

$$n \simeq N_C \frac{4}{3\sqrt{\pi}} \cdot \eta^{3/2} \tag{II.8}$$

En reprenant l'expression de N_C [Sze81], la densité d'électrons peut s'expliciter comme (II.9). Elle permet d'aboutir à une expression simple de ν_F qui met en évidence son indépendance en fonction de la température [Mathieu01] (II.10).

$$n \simeq \frac{8\pi}{3h^3}\left(2m_C\right)^{3/2} \cdot \left(E_F - E_C\right)^{3/2} \tag{II.9}$$

$$\nu_F = \frac{\hbar}{m}\left(3\pi^2 n\right)^{1/3} \tag{II.10}$$

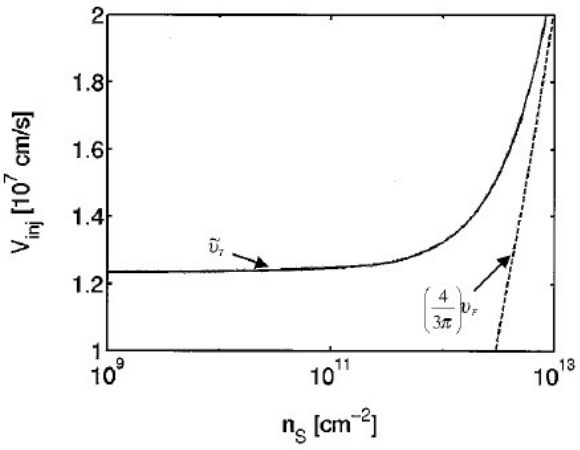


Fig. II.4 – Évolution de la vitesse thermique en fonction de la dégénérescence. Pour un semiconducteur très dégénéré, ν_{th} tend vers la vitesse de Fermi qui ne dépend plus de la température, d'après [Lundstrom02a]

II.1.5.a La connaissance de la charge d'inversion à la source : une difficulté pour tous

La charge d'inversion à la source virtuelle (notée Q_s) définit la densité de porteurs injectée dans le canal. La technique de mesure capacitive *split*-CV [Koomen73, Romanjek04b] permet d'extraire la capacité effective en inversion C_{eff} ainsi que charge d'inversion moyenne du canal en l'absence de champ électrique $\mathcal{E}_y = 0$ (II.11).

$$Q_{inv}(V_{gs}) = \int_{-\infty}^{V_{gs}} C_{gc}(u) \cdot du \tag{II.11}$$

Si la technique *split*-CV donne une valeur précise de $Q_{inv} \simeq Q_s$ en régime linéaire (pour des V_{ds} faibles), elle ne permet pas de mesurer la charge en régime de saturation. La relation (II.12) permet d'approximer la charge Q_s en régime de forte inversion, où C_{eff} représente la capacité effective en inversion et $V_{th}\,(V_{ds})$ la tension de seuil dans laquelle les effets de canal court (DIBL) sont pris en compte [Natori94]. Elle fournit des valeurs équivalentes à la méthode de Lochtefeld [Lochtefeld01] lorsque l'influence de la fuite de grille peut être négligée sur la réponse capacitive du transistor long. Finalement, l'expression (II.12) est utilisée dans de nombreuses études expérimentales [Chen02, Barral07] et théoriques [Rahman03, Natori94].

$$Q_{inv} \simeq C_{eff} \cdot (V_{gs} - V_{th}\,(V_{ds})) \tag{II.12}$$

La valeur de Q_s issue de (II.12) semble ne pas être égale à la valeur extraite depuis les simulations Monte-Carlo [Zilli07b]. Néanmoins, la position de la source virtuelle semble être difficile à définir précisément, répercutant une incertitude sur Q_s. Il est donc difficile de juger la précision de la relation (II.12) en utilisant les profils de densité de charge d'inversion simulés par cette technique.

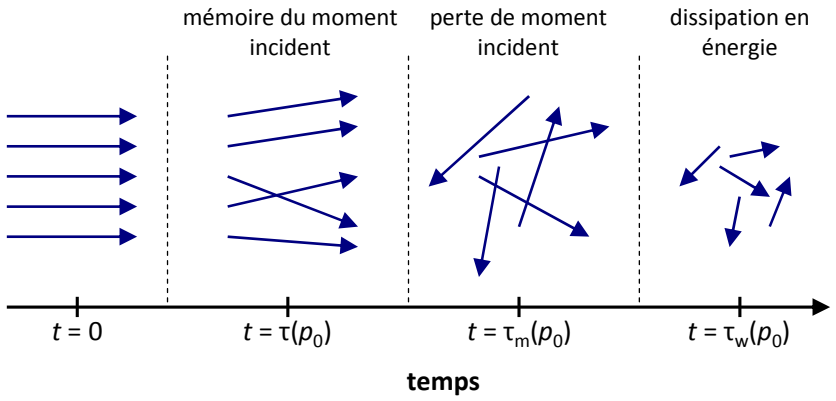


Fig. II.5 – Illustration de la manière dont les collisions élastiques et inélastiques affectent les trajectoires et les énergies d'un groupe d'électrons injecté à $t = 0$ avec un moment incident, d'après [Lundstrom00].

II.2 Le transport stationnaire

On parle de transport stationnaire lorsque le système est à l'équilibre et peut être régi par des équations indépendantes du temps. A l'échelle du transport la stationnarité est observée lorsque le temps de transit des porteurs dans le canal (τ_{ds}) est grand par rapport au temps moyen de relaxation en énergie (τ_w). Dans ce cas, les mécanismes de relaxation en énergie sont fréquents et les porteurs sont dit "thermalisés" avec le réseau de silicium : c'est l'équilibre.
Les équations du transport dérive-diffusion sont basées sur des calculs à l'équilibre et présupposent que les porteurs subissent statistiquement un grand nombre de collisions avec les centres de diffusion dans le canal avant de rejoindre le drain. Ces collisions peuvent faire intervenir un transfert d'énergie, elles sont alors qualifiées d'inélastiques et contribuent à la thermalisation des porteurs avec le réseau de silicium. La constante de temps τ_w quantifie alors le temps moyen pour qu'un porteur retrouve l'équilibre thermique avec le réseau [Seeger91]. A l'opposé, les collisions qui ne font pas intervenir un transfert d'énergie sont dite élastiques car elles modifient uniquement le moment des porteurs (l'orientation de leur trajectoire p) sans affecter leur énergie. Le temps de relaxation en moment (τ_m) représente alors le temps moyen nécessaire pour rendre le moment d'un groupe de porteurs aléatoire, c'est à dire pour perdre la directionalité du transport [Lundstrom00]. Finalement, nous noterons que τ_w est généralement beaucoup plus long que τ_m car si plusieurs collisions inélastiques sont généralement nécessaires pour retrouver l'équilibre, une seul collision inélastique permet de perdre la directionalité du transport [Lundstrom06b]. La figure II.5 illustre la manière dont les collisions élastiques et inélastiques affectent les trajectoires et les énergies d'un groupe d'électrons en transit sous l'influence d'un champ électrique.

La mobilité à bas champ (V_{ds} faible) permet alors de rendre compte de la vitesse de dérive moyenne acquise par les porteurs qui subissent des collisions élastiques à une occurrence $1/\tau_m$, il est alors possible de préciser les termes dans la relation (II.2) : cf. (II.13).

$$\nu_d = \mu_0 \cdot \mathcal{E}_y, \quad \mu_0 = \frac{e \cdot \tau_m}{m^\star} \tag{II.13}$$

Dans cette relation, l'influence des collisions inélastiques est négligée et il est supposé que τ_m est indépendant du champ électrique ($\mathcal{E}_y$) [Grove67]. Ces hypothèses sont raisonnables à bas champ, tant que la vitesse de dérive des porteurs reste faible devant la vitesse thermique (environ 10^7cm/s). Néanmoins la relation (II.13) perd sa validité pour des forts champs électriques V_{ds}

pour lesquelles les collisions inélastiques provoquent une déviation à la loi d'Ohm. Ce point sera expliqué en détail au §II.2.3.

II.2.1 Notion de mobilité effective et de champ effectif

Nous venons de voir que la mobilité est un paramètre critique puisqu'il régit le transport dans le transistor MOS. Il s'agit donc de savoir comprendre, analyser et comparer ce paramètre entre différentes options technologiques pour comprendre les mécanismes physiques qui le gouvernent.

II.2.1.a La notion de mobilité effective μ_{eff}

Si l'on considère que la vitesse de dérive des porteurs résulte de plusieurs mécanismes de collision, il est possible d'étendre la relation (II.13) en introduisant le paramètre de mobilité effective (μ_{eff}) qui rend compte de la mobilité moyenne vue par la couche d'inversion.

$$\nu_d = \mu_{eff} \cdot \mathcal{E}_{\|} \tag{II.14}$$

Dans le cas général, si tous les mécanismes de collision sont considérés comme indépendants les uns aux autres, μ_{eff} peut s'écrire loi de Matthiessen pondérée (II.15) [Fischetti02]. Dans cette expression, $\mu_{i,j,s}$ réfère à la mobilité résultant mécanisme s sur les $n_{i,j}$ porteurs de la j-ème vallée de la i-ème sous-bande.

$$\frac{1}{\mu_{eff}} = \frac{\sum_s \sum_{i,j} \left(n_{i,j} \cdot \frac{1}{\mu_{i,j,s}} \right)}{\sum_{i,j} n_{i,j}} \tag{II.15}$$

Dans le cas du premier niveau d'énergie et d'une seule vallée, (II.15) se simplifie en la relation (II.16)

$$\frac{1}{\mu_{eff}} = \sum_s \frac{1}{\mu_s} \tag{II.16}$$

Dans certains cas (basse température, matériau InGaAs, contrainte mécanique), la loi de Matthiessen n'est plus valide. Elle retourne alors une valeur de la mobilité effective surestimée par rapport à la valeur réelle qui caractérise la couche d'inversion [Jeon89, Fischetti02, Takeda81]. Ce phénomène traduit les limites de l'hypothèse d'indépendance des mécanismes de collision que suppose la loi de Matthiessen.

II.2.1.b La notion de champ effectif $\mathcal{E}_{eff}$

La mobilité dépend du champ transverse vu par les porteurs $\mathcal{E}_{eff}$ et n'est donc pas constante en fonction de la profondeur (x) choisie pour effectuer le calcul électrostatique. Dans la couche d'inversion, le gaz d'électrons est étalé et respecte une distribution dans un puits quasi-triangulaire qui offre un confinement dans la direction Ox. En considérant une seule direction de confinement, la densité de charges d'inversion peut s'écrire en fonction de x : $n_{inv}(x)$. Le champ électrique en tout point de la couche d'inversion $\mathcal{E}_x(x), x \in [0, x_{inv}]$) est donné par le théorème de Gauss et peut s'exprimer par (II.17), où Q_{dep} représente la densité surfacique de charges de désertion.

$$\mathcal{E}_x(x) = \frac{Q_{dep}}{\varepsilon_{Si}} + \frac{q}{\varepsilon_{Si}} \cdot \int_x^{x_{inv}} n_{inv}(u) \cdot du \tag{II.17}$$

Dans ce calcul, la contribution des charges de désertion est considérée constante dans la profondeur de la couche d'inversion ($0 < x < x_{inv}$). Cela revient à négliger la contribution des charges de désertion contenues dans l'intervalle $x \in [0, x_{inv}]$, hypothèse justifiée tant que la couche de désertion s'étale beaucoup plus en profondeur que la couche d'inversion dans le silicium. Dans ce cas les charges se situant au delà de x_{inv} sont responsables de la contribution principale au champ.
Le champ effectif $\mathcal{E}_{eff}$ est défini comme la moyenne du champ électrique transverse prise sur la distribution de la couche d'inversion (II.18) [Ando82, Sabnis79], où η est un paramètre qui rend compte de l'étalement de la couche d'inversion dans le silicium qui peut être calculé à partir de la relation (II.19), et Q_{inv} la densité surfacique de charges d'inversion (II.20).

$$\mathcal{E}_{eff} = \langle \mathcal{E}_x \rangle = \frac{Q_{dep} + \eta \cdot Q_{inv}}{\varepsilon_{Si}} \tag{II.18}$$

Le paramètre η peut être considéré comme une sorte de barycentre électrostatique propre à la distribution de la couche d'inversion.

$$\eta = \frac{1}{Q_{inv} \cdot x_{inv}} \cdot \int_{x=0}^{x_{inv}} \left(\int_{u=x}^{x_{inv}} n_{inv}(u) \cdot du \right) \cdot dx \tag{II.19}$$

$$Q_{inv} = \int_{0}^{x_{inv}} n_{inv}(u) \cdot du \tag{II.20}$$

II.2.1.c La valeur de η : un choix crucial

Si l'égalité $\eta = 1$ est injectée dans (II.17), le champ effectif correspond au champ calculé en surface du canal (interface oxyde-canal) : $\eta = 1 \Rightarrow \mathcal{E}_{eff} = \mathcal{E}_x(0)$. Cette valeur n'est pas représentative du champ moyen vu par les porteurs de la couche d'inversion car elle surestime la contribution de Q_{inv} dans le calcul du champ. La conséquence de l'utilisation de $\mathcal{E}_x(0)$ plutôt que de $\mathcal{E}_{eff}$ est l'observation d'une variation *apparente* de la mobilité effective (prise à champ constant) lorsque la polarisation V_{bs} change (équivalent à une variation du dopage). Ce résultat est incohérent car la mobilité en forte inversion (toujours prise à champ constant) doit dépendre uniquement de la forme du potentiel en surface et non du dopage [Sabnis79]. La prise en compte de la relation (II.19) permet donc d'inclure cet effet de V_{bs} "parasite" dans le modèle de mobilité en pondérant la contribution de Q_{inv} de manière à calculer un champ proche de la valeur moyenne vue par les porteurs de la couche d'inversion. La valeur de η peut être estimée empiriquement en effectuant des mesures de courbe $\mu_{eff}(\mathcal{E}_{eff})$ à plusieurs tensions V_{bs} ou plusieurs dopages de canal [Sabnis79, Takagi94a]. Aucune variation ne doit être observée entre les courbes lorsque la bonne valeur de η est trouvée.

Non-universalité du paramètre η — Historiquement, la valeur de η a été déterminée (et validée) à environ 1/2 pour les électrons [Sabnis79, Sun80, Takagi94a] et 1/3 pour les trous à 300K [Takagi94a, Arora87]. Ces valeurs ne sont pourtant pas universelles, elles doivent être mesurées pour une technologie donnée avant d'analyser le comportement de $\mu_{eff}(\mathcal{E}_{eff})$ (par exemple avec la technique décrite dans [Bennamane08, Bennamane09]). Certaines publications montrent en effet que η peut varier en fonction de la température, de l'épaisseur de l'oxyde, de la contrainte mécanique ou encore de la tension de grille [Emrani94, Arora87, Bennamane08]. Cela justifie de la précaution nécessaire à prendre avant d'utilisation des valeurs de η *historiques* parfois injustifiées dans la relation (II.18).

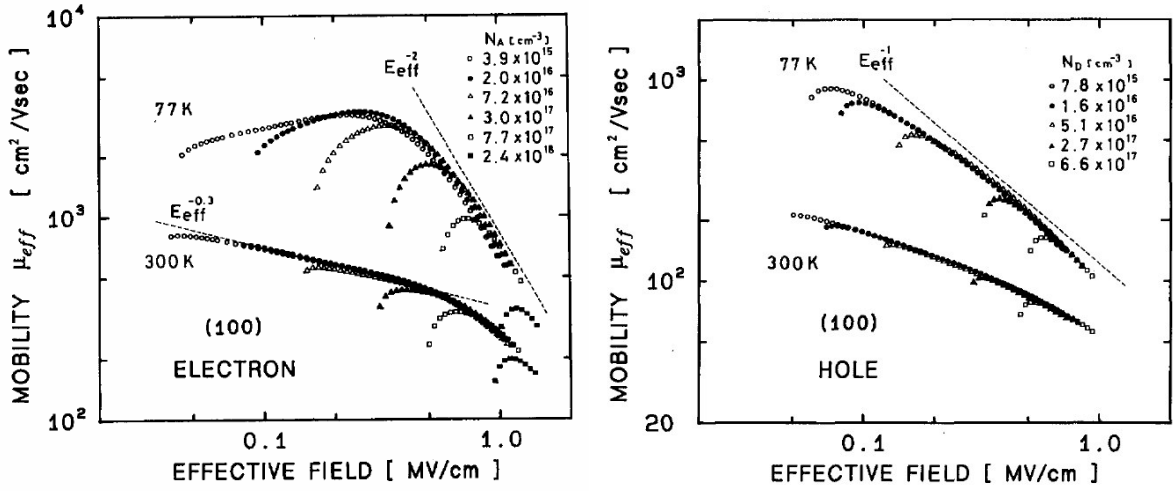


Fig. II.6 – Variation de la mobilité effective en fonction du champ effectif pour différentes valeurs de dopage pour les électrons (à gauche) et les trous (à droite). La mobilité des deux types de porteurs suit une loi universelle (indépendante du dopage) à fort champ électrique [Takagi94a] lorsque η est ajusté correctement.

II.2.1.d La mobilité universelle

Lorsque le paramètre η est choisi de manière à ce que sa valeur respecte la relation (II.19), la mobilité à champ fort suit une loi universelle en fonction de $\mathcal{E}_{eff}$ [Takagi94a] qui ne dépend pas du dopage (ou de V_{bs}), cf. II.6. La déviation à la tendance universelle observée à faible champ sur la figure II.6 provient de la réduction de mobilité causée par des collisions des porteurs avec les impuretés chargées du canal (le dopage) qui se voient principalement à faible charges d'inversion.

Les principaux mécanismes de collision (cf. II.7) dans le transistor MOS peuvent être classés en trois catégories dont la description fera l'objet des prochains paragraphes :

- les collisions avec les impuretés chargées (dopant ionisés, charges d'interface et/ou dans l'oxyde),
- les collisions avec les vibrations du réseau (phonons),
- les collisions dues à la rugosité de surface.

Si le concept de mobilité universelle parait attrayant nous resterons prudent quant à l'utilisation de ce terme. En effet, les procédés technologiques utilisés pour élaborer les transistors courts ($L_{eff} \leq$ 100nm) ouvrent la voie à de nouveaux mécanismes (contraintes mécaniques, défauts dans les empilements de grilles avancés) qui ne permettent plus de respecter les lois universelles proposées dans [Takagi94a, Cheng96].

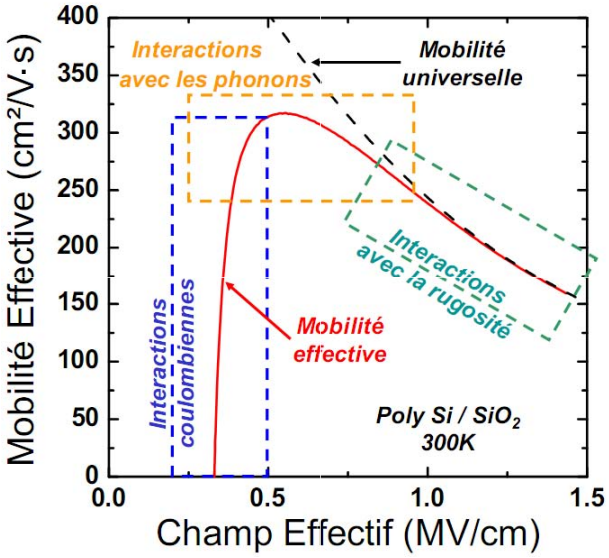


Fig. II.7 – Principaux mécanismes de réduction de la mobilité dans les transistors MOS avec un empilement de grille de type Si/SiO_2.

II.2.2 Interactions avec les centres de diffusion chargés

Le dopage de canal est constitué d'un ensemble d'atomes chargés, en position substitutionnelle dans le réseau de silicium. Ces particules fixes génèrent un potentiel électrique susceptible de perturber la trajectoire d'un électron dans le canal : c'est l'interaction de Coulomb. Le traitement de cette interaction dans le cas du transistor MOS ne peut pas se résumer à un problème à deux corps, il faut prendre en compte l'influence des toutes les charges mobiles de la couche d'inversion qui écrantent l'interaction entre un porteur ciblé et son centre de diffusion chargé [Stern67, Masetti83], comme représenté sur la figure II.8.

La longueur de Debye (λ_D) permet de donner un rayon d'action approximatif au delà duquel le potentiel perturbateur de l'impureté ionisée est écranté par le nuage électronique environnant. Elle est définie par la relation (II.21) et vaut environ 10nm pour les niveaux de dopage usuellement rencontrés dans les transistors MOS sur silicium massif ($N_I \sim 10^{17}$–10^{18}cm^{-3}).

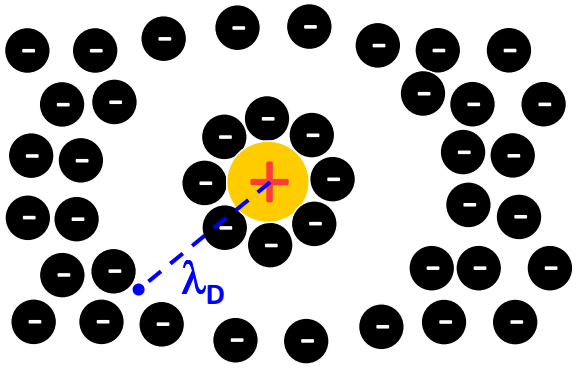


Fig. II.8 – Illustration du phénomène d'écrantage dune charge positive fixe par des porteurs chargés négativement.

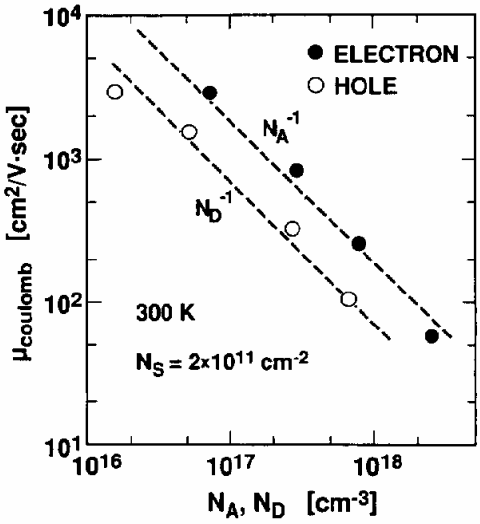


Fig. II.9 – Évolution de la contribution de la mobilité limitée par les interactions avec les impuretés du canal en fonction du dopage en surface [Takagi94a].

$$\lambda_D = \sqrt{\frac{\varepsilon_{Si} \cdot k_B \cdot T}{q^2 \cdot N_D}} \sim 10\text{nm} \tag{II.21}$$

Un modèle analytique de mobilité associée à cette interaction μ_I (I pour *Impurity*) est donné par [Sze81] en fonction de la masse effective $m^\star$, du dopage N_I et de la température T (II.22). Elle donne en particulier les tendances de variation de μ_{cb} en fonction de ces trois paramètres qui peuvent être simplifiés par (II.23), en accord avec [Takagi94a] et comme l'illustre la figure (II.9).

$$\mu_{cb} = \frac{64\sqrt{\pi}\varepsilon_{Si}^2 (2k_BT)^{3/2}}{N_I q^3 m^{\star 1/2}} \cdot \left\{ \ln\left[1 + \left(\frac{12\pi\varepsilon_{Si}k_BT}{q^2 {N_I}^{1/3}}\right)^2\right]\right\}^{-1} \tag{II.22}$$

$$\mu_{cb} \propto (m^\star)^{-1/2} \cdot {N_I}^{-1} \cdot T^{3/2} \tag{II.23}$$

II.2.2.a Prise en compte des défauts d'interface et de l'écrantage dans l'expression de μ_{cb}

Comme le montre la relation (II.22), les interactions avec les impuretés ionisées du canal est un mécanisme qui se manifeste principalement à basse température et pour de niveaux de dopage élevés [Jeon89]. Néanmoins, cette relation ne prend pas en compte la diffusion par les défauts d'interface chargés, de densité surfacique Q_{ss} et qui jouent un rôle similaire dans l'expression de μ_{cb} [Sun80]. L'écrantage de l'interaction par le nuage électronique de la couche d'inversion doit également être pris en compte pour une bonne modélisation sur toute la gamme de $\mathcal{E}_{eff}$. La relation (II.22) peut alors être améliorée pour prendre en compte les variations de μ_{cb} engendrées ces deux mécanismes : (II.24) [Cheng97a, Sze81].

$$\mu_{cb} \propto (m^\star)^{-1/2} \cdot \frac{Q_{inv}}{N_I + Q_{ss}/x_{inv}} \cdot T^{3/2} \tag{II.24}$$

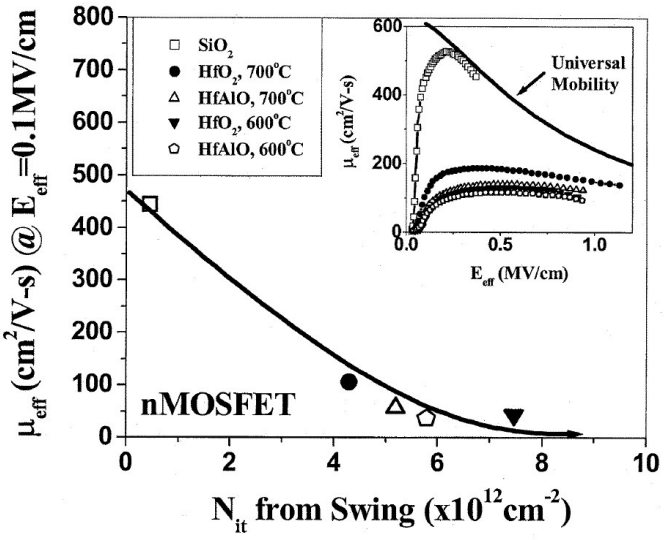


Fig. II.10 – Mobilité expérimentale pour des dispositifs possédant des empilements de grille avancés et standard, d'après [Zhu04].

Cette relation montre clairement que le mécanisme de collisions avec les défauts chargés est influent à faible charges d'inversion, lorsque l'écrantage est peu efficace.

II.2.2.b L'interaction de Coulomb par des centres de diffusion distants : le *remote Coulomb scattering*

L'interaction de Coulomb est suffisamment forte pour que des centres chargés situés à quelques dizaines d'angströms de la couche d'inversion puissent avoir une influence sur le transport électronique dans le canal. La réduction des dimensions, et plus précisément l'amincissement de l'oxyde de grille parallèlement à l'introduction de diélectriques à haute permittivité (high-κ) et d'une grille métallique sont ainsi responsables d'un nouveau phénomène : le *remote Coulomb scattering*. Ce terme anglophone désigne l'interaction additionnelle provenant de centres de diffusion chargés distants, qui proviennent de la forte densité d'états d'interface caractéristique des empilements de grille high-κ, mais également de la grille dans le cas d'un oxyde SiO_2 [Casse06]. La réduction de mobilité correspondante à été mise en évidence expérimentalement en comparant la mobilité de dispositifs ayant un tel empilement de grille avec des transistors pour lesquels l'empilement est *standard*, c'est-à-dire composé d'un dioxyde de Silicium (SiO_2) et d'une grille en silicium polycristallin (cf. Fig.II.10) [Zhu04]. L'origine de cette réduction de mobilité semble provenir d'une densité de charges élevée à l'interface HfO_2/SiO_2, pour les transistors dont la grille est à base de nitrure de titane TiN [Casse06, Weber06].

II.2.3 Interactions avec les vibrations du réseau : les phonons

Les atomes de silicium formant le réseau ne sont pas strictement fixes. Ils oscillent autour d'une position d'équilibre définie par le minimum de puits de potentiel dans lequel ils se trouvent avec une amplitude d'autant plus grande que la température du réseau T_L est élevée. Les atomes vibrent en groupe selon des modes (longitudinaux et transverses) possédant chacun une pulsation propre ω_0. Ces vibrations sont assimilées à des quasi-particules susceptibles d'interagir avec

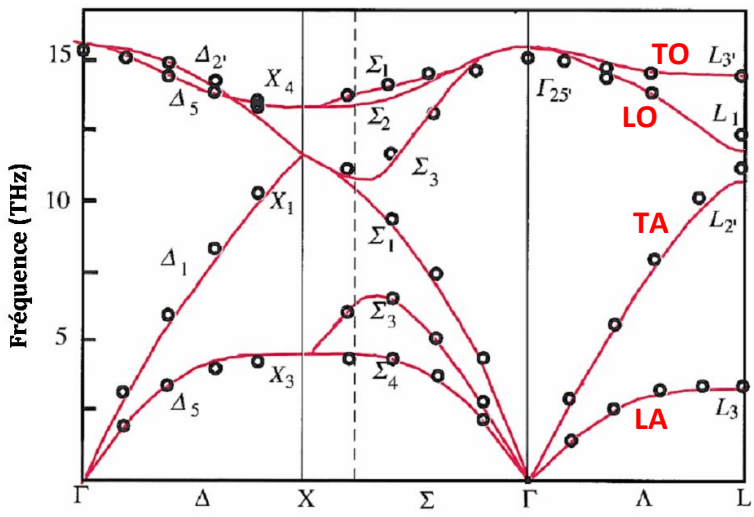


Fig. II.11 – Relation de dispersion des phonons dans le silicium le long des axes de symétrie, d'après [Yu05]

les électrons de la couche d'inversion : un phonon peut être absorbé ou émis par un porteur qui échange de ce fait un quantum d'énergie de valeur $\Delta E = \hbar \cdot \omega_0$ et un moment Δp. Ces collisions tendent à limiter la mobilité dans le canal et, puisqu'elles résultent de phénomènes physiques intrinsèques à la structure cristalline du silicium, ne peuvent pas être supprimées ou atténuées simplement. L'influence des phonons sur la mobilité est un des principaux facteurs limitant le transport dans les transistors MOS à température ambiante, le mécanisme d'interaction porteurs-phonons doit donc être compris et modélisé justement. Ceci fait l'objet des paragraphes suivants.

II.2.3.a Les différents types de phonons

Dans un cristal à trois dimensions, il est nécessaire de distinguer les vibrations s'effectuant dans le sens du transport électronique (appelées longitudinales et dénotées par un L), des 2 modes transverses polarisés à 90 ° l'un de l'autre qui caractérisent les vibrations dans le sens perpendiculaire au transport. D'une manière générale une maille élémentaire contenant p atomes engendre $3p-3$ branches optiques et 3 branches acoustiques. Dans le cas du silicium la symétrie impose aux modes transverses d'être dégénérés, la maille primitive contient 2 atomes (cf. Annexe B) et la relation de dispersion contient 4 branches (cf. Fig.II.11) :

- longitudinale optique (LO),
- transverse optique (TO),
- longitudinale acoustique (LA),
- transverse acoustique (TA).

Les modes optiques sont par définition caractérisés par des fréquences d'oscillation plus élevées (et par conséquent une énergie propre plus importante) que dans les modes acoustiques. Le mécanisme d'interaction avec les phonons optiques va donc se faire pour des énergies plus élevées (i.e. des champs plus élevés), qui permettront aux porteurs de pouvoir relaxer leur excédant d'énergie. Précisons également que les branches LO et TO sont dégénérées au centre de la zone de Briouillin.

Le processus de limitation de la mobilité par les phonons A l'équilibre thermique, les porteurs émettent et absorbent un nombre identiques de phonons, si bien que le flux énergétique moyen est nul. Au contraire, en présence d'un champ électrique les porteurs acquièrent de l'énergie cinétique qu'ils relaxent en émettant plus de phonons qu'ils n'en absorbent : l'équilibre est rompu.

II.2.3.b Diffusion par les phonons acoustiques

Les énergies faibles associées aux premiers modes de vibration justifient du traitement quasi-élastique qui peut être employé pour traiter le mécanisme de collision avec les phonons acoustiques. Si l'échange d'énergie est faible les probabilités d'émission et d'absorption restent élevées, et à chaque interaction le moment de la particule est modifié. Au final, la qualité du transport dans le canal est détériorée. Selon [Sze81], la mobilité résultant des interactions des porteurs avec les phonons acoustiques peut être modélisée par la relation (II.25). Le nombre de phonons (i.e. la dégradation du transport) augmente avec la température et résulte en une dépendance théorique de la mobilité résultant des interractions avec les phonons en fonction de la température qui respecte $\mu_{ph} \propto T^{-3/2}$. Dans cette dernière expression, C_{11} est la constante élastique moyenne du silicium.

$$\mu_{ph} = \frac{\sqrt{8\pi} q \hbar C_{11}}{3 E_{ds} m^{\star 5/2} (k_B T)^{3/2}} \tag{II.25}$$

II.2.3.c Rôle des phonons optiques

A fort champ électrique longitudinal, l'énergie cinétique des porteurs devient comparable à l'énergie propre des phonons optiques ($E_{opt} = \hbar \cdot \omega_{LO} \simeq 61$meV [Sze81, Crowel66, Ning77]). La probabilité d'émission de phonons optiques par les porteurs est donc élevée et, contrairement au cas des phonons acoustiques traité au II.2.3.b, la collision fait intervenir un quantum d'énergie qui ne peut plus être négligé devant l'énergie cinétique de la particule. Le caractère inélastique des ces collisions doit donc être pris en compte [Seeger91], ce type de collision est en effet le principal mécanisme de relaxation en énergie des porteurs à fort champ (l'influence des phonons acoustiques peut être négligée [Wolf71, Jacoboni77, Sze81]).

Dans le cas des vibrations longitudinales du réseau (phonons optiques, branche LO), les plans atomiques se déplacent à l'encontre l'un de l'autre, tel que l'illustre la figure II.12. Lorsqu'un électron se déplace d'un atome au suivant, l'oscillation caractéristique de la vibration peut modifier son énergie par une quantité proportionnelle au déplacement atomique a. Le temps de relaxation du moment est donné par la relation (II.26) [Seeger91, Kiréev75] calculé (comme dans [Moll64]) en résolvant l'équation des moments à l'équilibre. Dans (II.26), N_q représente la dépendance en température du mécanisme.

$$\frac{1}{\tau_m} = \frac{2 m^{*3/2} D^2 N_q t^{3/2}}{3\pi^{3/2} \rho \hbar^2 (k_B T_D)^{1/2}} \cdot \left[e^t \left(K_2(t) - K_1(t) \right) + e^{\frac{T_D}{T} - t} \left(K_2(t) + K_1(t) \right) \right], \quad t = \frac{T_D}{2T_e} \tag{II.26}$$

$$N_q = \frac{1}{\exp\left(\dfrac{\hbar \omega_0}{k_B T} \right) - 1} \tag{II.27}$$

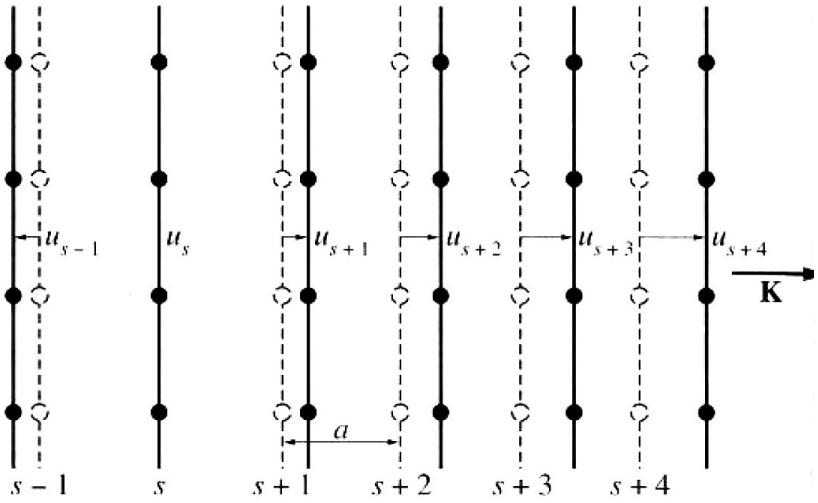


Fig. II.12 – Illustration du déplacement des plans atomiques caractéristique du mode de vibration longitudinal, d'après [Kittel98].

Le temps de relaxation du moment décroit lorsque l'énergie augmente, il traduit donc une probabilité d'interaction et donc de relaxation d'énergie qui augmente à mesure que la vitesse de la particule (i.e. son énergie cinétique) augmente. Ce phénomène résulte en une déviation de la loi d'Ohm ($v_d < \mu \cdot \mathcal{E}_y$) qui provoque une saturation de la vitesse de dérive des porteurs : c'est la notion de vitesse de saturation.

Notion de vitesse de saturation Dans le silicium et pour des champs électriques longitudinaux très grands (typiquement $\mathcal{E}_y \gg 10^5$ V/cm), la vitesse de dérive n'augmente plus car la relaxation en énergie avec les phonons optiques limite le gain en énergie cinétique des porteurs : c'est la vitesse de saturation ν_{sat}. Au premier ordre, la valeur de ν_{sat} peut être estimée comme étant la vitesse limite pour laquelle le porteur acquiert une énergie cinétique du même ordre de grandeur que l'énergie propre des phonons LO ($\hbar \cdot \omega_{LO}$). Une expression simplifiée est donnée dans [Shockley51, Jacoboni77] (II.28).

$$\nu_{sat} \sim \sqrt{\frac{8}{3\pi} \cdot \frac{\hbar\omega_{LO}}{m^\star}} \sim 10^7 \text{cm/s} \tag{II.28}$$

Une expression plus complète issue de la résolution des équations à l'équilibre est également donnée dans [Seeger91] (II.29). Elle fournit une précision supplémentaire sur la dépendance en température de ν_{sat}, représentée sur la figure II.13.

$$\nu_{sat} \simeq \sqrt{\frac{3\hbar\omega_{LO}}{4m^*} \cdot \tanh\left(\frac{\hbar\omega_{LO}}{2k_B T}\right)} \tag{II.29}$$

Nous noterons toutefois que les expressions (II.28) et (II.29) fournissent uniquement une valeur approchée de la vitesse de saturation. En effet ν_{sat} résulte en réalité de l'interaction avec plusieurs branches de phonons et nécessite la prise en compte du phénomène d'absorption.

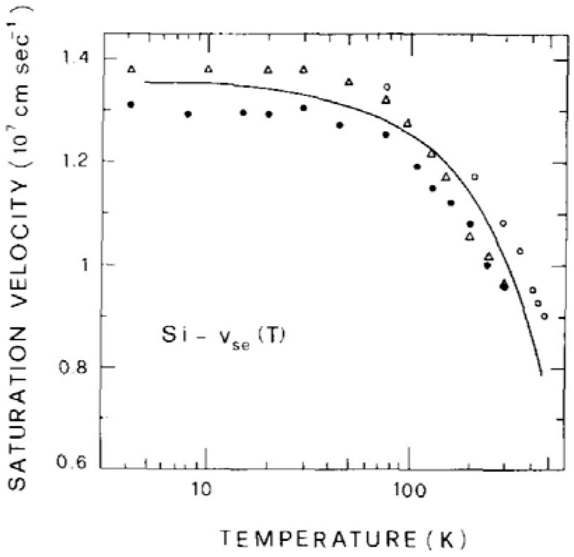


Fig. II.13 – Évolution de la vitesse de saturation en fonction de la température dans le silicium, d'après [Jacoboni77].

Auteur(s)	λ_{opt} électrons (Å)	λ_{opt} trous (Å)
Ning *et al.* [Ning77]	90.5	-
Marsland [Marsland91]	71.2	65
Ahmad et Arora [Ahmad86]	70	-
Slotboom *et al.* [Slotboom87]	89	-
Tam *et al.* [Tam84]	92.5	-
Crowel et Sze [Crowel66]	64	55
Sze [Sze81]	76	(55)
Henning *et al.* [Henning87]	87.9	43.1
Moyenne	**80.1**	**54.4**

Tab. II.1 – Valeurs de λ_{opt} issues de la littérature

Le libre parcours moyen dans l'interaction avec les phonons optiques Le libre parcours moyen des électrons et des trous dans l'interaction avec les phonons optiques λ_{opt} a été estimé dans de nombreuses publications. Les valeurs expérimentales sont reportées dans le tableau II.2.3.c. En moyenne les électrons et les trous ont un libre parcours moyen de 80Å et 54.5Å respectivement.

Certains travaux montrent que la valeur du libre parcours moyen en surface est beaucoup plus faible que la valeur dans la profondeur (que nous appellerons λ_{opt}^{bulk}). En particulier, [Slotboom87] fournit une expression analytique à cette réduction de libre parcours moyen (cf. (II.30)) dans laquelle $\Delta\lambda = 50$Å et $x_0 = 50nm$ sont des grandeurs empiriques, et x_{inv} représente la distance qui sépare la couche d'inversion de l'interface oxyde-silicium.

$$\lambda_{opt}(z) = \lambda_{opt}^{bulk} - \Delta\lambda \cdot \exp\left(-\frac{x_{inv}}{x_0}\right) \tag{II.30}$$

De manière générale, il a été démontré que la valeur effective du libre parcours moyen dans le MOSFET est donnée par : $\lambda_{opt}^{eff} \simeq \alpha \cdot \lambda_{opt}^{bulk}$, où $0.54 < \alpha < 0.64$ d'après [Slotboom87, Thurgate85,

profondeur	L_{eff} (nm)	$\mathcal{P}_{bal}$ électrons	$\mathcal{P}_{bal}$ trous
bulk	20	8.2%	2.5%
surface [Slotboom87]	20	2.0%	0.3%
bulk	10	28%	16%
surface [Slotboom87]	10	**14%**	**5.7%**

Tab. II.2 – Probabilité de passage balistique d'un porteur dans un transistor MOS, en considérant que seule l'interaction avec les phonons optiques limite le transport à fort champ latéral.

Henning87].

La probabilité qu'un électron puisse parcourir une distance d sans subir aucune interaction est donnée par W. Shockley [Shockley51], en fonction du libre parcours moyen. Elle permet de calculer une probabilité de passage balistique $\mathcal{P}_{bal}$ à fort champ latéral, sous l'hypothèse que seul le mécanisme d'interaction avec les phonons optiques est limitant. Les valeurs ainsi calculées en profondeur et en surface [Slotboom87] sont reportées dans le tableau II.2.3.c pour des longueurs de canal de 10nm et 20nm.

$$\mathcal{P}_{bal} = \exp\left(-\frac{d}{\lambda_{opt}}\right) \tag{II.31}$$

Les probabilités semblent trop faibles pour expliquer les taux de balisticité reportés dans la littérature. Ce phénomène soulève un problème dans la compréhension à l'état de l'art des mécanismes de transport à fort champ dans les transistors courts. Il est effet possible que les taux de balisticité reportés dans la littérature soient surestimés du fait d'avoir négligé l'influence des interactions inélastiques, toujours fortement présentes jusqu'à $L_{eff} = 10$nm.

Évolution de λ_{opt} avec la température La dépendance en température de λ_{opt} à une température T est estimée grâce à la relation (II.32) issue de [Marsland91], où $n(T)$ est le nombre de Bose-Einstein (cf. [Kittel98]). Cette expression est analytiquement identique à celle trouvée dans [Crowel66, Sze81] et montre un bon accord avec des mesures expérimentales effectuées entre 77K et 360K [Ning77].

$$\lambda_{op}(T) = \frac{2n(300K)+1}{2n(T)+1} \cdot \lambda_{opt}(300K), \quad n(T) = \frac{1}{\exp\left(\frac{E_{opt}}{kT}\right) - 1} \tag{II.32}$$

II.2.4 L'influence de la rugosité de surface sur la mobilité

A fort champ électrique transverse $\mathcal{E}_x$, les porteurs de la couche d'inversion sont confinés à proximité de l'interface oxyde/canal. La rugosité (les fluctuations géométriques) de cette interface influe sur le nombre de collisions que subissent les porteurs et dégrade la mobilité. La rugosité de surface peut être modélisée par deux paramètres :

- la longueur de corrélation des fluctuations géométriques Λ
- la rugosité r.m.s. de l'interface Δ

Ces deux paramètres interviennent dans la densité spectrale $S(q)$ de la rugosité. Elle rend compte de la morphologie de l'interface et se calcule à partir de la transformée de Fourier de la fonction d'auto-covariance de la rugosité. Historiquement elle fut déduite d'une fonction d'auto covariance de type Gaussienne (II.33) [Prange68], puis exponentielle (II.34) [Goodnick85]. En pratique les deux formes de $S(q)$ permettent de rendre compte de variations de mobilité

expérimentales [Pirovano00] en ajustant le couple de paramètres (Λ, Δ). Il parait donc impossible d'utiliser des mesures de mobilité pour en déduire la morphologie de l'interface sans connaitre au préalable la fonction d'auto-covariance.

$$S(q) = \pi\Lambda^2\Delta^2 \cdot \exp\left\{-\frac{(\Lambda q)^2}{4}\right\} \tag{II.33}$$

$$S(q) = \frac{\pi\Lambda^2\Delta^2}{1 + \left(\frac{\Lambda^2 q^2}{2}\right)^{3/2}} \tag{II.34}$$

II.2.5 Les mécanismes additionnels de réduction de la mobilité

Nous venons de décrire les trois mécanismes historiques qui permettent de comprendre le comportement de la mobilité sur des transistors MOS longs dont l'empilement de grille est composé de silicium polycristallin et d'un oxyde de type SiO_2. Le comportement de la mobilité en fonction du champ effectif est alors illustré schématiquement par la figure II.14 issue de [Takagi94a].

II.2.5.a Dégradation de mobilité provoquée par l'empilement de grille *high*-κ-métal

L'introduction récente de nouveaux matériaux dans l'empilement de grille (oxydes high-κ, métal de grille) ainsi que la réduction de la longueur effective vient corrompre ce modèle simple. En effet des composantes additionnelles liées à de nouveaux mécanismes physiques de réduction de la mobilité sont à prendre en compte pour justifier de la dégradation supplémentaire de mobilité observée sur ces dispositifs [Cros06b, Zhu04]. La figure II.15 présente en effet les différents mécanismes de diffusion susceptibles de dégrader la mobilité de dispositifs à empilement de grille HK/MG de faible longueur de grille. Nous noterons en particulier :

- le couplage entre les porteurs de la couche d'inversion et les phonons optiques de faibles énergies, propres à la polarisation des matériaux high-κ [Fischetti01, Ren03, Chau04],
- l'interaction de coulomb que subissent les porteurs avec des charges à l'interface SiO_2/high-κ ou dans la couche high-κ [Casse06, Saito03, Villa98]
- la modification de la rugosité de surface en fonction du procédé utilisé pour déposer le métal de grille (TiN) [Thevenod05]

II.2.5.b Couplage entre les électrons et les phonons "mous"

En 2001, Fischetti et al. ont proposés une étude théorique visant à expliquer l'origine physique de la dégradation de mobilité couramment observée dans les dispositifs ayant un empilement utilisant un oxyde de grille de type high-κ [Fischetti01]. Les électrons subiraient en effet un fort couplage avec les phonons optiques peu énergétiques de l'interface oxyde/canal, caractéristiques des matériaux high-κ (ils sont notés SO pour *soft optical phonons*). L'intensité de cette interaction dépend des propriétés intrinsèques de l'oxyde : sa permittivité qui traduit une forte polarisation des liaisons dans le matériau qui écrantent les champs électriques extérieurs. Contrairement au cas des liaisons Si-O du SiO_2 (ou SiON), les liaisons sont peu covalentes et ont un caractère fortement ionique. Les atomes oscillent autour de leur position d'équilibre et engendrent des phonons de faible énergie qui peuvent interagir avec les porteurs de la couche d'inversion.

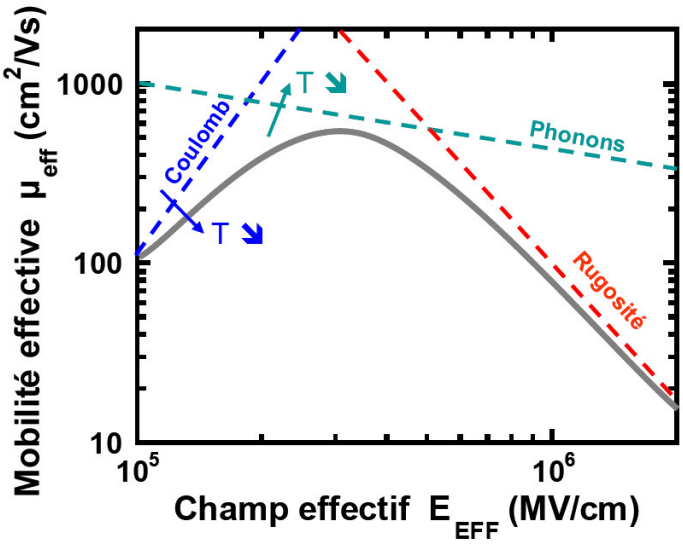


Fig. II.14 – Evolution de la mobilité effective des porteurs en fonction du champ effectif et de la température, d'après [Takagi94a].

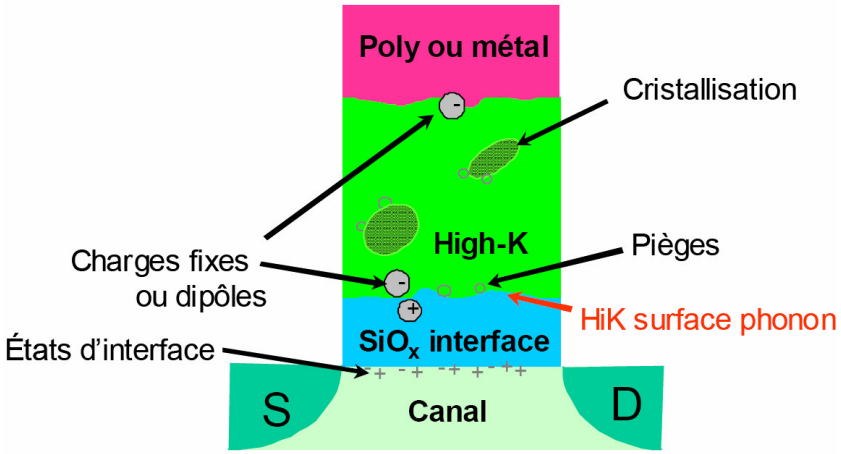


Fig. II.15 – Illustration des différents mécanismes qui peuvent dégrader la mobilité dans un transistor ayant un empilement de grille *high-*κ-métal, d'après [Weber05].

Ce mécanisme est peu influent pour les oxydes SiO_2 car la force de couplage ionique de la liaison est faible et que les phonons ont une énergie trop élevée ($\hbar\omega \simeq 150$meV) pour permettre des phénomènes d'émissions et d'absorptions fréquents à température ambiante (le nombre de phonons est trop faible à 300K). A L'opposé, l'oxyde HfO_2 possède une forte constante de couplage ionique et l'énergie des phonons associés est faible ($\hbar\omega = 12.4$ et 48.35meV pour les deux modes TO) permettant un nombre de phonons élevé à température ambiante et une probabilité d'interaction élevée par rapport au SiO_2. On parle alors d'interaction avec les phonons à distance ou *remote phonon scattering* en anglais. Dans la théorie, ce mécanisme de diffusion doit engendrer une forte dégradation de la mobilité (de l'ordre de 40% pour une permittivité de $20\varepsilon_0$) [Fischetti01, Ren03, Fischetti07]. Remarquons que l'influence des phonons SO est écrantée par le gaz d'électrons de la charge d'inversion en forte inversion [Fischetti07]. Les valeurs de dégradation liées à ce mécanisme reportées dans la littérature semblent toutefois être surestimées puisque des études expérimentales évaluent entre 13% à 16% la dégradation de mobilité sur des dispositifs ayant une grille métallique [Weber06]. Il est en effet démontré que l'application d'une contrainte tensive et que l'écrantage induit par la forte densité de charges libres présentes dans la grille métallique améliorent la mobilité résultant de l'interaction avec les phonons SO (μ_{SO}) [Datta03, Weber06].

II.2.5.c Les défauts neutres résultants de l'implantation

Citer Sclar et McGill [Sclar56, McGill75]

Plusieurs études ont récemment démontrées que la mobilité était dégradée sur les transistors courts [Andrieu05, Cros06b]. Cette dégradation affecte aussi bien les électrons que les trous et semble être présente sur toutes les architectures indépendamment :

- de la présence d'une grille métallique [Bidal08],
- de la présence d'un oxyde high-κ [Pham-Nguyen08],
- de la présence d'une contrainte mécanique [Andrieu05, Bidal08],
- de la présence d'un dopage de canal [Cros06b, Pham-Nguyen08]
- du type d'architecture (FDSOI [Gallon04, Pham-Nguyen08], sur silicium massif [Andrieu05, Cros06b], à grille enrobante GAA [Cros06b] ou FinFET [Ramos06].

Cette dégradation de mobilité qui affecte grandement les propriétés de transport dans les canaux courts semble ne pas avoir de dépendance en température et pourrait résulter de multiples collisions entre les porteurs et des défauts électriquement neutres. Le mécanisme théorique est détaillé dans [Sclar56, McGill75] et les étude expérimentales [Cros06b, Dupre07] semblent confirmer cette hypothèse. Comme suggéré dans [Cros06b], ces défauts pourrait être de nature interstitielle, résultant d'une guérison incomplète des défauts créés lors de l'implantation. Leur détection par le biais de la caractérisation physique serait alors impossible par voie directe et leur inactivité électrique les rendrait invisible sur les tensions de seuil ou de bandes-plates. La figure II.16 illustre la localisation présumée de ces défauts pour un architecture avec et sans poches. Le mécanismes de collosion des porteurs avec des défautss neutres est traité dans [Sclar56, McGill75, Kiréev75]

Pour plus d'informations à ce sujet, le lecteur pourra se reporter au travail de thèse d'A. Cros [Cros06a] et à l'étude expérimentale menée dans la partie §II.2 dans laquelle nous modéliserons l'influence de ce mécanisme sur le profil $\mu(L_{eff})$.

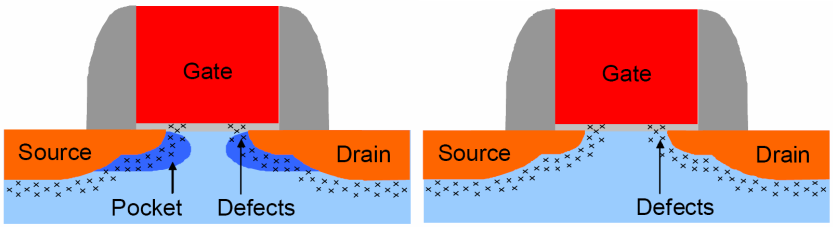


Fig. II.16 – Localisation présumée des défauts neutres résultant d'une guérison incomplète des défauts d'implantation, pour un transistor avec et sans poches de surdopage, d'après [Cros06b].

II.2.6 Influence de la température et du champ sur la mobilité

Comme l'ont montrées les parties précédentes, chaque mécanisme possède une signature caractéristique en fonction de la température et du champ électrique. Ces deux éléments fournissent donc en général un moyen pour détecter la présence et quantifier l'influence d'un mécanisme de collision donné dans le transport dans le canal. Les dépendances en champ, dopage et température des principaux mécanismes de collision sont données dans le tableau II.2.6.

Type d'interaction	dépendance champ/dopage	dépendance T
Coulomb (μ_{cb}) [Jeon89, Sze81]	$1/N_I$	$T^{1-1.5}$
Rugosité de surface (μ_{sr})	${\mathcal{E}_{eff}}^{-1}$ (e) ${\mathcal{E}_{eff}}^{-2}$ (h)	$T^{0.5}$
Phonons acoustiques (μ_{ph})	${\mathcal{E}_{eff}}^{-1/3}$	$T^{-3/2}$
Phonons optiques (μ_{opt})	${\mathcal{E}_{eff}}^{-1/b}, 3 \leq b \leq 6$	$T^{-a}, 1 \leq a \leq 3/2$
Défauts neutres (μ_N) [Kiréev75, Sclar56]	$1/N_N$	

Tab. II.3 – Dépendance en température et en champ des différentes contributions à la mobilité bas champ.

II.3 Le transport non-stationnaire

Nous venons de voir les mécanismes qui permettent de comprendre le transport stationnaire, valable lorsque le nombre de collisions dans le canal est statistiquement grand. Lorsque la longueur de grille est réduite en deçà de 100nm, les phénomènes de transport hors d'équilibre qui étaient négligeables sur les transistors longs deviennent influents et doivent être pris en compte. Dans cette partie, nous allons donc détailler les deux approches de transport non-stationnaire qui permettent d'expliquer les performances électriques des dispositifs courts :

- Le phénomène de survitesse, qui résulte du fait que la relaxation en énergie des porteurs n'est pas instantanée et permet d'atteindre des vitesses de transport supérieures à l'équilibre.
- Le transport balistique qui prévoit que la probabilité qu'un porteur passe sans interaction de la source au drain devient significative lorsque la longueur de canal est réduite.

II.3.1 Les phénomènes de survitesse : une théorie vite oubliée

Dans la partie §II.2.3, nous avons vu que le taux de collisions avec des phonons optiques augmente lorsque le champ électrique latéral est élevé (typiquement $\mathcal{E}_{\parallel} > 10^6$V/m). Ce phénomène plafonne la vitesse de dérive des porteurs : c'est le phénomène de vitesse de saturation [Jacoboni77]. Ce concept reste entièrement valable pour les transistors longs, pourvu que le champ latérale suffise. Le profil de vitesse et d'énergie des porteurs en fonction de la position dans le canal peut être obtenu par simulation Monte-Carlo [Lundstrom06b], comme le montre la figure II.17.

Le profil de vitesse dans la zone d'entrée du canal attire notre attention, elle constitue une zone de transition dans laquelle la vitesse des porteurs dépasse la valeur à l'équilibre ($\nu_{sat} \sim 10^7$cm/s). Dans cette zone d'environ 100nm de long, le temps de transit des porteurs est inférieur au temps moyen de relaxation en énergie (τ_w) qui représente le temps moyen nécessaire aux porteurs pour se thermaliser à la température du réseau. En d'autres termes, les porteurs ne subissent pas suffisamment de collisions inélastiques (avec les phonons optiques) pour atteindre l'état d'équilibre dès leur entrée dans le canal et ils acquièrent une vitesse supérieure à ν_{sat} [Ge01]. Ce phénomène est une conséquence directe du fait que le temps de relaxation moyen en moment et en énergie sont différents ($\tau_w \neq \tau_m$) [Roldan97], nous reviendrons sur ce point dans la partie expérimentale (chapitre V). Comme les longueurs de canaux des dispositifs modernes sont bien en deçà de 100nm, les champs sont bien supérieurs au champ critique nécessaire pour observer la saturation de la vitesse de dérive dans du silicium massif et le canal est une zone de transition où l'équilibre n'est jamais atteint. On peut donc s'attendre à ce que le phénomène de survitesse soit très présent pour les faibles longueurs de canal et assimilable à de la quasi-balisticité du transport [Ge01].

II.3.1.a Influence des phénomènes de survitesse sur le courant de drain

Si les phénomènes de survitesse ont une influence directe sur le temps de transit des porteurs et donc sur la fréquence de fonctionnement maximale du transistor, leur action sur le courant de drain est plus délicate à analyser. En effet, l'observation d'une survitesse proche du drain d'un transistor dont l'électrostatique est parfaitement contrôlée par la grille (i.e. DIBL=0) n'influence pas le courant car la vitesse et la charge à la source restent identiques. C'est donc la source qui contrôle le courant. Celui-ci respecte la relation de continuité en tout point du canal

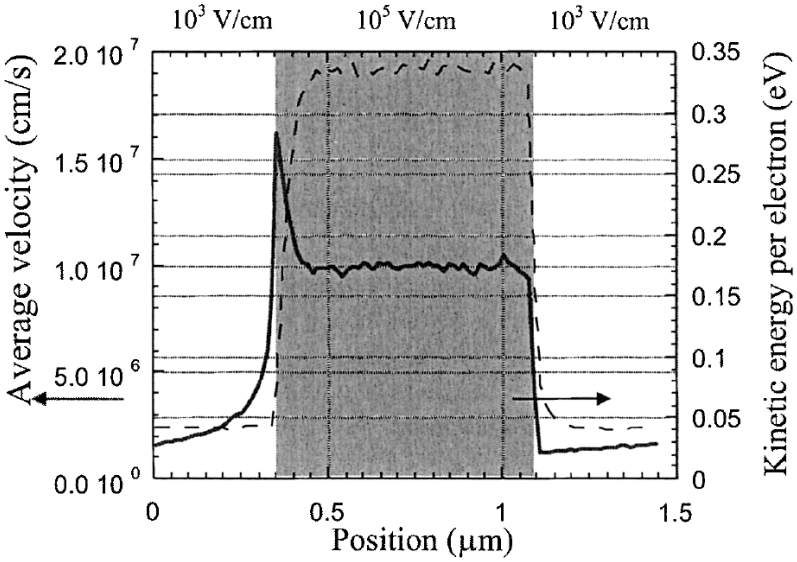


Fig. II.17 – Vitesse moyenne et énergie cinétique des électrons injectés dans un morceau de silicium sous un fort champ électrique, d'après [Lundstrom06b]

[Lochtefeld02] : un phénomène de survitesse au drain entraîne donc simplement une modification de la distribution des charges dans le canal sans changer le courant. En réalité, le contrôle électrostatique de la grille n'est jamais parfait (cf. I.2.1. Cela signifie que les mécanismes de transport après la source influencent l'électrostatique à la source [Saint-Martin04, Lundstrom02a]. Cela justifie donc que lorsque la survitesse intervient à proximité de la source ou que le DIBL est grand (cas des transistors sur silicium massif) le courant du transistor est accru [Lochtefeld02]. La corrélation entre la vitesse à la source et le DIBL à été mis en évidence par des simulations et des mesures [Hu95, Lochtefeld01] comme le montre la figure II.18.

Ces études présentent des résultats basés sur l'analyse de la vitesse effective de dérive (ν_{eff}) en utilisant la relation du même type que (II.35) [Sodini84]. Cette dernière retourne la vitesse moyenne du groupe de porteurs tous mécanismes confondus, et ne donne pas accès directement à l'extraction de la vitesse asymptotique qui plafonne l'augmentation de la ν_d : $\nu_{sat,eff}$. Il n'est donc pas surprenant que les conclusions de ces études évoquent la nécessité de fortes valeurs de DIBL (pour lesquels le transistor est en perçage) pour que la vitesse effective des porteurs dépasse ν_{sat} [Lochtefeld02].

$$\nu_{eff} = \frac{g_m}{W \cdot C_{eff}} \tag{II.35}$$

La modélisation des phénomènes de survitesse est basée sur l'équation de transport de Boltzmann. Un modèle relativement simple est donnée dans [Ge01] (II.36), où $d\mathcal{E}_y/dy$ représente les variations de champ électrique au premier ordre.

$$\nu_{sat,eff} = \nu_{sat} \cdot \left(1 + \frac{k_B T \mu_{eff}}{q\mathcal{E}_y \nu_{sat}} \cdot \frac{d\mathcal{E}_{lat}}{dy}\right) \tag{II.36}$$

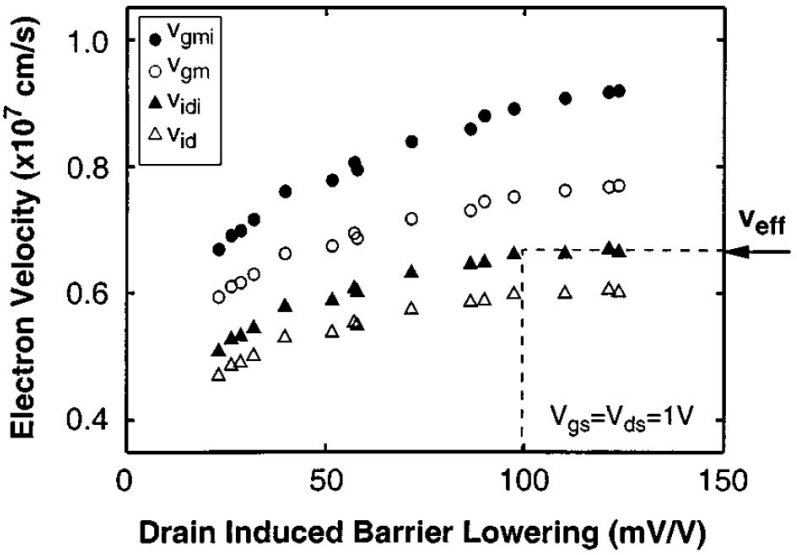


Fig. II.18 – Corrélation entre la vitesse de dérive effective ν_{eff} (extraite par différentes méthodes) et le DIBL, d'après [Lochtefeld01].

II.3.2 Le modèle de transport quasi-balistique : généralités et approximations

Cette partie a pour but d'établir l'expression analytique du courant de drain dans un approche quasi-balistique la plus générale possible. En particulier, nous prendrons soin d'avancer peu d'hypothèses sur la nature du transport, de manière à conserver la compatibilité avec les théories du transport de M. Lundstrom [Lundstrom97, Lundstrom01, Lundstrom02a] et de K. Natori [Natori94, Natori01, Tsuchiya06]. Une fois l'expression générale de I_d établie, nous effectuerons les simplifications possibles et évaluerons leur validité de manière à retrouver les relations analytiques usuelles de la littérature.

II.3.2.a Description du modèle quasi-balistique

Nous considérons un schéma de transport tel que présenté sur la figure II.19. Il illustre les deux mécanismes fondamentaux dans la théorie du transport quasi-balistique :

- **L'injection** : grâce à l'énergie thermique, les flux de porteurs respectifs Φ_s^{out} et Φ_d^{out} peuvent s'échapper des réservoir source et drain. Les vitesses de déplacement associées à ces flux sont notées ν_s^{out} et ν_d^{out},
- **La rétrodiffusion** : Les porteurs quittant les réservoirs peuvent être rétrodiffusés à leur point d'origine avec une probabilité r_s et une vitesse ν_s^{back} pour la source, et respectivement r_d et ν_d^{back} pour le drain. Les parties non rétrodiffusés ont par conséquent les probabilités $(1-r_s)$ et $(1-r_d)$ de rejoindre le drain ou la source.

Nous verrons que la vitesse de retour des porteurs suite à une rétrodiffusion du coté du drain n'intervient pas dans le calcul du courant, pour plus de simplicité nous admettrons donc $\nu_s^{back} = \gamma \cdot \nu_s^{out}$ dans les équations qui suivront. Le paramètre γ est ainsi égal au rapport de la vitesse moyenne du flux rétrodiffusé $(r_s \cdot \Phi_s^{out})$ sur celle du flux incident Φ_s^{out} et prend des valeurs comprises entre 0 et 1. Nous remarquerons également que $(1-\gamma^2)$ représente la perte relative moyenne d'énergie cinétique associée à la rétrodiffusion des porteurs. La prise en compte de $\gamma < 1$ revient donc à envisager que des mécanismes de diffusion inélastiques peuvent intervenir dans le processus de rétrodiffusion.

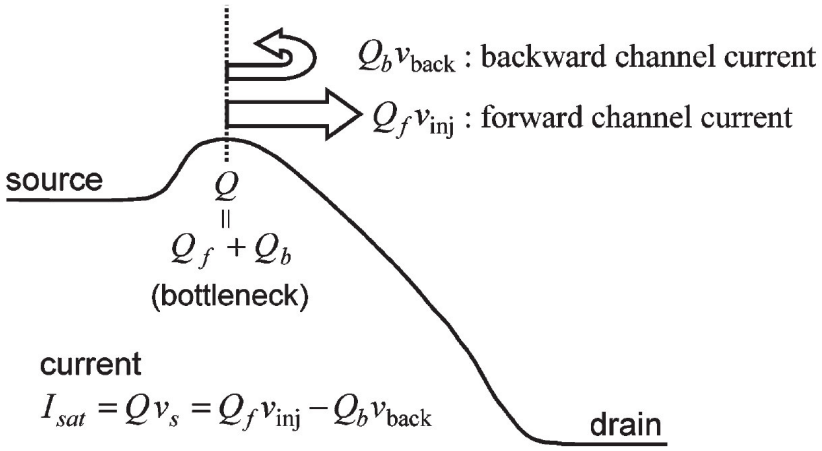


Fig. II.19 – Illustration du mécanisme d'injection à la source, d'après [Tsuchiya06]. Une fraction des porteurs qui franchissent le maximum de barrière de potentiel peuvent être rétrodiffusé avec une vitesse ν_{back} différente de ν_{inj}.

II.3.2.b Établissement de l'équation générale de I_{bal}

Nous cherchons dans un premier temps à définir l'expression de la charge à la source virtuelle Q_s. Tout d'abord, remarquons que la continuité des flux impose la relation (II.37) pour le flux de porteurs rétrodiffusés Φ_{back} (où Q_f et Q_{back} représentent respectivement les densités de charges incidente et rétrodiffusée). Remarquons également que la relation $\Phi_s^{back} = r_s \cdot \Phi_s^{out}$ est toujours conservée.

$$\Phi_{back} = r_s \cdot Q_s \cdot \nu_{inj} = Q_{back} \cdot \nu_{back} \tag{II.37}$$

Cette relation permet alors d'exprimer la densité de charges rétrodiffusées à la source virtuelle en fonction de ν_{inj}, Φ_s^{out} et r_s (II.38). Cette dernière expression permet d'établir le lien entre la théorie quasi-balistique de Lundstrom (élastique : $\gamma = 1$) [Lundstrom97, Lundstrom02a] et celle de Natori (inélastique : $\gamma < 1$).

$$Q_{back} = r_s \cdot \frac{\Phi_s^{out}}{\gamma \cdot \nu_s^{out}} \tag{II.38}$$

La charge au niveau de la source virtuelle se calcule en fonction des différents flux de porteurs entrant et sortant à ce niveau :

- des porteurs émis par le drain vers la source qui ne subissent pas de rétrodiffusion : $(1 - r_d) \cdot \Phi_d^{out}$,
- des porteurs émis par la source vers le drain qui ne subissent pas de rétrodiffusion : $(1 - r_s) \cdot \Phi_s^{out}$,
- des porteurs émis par la source et rétrodiffusés : Q_{back}, dont l'expression analytique est donnée en (II.38).

Q_s peut alors s'écrire comme (II.39) qui peut être mis sous la forme (II.40) afin d'en extraire une expression analytique de Φ_s^{out} en fonction des autres variables.

$$Q_s = q\left(\frac{\Phi_s^{out}}{\nu_s^{out}} - r_s \cdot \frac{\Phi_s^{out}}{\gamma \cdot \nu_s^{out}} - (1-r_d)\frac{\Phi_d^{out}}{\nu_d^{out}}\right) \tag{II.39}$$

$$Q_s = q \cdot \Phi_s^{out} \cdot \left[\left(\frac{1}{\nu_s^{out}} - \frac{r_s}{\gamma \cdot \nu_s^{out}}\right) - \frac{1-r_d}{\nu_d^{out}} \cdot \frac{\Phi_d^{out}}{\Phi_s^{out}}\right] \tag{II.40}$$

Une nouvelle fois, la relation de conservation des flux permet d'exprimer le courant de drain en fonction des flux entrant et sortant au niveau de la source :

$$I_{bal} = W \cdot q \cdot \left(\Phi_s^{out} - \Phi_s^{in}\right) \tag{II.41}$$

Ou encore, en sortant Φ_s^{out} du terme de droite :

$$I_{bal} = W \cdot q \cdot \Phi_s^{out}\left(1 - \frac{\Phi_s^{in}}{\Phi_s^{out}}\right) \tag{II.42}$$

Nous pouvons éliminer Φ_s^{in} en l'explicitant en fonction du flux provenant du drain et du flux de porteurs rétrodiffusés à la source :

$$\Phi_s^{in} = r_s \cdot \Phi_s^{out} + (1-r_d) \cdot \Phi_d^{out} \tag{II.43}$$

En remplaçant Φ_s^{out} par son expression analytique issue de (II.40), nous aboutissons à une nouvelle expression de I_{bal} en fonction de la charge Q_s et du rapport des flux incident et émergent.

$$I_{bal} = W \cdot Q_s \cdot \left[\frac{(1-r_s) - (1-r_d) \cdot \dfrac{\Phi_d^{out}}{\Phi_s^{out}}}{\left(\dfrac{1}{\nu_s^{out}} + \dfrac{r_s}{\gamma \cdot \nu_s^{out}}\right) + \dfrac{1-r_d}{\nu_d^{out}} \cdot \dfrac{\Phi_d^{out}}{\Phi_s^{out}}}\right] \tag{II.44}$$

Finalement, en sortant ν_s^{out} cette dernière équation s'écrit comme (II.45). **Elle constitue l'expression analytique du courant de drain balistique dans le cas général.**

$$I_{bal} = W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s}\right) \cdot \left[\frac{1 - \left(\dfrac{1-r_d}{1-r_s}\right) \cdot \dfrac{\Phi_d^{out}}{\Phi_s^{out}}}{\left(\dfrac{1+r_s/\gamma}{1+r_s}\right) + \left(\dfrac{1-r_d}{1+r_s}\right) \cdot \dfrac{\nu_s^{out}}{\nu_d^{out}} \cdot \dfrac{\Phi_d^{out}}{\Phi_s^{out}}}\right] \tag{II.45}$$

II.3.2.c Expression analytique des flux et des vitesses à la source et au drain

Les flux de porteurs Φ_s^{out} et Φ_s^{out} s'expriment en fonction de la densité de charges d'inversion à la source virtuelle (N_{2D}) et des intégrales de Fermi-Dirac (II.47).

$$\Phi_s^{out} = N_{2D} \cdot \nu_s^{out} \cdot \mathcal{F}_0(\eta_F) \tag{II.46}$$

$$\Phi_d^{out} = N_{2D} \cdot \nu_d^{out} \cdot \mathcal{F}_0(\eta_F - \frac{qV_{ds}}{kT}) \tag{II.47}$$

Si l'on considère que les vitesses des porteurs incidents et émergents sont égales à la vitesse thermique prenant en compte la dégénérescence, leurs expressions analytiques sont données par (II.49).

$$\nu_s^{out} = \nu_{th}\frac{\mathcal{F}_{1/2}(\eta_F)}{\mathcal{F}_0(\eta_F)} \tag{II.48}$$

$$\nu_d^{out} = \nu_{th}\frac{\mathcal{F}_{1/2}(\eta_F - V_{ds}/k_BT)}{\mathcal{F}_0(\eta_F - V_{ds}/kT)} \tag{II.49}$$

II.3.2.d Approximation de I_{bal} pour les tensions de drain élevées

Lorsque $q \cdot V_{ds} \gg kT$, il est infiniment peut probable que des porteurs issus du drain puissent franchir la chute de potentiel et rejoindre la source. Ils ont donc toutes les chances d'être rétrodiffusés et l'équation (II.45) peut être approximée en posant $r_d \simeq 1$ (sachant que le rapport des densités de charges $(\Phi_d^{out}/\nu_d^{out})/(\Phi_s^{out}/\nu_s^{out})$ est faible). Nous aboutissons à une expression analytique du courant de drain balistique (II.50) identique à [Tsuchiya06] et cohérente avec [Lundstrom97] si le rapport des vitesses γ est supposé égal à 1.

$$I_{bal}^{sat} \simeq W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s/\gamma}\right), \quad V_{ds} \gg \frac{kT}{q} \tag{II.50}$$

Précisons que la relation (II.50) est valable quelque soit la valeur de η_F et donc dans le cas dégénéré. La seule hypothèse est la suivante : $(1-r_d)/(1-r_s)$ est très proche de 0 car les porteurs en provenance du drain ne peuvent pas passer la barrière de potentiel imposé par V_{ds}. Cette relation aurait été beaucoup plus difficile à obtenir rigoureusement si nous avions posé $r_s = r_d$ dès le départ puisque la dissymétrie dans le transport imposée par V_{ds} ne peut pas être prise en compte dans un unique coefficient de rétrodiffusion.

II.3.2.e Approximation de I_{bal} pour les tensions de drain faibles ($V_{ds} \ll kT/q$)

Lorsque $q \cdot V_{ds} < kT$, il parait justifié de supposer que les porteurs aient la même probabilité de rétrodiffusion du coté de la source que du coté du drain. De la même manière, la dégénérescence étant la même au niveau de chaque réservoir, les vitesses d'injection au drain et à la source pourrons être supposées égales. Les égalités $r_d = r_s$ et $\nu_d^{out} = \nu_s^{out}$ pourront donc être utilisées en première approximation pour simplifier la relation (II.45).

$$I_{bal}^{lin} \simeq W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s}\right) \cdot \left[\frac{1-\dfrac{\Phi_d^{out}}{\Phi_s^{out}}}{\left(\dfrac{1+r_s/\gamma}{1+r_s}\right) + \left(\dfrac{1-r_s}{1+r_s}\right)\cdot\dfrac{\Phi_d^{out}}{\Phi_s^{out}}}\right] \tag{II.51}$$

En rappelant que $0 < \gamma \leq 1$ et que la condition $V_{ds} < kT$ implique d'être proche de l'équilibre thermique, il parait légitime de supposer que le rapport de flux $\Phi_s^{out}/\Phi_d^{out}$ est proche de 1 (cela revient à supposer que le courant balistique à faible V_{ds} provient d'un léger déséquilibre par rapport à la condition d'équilibre thermique). Cette lourde simplification permet d'aboutir à une expression simplifiée de I_{bal}^{lin} (II.52). Elle est d'autant plus valide que le rapport $q \cdot V_{ds}/kT$ est proche de 0.

$$I_{bal}^{lin} \simeq W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s/\gamma}\right) \cdot \left(1 - \frac{\Phi_d^{out}}{\Phi_s^{out}}\right), \quad V_{ds} < \frac{kT}{q} \tag{II.52}$$

En explicitant Φ_s^{out} et Φ_d^{out} par leurs expressions analytiques, la relation devient

$$I_{bal}^{lin} \simeq W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s/\gamma}\right) \cdot \left(1 - \frac{\mathcal{F}_{1/2}(\eta_F - V_{ds}/kT)}{\mathcal{F}_{1/2}(\eta_F)}\right), \quad V_{ds} < \frac{kT}{q} \tag{II.53}$$

Dans le cas ou V_{ds} peut être considéré comme négligeable devant kT/q, il est possible de développer l'intégrale de Fermi à l'ordre 1 autour de η_F en utilisant le théorème de Taylors (II.54) la relation se simplifie et devient conforme à [Lundstrom02a]

$$\mathcal{F}_{1/2}\left(\eta_F - \frac{qV_{ds}}{kT}\right) \simeq \mathcal{F}_{1/2}\left(\eta_F\right) - \frac{qV_{ds}}{kT} \cdot \mathcal{F}_{-1/2}\left(\eta\right), \quad V_{ds} \ll \frac{kT}{q} \tag{II.54}$$

$$I_{bal}^{lin} \simeq W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s/\gamma}\right) \cdot \frac{qV_{ds}}{kT} \cdot \frac{\mathcal{F}_{-1/2}(\eta)}{\mathcal{F}_{1/2}(\eta)}, \quad V_{ds} \ll \frac{kT}{q} \tag{II.55}$$

Dans l'approximation de Boltzmann ($\eta < 2$) l'intégrale de Fermi-Dirac peux s'approximer par la fonction Γ comme en (II.56) dont la démonstration est faite en annexe A.

$$\mathcal{F}_j\left(\eta\right) \simeq e^{\eta} \cdot \Gamma\left(j+1\right), \quad \eta < 2 \tag{II.56}$$

Cette approximation permet de simplifier la relation I_{bal}^{lin} (cf. II.57), qui devient conforme à l'expression dans [Lundstrom01] en posant $\gamma = 1$ et en supposant $r_s \simeq 1$.

$$I_{bal}^{lin} \simeq W \cdot Q_s \cdot \nu_s^{out} \cdot \left(\frac{1-r_s}{1+r_s/\gamma}\right) \cdot \frac{qV_{ds}}{kT}, \quad V_{ds} \ll \frac{kT}{q} \text{ et } \eta < 2 \tag{II.57}$$

II.3.3 La notion de distance critique (kT-*layer*) et de rétrodiffusion

La distance critique (ou *kT-layer*) est définie par Lundstrom comme étant la zone proche de la source dans laquelle le potentiel de surface varie de quelques unités de potentiel thermique (kT/q), permettant ainsi au porteur d'être rétrodiffusé à la source dans le cas d'une interaction. Passé cette zone, le flux de porteurs est supposé rejoindre le drain. Dans la théorie quasi-balistique de Lundstrom [Lundstrom02a], la *kT-layer* constitue donc la seul limitation à l'injection des porteurs depuis la source qui va faire que la vitesse d'injection sera inférieure à la vitesse thermique dégénérée. La probabilité de rétrodiffusion est alors modélisée par le paramètre r, supposé rendre compte du taux d'interaction des porteurs avec les centres de diffusion au sein de la *kT-layer*.

II.3.3.a Les éléments influents sur le taux de rétrodiffusion

Si l'on admet l'hypothèse de la *kT-layer*, il semble pertinent de se demander quels sont les mécanismes d'interactions qui limitent l'injection des porteurs de la source au canal. Des simulations Monte-Carlo ont été effectuées dans [Ponton06] précisément dans cet objectif. Elles montrent que la diffusion des porteurs avec la rugosité de surface contribue à hauteur de 80% (silicium massif) et 70% (FDSOI) dans le taux de rétrodiffusion, en rapport aux autres contributions considérées (phonons optiques et acoustiques). Néanmoins, un point semble gênant : l'interaction des porteurs avec la rugosité de surface est précisément le paramètre d'ajustement qui permet au simulateur de reproduire convenablement le comportement électrique des dispositifs. Cette incohérence ne remet-elle pas en cause la viabilité de toute l'étude ?
Si toutefois les conclusions de cette étude s'avèrent justes, cela signifie que les dispositifs sur

silicium contraint devraient avoir un taux de rétrodiffusion plus faible car les collisions avec la rugosité surface sont moins influentes [Evans05]

II.4 Vers l'unification des mécanismes de transport

Cette partie a pour but de proposer un modèle complet, qui repose sur le concept de vitesse limite [Fleury09a], afin d'unifier les théories de transport dérive-diffusion et quasi-balistique.

II.4.1 Expression du courant de drain dans l'approche dérive-diffusion

Nous allons montrer comment intégrer dans une même équation les concepts de vitesse de saturation et de vitesse d'injection (théorie quasi-balistique). Pour cela, nous allons partir des équations générales dérive-diffusion en régime linéaire et saturé (hypothèse champ contant dans le canal : $\mathcal{E} \simeq V_{ds}/L$)

$$I_{dd} = \frac{W}{L} V_{ds} \int_{\psi_s=0}^{V_{ds}} \mu_{dd} \cdot Q_{inv}(V_{gt} - \psi_s) \cdot \mathrm{d}\psi_s \tag{II.58}$$

La charge d'inversion varie linéairement avec le potentiel de surface par le biais de la relation (II.59) :

$$Q_{inv}(\psi_s) \simeq C_{eff}\,(V_{gt} - \psi_s) \tag{II.59}$$

Et la relation (II.58) peut s'approximer par (II.60)

$$I_{dd} \simeq \frac{W}{L} V_{ds} \cdot \mu_{dd} \cdot C_{eff} \left(V_{gt} - \frac{V_{ds}}{2} \right) \tag{II.60}$$

Lorsque la saturation est atteinte ($V_{ds} \geq V_{gt}$), le courant n'évolue plus et V_{ds} doit être remplacer par V_{gt} dans l'équation (II.60). Il est alors possible d'établir les expressions des courants de drain dérive diffusion en régime linéaire (I_{dd}^{lin}) et en régime saturé (I_{dd}^{sat}), dans la zone de forte inversion ($V_{gt} \gg kT/q$) en fonction de la charge à la source $Q_s = C_{eff} V_{gt}$.

$$I_{dd}^{lin} \simeq W \cdot \frac{\mu_{dd} V_{ds}}{L} \cdot Q_s \left(1 - \frac{V_{ds}}{2V_{gt}} \right), \quad V_{ds} < V_{gt} \tag{II.61}$$

$$I_{dd}^{sat} \simeq W \cdot \frac{\mu_{dd} V_{gt}}{L} \cdot \frac{Q_s}{2}, \quad V_{ds} \geq V_{gt} \tag{II.62}$$

Lorsque le rapport V_{ds}/V_{gt} est petit devant 1 le terme entre parenthèse dans l'équation (II.62) devient très proche de 1 et le courant de drain en régime linéaire peut être approximé par la relation (II.63).

$$I_{dd}^{lin} \simeq W \cdot \frac{\mu_{dd} V_{ds}}{L} \cdot Q_s, \quad V_{ds} \ll V_{gt} \tag{II.63}$$

II.4.2 Le concept de mobilité apparente

Quelques soient les mécanismes de transports mis en jeu dans le transport des porteurs, il est possible de définir des mobilités *apparentes*. Les grandeurs μ_{app}^{lin} et μ_{app}^{sat} représenteront cette mobilité dans les régimes linéaire et de saturation. Leur expression mathématique est donnée en (II.64) et (II.65)

$$\mu_{app}^{lin} = I_d^{lin} \cdot \frac{L}{W C_{eff} V_{gt} V_{ds}} \tag{II.64}$$

$$\mu_{app}^{sat} = I_d^{sat} \cdot \frac{2L}{WC_{eff}(V_{gt})^2} \tag{II.65}$$

En utilisant les relations précédantes il en vient :

$$\frac{1}{\mu_{app}^{lin}} = \frac{2kT}{\nu_{inj}} + \frac{L}{\mu_{dd}}, \quad V_{ds} < \frac{2kT}{q} \tag{II.66}$$

$$\frac{1}{\mu_{app}^{sat}} = \frac{V_{gt}}{2\nu_{inj}} + \frac{L}{\mu_{dd}}\left(1 + \frac{\mu_{dd}V_{gt}}{L\nu_{lim}}\right), \quad V_{ds} < \frac{2kT}{q} \tag{II.67}$$

En combinant ces deux dernières équations, il est possible d'éliminer μ_{dd} et d'aboutir à une relation qui permet l'extraction de ν_{inj} et ν_{lim} :

$$\frac{L}{\mu_{app}^{lin}} - \frac{L}{\mu_{app}^{sat}} = \frac{2kT}{\nu_{inj}} - \frac{V_{gt}}{\nu_{lim}} \tag{II.68}$$

II.4.3 L'unification des théories de transport

Toutes les méthodologies expérimentales destinées à extraire le taux de balisticité dans les transistors présupposent que le modèle balistique s'applique aux dispositifs. Nous décidons de prendre une approche plus globale, et par l'unification des théories de transport dérive-diffusion et balistique, d'établir une nouvelle méthodologie d'extraction qui ne présuppose pas la nature du transport. Le but ultime de cette nouvelle méthode est de pourvoir extraire la nature de la vitesse limitante en régime saturé. Tout d'abord, montrons que le formalisme de Lundstrom exposé au §II.3.2 revient à une somme de type Matthiessen fonction des contributions balistique et dérive-diffusion, aussi bien en régime linéaire que saturé.

II.4.3.a Le facteur de rétrodiffusion r en régime linéaire et saturé

En se basant sur la théorie des flux de McKelvey [McKelvey61, Shockley62], Lundstrom et al. ont pu démontrer qu'il était possible de modéliser l'influence des collisions dans le régime quasi-balistique par le facteur de rétrodiffusion r, défini comme la rapport du flux de porteurs rétro-diffusés vers la source par rapport aux flux injecté depuis celle-ci [Lundstrom97, Lundstrom02a]. Une expression relativement simple, basée sur la description de l'émission thermoïonique donnée en [Tanaka95] est alors proposée pour le régime linéaire et le régime saturé, cf. (V.7) et (II.70).

$$r^{lin} \stackrel{\text{def}}{=} \frac{L_{eff}}{L_{eff} + \lambda_\mu}, \quad V_{ds} \leq \frac{kT}{q} \tag{II.69}$$

$$r^{sat} \stackrel{\text{def}}{=} \frac{l_{kT}}{l_{kT} + \lambda_\mu}, \quad V_{ds} \gg \frac{kT}{q} \tag{II.70}$$

Dans ces équations, λ_μ est le libre parcours moyens des porteurs caractéristique du mécanisme de rétrodiffusion qui est approximé à partir de la mobilité faible champ du transistor long comme en (II.71) [Lundstrom01, Palestri06]. Cette définition diffère d'un facteur deux de celle donnée dans [Shockley62], résultant de la convention choisie et qui n'influe pas sur le résultat final [Clerc06].

$$\lambda_\mu \stackrel{\text{def}}{=} \frac{2\mu kT}{q\nu_{th}} \tag{II.71}$$

Par ailleurs, si l'introduction de la mobilité à bas champ dans l'expression d'un paramètre propre au régime de transport quasi-balistique, elle confirme l'influence que joue toujours la mobilité sur le courant des transistors courts et la corrélation qui existe entre les variations relatives de courants et celle de mobilité observées expérimentalement [Lin06, Chen06]. Dans une approche quasi-balistique (le concept reste vrai dans le cas d'un saturation de vitesse), le facteur de corrélation n'est autre que le taux de balisticité $B^{sat} \sim 0.5$ [Lundstrom01].

La grandeur l_{kT} est la distance critique prise depuis la position de la source virtuelle qui caractérise une chute de potentiel de surface de kT/q : $\psi_s(l_{kT}) = kT/q$. Baptisée *kT-layer* par Berz, elle définit la limite dans laquelle les porteurs peuvent encore joindre la source lors du mécanisme de rétrodiffusion [Berz85]. Une expression analytique est donnée par Lundstrom [Lundstrom97] en fonction du champ électrique qui règne à la source virtuelle $\mathcal{E}_x(0)$ (II.72).

$$l_{kT} = \frac{kT}{q\mathcal{E}_x(0)} \tag{II.72}$$

Il est possible de démontrer que dans le cas du transistor MOS, en régime de saturation en inversion forte, le profil de potentiel dans la théorie dérive diffusion conduit à définir $\mathcal{E}_y(0)$ et l_{kT} comme en (II.72), cf [Clerc06]. Le lecteur pourra se réfèrer à l'Annexe C pour la démonstration de ce résultat.

$$\mathcal{E}_y(0) \simeq \frac{V_{gt}}{2L_{eff}}, \quad l_{kT} \simeq \frac{2kT \cdot L_{eff}}{qV_{gt}} \tag{II.73}$$

Notons cependant que la relation II.73 repose sur une approche dérive-diffusion qui considère que les distributions de porteurs sont à l'équilibre thermique, i.e. décrites par une forme Maxwellienne. En 1985, Berz démontra que la distribution Maxwellienne ne permettait pas une représentation fidèle de la distribution de porteurs à la jonction et que la définition de la *kT-layer* comme une distance caractérisée par une chute de potentiel de $2kT/q$ permet une meilleur modélisation de l'émission thermoïonique. Ce point en en accord avec les simulations Monte-Carlo [Palestri06] et, dans le cas des transistors MOS, et peut être modélisé en considérant le caractère non Maxwellien de la distribution de porteurs chauffés arrivant au drain [Clerc08].

Finalement, précisons qu'une expression plus générale, issue de la résolution de l'équation de transport de Boltzmann, permet d'établir la continuité du facteur de rétrodiffusion entre les régime linéaire et de saturation [Clerc06].

$$r(V_{ds}) = \frac{l_{kT} \cdot \left(1 - \exp\left(-\frac{qV_{ds}}{kT}\right)\right)}{\lambda_\mu + l_{kT} \cdot \left(1 - \exp\left(-\frac{qV_{ds}}{kT}\right)\right)} \tag{II.74}$$

En réutilisant le résultat (II.73), il est possible d'expliciter les relations (V.7) et (II.70) en fonction des autres variables du système, cf. (II.75) et (II.76).

$$r^{lin} = \left(1 + \frac{2kT\mu}{q\nu_{inj}L_{eff}}\right)^{-1} \tag{II.75}$$

$$r^{sat} = \left(1 + \frac{\mu V_{gt}}{\nu_{inj}L_{eff}}\right)^{-1} \tag{II.76}$$

Le taux de balisticité peut alors s'écrire comme () On peut alors définir les taux de baliticité comme ()

$$BR^{lin} = \left(1 - r^{lin}\right) = \left(1 + \frac{L_{eff}\nu_{inj}}{\mu kT}\right)^{-1} \tag{II.77}$$

$$BR^{sat} = \frac{1 - r^{sat}}{1 + r^{sat}} = \left(1 + \frac{2L\nu_{inj}}{\mu V_{gt}}\right)^{-1} \tag{II.78}$$

Lorsque ces expressions sont injectées dans les expressions du courant I_{lin} et I_{on}, les relations Matthiessen suivantes sont obtenues. Ces dernières témoignent de la continuité qui existe entre les approches dérive-diffusion et quasi-balistique, conformément à [Lundstrom97, Lundstrom02b].

$$\frac{1}{I_{lin}} = \frac{1}{I_{bal}^{lin}} + \frac{1}{I_{dd}^{lin}} \tag{II.79}$$

$$\frac{1}{I_{on}} = \frac{1}{I_{bal}^{on}} + \frac{1}{I_{dd}^{on}} \tag{II.80}$$

Il est alors possible de définir la vitesse effective de transport dans le cas du régime de saturation sous une forme plus traditionnelle (II.81), qui rappelle la relation de Caughey-Thomas utilisée pour modéliser la déviation de la loi d'Ohm provoquée par ν_{sat} [Sodini84]. Cette relation est conforme avec [Lundstrom97, Dobbie08] et reste entièrement compatible avec le formalisme quasi-balistique de Lundstrom, conformément aux simulations Monte-Carlo [Palestri06, Clerc08].

$$\nu_{eff}^{sat} = \frac{\mu\mathcal{E}_x(0)}{2 + \dfrac{\mu\mathcal{E}_x(0)}{\nu_{inj}}} \tag{II.81}$$

II.4.3.b Utilisation du concept de mobilité apparente pour unifier les théories

Afin d'assurer la continuité entre les théories de transport dérive-diffusion et balistique, tout en conservant la validité des équations de Lundstrom, nous proposons d'utiliser le concept de mobilité apparente. Dans un premier temps, nous étendons le concept de mobilité balistique en régime linéaire introduit par Shur [Shur02] au régime de saturation, par analogie avec les équations dérive-diffusion.

$$\mu_{bal}^{lin} = \frac{I_d^{lin} \cdot L_{eff}}{WQ_s \cdot V_{ds}} = \frac{qL_{eff}\nu_{inj}}{2kT} \tag{II.82}$$

$$\mu_{bal}^{sat} = \frac{2I_d^{sat} \cdot L_{eff}}{WQ_s \cdot V_{gt}} = \frac{qL_{eff}\nu_{inj}}{V_{gt}} \tag{II.83}$$

D'autre part, toujours par analogie avec les équations du transport dérive-diffusion, les mobilités apparentes en état linéaire et saturé peuvent s'exprimer comme (II.84) et (II.85), respectivement.

$$\mu_{app}^{lin} = I_d^{lin} \cdot \frac{L}{WC_{eff}V_{gt}V_{ds}} \tag{II.84}$$

$$\mu_{app}^{sat} = I_d^{sat} \cdot \frac{2L}{WC_{eff}(V_{gt})^2} \tag{II.85}$$

Finalement, nous pourrons utiliser le résultat de la partie précédente qui prouve que le formalisme de Lundstrom peut s'assimiler à une loi d'addition inverse des courants (de type

Matthiessen). Il est donc possible, à partir des relations que nous venons de définir, d'exprimer la mobilité apparente du courant sous une forme conviviale, cf. (II.86) et (II.87)

$$\frac{1}{\mu_{app}^{lin}} = \frac{2kT}{\nu_{inj}} + \frac{L}{\mu_{dd}}, \quad V_{ds} \leq \frac{2kT}{q} \tag{II.86}$$

$$\frac{1}{\mu_{app}^{sat}} = \frac{V_{gt}}{2\nu_{inj}} + \frac{L}{\mu_{dd}}\left(1 + \frac{\mu_{dd}V_{gt}}{L\nu_{lim}}\right), \quad V_{ds} \gg \frac{2kT}{q} \tag{II.87}$$

Cette relation permet l'introduction du paramètre $\nu_{lim} \stackrel{\text{def}}{=} \min(\nu_{inj}, \nu_{sat})$ qui permet d'unifier les deux mécanismes de limitation de la vitesse : vitesse de saturation et par la vitesse d'injection. La notion de mobilité apparente sera ré-utilisée dans le chapitre V pour construire une méthodologie d'extraction et analyser les propriétés de transport en régime de saturation.

II.5 Bilan et discussions

Ce chapitre à permis de poser les bases pour mieux comprendre les mécanismes de transport qui régissent le courant des transistors d'hier et d'aujourd'hui.

Dans un premier temps, nous avons pu détailler le rôle des mécanismes de collisions sur le transport, modélisé par l'approche dérive-diffusion *hitorique*. La conaissance du libre parcours moyen permet une estimation fiable de la qualité de transport. Il témoigne de la forte présence des collisions dans les canaux courts et laisse peu d'espoir à l'observation d'un transport purement balistique. D'un autre coté, les faibles longueurs de canal propres aux transistors modernes résultent en un champ latéral fort et une plus grande probabilité de transport balistique : **le transport ne peut plus être considéré comme stationnaire**. Les phénomènes de transport hors d'équilibre (survitesse et de quasi-balisticité) sont donc à considérer avec une grande attention.

L'utilisation du concept de vitesse limite, englobant à la fois les phénomènes de saturation de vitesse, de survitesse et d'injection à la source permet, via le concept de mobilité apparente, d'unifier les théories de transport. Il a ainsi été possible de définir des expressions générales, qui ne présupposent pas le mécanisme dominant dans le transport des porteurs et permet une continuité de l'approche dérive-diffusion au transport purement balistique.

Cette unification nous sera particulièrement utile par la suite. Nous verrons au chapitre V qu'elle permettra de construire une nouvelle méthode d'extraction expérimentale qui servira à l'analyse des propriétés de transport en régime de saturation.

Chapitre III

Méthodologies d'extraction des paramètres électriques sur des transistors MOS

Sommaire

III.1 Les outils nécessaires à la caractérisation électrique en milieu industriel . . . 78

III.2 Le cœur de l'extraction : un modèle de courant adapté au comportement du transistor MOS moderne en régime linéaire et en inversion forte . . . 78

III.2.1 Influence et correction de la fuite de grille . . . 80

III.2.2 Prise en compte de la résistance série dans le modèle . . . 81

III.2.3 Discussion : les limites de l'extraction en inversion forte . . . 83

III.3 Les méthodes d'extraction de la tension de seuil (V_{th}) . . . 84

III.3.1 L'extraction de V_{th} par mesures capacitives . . . 84

III.3.2 La méthode du courant constant . . . 86

III.3.3 La méthode du V_{th} extrapolé depuis le maximum de transconductance . . . 86

III.3.4 Les méthodes utilisant deux dérivations successives du courant . . . 87

III.3.5 Proposition d'une nouvelle méthode : le V_{th} extrapolé en inversion faible 89

III.4 L'extraction de tous les paramètres du modèle V_{th}, β, Θ_1 et Θ_2 . . . 91

III.4.1 Les techniques utilisant la fonction Y : des valeurs sûres . . . 91

III.4.2 L'état de l'art des améliorations de la fonction Y pour les transistors courts ($L_{eff} < 100$nm) . . . 94

III.4.3 Proposition d'une amélioration de la fonction Y : la méthode de la fonction ξ . . . 97

III.4.4 Nouvelle méthodologie d'extraction utilisant les régimes de forte et de faible inversion . . . 101

III.5 Robustesse des méthodes d'extraction . . . 102

III.5.1 Immunité des résultats vis-à-vis des variations de R_{sd} . . . 102

III.5.2 Immunité des résultats aux variations de paramètres d'entrée . . . 106

III.6 L'extraction des résistances séries . . . 107

III.6.1 État de l'art : la méthode $\Theta(\beta)$ . . . 108

III.6.2 Proposition d'une nouvelle méthode d'extraction : $R_{tot}(1/\beta)$ . . . 110

III.6.3 "Le Graal" : vers l'extraction de R_{sd} sur un seul transistor ? . . . 113

III.7 Discussion et perspectives 115

RÉSUME — Les méthodologies d'extraction sont au cœur de la caractérisation électrique. Elles s'appuient sur des modèles dont les paramètres, une fois extraits, peuvent permettre d'analyser et de comprendre la physique du transistor. Si un grand nombre de méthodologies permettent d'extraire des paramètres depuis les caractéristiques courant-tension, nous allons voir que la construction d'une technique robuste pose aujourd'hui un certain nombre de problèmes.

⋆ Dans une première partie, nous détaillerons les outils qui permettent de développer la caractérisation électrique dans le milieu industriel.

⋆ La deuxième partie détaillera les modèles sur lesquels s'appuient les techniques d'extraction et dont la validité conditionne directement la viabilité de l'extraction.

⋆ Les différentes méthodes de la littérature qui permettent d'extraire la tension de seuil (V_{th}) seront détaillées dans une troisième partie. En particulier nous y détaillerons la nouvelle méthode d'extraction sous le seuil qui a été développée durant cette thèse.

⋆ Nous aborderons ensuite la description des méthodes d'extraction qui permettent d'extraire tous les paramètres du modèle de courant. Nous porterons en particulier notre attention sur les techniques construites autour de la fonction Y, qui fournissent des résultats peu dépendants de la résistance série du transistor.

⋆ La robustesse des techniques d'extraction présentées dans ce chapitre sera analysée sur deux critères : l'immunité des résultats par rapport aux variations de R_{sd} et l'immunité des résultats par rapport à un décalage de la plage d'extraction.

⋆ Finalement, les méthodes d'extraction de la résistance série seront détaillées dans une dernière partie. Nous y présenterons la nouvelle méthode $R_{tot}(1/\beta)$ qui a été développée durant cette thèse et offre une précision accrue.

III.1 Les outils nécessaires à la caractérisation électrique en milieu industriel

La caractérisation électrique des transistors MOS dans un contexte industriel nécessite des mesures statistiques, qui permettent d'estimer la moyenne d'un paramètre avec une certaine incertitude pour un intervalle de confiance donné. Dans cette thèse, des bancs de test entièrement automatisés (de type Accretech UF3000) équipé d'une matrice de connexion Agilent 4073B et pilotés par un système d'exploitation propriétaire sont utilisés. Cette configuration permet d'exécuter des tests électriques à large échelle (plusieurs puces et jusqu'à 25 plaques) sur plusieurs structures de test ayant des connectiques différentes. L'essentiel des mesures est réalisé sur des plaques de 300mm (12 pouces) de diamètre composées en moyenne de 65 puces qui, lorsque cela possible, sont toutes testées afin de déterminer les paramètres avec un intervalle de confiance le meilleur possible. Les mesures capacitives sont effectuées à moyenne fréquence ($f \sim$ 500kHz - 1MHz) et réalisées grâce à un capacimètre HP4284 (ou HP4980). Les mesures de courant sont effectuées grâce à un système de mesure Agilent 4073B. Le cout de ces équipements est élevé et le temps de test doit être optimisé au maximum, dans un compromis qui oppose la précision et le temps de mesure.

Outre la mesure, des outils de traitement sont également nécessaires afin de permettre une automatisation des méthodes d'extraction. Le logiciel SciLab a ainsi été massivement utilisé pour le traitement des courbes (courant-tension ou capacitives) au cours de cette thèse. Une interface utilisateur a été crée afin d'avoir un maximum d'ergonomie dans l'utilisation et l'application créée permet le traitement d'un nombre de courbes important en utilisant une large panoplie de méthodes d'extraction (il est par exemple possible de traiter jusqu'à 20000 courbes automatiquement).

Finalement, l'intégration d'une méthode d'extraction sous forme de test en ligne constitue la dernière étape d'un développement de méthodologie d'extraction. Elle permet effectivement d'extraire directement le paramètre sans avoir à mesurer une courbe entière qui sera post-traitée ultérieurement.

III.2 Le cœur de l'extraction : un modèle de courant adapté au comportement du transistor MOS moderne en régime linéaire et en inversion forte

L'extraction des paramètres électriques nécessite de se rapporter à un modèle analytique qui (idéalement) rend compte du comportement du transistor. Il faut donc trouver les équations qui respectent les critères suivants :

- les paramètres du modèle doivent conserver une signification physique afin de pouvoir analyser et comprendre le comportement électrique du MOSFET,
- Les équations et algorithmes de calcul mis en jeu doivent être simples afin d'envisager l'automatisation de la technique d'extraction.

Le modèle de courant simplifié que nous utilisons est basé sur le modèle semi-empirique BSIM3v3 [Cheng97b]. Il permet de reproduire fidèlement les caractéristiques I_d-V_{gs} des transistors MOS tout en conservant une simplicité mathématique qui facilite la construction des

méthodes d'extraction. Les paramètres qui le composent ont une signification physique et permettent ainsi d'analyser les propriétés de transport et le comportement électrostatique du transistor. L'équation de ce modèle est présentée en (III.1), elle est construite autour de 4 paramètres dont la description est donnée ci-dessous :

- la tension de seuil (V_{th}) : elle définit la tension pour laquelle le canal du transistor passe de l'inversion faible à l'inversion forte ($\psi_s = 2\phi_F$, cf. §I.1.2) et respecte la relation $Q_{inv} = C_{\text{eff}} \cdot (V_{\text{gs}} - V_{\text{th}})$ en inversion forte. La plupart du temps nous utiliserons la variable réduite $V_{\text{gt}} \stackrel{\text{def}}{=} V_{gs} - V_{th}$ pour simplifier l'écriture des équations ;
- les dimensions effectives du canal W_{eff} et L_{eff} qui représentent la largeur et la longueur de la couche d'inversion (cf. § IV.4) ;
- La capacité effective d'oxyde en inversion forte C_{eff}, dont la définition est donnée dans le chapitre I.
- la mobilité effective μ_{eff} qui quantifie la capacité des porteurs à se mouvoir et est fonction du champ effectif transverse $\mathcal{E}_{eff}$ (cf. §II.1.2).

$$I_d = \mu_{\text{eff}} \cdot C_{\text{eff}} \cdot \frac{W_{\text{eff}}}{L_{\text{eff}}} \cdot (V_{\text{gs}} - V_{\text{th}}) \cdot V_{ds} \tag{III.1}$$

La variation de mobilité en avec Q_{inv} peut être mise sous la forme d'une série entière et μ_{eff} peut être écrite sous la forme (III.2).

$$\mu_{eff}(Q_{inv}) \simeq \frac{\mu_0}{\sum_{i=0}^{N} \lambda_i \cdot (Q_{inv})^i}, \quad \lambda_0 = 1 \tag{III.2}$$

Le paramètre μ_0 est alors défini comme la valeur de la mobilité en inversion forte extrapolée à $Q_{inv} = 0$ (III.3). Il est appelé "mobilité à bas champ" et est couramment utilisé comme un indicateur de qualité du transport dans le canal du transistor.

$$\mu_0 \stackrel{\text{def}}{=} \lim_{Q_{inv} \to 0^+} (\mu_{eff}) \tag{III.3}$$

En général, la modélisation de $\mu_{eff}(Q_{inv})$ se limite à l'ordre 1 ou 2 et au régime d'inversion forte dans lequel la charge d'inversion peut être approximée par : $Q_{inv} \simeq C_{eff} \cdot (V_{gs} - V_{th})$. Cela permet de réécrire (III.2) en fonction de variables directement accessibles par la mesure de courant (contrairement à Q_{inv} qui nécessite des mesures capacitives). La capacité en inversion est alors intégrée dans les coefficients de la série entière en effectuant le changement de variable suivant : $\Theta_i = \lambda_i / C_{eff}, i = 1, 2$ et l'expression de μ_{eff} en fonction de V_{gt} est donnée par (III.4)[1]. Ce changement de variable fait apparaître C_{eff} dans l'expression des paramètres Θ_i, conformément à [Reichert96, Gautier03].

$$\mu_{eff}(V_{gt}) \simeq \frac{\mu_0}{1 + \Theta_1 \cdot V_{gt} + \Theta_2 \cdot {V_{gt}}^2} \tag{III.4}$$

Finalement, les paramètres μ_0, C_{eff}, L_{eff} et W_{eff} sont regroupés dans une variable unique β, appelé paramètre de gain en transconductance du transistor, qui définit la quantité de courant que pourra débiter ce dernier sous l'application d'un couple de polarisation (V_{gt},V_{ds}), cf. relation (III.5)

1. Le lecteur pourra trouver plus d'informations sur la nature physique des paramètres Θ_1 et Θ_2 dans [Reichert96]

$$\beta \overset{\text{def}}{=} \mu_0 \cdot C_{eff} \cdot \frac{W_{eff}}{L_{eff}} \tag{III.5}$$

L'expression du courant I_d peut alors être réécrite sous une forme qui facilite l'extraction des paramètres électriques (III.6).

$$I_d = \frac{\beta \cdot V_{ds} \cdot V_{\text{gt}}}{1 + \Theta_1 \cdot V_{\text{gt}} + \Theta_2 \cdot {V_{\text{gt}}}^2}, \quad V_{gs} \gg V_{th} \tag{III.6}$$

Remarque : Le modèle (III.4) n'est valable qu'en régime d'inversion forte. Il ne permet pas de prendre en compte les contributions de mobilité qui caractérisent les régimes d'inversion modérée ou faible. C'est en particulier les domaines dans lequel les collisions avec les centres chargés sont les plus influents (cf. §II.2.2).

III.2.1 Influence et correction de la fuite de grille

La densité (surfacique) de courant de grille augmente exponentiellement avec la diminution de l'épaisseur du diélectrique [Sze81]. Lorsque la surface équivalente est grande, la fuite de grille ne peut plus être négligée sur les caractéristiques I_d-V_{gs} : une correction est nécessaire. C'est en particulier le cas pour les transistors de grande surface ($W_{\text{eff}} \times L_{\text{eff}}$) ou pour les structures en batterie à grille commune (N transistors connectés en parallèle). Dans ce cas, le courant mesuré doit être corrigé selon (III.7), où α est le coefficient de partage de grille défini dans [Romanjek04a]. Cette relation permet une correction de manière à retrouver un courant de drain représentatif du canal et non parasité par les fuites à travers le diélectrique, cf. figure III.1. Le principe du partage de grille est présenté dans la partie de droite de la même figure.

$$I_d \simeq \alpha \, (I_{d,meas} - I_{s,meas}) \tag{III.7}$$

Pour des tensions V_{ds} très faibles, la couche d'inversion peut être supposée uniforme dans le canal et $\alpha \simeq 1/2$ [Romanjek03, Romanjek04a]. Cette correction sera systématiquement appliquée à nos mesures afin de s'affranchir de l'influence de la fuite de grille dans l'extraction.

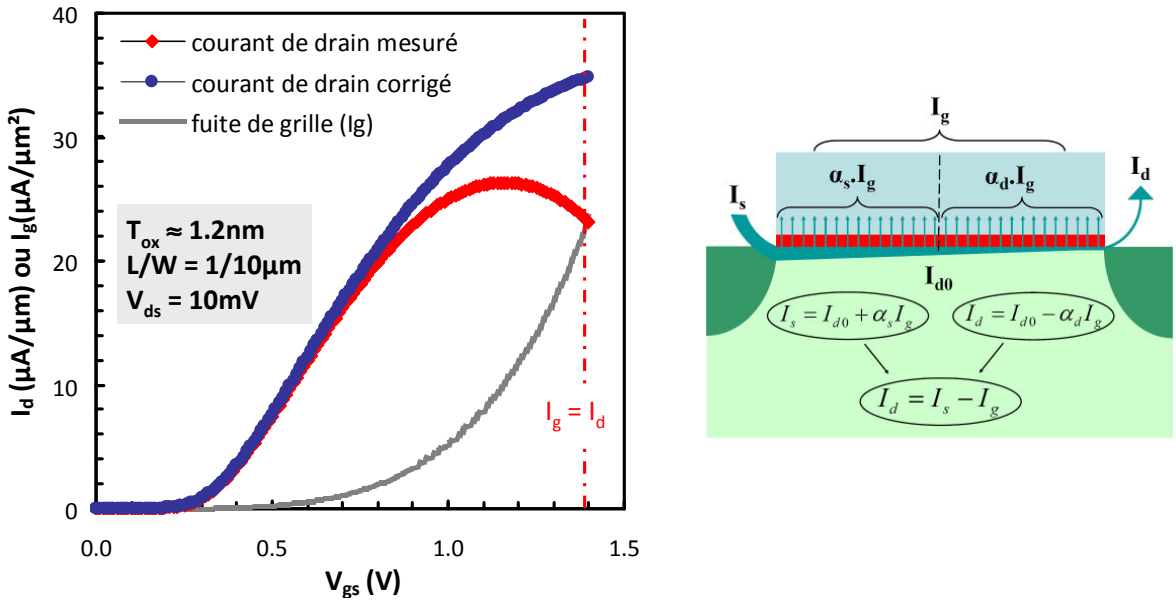


Fig. III.1 – A gauche : correction du courant de grille en utilisant la relation (III.7) sur un transistor à oxyde de grille mince ($T_{ox} \simeq 12$Å) de grande surface ($W \times L = 10\mu\text{m}^2$). A droite : illustration du mécanisme de partage de courant de grille, d'après [Romanjek04a].

III.2.2 Prise en compte de la résistance série dans le modèle

Le modèle en (III.6) permet de reproduire la caractéristique $I_d(V_{\text{gs}})$ d'un transistor MOS en l'absence des résistances séries ($R_{sd} = 0$). La chute de potentiel sur V_{ds} que provoque le passage du courant dans les résistances séries du transistor peut être prise en compte en effectuant le changement de variable $V_{ds} \rightarrow (V_{ds} - R_{sd}(V_{gs}) \cdot I_d)$ dans les équations. La relation (III.6) peut alors s'écrire d'une manière similaire (III.8), où le terme additionnel $\beta \cdot R_{sd}(V_{gs})$ rend compte de l'influence de R_{sd} sur le courant effectif $I_{d,eff}$.

$$I_{d,eff} = \frac{\beta \cdot V_{ds} \cdot V_{\text{gt}}}{1 + [\Theta_1 + \beta \cdot R_{sd}(V_{gs})] \cdot V_{\text{gt}} + \Theta_2 \cdot {V_{\text{gt}}}^2} \tag{III.8}$$

Remarquons que la prise en compte de R_{sd} en régime linéaire revient à appliquer la règle de proportionnalité en (III.9), où $I_{d,eff}$ représente le courant prenant en compte l'influence des résistances séries et I_d représente le courant intrinsèque du canal.

$$I_{d,eff} = I_d \times \frac{V_{\text{ds}}}{V_{\text{ds}} + I_d \cdot R_{sd}(V_{gs})} \tag{III.9}$$

Si les variations de R_{sd} avec V_{gs} se limitent au premier ordre (en accord avec [Kim02a, Cros05], cf. figure III.2), l'expression analytique de R_{sd} peut s'écrire comme (III.10).

$$R_{sd} = R_{sd,0} + \lambda_{R_{sd}} \cdot V_{\text{gs}} \tag{III.10}$$

Les composantes $R_{sd,0}$ et $\lambda_{R_{sd}}$ de R_{sd} peuvent alors être intégrées dans les paramètres modifiés $\Theta_{1,eff}$ et $\Theta_{2,eff}$ [Ghibaudo88, Cros05, Tanaka05, Fleury08a] (III.12), permettant d'écrire (III.8) sous une forme facilement exploitable (III.11).

$$I_{d,eff} = \frac{\beta \cdot V_{\text{ds}} \cdot V_{\text{gt}}}{1 + \Theta_{1,eff} \cdot V_{\text{gt}} + \Theta_{2,eff} \cdot {V_{\text{gt}}}^2} \tag{III.11}$$

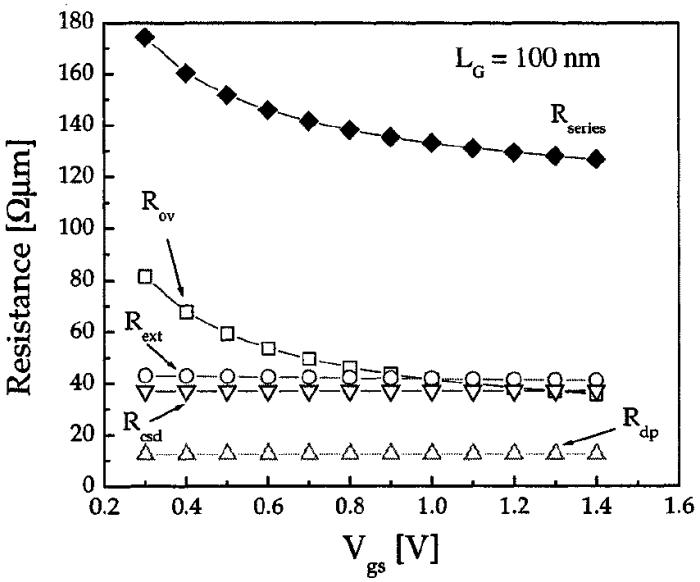


Fig. III.2 – Variation de la résistance série R_{sd} et de ses composantes en fonction de la tension de grille $V_{\rm gs}$. La résistance de recouvrement (R_{ov}) est la contribution majeure qui provoque la variation de R_{sd} avec $V_{\rm gs}$, d'après [Kim02b]

$$\begin{cases} \Theta_{1,eff} & = \Theta_1 + \beta \cdot R_{sd,0} \\ \\ \Theta_{2,eff} & = \Theta_2 + \beta \cdot \lambda_{R_{sd}} \end{cases} \tag{III.12}$$

Nous verrons par la suite que la variation $R_{sd}(V_{gs})$ peut être extraite à partir des paramètres $\Theta_{1,eff}$, $\Theta_{2,eff}$ et β grâce à (III.12).

III.2.2.a Influence de la résistance série sur la tension V_{gs}

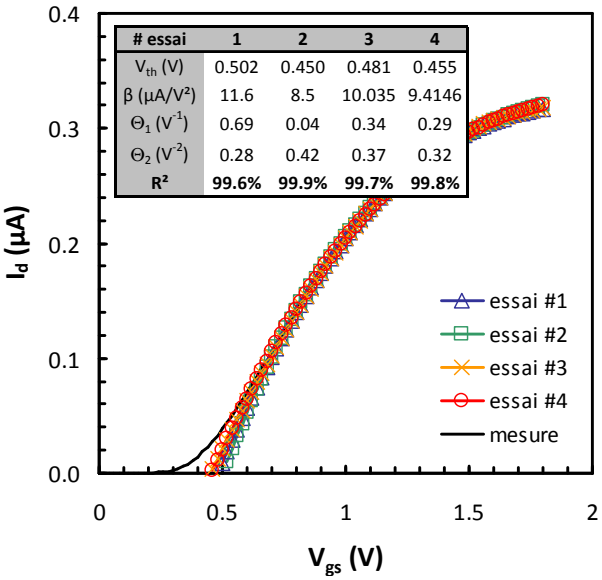
La résistance de source R_s provoque une chute de potentiel sur la tension V_{gs} appliquée. De la même manière que pour l'influence de R_{sd} sur V_{ds}, l'influence de R_s peut être prise en compte en effectuant la substitution suivante des les équations du courant : $V_{gs} \rightarrow (V_{gs} - R_s(V_{gs}) \cdot I_d)$. Pour simplifier les calculs, l'influence des résistances séries sur le courant peut se limiter au premier ordre et on peut considérer que le produit des chutes relatives de potentiel est petit devant 1, cf. (III.13).

$$\left(\frac{R_s \cdot I_d}{V_{gt}}\right) \cdot \left(\frac{R_{sd} \cdot I_d}{V_{ds}}\right) \ll 1 \tag{III.13}$$

Si (III.13) est justifiée, la prise en compte de la chute de potentiel sur V_{gs} revient alors à réécrire (III.8) par (III.14).

$$I_{d,eff} \simeq \frac{\beta \cdot V_{\rm ds} \cdot V_{\rm gt}}{1 + \beta \cdot R_s(V_{gs}) \cdot V_{ds} + \Theta_{1,eff} \cdot V_{\rm gt} + \Theta_{2,eff} \cdot {V_{\rm gt}}^2} \tag{III.14}$$

En général, l'influence du terme $R_s \cdot V_{ds}$ est négligeable en régime linéaire, pour de faibles tensions de drain.



# essai	1	2	3	4
V_{th} (V)	0.502	0.450	0.481	0.455
β (µA/V²)	11.6	8.5	10.035	9.4146
Θ_1 (V^{-1})	0.69	0.04	0.34	0.29
Θ_2 (V^{-2})	0.28	0.42	0.37	0.32
R^2	99.6%	99.9%	99.7%	99.8%

Fig. III.3 – Modélisation de la caractéristique $I_d(V_{gs})$ d'un transistor nMOS en régime linéaire, dans la zone d'inversion forte. Plusieurs combinaisons de paramètres (V_{th}, β, Θ_i) permettent de reproduire les variations du courant avec $R^2 \geq 99,6\%$.

III.2.3 Discussion : les limites de l'extraction en inversion forte

Gardons à l'esprit que le modèle de courant en inversion forte (III.6) ne permet pas de modéliser le maximum de transconductance (pas de point d'inflexion dans le modèle). Mathématiquement, il y a donc une infinité de combinaisons de paramètres $(\beta, V_{th}, \Theta_i)$ qui permettent d'ajuster le modèle sur une même courbe expérimentale $I_d(V_{gs})$ tout en offrant un coefficient de régression R^2 proche de 1 [Press07]. Ce point peut être facilement démontré en effectuant une optimisation non linéaire du modèle sur une mesure $I_d(V_{gs})$. La recherche de l'erreur minimale peut être réalisée par la biais d'un algorithme génétique, programmé sous le logiciel SciLab [INRIA]. Afin d'obtenir des valeurs *raisonnables*, des limites hautes et basses sont imposées pour chaque variation de paramètre :

- $0.5\max(g_m) \leq \beta \leq 2\max(g_m)$, où $\max(g_m)$ est le maximum de transconductance qui est proche de la valeur de β sur un transistor long, dans lequel la résistances séries à peu d'influence,
- $V_{th,ext} - 100\text{mV} \leq V_{th} \leq V_{th,ext} + 100\text{mV}$, où $V_{th,ext}$ est la valeur de V_{th} obtenue par la méthode de l'extrapolation en inversion forte à partir du $\max(g_m)$,
- $0 \leq \Theta_i \leq 1\text{V}^{-i}, \quad i = 1, 2$: la mobilité ne peut pas augmenter lorsque V_{gt} croit.

La figure III.3 montre ainsi 4 ajustements obtenus grâce à la méthode d'optimisation décrite ci-dessus et en respectant les conditions imposées sur chaque paramètre. Bien différentes, toutes les combinaisons de paramètres permettent de reproduire les variations du courant du transistor tout en offrant une valeur de $R^2 \sim 1$. Cela signifie qu'une simple régression non linéaire ne suffit pas à retirer l'information physique cachée dans la courbe $I_d(V_{gs})$. Le modèle doit donc être exploité de manière *intelligente* afin d'extraire les informations physiques sur le dispositif.

III.3 Les méthodes d'extraction de la tension de seuil (V_{th})

La définition de la tension de seuil est donnée au §I.1.2. Ce paramètre correspond conventionnellement à la tension de grille $V_{gs} = V_{\mathrm{th}}$ pour laquelle la couche d'inversion apparaît dans le canal du transistor. Elle est un paramètre clé dans la modélisation et la caractérisation des transistors MOS car elle représente la tension de grille à laquelle s'effectue la transition entre l'inversion faible et l'inversion forte dans le canal. Si la littérature compte un grand nombre de techniques pour mesurer V_{th}, la plupart est basée sur la caractéristique courant-tension $I_d(V_{\mathrm{gs}})$ du transistor. Une évaluation des méthodes d'extraction a déjà été faite dans [Ortiz-Conde02, Liou97, Dobrescu00, Tsuno99], ainsi, nous nous bornerons uniquement à l'étude des techniques qui présentent un intérêt pour le travail effectué durant la thèse.

De notre point de vue, pour être précise, une procédure d'extraction du V_{th} doit respecter les critères suivants :

- indépendance des résultats vis-à-vis des variations de la résistance série,
- renvoyer des résultats cohérents et stables pour toutes les longueurs de canal,
- indépendance vis-à-vis de la mobilité des porteurs (μ_{eff}).

De plus, pour être industrialisable, l'extraction de V_{th} doit être adaptée à l'utilisation de mesures rapides, excluant l'utilisation des techniques capacitives qui requièrent un long temps d'intégration et éventuellement des structures de test spécifiques.

III.3.1 L'extraction de V_{th} par mesures capacitives

Les mesures capacitives permettent d'extraire la tension de seuil sur des MOSFET. L'intérêt de cette technique de mesure est qu'elle ne nécessite pas l'application d'un V_{ds} et, comme aucun courant ne circule entre la source et le drain. Comme nous allons le voir, les résultats sont insensibles à la valeur des résistances série.

III.3.1.a Extraction de V_{th} depuis la dérivée première de la capacité

La mesure de la capacité grille-canal C_{gc} (cf. III.16) permet d'évaluer un valeur de la tension de seuil proche de la définition théorique donnée en (I.1), conformément à [Sze81, Sodini82, Wong05] (III.15).

$$V_{gs} = V_{th,2\phi_F} \Leftrightarrow \psi_s = 2\phi_f \tag{III.15}$$

$$C_{gc} = \frac{\partial Q_{inv}}{\partial V_{\mathrm{gs}}} \tag{III.16}$$

Le maximum de la dérivée dC_{gc}/dV_{gs} est un indicateur relativement fiable du point de polarisation pour lequel la condition $\psi_s \simeq 2\phi_f$ est atteinte. Lau et al. ont également proposé d'utiliser l'écrantage de la capacité de désertion caractéristique de la capacité grille-substrat C_{gb} (cf. III.17) pour extraire V_{th} [Lau99, Lau01]. Le maximum de la pente dC_{gb}/dV_{gs} dans cette zone d'écrantage résulte de l'apparition des charges d'inversion et correspond donc à la même définition de V_{th} que celle utilisée par la méthode basée sur la mesure de la pente maximale ($\max(dC_{gc}/dV_{gs})$). L'extraction du $V_{th,2\phi_F}$ par ces deux techniques est présentée sur la figure III.4. Sans surprise les valeurs sont identiques pour les deux méthodes. En revanche, la valeur

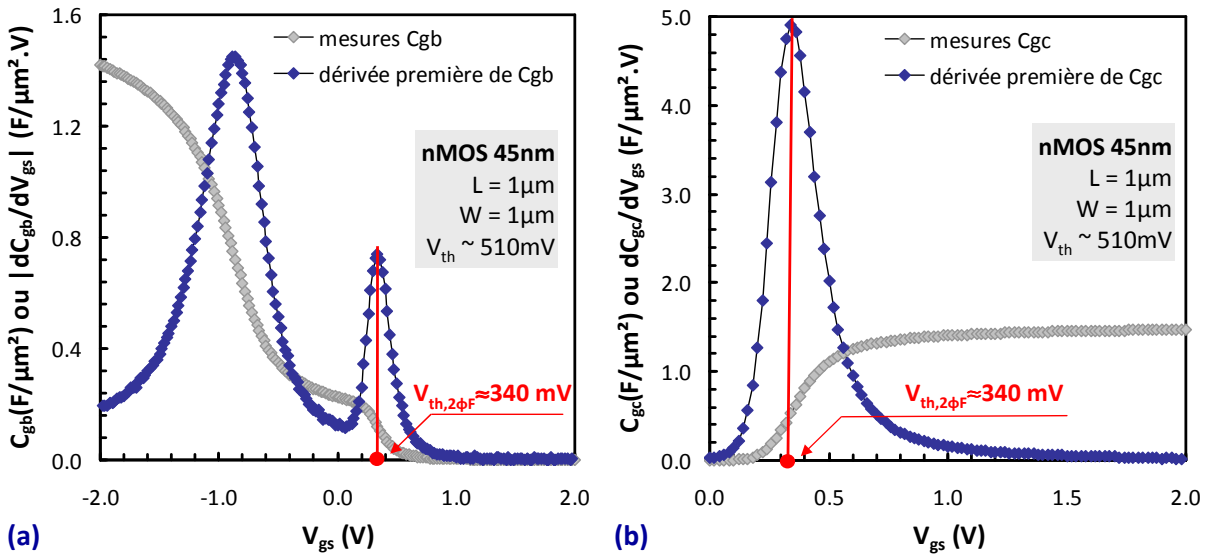


Fig. III.4 – Extraction de la tension de seuil par mesure capacitive, **(a)** en utilisant la méthode de la dérivée première de la capacité grille-substrat dC_{gb}/dV_{gs} [Lau99, Lau01] et **(b)** en utilisant la méthode de la dérivée première de la capacité grille-canal dC_{gc}/dV_{gs}.

de $V_{th,2\phi_F}$ sous-estime le V_{th} issu de l'extraction depuis les courbes $I_d(V_{gs})$ de 100 à 150mV, en accord avec les résultats dans [Sodini82, Lau01].

$$C_{gb} = -\frac{\partial Q_{sub}}{\partial V_{\mathrm{gb}}} \tag{III.17}$$

III.3.1.b Extraction de V_{th} depuis la mesure de la charge d'inversion

Il est également possible d'extraire V_{th} connaissant l'évolution de la charge d'inversion qui peut être extraite depuis des mesures de capacité grille-canal C_{gc} [Sodini82, Liou97]. Cette technique renvoie une valeur plus proche des valeurs de V_{th} extraites depuis les courbes $I_d(V_{gs})$ car elle consiste à extrapoler la valeur de V_{gs} à charge nulle ($Q_{inv} = 0$) à partir des mesures en inversion forte : $Q_{inv} \simeq C_{eff} \cdot (V_{gs} - V_{th})$. L'extraction est présentée sur la figure III.5. Le V_{th} extrait est sensiblement plus élevé que le résultat issu de la courbe $I_d(V_{gs})$ associée ($\Delta V_{th} \simeq 40mV$).

III.3.1.c Avantages et inconvénients de l'extraction de V_{th} par mesure capacitive.

Si l'extraction de V_{th} basée sur les mesures capacitives semble renvoyer des résultats précis sans être influencée par la valeur des résistances séries ou par la décroissance de la mobilité avec V_{gs}, ces méthodes ont des inconvénients majeurs :

- La mesure d'une capacité nécessite des temps de test beaucoup plus longs que pour des courants, d'où la difficulté d'un transfert industriel sous forme de test automatique ;
- Les mesures de C_{gb} et de C_{gc} sont très bruitées sur des transistors sub-100nm qui nécessitent une correction des capacités parasites inhérentes à la structure MOS. Cela impose donc des temps de mesure longs et un post-traitement complexe pour effectuer la correction, qui est impossible sur des petites surfaces ;
- Les valeurs extraites depuis les techniques capacitives ne sont pas en ligne avec les méthodes basées sur la mesure $I_d(V_{gs})$;

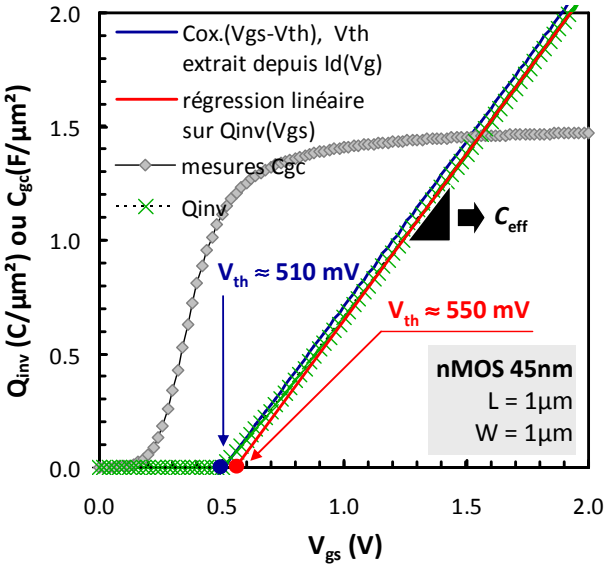


Fig. III.5 – Extraction de la tension de seuil depuis la courbe $Q_{inv}(V_{gs})$. La valeur extraite par ce biais est légèrement plus élevée que celle issue de l'extraction depuis la courbe $I_d(V_{gs})$ associée.

– Pour les méthodes basées sur la dérivée première de la capacité, le pas en V_{gs} doit être choisi dans un compromis avec le temps de mesure et la sensibilité de la mesure. Il est souvent grand (20-50mV) et limite la précision sur la valeur de V_{th} extraite.

L'extraction de V_{th} sur les caractéristiques $I_d(V_{gs})$ est donc généralement préférée en caractérisation électrique. Leur description fera l'objet des prochains paragraphes.

III.3.2 La méthode du courant constant

La méthode d'extraction du $V_{\rm th}$ par la méthode du courant constant consiste à évaluer la valeur de la tension de grille V_{gs} pour laquelle le courant atteint un seuil (arbitraire) I_{cc}, représentatif du passage entre l'inversion faible et l'inversion forte. Historiquement cette valeur fut définie par la relation (III.18) qui nécessite la connaissance de la longueur effective pour pouvoir extraire une tendance $V_{th}(L_{eff})$ précisément.

$$I_{cc} \simeq 100\text{nA} \times \frac{W_{eff}}{L_{eff}} \tag{III.18}$$

La technique du courant constant possède pourtant des avantages indéniables devant les autres techniques :

– elle utilise la gamme de courant proche du seuil, peu dépendante de la valeur des résistances séries (à l'inverse de la zone de forte inversion utilisée par beaucoup d'autres techniques),
– le seuil de courant I_{cc} peut être ajusté de manière à renvoyer une valeur en ligne avec des méthodes plus précises (fonction Y),
– elle permet une estimation rapide de V_{th} et peut être utilisée pour des tests à grande échelle (simple mesure de courant) car elle repose sur la mesure d'un niveau de courant élevé, en inversion modérée.

III.3.3 La méthode du V_{th} extrapolé depuis le maximum de transconductance

La méthode du V_{th} depuis le maximum de transconductance consiste à supposer que le maximum de la pente de la caractéristique $I_d(V_{gs})$ correspond au paramètre β et qu'en ce point,

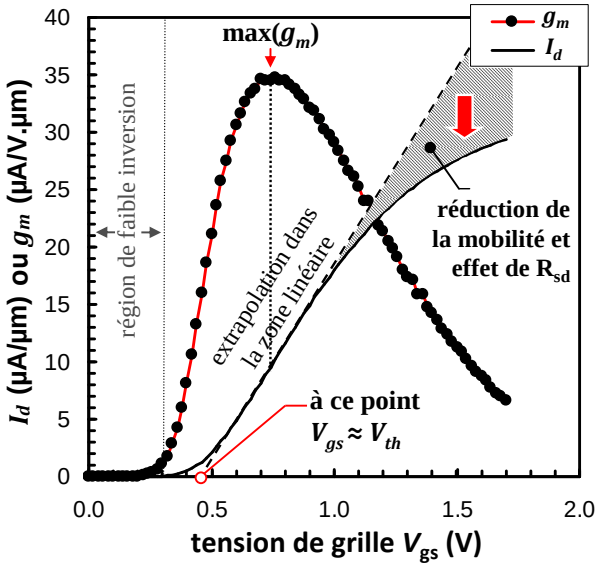


Fig. III.6 – Illustration de l'influence des résistances séries sur l'extraction de la tension de seuil en utilisant la méthode du V_{th} extrapolé depuis l'inversion forte.

le modèle peut être simplifié par une relation linéaire (modèle de mobilité III.2 à l'ordre 0 : Θ_1 et Θ_2 sont négligés (III.19).

$$g_m = \max(g_m) \Leftrightarrow I_d \simeq \beta \cdot V_{ds} \cdot (V_{gs} - V_{th}) \tag{III.19}$$

La valeur du V_{th} peut alors être extraite simplement, à partir du point d'intersection de la régression linéaire effectuée sur l'$I_d(V_{gs})$[2], dans la zone ou la pente maximale a été repérée, cf. figure III.6.

En pratique, cette technique d'extrapolation en inversion forte ne prend pas en compte la réduction de la mobilité et l'influence de la résistance série, qui influent sur la valeur $\max(g_m)$. Ainsi, la valeur de β obtenue par cette technique est largement sous-estimée (d'environ 50% pour un dispositif de longueur nominale issu la filière CMOS 45nm) et la valeur de V_{th} extraite est généralement sous-estimée. Cette méthode fournit néanmoins un moyen efficace d'extraire une faible différence de V_{th} entre deux dispositifs aux comportements proches, comme dans les applications de variabilité [Cathignol08].

III.3.4 Les méthodes utilisant deux dérivations successives du courant

Les méthodes utilisant des fonctions basées sur des dérivées de deuxième ordre sont appréciées car elle permettent une implémentation simple et, dans certains cas, s'affranchissent de l'influence de la résistance série et de la réduction de mobilité avec V_{gs} [McLarty95]. Nous allons présenter les deux méthodes les plus courantes qui utilisent une dérivation à l'ordre deux pour extraire V_{th}.

III.3.4.a La méthode de la dérivée seconde

La méthode de la dérivée seconde [Wong87] se base sur un modèle de courant simplifié qui prédit qu'en régime linéaire, I_d doit être nul sous le seuil et être proportionnel en V_{gt} pour

2. historiquement le point d'intersection retourne la valeur $V_{th} - V_{ds}/2$ mais notre convention inclue l'influence du V_{ds} directement dans la valeur de V_{th}

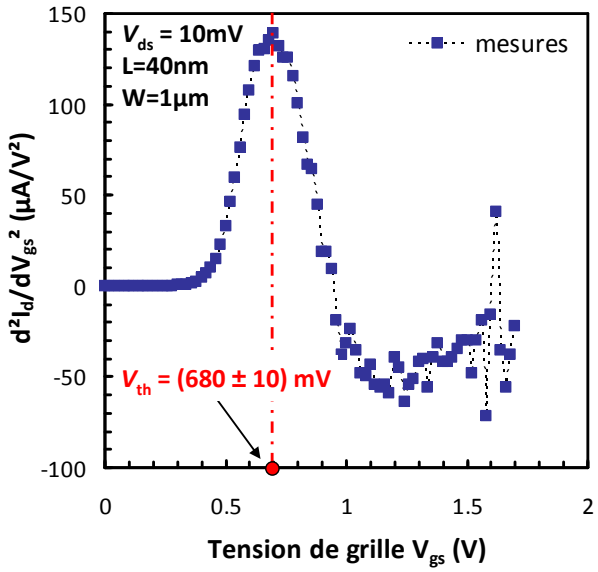


Fig. III.7 – Extraction de la tension de seuil par la méthode de la dérivée seconde. L'amplification du bruit de mesure liée aux deux dérivations successives ainsi que le pas de mesure limitent la précision de cette technique.

$V_{gs} > V_{th}$. La dérivée dI_d/dV_{gs} doit donc être une fonction échelon et sa dérivée doit se comporter comme un Dirac dont le pic indique la valeur de V_{th} (III.20). Il est donc proposé d'utiliser la dérivée seconde du courant pour extraire la tension de seuil qui sépare le régime sous le seuil (qui se comporte exponentiellement) et la partie de la courbe au delà du seuil (qui se comporte linéairement avec V_{gt}) [Wong87].

$$\frac{d^2 I_d}{dV_{gs}{}^2} = \begin{cases} +\infty, & V_{gs} = V_{th} \\ 0, & V_{gs} \neq V_{th} \end{cases} \tag{III.20}$$

La dérivée seconde du courant en fonction de V_{gs} est représentée sur la figure III.7 pour un transistor issu de la technologie 45nm (L_{eff} = 30nm). L'extraction n'est pas aussi simple que dans la théorie et la précision sur le résultat est principalement dégradé par :

- l'amplification du bruit de mesure causée par les deux dérivations successives,
- la limitation sur la précision du résultat imposé par le pas de mesure fixe, qui ne peut pas être réduit sans augmenter fatalement le temps de mesure.
- le modèle simplifié sur lequel repose la méthode ne prend pas en compte l'influence des résistances séries et la réduction de la mobilité (Θ_1 et Θ_2 dans le modèle). Les résultats sont donc dépendants des valeurs de ces paramètres.

III.3.4.b La méthode de la dérivée seconde de $\log(I_d)$

Une variante de la méthode présentée dans le paragraphe précédant (III.3.4.a) consiste à effectuer les deux dérivations successives sur le logarithme du courant $I_d = \log(I_d)$ [Aoyama95, Dobrescu00], cf. figure III.8. Elle ne présente pas un grand avantage par rapport à la méthode décrite précédemment et possède les mêmes inconvénients concernant la précision d'extraction. Nous n'aborderons pas la description de méthodes utilisant trois dérivations successives du courant de drain [Wong01] qui sont en général inapplicables à des mesures *réelles*.

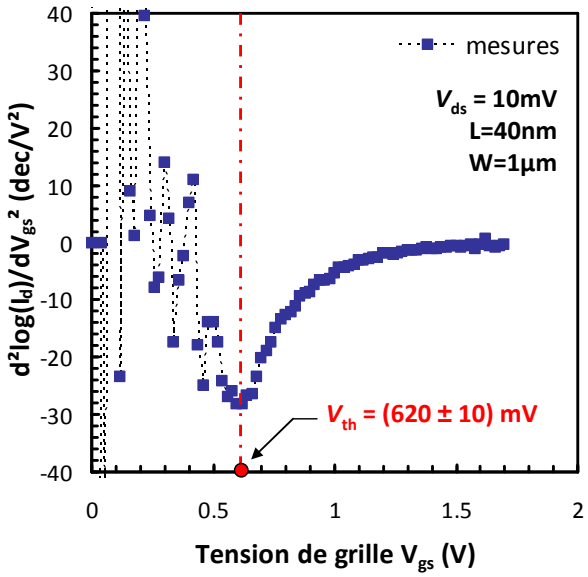


Fig. III.8 – Extraction de la tension de seuil par la méthode de la dérivée seconde du logarithme du courant [Aoyama95]. Comme pour la technique décrite au paragraphe III.3.4.a, l'amplification du bruit de mesure liée aux deux dérivations successives ainsi que le pas de mesure (en V_{gs}) limitent fortement la précision sur V_{th}.

III.3.5 Proposition d'une nouvelle méthode : le V_{th} extrapolé en inversion faible

Les méthodes d'extraction à courant constant et $V_{\rm th}$ extrapolé en inversion forte sont préférées pour les test en ligne de production, la calibration et les hypothèses nécessaires sont parfois négligées. Une nouvelle méthode a donc été développée durant la thèse. Elle est certe moins rapide et automatisable que la technique de l'extrapolation en inversion forte car elle nécessite la mesure d'un niveau de courant plus faible (inversion faible) mais fournit une précision et une fiabilité accrues. Cette technique utilise à la fois les régimes sous le seuil et au delà du seuil. Elle utilise les mêmes bases que la méthode du quasi-courant constant développée en [Yan91], mais permet, par une utilisation judicieuse de l'équation du courant sous le seuil de limiter l'influence des résistances séries sur la valeur de V_{th}.

III.3.5.a Description de la méthodologie

Le courant de drain en inversion faible est donné par la relation (III.21), où C_{dep} est la capacité de désertion, $C_{\rm eff}$ la capacité effective en inversion et A un rapport de capacités qui peut être calculé comme $A \simeq C_{\rm eff}/\left(C_{\rm eff}+C_{dep}\right)$ si l'influence des défauts d'interface est négligée [Ghibaudo97, Sze81]. L'exposant ^{wi} (pour *weak inversion*) réfère au domaine d'inversion faible de la caractéristique $I_d(V_{gs})$ ($V_{\rm gs} \ll V_{\rm th}$).

$$I_d^{wi} = \beta \cdot \frac{C_{dep}}{C_{\rm eff}} \left(\frac{kT}{q}\right)^2 \left[1-\exp\left(-\frac{qV_{\rm ds}}{kT}\right)\right] \cdot \exp\left(\frac{qA}{kT}V_{\rm gt}\right) \qquad \text{(III.21)}$$

Le paramètre A est lié à la pente sous le seuil du dispositifs et peut être extrait grâce à la relation (III.22), tandis que le paramètre Γ est une constante de la faible inversion définie par la relation (III.23).

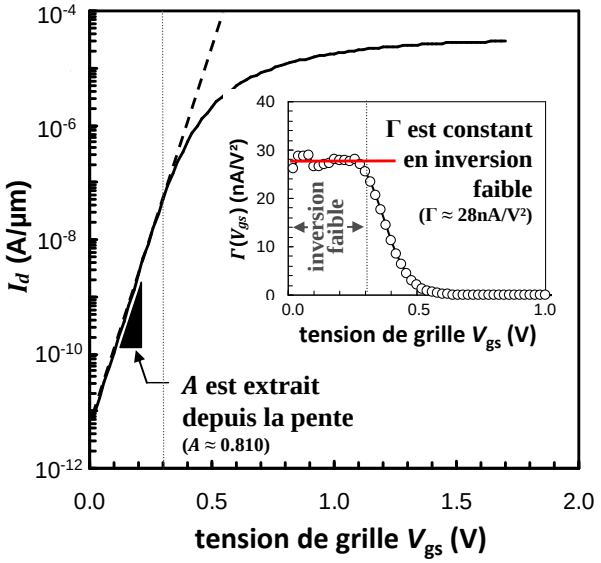


Fig. III.9 – Extraction des paramètres A et Γ sous le seuil, en régime de faible inversion.

$$A = \frac{kT}{q} \cdot \frac{\partial \ln\left(I_d^{wi}\right)}{\partial V_{\mathrm{gs}}} \tag{III.22}$$

$$\Gamma \stackrel{\mathrm{def}}{=} \frac{I_d \cdot \exp\left(-\dfrac{qA}{kT} V_{\mathrm{gs}}\right)}{\dfrac{C_{dep}}{C_{\mathrm{eff}}} \cdot \left(\dfrac{kT}{q}\right)^2 \cdot \left[1 - \exp\left(-\dfrac{qV_{\mathrm{ds}}}{kT}\right)\right]} \tag{III.23}$$

L'utilisation de (III.22) et (III.23) permet d'exprimer V_{th} en fonction des paramètres A, Γ et β (qui reste la seule inconnue dans le système), cf. (III.24)

$$V_{\mathrm{th}} = \frac{kT}{qA} \cdot \ln\left(\frac{\beta}{\Gamma}\right) \tag{III.24}$$

L'intérêt de cette méthode réside dans la présence de la fonction logarithme qui intervient dans le calcul de V_{th}. En effet, un calcul d'erreur montre que la corrélation entre β et V_{th} est fortement minimisée en comparaison de la méthode du V_{th} extrapolé en inversion forte. Un estimateur grossier de la valeur de β, tel que le maximum de transconductance ($\max\left(g_m\right)$), permet donc d'extraire V_{th} avec une précision d'environ $\pm$10mV, là où la méthode du V_{th} extrapolé fournit un résultat décalé de 60mV par rapport à la valeur de référence.

$$\Delta V_{\mathrm{th}} = \frac{kT}{qA} \cdot \ln\left(1 + \frac{\Delta\beta}{\beta}\right), \quad \frac{kT}{qA} \sim 30 - 40\mathrm{mV} \tag{III.25}$$

La méthode fournit une valeur de V_{th} proche de $V_{th,2\phi F}$ qui peut être extrait par voie capacitive.

Les principales limitations de cette méthode d'extraction sont :

- le temps de mesure est allongé par les faibles niveaux de courant caractéristiques de l'inversion faible,
- la technique nécessite de pouvoir analyser correctement les propriétés électriques sous le seuil. Elle ne fonctionne donc pas sur des dispositifs ayant un mauvais contrôle électrostatique (perçage).

III.4 L'extraction de tous les paramètres du modèle V_{th}, β, Θ_1 et Θ_2

III.4.1 Les techniques utilisant la fonction Y : des valeurs sûres

La fonction Y fut introduite par Ghibaudo en 1988 [Ghibaudo88]. Elle repose sur une particularité mathématique du modèle de courant (III.6) qui permet de construire une fonction $Y(V_{gs})$ insensible aux variations du paramètre Θ_1 (ou de $\Theta_{1,eff}$ si l'effet des résistances séries est pris en compte). La définition de la fonction Y est donnée en (III.26) qui présente également l'expression analytique de la fonction dans le cas général, en injectant (III.6).

$$Y \overset{\text{def}}{=} \frac{I_d}{\sqrt{g_m}} = \sqrt{\frac{\beta \cdot V_{ds}}{1 - \Theta_2 \cdot {V_{gt}}^2}} \cdot V_{gt} \tag{III.26}$$

L'absence de Θ_1 procure les avantages suivants à l'extraction depuis la fonction Y :

- l'extraction des paramètres est indépendante vis-à-vis des variations de la résistance série (cf. figure III.10),
- la fonction est indépendante de la réduction de la mobilité effective avec V_{gt} au premier ordre (le paramètre Θ_1 n'apparait plus dans l'expression de Y).

La relation (III.26) montre clairement une dépendance de Y en fonction de Θ_2 qui rend la fonction non linéaire à fort V_{gt} (on considérera $\Theta_{2,eff}$ si la variation de la résistance série avec V_{gs} doit être prise en compte). Pour mieux comprendre, il est nécessaire de se replacer dans le contexte de l'époque : les dispositifs possédaient des oxydes de grille trop épais pour que les effets de Θ_2 soient gênant dans l'application de la procédure d'extraction dans la plage de V_{gs} utile. Lorsque l'influence de Θ_2 peut être négligée (relation III.2 à l'ordre 1, (III.26) peut se réduire à (III.27) qui fournit un moyen rapide et simple pour extraire V_{th} et β.

$$Y \simeq \sqrt{\beta \cdot V_{ds}} \cdot V_{gt}, \quad \Theta_2 \cdot {V_{gt}}^2 \ll 1 \tag{III.27}$$

Pour des transistors issus des technologies sub-250nm, l'influence du paramètre Θ_2 ne peut plus être négligée dans le calcul de la fonction-Y, et plus globalement dans l'extraction des paramètres électriques des MOSFETs [Mourrain00]. La cause vient principalement de l'amincissement de l'oxyde de grille qui tend à augmenter le champ électrique (pour un V_{gt} donné) et à accroitre l'influence de la rugosité de surface à fort champ transverse ($\mathcal{E}_x$) [Reichert96]. Deux solutions sont envisageables pour résoudre ce problème :

- définir une plage d'extraction en début d'inversion forte, pour laquelle le modèle reste juste et l'approximation $\Theta_2 \cdot {V_{gt}}^2 \ll 1$ est justifiée,
- améliorer la technique d'extraction de manière à prendre en compte les effets de Θ_2 et à permettre l'extraction sur toute la plage d'inversion forte.

La première solution nécessite de cibler la bonne plage de tensions, opération que devient de plus en plus difficile à réaliser pour les transistors modernes, comme le montre la figure III.11. Le choix de cette plage conditionne le résultat et répercute une forte incertitude sur les valeurs extraites.

III.4.1.a La définition des plages d'extraction : un point critique

De manière générale, les méthodes d'extraction qui s'appuient sur le modèle de courant en inversion forte (III.6) nécessitent la définition d'au moins une plage d'extraction (des valeurs

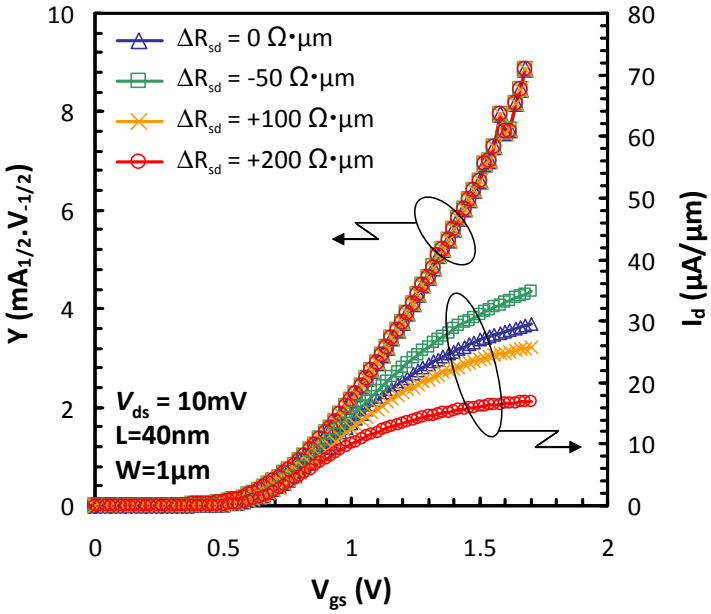


Fig. III.10 – Par définition, la fonction Y est insensible aux variations de $R_{sd}(V_{gs})$ au premier ordre.

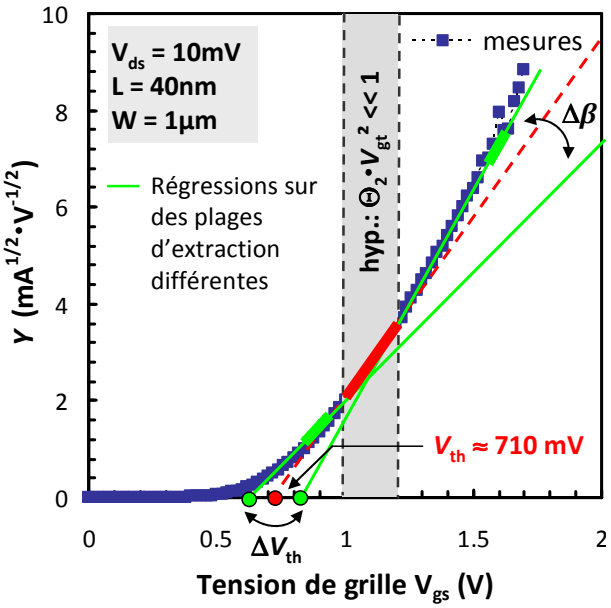


Fig. III.11 – Extraction de V_{th} depuis une régression sur la fonction Y [Ghibaudo88]. Le choix de la plage d'extraction où l'influence de Θ_2 peut être négligée conditionne la valeur du résultat. La régression linéaire en traits pointillés présente l'extraction sur la plage calibrée pour cette technologie. Les extractions présentées par des traits pointillés présentent d'autres choix possibles, soulignant l'incertitude sur la valeur de V_{th}, d'après [Fleury08a].

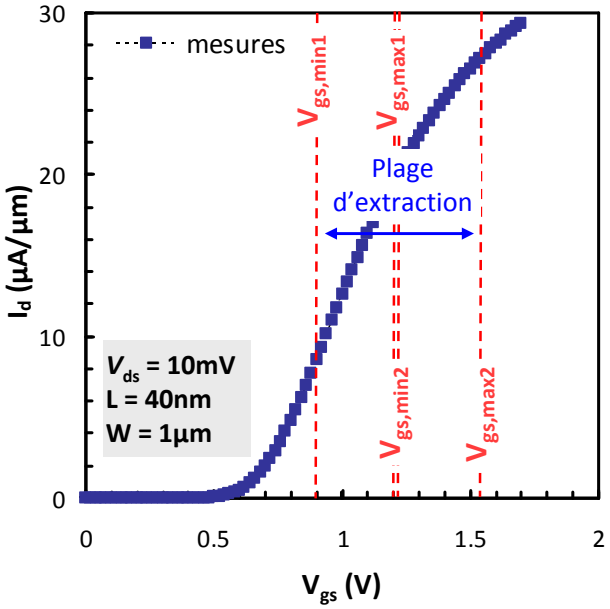


Fig. III.12 – Illustration des deux plages de tensions communes à toutes les méthodologies d'extraction, qui définissent la zone dans laquelle la fonction Y peut s'appliquer ($V_{gs} \in [V_{gs,min1}, V_{gs,max1}]$) et celle où le paramètre Θ_2 peut être extrait ($V_{gs} \in [V_{gs,min2}, V_{gs,max2}]$).

de V_{gs}) qui définit un domaine de validité pour l'hypothèse d'inversion forte du modèle. Les bornes de cette plage doivent être fixées par l'utilisateur. Elles influent plus ou moins sur les résultats, suivant la sensibilité propre à la technique utilisée. Afin d'homogénéiser le choix des plages entre les différentes techniques d'extraction que nous allons détailler dans les prochaines parties, nous définissons deux zones caractéristiques de la courbe $I_d(V_{gs})$ d'un transistor MOS telles que l'illustre la figure III.12) [Cros05] :

- une zone de faible champ en inversion forte, dans laquelle l'influence de Θ_2 peut être négligée, permettant l'extraction de V_{th} et β : $V_{gs} \in [V_{gs,min1}, V_{gs,max1}]$,
- une zone à fort champ transverse dans laquelle le paramètre Θ_2 est particulièrement influent et peut être extrait avec plus de facilité : $V_{gs} \in [V_{gs,min2}, V_{gs,max2}]$.

Notons que la condition $V_{gs,min1} \leq V_{gs} \leq V_{gs,max2}$ délimite la gamme de tensions pour laquelle l'hypothèse d'inversion forte est admise. Cette plage est commune à la plupart des méthodes d'extraction que nous verrons par la suite. En général, les plages d'extraction sont définies en fonction d'une valeur *approchée* de la tension de seuil, permettant ainsi de s'affranchir des variations de $V_{th}(L)$ dans les résultats de l'extraction. Ainsi, les plages d'extraction pour la technologie 45nm sont définies selon la relation (III.28), où $V_{th}{}^{\star}$ représente une valeur approchée de la tension de seuil, qui peut par exemple provenir de la technique d'extrapolation en inversion forte décrite au §III.3.3.

$$\begin{pmatrix} V_{gs,min1} \\ V_{gs,max1} \\ V_{gs,min2} \\ V_{gs,max2} \end{pmatrix} - V_{th}{}^{\star} = \begin{pmatrix} 0.35 \\ 0.55 \\ 0.55 \\ 0.9 \end{pmatrix} \quad \text{(V)} \tag{III.28}$$

Ces paramètres seront conservés fixes dans tous les résultats qui seront présentés afin de permettre une comparaison *juste* entre des différentes méthodes. L'influence de leurs choix sur les résultats sera évaluée pour la plupart des méthodologies d'extraction au §III.5.2.

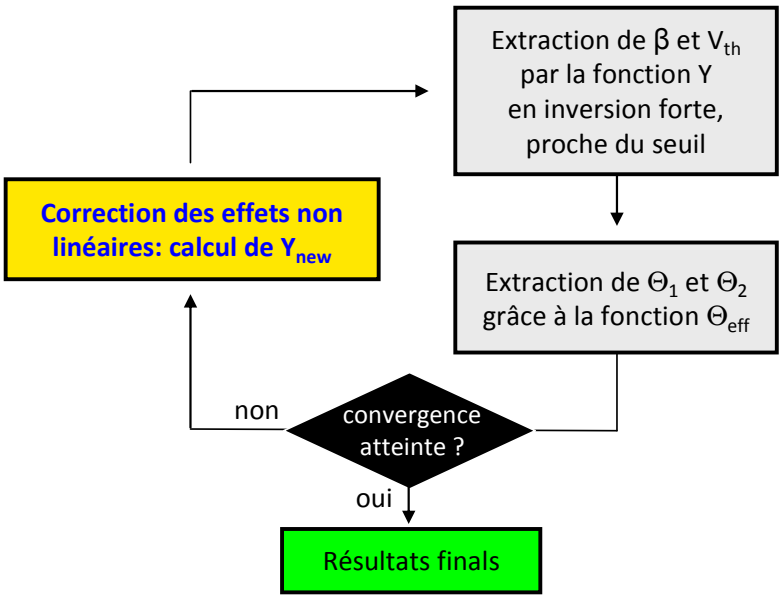


Fig. III.13 – Algorithme d'extraction itératif proposé par Mourrain dans [Mourrain00].

III.4.2 L'état de l'art des améliorations de la fonction Y pour les transistors courts ($L_{eff} < 100$nm)

III.4.2.a La méthode fonction Y itérative de Mourrain

En 2000 Mourrain propose une méthode qui corrige la non linéarité de la fonction Y grâce à une procédure itérative pour corriger l'effet indésirable de Θ_2 dans l'extraction [Mourrain00]. Cette méthodologie, dont l'algorithme est présenté sur la figure III.13, se compose des étapes suivantes :

1. effectuer une première extraction des paramètres V_{th}, β en inversion forte proche du seuil selon la méthode de [Ghibaudo88], en utilisant la relation (III.27).
2. extraire des paramètres Θ_1 et Θ_2 (connaissant V_{th} et β), en utilisant la fonction $\Theta_{eff}(V_{gs})$ définie par (III.29)

$$\Theta_{\text{eff}} \overset{\text{def}}{=} \left(\frac{\beta \cdot V_{\text{ds}}}{I_d} - \frac{1}{V_{\text{gt}}} \right) = \Theta_1 + \Theta_2 \cdot V_{\text{gt}} \tag{III.29}$$

3. utiliser la valeur de Θ_2 extraite en 2/ pour corriger la fonction Y des effets non linéaires en utilisant la relation (III.30) et répéter l'opération jusqu'à convergence de tous les paramètres.

$$Y_{new} = Y_{old} \times \sqrt{1 - \Theta_2 \cdot {V_{gt}}^2} \tag{III.30}$$

L'évolution typique des paramètres au cours du processus itératif est représentée sur la figure III.14. Celle-ci montre l'évolution de l'erreur entre la valeur des paramètres en sortie de l'itération n et la valeur stable obtenue au bout d'un grand nombre d'itérations. La convergence de l'algorithme est absolue (au sens mathématique du terme) et une erreur négligeable est obtenue au bout de seulement 5 itérations (limite indiquée par la droite verticale en pointillé). Nous verrons néanmoins au §III.5.2 que si cette correction itérative semble résoudre le problème de la non linéarité de $Y(V_{gs})$, les résultats sont très dépendants de la plage d'extraction choisie pour effectuer l'extraction des paramètres β et V_{th}.

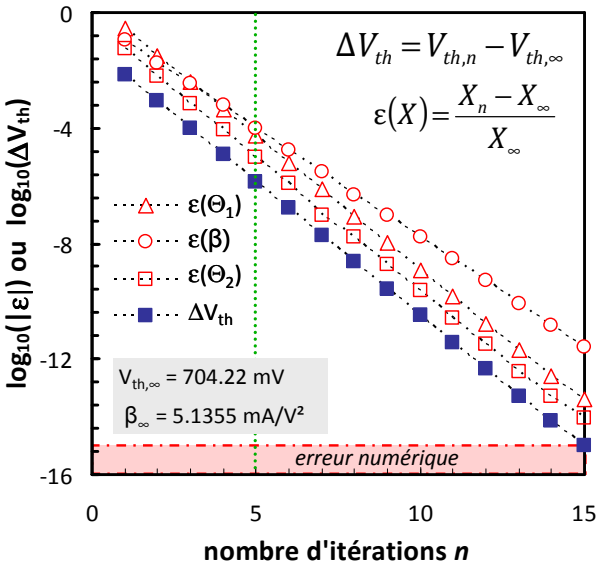


Fig. III.14 – Evolution de l'erreur sur les paramètres du modèle en fonction du nombre d'itérations n, lors de l'application de la technique décrite dans [Mourrain00], donnée par l'algorithme de la figure III.13. Transistor nMOS issu de la technologie 45nm ($L = 40nm$, $W = 1\mu$m).

Les courbes $Y(V_{gs})$ et $\Theta_{eff}(V_{gt})$ avant correction et après 5 itérations sont données sur la figure III.15. Cette dernière illustre très clairement l'effet de linéarisation de la fonction Y apporté par l'algorithme, ainsi que les modifications des valeurs $\Theta_{1,2}$, en accord avec [Mourrain00].

III.4.2.b La méthode Y-Φ introduite par Tanaka

Dans la même optique de [Mourrain00], en 2007, Tanaka propose une méthodologie d'extraction qui repose sur la linéarisation de $Y(V_{gs})$ par l'ajustement d'un facteur Φ et de la fonction F (III.31).

$$\begin{cases} F(\Phi, Y) = \dfrac{1}{\sqrt{1/Y^2 + \Phi}} \\ \Phi = \dfrac{\Theta_2}{\beta \cdot V_{ds}} \end{cases} \tag{III.31}$$

La détermination de Φ se fait grâce à un algorithme d'optimisation : dans notre cas nous utilisons l'algorithme de résolution non-linéaires de Scilab [INRIA] (basé sur un algorithme de type Levenberg-Marquardt [Press07]). Même avec une optimisation efficace et précise, cette méthode de linéarisation induit une forte erreur sur les résultats par rapport à la fonction Y classique ou au V_{th} extrapolé depuis le maximum de g_m [Ghibaudo88]. C'est en particulier ce que montre la figure III.16 qui compare les écart-types σ sur V_{th} obtenus par ces méthodes d'extraction pour différentes longueurs. Un facteur 2 est induit par le procédé d'optimisation dans le cas de la méthode Y-Φ. Nous noterons également qu'une optimisation correcte nécessite des mesures précises (longues) et un algorithme d'optimisation performant qui ne sont pas adaptés à une intégration en tests automatiques.

III.4.2.c La méthode de Faynot-McLarty

La méthode de Faynot-McLarty [McLarty95, Faynot94] propose d'exploiter les propriétés du modèles (III.6) pour construire une fonction insensible aux variations des paramètres Θ_1 et

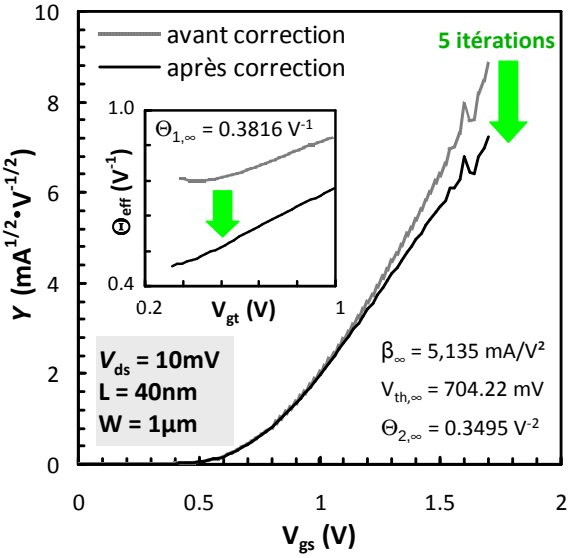


Fig. III.15 – Modification des courbes $Y(V_{gs})$ et $\Theta_{eff}(V_{gt})$ avant et après 5 itérations successives.

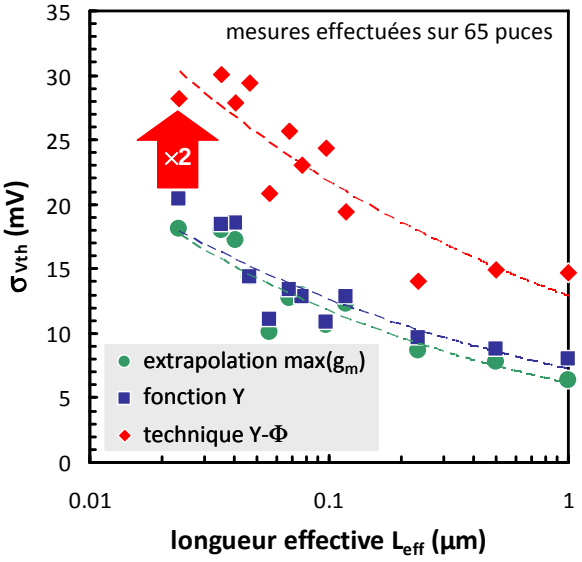


Fig. III.16 – Augmentation de l'écart-type sur la valeur de V_{th} résultant du procédé d'optimisation dans la méthode Y-Φ [Tanaka07]. Un facteur deux sur l'étalement de la distribution de V_{th} résulte de la procédure de linéarisation dans le cas de la méthode Y-Φ.

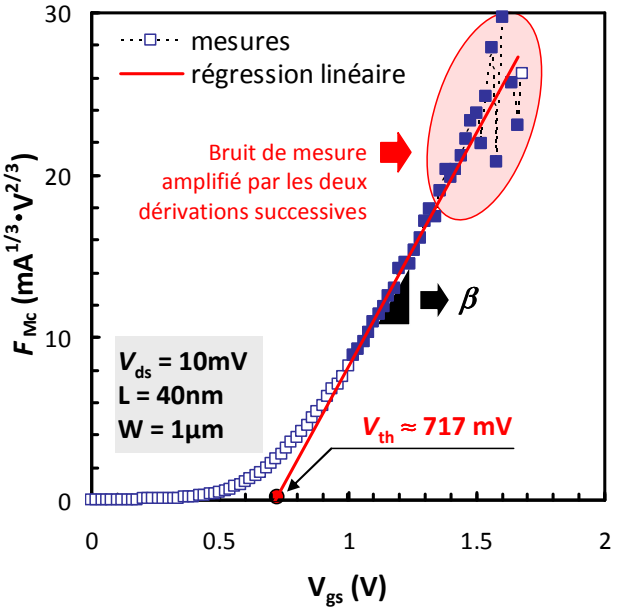


Fig. III.17 – Extraction à partir de la fonction de Faynot-McLarty (F_{Mc}). L'amplification du bruit de mesure en forte inversion dégrade la précision sur les paramètres V_{th} et β.

Θ_2 (et donc également aux variations du premier et du second ordre de la résistance série). La fonction de Faynot-McLarty est définie par la relation (III.32) qui, lorsque I_d est remplacé par son expression analytique (III.6), a une dépendance linéaire en V_{gt} et permet d'extraire V_{th} et β par une simple régression linéaire.

$$F_{Mc} \stackrel{\text{def}}{=} \left[\frac{\partial^2}{\partial {V_{gs}}^2}\left(\frac{1}{I_d}\right)\right]^{-1/3} = \left(\frac{\beta \cdot V_{ds}}{2}\right)^{1/3} \cdot V_{gt} \qquad \text{(III.32)}$$

L'extraction de V_{th} et β par la fonction F_{Mc} est illustrée sur la figure III.17 pour un transistor nMOS de la technologie 45nm de longueur de grille $L = 40$nm. L'inconvénient majeur de cette technique est dans l'usage d'une dérivée seconde qui amplifie le bruit de la mesure. Ce dernier est particulièrement présent dans la zone de forte inversion, précisément pour les valeurs de V_{gs} qui permettent la validité de l'hypothèse de forte inversion valide. Pour un temps de mesure donné, il en résulte une incertitude plus grande sur le résultat issu de F_{Mc} en comparaison de ce qu'il est possible d'obtenir avec les techniques de type fonction Y qui ne nécessitent qu'une seule dérivation.

III.4.3 Proposition d'une amélioration de la fonction Y : la méthode de la fonction ξ

Comme nous venons de le voir, la technique d'extraction fonction-Y est l'une des rares qui permet de s'affranchir de l'influence de la résistance série (qui est prise en compte dans le paramètre Θ_1) et des variations de la mobilité avec la tension de grille μ_{eff} ou la longueur de canal. Le problème principal réside dans la définition de la zone linéaire dans laquelle l'hypothèse de forte inversion est requise par le modèle et l'influence du paramètre Θ_2 est supposée négligeable. Tanaka a proposé de considérer l'influence du Θ_2 à travers de nouvelles fonctions basées sur le modèle complet (méthode 6c-5p) [Tanaka05], puis d'introduire un paramètre correctif Φ ajusté de manière à linéariser la fonction Y (méthode Y–Φ) [Tanaka07]. Cette méthode fournit des ré-

sultats instables et dépendants des paramètres d'entrée choisis, si bien que plusieurs extractions effectuées sur une même courbe $I_d(V_{\text{gs}})$ par des personnes différentes fourniront des paramètres électriques différents.

Ainsi, nous proposons une nouvelle méthodologie d'extraction, nommée fonction ξ et basée sur un algorithme récursif dont les critères de convergence et la stabilité seront étudiés en détail dans les prochaines parties. Le schéma d'extraction est défini par les trois points suivants :

1. extraction de V_{th} et β sur le modèle complet en utilisant un algorithme récursif,
2. extraction des facteurs de réduction de la mobilité effective Θ_1 et Θ_2 depuis la fonction Θ_{eff}, comme dans [Ghibaudo88, Mourrain00],
3. extraction éventuelle du paramètre ΔV_{th} pour assurer la compatibilité de la nouvelle méthode avec la technique décrite dans [Cros05], adaptée aux basses températures.

III.4.3.a Définition de l'algorithme récursif permettant l'extraction de V_{th} et β

Dans notre schéma d'extraction, V_{th} et β sont les premiers paramètres du modèle à être extraits. Comme dans les autres méthodologies, ils sont ensuite réutilisés pour permettre l'extraction des autres paramètres (Θ_1 et Θ_2) et doivent justifier d'une précision élevée pour éviter de propager une erreur trop importante sur ces valeurs. La méthode itérative décrite dans [Mourrain00] permet de linéariser la fonction-Y pour effectuer l'extraction comme dans [Ghibaudo88]. En revanche, la convergence de l'algorithme mis en place vers des valeurs précises semble difficile [Romanjek04b] car instable et fortement dépendante des paramètres d'entrée. Nous proposons donc un nouvel algorithme permettant une convergence plus stable, basé sur la fonction ξ définie comme dans (III.33).

$$\xi \overset{\text{def}}{=} \frac{1}{Y^2} \tag{III.33}$$

Dans un premier temps, nous considérons que V_{th} n'est pas connu avec précision et qu'il peut seulement être estimé avec une erreur ε par rapport à la valeur exacte V_{th}. Nous notons cette estimateur ${V_{\text{th}}}^\star$ et définissons ${V_{\text{gt}}}^\star$ comme une estimation (à ε près) de la tension de grille au delà du seuil. La fonction ξ peut alors être réécrite comme dans (III.34) et peut être approximée par un développement de Taylor pour aboutir à une fonction polynomiale en $1/{V_{\text{gt}}}^\star$, valable lorsque l'erreur relative $\varepsilon/{V_{\text{gt}}}^\star$ est petite (III.35).

$$\xi = \frac{1}{\beta \cdot V_{\text{ds}}} \cdot \left\{ \frac{1}{\left({V_{\text{gt}}}^\star + \varepsilon\right)^2} - \Theta_2 \right\} \tag{III.34}$$

$$\frac{\varepsilon}{{V_{\text{gt}}}^\star} \simeq 0 \Rightarrow \frac{1}{\left({V_{\text{gt}}}^\star + \varepsilon\right)^2} \simeq 1 - \frac{2\varepsilon}{{V_{\text{gt}}}^\star} + o\left(\frac{1}{{V_{\text{gt}}}^\star}\right) \tag{III.35}$$

La fonction ξ sous sa forme polynomiale approchée est donnée par (III.36). Elle fournit un système autoconsistant qui permet la détermination de V_{th} et β après quelques itérations.

$$\xi \simeq \frac{1}{\beta \cdot V_{\text{ds}}} \cdot \left\{ \frac{1}{{V_{\text{gt}}}^{\star 2}} - \frac{2\varepsilon}{{V_{\text{gt}}}^{\star 3}} - \Theta_2 \right\} \tag{III.36}$$

Comme détaillé dans [Ghibaudo88, Mourrain00], lorsque V_{th} et β sont connus, il est relativement simple d'extraire les paramètres Θ_1 et Θ_2 à partir de la fonction Θ_{eff} définie en (III.29).

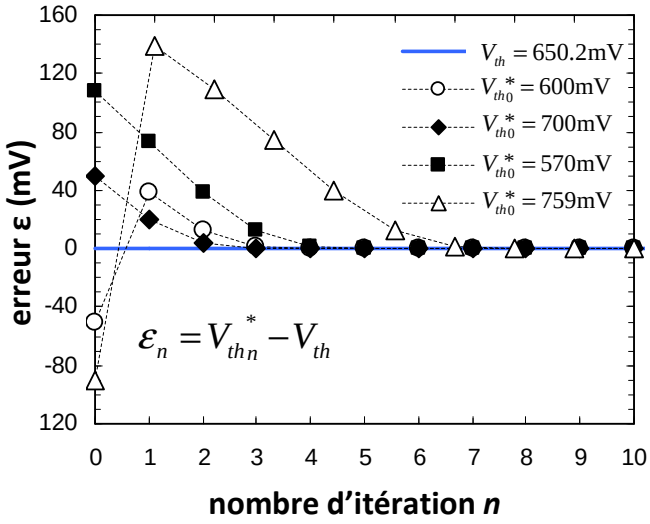


Fig. III.18 – Evolution de l'erreur ε_n en fonction du nombre d'itérations n, pour différents choix de valeur initiale $V_{\mathrm{th}0}$. L'algorithme permet une convergence quasi-absolue quand l'erreur sur la valeur initiale est comprise entre -80mV et 100mV. ε_n devient de l'ordre de la précision de calcul après 7 itérations.

L'erreur entre le modèle et la mesure est inférieure à 1% et bénéficie de la justesse dans la détermination du couple de paramètres (β, V_{th}).

III.4.3.b Convergence de l'algorithme

Nous allons étudier la convergence de l'algorithme en fonction du nombre d'itérations (noté n). Nous ajouterons donc un indice $_n$ à l'estimateur $V_{\mathrm{th}}{}^{\star}$ et à l'erreur associée : $\varepsilon_n = V_{\mathrm{th}n}^{\star} - V_{\mathrm{th}}$. L'évolution de l'erreur en fonction du nombre d'itérations est présentée sur la figure III.18. La convergence est réalisée même lorsque l'erreur sur la valeur initiale est élevée (typiquement -80mV $\leq \varepsilon_0 \leq$ 100mV). Une erreur négligeable est atteinte après quelques itérations et la limite de la précision de calcul est atteinte au delà de 7 itérations ($\varepsilon_{n>7} \sim 10^{-12}$V). La tension de seuil converge vers une valeur stable après 7 itérations : 650,2mV dans notre cas (III.37).

$$V_{\mathrm{th}n>7}^{\star} = V_{\mathrm{th}}, \quad -80\mathrm{mV} \leq \varepsilon_0 \leq 100\mathrm{mV} \tag{III.37}$$

Compte tenu de la souplesse possible dans le choix de la valeur initiale $V_{\mathrm{th}}{}^{\star}{}_0$, il est possible d'utiliser la technique du V_{th} extrapolé depuis l'inversion forte comme point d'entrée dans notre algorithme. En effet, cette valeur est un bon estimateur de V_{th} à quelques dizaines de mV près qui résultent de l'influence de la résistance série et de la décroissance de la mobilité avec V_{gs}.

III.4.3.c Cohérence avec la méthodologie d'extraction décrite par Cros et al.

Dans [Cros05], Cros et al. ont démontré l'intérêt d'ajouter un paramètre supplémentaire au modèle de courant pour que la prise en compte des V_{gs} forts dans l'extraction de type fonction Y par partie décrite dans [Ghibaudo88, Mourrain00]. Un paramètre ΔV_{th} est alors introduit dans l'équation de [Ong87] (III.38), il permet de paramétrer un déclenchement plus tardif (en terme d'intensité de champ électrique transverse) de la réduction de la mobilité due à la rugosité de surface.

$$I_d = \frac{\beta \cdot V_{\mathrm{ds}} \cdot V_{\mathrm{gt}}}{1 + \eta_1 \cdot V_{\mathrm{gt}} + \eta_2 \cdot (V_{\mathrm{gt}} - \Delta V_{\mathrm{th}})^2} \tag{III.38}$$

La concordance de (III.38) et (III.6) au premier ordre est assurée par la conservation du paramètre β qui respecte ainsi la relation (III.39) proche du seuil.

$$I_d = \beta \cdot V_{\mathrm{ds}} \cdot V_{\mathrm{gt}} + o\left(V_{\mathrm{gt}}\right), \quad V_{\mathrm{gt}} \sim 0^+ \tag{III.39}$$

Mathématiquement, il est alors trivial de faire la correspondance entre les paramètres η_1, η_2 et β' de (III.38) et ceux de notre modèle initial. Cela montre que l'introduction d'un ΔV_{th} dans un modèle continu (extraction effectuée sur la gamme de V_{gs} en inversion) revient juste à effectuer un changement de variable : une redistribution de l'influence des paramètres Θ_1 et Θ_2

$$\begin{cases} \eta_2 & = \dfrac{\Theta_1}{1 - \Theta_2 \cdot {\Delta V_{\mathrm{th}}}^2} \\ \\ \eta_1 & = \dfrac{\Theta_1 + 2\Theta_2 \cdot \Delta V_{\mathrm{th}}}{1 - \Theta_2 \cdot {\Delta V_{\mathrm{th}}}^2} \end{cases} \tag{III.40}$$

Il faut donc connaitre ΔV_{th} pour établir la correspondance entre nos paramètres et ceux utilisés dans un modèle utilisant un V_{th2} dans une extraction par domaine, comme dans [Cros05]. Au premier ordre, cela revient à dire que le modèle doit présenter un bon agrément avec la mesure même lorsque le paramètre Θ_2 est fixé à 0. Notons par $I_d^{(1)}$ et $g_m^{(1)}$ le courant et la transconductance issus du modèle III.6 dans lequel Θ_2 est volontairement fixé à 0. Dans ce cas, l'écart entre la mesure et le modèle illustre l'écart qui résulte indirectement de ΔV_{th}. Dans une extraction fonction-Y par domaines, les valeurs expérimentales issues du modèle de $g_m^{(1)}$ doivent concorder près du maximum de transconductance (noté $V_{\max(g_m)}$), cf. figure III.19. Nous pouvons donc établir une expression analytique permettant de déterminer la valeur de ΔV_{th}, cf. (III.41).

$$\Delta V_{\mathrm{th}} \simeq \frac{\sqrt{\dfrac{\beta V_{\mathrm{ds}}}{g_m^{\mathrm{meas}}}} - 1 - \Theta_1 \cdot V_{\mathrm{gt}}}{2\Theta_2 \cdot V_{\mathrm{gt}}}, \quad V_{\mathrm{gs}} \simeq V_{\max(g_m)} \tag{III.41}$$

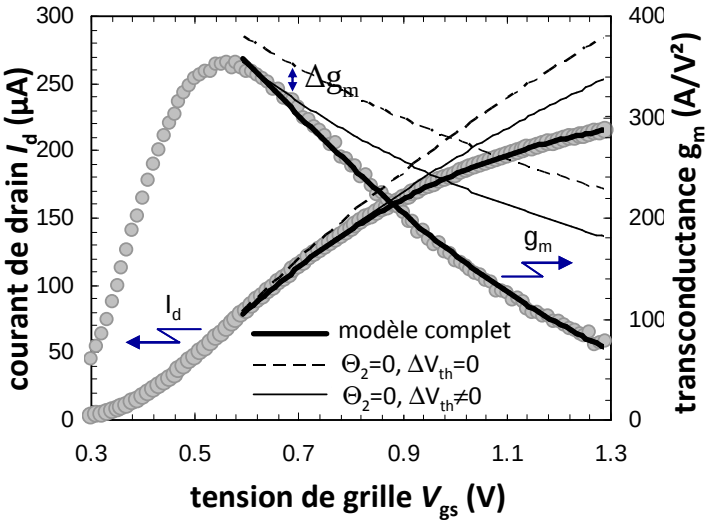


Fig. III.19 – Détermination de ΔV_{th} pour permettre la concordance avec la méthodologie d'extraction par partie décrite dans [Cros05].

III.4.4 Nouvelle méthodologie d'extraction utilisant les régimes de forte et de faible inversion

Une nouvelle méthodologie d'extraction peut être définie pour extraire tous les paramètres du modèle en utilisant à la fois les régimes sous le seuil et au dessus du seuil. En effet, la relation (III.24) peut être utilisée pour supprimer un degré de liberté en utilisant les propriétés sous le seuil (la suppression du degré de liberté vient de la relation $V_{th} = f(\beta)$). Il est alors possible d'effectuer une régression non-linéaire du modèle (III.6) sur le régime d'inversion forte de la courbe $I_d(V_{gs})$ en imposant une relation entre V_{th} et β extraite à partir des propriétés sous le seuil (III.24). Contrairement à ce que nous avons vu au §III.2.3, la suppression d'un degré de liberté dans la résolution stabilise l'extraction et réduit le nombre d'ensembles de paramètres qui donnent un bon accord avec les mesures. La régression peut être effectuée grâce à un algorithme de Levenberg-Marquardt [Press07], par exemple.

Pour s'affranchir de la définition d'une plage fixe dans laquelle est effectuée la régression, il est possible de faire la régression entre V_{th} (qui doit être déterminé) et le maximum de V_{gs}. Il est alors préférable de pondérer l'erreur à faible V_{gt} de manière à limiter d'influencer l'extraction par la zone d'inversion modérée qui n'est pas prédite correctement par les modèles d'inversion faible et forte. L'extraction sur un dispositif nMOS $L_{mask} = 40$nm de la technologie CMOS45nm LSTP est présentée sur la figure III.20. Finalement, précisons que si cette technique semble fournir des résultats précis et stable en post-traitement, elle est difficilement transférable en test en ligne du fait de la complexité des algorithmes de régression mis en jeu pour la détermination des paramètres.

Les principaux inconvénients de cette méthodologie d'extraction sont :
- elle nécessite une mesure en inversion faible (le temps de mesure est donc accru),
- elle est sensible aux dégradations de la pente sous le seuil (perçage, DIBL élevé).

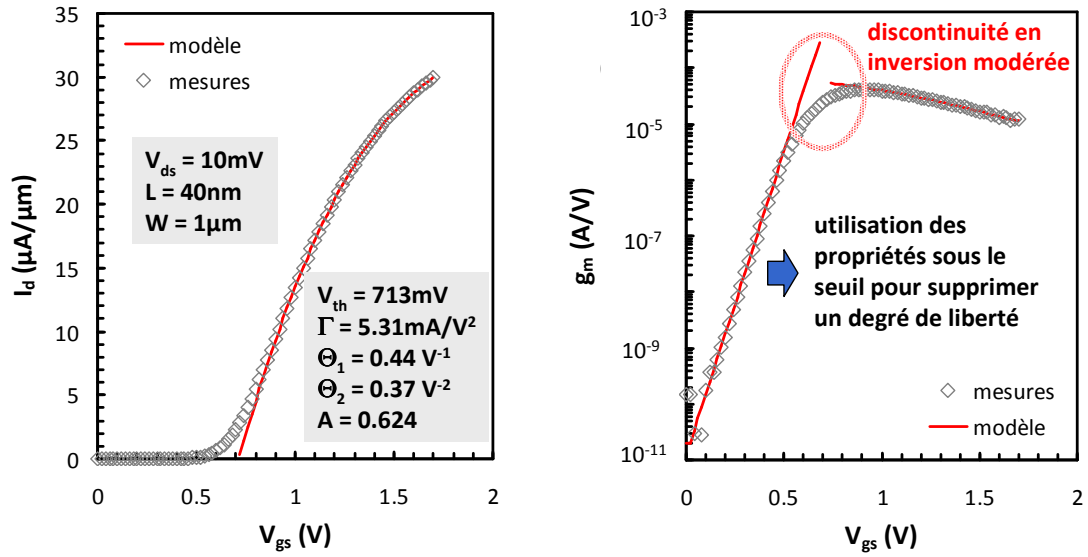


Fig. III.20 – Extraction des paramètres du modèle grâce à une régression non-linéaire (de type Levenberg-Marquardt) dans la zone d'inversion forte. L'utilisation des propriétés sous le seuil fournit une relation du type $V_{th} = f(\beta)$ qui permet supprimer un degré de liberté dans la résolution. L'extraction est ainsi stabilisée par rapport à ce qui a été vu au §III.2.3.

III.5 Robustesse des méthodes d'extraction

Comme nous venons de le voir, il existe plusieurs méthodes qui permettent d'extraire les paramètres du modèle de courant (III.6). Dans cette partie, nous proposons d'évaluer la précision et la fiabilité de ces techniques sur deux critères qui paraissent majeurs vis-à-vis de l'extraction de paramètres sur les MOSFETs :

- l'immunité des résultats vis-à-vis des variations de R_{sd}
- l'immunité des résultats vis-à-vis des paramètres d'entrée (bornes des plages d'extraction)

III.5.1 Immunité des résultats vis-à-vis des variations de R_{sd}

L'immunité des résultats de l'extraction vis-à-vis des variations de R_{sd} est critique. Ce point permet en effet de justifier de la cohérence entre les résultats obtenus sur des transistors de différentes longueurs, pour lesquels l'influence de R_{sd} n'est pas la même. Nous utiliserons la relation (III.42) qui permet d'ajouter virtuellement une composante ΔR_{sd} à la valeur initiale de la résistance série sur une caractéristique $I_d(V_{gs})$ d'un transistor en régime linéaire.

$$I_{d,\Delta R_{sd}} = \frac{V_{\rm ds}}{\left(\frac{V_{\rm ds}}{I_d}\right) + \Delta R_{sd}} \tag{III.42}$$

Dans un premier temps, analysons l'influence d'un ajout ΔR_{sd} allant jusqu'à $200\Omega \cdot \mu$m sur V_{th} extrait grâce aux différentes techniques évoquées dans ce chapitre. En effet, la tension de seuil semble être un indicateur fiable qui présente l'avantage d'être commun à toutes les techniques. La figure III.21 présente les dérives de V_{th} obtenues en fonction de ΔR_{sd}. Sans surprise, les méthodes d'extraction qui utilisent les propriétés de la caractéristique $I_d(V_{gs})$ au dessus du seuil (sauf celles basées sur la fonction Y) sont plus sensibles aux variations de R_{sd}. C'est en particulier le cas de la méthode de l'extrapolation en inversion forte qui subit à la fois la dérive de la tension caractéristique du $\max(g_m)$, mais également une modification de la pente utilisée

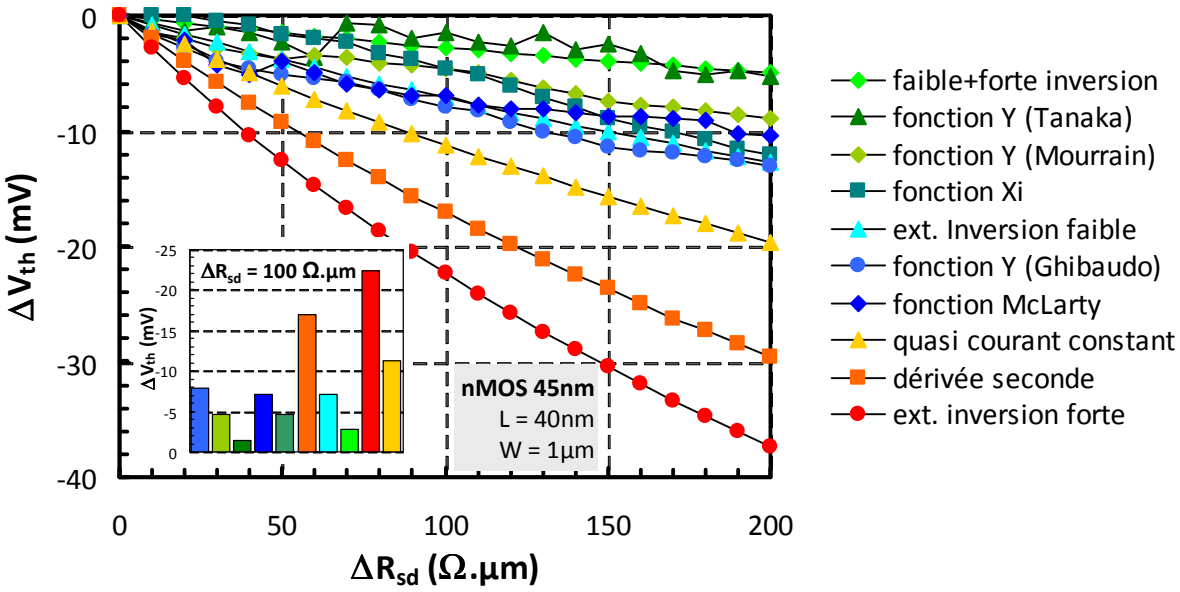


Fig. III.21 – Dérive de l'extraction de V_{th} lors de l'ajout d'une composante de résistance série ΔR_{sd}, pour les différentes méthodes d'extraction évoquées dans ce chapitre.

pour extrapoler V_{th}. Cette méthode présente un décalage de presque 40mV qui est inacceptable et fournit un profil $V_{th}(L)$ erroné sur cette technologie. Les valeurs issues des techniques d'extraction basées sur la fonction Y se détachent nettement de l'ensemble : les variations de V_{th} sont inférieures à 13mV dans le pire cas. Le V_{th} extrapolé depuis l'inversion faible et la méthode utilisant les deux régimes présentent également une bonne immunité aux variations de R_{sd}.

La figure III.22 présente l'évolution de l'erreur statistique sur V_{th} en fonction de ΔR_{sd} pour les mêmes techniques d'extraction que sur la figure III.21. Si aucune tendance de variation en fonction de ΔR_{sd} ne semble observable, l'instabilité de l'extraction proposée par Tanaka [Tanaka07] peut devenir problématique, suivant la précision requise.

Pour finir, notons la bonne immunité que présentent les méthodes basées sur les propriétés sous le seuil qui, par construction, exploitent une zone de la caractéristique $I_d(V_{gs})$ qui est peu influencée par R_{sd}. En particulier, la méthode à courant constant présente une variation $\Delta V_{th} < 1.5$mV lorsque ΔR_{sd} varie de 0 à $200\Omega \cdot \mu$m (le seuil de courant fixé à 40nA$\times W/L$). La méthode d'extraction en inversion faible détaillée au §III.3.5 présente elle aussi cet avantage même si elle reste faiblement influencée par les variations de $\max(g_m)$, qui intervient dans le calcul de V_{th} (III.24). Finalement, la meilleure immunité (mis à part la méthode à courant constant) est obtenue pour la fonction utilisant à la fois la forte et la faible inversion, pour laquelle un décalage de V_{th} inférieure à 5mV pour la gamme de valeurs ΔR_{sd} appliquées.

III.5.1.a Dérive du V_{th} extrapolé en inversion forte et répercutions sur les méthodes de type fonction Y

Comme le montre la figure III.21, la dérive du V_{th} extrapolé en inversion forte lors d'une modification de R_{sd} est la plus élevée en comparaison des autres méthodes d'extraction ($\Delta V_{th} \simeq$ 40mV pour $\Delta R_{sd} \simeq 200\Omega \cdot \mu$m). Cette dérive provient du fait que, lorsque R_{sd} change, l'amplitude du maximum de transconductance ($\max(g_m)$) ainsi que la valeur de V_{gs} pour lequel cette valeur est atteinte sont modifiées, de telle sorte que les variations ne se compensent pas et influent

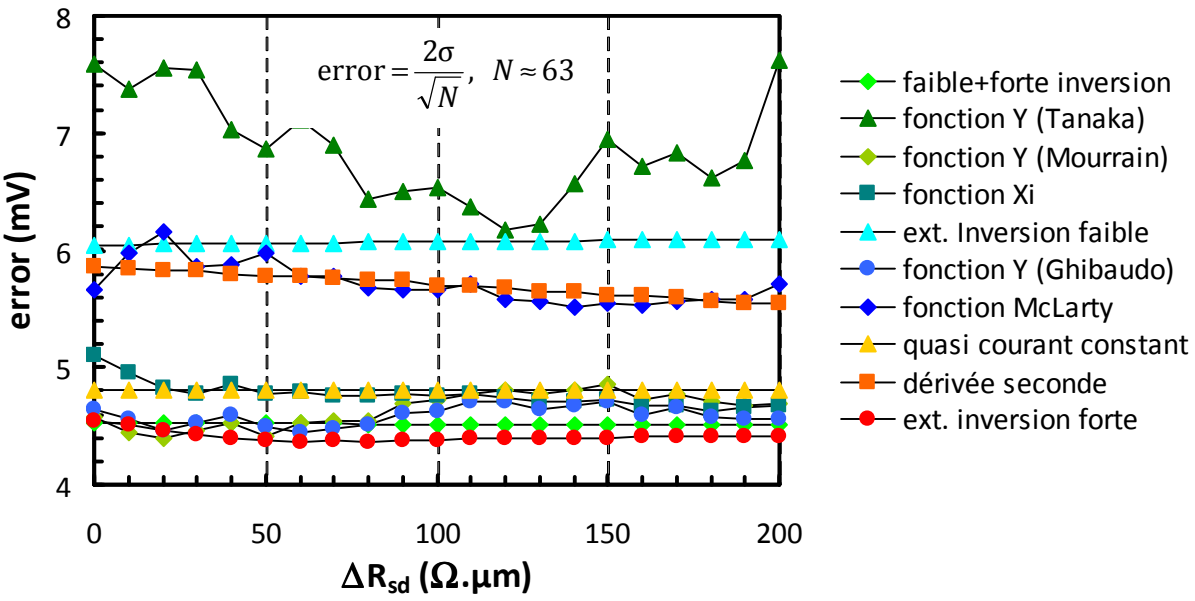


Fig. III.22 – Evolution de l'erreur statistique sur V_{th} lors de l'ajout d'une composante de résistances séries ΔR_{sd}.

fortement la valeur de V_{th} (cf. figure III.23).

Cette dérive est problématique car le V_{th} extrapolé en inversion forte est souvent utilisé comme un point d'entrée dans les techniques basées sur la fonction Y, afin de définir des plages d'extraction fixes en V_{gt} et non en V_{gs} (permet de s'affranchir des variations $V_{th}(L)$). Lorsque le V_{th} extrapolé en inversion forte change, il modifie implicitement les plages d'extraction et le résultat de l'extraction donné par la fonction Y (pourtant insensible aux variations de R_{sd} par construction). Ceci explique la dérive $V_{th}(\Delta R_{sd})$ observée pour les extractions de type fonction Y sur la figure III.21. Cette dépendance parasite peut bien sûr être supprimée en définissant des plages de tension fixes mais le problème d'une définition correcte de plages communes à toutes les longueurs se pose alors.

Comme solution, nous proposons d'utiliser une technique d'extraction moins sensible aux variations de R_{sd} pour fournir des plages d'extraction fixes en V_{gt} plus stables. La méthode d'extraction en inversion faible présentée au §III.3.5 semble remplir cette fonction et présente l'avantage de rester aussi simple à implémenter que l'extrapolation en inversion forte. De la même manière que pour la figure III.21, la figure III.24 montre les variations $V_{th}(\Delta R_{sd})$ qui proviennent des méthodes d'extraction basées sur la fonction Y lorsque le V_{th} extrapolé depuis l'inversion faible est utilisé pour définir les plages d'extraction en V_{gt} (III.43)[3]. Les valeurs obtenues sur la figure III.21 sont également reportées pour permettre la comparaison des résultats.

$$\begin{pmatrix} V_{gs,min1} \\ V_{gs,max1} \\ V_{gs,min2} \\ V_{gs,max2} \end{pmatrix} - V_{th,wi} = \begin{pmatrix} 0.55 \\ 0.75 \\ 0.75 \\ 1.1 \end{pmatrix} \text{ [V]} \qquad \text{(III.43)}$$

Si l'introduction d'un V_{th} plus stable dans la définition des plages d'extraction ne semble pas

3. Le décalage observé entre les valeurs de V_{th} extrapolé depuis les deux méthodes est compensé par un décalage de 200mV sur les tensions des plages.

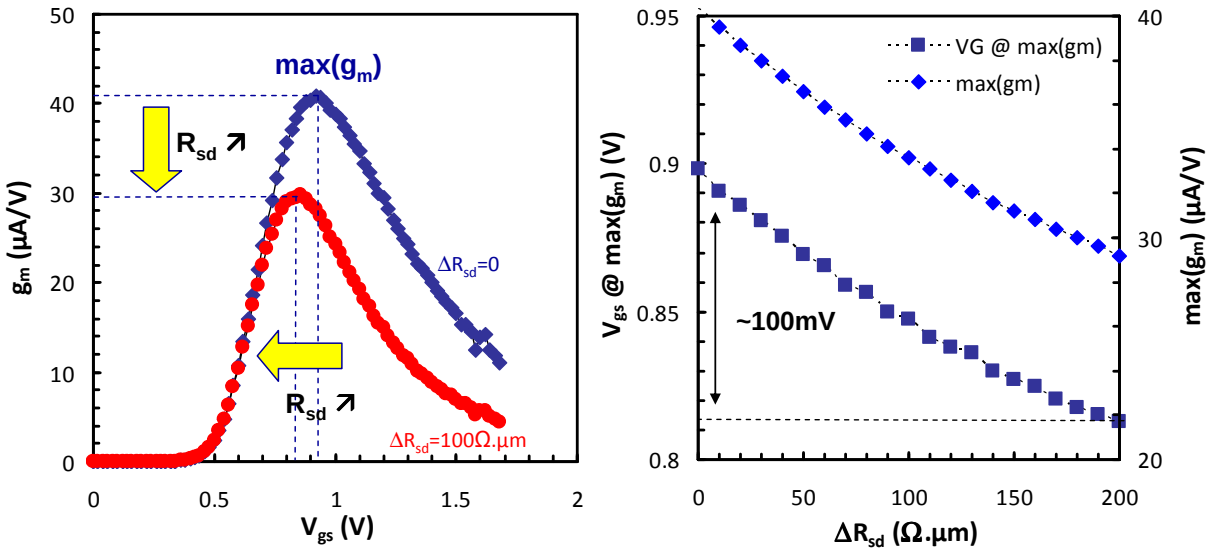


Fig. III.23 – Dérive du maximum de transconductance et de la tension de grille associée lorsque R_{sd} augmente.

améliorer l'insensibilité des résultats aux variations de R_{sd} pour les méthodes de type itératives [Mourrain00, Tanaka07, Fleury08a], cette technique permet de diminuer par un facteur 2 la dérive observée pour les techniques non itératives [Ghibaudo88, McLarty95].

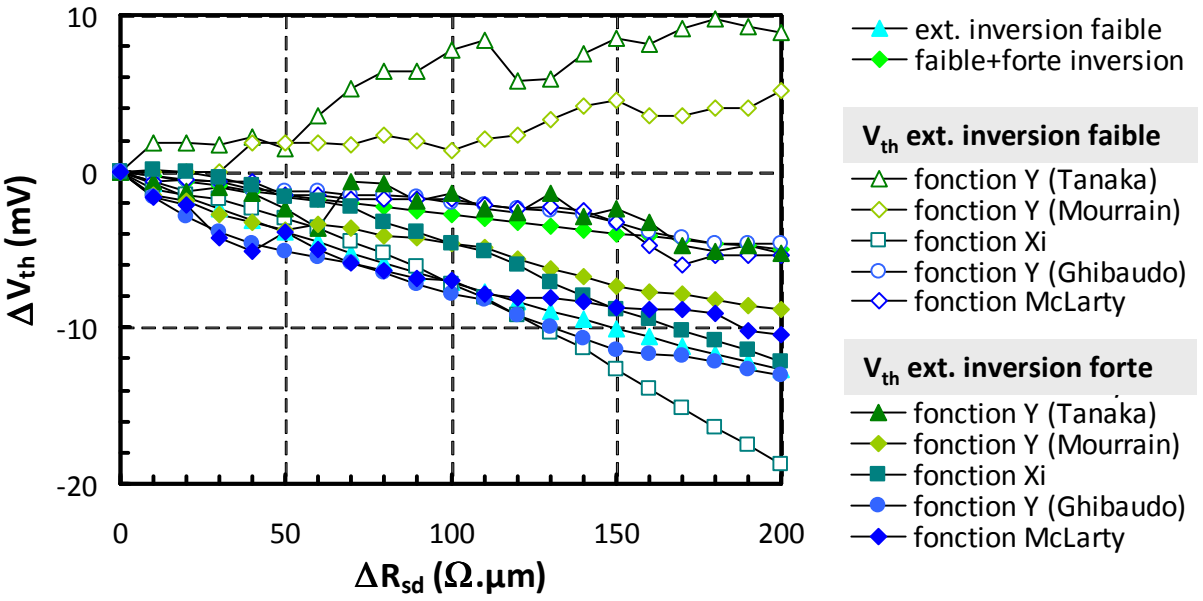


Fig. III.24 – Dérive de l'extraction de V_{th} lors de l'ajout d'une composante de résistance série ΔR_{sd}, pour les méthodes de type fonction Y, suivant l'origine du V_{th} utilisé pour définir les plages d'extraction.

III.5.2 Immunité des résultats aux variations de paramètres d'entrée

Les méthodes d'extraction qui reposent sur la fonction Y nécessitent au minimum la définition d'une plage d'extraction dans laquelle l'hypothèse de forte inversion, requise par le modèle, est admise. Nous avons vu que les faibles dépendances $V_{th}(\Delta R_{sd})$ provenant de la méthode de l'extrapolation en inversion forte utilisée dans la définition des plages d'extraction suffit déjà à influencer le résultat. Il parait donc pertinent d'étudier la sensibilité des méthodes d'extraction vis-à-vis des variations des bornes de la plage d'extraction définissant l'inversion forte.

La valeur $V_{gs,min1}$ semble la plus critique car elle définit la transition entre l'inversion modérée et l'inversion forte et donc la tension à partir de laquelle l'hypothèse d'inversion forte (i.e. la validité du modèle) est pertinente. La figure III.25 présente les variations sur les paramètres V_{th} et β lorsque $V_{gs,min1}$ est modifié de $\pm$100mV, de manière similaire à [Fleury08a][4]. Toutes les méthodes fournissent des variations de paramètres similaires. Nous noterons toutefois une déviation des résultats plus marquée pour la fonction ξ lorsque $V_{gs,min1}$ est décalé vers l'inversion modérée, et une instabilité des résultats issus de la méthode itérative de Mourrain [Mourrain00].

Les variations sont nulles concernant la méthode utilisant à la fois les régimes d'inversion faible et forte qui ne nécessitent pas de définition de plage d'extraction (la plage d'extraction est ajustée automatiquement au cours de la régression non linaire en réinjectant V_{th}). Ce résultat démontre l'avantage qui ressort de l'utilisation de cette nouvelle méthode qui permet une immunité accrue vis-à-vis des variations de $V_{gs,min1}$ et de R_{sd}.

4. Dans [Fleury08a], l'extraction était associée à une régression polynomiale d'ordre 4 qui permet d'atténuer l'influence de l'inversion modérée sur la courbe et ainsi de limiter les variations de paramètres lors d'une modification de $V_{gs,min1}$. Cette astuce n'est pas utilisée ici afin de permettre une comparaison plus objective entre les différentes méthodes.

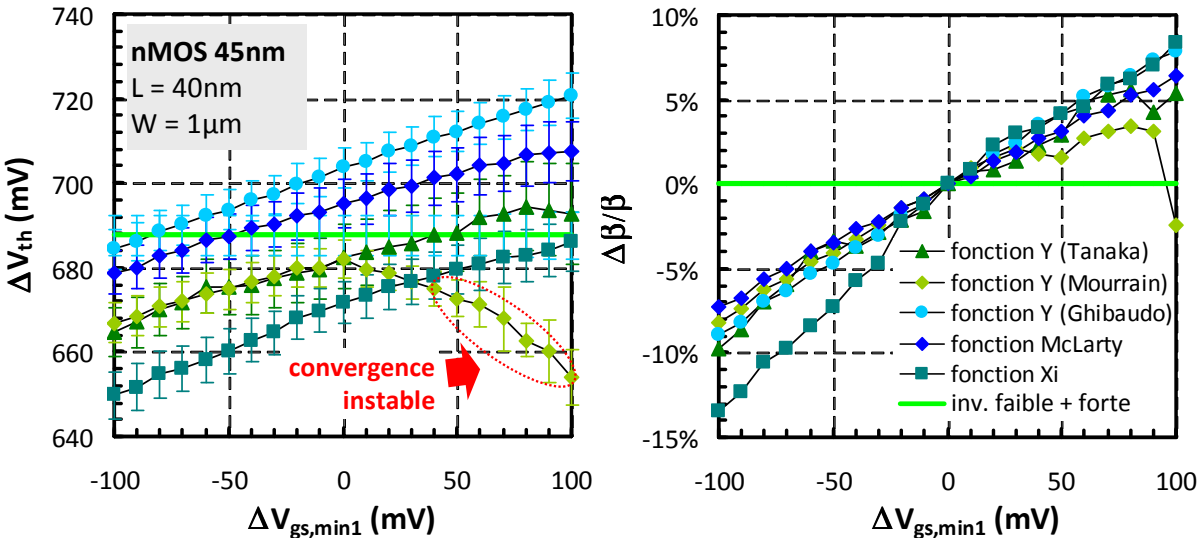


Fig. III.25 – Dérive des valeurs V_{th} et β issues des méthodes de type fonction Y, en fonction du décalage de $V_{gs,min1}$, utilisé dans la définition de la plage d'extraction. L'avantage de la nouvelle méthode "inversion faible + forte", qui ne nécessite pas de plage d'extraction, ressort clairement.

III.6 L'extraction des résistances séries

A mesure que les dimensions des transistors sont réduites, la réduction de la profondeur de jonction et la diminution de la longueur effective tendent à augmenter l'importance des résistances séries ($R_{sd} = R_s + R_d$) devant la résistance de canal (R_{ch}). La correction de R_{sd} est donc nécessaire afin d'isoler la contribution de R_{ch} dans la mesure et d'analyser pertinemment les propriétés de transport dans le canal du transistor.

La résistance totale R_{tot} peut se décomposer en fonction de R_{ch} et R_{sd} comme dans [Hu87] (III.44). Dans cette expression, la valeur R_{sd} inclut la résistance de la source et du drain, ainsi que la résistance de contact et du câblage (qui peut être négligée pour des dispositifs étroits).

$$R_{tot}(V_{gs}) \stackrel{\text{def}}{=} \frac{V_{\text{ds}}}{I_d} = R_{ch}(V_{gt}) + R_{sd}(V_{gs}) \qquad \text{(III.44)}$$

La tension de grille module la concentration en porteur des zones de recouvrement des extensions S/D sous la grille, et dans le reste des extensions par le biais du champ électrique. La résistance série présente donc une variation en fonction de V_{gs} qui doit être prise en compte dans la relation (III.44) [Sheu84, Antognetti81]. L'extraction de la résistance série est un point clé dans l'étude du transistor MOS et de nombreuses méthodes ont déjà été publiées à ce sujet dans la littérature. La plupart d'entre elles sont basées sur la relation de proportionnalité qui existe entre R_{tot} et L_{eff} lorsque la mobilité est supposée ne pas dépendre de la longueur du dispositif [Sheu84, Taur92, Hu87, Niu00, Chang07]. Ces techniques renvoient des valeurs erronées lorsqu'elles sont appliquées sur des dispositifs à canaux courts car des mécanismes additionnels (poches de surdopage, contraintes mécaniques, dégradation de la mobilité) font varier la mobilité avec L_{eff} et corrompent l'hypothèse de proportionnalité sur laquelle reposent les techniques [Andrieu05, Cros06b]. De la même manière, la méthode *Shift and Ratio* et ses améliorations développées dans [Taur92, Cretu01] renvoient des résultats erronés lorsque la mobilité ne peut pas être considérée comme constante (en fonction de L). En particulier pour les technologies

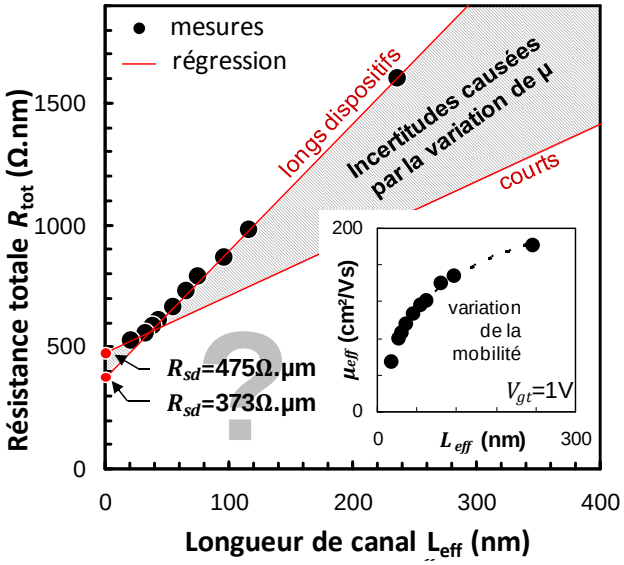


Fig. III.26 – Incertitude sur l'extraction de la résistance série avec la méthode $R_{tot}(L)$ lorsque les variation de mobilité ne sont pas compensées. La valeur correcte est estimée à environ $100\Omega \cdot \mu$m pour cette technologie.

ayant des poches de surdopage pour contrer les effets de canal court [Meer00].

La figure III.26 illustre l'incertitude qui se répercute sur l'extraction de R_{sd} lorsque l'hypothèse de mobilité constante est injustement admise dans les techniques du type $R_{tot}(L)$. Si la correction des variations de mobilité pourrait être envisagée, l'extraction de L_{eff} et μ_{eff} nécessite un lourd protocole expérimentale (équipements, temps de test) [Fleury08a], et la procédure de correction peut nécessiter des algorithmes complexes. Ces derniers reposent sur des hypothèses difficiles à vérifier telles que l'invariance de la mobilité à fort champ [Lin07] ou la convergence des calculs itératifs [Chen06, Kim08], qui rendent ces méthodes peu adaptées à un environnement industriel.

La technique $R_{tot}(L)$ peut être un moyen efficace de détecter des variations de résistances séries entre deux familles de transistors qui ont **exactement la même mobilité et la même longueur effective**. Ces conditions sont rarement respectées car un changement dans l'architecture S/D du transistor s'accompagne fréquemment d'une modification des propriétés de transport du canal, en particulier pour les dispositifs courts. Ainsi, dans la majorité des cas la méthode fournit des écarts relatifs erronés qui ne permettent pas de conclure objectivement.

III.6.1 État de l'art : la méthode $\Theta(\beta)$

La technique d'extraction de R_{sd}, basée sur l'utilisation de la fonction Y, parait la plus apte à extraire la valeur de la résistance série dans les transistors modernes, car elle ne requiert pas d'hypothèse sur la mobilité. Elle repose sur les équations du modèle analytique $I_d(v_{gs})$ développé au §III.2 et utilise le fait que les paramètres $\Theta_{1,eff}$ et $\Theta_{2,eff}$ contiennent les termes du développement de $R_{sd}(V_{\text{gs}})$ à l'ordre 1 en régime d'inversion forte, cf. (III.9). En effectuant l'extraction des paramètres du modèle sur un ensemble de transistors ayant différentes longueurs (et donc des valeurs de β différentes), il est possible de séparer la contribution de la résistance

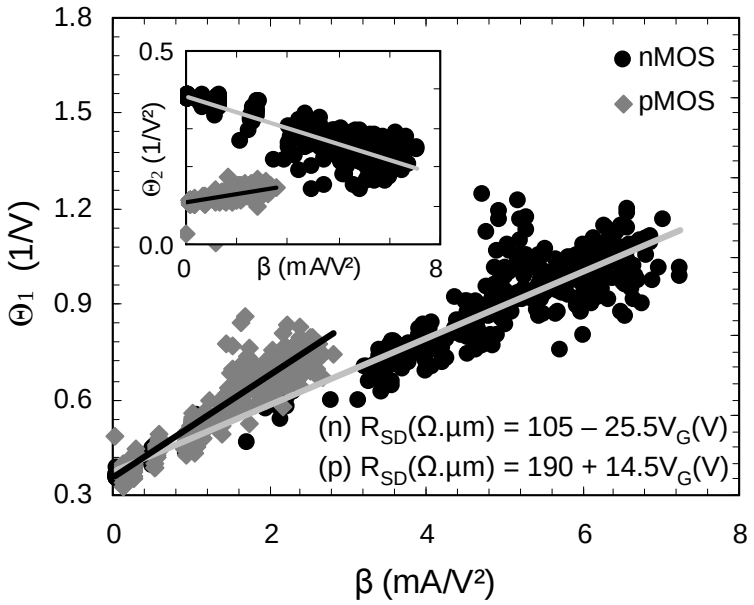


Fig. III.27 – Extraction de $R_{sd}(V_{\text{gs}})$ pour les nMOS et pMOS de la technologie 45nm. Les longueurs de grille vont de 35nm à 1μm, la largeur étant fixée à $W = 1\mu$m.

série des paramètres intrinsèques de la réduction de mobilité : Θ_1 et Θ_2 Ainsi, le terme constant de R_{sd} ($R_{sd,0}$) peut être extrait depuis une régression linéaire effectuée sur les points $\Theta_{1,eff}(\beta)$, et sa dépendance en tension $\lambda_{R_{sd}}$ peut être déduite de la même manière à partir des valeurs $\Theta_{2,eff}(\beta)$ [Ghibaudo88, Cros05]. L'extraction de $R_{sd,0}$ et $\lambda_{R_{sd}}$ est présentée sur la figure III.27 pour des transistors MOS issus de la technologie 45nm. Les pMOS montrent un paramètre $\lambda_{R_{sd}}$ positif qui semble indiquer une augmentation de la résistance série avec V_{gs}. Ce point semble en désaccord avec la littérature [Kim02a, Cros05] et reste inexpliqué à ce jour.

III.6.1.a Précision et fiabilité des résultats

La précision de la méthode $\Theta(\beta)$ est fortement dégradée par les difficultés dans l'extraction du paramètre Θ_1. Ce dernier est en effet corrélé de très près aux variations de tension de seuil et à la valeur de β et présente une erreur type plus large que tous les autres paramètres du modèle. Si la méthode $\Theta_{1,eff}(\beta)$ semble fournir des valeurs précises en relatif, la précision sur la valeur absolue de R_{sd} semble lourdement affectée par l'incertitude sur la valeur de $\Theta_{1,eff}$. C'est ce qui est illustré sur la figure III.30 qui présente l'extraction de R_{sd} depuis le graphique $\Theta_{1,eff}/\beta(1/\beta)$. La large incertitude sur le paramètre $\Theta_{1,eff}$ se répercute sur l'extraction de R_{sd} qui devient difficile à estimer précisément pour cette technologie (CMOS 45nm LSTP).

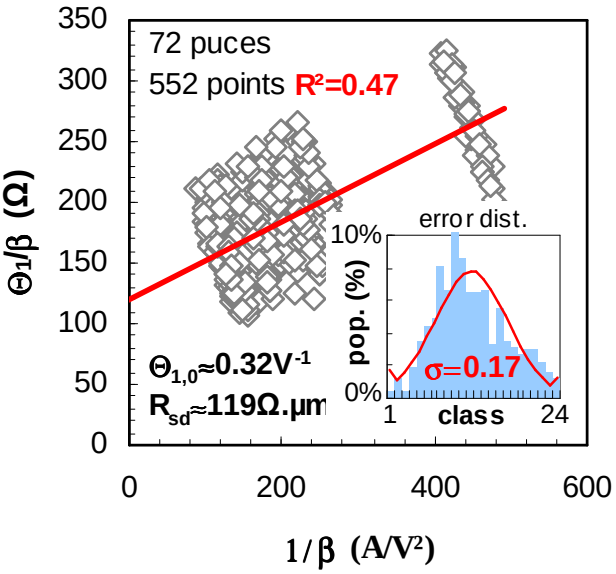


Fig. III.28 – L'incertitude sur l'extraction du paramètre $\Theta_{1,eff}$ se répercute sur la valeur extraite de R_{sd}.

III.6.2 Proposition d'une nouvelle méthode d'extraction : $R_{tot}(1/\beta)$

Comme nous venons de le voir, la méthode $\Theta_1(\beta)$ est à ce jour la seule qui permet une extraction précise de R_{sd} sur les technologies récentes dans lesquelles la mobilité ne peut plus être considéré comme invariante avec L_{eff}. Néanmoins cette méthode présente des inconvénients qui laissent de la place à des améliorations. En particulier, elle nécessite une extraction précise des paramètres $\Theta_{1,2}$ qui ne peut se faire que grâce à des méthodes avancées (fonctions Y et ξ [Ghibaudo88, Fleury08a]), difficilement intégrables dans des procédures de test en ligne.

Nous proposons donc une nouvelle méthode qui conserve la simplicité de la méthode $R_{tot}(L)$ [Taur98] où L est remplacé par le paramètre β. Cette méthode est donc similaire à celle présentée dans [Katto97] mais ne nécessite pas de négliger le facteur quadratique de réduction de mobilité Θ_2 et permet de prendre en compte la variation de R_{sd} avec V_{gs} [Fleury09b, Fleury09c]. Elle est basée sur le modèle $I_d(V_{gs})$ présenté précédemment (III.6) qui peut être réécrit comme (III.45) en utilisant la relation (III.44).

$$R_{tot} = \frac{1}{\beta} \cdot \left(\frac{1 + \Theta_1 \cdot V_{\mathrm{gt}} + \Theta_2 \cdot {V_{\mathrm{gt}}}^2}{V_{\mathrm{gt}}} \right) + R_{sd} \qquad \text{(III.45)}$$

Si la valeur de V_{gt} est fixée une fois pour toutes pour l'ensemble des dispositifs, la relation (III.45) fournit un moyen d'exprimer la dépendance entre R_{tot} à β et d'extraire R_{sd} au point d'interception de l'axe des ordonnées avec la régression linéaire (de type moindre carré [Press07]). Similairement à l'extraction de type $R_{tot}(L)$, l'extraction de R_{sd} par cette méthode consiste à trouver la valeur de R_{tot} extrapolée à un gain infini ($\beta \to \infty$), tandis que la pente de la régression retourne la réduction de mobilité $1 + \Theta_1 \cdot V_{\mathrm{gt}} + \Theta_2 \cdot {V_{\mathrm{gt}}}^2$.

Cette méthode combine à la fois la simplicité de la méthode $R_{tot}(L)$ et la justesse de la méthode $\Theta_1(\beta)$ (pas d'hypothèse sur la mobilité). En évitant l'utilisation du paramètre Θ_1, elle permet d'obtenir une précision accrue dans l'extraction de R_{sd} et reste compatible avec une intégration sous forme de test en ligne.

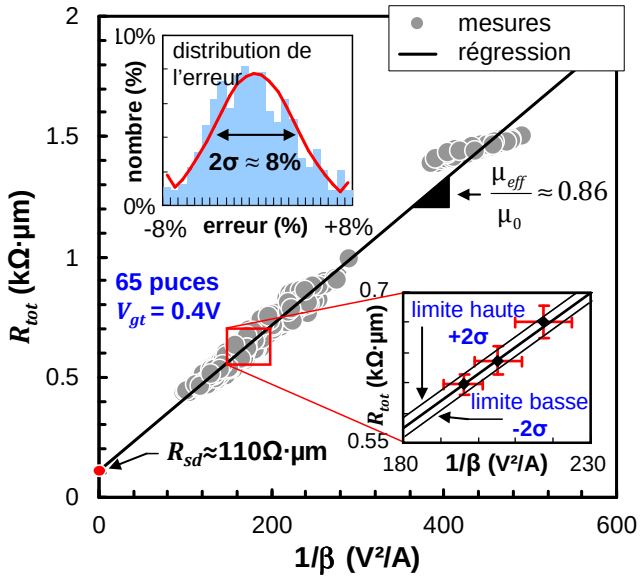


Fig. III.29 – Extraction de R_{sd} sur des transistors nMOS sur silicium massif en utilisant la nouvelle technique, pour $V_{gt} = 0.4$V. La valeur de R_{sd} est lue au point d'intersection tandis que la pente permet d'extraire l'amplitude de la réduction de mobilité pour cette polarisation. La distribution de l'erreur par rapport à la régression linéaire et les limites hautes et basses $\pm 2\sigma$ sont indiquées en insertion.

III.6.2.a Résultats expérimentaux

La technique d'extraction à été validée sur des dispositifs de la technologie 45nm LSTP (CET = 22Å) ayant des longueurs effectives variant entre 22nm et 240nm. La mesure de courbes $I_d(V_{\rm gs})$ à grande échelle (plaque entière : 65 puces) rend possible une analyse statistique des résultats et une étude poussée de la précision des résultats obtenus. La tension de seuil et le paramètre de gain (β) on été extraits grâce à la fonction ξ (cf. §III.4.3 [Fleury08a]) et la valeur de R_{tot} a été mesurée pour plusieurs tensions $V_{\rm gt}$ en utilisant la relation (III.44).

La figure III.29 présente l'extraction de R_{sd} pour $V_{\rm gt} = 0.4$V. Les points ont été préalablement filtrés de manière récursive avec une tolérance de $\pm 3\sigma$ (bande de confiance 99%), permettant de supprimer les points non-significatifs (transistors non fonctionnels ou mesures en dehors de la plaque). La valeur extraite dans notre cas ($R_{sd} = 110\Omega \cdot \mu$m) est cohérente avec la valeur obtenue depuis l'application de la méthode $\Theta_1(\beta)$ mais présente une meilleur précision du fait qu'elle n'utilise pas le paramètre Θ_1 qui propage une large erreur sur R_{sd}. Pour $V_{\rm gt} = 0.4$V, l'erreur relative entre la mesure et la régression est distribuée normalement et présente un faible écart type ($\sigma_{err} \simeq 4\%$, cf. histogramme de la figure). En comparaison, les résultats provenant de la technique $\Theta_{1,eff}(\beta)$ présente une erreur beaucoup plus dispersée : $\sigma_{err} \simeq 16\%$.

Finalement l'extraction a été étendue aux deux types de transistors (n et p), ainsi qu'à une architecture FDSOI [Aime07] pour vérifier la cohérence des résultats avec l'extraction de type $\Theta_1(\beta)$. Les valeurs sont reportées dans le tableau III.1. Elles confirment que les résultats sont en ligne pour les deux techniques, mais également que notre nouvelle méthodologie permet de fortement réduire l'imprécision sur l'extraction de R_{sd}. Cette différence est particulièrement bien mise en valeur sur l'architecture FDSOI et en général pour les transistors de type p.

$R_{sd}@V_{gs} = 1.1V(\Omega \cdot \mu m)$	from this work	from [Ghibaudo88]
nMOS sur silicium massif [Josse06]	110 ± 9	119 ± 22
pMOS sur silicium massif [Josse06]	170 ± 14	160 ± 34
nMOS FDSOI [Aime07]	160 ± 24	190 ± 92
pMOS FDSOI [Aime07]	270 ± 41	320 ± 100

Tab. III.1 – extraction de R_{sd} à $V_{gs} = 1.1V$ pour des dispositifs sur silicium massif et FDSOI

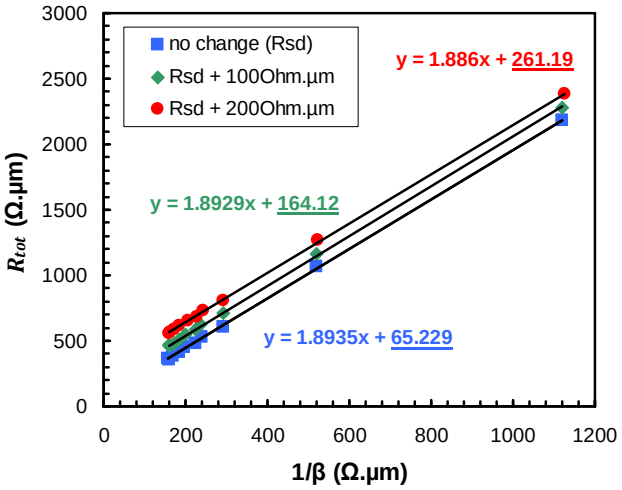


Fig. III.30 – Extraction de R_{sd} par la méthode $R_{tot}(1/\beta)$ en appliquant volontairement un décalage de R_{sd} sur la courbe $I_d(V_{gs})$. Les valeurs extraites reproduisent bien les variations appliquées, confirmant la précision relative de la nouvelle méthode d'extraction.

III.6.2.b Robustesse des résultats

Finalement, la robustesse des résultats a été testée en modifiant la valeur de la résistance série artificiellement (comme au §III.2.2), afin de vérifier que le résultat de l'extraction suit les changements et offre donc une bonne précision relative. Les résultats sont présentés sur la figure III.30 pour un décalage de la valeur de R_{sd} allant jusqu'à $200\Omega \cdot \mu$m. Ils confirment que la méthode permet une excellente précision relative et confirme également la stabilité du paramètre β issu de l'extraction par la fonction ξ.

III.6.2.c Le choix d'une gamme de longueur dans l'extraction de R_{sd}

L'implantation des poches de surdopage, utilisées pour augmenter le dopage de canal sur les petits transistors et contrecarrer les effets de canaux courts, modifie fortement la réponse électrique d'un transistor court par rapport à un transistor long. Le *contraste* de dopage entre les poches et le canal est si élevé (environ 2 décades) qu'il devient difficile de considérer que les transistors longs et les transistors courts sont de même architecture. En particulier, les poches rendent le dopage de canal inhomogène sur des transistors longs et l'écart de dopage sur la gamme entière de longueur est tel qu'il invalide l'hypothèse d'une dégradation de mobilité constante avec la longueur, nécessaire pour la validité de l'extraction.

La solution est de restreindre la gamme de longueurs à des dispositifs pour lesquels les poches de surdopages se rejoignent et forment un dopage quasi-constant et homogène dans le canal. Pour les dispositifs testés dans cette étude, cette distance critique est estimée à $L_{eff} \sim 240$nm en ce

basant sur des simulations TCAD (*Technology Computer Aided Design*). Cela explique la raison pour laquelle toutes nos extractions de R_{sd} étaient jusqu'à présent restreintes à la gamme de longueurs $L_{eff} \leq 240$nm. Bien sûr, ce phénomène n'est pas présent sur les dispositifs totalement désertés (type FDSOI) pour lesquels le canal est non dopé et lorsque l'architecture ne comporte pas de poches de surdopage.

Pour aller plus loin, nous pourrions envisager d'extraire R_{sd} sur une seule longueur, permettant ainsi de s'affranchir des changements d'architectures qui dépendent de la longueur. Ceci reste pour l'instant impossible avec les méthodes que nous venons de détailler.

III.6.3 "Le Graal" : vers l'extraction de R_{sd} sur un seul transistor ?

Le paramétrage sur la longueur de grille pose des problèmes lorsque le comportement de deux transistors de L_{eff} voisins n'ont plus le même comportement. Pour résoudre ce problème, Chen et al. proposent un algorithme d'optimisation qui permet d'extraire directement R_{sd} et μ_{eff} sur une petite gamme de longueurs [Chen06]. La méthode consiste à optimiser un modèle non linéaire à 4 paramètres sur la partie de forte inversion de la courbe. Compte tenu de ce que nous avons vu au §III.2.3, il parait donc impossible de pouvoir séparer les variations mobilité de l'influence de R_{sd} sur la caractéristique du transistor avec précision, d'autant plus que la dépendance de R_{sd} avec V_{gs} est négligée [Kim02a].

Dans [Lin07], une nouvelle méthode est proposée pour permettre l'extraction de la résistance série sur un seul dispositif depuis des mesures de courant I_d à différents V_{bs}. Cette technique repose sur l'hypothèse qu'à fort champ, la mobilité à champ effectif fixé reste la même quelque soit V_{bs} : c'est la définition même du champ effectif et de la mobilité universelle introduite par [Sabnis79, Takagi94a, Sun80]. Le calcul dans [Lin07] repose sur la relation qui permet d'exprimer le champ effectif $\mathcal{E}_{\text{eff}}$ en fonction de Q_{dep} et Q_{inv}. L'utilisation d'expressions analytiques simplifiées, valables en inversion forte, permet alors de travailler à champ constant (III.46) [Lundstrom06a, Skotnicki00] et d'aboutir à une relation simple pour déterminer R_{sd} à partir de deux $I_d(V_{gs})$ effectuées à des tensions V_{bs} différentes (III.47)[5]. Dans cette expression $A = V_{gs}^{V_{b1}} - V_{th}^{V_{b1}} - 0.5V_{ds}$ et $B = V_{gs}^{V_{b1}} + (\eta+1)\cdot V_{th}^{V_{b1}} - \eta \cdot V_{th}^{V_{b2}} - 0.5V_{ds}$.

$$\mathcal{E}_{\text{eff}} = \frac{C_{\text{eff}}}{\varepsilon_{Si}} \cdot \frac{V_{\text{gs}} + (\eta-1)\,V_{\text{th}} - \eta V_{fb} - 2\eta\Psi_b}{\eta}, \quad V_{\text{gs}} \gg V_{\text{th}} \tag{III.46}$$

$$R_{sd} = \left(\frac{B}{I_d^{V_{b2}}} - \frac{A}{I_d^{V_{b1}}}\right) \cdot \frac{V_{ds}}{\eta\left(V_{th}^{V_{b1}} - V_{th}^{V_{b2}}\right)} \tag{III.47}$$

Cette méthode devrait permettre d'extraire R_{sd} sur un dispositif isolé par le biais de la relation mais en réalité, les résultats qu'elle fournit semblent extrêmement dépendant du choix de la valeur de η, qui est mal connu et difficile à estimer avec précision. De plus, cette méthode ne permet pas de prendre en compte la dépendance de la résistance série avec V_{gs} et requiert des champs effectifs suffisamment élevés pour que l'hypothèse de mobilité constante soit justifiée. Ce dernier point est limitant dans l'utilisation de cette technique car c'est la valeur de la résistance série au point de fonctionnement du transistor qui est du plus grand intérêt (en générale $\mathcal{E}_{\text{eff}}(V_{dd}) \sim 1$MV/cm). Cette valeur est largement en deçà des champ critiques nécessaires

5. La validité de la technique requiert également que l'écart entre le potentiel de Fermi extrinsèque et intrinsèque et la tension de bande plate restent constants lorsque la polarisation change.

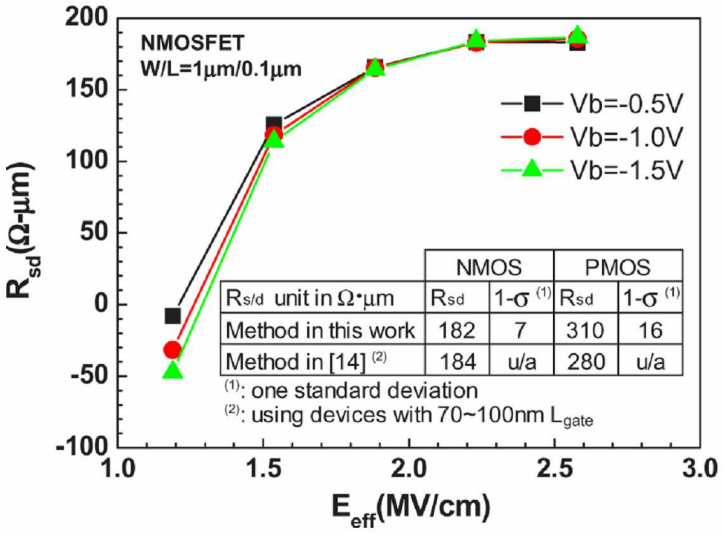


Fig. III.31 – Extraction de la résistance série en fonction du champ effectif, d'après [Lin07]

pour permettre l'observation d'une mobilité qui ne dépend plus de Q_{dep} (en particulier pour une technologie LSTP), du fait de la forte influence des collisions avec les centres chargés (cf. §II.2.2).

III.7 Discussion et perspectives

La construction et l'utilisation des méthodologies d'extraction ont été abordées dans ce chapitre. La caractérisation électrique ne consiste pas à extraire des paramètres d'un modèle quelconque, mais à pouvoir extraire de l'information physique (sur l'architecture, l'électrostatique ou le transport) depuis les paramètres extraits : **c'est là toute la difficulté !**
Dans cet objectif, les méthodologies présentées dans ce chapitre (et utilisées ultérieurement) sont basées sur des modèles semi-empiriques qui permettent de conserver une information physique dans les paramètres.

Que ce soit pour extraire de la tension de seuil ou des paramètres liés au transport (gain en transconductance, réduction de mobilité), les mêmes difficultés surgissent lorsqu'il s'agit d'analyser le transistor court. **Il faut s'affranchir de l'influence des résistances séries et assurer une robustesse suffisante pour permettre une comparaison entre des transistors ayant subit des procédés différents** (entrainant potentiellement un changement de mobilité, de longueur effective etc.). Ce point constitue le problème majeur que les techniques basées sur la fonction Y semblent résoudre, en particulier grâce aux innovations qui ont été apportées au cours de ce travail de thèse.

La connaissance des résistances séries est nécessaire pour évaluer les performances d'une technologie, mais également pour permettre une correction afin de remonter aux propriétés intrinsèques du canal (vitesse, mobilité par technique *split*-CV). **La nouvelle méthode que nous avons introduite semble aujourd'hui être le moyen le plus fiable pour évaluer la valeur de R_{sd} sur des transistors MOS en milieu industriel**. Le paramétrage sur les longueurs de grille (nécessaire pour faire varier β et permettre une régression) semble toutefois inévitable. Il devient problématique sur des technologies dans lesquelles des transistors courts ($L \leq 250$nm) et longs ($L \geq 500$nm) ont des comportements électriques radicalement différents (le dopage canal varie de plus d'une décade). La perspective d'une méthode d'extraction qui ne nécessite pas de paramétrage en longueur et permet d'évaluer R_{sd} sur un unique transistor suscite donc un grand intérêt dans la communauté scientifique. Cela restera pourtant impossible sans un levier supplémentaire qui permettra de décorréler l'influence de R_{sd} de la réduction de mobilité qui agit de manière similaire sur une courbe $I_d(V_{gs})$.

Quelques *briques* contribuant à l'innovation dans les méthodes de caractérisation ont été posées dans ce chapitre. **Elles ouvrent de multiples axes de recherches et de développement en guise de perspectives**. En particulier, l'utilisation de la polarisation V_{bs} comme un levier supplémentaire semble prometteuse. Elle fournirait un moyen de décorréler des variables étroitement liées (mobilité et longueur effective, R_{sd}) et pourrait fournir une solution pour extraire R_{sd} sur un seul dispositif. Néanmoins l'influence de V_{bs} sur la mobilité, la longueur effective ou encore la résistance série nécessite une modélisation complexe qui reste à construire.

Chapitre IV

La modélisation et la caractérisation de la capacité MOS : innovations et nouvelles méthodologies

Sommaire

IV.1 La modélisation analytique de la réponse capacitive des transistors MOS . . . 119
IV.1.1 Introduction : la mesure de la réponse capacitive . . . 119
IV.1.2 Construction d'un modèle analytique simple (transistor long) . . . 121
IV.1.3 Prise en compte des capacités parasites de la structure MOS . . . 123
IV.2 Modélisation numérique 1D pour le calcul de la réponse capacitive 130
IV.2.1 Expression analytique de la charge d'espace . . . 131
IV.2.2 Expression analytique de la densité surfacique de charges . . . 132
IV.2.3 Décomposition de la charge d'espace en contributions élémentaires . . . 133
IV.2.4 Décomposition de la charge totale en contributions élémentaires . . . 133
IV.2.5 Calcul de la capacité dynamique d'une distribution de charges . . . 135
IV.2.6 Calcul numérique et distribution de Fermi-Dirac . . . 136
IV.2.7 Prise en compte des effets quantiques dans la modélisation . . . 137
IV.3 Utilisation des mesures capacitives pour l'extraction de la charge et de la mobilité . . . 142
IV.3.1 L'extraction expérimentale de la charge . . . 142
IV.3.2 Cas des transistors courts : influence des capacités parasites . . . 143
IV.3.3 Extraction de la mobilité par la mesure *split*-CV . . . 144
IV.4 L'extraction de la longueur effective (L_{eff}) . . . 146
IV.4.1 Etat de l'art des méthodes d'extraction . . . 146
IV.4.2 Définition précise de la longueur effective de canal . . . 146
IV.4.3 Proposition d'une nouvelle méthodologie d'extraction . . . 147
IV.4.4 Automatisation de l'extraction . . . 150
IV.4.5 Influence de la fuite de grille sur l'extraction . . . 152
IV.4.6 La mesure de L_{eff} : applications pratiques . . . 153
IV.5 Conclusion et perspectives . . . 158

RÉSUME — La compréhension de la réponse capacitive est primordiale car elle permet de recueillir de l'information physique (charges, longueurs) indépendamment des propriétés de transport qui peuvent être mal connues. Il est donc indispensable d'exploiter au maximum les différentes configurations de mesures qui permettent de récupérer de l'information physique sur le dispositif.

- ⋆ La première partie (§IV.1) se consacrera à la modélisation analytique des différentes réponses capacitives. Elle permettra de mieux comprendre la mesure capacitive et en particulier l'influence des capacités parasites propres aux dispositifs courts.
- ⋆ Dans une seconde partie (§IV.2), nous exposerons les méthodes qui permettent de simuler la réponse capacitive, en utilisant une résolution approchée du système Poisson-Schrödinger.
- ⋆ La troisième partie (§IV.3) détaillera la manière dont les mesures capacitives peuvent être utilisées pour extraire les charges du canal et la mobilité effective des porteurs.
- ⋆ Finalement, la quatrième partie (§IV.4) exposera les innovations qui ont été apportées à la mesure de la longueur effective par voie capacitive, grâce à une meilleure compréhension des capacités parasites de la structure. Nous donnerons des exemples particulier de l'utilisation de la mesure de longueur effective en fin de partie afin de mieux se rendre compte de l'impact que l'introduction d'une méthode fiable et robuste représente dans un milieu industriel.
- ⋆ Nous conclurons finalement sur les résultats et les perspectives de cette études au IV.5.

IV.1 La modélisation analytique de la réponse capacitive des transistors MOS

IV.1.1 Introduction : la mesure de la réponse capacitive

Par définition, l'électromagnétisme définit la capacité comme la propriété d'un matériau à stocker une quantité de charges Q. Elle se calcule en mesurant la variation de charges ΔQ résultant d'une variation de potentiel ΔV aux électrodes du matériau, via la relation suivante : $\Delta Q = C \times \Delta V$. Il en vient la relation générale d'une capacité dynamique que nous utiliserons dans la suite de ce chapitre (IV.1).

$$C \stackrel{\text{def}}{=} \frac{\partial Q}{\partial V} \tag{IV.1}$$

La mesure des capacités entres les différentes électrodes du transistor MOS permet de remonter à des informations essentielles sur l'électrostatique de la structure, la charge et la présence éventuelle de défauts chargés. La technique de mesure a été particulièrement enrichie par la méthode dite *split-CV* [Koomen73, Romanjek04b] qui permet de séparer la contribution des trous de celle des électrons dans la réponse capacitive totale $C_{tot}(V_{gs})$. A l'influence près de la capacité de l'oxyde de grille, il est ainsi possible de mesurer séparément (cf. figure IV.1) :

- la réponse de la couche d'inversion : mesure $C_{gc}(V_{gs})$,
- la réponse de la couche d'accumulation et de la zone de désertion : $C_{gb}(V_{gs})$,
- la réponse de toutes les charges (libres et de désertion) : $C_{tot}(V_{gs})$.

La relation (IV.2) établit le lien entre ces trois valeurs lorsque les capacités parasites $C_{par}(V_{gs})$ peuvent être négligées (cas des transistors longs)

$$C_{tot}(V_{gs}) = C_{gc}(V_{gs}) + C_{gb}(V_{gs}), \quad C_{par}(V_{gs}) \ll C_{ox} \tag{IV.2}$$

La chute brutale qui s'observe dans la partie accumulation de la courbe $C_{gb}(V_{gs})$ ($V_{gs} \sim 0$V) résulte de la formation de la couche d'inversion, assimilable à une plaque métallique connectée à la masse, qui écrante la réponse des charges de désertion [Lau99] selon le schéma de la figure IV.2. Ce phénomène d'écrantage sera abordé plus en détail et modélisé analytiquement dans la suite de cette partie.

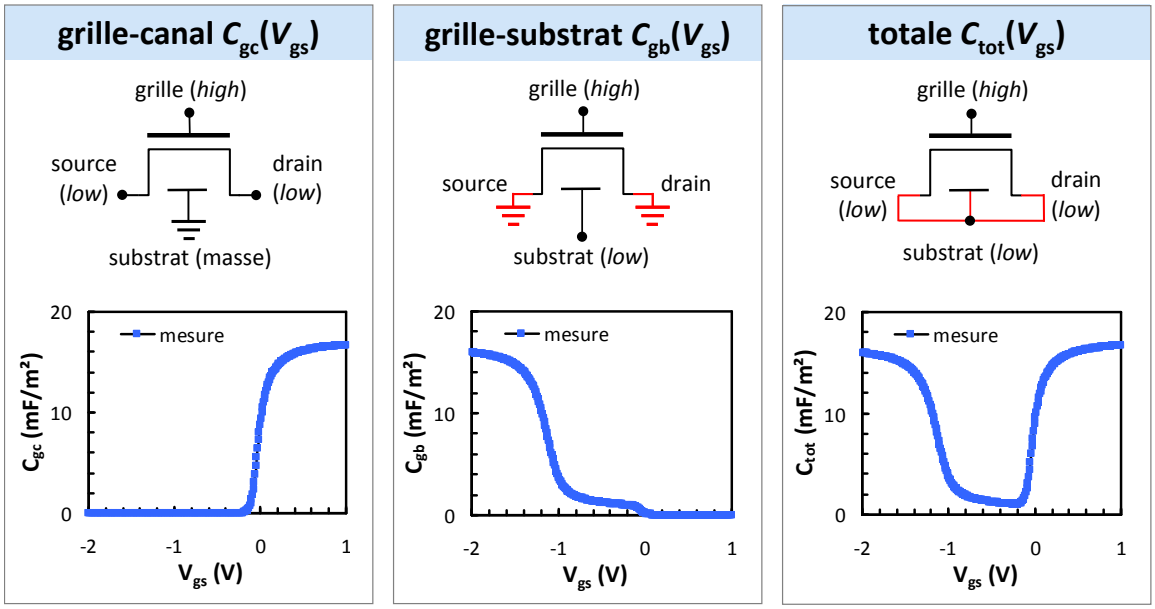


Fig. IV.1 – Différents schémas de câblage possibles pour mesurer la réponse capacitive des transistors MOS.

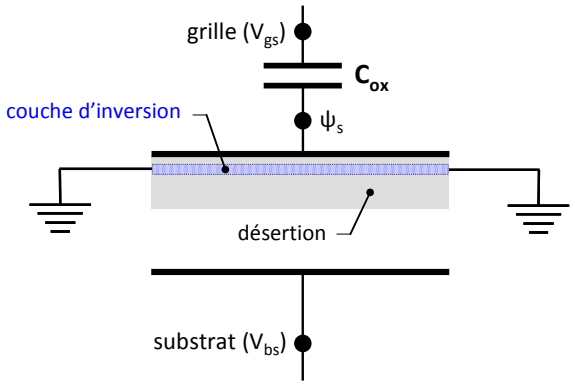


Fig. IV.2 – Schéma équivalent de la capacité C_{gb} en régime d'inversion. La formation d'une couche d'inversion reliée à la masse vient écranter la réponse des charges de désertion.

IV.1.2 Construction d'un modèle analytique simple (transistor long)

Le but de cette partie est de construire un modèle analytique simple qui permettra de décomposer les différents schémas capacitifs présentés sur la figure IV.1 en fonction des contributions élémentaires propres à l'architecture du transistor MOS. Dans un premier temps nous négligeons les capacités parasites et portons notre attention sur la modélisation de la réponse intrinsèque du transistor (charges libres, de désertion et capacité de l'oxyde de grille). Cette approximation reste valable pour les transistors longs ($L \geq 1\mu$m).

La mesure de la capacité totale permet de voir les contributions de toutes les charges (accumulation, inversion, désertion) [Sze81]. La réponse $C_{tot}(V_{gs})$ peut donc s'écrire comme (IV.3) qui peut être développée en fonction des différentes contributions capacitives de la structure (IV.4) si Q_{tot} est décomposée comme suit : $Q_{tot} = Q_{inv} + Q_{acc} + Q_{dep}$.

$$C_{tot}(V_{gs}) \stackrel{\text{def}}{=} \left[(C_{ox})^{-1} + \left(-\frac{\partial Q_{tot}}{\partial V_{gs}} \right)^{-1} \right]^{-1} \tag{IV.3}$$

$$C_{tot}(V_{gs}) = \left[(C_{ox})^{-1} + (C_{inv} + C_{acc} + C_{dep})^{-1} \right]^{-1} \tag{IV.4}$$

De la même manière, la réponse capacitive grille-canal $C_{gc}(V_{gs})$ est caractérisée par la réponse des charges d'inversion et peut simplement s'écrire comme (IV.5) : les variations de charges de désertion et d'accumulation ne sont pas visibles dans ce schéma de mesure.

$$C_{gc}(V_{gs}) = \frac{C_{ox} \cdot C_{inv}}{C_{ox} + C_{inv}}, \quad C_{inv} \stackrel{\text{def}}{=} -\frac{\partial Q_{inv}}{\partial \psi_s} \tag{IV.5}$$

Finalement, la réponse capacitive grille-substrat $C_{gb}(V_{gs})$ ne peut pas s'écrire sous une forme analogue à (IV.5) car l'écrantage de la capacité de désertion de la couche de désertion ne serait pas pris en compte dans un tel modèle [Lau99]. Il est néanmoins possible d'obtenir une expression analytique en partant de l'égalité (IV.2) qui permet d'écrire (IV.6). Comme le montre la figure IV.3, cette relation offre un bon accord avec la mesure.

$$C_{gb}(V_{gs}) = C_{tot}(V_{gs}) - C_{gc}(V_{gs}) \tag{IV.6}$$

IV.1.2.a Ecrantage de la capacité de désertion dans la mesure C_{gb}

La relation (IV.6) n'est pas satisfaisante : l'écrantage de la capacité de désertion doit être modélisée analytiquement, afin d'expliciter C_{gb} en fonction des autres composantes capacitives de la structure. Nous définissons donc la fonction d'écrantage $F_{shield}(V_{gs})$ qui nous permet d'écrire $C_{gb}(V_{gs})$ comme (IV.7) via l'opération suivante : $C_{dep} \rightarrow F_{shield} \cdot C_{dep}$.

$$C_{gb}(V_{gs}) = \left[(C_{ox})^{-1} + (C_{acc} + F_{shield}(V_{gs}) \cdot C_{dep})^{-1} \right]^{-1} \tag{IV.7}$$

En réutilisant la relation (IV.6) et en remplaçant $C_{tot}(V_{gs})$ et $C_{gc}(V_{gs})$ par (IV.4) et (IV.5) il est possible d'aboutir à l'expression analytique de $F_{shield}(V_{gs})$. La condition $C_{acc} = 0$ est posée dans le calcul afin de supprimer cette composante nulle en inversion et qui ne contribue donc pas au phénomène d'écrantage dans cette région de la courbe, cf. (IV.8).

$$F_{shield} = \frac{(C_{ox})^2}{(C_{ox} + C_{inv})^2 + C_{dep} \cdot C_{inv}} \tag{IV.8}$$

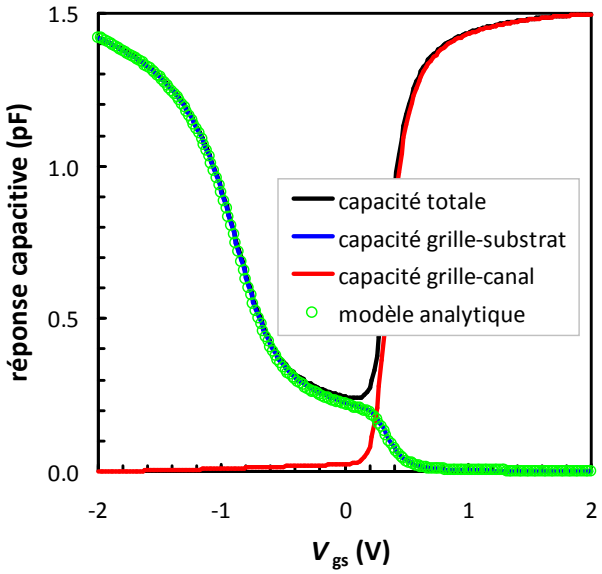


Fig. IV.3 – Comparaison entre modèle analytique (IV.6) et la mesure directe de la réponse capacitive $C_{bg}(V_{gs})$. Le modèle donne un bon agrément avec la mesure, en accord avec [Lau99] et permettant de confirmer sa validité pour notre technologie CMOS 45nm.

Notons que l'expression de $F_{shield}(V_{gs})$ respecte bien les conditions aux limites (IV.9) qui découlent du bon sens physique.

$$F_{shield} = \begin{cases} 1 & \text{si} \quad C_{inv} \to 0 \\ 0 & \text{si} \quad C_{inv} \gg C_{dep} \end{cases} \tag{IV.9}$$

L'expression analytique $F_{shield}(V_{gs})$ est comparée à l'écrantage observé sur la figure IV.4, grâce à des simulations numériques. Le modèle montre un bon accord avec la relation (IV.6), justifiant de la validité de notre relation.

Finalement, la relation (IV.8) peut être généralisée à l'écrantage d'une capacité C quelconque par la réponse capacitive des charges dans le canal (C_s), cf. (IV.10).

$$F_{shield}(C, C_s) = \frac{(C_{ox})^2}{(C_{ox} + C_s)^2 + C \cdot (C_s)} \tag{IV.10}$$

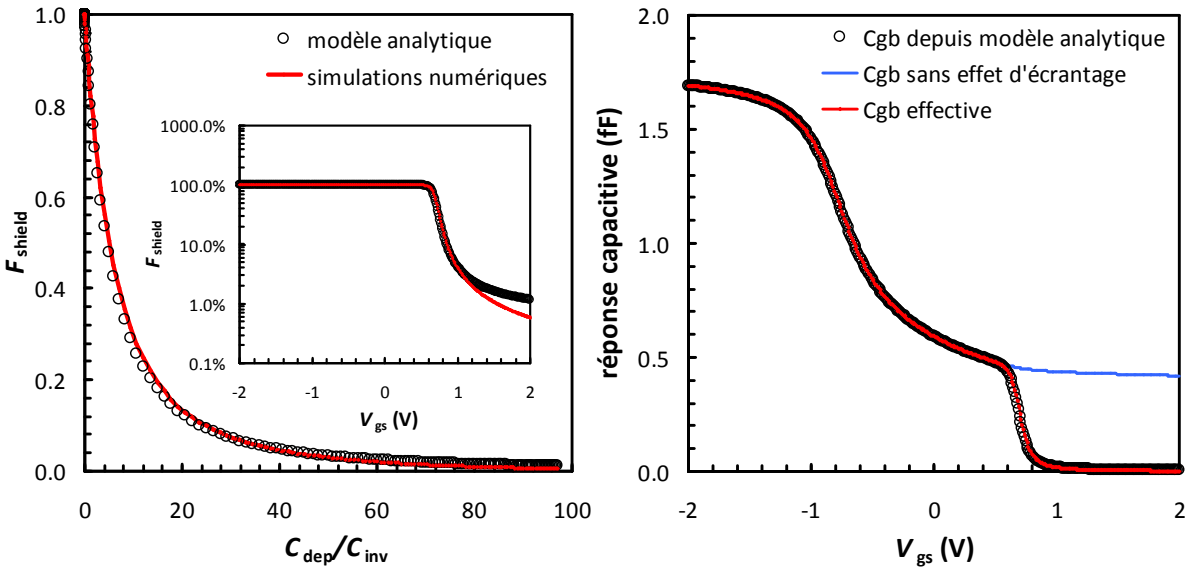


Fig. IV.4 – Validation du modèle analytique (IV.8) grâce à des simulations numériques. La fonction d'écrantage F_{shield} donne un bon accord avec le comportement $C_{gb}(V_{gs})$ et permet de reproduire fidèlement l'écrantage de la capacité de désertion en inversion.

IV.1.3 Prise en compte des capacités parasites de la structure MOS

Les capacités parasites jouent un rôle crucial sur les performances d'une technologie [Khakifirooz08, Deng08], justifiant de la nécessité d'une compréhension solide de la physique sous-jacente. Au delà de l'intérêt technologique, la correction des capacités parasites est souvent nécessaire car la plupart des méthodes d'extraction basées sur des mesures C-V ne tiennent pas compte de ces contributions qui peuvent fausser le résultat. De plus, si des techniques de *de-embedding* permettent de s'affranchir des capacités parasites propres au *backend* du transistor (via des structures de test spécifiques), il est impossible de corriger les capacités parasites inhérentes à l'architecture du transistor par ce biais. Cette partie fera l'état de l'art des modélisations des capacités parasites et exposera les innovations qui ont été apportées aux modèles préexistants.

Deux natures de capacités parasites sont ainsi distinguées dans ce manuscrit :

- **Les capacités parasites de nature extrinsèque**, qui ne sont pas liées à l'architecture du MOSFET mais à l'arrangement des interconnections qui constituent le *backend*. Elles ne dépendent pas (ou très peu) des tensions appliquées au transistor et peuvent être corrigées de la mesure en utilisant des structures de test prévues à cet effet (structures de *de-embedding*).
- **Les capacités parasites de nature intrinsèque**, qui dépendent de l'architecture, i.e. des matériaux et de l'arrangement géométrique des différents éléments de la structure MOSFET (grille, source/drain, espaceurs, etc...). Elles possèdent une forte dépendance avec les tensions appliquées et ne peuvent pas être corrigées simplement en utilisant des structures de *de-embedding*. La valeur des capacités parasites intrinsèque sera appelée C_{par} dans la suite du manuscrit.

Contrairement à la réponse capacitive du canal, l'amplitude de C_{par} ne suit pas une règle de proportionnalité avec la longueur des transistors. L'influence de C_{par} est donc plus importante à mesure que la longueur des dispositifs est réduite, rendant difficile l'analyse des propriétés du

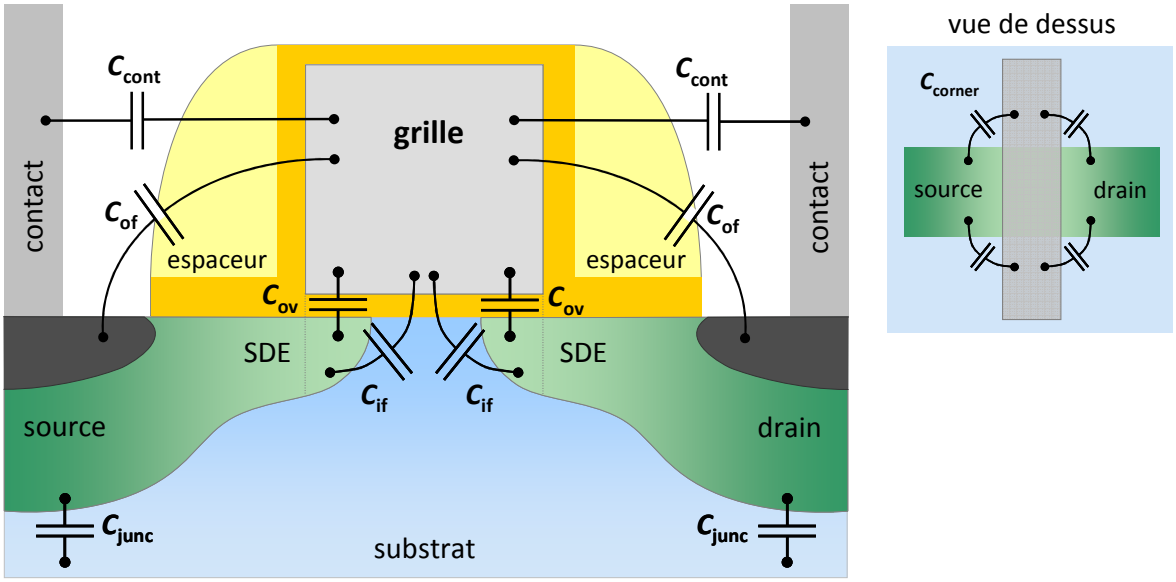


Fig. IV.5 – Schéma illustrant les différentes contributions à la capacité parasite C_{par}, inhérentes à l'architecture du transistor MOS, d'après [Prégaldiny02, Khakifirooz08, Wei09].

canal par voie capacitive. Par analogie avec les résistances séries traitées au §III.6, les capacités parasites doivent elles aussi être mesurées et corrigées afin de permettre une analyse précise des propriétés du canal (longueur effective, charge, mobilité).

IV.1.3.a Un modèle analytique simple pour la prise en compte de C_{par}

Nous ne nous intéresserons pas aux capacités parasites extrinsèques, mais seulement aux contributions qui sont inhérentes à la structure du transistor MOS. Ainsi, un modèle équivalent proposé dans [Prégaldiny02, Shrivastava82] et illustré par la figure IV.5 permet de rendre compte des principales contributions qui s'ajoutent à la réponse intrinsèque du MOSFET. La valeur de C_{par} par unité de largeur est alors donnée par l'expression analytique (IV.11), qui présuppose que les différentes composantes C_{if}, C_{of} et C_{ov} sont indépendantes. La longueur de recouvrement est définie comme la longueur totale des régions où les extensions sont recouvertes par la grille : $L_{ov} \stackrel{\text{def}}{=} L_{gate} - L_{eff}$.

$$C_{par}(V_{gs}) = 2 \times (L_{ov} \cdot C_{ov}(V_{gs}) + C_{if}(V_{gs}) + C_{of}) \tag{IV.11}$$

Dans cette approche :

- $C_{ov}(V_{gs})$ est la capacité planaire prise dans la zone de recouvrement, entre la grille et les extensions source/drain (SDE). Elle possède une dépendance en V_{gs} analogue à celle d'une capacité MOS,
- $C_{if}(V_{gs})$ est la capacité de bord interne, prise entre la grille et les extensions S/D en passant pas le canal. Elle dépend de V_{gs} car l'environnement est modifié lors de l'apparition d'une couche d'inversion ou d'accumulation (effet d'écrantage),
- C_{of} est la capacité de bord externe (couplage entre la grille et les régions source/drain en passant par les espaceurs). Elle possède une très faible dépendance en V_{gs} [Prégaldiny03] qui sera négligée dans notre modèle.
- C_{cont} est la capacité qui caractérise le couplage entre les contacts et la grille, en passant

par les espaceurs,
- C_{junc} est la capacité de jonction qui caractérise le couplage capacitif entre le substrat et les source/drain,
- C_{corner} représente la capacité entre les bords de la grille et ceux de la source et du drain.

Des modélisations analytiques ont déjà été établies pour C_{cont}, C_{junc} et C_{corner} dans [Wei09, Liu09]. Ces contributions ne dépendent pas de la polarisation V_{gs} appliquée, et de ce fait (nous le verrons par la suite) n'ont pas d'influence sur les résultats issus de nos méthodes d'extraction. La suite de cette partie se focalisera donc sur la modélisation analytique des trois autres composantes : C_{if}, C_{ov} et C_{of}.

IV.1.3.b Modélisation de la capacité de bord externe : C_{of}

La capacité de bord externe résulte du couplage électrostatique entre la grille et les régions S/D. Elle devient importante lorsque l'épaisseur du métal de grille est grande devant celle de l'oxyde (t_{ox}). Une expression analytique relativement simple C_{of} a été établie dans [Elmasry82, Shrivastava82] et utilisée dans de nombreux travaux [Wiatr01, Prégaldiny02, Kim06, Wei09]. Elle est basée sur l'hypothèse que les lignes de champ entre les faces latérales de la grille et le substrat ont une forme circulaire, hypothèse qui devient erronée à proximité de la partie inférieure de la grille, où les lignes de champs ont une forme quasi-linéaire [Suzuki99]. Nous préférerons donc utiliser le modèle plus complet détaillé dans [Mohapatra03], qui corrige ce point sans faire intervenir d'itérations complexes comme dans [El Kamchouchi75, Greeneich83]. L'expression de C_{of} par unité de largeur est alors donnée par (IV.12)[1].

$$C_{of} = \beta_1 \cdot \frac{\varepsilon_{ox}}{\pi} \ln\left[\left(K^2-1\right)\cdot\left(\frac{K^2}{K^2-1}\right)^{K^2}\right], \quad K^2 = \frac{L_{spacer}}{t_{ox}} \tag{IV.12}$$

Dans cette expression, β_1 est un paramètre d'ajustement qui permet d'ajuster le modèle qui suppose que les lignes de champ entre la partie inférieure de la grille et le substrat sont parfaitement linéaire. La valeur $\beta_1 \simeq 0.8$ semble donner un bon accord avec la simulation dans [Mohapatra03].

IV.1.3.c Modélisation de la capacité de recouvrement $C_{ov}(V_{gs})$

La contribution capacitive des zones de recouvrement $C_{ov}(V_{gs})$ est traitée comme une capacité planaire prise entre la grille et les extensions S/D, sur la longueur de recouvrement totale L_{ov}. Cette capacité est équivalente à une structure MOS dans laquelle le silicium (i.e. les extensions S/D) est fortement dopé (en général $10^{20-21}/\text{cm}^3$) et pour laquelle la tension de bande plate est proche de 0 ($V_{fb,SDE} \sim 0\text{V}$), car les dopages des extensions et de la grille sont du même type [Prégaldiny03]. Le diagramme de bande caractéristique de la zone de recouvrement en régime d'inversion est représenté sur la figure IV.6.

Si le modèle BSIM3v3 fournit une expression analytique de $C_{ov}(V_{gs})$ [Wakita00], cette dernière est extrêmement simplifiée car basée sur une statistique de Boltzmann invalide pour les dopages élevés en régime d'inversion (i.e. lorsque les extensions sont en accumulation). La résolution de C_{ov} est en réalité aussi complexe que celle de la capacité MOS intrinsèque et nécessite

1. Cette dernière expression est valable dans l'hypothèse où $t_{gate} + t_{ox} \geq L_{spacer}$, dans le cas contraire K^2 s'écrit comme dans [Suzuki99] : $K^2 = 1 + t_{gate}/t_{ox}$

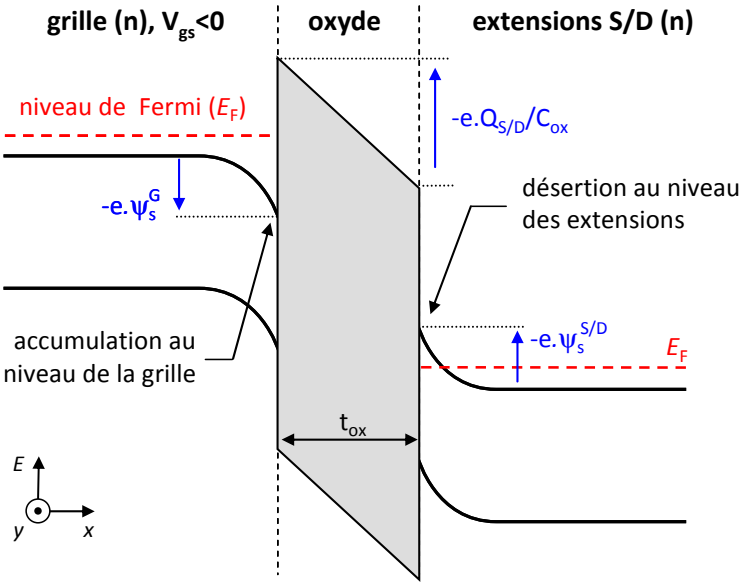


Fig. IV.6 – Diagramme de bande d'une coupe prise dans la largeur de la zone de recouvrement d'un transistor nMOS en régime d'inversion. L'application d'une tension positive sur la grille provoque la désertion des extensions S/D et augmente la valeur de $C_{ov}(V_{gs})$.

une résolution de type Poisson-Schrödinger pour être parfaitement exacte sur toute la plage de V_{gs}. Nous ne chercherons donc pas à établir un modèle analytique pour calculer $C_{ov}(V_{gs})$ et utiliserons le même algorithme de résolution que celui qui sera développé pour résoudre la capacité MOS intrinsèque.

Influence de la capacité de recouvrement C_{ov} – Dans une structure de transistor nMOS avec recouvrement, la désertion des extensions est responsable de la décroissance de la réponse $C_{gc}(V_{gs})$ en régime d'accumulation. Si on considère des jonctions S/D d'abrupteté finie, le phénomène de désertion qui provoque les variations $C_{ov}(V_{gs})$ peut être décomposées en deux mécanismes [Prégaldiny02] :

- la désertion verticale des extensions loin de la jonction, lorsque $V_{gs} \leq V_{fb,SDE}$, où $V_{fb,SDE}$ représente la tension de bande plate caractéristique de l'empilement grille-oxyde-SDE (en général $V_{fb,SDE} \sim 0$V). Ce phénomène augmente l'épaisseur d'oxyde apparente dans le calcul de C_{ov}.
- La désertion des extensions proches de la jonction metallurgique qui tend à diminuer L_{ov} pour des tensions $V_{gs} \leq V_{fb}$, lorsque le canal entre en régime d'accumulation.

IV.1.3.d Modélisation de la capacité de bord interne $C_{if}(V_{gs})$

La capacité de bord interne C_{if} est le résultat d'un couplage électrostatique entre la grille et les extensions S/D, en passant par le canal du transistor (cf. figure IV.5) [Prégaldiny02, Khakifirooz08]. L'expression analytique de C_{if} a été établie pour l'architecture du transistor MOS en condition de bandes plates, lorsque le canal ne contient pas de charge [Shrivastava82], cf. (IV.13). Cette expression résulte d'un calcul d'électrostatique simple et fait intervenir des grandeurs géométriques propres à l'architecture, telles que la profondeur des jonctions X_j (extensions dans notre cas) et l'épaisseur de l'oxyde de grille t_{ox}.

$$C_{if}(V_{fb}) = \frac{\epsilon_{ox}}{\beta} \ln\left[1 + \frac{X_j \sin(\beta)}{t_{ox}}\right], \quad \beta = \frac{\pi}{2}\frac{\epsilon_{ox}}{\epsilon_{Si}} \tag{IV.13}$$

La capacité de bord interne est très dépendante de V_{gs} car elle est écrantée par la forte densité de charges libres qui caractérise les régimes d'accumulation ($V_{gs} \leq V_{fb}$) et d'inversion ($V_{gs} \geq V_{th}$) [Romanjek04b, Severi06, Prégaldiny02]. Dans le régime intermédiaire ($V_{fb} \leq V_{gs} \leq V_{th}$), le canal est en régime de désertion et C_{if} atteint sa valeur maximale car la densité de charges est trop faible pour permettre un écrantage "efficace". Ces explications sont illustrées sur la figure IV.7 qui présente une mesure $C_{gc}(V_{gs})$ effectuée sur une structure spécifique, sans recouvrement ($L_{ov} = 0 \Rightarrow C_{ov} = 0$), qui permet de distinguer clairement les variations de C_{if} dans la zone $V_{gs} \leq V_{th}$.

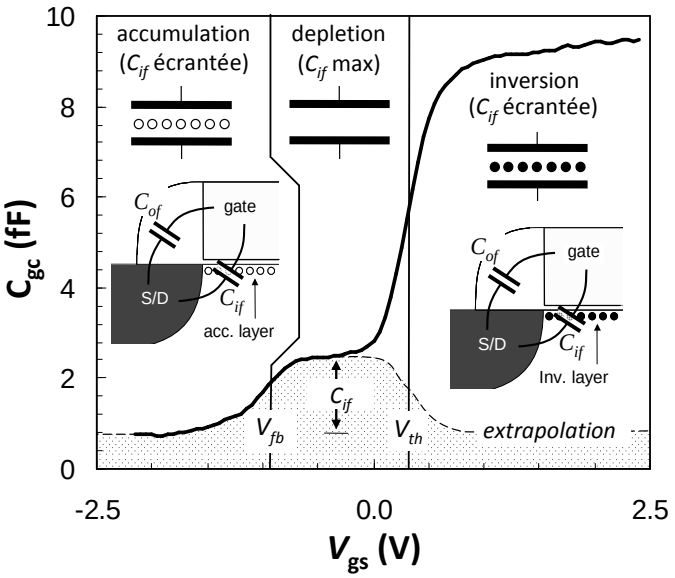


Fig. IV.7 – Mesure $C_{gc}(V_{gs})$ effectuée sur une structure sans recouvrement afin d'observer le comportement de $C_{if}(V_{gs})$ dans la gamme de tensions $V_{gs} < V_{th}$. Le transistor est issu de la technologie 45nm LSTP (L_{mask} = 45nm, $W = 1\mu$m). Lorsque la couche d'accumulation fait son apparition dans le canal ($V_{gs} \leq V_{fb}$) la capacité C_{if} est écrantée, phénomène analogue à l'écrantage de C_{dep} observé sur la mesure C_{gb}. La valeur maximale de $C_{if}(V_{gs})$ est observée dans la gamme $V_{fb} \leq V_{gs} \leq V_{th}$ qui présente un plateau caractéristique.

Le résultat obtenu concernant l'écrantage de la capacité de désertion et peut être réutilisé pour établir l'expression de $C_{if}(V_{gs})$. La capacité responsable du phénomène d'écrantage est alors défini comme $C_s = C_{acc} + C_{inv} + \alpha \cdot C_{dep}$, où α est un paramètre d'ajustement qui pondère l'efficacité de l'écrantage causé par les charges de désertion.

$$C_{if}(V_{gs}) = C_{if,max} \times F_{shield}(C_{if,max}, C_s) \tag{IV.14}$$

La figure IV.8 présente l'évolution d'une courbe $C_{gc}(V_{gs})$ obtenue par simulations numériques pour un transistor de longueur nominale sans recouvrement ($L_{ov} = 0$) de la technologie 45nm, pour différentes valeurs de α. La valeur $\alpha = 1$ semble fournir le meilleur accord avec la mesure expérimentale de la figure IV.7.

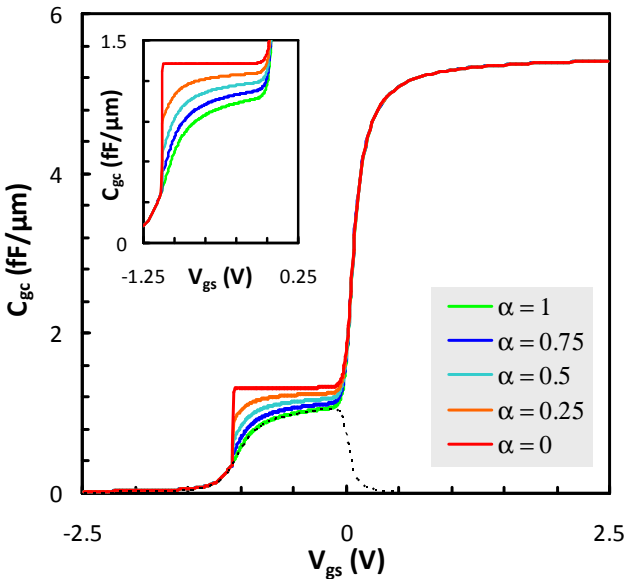


Fig. IV.8 – Simulation d'une courbe $C_{gc}(V_{gs})$ d'un transistor nMOS de longueur nominal de la technologie 45nm (sans recouvrement), pour différentes valeurs de α.

IV.1.3.e Comprendre l'allure de la mesure capacitive grille-canal

Par opposition avec la mesure de capacité C_{tot}, la réponse des charges d'accumulation et de désertion ne se voit pas sur la courbe $C_{gc}(V_{gs})$, permettant d'analyser le comportement des capacités parasites C_{if} et C_{ov} dans la région $V_{gs} \leq V_{fb}$. En effet, en régime d'accumulation, les jonctions entre les extensions S/D et le canal se désertent, empêchant les porteurs majoritaires du canal de pouvoir s'évacuer (cf. figure IV.9). La zone de désertion empêche un contact ohmique entre le canal et les extensions pour ces porteurs. Ainsi, une capacité de jonction (C_{junc}) est en série avec la réponse des charges d'accumulation C_{acc}, selon (IV.15).

$$C_{gc,acc} = \left(\frac{1}{C_{acc}} + \frac{1}{C_{junc}} \right)^{-1} \tag{IV.15}$$

Le comportement $C_{junc}(V_{gs})$ a été simulé pour plusieurs valeurs de dopages N_{sub} comprises entre $10^{21}/\mathrm{m}^3$ et $10^{23}/\mathrm{m}^3$ (cf. insert de la figure IV.16). La valeur de la capacité C_{junc} est 20 fois plus faible que celle de C_{acc} quelque soit la valeur de N_{sub}. Cela justifie le fait que C_{acc} ne se voit pas sur la mesure C_{gc} : son amplitude est fortement atténuée et a une valeur faible devant le niveau de capacité parasite de la structure. De plus, cette modélisation repose sur l'hypothèse des jonctions abruptes, inappropriée pour des jonctions très dopées et qui surestime les valeurs de C_{junc}. Les valeurs typiques de dopages ($N_{sub} \sim 10^{22}/\mathrm{m}^3$, $N_{SDE} \sim 10^{26}/\mathrm{m}^3$) et d'abrupteté ($\sim 10^{34}/\mathrm{m}^4$) sur les technologies 45nm et 65nm résultent en un écrantage beaucoup plus efficace de C_{acc}, rendant cette capacité indétectable dans la mesure C_{gc}.

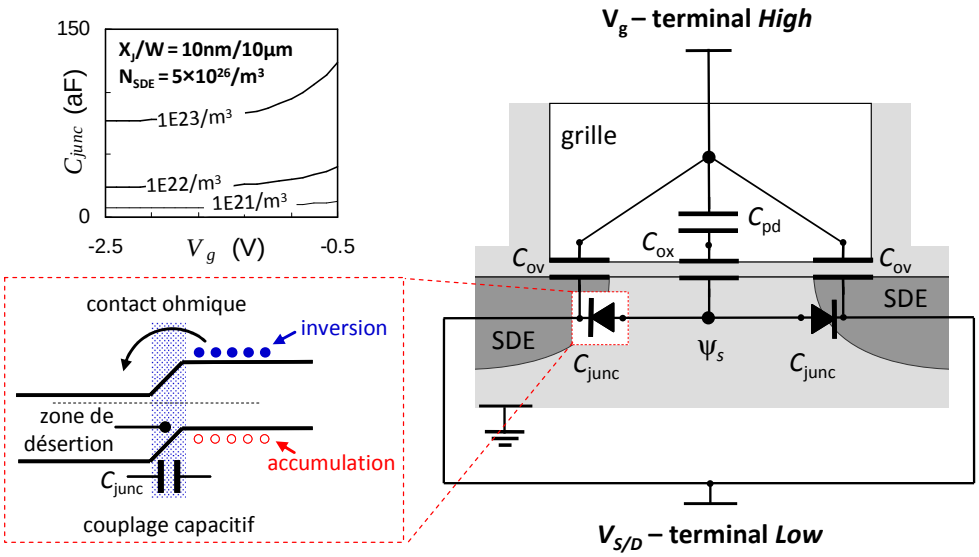


Fig. IV.9 – Schéma équivalent d'un MOSFET au cours d'une mesure C_{gc} (la capacité de bord interne n'est pas représentée). En haut à gauche : évolution de la capacité de jonction $C_{junc}(V_{gs})$ en fonction du niveau de dopage substrat N_{sub}, le dopage des extensions N_{SDE} étant fixé à $5 \times 10^{26}/\mathrm{m}^3$.

IV.2 Modélisation numérique 1D pour le calcul de la réponse capacitive

Cette partie a pour but d'établir les expressions de la charge et de la courbure de bande ψ_s des différents éléments d'une structure MOS (composée d'une grille en silicium polycristallin) correspondant au diagramme d'énergie de la figure IV.10. Le but final est de pouvoir calculer la valeur de la réponse capacitive dans les différentes parties de l'architecture MOSFET afin de reproduire (par simulations) les mesures présentées dans la partie précédente et d'améliorer la compréhension des capacités parasites. Les éléments suivants doivent être considérés :

- La faible épaisseur de l'oxyde de grille amène à prendre en compte les effets de quantification des niveaux d'énergie en surface du canal, ainsi que la désertion de grille. Les effets quantiques doivent donc être considérés dans notre modèle.
- Le modèle doit pouvoir résoudre les charges et les potentiels de l'empilement dans les zones de recouvrement (accumulation et désertion) afin de modéliser $C_{ov}(V_{gs})$. Le silicium dans ces zones est généralement dopé à dégénérescence ce qui invalide l'approximation de Boltzmann et impose de conserver l'intégrale de Fermi-Dirac dans les expressions.

Dans un premier temps, le modèle dit "classique" (i.e. sans prise en charge des effets quantiques) sera développé. Nous détaillerons ensuite la manière dont les effets quantiques seront pris en compte pour permetttre une modélisation plus réaliste.

Nous garderons à l'esprit les points suivants, qui restent valables tout au long de notre exposé :

- Le potentiel de surface, noté ψ_s, correspond à la différence entre la valeur du potentiel en surface du matériau considéré et celle en profondeur : $\psi_s \stackrel{\text{def}}{=} \psi(0) - \psi(\infty)$,
- L'influence des états de charges dans l'oxyde est négligée dans : $C_{ss} = 0$,
- La modélisation porte sur une structure de type silicium-oxyde-silicium

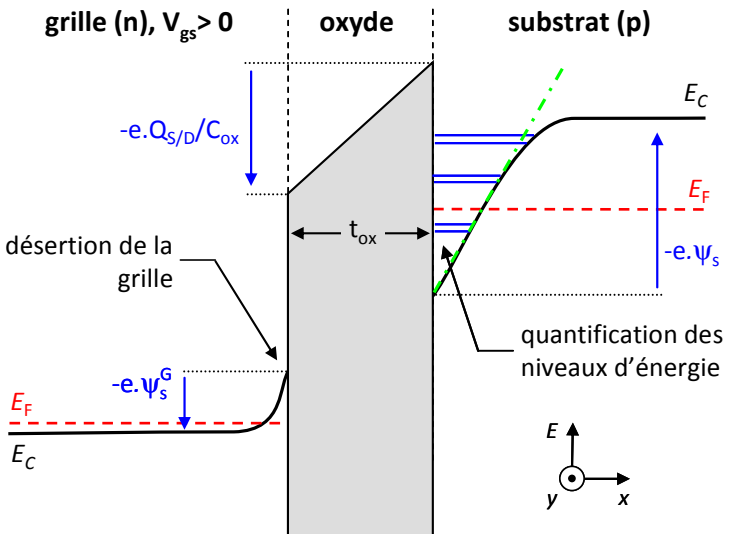


Fig. IV.10 – Diagramme d'énergie de la structure Silicium-Oxyde-Silicium considérée dans notre étude.

IV.2.1 Expression analytique de la charge d'espace

La charge d'espace $\rho(\psi)$ quantifie le surplus d'électrons et de trous par unité de volume en référence au régime de *bandes plates* pour lequel $\psi(x) = 0, \forall x$ et qui caractérise les densités d'électrons et de trous à l'équilibre : n_0 et p_0. Par définition, l'expression de $\rho(\psi)$ est donc donnée par (IV.16), cf. [Sze81]

$$\rho(\psi) = q \cdot [(p(\psi) - p_0) - (n(\psi) - n_0)] \tag{IV.16}$$

Les concentrations $n(\psi)$ et $p(\psi)$ peuvent s'expliciter en fonction des niveaux d'énergie(IV.17) et (IV.18), via la distribution de Fermi-Dirac d'ordre $\frac{1}{2}$ définie en annexe A.

$$n(\psi) = N_C \cdot \mathcal{F}_{1/2}\left(\eta_{nF} - \frac{q\psi}{kT}\right) \tag{IV.17}$$

$$p(\psi) = N_V \cdot \mathcal{F}_{1/2}\left(\eta_{pF} + \frac{q\psi}{kT}\right) \tag{IV.18}$$

Dans ces dernières relations, η_{nF} et η_{pF} représentent l'écart (en unité d'énergie thermique kT) entre le niveau de Fermi (E_F) et la bande de conduction (resp. bande de valence), cf. (IV.19) et (IV.20).

$$\eta_{nF} \overset{\text{def}}{=} \frac{E_F - E_C}{kT} \tag{IV.19}$$

$$\eta_{pF} \overset{\text{def}}{=} \frac{E_V - E_F}{kT} = -E_G - \eta_{nF} \tag{IV.20}$$

Les valeurs N_C et N_V quantifient la densité d'états disponibles dans la bande de conduction et dans la bande de valence. Les valeurs numériques pour le silicium à 300K sont données par (IV.21) et (IV.22). Les variables m_c^{DOS} et m_v^{DOS} représentent les masses effectives de densité d'états dans la bande de conduction et dans la bande de valence. Dans le silicium à 300K elles vallent $m_c^{DOS} \simeq 1.08 m_0$ et $m_v^{DOS} \simeq 0.59 m_0$ (cas 3D non quantifié). Remarquons que ces valeurs

diffèrent des masses de conduction ($m_{ce} \simeq 0.26m_0$, $m_{ch} \simeq 0.39m_0$). Nous prendrons donc soin de ne pas les confondre dans les prochains paragraphes.

$$N_C = 2\left(\frac{2\pi m_c^{DOS} kT}{h^2}\right) \simeq 2.65 \times 10^{25}\ \mathrm{m}^{-3} \tag{IV.21}$$

$$N_V = 2\left(\frac{2\pi m_v^{DOS} kT}{h^2}\right) \simeq 1.10 \times 10^{25}\ \mathrm{m}^{-3} \tag{IV.22}$$

En injectant (IV.17) et (IV.18) dans (IV.16),la charge d'espace peut être explicitée en fonction des différents niveaux d'énergie du diagramme de bande (IV.23).

$$\rho(\psi) = q\left[N_V\left(\mathcal{F}_{1/2}\left(\eta_{pF} + \frac{q\psi}{kT}\right) - \mathcal{F}_{1/2}\left(\eta_{pF}\right)\right) - N_C\left(\mathcal{F}_{1/2}\left(\eta_{nF} - \frac{q\psi}{kT}\right) - \mathcal{F}_{1/2}\left(\eta_{nF}\right)\right)\right] \tag{IV.23}$$

IV.2.2 Expression analytique de la densité surfacique de charges

Par définition, la densité surfacique de charges Q s'obtient en intégrant la charge d'espace ρ dans la profondeur du silicium (IV.24)

$$Q(x) \overset{\text{def}}{=} \int_{u=x}^{\infty} \rho(u) \cdot du \tag{IV.24}$$

Néanmoins, cette définition est difficilement utilisable puisque nous ne connaissons à priori pas la relation $\psi(x)$. Une solution plus simple consiste à appliquer le théorème de Gauss à l'interface oxyde-semiconducteur pour établir un lien direct entre le champ électrique vertical $\mathcal{E}_x$ et la charge totale contenue dans le canal (IV.25).

$$d(\epsilon_{Si} \cdot \mathcal{E}_x) = \rho(x) \cdot dx = \rho(x) \cdot \left(\frac{dx}{d\psi}\right) \cdot d\psi \tag{IV.25}$$

Cette relation peut s'écrire sous une forme simplifiée (IV.26) qui mène directement à l'intégration pour le calcul de la charge $Q(x)$ (IV.27).

$$d\left(\epsilon_{Si} \cdot \mathcal{E}^2\right) = 2\rho(\psi) \cdot d\psi \tag{IV.26}$$

$$Q^2(x) = -2\epsilon_{Si} \cdot \int_0^{\psi(x)} \rho(\psi) \cdot d\psi \tag{IV.27}$$

En surface ($x = 0$), cette dernière équation renvoie la valeur de la densité totale de charges par unité de surface dans le canal Q_s (IV.28), conformément à [Sze81].

$${Q_s}^2 = -2\epsilon_{Si} \cdot \int_0^{\psi_s} \rho(\psi) \cdot d\psi \tag{IV.28}$$

Finalement, en injectant l'expression de ρ (IV.23), il est possible d'expliciter Q^2 en fonction des variables η_{nF} et η_{pF} (IV.29).

$${Q_s}^2 = 2kT \cdot \epsilon_{Si}\left[N_V\left(\mathcal{I}_{1/2}\left(\eta_{pF} + \frac{q\psi_s}{kT}\right) - \mathcal{I}_{1/2}\left(\eta_{pF}\right) - \frac{q\psi_s}{kT}\mathcal{F}_{1/2}\left(\eta_{pF}\right)\right) + N_C\left(\mathcal{I}_{1/2}\left(\eta_{nF} - \frac{q\psi_s}{kT}\right) - \mathcal{I}_{1/2}\left(\eta_{nF}\right) + \frac{q\psi_s}{kT}\mathcal{F}_{1/2}\left(\eta_{nF}\right)\right)\right] \tag{IV.29}$$

où $\mathcal{I}_{1/2}(\psi)$ est ici définie comme l'intégrale de la fonction de Fermi-Dirac d'ordre $\frac{1}{2}$ entre les bornes 0 et ψ (IV.30)

$$\mathcal{I}_{1/2}(\psi) \stackrel{\text{def}}{=} \int_0^{\psi} \mathcal{F}_{1/2}(u) \cdot du \tag{IV.30}$$

Dans l'approximation de Boltzmann, cette expression se réduit à (IV.31), conformément à [Sze81, Mathieu01].

$$\begin{aligned} {Q_s}^2 = -2\varepsilon_{Si} q \Bigg[& N_V \cdot \exp(\eta_{pF}) \left(\frac{kT}{q} \left(\exp\left(\frac{q\psi_s}{kT} \right) - 1 \right) - \psi_s \right) \\ & - N_C \cdot \exp(\eta_{nF}) \left(-\frac{kT}{q} \left(\exp\left(-\frac{q\psi_s}{kT} \right) - 1 \right) - \psi_s \right) \Bigg] \end{aligned} \tag{IV.31}$$

Dans la suite nous omettrons l'indice $_s$ et désignerons la charge totale par unité de surface du canal par la variable Q.

IV.2.3 Décomposition de la charge d'espace en contributions élémentaires

Afin de permettre la modélisation des différentes contributions capacitives du transistor MOS, il est nécessaire de pouvoir décomposer les expressions globales de Q et ρ données en (IV.16) et (IV.28) en une somme de plusieurs contributions propres à la nature des charges (porteurs libres minoritaires, majoritaires ou charges fixes). Le principe de superposition s'applique à la charge d'espace et rend la tâche triviale : lorsque le semiconducteur est en inversion ($\psi_s < 0$ pour un nMOS), la charge libre est constituée d'électrons (proportionnelle à N_C) tandis que la charge fixe est constituée par des accepteurs non compensés (c'est à dire correspondant à la sous-population de trous $p_0 - p(\psi)$). Le raisonnement reste le même en ce qui concerne le régime d'accumulation. Il en vient les quatres charges d'espace élémentaires (IV.32–IV.35) pour lesquelles la conservation de la charge est évidemment respectée : $\rho = \sum_i \rho_i$.

$$\rho_{N_d^+} = -qN_C \left[\mathcal{F}_{1/2}\left(\eta_{nF} - \frac{\psi_s}{U_{th}} \right) - \mathcal{F}_{1/2}(\eta_{nF}) \right] \cdot H(\psi_s > 0) \tag{IV.32}$$

$$\rho_e = -qN_C \left[\mathcal{F}_{1/2}\left(\eta_{nF} - \frac{\psi_s}{U_{th}} \right) - \mathcal{F}_{1/2}(\eta_{nF}) \right] \cdot H(\psi_s < 0) \tag{IV.33}$$

$$\rho_{N_a^-} = qN_V \left[\mathcal{F}_{1/2}\left(\eta_{pF} + \frac{\psi_s}{U_{th}} \right) - \mathcal{F}_{1/2}(\eta_{pF}) \right] \cdot H(\psi_s < 0) \tag{IV.34}$$

$$\rho_h = qN_V \left[\mathcal{F}_{1/2}\left(\eta_{pF} + \frac{\psi_s}{U_{th}} \right) - \mathcal{F}_{1/2}(\eta_{pF}) \right] \cdot H(\psi_s > 0) \tag{IV.35}$$

Dans les expressions précédentes, $H(x)$ représente la fonction de Heaviside (ou fonction échelon) définie par (IV.36).

$$\forall x \in \mathbb{R},\ H(x) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{si} \quad x < 0 \\ 1 & \text{si} \quad x \geq 0. \end{cases} \tag{IV.36}$$

IV.2.4 Décomposition de la charge totale en contributions élémentaires

Le principe de superposition ne s'applique pas à la densité de charge surfacique (IV.28). Cela rend l'identification des contributions élémentaires plus complexes que dans le cas de la charge d'espace traité dans la partie précédente. Cette décomposition est néanmoins nécessaire pour le calcul des capacités élémentaires dont nous aurons besoin plus tard.

IV.2.4.a Identification des contribution en fonction du signe des charges

Dans un premier temps, la contribution des charges positives et négatives peut être isolée dans l'expression de Q^2. En effet, un simple raisonnement d'électrostatique permet de conclure que quand $\psi_s > 0$:

- les électrons sont repoussés de la surface du semiconducteur, laissant place à des donneurs non compensés et ionisés positivement : c'est la charge de désertion,
- les trous (de charge positive) sont attirés à la surface du semiconducteur : c'est le phénomène d'inversion.

On peut donc simplement affirmer que, selon notre définition de ψ_s, un potentiel de surface positif entraine une charge positive et de manière analogue, qu'un potentiel de surface négatif entraine une charge négative dans le canal. Les expressions de la charge négative Q_{neg} et positive Q_{pos} sont ainsi données par (IV.37) et (IV.38) en suivant ce raisonnement.

$$Q_{neg}(\psi) = \sqrt{Q^2} \cdot H\left(\psi \leq 0\right) \tag{IV.37}$$

$$Q_{pos}(\psi) = \sqrt{Q^2} \cdot H\left(\psi \geq 0\right) \tag{IV.38}$$

Remarquons que la conservation de la charge est respectée : $(Q_{neg} + Q_{pos})^2 = Q^2$ car la définition de fonction de Heaviside imposent $H(x) \cdot H(-x) \overset{\text{def}}{=} 0, \forall x \in \mathbb{R}$.

IV.2.4.b Identification des contribution provenant des charges de désertion

Il est tout à fait légitime d'affirmer que la charge induite par les donneurs non compensés N_D^+ provient d'une lacune électronique, donc d'un terme faisant appel à la densité d'état N_C dans les équations et réciproquement, que la charge induite par les accepteurs N_A^- provient du terme contenant N_V. En reprenant l'expression de Q^2 (IV.28), il est alors facile d'établir l'expression des contribution qui proviennent des charges fixes $Q_{N_D^+}$ et $Q_{N_A^-}$ (IV.39) et (IV.40)

$$Q_{N_D^+}(\psi) = \left[2kT \cdot \epsilon_{Si} N_C \left(\mathcal{I}_{1/2}\left(\eta_{nF} - \frac{q\psi}{kT}\right) - \mathcal{I}_{1/2}\left(\eta_{nF}\right) + \frac{q\psi}{kT}\mathcal{F}_{1/2}\left(\eta_{nF}\right)\right)\right]^{1/2} \cdot H\left(\psi \geq 0\right) \tag{IV.39}$$

$$Q_{N_A^-}(\psi) = -\left[2kT \cdot \epsilon_{Si} N_V \left(\mathcal{I}_{1/2}\left(\eta_{pF} + \frac{q\psi}{kT}\right) - \mathcal{I}_{1/2}\left(\eta_{pF}\right) - \frac{q\psi}{kT}\mathcal{F}_{1/2}\left(\eta_{pF}\right)\right)\right]^{1/2} \cdot H\left(\psi \leq 0\right) \tag{IV.40}$$

Ces expressions peuvent s'approximer par des relations plus simples qui restent valables quelques soit le dopage (IV.41), (IV.42) [Sze81]

$$Q^2_{N_D^+}(\psi) \simeq \left(2q \cdot \varepsilon_{Si} \cdot N_D^+ \cdot \psi\right) \cdot H\left(\psi \geq 0\right) \tag{IV.41}$$

$$Q^2_{N_A^-}(\psi) \simeq \left(2q \cdot \varepsilon_{Si} \cdot N_A^- \cdot \psi\right) \cdot H\left(\psi \leq 0\right) \tag{IV.42}$$

IV.2.4.c Identification des contribution provenant des charges libres

Nous sommes maintenant capables de séparer les contributions des charges libres des fixes et des charges négatives des positives dans l'expression de la charge totale. L'expression des charges induites par les électrons et les trous Q_e et Q_h qui manquent à notre identification peuvent alors se déduire par (IV.43) et (IV.44).

$$Q_e = Q_{neg} - Q_{N_A^-} \tag{IV.43}$$

$$Q_h = Q_{pos} - Q_{N_D^+} \tag{IV.44}$$

Ces expressions ne peuvent pas se réduire à une forme analytique simple et resteront sous cette forme. Elles permettent d'isoler la contribution de chaque type de charge dans l'expression de Q. Remarquons que la conservation de la charge est toujours respectée (IV.45) dans notre raisonnement.

$$\left(\sum_i Q_i\right)^2 = Q^2 \tag{IV.45}$$

IV.2.5 Calcul de la capacité dynamique d'une distribution de charges

L'expression générale de la capacité dynamique est donnée en (IV.1). Repartons donc de l'expression de la charge Q^2 (IV.27) et calculons sa dérivée première par rapport au potentiel de surface ψ_s (IV.46).

$$\frac{d(Q^2)}{d\psi_s} = 2kT\epsilon_{Si}\left[N_V\left(\frac{d}{d\psi_s}\mathcal{I}_{1/2}\left(\eta_{pF}+\frac{q\psi_s}{kT}\right)-\frac{d}{d\psi_s}\mathcal{I}_{1/2}\left(\eta_{pF}\right)\right)\right.$$
$$\left.+N_C\left(\frac{d}{d\psi_s}\mathcal{I}_{1/2}\left(\eta_{nF}-\frac{q\psi_s}{kT}\right)+\frac{d}{d\psi_s}\mathcal{I}_{1/2}\left(\eta_{nF}\right)\right)\right] \tag{IV.46}$$

En utilisant la propriété (IV.47), cette expression se réduit à (IV.48) qui donne l'expression analytique de la réponse capacitive engendrée par toutes les charges du canal.

$$\frac{d\mathcal{I}_{1/2}(u)}{du} = \mathcal{F}_{1/2}(u) \tag{IV.47}$$

d'où

$$\frac{d(Q^2)}{d\psi_s} = 2q\epsilon_{Si}\left[N_V\left(\mathcal{F}_{1/2}\left(\eta_{pF}-\frac{\psi_s}{U_{th}}\right)-\mathcal{F}_{1/2}\left(\eta_{pF}\right)\right)\right.$$
$$\left.-N_C\left(\mathcal{F}_{1/2}\left(\eta_{nF}+\frac{\psi_s}{U_{th}}\right)-\mathcal{F}_{1/2}\left(\eta_{nF}\right)\right)\right] = 2\epsilon_{Si}\rho(\psi_s) \tag{IV.48}$$

En reprenant l'expression de $\rho(\psi)$ (IV.23), cette expression se réduit à une forme extrêmement simple faisant intervenir le rapport de densité de charges surfaciques sur la charge d'espace (IV.49).

$$C = -\epsilon_{Si}\frac{\rho(\psi_s)}{Q(\psi_s)} \tag{IV.49}$$

C'est l'expression générale de la capacité dynamique (en accord avec [Mathieu01]) qui permet d'établir les expressions analytiques des contributions capacitives des différents types de porteurs. Nous garderons à l'esprit que le terme ϵ_{Si}/Q est égal au champ électrique en surface du canal ($\mathcal{E}_s$) et résulte de l'ensemble des charges. Il ne doit donc pas être modifié dans le calcul d'une contribution capacitive (IV.50).

$$C_i = -\epsilon_{Si} \cdot \frac{\rho(\psi_s)}{Q(\psi_s)} = -\epsilon \cdot \frac{\rho_i}{\mathcal{E}_s} \tag{IV.50}$$

IV.2.6 Calcul numérique et distribution de Fermi-Dirac

L'intégrale de Fermi-Dirac $\mathcal{F}_{1/2}$ requiert des temps de calcul importants et des méthodes d'intégration numériques évoluées [Press07] pour estimer la charge Q. De plus, nous ne pouvons pas utiliser l'approximation de Boltzmann, invalide à basse température [Chen97] et pour notre résolution qui comporte plusieurs zones dopées à dégénérescence (extensions S/D, grille). Afin d'alléger la simulation numérique en terme de temps un certains nombre d'approximations par des développements en série [Selvakumar82, Van Halen85, Sobhan85, Cong91] ou des polynômes de Tchebychev [Werner63, Jones66] ont été développées. Nous utiliserons l'approximation par des polynômes de Tchebychev effectuée dans [Werner63] pour sa simplicité et sa précision. Elle permet en effet une estimation de ($\mathcal{F}_{1/2}$) avec une erreur relative inférieure à 0.02%, tout en offrant la simplicité de calcul recherchée (IV.51).

$$\mathcal{F}_{1/2}(\eta) \simeq \begin{cases} \exp(\eta) \cdot \sum_{\nu=0}^{5} a_\nu \cdot \exp(\nu \cdot \eta) & \text{si} \quad -\infty < \eta \leq 1 \\ \eta^{3/2} \cdot \left[\frac{2}{3} + \sum_{\nu=0}^{5} \frac{b_\nu}{\eta^{2\nu+2}} \right] & \text{si} \quad 1 < \eta < \infty \end{cases} \tag{IV.51}$$

Les coefficients des polynômes de Tchebyshev sont donnés dans le tableau IV.1.

Tab. IV.1 – Coefficients de Tchebyshev permettant d'approximer la fonction $F_{1/2}(\eta), \forall \eta \in \mathbb{R}$ grâce à la relation (IV.51). D'après [Werner63].

ν	a_ν	b_ν
0	+0.88607596	+0.843500
1	−0.30871705	+0.710809
2	+0.14638520	−3.712456
3	−0.05843877	+6.705628
4	+0.01431771	−5.594877
5	−0.00150176	+1.777787

IV.2.7 Prise en compte des effets quantiques dans la modélisation

L'amincissement de l'oxyde de grille qui accompagne la réduction d'échelle des transistors MOS amène à considérer la couche d'inversion non plus comme "collée" à l'interface Si/SiO_2, mais comme une distribution de charges qui s'étale sur une profondeur non négligeable devant t_{ox}. Dans une approche classique, l'étalement des charges libres dans la profondeur du canal provient de l'effet d'écrantage qui donne une allure exponentiellement décroissante à la distribution de porteurs. La longueur de Debye λ_D est une grandeur caractéristique de la distribution dans le gaz de porteurs libres (IV.52) et la distribution de charges $\rho(x)$ présente un pic à l'interface Si/SiO_2, où la courbure de bande et le champ électrique transverse atteignent leur valeur maximale.

$$\lambda_D(\psi) = \sqrt{\frac{k_B T}{q} \cdot \frac{\varepsilon_{Si}}{\rho(\psi)}} \tag{IV.52}$$

La mécanique quantique (i.e. le principe d'exclusion de Pauli) interdit aux fermions (i.e. aux électrons) d'occuper la même position dans le même état quantique [Cohen-Tannoudji06]. Les porteurs sont traités sous un aspect ondulatoire : leur distribution est caractérisée par une fonction d'onde $\xi(x)$ et une probabilité de présence $|\xi(x)|^2$. Les énergies accessibles sont quantifiées et, contrairement au cas classique, le pic de la distribution $\rho(x)$ n'est pas situé à l'interface mais à une certaine profondeur du canal, proche du barycentre des charges X_{dc} défini par (IV.53).

$$X_{dc} \overset{\text{def}}{=} \frac{\int_0^\infty x \cdot \rho(x) \cdot dx}{Q} \tag{IV.53}$$

L'amplitude de la réponse capacitive peut alors se modéliser en ajoutant une capacité C_{QM} en série à C_{ox} dans les équations concernées (IV.54) [Takagi95, Takagi08], ce qui revient à considérer que l'épaisseur d'oxyde est artificiellement augmentée par les effets quantiques. Ceux-ci se manifestent pas une diminution de la charge dans le canal et une augmentation de la tension de seuil [Hareland98a].

$$C_{ox} \rightarrow \left(\frac{1}{C_{ox}} + \frac{1}{C_{QM}(V_{gs})} \right)^{-1} \tag{IV.54}$$

Remarque : lorsque V_{gs} est différent de la tension de bande plate, la densité de porteurs libres devient si grande que l'effet d'écrantage classique devient négligeable devant l'influence des effets quantiques.

IV.2.7.a Résolution des équations dans le cas quantique

La prise en compte des effets quantiques dans le calcul de la charge et des potentiels de la structure MOS requiert la résolution couplée des équations de Poisson (IV.55) et de Schrödinger (IV.56).

$$\frac{d^2\psi}{dx^2} + \frac{q}{\varepsilon_{Si}} \cdot \left[p(x) - n(x) + N_D^+(x) - N_A^-(x) \right] = 0 \tag{IV.55}$$

$$\frac{\hbar^2}{2m_x^\star} \cdot \frac{d^2\xi_j}{dx^2} + [E_j + q \cdot \psi(x)] \cdot \xi_j = 0 \tag{IV.56}$$

Dans ces équations, l'indice j caractérise la j-ème sous bande (associée à l'énergie propre E_j) et $m_x^\star$ est la masse effective des porteurs dans la direction du confinement (x dans notre cas). La résolution complète Poisson-Schrödinger multi-sousbandes est la seule approche qui fournit une résolution exacte du problème et permet d'obtenir la distribution des charges dans le canal. Malheureusement, cette technique est couteuse en termes de temps de calcul et complexe à implémenter, même lorsque la résolution se limite aux trois premières sous-bandes [Hareland96]. Plusieurs modèles analytiques ont été établis dans le but de corriger la tension de seuil ou la capacité de l'influence des effets quantiques [Hansch89, van Dort94]. Ces derniers reposent sur des hypothèses assez lourdes et ne conviennent pas à notre application qui nécessite la connaissance de Q_s et de ψ_s pour calculer les réponses capacitives.

IV.2.7.b Les méthodes approchées pour la prise en compte des effets quantiques dans l'équation de Poisson

Seule la résolution numérique permet, par le biais d'une intégration auto-consistante, de résoudre le système d'équation Poisson-Schrödinger (PS) décrit précédemment [Ando82, Spinelli98]. Cette itération est lourde en temps de calcul et nécessite une résolution de type éléments finis qui sort du cadre de notre étude. De nombreuses approximations ont été apportées dans la littérature pour tenter d'intégrer les effets quantiques à la résolution de l'équation de Poisson de manière plus simple. Certaine méthodes proposent donc la prise en compte des effets quantiques par des ajustements *ad hoc*, rendant compte des effets quantiques [Hansch89, van Dort94, Hareland97, Hareland98b], souvent à travers la modification de la tension de seuil ou l'élargissement du gap du silicium. Malheureusement, ces techniques ne fournissent pas d'information fiable sur la distribution spatiale des porteurs dans la profondeur du silicium et doivent être calibrées pour une technologie donnée grâce à une résolution PS complète. La plupart du temps, ces corrections ne permettent pas de considérer la distribution spatiale des charges et ont un intérêt limité car elles sont incapables de prédire l'éloignement des charges par rapport à l'interface Si/SiO_2 [Ancona97, Leong98, Li03].
L'approximation MLDA (pour *Modified Local Density Approximation*) est basée sur la résolution exacte de l'équation de Schrödinger dans un puits de potentiel carré de hauteur infini. Elle permet d'aboutir à une expression de la densité électronique prenant en compte les effets quantiques [Paasch82] et constitue une amélioration de [Hansch89] qui permet une résolution approchée plus précise du système Poisson-Schrödinger [Leong98, Li03]. Les temps de calculs mis en jeu restent pourtant trop longs pour notre application.
Enfin, l'approximation du puits de potentiel triangulaire que nous utiliserons dans notre étude permet d'avoir une solution analytique à l'équation de Schrödinger, et donc une résolution auto-cohérente simplifiée du système d'équations. Cette approximation donne des résultats en ligne avec la résolution PS complète, en ce qui concerne le calcul de la charge Q et de la courbure de potentiel ψ_s [Ma00, Hareland96]. Ce sont ces valeurs qui nous permettent de modéliser la réponse capacitive et il nous importe peu que la forme de la distribution de charges soit moins bien estimée [Ma00]. L'approximation du puits de potentiel triangulaire est justifiée tant que la courbure du potentiel induit par les charges de désertions reste grande devant l'influence des charges libres [Mathieu01]. Elle semble donc bien adaptée à nos dispositifs ayant un dopage du canal élevé (environ $10^{18}/cm^3$ pour les dispositifs les plus courts). Nous verrons par la suite que certaines corrections permettent de rendre l'approximation encore plus proche des résultats de calcul numérique complet [Ferrier06, Hareland98a].

IV.2.7.c L'approximation du puits de potentiel triangulaire

Pour simplifier la résolution du système d'équations Poisson-Schrödinger, il est possible d'approximer la forme du potentiel $\psi(x)$ à un puits triangulaire de profondeur infinie. Cette forme de potentiel permet d'obtenir des solutions analytiques à l'équation de Schrödinger qui permet le calcul des fonctions d'onde et de la distribution des porteurs dans le canal. L'approximation de $\psi(x)$ par une fonction triangulaire est données en (IV.57)

$$E(x) = -q\psi(x) \simeq \begin{cases} +\infty & \text{si} \quad x \leq 0 \\ \\ x \cdot q\mathcal{E}_s & \text{si} \quad x > 0 \end{cases} \tag{IV.57}$$

Les fonctions d'onde $\xi_{i,j}(x)$ sont données en fonction des fonctions d'Airy $\mathrm{Ai}(x)$ [Weisstein] et les énergies propres de la sous-bande j de l'orbitale i sont données en (IV.58) (les rappels sur la structure de bande du silicium sont donnés dans l'Annexe B).

$$\xi_{i,j}(x) = \mathrm{Ai}\left(u\right), \quad u = \left(\frac{2m^{\star}_{x,i} \cdot q\mathcal{E}_s}{\hbar^2}\right)^{1/3} \cdot \left(x - \frac{E_{i,j}}{q\mathcal{E}_s}\right) \tag{IV.58}$$

$$E_{i,j} = \left(\frac{\hbar}{2m^{\star}_{x,i}}\right)^{\frac{1}{3}} \left[\frac{3}{2}\pi q\mathcal{E}_s \left(j + \frac{3}{4}\right)\right]^{\frac{2}{3}} \tag{IV.59}$$

où $\mathcal{E}_s$ représente le champ en surface dans le canal[2].

$$\mathcal{E}_s \overset{\text{def}}{=} \frac{Q(\psi_s)}{\epsilon_{Si}} \tag{IV.60}$$

IV.2.7.d Les améliorations de l'approximation du puits triangulaire

L'approximation du puits de potentiels triangulaire est fréquemment utilisée pour rendre compte des effets quantiques dans la structure MOS. Elle donne des résultats en bon accord avec le calcul numérique auto-consistant en ce qui concerne le calcul de la charge Q_s et du potentiel de surface ψ_s dans les domaines d'accumulation ou d'inversion forte [Ma00]. Elle semble donc bien adaptée à notre utilisation, où la capacité est calculée en connaissant Q_s et ψ_s. En revanche, cette approximation renvoie une estimation imprécise du barycentre des charges libres du fait que les fonctions d'Airy ne constituent qu'une approximation grossière des fonctions d'ondes exactes [Stern72, Ando82]. De plus, l'approximation du puits triangulaire infini devient erronée à proximité de la tension de bandes plates, région dans laquelle le phénomène de *darkspace* doit être pris en compte pour rendre compte des effets quantiques [Ma00, Pacelli99, Spinelli02]. Les phénomènes suivants sont également négligés :

- la pénétration des fonctions d'onde dans l'oxyde qui résulte de la hauteur finie de la barrière de potentiel Si/SiO_2 ($\Phi_{ox} \simeq 3$eV),
- la non linéarité du potentiel créé par les charges de désertion $\psi_{dep}(x)$, cf. [Mathieu01],
- l'utilisation du champ électrique de surface $\mathcal{E}_s$ qui n'est pas représentatif du champ moyen vu par les porteurs d'un sous bande donnée.

Une amélioration a donc été proposée dans [Ferrier06, Ferrier07] afin de corriger ces lacunes grâce à des termes correctifs qui s'appliquent au calcul des niveaux d'énergie :

2. Certaines études définissent $\mathcal{E}_s = \mathcal{E}_{eff}$, cf. [Ma00]

– Le terme quadratique dans le calcul du potentiel créé par les charges de désertion [Mathieu01] est pris en compte par le terme correctif $\Delta E_{i,j}^{NL}$, cf. (IV.61) ;

$$\Delta E_{i,j}^{NL} = \frac{4}{15} \cdot \frac{q \cdot N_D \cdot E_i}{\varepsilon_{Si} \cdot \mathcal{E}_{eff,i,j}} \cdot \frac{\left[\frac{3\pi}{2} \cdot \left(j + \frac{3}{4}\right)\right]^{2/3}}{\left(\frac{2m_i^\star \cdot q \cdot \mathcal{E}_{eff,i,j}}{\hbar^2}\right)^{2/3}} \tag{IV.61}$$

– La pénétration des fonctions d'onde dans l'oxyde (ce phénomène résulte du fait que la barrière de potentiel Si/SiO_2 n'est pas infinie en réalité) est prise en compte par le terme $\Delta E_{i,j}^{WFP}$:

$$\Delta E_{i,j}^{WFP} = q \cdot \frac{\hbar \cdot m_{ox} \cdot \mathcal{E}_{eff,i,j}}{m_i^\star \cdot \sqrt{2m_{ox} \cdot \left(q\Phi_{ox} - E_{i,j}^{\infty}\right)}} \tag{IV.62}$$

– L'introduction du concept de champ effectif propre à chaque sous bande $\mathcal{E}_{i,j}$, représentant le champ moyen vu par les porteurs. La définition est enrichie par l'introduction d'un terme de pondération similaire à [Ando82, Sabnis79] et propre à chaque couple (i, j). Ce calcul renvoie donc une valeur plus proche d'un barycentre que d'une moyenne, comme défini historiquement.

$$\mathcal{E}_{i,j}^\star \overset{\text{def}}{=} \frac{\int_0^\infty \rho_{i,j}(x)\mathcal{E}(x)dx}{\int_0^\infty \rho_{i,j}(x)dx}, \quad \mathcal{E}(x) \overset{\text{def}}{=} \left|\frac{d\psi}{dx}\right| \tag{IV.63}$$

En appliquant le théorème de Gauss et en négligeant les variations de la charge de désertion sur l'épaisseur de la couche d'inversion, (IV.63) peut se mettre sous une forme plus usuelle (IV.64). Le paramètre η a été remplacé par un nouveau paramètre $\eta_{i,j}^\star$ propre à chaque couple orbitale/sous-bande.

$$\mathcal{E}_{i,j}^\star = \frac{Q_{dep} + \eta_{i,j}^\star \cdot Q_{fc}^{2D}}{\varepsilon_{Si}}, \quad \eta_{i,j}^\star = \frac{1}{Q_{fc}^{2D}} \cdot \frac{\int_0^\infty \rho(x) n_{inv}(x)dx}{\int_0^\infty n_{inv}(x)} \tag{IV.64}$$

Dans cette dernière expression, Q_{fc}^{2D} représente la densité surfacique de charges libres (électrons ou trous) dans le cas d'une quantification selon deux dimensions (IV.65), qui doit être couplée au calcul de $E_{i,j}$ incluant les corrections (IV.66) pour obtenir l'autoconsistance dans la résolution.

$$\begin{aligned} Q_{fc}^{2D}(\psi_s) = Q_e^{2D} + Q_h^{2D} = & q\sum_{i,j} \frac{kT \cdot m_{i,j}^{DOSv}}{\pi\hbar^2} \cdot \ln\left(1 + \exp\left(\eta_{Fv} - \frac{E_{i,j}(\psi_s) - q\psi_s}{k_B T}\right)\right) \\ & - q\sum_{i,j} \frac{kT \cdot m_{i,j}^{DOSc}}{\pi\hbar^2} \cdot \ln\left(1 + \exp\left(\eta_{Fc} - \frac{E_{i,j}(\psi_s) + q\psi_s}{k_B T}\right)\right) \end{aligned} \tag{IV.65}$$

$$E_{i,j} = \left(\frac{\hbar}{2m_i^\star}\right)^{\frac{1}{3}} \left[\frac{3}{2}\pi q \mathcal{E}_{i,j}^\star \left(j + \frac{3}{4}\right)\right]^{\frac{2}{3}} - \Delta E_{i,j}^{NL} - \Delta E_{i,j}^{WFP} \tag{IV.66}$$

Signalons finalement que des corrections permettent par exemple de corriger la forme de la fonction d'onde afin de considérer la queue de distribution des trous dans la profondeur

du silicium [Hareland98a]. Néanmoins, l'ajustement des paramètres du modèle nécessite une calibration à partir d'une résolution complète PS, calibrée sur la technologie étudiée. Ce type d'amélioration ne sera donc pas utilisé dans notre étude.

IV.2.7.e Prise en compte des effets quantiques pour régime de désertion

Lorsque le champ électrique de confinement est faible l'hypothèse du puits de potentiel triangulaire n'est plus valable car la profondeur du puits devient faible, voire nulle. Par contre, la discontinuité du potentiel à l'interface Si/SiO_2 cause un éloignement des porteurs de quelques nanomètres et fait apparaitre une zone proche de l'interface dans laquelle la densité de porteurs majoritaires est plus faible que dans la profondeur du substrat [Pacelli99, Ma00]. Cet effet quantique provoque une atténuation de la réponse capacitive à proximité de la tension de bande plate (V_{fb}), particulièrement influente lorsque le dopage du canal dépasse $10^{17}/\text{cm}^3$ et qui peut être attribuée à une augmentation apparente de l'épaisseur d'oxyde. Plus précisément, l'effet de *darkspace* est causé par la quantification de l'énergie des fonctions d'onde qui contribuent à la charge dans le canal. La valeur du *darkspace* λ_{dsk} peut alors être donnée par la distance qui sépare l'interface du premier pic de la fonction d'onde la plus énergétique, cf. (IV.67).

$$\lambda_{dsk} = \frac{\pi\hbar}{2\sqrt{2m_x^\star \cdot kT}} \tag{IV.67}$$

Pratiquement, le *darkspace* est influent sur une caractéristique C-V lorsque la valeur λ_{dsk} est supérieure ou égale à la longueur de Debye λ_D, qui caractérise la longueur d'écrantage dans le silicium [Sze81]. En reprenant l'expression de λ_D évoquée au chapitre II, il est possible de définir une condition sur la valeur du dopage nécessaire à la prédominance de l'effet de *darkspace* sur l'écrantage classique (IV.68). Pour les électrons à 300K et une orientation (100)/<110>, on trouve $N_D \geq 5\times10^{18}/\text{cm}^3$. Bien sûr l'effet du *darkspace* est visible pour des dopages plus faibles et n'est strictement nul que lorsque $\lambda_D \gg \lambda_{dsk}$.

$$\lambda_{dsk} \geq \lambda_D \quad \Leftrightarrow \quad N_D \geq \frac{8m_x^\star \cdot (kT)^2 \cdot \varepsilon_{Si}}{(\pi \cdot \hbar)^2} \tag{IV.68}$$

Remarquons finalement que cet effet est très marqué dans les extensions et dans la grille qui sont des zones caractérisées par une concentration de dopage élevée (donc un faible λ_D).

Le décalage du barycentre de la distribution de porteurs par une distance λ_{dsk} est généralement comprise entre 8Å et 16Å pour les nMOS et entre 12Å et 24Å pour les pMOS [Skotnicki03]. Ce phénomène peut être assimilé à une augmentation apparente de l'épaisseur physique d'oxyde t_{ox} selon la relation (IV.69).

$$t_{ox,eff} = t_{ox} + \frac{\epsilon_{ox}}{\epsilon_{Si}} \cdot \lambda_{dsk} \tag{IV.69}$$

IV.3 Utilisation des mesures capacitives pour l'extraction de la charge et de la mobilité

Les mesures capacitives semblent depuis longtemps être le moyen le plus fiable pour extraire les charges (de désertion et d'inversion) dans le transistor MOS et peut permettre de calculer la mobilité effective des porteurs. Dans cette partie, nous allons donc décrire les méthodes de bases qui permettent d'extraire tous les paramètres pour analyser le comportement $\mu_{eff}(\mathcal{E}_{eff})$ du MOSFET.

IV.3.1 L'extraction expérimentale de la charge

IV.3.1.a Extraction de la charge d'inversion

En reprenant la définition de la capacité dynamique introduite au §IV.1, il est facile d'écrire la charge d'inversion sous sa forme intégrale, en fonction de C_{inv} et ψ_s (IV.70).

$$Q_{inv}(\psi) \stackrel{\text{def}}{=} \int_{-\infty}^{\psi} C_{inv}(\psi_s)\, d\psi_s \tag{IV.70}$$

Le potentiel de surface n'est pas une grandeur directement accessible par la mesure et il est préférable de réécrire cette expression en fonction de V_{gs} donc l'expression en fonction des potentiels de surfaces et des charges est rappelée en (IV.71).

$$V_{gs} = \psi_s^G - \psi_s - \frac{(Q_{inv} + Q_{dep})}{C_{ox}} \tag{IV.71}$$

La relation d'égalité de la charge dans la structure MOS en inversion ($Q_{gate}+Q_{inv}+Q_{dep} = 0$) permet d'aboutir à la relation (IV.72). Pour information, les variations des contribution Q_{dep} et Q_{inv} sont illustrées sur la figure IV.11 en accord avec [Sze81]. L'erreur résultant de la prise en compte ou non de $\partial Q_{dep}/\partial Q_{inv}$ dans le calcul est illustrée en insertion.

$$C_{gc} \stackrel{\text{def}}{=} \frac{\partial Q_{inv}}{\partial V_{gs}} \tag{IV.72}$$

Cette relation permet d'exprimer Q_{inv} sous une forme intégrale faisant intervenir C_{gc} et V_{gs} (IV.73).

$$Q_{inv}(V_{gs}) = \int_{-\infty}^{V_{gs}} C_{gc}(u)\, du \tag{IV.73}$$

Remarquons finalement que cette relation ne fait pas d'hypothèse sur la désertion de grille ou sur la quantification des porteurs dans le canal.

IV.3.1.b Extraction de la charge de désertion

De la même manière que pour l'extraction de la charge d'inversion, la charge de désertion Q_{dep} peut être extraite depuis la mesure de la capacité grille-substrat. En suivant le même raisonnement que dans le paragraphe précédent, en en admettant que $\forall V_{gs} \geq V_{fb}, Q_{acc} = 0$ (nMOS), on aboutit facilement à l'expression (IV.74), en accord avec [Romanjek04b].

$$Q_{dep}(V_{gs}) = \int_{u=V_{fb}}^{V_{gs}} C_{gb}(u)\, dV_u \tag{IV.74}$$

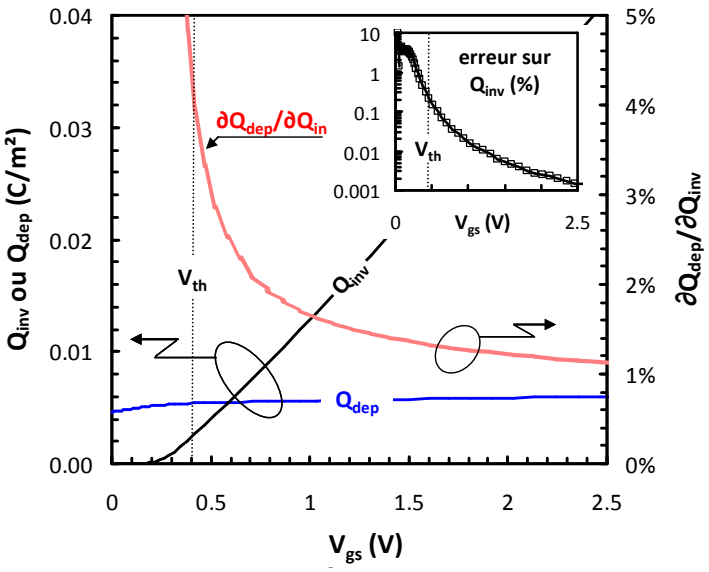


Fig. IV.11 – Evolution de Q_{inv}, Q_{dep} et du rapport $\partial Q_{dep}/\partial Q_{inv}$ en fonction de V_{gs} pour un transistor simulé de 17.5Å d'épaisseur d'oxyde. En inversion forte, la variation de Q_{dep} faible devant celle de Q_{inv} et peut être négligée dans les calculs. L'erreur résultant de la prise en compte ou non de $\partial Q_{dep}/\partial Q_{inv}$ dans le calcul est illustré en insertion.

IV.3.2 Cas des transistors courts : influence des capacités parasites

Les expressions permettant le calcul de Q_{inv} et Q_{dep} que nous venons d'introduire ne sont valables que dans le cas où les capacités parasites peuvent être négligées devant la réponse intrinsèque du transistor. Dans le cas contraire une correction des capacités parasites reposant sur le modèle introduit au §IV.1.3 est nécessaire [Romanjek04b]. Si la correction de la contribution parasite est bien établie en ce qui concerne la mesure $C_{gc}(V_{gs})$ [Fleury08b], la correction de la capacité grille-substrat n'est pas traitée dans la littérature. La mesure $C_{gb}(V_{gs})$ comporte également une difficulté d'ordre pratique car elle est caractérisée par un rapport signal-sur-bruit beaucoup plus faible que pour la mesure C_{gc}, la rendant peu adaptée à la mesure de transistors de petite surface.

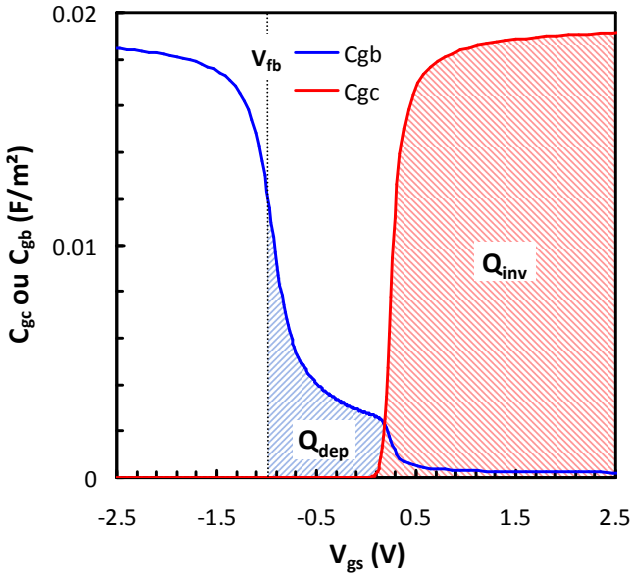


Fig. IV.12 – Les charges d'inversion (Q_{inv}) et de désertion (Q_{dep}) sont définient par l'intégration respectives de la mesure capacitive $C_{gc}(V_{gs})$ et $C_{gb}(V_{gs})$.

IV.3.3 Extraction de la mobilité par la mesure *split*-CV

La méthode *split*-CV introduite dans [Koomen73, Sodini82] et son adaptation pour les transistors courts [Romanjek04b] sont largement employées pour déterminer la mobilité effective des porteurs. Elles consistent d'une part en la mesure de Q_{inv} et de Q_{dep} pour le calcul du champ effectif en utilisant les relations (IV.74) et (IV.73) et d'autre part, permet d'extraire la mobilité effective depuis une mesure $I_d(V_{gs})$ effectuée sur le même transistor, via la relation (IV.75).

$$\mu_{eff} \stackrel{\text{def}}{=} \frac{L_{eff}}{W_{eff}} \cdot \frac{I_d}{Q_{inv} \cdot V_{ds}} \tag{IV.75}$$

Cette technique renvoie une mobilité nulle à faible charge ce qui est aberrant d'un point de vue physique et démontre l'invalidité de la technique en dessous de V_{th}. Pour des tensions respectant $V_{gs} \gg V_{th}$, les variations de mobilité $\mu_{eff}(Q_{inv})$ ou $\mu_{eff}(\mathcal{E}_{eff})$ peuvent être extraites. Les résultats obtenus sont en ligne avec ceux issus de l'extraction de type fonction Y (ou ξ) décrite au §III.4.3 (cf. figure IV.13) [Romanjek04a]. La divergence à faible charge provient de l'inégalité entre les tensions de seuil issues de la mesure capacitive et du courant.

IV.3.3.a Cas particulier des canaux courts

L'extraction de la mobilité par la technique *split*-CV sur les transistors courts (technologies sub-100nm) est complexe [Romanjek04b]. Pour que cette méthode puisse fournir des résultats précis, il est en effet nécessaire de :

- corriger l'effet de la résistance série sur le courant I_d,
- extraire la longueur effective du dispositif (ce point sera détaillé dans la partie §IV.4),
- corriger la mesure de l'influence des capacités parasites [Romanjek04b, Andrieu05],
- avoir un rapport signal-sur-bruit élevée (c'est en particulier un point critique pour la mesure $C_{gb}(V_{gs})$).

La prise en compte de ces points nécessite des techniques de caractérisation avancées qui ont rarement été developpées dans la littérature.

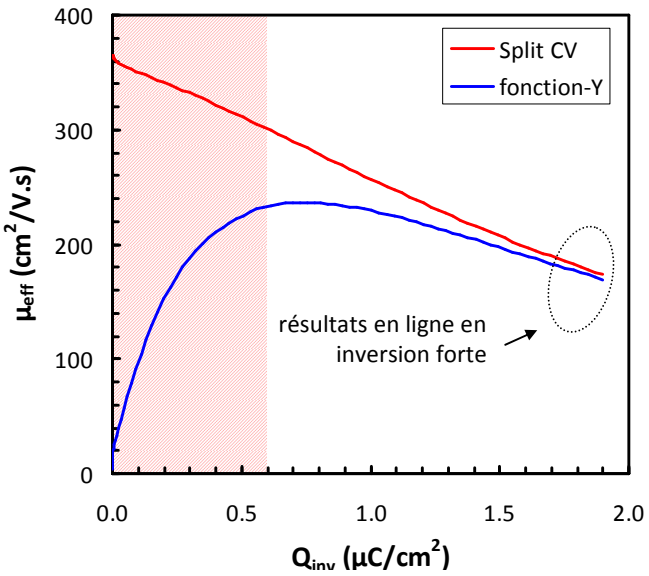


Fig. IV.13 – La mobilité effective extraite par la fonction Y et la technique *split*-CV fournissent des résultats en ligne en forte inversion.

IV.3.3.b Cas particulier des films minces totalement désertés

Il est impossible de mesurer la réponse $C_{gb}(V_{gs})$ sur des transistors à films minces totalement désertés (architecture FDSOI, cf. chapitre I). Dans ce cas particulier l'extraction de la charge de désertion pourra se faire en connaissance de l'épaisseur du film de silicium t_{Si} et de son dopage N_D, par l'intermédiaire de la relation (IV.76)

$$Q_{dep} = qN_D \cdot t_{Si} \tag{IV.76}$$

Dans le cas d'un transistor à double grille, cette expression se réduit à (IV.77)

$$Q_{dep} = qN_D \cdot \frac{t_{Si}}{2} \tag{IV.77}$$

IV.4 L'extraction de la longueur effective (L_{eff})

La longueur effective du canal (L_{eff}) est un paramètre primordial car il gouverne le courant débité par le transistor ($I_d \propto 1/L_{eff}$) et permet de comparer les propriétés électriques des canaux longs à celles des canaux courts. La réduction d'échelle effectuée ces dernières années engendre des longueurs effectives qui atteignent parfois la moitié de la longueur de grille, définie par le masque de lithographie. Il en résulte une erreur importante lorsque la valeur de L_{eff} n'est pas mesurée correctement ou assimilée par erreur à L_{mask}. La longueur effective joue en effet un rôle important sur les performances des dispositifs courts dans lesquels elle gouverne l'amplitude des effets de canaux courts (cf. §I.1.1). De plus, elle dépend généralement de manière complexe d'un ensemble de procédés technologiques (lithographie, implantation, etc...). En conséquence, cette grandeur doit être mesurée au cours de développement du procédé de fabrication afin d'assurer un bon contrôle des performances finales de la technologie. Dans cette partie, nous allons voir comment extraire pratiquement L_{eff} par le biais de mesures capacitives et d'une méthodologie adaptée aux transistors courts.

IV.4.1 Etat de l'art des méthodes d'extraction

Depuis plusieurs années, la mesure de L_{eff} a suscité l'intérêt de la communauté scientifique et de nombreuses méthodes ont été développées pour extraire ce paramètre expérimentalement. Certaines de ces techniques ont une validité limitée car elles reposent sur des hypothèses qui ne sont plus vérifiées lorsque la longueur de grille est inférieure à 100nm. En particulier, les techniques Shift-and-Ratio (S&R) [Taur92, Taur00, Cretu01] et de la résistance totale [Yamaguchi98, Takeuchi96, Brut97] proposent d'extraire L_{eff} à partir d'une mesure de courant $I_d(V_{gs})$ en supposant que c'est le seul paramètre qui défini la quantité de courant qu'un transistor court délivre en référence à un long. Cette approche perd sa validité sur les transistors sub-100nm pour lesquels la mobilité (qui influe sur le courant de drain) dépend de L_{eff} (cf. §II.2.5 et [Cros06b, Andrieu05, Bidal09a]). Au contraire, les techniques d'extraction basée sur des mesures capacitives $C_{gb}(V_{gs})$ [Heish01, Scholten01] ou $C_{gc}(V_{gs})$ [Yao86, Huang96, Romanjek04b, Severi06] offrent un moyen efficace et fiable d'extraire L_{eff} sans faire d'hypothèse sur le transport dans le canal ($V_{ds} = 0$). Néanmoins, elles nécessitent des mesures longues ainsi qu'une bonne connaissance des capacités parasites inhérentes à la structure du transistor. De plus, ces techniques ne fournissent pas de solution pour faire des mesures rapides et précises, adaptées à un environnement industriel.

IV.4.2 Définition précise de la longueur effective de canal

Le terme "longueur de canal" est généralement utilisé comme un terme générique dans la littérature et désigne des grandeurs physiques différentes. Une définition des différentes longueurs qui caractérisent l'architecture du transistor MOS est donnée sur la figure IV.14 :

- la longueur de masque L_{mask} correspond à la longueur de grille telle que dessinée sur le masque de lithographie,
- La longueur de grille réelle L_{gate} est la longueur physique de la grille suite aux procédés de lithographie et de gravure. Elle peut différer de L_{mask} à cause de la variabilité de ces étapes dans le procédé de fabrication et de l'OPC (*Optical Proximity Correction*),
- La longueur métallurgique L_{met} est définie comme la distance qui sépare les jonctions source-canal et drain-canal quand toutes les polarisations du transistor sont fixées à 0,

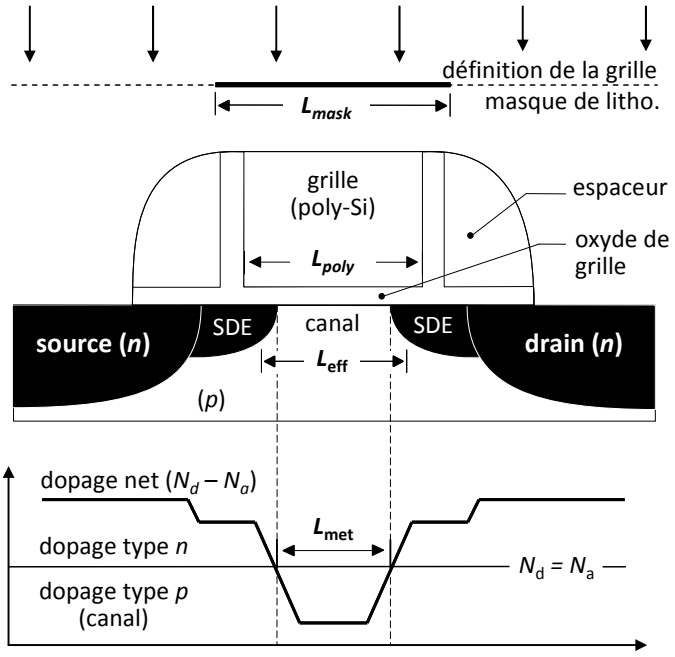


Fig. IV.14 – Illustration des différentes longueurs qui peuvent être définies dans l'architecture du transistor MOS.

– dans ce manuscrit, la longueur effective de canal L_{eff} est définie comme la longueur moyenne de la couche d'inversion.

La grandeur L_{met} découle directement des profils de jonctions du côté de la source et du drain. Dans le cas de jonctions non abruptes L_{met} est généralement plus petit que la longueur de grille : $L_{met} < L_{gate}$ et, dans le cas d'un canal très dopé, reste relativement proche de L_{eff}. Nous devons pourtant considérer L_{met} comme une grandeur dangereuse et une approximation de L_{eff} qui n'a pas de sens physique lorsqu'elle est utilisée pour définir la longueur de la couche d'inversion. En effet, sa valeur n'est pas corrélée avec le niveau de courant et est réduite à 0 dans le cas des canaux non dopés ou des architectures à canal enterré qui, par construction, ne possèdent pas de jonctions [Taur00]. Au contraire, L_{eff} gouverne physiquement le courant délivré par le transistor, comme décrit dans les modèles compactes BSIM3v3 [Cheng97b] et MOS model 9 [NXP07]. Dans les paragraphes suivants, les différences entre L_{mask} et L_{eff} seront regroupées dans une grandeur unique $\Delta L = L_{mask} - L_{eff}$, appelée réduction de la longueur de canal.

IV.4.3 Proposition d'une nouvelle méthodologie d'extraction

L'extraction de L_{eff} depuis les mesures capacitives repose sur l'hypothèse que l'amplitude de la réponse capacitive mesurée est proportionnelle à la surface effective du transistor ($W_{eff} \times L_{eff}$). En général, nous choisirons des transistors larges afin de pouvoir négliger les éventuelles interactions croisées du type $W_{eff}(L_{eff})$ et de pouvoir admettre $W_{eff} = W_{mask} = W$. Pratiquement, la mesure de capacité grille-canal $C_{gc}(V_{gs})$ est effectuée en connectant le terminal *High* du capacimètre à la grille du transistor et le terminal *Low* à la source et au drain court-circuités tandis que le plot de substrat reste à la masse [Yao86]. La mesure C_{gc} donne ainsi un moyen pratique d'obtenir la réponse capacitive de la couche d'inversion et d'extraire L_{eff} [Yao86, Huang96]. La figure IV.15 présente un réseau de courbes $C_{gc}(V_{gs})$ mesurées sur des transistors de longueurs variées issus de la technologie CMOS 65nm.

La méthodologie d'extraction (i.e. la validité des valeurs de L_{eff} extraites) repose sur le

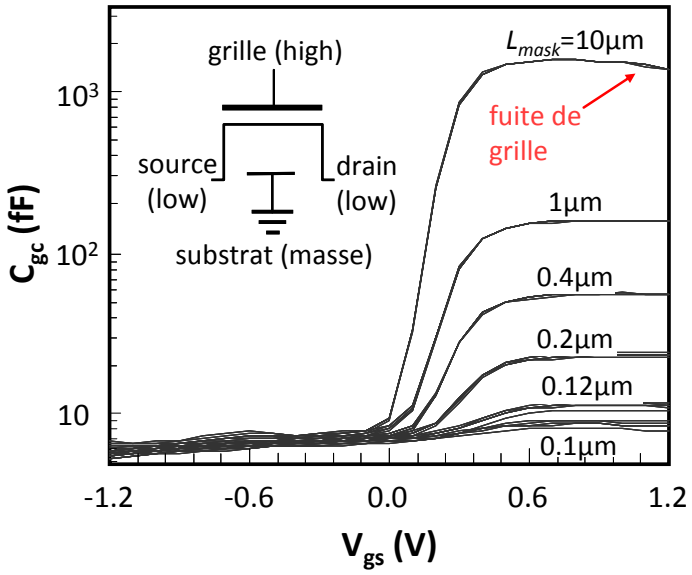


Fig. IV.15 – Courbes $C_{gc}(V_{gs})$ effectuées pour différentes longueurs de transistors ($W = 1\mu$m).

modèle introduit au §III.2 et sur notre faculté à l'utiliser pour extraire le niveau de capacité intrinsèque, noté C_{int}. Comme dans [Yao86], C_{int} est mesurée en un point unique dans le régime d'inversion forte : $C_{int} = C_{gc}(V_{gt}) - C_{par}$, cf. figure IV.16. L'utilisation de V_{th} (dans V_{gt}) permet de compenser les éventuelles variations de tension de seuil entre les longueurs. La tension de seuil considérée ici est issue de la mesure capacitive, en utilisant la méthode décrite au §III.3.1, en accord avec [Sze81].

L'expérience montre que le rapport des capacités intrinsèques reste constant tant que la fuite de grille est négligeable [Huang96, Schmitz03a]. Deux techniques permettent alors d'extraire L_{eff} (et ΔL) : La technique $\Delta L(L_{mask})$ et celle du ΔL constant.

IV.4.3.a La technique $\Delta L(L_{mask})$

Si un ensemble de transistors de longueurs variées comprend un dispositif suffisamment long pour que l'approximation $L_{eff} \simeq L_{mask}$ soit justifiée (typiquement $L_{mask} \geq 1\mu$m), ce dernier peut servir de référence et fournit la valeur de la capacité intrinsèque surfacique C_{int}^{ref}, valable pour toutes les autres longueurs. Dans ce cas, la longueur effective peut être extraite depuis la relation (IV.78), dans laquelle ^{ref} et * réfèrent respectivement aux paramètres du transistors choisi comme référence et dont nous cherchons à extraire L_{eff}.

$$L_{eff}^{\star}(V_{gt}) = L_{eff}^{ref} \times \frac{C_{int}^{\star}(V_{gt})}{C_{int}^{ref}(V_{gt})} \tag{IV.78}$$

Cette méthode offre l'avantage de pouvoir extraire la réduction de longueur de grille pour chaque dispositif : $\Delta L(L_{mask})$. En particulier, elle permet d'extraire les éventuelles variations de réduction de longueur de canal avec L_{mask}, qui peuvent résulter des procédés de gravure et de lithographie [Fleury08a].

Estimation de l'erreur sur L_{eff} : Pour estimer l'erreur sur L_{eff}, deux élements sont à prendre en compte :

– L'hypothèse de départ $L_{mask}^{ref} \simeq L_{eff}^{ref}$ répercute une erreur relative sur les valeurs de L_{eff} extraites,

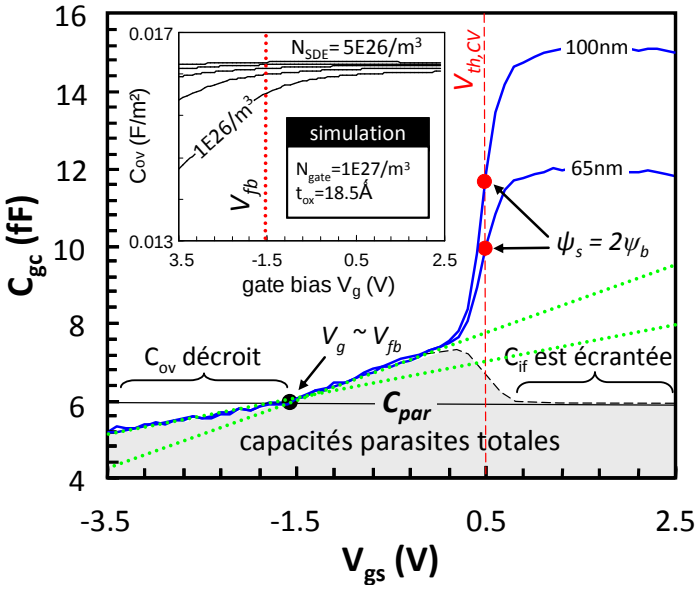


Fig. IV.16 – Courbes $C_{gc}(V_{gs})$ issues de mesures sur des transistors nMOS de la technologie 65nm ($t_{ox} = 18.5$Å, $W = 10\mu$m). Le niveau de capacité parasite a été volontairement extrapolé dans la gamme de tensions en inversion afin d'illustrer nos explications (ligne discontinue). Le changement de pente indiqué est caractéristique de l'écrantage de la capacité C_{if}. L'évolution de la capacité de recourvrement en fonction de V_{gs} a été simulé numériquement pour différentes valeurs de dopages, en insertion de la figure.

– Le bruit de mesure introduit une erreur sur le résultat qui devient plus influent à mesure que les dimensions sont réduites.

La relation (IV.79) permet une évalutation analytique de l'erreur. Le premier terme est lié à la méthode d'extraction et introduit en général une erreur relative de 3% lorsque $L_{mask}^{ref} \geq 1\mu$m. Le terme de droite est lié à la précision de la mesure et à la variabilité du procédé de fabrication. L'incertitude absolue est d'environ $\pm$1nm, cela correspond à une erreur relative d'environ 5% sur les dispositifs les plus courts de cette étude ($L_{eff} \sim 20$nm).

$$\frac{\Delta L_{eff}}{L_{eff}} \leq \frac{\Delta L_{eff}^{ref}}{L_{eff}^{ref}} \cdot \left(\frac{C_{int}}{C_{int}^{ref}}\right) + \frac{L_{eff}^{ref}}{C_{int}^{ref}} \cdot \left(\frac{\Delta C_{int}}{C_{inv}}\right) \tag{IV.79}$$

IV.4.3.b La méthode du ΔL constant

Lorsqu'il n'est pas possible d'utiliser un transistor long vérifiant $L_{mask} \geq 1\mu$m (cas particulier des oxydes de grille fuiteux), il est toujours possible d'extraire une valeur moyenne $\overline{\Delta L}$ sur une gamme de faibles longueurs de grille. La solution est d'effectuer une régression linéaire sur les points $C_{int}(L_{mask})$, qui renvoie la valeur de $\overline{\Delta L}$ à l'intersection avec l'axe des x et la valeur de C_{eff} à partir de la pente (IV.80), comme dans [Yao86].

$$C_{inv} = C_{eff} \times \left(L_{mask} - \overline{\Delta L}\right) \tag{IV.80}$$

La méthode du ΔL constant produit une erreur significative sur les valeurs de L_{eff} extraites car $\Delta L(L)$ n'est en général pas constant pour des technologies sub-100nm. Ce phénomène est le résultat des limites de la lithographie pour les très petites dimensions. $\Delta L(L)$ peut présenter des variations de plus de 20nm pour certaines technologies, répercutant ainsi une erreur de plus de 10% sur la valeur finale de L_{eff}. Une solution consiste à effectuer la régression sur

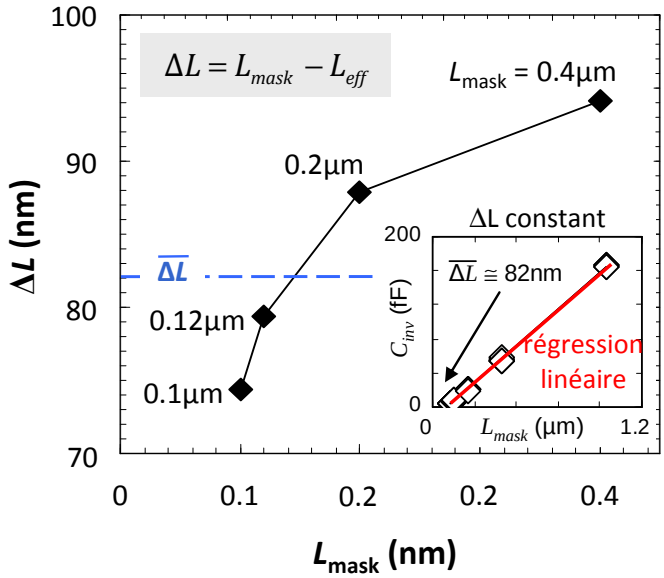


Fig. IV.17 – Comparaison entre les valeurs de ΔL extraites avec la méthode $\Delta L(L_{mask})$ et celles extraites grâce à la méthode du ΔL constant. L'extraction a été effectuée sur des transistors nMOS précurseurs à la technologie 45nm [Dumont07, Pouydebasque05] (t_{ox} = 12Å, $W = 1\mu$m). Dans ce résultat, la valeur de la capacité en inversion à été mesurée pour V_{gt} = 850mV.

une gamme de longueurs proches les unes des autres sur lesquelles il semble plus pertinent de supposer ΔL constant du fait d'une meilleure optimisation du procédé de fabrication (par exemple $L_{mask} \leq$ 100nm pour les technologies 45nm et 65nm).

IV.4.3.c Comparaison des deux méthodes d'extraction

L'extraction de la longueur effective a été effectuée sur des dispositifs précurseurs à la technologie 45nm [Dumont07, Pouydebasque05], en utilisant les deux méthodes présentées précedemment. Les réductions de longueur de grille ainsi obtenues sont illustrées sur la figure IV.17 pour des L_{mask} allant de 100nm à 0.4μm, sachant que $L_{mask} = 1\mu$m a été choisi comme référence dans l'extraction. La méthode "ΔL constant" renvoie une valeur unique qui correspond à la moyenne des valeurs obtenues par la méthode $\Delta L(L_{mask})$. Par ailleurs, les deux méthodes fournissent des valeurs de ΔL élevées, qui traduisent l'utilisation volontaire d'un procédé de *trimming* pour racourcir les longueurs de grille [Dumont07].

IV.4.4 Automatisation de l'extraction

La mesure d'une courbe $C_{gc}(V_{gs})$ complète requiert trop de temps et n'est pas compatible avec les exigences des tests automatiques en industrie. Dans cette partie, nous allons voir comment réduire ce temps de mesure afin de permettre une extraction de L_{eff} à grande échelle et une intégration dans les procédures d'extraction systématique pour optimiser la technologie au cours de son développement.

IV.4.4.a Améliorations apportées concernant les structures de test

Lorsqu'une matrice de connexion est utilisée, la mesure de capacité requiert des surfaces supérieures à 50μm^2 pour obtenir des rapports signal-sur-bruit, nécessitant des largeurs de plus de 1000μm pour des transistors nominaux. Nous proposons donc d'utiliser des structures composées

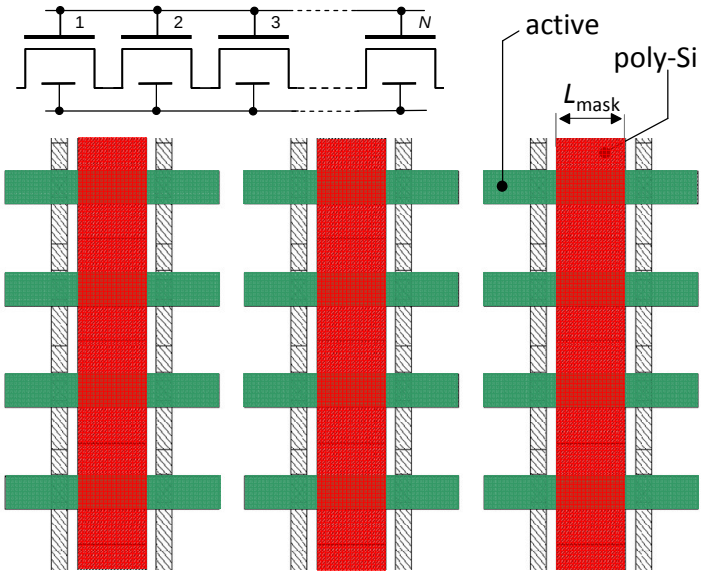


Fig. IV.18 – Structure de test matricées, utilisées pour mesurer L_{eff} de manière automatique en utilisant une matrice de connexion. La structure présentée ici est composées de 1980 transistors de dimensions $L_{mask} = 0.15\mu$m et $W_{mask} = 0.38\mu$m, fournissant une surface équivalente de 113μm^2.

de N transistors matricés (connectés en parallèle) qui permettent d'obtenir une surface équivalente identique pour toutes les longueurs, et autorisent la mesure sur des transistors courts dans des installations utilisant une matrice de connexion, cf. figure IV.18. Pour une largeur W donnée, une batterie de transistors aux longueurs variables est disponible. Le nombre de transistors N est ajusté pour chaque longueur de manière à obtenir une surface équivalente $N \times W \times L_{mask} \sim 100\mu$m^2, qui permet d'effectuer une mesure complète $C_{gc}(V_{gs})$ en 2-3 minutes. Ce type de structures est parfaitement adapté à la technologie CMOS LSTP ayant un oxyde de grille $t_{ox} \geq 15$Å [Fleury08a].

IV.4.4.b Optimisation du temps de mesure

Il n'est pas nécessaire de mesurer une caractéristique $C_{gc}(V_{gs})$ complète pour extraire L_{eff} et par conséquent le nombre de points peut être réduit, en réduisant les plages de V_{gs} uniquement aux zones de la courbe où les paramètres ($C_{gc,int}$, C_{par} et V_{th}) doivent être extraits. La figure IV.19 présente les caractéristiques $C_{gc}(V_{gs})$ typiques ainsi que leur dérivée première, obtenues sur des structures matricées (ici $W = 1\mu$m et $N = 5000$, technologie CMOS 45nm LSTP). En premier lieu, remarquons que V_{th} peut être facilement mesurée grâce à une dizaine de points dans une plage de V_{gs} calibrée à l'avance pour la technologie. C_{par} est mesurée en un point donné, proche du régime d'accumulation, dans la zone où C_{if} est en cours d'écrantage et où la désertion des extensions est encore peu influente. Une valeur ΔV_{par} est définie et calibrée pour la technologie de manière à respecter l'égalité $C_{gc}(V_{th} - \Delta V_{par}) = C_{par}$, comme illustré sur la figure IV.19. Finalement, le niveau de capacité en inversion est mesuré à une tension $V_{th} + \Delta V_{inv}$ qui correspond à la tension V_{gs} pour laquelle l'extraction de L_{eff} est souhaitée : $C_{int} = C_{gc}(V_{th} + \Delta V_{inj}) - C_{gc}(V_{th} - \Delta V_{par})$. Il est alors possible d'effectuer la mesure de C_{int} sur un dispositif court en une dizaine de points et de réduire les temps de mesure à quelques secondes par dispositif ($t = 4 \pm 2$s). Ce temps dépend notamment du nombre de points utilisés par l'algorithme d'extraction, de l'équipement de mesure et du temps d'intégration configuré sur celui-ci.

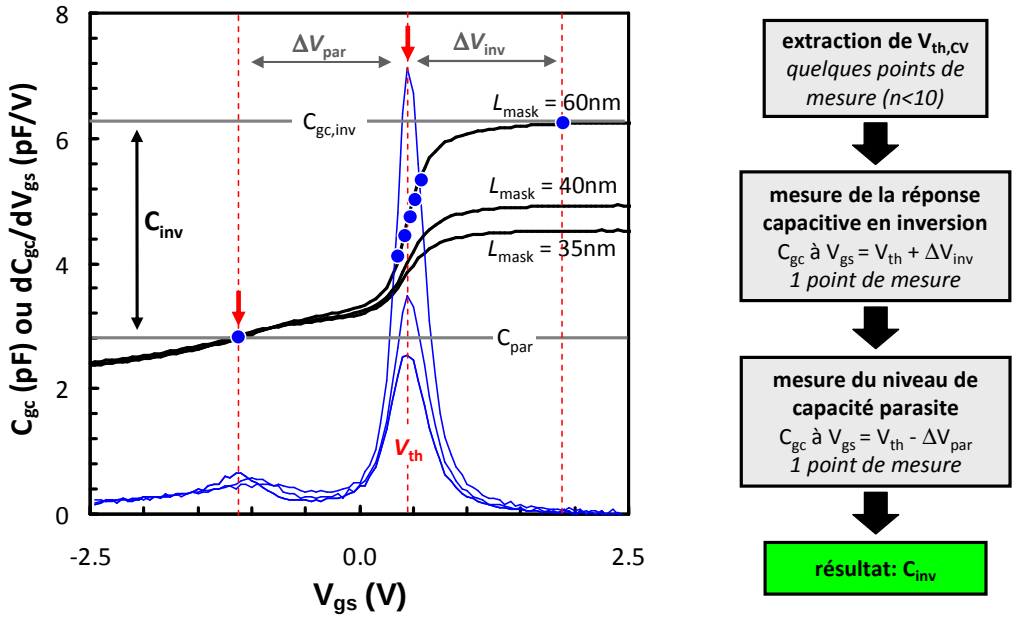


Fig. IV.19 – Illustration de la procédure d'extraction automatique mise en place. Quelques points de mesure ($n \leq 10$) permettent d'extraire la tension de seuil qui est ensuite réutilisée pour extraire la valeur de la capacité en inversion corrigée des contributions parasites : C_{int}. Les valeurs ΔV_{inj} et ΔV_{par} sont étalonnées préalablement pour la technologie.

IV.4.5 Influence de la fuite de grille sur l'extraction

Pour des tensions V_{gs} élevées, de grande surfaces effectives et des oxydes fins, la réponse capacitive peut être sujette aux fuites de grille qui se répercutent généralement sur le signal capacitif comme le montre la figure IV.20. Pour expliquer ce phénomène il est nécessaire de comprendre le fonctionnement d'un capacimètre. Ce dernier mesure une impédance et se base sur un schéma équivalent R_{gate}-C ou la résistance de la grille (R_{gate}) et la capacité de la structure sont connectées en série. Lorsque la fuite de grille devient importante et que la résistance tunnel, caractéristique du passage du courant à travers l'oxyde, devient du même ordre de grandeur que la résistance de grille le modèle n'est plus respecté et l'extraction de la valeur de la capacité depuis Z renvoit une valeur fausse. En réalité un modèle plus juste consisterait à introduire une résistance R_s câblée en parallèle avec le capacité C_s comme le montre le schéma équivalent de gauche sur la figure IV.20.

Si R_s est négligée et que le modèle $R_p \parallel C_p$ est valide, l'inductance complexe peut s'écrire comme (IV.81), où f représente la fréquence de mesure.

$$Z = \frac{1}{i \cdot 2\pi C_p \cdot f} + R_p \tag{IV.81}$$

Il est alors possible de définir un paramètre D appelé facteur de dissipation, défini par (IV.82).

$$D \overset{\text{def}}{=} \frac{\mathcal{I}m\{Z\}}{\mathcal{R}e\{Z\}} = -\frac{1}{2\pi f C_p R_p} \tag{IV.82}$$

Ce paramètre est assez pratique puisqu'il peut facilement être utilisé pour définir la précision sur la mesure de capacité (IV.83).

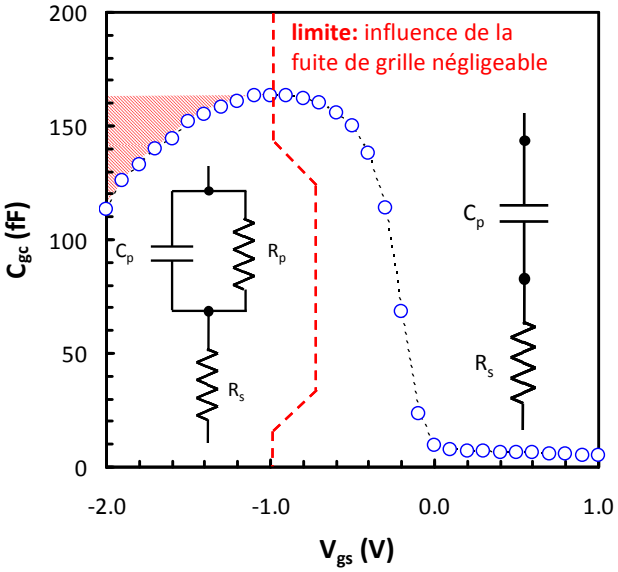


Fig. IV.20 – Influence de la fuite de grille sur la mesure $C_{gc}(V_{gs})$ d'un transistor pMOS. L'apparition d'un courant au travers de l'oxyde impose ici de considérer une conductance $(1/R_p)$ non nulle dans le schéma équivalent de l'appareil de mesure (HP4284).

$$\frac{\Delta C}{C} = A_0 \cdot \sqrt{1 + D^2} \tag{IV.83}$$

Cette relation montre clairement que plus la fuite de grille est importante (i.e. plus la résistance R_p est faible), plus la précision de la mesure sera altérée : une faible erreur sur l'estimation de la phase du signal sera amplifiée par le facteur $1/R_p$ dans le calcul de C_p. Le terme A_0 est un paramètre qui quantifie l'incertitude propre à l'équipement de mesure. Il est clair que D doit être le plus faible possible (fuite de grille négligeable) et inférieur à 1 pour conserver une incertitude minimale (proche de A_0).

Il n'existe pas à ce jour de méthode fiable pour corriger l'influence d'une fuite de grille importante sur la réponse capacitive. En effet, la plupart des techniques de correction actuelles (ex : [Barlage00]) perdent leur validité lorsque la fuite de grille est importante. En particulier, le *debiasing* qui caractérise les dispositifs fuiteux ne peut pas être modélisé précisément. Une solution consiste à effectuer des mesures à hautes fréquences (appelées HF-CV) qui permettent de supprimer l'influence de la fuite de grille mais nécessitent des structures et des équipements de tests spécifiques. Le lecteur pourra se référer aux travaux dans [Schmitz03a, Schmitz03b, Schmitz04a, Schmitz04b, Andres06] pour plus de détails.

IV.4.6 La mesure de L_{eff} : applications pratiques

IV.4.6.a Extraction de la longueur de recouvrement L_{ov}

La longueur de recouvrement est définie par la relation $L_{ov} \stackrel{\text{def}}{=} L_{gate} - L_{eff}$ et peut être extraite en connaissance de la longueur effective, qui peut être mesurée par voie capacitive, et de la longueur de grille réelle L_{gate} qui peut être mesurée par voie optique en utilisant des techniques de microscopie électronique à balayage. Ces deux grandeurs ont été extraites pour des dispositifs nMOS de la technologie 45nm LSTP et les réductions de longueurs de grille ($L_{mask} - L_{gate}$ et $\Delta L \stackrel{\text{def}}{=} L_{mask} - L_{eff}$) ont été reportées sur la figure IV.21. Sans surprise, L_{eff}

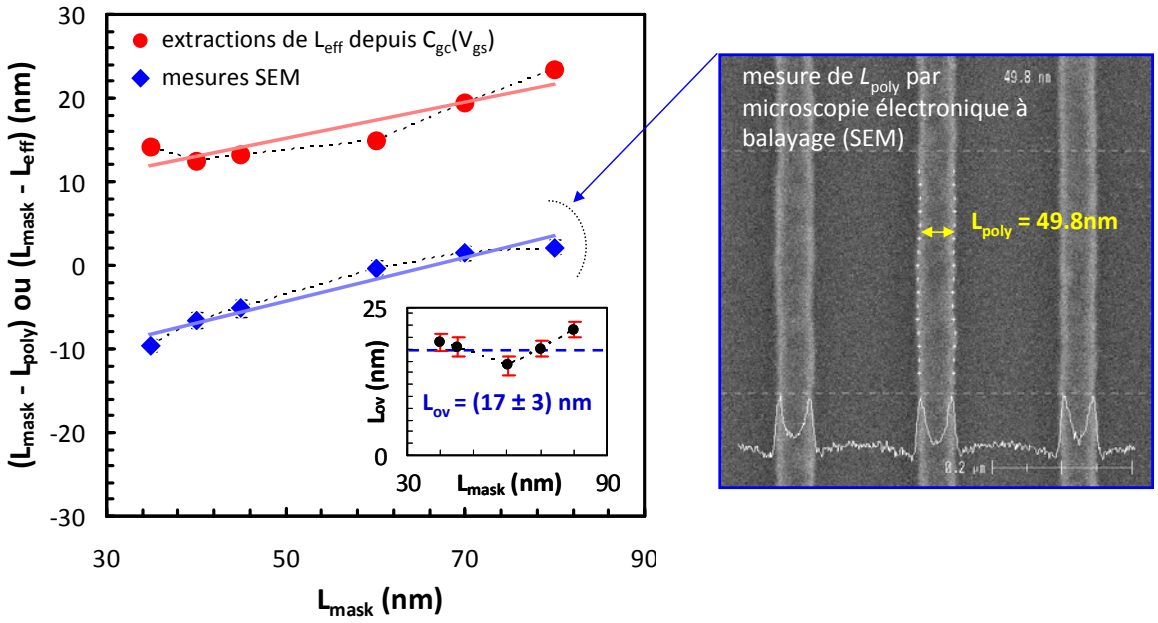


Fig. IV.21 – Comparaison entre la valeur de la réduction de la longueur de canal $L_{mask} - L_{eff}$ obtenue par mesures capacitive et la valeur $L_{mask} - L_{gate}$ mesurée grâce à des techniques de microscopie électronique à balayage. Les deux longueurs L_{eff} et L_{gate} présentent des variations similaires et la longueur de recouvrement $L_{ov}(L_{mask})$ est quasi-constante ($L_{ov} = 17 \pm 3$nm), en accord avec le fait que l'architecture de jonction reste la même quelque soit la longueur du dispositif.

et L_{gate} ont des tendances similaires et la longueur de recouvrement (cf. insertion de la figure) dépend peu de L_{mask}, ce qui semble cohérent avec le fait que l'architecture de jonction est la même pour toutes les longueurs. L'extraction permet ainsi d'estimer $L_{ov} = 17 \pm 3$nm, valeur qui semble cohérente pour cette technologie.

L'extraction de L_{ov} reposant sur ces deux techniques de mesures permet de valider la précision de notre méthode d'extraction par voie capacitive. Précisons que L_{met} ne peux pour l'instant pas être mesurée directement par imagerie électronique (SEM ou TEM) et nécessite des techniques innovantes telles que l'holographie TEM [Cooper08] ou SSRM (*Scanning Spreading Resistance Microscopy*) [Zhang08], qui à ce jour semblent tout juste pouvoir atteindre une précision de ±1nm.

IV.4.6.b Application aux mesures de durée de vie porteurs chauds (HCI)

La longueur effective de canal joue un rôle fondamental dans la dégradation induite par les porteurs chauds car elle définit l'amplitude du champ électrique longitudinal $\mathcal{E}_y \propto 1/L_{eff}$ et donc l'énergie cinétique des porteurs [Hu85]. En effet les porteurs ayant une énergie cinétique élevée peuvent passer au travers du diélectrique ou générer une paire électron-trou qui augmente le rapport des courants I_b/I_d, où I_b représente le courant de substrat. Dans les deux cas, un porteur peut être injecté dans l'oxyde et dégrader les propriétés du dispositif (c'est le phénomène de HCI, pour *Hot Carrier Injection*). Dans notre cas, la condition de dégradation est atteinte lorsque le courant de saturation I_d^{sat} sort des limites de fonctionnalité fixées pour la technologie et la durée de vie représente le temps écoulé pour atteindre cette condition de défaillance.

Les extractions de durée de vie HCI pire cas ($V_{gs} = V_d$) ont été effectuées sur deux types de dispositifs issus de la technologies CMOS 65nm LSTP, ayant un oxyde de grille SiON $t_{ox} \simeq 18.5$Å.

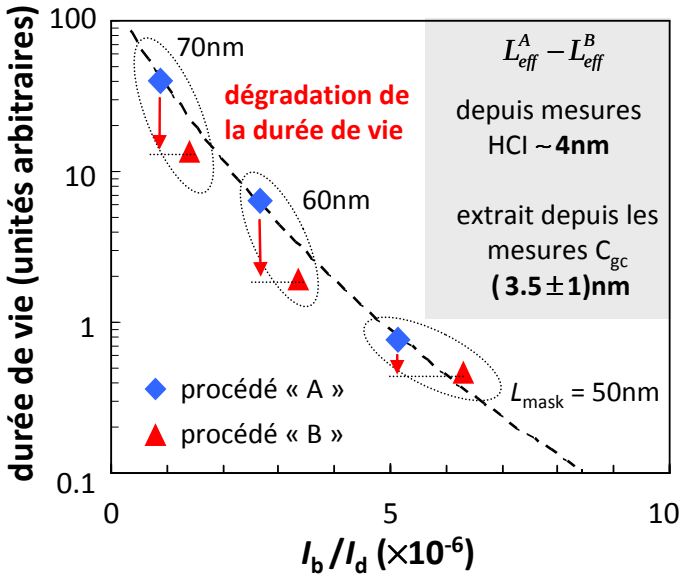


Fig. IV.22 – Evolution de la durée de vie lors de dégradation par porteurs chauds en fonction du rapport I_b/I_d (proportionnel au taux d'ionisation par impact) pour deux procédés de fabrication différents "A" et "B", entre lesquels L_{eff} est susceptible d'avoir changé. La dégradation de la durée de vie pour $L_{mask} = 70$nm peut être attribuée à un modification de la longueur effective d'environ 4nm, pour que les courbes $I_b/I_d(L_{mask})$ des deux procédés concordent. L'extraction de L_{eff} par mesure capacitive donne $L_{eff}^A - L_{eff}^B = (3.5 \pm 1)$nm qui valide la valeur extrapolée depuis les courbes $I_b/I_d(L_{mask})$.

Ces deux procédés de fabrications présente une modification des durées de vie HCI qui peut être attribuée à une modification de la longueur effective d'environ 4nm si ce paramètre est considéré comme le seul responsable de la modification de la durée de vie. Cet écart de 4nm est en effet nécessaire pour joindre les deux tendances I_{bulk}/I_d de la figure IV.22. L'extraction de la longueur effective en utilisant la technique décrite au §IV.4.3 permet de vérifier cette hypothèse. L'extraction, effectuée sur 20 puces, permet d'extraire un écart $L_{eff}^A - L_{eff}^B = 4 \pm 0.5$nm qui semble concorder parfaitement avec l'estimation faite depuis la figure IV.22. L'utilité d'une mesure de L_{eff} au cours du procédé de développement est donc démontrée.

IV.4.6.c *Monitoring* de L_{eff} avec des changements de procédé

La longueur effective est influencée par de nombreuses étapes du procédé de fabrication et le meilleur moyen pour éviter une dérive de cette grandeur est de la mesurer systématiquement au cours du développement de la technologie, grâce à la méthodologie décrite au §IV.4.3. Pour illustrer la dérive que peut subir L_{eff} lors d'une modification du procédé de fabrication, des extractions de L_{eff} ont été appliquées à deux types de dispositifs de la technologie CMOS 45nm LSTP pour lesquels la température de recuit d'activation (*Rapid Thermal Process*, RTP) à été changée de 30°C (980°C → 1010°C).

S'il semble logique de considérer qu'une augmentation du budget thermique provoque une diffusion accrue des dopants des extensions S/D, l'extraction permet de quantifier cette diffusion supplémentaire et de mesurer précisément le décalage de L_{eff} entre les deux variantes du procédé. Les résultats sont représentés sur la figure IV.23. Cette dernière montre une réduction de la longueur de grille de 2nm lorsque la température de ce recuit passe de 980°C à 1010°C,

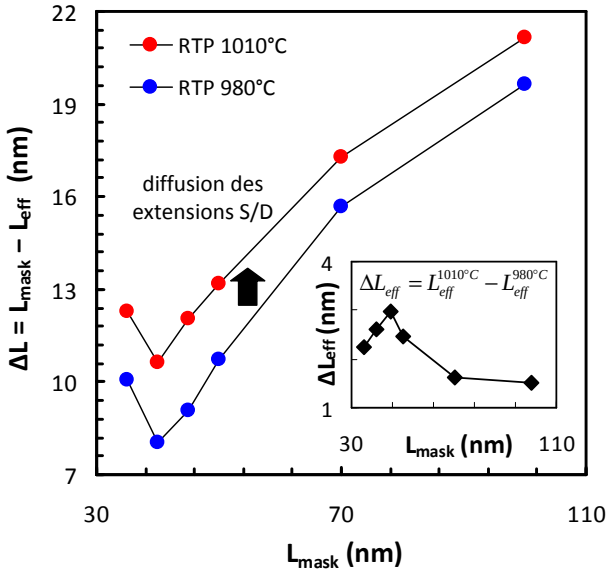


Fig. IV.23 – Variation de la longueur effective due à la diffusion des dopants lorsque la température du recuit est changée dans le procédé de fabrication. Dispositifs issus de la technologie CMOS 45nm LSTP.

validant par la même occasion l'hypothèse d'une diffusion accrue des dopants des extensions (Arsenic dans ce cas). La connaissance de cette dérive peut être facilement corrigée au cours du développement du procédé en réajustant la longueur de grille L_{gate} et permet ainsi d'éviter une dérive des performances (I_{on}, fiabilité HCI).

Une étude similaire peut être menée concernant la dose des poches dans le dispositif. En effet, les poches permettent, par une augmentation du dopage du canal pour les petits L_{eff}, de rétablir un bon contrôle électrostatique dans les transistors courts. Au fur et à mesure que le dopage des poches augmente, la longueur effective augmente car les jonctions métallurgiques sont décalées vers l'extérieur du canal. La figure IV.23 présente le profil $L_{eff}(L_{mask})$ obtenu pour des dispositifs de nMOS 45nm LSTP ayant différentes doses de poches, en utilisant la technique d'extraction $\Delta L(L_{eff})$. Les résultats montrent une augmentation importante de L_{eff} lorsque le dosage des poches augmente ($2 \times 10^{13}/\text{cm}^3 \rightarrow 4 \times 10^{13}/\text{cm}^3$), puis semble se stabiliser au delà de $4 \times 10^{13}/\text{cm}^3$.

Il semble donc que la mesure de L_{eff} en fonction du dosage des poches fasse apparaitre un compromis entre la mobilité, qui sera plus élevée pour un faible dosage des poches et les performances en fiabilité (durée de vie porteur chaud) qui seront dégradées si la dose de poche est réduite.

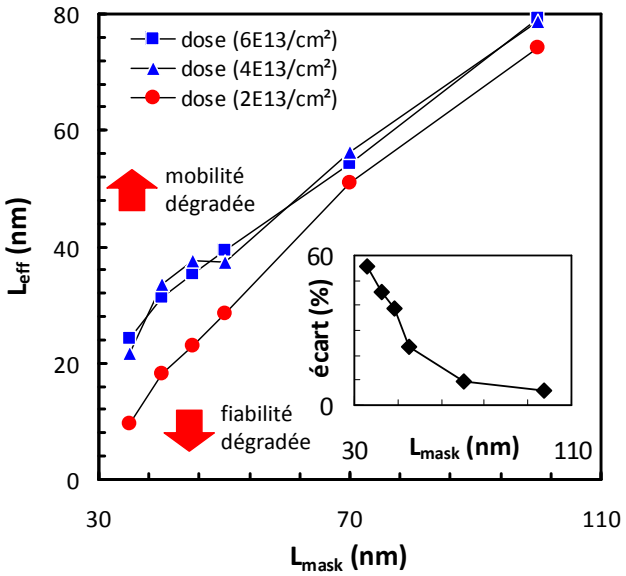


Fig. IV.24 – Evolution de la longueur effective des transistors pour différentes doses de poche. Cette analyse semble faire apparaitre un compromis entre les performances en courant (mobilité) et la fiabilité des dispositifs courts.

IV.5 Conclusion et perspectives

Si la méthode d'extraction de la longueur effective de canal reste relativement simple, son application sur des transistors courts requièrt une méthode de correction des contributions parasites qui reste complexe et requièrt une bonne compréhension physique de la structure. Les avancées apportées par les travaux de thèse sur la compréhension de la capacité parasites ont été exposées dans ce chapitre. Elles ont permis d'établir une méthode robuste qui permet une estimation de L_{eff} avec une précision record, sur des technologies modernes qui posent généralement plusieurs problèmes avec les autres techniques. L'extraction de L_{eff} reste néanmoins lourde à mettre en place. Elle nécessite des structures de test spécifiques et des temps de mesure longs, en comparaison des mesures de courant. Son utilité est pourtant démontrée, en particulier pour les aspects fiabilité HCI pour lesquels cette grandeur devient critique.

L'innovation dans les architectures des transistors de technologie 32nm et en deçà ne joue pas en notre faveur. L'intégration de nouvelles briques technologiques (canal enterré, empilement *high-*κ/métal complexes, canal germanuré) modifie l'allure de la réponse capacitive et nécessite une réévaluation des concepts posés, pour justifier de la validité de la méthode d'extraction. Une modélisation prenant en compte des effets quantiques semble indispensable. Il s'agira alors de réutiliser les briques élémentaires qui ont été posées dans la partie IV.2. **Le travail ne manque pas !**

Chapitre V

Étude expérimentale du transport dans le transistor MOS

Sommaire

V.1 Évaluation des propriétés de transport en régime linéaire 161
V.1.1 L'extraction des paramètres essentiels : L_{eff}, R_{sd}, V_{th} 161
V.1.2 Extraction de la mobilité en régime linéaire 163
V.1.3 Evaluation expérimentale de la réduction de mobilité 164
V.1.4 Proposition d'un nouveau modèle de dégradation de mobilité avec L_{eff} 165
V.1.5 L'extraction de coefficient de rétrodiffusion en régime linéaire 166
V.1.6 Influence de la balisticité dans la dégradation de mobilité 166
V.1.7 Bilan sur les propriétés de transport à bas champ 167
V.2 A la recherche de la balisticité 168
V.2.1 Un enthousiasme dans le domaine de la simulation 170
V.2.2 Les méthodes expérimentales pour évaluer la balisticité du transport . . 170
V.2.3 Les points faibles des méthodes existantes 175
V.2.4 Proposition d'une nouvelle méthodologie d'extraction 176
V.2.5 Protocole expérimental suivi dans la thèse 178
V.3 Évaluation des propriétés de transport en régime de saturation . . 182
V.3.1 Extraction de la vitesse effective 183
V.3.2 Extraction de ν_{lim} et détermination de sa nature 185
V.3.3 Analyse de nos résultats en s'appuyant sur un modèle de survitesse . . . 188
V.4 Bilan et discussions 191

RÉSUME — Si de nombreux ouvrages relatent des avantages d'un transport balistiques, la plupart ignore les difficultés que l'industrie rencontre aujourd'hui pour concevoir des transistors ayant de bonnes propriétés de transport. De ce fait, la question de l'influence des mécanismes de collisions dans le transport des porteurs de la source au drain reste ouverte. Le meilleur moyen de répondre à ces interrogations reste encore l'étude expérimentale du transport (de la mobilité et de la vitesse des porteurs) qui fait l'objet de ce chapitre, et complète ainsi les bases théoriques posées au chapitre I.

⋆ En premier lieu, nous étudierons en détail les propriétés de transport qui caractérisent le régime linéaire de nos transistors (§V.1). Cette partie permettra d'évaluer la qualité du transport dans nos dispositifs et de déterminer les mécanismes de collisions les plus limitant dans le transport.

⋆ Dans une seconde partie (§V.2), nous détaillerons les méthodes expérimentales qui permettent à ce jour d'extraire le taux de balisticité. Nous introduirons notre nouvelle méthodologie (en le nouveau protocole d'extraction qui lui est associé) qui permet de résoudre les problèmes engendrés par l'utilisation des autres méthodes.

⋆ La troisième partie de ce chapitre (§V.3) consistera en l'étude des propriétés de transport en régime saturé. En faisant le lien avec les résultats de la partie (§V.1) et en utilisant la nouvelle méthodologie d'extraction du §V.1, nous tenterons d'élucider la nature du mécanisme qui limite le courant I_{on} sur les dispositifs courts : ν_{sat} ou ν_{inj} ?

⋆ La dernière partie §V.4 nous permettra de conclure quant à la nature du transport dans nos dispositifs et aux évolutions qui viendront dans le futur avec la réduction des longueurs de grille.

V.1 Évaluation des propriétés de transport en régime linéaire

L'analyse des propriétés de transport en régime linéaire ($V_{ds} \ll V_{gt}$) est primordiale pour la compréhension des mécanismes de transport dans le canal. Elle permet d'extraire la mobilité qui donne une indication sur la capacité des électrons à se mouvoir, et permet une estimation du libre parcours moyen (λ_μ). Par ailleurs, la mobilité *dérive-diffusion* est expérimentalement corrélée avec le courant I_{on} [Lundstrom01, Lochtefeld01]. L'analyse du lien qui existe entre les propriétés de conduction du régime linéaire et de saturation permet enfin la compréhension des mécanismes de saturation de vitesse (ν_{lim} et ν_{sat}) qui se manifestent à forts champs.

V.1.1 L'extraction des paramètres essentiels : L_{eff}, R_{sd}, V_{th}

Extraction de L_{eff} – Elle est réalisée grâce à notre méthode automatisée [Fleury08b], décrite au §IV.4.4. La partie gauche de la figure V.1 présente la variation de la réduction de longueur de canal en fonction de la L_{mask}. Le paramètre ΔL augmente lorsque la longueur de grille diminue, signe d'une influence des procédés de gravure et/ou de lithographie. Ce décentrement est particulièrement utile dans notre étude car il permet d'atteindre des valeurs de L_{eff} proches 20nm, ce qui aurait été impossible sur une plaque *centrée*. La partie droite de la figure V.1 présente les variations $L_{eff}(T)$ et $C_{eff}(T)$. Si la valeur de la capacité en inversion reste quasi-constante quelque soit la température (ce qui est attendu), la longueur effective présente une légère variation lorsque la température change. Ce phénomène peut résulter d'une légère déviation des points d'extraction dans la méthodologie d'extraction automatique (choix de ΔV_{par} pour la correction des capacités parasites par exemple).

Extraction de la tension de seuil – La tension de seuil a été extraite en régime linéaire ($V_{ds} = 10$mV) via la technique détaillée au §III.5. Le DIBL a été mesuré à courant constant, sous le seuil, et retranché de la valeur de V_{th}^{lin} extraite en régime linéaire pour calculer V_{th}^{sat} selon la relation (V.1), comme dans [Barral07, Barral09b].

$$V_{th}^{sat} = V_{th}^{lin} - DIBL \tag{V.1}$$

Précisons que si les techniques du type V_{th} extrapolé en régime de saturation peuvent fonctionner sur les transistors longs, l'observation que nous venons de faire au §V.1.5 montre clairement que le courant I_{on} des transistors courts ne suit pas une règle de proportionnalité en V_{gt}^2. Il est donc impossible d'aboutir à une technique d'extraction fiable depuis la caractéristique $\sqrt{I_{on}}(V_{gs})$, comme proposé dans [Ortiz-Conde02].

Extraction de R_{sd} – La résistance série $R_{sd}(V_{gs})$ a été extraite en utilisant la technique $R_{tot}(1/\beta)$ [Fleury09b, Fleury09c]. Cette dernière permet une précision accrue et la prise en compte la dépendance $R_{ds}(V_{gs})$ pour une correction plus fine. Nous n'utiliserons pas les valeurs proches du seuil qui peuvent être erronées à cause de la proximité du seuil et de la validité limitée des modèles sur lesquels reposent la technique d'extraction. La dépendance de R_{sd} avec la température est représentée pour différentes valeurs de V_{gt} sur la partie droite de la figure V.2. La dépendance $R_{sd}(T)$ est faible et *noyée* dans une dispersion qui provient de l'extraction et probablement de la résistance de pointe qui fluctue au cours du test. Les valeurs utilisées seront donc moyennées sur la gamme de températures $T \leq 350K$ (qui semble la plus fiable), cf. partie droite de la figure V.2.

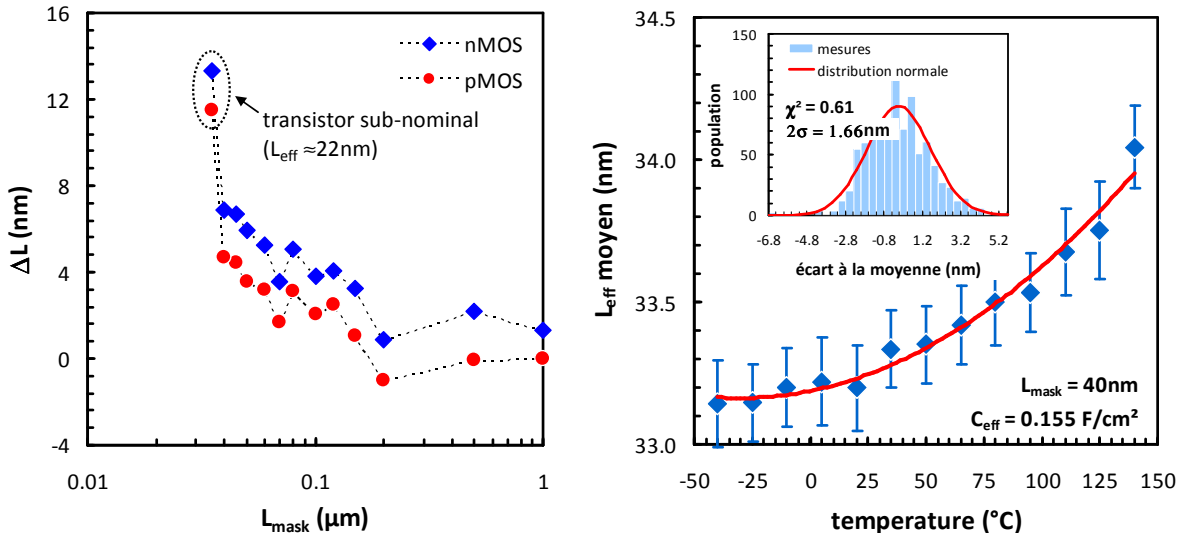


Fig. V.1 – A gauche : extraction de la réduction de longueur de canal en fonction de L_{mask} pour les deux types de transistors. A droite : évolution L_{eff} en fonction de la température. La précision de la mesure est de ±1.7nm dans un intervalle de confiance de 95% (cf. histogramme en insertion).

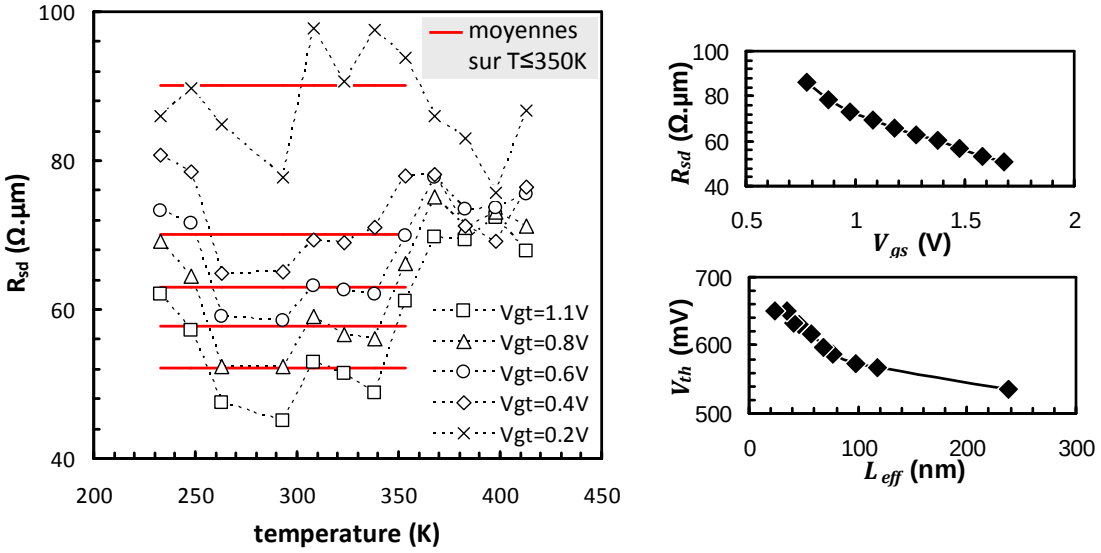


Fig. V.2 – A gauche : extraction de la valeur R_{sd} pour différentes tensions V_{gt}. La précision de la méthode ne permet pas d'extraire une dépendance en température et une valeur constante, mesurée sur la gamme 210K–350K est considérée. A droite : évolution $R_{sd}(V_{gT})$ et $V_{th}(L_{eff})$ pour ces dispositifs.

V.1.2 Extraction de la mobilité en régime linéaire

La mobilité à bas champ (μ_0) peut être extraite depuis les courbes $I_d(V_{gs})$, en utilisant la méthodologie décrite au §III.4.3. Les résultats montrent une chute de mobilité caractéristique sur les dispositifs courts qui peut être attribuée aux mécanismes additionnels de réduction de la mobilité (poches ou défauts neutres), détaillés au §II.2.5. La mobilité bas champ a été extraite depuis les courbe $I_d(V_{gs})$ en utilisant la fonction d'extraction ξ et de longueur effective détaillée au §III.4.3 et §IV.4.3. La figure V.3 présente les profils $\mu_0(L_{eff})$ obtenus à $T = 230$K et $T = 410$K. Cette dégradation de mobilité peut être modélisée en utilisant le modèle développé en collaboration avec G. Bidal [Bidal09a] par une relation de la forme (V.2), similaire à [Zilli07a].

$$\frac{1}{\mu(L_{eff})} = \frac{1}{\mu_{max}} + \frac{\alpha_\mu}{L_{eff}} \tag{V.2}$$

la figure V.3 présente ce profil à température ambiante sur lequel à été superposé le modèle détaillé dans [Bidal09a] afin de vérifier la cohérence du paramètre α_μ par rapports aux autres technologies. Le couple de paramètre (μ_{max}, α_μ) semble en ligne avec celles données des autres lots qui sont répertoriés dans [Bidal09a].

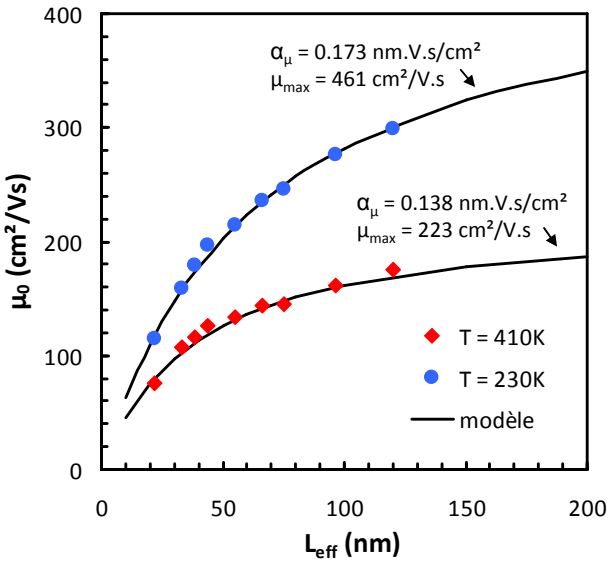


Fig. V.3 – Evolution de la mobilité à bas champ (μ_0) en fonction de la longueur effective de canal pour deux températures : 230K et 410K. Le modèle "linéaire" de dégradation de mobilité issue de [Bidal09a] a été ajusté sur les points expérimentaux.

V.1.3 Evaluation expérimentale de la réduction de mobilité

Dans cette partie nous utiliserons la mobilité effective, mesurée à charge d'inversion constante (i.e. $V_{gt} = 1.1$V) qui renvoie une image fidèle de la qualité de transport en régime d'inversion forte. Admettons aussi que les composantes de la mobilité associées aux mécanismes d'interaction avec les phonons acoustiques et de la rugosité de surface ne dépendent pas de la L_{eff}, ce qui parait une hypothèse raisonnable compte tenu de ce qui a été dit au chapitre II. Rappelons également que la mobilité extraite depuis la fonction-ξ ne permet pas de voir la contribution des charges fixes car ces dernières sont écrantées par la forte densité de porteurs libres de l'inversion forte (cf. §II.2.2). La valeur de la contribution de la mobilité responsable de la dégradation observée pour les transistors courts peut alors être estimée par le biais de la relation de Matthiessen (V.3). Nous nommerons cette contribution additionnelle $\mu_N(L_{eff})$ en vertu des travaux publiés dans [Cros06b]. Dans cette relation, μ_∞ représente la mobilité bas champ d'un transistor infiniment long, dans lequel les effets de dégradation de mobilité ne sont pas présents et peuvent être négligés. Dans l'expérience, nous choisirons $\mu_\infty = \mu_0(L = 1\mu\text{m})$.

$$\frac{1}{\mu_N(L_{eff})} \simeq \frac{1}{\mu_0(L_{eff})} - \frac{1}{\mu_\infty} \tag{V.3}$$

La figure V.4 présente l'allure $\mu_N(L_{eff})$ obtenue depuis (V.3). Elle confirme une tendance de variation exponentielle comme observé dans [Cros06b]. La faible variation $\mu_N(T)$ peut provenir d'une contribution de coulomb résiduelle qui n'est pas compensée correctement par la soustraction de la mobilité du transistor à canal long. Nous privilégierons donc les valeurs à température élevées qui permettent de s'assurer d'un écrantage plus efficace de l'influence des charges fixes. Dans tous les cas, l'influence de ces variations est négligeable à l'échelle de l'extraction de la tendance de $\mu_N(L_{eff})$.

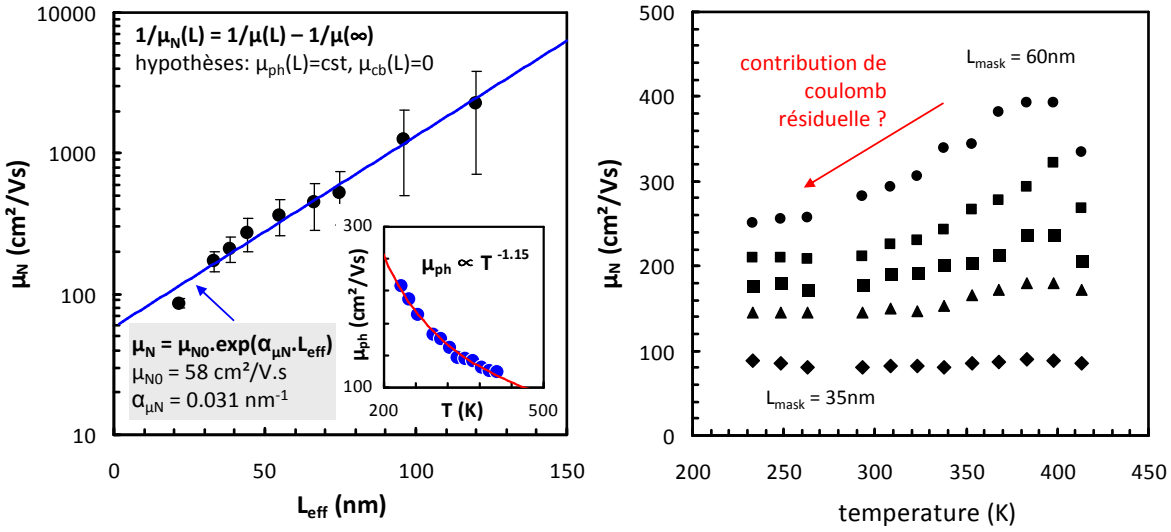


Fig. V.4 – A gauche : extraction de la dépendance en longueur de la composante neutre de mobilité : μ_N semble suivre une loi exponentielle. A droite : évolution $\mu_N(L_{eff})$ obtenue sur nos dispositifs. μ_N ne dépend pas de la température, justifiant de la validité des hypothèses utilisées pour faire ce calcul.

V.1.4 Proposition d'un nouveau modèle de dégradation de mobilité avec L_{eff}

L'allure exponentielle de $\mu_N(L_{eff})$ suggère d'introduire un nouveau modèle qui peut s'écrire comme (V.4).

$$\frac{1}{\mu(L_{eff})} = \frac{1}{\mu_\infty} + \frac{1}{\mu_{N0}} \cdot \exp\left(-\alpha_{\mu_N} \cdot L_{eff}\right) \tag{V.4}$$

La comparaison entre le nouveau modèle et l'ancien modèle est donnée sur la figure V.5 pour nos dispositifs, aux températures extrêmes de cette étude. Si les tendances $\mu_N(L_{eff})$ semblent mieux reproduites par le modèle linéaire, notre nouveau modèle a l'avantage de saturer dans le cas limite $L_{eff} \to 0$, là où l'ancien modèle retournait une valeur non physique ($\mu_N \to 0$).

Les paramètres du nouveau modèle semblent fournir une information physique concernant le mécanisme de dégradation de la mobilité. Ainsi, μ_{N0} représente la limitation de mobilité dans un canal saturé en défauts neutres (extrapolation de μ_N pour $L_{eff} \to 0$) tandis que α_{μ_N} est un paramètre lié à la distribution des défauts dans le canal. Il est corrélé à la distance critique définie dans [Cros06b] par le biais de la relation (V.5). En général, la dégradation de mobilité devient significative lorsque μ_N atteint environ 50 fois sa valeur à saturation μ_{N0} d'où $L_c \sim \ln(50)/\alpha_{\mu_N}$ μ_{N0}. Dans le cas des dispositifs de [Cros06b] cette formule renvoie $L_c \simeq 100$nm, en accord avec la valeur donnée dans la publication.

$$L_c \propto \frac{1}{\alpha_{\mu_N}} \tag{V.5}$$

Le paramètre μ_{N0} est lié à la physique du mécanisme d'interaction des porteurs avec les défauts neutres. En utilisant l'expression donnée dans [Kiréev75], on peut extraire une valeur approchée de la concentration en défauts neutres N_N à saturation (cas $L_{eff} \to 0$). L'expression de N_N est donnée en (V.6) et renvoie $N_N \simeq 8 \times 10^{19}/\text{cm}^3$ pour nos dispositifs et ceux ayant servi à l'étude dans [Cros06b]. Cette valeur est parfaitement en accord avec les simulations dans [Ghibaudo09] et semble cohérente avec l'hypothèse de collisions avec les défauts interstitiels

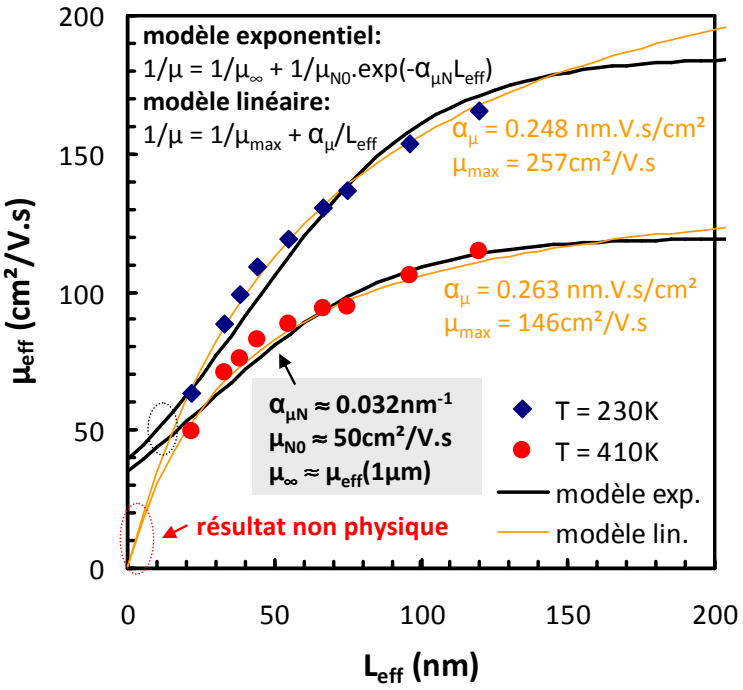


Fig. V.5 – Modélisation de la dégradation de mobilité à deux températures en utilisant le modèle "exponentiel" et le modèle "linéaire".

avancée dans [Cros06b, Dupre07].

$$N_N \simeq \frac{m^\star q^3}{20\varepsilon_{Si}\hbar^3 \cdot \mu_{N0}}. \tag{V.6}$$

Ce résultat est très intéressant car il montre une convergence des résultats expérimentaux et des modélisations. Il semble valider l'hypothèse d'une mobilité dégradée par des défauts interstitiels qui résulterait de l'implantation ionique utilisée pour les extensions source et drain. Ces défauts ne seraient pas entièrement guéris par le recuit de *guérison* qui possède un budget thermique trop faible.

V.1.5 L'extraction de coefficient de rétrodiffusion en régime linéaire

En théorie, l'utilisation du régime linéaire doit permettre de s'affranchir du problème d'identification entre ν_{inj} et ν_{sat} qui se pose en régime de saturation. En effet, pour de faibles tensions de drain (i.e. faible champ) l'influence de la vitesse de saturation peut être négligée et seule ν_{lim} peut être considéré comme responsable des éventuelles déviations à la loi d'Ohm qui peuvent être observées dans ce régime. Une extraction de r^{lin} est proposée dans [Pappas09] dans l'approximation d'une seule sous-bande. L'expression du coefficient de rétrodiffusion peut alors se réduire à (V.7).

$$r^{lin} = 1 - \frac{\mu^{lin}_{app}}{\mu^{lin}_{bal}} \tag{V.7}$$

V.1.6 Influence de la balisticité dans la dégradation de mobilité

De la même manière que dans [Cros06b], il est possible de corriger la mobilité apparente de la contribution balistique (la mobilité balistique μ_{bal} [Shur02]) afin d'estimer la composante

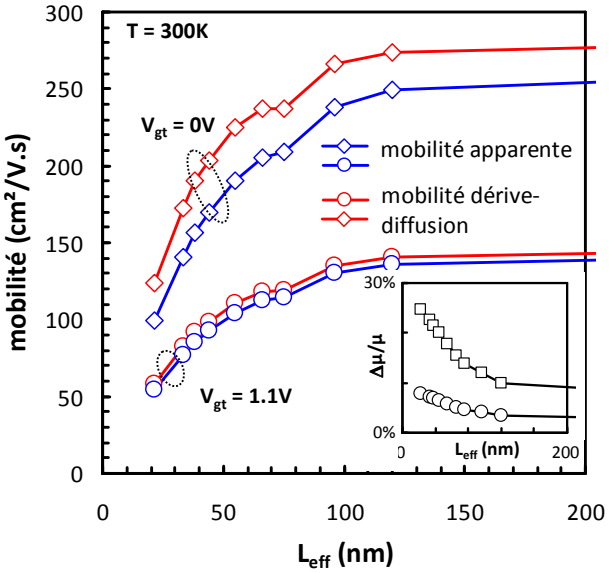


Fig. V.6 – Représentation de la mobilité apparente (mesurée expérimentalement) et de sa valeur corrigée de l'influence de la balisticité. Il a été supposé $\nu_{inj} \simeq 1.2 \times 10^7$cm/s. Ce résultat confirme le rôle mineur que joue la balisticité dans la dégradation de mobilité observée pour les courtes longueurs de grille, en accord avec [Cros06b, Zilli07a].

dérive-diffusion (μ_{dd}) et de confirmer l'origine de la dégradation observée sur les faibles longueurs de grille. La figure V.6 présente les valeurs μ_{dd} et μ_{app} pour $V_{gt} = 0$ et $V_{gt} = 1.1$V et l'insertion présente la variation relative de la mobilité liée à la correction de l'influence de μ_{bal}. Si la correction semble significative sur l'extrapolation à charge nulle, elle devient inférieure à 5% forte inversion car la valeur de μ_{dd} est déjà trop faible. Ce résultat, en accord avec [Zilli07a], confirme que la la balisticité ne permet pas à elle seule d'expliquer la dégradation de mobilité observée sur les transistors courts.

V.1.7 Bilan sur les propriétés de transport à bas champ

Cette partie nous a permis d'évaluer la qualité du transport à bas champ dans le transistor. En particulier, il a été possible de mettre en évidence la dégradation de mobilité qui affecte les canaux courts et pourrait résulter du procédé d'implantation des dopants de extensions S/D [Cros06b]. Cette réduction de mobilité traduit une augmentation du nombre de collisions élastiques et un libre parcours moyen réduit (cf. figure V.10). En particulier, la valeur $\lambda_\mu \simeq 3 \pm 2$nm pour $L_{eff} = 22$nm est obtenue en inversion forte ($V_{gt} = 1.1$V), en supposant un vitesse d'injection $\nu_{inj} = (1.8 \pm 0.3) \times 10^7$cm/s. Cette valeurs implique une probabilité de passage balistique négligeable et permet d'éliminer d'emblée cette hypothèse pour expliquer les performances de nos dispositifs en régime linéaire.

Qu'en est-il du régime de saturation ? Nous devrions nous attendre à retrouver l'influence des collisions élastiques qui dégradent le transport en régime ohmique. La partie suivante étendra donc notre analyse au régime de saturation et tentera de conclure quant à la nature des mécanismes qui limitent l'amplitude du courant dans les dispositifs courts.

V.2 A la recherche de la balisticité

Depuis l'apparition des premières théories du transport balistique de Natori [Natori94] et quasi-balistique de Lundstrom [Lundstrom97], la recherche de la balisticité dans le transistor MOS a subi un engouement croissant dans la littérature. C'est en particulier ce qui est illustré sur la Figure V.7 qui présente l'évolution du nombre de publications au sujet de la balisticité du transport au cours des 10 dernières années. Comme évoqué dans le partie §II.3.2, le terme "balistique" s'emploie pour désigner un transport sans collision dans le canal. Un tel mode de transport doit en théorie permettre d'atteindre des niveaux de courant et des fréquences de fonctionnement élevés, tout en réduisant la tension d'alimentation des dispositifs. La figure V.8 présente l'évolution du courant I_{on} et de la fréquence de coupure f_T entre le cas balistique et dérive-diffusion. Ces résultats, obtenus depuis des simulations Monte-Carlo [Palestri05, Eminente07], illustrent clairement le gain apporté par la balisticité et l'enjeu représenté par ce mécanisme de transport pour les technologies futures.

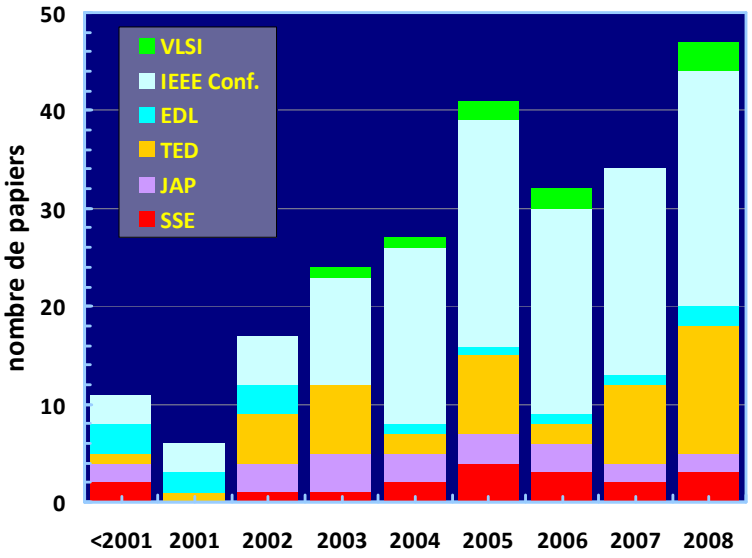


Fig. V.7 – Evolution du nombre de publications sur le thème de la balisticité dans le transistor MOS au cours de ces dix dernières années, dans les journaux et conférences les plus réputées.

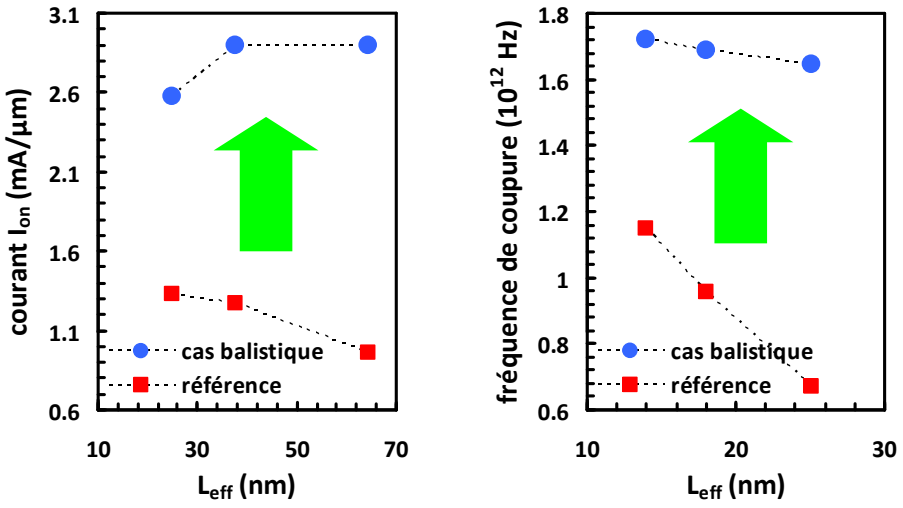


Fig. V.8 – Evolution du courant I_{on} et de la fréquence de coupure f_T en fonction de la longueur de grille des transistor, dans le cas d'un transport balistique et d'un transport dérive-diffusion d'après [Palestri05, Eminente07].

V.2.1 Un enthousiasme dans le domaine de la simulation

Si les études expérimentales portant sur la balisticité du transport sont relativement rares, les publications d'études portant sur des résultats de simulation Monte-Carlo sont nombreuses. Ces dernières ont en effet permis de vérifier la validité de la théorie quasi-balistique de Lundstrom [Lundstrom97, Lundstrom02a] et d'en comprendre les limitations [Palestri05]. Cependant, la plupart de ces études reposent sur la simulation d'une structure idéale qui possède les propriétés suivantes :

- une intégrité électrostatique quasi-parfaite (architecture double grille à canal mince) [Saint-Martin04],
- pas de dégradation de mobilité sur les transistors courts (les observations de [Cros06b, Dupre07] ne sont pas prises en compte)
- un canal non dopé (pas de poches de surdopage pour contrecarrer les effets de canaux courts)
- un grand libre parcours moyen (peu de collisions dans le canal)

Toutes ces conditions sont rarement réunies dans des dispositifs réels, dans lesquels le transport est généralement dégradé par des mécanismes additionnels (par exemple poches [Cao99] et/ou défauts neutres [Cros06b]). Il est donc extrêmement difficile de faire le lien entre ces études et les performances observées sur nos dispositifs, dont l'architecture et le comportement électrique sont radicalement différents.

V.2.2 Les méthodes expérimentales pour évaluer la balisticité du transport

Expérimentalement, la mesure du taux de balisticité dans un dispositif donné permet d'évaluer l'écart de performance qui la sépare d'un cas idéal purement balistique et permet de prévoir les évolutions de performance qui pourront être atteintes. Il est donc critique de disposer de méthode d'extraction qui permettent de savoir avec certitude quelle est l'influence de la balisticité dans nos transistors.

Il existe malheureusement un panel de méthodologies d'extraction très restreint pour évaluer la part de balisticité dans le transport des transistors MOS. Ces méthodes font souvent appel à des hypothèses qui peuvent se révéler difficiles à vérifier et qui conditionnent la validité des résultats. Elles sont basées sur le modèle quasi-balistique qui stipule que la saturation du courant à forts champs provient de la part croissante de balisticité dans le transport et éliminent d'emblée l'hypothèse d'un autre mécanisme (i.e. la vitesse de saturation). Pourtant, à l'heure actuelle nous n'avons aucune preuve qui nous permet de conclure objectivement sur la nature de la limitation du transport à forts champs.

Dans les paragraphes suivants, nous allons détailler les méthodologie d'extraction à l'état de l'art qui permettent d'extraire le taux de balisticité (BR) dans les transistors MOS.

V.2.2.a La méthode d'extraction introduite par A. Lochtefeld

La méthode d'extraction proposée par A. Lochtefeld dans [Lochtefeld01, Lochtefeld02] et utilisée dans [Dobbie08] permet d'extraire le coefficient de transmission des porteurs T[1] en couplant la mesure de la vitesse effective ν_{eff} au calcul de la vitesse thermique dégénérée ν_{inj}

1. La définition de T dans [Lochtefeld01] est identique celle de Lundstrom dans [Lundstrom97] et s'explicite en fonction du coefficient de rétrodiffusion r : $T \stackrel{\text{def}}{=} 1 - r$.

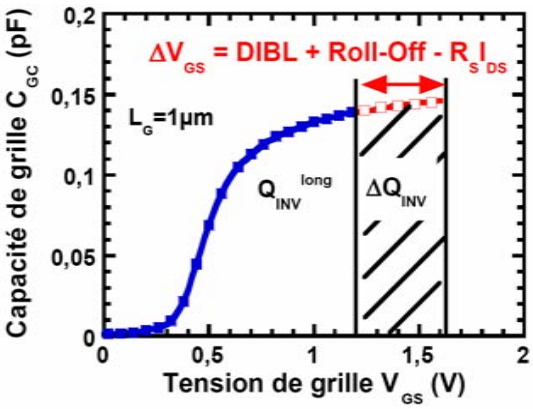


Fig. V.9 – Illustration de la méthode utilisée dans [Lochtefeld01] pour calculer la charge à la source virtuelle des transistors courts, à partir de la mesure $C_{gc}(V_{gs})$ effectuée sur un transistor long (pour lequel l'influence des capacités parasites peut être négligée).

[Assad99]. La valeur ν_{eff} est donnée par (V.8) [Lundstrom02a] qui, à l'instar de la relation utilisant le maximum de transconductance [Sodini84, Shahidi88] permet d'extraire une vitesse au niveau de la source virtuelle (position $y = 0$)

$$\nu_{eff}(0) = \frac{I_{on}}{W \cdot Q_s} \tag{V.8}$$

Toute la difficulté de cette méthode consiste donc en :

– la détermination précise de la charge d'inversion au niveau de la source virtuelle,

– la calcul d'une valeur de ν_{inj} fidèle à la vitesse d'injection réelle dans le transistor.

Concernant le premier point, la méthode de calcul de la charge proposée dans [Lochtefeld01] revient à faire l'approximation suivante $Q_s \simeq C_{eff} \cdot (V_{gs} - V_{th}(L_{eff}))$, comme dans [Lundstrom02a]. Cette expression est valable en inversion forte (lorsque le courant de diffusion peut être négligé) et revoie une valeur fidèle à Q_s tant que V_{th} reste représentatif de la hauteur de la barrière de potentiel au niveau de la source virtuelle. Nous noterons que dans [Lochtefeld01], l'utilisation d'une courbe $C_{gc}(V_{gs})$ effectuée sur un transistor long pour le calcul de Q_s plutôt que de la valeur C_{eff} en utilisant la formule du §IV.3.1 n'apporte à priori rien sauf une illustration visuelle (cf. figure V.9). Au contraire, l'utilisation de cette méthode peut introduire une erreur supplémentaire qui peut résulter de la fuite de grille.

En injectant la valeur issue de la résolution de type Poisson-Shrödinger multi sous-bandes détaillée dans [Assad99] ($\nu_{inj} = 1.7 \times 10^7$cm/s) les auteurs constatent que les transistors MOS sur silicium massif fonctionnent à 40% de leur limite balistique (longueur de grille : 50nm). Néanmoins, nous noterons que l'utilisation d'une valeur de ν_{inj} non expérimentale conduit à négliger l'influence de la rétrodiffusion sur la vitesse d'injection et répercute un erreur conséquente sur la valeur de r extraite. Les lacunes de cette techniques concernent donc les points suivants :

– la vitesse d'injection ν_{inj} est supposée être le mécanisme limitant le courant du transistor en régime de saturation (hypothèse d'un transport quasi-balistique),

– le taux de rétrodiffusion est extrait depuis une valeur de ν_{inj} calculée numériquement. L'autoconsistance du transport ($\nu_{inj} = f(r^{sat})$) ne peut pas être pris en compte et amène à extraire des valeurs de r^{sat} erronées [Palestri05],

– la correction des variations $V_{th}(L_{eff})$ et de l'influence de R_{sd} est basée sur une technique capacitive sensible à l'influence de la fuite et/ou de la désertion de grille sur la caractéristique $C_{gc}(V_{gs})$.

V.2.2.b La méthode d'extraction développée au CEA-LETI

La méthode d'extraction proposée par V. Barral et al. dans [Barral07, Barral09b, Barral09a] permet de prendre en compte la dégénérescence et plusieurs sous-bandes. Elle est particulièrement adaptée à l'extraction de BR sur des technologies à film mince, pour lesquelles la répartition des porteurs sur les différentes sous-bandes est très sensible à l'épaisseur du film de silicium T_{Si} [Barral09b]. Afin d'épargner à l'utilisateur de s'engager dans des calculs fastidieux (résolution numérique de Poisson-Shrödinger), la méthode est livrée avec plusieurs abaques qui fournissent des facteurs de correction pour la prise en compte de plusieurs sous-bandes et de la dégénérescence. La méthode offre l'avantage de fournir une voie d'extraction de r^{sat} qui ne repose pas sur le concept de distance critique (l_{kT}) qui peut être discutable [Palestri06, Clerc06]. Ainsi, des calculs relativement simples permettent d'aboutir à la relation (V.9) pour l'extraction de r^{sat} en connaissance de la charge d'inversion à la source virtuelle[2]. Une relation similaire que nous ne détaillerons pas permet l'extraction de r^{lin} en régime linéaire [Barral09b]. **Nous remarquons que, le fait d'avoir supposé dès le départ la validité du modèle quasi-balistique de Lundstrom conduit à extraire un taux de balistiticté non nul quelque soit la nature du transport (en particulier dans le cas d'un transistor dont le courant est limité par la vitesse de saturation comme dans [Sodini84]).**

$$\frac{1-r^{sat}}{(1+r^{sat})^{3/2}} = \frac{I_d}{W \cdot f_{sat} \cdot B \cdot {Q_s}^{3/2}}, \quad B \stackrel{\text{def}}{=} \frac{8\hbar}{3\sqrt{\pi q \cdot g_0 \cdot m_x^\star}} \tag{V.9}$$

Le paramètre f_{sat} (f_{lin} en régime linéaire) est ici un facteur de correction, déterminé par le biais des abaques, qui permet de prendre en compte la répartition des porteurs sur les différentes sous-bandes et de corriger l'approximation de forte dégénérescence utilisée dans le calcul. L'estimation de la charge à la source virtuelle Q_s et la correction des résistances séries sont reprises de [Lochtefeld01]. Par ailleurs, l'auteur préconise de ne pas utiliser cette méthode pour caractériser des dispositifs ayant un DIBL supérieur à environ 100mV/V, pour lesquels une mauvaise estimation de la charge peut considérablement fausser le résultat. Par conséquent cette méthode n'est pas adaptée aux dispositifs courts sur silicium massif qui sont caractérisés par une *mauvaise* intégrité électrostatique (i.e. des valeurs de DIBL qui vont au delà de 100mV/V). Cette méthode récente a été publiée pour la première fois en 2007 [Barral07] et est pour l'instant peu utilisée. Elle semble toutefois fournir des résultats pertinents lorsqu'elle est appliquée sur des dispositifs FDSOI dont le T_{Si} est inférieur à 10nm [Barral09a].

V.2.2.c La méthode d'extraction développée par l'Université de Taïwan

La méthode développée par le Département d'Ingénierie Électronique de Taïwan et publiée en 2002 dans [Chen02], constitue la première méthode d'extraction **entièrement expérimentale** du coefficient de r^{sat}. De construction assez simple, elle ne nécessite que la mesure du courant de drain en régime saturé, à deux températures proches. Jusqu'à maintenant, elle est la plus utilisée pour extraire r^{sat} sur des architectures variées :

- des transistors sur substrat massif [Chen02]
- des architectures contraintes par une couche CESL [Lin05c, Lin05b, Lin06]
- des architectures avec des source/drain en carbure de silicium (SiC) [Ang06, Ang07] ou en SiGe [Lin05a, Lin06],
- des orientations cristallines différentes [Tsai07, Chung07],

2. Le lecteur pourra se rapporter à [Barral09b, Barral09a] pour les détails de calculs menant à ce résultat.

– des largeurs variables [Liao08, Chang09],
– des transistors constitués d'un canal en silicium polycristallin [Suk09].

La méthode est construite autours du modèle des équations de transport quasi-balistique établi par Lundstrom [Lundstrom97], détaillées au §II.3.2. Le logarithme de l'expression analytique du courant en régime de saturation peut ainsi être dérivé par rapport à la température comme en (V.10).

$$\alpha \stackrel{\text{def}}{=} \frac{\partial}{\partial T}\ln\left(I^{on}\right) = \left[\frac{1}{2T} - \left(\frac{1}{1+r^{sat}} + \frac{1}{1-r^{sat}}\right)\frac{\partial r^{sat}}{\partial T} - \frac{\partial}{\partial T}\ln\left(Q_{inv}\right)\right] \tag{V.10}$$

En remplaçant r^{sat} par son expression donnée dans [Lundstrom97] $r^{sat} \stackrel{\text{def}}{=} \lambda_\mu/l_{kt}$, le rapport λ_μ/l_{kt} peut être exprimé en fonction de α (grandeur mesurable) comme en (V.12), où les termes γ_i représentent les dépendances en puissance de μ, ν_{inj} et l_{kT} en fonction de la température (ex : $\mu \propto T^{\gamma_\mu}$).

$$\left.\frac{\lambda_\mu}{l_{kT}}\right|_{\eta\ll 2} \simeq \frac{-2\left(1+\gamma_\mu-\gamma_{\nu_{inj}}-\gamma_{l_{kt}}+\right)}{\gamma_{\nu_{inj}}-(\alpha-\beta)\cdot T} - 2, \quad \beta \stackrel{\text{def}}{=} \frac{\partial}{\partial T}\ln\left(Q_{inv}\right) \tag{V.11}$$

Initialement, dans [Chen02], ces dépendances en températures sont supposées approximées à $\gamma_\mu = -1.5$, $\gamma_{\nu_{inj}} = 0.5$ et $\gamma_{l_{kT}} = 1$. La charge d'inversion est obtenue de manière classique grâce à la relation suivante : $Q_s = C_{eff}\cdot(V_{gs}-V_{th})$, permettant ainsi d'écrire $\beta = -\eta_{V_{th}}/V_{gt}$, où $\eta_{V_{th}}$ représente la dépendance en température de V_{th} : $\eta_{V_{th}} \stackrel{\text{def}}{=} \partial V_{th}/\partial T$. Lorsque ces hypothèses sont justifiées, la relation (V.11) prend une forme simple (V.12) qui permet d'extraire le rapport λ_μ/l_{kT} par une simple mesure de courant effectuée à deux températures.

$$\left.\frac{\lambda_\mu}{l_{kT}}\right|_{\eta\ll 2} \simeq \frac{4}{\frac{1}{2} - \left(\frac{\eta_T}{V_{gt}} + \alpha\cdot T\right)} - 2 \tag{V.12}$$

Ainsi, si la relation (V.11) reste relativement juste[3], le choix arbitraire des dépendances en température des différents paramètres (y compris la charge) semble peu fondé et constitue une source d'erreur importante dans le modèle.

Les limites de la méthode d'extraction – Si en 2002 cette méthode bénéficiait de *l'avantage de la nouveauté*, elle est basée sur plusieurs hypothèses et concepts non vérifiés (et infondés dans la plupart des cas). Il nous semble donc totalement déraisonnable d'espérer obtenir des résultats fiables en voulant l'appliquer comme dans [Chen02]. Une étude détaillée de la validité de la méthode à déjà été faite dans [Zilli07b] en s'appuyant sur des simulations Monte-Carlo. Il en ressort les défaillances suivantes :

– L'hypothèse de Lundstrom $\gamma_{l_{kT}} = 1$ n'est valable que si la bande de conduction évolue linéairement (dans la longueur) à proximité de la source virtuelle, indépendamment de la température. Les simulations MC montrent que cette hypothèse n'est pas valide et que $\gamma_{l_{kT}}$ prend des valeurs comprises entre 0.7 et 1.24 suivant la longueur du transistor. Notons également que la définition de la longueur critique (kT-layer) ne reste fondée que lorsque

3. elle repose sur la définition du libre parcours moyen de [Lundstrom97] qui peut être discutée. En effet, la définition du libre parcours moyen λ_μ est délicate puisqu'elle dépend de l'architecture et du champ transverse considéré, cf. [Palestri06].

l'équilibre thermodynamique est atteint. Sa validité peut donc être remise en question dans le cas d'un transport quasi-balistique, cf. [Clerc06].

– La dépendance en température de la mobilité dérive-diffusion $\gamma_\mu = -1.5$ dans [Chen02] est infondée. Elle semble être tirée de la dépendance en température établie pour la mobilité résultant des collisions avec les phonons acoustiques ($\gamma_{\mu ph}$ varie entre -1 et -1.75 dans la littérature [Takagi94b, Jeon89, Jacoboni77]) et néglige l'influence de la rugosité de surface, très présente à forts champs transverses. Cette valeur semble largement surestimée et est considérée comme un paramètre ajustable dans d'autres publications du même auteur [Chen04, Chen07].

– L'impossibilité de prendre en compte l'influence d'une dégénérescence modérée dans le calcul constitue une sérieuse limitation dans méthodologie d'extraction. En effet, en régime d'inversion forte le niveau de Fermi dépasse la bande de conduction (nMOS), invalidant l'approximation de Boltzmann dans les calculs sans pour autant permettre l'utilisation des formules établies pour le cas complètement dégénéré ($\eta_F \to \infty$) dans [Lin06]. L'approximation de Boltzmann conduit à une sous-estimation de la vitesse d'injection dans le calcul et maximise le taux de balisticité BR^{sat} (ceci explique éventuellement les valeurs élevées reportées dans [Chen02]).

– Finalement, selon [Zilli07b], l'approximation $Q_s = C_{eff} \cdot (V_{gs} - V_{th})$ ne permet pas d'extraire la dépendance en température de Q_s. Cette approximation constitue la source d'erreur la plus critique dans la méthodologie d'extraction. Nous noterons toutefois qu'un V_{th} à faible V_{ds} semble avoir été considéré dans [Zilli07b] alors que la tension de seuil en régime de saturation doit normalement être corrigée des effets de canaux courts (DIBL) comme dans [Lochtefeld01].

Une variante de la méthode d'extraction permettant de prendre en compte une dégénérescence intermédiaire a été publiée dans [Lin06, Chen06]. Malheureusement elle fait appel a des calculs complexes qui n'ont de solution analytique que dans le cas complètement dégénéré, non justifiée dans nos mesures. L'expression du rapport λ_μ / l_{kT} est alors donnée par (V.13).

$$\left.\frac{\lambda_\mu}{l_{kT}}\right|_{\eta\to\infty} \simeq -\left[\left(\frac{5}{T}+\frac{\eta_T}{V_{gt}}\right) / \left(\alpha+\frac{3}{2}\frac{\eta_T}{V_{gt}}\right)+2\right] \tag{V.13}$$

Pour résoudre le problème lié à la dégénérescence, il a été proposé d'utiliser des simulations numériques de type MC pour connaitre la valeur de ν_{inj} et de reprendre la méthode d'extraction décrite précédemment en considérant les paramètres γ_i comme des inconnues à déterminer [Chen04, Chen07]. Cette variante de la méthode donne des résultats plus cohérents avec les simulations [Chen04] mais perd l'avantage d'être totalement expérimentale puisque la simulation est nécessaire pour trouver les dépendances en température de ν_{inj} et Q_s. Notons également que cette technique ne permet pas de prendre en compte l'autoconsistance du transport ($\nu_{inj}(r^{sat})$) dans le calcul (cf. chapitre II). Nous pouvons nous attendre à ce que cette négligence répercute une erreur importante sur les résultats, en particulier sur les dispositifs sur silicium massif ayant un DIBL est élevé.

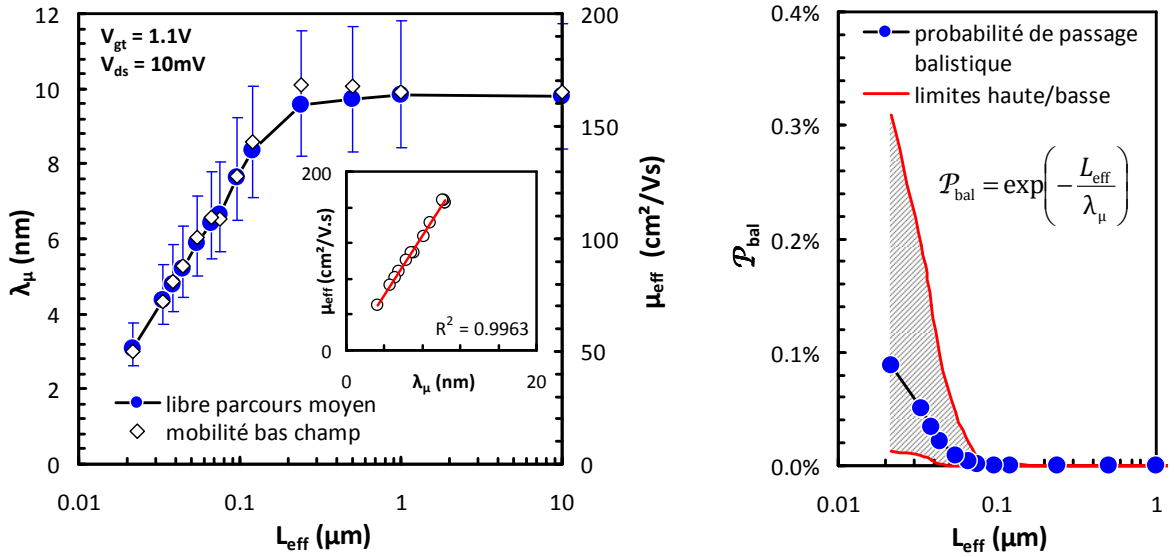


Fig. V.10 – A gauche : Evolution de la mobilité effective et du libre parcours moyen en fonction de L_{eff} pour $V_{gt} = 1$V. L'insertion montre une forte corrélation entre μ_{eff} et λ_μ ($R^2 \simeq 1$), permettant de justifier la cohérence de nos calculs. A droite : probabilité de passage balistique $\mathcal{P}_{bal}$ calculée grâce à la formule décrite dans [Leo94], en supposant $\nu_{inj} = (1.8 \pm 0.3) \times 10^7$cm/s.

V.2.3 Les points faibles des méthodes existantes

Mis à part l'extraction de ν_{eff} proposée dans [Lochtefeld01], toutes les méthodes d'extraction postulent que le modèle quasi-balistique est adapté aux dispositifs à analyser. Cette hypothèse peut être remise en question sur des dispositifs comme les nôtres pour lesquels $\lambda_\mu \ll L_{eff}$ (cf. figure V.10) ce qui ne permet pas aux porteurs de transiter sans interaction de la source au drain (nous rappelons que la probabilité de passage balistique est donnée par $\mathcal{P}_{bal} = \exp(-L_{eff}/\lambda_\mu)$ [Leo94]). La méthode du CEA-LETI qui semble la plus élaborée [Barral07, Barral09b] ne s'applique pas à des dispositifs sur silicium massifs (qui présente un DIBL relativement élevé et pour lesquels les abaques de corrections ne sont pas fournis). Il y a donc une lacune dans le panel de méthodologies d'extraction disponibles : aucune d'entre elles ne semblent capable d'extraire de manière fiable la vitesse limitant le transport en régime de saturation (ν_{inj} ou ν_{sat}) et, dans le cas d'un transport quasi-balistique, la valeur de BR^{sat} sur des architectures sur silicium massif. Ainsi, le prochain paragraphe présente la nouvelle méthodologie d'extraction qui a été développée durant cette thèse. Elle permet une analyse du mécanisme limitant I_{on} dans de tels dispositifs et ne requiert pas d'hypothèse sur la nature de ce dernier [Fleury09a].

Signification du taux de balisticité BR – si l'utilisation du taux de balisticité permet d'évaluer les performances d'une technologie, nous prendrons garde à la signification physique de cette variable. En effet, dans une approche où le mécanisme de transport est effectivement de type quasi-balistique, le taux de balisticité a une signification physique puisque qu'il est directement relié à la proportion de porteurs qui rejoint le drain depuis la source sans interaction dans le canal. Au contraire, lorsque l'hypothèse du transport balistique n'est pas vérifiée, le taux de balisticité quantifie le niveau de performances de la technologie en rapport aux performances idéales qu'elle aurait dans le cas balistique, sans lien avec le nombre de porteurs balistiques.

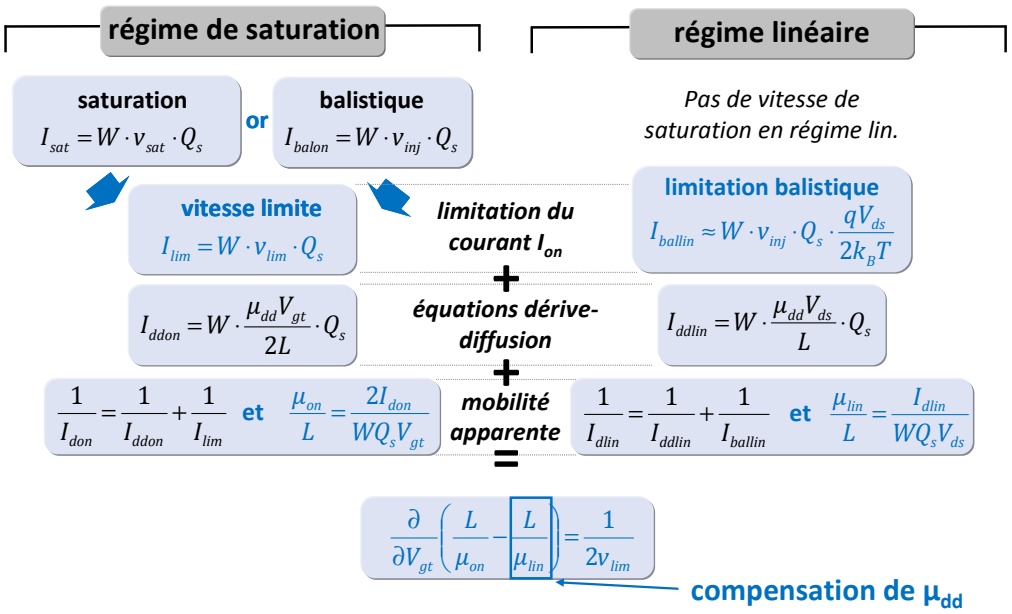


Fig. V.11 – Représentation schématique du calcul permettant d'aboutir à la relation (V.14) à partir des équations présentées au chapitre II.

V.2.4 Proposition d'une nouvelle méthodologie d'extraction

En réutilisant le concept de mobilité apparente et les relations introduites au §II.4.3, il est possible d'établir une fonction qui permet d'extraire la vitesse limitante (baptisée ν_{lim}) dans le transport, cf. (V.14). Cette relation offre l'avantage de permettre une extraction entièrement expérimentale, sans nécessiter de mesure de mobilité ou de L_{eff}. En utilisant astucieusement le concept de mobilité apparente [Pappas09], elle fournit une estimation de l'écart à la loi d'Ohm que présente la caractéristique d'un transistor, en couplant le régime linéaire et le régime de saturation dans une unique relation.

$$\nu_{lim} = \frac{1}{2} \cdot \left[\frac{\partial}{\partial V_{gt}} \left(\frac{L}{\mu_{app}^{sat}} - \frac{L}{\mu_{app}^{lin}} \right) \right]^{-1} \tag{V.14}$$

La construction de (V.14) à partir des équations du courant en régime linéaire et de saturation est schématisée par la figure V.11. Le régime de saturation est caractérisé par deux mécanismes de saturation de vitesse (ν_{sat} et ν_{lim}), qui sont regroupés dans une unique variable $\nu_{lim} = \min(\nu_{sat}, \nu_{inj})$. A l'inverse, le phénomène de vitesse de saturation n'est pas présent à champ faible, et seul ν_{inj} caractérise le régime linéaire. Le courant mesuré est décomposé fonction des contributions I_{bal}, I_{sat} et I_{dd} en utilisant les relations de Matthiessen dont la validité est démontrée au §II.4. Finalement les mobilités apparentes des régimes linéaires et de saturation sont calculées et utilisées de manière à éliminer le terme dérive-diffusion et à ne conserver que la partie du courant qui provoque la déviation par rapport à la loi d'Ohm.

V.2.4.a Protocole d'extraction de ν_{lim}

Comme pour les méthodologies d'extraction exposées précédemment, il est impératif de corriger la mesure de l'influence des résistances séries. La valeur de R_{sd} peut être mesurée par la technique $R_{tot}(1/\beta)$ [Fleury09b, Fleury09c] décrite au §III.6.2 et être corrigée en suivant les relations données au §III.2.2. même si nos transistors MOS ont des effets de canaux courts intenses

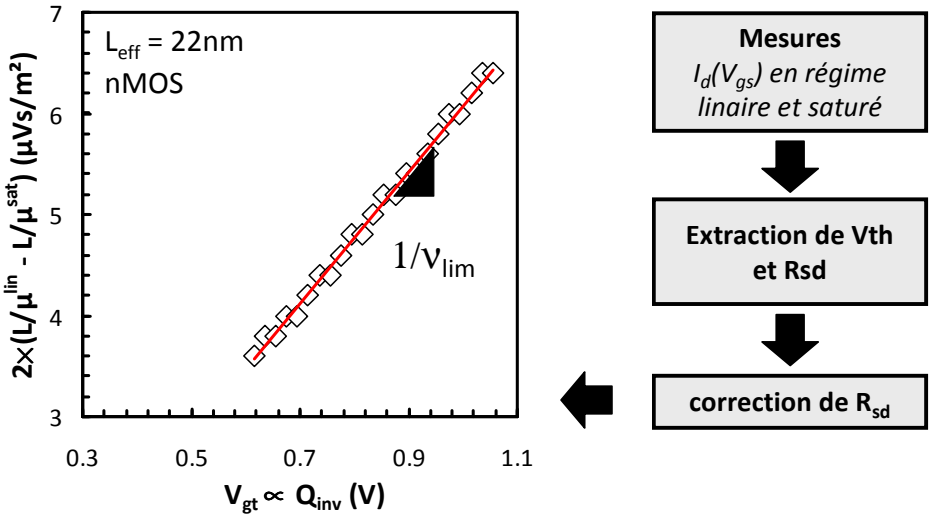


Fig. V.12 – Protocole d'extraction permettant la mesure de ν_{lim}.

et ne correspondent pas exactement à la définition *well tempered* de Lochtefeld [Lochtefeld01], la charge reste principalement contrôlée par la grille et varie peu avec la tension de drain V_{ds} [Lundstrom02a, Ren03]. Elle peut donc être approximée par la relation (V.15) où l'influence de V_{ds} est prise en compte dans V_{th} qui est supposée reproduire fidèlement les variations de Q_s : $\partial Q_s = C_{eff} \cdot \partial V_{th}$. Soulignons que la notation $V_{th}(V_{ds})$ est utilisée dans la relation (V.15) pour mettre en évidence la prise en compte du DIBL dans l'extraction.

$$Q_s(V_{ds}) \simeq C_{eff} \cdot (V_{gs} - V_{th}(V_{ds})) \tag{V.15}$$

Le protocole à suivre pour réaliser l'extraction est schématisé sur la figure V.12. Il comporte les étapes suivantes :

1. mesure des courbes $I_d(V_{gs})$ en régime linéaire ($V_{ds} = 10$mV) et de saturation ($V_{ds} = 1.1$V),
2. extraction de R_{sd} et de V_{th}, correction de l'influence de R_{sd},
3. calcul des mobilités apparentes et extraction de ν_{lim} en forte inversion en utilisant la relation (V.14).

V.2.4.b Les points forts de la nouvelle méthode

Par rapport aux autres méthodes d'extraction, les avantages dans l'utilisation de notre nouvelle méthode sont multiples :

- elle permet d'extraire ν_{lim} sans faire d'hypothèse sur la nature du phénomène limitant, qui provoque la saturation du courant I_{on},
- la correction de la mobilité μ_{dd} s'effectue implicitement grâce au concept de mobilité apparente du §II.4.3, en couplant à la fois des mesures à fort et faible V_{ds}. L'application de la méthode ne requiert pas la connaissance de la mobilité, ni le calcul numérique de ν_{inj} comme dans [Lochtefeld01, Pappas09],
- l'autoconsistance est implicitement prise en compte dans la vitesse extraite. Dans le l'hypothèse d'un transport quasi-balistique, la valeur de r^{sat} peut être obtenue en injectant la valeur de ν_{inj} mesurée dans l'extraction de [Lochtefeld01], résolvant ainsi le problème d'autoconsistance.

– La technique repose sur des mesures et des extractions simples ($I_d(V_{gs})$, R_{sd}, V_{th}) qui peuvent être implémentées de manière automatique pour obtenir des résultats statistiques.

V.2.5 Protocole expérimental suivi dans la thèse

L'étude du transport qui sera détaillée dans les deux prochaines parties a été effectuée sur des transistors nMOS issus de la technologie CMOS 45nm de STMicroelectronics (épaisseur physique d'oxyde : $t_{ox} \simeq 17.5$Å). Une plaque a été choisie pour son décentrement vis-à-vis du procédé de fabrication qui lui procure des longueurs de grille plus faible qu'à la normale, sur lesquelles les phénomènes de balisticité peuvent être supposés plus influents et mieux détectables. Nous choisirons ainsi de travailler sur des structures de test de 1μm de large (pour les raisons indiquées au chapitre III) et de longueur de masque variable allant de 35nm (longueur sub-nominale) jusqu'à 10μm.

Toutes les mesures ont été faites sur 65 puces afin d'augmenter la précision sur les résultats, pour plusieurs températures allant de 230K à 410K. Les mesures de courant $I_d(V_{gs})$ en régime saturé sont effectuées pour $V_{ds} = 1.1$V. Les mesures en régime linéaire sont effectuées à $V_{ds} = 10$mV pour les raisons suivantes :

– vérifier la condition $V_{ds} \leq 2kT/q$ pour toutes les températures et permettre l'utilisation du modèle simplifié présenté en §II.3.2,
– pouvoir corriger la fuite de grille comme au §III.2.1, en supposant le partage de grille symétrique entre la source et le drain,
– offrir une précision correcte vis-à-vis des spécifications de l'équipement de mesure. En particulier nous n'adopterons pas le choix $V_{ds} = 1$mV fait dans [Barral07, Barral09a], qui selon nous peut conduire à de larges incertitudes sur les résultats (car l'équipement induit une incertitude sur V_{ds}).
– de minimiser le champ dans les transistors courts ($\max(\mathcal{E}_y) = 5 \times 10^3$V/cm) et de pouvoir négliger la déviation de la loi d'Ohm causée par des collisions de nature inélastique, cf. [Sodini84].

V.2.5.a La température : un levier essentiel pour identifier les propriétés de transport

Comme expliqué dans [Fleury09a, Pappas09] et illustré sur la figure V.13, les deux mécanismes susceptibles de corrompre la loi d'Ohm $I_{on} \propto 1/L_{eff}$ sur les petits dispositifs se manifestent de la même manière. A 300K les valeurs de ν_{sat} et ν_{inj} sont proches et seule la température semble pouvoir donner l'information essentielle qui permettra d'identifier le mécanisme influent. En effet, le phénomène de saturation est une conséquence des collisions avec les phonons optiques et décroit lorsque T augmente, tandis que la vitesse d'injection résulte de l'énergie thermique dégénérée et augmente avec T[4].

L'évolution de ν_{inj} et ν_{sat} [Jacoboni77, Zilli07b] est donnée en fonction de la température sur la figure V.14 pour illustrer notre propos.

4. Dans le cas complètement dégénéré, ν_{inj} tend vers la vitesse de Fermi ν_F qui ne dépend plus de la température [Lundstrom02a, Takagi08]

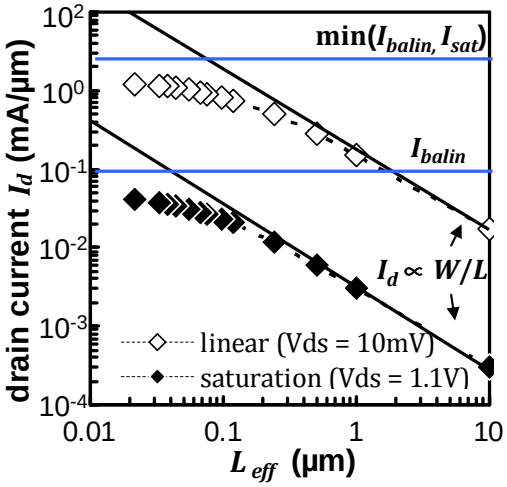


Fig. V.13 – Il est difficile d'identifier le mécanisme limitant car ν_{inj} ou ν_{sat} ont des valeurs proches. Ces mesures ne sont pas corrigées de la valeur des résistances séries qui contribue à la saturation des courants dans les deux régimes, à fort V_{gs}.

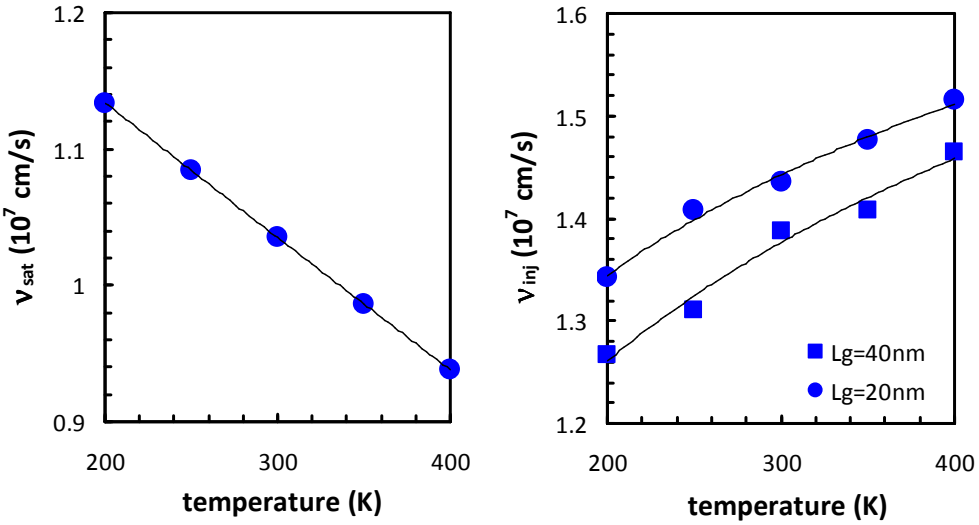


Fig. V.14 – Evolution de ν_{sat} et de ν_{inj} en fonction de la température dans la gamme (200K – 400K). Les dépendances de ν_{sat} et ν_{inj} sont opposées et peuvent permettre l'identification du mécanisme responsable de la saturation de la vitesse. Remarque : les valeurs de ν_{sat} sont issues de [Jacoboni77] et mesurées dans le silicium massif, celles de ν_{inj} sont issues de simulations Monte-Carlo [Zilli07b].

V.2.5.b Premiers résultats : des observations et des modèles contradictoires...

La partie gauche de la figure V.15 présente l'évolution du courant I_{on} (corrigé des résistances séries) en fonction de la longueur effective mesurée suivant la méthode décrite au §IV.4, pour une température de 233K et 413K. La représentation en échelle log.–log. des valeurs permet d'évaluer facilement le niveau du courant I_{lim} responsable de la saturation du courant (qui provoque la déviation à la loi d'Ohm). L'évolution du niveau de courant, représentatif de ν_{lim} attire notre attention car il présente une contradiction avec l'hypothèse de la quasi-balisticité qui prévoit un seuil à la saturation qui suit les variations $\nu_{inj}(T)$ (i.e. proportionnel à T^{α} avec $\alpha \leq 1$). Au contraire le comportement de $I_{lim}(T)$ semble indiquer la présence d'un mécanisme de collision (la vitesse de saturation par exemple) et montre une forte similitude avec les observations de [Jacoboni77].

La partie droite de la figure V.15 présente l'évolution du rapport des mobilités apparentes entre 233K et 410K en fonction L_{eff}. Ce rapport décroit et sature pour les faibles longueurs de grille, en accord avec les observations et les simulations dans [Cros06b, Ghibaudo09]. Par ailleurs, le rapport des μ_{app} n'est jamais inférieur à 1 et semble également en contradiction avec l'hypothèse d'un transport de type quasi-balistique.

Les deux graphiques de la figure V.15 sont un point de départ pour l'analyse poussée que nous allons entreprendre. En effet, si les méthodes d'extraction telles que celle développée par l'université de Taïwan retournent des taux de balisticité élevés, l'observation la plus basique semble déjà relever une incohérence entre les mesures et la théorie quasi-balistique de Lundstrom. De plus, les méthodes d'extraction existantes ne permettent pas de déceler le mécanisme responsable de la saturation du courant et la validité des résultats qu'elles retournent peut être remise en cause.

Dans les prochaines parties, nous analyserons des propriétés de transport en régime linéaire et saturé, en utilisant les outils détaillés dans le chapitre III et la nouvelle méthode d'extraction de ν_{lim} présentée au §V.2.4. Nous tenterons de comprendre l'origine du mécanisme de saturation de I_{on} et, par la même occasion, de confirmer ou d'infirmer le caractère quasi-balistique du transport dans nos dispositifs.

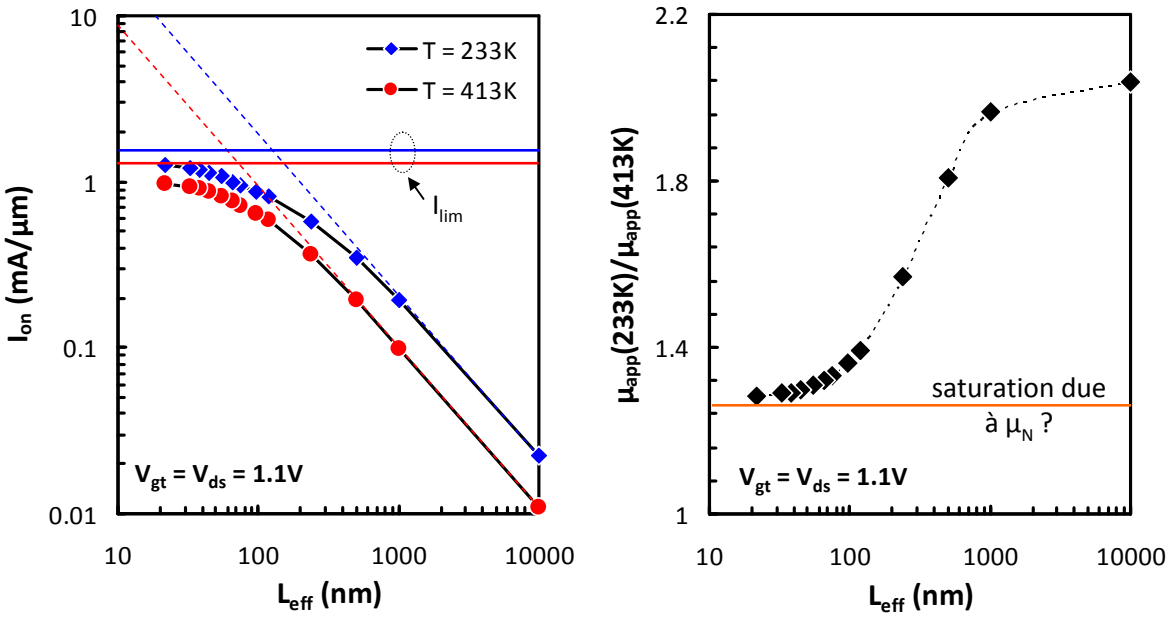


Fig. V.15 – A gauche : évolution du courant I_{on} en fonction de L_{eff}. La saturation de courant semble avoir une dépendance en température en contradiction avec l'hypothèse de la quasi-balisticité. A droite : évolution du rapport des mobilités apparentes à 233K et 413K en fonction de la longueur de grille. La saturation pour les dispositifs courts semble en accord avec l'influence croissante d'une composante de la mobilité indépendante de la température, comme observée dans [Cros06b].

V.3 Évaluation des propriétés de transport en régime de saturation

Comme nous l'avons vu au §II.2.3, le phénomène de saturation de la vitesse des porteurs est responsable d'une déviation à la loi d'Ohm, i.e. d'un courant plus faible que ce que retourne l'extrapolation à partir des valeurs de la mobilité à bas champ. Ce phénomène de saturation est particulièrement visible lorsque les courants normalisés $I_{on}(V_{gt}) \times L_{eff}/V_{gt}$ et $I_{lin}(V_{gt}) \times L_{eff}/V_{ds}$ (corrigés de l'influence de R_{sd}) sont représentés en fonction de la mobilité effective $\mu_{eff}(V_{gt})$ (mesurées en régime linéaire $V_{ds} = 10$mV). La figure V.16 montre ainsi une forte corrélation courant-mobilité pour les deux régimes qui semble en accord avec les conclusions de l'étude de transport en régime linéaire : les collisions ont une grande influence sur le courant. Néanmoins, la valeur de la pente dans le cas du régime de saturation est inférieure à 1, traduisant une perte d'influence de la mobilité sur le courant I_{on}. Cette tendance est de plus confirmée par l'insertion de la figure qui montre que le courant I_{on} tend à être de moins en moins dépendant de I_{lin} à mesure que la longueur de grille est faible : nous sommes bien en présence d'un phénomène de saturation caractérisé par une déviation de la loi d'Ohm.

Si la compréhension des mécanismes de collisions semble solidifiée par l'analyse de la partie §V.2 en régime ohmique, l'origine physique du phénomène de saturation reste inconnue (ν_{inj} ou ν_{sat} ?). Pour comprendre la physique qui régit le comportement en courant du régime de saturation il convient donc de distinguer la nature de ν_{lim}, que nous proposons d'entreprendre en deux étapes :

- quantifier la valeur de ν_{lim},
- analyser la dépendance en température $\nu_{lim}(T)$ pour tenter d'identifier le mécanisme physique sous-jacents [Lundstrom02a, Fleury09a].

Ces points feront l'objet des paragraphes suivants.

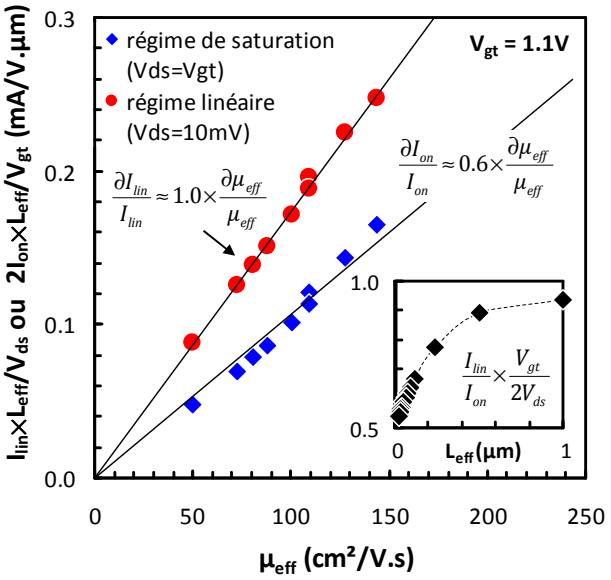


Fig. V.16 – Influence de la mobilité sur le courant en régime linéaire et de saturation. La perte d'influence en régime de saturation est caractéristique d'un phénomène de saturation de vitesse. En suivant l'approche de Lundstrom [Lundstrom01], il est possible d'extraire un taux de balisticité d'environ 40% sur ces dispositifs, en accord avec les résultats de la littérature.

V.3.1 Extraction de la vitesse effective

Quelque soit le mécanisme de transport mis en jeu dans le canal, il est toujours possible de mesurer la vitesse effective des porteurs (ν_{eff}) en utilisant les méthodes décrites dans [Sodini84, Lochtefeld01]. La figure V.17 montre l'évolution $\nu_{eff}(L_{eff})$ qui présente une saturation dans les deux régimes. Cette saturation peut s'expliquer par plusieurs mécanismes[5] :

- la dégradation de la mobilité dans les canaux courts,
- la vitesse de saturation ν_{sat} qui résulte des interactions entre les électrons et les phonons optiques dans le canal à forts champs électriques $\mathcal{E}_y$ (uniquement en régime de saturation),
- la vitesse d'injection à la source qui peut limiter le flux de porteurs lorsque la mobilité n'est plus limitante.

Le problème de dégradation de mobilité soulevé dans la partie §V.2, révèle une limitation dans l'analyse de ν_{eff} : elle ne permet pas de décorréler l'influence de μ_{eff} de celle d'une saturation de la vitesse provoquée par ν_{inj} ou ν_{sat}. Dans les prochaines parties, nous préfèrerons donc utiliser la méthode développée durant cette thèse (cf. §V.1.4) qui permet d'extraire directement ν_{lim}.

V.3.1.a Extraction du taux de balisticité et du paramètre de rétrodiffusion

En supposant que le modèle quasi-balistique s'applique à nos dispositifs ($\nu_{lim} = \nu_{inj}$), il est toujours possible d'extraire un taux de balisticité et, par conséquent, la valeur du coefficient de rétrodiffusion. Nous utilisons donc les méthodes décrites dans [Lochtefeld01, Chen02] pour extraire r et BR dans les deux régimes lorsque cela est possible. La valeur de ν_{inj}, nécessaire pour l'extraction de Lochtefeld, est estimée de manière numérique environ 2×10^7cm/s lorsque $V_{gt} = 1.1$V.

Les résultats sont présentés sur la figure V.18. Ils permettent d'analyser et de comparer les

5. les résultats sont déjà corrigés de l'influence de R_{sd}

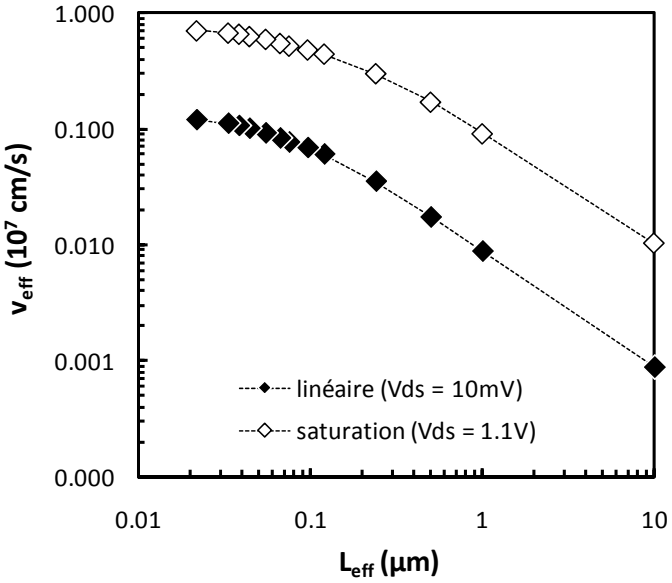


Fig. V.17 – Extraction de la vitesse de transport effective ν_{eff}en régime linéaire et saturé, d'après la méthode de [Lochtefeld01].

résultats retournés par les deux méthodes d'extraction. En premier lieu, nos valeurs (toutes méthodes confondues) sont cohérentes avec les résultats de la littérature [REF] ($BR^{lin} \simeq 5\%$ et $BR^{sat} \sim 40\%$). Nos dispositifs ne semblent donc pas avoir un comportement anormal. Néanmoins, La méthode de [Chen02] semble fournir des valeurs de BR largement surestimées, en particulier pour les dispositifs longs pour lesquels la condition aux limites $BR^{sat} \rightarrow 0$ n'est pas respectée. Ce résultat est en accord avec les observations dans [Zilli07b] et ne nous surprend donc pas. La surestimation du paramètre BR^{sat} sur les transistors courts peut en effet provenir de l'absence de prise en compte de la dégénérescence dans le modèle. Elle est ici indispensable compte tenu de la forte densité de charges en inversion forte $Q_s \sim 10^{12}/\text{cm}^3$.

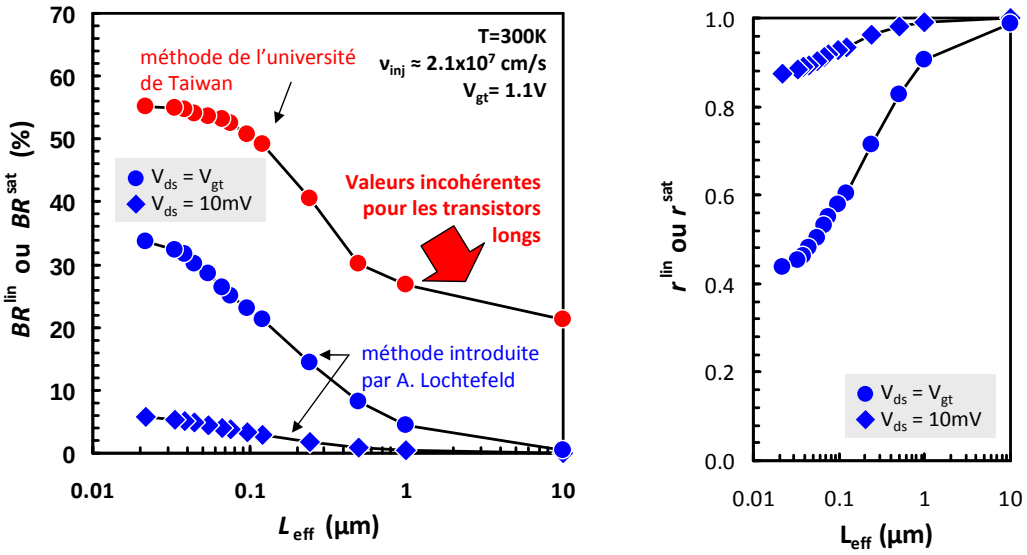


Fig. V.18 – Extraction de la vitesse de transport effective ν_{eff} en régime linéaire et saturé, d'après la méthode de [Lochtefeld01].

V.3.2 Extraction de ν_{lim} et détermination de sa nature

Si les méthodes d'extraction à l'état de l'art permettent d'extraire un taux de balisticité, la détermination de la nature de la vitesse limitante reste impossible et le résultat peu fiable. Nous proposons donc d'utiliser la méthode développée au cours de cette thèse [Fleury09a], qui permet une détermination directe de la vitesse limitant le transport ν_{lim}.

La figure V.19 présente les profils $\nu_{lim}(L_{eff})$ obtenus, pour différentes températures. Deux phénomènes sont clairement identifiables :

- ν_{lim} augmente pour les faibles longueurs de grille (d'environ 40%),
- ν_{lim} augmente lorsque la température diminue.

Les paragraphes suivants vont tenter d'expliquer ce comportement qui semble contradictoire avec l'hypothèse d'un transport quasi-balistique.

V.3.2.a Comportement en fonction de la température : détermination de la nature de ν_{lim}

Le comportement $\nu_{lim}(T)$ est représenté sur la figure V.20, pour deux longueurs de grille, et superposé aux dépendances en température de $\nu_{inj}(T)$ et $\nu_{sat}(T)$ reportées dans la littérature [Jacoboni77, Zilli07b] (une dépendance $\nu_{inj}(T)$ similaire est observée dans [Barral09a])[6]. La dépendance $\nu_{lim}(T)$ observée est surprenante car elle s'oppose à l'hypothèse d'une limitation par ν_{inj} et montre une grande similitude avec la vitesse de saturation, qui pour rappel, traduit l'influence des collisions inélastiques. Pour l'instant, si nous ne pouvons pas conclure objectivement car la dépendance en L_{eff} n'est toujours pas expliquée, **nous pouvons affirmer que le mécanisme dominant de transport ne peut pas être quasi-balistique (et donc limité par ν_{inj}) puisqu'il semble limité par une vitesse qui ne présente par les dépendances en température caractéristiques de ν_{inj}.**

La dépendance $\nu_{lim}(L_{eff})$ est complexe à analyser car plusieurs paramètres dépendant de

6. Précisons que la dépendance $\nu_{sat}(T)$ à été volontairement décalée dans le cas des transistors courts malgré que la dépendance $\nu_{sat}(L_{eff})$ qui ne soit originellement pas expliquée dans [Jacoboni77].

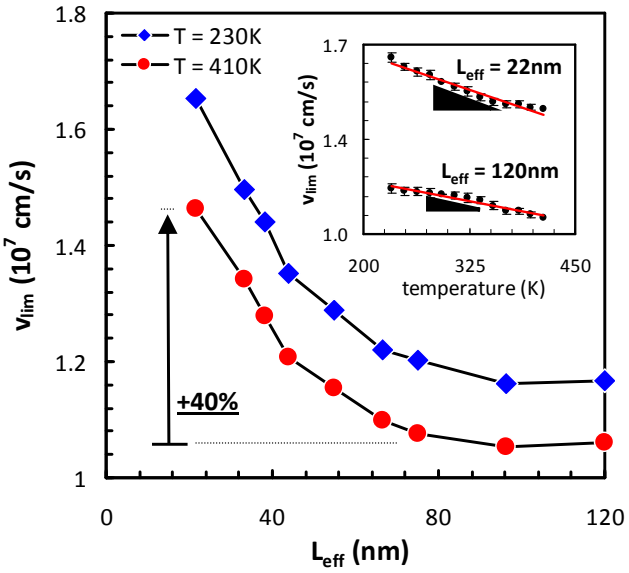


Fig. V.19 – La vitesse limite des porteurs à la source montre une augmentation de plus de 40% sur les dispositifs courts (en comparaison des longs). La dépendance en température, signature du mécanisme physique sous-jacent, est sensiblement la même quelque soit L_{eff}.

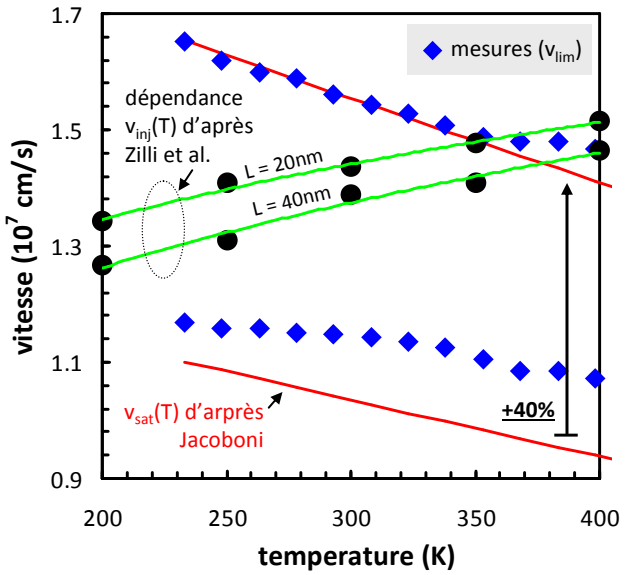


Fig. V.20 – Extraction de la vitesse de transport effective ν_{eff} en régime linéaire et saturé, d'après la méthode de [Lochtefeld01].

L_{eff} peuvent influencer ν_{lim}. Nous noterons entre autre :
- les contraintes mécaniques [Payet08],
- les phénomènes de survitesse [Ge01, Roldan97]

Ces deux mécanismes peuvent expliquer l'augmentation $\nu_{lim}(L_{eff})$ mais doivent rester compatibles avec la dépendance en température qui tend à assimiler ν_{lim} à ν_{sat}. Ce point nous permet d'abandonner la première hypothèse (contraintes mécaniques) qui ne permet pas d'expliquer une augmentation de ν_{sat}. Il est donc impossible de poser directement $\nu_{lim} = \nu_{sat}$ et d'attribuer la variation $\nu_{sat}(L_{eff})$ à l'influence des contraintes mécaniques.

Influence des contraintes mécaniques sur ν_{sat} : nos dispositifs sont contraints mécaniquement par l'utilisation d'une couche tensive CESL (*Contact Etch Stop Layer*) et d'une contrainte SMT (*Stress Memory Technique*) qui pourrait éventuellement influencer ν_{lim} et expliquer le comportement $\nu_{lim}(L_{eff})$. Si l'influence de la contrainte mécanique sur ν_{inj} par l'intermédiaire de $m^\star$ est bien comprise, l'influence sur ν_{sat} est faible [Bufler97, Bufler02, Fossum03] et ne permet pas d'expliquer la variation $\nu_{lim}(L_{eff})$ observée. En effet une variation de l'ordre de 10% est reportée entre un cas contraint et non contraint, valeur qui reste faible devant la variation de 40% observée dans notre étude entre les transistors longs et courts.

V.3.2.b Le phénomène de survitesse

Nous avons déjà détaillé le phénomène de survitesse au paragraphe §II.3.1. Pour rappel, ce dernier provient de l'inégalité des temps de relaxation en moment et en énergie ($\tau_w \neq \tau_m$). Lorsque le champ électrique $\mathcal{E}_y$ évolue rapidement, les porteurs peuvent ainsi acquérir une vitesse supérieure à la vitesse de saturation caractéristique de l'état d'équilibre. Les phénomènes de survitesse sont particulièrement présents au niveau du drain, où les variations $\partial\mathcal{E}_y/\partial y$ sont les plus fortes. Néanmoins, une augmentation de la vitesse des porteurs au drain provoque juste une redistribution de la charge dans le canal et a une faible influence au niveau de la source dans des dispositifs *well tempered* [Lochtefeld01].

Les faibles longueurs de grille de nos dispositifs permettent d'atteindre des gradients de champ susceptibles de permettre une survitesse *efficace* à proximité de la source et qui influe sur le courant. Il est ainsi possible d'observer des vitesses effectives de saturation ($\nu^\star_{sat}$)[7] de l'ordre de 2×10^7cm/s [Bude00] pour des dispositifs de 40nm. Ce résultat semble en bon accord avec nos résultats.

L'influence des phénomènes de survitesse à proximité de la source sur la vitesse des porteurs à la source virtuelle reste liée à l'intégrité électrostatique du dispositif. Le DIBL quantifie ainsi la capacité qu'un événement se produisant en un point y du canal, d'influencer l'électrostatique au niveau de la source. Une corrélation peut d'ailleurs être mise en évidence en traçant ν_{lim} en fonction du DIBL comme sur la figure V.21.

7. Dans la suite nous utiliserons le terme "vitesse de saturation effective" (comme dans [Ge01]), notée $\nu^\star_{sat}$, afin de prévenir toute confusion avec la vitesse de saturation à l'équilibre dans le silicium massif que nous continuerons à noter ν_{sat}.

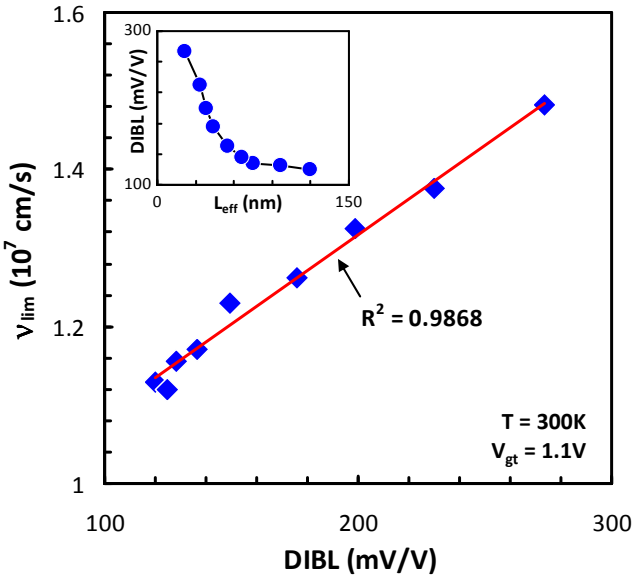


Fig. V.21 – Corrélation entre l'évolution de ν_{lim} et la valeur du DIBL (intégrité électrostatique).

V.3.3 Analyse de nos résultats en s'appuyant sur un modèle de survitesse

Réutilisons les résultats obtenus pour le transistor le plus court de la série testée ($L_{eff} \simeq$ 22nm). L'étude de la mobilité effective du §V.2 révèle que la présence de nombreuses collisions élastiques qui *interdisent* la possibilité d'un transport balistique car $\lambda_\mu \ll L_{eff}$. En effet, à température ambiante, λ_μ est de l'ordre de 3-4nm[8] et la probabilité de passage sans interaction $\mathcal{P}_{bal}$ est inférieure à 0.3%. **La balisticité joue donc un rôle infime dans le transport de nos transistors**.

Paradoxalement, les collisions inélastiques du canal proviennent majoritairement des échanges d'énergie avec les phonons optiques, dont le libre parcours moyen est du même ordre de grandeur que L_{eff} : $\lambda_{opt} \simeq$ 8nm (cf. §II.2.3). Il semble donc que toutes les conditions soient réunies pour observer un transport fortement non stationnaire : $\tau_w \neq \tau_m$ et $\lambda_{opt} \sim L_{eff}$. En nous appuyant sur le modèle de survitesse développé dans [Ge01], nous allons montrer que le phénomène de survitesse localisé à la source est possible et joue un rôle fondamental sur la valeur du courant I_{on} des transistors courts ($L_{eff} \leq$ 100nm). En utilisant l'équation de Boltzmann, Fossum et al. aboutissent à une relation physique qui relie l'augmentation de la vitesse de saturation apparente (i.e. $\nu^\star_{sat}$) en fonction du champ électrique, de son gradient et du temps de relaxation en énergie τ_w, cf (V.16).

$$\nu^\star_{sat}(y) = \nu_{sat} \cdot \left(1 + \frac{\nu_{sat} \cdot \tau_w}{\mathcal{E}_y} \cdot \frac{\partial \mathcal{E}_y}{\partial y}\right) \tag{V.16}$$

En utilisant l'expression analytique de $\psi_s(y)$ développée dans l'annexe C et valable pour le régime dérive-diffusion en forte inversion, il est possible d'aboutir à une expression simple de la vitesse de saturation apparente à l'entrée du canal ($\nu^\star_{sat}(0)$) : (V.17).

$$\nu^\star_{sat}(0) = \nu_{sat} \cdot \left(1 + \frac{\nu_{sat} \cdot \tau_w}{3L_{eff}}\right) \tag{V.17}$$

8. En fonction de la dégénérescence, nous supposons que la vitesse d'injection thermique peut varier entre 1.5×10^5 et 2.1×10^5cm/s, en accord avec [REF]

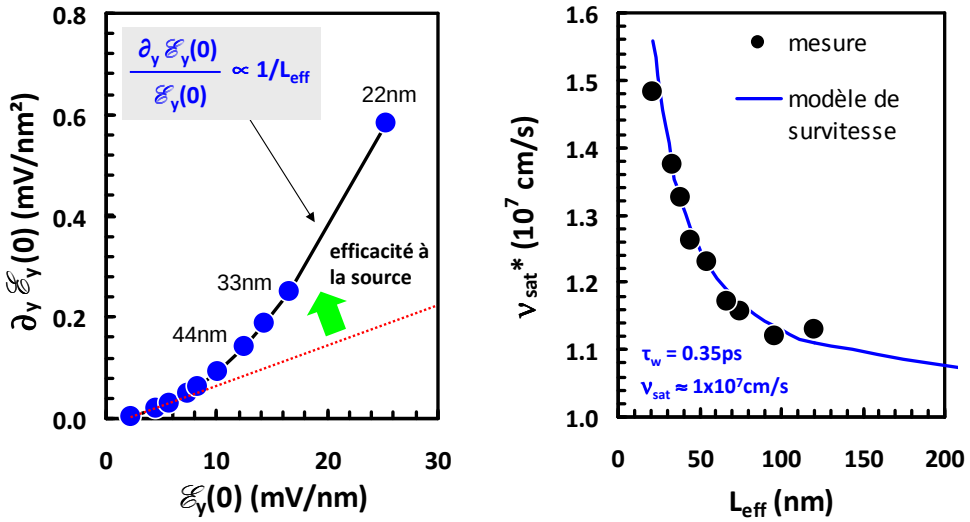


Fig. V.22 – A gauche : évolution du gradient du champ en fonction du champ au niveau de la source. L'évolution du potentiel dans les canaux courts permet au phénomène de survitesse d'être efficace dès la source. A droite : utilisation du modèle de survitesse introduit dans [Ge01] pour reproduire la tendance expérimentale $\nu_{lim}(L_{eff})$.

La partie gauche de la figure V.22 montre l'évolution de $\partial_y \mathcal{E}_y$[9] en fonction de $\mathcal{E}_y$ pris au niveau de la source $y = 0$ pour les différentes longueurs testées dans cette étude. A mesure que L_{eff} devient petit, le rapport du gradient du champ sur le champ augmente et favorise l'efficacité de la survitesse à la source.

En calculant la vitesse de saturation effective grâce à la relation (V.17), il est possible de reproduire les variations $\nu_{lim}(L_{eff})$. Le graphique de droite de la figure V.22 présente le résultat pour $T = 300$K, en choisissant τ_w et ν_{sat} comme des paramètres d'ajustement ($\tau_w \simeq 0.4$ps est reporté dans [Jacoboni77] à cette température).

En répétant l'opération pour plusieurs températures il est possible d'extraire la dépendance $\nu_{sat}(T)$ et de vérifier son bon accord avec résultats publiés dans [Jacoboni77, Duh67]. La partie gauche de la figure V.23 présente ainsi la dépendance obtenue pour $\tau_w = 0.35ps$ qui est gardé constant en fonction de T (ce qui semble en accord avec les explications de [Jacoboni77] dans notre gamme de température). Pour chaque température, l'extraction de ν_{sat} est effectuée, puis comparé au modèle de [Jacoboni77] (équivalent aux mesures de [Duh67]). La dépendance en température est en bon accord avec le modèle à $\pm 5\%$ et permet de valider que ν_{lim} résulte ici d'un phénomène qui s'apparente à la vitesse de saturation historique à laquelle s'ajoute une non-localité du transport (car $\lambda_{opt} \sim L_{eff}$).

9. Nous utilisons les notations réduites pour les dérivées partielles : $\partial_x \stackrel{\text{def}}{=} \partial/\partial x$

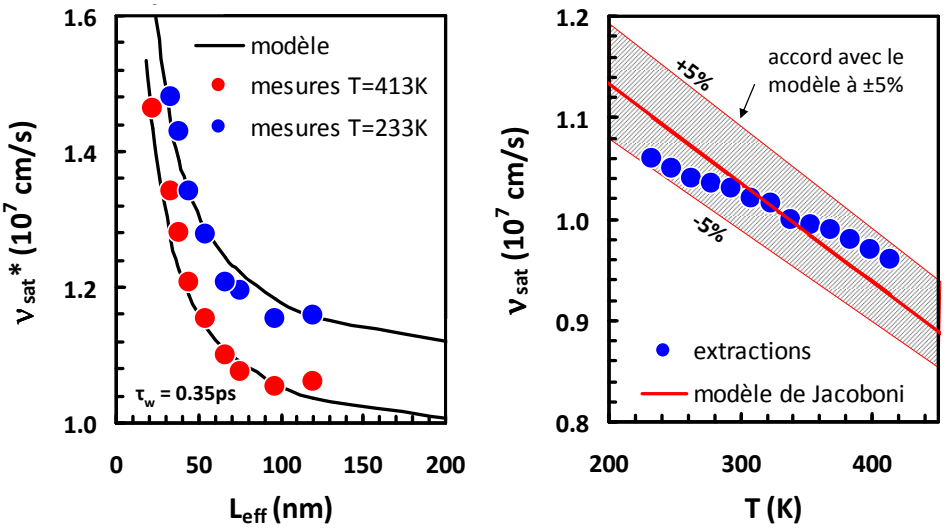


Fig. V.23 – A gauche : le modèle de survitesse est ajusté sur les points expérimentaux $\nu_{lim}(L_{eff})$ pour les deux températures extrêmes de cette étude. Un bon accord est trouvé lorsque le temps de relaxation en énergie est fixé à $\tau_w = 0.35$ps. A droite : évolution de la vitesse de saturation stationnaire (extrapolée depuis le modèle de survitesse) en fonction de la température. Les points extraits présentent un bon accord (valeurs et pente) avec la dépendance en température trouvée dans [Jacoboni77], compte tenu des incertitudes sur nos valeurs.

V.4 Bilan et discussions

Ces résultats nous ont permis d'investiguer et de mieux comprendre les mécanismes de transport dans le transistor MOS sur silicium massif. En particulier, la dégradation de mobilité observée sur les transistors courts bouleverse la plupart des théories de transport qui pouvaient paraître pertinentes pour expliquer les caractéristiques courant tension. La faible valeur de λ_μ ne permet pas d'observer un transport balistique. Cette constatation invalide les notions de longueur critique (kT-layer) [Lundstrom97] et de rétrodiffusion présentes dans la théorie du transport quasi-balistique et justifie le caractère *dérive-diffusion* du transport.

Les forts champs électriques qui règnent à proximité de la source ainsi que la disparité entre les temps de relaxation en énergie et en moment permettent l'apparition de phénomènes non stationnaires à proximité de la source. La notion de vitesse de saturation telle que décrite dans [Jacoboni77, Sodini84] perd alors tout son sens dans un transistor court. La théorie de la survitesse permet de retrouver une explication physique à nos résultats $\nu_{lim}(L_{eff}, T)$. Elle montre que le rapport entre le gradient de champ et le champ lui-même est si fort sur les petits dispositifs, que la survitesse due à la non-localité du transport au niveau de la source permet d'expliquer les augmentations de 40% mesurées.

Le transport dans les transistors courts semble donc limité par une quasi-saturation du courant plutôt que par une quasi-balisticité, dans le sens ou les collisions ont toujours un rôle majeur dans le transport et ou l'observation d'un passage balistique (sans interaction) est fortement improbable. **Les collisions inélastiques avec les phonons optiques constituent sûrement le mécanisme principal qui, additionné de la non localité du transport, permet d'expliquer le comportement des transistors modernes sur silicium massif dont la longueur de grille n'est pas inférieure à 20nm.**

Conclusion et perspectives

"Nous n'avons recours à ce qu'on appelle réflexion
que parce que la connaissance nous échappe."
Platon
Le Banquet

Notre étude fournit quelques éléments de réponse pour mieux comprendre la physique complexe qui régie le transistor MOS moderne. Nous avançons sur un terrain inconnu mais des axes de recherches prometteurs ont pu être dégagés au cours des cinq derniers chapitres.

L'extraction des paramètres électriques sur des transistors courts devient de plus en plus complexe. Elle doit être basée sur un modèle semi-empirique qui reproduit fidèlement le comportement électrique des transistors et à partir duquel il est possible d'analyser la physique du transistor.
La plupart des méthodes sont basées sur le modèle en inversion forte (type BSIM3v3) et les limites d'une telle approche commencent à devenir perceptibles : il est de plus en plus difficile de décorréler l'influence des résistances séries de celle de la dégradation de mobilité. Ce problème vient en partie du paramétrage en longueur, utilisé pour extraire R_{sd}, qui devient peu précis sur des dispositifs, dont l'architecture change d'une longueur à l'autre (à cause des poches de surdopage par exemple). Un levier supplémentaire, telle que l'utilisation de la tension V_{bs}, semble nécessaire pour permettre une extraction plus fiable dans le futur.

Une avancée a été faite dans la compréhension des capacités parasites inhérentes à l'architecture du transistor MOS. Cela nous a permis de développer une nouvelle méthode d'extraction de la longueur effective de canal, plus fiable et plus précise. Il est aujourd'hui possible de mesurer toutes les dérives de L_{eff} en cours du développement de procédé grâce à un suivi systématique en ligne de production.
Un lourd travail de modélisation est également à prévoir en guise de perspective. Ce point est complexe techniquement (prise en compte des effets quantiques, inhomogénéités de dopage) mais fournit un moyen prometteur pour mieux comprendre l'allure de la réponse capacitive des transistors des nœuds 32nm et en deçà, dans lesquels de nouveaux éléments seront introduits (canal enterré, contraintes mécaniques). C'est également par ce biais qu'il deviendra possible de mesurer les paramètres aujourd'hui difficilement mesurables sur des canaux courts : dopage du canal, tension de bandes plates.

La compréhension des mécanismes de transport dans les transistors MOS permet d'anticiper les gains de performance des générations à venir. Les analyses présentées dans le dernier chapitre

ont montré le rôle majoritaire qu'ont les mécanismes de collision dans tous les régimes (linéaire et saturation). Le taux de balisticité élevé que renvoient certaines techniques d'extraction semble inapproprié pour qualifier le transport dans nos transistors. Ces dispositifs ont un canal dans lequel des collisions interviennent fréquemment et sont caractérisés par une survitesse influente dès la source.

Le libre parcours moyen semble être la grandeur pertinente pour évaluer la qualité du transport (i.e. les performances) d'une technologie. Sa faiblesse explique actuellement l'écart observé avec les prédictions de performances balistiques sur les canaux courts, qui sont fortement impactées par des mécanismes additionnels qui dégrade la mobilité (défauts d'implantation ?). L'amélioration des performances des dispositifs de demain nécessitera donc de rehausser ce libre parcours moyen, i.e. de trouver des solutions technologiques qui limiteront la dégradation de mobilité (ex : dopage des extensions S/D in-situ si l'hypothèse des défauts d'implantation est fondée). **C'est lorsque cette condition sera remplie que les gains tant attendus du transport balistique deviendront observables.**

Notre travail sera-t-il utile demain ? Jusqu'où l'évolution de la technologie CMOS sera-t-elle possible et quelles sont les innovations technologiques qui permettront de concevoir les circuits numériques de demain sans nécessiter des solutions complémentaires *more than* Moore[10] ? C'est une question à laquelle de nombreux groupes de recherches et industriels tentent de répondre chaque jour en développant de nouvelles solutions technologiques pour permettre d'envisager le CMOS de demain. La complexification technologique risque donc de se poursuivre et la compréhension physique, via la caractérisation électrique, sera le guide de cette évolution.

Si la technologie CMOS a encore de beaux jours devant elle, la co-intégration avec des architectures radicalement différentes (cf. figure V.24) semble inévitable. Cette révolution est déjà en marche via le SoC *System-On-Chip* dans les domaines de l'imagerie et de la radiofréquence. **La compréhension de la physique du transistor MOS d'aujourd'hui permet donc d'anticiper les bons choix technologiques pour concevoir des blocs logiques performants dans les circuits intégrés de demain.**

10. Le *more than* Moore à été introduit par l'ITRS pour qualifier l'ajout de fonctionnalités complémentaires (optoélectronique, microsystèmes, biométrie, imagerie) qui se différentient de la technologie CMOS traditionnelle.

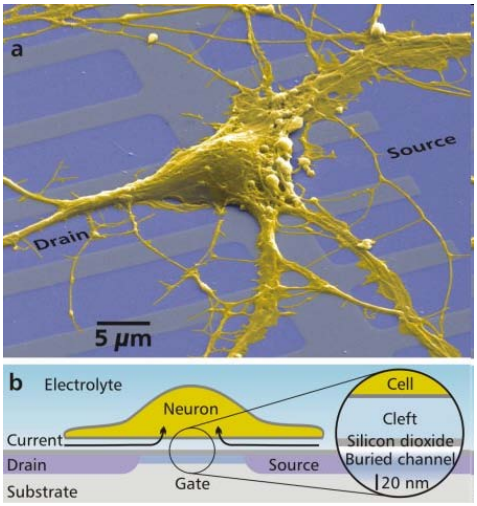


Fig. V.24 – Neurone de rat interfaçant un dispositif électronique de type transistor MOS, expérience réalisée au Max Planck *Institute for Biochemistry* [Voelker05]. Cette application ouvre la voie à de nouvelles possibilités de calcul (non-déterminisme) sans pour autant interdire une co-intégration avec un bloc logique "traditionnel" : **l'avenir ?**

Annexe A

Approximation de l'intégrale de Fermi-Dirac : cas général

L'intégrale de Fermi-Dirac d'ordre j est donnée par la relation (A.1)

$$\mathcal{F}_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{\epsilon^j}{\exp(\epsilon - \eta) + 1} d\epsilon \tag{A.1}$$

Lorsque $\eta < 2$ l'approximation suivante peut être effectuée :

$$\frac{1}{1 + \exp(\epsilon - \eta)} \simeq \exp(\epsilon - \eta) \tag{A.2}$$

En l'injectant dans (A.1) elle rend possible le calcul analytique de l'intégrale :

$$\mathcal{F}_j(\eta) \simeq \exp(\eta), \quad \eta < 2 \tag{A.3}$$

Annexe B

Rappel sur la structure crystalline du silicium

La structure crystalline de type "diamant" constitute l'état stable de composés Si, Ge et C. Dans une telle structure chaque atome de la maille est engagé dasn les liaisons de valence avec ses quatre plus proches voisins. Ces liaisons sont orientées en position tétraédrique et fonf entre elle un angle de 109.28 degrés . On obtient ainsi la structure diamant, composée de deux structures cfc décalées l'une par rapport à l'autre d'un quart de la diagonale du cube, comme représenté dans la figure Fig.B.1. La maille primitive est de type rombohédrique et contient 2 atomes (1 pour chaque réseau cfc). Les propriétés de la structure cristalline du silicium sont résumées dans Tab.B.1 [Gautier03].

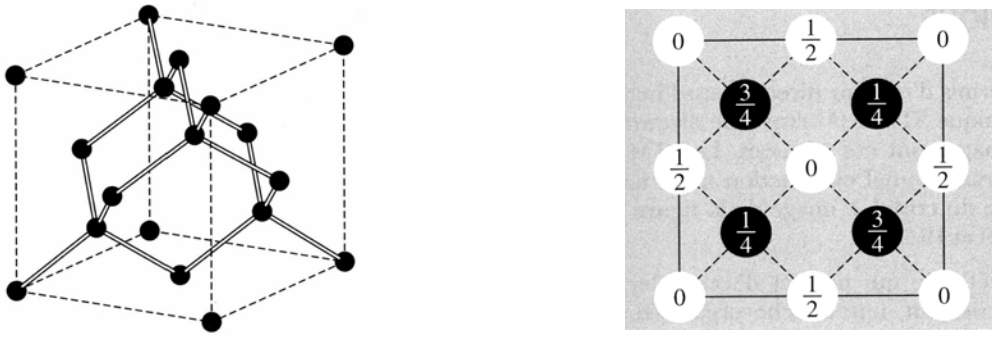


Fig. B.1 – Structure cristalline du silicium montrant les liaisons tétraédriques avec les proches voisins (gauche) et position atomique des atomes dans la maille en projection sur une face (droite) [Kittel98]

Tab. B.1 – Caractéristiques crystallines du silicium

paramètre de maille (Å)	5.43
masse atomique	28.09
densité d'atome (cm^{-3})	5.0×10^{22}
masse volumique ($g.cm^{-3}$)	2.33
distance entre ppv (Å)	2.35

Annexe C

Expression analytique de la kT-layer dans l'approximation dérive diffusion, inversion forte

Nous allons démontrer que la longueur de la kT-layer peut être estimée à (C.1) dans le cas d'un transistor MOS en inversion forte dans l'approche d'un transport de type dérive-diffusion.

$$l_{kT} = \frac{2kT \cdot L}{V_{gt}} \ ? \tag{C.1}$$

Considérons un transistor MOS long en régime de saturation et recherchons la grandeur l_{kT} qui repecte la définition donnée par Lundstrom [Lundstrom97] : $\psi_s(l_{kT}) \stackrel{\text{def}}{=} kT/q$.

Dans un premier temps, calculons $\psi_s(y)$ où y représente la position le long du canal qui vaut 0 à la source et L au drain. L'équation du courant dans l'approche dérive-diffusion, en régime de saturation est donnée par (C.2).

$$I_{dd}^{sat} = \frac{W}{L}\mu C_{eff}\frac{{V_{dsat}}^2}{2}, \quad V_{dsat} \stackrel{\text{def}}{=} V_{gt} \tag{C.2}$$

Avant le pincement du canal et en régime d'inversion forte (qui permet de négliger le courant de diffusion et d'approximer le quasi niveau de Fermi par la potentiel de surface), le courant est donné par la relation ().

$$I_{dd}^{sat} = W\mu C_{eff}\left(V_{dsat} - \psi_s(y)\right)\frac{d\psi_s}{dy} \tag{C.3}$$

Cette équation peut se réécrire sous une forme intégrée (C.4) qui, en posant (C.2)=(C.4) nous amène à une équation du second degré en ψ_s (C.5) dont la résolution nous permet de connaitre le profil de potentiel dans le canal (C.6).

$$I_{dd}^{sat} \cdot y = W\mu C_{eff}\left(V_{dsat}\psi_s - \frac{1}{2}{\psi_s}^2\right) \tag{C.4}$$

$${\psi_s}^2 - 2V_{dsat}\psi_s + {V_{dsat}}^2\frac{y}{L} = 0 \tag{C.5}$$

$$\psi_s(y) = V_{dsat}\left[1 - \sqrt{1 - \frac{y}{L}}\right] \tag{C.6}$$

Au voisinage de la source, le profil de potentiel est donné par $\psi_s(y) \simeq y/(2L)$ qui nous amène à la vérification de la relation (C.7).

$$\psi_s(l_{kT}) = kT/q \Rightarrow l_{kT} = \frac{2kT \cdot L}{V_{gt}} \quad \text{CQFD} \tag{C.7}$$

Bibliographie

[Ahmad86] N. Ahmad et V.K. Arora. *Velocity-field profile of n-silicon : a theoretical analysis.* IEEE Transactions on Electron Devices, vol. ED-33, no. 7, pages 1075 – 7, 1986.

[Aime07] D. Aime, C. Fenouillet-Beranger, P. Perreau, S. Denorme, J. Coignus, A. Cros, D. Fleury, O. Faynot, A. Vandooren, R. Gassilloud, F. Martin, S. Barnola, T. Salvetat, G. Chabanne, L. Brevard, M. Aminpur, F. Leverd, R. Gwoziecki, F. Boeuf, C. Hobbs, A. Zauner, M. Muller, V. Cosnier, S. Minoref, D. Bensahel, M. Orlowski, H. Mingam, A. Wild, S. Deleonibus et T. Skotnicki. *Fully-depleted SOI CMOS technology using W_xN metal gate and $HfSi_xO_yN_z$ high-κ dielectric.* In Proceedings of the 37th European Solid-State Device Research Conference (ESSDERC), pages 255 – 8, Munich, Germany, 2007.

[Ancona97] M.G. Ancona, Z. Yu, W.-C. Lee, R.W. Dutton et P.V. Voorde. *Density-gradient simulations of quantum effects in ultra-thin-oxide MOS structures.* In Proceedings of the 1997 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pages 97 – 100, New York, NY, USA, 1997.

[Ando82] Tsuneya Ando, Alan B. Fowler et Frank Stern. *Electronic properties of two-dimensional systems.* Rev. Mod. Phys., vol. 54, no. 2, pages 437–672, Apr 1982.

[Andres06] E.S. Andres, L. Pantisano, J. Ramos, S. Severi, L. Trojman, S. De Gendt et G. Groeseneken. *RF split capacitance-voltage measurements of short-channel and leaky MOSFET devices.* IEEE Electron Device Letters, vol. 27, no. 9, pages 772 – 4, Sept. 2006.

[Andrieu05] F. Andrieu, T. Ernst, C. Ravit, M. Jurczak, G. Ghibaudo et S. Deleonibus. *In-depth characterization of the hole mobility in 50-nm process-induced strained MOSFETs.* IEEE Electron Device Letters, vol. 26, no. 10, pages 755–757, Oct. 2005.

[Ang06] Kah-Wee Ang, Hock-Chun Chin, King-Jien Chui, Ming-Fu Li, Ganesh Samudra et Yee-Chia Yeo. *Carrier backscattering characteristics of strained N-MOSFET featuring silicon-carbon source/drain regions.* In Proceedings of the 36th European Solid-State Device Research Conference (ESSDERC), pages 89 – 92, Montreux, Switzerland, 2006.

[Ang07] Kah-Wee Ang, Hock-Chun Chin, King-Jien Chui, Ming-Fu Li, Ganesh S. Samudra et Yee-Chia Yeo. *Carrier backscattering characteristics of strained silicon-on-insulator n-MOSFETs featuring silicon-carbon source/drain regions.* Solid-State Electronics, vol. 51, no. 11-12, pages 1444 – 1449, 2007.

[Antognetti81] P. Antognetti, C. Lombardi et D. Antoniadis. *Use of process and 2-D MOS simulation in the study of doping profile influence on S/D resistance in short channel MOSFET's.* International Electron Devices Meeting (IEDM). Technical Digest, pages 574 – 7, New York, NY, USA, 1981.

[Aoyama95] K. Aoyama. *A Method for Extracting the Threshold Voltage of MOSFETs Based on Drain Current Components.* In Simulation of Semiconductor Devices and Processes, volume 6, pages 118–121, September 1995.

[Arora87] N.D. Arora et G.Sh. Gildenblat. *A semi-empirical model of the MOSFET inversion layer mobility for low-temperature operation.* IEEE Transactions on Electron Devices, vol. 34, no. 1, pages 89–93, Jan 1987.

[Ashcroft76] Neil W. Ashcroft et David N. Mermin. *Solid State Physics.* Brooks Cole. January 1976.

[Assad00] F. Assad, Zhibin Ren, D. Vasileska, S. Datta et M. Lundstrom. *On the performance limits for Si MOSFETs : a theoretical study.* IEEE Transactions on Electron Devices, vol. 47, no. 1, pages 232 – 40, 2000.

[Assad99] F. Assad, Zhibin Ren, S. Datta, M. Lundstrom et P. Bendix. *Performance limits of silicon MOSFET's.* International Electron Devices Meeting (IEDM). Technical Digest, pages 547 – 50, Washington DC, USA, 1999.

[Baccarani84] G. Baccarani, M.R. Wordeman et R.H. Dennard. *Generalized scaling theory and its application to a 1/4 micrometer MOSFET design.* IEEE Transactions on Electron Devices, vol. ED-31, no. 4, pages 452 – 62, 1984.

[Barlage00] D.W. Barlage, J.T. O'Keeffe, J.T. Kavalieros, M.M. Nguyen et R.S. Chau. *Inversion MOS capacitance extraction for high-leakage dielectrics using a transmission line equivalent circuit.* IEEE Electron Device Letters, vol. 21, no. 9, pages 454 – 6, 2000.

[Barral07] V. Barral, T. Poiroux, F. Rochette, M. Vinet, S. Barraud, O. Faynot, L. Tosti, F. Andrieu, M. Casse, B. Previtali, R. Ritzenthaler, P. Grosgeorges, E. Bernard, G. LeCarval, D. Munteanu, J.L. Autran et S. Deleonibus. *Will strain be useful for 10 nm quasi-ballistic FDSOI devices ? An experimental study.* Symposium on VLSI Technology, pages 198 – 9, Kyoto, Japan, 2007.

[Barral09a] V. Barral, T. Poiroux, D. Munteanu, J.-L. Autran et S. Deleonibus. *Experimental investigation on the quasi-ballistic transport : Part II. Backscattering coefficient extraction and link with the mobility.* IEEE Transactions on Electron Devices, vol. 56, no. 3, pages 420 – 30, 2009.

[Barral09b] V. Barral, T. Poiroux, J. Saint-Martin, D. Munteanu, J.-L. Autran et S. Deleonibus. *Experimental investigation on the quasi-ballistic transport. Part I. Determination of a new backscattering coefficient extraction methodology.* IEEE Transactions on Electron Devices, vol. 56, no. 3, pages 408 – 19, 2009.

[Bennamane08] K. Bennamane, M. DeMichielis, G. Ghibaudo et D. Esseni. *Extraction of n parameter characterising eff against Eeff curves in strained Si nMOS devices.* IEE Electronics Letters, vol. 44, no. 20, pages 1219 – 20, 2008.

[Bennamane09] K. Bennamane, G. Ghibaudo et A. Benfdila. *Method for extraction of n parameter characterising neff against Eeff curves in FD-SOI Si MOS devices.* IEE Electronics Letters, vol. 45, no. 12, pages 655 – 7, 2009.

[Berz85] F. Berz. *The Bethe condition for thermionic emission near an absorbing boundary.* Solid-State Electronics, vol. 28, no. 10, pages 1007 – 13, 1985.

[Bidal08] G. Bidal, F. Boeuf, F. Payet, S. Denorme, N. Loubet, P. Perreau, C. Mezzomo, M. Marin, D. Fleury, C. Leyris, F. Leverd, P. Gouraud, C. Laviron, R. Beneyton, A. Torres, B. Imbert, D. Dellile, L. Clement, G. Ghibaudo et T. Skotnicki. *Pushing Bulk Transistor with Conventional SiON Gate Oxide for Low Power Applications.* In Proceeding of the 2008 International Conference on Solid State Devices and Materials (SSDM), 2008.

[Bidal09a] G. Bidal. *Guidelines for MOSFET Device Optimization accounting for L-dependent Mobility Degradation.* In Proceedings of the 2009 Silicon Nanoelectronics Workshop (SNW), Kyoto, Japan, 2009.

[Bidal09b] G. Bidal. *Intégration et caractérisation de nouveaux modules technologiques pour les applications CMOS à basse consommation.* PhD thesis, 2009.

[Bohr07] M.T. Bohr, R.S. Chau, T. Ghani et K. Mistry. *The high-k solution [Intel's Core 2 microprocessors].* IEEE Spectrum, vol. 44, no. 10, pages 29 – 35, 2007.

[Brut97] Hugues Brut, Andre Juge et Gerard Ghibaudo. *New Approach for the Extraction of Gate Voltage Dependent Series Resistance and Channel Length Reduction in CMOS Transistors.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 188–193, Monterey, USA, Mars 1997.

[Bude00] J.D. Bude. *MOSFET modeling into the ballistic regime.* In Proceedings of the 2000 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pages 23 – 6, 2000.

[Bufler02] F. M. Bufler et W. Fichtner. *Hole and electron transport in strained Si : Orthorhombic versus biaxial tensile strain.* Applied Physics Letters, vol. 81, pages 82–84, 2002.

[Bufler97] F. M. Bufler, P. Graf, S. Keith et B. Meinerzhagen. *Full band Monte Carlo investigation of electron transport in strained Si grown on Si[sub 1 - x]Ge[sub x] substrates.* Applied Physics Letters, vol. 70, no. 16, pages 2144–2146, 1997.

[Cao99] Kanyu Mark Cao, Weidong Liu, Xiaodong Jin, K. Vashanth, K. Green, J. Krick, T. Vrotsos et Chenming Hu. *Modeling of pocket implanted MOSFETs for anomalous analog behavior.* International Electron Devices Meeting (IEDM). Technical Digest, pages 171 – 4, Washington DC, USA, 1999.

[Cartier09] E. Cartier, M. Steen, B.P. Linder, T. Ando, R. Iijima, M. Frank, J.S. Newbury, Y.H. Kim, F.R. McFeely, M. Copel, R. Haight, C. Choi, A. Callegari, V.K. Paruchuri et V. Narayanan. *pFET Vt control with HfO2/TiN/poly-Si gate stack using a lateral oxygenation process.* Symposium on VLSI Technology. Digest of Technical Papers, pages 42 – 3, Kyoto, Japan, 2009.

[Casse06] M. Casse, L. Thevenod, B. Guillaumot, L. Tosti, F. Martin, J. Mitard, O. Weber, F. Andrieu, T. Ernst, G. Reimbold, T. Billon, M. Mouis et F. Boulanger. *Carrier transport in HfO2/metal gate MOSFETs : physical insight into critical parameters.* IEEE Transactions on Electron Devices, vol. 53, no. 4, pages 759 – 68, 2006.

[Cathignol08] Augustin Cathignol. *Caractérisation et modélisation des fluctuations locales des paramètres électriques du transistor des filières CMOS sub-65 nm.* PhD thesis, 2008.

[Celler03] G.K. Celler et S. Cristoloveanu. *Frontiers of silicon-on-insulator.* Journal of Applied Physics, vol. 93, no. 9, pages 4955 – 78, 2003.

[Chang07] Y.H. Chang, Y.F. Wu et C.S. Ho. *A Simple Method to Extract Source/Drain Series Resistance for Advanced MOSFETs.* IEEE Conference on Electron Devices and Solid-State Circuits (EDSSC). Proceedings, pages 87–90, Dec. 2007.

[Chang09] Shu-Tong Chang, Ming-Han Liao, Chang-Chun Lee, Jacky Huang, Wei-Ching Wang et Bing-Fong Hsieh. *Carrier backscattering characteristics of nanoscale strained complementary metal-oxide-semiconductor devices featuring the optimal stress engineering.* volume 27, pages 1261–1266. AVS, 2009.

[Chau04] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros et M. Metz. *High-k/metal-gate stack and its MOSFET characteristics.* IEEE Electron Device Letters, vol. 25, no. 6, pages 408 – 10, 2004.

[Chen02] Ming-Jer Chen, Huan-Tsung Huang, Kuo-Chuan Huang, Po-Nien Chen, Chih-Sheng Chang et C.H. Diaz. *Temperature dependent channel backscattering coefficients in nanoscale MOSFETs.* International Electron Devices Meeting (IEDM). Technical Digest, pages 39 – 42, San Francisco, CA, USA, 2002.

[Chen04] Ming-Jer Chen, Huan-Tsung Huang, Yi-Chin Chou, Rong-Ting Chen, Yin-Ta Tseng, Po-Nien Chen et C.H. Diaz. *Separation of channel backscattering coefficients in nanoscale MOSFETs.* IEEE Transactions on Electron Devices, vol. 51, no. 9, pages 1409 – 15, 2004.

[Chen06] W.P.N. Chen, P. Su, J.S. Wang, C.H. Lien, C.H. Chang, K. Goto et C.H. Diaz. *A New Series Resistance and Mobility Extraction Method by BSIM Model for Nano-Scale MOSFETs.* International Symposium on VLSI Technology Systems and Applications (VLSI-TSA). Proceedings, pages 1–2, April 2006.

[Chen07] Ming-Jer Chen, S.-G. Yan, Rong-Ting Chen, Chen-Yu Hsieh, Pin-Wei Huang et Han-Ping Chen. *Temperature-oriented experiment and simulation as corroborating evidence of MOSFET backscattering theory.* IEEE Electron Device Letters, vol. 28, no. 2, pages 177 – 9, 2007.

[Chen97] Kai Chen, George Zhang, Jon Duster, Chenming Hu, Jianhui Huang, Zhihong Liu et Ping K. Ko. *MOSFET inversion layer capacitance model based on Fermi-Dirac statistics for wide temperature range.* Solid-State Electronics, vol. 41, no. 3, pages 507 – 509, 1997.

[Cheng96] B. Cheng et J. Woo. *Measurement and modeling of the n-channel and p-channel MOSFET's inversion layer mobility at room and low temperature operation.* Journal de Physique IV (Colloque), volume 6, pages 43 – 7, France, 1996.

[Cheng97a] Baohong Cheng et J. Woo. *A temperature-dependent MOSFET inversion layer carrier mobility model for device and circuit simulation.* IEEE Transactions on Electron Devices, vol. 44, no. 2, pages 343–345, Feb 1997.

[Cheng97b] Yuhua Cheng, Min-Chie Jeng, Shihong Liu, Jianhui Huang, Mansun Chan, Kai Chen, Ping Keung Ko et Chenming Hu. *A Physical and Scalable I-V Model in BSIM3v3 for*

Analog/Digital Circuit Simulation. IEEE Transactions on Electron Devices, vol. 44, pages 277–287, Février 1997.

[Chou85] S.Y. Chou, D.A. Antoniadis et H.I. Smith. *Observation of electron velocity overshoot in sub-100-nm-channel MOSFETs in silicon.* IEEE Electron Device Letters, vol. ED-6, no. 12, pages 665 – 7, 1985.

[Chung07] S.S. Chung, Y.J. Tsai, C.H. Tsai, P.W. Liu, Y.H. Lin, C.T. Tsai, G.H. Ma, S.C. Chien et S.W. Sun. *Technology roadmaps on the ballistic transport in strain engineered nanoscale CMOS devices.* Conference on Electron Devices and Solid-State Circuits (EDSSC), pages 23 – 6, 2007.

[Clerc06] R. Clerc, P. Palestri et L. Selmi. *On the physical understanding of the kT-layer concept in quasi-ballistic regime of transport in nanoscale devices.* IEEE Transactions on Electron Devices, vol. 53, no. 7, pages 1634 – 40, 2006.

[Clerc08] R. Clerc, P. Palestri, L. Selmi et G. Ghibaudo. *Back-scattering in quasi ballistic nano-MOSFETs : the role of non thermal carrier distributions.* 9th International Conference on Ultimate Integration on Silicon, pages 125 – 8, 2008.

[Cohen-Tannoudji06] Claude Cohen-Tannoudji, Bernard Diu et Frank Laloe. *Quantum Mechanics (2 vol. set).* Wiley-Interscience. October 2006.

[Cong91] Huynh Van Cong. *New series representation of Fermi-Dirac integral Fj(-a) for arbitrary j-1, and its effect on Fj(a0+) for integer j0.* Solid-State Electronics, vol. 34, no. 5, pages 489 – 92, 1991.

[Cooper08] D. Cooper, C. Ailliot, R. Truche, J.-P. Barnes, J.-M. Hartmann et F. Bertin. *Experimental off-axis electron holography of focused ion beam-prepared Si p-n junctions with different dopant concentrations.* Journal of Applied Physics, vol. 104, no. 6, pages 064513 (8 pp.) –, 2008.

[Cretu01] B. Cretu, T. Boutchacha, G. Ghibaudo et F. Balestra. *New ratio method for effective channel length and threshold voltage extraction in MOS transistors.* IEEE Electron Device Letters, vol. 37, pages 717–719, Mai 2001.

[Cros05] A. Cros, S. Harrison, R. Cerutti, P. Coronel, G. Ghibaudo et H. Brut. *New extraction method for gate bias dependent series resistance in nanometric double gate transistors.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 69 – 74, Leuven, Belgium, 2005.

[Cros06a] A. Cros. *Caractérisation électrique des transistors MOS à grille enrobante pour les technologies CMOS sub-45nm.* PhD thesis, 2006.

[Cros06b] A. Cros, K. Romanjek, D. Fleury, S. Harrison, R. Cerutti, P. Coronel, B. Dumont, A. Pouydebasque, R. Wacquez, B. Duriez, R. Gwoziecki, F. Boeuf, H. Brut, G. Ghibaudo et T. Skotnicki. *Unexpected mobility degradation for very short devices : A new challenge for CMOS scaling.* International Electron Devices Meeting (IEDM). Technical Digest, pages 1–4, San Francisco, CA, USA, Dec. 2006.

[Crowel66] C.R. Crowel et S.M. Sze. *Temperature dependence of avalanche multiplication in semiconductors".* Applied Physics Letters, vol. 9, page 242, 1966.

[Datta03] S. Datta, G. Dewey, M. Doczy, B.S. Doyle, B. Jin, J. Kavalieros, R. Kotlyar, M. Metz, N. Zelick et R. Chau. *High mobility Si/SiGe strained channel MOS transistors with HfO2/TiN gate stack.* International Electron Devices Meeting (IEDM). Technical Digest, pages 28 – 1, Washington DC, USA, 2003.

[Deng08] Jie Deng, Lan Wei, Li-Wen Chang, Keunwoo Kim, Ching-Te Chuang et H.-S.P. Wong. *Extending technology roadmap by selective device footprint scaling and parasitics engineering.* International Symposium on VLSI Technology Systems and Applications (VLSI-TSA). Proceedings, pages 159 – 60, Hsinchu, Taiwan, 2008.

[Dennard74] R.H. Dennard, F.H. Gaensslen, Hwa-Nien Yu, V.L. Rideout, E. Bassous et A.R. LeBlanc. *Design of ion-implanted MOSFET's with very small physical dimensions.* IEEE Journal of Solid-State Circuits, vol. sc-9, no. 5, pages 256 – 68, 1974.

[Dobbie08] A. Dobbie, B. De Jaeger, M. Meuris, T.E. Whall, E.H.C. Parker et D.R. Leadley. *Channel backscattering characteristics of high performance germanium pMOSFETs.* 9th International Conference on Ultimate Integration on Silicon, pages 7 – 10, 2008.

[Dobrescu00] L. Dobrescu, M. Petrov, D. Dobrescu et C. Ravariu. *Threshold voltage extraction methods for MOS transistors.* International Semiconductor Conference. Proceedings, vol. 1, pages 371–374 vol.1, 2000.

[Drude00] P. Drude. *Zur Elektronentheorie der Metalle ; II. Teil. Galvanomagnetische und thermomagnetische Effecte.* Annalen der Physik, vol. 308, pages 369–402, 1900.

[Duh67] C.Y. Duh et J.L. Moll. *Electron drift velocity in avalanching silicon diodes.* IEEE Transactions on Electron Devices, vol. 14, no. 1, pages 46–49, Jan 1967.

[Dumont07] Benjamin Dumont. *Etude et intégration de jonctions ultra-fines pour les technologies CMOS 45 nm et en deçà.* PhD thesis, 2007.

[Dupre07] C. Dupre, T. Ernst, J.-M. Hartmann, F. Andrieu, J.-P. Barnes, P. Rivallin, O. Faynot, S. Deleonibus, P.F. Fazzini, A. Claverie, S. Cristoloveanu, G. Ghibaudo et F. Cristiano. *Carrier mobility degradation due to high dose implantation in ultrathin unstrained and strained silicon-on-insulator films.* Journal of Applied Physics, vol. 102, no. 10, pages 104505 – 1, 2007.

[El Kamchouchi75] H. El Kamchouchi et A.A. Zaky. *A direct method for the calculation of the edge capacitance of thick electrodes.* Journal of Physics D (Applied Physics), vol. 8, no. 12, pages 1365 – 71, 1975.

[Elmasry82] M.I. Elmasry. *Capacitance calculations in MOSFET VLSI.* IEEE Electron Device Letters, vol. 3, no. 1, pages 6–7, Jan 1982.

[Eminente07] S.. Eminente, N.. Barin, P.. Palestri, C.. Fiegna et E.. Sangiorgi. *Small-signal analysis of decananometer bulk and SOI MOSFETs for analog/mixed-signal and RF applications using the time-dependent Monte Carlo approach.* IEEE Transactions on Electron Devices, vol. 54, no. 9, pages 2283 – 92, 2007.

[Emrani94] A. Emrani, G. Ghibaudo et F. Balestra. *On the universal electric field dependence of the electron and hole effective mobility in MOS inversion layers.* Solid-State Electronics, vol. 37, no. 1, pages 111 – 13, 1994.

[Evans05] M.H. Evans, M. Caussanel, R.D. Schrimpf et S.T. Pantelides. *First-principles modeling of double-gate UTSOI MOSFETs.* International Electron Devices Meeting (IEDM). Technical Digest, pages 4 pp. –, Washington DC, USA, 2005.

[Faynot94] O. Faynot, S. Cristoloveanu, P. McLarty, C. Raynaud et J. Gautier. *A new parameter extraction method for ultra-thin oxide SOI MOSFET's.* IEEE International SOI Conference. Proceedings, pages 17 – 18, Nantucket Island, MA, USA, 1994.

[Ferrier06] M. Ferrier, R. Clerc, G. Ghibaudo, F. Boeuf et T. Skotnicki. *Analytical model for quantization on strained and unstrained bulk nMOSFET and its impact on quasi-ballistic current.* Solid-State Electronics, vol. 50, no. 1, pages 69 – 77, 2006.

[Ferrier07] M. Ferrier. *Modélisation analytique du transport balistique et quasi-balistique dans les MOSFETs avancés.* PhD thesis, 2007.

[Fischetti01] M.V. Fischetti, D.A. Neumayer et E.A. Cartier. *Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator : The role of remote phonon scattering.* Journal of Applied Physics, vol. 90, no. 9, pages 4587 – 608, 2001.

[Fischetti02] M. V. Fischetti, F. Gamiz et W. Hansch. *On the enhanced electron mobility in strained-silicon inversion layers.* Journal of Applied Physics, vol. 92, no. 12, pages 7320–7324, 2002.

[Fischetti07] M.V. Fischetti, T.P. O'Regan, Sudarshan Narayanan, C. Sachs, Seonghoon Jin, Jiseok Kim et Yan Zhang. *Theoretical study of some physical aspects of electronic transport in nMOSFETs at the 10-nm gate-length.* IEEE Transactions on Electron Devices, vol. 54, no. 9, pages 2116 – 36, 2007.

[Fleury08a] D. Fleury, A. Cros, H. Brut et G. Ghibaudo. *New Y-function-based methodology for accurate extraction of electrical parameters on nano-scaled MOSFETs.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 160–165, March 2008.

[Fleury08b] D. Fleury, A. Cros, K. Romanjek, D. Roy, F. Perrier, B. Dumont, H. Brut et G. Ghibaudo. *Automatic extraction methodology for accurate measurements of effective channel length on 65-nm MOSFET technology and below.* IEEE Transactions on Semiconductor Manufacturing, vol. 21, no. 4, pages 504 – 12, 2008.

[Fleury09a] D. Fleury, G. Bidal, A. Cros, F. Boeuf, T. Skotnicki et G. Ghibaudo. *New experimental insight into ballisticity of transport in strained bulk MOSFETs.* Symposium on VLSI Technology. Digest of Technical Papers, pages 16 – 17, Kyoto, Japan, 2009.

[Fleury09b] D. Fleury, A. Cros, G. Bidal, H. Brut, E. Josse et G. Ghibaudo. *A new technique to extract the gate bias dependent S/D series resistance of sub-100 nm MOSFETs.* International Symposium on VLSI Technology Systems and Applications (VLSI-TSA). Proceedings, pages 109 – 10, Hsinchu, Taiwan, 2009.

[Fleury09c] D. Fleury, A. Cros, G. Bidal, J. Rosa et G. Ghibaudo. *A New Technique to Extract the Source/Drain Series Resistance of MOSFETs.* IEEE Electron Device Letters, vol. 30, no. 9, pages 975–977, Sept. 2009.

[Fossum03] J.G. Fossum et Weimin Zhang. *Performance projections of scaled CMOS devices and circuits with strained Si-on-SiGe channels.* IEEE Transactions on Electron Devices, vol. 50, no. 4, pages 1042–1049, April 2003.

[Gallon04] C. Gallon, C. Fenouillet-Beranger, Y.M. Meziani, J.P. Cesso, J. Lusakowski, F. Teppe, N. Dyakonova, A. Vandooren, W. Knap, G. Ghibaudo, D. Delille, S. Cristoloveanu et T. Skotnicki. *New magnetoresistance method for mobility extraction in scaled fully-depleted SOI devices.* IEEE International SOI Conference. Proceedings, pages 153 – 5, New York NY, USA, 2004.

[Gautier03] J. Gautier. *Physique des dispositifs pour circuites intégrés silicium.* Lavoisier. 2003.

[Ge01] Lixin Ge, J.G. Fossum et Bin Liu. *Physical compact modeling and analysis of velocity overshoot in extremely scaled CMOS devices and circuits.* IEEE Transactions on Electron Devices, vol. 48, no. 9, pages 2074 – 80, 2001.

[Ghibaudo09] G. Ghibaudo, M. Mouis, L. Pham-Nguyen, K. Bennamane, I. Pappas, A. Cros, G. Bidal, D. Fleury, A. Claverie, G. Benassayag, P.-F. Fazzini, C. Fenouillet-Beranger, S. Monfray, F. Boeuf, S. Cristoloveanu, T. Skotnicki et N. Collaert. *Electrical transport characterization of nanoCMOS devices with ultra-thin silicon film.* International Workshop on Junction Technology (IWJT), pages 58 – 63, Kyoto, Japan, 2009.

[Ghibaudo88] Gérard Ghibaudo. *New method for the extraction of MOSFET parameters.* IEE Electronics Letters, vol. 24, pages 543–545, Avril 1988.

[Ghibaudo97] G. Ghibaudo. *Critical MOSFETs operation for low voltage/low power IC's : ideal characteristics, parameter extraction, electrical noise and RTS fluctuations.* In Proceedings of the first session on Low-power, low-voltage integrated circuits : technology and design, pages 31–57, Amsterdam, The Netherlands, 1997. Elsevier Science Publishers B. V.

[Goodnick85] S.M. Goodnick, D.K. Ferry, C.W. Wilmsen, Z. Liliental, D. Fathy et O.L. Krivanek. *Surface roughness at the Si(100)-SiO2 interface.* Physical Review B (Condensed Matter), vol. 32, no. 12, pages 8171 – 86, 1985.

[Greeneich83] E.W. Greeneich. *An analytical model for the gate capacitance of small-geometry MOS structures.* IEEE Transactions on Electron Devices, vol. 30, no. 12, pages 1838–1839, Dec 1983.

[Grove67] A. S. Grove. *Physics and Technology os Semiconductor Devices.* John Wiley & Sons. 1967.

[Hansch89] W. Hansch. *Carrier transport near the Si/SiO2 interface of a MOSFET.* Solid State Electronics, vol. 32, pages 839–849, Octobre 1989.

[Hareland96] S.A. Hareland, S. Krishnamurthy, S. Jallepalli, Choh-Fei Yeap, K. Hasnat, Jr. Tasch A.F. et C.M. Maziar. *A computationally efficient model for inversion layer quantization effects in deep submicron N-channel MOSFETs.* IEEE Transactions on Electron Devices, vol. 43, no. 1, pages 90 – 6, 1996.

[Hareland97] S.A. Hareland, S. Jallepalli, G. Chindalore, W.-K. Shih, Jr. Tasch A.F. et C.M. Maziur. *A simple model for quantum mechanical effects in hole inversion layers in silicon PMOS devices.* IEEE Transactions on Electron Devices, vol. 44, no. 7, pages 1172 – 3, 1997.

[Hareland98a] S.A. Hareland, S. Jallepalli, Wei-Kai Shih, Haihong Wang, G.L. Chindalore, A.F. Tasch et C.M. Maziar. *A physically-based model for quantization effects in hole inversion layers.* IEEE Transactions on Electron Devices, vol. 45, no. 1, pages 179 – 86, 1998.

[Hareland98b] S.A. Hareland, M. Manassian, W.-K. Shih, S. Jallepalli, H. Wang, G.L. Chindalore, Al.F. Tasch et C.M. Maziar. *Computationally efficient models for quantization effects in MOS electron and hole accumulation layers.* IEEE Transactions on Electron Devices, vol. 45, no. 7, pages 1487 – 93, 1998.

[Heish01] T.S. Heish, Y.W.Chang, W.J. Tsai et T.C. Lu. *A new Leff Extraction approach for Devices with Pocket Implants.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 15–18, Kobe, Japan, Mars 2001.

[Henning87] A.K. Henning, N.N. Chan, J.T. Watt et J.D. Plummer. *Substrate current at cryogenic temperatures : Measurements and a two-dimensional model for CMOS technology.* IEEE Transactions on Electron Devices, vol. 34, no. 1, pages 64–74, Jan 1987.

[Hobbs03] C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White et P. Tobin. *Fermi level pinning at the polySi/metal oxide interface.* Symposium on VLSI Technology. Digest of Technical Papers, pages 9 – 10, Kyoto, Japan, 2003.

[Hu85] C. Hu, S.C. Tam, F.-C. Hsu, P.-K. Ko, T.-Y. Chan et K.W. Terrill. *Hot-electron-induced MOSFET degradation-model, monitor, and improvement.* IEEE Transactions on Electron Devices, vol. ED-32, no. 2, pages 375 – 85, 1985.

[Hu87] G.J. Hu, C. Chang et Y.-T. Chia. *Gate-voltage-dependent effective channel length and series resistance of LDD MOSFETs.* IEEE Transactions on Electron Devices, vol. ED-34, no. 12, pages 2469 – 75, 1987.

[Hu95] Hang Hu, J.B. Jacobs, L.T. Su et D.A. Antoniadis. *A study of deep-submicron MOSFET scaling based on experiment and simulation.* IEEE Transactions on Electron Devices, vol. 42, no. 4, pages 669 – 77, 1995.

[Huang09] J. Huang, D. Heh, P. Sivasubramani, P.D. Kirsch, G. Bersuker, D.C. Gilmer, M.A. Quevedo-Lopez, M.M. Hussain, P. Majhi, P. Lysaght, H. Park, N. Goel, C. Young, C.S. Park, C. Park, M. Cruz, V. Diaz, P.Y. Hung, J. Price, H.-H. Tseng et R. Jammy. *Gate first high-k/metal gate stacks with zero SiOx interface achieving EOT = 0.59 nm for 16 nm application.* Symposium on VLSI Technology. Digest of Technical Papers, pages 34 – 5, Kyoto, Japan, 2009.

[Huang96] Cheng-Liang Huang, J.V. Faricelli, D.A. Antoniadis, N.A. Khalil et R.A. Rios. *An accurate gate length extraction method for sub-quarter micron MOSFET's.* IEEE Transactions on Electron Devices, vol. 43, no. 6, pages 958 – 64, 1996.

[INRIA] INRIA. *Scilab Home Page.* http://www.scilab.org/.

[Jacoboni77] C. Jacoboni, C. Canali, G. Ottaviani et A. Alberigi Quaranta. *A review of some charge transport properties of silicon.* Solid-State Electronics, vol. 20, no. 2, pages 77 – 89, 1977.

[Jeon89] D.S. Jeon et D.E. Burk. *MOSFET electron inversion layer mobilities-a physically based semi-empirical model for a wide temperature range.* IEEE Transactions on Electron Devices, vol. 36, no. 8, pages 1456 – 63, 1989.

[Jones66] E.L. Jones. *Rational Chebyshev approximation of the Fermi-Dirac integrals.* Proceedings of the IEEE, vol. 54, no. 4, pages 708–709, April 1966.

[Josse06] E. Josse, S. Parihar, O. Callen, P. Ferreira, C. Monget, A. Farcy, M. Zaleski, D. Villanueva, R. Ranica, M. Bidaud, D. Barge, C. Laviron, N. Auriac, C. Le Cam, S. Harrison, S. Warrick, F. Leverd, P. Gouraud, S. Zoll, F. Guyader, E. Perrin, E. Baylac, J. Belledent, B. Icard, B. Minghetti, S. Manakli, L. Pain, V. Huard, G. Ribes, K. Rochereau, S. Bordez, C. Blanc, A. Margain, D. Delille, R. Pantel, K. Barla, N. Cave et M. Haond. *A cost-effective low power platform for the 45-nm technology node.* International Electron Devices Meeting (IEDM). Technical Digest, pages 4 pp. –, San Francisco, CA, USA, 2006.

[Jurczak99] M. Jurczak, T. Skotnicki, M. Paoli, B. Tormen, J.-L. Regolini, C. Morin, A. Schiltz, J. Martins, R. Pantel et J. Galvier. *SON (silicon on nothing)-a new device architecture for the ULSI era.* Symposium on VLSI Technology. Digest of Technical Papers, pages 29 – 30, Honolulu, Hawaii, 1999.

[Katto97] H. Katto. *Device Parameter Extraction in the Linear Region of MOSFET's.* IEEE Transactions on Electron Devices, vol. 18, pages 408–410, Septembre 1997.

[Khakifirooz08] A. Khakifirooz et D.A. Antoniadis. *MOSFET performance scaling-Part II : future directions*. IEEE Transactions on Electron Devices, vol. 55, no. 6, pages 1401 – 8, 2008.

[Kim02a] Seong-Dong Kim, Cheol-Min Park et J.C.S. Woo. *Advanced model and analysis of series resistance for CMOS scaling into nanometer regime. I. Theoretical derivation*. IEEE Transactions on Electron Devices, vol. 49, no. 3, pages 457–466, Mar 2002.

[Kim02b] Seong-Dong Kim, Cheol-Min Park et J.C.S. Woo. *Advanced model and analysis of series resistance for CMOS scaling into nanometer regime. II. Quantitative analysis*. IEEE Transactions on Electron Devices, vol. 49, no. 3, pages 467–472, Mar 2002.

[Kim06] Seung-Hwan Kim, J.G. Fossum et Ji-Woon Yang. *Modeling and significance of fringe capacitance in nonclassical CMOS devices with gate-source/drain underlap*. IEEE Transactions on Electron Devices, vol. 53, no. 9, pages 2143 – 50, 2006.

[Kim08] Junsoo Kim, Jaehong Lee, Ickhyun Song, Yeonam Yun, Jong Duk Lee, Byung-Gook Park et Hyungcheol Shin. *Accurate Extraction of Effective Channel Length and Source/Drain Series Resistance in Ultrashort-Channel MOSFETs by Iteration Method*. IEEE Transactions on Electron Devices, vol. 55, no. 10, pages 2779–2784, Oct. 2008.

[Kiréev75] P. Kiréev. *La Physique des Semiconducteurs*. Edition Mir. 1975.

[Kittel98] C. Kittel. *Physique de l'état solide (7eme édition)*. Dunod. 1998.

[Koomen73] J. Koomen. *Investigation of the MOST channel conductance in weak inversion*. Solid-State Electronics, vol. 16, no. 7, pages 801 – 10, 1973.

[Lau01] M.M. Lau, C.Y.T. Chiang, Y.T. Yeow et Z.Q. Yao. *A new method of threshold voltage extraction via MOSFET gate-to-substrate capacitance measurement*. IEEE Transactions on Electron Devices, volume 48, pages 1742 – 4, USA, 2001.

[Lau99] M.M. Lau, C.Y.T. Chiang, Y.T. Yeow et Z.Q. Yao. *Measurement of VT and Leff using MOSFET gate-substrate capacitance*. IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 152 – 5, Goteborg, Sweden, 1999.

[Lee95] T.-Y. Lee et R.M. Fox. *Extraction of thermal resistance for fully-depleted SOI MOSFETs*. IEEE International SOI Conference. Proceedings, pages 78 – 9, Tucson, Arizona, USA, 1995.

[Leo94] William R Leo. *Techniques for nuclear and particle physics experiments : a how-to approach*. Springer. 1994.

[Leong98] M.K. Leong, R. Logan et J. Slinkman. *Efficient quantum correction model for multi-dimensional CMOS simulations*. In Proceedings of the 1998 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pages 129 – 32, Wien, Austria, 1998.

[Li03] Yiming Li, Shao-Ming Yu, Chien-Shao Tang et Teng-Sheng Chao. *Comparison of quantum correction models for ultratin oxide single- and double-gate MOS structures under the inversion conditions*. Third IEEE Conference on Nanotechnology. IEEE-NANO 2003. Proceedings, volume vol.2, pages 36 – 9, 2003.

[Liao08] M. H. Liao, C. W. Liu, Lingyen Yeh, T.-L. Lee et M.-S. Liang. *Gate width dependence on backscattering characteristics in the nanoscale strained complementary metal-oxide-semiconductor field-effect transistor*. Applied Physics Letters, vol. 92, no. 6, page 063506, 2008.

[Lin05a] Hong-Nien Lin, Hung-Wei Chen, Chih-Hsin Ko, Chung-Hu Ge, Horng-Chih Lin, Tiao-Yuan Huang et Wen-Chin Lee. *Channel backscattering characteristics of strained PMOSFETs with embedded SiGe source/drain*. International Electron Devices Meeting (IEDM). Technical Digest, pages 4 pp. –, Washington DC, USA, 2005.

[Lin05b] Hong-Nien Lin, Hung-Wei Chen, Chih-Hsin Ko, Chung-Hu Ge, Horng-Chih Lin, Tiao-Yuan Huang et Wen-Chin Lee. *Channel backscattering characteristics of uniaxially strained nanoscale CMOSFETs*. IEEE Electron Device Letters, vol. 26, no. 9, pages 676 – 8, 2005.

[Lin05c] Hong-Nien Lin, Hung-Wei Chen, Chih-Hsin Ko, Chung-Hu Ge, Horng-Chih Lin, Tiao-Yuan Huang, Wen-Chin Lee et D.D. Tang. *The impact of uniaxial strain engineering on channel backscattering in nanoscale MOSFETs*. Symposium on VLSI Technology. Digest of Technical Papers, pages 174 – 5, Kyoto, Japan, 2005.

[Lin06] Hong-Nien Lin, Hung-Wei Chen, Chih-Hsin Ko, Chung-Hu Ge, Horng-Chih Lin, Tiao-Yuan Huang et Wen-Chin Lee. *Characterizing the Channel Backscattering Behavior in Nanoscale Strained Complementary Metal Oxide Semiconductor Field-Effect Transistors.* Japanese Journal of Applied Physics, vol. 45, no. 11, pages 8611–8617, 2006.

[Lin07] Da-Wen Lin, Ming-Lung Cheng, Shyh-Wei Wang, Chung-Cheng Wu et Ming-Jer Chen. *A Constant-Mobility Method to Enable MOSFET Series-Resistance Extraction.* IEEE Electron Device Letters, vol. 28, no. 12, pages 1132–1134, 2007.

[Liou97] J.J. Liou, A. Ortiz-Condez et F.G. Sanchez. *Extraction of the threshold voltage of MOSFETs : an overview.* Hong Kong Electron Devices Meeting. Proceedings, pages 31 – 8, New York, NY, USA, 1997.

[Liu09] Xi Liu, Xiaoshi Jin et Jong-Ho Lee. *A compact model of fringing field induced parasitic capacitance for deep sub-micrometer MOSFETs.* Solid-State Electronics, vol. 53, no. 9, pages 1041 – 1045, 2009.

[Lochtefeld01] A. Lochtefeld et D.A. Antoniadis. *On experimental determination of carrier velocity in deeply scaled NMOS : how close to the thermal limit ?* IEEE Electron Device Letters, vol. 22, no. 2, pages 95 – 7, 2001.

[Lochtefeld02] A. Lochtefeld, I.J. Djomehri, G. Samudra et D.A. Antoniadis. *New insights into carrier transport in n-MOSFETs.* IBM Journal of Research and Development, vol. 46, no. 2-3, pages 347 – 57, 2002.

[Lundstrom00] Mark Lundstrom. *Fundamentals of Carrier Transport (second edition).* Cambridge University Press. 2000.

[Lundstrom01] M.S. Lundstrom. *On the mobility versus drain current relation for a nanoscale MOSFET.* IEEE Electron Device Letters, vol. 22, no. 6, pages 293 – 5, 2001.

[Lundstrom02a] M. Lundstrom et Z. Ren. *Essential physics of carrier transport in nanoscale MOSFETs.* IEEE Transactions on Electron Devices, vol. 49, no. 1, pages 133 – 41, 2002.

[Lundstrom02b] M. Lundstrom et J.-H. Rhew. *A Landauer Approach to Nanoscale mosfets.* Journal of Computational Electronics, vol. 1 (4), pages 481–489, 2002.

[Lundstrom06a] Mark Lundstrom. *ECE 612 Lecture 14 : Effective Mobility*, Oct 2006.

[Lundstrom06b] Mark Lundstrom. *Nanoscale Transistors : Device Physics, Modeling and Simulation.* Springer. 2006.

[Lundstrom97] M. Lundstrom. *Elementary scattering theory of the Si MOSFET.* IEEE Electron Device Letters, vol. 18, no. 7, pages 361 – 3, 1997.

[Ma00] Yutao Ma, Litian Liu, Z. Yu et Zhijian Li. *Validity and applicability of triangular potential well approximation in modeling of MOS structure inversion and accumulation layer.* IEEE Transactions on Electron Devices, vol. 47, no. 9, pages 1764 – 7, 2000.

[Marsland91] J.S. Marsland. *Temperature dependence of ionisation coefficients in silicon derived from physical model [MOSFETs].* IEE Electronics Letters, vol. 27, no. 22, pages 1997 – 8, 1991.

[Masetti83] G. Masetti, M. Severi et S. Solmi. *Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon.* IEEE Transactions on Electron Devices, vol. ED-30, no. 7, pages 764 – 9, 1983.

[Mathieu01] Henry Mathieu. *Physique des semiconducteurs et des composants électroniques (5e édition).* Dunod. 2001.

[Mayuzumi09a] S. Mayuzumi, S. Yamakawa, Y. Tateshita, T. Hirano, M. Nakata, S. Yamaguchi, K. Tai, H. Wakabayashi, M. Tsukamoto et N. Nagashima. *High-performance metal/high-k n- and p-MOSFETs with top-cut dual stress liners using gate-last damascene process on (100) substrates.* IEEE Transactions on Electron Devices, vol. 56, no. 4, pages 620 – 6, 2009.

[Mayuzumi09b] S. Mayuzumi, S. Yamakawa, Y. Tateshita, M. Tsukamoto, H. Wakabayashi, T. Ohno et N. Nagashima. *Stress-enhancement technique in narrowing NMOSFETs with damascene-gate process and tensile liner.* International Symposium on VLSI Technology Systems and Applications (VLSI-TSA). Proceedings, pages 20 – 1, Hsinchu, Taiwan, 2009.

[McGill75] T.C. McGill et R. Baron. *Neutral impurity scattering in semiconductors.* Physical Review B, vol. 11, pages 5208–5210, 1975.

[McKelvey61] J. P. McKelvey, R. L. Longini et T. P. Brody. *Alternative Approach to the Solution of Added Carrier Transport Problems in Semiconductors.* Phys. Rev., vol. 123, no. 1, pages 51–57, Jul 1961.

[McLarty95] P.K. McLarty, S. Cristoloveanu, O. Faynot, V. Misra, J.R. Hauser et J.J. Wortman. *A simple parameter extraction method for ultra-thin oxide MOSFETs.* Solid-State Electronics, vol. 38, no. 6, pages 1175 – 7, 1995.

[Meer00] Hans Van Meer, Kirklen Henson, Jeong-Ho Lyu, Maarten Rosmeulen, Stefan Kubicek, Nadine Collaert et Krinstine De Meyer. *Limitation of Shift-and-Ratio Based L_{eff} Extraction Techniques for MOS Transistors with Halo or Pocket Implants.* IEEE Transactions on Electron Devices, vol. 21, pages 133–136, Mars 2000.

[Mistry07] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier et Buehler. *A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100Pb-free Packaging.* In Electron Devices Meeting, 2007, 2007.

[Mohapatra03] N.R. Mohapatra, M.P. Desai, S.G. Narendra et V. Ramgopal Rao. *Modeling of parasitic capacitances in deep submicrometer conventional and high-K dielectric MOS transistors.* IEEE Transactions on Electron Devices, vol. 50, no. 4, pages 959 – 66, 2003.

[Moll64] J. L. Moll. *Physics of Semiconductors.* McGraw-Hill Book Co. 1964.

[Moore95] G.E. Moore. *Lithography and the future of Moore's law.* International Society for Optical Engineering (SPIE). Proceedings, volume 2440, pages 2 – 17, San Diego, CA, USA, 1995.

[Mourrain00] C. Mourrain, B. Cretu, G. Ghibaudo et P. Cottin. *New method for parameter extraction in deep submicrometer MOSFETs.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 181–186, Monterey, CA, USA, Mars 2000.

[Na02] M.H. Na, E.J. Nowak, W. Haensch et J. Cai. *The effective drive current in CMOS inverters.* International Electron Devices Meeting (IEDM). Technical Digest, pages 121 – 4, San Francisco, CA, USA, 2002.

[Natori01] K. Natori. *Scaling Limit of the MOS Transistor – A Ballistic MOSFET (invited paper).* In IEICE Trans. Electron., volume E84-C, pages 1029–1036, August 2001.

[Natori94] K. Natori. *Current-voltage characteristics of silicon on insulator metal oxide semiconductor field effect transistors in ballistic mode.* Japanese Journal of Applied Physics, Part 1 (Regular Papers & Short Notes), vol. 33, pages 554–557, 1994.

[Ning77] T.H. Ning, C.M. Osburn et H.N. Yu. *Emission probability of hot electrons from silicon into silicon dioxide.* Journal of Applied Physics, vol. 48, no. 1, pages 286 – 93, 1977.

[Niu00] Guofu Niu, S.J. Mathew, J.D. Cressler et S. Subbanna. *A novel channel resistance ratio method for effective channel length and series resistance extraction in MOSFETs.* Solid-State Electronics, vol. 44, no. 7, pages 1187 – 9, 2000.

[NXP07] NXP. *MOS model 9*, 2007.

[Ong87] Tong-Chern Ong, Ping K. Ko et Chenming Hu. *50-A Gate-Oxide MOSFET's at 77 K.* IEEE Transactions on Electron Devices, vol. 34, pages 2129–2135, Octobre 1987.

[Ortiz-Conde02] A. Ortiz-Conde, F.J. Garcia Sanchez, J.J. Liou, A. Cerdeira, M. Estrada et Y. Yue. *A review of recent MOSFET threshold voltage extraction methods.* Microelectronics Reliability, vol. 42, no. 4-5, pages 583 – 96, 2002.

[Paasch82] G. Paasch et H. Ubensee. *A modified local density approximation. Electron density in inversion layers.* Physica Status Solidi B, vol. 113, no. 1, pages 165 – 78, 1982.

[Pacelli99] A. Pacelli, A.S. Spinelli et L.M. Perron. *Carrier quantization at flat bands in MOS devices.* IEEE Transactions on Electron Devices, vol. 46, no. 2, pages 383 – 7, 1999.

[Palestri05] P. Palestri, D. Esseni, S. Eminente, C. Fiegna, E. Sangiorgi et L. Selmi. *Understanding quasi-ballistic transport in nano-MOSFETs : part I-scattering in the channel and in the drain.* IEEE Transactions on Electron Devices, vol. 52, no. 12, pages 2727 – 35, 2005.

[Palestri06] P. Palestri, R. Clerc, D. Esseni, L. Lucci et L. Selmi. *Multi-subband-Monte-Carlo investigation of the mean free path and of the kT layer in degenerated quasi ballistic nanoMOSFETs.* International Electron Devices Meeting (IEDM). Technical Digest, pages 4 pp. –, San Francisco, CA, USA, 2006.

[Pappas09] I. Pappas, G. Ghibaudo, C.A. Dimitriadis et C. Fenouillet-Béranger. *Backscattering coefficient and drift-diffusion mobility extraction in short channel MOS devices.* Solid-State Electronics, vol. 53, no. 1, pages 54 – 56, 2009.

[Payet08] F. Payet, F. Boeuf, C. Ortolland et T. Skotnicki. *Nonuniform mobility-enhancement techniques and their impact on device performance.* IEEE Transactions on Electron Devices, vol. 55, no. 4, pages 1050 – 7, 2008.

[Pham-Nguyen08] L. Pham-Nguyen, C. Fenouillet-Beranger, A. Vandooren, A. Wild, G. Ghibaudo et S. Cristoloveanu. *Direct comparison of Si/high-K and Si/SiO2 channels in advanced FD SOI MOSFETs.* IEEE International SOI Conference. Proceedings, pages 25 – 6, New Paltz, NY, USA, 2008.

[Pirovano00] A. Pirovano, A.L. Lacaita, G. Zandler et R. Oberhuber. *Explaining the dependences of the hole and electron mobilities in Si inversion layers.* IEEE Transactions on Electron Devices, vol. 47, no. 4, pages 718 – 24, 2000.

[Ponton06] D. Ponton, L. Lucci, P. Palestri, D. Esseni et L. Selmi. *Assessment of the impact of biaxial strain on the drain current of decanometric n-MOSFET.* In Proceedings of the 36th European Solid-State Device Research Conference (ESSDERC), pages 166 – 9, Montreux, Switzerland, 2006.

[Pouydebasque05] A. Pouydebasque, B. Dumont, F. Wacquant, A. Halimaoui, C. Laviron, D. Lenoble, R. El-Farhane, B. Duriez, F. Arnaud, V. Carron, C. Rossato, S. Pokrant, F. Salvetti, A. Dray, F. Boeuf et T. Skotnicki. *NMOS-junction integration study with ultra-high temperature non-diffusive laser annealing for the 45 nm node and below.* In Junction Technology, 2005. Extended Abstracts of the Fifth International Workshop on, pages 15–18, June 2005.

[Prange68] R. E. Prange et Tsu-Wei Nee. *Quantum Spectroscopy of the Low-Field Oscillations in the Surface Impedance.* Physical Review, vol. 163, pages 779–786, 1968.

[Press07] William Press, Saul Teukolsky, William Vetterling et Brian Flannery. *Numerical Recipes - the Art of Scientific Computing.* Cambridge University Press. 2007.

[Prégaldiny02] Fabien Prégaldiny, Christophe Lallement et Daniel Mathiot. *A simple efficient model of parasitic capacitances of deep-submicron LDD MOSFETs.* Solid State Electronics, vol. 46, pages 2191–2198, Juin 2002.

[Prégaldiny03] Fabien Prégaldiny. *Etude et modélisation du comportement électrique des transistors MOS fortement submicroniques.* PhD thesis, 2003.

[Rafik08] M. Rafik. *Caractérisation et modélisation de la fiabilité des transistors avancés à diélectriques de haute permittivité et à grille métallique.* PhD thesis, 2008.

[Rahman03] A. Rahman, Jing Guo, S. Datta et M.S. Lundstrom. *Theory of ballistic nanotransistors.* IEEE Transactions on Electron Devices, vol. 50, no. 9, pages 1853 – 64, 2003.

[Ramos06] J. Ramos, S. Severi, E. Augendre, C. Kerner, T. Chiarella, A. Nackaerts, T. Hoffmann, N. Collaert, M. Jurczak et S. Biesemans. *Effective mobility extraction based on a split RF C-V method for short-channel FinFETs.* In Proceedings of the 36th European Solid-State Device Research Conference (ESSDERC), pages 363 – 6, Montreux, Switzerland, 2006.

[Reichert96] Guenter Reichert et Thierry Ouisse. *Relationship Between Empirical and Theoritical Mobility Models in Silicon Inversion Layers.* IEEE Transactions on Electron Devices, vol. 43, pages 1394–1398, Septembre 1996.

[Ren03] Z. Ren, M.V. Fischetti, E.P. Gusev, E.A. Cartier et M. Chudzik. *Inversion channel mobility in high-k high performance MOSFETs.* International Electron Devices Meeting (IEDM). Technical Digest, pages 33 – 2, Washington DC, USA, 2003.

[Rochette08] F. Rochette. *Etude et caractérisation de l'influence des contraintes mécaniques sur les propriétés du transport électronique dans les architectures MOS avancées.* PhD thesis, 2008.

[Roldan97] J.B. Roldan, F. Gamiz, J.A. Lopez-Villanueva et J.E. Carceller. *Modeling effects of electron-velocity overshoot in a MOSFET.* IEEE Transactions on Electron Devices, vol. 44, no. 5, pages 841 – 6, 1997.

[Romanjek03] K. Romanjek, F. Lime, G. Ghibaudo et C. Leroux. *New approach for the gate current source-drain partition modeling in advanced MOSFETs.* Solid-State Electronics, vol. 47, no. 10, pages 1657 – 1661, 2003.

[Romanjek04a] K. Romanjek. *Caractérisation et modélisation des transistors CMOS des technologies 50nm et en deçà*. PhD thesis, 2004.

[Romanjek04b] K. Romanjek, F. Andrieu, T. Ernst et G. Ghibaudo. *Improved split C-V method for effective mobility extraction in sub-0.1-µm Si MOSFETs*. IEEE Electron Device Letters, vol. 25, no. 8, pages 583–585, Aug. 2004.

[Sabnis79] A.G. Sabnis et J.T. Clemens. *Characterization of the electron mobility in the inverted <100> Si surface*. In Electron Devices Meeting, 1979 International, volume 25, pages 18–21, 1979.

[Saint-Martin04] J. Saint-Martin, A. Bournel et P. Dollfus. *On the ballistic transport in nanometer-scaled DG MOSFETs*. IEEE Transactions on Electron Devices, vol. 51, no. 7, pages 1148 – 55, 2004.

[Saito03] S. Saito, D. Hisamoto, S. Kimura et M. Hiratani. *Unified mobility model for high-k gate stacks [MISFETs]*. International Electron Devices Meeting (IEDM). Technical Digest, pages 33 – 3, Washington DC, USA, 2003.

[Samavedam03] S.B. Samavedam, L.B. La, P.J. Tobin, B. White, C. Hobbs, L.R.C. Fonseca, A.A. Demkov, J. Schaeffer, E. Luckowski, A. Martinez, M. Raymond, D. Triyoso, D. Roan, V. Dhandapani, R. Garcia, S.G.H. Anderson, K. Moore, H.H. Tseng, C. Capasso, O. Adetutu, D.C. Gilmer, W.J. Taylor, R. Hegde et J. Grant. *Fermi level pinning with sub-monolayer MeOx and metal gates [MOSFETs]*. International Electron Devices Meeting (IEDM). Technical Digest, pages 13 – 1, Washington DC, USA, 2003.

[Schmitz03a] J. Schmitz, F. N. Cubaynes, R. J. Havens, R. de Kort, A. J. Scholten et L. F. Tiemeijer. *RF Capacitance-Voltage Characterization of MOSFETs With High Leakage Dielectrics*. IEEE Electron Device Letters, vol. 24, pages 37–39, Janvier 2003.

[Schmitz03b] J. Schmitz, F.N. Cubaynes, R.J. Havens, R. de Kort, A.J. Scholten et L.F. Tiemeijer. *Test structure design considerations for RF-CV measurements on leaky dielectrics*. IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 181 – 5, Monterey, CA, USA, 2003.

[Schmitz04a] J. Schmitz, F.N. Cubaynes, R. de Kort, R. Havens, A.J. Scholten et L.F. Tiemeijer. *The RF-CV method for characterization of leaky gate dielectrics*. Microelectronic Engineering, volume 72, pages 149 – 53, Netherlands, 2004.

[Schmitz04b] J. Schmitz, F.N. Cubaynes, R.J. Havens, R. de Kort, A.J. Scholten et L.F. Tiemeijer. *Test structure design considerations for RF-CV measurements on leaky dielectrics*. IEEE Transactions on Semiconductor Manufacturing, vol. 17, no. 2, pages 150 – 4, 2004.

[Scholten01] A.J. Scholten, R. Duffy, R. van Langevelde et D.B.M. Klaassen. *Compact modelling of pocket-implanted MOSFETs*. In Proceedings of the 31th European Solid-State Device Research Conference (ESSDERC), pages 311–314, Nuremberg, Germany, Septembre 2001.

[Sclar56] N. Sclar. *Neutral Impurity Scattering in Semiconductors*. Phys. Rev., vol. 104, no. 6, pages 1559–1561, Dec 1956.

[Seeger91] K. Seeger. *Semiconductors Physics : an introduction (fifth edition)*. Springer-Verlag. 1991.

[Selvakumar82] C.R. Selvakumar. *Approximations to Fermi-Dirac integrals and their use in device analysis*. Proceedings of the IEEE, vol. 70, no. 5, pages 516 – 18, 1982.

[Severi06] S. Severi, G. Curatola, C. Kerner et K. De Meyer. *Accurate channel length extraction by split C-V measurements on short-channel MOSFETs*. IEEE Electron Device Letters, vol. 27, no. 7, pages 615–618, July 2006.

[Shahidi88] G.G. Shahidi, D.A. Antoniadis et H.I. Smith. *Electron velocity overshoot at room and liquid nitrogen temperatures in silicon inversion layers*. IEEE Electron Device Letters, vol. 9, no. 2, pages 94 – 6, 1988.

[Sheu84] B.J. Sheu et P. K. Ko. *A capacitance method to determine channel lengths for conventional and LDD MOSFET's*. IEEE Electron Device Letters, vol. 5, pages 491–493, Novembre 1984.

[Shockley51] W. Shockley. *Hot electrons in germanium crystal and Ohm's law*. Bell System Technical Journal, vol. 30 (10), pages 990–1034, 1951.

[Shockley62] W. Shockley. *Diffusion and Drift of Minority Carriers in Semiconductors for Comparable Capture and Scattering Mean Free Paths.* Phys. Rev., vol. 125, no. 5, pages 1570–1576, Mar 1962.

[Shrivastava82] R. Shrivastava et K. Fitzpatrick. *A simple model for the overlap capacitance of a VLSI MOS device.* IEEE Transactions on Electron Devices, vol. ED-29, no. 12, pages 1870 – 5, 1982.

[Shur02] M.S. Shur. *Low ballistic mobility in submicron HEMTs.* IEEE Electron Device Letters, vol. 23, no. 9, pages 511–513, Sep 2002.

[Skotnicki00] T. Skotnicki. *Heading for decananometer CMOS - is navigation among icebergs still a viable strategy ?* In Proceedings of the 30th European Solid-State Device Research Conference (ESSDERC), pages 19 – 33, Gif-sur-Yvette, France, 2000.

[Skotnicki03] T Skotnicki et F Boeuf. T Skotnicki et F Boeuf, chapitre Physique des dispositifs pour circuits intégrés sur silicium. EGEM, Encyclopédie Hermès, 2003.

[Skotnicki08] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Boeuf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J.-P. Schoellkopf, E. Perea, R. Ferrant et H. Mingam. *Innovative materials, devices, and CMOS technologies for low-power mobile multimedia.* IEEE Transactions on Electron Devices, vol. 55, no. 1, pages 96 – 130, 2008.

[Slotboom87] J.W. Slotboom, G. Streutker, G.J.T. Davids et P.B. Hartog. *Surface impact ionization in silicon devices.* International Electron Devices Meeting (IEDM). Technical Digest, pages 494 – 7, New York, NY, USA, 1987.

[Sobhan85] M.A. Sobhan et S. NoorMohammad. *Approximation for the Fermi-Dirac integral with applications to the modeling of charge transport in heavily doped semiconductors.* Journal of Applied Physics, vol. 58, no. 7, pages 2634 – 7, 1985.

[Sodini82] C.G. Sodini, T.W. Ekstedt et J.L. Moll. *Charge accumulation and mobility in thin dielectric MOS transistors.* Solid-State Electronics, vol. 25, no. 9, pages 833 – 41, 1982.

[Sodini84] C.G. Sodini, P.-K. Ko et J.L. Moll. *The effect of high fields on MOS device and circuit performance.* IEEE Transactions on Electron Devices, vol. ED-31, no. 10, pages 1386 – 93, 1984.

[Spinelli02] A.S. Spinelli, R. Clerc et G. Ghibaudo. *An analytical model for flat-band polysilicon quantization in MOS devices.* IEEE Transactions on Electron Devices, vol. 49, no. 7, pages 1314 – 16, 2002.

[Spinelli98] A.S. Spinelli, A. Benvenuti et A. Pacelli. *Self-consistent 2-D model for quantum effects in n-MOS transistors.* IEEE Transactions on Electron Devices, vol. 45, no. 6, pages 1342 – 9, 1998.

[Stern67] Frank Stern et W. E. Howard. *Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit.* Phys. Rev., vol. 163, no. 3, pages 816–835, Nov 1967.

[Stern72] F. Stern. *Self-consistent results for n-type Si inversion layers.* Physical Review B (Solid State), vol. 5, no. 12, pages 4891 – 9, 1972.

[Suk09] Sung Dae Suk, Ming Li, Yun Young Yeoh, Kyoung Hwan Yeo, Jae Kyu Ha, Hyunseok Lim, HyunWoo Park, Dong-Won Kim, TaeYoung Chung, Kyung Seok Oh et Won-Seong Lee. *Characteristics of sub 5 nm tri-gate nanowire MOSFETs with single and poly Si channels in SOI structure.* Symposium on VLSI Technology. Digest of Technical Papers, pages 142 – 3, Kyoto, Japan, 2009.

[Sun80] S.C. Sun et J.D. Plummer. *Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces.* IEEE Transactions on Electron Devices, vol. ED-27, no. 8, pages 1497 – 508, 1980.

[Suzuki99] K. Suzuki. *Parasitic capacitance of submicrometer MOSFET's.* IEEE Transactions on Electron Devices, vol. 46, no. 9, pages 1895–1900, Sep 1999.

[Sze81] S. M. Sze. *Physics of Semiconductor Devices.* John Wiley and Sons. 1981.

[Takagi08] S. Takagi, T. Iisawa, T. Tezuka, T. Numata, S. Nakaharai, N. Hirashita, Y. Moriyama, K. Usuda, E. Toyoda, S. Dissanayake, M. Shichijo, R. Nakane, S. Sugahara, M. Takenaka et

N. Sugiyama. *Carrier-transport-enhanced channel CMOS for improved power consumption and performance.* IEEE Transactions on Electron Devices, vol. 55, no. 1, pages 21 – 39, 2008.

[Takagi94a] S. Takagi, A. Toriumi, M. Iwase et H. Tango. *On the universality of inversion layer mobility in Si MOSFET's : Part I-effects of substrate impurity concentration.* IEEE Transactions on Electron Devices, vol. 41, no. 12, pages 2357–2362, Dec 1994.

[Takagi94b] S. Takagi, A. Toriumi, M. Iwase et H. Tango. *On the universality of inversion layer mobility in Si MOSFET's : Part II-effects of surface orientation.* IEEE Transactions on Electron Devices, vol. 41, no. 12, pages 2363–2368, Dec 1994.

[Takagi95] S. Takagi et A. Toriumi. *Quantitative understanding of inversion-layer capacitance in Si MOSFET's.* IEEE Transactions on Electron Devices, vol. 42, no. 12, pages 2125 – 30, 1995.

[Takeda81] Y. Takeda et T.P. Pearsall. *Failure of Matthiessen's rule in the calculation of carrier mobility and alloy scattering effects in Ga0.47In0.53As.* IEE Electronics Letters, vol. 17, no. 16, pages 573–574, 6 1981.

[Takeuchi96] K. Takeuchi, N. Kasai, T. Kunio et K. Terada. *An effective channel length determination method for LDD MOSFETs.* IEEE Transactions on Electron Devices, vol. 43, no. 4, pages 580 – 7, 1996.

[Tam84] S. Tam, P.K. Ko et C. Hu. *Lucky-electron model of channel hot-electron injection in MOSFETs.* IEEE Transactions on Electron Devices, vol. ED-31, no. 9, pages 1116 – 25, 1984.

[Tanaka05] Takuji Tanaka. *Novel Extraction Method for Size-Dependent Mobility Based on BSIM3-Like Compact Model.* Japanese Journal of Applied Physics, vol. 44, pages 2424–2427, 2005.

[Tanaka07] Takuji Tanaka. *Novel parameter extraction method for low field drain current of nano-scaled MOSFETs.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 265–267, Tokyo, Japan, Mars 2007.

[Tanaka95] S. Tanaka et M.S. Lundstrom. *A flux-based study of carrier transport in thin-base diodes and transistors.* IEEE Transactions on Electron Devices, vol. 42, no. 10, pages 1806–1815, Oct 1995.

[Taur00] Y. Taur. *MOSFET Channel Length : Extraction and Interpretation.* IEEE Electron Device Letters, vol. 47, pages 160–170, Janvier 2000.

[Taur92] Y. Taur, D.S. Zicherman, D.R. Lombardi, P.J. Restle, C.H. Hsu, H.I. Nanafi, M.R. Wordeman, B. Davari et G.G. Shahidi. *A new 'shift and ratio' method for MOSFET channel-length extraction.* IEEE Electron Device Letters, vol. 13, no. 5, pages 267 – 9, 1992.

[Taur98] Y. Taur et Tak H. Ning. *Fundamentals of Modem VLSI Devices.* Cambridge University Press. 1998.

[Thevenod05] L. Thevenod, M. Casse, M. Mouis, G. Reimbold, F. Fillot, B. Guillaumot et F. Boulanger. *Influence of TiN metal gate on Si/SiO2 surface roughness in N and PMOSFETs.* Microelectronic Engineering, vol. 80, pages 11 – 14, 2005.

[Thurgate85] T. Thurgate et N. Chan. *An impact ionization model for two-dimensional device simulation.* IEEE Transactions on Electron Devices, vol. ED-32, no. 2, pages 400 – 4, 1985.

[Tsai07] Y.J. Tsai, S.S. Chung, P.W. Liu, C.H. Tsai, Y.H. Lin, C.T. Tsai, G.H. Ma, S.C. Chien et S.W. Sun. *The channel backscattering characteristics of sub-100 nm CMOS devices with different channel/substrate orientations.* International Symposium on VLSI Technology Systems and Applications (VLSI-TSA). Proceedings, pages 154 – 5, Hsinchu, Taiwan, 2007.

[Tsuchiya06] Hideaki Tsuchiya, Kazuya Fujii, Takashi Mori et Tanroku Miyoshi. *A Picture of Quasi-Ballistic Transport in Nanoscale MOSFETs.* In Extended Abstracts of the 2006 International Conference on Solid State Devices and Materials, pages 350–351, Yokohama, 2006.

[Tsuno99] M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Miura-Mattausch et M. Hirose. *Physically-based threshold voltage determination for MOSFET's of all gate lengths.* IEEE Transactions on Electron Devices, vol. 46, no. 7, pages 1429 – 34, 1999.

[Uchida02] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata et S. Takagi. *Experimental study on carrier transport mechanism in ultrathin-body SOI nand p-MOSFETs with SOI thickness less than 5 nm.* International Electron Devices Meeting (IEDM). Technical Digest, pages 47 – 50, San Francisco, CA, USA, 2002.

[van Dort94] M.J. van Dort, P.H. Woerlee et A.J. Walker. *A simple model for quantisation effects in heavily-doped silicon MOSFETs at inversion conditions.* Solid-State Electronics, vol. 37, no. 3, pages 411 – 14, 1994.

[Van Halen85] P. Van Halen et D.L. Pulfrey. *Accurate, short series approximations to Fermi-Dirac integrals of order -1/2, 1/2, 1, 3/2, 2, 5/2, 3, and 7/2.* J. Appl. Phys. (USA), vol. 57, no. 12, pages 5271 – 4, 1985.

[Villa98] S. Villa, A.L. Lacaita, L.M. Perron et R. Bez. *A physically-based model of the effective mobility in heavily-doped n-MOSFETs.* IEEE Transactions on Electron Devices, vol. 45, no. 1, pages 110 – 15, 1998.

[Voelker05] M. Voelker et P. Fromherz. *Signal transmission from individual mammalian nerve cell to field-effect transistor.* Small, vol. 1, no. 2, pages 206 – 10, 2005.

[Wakabayashi03] H. Wakabayashi, S. Yamagami, N. Ikezawa, A. Ogura, M. Narihiro, K. Arai, Y. Ochiai, K. Takeuchi, T. Yamamoto et T. Mogami. *Sub-10-nm planar-bulk-CMOS devices using lateral junction control.* International Electron Devices Meeting (IEDM). Technical Digest, pages 20 – 7, Washington DC, USA, 2003.

[Wakita00] Naoki Wakita et Naoyuki Shigyo. *Verification of overlap and fringing capacitance models for MOSFETs.* Solid-State Electronics, vol. 44, no. 6, pages 1105 – 1109, 2000.

[Weber05] O. Weber. *Etude, Fabrication et Propriétés de Transport de Transistors CMOS associant un Diélectrique Haute Permittivité et un Canal de Conduction Haute Mobilité.* PhD thesis, 2005.

[Weber06] O. Weber, M. Casse, L. Thevenod, F. Ducroquet, T. Ernst et S. Deleonibus. *On the mobility in high-k/metal gate MOSFETs : Evaluation of the high-k phonon scattering impact.* Solid-State Electronics, vol. 50, no. 4, pages 626 – 31, 2006.

[Wei09] Lan Wei, F. Boeuf, T. Skotnicki et H.-S.P. Wong. *CMOS technology roadmap projection including parasitic effects.* International Symposium on VLSI Technology Systems and Applications (VLSI-TSA). Proceedings, pages 78 – 9, Hsinchu, Taiwan, 2009.

[Weisstein] Eric W. Weisstein. *Airy Functions.* From MathWorld–A Wolfram Web Resource.

[Werner63] H. Werner et G. Raymann. *An Approximation to the Fermi Integral F1/2(x).* Mathematics of Computation, vol. 17, no. 82, pages 193–194, 1963.

[Wiatr01] Maciej Wiatr, Peter Seegebrecht et Helge Peters. *Charge based modeling of the inner fringing capacitance of SOI-MOSFETs.* Solid-State Electronics, vol. 45, no. 4, pages 585 – 592, 2001.

[Wolf71] H. F. Wolf. *Semiconductors.* John Wiley & Sons. 1971.

[Wong01] J.S. Wong, J.G. Ma, K.S. Yeo et M.A. Do. *A New Approach for the Extraction of Threshold Voltage for MOSFET's.* In Technical Proceedings of the 2001 International Conference on Modeling and Simulation of Microsystems, volume 1, pages 534 – 537, 2001.

[Wong05] Man Wong et Xuejie Shi. *On the threshold voltage of metal-oxide-semiconductor field-effect transistors.* Solid-State Electronics, vol. 49, no. 7, pages 1179 – 84, 2005.

[Wong87] Hon-Sum Wong, M.H. White, T.J. Krutsick et R.V. Booth. *Modeling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFETs.* Solid-State Electronics, vol. 30, no. 9, pages 953 – 68, 1987.

[Yamaguchi98] K. Yamaguchi, H. Amishiro, M. Yamawaki et S. Asai. *A new variational method to determine effective channel length and series resistance of MOSFET's.* IEEE International Conference on Microelectronic Test Structures (ICMTS). Proceedings, pages 123 – 6, Kanazawa, Japan, 1998.

[Yan91] Z.X. Yan et M.J. Deen. *Physically-based method for measuring the threshold voltage of MOSFETs.* Circuits, Devices and Systems, IEE Proceedings G, vol. 138, no. 3, pages 351–357, Jun 1991.

[Yao86] C.T. Yao, I.A. Mack et H.C. Lin. *Accuracy of effective channel-length extraction using the capacitance method.* IEEE Electron Device Letters, vol. ED-7, no. 4, pages 268 – 70, 1986.

[Yu05] P. Yu et M. Cardona. *Fundamentals of Semiconductors.* Springer. 2005.

[Zhang08] Li Zhang, H. Tanimoto, K. Adachi et A. Nishiyama. *1-nm spatial resolution in carrier profiling of ultrashallow junctions by scanning spreading resistance microscopy.* IEEE Electron Device Letters, vol. 29, no. 7, pages 799 – 801, 2008.

[Zhu04] Wenjuan Zhu, Jin-Ping Han et T.P. Ma. *Mobility measurement and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics*. IEEE Transactions on Electron Devices, vol. 51, no. 1, pages 98 – 105, 2004.

[Zilli07a] M. Zilli, D. Esseni, P. Palestri et L. Selmi. *On the apparent mobility in nanometric n-MOSFETs*. IEEE Electron Device Letters, vol. 28, no. 11, pages 1036 – 9, 2007.

[Zilli07b] M. Zilli, P. Palestri, D. Esseni et L. Selmi. *On the experimental determination of channel back-scattering in nanoMOSFETs*. International Electron Devices Meeting (IEDM). Technical Digest, pages 105 – 8, Washington DC, USA, 2007.

Publications de l'auteur

D. Fleury, A. Cros, K. Romanjek, D. Roy, F. Perrier, B. Dumont et H. Brut
Automatic extraction methodology for accurate measurement of effective channel length on 65nm MOSFET technology and below, International Conference on Microelectronics Test Structures (ICMTS), 2007, Tokyo (Japan)

D. Fleury, A. Cros, H. Brut et G. Ghibaudo
New Y -Function-Based Methodology for Accurate Extraction of Electrical Parameters on Nano-Scaled MOSFETs, International Conference on Microelectronics Test Structures (ICMTS), 2008, Edinburgh (UK)

D. Fleury, A. Cros, K. Romanjek, D. Roy, B. Dumont, H. Brut and G. Ghibaudo
Automatic extraction methodology for accurate measurements of effective channel length on 65-nm MOSFET technology and below, IEEE Transaction on Semiconductor Manufacturing (T-SM) v.21, n.4, Nov. 2008

D. Fleury, A. Cros, G. Bidal, H. Brut, E. Josse and G. Ghibaudo
A New Technique to Extract the Gate Bias Dependent S/D Series Resistance of Sub-100nm MOSFETs, International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA), 2009, Hsinchu (Taiwan) — **Récompense du meilleur papier étudiant de l'année 2009**

D. Fleury, G. Bidal, A. Cros, F. Boeuf, T. Skotnicki and G. Ghibaudo
New Experimental Insight into Ballisticity of Transport in Strained Bulk MOSFETs, Symposium on VLSI Technology, 2009, Kyoto (Japan)

D. Fleury, A. Cros, G. Bidal, J. Rosa and G. Ghibaudo
A New Technique to Extract the Gate Bias Dependent S/D Series Resistance of Sub-100nm MOSFETs, IEEE Electron Device Letters (EDL), v.30, N.9, Sept. 2009

D. Fleury, G. Bidal, A. Cros, F. Boeuf, T. Skotnicki and G. Ghibaudo
Etude expérimentale de la balisticité du transport dans les transistors nMOS contraints sur silicium massif, GdR Nano, 2009, Paris-Orsay (France)

A. Cros, K. Romanjek, **D. Fleury** et al.
Unexpected mobility degradation for very short devices : a new challenge for CMOS scaling, International Electron Devices Meeting (IEDM) 2006, San Fransisco CA (USA)

R. Wacquez, R. Cerruti, P. Coronel, A. Cros, **D. Fleury** et al.
A Novel Self Aligned Design Adapted Gate All Around (SADAGAA) MOSFET including two stacked Channels : A High Co-Integration Potential, International Conference on Solid State Devices and Materials (SSDM) 2006, Yokohama (Japan)

D. Aimé, C. Fenouillet-Beranger, [...], **D. Fleury** et al.
Fully-depleted SOI CMOS technology using W_xN metal gate and $HfSi_xO_yN_z$ high-κ dielectric, European Solid-State Device Research Conference (ESSDERC) 2007, Munich (Germany)

A. Pouydebasque, S. Denorme, [...], **D. Fleury** et al.
High Performance High-K / Metal Planar Self-Aligned Gate-All-Around CMOS Devices for 32 nm Technologies and Beyond, Silicon Nanoelectronics Workshop (SNW) 2007, Kyoto (Japan)

M. Muller, C. Hobbs, [...], **D. Fleury** et al.
Highly Manufacturable and Cost-effective Single Ta_xC / $Hf_xZr_{(1-x)}O_2$ Gate CMOS Bulk Platform for LP Applications at the 45nm Node and Beyond, International Conference on Solid State Devices and Materials (SSDM) 2007, Ibaraki (Japan)

G. Bidal, F. Boeuf, [...], **D. Fleury** et al.
Planar bulk+ technology using TiN/Hf-based gate stack for low power applications, Symposium on VLSI Technology, 2008, Hawaii (USA)

G. Bidal, N. Loubet, [...], **D. Fleury** et al.
Folded fully depleted Bulk+ technology as a highly W-scaled planar solution, European Solid-State Device Research Conference (ESSDERC) 2008, Edinburgh (UK)

G.Bidal, F. Boeuf, F. Payet, [...], **D. Fleury** et al.
Pushing Bulk Transistor with Conventional SiON Gate Oxide for Low Power Applications, International Conference on Solid State Devices and Materials (SSDM) 2008, Ibaraki (Japan)

G. Ghibaudo, M. Mouis, [...], **D. Fleury** et al.
Electrical transport characterization of nanoCMOS devices with ultra-thin silicon film, International Workshop on Junction Technology (IWJT) 2009, Kyoto (Japan)

G. Bidal, **D.Fleury**, G. Ghibaudo, F. Boeuf and T. Skotnicki
Guidelines for MOSFET Device Optimization accounting for L-dependent mobility degradation, Silicon Nanoelectronics Workshop (SNW) 2009, Kyoto (Japan)

G. Bidal, F. Boeuf, [...], **D. Fleury** et al.
High Velocity Si-Nanodot : A Candidate for SRAM Applications at 16nm Node and Below Symposium on VLSI Technology 2009, Kyoto (Japan)

G. Bidal, F. Boeuf, [...], **D. Fleury** et al.
First CMOS integration of Ultra Thin Body and BOX (UTB2) structures on Bulk Direct Silicon Bonded (DSB) wafer with multi surface orientations, International Electron Device Meeting, 2009, Baltimore MD (USA)

Contribution à l'étude expérimentale du transport dans les transistors de dimensions déca-nanométriques des technologies CMOS sub-45nm

Résumé : La miniaturisation des composants électroniques qui permet aujourd'hui une intégration à grande échelle a été possible grâce aux innovations des procédés de fabrication. Ces modifications affectent profondément le comportement électrique des transistors MOS lorsque la longueur de grille devient inférieure à 100nm, altérant notre compréhension physique de ce dispositif. Ce travail de thèse se situe dans le domaine de l'étude des performances des transistors fabriqués dans les filières avancées (technologies sub-45nm) et l'analyse de leur réponse électrique. Il propose d'améliorer les méthodologies existantes et apporte de nouvelles techniques d'extraction qui permettent une analyse des paramètres électriques valide dans un environnement industriel, sur des transistors courts. L'utilisation des ces nouvelles techniques permet une compréhension physique plus juste, utile pour prédire les performances des technologies futures.

Mots clés : transistor MOS, caractérisation électrique, méthodologie d'extraction, longueur effective, mobilité, transport balistique, dérive-diffusion, fonction-Y, performance.

Contribution to the experimental study of transport in decananometric transistors of sub-45nm CMOS technologies

Abstract: The downscaling of electronic devices which allows a large-scale integration has been feasible thanks to many innovations regarding the fabrication processes. These changes deeply modify the electrical behavior of MOS transistors when the gate length becomes shorter than 100nm, altering the physical understanding of this device. This work deals with the study about advanced devices performances (sub-45nm technologies) and the analyze of electrical characteristics. Improvements of state-of-the-art methodologies and new extraction techniques are proposed for enabling the analysis of electrical parameters to be adapted to an industrial context, on very short devices. The use of these new techniques provides a better physical understanding which is required to predict the performances of future technologies.

Keywords: MOS transistor, electrical characterization, extraction methodology, effective channel length, ballistic transport, drift-diffusion, Y-function, performances.

contact: dominique.fleury@ieee.org