HAL
open science

Etude et modélisation du comportement électrique des transistors MOS fortement submicroniques

Fabien Prégaldiny

To cite this version:

Fabien Prégaldiny. Etude et modélisation du comportement électrique des transistors MOS fortement submicroniques. Micro et nanotechnologies/Microélectronique. Université Louis Pasteur - Strasbourg I, 2003. Français. NNT : . tel-00004312

HAL Id: tel-00004312

https://theses.hal.science/tel-00004312v1

Submitted on 26 Jan 2004

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.

N° d'ordre : 4460

École Doctorale Sciences Pour l'Ingénieur

Université Louis Pasteur

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université Louis Pasteur – Strasbourg I
Discipline : Sciences pour l'Ingénieur
Spécialité : Microélectronique

par

Fabien Prégaldiny

Étude et modélisation du comportement électrique des transistors MOS fortement submicroniques

Soutenue publiquement le 12 décembre 2003

Membres du jury

Président, rapporteur interne :	M. Jean-Louis Balladore, Professeur, ULP
Directeur de thèse :	M. Daniel Mathiot, Professeur, ULP
Co-Directeur de thèse :	M. Christophe Lallement, Professeur, ULP
Rapporteur externe :	M. Pierre Gentil, Professeur, INPG
Rapporteur externe :	M. Christian C. Enz, Professeur, EPFL
Examinateur :	M. Jean-Michel Sallese, Docteur, EPFL

Laboratoire PHASE **UPR 292**

Remerciements

En premier lieu, un grand merci à mes deux directeurs de thèse, Daniel Mathiot, directeur du laboratoire PHASE, et Christophe Lallement, pour m'avoir offert l'opportunité d'effectuer cette thèse au sein de leur équipe. J'ai beaucoup profité de la rigueur scientifique de Daniel, de son sérieux et de son esprit physique. Quant à Christophe, sa longue expérience du domaine de la modélisation compacte et son extrême rigueur scientifique m'ont permis de m'investir sereinement et efficacement dans ce travail de thèse.

Je tiens à remercier chaleureusement Pierre Gentil, Christian C. Enz, Jean-Louis Balladore et Jean-Michel Sallese d'avoir accepté d'être rapporteur de mon travail.

Ma reconnaissance va également à Jean-Baptiste Kammerer, tout d'abord pour son amitié, et pour ses compétences en langage de haut niveau. Je remercie aussi chaleureusement Ronald van Langevelde et Wladek Grabinski, sans qui l'accomplissement de cette thèse n'aurait certainement pas été possible.

Mes remerciements s'adressent également à mes collègues du laboratoire PHASE, Alfred Goltzené, Anne-Sophie Cordan, Yann Leroy, Yannick Hervé, Thomas Heiser et Sébastien Snaidero entre autres. Je tiens à remercier tout particulièrement Yann Leroy, pour son aide précieuse en informatique, et Martine Brutt pour sa gentillesse et son efficacité.

Avant de conclure, je remercierai une fois encore Christophe Lallement, mais cette fois-ci dans un contexte plus personnel. Sa vision poétique du monde et notre passion commune des plaisirs de la vie ont en effet largement contribué à égayer notre quotidien, et en ce sens, à rendre possible le bon déroulement de cette thèse.

Je terminerai par une mention spéciale pour ma petite Estelle, sans qui je n'aurais jamais eu l'idée de faire cette thèse…

Table des matières

1 Introduction **1**
1.1 Évolution de la technologie CMOS 1
1.2 Contraintes pour les générations futures 2
1.3 La modélisation du transistor MOS 5
1.4 Objectifs de la thèse et présentation du manuscrit 7
Bibliographie 10

2 Les modèles compacts **11**
2.1 Introduction 11
2.2 Approche physique 14
2.2.1 Le fonctionnement du transistor MOS 14
2.2.2 Définition du potentiel de surface 17
2.2.3 Modélisation en feuille de charge 19
2.2.3.1 Courant de drain 20
2.2.3.2 Approximation de la feuille de charge 21
2.2.3.3 Capacités et paramètres petits signaux 25
2.2.4 Points essentiels de l'approche physique 27
2.3 Approche classique 28
2.3.1 Les modèles SPICE de première génération 29
2.3.2 Les modèles de deuxième génération 31
2.3.3 Les modèles de troisième génération 32
2.4 Approches alternatives 35
2.4.1 Le modèle EKV 35
2.4.1.1 Le modèle originel 35
2.4.1.2 La version 2.6 37
2.4.1.3 Le nouveau formalisme 37
2.4.2 Le modèle MM11 38
2.4.3 Le modèle SP 39
2.4.4 Synthèse des caractéristiques des modèles avancés 40
2.5 Conclusion 41
Bibliographie 43

3 Aspects extrinsèques du transistor MOS **47**
3.1 Introduction 47
3.2 Présentation du dispositif MOSFET 48

3.2.1 La structure MOSFET 48
3.2.2 Le régime extrinsèque 49
3.2.3 Les capacités parasites 50
3.3 Simulations numériques 2D 52
3.3.1 Mise en évidence de la capacité extrinsèque 53
3.3.2 Influence des différentes composantes de la capacité extrinsèque 56
3.3.2.1 La capacité d'overlap 56
3.3.2.2 La capacité de bord interne 57
3.3.2.3 Résultats de simulations 2D 58
3.4 Modélisation de la capacité extrinsèque 64
3.4.1 Nécessité d'un nouveau modèle 64
3.4.2 Formulation du nouveau modèle 64
3.4.2.1 Capacité d'overlap 65
3.4.2.2 Capacité de bord externe 66
3.4.2.3 Capacité de bord interne 67
3.4.2.4 Modèle complet 71
3.4.3 Résultats 71
3.4.4 Extension du modèle 75
3.5 Conclusion 77
Bibliographie 79

4 Modélisation analytique des effets de mécanique quantique 81
4.1 Introduction 81
4.1.1 Signification du terme effets quantiques 81
4.1.2 Origine physique du confinement quantique 81
4.2 Influence des effets quantiques 83
4.2.1 Régime d'inversion 83
4.2.2 Régime d'accumulation 85
4.3 Modélisation physique 86
4.4 Méthodes d'approximations analytiques 90
4.4.1 L'approximation variationnelle 90
4.4.2 L'approximation du puits de potentiel triangulaire 92
4.5 État de l'art 93
4.5.1 Travaux pionniers 94
4.5.2 Approches en tension de seuil 101
4.5.3 Approches en potentiel de surface 104
4.5.3.1 Le modèle SP 106
4.5.3.2 Le modèle MM11 109
4.5.4 Approche alternative : le modèle EKV 111
4.5.5 Bilan et intérêt d'un nouveau modèle 113
4.6 Formulation du nouveau modèle : déplétion et inversion 115
4.6.1 Modélisation explicite quantique du potentiel de surface 116
4.6.1.1 Modèle explicite classique 116
4.6.1.2 Modèle explicite quantique :
Approximation de l'inversion modérée 121

4.6.2 Modélisation analytique du phénomène de polydéplétion 133
4.7 Validation du modèle : déplétion et inversion 138
4.7.1 Modélisation des charges et des capacités 138
4.7.2 Modélisation du courant de drain 141
4.7.3 Conclusion 148
4.8 Formulation du nouveau modèle : accumulation 149
4.8.1 Modèle explicite classique 149
4.8.2 Modèle explicite quantique 153
4.9 Modèle explicite quantique complet 155
4.10 Validation du modèle complet 156
4.11 Conclusion 162
Bibliographie 165

5 Conclusion et perspectives 169
5.1 Conclusion 169
5.2 Perspectives 174

A Extraction de la tension de seuil à partir des caractéristiques *C–V* 175
Bibliographie 177

B Code VHDL-AMS du modèle de potentiel de surface quantique 179

Publications associées à ce travail 185

Liste des tableaux

1.1 Prévisions SIA de l'évolution de la technologie CMOS 3

2.1 Liste des paramètres du modèle SPICE Level 1 30
2.2 Caractéristiques des principaux modèles compacts actuellement utilisés 40

3.1 Paramètres technologiques des TMOS LDD conçus en simulation de procédés . 52
3.2 Longueur de diffusion latérale en fonction de la dose LDD 60
3.3 Principaux paramètres du modèle de la capacité de bord interne 70

4.1 Différentes utilisations du modèle drift–diffusion de courant de drain 144
4.2 Influence des effets quantiques et de la polydéplétion sur le calcul du courant de drain et de la transcapacité de grille 146

Table des figures

1.1 Évolution de la technologie CMOS, loi de Moore 2
1.2 Évolution des grandeurs caractéristiques de la technologie CMOS 4

2.1 Développement des modèles compacts, nombre de paramètres 13
2.2 Structure basique d'un transistor MOS de type n 15
2.3 Distribution des charges et diagrammes des bandes d'énergie correspondants, dans un transistor n-MOS 16
2.4 Diagramme des bandes d'énergie et grandeurs électriques associées dans le cas d'un transistor n-MOS en régime d'inversion 17
2.5 Courant de drain calculé dans le cadre de l'approximation drift–diffusion . . . 24
2.6 Courant de drain obtenu avec un modèle compact classique de troisième génération 34

3.1 Structure n-MOSFET LDD et caractéristiques technologiques associées 49
3.2 Courbes $C_{dg} + C_{sg}$ simulées en fonction de la tension de grille 54
3.3 Schéma illustrant les différentes composantes de la capacité extrinsèque 55
3.4 Capacité extrinsèque normalisée en fonction de la tension de grille 59
3.5 Influence de la longueur de grille sur la capacité extrinsèque normalisée 61
3.6 Influence de la polarisation du drain sur la capacité extrinsèque normalisée . . . 62
3.7 Influence de la polydéplétion sur le comportement capacitif global 63
3.8 Modélisation analytique de la capacité de bord interne 70
3.9 Comparaison entre le nouveau modèle extrinsèque et la simulation 2D 72
3.10 Paramètre λ en fonction de la longueur de diffusion latérale L_d 73
3.11 Modélisation complète ($\forall\ V_{gb}$) et simulation 2D de la capacité $C_{dg} + C_{sg}$ 75
3.12 Capacité extrinsèque vs. la tension de grille, à tension de drain fixée 76
3.13 Capacité extrinsèque vs. la tension de drain, à tension de grille fixée 77

4.1 Diagramme des bandes d'énergie et distribution des porteurs d'un transistor n-MOS, en tenant compte des effets quantiques 82
4.2 Densité de charge d'inversion $|Q_{inv}|$ en fonction de la tension de grille V_{gb} . . . 84
4.3 Potentiel de surface ϕ_s en fonction de la tension de grille V_{gb} 85
4.4 Tensions de seuil quantique et classique, en fonction du dopage substrat 98
4.5 Modification du modèle de tension de seuil quantique de van Dort 100
4.6 Comparaison d'approches quantiques pour simuler le courant de drain 102
4.7 Potentiel de surface (calcul implicite) en fonction de la tension de grille, en régime d'inversion faible 117

4.8 Potentiel de surface (calcul implicite) en fonction de la tension de grille, en régime d'inversion forte . . . 118
4.9 Calcul explicite du potentiel de surface, de la déplétion à l'inversion forte . . . 120
4.10 Calcul explicite du potentiel de surface, avec le modèle de Clerc et al. . . . 124
4.11 Potentiel de surface calculé avec le nouveau modèle analytique (1/2) . . . 129
4.12 Potentiel de surface calculé avec le nouveau modèle analytique (2/2) . . . 130
4.13 Potentiel de surface en fonction des tensions de grille et de drain . . . 132
4.14 Influence respective du dopage substrat et de l'épaisseur d'oxyde de grille sur le potentiel de surface . . . 132
4.15 Calcul du potentiel de surface pour des technologies futures . . . 133
4.16 Schéma illustrant le phénomène de polydéplétion . . . 134
4.17 Influence de la polydéplétion sur la tension de grille effective . . . 137
4.18 Comparaison du nouveau modèle avec des mesures capacitives (*C*–*V*) . . . 140
4.19 Densité de charge d'inversion en fonction de la tension de grille . . . 142
4.20 Simulation du courant de drain avec le modèle quantique analytique . . . 146
4.21 Validation expérimentale du modèle de courant de drain (Philips) . . . 147
4.22 Influence des effets quantiques sur la transconductance g_m . . . 148
4.23 Validation expérimentale du modèle de courant de drain (Motorola) . . . 148
4.24 Impact de l'utilisation d'une fonction de lissage sur le potentiel de surface en inversion faible . . . 152
4.25 Calcul explicite du potentiel de surface, de la déplétion à l'inversion faible . . . 152
4.26 Calcul explicite du potentiel de surface, de l'accumulation à l'inversion forte . . . 157
4.27 Simulation analytique de la transcapacité de grille en fonction de la tension de grille, de l'accumulation à l'inversion forte . . . 158
4.28 Transcapacité de grille en fonction de V_{gb}, N_a et t_{ox} . . . 160
4.29 Comparaison expérimentale du nouveau modèle avec des mesures *C*–*V* . . . 161

A.1 Principe de notre méthode d'extraction de la tension de seuil . . . 176
A.2 Réseau de caractéristiques *C*–*V* correspondant à l'expérience de van Dort . . . 176
A.3 Validation de notre méthode d'extraction de la tension de seuil . . . 177

Chapitre 1

Introduction

1.1 Évolution de la technologie CMOS

Le développement de la microélectronique depuis ces 30 dernières années est véritablement spectaculaire. Ce succès résulte en grande partie d'un savoir-faire et d'une maîtrise technologique de plus en plus poussés de l'élément fondamental de la microélectronique : le silicium. Le transistor MOS (Metal Oxide Semiconductor) est à la fois le principal acteur et le vecteur de cette évolution technologique. Il est la base de la conception des circuits intégrés à très large et ultra large échelle (VLSI–ULSI), et a mené la technologie CMOS (Complementary MOS) au rang incontesté de technologie dominante de l'industrie du semi-conducteur. Au fil des années, la complexité des circuits intégrés a augmenté de façon continue, principalement grâce aux performances accrues des nouvelles générations de transistors MOS (TMOS). La réduction constante des dimensions de ces composants est le moteur de cette course à la performance ; en fait, c'est cette volonté de toujours réduire la taille des transistors MOS qui a entraînée toute l'industrie du semi-conducteur à se surpasser et à se projeter en permanence dans le futur.

En 1973, G. MOORE, l'un des co-fondateurs d'Intel avait observé que le nombre de transistors intégrés sur une même puce doublait tous les 18 mois. Cette observation l'avait alors conduit à prédire que le nombre de transistors intégrés sur une puce continuerait à doubler tous les 18 mois, jusqu'à ce que les limites physiques soient atteintes. La véracité de sa prédiction durant ces 30 dernières années a été telle que l'on s'y réfère maintenant en tant que « Loi de Moore ». La Fig. 1.1 illustre la validité de cette prévision, pourtant originellement dérivée d'un simple constat. Aujourd'hui, des circuits intégrés (IC) comprenant plus de 40 millions de transistors sont produits de façon industrielle (microprocesseurs). La longueur de grille des

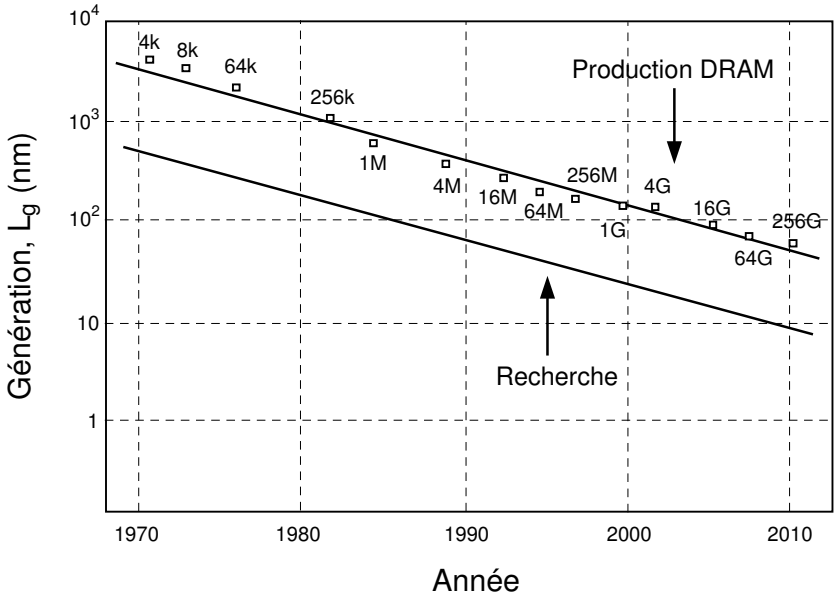


FIG. 1.1 : *Réduction d'échelle de la technologie CMOS, en accord avec la loi de Moore.*

TMOS utilisés pour ces dernières générations de microprocesseurs est égale à 0.13 μm, tandis que la surface de la puce varie de 80 à 150 mm^2. En fait, la diminution de longueur de grille des dispositifs a deux avantages décisifs pour les fabricants : d'une part, à puissance égale elle permet de réduire la surface de silicium de la puce, ce qui en termes de coût est bénéfique, et d'autre part, elle permet d'augmenter la fréquence des circuits, cette dernière étant inversement proportionnelle à la longueur de grille.

1.2 Contraintes pour les générations futures

À chaque nouvelle génération de transistor, la réalisation du défi lancé par la loi de Moore apparaît comme un casse-tête de plus en plus difficile à réaliser. Un compromis complexe entre la physique, la technologie et la rentabilité concentre ainsi toute l'attention des ingénieurs et des chercheurs. Des paramètres et contraintes souvent contradictoires, tels que la performance, la consommation et la fiabilité sont à prendre en compte [1–3]. Pour résumer, disons que le jeu consiste à augmenter les performances en diminuant les dimensions, sans trop augmenter la puissance dissipée à l'état bloqué du transistor.

Parier sur une croissance au rythme de la loi de Moore pour la décennie à venir relève du défi ambitieux. De plus, les architectures devenant très complexes, la conception, la fabrication

et la vérification voient leurs coûts croître exponentiellement. Il est actuellement admis que la Loi de Moore sera encore valide pour les 10–12 ans à venir i.e. pour 3 à 4 générations de microprocesseurs. En effet, les projections industrielles pour le développement de la technologie CMOS suggèrent que cette dernière est proche des limites fondamentales physiques. L'association de l'industrie du semi-conducteur : SIA (Semiconductor Industry Association), publie depuis 1998 « The International Technology Roadmap for Semiconductors, ITRS » qui est un guide de référence pour l'industrie mondiale du semi-conducteur [4] (voir Tableau 1.1).

TAB. 1.1 : *Prévisions SIA de l'évolution de la technologie CMOS [4].*

Année	**1999**	**2002**	**2005**	**2008**	**2011**	**2014**
L_g (nm)	180	130	100	70	50	35
V_{dd} (V)	1.5–1.8	1.2–1.5	0.9–1.2	0.6–0.9	0.5–0.6	0.3–0.6
V_{th} (V)	0.5	0.4	0.35	0.3	0.25	0.2
t_{ox} (nm)	1.9–2.5	1.5–1.9	1.0–1.5	0.8–1.2	0.6–0.8	0.5–0.6
N_a (cm^{-3})	$< 10^{18}$	$< 10^{18}$	10^{18}	10^{18}	10^{18}	10^{18}
X_j (nm)	45–70	30–50	25–40	20–28	13–20	10–14
E (MV/cm)	< 5	5	> 5	>5	> 5	> 5
I_{on} (μA/μm)	750/350	750/350	750/350	750/350	750/350	750/350
I_{off} (μA/μm)	5	10	20	40	80	160
$\varnothing$ du wafer	200	300	300	300	300	450

Selon l'édition 1999, malgré l'utilisation de nouveau matériel, il sera difficile de maintenir l'augmentation des performances électriques des composants au rythme de la loi de Moore. Il convient cependant de rappeler que les données du Tableau 1.1 sont basées sur de simples projections des progrès passés. Ceci ne garantit pas forcément qu'un dispositif plus court pourra être fabriqué, ni qu'il présentera les mêmes performances. Par exemple au niveau des DRAM (Fig. 1.1), les prévisions indiquent 256 Go en 2010, ce qui impliquerait de réduire la surface de silicium requise pour le stockage d'un simple bit au-delà de ce que les dernières innovations technologiques espèrent approcher.

La Fig. 1.2 illustre graphiquement l'évolution espérée des principales caractéristiques des TMOS, à savoir, la longueur de grille (L_g), la tension d'alimentation (V_{dd}), l'épaisseur d'oxyde de grille (t_{ox}) et les profondeurs de jonctions des extensions de source et drain (X_j). Une première analyse de ces valeurs permet d'annoncer quelques possibles limitations et freins

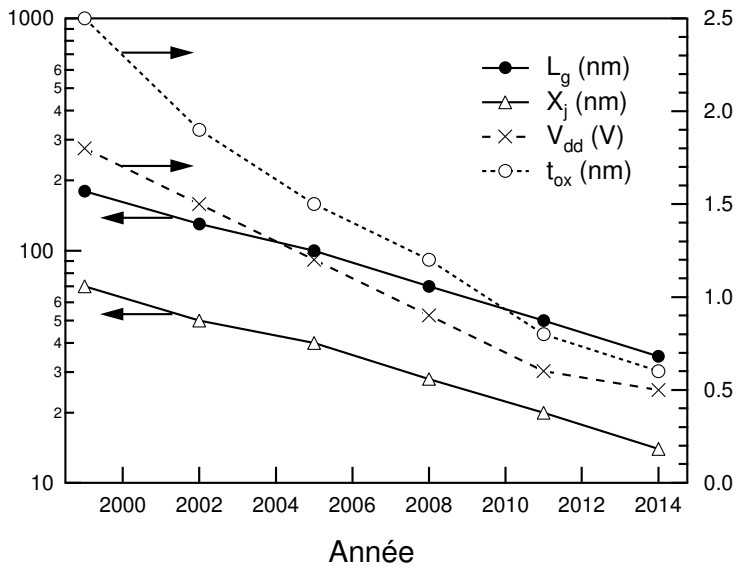


FIG. 1.2 : *Évolution des grandeurs caractéristiques de la technologie CMOS, selon les prévisions de la SIA [4].*

technologiques à la réduction d'échelle énoncée selon les critères de la SIA :

- La diminution de la longueur de grille en dessous de 50 nm semble difficile, compte tenu du contrôle nécessaire du courant de fuite à l'état bloqué du transistor.

- En raison de la réduction de la résistance du canal à l'état passant, il faut veiller à ce que les résistances source/drain, placées en série avec celle du canal, soient suffisamment faibles pour ne pas dégrader sérieusement les performances du composant. Cette contrainte impose donc de ne pas choisir des profondeurs de jonctions X_j trop faibles, et conduit à adopter un très fort dopage de source et de drain. Cela est cependant défavorable du point de vue des effets canaux courts car la réduction des profondeurs de jonctions source/drain permet en fait un meilleur contrôle de la charge du canal à l'état bloqué [5].

- La tension d'alimentation ne peut que difficilement être réduite en dessous de 0.6 V, en raison de la nécessité du maintien de la tension de seuil (V_{th}) à un niveau suffisant pour garantir des marges de bruit acceptables dans les circuits logiques [6].

- La réduction de l'épaisseur d'oxyde en dessous de 2 nm résulte en un important courant tunnel, or vu les épaisseurs annoncées (cf. Fig. 1.2) de sérieux problèmes risquent de

se poser au niveau de la consommation statique. Il est admis que pour une tension d'alimentation de 1 V, la limite maximale admise pour le courant de fuite de grille est de l'ordre de 1 A/cm^2, ce qui situe l'épaisseur minimale d'oxyde aux environs de 1.8 nm [7]. Cependant, on sait que ces courants de fuite ne perturberont pas le fonctionnement élémentaire des transistors MOS de longueur de canal inférieure à 1 μm, mais en revanche, augmenteront la puissance dissipée à l'état bloqué [8–10]. Par ailleurs, il est clair également que la réduction des dimensions ne peut se faire sans réduire l'oxyde de grille, sous peine de ne plus parvenir à contrôler les effets canaux courts [11].

En résumé, comme nous l'avons précédemment mentionné, la loi de Moore continuera d'être valide pour les 3 ou 4 prochaines générations de microprocesseurs. Le maintien de l'augmentation exponentielle du nombre de transistors deviendra cependant extrêmement difficile et coûteux. Le problème du coût risque d'ailleurs de devenir plus important que celui des limitations technologiques, puisque ce dernier dicte les différents choix d'investissement en recherche et développement. Cependant, bien qu'il soit admis que la loi de Moore ne durera pas indéfiniment, le fait que les ordinateurs continueront à devenir de plus en plus rapides et puissants sera probablement vrai pour de nombreuses années encore. Ainsi, il reste encore un travail important pour faire progresser la technologie CMOS jusqu'à son apogée. Cet effort de recherche sera bien-entendu multidisciplinaire. Deux axes complémentaires, sont et continueront à être essentiels au bon développement du transistor MOS. L'axe amont concerne les concepteurs du dispositif lui-même, c'est-à-dire les technologues ; l'axe aval regroupe les personnes de la modélisation, au sens large. Ce second axe est tout aussi important que le premier ; en effet, la modélisation précise des transistors MOS est le point clé de la validité de toute simulation de circuits intégrés, et donc de la conception de tout circuit.

1.3 La modélisation du transistor MOS

Les modèles de transistors décrivent le comportement du composant en termes de caractéristiques électriques, principalement courant-tension (*I–V*) et capacité-tension (*C–V*), ainsi que le processus de déplacement des porteurs dans le dispositif [12]. Ces modèles doivent donc refléter le comportement du composant dans toutes les régions de fonctionnement. Dans le cas du transistor MOS, un modèle doit alors être valide de l'accumulation à l'inversion. Deux principales catégories de modèles coexistent : les modèles physiques et les modèles compacts destinés à la simulation de circuits.

Les modèles physiques de dispositifs sont basés sur une définition rigoureuse des paramètres géométriques, des profils de dopage, des équations du transport des porteurs (*équations du semi-conducteur*) et des caractéristiques des matériaux utilisés. Ces modèles peuvent être utilisés pour prédire les caractéristiques électriques du TMOS, ainsi que les différents phénomènes liés au transport des porteurs. Leur utilisation nécessite une grande puissance de calcul — associée à des méthodes numériques performantes — en raison de la résolution complexe des équations couplées de la physique du semi-conducteur, dans un espace à 2 voire 3 dimensions. L'intérêt de ces modèles est qu'ils fournissent une compréhension détaillée de l'aspect physique du fonctionnement du TMOS, et qu'ils ont une réelle capacité à prédire les caractéristiques électriques des dispositifs futurs. Pour ces raisons, ils sont largement employés, soit pour étudier la physique et la conception du dispositif, soit comme moyen de validation, des modèles analytiques par exemple. Ils sont par contre inadaptés à la simulation de circuits intégrés car ils nécessitent un temps de calcul beaucoup trop élevé.

Du fait de la nature 2D voire 3D des effets physiques régissant le comportement électrique des transistors MOS, il est très difficile d'obtenir une formulation analytique qui soit valide dans tous les régimes de fonctionnement. Cependant, il est possible d'obtenir des formulations analytiques, basées sur la physique du dispositif, mais généralement limitées à un domaine particulier de fonctionnement. En dépit de cette limitation, ce sont de tels modèles qui sont utilisés pour la simulation de circuits, en raison de leur rapidité d'exécution. Les modèles de circuits équivalents — ou modèles compacts — décrivent les propriétés électriques du dispositif en connectant les éléments du circuit de façon à ce que le modèle fournisse le comportement électrique aux différents terminaux. Ces modèles reposent sur des expressions analytiques plus ou moins fondées sur la physique du semi-conducteur, et sur un degré d'empirisme variable.

Actuellement, il existe différentes approches de modélisation du TMOS, chacune reflétant une solution technique particulière mis en oeuvre lors du développement du modèle. Dans le cadre de cette thèse, nous nous intéresserons tout particulièrement aux familles de modèles appartenant au groupe des *modèles compacts*. Le chapitre 2 sera d'ailleurs entièrement consacré à leur étude. Un point fondamental est que l'utilité des simulateurs comme outils aidant à la conception et à l'analyse de circuits dépend directement de la qualité et de l'adéquation des modèles de composants utilisés dans le simulateur. Aussi, le choix d'un modèle adapté est une étape importante lors de la simulation d'un circuit donné ; ce choix doit être basé sur le compromis entre la capacité du modèle à prédire correctement les caractéristiques

électriques du dispositif, et son efficacité en termes de temps de calcul. Comme la taille et la complexité des circuits modernes augmentent, le choix d'un modèle approprié devient de plus en plus critique. Finalement, en raison de la nature évolutive de la technologie CMOS, le développement d'un modèle précis, efficace et adapté à la simulation de circuits ayant une très haute densité d'intégration est un défi permanent.

Pour conclure, voici quelques idées phares, que tout concepteur de modèles se doit d'avoir à l'esprit. Il faut tout d'abord être bien conscient que le monde du développement de la technologie et celui de la conception de circuits ont longtemps été séparés [13]. Il est pourtant clair qu'un modèle de TMOS est le véhicule privilégié pour l'échange d'informations entre le fondeur et le concepteur de circuits intégrés. Le rôle d'un modèle est donc multiple, et est souvent caractérisé par des impératifs contradictoires. Idéalement, un modèle de TMOS devrait avoir une structure simple, être efficace en termes de temps de calcul, fournir une grande précision, et avoir un nombre minimum de paramètres. De plus, un modèle idéal devrait aussi être raisonnablement prédictif, par exemple le changement d'une quantité physique telle que l'épaisseur de l'oxyde de grille, devrait résulter en un relativement bon changement des autres grandeurs électriques liées à cette dernière. Cet exemple, qui peut au premier abord sembler trivial, est pourtant l'un des grands défis de la modélisation compacte, à savoir la notion de prédictivité [14]. Aussi, au fil de ce manuscrit, dans le but de satisfaire au mieux cette notion de prédictivité, nous tenterons lors du développement de nos modèles, de mettre au maximum en exergue le trait d'union *paramètres technologiques–équations du modèle*.

1.4 Objectifs de la thèse et présentation du manuscrit

Cette thèse se positionne dans le cadre de la modélisation compacte du transistor MOS, et concerne plus particulièrement celle du TMOS fortement submicronique. Notre démarche consiste dans un premier temps à réaliser des études physiques de différents phénomènes perturbant le fonctionnement classique du TMOS, puis dans un second temps, à définir des modèles prenant en compte ces différents effets. D'un point de vue général, notre approche a pour but la réalisation de modèles purement analytiques mais fortement liés à la physique du composant, ce qui sous-entend bien entendu un degré d'empirisme réduit au minimum. Les besoins de la modélisation compacte étant multiples, il existe de nombreuses voies de recherche, distinctes, mais toujours complémentaires. Au cours de cette thèse, nous nous intéresserons à deux voies de recherche différentes, mais toutes intimement liées aux problèmes

et perturbations engendrées par la réduction d'échelle drastique des dispositifs. Précisons qu'en aucun cas l'ambition de ce travail ne consiste en la définition d'un nouveau modèle compact ; notre objectif est de développer de nouvelles équations facilement implémentables au sein des modèles existants, afin de les améliorer et d'augmenter leur contenu physique.

En tant que suite logique à ce chapitre d'introduction, le chapitre 2 sera consacré à un état de l'art des différents modèles compacts de TMOS utilisés en CAO (Conception Assistée par Ordinateur). Après à un bref historique, nous aborderons les différentes approches co-existantes. Nous commencerons par définir une approche physique a priori idéale, mais cependant non adaptée aux besoins réels de la modélisation compacte. Dans un second temps nous présenterons les différentes approches classiques, i.e. celles utilisées dans les modèles basés sur la formulation d'une *tension de seuil*. De nouveaux concepts de modélisation, principalement basés sur la formulation du *potentiel de surface* seront ensuite abordés. Enfin, nous définirons l'approche de modélisation choisie dans le contexte de cette thèse.

Le chapitre 3 discutera des aspects extrinsèques du TMOS. Nous nous intéresserons tout particulièrement à la prise en compte des phénomènes parasites négligés, ou du moins mal pris en compte dans les modèles existants. Il s'agit de ce que nous appelons les *phénomènes extrinsèques* qui perturbent le fonctionnement classique des transistors MOS. Avec la miniaturisation extrême des composants, le caractère extrinsèque affecte de plus en plus les caractéristiques du transistor, le cas des capacités parasites en régime sans courant (état bloqué du transistor) en est un exemple remarquable. Au cours de ce chapitre, nous présenterons un nouveau modèle analytique décrivant les différentes capacités parasites du TMOS. Un intérêt tout particulier de la modélisation réalisée est d'être immédiatement implémentable au sein des différents types de modèles compacts existants.

Le chapitre 4 sera consacré au développement d'un nouveau modèle incluant les effets de mécanique quantique au sein des structures TMOS. La nécessité de réaliser un tel modèle résulte de la réduction d'échelle continue des composants, qui induit de nouveaux choix technologiques. Dans les technologies CMOS modernes, l'utilisation conjointe de forts niveaux de dopage substrat (N_a) et de très faibles épaisseurs d'oxyde de grille (t_{ox}) conduit à une importante courbure des bandes d'énergie à l'interface Si–SiO_2. Le puits de potentiel résultant peut ainsi devenir suffisamment étroit pour quantifier le mouvement des porteurs dans la direction perpendiculaire à l'interface. De nos jours, un certain nombre de modèles compacts

analytiques prennent en compte les effets quantiques, avec plus ou moins de succès. Cependant, tous sont basés sur de simples *corrections quantiques* dont l'intégration au sein des modèles classiques repose sur des fondements physiques incertains. Par exemple, aucun modèle valide de l'accumulation à l'inversion n'a été développé en considérant les meilleures solutions analytiques disponibles dans chaque région. Dans ce chapitre, nous présenterons de nouveaux concepts permettant de modéliser qualitativement et quantitativement l'influence des effets quantiques sur le comportement électrique du TMOS. Précisons déjà que le modèle quantique développé ici est un modèle dit en *potentiel de surface*.

Le chapitre 5 dressera le bilan du travail présenté dans ce manuscrit. L'intérêt et les applications potentielles des modèles développés seront discutés. En particulier, nous conclurons par le fait qu'une approche simple et cohérente, en un mot pragmatique, de la modélisation compacte est réellement envisageable. Cette thèse n'étant pas, fort heureusement, une fin en soi, nous présenterons les perspectives liées à ce travail, envisageables à court et moyen termes.

Bibliographie

[1] Y. Taur *et al.*, "CMOS scaling into the nanometer regime," in *Proc. IEEE*, vol. 85, no. 4, pp. 486–504.

[2] C. Caillat, "Etude des dispositifs MOS de longueur de grille inférieure à 0.1 μm, vers les limites de l'intégration du CMOS sur silicium," Ph.D. dissertation, INPG, nov 1999.

[3] G. Timp *et al.*, "Progress toward 10nm CMOS devices," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, 1998, pp. 615–618.

[4] Semiconductor Industry Association, "The international roadmap for semiconductors," update 1999, http ://www.semichips.org.

[5] H.-S. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. J. Welser, "Nanoscale CMOS," in *Proc. IEEE*, vol. 87, no. 4, pp. 537–569.

[6] S. Thomson, P. Packan, and M. Bohr, "MOS scaling : Transistor challenges for the 21st century," 1998, Intel Corp., http ://www.chips.com/technology/ith/q31998/articles/.

[7] S.-H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's," *IEEE Electron Device Lett.*, vol. 18, no. 5, pp. 209–211, May 1997.

[8] H. S. Mosmose, T. Morimoto, Y. Osawa, K. Yamabe, and H. Iwai, "Electrical characteristics of rapid thermal nitrited-oxide gate n- and p- MOSFET's with less than 1 atom% nitrogen concentration," *IEEE Trans. Electron Devices*, vol. 41, pp. 546–552, 1994.

[9] H. S. Mosmose *et al.*, "1.5 nm direct-tunneling gate oxide Si MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, pp. 1233–1241, 1996.

[10] H. Iwai, "CMOS technology – Year 2010 and beyond," *IEEE Journ. of Solid-State Circuits*, vol. 34, no. 3, pp. 357–366, 1999.

[11] Y. Taur, C. H. Wann, and D. J. Frank, "25 nm CMOS design considerations," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, 1998, pp. 789–792.

[12] N. Arora, *MOSFET Models for VLSI Circuit Simulation. Theory and Practice*. New York : Springer-Verlag, 1993, ISBN 3-211-82395-6.

[13] M. Bucher, "Analytical MOS transistor modelling for analog circuit simulation," Ph.D. dissertation, EPFL, 1999, thèse no. 2114.

[14] C. Lallement, "Le transistor MOSFET : Etudes, modélisation, et applications dans les S.O.C." 2002, habilitation à Diriger des Recherches, ULP, Strasbourg.

Chapitre 2

Les modèles compacts

2.1 Introduction

La modélisation des transistors MOS pour la conception et la simulation de circuits est un défi constant en raison de l'évolution incessante de la technologie CMOS. La modélisation avec maintien du sens physique est un défi encore plus grand pour le monde de la conception et de la simulation des circuits analogiques. Dans ce cas, sont importantes, non seulement les caractéristiques *grand signal* du composant, mais aussi les caractéristiques *petit signal*. La modélisation du TMOS nécessite donc une compréhension claire et approfondie du fonctionnement complexe et fortement non-linéaire de ce composant.

Ce sont des modèles formulés de manière analytique qui sont utilisés le plus souvent pour aider à la conception de circuits analogiques et mixtes. Les modèles disponibles dans les simulateurs de circuits standards contiennent d'une part des expressions basées sur la physique et d'autre part un certain degré d'empirisme. Ces modèles peuvent être adaptés aux différentes technologies CMOS à l'aide d'un certain nombre de paramètres, dans le but de décrire correctement les caractéristiques électriques du composant. Afin de rendre pratique l'utilisation d'un modèle, ce dernier doit être complété par des méthodes d'extraction de paramètres. Un modèle de transistor MOS représente généralement un compromis entre les aspects de simplicité et de complexité, les notions physiques et empiriques, le nombre d'effets inclus, le nombre de paramètres, l'adaptabilité aux diverses technologies, et enfin, l'efficacité de calcul.

Différentes approches de modélisation du TMOS sont actuellement utilisées. La famille de modèles nous intéressant tout particulièrement est celle des *modèles compacts*, dont les principales caractéristiques viennent d'être définies au paragraphe précédent.

Il existe différents types de modèles compacts [1,2]. Nous pouvons commencer par citer les modèles empiriques, qui ont l'avantage de la simplicité au détriment de la précision des simulations. Ce type de modèle ne permet pas ou très difficilement de s'adapter à des technologies différentes. La catégorie de modèle la plus répandue est celle des modèles analytiques plus ou moins basés sur la physique du dispositif, et ayant un degré d'empirisme variable. Une troisième catégorie regroupe les modèles physiques numériques, offrant une grande précision mais au détriment de l'avantage que représente une formulation analytique. Le principal inconvénient de cette troisième famille de modèle est qu'ils engendrent un important temps de calcul additionnel en comparaison aux modèles analytiques. Dans le but de bien fixer les idées, précisons qu'une autre catégorie de modèle, n'entrant pas dans la famille des modèles compacts, est constituée de modèles purement numériques utilisés pour la simulation de dispositifs. Au sein de tels modèles, les équations du semi-conducteur sont résolues à l'intérieur d'un maillage spatial, en 2 ou 3 dimensions. Cette dernière catégorie est intéressante pour obtenir des informations fiables sur la physique exacte du dispositif, mais est clairement inadaptée à la conception/simulation de circuits en raison de temps de calcul prohibitifs. Elle est cependant très utile dans le domaine de la validation des modèles compacts, en particulier lorsque des mesures expérimentales ne sont pas disponibles.

Précisons que probablement la meilleure compilation sur les problèmes de la modélisation du TMOS (ou les exigences de base d'un modèle de TMOS) se trouve dans les travaux de Y. TSIVIDIS [3].

Depuis quelques années, les concepteurs de circuits utilisent des modèles compacts de TMOS dits de troisième génération (Fig. 2.1). Brièvement, les caractéristiques principales de cette génération sont [4] :

- L'intention originelle de simplicité (en contraste avec les modèles de deuxième génération : BSIM, BSIM2, HSPL28),
- un petit nombre de paramètres (théoriquement étroitement liés à la physique),
- un conditionnement mathématique amélioré,
- une seule équation de modèle valable pour toutes les régions de fonctionnement du dispositif,
- l'utilisation de fonctions de lissage.

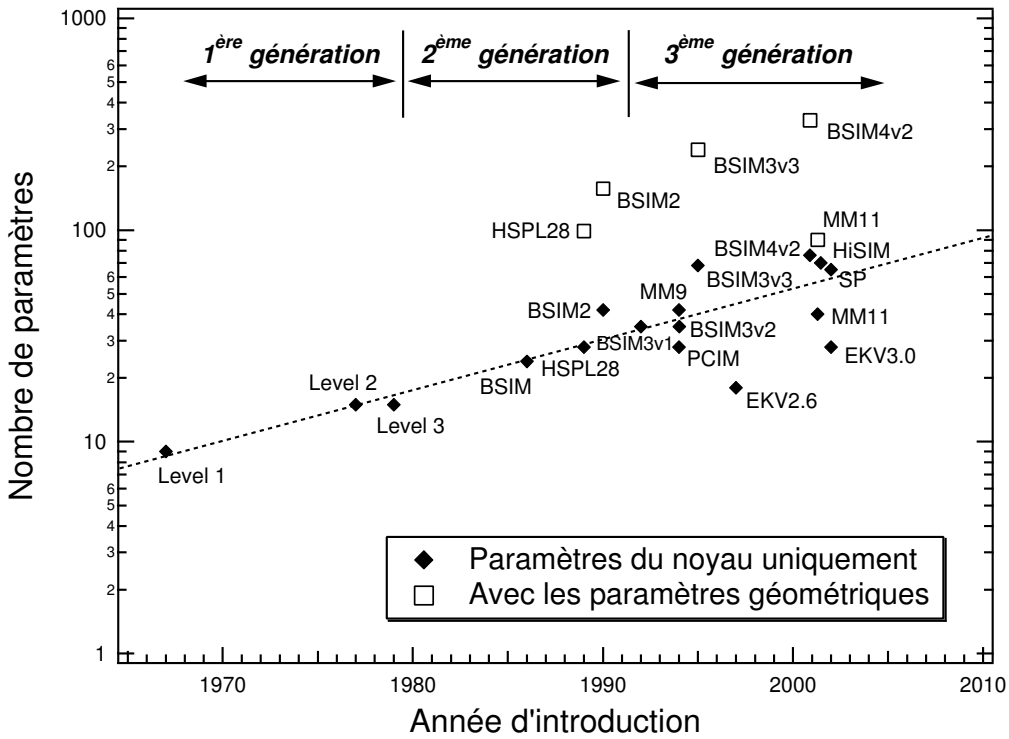


FIG. 2.1 : *Nombre de paramètres des modèles compacts en fonction de leur année d'introduction.*

Les modèles principaux de cette troisième génération sont BSIM3v3 [4], BSIM4v2 [5], MM9 [6], EKV2.6 [7,8], MM11 [9,10], ainsi que les nouveaux venus SP [11,12], HiSIM [13,14] et prochainement EKV3.0 [15,16]. Tous ces modèles sont dédiés aux transistors submicroniques, et fortement submicroniques. Les différents modèles BSIM ont été développés à l'Université de Californie à Berkeley (BSIM3v3/BSIM4), les modèles MM9/MM11 à Philips Eindhoven, le modèle EKV (Enz-Krummenacher-Vittoz) à l'EPFL-Lausanne, le modèle SP à l'Université de Pennsylvanie, et le modèle HiSIM à l'Université d'Hiroshima.

Une lecture rapide de la Fig. 2.1 indique malheureusement que pour le plus utilisé de ces modèles, à savoir BSIM3/4, *l'intention originelle* de simplicité et d'un petit nombre de paramètres a été définitivement abandonnée. Ceci peut s'expliquer par la quête d'une précision à l'extrême et d'une modélisation du tout petit effet physique, même insignifiant, aboutissant ainsi en une structure finale très complexe et in fine peu fiable.

Au cours de ce manuscrit (cf. chapitres 3 et 4), nous tenterons de mettre un point d'honneur à développer des modèles de transistor MOS ayant un rapport efficacité/simplicité élevé, tout en maintenant un lien étroit avec la physique du dispositif, au travers des équations utilisées.

2.2 Approche physique

Avant de comparer les principales approches concurrentes employées habituellement dans les simulateurs de circuits intégrés, nous allons tout d'abord décrire les bases physiques, a priori idéales, utiles à la description du fonctionnement du TMOS. Les modèles reposant sur ces principes fondamentaux sont généralement appelés modèles physiques *numériques*, car ils nécessitent l'utilisation de processus itératifs pour l'obtention des caractéristiques électriques du TMOS. De ce fait ils sont peu utilisés en simulation de circuits, car ils engendrent des temps de calcul prohibitifs. Cependant, les autres approches de la modélisation compacte (approches classiques et alternatives, cf. sections suivantes) étant plus ou moins dérivées des concepts caractérisant ces modèles physiques, il nous paraît essentiel d'en présenter les principes fondamentaux.

2.2.1 Le fonctionnement du transistor MOS

Comme le nom MOSFET (Metal-Oxide-Semiconductor Field-Effect Transistor) le suggère, le transistor MOS consiste en un substrat semi-conducteur sur lequel repose une fine couche d'oxyde isolant (SiO_2), d'épaisseur t_{ox}. Une couche conductrice (métal ou polysilicium fortement dopé) appelée l'électrode de grille est aussi déposée sur l'oxyde. Enfin, deux régions fortement dopées de profondeurs X_j, appelées source et drain, sont formées dans le substrat de part et d'autre de la grille. La structure basique d'un transistor n-MOS[1] est représentée à la Fig. 2.2. En raison du procédé de fabrication, la grille recouvre légèrement les régions de source et de drain. La région entre les jonctions de source et de drain est appelée la région du canal, et est définie par sa longueur L_{eff} et sa largeur W. Nous nous limiterons pour l'instant à cette description très succincte du dispositif MOSFET.

Avec l'aide de la Fig. 2.3, nous allons rappeler le principe des différents modes de fonctionnement du TMOS. Lorsqu'une tension V_{gb} est appliquée entre la grille et le substrat, la structure de bande, près de l'interface Si–SiO_2 est modifiée. Pour le moment, nous supposons que la source et le drain sont à la masse ($V_{sb} = V_{db} = 0$) ; dans ce cas, trois situations peuvent être distinguées (dans la région du canal) : *accumulation, déplétion et inversion*, comme indiqué aux Figs. 2.3a-b-c, respectivement.

[1] Avec un changement de signe approprié des charges et des potentiels, le raisonnement suivant est aussi valable pour des dispositifs de type p, fabriqués à partir d'un dopage substrat de type n.

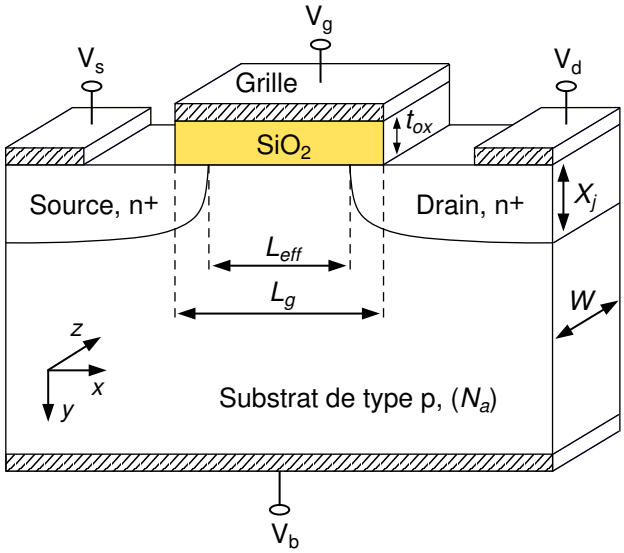
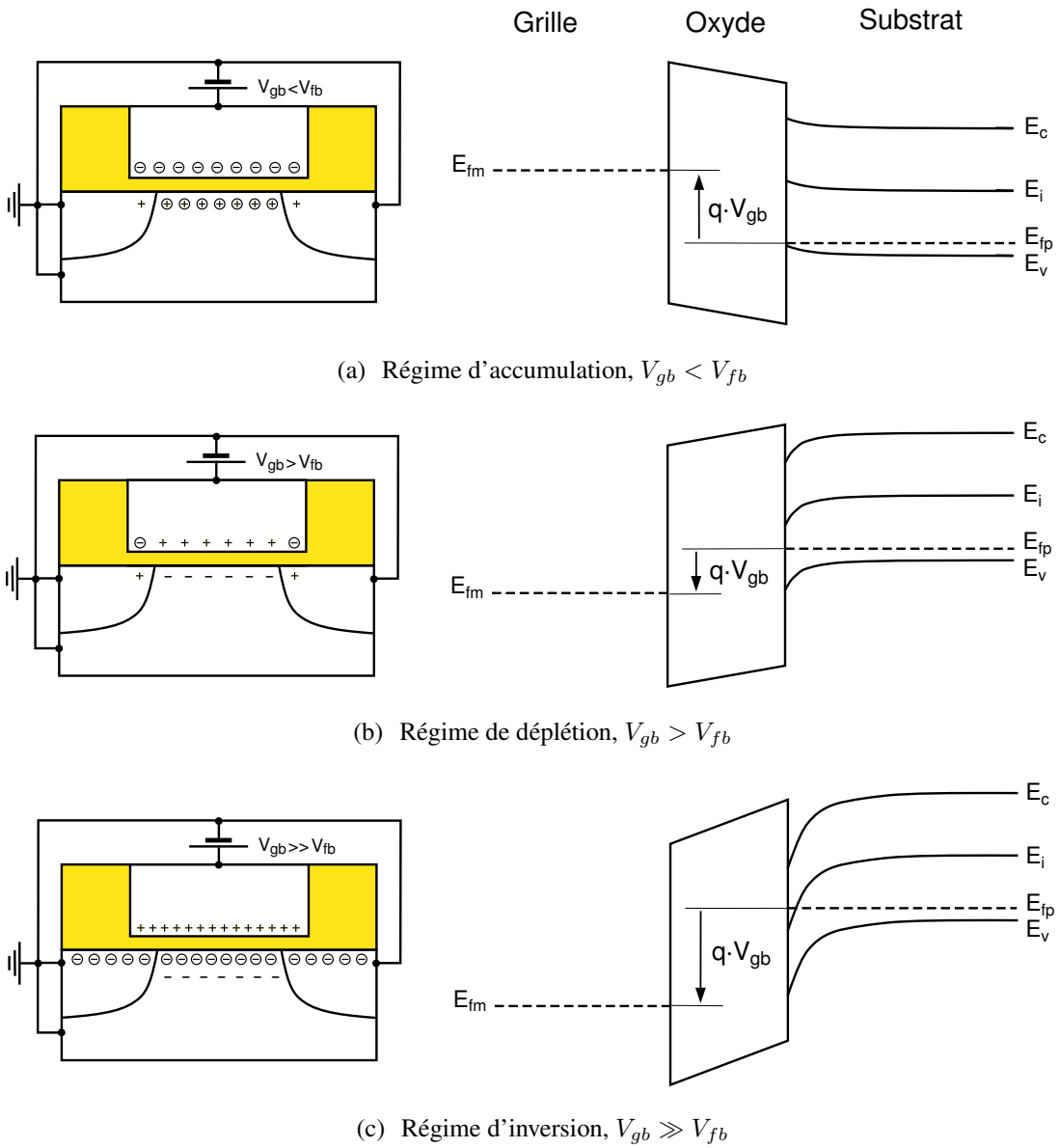


FIG. 2.2 : *Structure basique d'un transistor MOS de type n.*

Pour des tensions de grille négatives, les trous sont attirés à la surface et une très fine couche de charges positives (la couche d'accumulation) est alors formée (Fig. 2.3a). Avec l'augmentation de V_{gb}, la courbure des bandes devient plus faible, jusqu'à une certaine valeur où il n'y a plus de courbure des bandes. Cette valeur particulière de tension de grille est appelée la tension de bandes plates V_{fb}.

Au-delà de ce point, la courbure des bandes est opposée à celle en accumulation, une charge négative est en train de se former. En fait, la charge positive à la grille repousse les trous de la surface du silicium et fait apparaître une charge négative (due aux ions accepteurs immobiles), appelée charge de déplétion (Fig. 2.3b).

Quand la tension de grille augmente encore plus, la courbure des bandes vers le bas devient plus prononcée. Cette courbure peut résulter en un croisement du niveau de Fermi intrinsèque E_i avec le niveau de Fermi E_{fp} (Fig. 2.3c). Dans cette situation, la surface du semi-conducteur se comporte comme un matériau de type n, d'où le nom de région d'inversion. Une couche conductrice composée de charges négatives mobiles (électrons) est alors formée : c'est la charge d'inversion. Cette charge écrantant la couche de déplétion, cette dernière n'est alors plus que faiblement dépendante de la polarisation de la grille. En conséquence le couplage entre l'extension de la courbure des bandes dans le silicium et l'augmentation de la tension de grille



(a) Régime d'accumulation, $V_{gb} < V_{fb}$

(b) Régime de déplétion, $V_{gb} > V_{fb}$

(c) Régime d'inversion, $V_{gb} \gg V_{fb}$

FIG. 2.3 : *Distribution des charges et diagrammes des bandes d'énergie correspondants, dans un transistor n-MOS, pour les différents modes de fonctionnement : accumulation, déplétion et inversion.*

est alors fortement réduit. On parle d'inversion forte lorsque la densité de charge mobile dans la couche d'inversion est supérieure à la densité de charge fixe dans la couche de déplétion.

La charge d'inversion peut alors être mise en contact via les régions de source et de drain, et ainsi, un courant peut circuler dans le canal lorsqu'une différence de potentiel V_{ds} est appliquée entre le drain et la source. Puisque la charge d'inversion dépend fortement du potentiel appliqué à la grille, cette dernière peut alors être utilisée pour moduler le niveau du courant circulant dans le canal.

2.2.2 Définition du potentiel de surface

Nous allons maintenant introduire une grandeur fondamentale, appelée le *potentiel de surface*, qui est en fait la grandeur clé utilisée pour décrire les caractéristiques électriques du TMOS au sein des modèles compacts numériques.

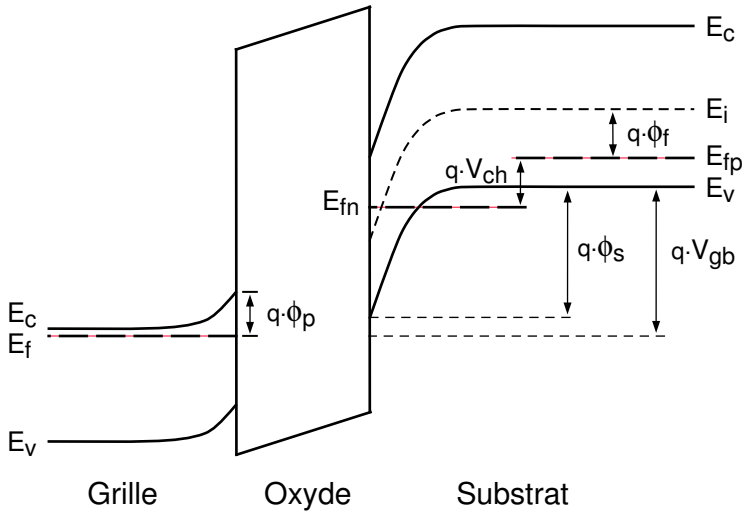


FIG. 2.4 : *Diagramme des bandes d'énergie du transistor n-MOS, en mode d'inversion. Différentes grandeurs électriques associées à cette représentation en bandes d'énergie sont aussi présentées : ϕ_s représente le potentiel de surface, ϕ_p la chute de potentiel due à la déplétion du polysilicium de grille, ϕ_f le potentiel de Fermi, et V_{ch} le potentiel de quasi-Fermi, i.e. la différence entre les potentiels des quasi niveaux de Fermi des porteurs majoritaires E_{fp} et minoritaires E_{fn} (canal).*

La définition du potentiel de surface est indiquée à la Fig. 2.4. Il représente l'importance de la courbure de bandes en termes de potentiel, c'est-à-dire le potentiel électrostatique à l'interface oxyde de grille/substrat défini par rapport à la zone neutre du substrat. Pour l'instant, nous supposerons que la grille est idéale et qu'aucune déplétion du polysilicium de grille ne se produit (i.e. $\phi_p = 0$ à la Fig. 2.4). Dans ce cas, le potentiel de surface ϕ_s peut être calculé en utilisant la méthode suivante.

Dans le cas d'un substrat de type p, une charge d'espace $\rho(x, y)$[1] est présente [17] :

$$\rho(x, y) = q \cdot (p(x, y) - n(x, y) - N_a) \tag{2.1}$$

où N_a est la concentration nette des dopants accepteurs. Les densités d'électrons et de trous, n et p respectivement, sont données — pour un semi-conducteur non dégénéré — par les statistiques

[1]Les axes x, y, z sont définis à la Fig. 2.2, page 15. La position $y = 0$ correspond à l'interface Si–SiO_2.

de Maxwell-Boltzmann :

$$n(x,y) \simeq N_a \cdot \exp\left[\frac{\phi(x,y) - V_{ch}(x) - 2 \cdot \phi_f}{\phi_t}\right] \tag{2.2}$$

$$p(x,y) \simeq N_a \cdot \exp\left[-\frac{\phi(x,y)}{\phi_t}\right] \tag{2.3}$$

où ϕ est le potentiel électrostatique par rapport à la zone neutre du substrat et ϕ_t la tension thermodynamique définie par kT/q. Le potentiel de Fermi intrinsèque ϕ_f, (voir Fig. 2.4) est défini comme $\phi_t \cdot \ln(N_a/n_i)$ (avec n_i la concentration intrinsèque de porteurs), et $V_{ch}(x)$ représente le potentiel de quasi-Fermi des électrons, qui varie de V_{sb} de la source ($x = 0$) à V_{db} au drain ($x = L$). L'équation de Poisson pour le potentiel électrostatique ϕ est alors écrite comme :

$$\nabla^2\phi = -\frac{\rho(x,y)}{\epsilon_{si}} \tag{2.4}$$

Dans le but d'obtenir une expression analytique approchant la solution de (2.4), nous faisons l'hypothèse habituelle que $\partial^2\phi/\partial x^2 \ll \partial^2\phi/\partial y^2$. Cette hypothèse est appelée *l'approximation du canal graduel*, et est valide, au sens strict, pour des dispositifs ayant un canal long. En conséquence, l'équation de Poisson peut être réécrite sous la forme suivante :

$$\frac{\partial^2\phi}{\partial y^2} \simeq \frac{qN_a}{\epsilon_{si}} \cdot \left[-\exp\left(\frac{-\phi}{\phi_t}\right) + \exp\left(\frac{\phi - V_{ch} - \phi_B}{\phi_t}\right) + 1\right] \tag{2.5}$$

avec $\phi_B = 2\,\phi_f$. Les conditions aux limites pour ϕ et $\partial\phi/\partial y$ sont choisies pour que ces deux quantités soient égales à zéro en profondeur, i.e. dans la zone neutre du substrat. En utilisant la relation $\partial^2\phi/\partial y^2 = \frac{1}{2} \cdot \partial(\partial\phi/\partial y)^2/\partial y$, la charge totale Q_{sc} (par unité de surface) dans le semi-conducteur est alors obtenue en appliquant la loi de Gauss :

$$\begin{aligned} Q_{sc} &= \epsilon_{si} \cdot \left.\frac{\partial\phi}{\partial y}\right|_{y=0} \\ &= \pm\,\gamma\, C_{ox} \left\{\phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t}\right) - 1\right] + \phi_t \cdot \exp\left(\frac{-V_{ch} - \phi_B}{\phi_t}\right)\left[\exp\left(\frac{\phi_s}{\phi_t}\right) - 1\right]\right\}^{1/2} \end{aligned} \tag{2.6}$$

où C_{ox} est la capacité de l'oxyde de grille par unité de surface, donnée par ϵ_{ox}/t_{ox}, et γ le facteur de substrat, égal à $\sqrt{2q\epsilon_{si}N_a}/C_{ox}$.

Le signe $\pm$ tient compte du fait que la densité de charge Q_{sc} est négative quand $V_{gb} > V_{fb}$ (inversion), et positive quand $V_{gb} < V_{fb}$ (accumulation).

En appliquant le théorème de Gauss à l'interface de l'oxyde, la charge Q_{sc} peut être reliée à la tension appliquée à la grille [1] :

$$Q_{sc} = -C_{ox} \cdot (V_{gb} - V_{fb} - \phi_s) \tag{2.7}$$

La combinaison des équations (2.6) et (2.7) permet d'obtenir une relation implicite pour le potentiel de surface $\phi_s(V_{gb}, V_{ch})$, donnée par :

$$\left(\frac{V_{gb} - V_{fb} - \phi_s}{\gamma}\right)^2 = \phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t}\right) - 1\right] + \phi_t \cdot \exp\left(\frac{-V_{ch} - \phi_B}{\phi_t}\right) \cdot \left[\exp\left(\frac{\phi_s}{\phi_t}\right) - 1\right] \tag{2.8}$$

Malheureusement, le potentiel de surface ne peut pas être résolu analytiquement à partir de (2.8), et doit donc être résolu *itérativement*. Ceci représente le principal inconvénient des modèles compacts dits numériques ou physiques. En effet, il faut recalculer de façon itérative le potentiel de surface pour chaque polarisation externe donnée, i.e. pour tous les couples (V_{gb}, V_{ch}). Cet inconvénient est réellement dommageable car d'un point de vue modélisation physique, cette formulation n'a que des avantages. Par exemple dans le contexte d'un *modèle en feuille de charge* (cf. § suivant), la connaissance précise du potentiel de surface permet d'obtenir immédiatement toutes les caractéristiques électriques fondamentales du TMOS. Enfin l'avantage majeur de la relation implicite (2.8) est d'être continue et valable quel que soit le régime de fonctionnement du TMOS. Ainsi, cette unique équation est suffisante pour déterminer la valeur des différentes charges — entre autres — de l'accumulation à l'inversion forte.

2.2.3 Modélisation en feuille de charge

Un modèle de transistor MOS adapté à la simulation de circuits est composé de deux principales parties : (a) un modèle statique (DC) où les tensions appliquées aux électrodes du dispositifs restent constantes, et (b) un modèle dynamique (AC) où les tensions varient avec le temps. Le modèle statique est lié à la détermination du courant de drain et des charges, tandis que le modèle dynamique est lié à la détermination des capacités et des paramètres petits signaux (transconductances). Précisons déjà que le comportement dynamique du TMOS étant dû aux effets capacitifs, il dépend donc directement des charges stockées dans la structure. Ainsi, le modèle dynamique peut être obtenu de façon systématique à partir de la définition du modèle statique.

Dans un premier temps, nous allons brièvement expliquer les approximations fondamentales nécessaires au développement du modèle de courant de drain. Ensuite, nous expliquerons en quoi consiste un modèle en feuille de charge, puis nous définirons les capacités et les principaux paramètres petits signaux du TMOS.

2.2.3.1 Courant de drain

Nous considérons un transistor n-MOS avec un dopage substrat uniforme N_a et dont les caractéristiques géométriques sont celles illustrées par la Fig. 2.2. Dans un souci de simplicité, nous supposons que le dispositif est long et large, de telle sorte que les effets de canaux courts et étroits puissent être négligés. Les caractéristiques statique et dynamique sont alors décrites par trois jeux d'équations différentielles couplées [1] :

1. L'équation de Poisson pour le potentiel électrostatique,
2. l'équation du courant pour les électrons,
3. les équations de continuité pour les électrons et les trous.

A priori, la résolution de ces équations devrait être réalisée en 3D, cependant pour diminuer la complexité du problème, nous pouvons traiter ce système comme un problème 2D, dans les directions x et y uniquement. Mais même en tant que problème à deux dimensions, les équations précédemment citées restent assez complexes, et peuvent seulement être résolues en utilisant des techniques numériques, identiques à celles mises en pratique dans les simulateurs de dispositifs 2D, tels que Atlas® [18], par exemple. Aussi, dans le but d'obtenir des solutions analytiques adaptées aux exigences des modèles compacts, de nouvelles approximations doivent être faites. Cinq approximations majeures sont alors nécessaires [1] :

1. Tout d'abord, nous supposons que la variation du champ électrique E_x dans la direction x (le long du canal) est beaucoup plus faible que la variation correspondante du champ E_y dans la direction y. C'est ce qu'on appelle l'approximation du canal graduel, dont nous avons déjà parlé au § 2.2.2, page 18. Cette approximation est généralement valable, excepté dans la zone proche du drain. En dépit de ce problème, elle est utilisée car elle réduit à une dimension le problème de la détermination du courant de drain.
2. Le courant de trous peut être négligé, c'est-à-dire que le calcul de la densité de courant J_n (courant d'électrons) est suffisant pour déterminer le courant de drain.

3. Les phénomènes de recombinaison et génération sont négligés. Cela signifie que la densité du courant de drain est un courant d'électrons de divergence nulle, et donc que le courant de drain total I_d est constant en n'importe quel point du canal.

4. Le courant circule dans la direction x uniquement (le long du canal). Cela implique que $\partial\phi_n/\partial y = 0$, et que donc le potentiel de quasi-Fermi des électrons ϕ_n est constant dans la direction y. Ces hypothèses sont d'ailleurs aussi utilisées dans les simulateurs de dispositifs 2D, et donnent entière satisfaction.

5. La mobilité des électrons μ_n est supposée constante — pour l'instant — pour des raisons de simplification de calcul d'intégrales[1]. En conséquence, il est possible de définir une densité de charge mobile Q_{inv} (densité de charge d'inversion), donnée par :

$$Q_{inv}(x) = q \int_0^{\infty} n(x, y)\, dy \tag{2.9}$$

Connaissant Q_{inv}, nous pouvons écrire le courant de drain sous la forme suivante [1] :

$$I_d(x)\, dx = -\mu_n \cdot W \cdot Q_{inv}(x)\, dV_{ch} \tag{2.10}$$

En supposant que l'approximation du canal graduel soit valide sur toute la longueur du canal et en intégrant (2.10) le long du canal, nous obtenons :

$$I_d(x) = -\mu_n \cdot \frac{W}{L} \int_{V_{sb}}^{V_{db}} Q_{inv}(x)\, dV_{ch} \tag{2.11}$$

Ainsi, pour calculer le courant I_d, nous avons besoin de connaître la densité de charge d'inversion Q_{inv}. C'est avec cet objectif que nous allons maintenant détailler le principe d'une modélisation en feuille de charge.

2.2.3.2 Approximation de la feuille de charge

L'analyse menée jusqu'à présent était très générale. Aucune hypothèse n'a encore été faite sur l'épaisseur de la couche d'inversion, ni sur la présence éventuelle de porteurs libres dans la région de déplétion. Supposons maintenant que la couche d'inversion ait une épaisseur nulle[2] (i.e. simplement une feuille de charge) de telle sorte qu'aucune chute de potentiel ne se produise

[1] En fait la mobilité dépend fortement des deux champs électriques E_x et E_y.

[2] Il convient de se rappeler que la région du canal est confinée dans une couche très mince, de l'ordre de 1–10 nm. Négliger l'épaisseur de la couche d'inversion est donc une très bonne approximation puisque cette épaisseur est au moins de deux ordres de grandeur inférieure à celle de la couche de déplétion.

au travers d'elle. De plus, plaçons nous également dans l'hypothèse où l'approximation de la déplétion totale est valide (i.e. la région de déplétion est dépourvue de porteurs mobiles). En résumé, cela signifie que la densité de charge mobile Q_{inv} résulte du phénomène d'inversion uniquement, d'où son nom usuel de densité de charge d'inversion.

La densité de charge du substrat Q_b est alors définie par :

$$Q_b(x) = -q \int_0^{\infty} (N_a - p)\, dy \tag{2.12}$$

En assumant l'approximation de la déplétion, il vient :

$$\begin{aligned} Q_b(x) &= \pm\, \gamma \cdot C_{ox} \cdot \sqrt{\phi_s(x) + \phi_t \cdot \left[\exp\left(\frac{-\phi_s(x)}{\phi_t}\right) - 1\right]} \\ &\simeq -\gamma \cdot C_{ox} \cdot \sqrt{\phi_s(x)} \quad \text{déplétion et inversion} \end{aligned} \tag{2.13}$$

En accumulation Q_b est positive et égale à la densité de charge d'accumulation, et en inversion, Q_b est négative et égale à la densité de charge de déplétion. La charge induite Q_{sc} dans le semi-conducteur étant la somme de la charge d'inversion Q_{inv} et de la charge du substrat Q_b (i.e. $Q_{sc} = Q_{inv} + Q_b$), en combinant les équations (2.7) et (2.13) l'expression de la densité de charge d'inversion est finalement donnée par :

$$Q_{inv}(x) = -C_{ox} \cdot (V_{gb} - V_{fb} - \phi_s(x)) - Q_b(x) \tag{2.14}$$

Maintenant, en exprimant (2.10) en termes de potentiel de surface, nous obtenons :

$$I_d(x) = -\mu_n \cdot W \cdot Q_{inv}(x) \cdot \frac{dV_{ch}}{d\phi_s} \frac{d\phi_s}{dx} \tag{2.15}$$

Cette expression peut alors être réarrangée mathématiquement. Cela est largement détaillé dans la littérature scientifique, citons par exemple les articles de référence de PAO ET SAH [19], et BREWS [20], ainsi que le livre d'ARORA [1]. Après simplification, le courant de drain peut se mettre sous sa forme bien connue :

$$\begin{aligned} I_d(x) &= -\mu_n \cdot W \cdot \left[Q_{inv}(x) \cdot \frac{d\phi_s}{dx} - \phi_t \cdot \frac{dQ_{inv}(x)}{dx}\right] \\ &= I_{\text{drift}} + I_{\text{diff}} \end{aligned} \tag{2.16}$$

L'équation (2.16), originellement introduite par BREWS, est connue en tant que modèle en feuille de charge [20]. Cette équation montre que le courant de drain est la somme de deux

composantes I_{drift} et I_{diff}, à savoir le courant de conduction (drift) et le courant de diffusion, respectivement. Bien que I_d soit constant, ses deux composantes sont des fonctions de la distance x le long du canal. Il convient aussi de remarquer que I_{drift} et I_{diff} sont des équations différentielles couplées, ce qui signifie qu'elles ne peuvent pas être intégrées séparément. Il est cependant possible de contourner cette difficulté en supposant que seule une des deux composantes est présente. Ainsi, en ne considérant que la composante de conduction, nous pouvons intégrer la partie conduction de (2.16) avec les conditions aux limites suivantes :

$$\phi_s(x) = \begin{cases} \phi_{s0} & \text{pour} \quad x = 0, \\ \phi_{sL} & \text{pour} \quad x = L. \end{cases} \tag{2.17}$$

Ensuite, en utilisant (2.14) pour la densité de charge d'inversion Q_{inv}, il vient :

$$\int_0^L I_{\text{drift}}\, dx = -\mu_n \cdot W \int_{\phi_{s0}}^{\phi_{sL}} Q_{inv}\, d\phi_s \tag{2.18}$$

ce qui mène à :

$$I_{\text{drift}} = \mu_n\, C_{ox}\, \frac{W}{L} \left[(V_g - V_{fb})\,(\phi_{sL} - \phi_{s0}) - \frac{1}{2}\left(\phi_{sL}^2 - \phi_{s0}^2\right) - \frac{2}{3}\,\gamma\, \left(\phi_{sL}^{3/2} - \phi_{s0}^{3/2}\right) \right] \tag{2.19}$$

où ϕ_{s0} et ϕ_{sL} sont les valeurs du potentiel de surface à la fin de la source et du drain, respectivement.

De façon similaire, si l'on suppose que seulement la composante diffusion est présente, le courant de drain (partie diffusion) peut être obtenu par intégration de I_{diff} de la source au drain :

$$\int_0^L I_{\text{diff}}\, dx = \mu_n \cdot W \cdot \phi_t \int_{\phi_{s0}}^{\phi_{sL}} dQ_{inv} \tag{2.20}$$

ce qui mène à :

$$I_{\text{diff}} = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \phi_t \cdot \left[\gamma \cdot (\phi_{sL}^{1/2} - \phi_{s0}^{1/2}) + (\phi_{sL} - \phi_{s0}) \right] \tag{2.21}$$

Le courant de drain total I_d est alors complètement défini en additionnant les équations (2.19) et (2.21). Le courant de drain obtenu dans le cadre de l'approximation drift–diffusion est illustré à la Fig. 2.5.

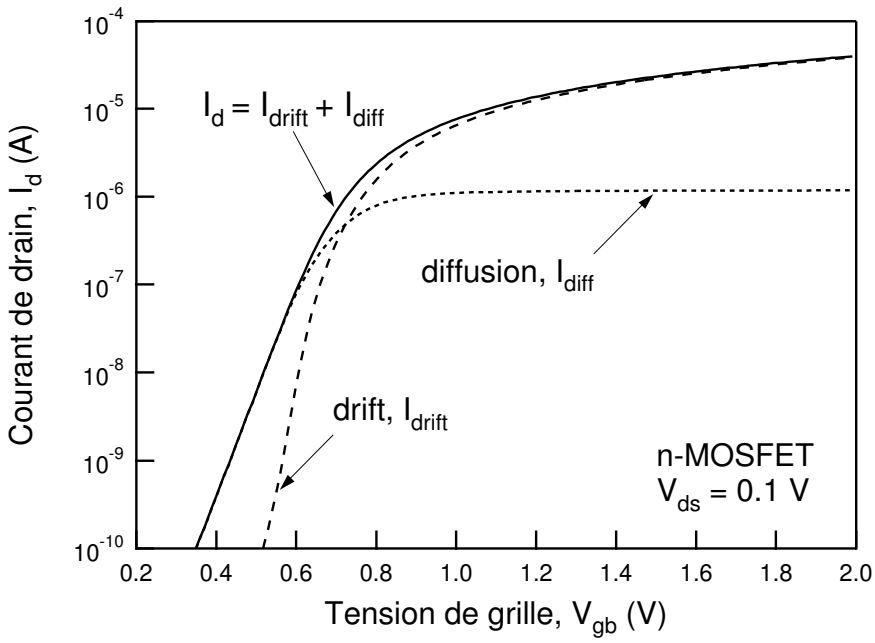


FIG. 2.5 : *Courant de drain I_d calculé dans le cadre de l'approximation drift–diffusion.*

L'inconvénient des modèles compacts physiques apparaît à ce stade du raisonnement. En effet, les valeurs du potentiel de surface ϕ_{s0} et ϕ_{sL}, nécessaires au calcul du courant de drain, doivent être obtenues numériquement, par la résolution de l'équation implicite (2.8), sous les conditions suivantes :

$$\phi_s(x) = \begin{cases} \phi_{s0} \text{ et } V_{ch}(x) = V_{sb} & \text{pour } x = 0, \\ \phi_{sL} \text{ et } V_{ch}(x) = V_{db} & \text{pour } x = L. \end{cases} \tag{2.22}$$

Dans la région d'inversion faible, où ϕ_{s0} est quasiment égal à ϕ_{sL}, de très faibles erreurs dans les valeurs de ϕ_{s0} et ϕ_{sL} peuvent engendrer d'importantes erreurs dans le calcul du courant de diffusion, puisqu'il dépend de la différence $\phi_{sL} - \phi_{s0}$. En conséquence, une détermination extrêmement précise du potentiel de surface est donc indispensable au bon comportement de ce modèle de courant de drain.

En guise de première conclusion sur les modèles en feuille de charge, il se dégage deux points forts. D'une part, ces modèles sont fortement liés à la physique du dispositif MOSFET, et donc apportent une précision de modélisation optimale. D'autre part, ils requièrent un temps de calcul coûteux, ce qui les rend peu adaptés à la simulation des circuits ayant une forte densité d'intégration (VLSI–ULSI).

2.2.3.3 Capacités et paramètres petits signaux

Le comportement dynamique du TMOS résulte d'effets capacitifs liés aux charges stockées dans la structure. Cette description des capacités s'ajoute alors au modèle statique précédemment développé. Dans un transistor MOS, les caractéristiques capacitives sont en fait la somme des capacités intrinsèques (région du canal) et des capacités extrinsèques (régions source/drain)[1].

Pour l'instant, nous allons surtout nous intéresser aux capacités intrinsèques, sans toutefois oublier d'introduire dans un second temps, la notion de paramètre *petit signal*. Il est utile de préciser que les modèles développés ici sont des modèles *quasi-statiques* ; c'est-à-dire que l'on suppose que les tensions appliquées au TMOS varient suffisamment lentement pour permettre aux charges stockées de suivre les variations des tensions.

Le point fondamental pour un calcul correct des capacités est d'avoir au préalable défini un modèle de charges très précis. En particulier, les variations des différentes charges en fonction des tensions appliquées au dispositif doivent être rigoureusement modélisées. Voici justement, un nouvel avantage d'un modèle en feuille de charge. Dans le cadre du modèle statique développé au paragraphe précédent, nous avons déjà obtenu une formulation physique des différentes charges.

Commençons par récapituler les différentes expressions des charges du TMOS ; nous avons :

- la charge totale de grille Q_g, qui est l'opposée de la charge totale induite dans le semi-conducteur Q_{sc} :

$$Q_g = C_{ox} \cdot (V_{gb} - V_{fb} - \phi_s) = -Q_{sc} \tag{2.23}$$

- la charge du substrat Q_b (déplétion et inversion) :

$$Q_b = -\gamma \cdot C_{ox} \cdot \sqrt{\phi_s} \tag{2.24}$$

- la charge d'inversion Q_{inv} :

$$Q_{inv} = -C_{ox} \cdot (V_{gb} - V_{fb} - \phi_s) - Q_b \tag{2.25}$$

[1] Le chapitre 3 sera entièrement consacré à la description des capacités extrinsèques ou parasites.

Le principe de conservation de la charge est donc bien respecté :

$$Q_g + Q_{inv} + Q_b = 0 \tag{2.26}$$

D'un point de vue simulation de circuits, il se pose alors un problème. En effet, nous avons besoin de connaître les différentes charges de la structure correspondant aux différents terminaux du TMOS, i.e. la charge associée à la grille (Q_g), au substrat (Q_b), à la source (Q_s) et au drain (Q_d). Si les charges de grille et de substrat, sont elles, parfaitement définies par (2.23) et (2.24), les charges attribuées à la source et au drain ne sont pas explicitement connues. En fait, seule la charge d'inversion est précisément connue. Cette charge est en contact avec les deux terminaux de source et de drain grâce au canal, c'est pourquoi il est possible (et indispensable) de la partitioner en une charge Q_s liée à la source, et en une charge Q_d liée au drain :

$$Q_{inv} = Q_s + Q_d \tag{2.27}$$

Différentes approches ont été proposées dans la littérature pour partitioner Q_{inv} en Q_s et Q_d ; une des plus efficaces est celle proposée par WARD [21]. Cette dernière conduit à un modèle de capacités qui est en très bon accord avec les résultats expérimentaux [22].

Maintenant que nous connaissons les différentes charges Q_i ($i = g, b, s, d$) en fonction des différentes tensions appliquées aux électrodes V_j ($j = g, b, s, d$), nous pouvons immédiatement obtenir les différentes capacités — aussi appelées *transcapacités* — par dérivation des charges par rapport aux tensions appliquées. Ainsi, nous obtenons [1,23] :

$$C_{ij} = \begin{cases} -\dfrac{\partial Q_i}{\partial V_j} & \text{si} \quad i \neq j, \\ \dfrac{\partial Q_i}{\partial V_j} & \text{si} \quad i = j. \end{cases} \tag{2.28}$$

Il résulte de cette description une matrice de 16 transcapacités, décrivant le comportement capacitif du TMOS. D'une manière générale, il est tout particulièrement utile de connaître les transcapacités C_{gg}, C_{bg}, C_{sg} et C_{dg} car ce sont ces dernières qui sont extraites lors des mesures effectuées sur des dispositifs réels (objectif de validation du modèle).

Avant de clore cette section traitant du modèle dynamique, nous allons introduire la notion importante de paramètres petits signaux. Lors du fonctionnement normal du TMOS, une tension appliquée à la grille, au substrat, au drain ou à la source, provoque une variation du courant de drain. Cela va nous permettre de définir le concept de *transconductance*.
À titre d'exemple, le rapport entre la variation de courant de drain sur la variation de la tension de grille — les autres tensions étant fixées — est appelé *transconductance de grille* ou plus simplement transconductance g_m. Ce rapport est défini comme :

$$g_m = \left.\frac{\partial I_d}{\partial V_{gs}}\right|_{V_{ds},V_{bs}} \tag{2.29}$$

La transconductance g_m est un des plus importants paramètres du TMOS, car c'est en réalité une mesure du gain du dispositif. D'une façon similaire au cas de la transconductance g_m, il est possible de définir la conductance de drain g_{ds} (aussi appelée conductance *de sortie*) :

$$g_{ds} = \left.\frac{\partial I_d}{\partial V_{ds}}\right|_{V_{gs},V_{bs}} \tag{2.30}$$

ou encore la transconductance de substrat g_{mbs} :

$$g_{mbs} = \left.\frac{\partial I_d}{\partial V_{bs}}\right|_{V_{gs},V_{ds}} \tag{2.31}$$

Ces trois grandeurs sont nécessaires pour réaliser une analyse petit signal du fonctionnement du TMOS. En particulier, la conductance de sortie g_{ds} et la transconductance g_m sont très importantes pour la conception des circuits analogiques [22].

2.2.4 Points essentiels de l'approche physique

Dans cette section, nous avons décrit les bases à partir desquelles sont développés les modèles compacts physiques. De par leur approche étroitement liée à la physique du TMOS, ces modèles sont naturellement précis dans la description des caractéristiques électriques du TMOS. Ils sont basés sur la description d'une grandeur clé : le potentiel de surface. C'est pourquoi ces modèles sont généralement appelés *modèles à base de potentiel de surface* (surface-potential-based MOSFET model). L'équation implicite (2.8) permet d'obtenir la valeur du potentiel de surface pour n'importe quelle polarisation appliquée au dispositif. L'avantage majeur d'une telle méthode est qu'elle fournit une description *continue* du potentiel de surface, du régime

d'accumulation à celui d'inversion. Ainsi, dans le contexte d'un modèle en feuille de charge, nous avons directement accès aux charges, et au courant de drain grâce à l'approximation drift–diffusion. Les charges et le courant de drain étant chacun formulés par une unique équation, continue sur toute la plage de fonctionnement, les dérivées de ces quantités sont aussi continues pour tous les modes d'opération. Il en résulte une description précise et cohérente des différentes transcapacités et transconductances.

Cependant en dépit de tous les avantages inhérents aux modèles formulés en potentiel de surface, ce type de modèle compact est peu adapté à la simulation de circuits (VLSI–ULSI). En effet, comme nous l'avons déjà mentionné, le temps de calcul requis par ces modèles étant important, leur utilité (en simulation de circuits) est finalement assez réduite. Le principal modèle compact numérique disponible à l'heure actuelle s'appelle HiSIM (Hiroshima-university STARC IGFET Model) [13]. C'est un modèle récent (2001), dans lequel une nouvelle solution a été mise en place dans le but de réduire le temps de simulation imputé à la résolution itérative de l'équation de Poisson.

2.3 Approche classique

Dans le but de ne pas avoir à résoudre l'équation implicite du potentiel de surface, les modèles compacts utilisés en simulation de circuits ont été développés autour de diverses approximations analytiques. Ces modèles sont encore très utilisés, d'une part pour des raisons historiques, et d'autre part en raison de leur simplicité. L'approche la plus classique consiste à modéliser de façon indépendante les différents régimes de fonctionnement du TMOS. En plus des trois régions classiques que sont l'accumulation, la déplétion et l'inversion, une distinction supplémentaire est généralement faite. Il s'agit de la séparation du régime d'inversion en deux sous-régimes : l'inversion faible et l'inversion forte. La limite couramment utilisée pour décrire le point de séparation entre ces deux régimes est la tension de seuil V_{th}. Pour cette raison, ce type de modèle compact analytique est généralement appelé *modèle à base de tension de seuil* (threshold-voltage-based MOSFET model).

Un exemple typique illustrant les principes de modélisation utilisés dans ces modèles est celui du courant de drain. La région d'inversion étant scindée en deux sous-régions, le modèle du courant de drain est alors naturellement composé de deux sous-modèles distincts : un pour l'inversion faible, l'autre pour l'inversion forte. De légitimes questions se posent alors quant à la validité du modèle au niveau de la transition entre ces régimes de fonctionnement.

2.3.1 Les modèles SPICE de première génération

L'acronyme SPICE signifie Simulation Program with Integrated Circuit Emphasis. Le noyau originel de SPICE a été développé à l'Université de Californie à Berkeley, à la fin des années 60 (cf. Fig. 2.1, page 13). Une raison probable du succès connu par SPICE depuis ses débuts est que ce simulateur a été rendu public, au bon endroit, au bon moment [23]. Sa sortie coïncida à la période de forte croissance de l'industrie du circuit intégré, et ainsi ce simulateur s'est rapidement imposé comme le standard de la simulation de circuits. Différents modèles de transistor MOS furent tour à tour intégrés à SPICE. Nous allons très brièvement faire l'historique de ces différents modèles (Level 1, 2 et 3), et donner quelques unes de leurs caractéristiques majeures [23,24].

Le modèle Level 1 (1967) est le modèle de TMOS original inclus dans tous les simulateurs SPICE. Il est important dans un contexte historique. L'approche mathématique utilisée dans ce modèle est à la base des modèles plus sophistiqués qui seront développés par la suite (en particulier à Berkeley). En ce sens, la compréhension du Level 1 est impérative, car elle va permettre de comprendre comment sont construits les modèles plus évolués. Des explications détaillées sur ce modèle peuvent être trouvées dans de nombreux ouvrages, une référence bien connue étant le livre de D. FOTY [23]. Précisons simplement que le nombre restreint de paramètres du Level 1 (cf. Tableau 2.1) reflète le caractère simpliste de ce modèle, où quasiment aucun effet (effets de petite géométrie, réduction de la mobilité, etc.) n'est pris en compte. Il convient cependant de reconnaître qu'un intérêt inhérent à cette nature simple est de permettre une extraction aisée des paramètres du modèle. Contrairement aux modèles plus complexes, l'extraction des paramètres du Level 1 ne nécessite pas l'utilisation de techniques sophistiquées. En commentaire final, nous dirons simplement que ce modèle est maintenant beaucoup trop simple pour être utilisé avec les technologies de transistors actuels : d'un point de vue conception industrielle, le Level 1 est obsolète.

Les principaux effets de petite géométrie étant ignorés dans le modèle Level 1, le modèle Level 2 a alors été développé pour prendre en compte ces oublis. L'approche fondamentale est de reprendre les bases du Level 1, et d'y ajouter de nouvelles équations et de nouveaux paramètres pour inclure des effets de petite géométrie, en tant que corrections du modèle de base. Une amélioration majeure concerne la prise en compte de la variation de la charge de déplétion le long du canal. Il en résulte une expression complexe mais aussi plus précise

TAB. 2.1 : *Liste des paramètres du modèle SPICE Level 1.*

Paramètres	**Unités**	**Description**
TPG		Type de matériau de grille
TOX	m	Épaisseur d'oxyde de grille
NSUB	cm^{-3}	Niveau de dopage substrat
XJ	m	Profondeur des jonctions source/drain
VTO	V	Tension de seuil (canal long et large)
UO	cm^2/Vs	Mobilité à champ faible
LAMDA	V^{-1}	Modulation de la longueur du canal
CGSO	F/m	Capacité grille–source (zero bias)
CGDO	F/m	Capacité grille–drain (zero bias)
CGBO	F/m	Capacité grille–substrat (zero bias)

du courant de drain. Parmi les effets canaux courts inclus, nous pouvons citer la modulation de la longueur du canal, la saturation de la vitesse des porteurs, la réduction de la mobilité des porteurs, etc. Le Level 2 cherche à prendre en compte tous ces phénomènes, mais malheureusement toutes ces additions le rendent très complexe mathématiquement. Au final, le modèle est plutôt inefficace, les effets canaux courts ne sont que partiellement considérés et des problèmes de convergence apparaissent souvent (discontinuités dans les dérivées).

Un nouveau modèle, le Level 3, a alors été développé pour pallier aux défauts du Level 2. L'approche de base est similaire à celle de ses prédécesseurs en ce sens que toute dépendance géométrique est codée dans les équations du modèle, et que la plupart des équations et paramètres sont les mêmes (ou quasiment les mêmes). Cependant, contrairement aux modèles Level 1 et 2, le Level 3 adopte une approche *semi-empirique*. De par sa simplicité et sa nature semi-empirique, ce modèle remporta un réel succès. En comparaison au Level 2, le Level 3 est plus précis, plus rapide, et rencontre beaucoup moins de problèmes de convergence [23,24]. Il est important de préciser qu'originellement ce dernier n'a pas été conçu pour être très précis à des grandes longueurs de canal, mais à la longueur minimale d'une technologie donnée. Ceci n'est pas gênant dans la conception des circuit digitaux, où seules les longueurs les plus courtes sont d'intérêt, et donc où une précision pour les grandes dimensions n'est pas exigée. Cependant, les circuits analogiques utilisent des dispositifs à grand canal, où le Level 3 n'est pas précis (discontinuité de la conductance de sortie). Précisons aussi qu'en dépit du fait que

le Level 3 ait été développé pour des longueurs de canal supérieures à 1 micron, moyennant certaines précautions il peut être étendu à des dimensions plus petites, d'où son succès en simulation numérique. Ainsi, une partie du succès du Level 3 est due au « binning » du modèle, i.e. la division de l'espace géométrique W/L en sous-modèles décrivant des régions plus étroites définies dans cet espace [23,24].

Au chapitre des défauts, comme dans le cas du Level 2, le modèle de courant sous le seuil n'est pas physiquement réaliste. Ceci limite donc l'intérêt du Level 3 pour la conception de circuits analogiques, et illustre la différence entre les exigences requises par la simulation analogique et la simulation digitale. Enfin, la pauvre dépendance géométrique intrinsèque du modèle Level 3 et son incapacité à modéliser proprement la conductance de sortie poussent au développement de la deuxième génération de modèles de MOSFET.

2.3.2 Les modèles de deuxième génération

Dans le même état d'esprit que précédemment, nous allons maintenant parler de la philosophie des modèles de deuxième génération, et non pas dresser un descriptif laborieux de leurs équations, qui de toute façon est aisément accessible via les références [1,4,23]. La seconde génération des modèles SPICE de MOSFET inclut BSIM et ses deux descendants à savoir : HSPICE Level 28, et BSIM2. Pour chacun d'eux une approche entièrement nouvelle est introduite. Dans les modèles de première génération, le TMOS de grande géométrie est défini comme fournissant un modèle de base (point de départ), les effets de petite géométrie étant considérés comme des corrections à apporter à ce modèle de base. Toutes les dépendances géométriques (longueur et largeur de canal) sont codées dans les équations du modèle ; il est donc supposé que le jeu de paramètres du modèle est valable pour la gamme de géométrie entière. Cependant, cette dernière hypothèse n'est généralement pas valide [23]. Les résultats du modèle — pour un jeu de paramètres donné — ne sont en fait généralement valables que pour un domaine restreint de l'espace W/L. Ailleurs que dans cette zone, les résultats vont se dégrader. Cette situation va donner lieu à une extension du modèle par binning, avec l'introduction d'un jeu de sous-modèles valides uniquement dans des gammes géométriques particulièrement étroites (cf. Level 3) .

Les modèles de deuxième génération ont une approche complètement différente. Un jeu d'équations est élaboré pour décrire le comportement du dispositif ; ces équations contiennent la longueur et la largeur de canal. Cette partie intrinsèque de la formulation du modèle est identique à la structure de base des modèles de première génération. Cependant, en net contraste

avec les modèles de première génération, une construction additionnelle, en fait la structure extrinsèque du modèle, est créée au-dessus de la structure intrinsèque. Cette structure contient une dépendance géométrique intégrée dans l'expression suivante [23] :

$$X = X_0 + \frac{LX}{L_{eff}} + \frac{WX}{W_{eff}} \tag{2.32}$$

où X représente un paramètre quelconque du modèle. Dans les modèles de première génération, les paramètres étaient considérés de façon isolée et indépendante. Dans les modèles de deuxième génération, X est en réalité un paramètre composé, i.e. synthétisé à partir de trois autres paramètres X_0, LX et WX, où LX et WX décrivent respectivement les variations en longueur et en largeur du paramètre X. Ainsi, au sein de ces modèles les différents paramètres n'apparaissent plus sous forme de singleton (X) mais sous forme de triplets (X_0, LX, WX), ce qui augmente considérablement le nombre total de paramètres. À titre d'exemple, le modèle BSIM de base exige 17 paramètres. Puisque presque tous ses paramètres sont en fait composés de triplets, le jeu de paramètres final contient en définitive 49 paramètres distincts.

2.3.3 Les modèles de troisième génération

Dans cette partie, nous allons décrire les approches traditionnellement utilisées dans les modèles compacts *classiques* de troisième génération (cf. Fig. 2.1, page 13). Nous regroupons sous le terme 'classique' les modèles de Berkeley (BSIM3v3 et supérieurs) [4,5] et le modèle MM9 de Philips [6]. Tous ces modèles sont basés sur la notion de tension de seuil, d'où leur qualificatif usuel de modèles à base de tension de seuil.

Les modèles Level 1, Level 2, et Level 3 représentent la première génération de modèle de transistor MOS. Ils se focalisent sur les descriptions analytiques du comportement du dispositif, lesquelles amènent à un petit nombre de paramètres, relativement facile à extraire. Les modèles de deuxième génération : BSIM, Level 28 (de HSPICE) et BSIM2 insistent plus sur le conditionnement mathématique des équations du modèle, dans le but de réaliser des simulations de circuits plus efficaces et plus robustes. Cependant, leurs équations sont construites sur une base semi-empirique, voire même complètement empirique. En conséquence, les paramètres des modèles sont largement empiriques et fournissent peu d'information quant à la physique (technologie) du procédé (process) qu'ils décrivent. Enfin, le nombre de paramètres est très élevé.

Les modèles BSIM3/4 et MM9 représentent l'émergence de la troisième génération de modèles SPICE de MOSFET. Cette génération reflète le souhait et l'effort de ré-introduire une base physique dans la forme du modèle et dans ses paramètres associés, tout en maintenant l'aptitude mathématique du modèle. L'objectif premier est de permettre une liaison plus concrète entre le jeu de paramètres et la technologie du procédé sous-jacent, tout en évitant l'utilisation abusive d'expressions polynômiales (au comportement parfois problématique) des modèles de deuxième génération.

BSIM3 lui-même a évolué au travers de trois versions :

- BSIM3v1 forme la base originale du modèle, mais souffre de sévères problèmes mathématiques.
- BSIM3v2 introduit de fortes corrections pour résoudre les difficultés mathématiques de BSIM3v1, et de nouveaux paramètres sont ajoutés.
- BSIM3v3 change significativement la forme du modèle pour garantir des équations lissées et continues ; un nombre d'expressions empiriques important est aussi introduit, au travers de nombreux paramètres additionnels.

Aujourd'hui la série de modèle BSIM en est à sa version 4v2, et même 4v3 depuis peu.

Les caractéristiques générales des modèles de transistor MOS de troisième génération peuvent se résumer ainsi :

- La structure de base est très similaire aux modèles de première génération ; toute dépendance géométrique est définie dans la structure intrinsèque du modèle, tandis que la structure géométrique extrinsèque des modèles de deuxième génération est écartée[1]. Les équations des modèles sont donc valides pour toutes les géométries.
- Les modèles contiennent *théoriquement* un petit nombre de paramètres (basé sur la physique) ; l'intention étant de fournir une meilleure description de la technologie du process sous-jacent, ainsi que des variations associées à ce process[2].
- L'utilisation de fonctions de lissage garantit un comportement continu et lissé des équations du modèle à travers toutes les régions de fonctionnement du dispositif. Ces fonctions permettent de conserver l'utilisation d'expressions distinctes (correspondant aux différentes régions de fonctionnement) pour décrire les caractéristiques électriques,

[1] L'exception à cette description est le modèle MM9 qui prolonge la structure extrinsèque des modèles de deuxième génération.

[2] L'exception à ce traitement est le modèle BSIM3v3, lequel a commencé avec cette structure, mais a depuis évolué vers une forme empirique, caractérisée par un nombre impressionnant de paramètres (≈ 240).

tout en fournissant une plus ou moins bonne région de transition entre les limites opérationnelles.

– Enfin, la plupart de ces modèles inclut de nombreux effets de petite géométrie et d'effets parasites ignorés dans le passé.

En conclusion, une avancée importante liée aux modèles de troisième génération peut être résumée par la Fig. 2.6. Ces modèles décrivent précisément des zones bien déterminées de la dynamique de fonctionnement du transistor. Prenons l'exemple montré à la Fig. 2.6 : il apparaît que le courant de drain est modélisé en deux parties, l'une correspondant à la région d'inversion faible et l'autre à celle d'inversion forte. Grâce à l'utilisation d'une fonction de lissage, la caractéristique globale du courant de drain est bien lissée et continue (y compris sa première dérivée), ce qui est très important pour le calcul des transconductances. Cependant une interrogation légitime se pose : qu'en est-il de la précision du modèle au niveau de la région de transition, c'est-à-dire au niveau de la région *d'inversion modérée* ?

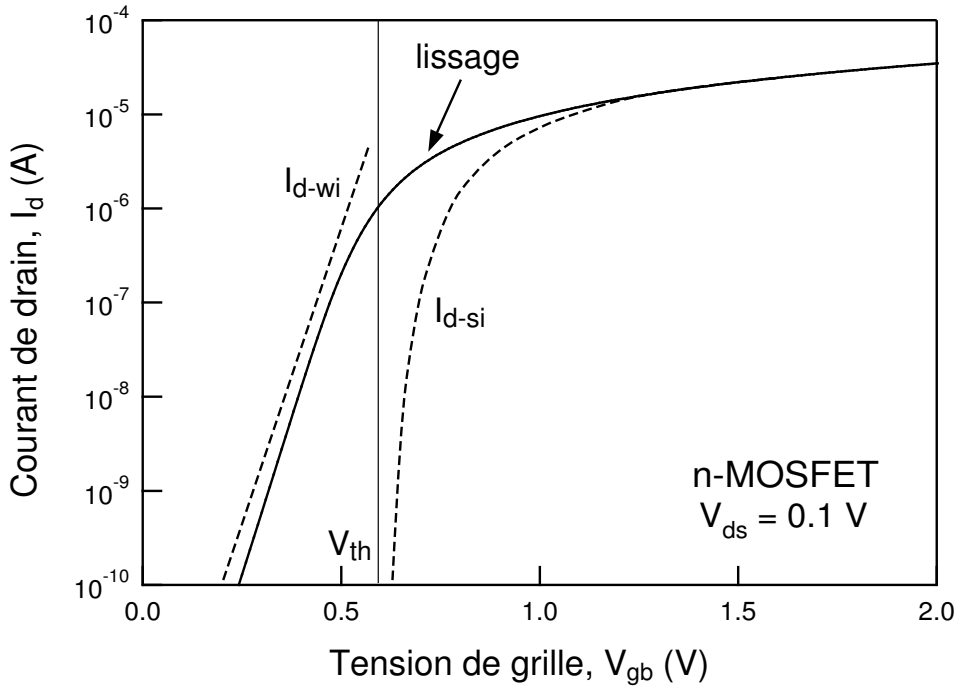


FIG. 2.6 : *Courant de drain I_d calculé suivant l'approche classique des modèles compacts conventionnels de troisième génération (BSIM3v3, BSIM4v2, MM9). V_{th} correspond à la tension de seuil, $I_{d\text{-}wi}$ à l'approximation du courant de drain en inversion faible (région sous le seuil) et $I_{d\text{-}si}$ à l'approximation du courant de drain en inversion forte (région au-delà du seuil).*

La transition lissée est en fait réalisée par une fonction mathématique purement empirique, et ne décrit donc pas le comportement physique du TMOS dans cette région. Lors de la sortie des modèles BSIM3v3 et MM9, cela n'était pas particulièrement problématique en raison de la technologie des dispositifs de l'époque, les concepteurs de circuits utilisant alors essentiellement le TMOS en régime d'inversion forte. Cependant cela n'est absolument plus vrai de nos jours, les analogiciens travaillant de plus en plus autour de la zone d'inversion modérée, région où le gain du transistor est plus élevé qu'en inversion forte.

En outre, avec la réduction d'échelle incessante, les tensions d'alimentation ont été constamment réduites (conséquence naturelle du "scaling") et ainsi la dynamique du TMOS aussi. En conséquence, l'importance de la région d'inversion modérée — en analogique — devient réellement capitale, et l'utilisation inconditionnelle de fonction de lissage pour décrire cette région est alors objectivement discutable. À l'heure actuelle, même la simulation des circuits numériques nécessite une description précise de cette région de fonctionnement, en raison des tensions d'alimentation qui deviennent extrêmement faibles (≈ 1.2 V).

Finalement, les contraintes sévères imposées à la modélisation compacte du TMOS ont conduit au développement de nouveaux modèles, reposant sur des concepts tout à fait innovants.

2.4 Approches alternatives : « Vers les modèles de quatrième génération ? »

2.4.1 Le modèle EKV

2.4.1.1 Le modèle originel

Le modèle de transistor MOS EKV [7] est un modèle compact de simulation lié à la géométrie et bâti sur les propriétés physiques fondamentales de la structure MOS. Les approches mises en oeuvre dans ce modèle diffèrent radicalement de celles employées par les modèles BSIM3/4 et MM9. Le modèle EKV a été initialement développé pour des applications basse tension – basse puissance dans lesquelles les transitions entre les différents modes de fonctionnement du TMOS doivent être véritablement bien décrites. En particulier, les exigences requises pour la simulation correcte d'un circuit analogique CMOS sont la précision et la continuité des caractéristiques petit et grand signaux. Jusqu'en 1994, date où le modèle EKV

a été présenté pour la première fois, aucun modèle continu n'était disponible dans le domaine public. Originellement, le modèle EKV était basé sur une interpolation mathématique du courant de drain, d'une part entre l'inversion faible et l'inversion forte, et d'autre part entre le régime de conduction et celui de saturation [7,8]. Ce modèle a introduit des concepts novateurs, menant à la définition de variables normalisées.

Un point capital est que contrairement aux autres modèles compacts de type SPICE, le modèle EKV définit toutes ses tensions par rapport au *substrat*, et non plus par rapport à la source. La référence au substrat permet alors au modèle de se comporter symétriquement par rapport à la source et au drain, symétrie qui est inhérente à toute structure MOSFET.

Clairement, des expressions maniables pour décrire les caractéristiques des dispositifs en termes de paramètres de process et de conditions de polarisation sont très souhaitables pour la conception de circuits analogiques. Dans ce contexte, le modèle EKV offre une alternative, non seulement par rapport aux modèles formulés en tension de seuil, mais aussi par rapport aux modèles formulés en potentiel de surface ; c'est un modèle basé sur le concept de *linéarisation de la charge* (linearization charge model) [7,22,25–29]. La première version originale d'EKV introduit le concept de linéarisation de la charge par rapport au potentiel de canal V_{ch} (i.e. le potentiel de quasi-Fermi) [7,30].

Dans un but comparatif, reprenons l'exemple du calcul du courant de drain, qui nous a déjà permis d'illustrer les différences entre les modèles compacts physiques et classiques. Avec la version originale du modèle EKV, le courant total est défini par deux expressions distinctes pour l'inversion faible et forte, reprenant en ce sens la philosophie des modèles à base de tension de seuil ; mais en réalité la ressemblance s'arrête à ce stade[1]. En effet, dans EKV, le courant de drain est directement dérivé de l'équation (2.10), que nous rappelons ici :

$$I_d = -\mu_n \cdot W \cdot Q_{inv} \cdot \frac{dV_{ch}}{dx} \tag{2.33}$$

Bien que cette équation soit valable dans toute la gamme de fonctionnement, et donc en

[1]Ici tout le raisonnement se positionne dans le contexte de la linéarisation de la charge d'inversion Q_{inv}, qui est approximée comme une fonction linéaire de V_{ch}. C'est alors qu'intervient la notion de tension de pincement V_p, qui en fait, représente la tension à appliquer au canal pour annuler l'effet de la tension de grille, autrement dit, $Q_{inv} = 0$ quand $V_{ch} = V_p$. Ces trois notions de tensions de canal, de pincement et de charge d'inversion sont en quelques sortes les points clés du modèle EKV. Toutes les informations et équations utiles à la compréhension de ce modèle sont accessibles dans les références [7,8,15,26,28,29].

inversion modérée, les expressions du modèle pour le courant de drain ne sont pas au sens strict valides en inversion modérée, en raison des courants de diffusion et de conduction qui ont été considérés séparément pour les zones d'inversion faible et d'inversion forte, respectivement. Similairement aux modèles BSIM3/4 et MM9, une fonction d'interpolation (ou de lissage) est définie pour assurer la continuité entre l'inversion faible et l'inversion forte. Cependant, contrairement aux modèles BSIM3 et BSIM4 la fonction d'interpolation utilisée ici est *indépendante* des dimensions du transistor et des paramètres technologiques [30].

2.4.1.2 La version 2.6

Les modèles compacts capacitifs ont un grave défaut car ils prédisent une non-conservation de la charge, phénomène pourtant très important dans des circuits tels que les mémoires dynamiques (RAM) et les filtres à capacités commutées par exemple. En fait, la non-conservation de la charge résulte généralement plus d'une intégration numérique incorrecte des simulateurs que d'une imprécision du modèle [1]. La nécessité de prendre en compte cette conservation de la charge a conduit au développement de nouveaux modèles compacts, connus sous le nom de *modèles en charge*. Le modèle EKV, dans sa version 2.6 est basé sur la description des charges, qui sont alors définies dans l'approximation de la feuille de charge [8,31]. En résumé, cette version EKV2.6 diffère nettement du modèle EKV précédent en ce sens qu'il s'agit d'un modèle en charge [31].

2.4.1.3 Le nouveau formalisme

Notons tout d'abord que de nombreux points communs existent entre les modèles EKV 2.6 (révision III) et 3.0 (encore en cours de développement), ces deux versions étant des modèles en charge. Les concepts de base du modèle dans son nouveau formalisme sont les suivants :

En contraste avec les premières versions, EKV3.0 utilise le concept de linéarisation de la charge par rapport au potentiel de surface ϕ_s, et non plus par rapport au potentiel de canal V_{ch} [16,25,26,29]. Il en résulte un modèle plus fortement lié à la physique du TMOS, en particulier pour le courant de drain et les transconductances. Ces caractéristiques sont maintenant définies de manière continue de l'inversion faible à l'inversion forte, et cela sans utiliser de fonction d'interpolation, contrairement à EKV2.6. En réalité lors du processus de linéarisation de la charge, une fonction de lissage est implicitement utilisée. Cependant son rôle est complètement différent de celui de la fonction d'interpolation de EKV2.6 ; en effet dans ce nouveau contexte la fonction de lissage est seulement utilisée pour améliorer le comportement du modèle, et non

pas pour décrire un domaine de fonctionnement du TMOS, où aucune équation physique ne serait définie [29]. Il est important de bien comprendre que ces approches sont radicalement différentes sur le fond, en dépit leur similarité sur la forme.

En conclusion, le modèle EKV3.0 offre une alternative aux modèles formulés en potentiel de surface. Il propose des expressions analytiques implicitement liées au potentiel de surface, permettant ainsi l'élaboration d'un modèle conservant un lien fort avec la physique du dispositif. En ce sens, cette dernière version du modèle EKV se distingue clairement des modèles classiques de troisième génération tels que BSIM3/4 ou MM9. Dès lors, le qualificatif de modèle de quatrième génération serait peut-être plus adapté pour parler d'un tel modèle.

Les deux modèles auxquels nous allons encore nous intéresser sont les modèles MM11 de Philips Electronics, et SP (Surface Potential) de l'Université de Pennsylvanie. Ce sont des modèles en *quasi-potentiel de surface*, c'est-à-dire que le noyau du modèle est lié à une approximation explicite (analytique) de l'équation implicite décrivant le potentiel de surface. Ces modèles sont donc par nature très proches des lois physiques régissant le fonctionnement du TMOS. À notre avis, MM11 et SP méritent clairement le qualificatif de modèles de quatrième génération, tant ils diffèrent des approches classiquement utilisées en modélisation compacte (modèles à base de tension de seuil).

2.4.2 Le modèle MM11

Ce modèle récent (2001) est le successeur du modèle de MOS MM9 [10], en dépit de leurs approches de modélisation radicalement opposées. Les principaux objectifs de ce nouveau modèle sont d'être efficace à la fois en simulation digitale, analogique et RF, et d'être adapté à la description des TMOS fortement submicroniques des technologies les plus avancées. Le nombre de paramètres du modèle et le temps de simulation sont relativement réduits, du même ordre que pour MM9, et l'extraction de paramètres est assez simple si nous la comparons à celle de BSIM4 par exemple. Enfin, le point fondamental est que MM11 est un modèle proche de la physique, puisque complètement basé sur la description du potentiel de surface.

Au sein de MM11, les équations décrivant les différentes caractéristiques électriques (courants, charges, etc.) sont toutes basées sur des formulations du potentiel de surface. Ainsi, ces équations sont valides dans tous les régimes de fonctionnement : accumulation, déplétion et inversion. Bien que généralement le potentiel de surface dépende d'une équation implicite

et requiert donc une résolution itérative, dans MM11 il est approximé par une expression explicite [9] ; d'où notre qualificatif de modèle en quasi-potentiel de surface. Les équations du noyau du modèle sont toutes référencées par rapport au substrat (comme EKV), ce qui est très bénéfique du point de vue de la symétrie du modèle. Dans ce contexte, le courant de drain, les transconductances, les charges et les transcapacitances sont calculés dans l'approximation du modèle en feuille de charge. Par exemple, le courant de drain est obtenu sans l'emploi d'aucune fonction de lissage, il est simplement égal à la somme de ses composantes conduction et diffusion (cf. § 2.2.3, page 22).

Au cours de ce manuscrit, nous décrirons les fondements physiques sur lesquels reposent les principales approximations développées dans MM11. En fait, le noyau de ce modèle nous a servi de base dans notre travail sur la prise en compte des effets quantiques dans les TMOS ; ceci sera détaillé au chapitre 4.

2.4.3 Le modèle SP

Le modèle SP (Surface Potential), récemment développé par GILDENBLAT et CHEN est un modèle dont le but est la recherche d'une précision maximale [11,12,32,33]. Ce modèle pourrait même être qualifié de modèle en potentiel de surface plutôt que de modèle en quasi-potentiel de surface. En effet, bien que SP ne nécessite pas de processus itératifs pour la résolution du potentiel de surface, une solution mathématique performante permet d'en obtenir une approximation très précise [11]. L'approche est très différente de celle utilisée dans MM11 ; elle résulte en une approximation globale du potentiel de surface valable pour les différents régimes de fonctionnement, sans avoir à utiliser de fonctions de lissage pour relier ces différents domaines. En résumé, SP est à la fois le modèle le plus proche de la physique et aussi le plus mathématique (parmi les modèles de quatrième génération). Au regard des solutions techniques choisies pour obtenir le potentiel de surface, un tel modèle ne peut pas être utilisé pour des calculs dits « à la main » tandis que cela est réellement possible avec EKV ou MM11. De plus s'il est évident que le lien entre la technologie (physique) et les caractéristiques électriques du TMOS est inhérent à ce modèle, il est néanmoins complètement occulté par l'arsenal mathématique employé. Cette approche conviendra cependant parfaitement aux concepteurs de circuits utilisant le simulateur comme une boîte noire.

Les principales caractéristiques de SP sont les suivantes. Tout d'abord, une approximation très précise (ordre de 10 nV) du potentiel de surface est définie à partir de l'équation implicite (2.8). Le second point essentiel consiste en une *linéarisation symétrique* de la charge d'inversion, basée sur la définition d'un potentiel de surface moyen ($\phi_m = 1/2 \cdot (\phi_{s0} + \phi_{sL})$). Ainsi la symétrie entre la source et le drain est totalement préservée. Une similitude entre les approches EKV et SP est donc liée à la linéarisation de la charge d'inversion. Cependant elle n'est pas du tout réalisée de la même façon. Dans EKV3.0, la charge est linéarisée par rapport au potentiel de surface, ce qui justement permet de s'affranchir de la connaissance de ce dernier : la linéarisation de la charge est en fait réalisée en termes de charges. Le modèle SP est quant à lui un véritable modèle en potentiel de surface, en conséquence la linéarisation de la charge est effectuée en termes de potentiel de surface. Enfin, précisons que l'utilisation de ce modèle dans le cadre d'un modèle en feuille de charge mène fort logiquement à de bons résultats, que ce soit pour le courant, les charges, ainsi que pour les différentes dérivées de ces grandeurs.

2.4.4 Synthèse des caractéristiques des modèles avancés

Le tableau ci-dessous récapitule les caractéristiques majeures des principaux modèles compacts dits de troisième et quatrième générations.

TAB. 2.2 : *Comparaison des caractéristiques fondamentales des modèles compacts les plus couramment utilisés. En ce qui concerne le 'référençage' des modèles (4ème ligne du tableau), nous faisons allusion aux équations du noyau des modèles, ce qui explique les possibles différences avec des comparatifs similaires existants [34,35].*

	BSIM 3v3	**BSIM 4**	**MM 9**	**MM 11**	**EKV 3.0**	**SP**
Méthode	Analytique	Analytique	Analytique	Analytique	Analytique	Analytique
Inversion	V_{th}-based	V_{th}-based	V_{th}-based	ϕ_s-based	Hybride	ϕ_s-based
Courant DC	Drift	Drift	Drift	Drift-diff.	Drift-diff.	Drift-diff.
Référence	Source	Source	Source	Substrat	Substrat	Substrat
Symétrie	Non	$\sim$ Oui	Non	Oui	Oui	Oui

2.5 Conclusion

Au fil de ce chapitre, nous avons dressé une liste des principaux modèles compacts destinés à la simulation de circuits. Cette liste est forcément non-exhaustive, puisque de multiples dérivations des modèles présentés ici existent ; cependant ces derniers définissent tous les principes fondamentaux généralement mis en oeuvre en modélisation compacte.

Les modèles compacts les plus utilisés actuellement sont les modèles développés à Berkeley, à savoir BSIM3 et BSIM4. Ces modèles sont conçus autour d'approximations régionales, jointes les unes aux autres par des fonctions de lissage empiriques, utilisées non seulement pour lisser les équations mais aussi pour décrire certains domaines de fonctionnement (cas de la zone d'inversion modérée). Au jour d'aujourd'hui, cela devient une approche très discutable, tout particulièrement en raison de la baisse drastique du niveau des tensions d'alimentation. En conséquence l'inversion modérée devrait être physiquement modélisée, aussi bien pour la conception analogique que numérique. Cela n'est pourtant pas le cas au sein des modèles formulés en tension de seuil, modèles qui en fait reposent plus sur leurs fonctions de lissage que sur les approximations physiques disponibles en inversion faible et forte. Finalement, ce genre d'approche a malheureusement mené à la tendance actuelle d'augmentation permanente de la complexité du modèle et du nombre de paramètres employés.

Les modèles basés sur une formulation explicite du potentiel de surface se positionnent donc en alternative aux modèles précédents. Ils sont très proches de la physique du dispositif et permettent de façon naturelle l'utilisation d'un modèle en feuille de charge. D'un point de vue général, ces modèles représentent une tentative visant à accroître le contenu physique du corps du modèle, et ce faisant à réduire sa complexité et son nombre de paramètres. Il nous paraît donc que les modèles EKV3.0, MM11 et SP sont l'avenir de la modélisation compacte, de par leur nature simple, physique et efficace. En conclusion, notre réponse à la question précédemment posée : « Approches alternatives, vers les modèles de quatrième génération ? » est très clairement, oui.

L'approche de modélisation choisie au cours du développement des modèles réalisés durant cette thèse est double. S'il est vrai que toute notre attention est maintenant focalisée sur les modèles en potentiel de surface, il convient de pas oublier que la grande majorité des simulateurs de circuits utilise actuellement des modèles plus classiques, en tension de seuil.

Dans cet esprit, nous avons développé une nouveau modèle de capacités extrinsèques qui est directement et facilement implémentable dans n'importe quel type de modèle : de BSIM3 à MM11, en passant par EKV. Ce modèle sera présenté au chapitre 3.
Dans un second temps, nous nous sommes entièrement investi dans les modèles à base de potentiel de surface. Notre objectif principal est la description et la prise en compte des effets de mécanique quantique au sein des TMOS, et cela de l'accumulation à l'inversion. Dans ce but, nous avons défini des nouveaux concepts analytiques, menant à l'élaboration d'un modèle formulé uniquement en termes de potentiel de surface. Cette étude, qui représente la majeure partie de ce manuscrit sera détaillée au chapitre 4.

Bibliographie

[1] N. Arora, *MOSFET Models for VLSI Circuit Simulation. Theory and Practice*. New York : Springer-Verlag, 1993, ISBN 3-211-82395-6.

[2] M. Bucher, "Analytical MOS transistor modelling for analog circuit simulation," Ph.D. dissertation, EPFL, 1999, thèse no. 2114.

[3] Y. Tsividis and K. Suyama, "MOSFET modeling for analog circuit CAD : Problems and prospects," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 210–216, 1994.

[4] W. Liu *et al.*, *BSIM3v3.2 MOSFET model – Users' manual*, Univ. of California, Berkeley, 1998. [Online]. Available : http ://www-device.eecs.berkeley.edu/˜bsim

[5] *BSIM 4.2 Manual*, Univ. of California, Berkeley, Apr. 2001. [Online]. Available : http ://www-device.eecs.berkeley.edu/˜bsim

[6] *MM9 Model*, Philips Semiconductors, 2001. [Online]. Available : http ://www.semiconductors.philips.com/philips_models/mos_models

[7] C. C. Enz, F. Krummenacher, and E. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *J. Analog Integrated Circuits and Signal Processing*, no. 8, pp. 83–114, 1995.

[8] M. Bucher, C. Lallement, C. C. Enz, F. Théodoloz, and F. Krummenacher, "The EPFL–EKV MOSFET model, version 2.6," EPFL, Tech. Rep., 1999. [Online]. Available : http ://legwww.epfl.ch/ekv/

[9] R. van Langevelde and F. M. Klaassen, "An explicit surface-potential-based MOSFET model for circuit simulation," *Solid-State Electron.*, vol. 44, pp. 409–418, 2000.

[10] D. B. M. Klaassen, R. van Langevelde, A. J. Scholten, *et al.*, *The MOS model, level 1101*, Philips Research Laboratories, 2002. [Online]. Available : http ://www.semiconductors.philips.com/philips_models/mos_models/model11

[11] T. L. Chen and G. Gildenblat, "Analytical approximation for the MOSFET surface potential," *Solid-State Electron.*, vol. 45, pp. 335–339, 2001.

[12] G. Gildenblat and T. L. Chen, "Overview of an advanced surface-potential-based MOSFET model (SP)," in *Proc. 5th Int. Conf. on Modeling and Simulation of Microsystems (MSM)*, 2002.

[13] H. J. Mattausch *et al.*, "HiSIM : The first complete drift-diffusion MOSFET model for circuit simulation," in *Proc. IEEE Int. Conf. Solid-state and Integrated Circuit Technology*, 2001, pp. 861–866.

[14] M. Miura-Mattausch *et al.*, "HiSIM : Self-consistent surface-potential MOS-Model valid down to sub-100nm technologies," in *Proc. 5th Int. Conf. on Modeling and Simulation of Microsystems (MSM)*, 2002.

[15] M. Bucher, J.-M. Sallese, and C. Lallement, *Proc. IEEE Int. Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*. New York : D. Tsoukalas, C. Tsamis, Springer Vienna, pp. 296–299, 2001, ISBN 3-211-83708-6.

[16] C. Lallement, J.-M. Sallese, M. Bucher, W. Grabinski, and P. Fazan, "Accounting for quantum effects and polysilicon depletion from weak to strong inversion, in a charge

based design-oriented MOSFET model," *IEEE Trans. Electron Devices*, vol. 50, no. 2, pp. 406–417, Feb. 2003.

[17] H. C. Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*. New York : Springer-Verlag, 1990, ISBN 0387821368.

[18] *Atlas User's Manual*, SILVACO International, 1997.

[19] H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electron.*, vol. 9, pp. 927–937, 1966.

[20] J. R. Brews, "A charge sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, pp. 345–355, 1978.

[21] D. Ward and R. W. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE Journ. of Solid-State Circuits*, vol. 13, pp. 703–707, 1978.

[22] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*. New York : McGraw-Hill Book Company, 1987, ISBN 0-19-517014-8.

[23] D. Foty, *MOSFET Modeling with Spice. Principles and Practice*. Prentice Hall, 1997, ISBN 0-13-227935-5.

[24] C. Lallement, "Modélisation du transistor MOSFET pour la conception de circuits et technologies fortement submicroniques," 2001, cours de DEA, ULP, Strasbourg.

[25] B. Iñíguez and E. García-Moreno, "A physically based C-continuous model for small-geometry MOSFET's," *IEEE Trans. Electron Devices*, vol. 42, no. 2, pp. 283–287, 1995.

[26] A. Cunha, M. C. Schneider, and C. Galup-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electron.*, vol. 38, pp. 1945–1952, 1995.

[27] A. I. A. Cunha, M. C. Schneider, and C. Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE Journ. of Solid-State Circuits*, vol. 33, no. 10, pp. 1510–1519, Oct. 1998.

[28] J.-M. Sallese, M. Bucher, and C. Lallement, "Improved analytical modeling of polysilicon depletion," *Solid-State Electron.*, vol. 44, pp. 905–912, 2000.

[29] J.-M. Sallese, M. Bucher, F. Krummenacher, and P. Fazan, "Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model," *Solid-State Electron.*, vol. 47, no. 4, pp. 677–683, Apr. 2003.

[30] C. Lallement, "Le transistor MOSFET : Etudes, modélisation, et applications dans les S.O.C." 2002, habilitation à Diriger des Recherches, ULP, Strasbourg.

[31] M. Bucher, J.-M. Sallese, C. Lallement, W. Grabinski, C. C. Enz, and F. Krummenacher, "Extended charges modeling for deep submicron CMOS," in *Proc. International Semiconductor Device Research Symposium (ISDRS)*, Dec. 1999, pp. 397–400.

[32] G. Gildenblat, T. L. Chen, and P. Bendix, "Analytical approximation for perturbation of MOSFET surface potential by polysilicon depletion layer," *Electronics Letters*, vol. 35, no. 22, pp. 1974–1976, Oct. 1999.

[33] ——, "Closed-form approximation for the perturbation of MOSFET surface potential by quantum-mechanical-effects," *Electronics Letters*, vol. 36, no. 12, pp. 1072–1073, June 2000.

[34] Http ://www.cs.berkeley.edu/˜hara/models/.

[35] F. Prégaldiny, C. Lallement, W. Grabinski, J.-B. Kammerer, and D. Mathiot, “An analytical quantum model for the surface potential of deep-submicron MOSFETs,” in *Proc. 10th International Conference on Mixed Design of Integrated Circuits and Systems (MIXDES)*, June 2003, pp. 98–104.

Chapitre 3

Aspects extrinsèques du transistor MOS

3.1 Introduction

La miniaturisation des transistors MOS résulte en l'apparition de nouveaux effets, ou plus précisément, disons que certains effets jusqu'alors négligeables doivent maintenant être pris en compte. Parmi les nouveaux phénomènes à prendre en compte, deux types sont à distinguer : d'une part, il y a les effets liés au fonctionnement normal du TMOS, dits effets *intrinsèques*, et d'autre part les effets *extrinsèques*, qui eux perturbent le fonctionnement intrinsèque.

Au cours de ce chapitre nous nous intéresserons à cette seconde catégorie, et plus exactement au cas des capacités extrinsèques — ou parasites — du TMOS fortement submicronique.

L'estimation des capacités parasites dans un dispositif MOSFET est très importante, et cela particulièrement pour la simulation des circuits mixtes et RF (Radio Frequency). En raison de la diminution de la taille des composants, la capacité extrinsèque[1] devient une fraction importante de la capacité d'oxyde de grille. Dans les TMOS de grande géométrie, la capacité extrinsèque C_{ext} était considérée comme constante ou nulle, ce qui pour un transistor de grande longueur ($L_g > 10$ μm) était une approximation relativement correcte. De nos jours, le caractère extrinsèque joue un rôle majeur dans toutes les technologies fortement submicroniques ($L_g < 0.25$ μm), et perturbe fortement le comportement intrinsèque conventionnel des dispositifs. Dès lors, la modélisation fiable du comportement dynamique du transistor MOS impose une description précise des deux types de capacités, intrinsèque et extrinsèque. Avec la réduction de longueur de canal du TMOS, les capacités liées aux régions

[1] À partir de maintenant, nous regroupons sous le terme *capacité extrinsèque* l'ensemble des trois capacités parasites du TMOS, à savoir la capacité de recouvrement (ou d'overlap) et les capacités de bord interne et externe.

de la source et du drain commencent à jouer un rôle essentiel pour expliquer le comportement capacitif global (*C–V*) des MOSFET de technologie LDD (Lowly Doped Drain) [1–3]. En d'autres termes, le caractère extrinsèque devient très important, en particulier dans le régime sans courant (état bloqué du TMOS). Vu les avancées technologiques rapides, il devient donc indispensable de développer des modèles basés sur la physique, capables de prédire au mieux la capacité extrinsèque. En particulier, un modèle performant devrait pouvoir décrire une grandeur donnée indépendamment de la technologie, d'où la nécessité d'augmenter le contenu physique des modèles. De plus, comme nous allons le voir dans la suite de ce chapitre, la capacité extrinsèque ne peut plus être considérée comme une constante, elle est en fait fortement dépendante de la polarisation de grille appliquée au TMOS.

3.2 Présentation du dispositif MOSFET

3.2.1 La structure MOSFET

Nous avons déjà succinctement présenté le TMOS au chapitre précédent, en particulier sa structure géométrique de base, à la Fig. 2.2. Nous allons maintenant décrire plus en détail ces caractéristiques technologiques. La Fig. 3.1 représente une vue en coupe du transistor n-MOS LDD qui nous a servi de référence lors de notre étude du comportement extrinsèque.

Les principales caractéristiques technologiques de ce TMOS sont les suivantes :

- L'épaisseur d'oxyde : t_{ox},
- l'épaisseur du polysilicium de grille : t_{poly},
- la longueur de grille : L_g,
- la profondeur de jonctions source/drain : X_j,
- la longueur de la diffusion latérale des régions de source/drain : L_d,
- la dose de dopants des zones source/drain : N_j,
- la dose de dopants des extensions des zones source/drain, i.e. des zones LDD : N_{ldd},
- le niveau de dopage du substrat[1] : N_b,
- et enfin, le niveau dopage du canal : N_s.

[1] Au cours de ce manuscrit nous appelons N_a le dopage substrat quand celui-ci est similaire à celui du canal ; dans le cas contraire, i.e. en présence d'un dopage clairement non-uniforme, le dopage substrat est dénoté par N_b.

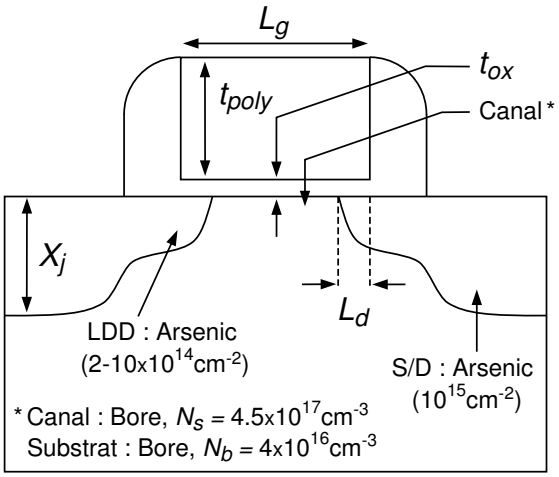


FIG. 3.1 : *Représentation en 2D d'un transistor n-MOS de technologie LDD.*

Les zones LDD sont un élément jouant un rôle majeur dans le comportement extrinsèque du TMOS. Ces zones plus faiblement dopées que les régions de source et drain, ont été introduites dans les technologies submicroniques dans le but d'améliorer la fiabilité et le vieillissement du dispositif. En effet, de par leur plus faible dopage, elles limitent le champ électrique à l'interface Si–SiO_2, réduisant ainsi le nombre de porteurs chauds susceptibles de dégrader l'oxyde de grille. Cependant, comme toujours en physique, ce que l'on gagne d'un côté, on le perd de l'autre : si les extensions LDD améliorent la fiabilité du TMOS, elles réduisent en contre-partie ses performances intrinsèques telles que la valeur maximale du courant de drain, la vitesse de commutation, etc. [4]. Elles sont responsables en particulier de l'augmentation de valeur des résistances séries et de la dépendance aux tensions externes de la capacité extrinsèque. Finalement, nous retrouvons dans cet exemple toute la difficulté des compromis à réaliser lors de la réduction d'échelle des composants.

3.2.2 Le régime extrinsèque

Le fonctionnement du transistor MOS est habituellement divisé en trois régions, la région d'accumulation, de déplétion et d'inversion (cf. Fig. 2.3, chapitre 2). De manière plus générale, nous pouvons scinder le fonctionnement du TMOS en deux régimes, intrinsèque et extrinsèque. Le premier correspond à la description classique du dispositif, c'est-à-dire qui ne tient pas compte des éléments parasites qui perturbent et modifient le fonctionnement intrinsèque. Contrairement à la distinction accumulation–déplétion–inversion, la distinction régime intrinsèque/extrinsèque ne repose pas uniquement sur une division de la plage de tension

de grille appliquée au dispositif. Illustrons ceci par un exemple simple : la comparaison du cas de la résistance série et de la capacité d'overlap. La résistance série (en fait la résistance extrinsèque) intervient lorsque le dispositif est actif, i.e. quand un courant circule entre la source et la drain, le TMOS est donc en mode d'inversion. De façon très différente, la capacité d'overlap — une composante majeure de la capacité extrinsèque — joue un rôle dans le comportement capacitif sur toute la plage de fonctionnement du TMOS, de l'accumulation à l'inversion. Par contre c'est sa modélisation en régime sans courant qui en réalité est un caractère extrinsèque, puisque sa valeur en régime d'inversion est parfaitement connue uniquement d'après les paramètres technologiques du procédé de fabrication. En résumé, pour la résistance série, régime extrinsèque = régime d'inversion, tandis que pour la capacité d'overlap, régime extrinsèque = régime sans courant (accumulation et déplétion). Ainsi, il n'est pas possible de définir au sens large la notion de région intrinsèque ou extrinsèque en termes de modes de fonctionnement (i.e. de tension de grille), ce type de dénomination étant directement liée à la grandeur physique ou électrique considérée. En conclusion, il faut donc avoir conscience que le parallèle classique tendant à assimiler régime sans courant et régime extrinsèque est dans l'absolu un raisonnement erroné.

3.2.3 Les capacités parasites

Comme nous l'avons déjà mentionné dans l'introduction, l'estimation des capacités parasites du TMOS est très importante, notamment en simulation de circuits mixtes. En effet, au sein des TMOS LDD fortement submicroniques, la capacité extrinsèque C_{ext} devient une fraction de plus en plus importante de la capacité totale de grille C_{ox}. Ainsi, une modélisation précise de cette grandeur est maintenant indispensable. Dans un souci de clarté, au cours de ce manuscrit, nous avons regroupé sous le terme *capacité extrinsèque* l'ensemble des trois principales capacités parasites, à savoir, la capacité d'overlap C_{ov}, la capacité de bord interne C_{if} (inner fringing capacitance) et la capacité de bord externe C_{of} (outer fringing capacitance). Cette distinction des différentes composantes de la capacité extrinsèque est importante, puisqu'elle est à la base d'une modélisation cohérente de C_{ext}. Il faut naturellement que chacune de ces composantes soit correctement décrite pour obtenir un modèle viable. C'est ce que nous nous proposons de faire au cours de cette étude.

De nombreux auteurs se sont déjà intéressés au traitement de la capacité extrinsèque [1–3,5–13]. Cependant, jusqu'à maintenant, aucun modèle simple et disponible dans la littérature

ne décrivait correctement toutes les composantes de la capacité extrinsèque. Un des premiers modèles prenant en compte la dépendance de la capacité d'overlap à la polarisation de la grille a été proposé par CETNER dans [5]. Les résultats de ce modèle ne sont toutefois pas valables pour des dispositifs submicroniques.

D'autres approches prenant en compte la dépendance de C_{ov} vis à vis de la tension de grille ont été décrites en utilisant une relation empirique simple, comme celle présentée dans [6,7]. En ce qui concerne la capacité de bord interne, seulement deux modèles considèrent les effets de C_{if}, mais ils ne sont pas valables pour les TMOS fortement submicroniques [8,9]. Bien que cette capacité joue aussi un rôle dans le comportement capacitif extrinsèque global, même un modèle aussi récent que BSIM4v2 n'en tient pas compte. Cela est d'autant plus surprenant que dans le manuel de ce même modèle [10], il est expliqué que cette capacité devrait être prise en compte, et qu'en plus elle est fortement dépendante de la tension appliquée à la grille. Comme nous le démontrerons par la suite, négliger la capacité de bord interne engendre des erreurs dans le calcul de la capacité extrinsèque quand le TMOS est en régime de déplétion.

Un modèle compact en charge développé pour les TMOS LDD a été proposé par KLEIN dans [1]. Ce modèle a ensuite été incorporé dans BSIM3v3.1 [2]. Bien que ce modèle en charge repose sur des hypothèses physiques tangibles et soit donc relativement précis, les résultats publiés montrent cependant clairement que la capacité de bord interne n'a pas été modélisée. De plus, ce modèle est très difficile à implémenter au sein d'un simulateur de circuits, ce qui rend son usage plus ou moins hypothétique. Par exemple, nous n'avons pas réussi à tester ce modèle sous Mathcad Professional®, pour diverses raisons : soit nous obtenions des résultats surréalistes, soit le simulateur ne convergeait pas.

Dans l'objectif de développer un nouveau modèle de capacité extrinsèque, simple et physiquement cohérent, nous avons tout d'abord grâce à de multiples simulations numériques 2D (procédé et dispositif), étudié le comportement capacitif extrinsèque global. Nous avons ensuite mis au point un nouveau modèle semi-empirique décrivant l'évolution de la capacité d'overlap. D'autre part, pour la première fois en modélisation compacte, nous avons présenté un modèle lié à la technologie du TMOS, et décrivant la capacité de bord interne et sa forte dépendance à la tension de grille [14].

3.3 Simulations numériques 2D

Nous avons dans un premier temps réalisé une campagne extensive de simulations numériques 2D, dans le but d'étudier les phénomènes physiques et électriques liés aux capacités parasites. Les simulations de *procédés* (process) ont été obtenues à partir d'un fichier décrivant les caractéristiques technologiques d'un n-MOSFET LDD de 0.25 μm. Toutes les données du procédé de fabrication de ce TMOS (traitements thermiques, énergies d'implantation ionique, etc.) sont dérivées d'une technologie réelle, développée au LETI. Nous remercions d'ailleurs sincèrement le LETI de nous avoir offert l'accès à ces données technologiques. Précisons que le simulateur de procédés employé lors de nos études est Athena®, et le simulateur électrique associé est Atlas®. Ces deux simulateurs sont des outils propriétaires de la société Silvaco [15,16].

La structure développée au cours des simulations de procédés correspond à celle précédemment montrée à la Fig. 3.1. Les paramètres technologiques associés à ce transistor sont résumés au Tableau 3.1.

TAB. 3.1 : *Paramètres technologiques des TMOS réalisés lors des simulations de procédés.*

L_g	Longueur physique de grille	0.2 à 10 μm
L_{eff}	Longueur effective de grille	$L_g - 2 \cdot L_d$
L_d	Longueur de diffusion latérale des source/drain	30 à 40 nm
t_{ox}	Épaisseur d'oxyde de grille	4.5 nm
t_{poly}	Épaisseur du polysilicium de grille	148 nm
X_j	Profondeur des jonctions source/drain	60 nm
N_s	Dopage du canal (bore)	4.5×10^{17} cm^{-3}
N_b	Dopage du substrat (bore)	4×10^{16} cm^{-3}
N_j	Dose de dopants des source/drain (As)	10^{15} cm^{-2}
N_{ldd}	Dose de dopants des zones LDD (As)	2×10^{14} à 10^{15} cm^{-2}

Les TMOS de différentes longueurs de grille (de 0.2 à 10 μm) que nous avons simulés sont tous basés sur les caractéristiques technologiques décrites dans le tableau ci-dessus. Cependant pour une longueur de grille fixée, nous avons créé plusieurs variantes d'une même structure, en modifiant la dose de dopants des extensions de source et drain (N_{ldd}). Cela explique les variations des deux autres grandeurs L_d et L_{eff} dans le tableau. En effet, plus le dopage des

zones LDD est élevé, plus cette région diffuse sous la grille, ce qui augmente la longueur de diffusion latérale L_d et donc réduit la longueur effective de grille ($L_{eff} = L_g - 2 \cdot L_d$).

3.3.1 Mise en évidence de la capacité extrinsèque

Les phénomènes parasites sont de façon inhérente présents dans tous les transistors MOS, et ceci indépendamment de leur dimension. L'influence de la réduction de la longueur de grille résulte juste en l'accentuation du caractère dominant de tel ou tel effet, par rapport au fonctionnement conventionnel (ou intrinsèque) du dispositif. En fait, ce qu'il est réellement important de considérer, bien avant la réduction de dimension en tant que telle, est l'évolution des choix technologiques qui accompagnent cette miniaturisation des composants. Ce sont ces mêmes choix qui entraînent l'importance accrue des divers phénomènes parasites. Prenons un exemple : soit à longueur de grille fixée, deux générations de TMOS différentes : une 1 μm (technologie 'a' non-LDD) et une 0.25 μm (technologie 'b' LDD). 'a' et 'b' se comportent alors distinctement face aux effets extrinsèques, en raison des choix technologiques différents mis en oeuvre lors de leur conception respective. Prenons alors deux TMOS de longueur de grille de 2 μm, réalisés en technologies 'a' et 'b' ; ces deux transistors auront chacun une capacité d'overlap non nulle en régime sans courant, mais cependant en raison de l'introduction des zones LDD dans la technologie 'b', le comportement de la capacité d'overlap du transistor 'b' sera fortement différent de celui du transistor 'a'. C'est principalement pour ce genre de raison que les modèles de transistor ont besoin d'évoluer au fil des nouvelles générations de TMOS, l'apparition de nouveaux choix technologiques obligeant à repenser les modèles compacts. De ce point de vue, les modèles basés sur la physique (technologie) du dispositif permettent d'obtenir très clairement de meilleurs résultats dans la durée.

Nous allons maintenant présenter les résultats des simulations électriques réalisées sur les TMOS générés par le simulateur de procédés 2D. La Fig. 3.2 représente la caractéristique $(C_{dg} + C_{sg})$[1] en fonction de la tension de grille V_{gb} pour un TMOS à canal très court (0.2 μm) et deux TMOS à canaux courts (0.5 et 1 μm). Le descriptif technologique des TMOS simulés correspond à celui décrit au Tableau 3.1, la dose de dopants LDD étant fixée à 2×10^{14} cm^{-2}, ce qui correspond aux paramètres standards de cette technologie 0.25 μm. Cette figure illustre clairement la séparation des régimes de fonctionnement intrinsèque et extrinsèque, d'un point de vue capacitif. Le point de croisement des trois courbes sépare la capacité extrinsèque de la

[1]Rappel : $C_{ij} = dQ_i/dV_j$ où i et j sont égales à g, b, s ou d (grille, substrat, source ou drain).

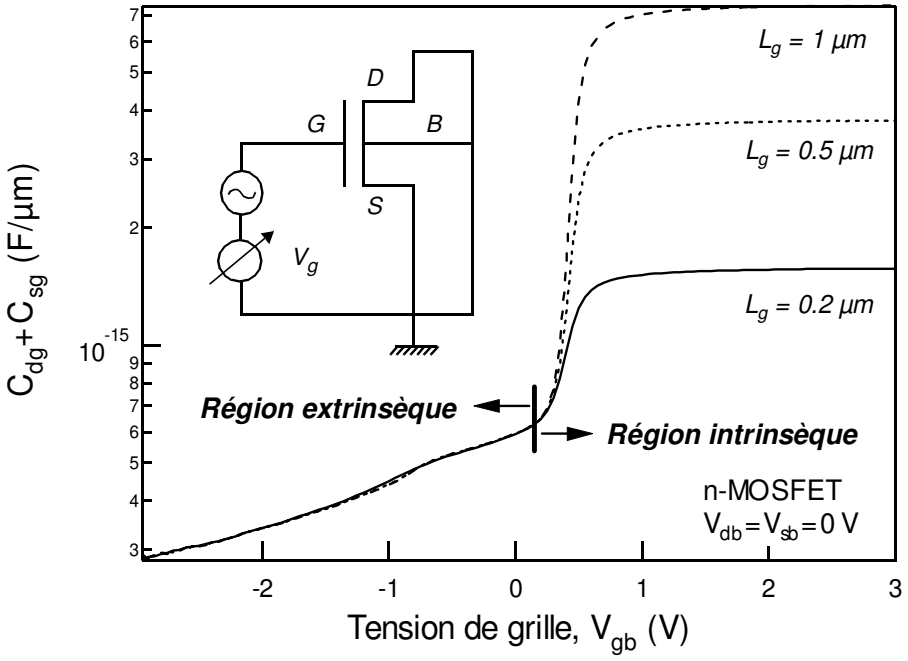


FIG. 3.2 : *Simulation des caractéristiques* $(C_{dg} + C_{sg})$ *en fonction de la tension de grille* V_{gb} *pour des TMOS LDD de différentes longueurs. Les courbes sont représentées en échelle semi-log. En insert, le principe des simulations électriques réalisées.*

capacité intrinsèque liée au canal. Comme nous le verrons par la suite, ce point correspond en fait au maximum de la capacité C_{ext}.

Il apparaît aussi un caractère fondamental de la capacité extrinsèque à la Fig. 3.2 : en régime sans courant (région extrinsèque), cette capacité est indépendante de la longueur du dispositif. Cependant, avec la décroissance de L_g, C_{ext} devient une fraction de plus en plus importante de la capacité $(C_{dg} + C_{sg})$, et donc ne peut plus être négligée. Cette figure illustre que même dans le cas d'un TMOS de grande géométrie, le fait de considérer la capacité C_{ext} comme une constante est faux.

Nous avons pour l'instant supposé que la capacité extrinsèque est dépendante de la polarisation de la grille, sans hypothèse particulière quant aux autres polarisations (tension de drain V_{db} ou source V_{sb}). Nous justifierons un peu plus tard que négliger la dépendance de C_{ext} à la polarisation du drain (ou de la source) est une approximation valable.

Nous définissons alors la capacité extrinsèque par l'expression suivante :

$$C_{ext} = C_{ext}(V_g, V_s, V_d) \simeq C_{ext}(V_g) \tag{3.1}$$

Dans un souci de lisibilité des équations, à partir de maintenant et tout au long de ce manuscrit, nous considérons que les potentiels sont référencés par rapport au substrat, i.e. $V_x \equiv V_{xb}$ avec $x = g, d$ ou s.

Pour comprendre et modéliser la dépendance de C_{ext} vis à vis de la tension de grille, il faut d'abord décrire toutes les composantes de cette capacité :

$$C_{ext}(V_g) = 2 \cdot (C_{ov}(V_g) + C_{if}(V_g) + C_{of}) \tag{3.2}$$

où C_{ov} est la capacité d'overlap plate associée au champ électrique allant de la grille à la région d'overlap du drain ou de la source, C_{if} est la capacité de bord interne associée au champ électrique provenant de la jonction métallurgique du drain ou de la source et se terminant sous la grille de polysilicium, et C_{of} est la capacité de bord externe associée au champ électrique provenant du côté de la grille de polysilicium, traversant l'espaceur et allant jusqu'à la région de source ou de drain. Précisons déjà que la capacité de bord externe C_{of} est indépendante des tensions appliquées à la structure MOSFET.

Les trois composantes de la capacité extrinsèque d'un n-MOSFET LDD sont présentées à la Fig. 3.3.a. La notation V_{fbldd} utilisée dans ce schéma correspond à la tension de bandes plates des régions LDD, qui au premier ordre est quasiment égale à zéro (cf. paragraphe suivant).

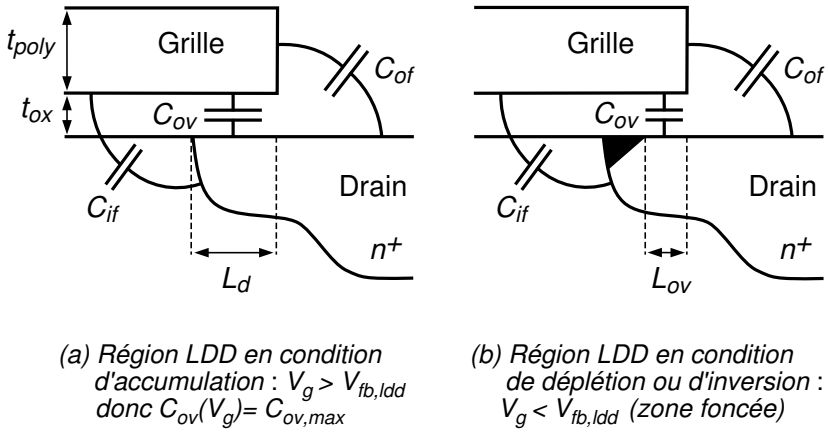


FIG. 3.3 : *Schéma illustrant (a) les différentes composantes de la capacité extrinsèque C_{ext} et (b) la capacité d'overlap C_{ov} comme une fonction de la longueur de diffusion latérale effective L_{ov}, qui est elle-même dépendante de la polarisation de la grille. Cette figure montre le cas de la région de drain ; le raisonnement étant totalement symétrique pour le cas de la source.*

3.3.2 Influence des différentes composantes de la capacité extrinsèque

3.3.2.1 La capacité d'overlap

Dans cette section, nous allons étudier le comportement de la capacité d'overlap dans le but de bien comprendre l'allure des caractéristiques *C–V* en régime extrinsèque [12,13]. Une méthode efficace pour décrire le comportement dynamique de la capacité d'overlap C_{ov} consiste à définir une longueur de diffusion latérale effective L_{ov} (cf. Fig. 3.3.b) telle que :

$$L_{ov} = L_{ov}(V_g) \quad \text{et,} \tag{3.3}$$

$$L_{ov}(V_g)\big|_{V_g \geqslant V_{fbldd} \simeq 0} = L_d \tag{3.4}$$

Dans (3.4), nous faisons l'hypothèse que la tension de bandes plates des régions LDD peut être approximée par zéro. Cette approximation a pour seul but de simplifier les équations du modèle ; son intégration au modèle étant immédiate, si besoin est (cf. § 3.4.4). Voici sa définition dans le cas d'un TMOS de type n :

$$V_{fbldd} = \phi_t \cdot \ln\left(\frac{\tilde{N}_{ldd}}{N_p}\right) \tag{3.5}$$

où $\tilde{N}_{ldd}$ et N_p représentent les dopages moyens des extensions LDD et du polysilicium de grille. Dans le cas de notre technologie, nous avons un dopage moyen LDD d'environ 5×10^{19} cm^{-3} et un dopage polysilicium moyen d'environ 6×10^{19} cm^{-3}, ainsi il est clair que (3.5) tend vers zéro.

Ceci étant maintenant défini, le comportement de la capacité d'overlap peut être physiquement et simplement décrit à partir de la Fig. 3.3 et des équations (3.3)–(3.4). En fonction de la polarisation de la grille, la nature de la charge stockée dans la région LDD varie. Cette charge consiste soit (a) en une accumulation d'électrons lorsque la région LDD est en mode d'accumulation, soit (b) en une charge positive fixe lorsque la région LDD est en mode de déplétion, ou bien même en une charge de trous délivrés par le substrat quand la région LDD entre en inversion. La Fig. 3.3 illustre ces deux situations (a) et (b).

Ainsi, lorsque la région LDD commence à être déplétée (Fig. 3.3.b), la longueur de diffusion latérale L_{ov} est réduite par rapport à sa valeur d'origine, i.e. sa valeur technologique L_d ; d'où le nom de diffusion latérale *effective* donnée à L_{ov}. En conséquence, la capacité d'overlap C_{ov} diminue quand la tension appliquée à la grille décroît de zéro vers des valeurs négatives. En

résumé, la valeur maximale (classique) de la capacité d'overlap est donnée par l'expression :

$$C_{ov,\max} = C_{ox} \cdot L_d \tag{3.6}$$

Cette équation a longtemps été employée pour modéliser la capacité d'overlap du TMOS. En réalité, elle entraîne d'importantes erreurs car elle est valable uniquement en régime intrinsèque, i.e. quand la capacité d'overlap est maximum et indépendante de la tension de grille appliquée au dispositif. Une modélisation plus fine, prenant en compte la diminution de la capacité d'overlap en régime extrinsèque est alors donnée par :

$$C_{ov}(V_g) = C_{ox} \cdot L_{ov}(V_g) \tag{3.7}$$

où $C_{ox} = \epsilon_{ox}/t_{ox}$ est la capacité d'oxyde de grille, par unité de surface. Notons que (3.6) et (3.7) sont exprimées par unité de largeur du dispositif.

L'objectif est maintenant de définir une expression modélisant la variation de L_{ov} en fonction de la polarisation appliquée à la grille. Nous présenterons notre nouvelle approche au § 3.4, page 64.

3.3.2.2 La capacité de bord interne

L'estimation de la capacité de bord interne est ignorée par tous les modèles compacts actuels. Différents modèles ont cependant déjà été proposés dans la littérature, dont un utilisable en modélisation compacte : celui de SCHRIVASTAVA [8]. Ce modèle est basé uniquement sur des considérations géométriques et ne prend pas en compte la dépendance de C_{if} à la tension de grille. Il mène à d'importantes erreurs dans l'estimation de cette capacité, qu'il surestime très nettement. Au cours du § 3.4, nous présenterons un nouveau modèle de capacité de bord interne, qui tient compte de cette dépendance à la tension de grille. En particulier, nous décrirons en détail le comportement de cette capacité au travers des différents régimes de fonctionnement du transistor.

Il est utile à ce point de noter que le comportement capacitif extrinsèque global est principalement dépendant de la capacité d'overlap C_{ov}. En effet, la capacité de bord interne C_{if} modifie le comportement de C_{ext} mais n'en altère pas la caractéristique générale. Quant à la capacité de bord externe C_{of}, elle est juste la cause d'un décalage constant de la somme

des deux autres capacités parasites. Cette dernière étant indépendante des tensions appliquées au dispositif, elle dépend donc uniquement des paramètres technologiques du TMOS. Elle ne contribue ainsi en aucune manière à la décroissance de la capacité extrinsèque en régime extrinsèque.

3.3.2.3 Résultats de simulations 2D

La simulation 2D est très utile à la compréhension des différents phénomènes physiques mis en jeu dans le comportement capacitif des transistors MOS. Nous avons réalisé différentes structures types lors de nos simulations de procédés (Athena), puis nous avons étudié les réponses électriques obtenues lors des simulations de dispositifs (Atlas). Les paramètres technologiques utilisés lors des simulations de procédés sont ceux spécifiés au Tableau 3.1.

Nous avons tout d'abord généré le TMOS typique de cette technologie, à savoir celui conçu avec une dose de dopants LDD égale à 2×10^{14} cm^{-2}. Ensuite nous avons généré d'autres structures ayant une dose LDD plus forte, allant jusqu'à 1×10^{15} cm^{-2}, ce qui in fine correspond quasiment à un TMOS HDD (classique), la dose des zones LDD étant alors similaire à celle des source/drain. Cependant rigoureusement parlant, cela n'est pas tout à fait exact car l'énergie d'implantation correspondant aux régions LDD est plus faible d'un facteur 2 par rapport à celle des régions source/drain. En conséquence, la profondeur des jonctions des régions LDD ne correspond donc pas à celle des source/drain, même dans le cas de l'implantation d'une dose de dopants équivalente. En ce qui concerne les énergies d'implantation, les extensions des source/drain ont été formées par implantation d'arsenic à 15 keV et les regions de source/drain par implantation d'arsenic à 30 keV.

La Fig. 3.4 présente une simulation de la capacité extrinsèque dans le cas d'un transistor de 0.2 μm de longueur de grille. Les différentes courbes correspondent aux différentes doses LDD définies lors des simulations de procédés. Ce graphique appelle à de nombreux commentaires. En premier lieu, il permet de se rendre compte de l'importance de la capacité extrinsèque normalisée[1]. Dans le cas de la technologie standard ($N_{ldd} = 2 \times 10^{14}$ cm^{-2}), il apparaît que pour une tension de grille nulle, la valeur de la capacité extrinsèque normalisée est proche de 40 %, tandis qu'en accumulation (forte) elle est de l'ordre de 20 %. Ainsi, négliger la composante 'capacité extrinsèque' dans la simulation du comportement capacitif global

[1]Par définition le terme générique de capacité extrinsèque normalisée C^*_{ext} fait référence au rapport de la capacité C_{ext} sur la capacité d'oxyde de grille C_{ox}.

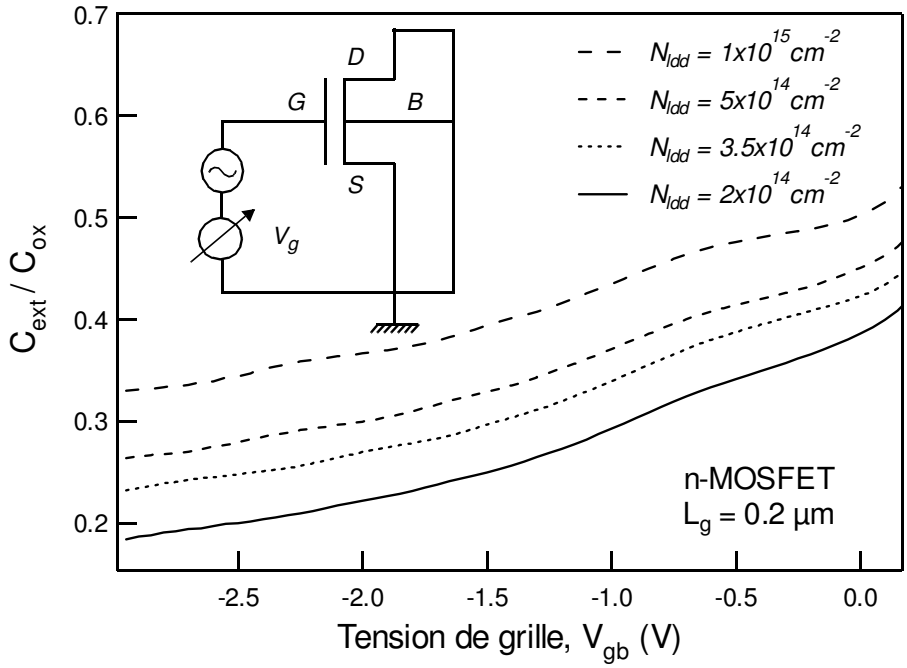


FIG. 3.4 : *Simulation de la capacité extrinsèque C_{ext} normalisée par rapport à la capacité d'oxyde de grille C_{ox}, en fonction de la tension de grille V_{gb}. La technologie standard correspond à la dose LDD la plus faible.*

entraînerait inévitablement d'importantes erreurs en simulation de circuits. Le second point mis en valeur par la Fig. 3.4 est la forte dépendance de la capacité extrinsèque normalisée au dopage des régions LDD. Plus les zones LDD sont fortement dopées, plus l'influence de C^*_{ext} est importante ; cela s'explique assez simplement par le fait que plus le dopage des LDD est élevé, plus la diffusion des dopants est importante, et donc plus la longueur de diffusion latérale L_d est importante. Nous allons expliquer cela plus en détail avec l'aide du Tableau 3.2.

Le Tableau 3.2 présente la longueur de diffusion latérale L_d comme une fonction de la dose LDD N_{ldd}. Les renseignements donnés, tels que la capacité d'overlap normalisée $A = C^*_{ov,\max}$ et la capacité extrinsèque minimale normalisée $B = C^*_{ext,\min}$ sont importants pour la compréhension du comportement capacitif extrinsèque. En particulier le calcul de la différence $(A - B)$ montre que cette différence ne dépend quasiment pas de la dose LDD. Cela signifie que la décroissance globale de la capacité extrinsèque avec la tension de grille est due principalement à la réduction de la capacité d'overlap. La capacité de bord externe étant indépendante de la tension de grille, cela nous indique déjà que la capacité de bord interne n'a pas une influence sur toute la plage de tension de grille.

TAB. 3.2 : *Influence de la dose LDD implantée sur la longueur de diffusion latérale L_d. Les valeurs maximales de la capacité d'overlap normalisée (A) ainsi que les valeurs minimales de la capacité extrinsèque normalisée (B) sont présentées. La dernière colonne indique la différence entre ces deux quantités.*

N_{ldd} (cm^{-2})	L_d (nm)	$A = C_{ov,\max}/C_{ox}$	$B = C_{ext,\min}/C_{ox}$	$A - B$
2×10^{14}	30	0.38	0.18	0.20
3.5×10^{14}	32	0.42	0.23	0.19
5×10^{14}	34.5	0.45	0.26	0.19
1×10^{15}	40	0.5	0.32	0.18

Il semble donc à ce stade de notre raisonnement que la capacité d'overlap soit la composante prépondérante du comportement capacitif extrinsèque ; en effet, d'une part elle est la contribution dominante à la capacité extrinsèque normalisée (parallèle Fig. 3.4–Tableau 3.2), et d'autre part elle est la cause majeure de la dépendance de C^*_{ext} à V_{gb}. Cependant, lors de la présentation du nouveau modèle, nous montrerons que la capacité de bord interne joue aussi un rôle non négligeable dans le comportement capacitif extrinsèque. En fait son influence est plus ciblée que celle de la capacité d'overlap et intéresse plus particulièrement le régime de déplétion du TMOS.

Avant d'aller plus loin dans l'étude séparée des différentes composantes de la capacité extrinsèque, il est instructif d'observer l'influence de la longueur de grille du TMOS sur la capacité extrinsèque normalisée. La Fig. 3.5 montre la valeur de C^*_{ext} pour quatre dispositifs de différentes longueurs, allant de 0.2 à 10 μm. Il ressort très clairement de cette figure que la dépendance à la tension de grille de la capacité extrinsèque est liée à la dimension du transistor. En fait, dans le cas d'un dispositif canal long (10 μm), cette dépendance devient négligeable en raison de la grande valeur des capacités intrinsèques de la structure. Pour un TMOS très court (0.2 μm), la décroissance de C^*_{ext} est au contraire très importante : pour V_{gb} allant de 0 à -2 V, la valeur de C^*_{ext} baisse de moitié. Cette différence de comportement entre les TMOS court et long s'explique par la valeur fixée de la longueur de diffusion latérale L_d. Quelle que soit la longueur de grille du dispositif, la grandeur L_d reste la même, puisqu'elle ne dépend que des paramètres de process. Ainsi quand la longueur de grille augmente, le rapport L_d/L_g diminue et donc C^*_{ext} aussi, en majeure partie via la baisse de sa composante C^*_{ov} ; la capacité d'overlap étant directement dépendante de la longueur de diffusion latérale (cf. Eqs. (3.6)–(3.7)). Il y a

aussi une diminution de C^*_{ext} due à la diminution de la capacité de bord externe normalisée C^*_{of}, cependant elle se ressent principalement en régime d'accumulation forte, région où la capacité d'overlap devient alors très faible.

En fait, en accumulation forte, la capacité extrinsèque tend vers sa composante 'capacité de bord externe' [1,9,14] :

$$C_{ext} \rightarrow 2 \cdot C_{of} \quad \text{pour} \quad V_{gb} \ll V_{fb} \tag{3.8}$$

En résumé, la Fig. 3.5 illustre l'importance croissante de la région de recouvrement de la grille (région d'overlap, de longueur L_d) par rapport la région intrinsèque (région du canal, de longueur L_{eff}). De plus, ce phénomène va s'accentuer avec l'arrivée des nouvelles technologies de TMOS, en raison des difficultés technologiques croissantes rencontrées lors de la fabrication des dispositifs. En particulier le contrôle précis de grandeurs clés telles que les profondeurs de jonctions — et donc la diffusion latérale des dopants — devient de plus en plus difficile, compte tenu des ordres de grandeurs visés [17]. Une modélisation correcte du comportement capacitif extrinsèque est donc un réel besoin, et va le devenir encore plus en raison de la réduction d'échelle des composants.

Jusqu'à présent, nous nous sommes uniquement intéressés à la dépendance de la capacité extrinsèque avec la polarisation de la grille. Comme indiqué à l'équation (3.1), nous avons

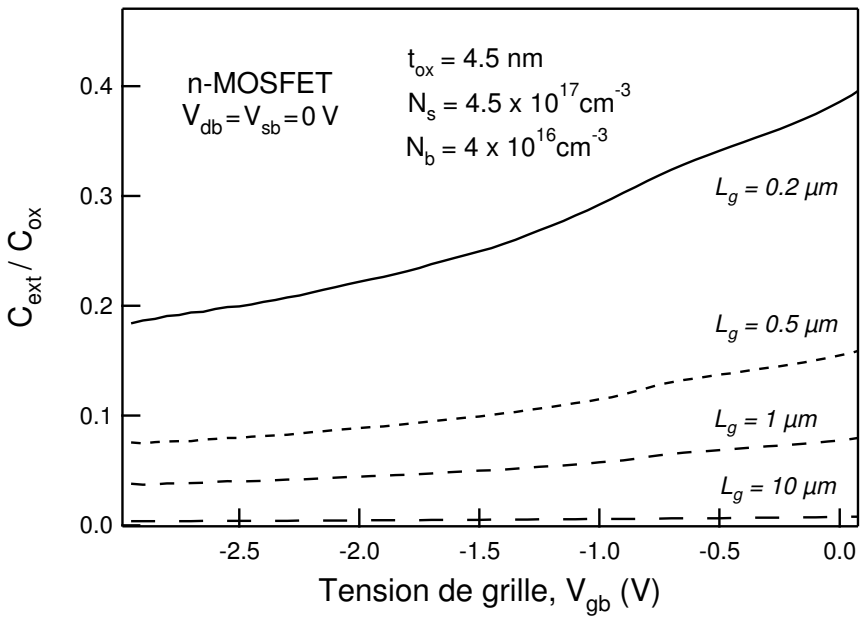


FIG. 3.5 : *Simulation de la capacité extrinsèque C^*_{ext} en fonction de la tension de grille V_{gb}, et de la longueur de grille L_g. Les TMOS LDD simulés ici correspondent tous aux critères technologiques standards, en particulier la dose LDD implantée est égale à 2×10^{14} cm^{-2}.*

supposé que C_{ext} était au premier ordre indépendant des tensions appliquées au drain ou à la source. Pour quantifier l'influence de la polarisation du drain V_d, nous avons réalisé différentes simulations 2D mettant en évidence son rôle dans le comportement capacitif extrinsèque. La Fig. 3.6 montre l'influence de V_d sur des dispositifs de 0.2 μm, conçus avec différents dopages LDD. Il apparaît que dans le cas de la dose de dopants la plus élevée ($N_{ldd} = 1 \times 10^{15}$ cm^{-2}) la tension de drain n'a aucun effet sur la capacité extrinsèque. Pour la dose LDD la plus faible ($N_{ldd} = 2 \times 10^{14}$ cm^{-2}), il y a une petite différence entre les courbes correspondant aux deux polarisations de drain ($V_d = 0$ et 1 V) lorsque le TMOS entre en régime de déplétion ($V_g > V_{fb} \simeq -1$ V). L'explication physique de ce phénomène est immédiate : à polarisation de drain non nulle, il se produit une augmentation de la taille de la région de déplétion au niveau de la jonction drain–substrat, ce qui en fait réduit la longueur de diffusion latérale effective L_{ov} et donc par voie de conséquence la capacité d'overlap C_{ov}.

Ainsi, la capacité extrinsèque est donc légèrement dépendante de la polarisation de drain (source) via sa composante 'capacité d'overlap'. Cependant, dans le pire des cas, l'écart entre les deux courbes présentées à la Fig. 3.6 est inférieur à 2 % du rapport C^*_{ext}. À ce stade du raisonnement, il est alors utile de rappeler que la réduction d'échelle continuelle des TMOS

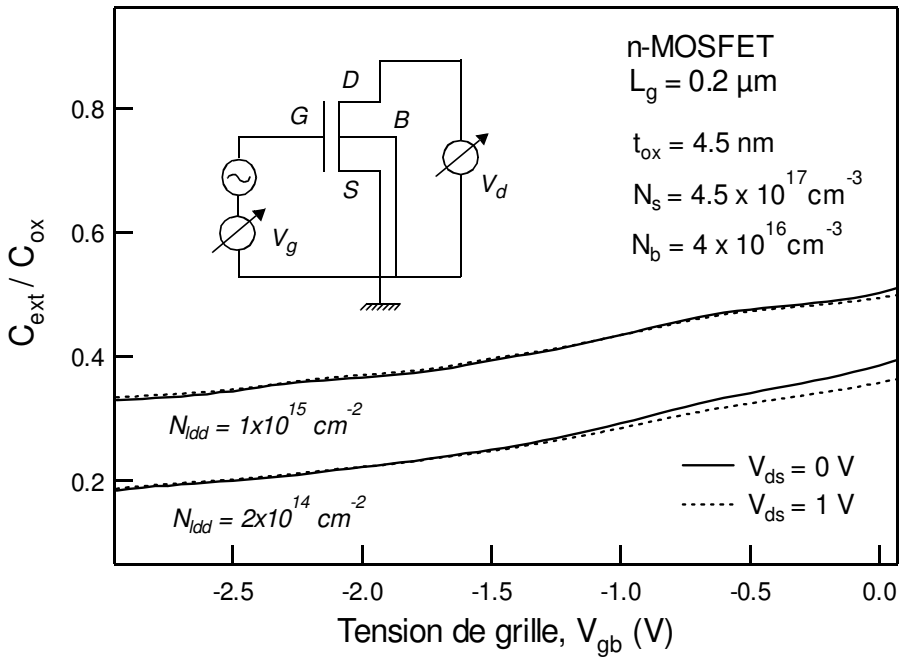


FIG. 3.6 : *Simulation de la capacité extrinsèque C^*_{ext} en fonction des tensions de grille V_{gb}, et de drain V_{ds} ($= V_{db} - V_{sb}$). Les résultats sont aussi montrés en tant que fonction de la dose de dopants implantée dans les régions LDD.*

nécessite une augmentation des niveaux de dopage de la région de canal et des extensions source/drain, dans le but d'améliorer les performances des dispositifs [17]. En conséquence, d'après nos simulations (Fig. 3.6), il apparaît que négliger l'influence de la polarisation du drain sur la capacité extrinsèque est une hypothèse valide. De toute façon, nous montrerons que si dans le cas d'une technologie CMOS particulière, il s'avère nécessaire de tenir compte des polarisations source/drain dans le calcul de C_{ext}, (i.e. en réalité dans le calcul de C_{ov}), le modèle que nous allons présenter peut très facilement être étendu pour prendre en compte non seulement l'influence de V_g, mais aussi celle de V_d ou V_s.

En dernier lieu, avant d'expliciter notre nouveau modèle, nous allons observer l'influence du phénomène de déplétion du polysilicium de grille (polydéplétion) sur la capacité extrinsèque. A priori, cet effet ne joue un rôle que pour des tensions de grille correspondant au régime intrinsèque [18,19]. Pour se convaincre définitivement que négliger la polydéplétion dans le calcul de la capacité extrinsèque est justifié, nous avons simulé deux dispositifs caractéristiques de notre technologie standard ($N_{ldd} = 2 \times 10^{14}$ cm^{-2}), l'un ayant une grille assimilée à un métal et l'autre une grille en polysilicium, ayant un dopage équivalent à $N_p = 8 \times 10^{19}$ cm^{-3}. Précisons que le dopage du polysilicium est réalisé avec de l'arsenic, i.e. le dopant des régions de source et de drain. La Fig. 3.7 présente les résultats de cette simulation. Il est clair que la polydéplétion joue un rôle négligeable en ce qui concerne le comportement capacitif extrinsèque. De plus, les niveaux de dopage polysilicium des technologies à venir n'étant pas amené à diminuer (les technologues espérant au contraire pouvoir l'augmenter), ce constat restera valable pour les générations futures.

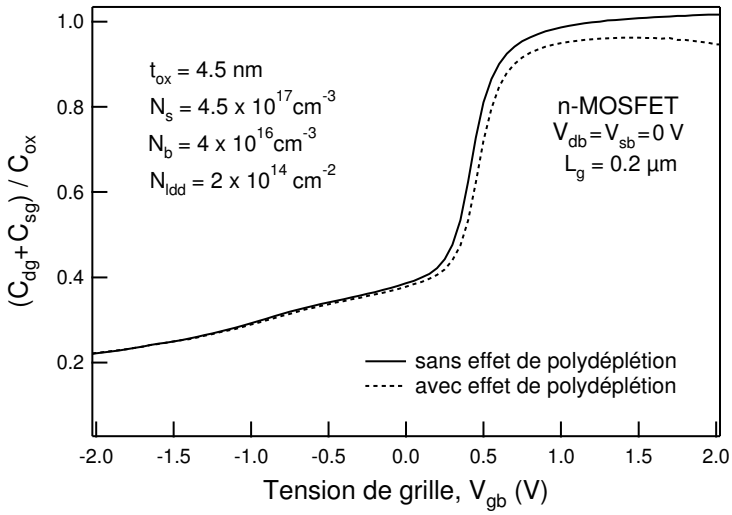


FIG. 3.7 : *Influence du phénomène de polydéplétion sur le comportement capacitif global. Il apparaît que la polydéplétion ne perturbe pas le comportement capacitif extrinsèque.*

3.4 Modélisation de la capacité extrinsèque

3.4.1 Nécessité d'un nouveau modèle

Les modèles existants dont nous avons parlé précédemment (§ 3.2.3) ne parviennent pas à réaliser un compromis important en modélisation compacte, à savoir le compromis *simplicité/efficacité*. Par exemple, un modèle performant tel que celui présenté dans [1,2] a une forte base physique, mais est extrêmement complexe à comprendre, et surtout à implémenter [14]. En outre, malgré son caractère physique, d'après les résultats publiés il apparaît clairement que la capacité de bord interne n'est pas prise en compte dans ce modèle. De nombreux autres modèles développés dans les années 80 ne sont maintenant plus valables en raison des importantes mutations technologiques ayant eu lieu depuis leur création. Cependant, les bases physiques de certaines approches précédemment formulées sont toujours d'actualité [8].

Notre nouveau modèle se veut avant tout pragmatique. Bien que reposant sur la physique du transistor MOS, les approximations développées le rendent très simple et compréhensible, sans pour autant sacrifier sa précision. Ce n'est donc pas un modèle purement physique, mais plutôt un modèle semi-empirique fondé sur la physique. Le degré d'empirisme du modèle est réduit au minimum, un unique paramètre d'ajustement étant nécessaire. Enfin un avantage non négligeable de notre modélisation est d'être facilement implémentable au sein de n'importe quels types de modèles compacts.

3.4.2 Formulation du nouveau modèle

Ce nouveau modèle est basé sur une description géométrique et physique des différentes composantes de la capacité extrinsèque. L'approche originale dont s'inspire notre étude repose sur les travaux présentés par SCHRIVASTAVA et FITZPATRICK [8]. Ce papier est l'un des pionniers en ce qui concerne les capacités parasites. En effet, les auteurs ont identifié les trois capacités parasites constituant la capacité extrinsèque, puis proposé un modèle indépendant pour chacune des trois capacités. Cependant, toutes sont considérées comme constantes, ce qui engendre d'importantes erreurs de simulation.

Reprenant certaines idées de ce modèle, nous avons développé de nouvelles expressions prenant en compte la dépendance de la capacité d'overlap et de la capacité de bord interne à la polarisation de la grille. Comme nous l'avons dit ci-dessus, en dépit des modifications réalisées sur le modèle de base, le nouveau modèle reste relativement abordable. En particulier,

la description de la capacité d'overlap est formulée de manière telle qu'un simple calcul manuel de cette capacité est tout à fait envisageable.

3.4.2.1 Capacité d'overlap

Le comportement de la capacité d'overlap C_{ov} a déjà été étudié au § 3.3.2.1. Nous allons en rappeler les points essentiels, en régimes extrinsèque et intrinsèque. Tout d'abord, intéressons nous au comportement de C_{ov} en régime intrinsèque, i.e. lorsque le TMOS est en mode d'inversion. Dans ce mode de fonctionnement, une modélisation classique de C_{ov} est justifiée. En effet, la capacité d'overlap est constante et égale à sa valeur maximum, et ne dépend donc plus de la polarisation de la grille. Cette modélisation classique est définie par :

$$C_{ov,\max} = C_{ox} \cdot L_d \tag{3.9}$$

Cette équation est donnée par unité de largeur du dispositif, la capacité d'oxyde de grille étant exprimée par unité de surface.

Au cours du § 3.3.2.1, nous avons interprété la dépendance de la capacité d'overlap à la polarisation de grille en termes de modulation de la longueur de diffusion latérale L_d. Ceci nous a permis de définir une longueur de diffusion latérale effective L_{ov}, dépendante de la polarisation de la grille, comme illustrée à la Fig. 3.3. Par commodité nous rappelons ici le principe de cette modélisation :

$$L_{ov} = L_{ov}(V_g) \quad \text{et,} \tag{3.10}$$

$$L_{ov}(V_g)\big|_{V_g \geqslant V_{fbldd} \simeq 0} = L_d \tag{3.11}$$

Ainsi la capacité d'overlap est reliée à la polarisation de la grille par l'expression :

$$C_{ov}(V_g) = C_{ox} \cdot L_{ov}(V_g) \tag{3.12}$$

Dans le but de décrire la variation de la longueur de diffusion latérale effective avec la tension de grille, nous avons défini une fonction nommée $A(V_g)$, de la façon suivante :

$$L_{ov}(V_g) = A(V_g) \cdot L_d \tag{3.13}$$

$A(V_g)$ nous permet de décrire empiriquement la décroissance de L_{ov} pour des valeurs négatives de la tension de grille V_g :

$$A(V_g) = \begin{cases} 1 & \text{pour } V_g \geqslant 0\,\text{V}, \\ \dfrac{1}{1-\lambda \cdot V_g} & \text{pour } V_g \leqslant 0\,\text{V}. \end{cases} \tag{3.14}$$

où λ est un paramètre d'ajustement dépendant du dopage du canal et des zones LDD, via la grandeur technologique L_d (longueur de diffusion). λ est l'unique paramètre d'ajustement de tout le modèle extrinsèque, et il est fixé pour un procédé technologique donné. L'extraction de ce paramètre sera discutée ultérieurement (cf. § 3.4.3, page 71).

Pour pouvoir être correctement utilisable en simulation de circuits, le modèle que nous souhaitons réaliser doit être continu et avoir des transitions douces entre toutes les régions de fonctionnement du TMOS. Nous avons donc utilisé une fonction de lissage hyperbolique pour obtenir une transition douce autour de $V_g = 0$ V dans la fonction $A(V_g)$. La robustesse de ce type de fonction de lissage est largement éprouvée, c'est en effet une des fonctions hyperboliques de base utilisée dans les modèles MM9 et MM11 de Philips [20,21]. Elle nous permet de faire converger la capacité d'overlap vers sa valeur maximum pour des tensions de grille positives :

$$V_g{}' = V_g - \frac{1}{2} \cdot \left[V_g + \sqrt{V_g^2 + 0.05} \right] \tag{3.15}$$

Finalement la capacité d'overlap est définie par :

$$C_{ov}(V_g) = W_g \cdot C_{ox} \cdot \frac{L_d}{1 - \lambda \cdot V_g{}'} \tag{3.16}$$

où W_g est la largeur du canal du dispositif. La détermination de la valeur du paramètre λ nécessitant la connaissance de la capacité de bord externe, nous reviendrons sur sa détermination après avoir décrit la modélisation de toutes les composantes de C_{ext}.

3.4.2.2 Capacité de bord externe

Le cas de cette capacité est le plus simple car elle est indépendante des polarisations externes appliquées au dispositif [1,2,9,10,13]. Elle peut être complètement définie à partir de considérations géométriques, comme cela a été expliqué en détail par SCHRIVASTAVA et FITZPATRICK [8]. Dans le cadre de notre modèle extrinsèque, la capacité de bord externe C_{of}

est modélisée par l'équation unanimement reconnue, originellement proposée dans [8] :

$$C_{of} = \frac{\varepsilon_0 \varepsilon_{ox}}{\pi/2} \cdot \ln\left(1 + \frac{t_{poly}}{t_{ox}}\right) \tag{3.17}$$

Cette équation est donnée par unité de largeur du dispositif. Il est important de se rendre compte de l'importance non négligeable de cette capacité pour des transistors canaux courts. Par exemple, dans le cas d'un TMOS de 0.14 μm de longueur de grille effective, le rapport $2(C_{of}/C_{ox})$ est supérieur à 10 %. Cela signifie qu'ignorer la capacité de bord externe conduirait au minimum à une erreur de 10 % dans l'estimation de la capacité $(C_{dg} + C_{sg})$, dans tous les régimes de fonctionnement, y compris celui d'inversion.

3.4.2.3 Capacité de bord interne

Le traitement de la capacité de bord interne C_{if} est nettement plus complexe que celui de la capacité de bord externe C_{of}, en raison de sa forte dépendance à la polarisation de la grille. Tous les modèles compacts courants ignorent cette capacité, étant donné l'absence de modèle valable disponible. Pourtant cette capacité existe bien, et joue un rôle dans le comportement capacitif extrinsèque. Il est d'ailleurs intéressant de constater qu'un modèle tel que BSIM4v2 inclut dans sa documentation des explications au sujet de C_{if}, insiste sur le fait que cette capacité est très liée à V_g, et in fine pose $C_{if} = 0$ pour tout le modèle ! Certes la capacité de bord interne n'a pas un rôle aussi dominant que la capacité d'overlap, cependant elle perturbe le comportement capacitif extrinsèque, tout particulièrement en régime de déplétion.

Une modélisation analytique classique de C_{if} est donnée dans [8] par :

$$C_{if} = \frac{\varepsilon_0 \varepsilon_{ox}}{\beta} \cdot \ln\left[1 + \frac{X_j}{t_{ox}} \cdot \sin\left(\frac{\pi}{2}\frac{\varepsilon_{ox}}{\varepsilon_{si}}\right)\right] \tag{3.18}$$

où β est défini par :

$$\beta = \frac{\pi}{2}\frac{\varepsilon_{ox}}{\varepsilon_{si}} \tag{3.19}$$

Cependant, l'équation (3.18) n'est pas valable dans le cas des dispositifs submicroniques, pour différentes raisons. Comme nous l'avons déjà mentionné, la capacité de bord interne n'est pas du tout constante. En fait lorsque la tension de grille induit une couche d'accumulation (forte) ou une couche d'inversion dans la région du canal, cette capacité est nulle :

$$C_{if} = 0 \qquad \text{(accumulation forte ou inversion)} \tag{3.20}$$

Quand le TMOS est en accumulation ($V_g < V_{fb}$), une couche de trous libres (n-MOSFET) électriquement déconnectée de la région de drain (source) écrante la composante du champ électrique liée à la capacité de bord interne. Ainsi cette capacité tend vers zéro quand la tension de grille devient très négative. Il en est de même lorsque le TMOS est en inversion en raison de la formation du canal d'inversion.

Lorsque la tension de grille augmente au-delà de la tension de bandes plates, le dispositif entre en régime de déplétion. C'est dans ce régime que la capacité de bord interne atteint sa valeur maximum. D'après nos multiples comparaisons avec les simulations 2D, la valeur prédite par (3.18) surestime largement la valeur maximum de la capacité de bord interne. Dans le but d'obtenir un bon accord entre les résultats obtenus en simulation 2D et les résultats du modèle analytique, nous avons défini $C_{if,\max}$ comme :

$$C_{if,\max} = \frac{1}{6} \cdot \frac{\varepsilon_0 \varepsilon_{ox}}{\beta} \cdot \ln\left[1 + \frac{(2/3)X_j}{t_{ox}} \cdot \sin\left(\frac{\pi}{2}\frac{\varepsilon_{ox}}{\varepsilon_{si}}\right)\right] \tag{3.21}$$

où le facteur $1/6$ a été déterminé par un ajustement avec les résultats numériques, et le facteur $2/3$ prend en compte la forme des jonctions drain–substrat et source–substrat dans le cas des transistors de technologie LDD. Généralement, dans le cas des dispositifs LDD, la profondeur de jonction des extensions de source et drain est définie par :

$$X_{jldd} = a \cdot X_j \tag{3.22}$$

où la valeur de a dépend du procédé technologique mis en oeuvre lors de la conception du TMOS. Une plage de valeur standard est :

$$1/2 \leqslant a \leqslant 2/3 \tag{3.23}$$

Dans le cas des structures générées par notre simulateur de procédés, nous avons trouvé $a = 2/3$, d'où le facteur $2/3$ devant X_j dans (3.21). Il convient de remarquer que le choix de la valeur $a = 1/2$ n'affecte pas la précision du modèle de $C_{if,\max}$.

Lorsque la tension de grille augmente plus encore ($V_g \gg V_{fb}$), le TMOS entre en régime d'inversion. La capacité de bord interne tend alors vers zéro en raison de la formation du canal d'inversion. D'après notre étude, C_{if} est maximum à la fin du régime de déplétion, avant l'inversion faible, c'est-à-dire quand le champ électrique émergeant de la jonction

métallurgique de drain ou de source et allant jusqu'à la face inférieure de la grille, est maximum. Nous avons trouvé que cette situation correspond à une tension de grille égale à la tension de bandes plates V_{fb} décalée d'un demi-potentiel de Fermi ϕ_f.

La synthèse de ces différents éléments nous a conduit à la modélisation suivante :

$$C_{if}(V_g) = C_{if,\max} \cdot \exp\left[-\left(\frac{V_g - V_{fb} - \phi_f/2}{3\phi_f/2}\right)^2\right] \tag{3.24}$$

où le terme $3\phi_f/2$ au dénominateur joue le rôle d'une variance, et décrit la gamme de tension de grille pour laquelle la capacité de bord interne est non nulle.

Un paramètre important à la compréhension approfondie du comportement de C_{if} n'a pas encore été abordé. Il s'agit comme dans le cas de la capacité d'overlap d'étudier sa dépendance à la longueur de diffusion latérale L_d. En effet, puisque L_d est modulée par la tension de grille, la forme de la région de déplétion des zones LDD varie forcément, et in fine la capacité de bord interne également. Lorsque le TMOS est en régime d'accumulation ou d'inversion, cette variation de forme de la région déplétée des LDD n'a aucune influence puisque C_{if} est alors nulle. Quand la capacité de bord interne n'est pas négligeable, notre approche prend en compte cette interaction implicitement, comme nous allons l'expliquer.

D'un côté, plus la tension de grille augmente, plus C_{if} devrait être élevée puisque la région LDD retrouve sa forme complète (cf. Fig. 3.3), et par conséquent une partie plus importante de la région LDD peut collecter le champ de bord interne. D'un autre côté, quand la tension de grille augmente, la densité d'électrons dans le canal devient non négligeable et donc induit une diminution rapide de la valeur de C_{if}. Nous considérons que ce dernier effet domine le premier dès que la tension de grille est supérieure à $V_{fb} + \phi_f/2$. Cela est bien en accord avec notre modélisation illustrée par (3.24).

Le Tableau 3.3 résume les paramètres utiles au nouveau modèle de capacité de bord interne. Ce tableau illustre un point important de notre modèle : il ne nécessite aucun paramètre qui ne soit déjà inclus dans un modèle compact standard.

TAB. 3.3 : *Principaux paramètres du modèle de la capacité de bord interne.*

Paramètres	Description
t_{ox}	Épaisseur d'oxyde de grille
X_j	Profondeur des jonctions source/drain
V_{fb}	Tension de bandes plates (substrat)
ϕ_f	Potentiel de Fermi (substrat)

Observons maintenant les résultats obtenus avec ce nouveau modèle. Une simulation de la capacité de bord interne pour un dispositif de 0.14 μm de longueur de grille effective est présentée à la Fig. 3.8. Notre modèle est alors comparé à deux approximations classiques, habituellement utilisées dans les modèles compacts. Il convient de remarquer que la capacité de bord interne totale est représentée à la Fig. 3.8, i.e. nous avons considéré les deux capacités de bord interne, côté drain et côté source. Il apparaît que notre modèle diffère significativement des modèles habituels. L'amélioration apportée par ce nouveau modèle sur le comportement capacitif extrinsèque global sera clairement démontrée par comparaison avec des simulations numériques 2D, à la Fig. 3.9.

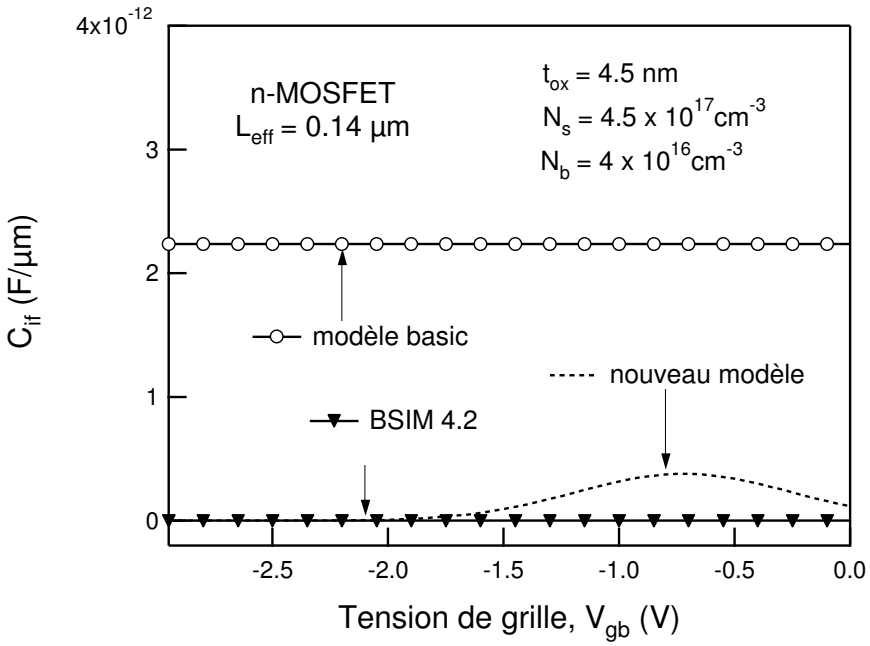


FIG. 3.8 : *Comparaison entre le nouveau modèle de capacité de bord interne et les modèles existants, pour un dispositif de 0.14 μm de longueur de grille effective. Le modèle basic est détaillé dans [8], BSIM4v2 dans [10] et notre modèle dans [14].*

3.4.2.4 Modèle complet

La capacité extrinsèque a été définie au début de ce chapitre comme étant la somme des différentes capacités parasites du transistor MOS (cf. Eq. (3.2), page 55). Compte tenu des expressions établies pour les trois composantes de la capacité extrinsèque, nous obtenons finalement une unique équation définissant C_{ext} :

$$C_{ext}(V_g) = 2 \cdot W_g \cdot \Bigg[C_{ox} \cdot \frac{L_d}{1 - \lambda \cdot V_g{}'} + C_{if,\max} \cdot \exp\left[-\left(\frac{V_g - V_{fb} - \phi_f/2}{3\phi_f/2} \right)^2 \right] + \frac{\varepsilon_0 \varepsilon_{ox}}{\pi/2} \cdot \ln\left(1 + \frac{t_{poly}}{t_{ox}} \right) \Bigg] \tag{3.25}$$

La capacité extrinsèque donnée par (3.25) est exprimée en Farad.

3.4.3 Résultats

Les résultats du modèle extrinsèque complet vont maintenant être comparés à ceux obtenus en simulations 2D. Un point important à souligner est qu'un unique jeu de paramètres est utilisé dans le modèle analytique, pour toutes les géométries d'une technologie de TMOS donnée, ainsi que pour toutes les conditions de polarisation.

La Fig. 3.9 présente un ensemble de comparaisons illustrant le bon comportement du nouveau modèle, et cela pour deux choix technologiques différents. Ces deux choix correspondent à deux doses différentes de dopants LDD, à savoir $N_{ldd} = 2 \times 10^{14}$ et 10^{15} cm^{-2}, ce qui correspond respectivement à des transistors de 0.14 et 0.12 μm de longueur de grille effective (transistor de longueur de grille de 0.2 μm). Quelle que soit la dose de dopants LDD employée, la dépendance de la capacité extrinsèque à la polarisation de la grille est correctement prédite. En outre, il apparaît que la prise en compte de la capacité de bord interne améliore significativement le comportement global du modèle capacitif extrinsèque.

Il convient de noter que ce bon accord entre le modèle analytique et les simulations numériques 2D est obtenu en n'utilisant qu'un seul paramètre d'ajustement (λ), fixé pour une technologie donnée, et donc indépendant de la longueur de grille du dispositif. Cela confirme la validité des bases physiques considérées lors du développement du modèle.

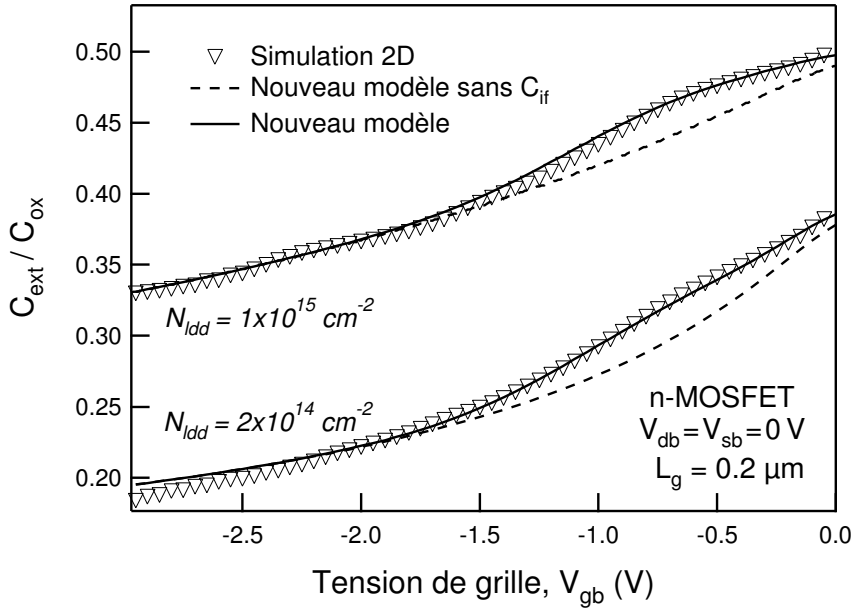


FIG. 3.9 : *Capacité extrinsèque normalisée en fonction de la tension de grille, pour deux technologies de MOSFET fortement submicroniques. La nouveau modèle (lignes) est comparé aux simulations 2D (symboles). Les pointillés représentent les résultats obtenus avec le nouveau modèle, mais sans tenir compte de la capacité de bord interne.*

L'utilisation du modèle capacitif extrinsèque requiert la connaissance de λ. Nous avons donc développé une procédure d'extraction, très simple, permettant d'obtenir de ce paramètre. Une seule valeur de la capacité extrinsèque, provenant d'un résultat de simulation numérique ou de mesure, est nécessaire à l'obtention de λ.

La procédure d'extraction est la suivante :

Pour une tension de grille $V_g \leqslant -2$ V, nous avons besoin d'une valeur de C_{ext}. Par exemple, supposons que $C_{ext}(V_g = -2) = M$.

En régime d'accumulation, la capacité de bord interne est nulle, i.e. $C_{if}(-2) = 0$. Ainsi, nous obtenons :

$$C_{ext}(-2) = M = 2 \cdot (C_{ov}(-2) + C_{of}) \tag{3.26}$$

La valeur de C_{of} est complètement connue, et donnée par (3.17). Quant à l'expression de la capacité d'overlap, elle est peut alors s'écrire comme :

$$C_{ov}(-2) = \frac{C_{ox} \cdot W_g \cdot L_d}{1 + 2 \cdot \lambda} \tag{3.27}$$

En combinant (3.26) et (3.27), nous obtenons la valeur du paramètre λ. Finalement il vient :

$$\lambda = \frac{C_{ox} \cdot W_g \cdot L_d}{M - 2 \cdot C_{of}} - \frac{1}{2} \tag{3.28}$$

Dans le cas de nos simulations, nous avons trouvé $\lambda = 0.74$ et $\lambda = 0.25$ pour des technologies correspondant à des doses LDD respectivement égales à 2×10^{14} et 10^{15} cm^{-2}.

Dans certains cas, lorsqu'aucun résultat de mesure ou de simulation 2D n'est disponible, il faut cependant connaître la valeur de λ. Pour pallier à cette éventualité, nous proposons une loi mathématique où λ est défini comme une fonction de la longueur de diffusion latérale L_d. En effet, au travers de la dépendance de λ à L_d, c'est la dépendance de λ aux dopages des source/drain et du substrat qui est prise en compte. Pour parvenir à cette loi, nous avons réalisé une série de simulations de procédés, où nous avons fait varier la dose de dopants LDD. Ainsi, à chaque dose correspond une longueur de diffusion bien déterminée, comme précédemment indiqué au Tableau 3.2. Cette dépendance de λ à L_d est illustrée à la Fig. 3.10.

Suite à un ajustement mathématique, nous avons trouvé une loi empirique décrivant cette dépendance :

$$\lambda(L_d) = 0.2 + \frac{10^{-5}}{(L_d - 0.0263)^2 - 5 \cdot 10^{-6}} \tag{3.29}$$

Les valeurs de λ calculées avec (3.29) sont représentées à la Fig. 3.10. Il apparaît que cet ajustement donne de relativement bons résultats en comparaison avec les valeurs de λ obtenues par la procédure d'extraction explicitée précédemment.

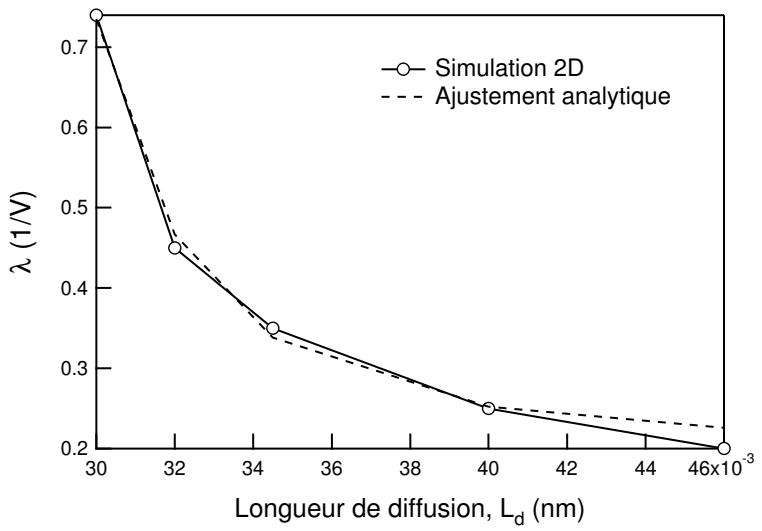


FIG. 3.10 : *Valeur du paramètre λ en fonction de la longueur de diffusion latérale L_d.*

Notre objectif initial était double au moment du développement de ce modèle de capacités parasites. D'une part, nous voulions réaliser un modèle précis, basé sur une description physique du fonctionnement du transistor MOS. D'autre part, ce modèle devait être simple, et donc bénéficier d'une grande capacité d'adaptabilité vis à vis des modèles compacts standards.

Pour tester ce second point, nous avons alors implémenté notre modèle capacitif extrinsèque dans le modèle compact EKV2.6 [22]. Plus précisément, l'équation finale (3.25) décrivant toutes les composantes de C_{ext} a été intégrée au sein du modèle capacitif intrinsèque de EKV2.6. Le modèle de capacité extrinsèque étant continu de l'accumulation à l'inversion, aucun problème de convergence ne s'est posé. En conséquence, nous disposons d'un modèle capacitif complet, continu sur toute la gamme de fonctionnement du TMOS. Ainsi, une description totale de la capacité $(C_{dg} + C_{sg})$ est disponible.

Une comparaison entre les résultats de simulations 2D et le modèle capacitif complet (extrinsèque + intrinsèque) est montrée à la Fig. 3.11. Des transistors avec différentes longueurs de grille $(0.2, 0.5$ et 1 μm) ont été simulés. Un unique jeu de paramètres a été employé pour toutes les simulations, en particulier le paramètre λ était fixé à 0.74. La transition entre les régimes de fonctionnement extrinsèque et intrinsèque est continue et lissée, respectant ainsi un impératif majeur de tout modèle compact moderne.

Les résultats présentés aux Figs. 3.9 et 3.11 ont été obtenus en ne prenant ni en compte le phénomène de polydéplétion, ni les effets de mécanique quantique. Comme nous l'avons montré au § 3.3.2.3, la polydéplétion n'influe pas sur le comportement capacitif extrinsèque. Quant aux effets quantiques, contrairement à la polydéplétion, leur influence est importante non seulement en inversion mais aussi en accumulation. Néanmoins, leur impact porte principalement sur les grandeurs intrinsèques du TMOS[1] [23,24]. En conséquence et dans un souci de simplicité, nous n'avons pas tenu compte de ces effets dans nos simulations.

[1]Le chapitre 4 sera entièrement consacré à l'étude et à la modélisation des effets quantiques au sein des TMOS ; en particulier, leur impact sur le comportement statique et dynamique sera quantifié et modélisé.

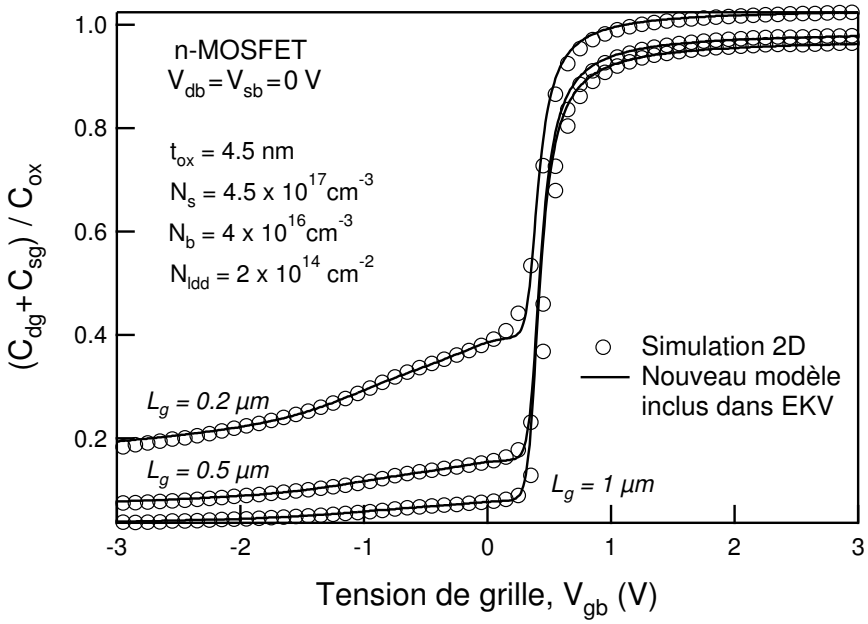


FIG. 3.11 : *Modélisation et simulation 2D de la capacité* $(C_{dg} + C_{sg})$ *pour des dispositifs de différentes longueurs de grille. Les résultats sont présentés pour toute la plage de fonctionnement du transistor MOS : accumulation, déplétion et inversion.*

3.4.4 Extension du modèle

Lors de l'étude de l'influence des composantes de la capacité extrinsèque (§ 3.3.2.1, page 56), nous avons supposé dans le but de simplifier les équations du modèle, que la tension de bandes plates des extensions LDD (V_{fbldd}) était environ égale à zéro. Cette approximation peut avoir une influence sur le modèle de capacité d'overlap, en fonction des choix technologiques décidés lors de la conception du TMOS. Par exemple, si le dopage du polysilicium est très différent de celui des régions LDD, V_{fbldd} ne tend plus vers zéro. La prise en compte de cette grandeur est cependant immédiate dans notre modèle, il suffit de réécrire (3.15) de la façon suivante :

$$V_g'' = (V_g - V_{fbldd}) - \frac{1}{2} \cdot \left[(V_g - V_{fbldd}) + \sqrt{(V_g - V_{fbldd})^2 + 0.05} \right] \tag{3.30}$$

L'étude des simulations 2D de la capacité extrinsèque nous a aussi conduit à négliger l'influence de la tension de drain (source) sur la valeur de la capacité d'overlap. Si la prise en compte de V_d ou V_s s'avère nécessaire dans le cas d'une technologie de TMOS bien particulière (par exemple, dans le cas d'un dopage LDD très faible), une simple modification de (3.30) est

nécessaire :

$$\tilde{V_g}(V_x) = \Delta(V_x) - \frac{1}{2} \cdot \left[\Delta(V_x) + \sqrt{\Delta(V_x)^2 + 0.05}\right] \tag{3.31}$$

avec

$$\Delta(V_x) = V_g - V_{fbldd} - V_x \quad \text{avec} \quad V_x = V_d \text{ ou } V_s \tag{3.32}$$

Pour bien fixer les idées, prenons un exemple. Considérons un TMOS avec une tension de bandes plates LDD non nulle, une polarisation de source nulle et une polarisation de drain non nulle. Le modèle complet de la capacité d'overlap est alors défini par :

$$C_{ov}(V_g) = W_g \cdot C_{ox} \cdot \left[\underbrace{\frac{L_d}{1 - \lambda \cdot \tilde{V_g}(0)}}_{\text{côté source}} + \underbrace{\frac{L_d}{1 - \lambda \cdot \tilde{V_g}(V_d)}}_{\text{côté drain}}\right] \tag{3.33}$$

Les résultats obtenus en incluant la dépendance à la tension de drain dans le modèle sont présentés aux Figs. 3.12 et 3.13. La Fig. 3.12 illustre le phénomène de variation de la longueur de diffusion latérale effective avec la tension de drain. Plus la tension de drain est élevée, plus L_{ov} est réduite en raison de l'élargissement de la zone de déplétion située à la jonction drain–substrat. Par conséquent, la valeur de la capacité d'overlap est réduite, provoquant ainsi une diminution globale de la capacité extrinsèque. La Fig. 3.13 illustre l'influence de la polarisation du drain en fonction du régime de fonctionnement du transistor MOS, i.e. de l'accumulation pour $V_g = -2$ V à l'inversion pour $V_g = 1$ V.

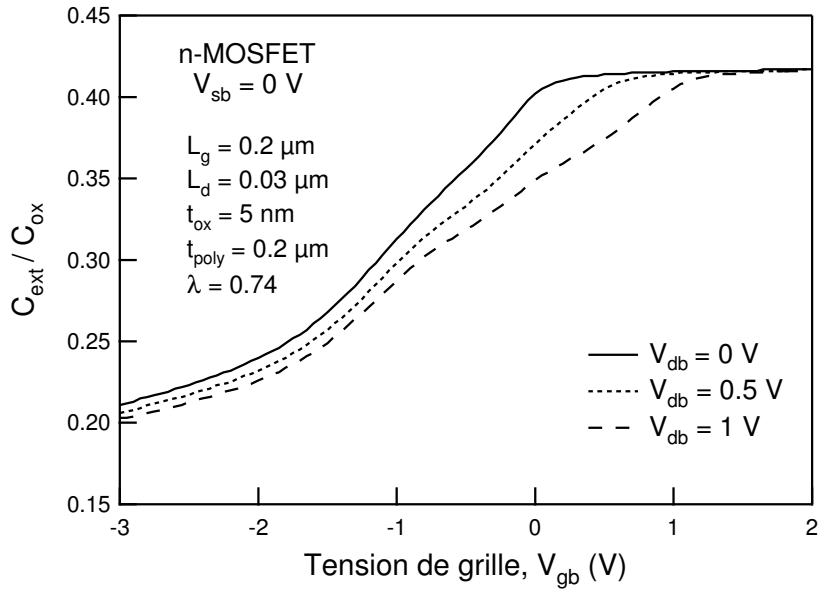


FIG. 3.12 : *Simulation de la capacité extrinsèque normalisée en fonction de la tension de grille, pour différentes tensions de drain.*

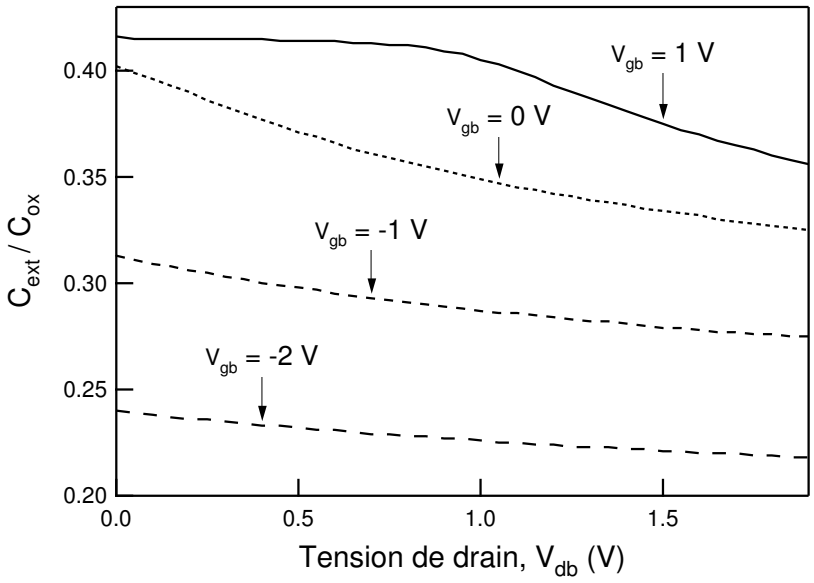


FIG. 3.13 : *Simulation de la capacité extrinsèque normalisée en fonction de la tension de drain, pour différentes tensions de grille. Les paramètres de simulation sont les mêmes qu'à la figure précédente ; en particulier* $V_{sb} = 0$ *V.*

Nous venons donc de démontrer que le modèle de capacité extrinsèque présenté au cours de ce chapitre est très facilement adaptable à différentes situations de simulation. Le modèle étant relativement simple, n'importe quel utilisateur potentiel est en mesure de redéfinir certaines équations du modèle, en fonction de ses propres besoins.

3.5 Conclusion

L'objectif principal de notre étude du comportement extrinsèque capacitif était de réaliser un nouveau modèle décrivant les capacités parasites du transistor MOS. La compréhension physique des résultats obtenus en simulations numériques nous a permis de développer un modèle basé sur la physique du TMOS, et ayant un faible degré d'empirisme. En effet, un unique paramètre additionnel est nécessaire par rapport à un modèle intrinsèque. La formulation du modèle complet est relativement simple, et mène à de bons résultats en comparaison aux simulations 2D.

Les caractéristiques essentielles de notre modèle extrinsèque reposent sur une nouvelle formulation de la capacité d'overlap et de la capacité de bord interne. La dépendance à la tension de grille de la capacité d'overlap est modélisée de façon semi-empirique, en termes de longueur de diffusion latérale. Cette dépendance à la longueur de diffusion latérale rend notre

formulation de la capacité d'overlap implicitement dépendante des paramètres technologiques du TMOS. En ce qui concerne la modélisation de la capacité de bord interne, pour la première fois en modélisation compacte, un modèle analytique prenant en compte l'influence de la polarisation de la grille est explicité. Le modèle final, incluant les trois capacités parasites du TMOS, améliore significativement les modèles existants. Chaque composante de la capacité extrinsèque est précisément décrite, et est exprimée comme une fonction des paramètres technologiques du TMOS ; cela permet au modèle d'avoir un caractère prédictif.

Enfin, un intérêt particulier de ce nouveau modèle réside dans sa facilité d'implémentation au sein des modèles capacitifs intrinsèques des modèles compacts actuels. Le modèle étant continu de l'accumulation à l'inversion, il ne souffre d'aucun problème de convergence. En outre, de par sa formulation analytique simple, son incorporation dans un modèle intrinsèque n'augmente pas le temps de calcul du modèle global ainsi formé, ce qui est un point important en modélisation compacte. Nous avons aussi montré comment faire évoluer le modèle en cas de besoins spécifiques à certaines technologies de TMOS. Par exemple, prendre en compte l'influence des polarisations de source et de drain — initialement négligées — s'obtient d'une manière tout à fait logique, et conduit à des résultats physiquement cohérents.

Une perspective intéressante à ce travail pourrait être de concevoir un modèle extrinsèque en charge. En particulier, la possibilité d'utiliser notre approche consistant à définir une longueur de diffusion latérale effective, serait une solution envisageable pour modéliser la charge d'overlap. D'autres approches basées sur une modélisation du potentiel de surface, permettent d'obtenir une description physique et précise de cette charge d'overlap. Les équations utilisées dans de tels modèles sont cependant beaucoup plus complexes. Il serait donc instructif de comparer les résultats obtenus entre ce dernier type de modèle et une évolution en charge du modèle extrinsèque ici présenté. Il conviendra alors de bien évaluer le critère simplicité/efficacité de chaque approche, critère important puisqu'il permet de déterminer vers quel type d'utilisation se destine un modèle ou un autre.

Bibliographie

[1] P. Klein, “A compact-charge LDD-MOSFET model,” *IEEE Trans. Electron Devices*, vol. 44, no. 9, pp. 1483–1490, 1997.

[2] F. Schuler, K. Hoffman, and P. Klein, “Source-drain-C(V)-behaviour of short channel LDD–MOSFETs,” in *Proc. ESSDERC*, 1998, pp. 108–111.

[3] K. Suzuki, “Parasitic capacitance of submicrometer MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 46, no. 9, pp. 1895–1900, 1999.

[4] N. Arora, *MOSFET Models for VLSI Circuit Simulation. Theory and Practice*. New York : Springer-Verlag, 1993, ISBN 3-211-82395-6.

[5] A. Cetner, “A simple formula for two-dimensional capacitance,” *Solid-State Electron.*, vol. 31, pp. 973–974, 1988.

[6] C. E. Bieber, M. L. Schmatz, *et al.*, “A non-linear microwave MOSFET model for spice simulators,” *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 5, pp. 604–610, 1998.

[7] C. C. Enz and Y. Cheng, “Issues in RF MOS transistors modeling,” in *Proc. 8th Workshop on Advances in Analog Circuit Design (AACD’99)*, Opio, France, Mar. 1999, pp. 108–111.

[8] R. Schrivastava and K. Fitzpatrick, “A simple model for the overlap capacitance of a VLSI MOS device,” *IEEE Trans. Electron Devices*, vol. 29, no. 12, pp. 1870–1875, 1982.

[9] V. Kol’dyaev, A. Clerix, *et al.*, “Characterisation of the overlap capacitance of submicron LDD MOSFETs,” in *Proc. ESSDERC*, 1995, pp. 757–760.

[10] *BSIM 4.2 Manual*, Univ. of California, Berkeley, Apr. 2001. [Online]. Available : http ://www-device.eecs.berkeley.edu/˜bsim

[11] J.-C. Guo, C. C.-H. Hsu, *et al.*, “An accurate decoupled C–V method for characterizing channel and overlap capacitances of miniaturized MOSFET,” in *Proc. VLSITSA*, 1993, pp. 256–260.

[12] T. Smedes and F. M. Klassen, “Effects of the lightly doped drain configuration on capacitance characteristics of submicron MOSFET’s,” in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, 1990, pp. 197–200.

[13] N. Wakita and N. Shigyo, “Verification of overlap and fringing capacitance models for MOSFET’s,” *Solid-State Electron.*, vol. 44, pp. 1105–1109, 2000.

[14] F. Prégaldiny, C. Lallement, and D. Mathiot, “A simple efficient model of parasitic capacitances of deep-submicron LDD MOSFETs,” *Solid-State Electron.*, vol. 46, no. 12, pp. 2191–2198, Dec. 2002.

[15] *Athena User’s Manual*, SILVACO International, 1997.

[16] *Atlas User’s Manual*, SILVACO International, 1997.

[17] Semiconductor Industry Association, “The international roadmap for semiconductors,” update 1999, http ://www.semichips.org.

[18] J.-M. Sallese, M. Bucher, and C. Lallement, “Improved analytical modeling of polysilicon depletion,” *Solid-State Electron.*, vol. 44, pp. 905–912, 2000.

[19] N. Arora, R. Rios, and C.-L. Huang, “Modeling the polysilicon depletion effect and its impact on submicromter CMOS circuit performance,” *IEEE Trans. Electron Devices*, vol. 42, no. 5, pp. 935–943, 1995.

[20] *MM9 Model*, Philips Semiconductors, 2001. [Online]. Available : http ://www.semiconductors.philips.com/philips_models/mos_models

[21] D. B. M. Klaassen, R. van Langevelde, A. J. Scholten, *et al.*, *The MOS model, level 1101*, Philips Research Laboratories, 2002. [Online]. Available : http ://www.semiconductors.philips.com/philips_models/mos_models/model11

[22] M. Bucher, C. Lallement, C. C. Enz, F. Théodoloz, and F. Krummenacher, “The EPFL–EKV MOSFET model, version 2.6,” EPFL, Tech. Rep., 1999. [Online]. Available : http ://legwww.epfl.ch/ekv/

[23] R. Rios, N. Arora, C.-L. Huang, *et al.*, “A physical compact MOSFET model, including quantum mechanical effects, for statistical circuit design applications,” in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, 1995, pp. 937–940.

[24] C. Lallement, J.-M. Sallese, M. Bucher, W. Grabinski, and P. Fazan, “Accounting for quantum effects and polysilicon depletion from weak to strong inversion, in a charge based design-oriented MOSFET model,” *IEEE Trans. Electron Devices*, vol. 50, no. 2, pp. 406–417, Feb. 2003.

Chapitre 4

Modélisation analytique des effets de mécanique quantique

4.1 Introduction

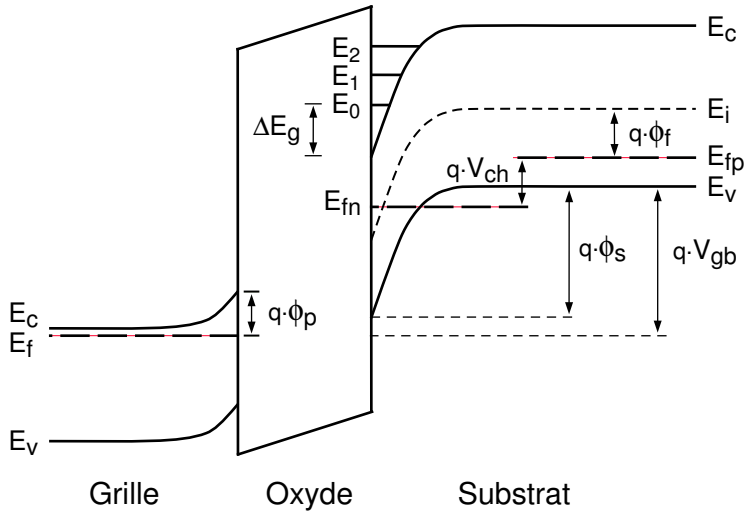
4.1.1 Signification du terme effets quantiques

Le terme *effets quantiques* a une connotation particulière en modélisation compacte. Nombre d'incompréhensions entre les concepteurs de modèles compacts et certains chercheurs plus axés sur la physique fondamentale résultent de cet abus de langage. Un exemple significatif concerne le cas de la prise en compte de l'effet tunnel, à travers les oxydes de grille ultra-fins des TMOS modernes. Au sens physique du terme, l'effet tunnel est un effet purement quantique, et pourtant en modélisation compacte, ce dernier n'est pas inclus dans ce que nous appelons les effets quantiques... Plus troublant encore est que l'effet tunnel peut être modélisé — dans un modèle compact — de façon quantique ou classique ! En réalité, cette distinction est liée au type de modélisation choisie pour décrire la densité de porteurs du canal d'inversion, dont dépend le courant tunnel. C'est cette dernière qui en fait peut être modélisée de manière quantique ou classique. Ainsi, en modélisation compacte, le terme *effets quantiques* fait référence au phénomène de quantification de l'énergie des porteurs à la surface du semi-conducteur, et aux conséquences qu'il entraîne sur les caractéristiques électriques du transistor MOS.

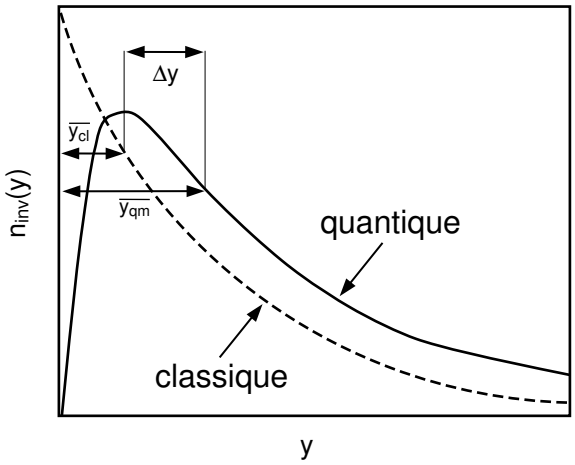
4.1.2 Origine physique du confinement quantique

Dans les technologies CMOS récentes, la combinaison de forts dopages substrat N_a et de faibles épaisseurs d'oxyde de grille t_{ox}, résulte en un champ électrique normal très élevé à l'interface Si–SiO_2. Ce fort champ provoque une importante courbure des bandes d'énergie à

la surface du semi-conducteur, qui résulte en un puits de potentiel suffisamment étroit pour quantifier le mouvement des porteurs dans la direction perpendiculaire à l'interface [1–3]. En conséquence, le spectre d'énergie consiste alors en un jeu discret de niveaux d'énergie et la distribution des porteurs à la surface du semi-conducteur est modifiée. La Fig. 4.1 illustre les deux principaux effets liés au confinement quantique, dans le cas d'un transistor n-MOS



(a) Mise en évidence de l'élargissement du gap.



(b) Modification de la distribution des porteurs.

FIG. 4.1 : *(a) Diagramme des bandes d'énergie d'un transistor n-MOS (direction transversale) en régime d'inversion. En raison des effets quantiques, le premier niveau d'énergie E_0 ne coïncide plus avec le bas de la bande de conduction E_c, résultant ainsi en une différence d'énergie ΔE_g. (b) Densité d'électrons $n_{inv}(y)$ en fonction de la position transversale y. Dans le cas quantique, la distance moyenne à l'interface $\bar{y}$ est augmentée d'une valeur Δy par rapport au cas classique.*

en régime d'inversion. En particulier la Fig. 4.1a schématise l'éclatement de la bande de conduction en sous-bandes discrètes, et met ainsi en évidence le phénomène d'élargissement du gap (bandgap widening) du semi-conducteur, défini par $\Delta E_g = E_0 - E_c$. La Fig. 4.1b illustre le phénomène de déplacement de la distribution des porteurs par rapport à l'interface Si–SiO_2. La densité d'électrons s'annule à la surface, ce qui augmente la distance moyenne $\bar{y}$ du centroïde d'inversion par rapport à la solution classique.
Nous verrons au fil de ce chapitre que l'élargissement du gap est le point clé de la modélisation analytique des effets quantiques.

4.2 Influence des effets quantiques

Les effets quantiques jouent un rôle très important dans le comportement électrique des transistors MOS. Leur présence perturbe le fonctionnement conventionnel du TMOS, ce qui entraîne des écarts significatifs entre les caractéristiques électriquement mesurées (*I–V*, *C–V*) et les valeurs attendues selon les prévisions (modèles) classiques [4–6]. Les effets quantiques affectent aussi les grandeurs non immédiatement mesurables du TMOS, telles que la densité de charge d'inversion, le potentiel de surface, la tension de seuil [7,8]. Une approche commode consistant à étudier leur influence comme une fonction du régime de fonctionnement du transistor, nous allons séparer le fonctionnement du TMOS en deux modes.

4.2.1 Régime d'inversion

Pour l'instant nous allons simplement énumérer les différentes grandeurs électriques du TMOS qui sont perturbées par les effets quantiques. Une description qualitative et quantitative détaillée sera présentée dans les sections suivantes.

Le premier impact des effets quantiques originellement discuté dans la littérature est la modification de la tension de seuil en comparaison au cas classique. En particulier les travaux de VAN DORT ont montré que la quantification de l'énergie des porteurs provoque une augmentation de la tension de seuil V_{th} [9,10]. Ainsi, l'inversion forte se produit pour une polarisation de grille plus élevée dans le cas d'un modèle quantique. D'un point de vue physique, c'est en réalité la densité de charge d'inversion Q_{inv} qui est modifiée lorsque les effets quantiques sont pris en compte ; la variation de la tension de seuil étant seulement une des conséquences directes de la modification de Q_{inv}. La seconde conséquence de la diminution quantique de Q_{inv} consiste en un changement de la relation classique $n_{inv} \simeq C_{ox} \cdot (V_g - V_{th})$, où

la pente n'est alors plus égale à C_{ox} mais à une valeur inférieure — et non constante —, définie comme étant une capacité d'oxyde effective C_{oxeff}.

La Fig. 4.2 illustre l'impact des deux effets quantiques ici discutés (ΔV_{th} et ΔC_{ox}), sur la densité de charge d'inversion Q_{inv}. Les résultats y sont présentés sur deux échelles, linéaire et logarithmique, dans le but de pouvoir observer distinctement l'impact des effets quantiques en inversion faible et en inversion forte.

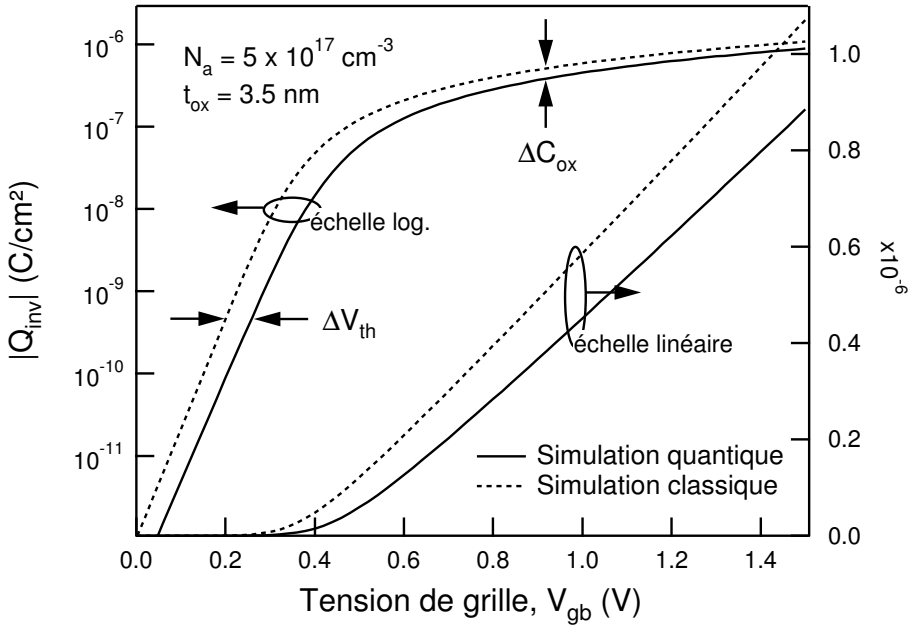


FIG. 4.2 : *Densité de charge d'inversion $|Q_{inv}|$ en fonction de la tension de grille V_{gb}, calculée en prenant en compte les effets quantiques (lignes continues) ou non (pointillés). Ces résultats ont été obtenus avec le nouveau modèle analytique développé dans ce chapitre.*

Une autre grandeur très importante modifiée par les effets quantiques est le potentiel de surface ϕ_s [4–7]. Comme le montre la Fig. 4.3, les effets quantiques provoquent une forte augmentation du potentiel de surface en régime d'inversion. Dans le cadre d'un modèle basé sur la définition du potentiel de surface, toutes les autres grandeurs dépendent de ϕ_s ; en conséquence, toute modification de ϕ_s due aux effets quantiques sera immédiatement répercutée sur le calcul des charges, des capacités, du courant de drain, etc.

En guise de première conclusion, il semble donc évident qu'en régime d'inversion (faible, modérée et forte), les effets quantiques ne peuvent raisonnablement plus être ignorés dans les modèles compacts destinés à la simulation des dispositifs actuels.

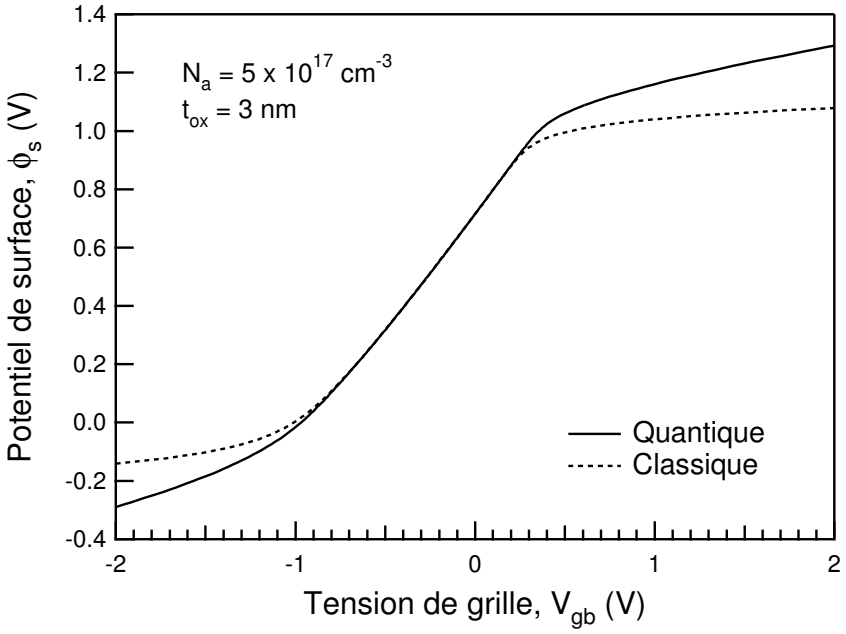


FIG. 4.3 : *Potentiel de surface ϕ_s en fonction de la tension de grille V_{gb}, calculé en prenant en compte les effets quantiques (résolution auto-cohérente Schrödinger–Poisson : ligne continue) ou non (résolution numérique de l'équation implicite décrivant ϕ_s : pointillés).*

4.2.2 Régime d'accumulation

Ce régime de fonctionnement correspond à l'état bloqué du TMOS. Principalement pour cette raison, l'influence des effets quantiques — dans le contexte des modèles compacts — y a moins été discutée dans la littérature qu'en régime d'inversion. Une autre raison est certainement aussi que la modélisation de ces effets en accumulation est nettement plus compliquée qu'en inversion [11].

En accumulation, les effets quantiques se traduisent par une diminution significative du potentiel de surface, par rapport au cas classique (voir Fig. 4.3). Il en résulte une réduction de la densité de charge de grille Q_g, ce qui entraîne une forte diminution de la transcapacité de grille $C_{gg} = dQ_g/dV_g$ (en présence d'effets quantiques C_{gg} devient très inférieure à la capacité d'oxyde de grille C_{ox}). Précisons que bien que la région d'accumulation soit d'un intérêt limité pour le concepteur de circuits, il est néanmoins important de bien modéliser le comportement électrique du TMOS dans ce régime, car c'est ici que se fait habituellement l'extraction de l'épaisseur d'oxyde de grille [12].

4.3 Modélisation physique

Une description physique exacte des effets quantiques nécessite forcément de résoudre l'équation de Schrödinger. En parallèle, il est nécessaire de résoudre l'équation de Poisson, il s'agit donc d'effectuer un calcul Schrödinger–Poisson couplé ou auto-cohérent. Une hypothèse classique faite sur les conditions aux limites du système étudié, est de supposer que l'énergie potentielle ne varie que dans une dimension, perpendiculaire à l'interface Si–SiO_2 (direction y, cf. notation Fig. 2.2, page 15). Cette hypothèse n'est cependant en toute rigueur valable que pour les capacités MOS. En fait, elle peut également s'étendre au cas des transistors MOS à canal long[1].

Nous allons maintenant présenter le principe général d'une simulation auto-cohérente Schrödinger–Poisson. En supposant le mouvement libre dans les directions x et z, on peut décomposer la fonction d'onde associée aux électrons de la façon suivante :

$$\Psi(\vec{r}) = \varphi(y) \cdot \frac{1}{\sqrt{\Omega}} \cdot \exp\left[i\left(k_x\, x + k_z\, z\right)\right] \tag{4.1}$$

où $\Omega^{-1/2}$ correspond au coefficient de normalisation adéquat [14]. L'équation de Schrödinger (indépendante du temps) s'écrit alors sous une forme simplifiée :

$$\left[-\frac{\hbar^2}{2m_y}\frac{d^2}{dy^2} + V(y)\right] \cdot \varphi(y) = \varepsilon \cdot \varphi(y) \tag{4.2}$$

où $V(y)$ désigne l'énergie potentielle extérieure, m_y la masse effective dans la direction y et ε la contribution à l'énergie totale (E) du système dans la direction y. En plus de cette approximation, nous supposerons également que Ψ est nulle à l'interface ($y = 0$), en négligeant l'effet tunnel. Finalement, en conséquence de la forme du potentiel, nous aurons pour les états les plus faibles en énergie (donc les plus peuplés) :

$$V(y \to \infty) > \varepsilon \tag{4.3}$$

ce qui signifie que les états les plus faibles en énergie, correspondent à des états liés, états dont l'énergie ε est quantifiée.

Nous désignerons par ε_i ces énergies quantifiées, et φ_i les fonctions d'onde dans la direction y correspondante. Ces grandeurs sont obtenues en résolvant (4.2), ce qui nécessite de connaître

[1] Une étude quantique rigoureuse des autres dispositifs nécessite de prendre en compte la nature 2D de l'énergie potentielle, ce qui complexifie considérablement le problème [13].

l'expression de l'énergie potentielle $V(y)$. Dans l'approximation de Hartree, et en négligeant les énergies d'échange et la polarisation de l'oxyde (force image)[1], l'énergie potentielle $V(y)$ se réduit tout simplement à une énergie potentielle électrostatique, qui s'obtient en résolvant l'équation de Poisson :

$$\frac{d^2V}{dy^2} = -\frac{q}{\epsilon_{si}} \cdot \rho(y) \tag{4.4}$$

où q représente la charge des particules considérées (électrons ou trous), ϵ_{si} la constante diélectrique du silicium et ρ la concentration de charges du milieu où évolue la particule considérée.

La concentration de charges ρ à la surface d'un semi-conducteur de type p est donnée par :

$$\rho(y) = -q \cdot [n(y) + N_a(y) - p(y) - N_d(y)] \tag{4.5}$$

où N_a et N_d sont les concentrations des dopants accepteurs et donneurs, et n et p sont les densités d'électrons et de trous, respectivement. Dans le cas d'un transistor n-MOS, le semi-conducteur est de type p et (4.5) se réduit alors à :

$$\rho(y) = -q \cdot [n(y) + N_a(y) - p(y)] \tag{4.6}$$

Intéressons nous maintenant plus en détail aux équations des densités d'électrons $n(y)$ et de trous $p(y)$. Nous donnerons ici les résultats pour les électrons (sachant qu'il est facile de l'étendre ensuite au cas des trous). Puisqu'en mécanique quantique, le module au carré de la fonction d'onde $\varphi_i(y)$ représente la densité de probabilité de présence, en supposant que l'équilibre thermodynamique soit respecté, nous obtenons simplement [14] :

$$n(y) = \sum_i N_i \cdot |\varphi_i(y)|^2 \tag{4.7}$$

N_i représente le nombre total d'électrons occupant un niveau i donné, par unité de surface (ici la surface de la capacité MOS). Comme les fonctions d'onde sont normées, nous avons par ailleurs :

$$\int_0^{+\infty} n(y)dy = \sum_i N_i \tag{4.8}$$

[1]Ces deux grandeurs se compenseraient approximativement [3].

La grandeur Q, donnée par :

$$Q = -q \cdot \sum_i N_i \tag{4.9}$$

représente alors la charge liée à la présence des électrons attirés (ou repoussés) à la surface de la structure MOS.

En supposant que l'équilibre thermodynamique soit vérifié, la distribution énergétique des porteurs est donnée par la fonction de Fermi–Dirac $f(E)$, fonction de l'énergie totale E des porteurs (et non pas seulement de l'énergie ε dans la direction y) et du niveau de Fermi local. Pour calculer la grandeur N_i, il faut pouvoir dénombrer les états susceptibles d'accueillir des électrons ayant une énergie totale E et occupant le niveau i, c'est-à-dire ayant une contribution à l'énergie ε_i liée au mouvement dans la direction y. Remarquons que ce problème revient à estimer le nombre d'états disponibles pour des électrons libres dans deux degrés de dimension de l'espace, et ayant une énergie $E_\parallel = E - \varepsilon_i$.

Compte tenu de la nature anisotrope de la bande de conduction du silicium, quand le potentiel ne dépend que de y, — ce qui représente la direction (100) — nous pouvons séparer les électrons en deux familles : ceux qui ont une masse effective dans la direction (100) de valeur m_t et en conséquence une masse de densité d'états 2D donnée par $m^*_{dl} = 4\sqrt{m_t\, m_l}$, et ceux qui ont une masse effective dans la direction (100) de valeur m_l et donc une masse de densité d'états 2D donnée par $m^*_{dt} = 2\, m_t$. Cette anisotropie donne lieu à des traitements physiques distincts pour chacune des deux familles d'électrons, puisque la masse effective à utiliser pour résoudre l'équation de Schrödinger (4.2) vaut soit m_t, soit m_l ('l' désigne la famille des électrons longitudinaux et 't' des électrons transversaux).

Le nombre d'électrons N_i appartenant au niveau i s'exprime alors comme :

$$N_i = \int_0^{+\infty} \frac{m^*_{dl}}{\pi \cdot \hbar^2} \cdot f(E) dE_\parallel + \int_0^{+\infty} \frac{m^*_{dt}}{\pi \cdot \hbar^2} \cdot f(E) dE_\parallel \tag{4.10}$$

ce qui mène à :

$$N_i = \frac{m^*_{dl}}{\pi \cdot \hbar^2} \cdot k \cdot T \cdot \ln\left[1 + \exp\left(\frac{E_f - \varepsilon_i^{(l)}}{k \cdot T}\right)\right] + \frac{m^*_{dt}}{\pi \cdot \hbar^2} \cdot k \cdot T \cdot \ln\left[1 + \exp\left(\frac{E_f - \varepsilon_i^{(t)}}{k \cdot T}\right)\right] \tag{4.11}$$

Pour les trous, il y a également deux familles, correspondant aux trous lourds (désignés ici par 'hh' pour heavy holes) et aux trous légers ('lh' pour light holes).

Finalement, les densités d'électrons $n(y)$ et de trous $p(y)$ sont données par :

$$n(y) = \sum_i \frac{m^*_{dl}}{\pi \cdot \hbar^2} \cdot k \cdot T \cdot \ln\left[1 + \exp\left(\frac{E_f - \varepsilon_i^{(l)}}{k \cdot T}\right)\right] \cdot |\varphi_i^{(l)}(y)|^2 + \sum_j \frac{m^*_{dt}}{\pi \cdot \hbar^2} \cdot k \cdot T \cdot \ln\left[1 + \exp\left(\frac{E_f - \varepsilon_j^{(t)}}{k \cdot T}\right)\right] \cdot |\varphi_j^{(t)}(y)|^2$$

$$p(y) = \sum_i \frac{m^*_{lh}}{\pi \cdot \hbar^2} \cdot k \cdot T \cdot \ln\left[1 + \exp\left(\frac{E_f - \varepsilon_i^{(lh)}}{k \cdot T}\right)\right] \cdot |\varphi_i^{(lh)}(y)|^2 + \sum_j \frac{m^*_{hh}}{\pi \cdot \hbar^2} \cdot k \cdot T \cdot \ln\left[1 + \exp\left(\frac{E_f - \varepsilon_j^{(hh)}}{k \cdot T}\right)\right] \cdot |\varphi_j^{(hh)}(y)|^2 \quad (4.12)$$

Connaissant les densités de porteurs n et p, il est alors possible de calculer la densité de charges ρ donnée par (4.6), et de déduire l'énergie potentielle V en intégrant l'équation différentielle (4.4). L'énergie potentielle étant maintenant connue, il est alors possible de résoudre l'équation de Schrödinger (4.2). Ceci illustre bien le caractère auto-cohérent d'une résolution Schrödinger–Poisson couplée. En effet, il apparaît que pour résoudre l'équation de Schrödinger, il faut déjà en connaître le résultat, puisque les fonctions d'onde et les niveaux d'énergie servent à établir l'équation dont ils sont issus, par le biais de (4.12). En conséquence, l'obtention de résultats par une simulation Schrödinger–Poisson couplée nécessite forcément une résolution numérique.

La résolution numérique de ce système d'équation a été longuement étudiée dans la littérature, durant ces trente dernières années. Tout d'abord par les physiciens, dont l'article de ANDO, FOWLER et STERN fait figure de conclusion [3], et ensuite dans le domaine de la physique des dispositifs [5,11,15].

Actuellement, différents simulateurs auto-cohérents Schrödinger–Poisson sont accessibles, entre autres citons le simulateur de Berkeley [16], NEMO [17] ou encore NCSU [18]. Une comparaison très instructive entre les principaux simulateurs quantiques disponibles a récemment été réalisée par RICHTER [19]. Nous reviendrons d'ailleurs sur les résultats de cette comparaison ultérieurement.

En conclusion, ces simulateurs numériques physiques sont donc un moyen très efficace pour modéliser précisément l'impact des effets quantiques sur les caractéristiques électriques des structures MOS. Ils donnent accès à différentes grandeurs électriques clés, telles que les caractéristiques capacitives *C–V*, la valeur de la charge d'inversion Q_{inv} en fonction de la polarisation de la grille ou encore le potentiel de surface ϕ_s en fonction de la polarisation de la grille. Cependant, en dépit de leurs avantages, ils ne sont pas du tout adaptés à la simulation de circuits, pour différentes raisons. Tout d'abord, les résolveurs Schrödinger–Poisson précédemment cités ne permettent pas de simuler le comportement électrique d'un transistor MOS, mais juste celui d'une capacité MOS. En outre, le temps de calcul requis par la résolution itérative des équations de Schrödinger et de Poisson est relativement important. Cette façon de prendre en compte les effets quantiques ne répond donc pas aux exigences de la modélisation compacte. Mais en définitive, malgré ces limitations, les simulateurs quantiques sont d'un intérêt majeur pour le développeur de modèles compacts. En effet, de par leur précision inhérente à leur nature physique, ils sont un outil idéal en ce qui concerne la validation des modèles analytiques quantiques destinés à la simulation de circuits.

4.4 Méthodes d'approximations analytiques

Dans le but de résoudre de manière approchée les équations couplées de Poisson et Schrödinger, deux principales méthodes d'approximations ont été développées [1,2,20]. Nous utiliserons d'ailleurs chacune de ces méthodes dans le cadre du modèle quantique de potentiel de surface réalisé dans cette thèse. La recherche de méthodes d'approximations se justifie par deux raisons fondamentales : d'une part, comme nous l'avons dit précédemment, la résolution auto-cohérente du système d'équations Schrödinger–Poisson nécessite du temps de calcul, et d'autre part l'obtention de formules simplifiées permet d'envisager la création de modèles compacts destinés à la simulation de circuits ou à la caractérisation.

4.4.1 L'approximation variationnelle

Parmi toutes les approches simplifiées, l'approximation variationnelle est de loin la méthode approchée la plus efficace [3,14,20,21]. Elle s'applique dans les cas où la prise en compte d'un seul niveau d'énergie est suffisante pour modéliser correctement le confinement quantique des porteurs, c'est-à-dire en régime d'inversion forte. Son domaine de validité s'étend néanmoins à l'intégralité du régime d'inversion, tout en conservant une précision acceptable.

Nous allons maintenant décrire les principaux résultats de cette méthode, qui nous seront utiles dans la suite de notre travail. Dans le cadre de l'approche variationnelle, une expression analytique simple de la fonction d'onde représentant le niveau fondamental E_0 est donnée par :

$$\Psi_b(y) = \frac{b^{3/2}}{\sqrt{2}} \cdot y \cdot \exp\left[\frac{-b \cdot y}{2}\right] \tag{4.13}$$

où b est un paramètre homogène à l'inverse d'une longueur, qui pour l'instant est non déterminé. Cette fonction d'onde n'étant qu'une approximation, l'équation de Schrödinger (4.2) devient :

$$\mathrm{H}\,\Psi_b \approx E(b)\,\Psi_b \tag{4.14}$$

où H est l'Hamiltonien du système et $E(b)$ son énergie. Le principe de la méthode variationnelle consiste à chercher parmi toutes les valeurs de b, celle pour laquelle la fonction d'onde Ψ_b s'approche le plus d'un état propre de l'Hamiltonien H. Ceci est alors vérifié par un extremum de la fonction $E(b)$, définie par :

$$E(b) = \langle \Psi_b \,|\, \mathrm{H} \,|\, \Psi_b \rangle = \int \Psi_b(y)^* \,\mathrm{H}[\Psi_b](y)\, dy \tag{4.15}$$

En déterminant l'extremum de $E(b)$, on peut alors calculer la fonction d'onde et l'énergie correspondante de manière approchée [14]. Le détail de ces calculs n'étant pas d'un intérêt fondamental pour notre étude, nous ne le présenterons pas ici ; il peut d'ailleurs être trouvé dans l'ouvrage de référence de MATHIEU [22].

Par contre, ce qui nous sera très utile est la connaissance de l'expression du paramètre b minimisant la fonction $E(b)$:

$$b(\phi_s) = \left[\frac{12 \cdot m^* \cdot q^2}{\epsilon_{si} \cdot \hbar^2} \cdot \left(\frac{n_{inv}(\phi_s)}{3} + n_{dep}(\phi_s)\right)\right]^{1/3} \tag{4.16}$$

où m^* représente la masse effective des porteurs libres, n_{inv} la densité de porteurs libres de la couche d'inversion et n_{dep} la densité d'atomes fixes ionisés dans la couche de déplétion.

Ainsi l'expression de l'énergie $E(b)$ pour la valeur particulière de b donnée par (4.16) correspond au niveau d'énergie fondamentale E_0. Son expression est la suivante :

$$E(b) = E_0 = \frac{3}{8}\frac{\hbar^2}{m^*} \cdot b^2 \tag{4.17}$$

Pour être précis, il convient de noter que la valeur de $E(b)$ obtenue avec (4.17) correspond à la différence d'énergie entre le premier niveau d'énergie quantifié et le bas de la bande de conduction E_c. L'équation (4.17) est donc une représentation approchée, mais cependant physique, du phénomène d'élargissement du gap ΔE_g, illustré en début de chapitre à la Fig. 4.1a, page 82.

Finalement, pour une valeur de potentiel de surface donnée, les relations (4.15), (4.16) et (4.17) forment un système d'équations auto-cohérentes, de solution $n_{inv}(\phi_s)$, $E(\phi_s)$ et $b(\phi_s)$. Il est ainsi possible de reproduire toutes les grandeurs obtenues par une simulation couplée Schrödinger–Poisson standard. En outre, un avantage majeur de cette méthode est qu'elle permet, moyennant certaines approximations, de développer des modèles simplifiés et complètement analytiques, ce que nous montrerons par la suite.

4.4.2 L'approximation du puits de potentiel triangulaire

Cette méthode qui consiste à approcher la forme du puits de potentiel par une droite, est l'une des plus employées en modélisation compacte [3,10,22–24]. D'un point de vue physique, l'approximation du potentiel triangulaire est réaliste tant que la charge de la couche d'inversion reste faible devant celle de la couche de déplétion [20,21]. Cette méthode est donc tout naturellement indiquée pour le calcul des couches d'accumulation, de déplétion et d'inversion faible. Cependant de nombreux auteurs l'utilisent aussi pour le domaine d'inversion tout entier, c'est-à-dire inversion forte comprise [5,7,10,25–27]. Bien que son utilisation en inversion forte permette d'obtenir des résultats corrects pour le calcul de la densité de porteurs de la couche d'inversion (n_{inv}), elle engendre d' importantes erreurs en ce qui concerne le calcul du potentiel de surface [5].

Comme son nom l'indique, le principe général de cette méthode est de supposer que la forme du puits de potentiel est triangulaire. De plus, si l'on admet que le puits est infiniment profond, l'équation de Schrödinger admet alors des solutions analytiques sous la forme de fonctions d'Airy A_i :

$$\Psi_i(y) = A_i \cdot \left(\frac{2 \cdot m^* \cdot q \cdot F_s}{\hbar^2} \right)^{1/3} \cdot \left[y - \frac{E_i}{q \cdot F_s} \right] \tag{4.18}$$

dont les énergies correspondantes sont données par :

$$E_i = \left(\frac{\hbar^2}{2 \cdot m^*}\right)^{1/3} \cdot \left[\frac{3}{2} \cdot \pi \cdot q \cdot F_s \left(i + \frac{3}{4}\right)\right]^{2/3} \tag{4.19}$$

pour un puits de potentiel de forme $V(y) = q \cdot F_s \cdot y$, où F_s désigne le champ électrique normal à l'interface Si–SiO_2 (souvent appelé champ de surface), qui est défini par :

$$F_s = -\frac{Q_{inv} + Q_b}{\epsilon_{si}} \tag{4.20}$$

Il est utile de remarquer que de façon similaire à l'équation (4.17), les niveaux d'énergie calculés par (4.19) sont mesurés par rapport au bas de la bande de conduction à l'interface Si–SiO_2. En ne considérant que le premier niveau d'énergie E_0, l'équation (4.19) devient à l'instar de (4.17) une représentation approchée du phénomène d'élargissement du gap ΔE_g.

Tout comme la méthode variationnelle, cette méthode est également auto-cohérente, dans la mesure où la pente de la droite $V(y) = q \cdot F_s \cdot y$ dépend de la charge répartie sur les différents niveaux, et réciproquement. Elle permet donc également de reproduire les grandeurs obtenues par une simulation couplée Schrödinger–Poisson standard. Pour conclure, rappelons que l'approximation du potentiel triangulaire est tout particulièrement destinée au traitement quantique des couches d'accumulation, puisqu'elle permet de facilement prendre en compte plusieurs niveaux d'énergie, ce qui est nécessaire à une description correcte de la charge totale d'accumulation [14].

4.5 État de l'art

Le développement de modèles analytiques prenant en compte les effets quantiques date d'environ une dizaine d'années. De nombreuses approches ont été élaborées, avec un degré d'empirisme très variable. L'état de l'art que nous allons dresser ne se veut pas exhaustif, — cela nécessiterait ce manuscrit tout entier — le but étant plutôt de montrer les différentes approches mises en oeuvre, ainsi que leur utilité réelle. Dans un premier temps, nous allons exposer les premiers modèles publiés, puis nous ferons une étude séparée, distinguant l'inclusion des effets quantiques au sein des modèles en tension de seuil et des modèles en potentiel de surface.

4.5.1 Travaux pionniers

Les effets quantiques ont originellement suscité de l'intérêt en raison de leur impact sur la tension de seuil du TMOS, cette caractéristique étant très importante aux yeux du concepteur. Les travaux de VAN DORT ont été d'une grande importance à cet égard [9,10]. En particulier, ils ont montré l'influence des forts niveaux de dopage substrat sur la tension de seuil.

Avant d'entrer dans le détail des recherches de VAN DORT, quelques explications physiques permettant de bien comprendre pourquoi les effets quantiques modifient la tension de seuil, sont nécessaires. Le phénomène physique en jeu est le suivant : les niveaux croissants de dopage substrat résultent en un puits de potentiel de plus en plus abrupte, ce qui accroît la séparation des niveaux d'énergie — pour toutes les températures —, et donc augmente le gap apparent du silicium. *En conséquence, un potentiel de surface plus élevé est nécessaire pour obtenir une population donnée dans la bande de conduction (i.e. une charge donnée dans le canal), ce qui forcément entraîne une augmentation de la tension de seuil en comparaison au cas classique.*

En se positionnant d'un côté plus conception de circuits et moins modélisation axée sur la physique, l'augmentation de la tension de seuil peut être expliquée plus simplement, ou du moins en des termes plus parlant pour le concepteur.
Dans une approche classique, le passage de la faible à la forte inversion correspond à une valeur de potentiel de surface ϕ_s généralement définie par :

$$\phi_B = 2\,\phi_f \quad \text{avec} \quad \phi_f = \phi_t \cdot \ln(N_a/n_i) \tag{4.21}$$

où n_i désigne la concentration intrinsèque de porteurs du silicium.
Étant donné que $(V_g - V_{fb} - \phi_s) = -Q_{sc}/C_{ox}$ (cf. (2.7)), la tension de seuil s'écrit comme[1] :

$$V_{th} = V_{fb} + \phi_B + Q_b(\phi_B)/C_{ox} = V_{fb} + \phi_B + \gamma\sqrt{\phi_B} \tag{4.22}$$

où γ est le coefficient de substrat et V_{fb} la tension de bandes plates. Il apparaît alors que (4.22) dépend de la concentration intrinsèque n_i, dont l'expression physique est donnée par :

$$n_i = \sqrt{N_c \cdot N_v} \cdot \exp\left[\frac{-E_g}{2 \cdot k \cdot T}\right] \tag{4.23}$$

où N_c désigne la densité d'états effective dans la bande de conduction ($N_c = 1.04 \times 10^{19}$ cm^{-3}),

[1] Q_{sc} est la densité de charge du semi-conducteur, égale à $Q_{inv} + Q_b$. Dans le contexte des modèles basés sur la définition d'une tension de seuil, il est généralement supposé que $Q_{inv}(\phi_B) = 0$, d'où l'équation (4.22).

N_v la densité d'états effective dans la bande de valence ($N_v = 2.8 \times 10^{19}$ cm^{-3}) et E_g la largeur de gap du silicium. Or nous avons vu qu'en raison de la quantification de l'énergie des bandes permises (ici la bande de conduction), le gap est augmenté d'une valeur ΔE_g correspondant à l'énergie du niveau fondamental E_0. Il est donc naturel de définir une concentration intrinsèque quantique de porteurs de la façon suivante :

$$n_{i[\mathrm{qm}]} = n_i \cdot \exp\left[\frac{-\Delta E_g}{2 \cdot k \cdot T}\right] \tag{4.24}$$

ce qui mène à une nouvelle définition de (4.21) :

$$\phi_{B[\mathrm{qm}]} = 2 \cdot \phi_t \cdot ln(N_a/n_{i[\mathrm{qm}]}) = \phi_B + \Delta E_g/q \tag{4.25}$$

Ainsi la tension de seuil n'est plus définie pour un potentiel de surface égal à ϕ_B, mais à une valeur supérieure, égale à $\phi_B + \Delta E_g/q$, d'où son augmentation par rapport au cas classique.

Ces quelques explications étant données, nous allons maintenant expliquer l'approche proposée par VAN DORT pour modéliser la modification du gap due aux effets quantiques. Dans son premier modèle, l'élargissement du gap ΔE_g est directement relié au dopage substrat N_a par l'expression empirique suivante [9] :

$$\Delta E_g = \beta \cdot \left[N_a \cdot \ln\left(\frac{N_a}{n_i}\right)\right]^{1/3} \tag{4.26}$$

où β est un paramètre d'ajustement égal à 4.3×10^{-8} eV·cm.

Une première limitation à cette expression est qu'elle est indépendante de la tension de grille du TMOS. Elle est donc utilisable uniquement à la condition d'inversion définie par $V_g = V_{th}$, i.e. au seuil d'inversion forte. De plus, la dépendance à l'épaisseur d'oxyde t_{ox} n'est pas prise en compte dans ce modèle, cette grandeur étant pourtant directement impliquée dans l'accroissement du champ électrique surfacique F_s. Bien que ce second point soit une réelle lacune de ce modèle d'élargissement du gap, si l'on se restreint au seul cas de l'étude du décalage de la tension de seuil, l'expression (4.26) est quand même valide. En fait, l'influence de l'épaisseur d'oxyde est d'une importance capitale uniquement en inversion modérée et forte, tandis qu'en inversion faible c'est le dopage substrat qui est le point clé dans l'étude des effets quantiques. Nous reviendrons en détail sur ce point, lorsque nous présenterons notre nouveau modèle.

Dans son modèle, VAN DORT propose en fait de définir un élargissement de gap effectif ΔE_{geff} qui inclut trois termes différents. À ΔE_g s'ajoute également un terme prenant en compte le déplacement quantique Δy de la distribution de porteurs $n_{inv}(y)$ par rapport à l'interface Si–SiO_2, défini par $qF_s\Delta y$. La valeur du champ surfacique est calculée au seuil[1], i.e. $F_s = F_s(\phi_B)$, avec $\phi_B = 2\,\phi_f$. Le terme Δy est quant à lui obtenu dans l'hypothèse de l'approximation du potentiel triangulaire, en supposant que seul le premier niveau d'énergie est occupé :

$$\Delta y = \frac{1}{q \cdot F_s(\phi_B)} \cdot \left(\frac{2 \cdot \Delta E_g}{3} - q \cdot \phi_t \right) \tag{4.27}$$

Le troisième terme composant l'expression de ΔE_{geff} compense légèrement les deux premiers, puisqu'il est défini comme un rétrécissement du gap ΔE_c (bandgap narrowing), a priori lié aux effets de forts dopages. La formule employée est celle proposée par SLOTBOOM [28] :

$$\Delta \tilde{E}_c = 0.009 \cdot \left[\ln\left(\frac{N_a}{10^{17}}\right) + \sqrt{\ln\left(\frac{N_a}{10^{17}}\right)^2 + 0.5} \right] \tag{4.28}$$

où N_a est exprimé en cm^{-3} et $\Delta \tilde{E}_c$ en V (ΔE_c en eV).

Finalement l'élargissement effectif du gap est donné par :

$$\Delta E_{geff} = \Delta E_g - q \cdot \Delta \tilde{E}_c + q \cdot F_s \cdot \Delta y \tag{4.29}$$

En d'autres mots, cela revient à dire qu'une courbure de bandes supplémentaire est nécessaire pour arriver à l'état d'inversion. Cet incrément du potentiel de surface s'écrit simplement comme :

$$\Delta \phi_s = \frac{\Delta E_{geff}}{q} = \frac{\Delta E_g}{q} - \Delta \tilde{E}_c + F_s \cdot \Delta y \tag{4.30}$$

En conséquence, la tension de seuil tenant compte des corrections quantiques est donnée par :

$$V_{th[\mathrm{qm}]} = V_{fb} + \phi_B + \Delta \phi_s + \gamma \sqrt{\phi_B + \Delta \phi_s} \tag{4.31}$$

ce qui implique, au premier ordre en $\Delta \phi_s / \phi_B$:

$$\Delta V_{th} = V_{th[\mathrm{qm}]} - V_{th[\mathrm{cl}]} \simeq \Delta \phi_s \cdot \left(1 + \frac{1}{2 \cdot C_{oxeff}} \cdot \sqrt{\frac{q \cdot \epsilon_{si} \cdot N_a}{\phi_f}} \right) \tag{4.32}$$

[1] Dans ce modèle le champ de surface ne tient compte que de la charge de déplétion, d'où : $F_s = F_{dep} = Q_b/\epsilon_{si}$.

avec

$$C_{oxeff} = \frac{\epsilon_{ox}}{t_{oxeff}} = \frac{\epsilon_{ox}}{t_{ox} + \frac{\epsilon_{ox}}{\epsilon_{si}} \cdot \Delta y} \tag{4.33}$$

Dans (4.32), $V_{th[\mathrm{cl}]}$ représente la tension de seuil classique, définie précédemment en (4.22).

Avant d'observer les résultats obtenus par ce modèle, nous allons présenter sa seconde version [10], puis nous comparerons alors la validité des deux modèles successivement proposés par VAN DORT.

Ce second modèle reprend en partie les résultats du précédent. Les principaux changements consistent en trois nouvelles définitions : celle de ΔE_g, de Δy et de ΔE_{geff}.
L'expression utilisée pour modéliser ΔE_g est dérivée de l'approximation du puits de potentiel triangulaire, en considérant seulement l'occupation du premier niveau d'énergie :

$$\Delta E_g = \beta \cdot \left(\frac{\epsilon_{si}}{4 \cdot k \cdot T}\right)^{1/3} \cdot \max(F_s, 0)^{2/3} \tag{4.34}$$

où la valeur de β[1] est maintenant fixée à 4.1×10^{-8} eV·cm. Cette équation dépend du champ électrique F_s, ce qui rend ΔE_g dépendant de la tension appliquée à la grille du TMOS. Ainsi écrite, cette formulation analytique de l'élargissement du gap n'est donc pas valable uniquement au niveau du seuil, mais pour toute la région d'inversion, ce qui représente un net progrès par rapport à la relation donnée par (4.26).

L'effet de déplacement de la distribution de porteurs (ou centroïde d'inversion) depuis l'interface est modélisé suivant une approche variationnelle, contrairement au premier modèle (cf. (4.27)). Après simplification, la formulation suivante est proposée [10] :

$$q \cdot F_s \cdot \Delta y = \frac{4}{9} \cdot \Delta E_g \tag{4.35}$$

En combinant les équations (4.34) et (4.35), une nouvelle expression pour l'élargissement effectif du gap est introduite :

$$\Delta E_{geff} = \Delta E_g + q \cdot F_s \cdot \Delta y = \frac{13}{9} \cdot \Delta E_g \tag{4.36}$$

Une différence notable entre ce second modèle caractérisé par (4.36) et le premier caractérisé par (4.29) est l'absence du terme prenant en compte le rétrécissement du gap lié aux effets de forts dopages. Nous approuvons d'ailleurs ce second choix car la possibilité réelle

[1] Dans (4.34), β représente en fait une constante physique dont la valeur théorique est égale à 4.4×10^{-8} eV·cm.

que le 'bandgap narrowing' puisse avoir une influence sur la tension de seuil n'est pas du tout évidente et n'a pas été clairement démontrée [23].

Nous allons maintenant analyser les résultats obtenus sur le calcul de la tension de seuil avec les deux modèles que nous venons de présenter. Une première constatation apparaît à la lecture de la Fig. 4.4 : le second modèle surestime nettement l'augmentation de la tension de seuil due aux effets quantiques, et cela quel que soit le niveau de dopage. Le premier modèle est proche des résultats obtenus en simulations numériques[1], mais surestime légèrement V_{th} pour des dopages substrat élevés.

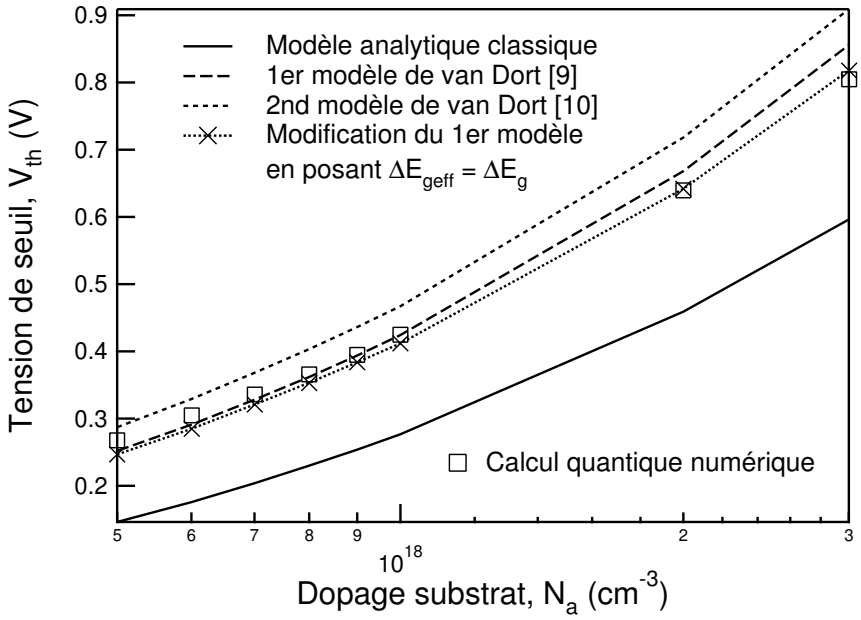


FIG. 4.4 : *Influence du dopage substrat sur la tension de seuil du transistor n-MOS. L'épaisseur d'oxyde t_{ox} est fixée à 2.3 nm. Les symboles représentent la tension de seuil quantique extraite des simulations auto-cohérentes Schrödinger–Poisson. Cette dernière est utilisée comme référence pour tester la validité des modèles analytiques.*

Nous avons une théorie pour expliquer cela. Tout d'abord, il ne nous semble pas réaliste de considérer le terme relatif au changement de forme de la distribution de porteurs dans l'expression de $\Delta\phi_s$ (quel que soit le modèle considéré). À notre sens, ΔE_g et Δy ne devraient pas être considérés de façon indépendante, ces deux représentations étant fortement auto-corrélées. Ainsi, ajouter simplement entre-eux ΔE_g et Δy dans l'expression de $\Delta\phi_s$, n'a

[1]Nous avons développé une méthode efficace d'extraction de la tension de seuil à partir des courbes *C–V* (classique ou quantique). Cette procédure est détaillée en Annexe, page 175.

pas réellement de justification physique, voire même de sens. Il suffit d'ailleurs de regarder le diagramme des bandes montré à la Fig. 4.1 (page 82), pour se convaincre que la courbure de bande supplémentaire $\Delta\phi_s$ — due aux effets quantiques — correspond simplement à $E_0 - E_c$, c'est-à-dire à ΔE_g. Cette hypothèse sera d'ailleurs validée dans le cadre du modèle que nous avons développé. Pour résumer notre point de vue, la considération fusionnelle des phénomènes d'élargissement du gap et de déplacement de la distribution des porteurs est une erreur. Finalement, puisque le modèle de VAN DORT traite le cas d'un modèle d'élargissement du gap, il serait donc logique de considérer uniquement le terme ΔE_g.

En second lieu, comme nous l'avons déjà dit et comme tous les chercheurs du domaine l'admettent[1], nous ne prendrions pas en compte l'hypothétique rétrécissement du gap lié aux effets de forts dopages (cas du premier modèle VAN DORT [9]).

En résumé, cela revient à dire, — en reprenant les notations précédentes — que nous supposons :

$$\Delta E_{geff} \equiv \Delta E_g \tag{4.37}$$

Revenons maintenant aux résultats du premier modèle (cf. Fig. 4.4), qui d'ailleurs sont relativement bons. A priori, ce constat peut sembler contradictoire au regard de notre raisonnement précédent... En fait, cela s'explique simplement : dans le premier modèle, le terme Δy prenant en compte le déplacement de la distribution de porteurs est compensé par le terme de rétrécissement du gap $\Delta\tilde{E}_c$ de telle sorte que $F_s\Delta y \simeq \Delta\tilde{E}_c$. C'est pour cette raison que le modèle est proche des résultats théoriques. Pour prouver la validité de cette hypothèse, nous avons calculé la tension de seuil (avec ce premier modèle) en considérant d'une part ΔE_g au lieu de ΔE_{geff} et d'autre part C_{ox} au lieu de C_{oxeff} dans (4.32). Les résultats de ce modèle modifié sont très satisfaisants, comme le montre la Fig. 4.4.

En ce qui concerne les résultats du second modèle, ils surévaluent clairement l'influence des effets quantiques sur le seuil. Pour tester à nouveau notre théorie précédente, nous avons modifié le second modèle en supprimant le terme en Δy dans (4.36), ce qui implique l'utilisation de la valeur classique de C_{ox} dans l'expression de $\Delta\phi_s$ (cf. (4.32))[2]. Les résultats présentés à la Fig. 4.5 montrent le bon accord obtenu avec les valeurs de tension de seuil extraites des simulations numériques (simulations couplées Schrödinger–Poisson). Il est clair

[1] C'est-à-dire les personnes s'occupant de la prise en compte des effets quantiques dans le cadre des modèles compacts de TMOS [8,10,16,21,23,27,29–32].

[2] Exactement comme dans le cas de la modification apportée au premier modèle.

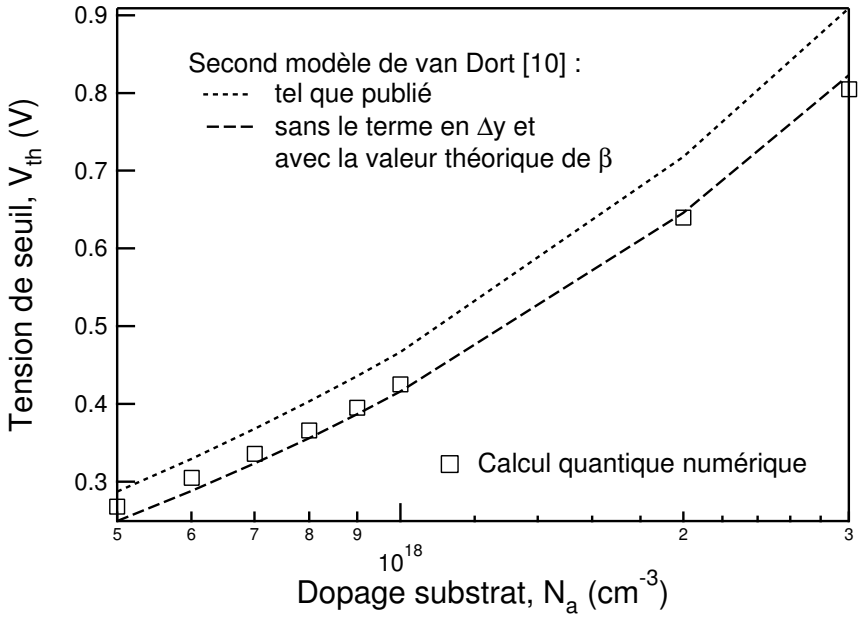


FIG. 4.5 : *Impact du terme Δy sur le calcul de la tension de seuil quantique. Les caractéristiques de simulation sont identiques à la Fig. 4.4. La valeur théorique de β utilisée dans le cas du modèle modifié (tirets) est égale à 4.4×10^{-8} eV·cm. Celle utilisée dans le modèle publié (pointillés) était ajustée à 4.1×10^{-8} eV·cm.*

que notre modification (simplification) améliore significativement le comportement du modèle. Enfin, un dernier point important à préciser est que nous avons utilisé la valeur théorique de β lors de la modification du modèle, ce qui en renforce le caractère physique.

Un article instructif à ce sujet est celui de BUCHER et al. [30]. Ce papier propose une méthodologie d'implémentation des effets quantiques au sein du modèle EKV, basée sur le second modèle de VAN DORT [10]. Un point intéressant de ce travail est que les auteurs ont dû ajuster le coefficient β pour réduire la valeur de $\Delta\phi_s$[1]...
Finalement, ils ont utilisé l'expression suivante :

$$\Delta\phi_s = \frac{13}{9}\cdot\beta^*\left(\frac{\epsilon_{si}}{4\cdot k\cdot T}\right)^{1/3}\cdot\frac{F_s^{2/3}}{q} = \frac{1}{q}\left[\frac{4}{9}\cdot\Delta E_g^* + \Delta E_g^*\right] \tag{4.38}$$

qui au regard de la valeur β^* choisie, pourrait en fait se réécrire comme :

$$\Delta\phi_s \simeq \frac{\Delta E_g}{q} \tag{4.39}$$

[1]Dans la but d'observer un bon accord avec les résultats expérimentaux.

ce qui in fine revient à ne pas considérer le terme relatif au déplacement de la distribution de porteurs dans l'expression de l'élargissement effectif du gap, de telle sorte que $\Delta E_{geff} \rightarrow \Delta E_g$. Sans le savoir, les conclusions de cet article nous confortent donc dans notre position vis à vis du traitement quantique de l'élargissement du gap.

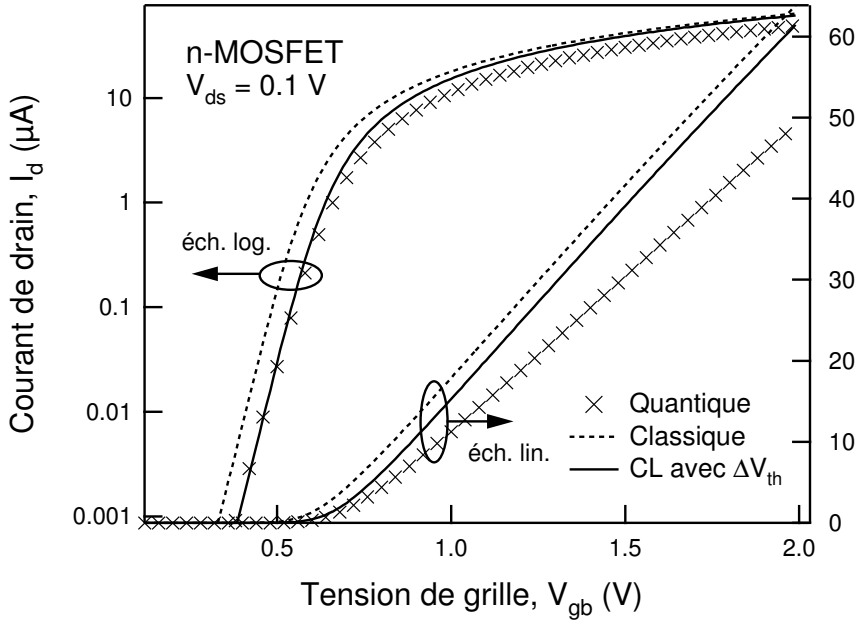
En conclusion, il semble évident de devoir reconnaître le mérite des travaux de VAN DORT. Ils tracèrent en effet la voie menant à la prise en compte des effets quantiques en modélisation compacte. Bien que certaines approches utilisées soient discutables, il est indéniable que ces modèles historiques ont servi de base de réflexion à de nombreux chercheurs ; un certain nombre d'auteurs ayant repris ces travaux, en les modifiant avec plus ou moins de réussite. Citons par exemple HARELAND qui a développé un modèle d'élargissement du gap pour les couches d'accumulation, simplement en modifiant la valeur du paramètre β [21]. Ce paramètre n'a alors plus aucun caractère physique : il est redéfini comme une fonction purement empirique donnée par $a/(1 + N_a/b)$, où a et b sont des paramètres d'ajustement du modèle.

4.5.2 Approches en tension de seuil

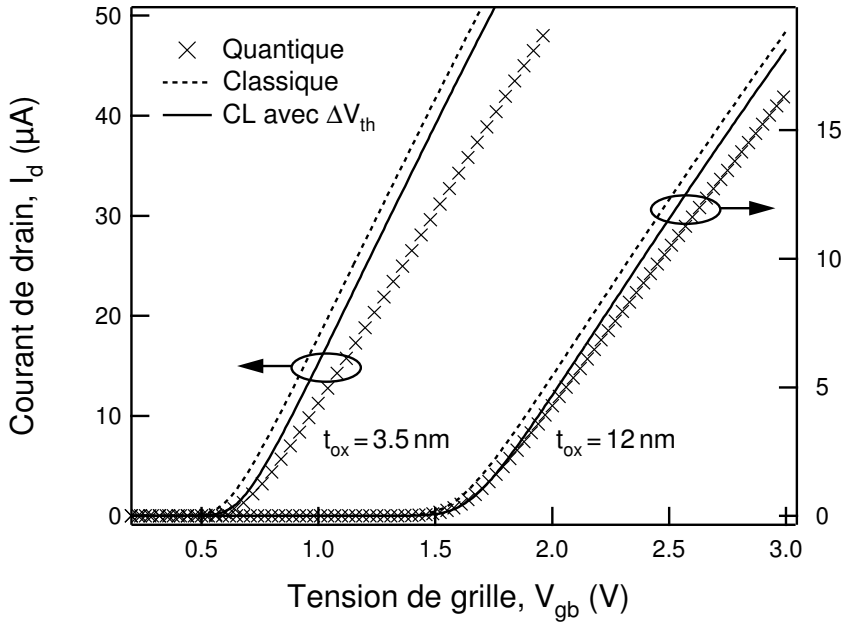
Les modèles compacts utilisant la définition d'une tension de seuil sont encore largement utilisés en simulation de circuits. C'est pourquoi, malgré notre intérêt prononcé pour les modèles dits en potentiel de surface, il nous semble quand même légitime d'étudier l'inclusion des effets quantiques dans ce type de modèle. Nous nous limiterons au cas emblématique du modèle de Berkeley, à savoir BSIM4v2 [32].

Au cours du paragraphe précédent, nous avons discuté de l'impact des effects quantiques sur la tension de seuil. En particulier, nous avons montré que le premier modèle de VAN DORT utilise une expression indépendante de la polarisation de la grille pour modéliser l'élargissement du gap [9], ce qui limite la validité du modèle de ΔE_g à la seule description du seuil d'inversion. Or comme nous l'avons déjà montré à la Fig. 4.2 (page 84), les effets quantiques sont la cause non seulement d'un décalage de la tension de seuil en inversion faible, mais aussi d'un changement de la pente de la densité de porteurs[1] en inversion modérée et forte. Ainsi la seule prise en compte du décalage du seuil ne permet pas de modéliser correctement le comportement électrique du TMOS en inversion modérée et forte, ce qui est encore plus flagrant dans le cas d'oxydes de grille ultra-fins. Ceci est illustré à la Fig. 4.6, en prenant l'exemple du calcul du courant de drain I_d. Les résultats présentés ont été obtenus par une simulation drift–diffusion.

[1]Et par suite logique, la densité de charge d'inversion Q_{inv} et le courant de drain I_d sont aussi modifiés.



(a) $t_{ox} = 3.5$ nm et $N_a = 5 \times 10^{17}$ cm^{-3}.



(b) $N_a = 5 \times 10^{17}$ cm^{-3}.

FIG. 4.6 : *Comparaison de différentes approches quantiques pour simuler le courant de drain. Dans tous les cas, les dimensions du dispositif sont $W/L = 10/10$ µm. La notation 'CL' signifie classique.*

Deux constats se dégagent de cette figure. D'une part, il est vrai que l'utilisation d'un modèle classique incluant le phénomène de décalage de la tension de seuil permet de calculer précisément la pente sous le seuil, même dans le cas d'oxydes fins. D'autre part, les régions d'inversion modérée et d'inversion forte sont très mal décrites, et cela particulièrement dans le

cas des faibles épaisseurs d'oxydes. D'un point de vue quantitatif, en se plaçant à $V_g = 2$ V l'erreur du modèle classique à décalage de seuil est de ∼10 % pour $t_{ox} = 12$ nm contre ∼25 % pour $t_{ox} = 3.5$ nm.

Pour résoudre ce problème, avant toute chose, il faut bien être conscient qu'il n'est pas réaliste de ne considérer que le décalage de la tension de seuil pour modéliser convenablement les effets quantiques sur toute la dynamique de tension de grille. En d'autres termes, il est donc indispensable d'utiliser une expression dépendante de V_g pour modéliser l'élargissement du gap. Par exemple, dans le cadre de l'approximation triangulaire, une équation de la forme

$$\Delta E_g = E_0 = \left(\frac{\hbar^2}{2 \cdot m^*}\right)^{1/3} \cdot \left[\frac{9}{8} \cdot \pi \cdot q \cdot F_s\right]^{2/3} \tag{4.40}$$

conviendrait ; et donc a fortiori son homologue obtenue avec l'approximation variationnelle. Dans cette équation, F_s représente le champ électrique normal à l'interface Si–SiO_2, qui dépend de la tension de grille.

Connaissant $\Delta E_g(F_s)$, il est facile de modéliser la quantité Δy — le déplacement de la distribution de porteurs de l'interface —, soit dans le cadre de l'approximation triangulaire, soit dans celui de l'approximation variationnelle [20]. Les effets quantiques peuvent alors être relativement bien pris en compte par la définition d'une capacité d'oxyde effective, qui est une fonction de la polarisation de la grille à travers le terme Δy. Dans un souci de clarté, nous en rappelons ici l'expression :

$$C_{oxeff} = \frac{\epsilon_{ox}}{t_{oxeff}} = \frac{\epsilon_{ox}}{t_{ox} + \frac{\epsilon_{ox}}{\epsilon_{si}} \cdot \Delta y} \tag{4.41}$$

C'est précisément ce qui est fait au sein des principaux modèles compacts formulés en tension du seuil, à savoir BSIM3v3 et BSIM4. D'une manière extrêmement simplifiée[1], voici l'approche choisie dans ces deux modèles : la définition de la capacité d'oxyde effective est utilisée pour le calcul des différentes caractéristiques électriques telles que la tension de seuil, le courant de drain (*I*–*V*), les capacités (*C*–*V*), etc.

Il apparaît alors immédiatement une limitation à cette méthode : l'augmentation quantique de la tension de seuil sera sous-évaluée, puisqu'elle ne tient compte que du terme C_{oxeff} dans sa

[1]En fait, BSIM fait cohabiter deux expressions différentes de capacité d'oxyde effective, la première devant être fournie par l'utilisateur et la seconde étant alors définie comme une fonction de la première [32] !

définition. Il faut d'ailleurs bien comprendre que ce problème résulte uniquement de l'obligation à considérer une valeur fixée de C_{oxeff} dans l'équation de V_{th}. Dans l'absolu, il est tout à fait possible de prendre en compte totalement les effets quantiques (i.e. de l'inversion faible à forte) en se basant uniquement sur la notion de capacité d'oxyde effective, *mais cela en aucun cas dans le contexte d'un modèle formulé en termes de tension de seuil.*

Dans le cas du dernier modèle BSIM en date (BSIM4v2), l'expression choisie pour modéliser le déplacement du centroïde d'inversion est donnée par [32] :

$$\Delta y = X_{dc} = \frac{1.9 \times 10^{-9}}{1 + \left(\frac{V_{gsteff} + 4 \cdot (V_{th0} - V_{fb} - \phi_s)}{2 \cdot t_{ox}} \right)^{0.7}} \tag{4.42}$$

où V_{gsteff} est une fonction de lissage empirique et complexe, utilisée non seulement pour améliorer la convergence du modèle mais aussi pour en compenser ses faibles bases physiques.

En guise de conclusion et toujours en se plaçant dans le contexte des modèles en tension de seuil, voici une solution envisageable pour prendre en compte plus proprement les effets quantiques sur les deux grandeurs ΔV_{th} et ΔC_{ox}. Il faudrait procéder de façon indépendante, ce qui a priori ne devrait pas être problématique avec ce type de modèle, puisque par nature, leurs équations sont définies en deux catégories indépendantes : en dessous et au-dessus du seuil. Nous proposerions l'approche suivante : il faudrait tout d'abord modifier la tension de seuil d'une façon similaire à celle que nous avons proposée (cf. notre modification du modèle de VAN DORT). Le décalage de la tension de seuil — qui correspond aux effets quantiques en inversion faible — serait donc convenablement modélisé. Enfin, pour une modélisation correcte de l'inversion forte, il suffirait de remplacer C_{ox} par C_{oxeff} dans l'équation de la charge d'inversion correspondant à ce même régime, i.e. $Q_{inv} \simeq q \cdot C_{oxeff} \cdot (V_g - V_{th[\mathrm{qm}]})$.

4.5.3 Approches en potentiel de surface

L'intérêt fondamental des modèles formulés potentiel de surface est lié au fort caractère physique inhérent à ce type d'approche. En particulier, l'absence de notion de tension de seuil simplifie considérablement la prise en compte des effets quantiques en comparaison aux modèles du même nom. Il n'y a plus à considérer d'une part, un décalage du seuil d'inversion (effets quantiques en inversion faible) et d'autre part, une capacité d'oxyde effective pour modéliser l'inversion modérée et forte. L'impact des effets quantiques peut alors être considéré de manière unifiée et cohérente de l'inversion faible à l'inversion forte. En termes de

compréhension physique cela est réellement avantageux, en dépit de la structure de base plus complexe des modèles en potentiel de surface. Rappelons enfin que la connaissance exacte du potentiel de surface permet à elle seule de simuler précisément toutes les autres caractéristiques électriques. Ainsi, prendre en compte directement l'impact des effets quantiques sur le potentiel de surface est une approche a priori prometteuse, mais cependant loin d'être évidente.

En fait, il existe bien une méthode simple et efficace pour inclure les effets quantiques dans un modèle en potentiel de surface. Supposons que nous ayons à notre disposition une expression modélisant correctement l'élargissement du gap ΔE_g[1]. Au cours du chapitre 2, nous avons introduit une équation importante : l'équation implicite décrivant le potentiel de surface en tout régime de fonctionnement (cf. (2.8)). Cette équation peut se réécrire sous la forme suivante :

$$(V_g - V_{fb} - \phi_s)^2 = \gamma^2 \cdot \left\{ \phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t}\right) - 1 \right] + \phi_t \cdot \exp\left(\frac{-V_{ch} - \phi_B}{\phi_t}\right) \cdot \left[\exp\left(\frac{\phi_s}{\phi_t}\right) - 1 \right] \right\} \tag{4.43}$$

Utilisant alors le concept de concentration intrinsèque effective de porteurs, la prise en compte des effets quantiques est immédiate. Par commodité, nous en rappelons ici l'expression :

$$n_{i[\mathrm{qm}]} = n_i \cdot \exp\left(\frac{-\Delta E_g}{2 \cdot k \cdot T}\right) = n_i \cdot \exp\left(\frac{-\Delta E_g/q}{2 \cdot \phi_t}\right) \tag{4.44}$$

Grâce à cette correction apportée à la formulation classique de n_i, il est possible de modifier (4.43), dans le but d'y inclure les effets quantiques [4] :

$$(V_g - V_{fb} - \phi_s)^2 = \gamma^2 \cdot \left\{ \phi_s + \phi_t \cdot \left[\exp\left(\frac{-\Delta E_g/q}{2 \cdot \phi_t}\right) \cdot \exp\left(\frac{-\phi_s}{\phi_t}\right) - 1 \right] + \phi_t \cdot \exp\left(\frac{-V_{ch} - \phi_B}{\phi_t}\right) \cdot \left[\exp\left(\frac{-\Delta E_g/q}{2 \cdot \phi_t}\right) \cdot \exp\left(\frac{\phi_s}{\phi_t}\right) - 1 \right] \right\} \tag{4.45}$$

Cette méthode résulte en une implémentation précise des effets quantiques, et donne de bons résultats en comparaison aux simulations Schrödinger–Poisson auto-cohérentes [4]. Malheureusement elle nécessite la résolution itérative du potentiel de surface, ce qui la rend inadaptée à la simulation de circuits. Une solution envisageable serait d'adapter cette méthode

[1] Il faut naturellement que l'expression analytique choisie soit dépendante du champ électrique F_s. De plus, elle doit être à la fois valable pour décrire le confinement quantique des trous et des électrons, ce qui d'un point de vue physique strict n'est pas possible avec une unique expression de ΔE_g.

au cas d'un modèle analytique de potentiel de surface, mais cependant réaliser une telle implémentation n'est pas une chose évidente [31]. C'est pourtant cette méthode qui a été choisie pour incorporer les effets quantiques au sein du modèle compact SP [6,33]. Dans un premier temps, nous allons expliquer le principe de cette modélisation, puis dans une seconde partie nous décrirons l'approche employée dans le modèle compact MM11 [31,34].

4.5.3.1 Le modèle SP

SP est un modèle de potentiel de surface analytique, basé sur une approximation mathématique performante mais opaque (en termes de paramètres technologiques), de l'équation implicite décrivant ϕ_s [35]. L'inclusion des effets quantiques dans le modèle est réalisée en postulant que :

$$\phi_{s[\mathrm{qm}]} = \phi_s + \delta\phi_s \tag{4.46}$$

où $\delta\phi_s$ représente l'écart entre le potentiel de surface calculé de façon quantique $\phi_{s[\mathrm{qm}]}$ et classique ϕ_s.

Il faut prêter une attention toute particulière au fait que la quantité $\delta\phi_s$ définie dans (4.46) correspond à la différence de courbure de bandes *à tension de grille fixée*, et non pas à densité de charge d'inversion fixée, contrairement aux quantités $\Delta\phi_s$ données par (4.30) et (4.39)[1]. En conséquence, deux points fondamentaux doivent être bien compris :

1. L'équivalence $\delta\phi_s = \Delta E_g/q$ est physiquement incohérente au sens classique de la définition de l'élargissement du gap, à savoir le décalage entre le premier niveau d'énergie permis et le bas de la bande de conduction. Pour s'en convaincre, il suffit de constater qu'en régime d'inversion faible le potentiel de surface n'est pas affecté par les effets quantiques[2], puisque la charge de déplétion domine alors très nettement celle d'inversion ($|Q_b| \gg |Q_{inv}|$). En conséquence $\delta\phi_s \approx 0$ bien que $\Delta E_g/q \neq 0$.
2. L'équivalence $\Delta\phi_s = \Delta E_g/q$ — qui est en fait une définition — n'est pas utilisable dans le contexte d'un modèle formulé intégralement en potentiel de surface, où ϕ_s est défini comme une fonction de V_g.

---

[1]Dans un souci de clarté, nous adopterons les notations suivantes : $\delta\phi_s$ correspond à l'augmentation de ϕ_s à V_g fixée, et $\Delta\phi_s$ correspond à l'augmentation de ϕ_s à Q_{inv} fixée.

[2]D'un point de vue rigoureux, le potentiel de surface est quand même perturbé par les effets quantiques en inversion faible, mais sa variation est infime, de l'ordre de 0.01 %.

En présence d'effets quantiques, nous avons vu que le potentiel de surface est plus élevé à tension de grille fixée que dans un schéma classique (cf. Fig. 4.3, page 85). Pour comprendre ce phénomène, il faut se rappeler que le confinement quantique joue un rôle majeur sur les densités de porteurs (accumulation et inversion) ; prenons l'exemple des électrons de la couche d'inversion d'un transistor n-MOS, en régime d'inversion forte. À tension de grille fixée, la population d'électrons est plus faible que lors d'une description classique, en raison du décalage du premier niveau d'énergie permis du bas de la bande de conduction. La charge d'inversion $|Q_{inv}|$ est alors réduite, de même que la charge totale $|Q_{sc}|$ du semi-conducteur. Or, si V_g est fixée et $|Q_{sc}|$ plus faible, la différence d'énergie résultante est forcément encaissée par le potentiel de surface ϕ_s qui finalement est plus élevé. Il suffit d'appliquer le théorème de Gauss à l'interface Si–SiO_2 de la structure pour s'en convaincre :

$$Q_{sc} = -C_{ox} \cdot (V_g - V_{fb} - \phi_s) \tag{4.47}$$

Une stratégie mathématique assez complexe est utilisée dans SP pour obtenir une expression analytique de $\delta\phi_s$. Cette dernière repose sur le modèle d'élargissement du gap originellement proposé par VAN DORT [10], qui pour l'occasion, est réécrit de la façon suivante [6] :

$$\Delta E_g = k_Q \cdot \left(\frac{V_g - V_{fb} - \phi_s}{t_{ox} \cdot \sqrt{T}} \right)^{2/3} \tag{4.48}$$

où k_Q est un coefficient numérique [6]. ϕ_s représente le potentiel de surface qui est obtenu par :

$$(V_g - V_{fb} - \phi_s)^2 = \gamma^2 \cdot (f \cdot \phi_s - \phi_t + \phi_t \cdot \Delta) \tag{4.49}$$

où f est un paramètre empirique, et Δ correspond à :

$$\Delta = \exp\left(\frac{\phi_s - \phi_B - V_{ch} - \Delta E_g}{\phi_t} \right) \tag{4.50}$$

L'équation (4.49) est donc assez proche de la définition implicite du potentiel de surface précédemment définie en (4.45). On peut cependant y distinguer quelques différences notoires. La plus importante est la suppression du facteur $1/2$ devant le terme ΔE_g par rapport à l'expression (4.45) proposée par RIOS et ARORA [4]. Aucune précision n'est fournie à ce sujet par GILDENBLAT et CHEN, deux des principaux auteurs de SP. La deuxième modification est d'avoir supposé $\phi_s \gg \phi_t$, ce qui implique une modélisation du potentiel de surface valable uniquement en régimes de déplétion et d'inversion. Enfin, l'ajout du terme f dans (4.49) est

assez ambigu.

À partir de ce point, la méthode développée pour obtenir $\delta\phi_s$ est une méthode purement mathématique, que nous n'exposerons pas ici, son développement complet étant détaillé dans [6]. Nous allons juste en préciser les conclusions. L'augmentation $\delta\phi_s$ du potentiel de surface est finalement donnée — suite à de nombreuses approximations — par :

$$\delta\phi_s \simeq q \cdot (p - q/p)^{-1} \tag{4.51}$$

où

$$p = 2 \cdot w + \gamma^2 \cdot \left[f + a \cdot \Delta^{(0)}\right] \quad \text{et} \quad q = -\gamma^2 \cdot \Delta^{(0)} \cdot \Delta E_g^{(0)} \tag{4.52}$$

avec

$$w = V_g - V_{fb} - \phi_s^{(0)} \quad \text{et} \quad a = 1 + \frac{2 \cdot \Delta E_g^{(0)}}{3 \cdot w} \tag{4.53}$$

où $\phi_s^{(0)}$ et $\Delta^{(0)}$ sont les valeurs de ϕ_s et Δ correspondant à une évaluation classique du potentiel de surface, c'est-à-dire en considérant que $\Delta E_g = 0$. Enfin, $\Delta E_g^{(0)}$ représente la valeur de ΔE_g calculée pour $\phi_s = \phi_s^{(0)}$.

En résumé, nous pouvons dire que cette approche est relativement physique sur le fond, et très mathématique sur la forme. Le terme « relativement » employé ici, est peut être un peu excessif, néanmoins il se justifie quand même par les commentaires suivants. Tout d'abord l'approximation choisie pour modéliser l'élargissement du gap est discutable. Elle repose en effet sur le concept du puits de potentiel triangulaire, concept qui, en inversion forte, mène à une estimation non optimale (voire erronée) des effets quantiques [14,20,21,36,37]. Plus important encore est la transformation du terme $\Delta E_g/2$ en ΔE_g dans l'expression implicite du potentiel de surface (cf. (4.49)–(4.50)). Une troisième remarque est qu'un facteur nommé f est utilisé en tant que paramètre d'ajustement dans l'équation implicite de ϕ_s. Enfin, les nombreuses approximations mises en oeuvre pour obtenir la formulation explicite finale de $\delta\phi_s$ laissent un peu perplexe quant à son réel caractère physique.

Pour en revenir à la forme du modèle, il est indiscutable qu'elle soit très mathématique. Cependant l'aspect mathématique externe de SP n'est pas un défaut en soi puisque les concepteurs de circuits utilisent en général les simulateurs comme des boîtes noires, et donc se moquent éperdument de ce genre de considération. Cette complexité mathématique est par contre un réel frein quant à l'utilisation « manuelle » du modèle. Il est aussi intéressant de noter

que l'expression finale du décalage quantique $\delta\phi_s$ est définie comme une fonction de ϕ_s. Son évaluation requiert donc au préalable la connaissance du potentiel de surface, ce qui forcément entraîne un temps de calcul supplémentaire.

4.5.3.2 Le modèle MM11

L'implémentation des effets quantiques dans MM11 se limite simplement à la définition d'une épaisseur d'oxyde équivalente t_{oxeff}. Contrairement au cas des modèles formulés en tension de seuil (BSIM3v3 et 4), cette approche est complètement envisageable dans le contexte d'un modèle en potentiel de surface, en raison de la non-utilisation de la grandeur 'tension de seuil'. Il n'y a pas à ce jour de publication détaillée sur la prise en compte des effets quantiques dans MM11. Ce sujet a juste été rapidement abordé dans un article de VAN LANGEVELDE [34], qui fournit les renseignements suivants : l'augmentation effective de l'épaisseur d'oxyde est définie selon la relation maintenant bien connue :

$$t_{oxeff} = t_{ox} + \frac{\epsilon_{ox}}{\epsilon_{si}} \cdot \Delta y \tag{4.54}$$

où Δy est le déplacement de la distribution d'électrons, donné par :

$$\Delta y \propto Q_b + Q_{inv}/3 \tag{4.55}$$

La forme de l'équation de Δy indique que l'approximation analytique choisie pour modéliser l'élargissement du gap découle de la méthode variationnelle.

Finalement, nous avons quand même réussi à obtenir des informations supplémentaires sur les méthodes employées dans MM11 [38]. En particulier, l'expression de l'élargissement du gap est de la forme :

$$\Delta E_g = \frac{3}{2} \cdot \left(\frac{3 \cdot q \cdot \hbar}{2 \cdot \epsilon_{si} \cdot \sqrt{m^*}} \right)^{2/3} \cdot \frac{Q_b + \frac{55}{96} \cdot Q_{inv}}{(Q_b + \frac{11}{32} \cdot Q_{inv})^{1/3}} \tag{4.56}$$

ce qui moyennant quelques simplifications devient :

$$\Delta E_g \simeq \frac{3}{2} \cdot \left(\frac{3 \cdot q \cdot \hbar \cdot F_{eff}}{2 \cdot \epsilon_{si}} \right)^{2/3} = q \cdot \frac{3}{5} \cdot QM \cdot (\epsilon_{si} \cdot F_{eff})^{2/3} \tag{4.57}$$

où QM est une constante physique ($QM_n = 5.952$ et $QM_p = 7.459$ pour les électrons et les trous, respectivement), et F_{eff} le champ électrique effectif perpendiculaire à l'interface Si–SiO_2,

donné par :

$$F_{eff} = -\frac{Q_b + Q_{inv}/3}{\epsilon_{si}} \tag{4.58}$$

Cette description de ΔE_g correspond bien aux résultats de l'approche variationnelle originellement présentée par STERN [20]. L'étape suivante consiste alors simplement à intégrer (4.57) dans l'équation du déplacement du centroïde d'inversion Δy, mais fait étrange, cette dernière est modélisée dans le cadre des résultats obtenus avec l'approximation du puits de potentiel triangulaire ! Cela est d'autant plus surprenant que Δy s'exprime aussi très simplement avec une approche variationnelle, une fois ΔE_g connu [1] ; ce point particulier est donc en quelque sorte un non-sens de cette modélisation.

L'expression prenant en compte la modification de forme de la distribution de porteurs est alors écrite comme :

$$\Delta y \simeq \frac{2}{3} \cdot \frac{\Delta E_g}{q \cdot F_{eff}} \tag{4.59}$$

Et finalement, l'augmentation effective de l'épaisseur d'oxyde Δt_{ox}[1] s'obtient en combinant (4.59) et (4.54) :

$$\frac{\Delta t_{ox}}{t_{ox}} = \frac{2}{3} \cdot \frac{\Delta E_g}{q} \cdot \frac{C_{ox}}{\epsilon_{si} \cdot F_{eff}} = QM_{t_{ox}} \cdot \left[\frac{C_{ox}}{\epsilon_{si} \cdot F_{eff}}\right]^{1/3} \tag{4.60}$$

où $QM_{t_{ox}}$ est égal à $(2/5 \cdot QM \cdot C_{ox}^{2/3})$.

En conclusion, les effets quantiques sont inclus dans MM11 par une approche limitée à la seule définition d'une capacité d'oxyde effective (i.e. d'une épaisseur d'oxyde effective), résultant en un modèle compréhensible et relativement intuitif.

En particulier, ce concept de capacité d'oxyde effective est utilisé pour la description des différentes charges de la structure, ce qui lors d'une modélisation en feuille de charge affecte de façon implicite les valeurs du courant de drain et des capacités (transcapacités, pour être précis), entre autres. Signalons aussi dans un but comparatif, que similairement au modèle SP, la prise en compte des effets quantiques dans MM11 nécessite au préalable l'évaluation du potentiel de surface, puisque la capacité d'oxyde effective est une fonction des charges qui elles-mêmes dépendent du potentiel de surface.

[1] $\Delta t_{ox} = t_{oxeff} - t_{ox}$.

4.5.4 Approche alternative : le modèle EKV

Le modèle EKV, dans sa version 3.0 — en voie de finalisation —, repose sur le principe de la linéarisation de la charge d'inversion par rapport au potentiel de surface, comme cela a été expliqué au § 2.4 du chapitre 2 [27,30,39]. D'un point de vue général « philosophie des modèles compacts », on pourrait dire que EKV représente un compromis entre la simplicité (théorique) des modèles en tension de seuil, et la précision inhérente aux modèles en potentiel de surface.

Les effets quantiques sont pris en compte au niveau du calcul des charges — à l'instar de MM11 —, ce qui est effectivement une approche pragmatique et cohérente. Dans [27], il est montré qu'ils affectent trois caractéristiques fondamentales du TMOS, à savoir le facteur de substrat γ[1], la tension de bandes plates V_{fb} et la capacité d'oxyde de grille C_{ox}. Dans la suite de ce paragraphe, lorsque les effets quantiques seront pris en compte, ces trois grandeurs seront affectées d'une étoile, par exemple : $\gamma \rightarrow \gamma^*$.

Comme dans le cas de tous les autres modèles, EKV utilise le phénomène d'élargissement du gap comme base fondamentale de sa modélisation. L'expression choisie correspond à celle calculée dans l'approximation du puits de potentiel triangulaire (cf. (4.40)). Elle est réécrite comme :

$$\Delta E_g \simeq 3.53 \cdot (-Q_b - Q_{inv})^{2/3} \tag{4.61}$$

À ce stade une nouvelle approximation est faite ; elle consiste à développer ΔE_g en une série de Taylor du second ordre, telle que :

$$\Delta E_g \simeq 3.53 \cdot \left[Q_{T0}^{2/3} + \frac{2}{3} \cdot Q_{T0}^{-1/3} \cdot (Q_T - Q_{T0}) - \frac{1}{9} \cdot Q_{T0}^{-4/3} \cdot (Q_T - Q_{T0})^2 \right] \tag{4.62}$$

où Q_T correspond à la charge totale du semi-conducteur (habituellement notée Q_{sc} dans ce manuscrit) et Q_{T0} est une valeur arbitrairement fixée du point de linéarisation, déterminée par un ajustement avec des résultats expérimentaux [27].

Exprimée en termes de potentiel de surface, (4.62) peut se mettre sous la forme :

$$\Delta\phi_s \simeq a_0 + a_1 \cdot (Q_b + Q_{inv}) + a_2 \cdot (Q_b + Q_{inv})^2 \tag{4.63}$$

[1] Le facteur de la grille en polysilicium γ_p l'est aussi, mais dans un souci de simplicité, nous ne considérerons pas le phénomène de polydéplétion au cours de cette description des effets quantiques dans EKV.

où a_0, a_1 et a_2 sont trois paramètres fixés et indépendants de la technologie du transistor. Ils dépendent par contre de la valeur choisie pour Q_{T0}.

Ensuite, une modification de l'équation classique $(V_g - V_{fb} - \phi_s) = -Q_{sc}/C_{ox}$ permet de prendre en compte l'influence des effets quantiques en termes de capacité d'oxyde de grille équivalente et de tension de bandes plates équivalente :

$$\begin{aligned} V_g &= V_{fb} + \phi_s + \Delta\phi_s - \frac{Q_b + Q_{inv}}{C_{ox}} \\ &= V_{fb}^* + \phi_s - \frac{Q_b + Q_{inv}}{C_{ox}^*} \end{aligned} \tag{4.64}$$

avec

$$C_{ox}^* = \frac{C_{ox}}{1 - a_1 \cdot C_{ox}} \tag{4.65}$$

$$V_{fb}^* = V_{fb} + a_0 \tag{4.66}$$

Un troisième paramètre interne au modèle est aussi modifié : le facteur de substrat γ, qui devient :

$$\gamma^* = \gamma \cdot (1 - a_1 \cdot C_{ox}) \tag{4.67}$$

Précisons aussi que lorsque la polydéplétion est prise en compte, le facteur de grille γ_p est reformulé comme :

$$\left(\frac{1}{\gamma_p^*}\right)^2 = \frac{1}{(1 - a_1 \cdot C_{ox})^2} \cdot \left[\frac{1}{\gamma_p^2} + a_2 \cdot C_{ox}^2\right] \tag{4.68}$$

Grâce aux nouvelles définitions C_{ox}^*, V_{fb}^*, γ^* et γ_p^* , différentes expressions spécifiques au modèle EKV (par exemple le facteur de normalisation n_Q) sont alors reformulées, affectant ainsi le modèle dans son ensemble (charges, courant, etc.) [27]. Il est important de bien distinguer cette définition d'une nouvelle capacité d'oxyde C_{ox}^*, de la notion de capacité d'oxyde effective C_{oxeff}, car contrairement à cette dernière C_{ox}^* est indépendante des polarisations externes au TMOS. Enfin, signalons que la nouvelle définition de la densité de charge de déplétion est, en dépit de sa forme différente, exactement la même que celle formulée sans tenir compte des effets quantiques, ce qui n'est pas réaliste puisque $Q_b \propto \sqrt{\phi_s}$. Néanmoins, dans le contexte spécifique de la nouvelle formulation d'EKV, il apparaît que cette simplification n'engendre pas d'erreur significative [27]. Il faut bien être conscient que dans le cadre d'un modèle en potentiel de surface, cette approximation serait la cause d'erreurs importantes, en particulier sur le calcul de la densité de charge d'inversion.

En conclusion et pour illustrer le point sensible dont nous avons discuté précédemment lors de l'inclusion des effects quantiques au sein du modèle SP (cf. § 4.5.3.1), il est instructif de préciser qu'une telle modélisation des effets quantiques mènerait à des résultats erronés dans *le contexte d'un modèle en potentiel de surface* (donc pas celui d'EKV), en raison de l'utilisation du concept $\Delta\phi_s = \Delta E_g/q$. Pour mieux se représenter la situation, prenons un exemple concret. Imaginons que nous ayons à notre disposition un modèle de potentiel de surface, analytique mais non quantique. Si pour prendre en compte les effets quantiques, nous devions modifier la valeur de la tension de bandes plates par rapport à sa valeur standard, cela entraînerait un décalage de ϕ_s pour toutes les valeurs de V_g, i.e. de l'accumulation à l'inversion, ce qui n'aurait aucun sens.

4.5.5 Bilan et intérêt d'un nouveau modèle

Quel que soit le modèle considéré, un caractère dominant de la modélisation des effets quantiques est lié la définition d'un élargissement du gap du silicium. Différentes définitions sont adoptées selon les modèles : SP et EKV3.0 se basent sur une équation de ΔE_g obtenue dans l'approximation du puits de potentiel triangulaire, tandis que MM11 se base d'une part sur les résultats de l'approximation variationnelle pour la définition de ΔE_g, et d'autre part sur ceux du puits de potentiel triangulaire pour la définition du déplacement du centroïde d'inversion Δy. En résumé, il n'existe pas de consensus à ce propos entre ces trois modèles. Pourtant il est démontré que l'approche variationnelle donne de meilleurs résultats pour la modélisation de l'inversion, en particulier modérée et forte [1,20,21,40].

À cette étape de notre raisonnement, une question essentielle se pose : y-a-t'il un intérêt à proposer de nouvelles solutions pour modéliser les effets quantiques au sein des modèles compacts ? Notre réponse est bien entendu positive, pour plusieurs raisons. Avant d'entrer dans les détails, il convient déjà de reconnaître que BSIM4, EKV3.0, MM11 et SP sont tout à fait capables de simuler les principales caractéristiques *I–V* et *C–V* avec une précision acceptable. Cependant cela n'est pas un critère suffisant pour juger de la qualité intrinsèque d'un modèle, en particulier d'un modèle quantique. En effet, un ajustement astucieux de certains paramètres du modèle standard peut suffire à obtenir un accord correct entre les courbes expérimentales et simulées [27,38]. Par exemple, fournir au modèle une épaisseur d'oxyde différente de l'épaisseur physique peut rendre bien des services [32,34]. Mais une telle approche n'a alors plus aucun caractère prédictif, i.e. elle a perdu son sens physique.

Notre but est de réaliser un modèle de potentiel de surface purement analytique et quantique. Avec un tel modèle, la description (quantique) des charges et du courant se ferait de manière transparente, sans faire intervenir les notions de capacité d'oxyde effective, de coefficient de substrat équivalent, etc. En bref, aucun des paramètres du noyau du modèle classique ne devrait être modifié, seul le potentiel de surface devrait être traité de façon quantique. Cette approche étant sans rapport avec celles développées dans BSIM4 et EKV3.0, la justification du développement d'un nouveau modèle est donc dans ce cas inutile. Par contre, force est de constater que cette approche est en fait celle proposée par le modèle SP. Le problème avec SP est que la transformation quantique du modèle de base est extrêmement lourde, mathématiquement complexe, si bien que les expressions obtenues ne sont plus du tout parlantes en termes de paramètres technologiques et de tensions appliquées au dispositif. De plus, le conditionnement mathématique du modèle a nécessité de nombreuses approximations, en addition de celles déjà faites au départ (substitution de $\Delta E_g/2$ par ΔE_g, choix de l'approximation triangulaire), ce qui in fine occulte la base physique réelle du modèle de départ, i.e. celle du modèle SP classique (non quantique) [35].

En ce qui concerne MM11, l'inclusion des effets quantiques est réalisée d'une manière assez intuitive, en utilisant uniquement le concept de capacité d'oxyde effective, définie comme une grandeur dépendante des tensions appliquées au TMOS. MM11 étant un modèle en potentiel de surface, il est cependant critiquable que les effets quantiques soient seulement inclus sous la forme d'une correction de capacité d'oxyde de grille, car finalement bien que ce modèle soit formulé en termes de potentiel de surface, il ne permet pas d'obtenir la valeur réelle du potentiel de surface en présence d'effets quantiques. Il n'est en effet pas possible d'utiliser la définition de la capacité d'oxyde effective au sein même des expressions analytiques décrivant le potentiel de surface, ce dernier n'étant quasiment pas modifié par les effets quantiques en inversion faible.

Jusqu'à présent, nous ne nous sommes intéressés qu'à la description des effets quantiques en mode d'inversion. Or comme nous l'avons montré au début de ce chapitre, ces derniers affectent aussi le comportement électrique du transistor MOS en régime d'accumulation (cf. Fig. 4.3, page 85). Il est vrai que ce régime de fonctionnement intéresse moins le concepteur de circuits, cependant il est important en ce qui concerne l'extraction de paramètres, et en particulier l'extraction de l'épaisseur d'oxyde de grille. De plus, dans l'absolu, un modèle rigoureux se devrait de modéliser précisément les caractéristiques électriques sur toute la dynamique de fonctionnement du TMOS, de l'accumulation à l'inversion.

Dans la littérature[1], il y a peu, voire pas de publication sur la prise en compte des effets quantiques en accumulation. Les modèles SP et EKV3.0 tels que décrits précédemment ne sont valables qu'à partir du régime de déplétion (au-delà de la tension de bandes plates). La formulation de MM11 permet elle de décrire le confinement quantique dans tous les régimes de fonctionnement ; cependant l'expression choisie pour modéliser l'élargissement du gap découle d'une approximation valide en régime d'inversion, d'où des questions légitimes sur le bon escient de son usage en accumulation. En outre les valeurs des masses effectives étant différentes pour les électrons et les trous, il faudrait considérer séparément chaque type de porteurs en fonction du régime où se trouve le TMOS. Aucune explication sur ces points importants n'est malheureusement fournie dans la documentation de MM11 [31] ou dans les publications associées à ce modèle [34,41].

Ayant tiré parti de toutes ces observations, nous avons développé de nouveaux concepts permettant d'inclure de façon cohérente les effets quantiques au sein d'un nouveau modèle compact de TMOS.
Les principales caractéristiques de ce modèle sont les suivantes :

- Formulé intégralement en potentiel de surface,
- analytique et quantique,
- valable de l'accumulation à l'inversion,
- aucun paramètre additionnel en comparaison au modèle classique.

4.6 Formulation du nouveau modèle : déplétion et inversion

Outre les principales caractéristiques énoncées ci-dessus, plusieurs points spécifiques à notre modélisation doivent être précisés dès maintenant. Ce nouveau modèle est construit autour des approximations analytiques du potentiel de surface utilisées dans le modèle de Philips, MM11. Nous les avons cependant modifiées, voire simplifiées dans la plupart des cas, pour ne tenir compte que des phénomènes majeurs relatifs aux différents modes de fonctionnement du TMOS. En particulier, nous avons dissocié la prise en compte du phénomène de polydéplétion de l'expression du potentiel de surface. Contrairement à MM11, les effets quantiques sont inclus ici directement en termes de potentiel de surface, c'est-à-dire qu'à polarisation externe fixée, notre modèle permet d'obtenir la véritable valeur quantique du potentiel de surface. En conséquence, cette approche ne modifie aucun paramètre technologique, la valeur de la capacité

[1] Dans le contexte précis des modèles compacts de transistors MOS.

d'oxyde de grille reste donc égale à sa valeur réelle, déterminée comme une fonction des seuls paramètres géométriques du transistor MOS (comme la valeur physique de l'épaisseur d'oxyde de grille, par exemple). Aucune modification de la tension de bandes plates ou du coefficient de substrat n'est non plus nécessaire.

En résumé, les effets quantiques ne sont pas vus ici comme une simple correction à apporter à un certain nombre de paramètres de tel ou tel modèle classique, mais plutôt considérés comme un aspect intrinsèque du modèle. À ce titre, une redéfinition complète des équations du noyau du modèle classique est nécessaire. Cela n'est pas un problème insurmontable dans le cadre d'un modèle en potentiel de surface, puisqu'en réalité une seule équation requiert une modification : celle du potentiel de surface. Un aspect intéressant de cette approche est que le modèle ainsi reformulé est complètement transparent vis à vis de l'incorporation des effets quantiques ; pour son utilisateur, c'est un véritable clone du modèle équivalent classique.

4.6.1 Modélisation explicite quantique du potentiel de surface

Nous allons pour l'instant traiter le cas des régimes de déplétion et d'inversion, en l'absence d'effets quantiques. Après avoir obtenu une expression précise, continue et explicite du potentiel de surface valable dans ces deux régimes de fonctionnement, nous détaillerons le nouveau concept *d'approximation de l'inversion modérée*, dont le but est la modélisation des effets quantiques, directement en termes de potentiel de surface.

4.6.1.1 Modèle explicite classique

L'objectif premier est d'obtenir une équation du potentiel de surface exprimée comme une simple fonction des paramètres technologiques et des polarisations du TMOS. Comme point de départ, nous connaissons l'équation implicite relative au potentiel de surface ; écrivons la sous la forme suivante :

$$V_g - V_{fb} - \phi_s = \gamma \cdot \left\{ \phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t} \right) - 1 \right] + \phi_t \cdot \exp\left(\frac{-V_{ch} - \phi_B}{\phi_t} \right) \cdot \left[\exp\left(\frac{\phi_s}{\phi_t} \right) - 1 \right] \right\}^{1/2} \tag{4.69}$$

Cette équation est valable quel que soit le mode de fonctionnement du dispositif, de l'accumulation à l'inversion. Elle peut être simplifiée si l'on ne considère que les régimes usuels de fonctionnement (déplétion et inversion). En effet, si ϕ_s est positif, et en plus si $\phi_s \gg \phi_t$,

(4.69) peut être réduite à :

$$V_g - V_{fb} - \phi_s = \gamma \cdot \left[\phi_s + \phi_t \cdot \exp\left(\frac{\phi_s - V_{ch} - \phi_B}{\phi_t}\right)\right]^{1/2} \tag{4.70}$$

Cette nouvelle expression est certes plus simple, mais elle conserve toujours son caractère implicite. En conséquence de nouvelles approximations doivent être proposées.

Une méthode efficace est de séparer l'étude du potentiel de surface en inversion forte, de celle en inversion faible.

Commençons par la région d'inversion faible, où le potentiel de surface est défini comme $0 < \phi_s < \phi_B + V_{ch}$. Dans ce cas, le terme en exponentiel dans (4.70) est négligeable, et nous pouvons alors écrire :

$$\phi_{swi}(V_g) = \left[\sqrt{V_g - V_{fb} + \gamma^2/4} - \gamma/2\right]^2 \tag{4.71}$$

Il apparaît donc que le potentiel de surface en inversion faible est quasiment proportionnel à $V_g - V_{fb}$, comme le montre la Fig. 4.7.

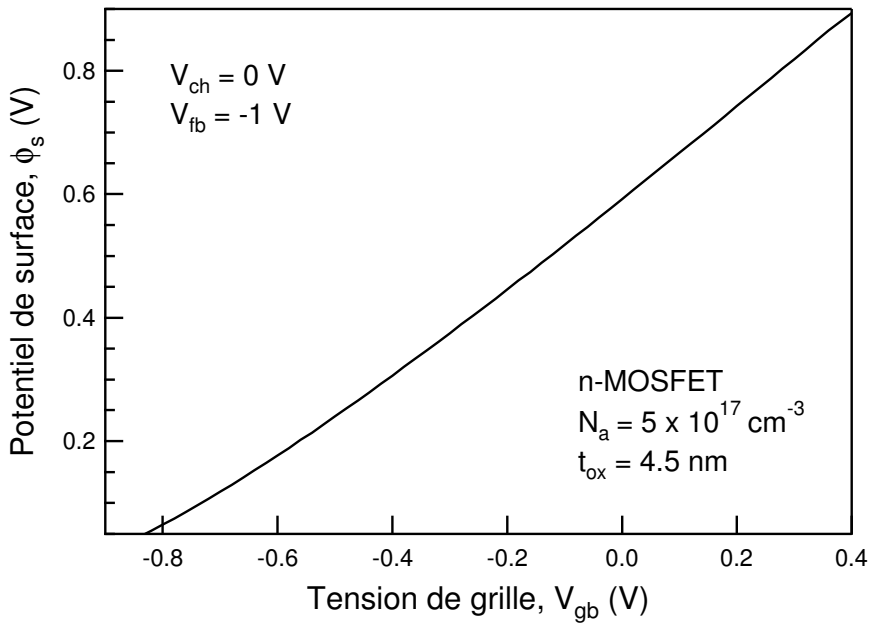


FIG. 4.7 : *Simulation numérique exacte du potentiel de surface ϕ_s en régimes de déplétion et d'inversion faible. La dépendance quasi-linéaire de ϕ_s avec V_g apparaît clairement.*

En régime d'inversion forte, le potentiel de surface est généralement considéré comme constant, indépendant de la tension de grille, et au premier ordre égal à $\phi_B + V_{ch}$, avec

$\phi_B = 2\,\phi_f$. Cette approximation mène cependant à d'importantes erreurs, comme l'illustre la Fig. 4.8, et c'est pourquoi très souvent un autre critère est utilisé. Il consiste à supposer que $\phi_s \simeq \phi_B + \alpha \cdot \phi_t + V_{ch}$ où $1 < \alpha < 6$, ce qui permet d'obtenir de meilleurs résultats à fortes polarisations de grille. Cependant, au niveau de la transition inversion faible / inversion forte, c'est-à-dire dans la région d'inversion modérée, les résultats demeurent incorrects.

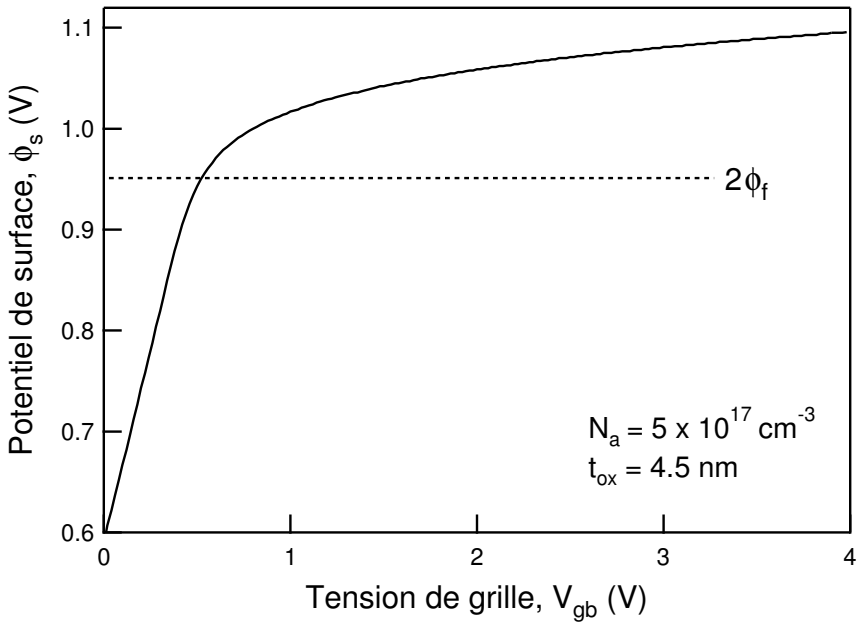


FIG. 4.8 : *Simulation numérique exacte du potentiel de surface ϕ_s en régime d'inversion forte. Le potentiel de surface n'est pas constant, il dépend de V_g, contrairement à ce qui est habituellement supposé. Le potentiel de quasi-Fermi V_{ch} est supposé nul.*

Une approche très intéressante pour modéliser le potentiel de surface en inversion forte est celle proposée par LANGEVELDE [41]. Une expression précise de ϕ_s peut être trouvée en supposant que le terme en exponentiel de (4.70) est dominant, et en gardant tous les autres termes de ϕ_s constants, à la valeur de $\phi_B + V_{ch}$. Dans ce cas, (4.70) est donnée par :

$$V_g - V_{fb} - (\phi_B + V_{ch}) = \gamma \cdot \left[\phi_B + V_{ch} + \phi_t \cdot \exp\left(\frac{\phi_s - V_{ch} - \phi_B}{\phi_t}\right)\right]^{1/2} \tag{4.72}$$

En isolant le terme de ϕ_s de l'exponentiel, nous obtenons une équation explicite du potentiel de surface en inversion forte :

$$\phi_{ssi} = \phi_B + V_{ch} + \phi_t \cdot \ln\left\{\frac{1}{\phi_t} \cdot \left[\left(\frac{V_g - V_{fb} - \phi_B - V_{ch}}{\gamma}\right)^2 - \phi_B - V_{ch} + \phi_t\right]\right\} \tag{4.73}$$

Dans le logarithme de (4.73), un terme ϕ_t à été ajouté de telle sorte que $\phi_{ssi} = \phi_{swi}$ lorsque la tension de grille est égale à $V_g = V_{fb} + \phi_B + \gamma\sqrt{\phi_B}$ (i.e. à la tension de seuil) ; son influence globale est cependant négligeable.

Bien que cette nouvelle expression du potentiel de surface soit nettement plus précise que l'approximation classique $\phi_{ssi} \simeq \phi_B$, elle donne cependant des résultats incorrects au niveau du seuil, c'est-à-dire dans la région d'inversion modérée [41]. C'est justement pour résoudre ce problème que l'approche proposée par LANGEVELDE est très utile. Selon cet auteur, cette erreur résulte de l'utilisation d'une valeur fixe de ϕ_B dans (4.73). Une simple observation du comportement du potentiel de surface en inversion forte a alors permis de développer une nouvelle stratégie de modélisation, basée sur le constat suivant : à forte polarisation de la grille, le potentiel de surface semble saturer à une valeur d'environ $4\,\phi_t$. En d'autres mots, il apparaît que le potentiel de surface varie approximativement de $\phi_B + V_{ch}$ au niveau du seuil, à $\phi_B + 4\,\phi_t + V_{ch}$ pour des tensions de grille élevées.

Une fonction empirique très simple peut alors être définie dans le but de modéliser cette variation de niveau du potentiel de surface :

$$\phi_B^* = \phi_B + V_{ch} + \frac{\phi_{swi} - \phi_B - V_{ch}}{\sqrt{1 + \left[\dfrac{\phi_{swi} - \phi_B - V_{ch}}{4 \cdot \phi_t}\right]^2}} \tag{4.74}$$

L'étape suivante consiste à remplacer les termes ϕ_B de la partie logarithmique de (4.73) :

$$\phi_{ssi} = \phi_B + V_{ch} + \phi_t \cdot \ln\left\{\frac{1}{\phi_t} \cdot \left[\left(\frac{V_g - V_{fb} - \phi_B^* - V_{ch}}{\gamma}\right)^2 - \phi_B^* - V_{ch} + \phi_t\right]\right\} \tag{4.75}$$

En remarquant que le terme quadratique du logarithme est dominant, il est possible de simplifier (4.75) en :

$$\phi_{ssi} = \phi_B + V_{ch} + \phi_t \cdot \ln\left\{\frac{1}{\phi_t} \cdot \left[\left(\frac{V_g - V_{fb} - \phi_B^* - V_{ch}}{\gamma}\right)^2 - \phi_B - V_{ch} + \phi_t\right]\right\} \tag{4.76}$$

ce qui en définitive résulte en une description extrêmement précise du potentiel de surface en inversion modérée et forte.

Une constatation intéressante est que si l'on remplace tous les termes $\phi_B + V_{ch}$ dans (4.76) par ϕ_{swi}, le potentiel de surface ϕ_{ssi} est alors égal à ϕ_{swi}, ce qui prouve bien la cohérence de l'approche proposée.

Finalement, une transition continue de l'inversion faible ($\phi_s = \phi_{swi}$) à l'inversion forte ($\phi_s = \phi_{ssi}$) est obtenue en remplaçant $\phi_B + V_{ch}$ dans (4.76) par une fonction qui change progressivement ϕ_{swi} en $\phi_B + V_{ch}$. Cette fonction, appelée f, est donnée par :

$$f(V_g, V_{ch}) = \frac{\phi_B + V_{ch}}{2} + \frac{\phi_{swi}}{2} - \frac{1}{2} \cdot \sqrt{(\phi_{swi} - \phi_B - V_{ch})^2 + 4 \cdot \delta^2} \tag{4.77}$$

où δ est une constante de lissage, fixée à 0.02.

L'expression analytique complète du potentiel de surface, valable du régime de déplétion à celui d'inversion forte, peut alors se mettre sous la forme :

$$\phi_s(V_g, V_{ch}) = f + \phi_t \cdot \ln \left\{ \left[\frac{1}{\gamma \cdot \sqrt{\phi_t}} \cdot \left(V_g - V_{fb} - f - \frac{\phi_{swi} - f}{\sqrt{1 + \left[\frac{\phi_{swi} - f}{4 \cdot \phi_t} \right]^2}} \right) \right]^2 - \frac{f}{\phi_t} + 1 \right\} \tag{4.78}$$

Les résultats obtenus avec (4.78) sont présentés à la Fig. 4.9. L'excellent accord entre la simulation numérique et le modèle analytique confirme donc la validité des différentes approximations mises en oeuvre lors du développement du modèle.

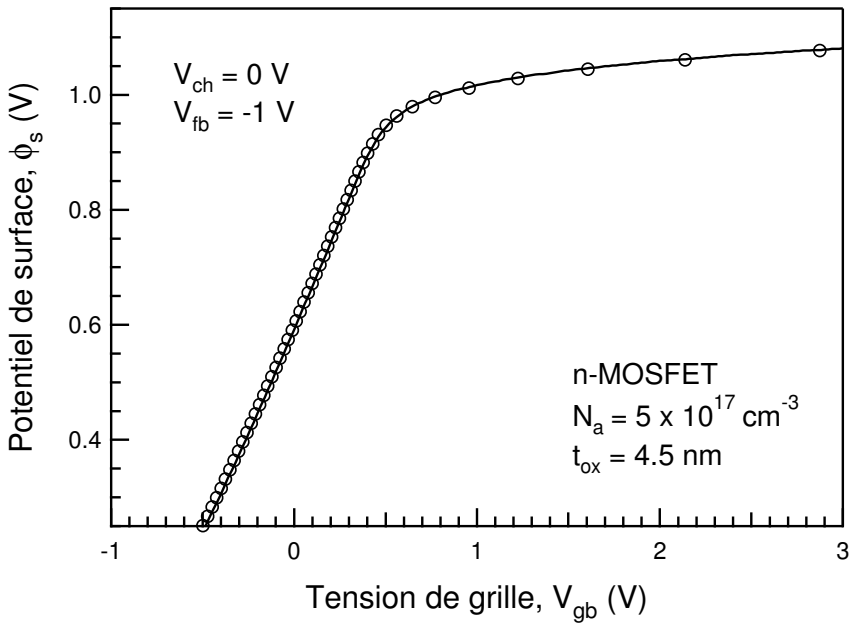


FIG. 4.9 : *Comparaison entre le potentiel de surface ϕ_s calculé numériquement (cercles) et analytiquement (ligne continue), dans un contexte classique i.e. sans effets quantiques.*

Nous avons maintenant à notre disposition une expression purement analytique décrivant le potentiel de surface en fonction des différentes polarisations externes au transistor MOS. Cette relation, définie par (4.78), permet une modélisation précise du potentiel de surface classique. Elle va être à la base du modèle quantique que nous allons expliciter au paragraphe suivant. Notre objectif est une transformation interne du modèle, c'est-à-dire que lorsque les effets quantiques seront pris en compte, le modèle final devrait idéalement être défini par une équation similaire à celle du modèle classique. En d'autres mots, nous souhaitons inclure les effets quantiques de façon complètement transparente vis à vis du modèle classique, en ne modifiant aucun paramètre du modèle originel.

4.6.1.2 Modèle explicite quantique : approximation de l'inversion modérée

Comme nous l'avons déjà mentionné à plusieurs reprises, dans le cadre de la théorie quantique, la bande de conduction ne peut plus être vue comme un continuum d'états, mais plutôt comme éclatée en sous-bandes discrètes. Cependant, il est largement reconnu qu'en régime d'inversion (et plus particulièrement en inversion forte), la contribution des porteurs appartenant à la sous-bande la plus basse, i.e. le niveau fondamentale E_0, est dominante [1,3,10,30,31]. C'est ce constat qui permet de définir une modélisation des effets quantiques basée sur un élargissement du gap du silicium. Ce point de départ, inhérent à toute modélisation des effets quantiques au sein des modèles compacts, sera aussi celui utilisé dans notre modèle.

Les résultats de la mécanique quantique nous apprennent que lorsque la contribution des porteurs du niveau fondamental est dominante, une approche variationnelle de la résolution des équations couplées Schrödinger–Poisson entraîne une estimation précise de l'énergie de la première sous-bande permise [20]. C'est pourquoi, à la différence de la plupart des modèles existants, nous avons choisi de modéliser l'élargissement du gap ΔE_g en utilisant les résultats de l'approximation variationnelle (cf. page 90). En effet, le plus souvent les modèles compacts se basent sur l'approximation du puits de potentiel triangulaire pour modéliser ΔE_g [5,7,10,25–27], bien que cette dernière ne soit pas précise que pour la modélisation des effets quantiques en inversion forte [1,20]. L'approximation du puits de potentiel triangulaire est en fait une bonne méthode quand il n'y a pas ou peu de charge dans la couche d'inversion, mais elle est incorrecte — d'un point de vue physique strict — lorsque la densité de charge d'inversion est comparable ou supérieure à celle de la couche de déplétion [20,21]. Elle permet néanmoins d'obtenir des résultats corrects dans les modèles formulés en charge [27], mais elle

engendre d'importantes erreurs dans le calcul du potentiel de surface [5]. En fait, elle permet aussi d'obtenir des résultats acceptables dans le cadre d'un modèle de potentiel de surface (cf. modèle SP), mais à condition d'utiliser certains termes de l'expression basique de ΔE_g en tant que paramètres d'ajustement (cf. (4.48)).

Dans le contexte de l'approximation variationnelle, la quantification de la couche d'inversion est calculée en utilisant une fonction d'onde (variationnelle) $\xi_b(y)$, qui s'annule à l'interface Si–SiO_2. Cette fonction d'onde est associée au niveau d'énergie le plus bas, et est donnée par [3,20] :

$$\xi_b(y) = \frac{b^{3/2}}{\sqrt{2}} \cdot y \cdot \exp\left[\frac{-b \cdot y}{2}\right] \tag{4.79}$$

où b est notre paramètre d'intérêt. Sa valeur théorique est choisie de telle façon qu'elle minimise l'énergie du premier niveau E_0 :

$$b(n_{inv}, \phi_s) = \left[\frac{12 \cdot m^* \cdot q^2}{\epsilon_{si} \cdot \hbar^2} \cdot \left(\frac{n_{inv}}{3} + n_{dep}(\phi_s)\right)\right]^{1/3} \tag{4.80}$$

où m^* est la masse effective longitudinale des électrons, égale à $m^* = 0.98\, m_0$ (m_0 est la masse de l'électron libre) puisque le silicium du substrat est orienté (100). Les quantités n_{inv} et n_{dep} représentent respectivement la densité d'électrons dans la couche d'inversion, et la densité de charge fixe dans la couche de déplétion.

Jusqu'à présent, les équations décrites par (4.79) et (4.80) sont exactement celles établies par l'approximation variationnelle. Elles permettent de définir en termes d'élargissement du gap (ΔE_g) l'augmentation du potentiel de surface ($\Delta\phi_s$) nécessaire à l'obtention d'une densité de charge d'inversion donnée. Le problème est que la modélisation que nous souhaitons réaliser est de la forme :

$$\phi_{s[\mathrm{qm}]}(V_g, V_{ch}) = \phi_s(V_g, V_{ch}) + \delta\phi_s(V_g, V_{ch}) \tag{4.81}$$

Comme nous l'avons expliqué au § 4.5.3.1 page 106, les grandeurs $\Delta\phi_s$ et $\delta\phi_s$, bien que corrélées, sont totalement différentes. $\delta\phi_s$ représente l'augmentation du potentiel de surface — due aux effets quantiques — pour une tension de grille donnée, et non pas pour une charge d'inversion donnée. En conséquence, si par définition nous avons bien $\Delta\phi_s = \Delta E_g/q$, nous ne pouvons pas écrire $\delta\phi_s = \Delta E_g/q$. C'est précisément dans le but de décrire l'incrément de potentiel de surface $\delta\phi_s$ que nous avons développé un nouveau concept : l'approximation de l'inversion modérée.

Dans la littérature, deux approximations des résultats de l'approche variationnelle sont généralement utilisées [8]. Ces approximations résultent en deux expressions différentes du paramètre b, l'une valable en inversion faible, et l'autre en inversion forte. L'approximation utilisée pour l'inversion faible est basée sur le fait que la contribution de la charge des porteurs libres est négligeable comparée à la charge totale ($Q_{inv}+Q_b \simeq Q_b$), tandis que celle utilisée pour l'inversion forte repose sur la relation linéaire liant la charge d'inversion à la tension de grille. Dans un article de CLERC, de telles approximations sont détaillées [8]. Moyennant l'utilisation d'une fonction de lissage (fortement dépendante de 2 paramètres empiriques), une modélisation analytique de la densité de porteurs libres, et donc de la charge d'inversion, est proposée.

Le problème de cette méthode est que la transition entre l'inversion faible et l'inversion forte est réalisée par une fonction de lissage n'ayant pas pour unique but d'améliorer cette transition (de la rendre douce), mais aussi de compenser la faiblesse de ce modèle au niveau de cette zone de transition, qui correspond à la région d'inversion modérée. Ainsi, la région d'inversion modérée est en fait modélisée par cette fonction de lissage[1] et non pas par les équations valides en inversion faible et forte. Il convient néanmoins de reconnaître qu'en dépit des incertitudes liées à cette fonction de lissage, ce modèle permet d'obtenir des résultats cohérents pour le calcul de la densité de charge d'inversion. Par contre, il n'en va pas de même pour le calcul du potentiel de surface, qui souffre d'un comportement non physique en inversion modérée.

La Fig. 4.10 illustre ce problème : il apparaît une sorte de bosse au niveau de la plage de tensions de grille correspondant à l'inversion modérée. Suite à de nombreux tests de ce modèle, nous avons constaté que l'importance de cette bosse est fortement dépendante de la valeur des paramètres empiriques utilisés dans la fonction de lissage incriminée. Un second problème est que de toute façon, le potentiel n'est pas correctement estimé pour toute la région au-delà du seuil, ce qui par exemple rend impossible toute estimation correcte du courant de drain dans le contexte d'un modèle en feuille de charge (approximation drift–diffusion).

L'étude de ce modèle est instructive car elle met en exergue deux points cruciaux. D'une part, elle illustre toute la difficulté à réaliser un modèle simple, compréhensible et permettant de prendre correctement en compte l'influence des effets quantiques sur le potentiel de surface. D'autre part, elle représente un avertissement quant à une utilisation (ir)raisonnée des fonctions de lissage.

[1] De plus, la dépendance de cette fonction de lissage aux deux paramètres empiriques dont elle est fonction est telle, que même une petite variation de l'un ou l'autre de ces paramètres perturbe inconsidérément les résultats obtenus avec ce modèle.

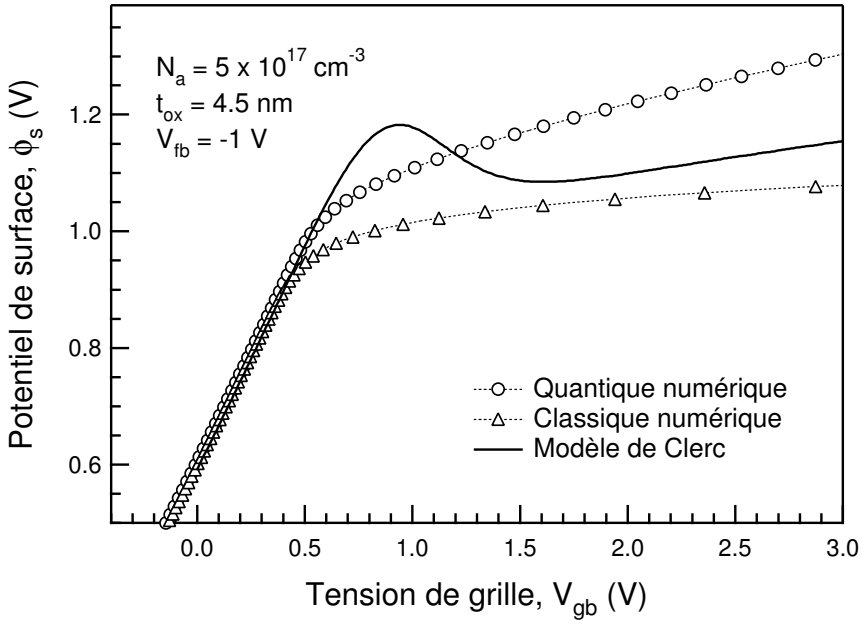


FIG. 4.10 : *Potentiel de surface simulé avec le modèle proposé par Clerc et al. [8]. Les paramètres empiriques utilisés sont ceux indiqués dans [8]. Les paramètres généraux tels que la tension de bandes plates ou la concentration intrinsèque de porteurs sont les mêmes que ceux utilisés dans les simulations numériques exactes.*

Il est toutefois inutile, voire même non justifié de faire le procès des fonctions de lissage ; les véritables questions à se poser étant plutôt :

- Dans quel cas est-il judicieux et physiquement réaliste d'utiliser une fonction de lissage ?
- Quel type de fonction mathématique est le plus adapté pour répondre à tel ou tel besoin ?

La réponse à cette seconde question nécessite d'être traitée au cas par cas. Nous pouvons cependant dire que les fonctions de type hyperbolique se comportent remarquablement bien, et point important, leurs dérivées aussi.

La première question est plus sujet à polémique. À notre sens, une fonction de lissage devrait avoir pour unique but d'améliorer la forme du modèle, et ne devrait en aucun cas en modifier le fond, i.e. son contenu physique. Prenons un exemple concret pour illustrer ce propos : soit une fonction $a(x)$, physiquement définie pour $x \geqslant x_0$ et telle que $a(x_0) = 0$. L'utilisation d'une fonction de lissage $\tilde{x} = f(x)$ peut tout à fait se justifier si son seul but est d'étendre la zone de définition de $a(x)$ à n'importe quelle valeur de x ; il suffit alors de supposer : $a(x < x_0) = 0$. Dès lors, remplacer $a(x)$ par $a(\tilde{x})$ dans un modèle donné aurait pour seul but d'assurer une bonne convergence des équations du modèle pour toute valeur de x.

Pour en revenir à la fonction de lissage utilisée dans le modèle de CLERC, sa fonction réelle est de modifier le fond du modèle. Son but est de réunir en une seule équation (soit c cette équation) deux équations asymptotiques a et b. Dans un tel cas de figure, l'emploi d'une fonction de lissage peut cependant être complètement légitime, mais à la condition express où en un point particulier x_1, les fonctions a et b seraient égales. Définir une fonction de lissage autour de ce point x_1 aurait donc pour seul rôle d'assurer la continuité du modèle, lors du passage *naturel* de $c = a$ à $c = b$. Par contre, quand il n'y a pas de point de transition x_1 tel que $a(x_1) = b(x_1)$, la fonction de lissage joue plutôt le rôle de contre-mesure vis à vis d'une modélisation discutable de l'une des grandeurs a ou b (cas du modèle présenté à la Fig. 4.10).

Cette parenthèse concernant l'utilisation des fonctions de lissage étant refermée, nous allons maintenant présenter les principes de notre modélisation des effets quantiques. Notre modèle est basé sur un nouveau concept, très simple, celui de l'approximation de l'inversion modérée. Ce concept est développé en partant du principe que l'augmentation $\delta\phi_s$ du potentiel due aux effets quantiques devrait idéalement être décrite comme une simple fonction des tensions appliquées au dispositif, i.e. $\delta\phi_s \propto V_g, V_{ch}$, où V_{ch} représente le potentiel de quasi-Fermi, variant de V_s à la source, à V_d au drain.

Les différentes étapes du développement de l'approximation de l'inversion modérée sont les suivantes. Tout d'abord, nous avons reformulé le paramètre b en négligeant le terme relatif à la densité de charge fixe présente dans la couche de déplétion du substrat :

$$b(V_g, V_{ch}) \simeq \left[\frac{12 \cdot m^* \cdot q^2}{\epsilon_{si} \cdot \hbar^2} \cdot \frac{n_{inv}(V_g, V_{ch})}{3}\right]^{1/3} \tag{4.82}$$

Il est clair que cette expression de b va naturellement sous-estimer l'influence des effets quantiques, en particulier au niveau de la région d'inversion modérée, puisque dans ce mode la densité de charge fixe est proche de celle des porteurs libres, c'est-à-dire $n_{inv} \simeq n_{dep}$.

À ce stade du raisonnement, une « fausse bonne » idée serait de réécrire b comme :

$$b(V_g, V_{ch}) \simeq \left[\frac{12 \cdot m^* \cdot q^2}{\epsilon_{si} \cdot \hbar^2} \cdot \left(\frac{n_{inv}}{3} + n_{inv}\right)\right]^{1/3} = \left[\frac{12 \cdot m^* \cdot q^2}{\epsilon_{si} \cdot \hbar^2} \cdot \frac{4 \cdot n_{inv}(V_g, V_{ch})}{3}\right]^{1/3} \tag{4.83}$$

puisqu'une telle expression mènerait alors à une surestimation de l'importance des effets quantiques, tout particulièrement en inversion forte puisque dans cette région la densité de porteurs libres est au premier ordre définie comme une fonction linéaire de V_g.

C'est précisément à ce moment qu'intervient le concept d'approximation de l'inversion modérée. Il consiste en fait à corriger l'équation (4.82) de façon à ne pas négliger la densité de charge de déplétion au niveau du seuil, sans pour autant la surestimer en inversion forte, à l'image de (4.83). Nous avons réalisé ce compromis en définissant une *densité de porteurs équivalente* n_{all}, prenant simultanément en compte la densité de porteurs libres dans la couche d'inversion et la densité de charge fixe dans la couche de déplétion. La densité de porteurs équivalente converge doucement vers zéro quand la tension de grille est inférieure à $V_{to} + V_{ch}$, suite à l'utilisation d'une fonction de lissage hyperbolique (cf. (4.86)).

Finalement cette approche résulte en une nouvelle formulation du paramètre b, valide dans toutes les régions de fonctionnement, et particulièrement précise en inversion modérée, définie par :

$$b(V_g, V_{ch}) \simeq \left[\frac{12 \cdot m^* \cdot q^2}{\epsilon_{si} \cdot \hbar^2} \cdot \frac{n_{all}(V_g{}', V_{ch})}{3}\right]^{1/3} \tag{4.84}$$

avec

$$n_{all}(V_g{}', V_{ch}) = \frac{2 \cdot C_{ox}}{q} \cdot (V_g{}' - V_{to} - V_{ch}) \tag{4.85}$$

et

$$V_g{}' = \frac{1}{2} \cdot \left[V_g + \sqrt{(V_g - V_{to} - V_{ch})^2 + 4 \cdot \varepsilon^2} + \sqrt{(V_{to} + V_{ch})^2 + 4 \cdot \varepsilon^2}\right] \tag{4.86}$$

où ε est une constante de lissage (fixée arbitrairement à 0.15) et V_{to} la tension de seuil canal-long classique donnée par $V_{fb} + \phi_B + \gamma\sqrt{\phi_B}$.

Une remarque importante concerne le choix de la valeur de ε :

- D'une part, ce choix est définitif,
- d'autre part, il est complètement indépendant de toute considération technologique.

Ainsi, ε ne sera en aucun cas utilisé comme un paramètre d'ajustement dans notre modèle.

Grâce à cette nouvelle formulation de b, nous pouvons décrire avec une grande précision le déplacement quantique de la bande conduction. Ce déplacement quantique s'exprime en fait comme un pseudo élargissement du gap :

$$E_w(V_g, V_{ch}) = \frac{3}{8}\frac{\hbar^2}{m^*} \cdot b(V_g, V_{ch})^2 \tag{4.87}$$

L'utilisation du qualificatif « pseudo » dans la phrase précédente ne résulte pas d'une simple formule de style, elle nous permet de différencier E_w de l'élargissement « classique » du gap,

noté ΔE_g et égal à $E_0 - E_c$. La grandeur E_w représente quant à elle la différence de courbure de bandes — pour une polarisation externe donnée V_g, V_{ch} — entre les cas classique et quantique. Il devrait bien entendu être noté que E_w est une fonction de ΔE_g ; ces deux grandeurs étant en fait étroitement liées par l'expression suivante :

$$E_w(V_g, V_{ch}) = \Delta E_g(b(n_{all})) \tag{4.88}$$

Finalement, à partir de (4.87), le pseudo élargissement du gap E_w est exprimé en termes d'incrément de potentiel de surface par la fonction $\delta\phi_s$:

$$\delta\phi_s(V_g, V_{ch}) = E_w(V_g, V_{ch})/q \tag{4.89}$$

Cette fonction $\delta\phi_s$ offre donc une relation explicite entre l'augmentation du potentiel de surface due aux effets quantiques et les différentes tensions appliquées au dispositif ($V_g, V_{ch} = V_s$ ou V_d). Ainsi, grâce à cette méthode, prendre en compte quantitativement les effets quantiques ne nécessite quasiment aucun temps de calcul additionnel par rapport à un modèle classique[1].

L'étape suivante du développement de notre modèle est capitale : il s'agit d'incorporer l'approximation de l'inversion modérée (i.e. $\delta\phi_s$) au sein du modèle classique analytique de potentiel de surface précédemment explicité (cf. § 4.6.1.1). Dans le but de préserver le caractère physique du modèle classique et de l'approximation de l'inversion modérée, nous ne souhaitons pas utiliser de fonction de lissage supplémentaire. Rappelons que le potentiel de surface quantique est obtenu dans notre approche par une relation de la forme :

$$\phi_{s[\mathrm{qm}]} = \phi_s + \delta\phi_s \tag{4.90}$$

Le simple fait d'additionner les expressions ϕ_s et $\delta\phi_s$ ensemble est cependant une approche trop simpliste — dans le cadre particulier de ce modèle — qui ne permet pas d'obtenir le meilleur modèle possible. Une telle méthode autorise malgré tout une description correcte de ϕ_s, mais pas des ses dérivées, ce qui est un problème classique en modélisation compacte. En fait, l'addition de deux fonctions sans utilisation de fonction de lissage entraîne souvent des problèmes de discontinuité, ou du moins un mauvais comportement des dérivées de la fonction résultante.

[1] Sous Mathcad Professional®, aucune différence notable de temps de calcul n'a pu être mesurée lorsque le terme $\delta\phi_s$ est pris en compte dans le calcul de ϕ_s (PC avec un processeur PII à 450 MHz).

Pour ces différentes raisons, et en accord avec notre objectif original, nous avons transformé le modèle classique de l'intérieur. Une simple modification de la fonction f (cf. (4.77)) du modèle classique est nécessaire. Sa réécriture quantique mène à la relation suivante :

$$f_{[\mathrm{qm}]}(V_g, V_{ch}) = \frac{\phi_B + V_{ch} + \delta\phi_s(V_g, V_{ch})}{2} + \frac{\phi_{swi}(V_g)}{2} - \frac{1}{2} \cdot \sqrt{[\phi_{swi}(V_g) - \phi_B - V_{ch} - \delta\phi_s(V_g, V_{ch})]^2 + 4 \cdot \delta^2} \tag{4.91}$$

De par cette simple reformulation de f, nous avons alors accès à une nouvelle expression du potentiel de surface, qui est à la fois quantique, analytique, et totalement dépendante de toutes les polarisations appliquées au TMOS.

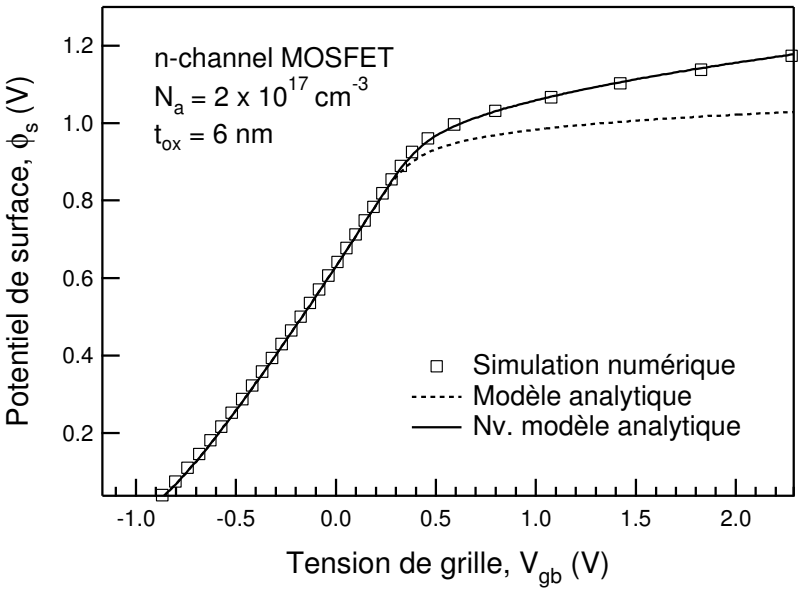
Cette expression, valable du régime de déplétion à celui d'inversion forte, est finalement donnée par :

$$\phi_{s[\mathrm{qm}]}(V_g, V_{ch}) = \phi_s(V_g, V_{ch}, f_{[\mathrm{qm}]}(V_g, V_{ch})) \tag{4.92}$$

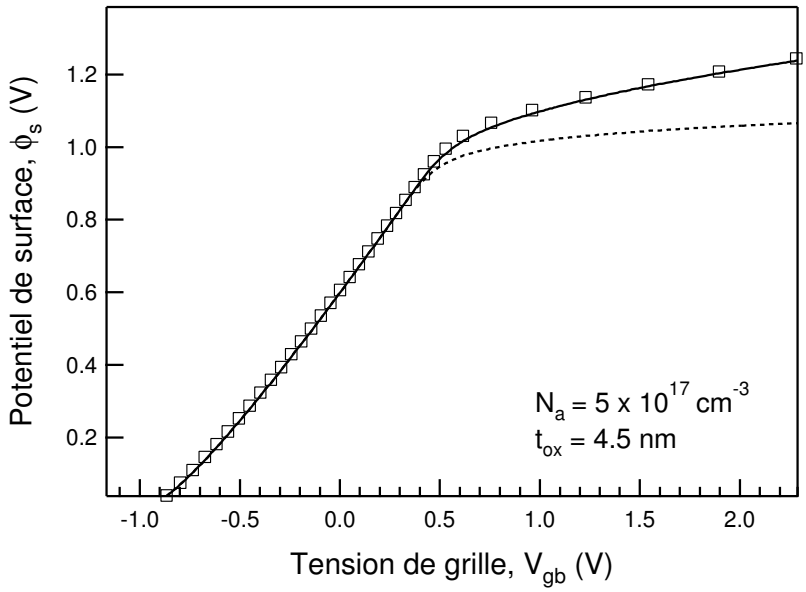
ce qui est équivalent à :

$$\phi_{s[\mathrm{qm}]}(V_g, V_{ch}) = f_{[\mathrm{qm}]} + \phi_t \cdot \ln\left\{\left[\frac{1}{\gamma \cdot \sqrt{\phi_t}} \cdot \left(V_g - V_{fb} - f_{[\mathrm{qm}]} - \frac{\phi_{swi} - f_{[\mathrm{qm}]}}{\sqrt{1 + \left[\frac{\phi_{swi} - f_{[\mathrm{qm}]}}{4 \cdot \phi_t}\right]^2}}\right)\right]^2 - \frac{f_{[\mathrm{qm}]}}{\phi_t} + 1\right\} \tag{4.93}$$

Un premier ensemble de résultats de simulation est présenté aux Figs. 4.11–4.12. Une comparaison entre les résultats obtenus avec notre nouveau modèle et ceux obtenus par une résolution numérique exacte des équations couplées de Schrödinger et de Poisson est montrée. Nous avons aussi simulé le potentiel de surface avec le modèle analytique classique, dans le but de mettre en évidence son inaptitude à décrire correctement ϕ_s en présence d'effets quantiques.

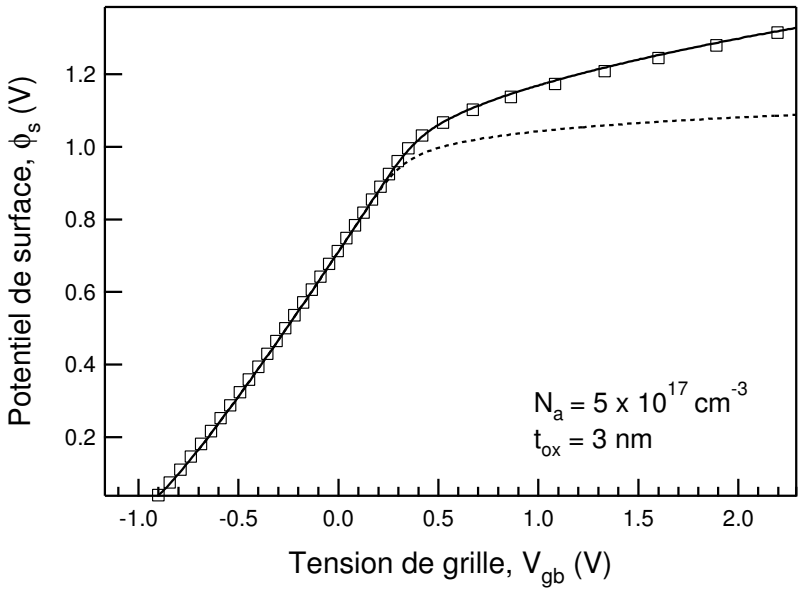


(a) Technologie CMOS 0.5–0.35 μm.

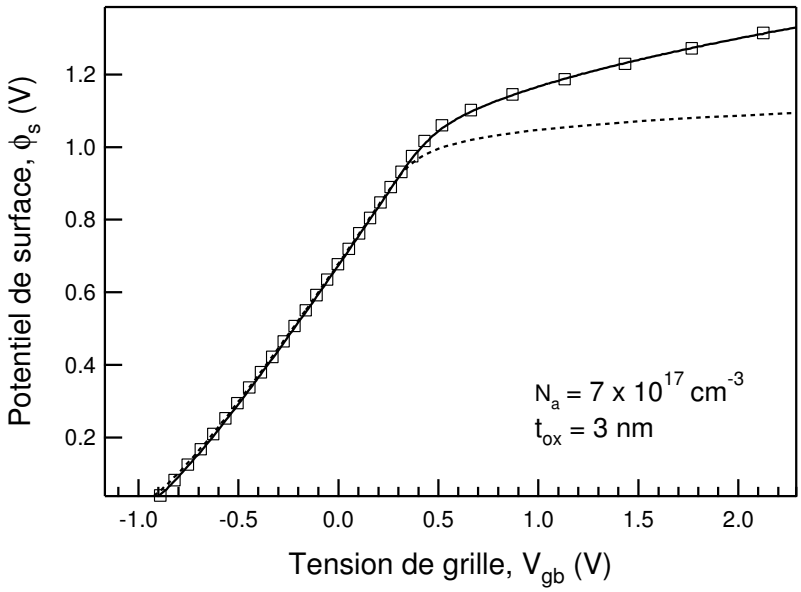


(b) Technologie CMOS 0.35–0.25 μm.

FIG. 4.11 : *Potentiel de surface ϕ_s en fonction de la tension de grille V_{gb}. Les symboles représentent les valeurs obtenues par une résolution auto-cohérente des équations Schrödinger–Poisson. Le trait plein et les pointillés correspondent aux résultats obtenus avec le nouveau modèle, cf. (4.93), et le modèle classique, cf. (4.78), respectivement.*



(a) Technologie CMOS 0.18 μm : $N_a = 5 \times 10^{17}$ cm^{-3}.



(b) Technologie CMOS 0.18 μm : $N_a = 7 \times 10^{17}$ cm^{-3}.

FIG. 4.12 : *Potentiel de surface ϕ_s en fonction de la tension de grille V_{gb}. Les résultats sont présentés pour une épaisseur d'oxyde de grille fixée à $t_{ox} = 3$ nm, et pour différents dopages substrat.*

Ces quatre simulations ont toutes été effectuées en posant $V_{ch} = 0$. Ce choix résulte d'une considération purement matérielle, le simulateur numérique Schrödinger–Poisson utilisé ici comme moyen de validation permettant uniquement de simuler une capacité MOS. À ce propos, les calculs numériques auto-cohérents ont été réalisés avec le simulateur QMCV de Berkeley (Quantum-Mechanical-C-V simulator) [16]. Son principe de fonctionnement est le

suivant : il calcule les distributions d'électrons et de trous en inversion et accumulation, suite à une résolution auto-cohérente des équations de Schrödinger et de Poisson avec la distribution de Fermi-Dirac. Comme le démontre l'article de RICHTER [19], ce simulateur est en très bon accord avec d'autres simulateurs quantiques tels que NEMO [17] ou NCSU [18]. La véracité de cette dernière affirmation se limite cependant au domaine d'inversion, car en accumulation aucun simulateur numérique ne fournit des valeurs identiques [19].

Venons en maintenant à l'étude des résultats obtenus. Les différents ordres de grandeurs d'épaisseur d'oxyde t_{ox} et de dopage substrat N_a choisis lors de nos simulations sont représentatifs des valeurs réellement utilisées par les fondeurs. En termes de longueur de grille, on pourrait dire que la Fig. 4.11a correspond à une technologie CMOS 0.5–0.35 μm, la Fig. 4.11b à une technologie 0.35–0.25 μm et les Figs. 4.12a–4.12b à une technologie 0.18 μm. Une série de résultats correspondant à des cas plus critiques (nouvelles et futures générations de TMOS, 130 et 90 nm) sera montrée ultérieurement. À la lecture des Figs. 4.11–4.12, il apparaît que le nouveau modèle améliore significativement le modèle classique. L'augmentation du potentiel de surface due aux effets quantiques est correctement décrite par notre modèle, tandis que le modèle classique sous-estime très nettement la valeur du potentiel de surface en inversion modérée et en inversion forte.

Les résultats présentés aux Figs. 4.11–4.12, bien que confirmant la validité des concepts développés au cours de notre modélisation, ne permettent cependant pas d'observer le comportement du modèle en tant que fonction des tensions de source ou de drain, i.e. quand $V_{ch} \neq 0$. Dans ce but, la Fig. 4.13 présente une comparaison entre le potentiel de surface calculé au niveau du drain avec les différents modèles explicites, c'est-à-dire quantique et analytique. Ce graphique illustre le bon comportement du modèle explicite quantique quelle que soit la polarisation du drain. En particulier, il souligne que l'inclusion d'un terme non nul pour le potentiel de quasi-Fermi V_{ch} autorise toujours une modélisation correcte du potentiel de surface, et cela de la déplétion à l'inversion forte.

Notons aussi que l'étude du potentiel de surface est intéressante en ce sens qu'elle permet de comprendre qui de l'épaisseur d'oxyde de grille ou du dopage substrat influe le plus sur l'importance des effets quantiques en inversion forte (cf. Fig. 4.14). Contrairement à ce qui est souvent supposé, l'augmentation du dopage substrat ne joue quasiment aucun rôle sur les effets quantiques en inversion forte. Le niveau de dopage substrat est en fait lié aux effets quantiques

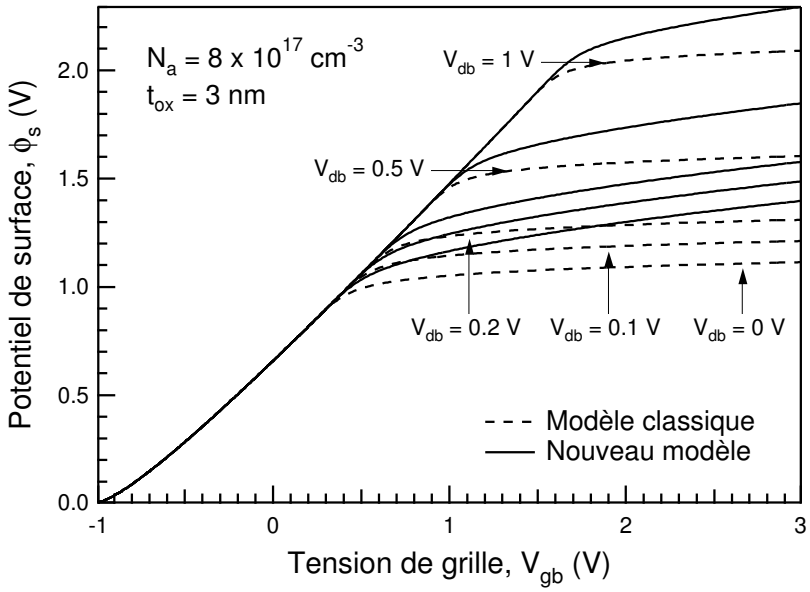


FIG. 4.13 : *Comparaison entre les modèles de potentiel de surface classique et quantique. Le potentiel de surface est calculé au niveau du drain en fonction de la tension de grille, pour différentes tensions de drain V_{db} allant de 0 à 1 V.*

en inversion faible, où il est la principale cause du décalage de la tension de seuil, comme nous l'avons vu au début de ce chapitre. A contrario, la diminution drastique de l'épaisseur d'oxyde de grille joue un rôle majeur dans l'augmentation quantique du potentiel de surface en inversion forte, et modifie donc considérablement la valeur du courant de drain et des capacités dans ce régime de fonctionnement. La Fig. 4.14 confirme sans ambiguïté ces propos, les résultats présentés provenant d'une résolution exacte des équations couplées Schrödinger–Poisson.

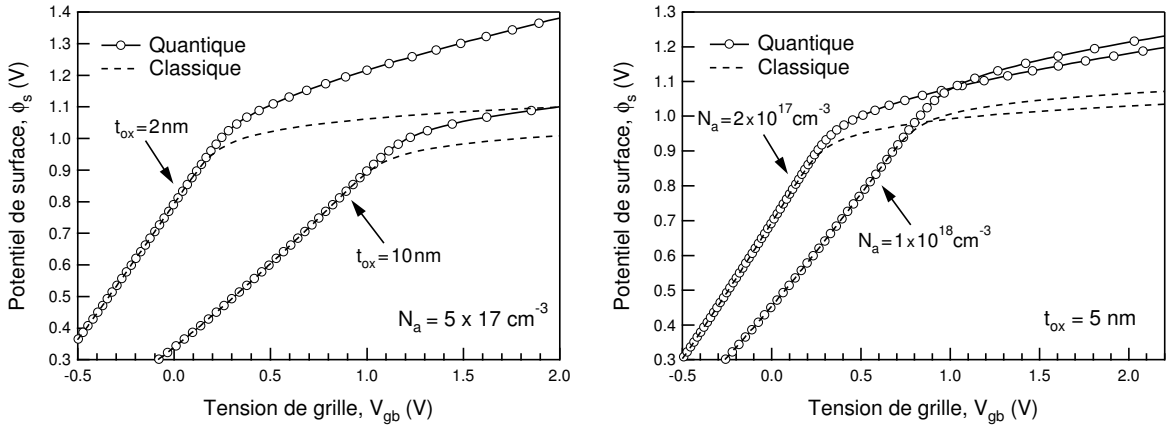


FIG. 4.14 : *Influence respective du dopage substrat et de l'épaisseur d'oxyde de grille sur le potentiel de surface. Graphique de gauche : N_a fixé, t_{ox} variant d'un facteur 5. Graphique de droite : t_{ox} fixée, N_a variant d'un facteur 5.*

Finalement, pour tester l'intérêt de notre modèle en termes de prédictivité, nous avons étudié l'impact d'une réduction d'échelle agressive du TMOS, sur le potentiel de surface. Une réduction d'échelle poussée à l'extrême correspond en fait à une forte diminution de l'épaisseur d'oxyde de grille t_{ox}, en dessous de la barre des 2 nm, et à une augmentation du dopage substrat au-delà de 10^{18} cm^{-3}. La Fig. 4.15 présente les résultats de telles simulations. Il apparaît que notre modèle quantique analytique est apte à la description des futurs dispositifs MOSFET, l'accord entre les résultats analytique et numérique étant très bon.

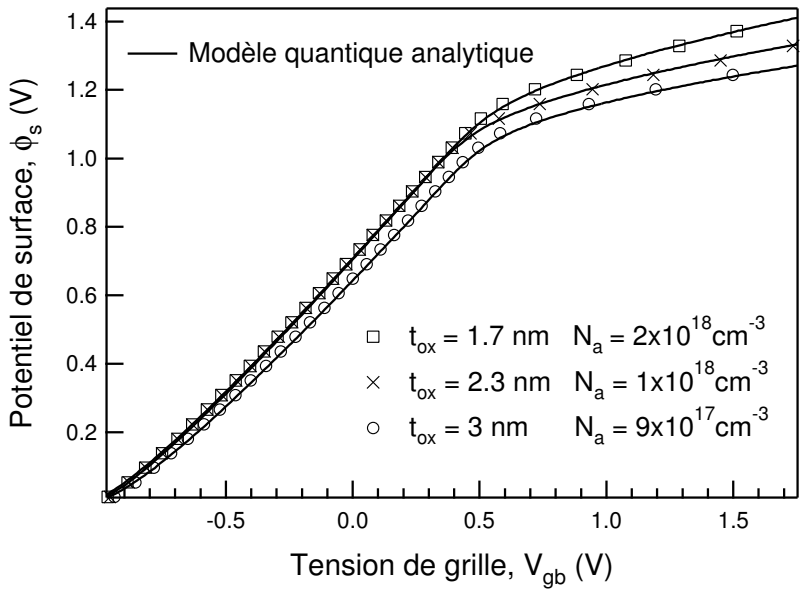


FIG. 4.15 : *Impact d'un scaling agressif de l'épaisseur d'oxyde de grille et du dopage substrat sur le potentiel de surface ϕ_s. Le potentiel de quasi-Fermi est mis à zéro. Les symboles représentent les résultats numériques obtenus par résolution auto-cohérente des équations de Schrödinger et de Poisson.*

4.6.2 Modélisation analytique du phénomène de polydéplétion

Une nouvelle modélisation du potentiel de surface, valable de la déplétion à l'inversion forte vient d'être explicitée et validée. La validation du modèle n'est cependant que partielle à ce stade, puisqu'aucune comparaison avec des résultats expérimentaux (*I–V*, *C–V*) n'a été présentée. Réaliser une telle comparaison n'est pas possible pour l'instant, car nous n'avons pas encore considéré dans notre modèle un phénomène majeur au sein des TMOS submicroniques : l'effet de polydéplétion. Dans ce paragraphe, nous allons décrire comment prendre en compte de manière physique et analytique le phénomène de polydéplétion.

L'utilisation d'une grille en polysilicium résulte en la formation d'une couche de déplétion à l'interface polySi–SiO_2, lorsque le TMOS est en régime d'inversion ($V_g > V_{fb}$). Une chute de potentiel ϕ_p s'établit au travers de cette couche de déplétion, réduisant ainsi la valeur effective de la tension de grille appliquée au TMOS. La Fig. 4.16 schématise cette chute de potentiel à l'intérieur de la grille. En conséquence tout le comportement électrique du TMOS est affecté, que ce soit au niveau des charges, du potentiel de surface, du courant, etc.

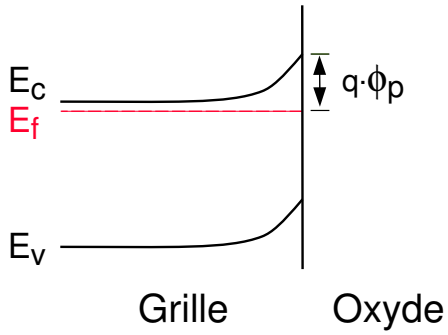


FIG. 4.16 : *Schéma illustrant le phénomène de polydéplétion. Les bandes d'énergie côté grille se courbent d'une valeur ϕ_p, en raison de la couche de déplétion formée au niveau de l'interface polySi–SiO_2.*

L'application du théorème de Gauss à l'interface Si–SiO_2 mène généralement à l'expression suivante (cf. chapitre 2, page 19) :

$$Q_{sc} = -C_{ox} \cdot (V_g - V_{fb} - \phi_s) \tag{4.94}$$

En incluant le phénomène de polydéplétion, (4.94) peut alors être écrite comme :

$$Q_{sc} = -C_{ox} \cdot (V_g - V_{fb} - \phi_s - \phi_p) \tag{4.95}$$

où la valeur de ϕ_p est pour l'instant indéterminée. Pour déterminer ϕ_p, nous allons utiliser un raisonnement semblable à celui présenté au chapitre 2, lors de la définition du potentiel de surface, page 17. Nos explications correspondent au cas d'un transistor n-MOS, ayant une grille en polysilicium de type n^+. Ce raisonnement est bien entendu aussi valable dans le cas d'un p-MOS, moyennant les changements adéquats de signes et de types de porteurs.

Étant en présence d'une grille fortement dopée de type n^+, nous allons supposer que la densité de trous p' dans la grille peut être négligée pour des conditions standards de

fonctionnement[1]. Dans ce cas, la charge d'espace $\rho'(x, y)$ dans la grille de polysilicium est donnée par :

$$\rho'(x, y) = q \cdot (N_p - n'(x, y)) \tag{4.96}$$

où N_p représente le dopage (concentration nette d'atomes donneurs) du polysilicium. La densité d'électrons n' du polysilicium est donnée par les statistiques de Maxwell-Boltzmann[2] :

$$n'(x, y) \simeq N_p \cdot \exp\left(-\frac{\psi(x, y)}{\phi_t}\right) \tag{4.97}$$

La grandeur ψ est définie comme le potentiel électrostatique par rapport à la zone neutre de la grille. Sous l'approximation du canal graduel, l'équation de Poisson 1D — dans la grille — peut être écrite comme :

$$\frac{\partial^2 \psi}{\partial y^2} \simeq \frac{qN_p}{\epsilon_{si}} \cdot \left[1 - \exp\left(\frac{-\psi}{\phi_t}\right)\right] \tag{4.98}$$

Les conditions aux limites pour ψ et $\partial\psi/\partial y$ sont choisies pour que ces deux quantités soient égales à zéro dans la zone neutre de la grille. En utilisant la relation $\partial^2\psi/\partial y^2 = \frac{1}{2} \cdot \partial(\partial\psi/\partial y)^2/\partial y$, la densité de charge de la grille Q_g est alors obtenue en appliquant la loi de Gauss :

$$\begin{aligned} Q_g &= \epsilon_{si} \cdot \left.\frac{\partial \psi}{\partial y}\right|_{y=-t_{ox}} \\ &= \pm\, \gamma_p \cdot C_{ox} \cdot \left\{\phi_p + \phi_t \cdot \left[\exp\left(\frac{-\phi_p}{\phi_t}\right) - 1\right]\right\}^{1/2} \end{aligned} \tag{4.99}$$

où γ_p est le facteur de corps du polysilicium, égal à $\sqrt{2q\epsilon_{si}N_p}/C_{ox}$. Comme indiqué dans (4.99), la densité de charge Q_g est calculée pour $y = -t_{ox}$, c'est-à-dire au niveau de l'interface polySi–SiO_2 ($y = 0$ correspondant à la position de l'interface Si–SiO_2).

En raison de la neutralité de la charge ($Q_g = -Q_{sc}$), les équations (2.6) , cf. page 18, (4.95) et (4.99) peuvent être égalées, résultant en deux expressions implicites à partir desquelles peuvent être extraites les deux grandeurs ϕ_s et ϕ_p. Dans le but de simplifier les choses, une distinction est alors faite entre les régions d'accumulation ($V_g < V_{fb}$) et d'inversion ($V_g > V_{fb}$).

[1]Cela revient à dire que la grille peut uniquement être en accumulation ou en déplétion. Ainsi il est supposé qu'aucune couche d'inversion ne se forme dans la grille de polysilicium, ce qui est vrai pour des conditions normales d'utilisation du transistor.

[2]En pratique, la grille est dégénérée et donc les statistiques de Fermi-Dirac devraient théoriquement être utilisées. Les résultats obtenus avec celles de Maxwell-Boltzmann sont néanmoins très satisfaisants.

En accumulation, une couche d'accumulation de trous est formée dans le substrat et une couche d'accumulation d'électrons est formée dans la grille. Dans ce cas, ϕ_s et ϕ_p sont négatifs, et en conséquence (2.6) peut être approximée par :

$$Q_{sc} \simeq \gamma \cdot C_{ox} \cdot \sqrt{\phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t}\right) - 1\right]} \tag{4.100}$$

ce qui nous permet d'écrire en égalant (4.99) et (4.100) :

$$\gamma \cdot \sqrt{\phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t}\right) - 1\right]} = \gamma_p \cdot \sqrt{\phi_p + \phi_t \cdot \left[\exp\left(\frac{-\phi_p}{\phi_t}\right) - 1\right]} \tag{4.101}$$

À ce point, sachant que $\gamma \ll \gamma_p$ (car $N_a \ll N_p$), il est clair que $|\phi_s| \gg |\phi_p|$. En conséquence, il est raisonnable de négliger l'influence de ϕ_p en accumulation, et donc de considérer que ϕ_p est approximativement égal à zéro.

En inversion, une couche d'inversion d'électrons et une couche de déplétion d'atomes accepteurs ionisés sont formées dans le substrat. De plus, une couche de déplétion d'atomes donneurs ionisés est aussi formée dans la grille. Dans ce cas, ϕ_s et ϕ_p sont positifs, et en conséquence (4.99) peut être approximée par :

$$Q_g \simeq \gamma_p \cdot C_{ox} \cdot \sqrt{\phi_p} \tag{4.102}$$

Le potentiel ϕ_p peut maintenant être calculé en égalant (4.95) et (4.102). Il en résulte une expression relativement simple de ϕ_p, valide dans toutes les régions de fonctionnement, et donnée par :

$$\phi_p = \begin{cases} 0 & \text{pour } V_g < V_{fb}, \\ \left[\sqrt{V_g - V_{fb} - \phi_s + \gamma_p^2/4} - \gamma_p/2\right]^2 & \text{pour } V_g > V_{fb}. \end{cases} \tag{4.103}$$

Il est intéressant de remarquer que cette expression est aussi valable dans le cas d'une grille en métal, car alors $\gamma_p \to \infty$ et $\phi_p = 0$, ce qui est cohérent.

Pour conclure, l'effet de polydéplétion est généralement modélisé en définissant une tension de grille effective V_{geff}, telle que :

$$V_{geff} = V_g - \phi_p = V_{fb} + \phi_s + \frac{\gamma_p^2}{2} \cdot \left[\sqrt{1 + \frac{4}{\gamma_p^2} \cdot (V_g - V_{fb} - \phi_s)} - 1\right] \tag{4.104}$$

Cette tension de grille effective est donc inférieure à la valeur appliquée à la grille, puisque ϕ_p est une quantité positive. En remplaçant V_g par V_{geff} dans les équations permettant d'obtenir les charges et le courant, une description précise de l'influence de la polydéplétion est finalement obtenue [31,32,42,43].

Une observation rapide de (4.104) montre cependant que le calcul de ϕ_p nécessite la connaissance du potentiel de surface ϕ_s. Dans les modèles formulés en tension de seuil, cette difficulté est contournée en posant $\phi_s = \phi_B$ dans l'équation de V_{geff}, ce qui en définitive n'est pas une trop mauvaise approximation, comme le montre la Fig. 4.17.

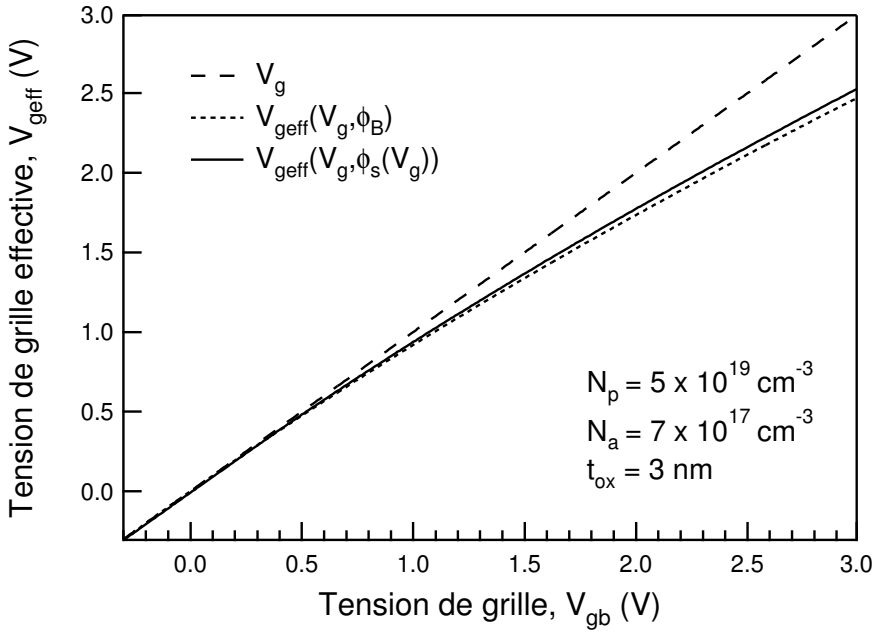


FIG. 4.17 : *Modélisation du phénomène de polydéplétion par la définition d'une tension de grille effective.*

En commentaire final, nous préciserons qu'il est envisageable de prendre en compte directement l'effet de polydéplétion au sein même des équations analytiques décrivant le potentiel de surface. Une telle approche complique cependant nettement les différentes expressions analytiques de ϕ_s (en inversion faible et forte), et nécessite en outre de faire de nouvelles approximations. Lors de la validation expérimentale de notre modèle, nous montrerons que la modélisation de la polydéplétion en termes de tension de grille effective donne des résultats tout à fait satisfaisants.

4.7 Validation du modèle : déplétion et inversion

Comme nous l'avons expliqué en détail au cours du chapitre 2, l'intérêt majeur d'un modèle formulé en potentiel de surface est double. Tout d'abord, un tel modèle offre une description naturelle et précise des charges, résultant de l'application du théorème de Gauss au niveau de l'interface Si–SiO_2 de la structure MOS. Ensuite, de par la formulation unifiée du potentiel de surface de la déplétion à l'inversion forte, le courant de drain peut être décrit de manière physique et continue grâce à l'utilisation d'un modèle drift–diffusion. Aucune interpolation au niveau de la région d'inversion modérée n'est nécessaire, contrairement aux modèles formulés en tension de seuil.

4.7.1 Modélisation des charges et des capacités

Nous allons commencer par décrire les différentes charges du TMOS, puisque la connaissance rigoureuse de la charge d'inversion est indispensable au calcul du courant de drain. Rappelons par commodité les trois charges caractéristiques d'une structure MOSFET :

– La densité de charge de grille :

$$Q_g = C_{ox} \cdot (V_g - V_{fb} - \phi_{s[\mathrm{qm}]}) = -(Q_{inv} + Q_b) \tag{4.105}$$

– La densité de charge du substrat :

$$\begin{aligned} Q_b &= \pm\, \gamma \cdot C_{ox} \cdot \sqrt{\phi_{s[\mathrm{qm}]} + \phi_t \cdot \left[\exp\left(\frac{-\phi_{s[\mathrm{qm}]}}{\phi_t}\right) - 1\right]} \\ &\simeq -\gamma \cdot C_{ox} \cdot \sqrt{\phi_{s[\mathrm{qm}]}} \quad \text{déplétion et inversion} \end{aligned} \tag{4.106}$$

– La densité de charge d'inversion :

$$Q_{inv} = -C_{ox} \cdot (V_g - V_{fb} - \phi_{s[\mathrm{qm}]}) - Q_b \tag{4.107}$$

Lorsque la polydéplétion est prise en compte, les équations décrivant Q_g et Q_{inv} peuvent être réécrites de la façon suivante[1] :

$$Q_g = C_{ox} \cdot (V_g - V_{fb} - \phi_{s[\mathrm{qm}]} - \phi_p) \tag{4.108}$$

[1] L'impact de la polydéplétion sur la charge de déplétion étant négligeable, celui-ci n'est pas pris en compte [31].

$$Q_{inv} = -C_{ox} \cdot (V_g - V_{fb} - \phi_{s[\mathrm{qm}]} - \phi_p) - Q_b \tag{4.109}$$

Écrire les charges de cette façon revient en fait à utiliser V_{geff} au lieu de V_g puisque $V_{geff} = V_g - \phi_p$. Cette simple description des charges, basée sur l'approximation de la feuille de charge (cf. chapitre 2, page 19), permet alors de facilement calculer les différentes transcapacités C_{ij}, définies par :

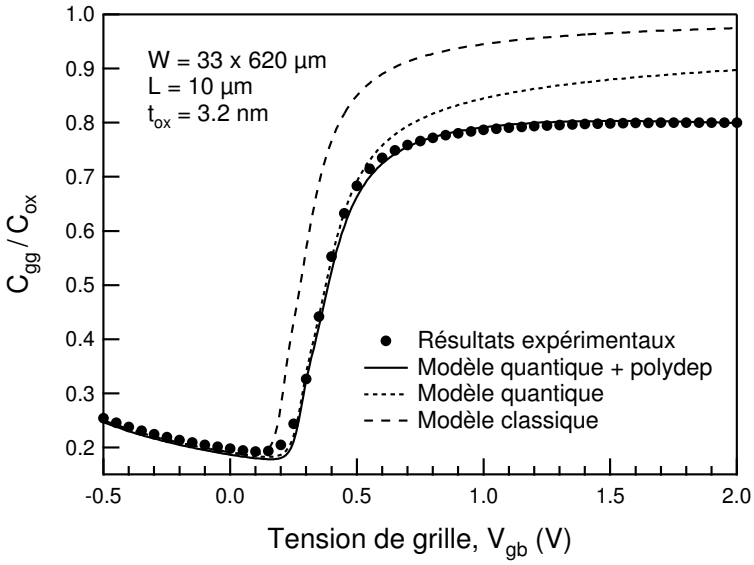
$$C_{ij} = \pm \frac{\partial Q_i}{\partial V_j} \tag{4.110}$$

où i et j dénotent l'électrode de grille g, de substrat b, de drain d ou de source s. Le signe négatif est utilisé si $i \neq j$ et le signe positif pour $i = j$.

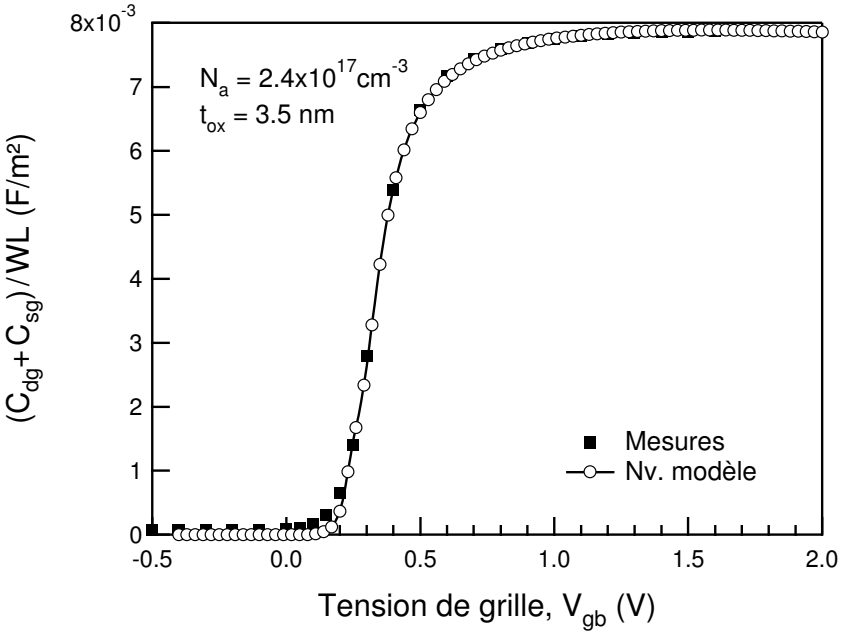
L'utilisation conjointe de ce modèle en feuille de charge avec notre modèle quantique analytique de potentiel de surface offre donc une solution simple et efficace pour simuler les transcapacités du TMOS. Dans le but de valider notre modèle, nous avons comparé ses résultats avec deux jeux de mesures expérimentales, provenant de Philips et de Motorola. Ces deux lots correspondent chacun à des technologies CMOS de 0.18 μm, actuellement en production.

Les caractéristiques technologiques associées aux mesures de Philips sont les suivantes : le dopage substrat est non-uniforme mais d'un ordre de grandeur de 6×10^{17} cm^{-3}, le dopage polysilicium est de l'ordre de 1.2×10^{20} cm^{-3} et l'épaisseur d'oxyde de grille est égale à 3.2 nm. En utilisant exactement les valeurs données ci-dessus, un très bon accord est observé entre le nouveau modèle et les résultats expérimentaux, comme le montre la Fig. 4.18a.

Concernant les caractéristiques technologiques de Motorola, nous avons obtenu les informations suivantes : le dopage substrat est légèrement supérieur à 2×10^{17} cm^{-3}, le dopage polysilicium est environ égal à 10^{20} cm^{-3} et l'épaisseur d'oxyde est égale à 3.5 nm.
Le meilleur ajustement de notre modèle a été obtenu pour une valeur de dopage substrat égale à 2.4×10^{17} cm^{-3}, tous les autres paramètres étant fixés aux valeurs indiquées ci-dessus. Une comparaison entre les mesures expérimentales de Motorola et nos simulations analytiques est présentée à la Fig. 4.18b. Un excellent agrément modèle–mesures est clairement montré.



(a) Résultats expérimentaux de Philips.



(b) Résultats expérimentaux de Motorola.

FIG. 4.18 : *Comparaison du nouveau modèle avec deux jeux de mesures capacitives (C–V). Précisons que $C_{gg} = dQ_g/dV_g$ (Fig. a) et que $(C_{dg} + C_{sg}) = dQ_{inv}/dV_g$ (Fig. b).*

Un point fondamental de cette modélisation des charges — et donc des capacités — est que l'épaisseur d'oxyde t_{ox} n'a pas été considérée comme un paramètre d'ajustement du modèle. Toutes nos simulations ont été réalisées en utilisant les valeurs exactes de t_{ox}, i.e. celles fournies par les deux fondeurs. Les bons résultats obtenus avec notre modèle confirment donc la validité de notre approche, et en outre démontrent que la définition d'une capacité d'oxyde effective n'est en rien obligatoire pour prendre en compte correctement les effets quantiques.

4.7.2 Modélisation du courant de drain

Le calcul du courant de drain avec une approche drift–diffusion requiert à la fois une description très précise du potentiel de surface et de la densité de charge d'inversion. Comme nous l'avons montré au paragraphe précédent, la densité de charge d'inversion dépend directement du potentiel de surface. L'équation (4.107) permet une description précise de Q_{inv} en inversion modérée et forte, mais souffre d'un mauvais comportement au début de l'inversion faible (avec ou sans considération des effets quantiques). Bien que cela n'ait aucune influence sur le calcul des transcapacités, cela pose quelques problèmes lors du calcul du courant de diffusion, où la pente sous le seuil n'est alors correcte que sur 2 ou 3 décades, à la fin de l'inversion faible.

L'équation usuelle de la densité de charge d'inversion est égale à :

$$Q_{inv} = -C_{ox} \cdot (V_g - V_{fb} - \phi_s) + \gamma \cdot C_{ox} \cdot \sqrt{\phi_s} \tag{4.111}$$

Pour améliorer le comportement de (4.111) en inversion faible, il est nécessaire de considérer l'équation implicite (4.70), que nous rappelons ici, dans un souci de clarté :

$$V_g - V_{fb} - \phi_s = \gamma \cdot \left[\phi_s + \phi_t \cdot \exp\left(\frac{\phi_s - V_{ch} - \phi_B}{\phi_t}\right)\right]^{1/2} \tag{4.112}$$

En remplaçant le terme $(V_g - V_{fb} - \phi_s)$ dans (4.111) par le membre droit de (4.112), nous obtenons alors une description très précise de la densité de charge d'inversion pour toute la région d'inversion faible. Lorsque les effets quantiques sont pris en compte, (4.112) peut être réécrite de la façon suivante :

$$V_g - V_{fb} - \phi_{s[\mathrm{qm}]} = \gamma \cdot \left[\phi_{s[\mathrm{qm}]} + \phi_t \cdot \exp\left(\frac{\phi_{s[\mathrm{qm}]} - V_{ch} - \phi_B - \delta\phi_s}{\phi_t}\right)\right]^{1/2} \tag{4.113}$$

Finalement, une expression améliorée de la densité de charge d'inversion est donnée par :

$$Q_{inv} = -\gamma \cdot C_{ox} \cdot \left\{\left[\phi_{s[\mathrm{qm}]} + \phi_t \cdot \exp\left(\frac{\phi_{s[\mathrm{qm}]} - V_{ch} - \phi_B - \delta\phi_s}{\phi_t}\right)\right]^{1/2} - \sqrt{\phi_{s[\mathrm{qm}]}}\right\} \tag{4.114}$$

La Fig. 4.19 montre une comparaison entre une simulation quantique numérique de la densité de charge d'inversion, et les résultats obtenus avec les différents modèles analytiques de potentiel de surface (classique et quantique) incorporés dans (4.114). Les résultats sont présentés sur deux échelles, linéaire et logarithmique, dans le but d'illustrer clairement le

comportement de (4.114) dans les différents régimes de fonctionnement : inversion faible, modérée et forte. La densité de charge d'inversion est particulièrement bien décrite en inversion modérée et forte. L'augmentation de la tension de seuil due aux effets quantiques est aussi bien définie, comme le montre la comparaison entre les modèles classique et quantique en régime d'inversion faible (échelle logarithmique). Au début de l'inversion faible, le modèle est légèrement moins précis, ceci résultant directement des hypothèses faites lors du développement de l'approximation de l'inversion modérée. Cette précision est cependant tout à fait suffisante dans le cadre d'un modèle compact destiné à la simulation de circuits. Nous montrerons d'ailleurs par la suite que les résultats obtenus pour la simulation du courant de drain sont excellents, et cela sur toute la dynamique de tension de grille, de l'inversion faible à l'inversion forte.

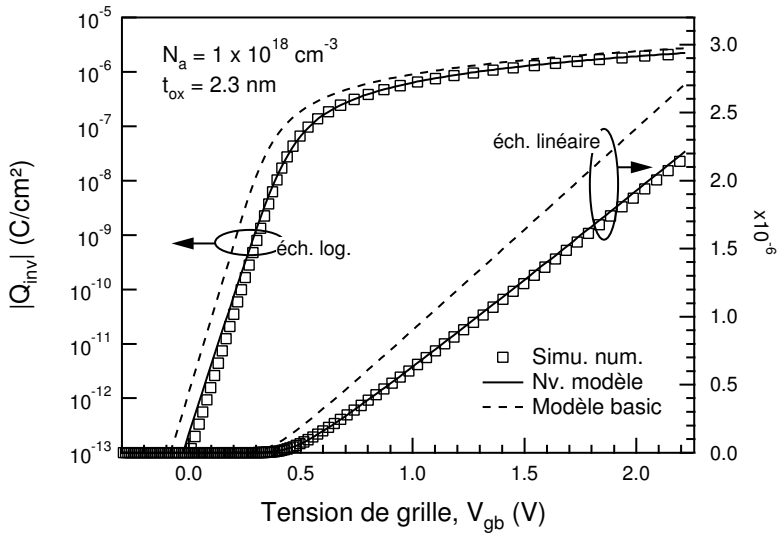


FIG. 4.19 : *Influence des effets quantiques sur la densité de charge d'inversion. Les symboles représentent les résultats des simulations auto-cohérentes Schrödinger–Poisson, et les lignes les résultats obtenus avec le modèle de potentiel de surface analytique (traits pleins : nouveau modèle quantique, tirets : modèle classique.)*

Dans notre modèle de courant de drain, la relation (4.114) sera utilisée pour modéliser la composante diffusion I_{diff} du courant de drain I_d. L'expression usuelle de Q_{inv} sera quant à elle utilisée pour décrire les autres principales caractéristiques électriques, à savoir le courant de conduction I_{drift} et toutes les transcapacités.

Rappelons tout d'abord que le principe de l'approche drift–diffusion a déjà été détaillée au cours du chapitre 2, page 23. Dans ce contexte, le courant de drain I_d est défini comme la somme de deux composantes : le courant de diffusion I_{diff} et le courant de conduction I_{drift}.

Le courant de conduction est exprimé comme :

$$\begin{aligned} I_{\text{drift}} &= -\mu_n \cdot \frac{W}{L} \cdot \int_{\phi_{s0}}^{\phi_{sL}} Q_{inv}\, d\phi_s \\ &= \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \int_{\phi_{s0}}^{\phi_{sL}} \left[V_g - V_{fb} - \phi_{s[\text{qm}]} - \gamma \cdot \sqrt{\phi_{s[\text{qm}]}} \right] d\phi_s \end{aligned} \tag{4.115}$$

Après intégration, il vient :

$$\begin{aligned} I_{\text{drift}} = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \Big[&(V_g - V_{fb}) \left(\phi_{sL[\text{qm}]} - \phi_{s0[\text{qm}]} \right) \\ &- \frac{1}{2} \cdot \left(\phi_{sL[\text{qm}]}^2 - \phi_{s0[\text{qm}]}^2 \right) - \frac{2}{3} \cdot \gamma \cdot \left(\phi_{sL[\text{qm}]}^{3/2} - \phi_{s0[\text{qm}]}^{3/2} \right) \Big] \end{aligned} \tag{4.116}$$

où ϕ_{s0} et ϕ_{sL} sont les valeurs du potentiel de surface à la fin de la source et du drain, respectivement. Ainsi, $\phi_{s0} = \phi_s(V_{ch} = V_s)$ et $\phi_{sL} = \phi_s(V_{ch} = V_d)$.

Le courant de diffusion est quant à lui donné par :

$$\begin{aligned} I_{\text{diff}} &= \mu_n \cdot \frac{W}{L} \cdot \phi_t \cdot \int_{\phi_{s0}}^{\phi_{sL}} dQ_{inv} \\ &= \mu_n \cdot \frac{W}{L} \cdot \phi_t \cdot (Q_{invL} - Q_{inv0}) \end{aligned} \tag{4.117}$$

En combinant (4.114) avec (4.117), nous obtenons une description très précise du courant de diffusion :

$$\begin{aligned} I_{\text{diff}} = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \phi_t \cdot \gamma \cdot \Bigg\{ &\left[\phi_{s0[\text{qm}]} + \phi_t \cdot \exp\left(\frac{\phi_{s0[\text{qm}]} - V_s - \phi_B - \delta\phi_s}{\phi_t} \right) \right]^{1/2} \\ - \sqrt{\phi_{s0[\text{qm}]}} - &\left[\left[\phi_{sL[\text{qm}]} + \phi_t \cdot \exp\left(\frac{\phi_{sL[\text{qm}]} - V_d - \phi_B - \delta\phi_s}{\phi_t} \right) \right]^{1/2} - \sqrt{\phi_{sL[\text{qm}]}} \right] \Bigg\} \end{aligned} \tag{4.118}$$

L'expression finale du courant de drain est donc de la forme :

$$I_d = I_{\text{drift}} + I_{\text{diff}} \tag{4.119}$$

Cette expression est réellement très pratique, car elle permet de simuler I_d pour différents cas de figure, comme l'illustre le Tableau 4.1.

La description complète du courant de drain d'un transistor MOS idéal, à canal long est maintenant réalisée. Le problème est que dans la pratique, le comportement électrique du

TAB. 4.1 : *Simulation du courant de drain selon différents cas de figure : CL=classique, PDE=polydéplétion et QME=quantique. Il suffit simplement d'utiliser V_{geff} au lieu de V_g et/ou $\phi_{s[\mathrm{qm}]}$ au lieu de ϕ_s, pour prendre en compte ou non les différents effets.*

I_d	CL	CL+PDE	QME	QME+PDE
V_g	×		×	
V_{geff}		×		×
ϕ_s	×	×		
$\phi_{s[\mathrm{qm}]}$			×	×

TMOS dévie considérablement du comportement idéal décrit par (4.119)[1]. Cette déviation est due à de nombreux effets, tels que la réduction de la mobilité des porteurs, la saturation de la vitesse des porteurs, les résistances série, la modulation de la longueur du canal, le DIBL (Drain Induced Barrier Lowering), etc. La modélisation de ces effets est cependant hors du cadre de cette thèse ; en fait, le traitement complet de tous ces effets mériterait à lui seul plusieurs ouvrages comme celui-ci. Les articles et ouvrages traitant de ces effets sont nombreux, nous pouvons entre autres citer les suivants [31,32,43–49].

Dès lors, il devient difficile de confronter notre modèle à des mesures expérimentales de courant de drain. Il est malgré tout possible de s'affranchir de la plupart de ces difficultés si les critères suivants sont respectés :

- Les TMOS utilisés lors des mesures doivent être longs : $L \geqslant 10\ \mu$m,
- les mesures doivent être réalisées à faible polarisation drain–source : $V_{ds} \leqslant 100$ mV.

Le seul problème restant concerne la réduction de la mobilité des porteurs μ_n, qui est un phénomène très important en inversion forte. Nous l'avons modélisé par une fonction similaire à celle employée dans les modèles SPICE (BSIM2, BSIM3) [50] :

$$\mu_n = \frac{\mu_0}{1 + \theta_1 \cdot (V_g - V_{to}) + \theta_2 \cdot (V_g - V_{to})^2} \tag{4.120}$$

où μ_0, θ_1 et θ_2 sont trois paramètres d'ajustement, à adapter en fonction des différentes technologies de TMOS. Cette modélisation, peu satisfaisante physiquement, permet cependant d'obtenir un bon accord avec les résultats expérimentaux.

[1]En particulier pour les TMOS à canaux courts.

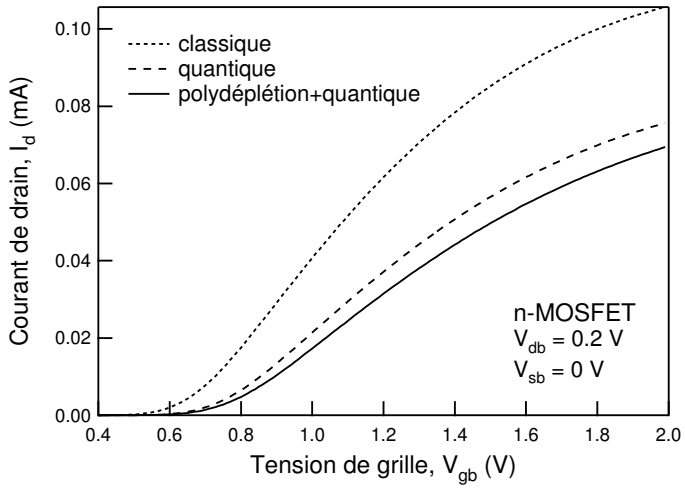
L'intérêt de valider un modèle par rapport à des courbes *C–V* expérimentales trouve ici tout son sens, la mobilité n'intervenant pas dans le calcul des charges. La validation d'un modèle formulé en potentiel de surface ne sera donc pas sujette à discussion si les transcapacités sont bien décrites. Bien évidemment, cette remarque n'a de sens que si les paramètres d'entrée du modèle (t_{ox}, N_a, etc.) sont strictement les mêmes que les paramètres technologiques du dispositif que l'on souhaite simuler. Dans le cas d'une validation avec des courbes *I–V* expérimentales, il est vrai qu'un simple ajustement des paramètres de (4.120) peut être suffisant pour décrire correctement le courant de drain en inversion forte. Par contre, il n'est pas possible de « tricher » pour décrire correctement la pente de I_d en inversion faible. C'est pourquoi, dans le but de lever toute ambiguïté, nous proposons dans ce manuscrit une double validation expérimentale du modèle : *I–V* et *C–V*. Le bon comportement de notre modèle du point de vue capacitif ayant déjà été prouvé (Fig. 4.18, page 140), nous allons maintenant étudier l'impact des effets quantiques sur le courant de drain.

Avant de comparer le modèle aux mesures, nous allons tout d'abord observer la Fig. 4.20 et le Tableau 4.2. La Fig. 4.20a décrit sur une échelle linéaire trois simulations drift–diffusion du courant de drain : la première correspond à l'utilisation du modèle classique, la deuxième y ajoute l'effet de polydéplétion, et la troisième correspond au modèle de potentiel de surface complet (quantique et polydéplétion). La Fig. 4.20b présente les mêmes simulations sur une échelle logarithmique, ce qui nous permet de distinguer le décalage du seuil résultant de l'utilisation du modèle quantique. Ces figures et ce tableau illustrent clairement la très forte influence des effets quantiques sur le courant de drain : leur prise en compte dans un modèle compact est donc indispensable. Le rôle de la polydéplétion apparaît quant à lui légèrement moins important en ce qui concerne la diminution du courant de drain en inversion modérée et forte. L'impact de la polydéplétion est en fait plus significatif pour le calcul des transcapacités, dans ce cas il est quasiment équivalent à celui des effets quantiques, en particulier quand le TMOS est en régime d'inversion forte (cf. Tableau 4.2).

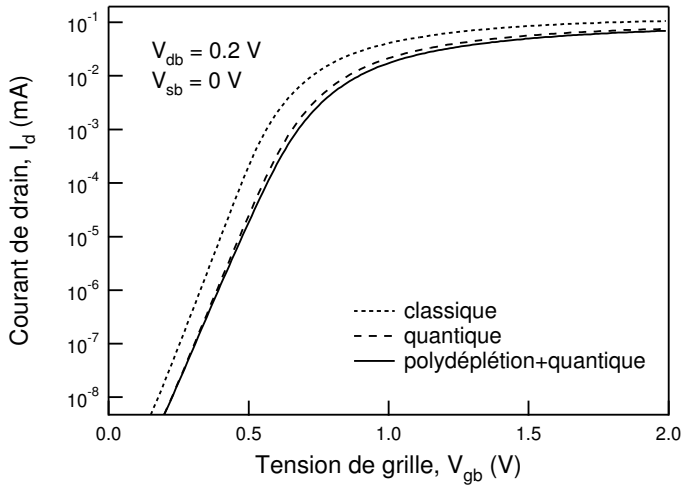
Une première validation *I–V* est réalisée par comparaison avec des résultats de Philips. Nous avons obtenu deux lots de mesures, les caractéristiques communes aux deux étant un dopage substrat uniforme égal à 5×10^{17} cm^{-3} et un dopage polysilicium légèrement supérieur à 1×10^{20} cm^{-3}. Chaque jeu de mesure est caractérisé par une épaisseur d'oxyde différente, à savoir $t_{ox} = 2$ et 5 nm.

TAB. 4.2 : *Erreur relative commise sur le calcul du courant de drain I_d et de la transcapacité de grille C_{gg}, lors de l'utilisation du modèle ϕ_s classique. Les erreurs sont exprimées en %. Les comparaisons ont été effectuées à $V_g = 1$ V, $V_d = 0.1$ V et $V_s = 0$ V. Le niveau de dopage polysilicium choisi dans ces simulations est $N_p = 8 \times 10^{19}$ cm^{-3}.*

N_a (cm^{-3})	t_{ox} (nm)	$\frac{I_d - I_{d[\mathrm{qm}]}}{I_{d[\mathrm{qm}]}}$	$\frac{I_d - I_{d[\mathrm{qm+pd}]}}{I_{d[\mathrm{qm+pd}]}}$	$\frac{C_{gg} - C_{gg[\mathrm{qm}]}}{C_{gg[\mathrm{qm}]}}$	$\frac{C_{gg} - C_{gg[\mathrm{qm+pd}]}}{C_{gg[\mathrm{qm+pd}]}}$
5×10^{17}	3.5	60	75	12	20
1×10^{18}	2	97	154	17.5	45



(a) Échelle linéaire.



(b) Échelle logarithmique.

FIG. 4.20 : *Influence des effets quantiques et du phénomène de polydéplétion sur le calcul du courant de drain. Les paramètres technologiques du TMOS simulé sont typiques d'une génération CMOS avancée : $t_{ox} = 2.5$ nm, $N_a = 8 \times 10^{17}$ cm^{-3}, $N_p = 8 \times 10^{19}$ cm^{-3}.*

La Fig. 4.21 montre que le courant de drain des deux dispositifs est précisément décrit par notre modèle. Le courant de drain est aussi représenté sur une échelle logarithmique dans le but de démontrer le bon comportement du modèle dans la région sous le seuil. Il apparaît que la pente en inversion faible est correctement prédite pour les deux dispositifs. Une précision importante est que le modèle n'est pas seulement valide pour le potentiel de surface ϕ_s mais aussi pour sa dérivée du premier-ordre $\partial\phi_s/\partial V_g$. Cela permet non seulement une estimation précise des transcapacités (cf. Fig. 4.18), mais aussi de la transconductance g_m (cf. Fig. 4.22). D'après la Fig. 4.22, la valeur maximum de g_m est fortement réduite par les effets quantiques, et est obtenue pour une tension de grille supérieure à celle prédite par une simulation classique. L'erreur engendrée par l'utilisation du modèle classique sur le pic de g_m est d'environ 20 %.

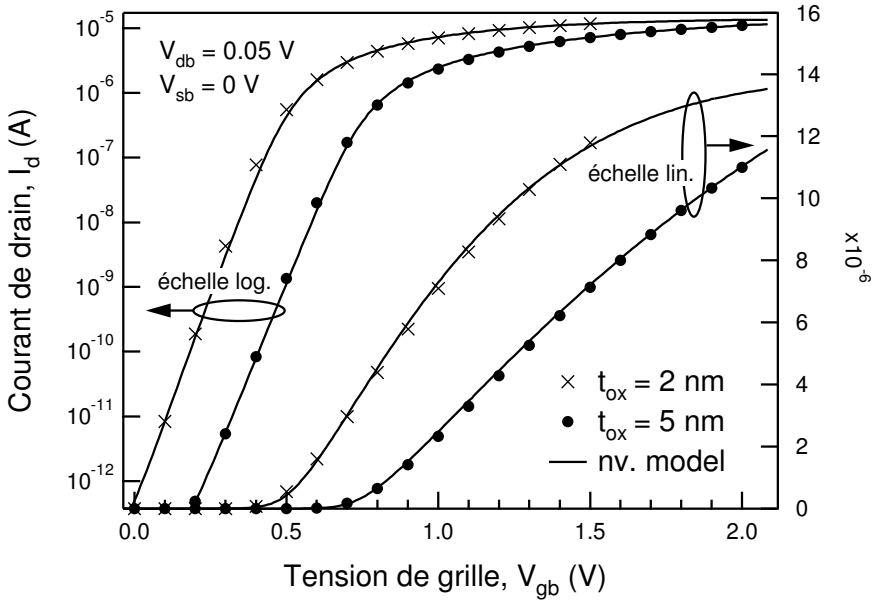


FIG. 4.21 : *Comparaison entre le courant de drain simulé par le nouveau modèle de potentiel de surface (quantique + polydéplétion) et les mesures expérimentales. Les mesures ont été effectuées à faible tension de drain ($V_{ds} = 50$ mV) sur des transistors n-MOS de grande géométrie ($W/L = 10/10$ μm).*

Une seconde validation *I–V* est réalisée par une comparaison du modèle avec des mesures issues d'une technologie CMOS 0.18 μm de Motorola[1]. La Fig. 4.23 montre l'excellent agrément entre le modèle quantique analytique et les résultats expérimentaux.

[1]Cette technologie est la même que celle décrite lors de la validation *C–V* du modèle, voir page 140. Ses principales caractéristiques sont $N_a = 2.4 \times 10^{17}$ cm^{-3}, $N_p = 1 \times 10^{20}$ cm^{-3} et $t_{ox} = 3.5$ nm.

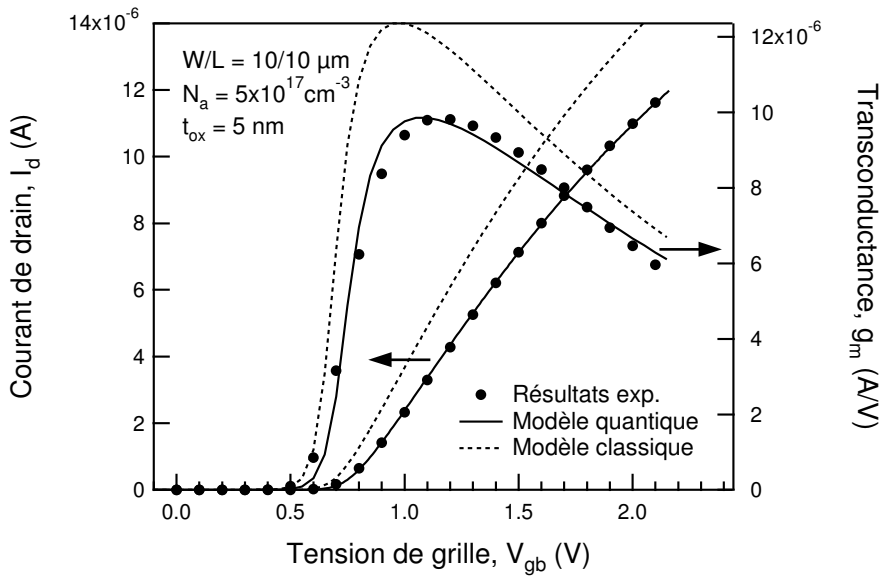


FIG. 4.22 : *Influence des effets quantiques sur le calcul de la transconductance g_m. Les mesures ont été effectuées à faible tension de drain ($V_{ds} = 50$ mV) sur des transistors n-MOS de grande géométrie ($W/L = 10/10$ μm).*

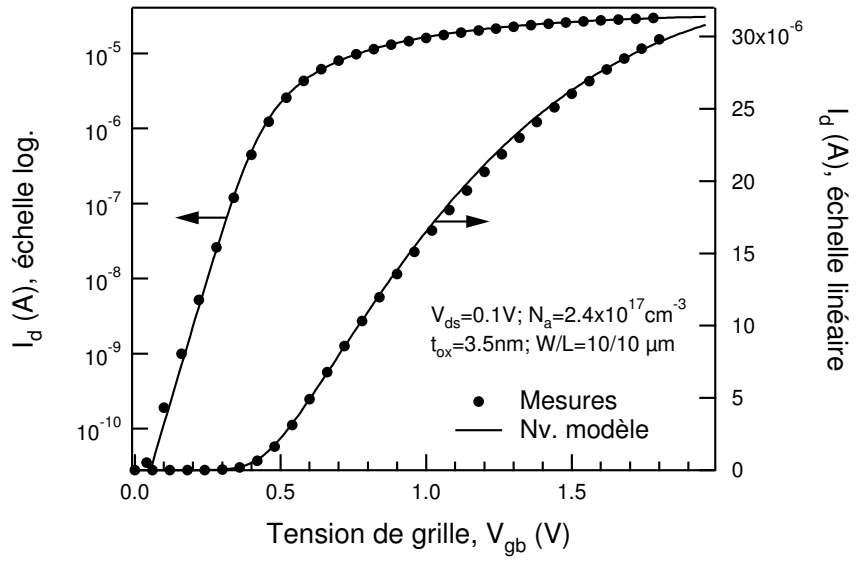


FIG. 4.23 : *Validation du modèle de courant de drain avec les mesures d'une technologie 0.18 μm de Motorola. Les caractéristiques technologiques sont indiquées sur la figure.*

4.7.3 Conclusion

Une formulation quantique et explicite du potentiel de surface, valide de la déplétion à l'inversion forte, vient d'être développée. L'inclusion des effets quantiques est basée sur un nouveau concept — l'approximation de l'inversion modérée — qui permet de prendre en compte analytiquement l'augmentation quantique du potentiel de surface en régime d'inversion.

Ce modèle est complètement dépendant des différentes tensions appliquées au dispositif et ne nécessite aucun paramètre d'ajustement. L'incorporation du modèle de potentiel de surface quantique au sein d'un modèle en feuille de charge autorise des simulations réalistes et rapides des caractéristiques électriques fondamentales du TMOS, telles que les charges, les transcapacités, le courant de drain, etc. Finalement de nombreuses comparaisons avec des simulations numériques auto-cohérentes Schrödinger–Poisson et des résultats expérimentaux ont confirmé la validité de notre approche de modélisation. En dernier lieu, dans le but de démontrer que notre modèle est approprié à la simulation de circuits, nous l'avons implémenté dans deux langages de haut niveau, à savoir Verilog-AMS et VHDL-AMS [51]. Le code VHDL-AMS de ce modèle est présenté dans son intégralité en annexe, page 179.

Bien que cette première partie du modèle soit la plus importante[1], elle est cependant incomplète. En effet, la connaissance du potentiel de surface en accumulation est obligatoire si l'on désire obtenir une caractéristique $C_{gg}(V_g)$ complète, ce qui est particulièrement utile dans le domaine de la caractérisation [12,14].

4.8 Formulation du nouveau modèle : accumulation

Nous allons dans un premier temps expliquer comment obtenir une modélisation analytique du potentiel de surface en accumulation. Après avoir explicité cette description classique, nous détaillerons notre méthode pour prendre en compte les effets quantiques dans cette région. De façon similaire au modèle d'inversion, l'augmentation du potentiel de surface (en valeur absolue) sera directement exprimée en termes d'incrément de potentiel de surface $\delta\phi_s$.

4.8.1 Modèle explicite classique

L'équation implicite (4.69) dont dépend le potentiel de surface peut facilement être approximée en région d'accumulation. Dans cette région, l'influence des trous est dominante et $\phi_s < 0$ puisque $V_g < V_{fb}$. En conséquence, en négligeant l'influence des électrons dans (4.69) nous obtenons :

$$V_g - V_{fb} - \phi_s = -\gamma \cdot \sqrt{\phi_s + \phi_t \cdot \left[\exp\left(\frac{-\phi_s}{\phi_t}\right) - 1\right]} \tag{4.121}$$

[1] Elle correspond en effet à l'état 'on' du TMOS.

Le terme exponentiel étant dominant pour $\phi_s < 0$, il est approprié de réécrire (4.121) de la façon suivante :

$$\phi_s = -\phi_t \cdot \ln\left[\frac{(V_g - V_{fb} - \phi_s)^2/\gamma^2 - \phi_s + \phi_t}{\phi_t}\right] \tag{4.122}$$

Nous allons maintenant procéder à un raisonnement complètement équivalent à celui proposé lors de la description du modèle d'inversion forte.

Dans (4.122), les termes ϕ_s dans le logarithme peuvent être remplacés par une constante appelée ψ_B (en quelque sorte l'équivalent de ϕ_B pour l'inversion forte), que nous allons dans un premier temps supposer nulle puisque $\phi_s \approx 0$ lorsque $V_g < V_{fb}$. La nouvelle expression du potentiel de surface est alors donnée par :

$$\begin{aligned}\phi_s &\simeq -\phi_t \cdot \ln\left[\frac{(V_g - V_{fb} - \psi_B)^2/\gamma^2 - \psi_B + \phi_t}{\phi_t}\right] \\ &\simeq -\phi_t \cdot \ln\left[\frac{(V_g - V_{fb})^2/\gamma^2 + \phi_t}{\phi_t}\right]\end{aligned} \tag{4.123}$$

D'une façon symétrique au cas de l'inversion où une valeur constante de ϕ_B ne donnait pas une modélisation correcte du potentiel de surface (cf. page 119), l'utilisation d'une valeur constante de ψ_B ne donne pas ici des résultats suffisamment précis. Cette difficulté peut être contournée en définissant une fonction faisant varier le potentiel de surface de zéro au niveau de la tension de bandes plates ($V_g = V_{fb}$), à environ $-4\,\phi_t$ pour des valeurs plus négatives de la tension de grille ($V_g \ll V_{fb}$). En suivant l'approche proposée dans MM11 [31], il est possible de définir une fonction empirique qui réalise ce changement :

$$\psi_B^* = \frac{\Gamma \cdot (V_g - V_{fb})}{\sqrt{1 + \left[\frac{\Gamma \cdot (V_g - V_{fb})}{4 \cdot \phi_t}\right]^2}} \tag{4.124}$$

où Γ est donnée par :

$$\Gamma = \left.\frac{\partial \phi_s}{\partial V_g}\right|_{V_g = V_{fb}} = \frac{1}{1 + \gamma/\sqrt{2 \cdot \phi_t}} \tag{4.125}$$

assurant que $\partial\psi_B/\partial V_g = \partial\phi_s/\partial V_g$ à $V_g = V_{fb}$. Cette constante garantit donc un comportement physique cohérent à ψ_B^* au niveau de la tension de bandes plates.

En combinant (4.123) et (4.124), nous obtenons finalement une expression décrivant précisément le potentiel de surface en accumulation :

$$\phi_s = -\phi_t \cdot \ln\left[\frac{(V_g - V_{fb} - \psi_B^*)^2/\gamma^2 - \psi_B^* + \phi_t}{\phi_t}\right] \tag{4.126}$$

Cette équation est cependant valide uniquement en régime d'accumulation. Dans le but d'obtenir une transition douce et continue de l'accumulation à la déplétion, il suffit de remplacer V_g par $(V_g - V_{g[\mathrm{acc}]})$ dans (4.124) et (4.126). La fonction de lissage $V_{g[\mathrm{acc}]}$ est définie de telle sorte qu'elle varie doucement de zéro lorsque $V_g \leqslant V_{fb}$ à $(V_g - V_{fb})$ lorsque $V_g \geqslant V_{fb}$. Elle est donnée par :

$$V_{g[\mathrm{acc}]} = \frac{1}{2} \cdot \left[V_g - V_{fb} + \sqrt{(V_g - V_{fb})^2 + 4 \cdot \kappa^2} \right] \tag{4.127}$$

où κ est une constante de lissage, définitivement fixée à 0.1.

À ce stade du développement, il est facile d'étendre la validité de ce modèle à la région d'inversion faible. L'expression classique du potentiel de surface en inversion faible est définie par (cf. page 117) :

$$\phi_{swi} = \left[\sqrt{V_g - V_{fb} + \gamma^2/4} - \gamma/2 \right]^2 \tag{4.128}$$

En remplaçant $(V_g - V_{fb})$ par $V_{g[\mathrm{acc}]}$, le domaine de définition de (4.128) est alors étendu aux valeurs de V_g inférieures à V_{fb} :

$$\phi^*_{swi} = \left[\sqrt{V_{g[\mathrm{acc}]} + \gamma^2/4} - \gamma/2 \right]^2 \tag{4.129}$$

La Fig. 4.24 illustre l'effet de l'utilisation de la fonction de lissage $V_{g[\mathrm{acc}]}$ sur l'approximation de ϕ_s en inversion faible. Grâce à cette fonction de lissage, l'approximation du potentiel de surface en inversion faible (Eq. (4.129)) converge doucement vers 0 quand $V_g < V_{fb}$.

Dès lors, une description fiable et continue du potentiel de surface, valide de l'accumulation à l'inversion faible est obtenue en combinant (4.126) et (4.129) :

$$\begin{aligned} \phi_{s[\mathrm{acc}\to\mathrm{wi}]} = & \left[\sqrt{V_{g[\mathrm{acc}]} + \gamma^2/4} - \gamma/2 \right]^2 \\ & - \phi_t \cdot \ln \left[\frac{(V_g - V_{g[\mathrm{acc}]} - V_{fb} - \psi^*_B)^2/\gamma^2 - \psi^*_B + \phi_t}{\phi_t} \right] \end{aligned} \tag{4.130}$$

La Fig. 4.25 montre les résultats obtenus avec l'approximation analytique (4.130). La comparaison entre ce modèle et un calcul numérique (résolution exacte de l'équation implicite dont dépend ϕ_s) indique que (4.130) est une excellente approximation de ϕ_s, de l'accumulation à l'inversion faible.

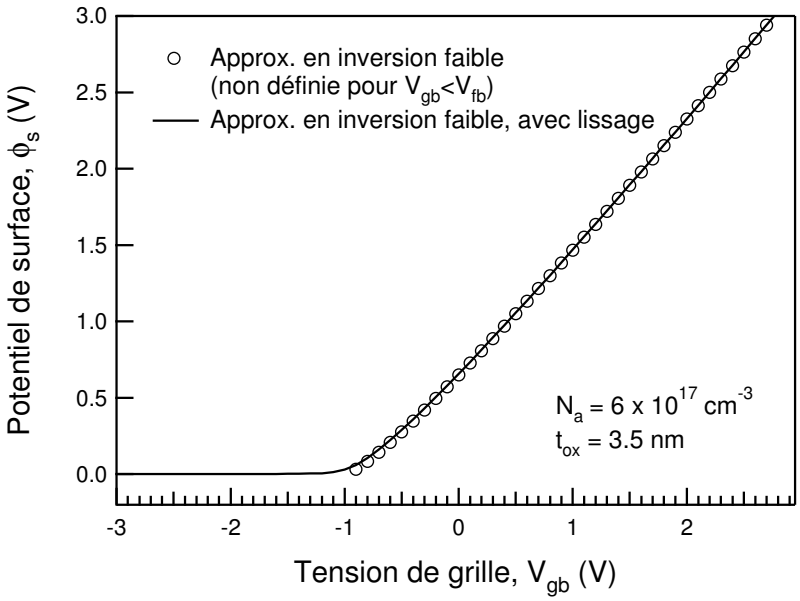


FIG. 4.24 : *Impact de l'utilisation de la fonction de lissage $V_{g[\mathrm{acc}]}$ sur l'approximation du potentiel de surface en inversion faible. Le comportement de ϕ_s en inversion faible n'est pas modifié, ce qui justifie d'un point de vue physique l'usage de $V_{g[\mathrm{acc}]}$.*

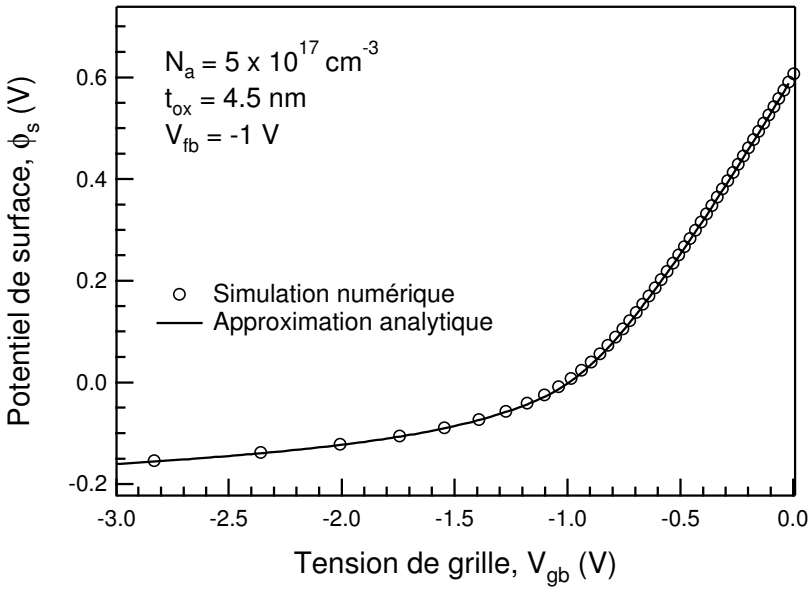


FIG. 4.25 : *Comparaison entre le potentiel de surface calculé numériquement (cercles) et analytiquement (ligne continue), dans un contexte classique i.e. sans effets quantiques.*

Nous avons maintenant à notre disposition une expression purement analytique décrivant le potentiel de surface dans la région d'accumulation. Cette relation, définie par (4.130) permet une modélisation précise du potentiel de surface classique, et sera à la base du modèle quantique que nous allons expliciter au paragraphe suivant.

4.8.2 Modèle explicite quantique

Le traitement quantique de la couche d'accumulation est de loin plus compliqué que celui de la couche d'inversion, en raison de la structure complexe de la bande de valence [11,52]. D'un point de vue physique strict, plusieurs niveaux d'énergie devraient être pris en compte pour modéliser les effets quantiques dans ce régime [14,52].

Dans le cas de notre étude (transistors n-MOS), la couche d'accumulation est constituée de trous, ce qui nous conduit à faire les commentaires suivants. Comme pour les électrons, en raison de l'équilibre thermodynamique nous savons que les niveaux d'énergie les plus faibles sont favorisés. Cependant, même si ces derniers sont favorisés, leur importance réelle dépend du nombre d'états dont ils disposent : c'est à ce moment qu'intervient la notion de masse effective de densité d'états. Pour les trous, les états de trous lourds ($m_{hh} = 0.53\, m_0$) sont également ceux qui ont la masse effective de densité d'états la plus forte, ainsi les états de trous lourds sont de loin les plus peuplés. En conséquence, le calcul de la densité de charge d'accumulation pourrait a priori être réalisé en ne considérant que les trous lourds, i.e. en négligeant les trous légers ($m_{lh} = 0.16\, m_0$). Une telle affirmation est cependant à nuancer, dans la mesure où ces résultats dépendent très fortement des valeurs de masses effectives utilisées lors de la résolution exacte des équations couplées de Schrödinger et de Poisson. En d'autres mots, l'affirmation précédente n'est valable que sous l'hypothèse où les valeurs des masses effectives sont correctes. Bien que pour les électrons l'approximation de la masse effective soit validée, la situation est plus confuse pour les trous. En fait, la structure de la bande de valence ne serait pas quasi-isotrope, et en conséquence d'autres valeurs de masses effectives que celles précédemment données seraient peut-être à prendre en compte [14]. Malheureusement, il n'existe aucun consensus à ce sujet dans la littérature.

En résumé, développer un modèle analytique, physique et prenant en compte les effets quantiques en accumulation est loin d'être évident. La principale difficulté est qu'il est nécessaire d'inclure plusieurs niveaux d'énergie pour calculer correctement la densité de porteurs libres de la couche d'accumulation. A priori ceci n'est pourtant pas un problème insurmontable ; en effet, nous avons vu précédemment dans ce chapitre (cf. page 92) que l'approximation du puits de potentiel triangulaire permet de calculer simplement les énergies des différentes sous-bandes discrètes se trouvant dans le puits de potentiel formé par la forte courbure de bandes à l'interface Si–SiO_2. Une telle approche ne permet malheureusement pas de définir un élargissement du gap comme cela est fait en inversion, ce qui comme nous l'avons vu, est une méthode efficace pour modéliser les effets quantiques.

Dans le but de simplifier le problème, nous avons développé une approche semi-empirique basée sur la considération d'un unique niveau d'énergie effectif $\tilde{E}_0$ au lieu de multiples niveaux ($E_0, E_1, E_2 \ldots$). Grâce à cette approche et toujours dans le cadre de l'approximation du puits de potentiel triangulaire, nous pouvons définir un pseudo élargissement du gap $E_{w[\mathrm{acc}]}$. Ce dernier est en fait une représentation du déplacement quantique de la bande de valence, et joue un rôle équivalent au pseudo élargissement du gap E_w défini lors du développement de notre modèle quantique dédié au mode d'inversion (cf. page 126).

Le concept majeur de ce nouveau modèle repose sur la définition d'une *densité équivalente de porteurs majoritaires* p_{acc} dont le rôle est la prise en compte du nombre de porteurs libres se trouvant dans la couche d'accumulation quantifiée [53]. Nous avons défini p_{acc} par la fonction suivante :

$$p_{acc}(\tilde{V_g}) = \frac{C_{ox}}{q} \cdot \left[\frac{V_{fb} - \tilde{V_g}}{a_1} + \frac{(V_{fb} - \tilde{V_g})^2}{a_2} \right] \tag{4.131}$$

où a_1 et a_2 sont deux constantes indépendantes des paramètres technologiques du transistor MOS, respectivement égales à 2 et 5. Les valeurs de ces deux constantes ont été déterminées par un ajustement du modèle avec des résultats numériques obtenus par résolution auto-cohérente des équations de Schrödinger et de Poisson.

La fonction de lissage $\tilde{V_g}$ utilisée dans (4.131) a pour but de faire converger la densité équivalente de porteurs p_{acc} vers zéro lorsque la tension de grille devient supérieure à la tension de bandes plates. Elle est donnée par :

$$\tilde{V_g} = V_g - \frac{1}{2} \cdot \left[V_g + \sqrt{(V_{fb} - V_g)^2 + 4 \cdot \kappa^2} + \sqrt{V_{fb}^2 + 4 \cdot \kappa^2} \right] \tag{4.132}$$

où κ est la même constante de lissage que celle utilisée dans l'équation de $V_{g[\mathrm{acc}]}$ (voir l'équation (4.127) du modèle classique, où $\kappa = 0.1$).

Grâce à ce concept de densité équivalente de porteurs et aux résultats de l'approximation du puits de potentiel triangulaire, nous pouvons alors décrire avec une précision satisfaisante le déplacement quantique de la bande de valence. Ce déplacement quantique s'exprime en fait comme un pseudo élargissement du gap :

$$E_{w[\mathrm{acc}]}(V_g) = \left(\frac{\hbar^2}{2 \cdot m_h^*} \right)^{1/3} \cdot \left[\frac{9}{8} \cdot \pi \cdot q \cdot F_{s[\mathrm{acc}]}(V_g) \right]^{2/3} \tag{4.133}$$

où m_h^* est la masse effective des trous — fixée à $0.16\, m_0$ dans notre modèle — et $F_{s[\mathrm{acc}]}$ le

champ électrique moyen normal à l'interface Si–SiO_2, donné par :

$$F_{s[\mathrm{acc}]}(V_g) = \frac{q \cdot p_{acc}(\tilde{V_g})}{2 \cdot \epsilon_{si}} \tag{4.134}$$

Finalement, à partir de (4.133), le pseudo élargissement du gap $E_{w[\mathrm{acc}]}$ est exprimé en termes de potentiel de surface par la fonction $\delta\phi_{s[\mathrm{acc}]}$, définie comme :

$$\delta\phi_{s[\mathrm{acc}]}(V_g) = E_{w[\mathrm{acc}]}(V_g)/q \tag{4.135}$$

Cette expression nous fournit donc une relation explicite entre le décrément quantique du potentiel de surface en accumulation et la tension de grille appliquée au dispositif. L'évaluation de $\delta\phi_{s[\mathrm{acc}]}$ ne nécessitant pas la connaissance préalable du potentiel de surface classique, elle n'est absolument pas coûteuse en temps de calcul. De plus, son intégration au sein du modèle classique de potentiel de surface est immédiate, et peut être réalisée sans paramètre additionnel ni fonction de lissage supplémentaire.

En fait, il suffit de combiner l'expression classique du potentiel de surface $\phi_{s[\mathrm{acc}\to\mathrm{wi}]}$ (cf. (4.130)) avec la fonction $\delta\phi_{s[\mathrm{acc}]}$ (cf. (4.135)) pour obtenir une description continue, analytique et quantique du potentiel de surface. Ce nouveau modèle, valable de l'accumulation au début de l'inversion faible est finalement donné par :

$$\phi^{\mathrm{qm}}_{s[\mathrm{acc}\to\mathrm{wi}]}(V_g) = \phi_{s[\mathrm{acc}\to\mathrm{wi}]}(V_g) - \delta\phi_{s[\mathrm{acc}]}(V_g) \tag{4.136}$$

4.9 Modèle explicite quantique complet

Nous avons maintenant à notre disposition deux expressions nous permettant de définir le potentiel de surface sur toute la plage de fonctionnement du TMOS, de l'accumulation à l'inversion forte. La dernière étape de ce travail consiste donc à unifier ces deux sous-modèles en un modèle global.

D'une part, une expression valide de l'accumulation au début de l'inversion faible est donnée par (4.136). Dans le but de simplifier l'écriture du modèle final, réécrivons cette équation sous la forme :

$$\phi_{s1}(V_g) = \phi^{\mathrm{qm}}_{s[\mathrm{acc}\to\mathrm{wi}]}(V_g) \tag{4.137}$$

D'autre part, une expression valide de la déplétion à l'inversion forte est donnée par (4.92).

Toujours dans un souci de clarté, il est préférable de réécrire cette équation comme :

$$\phi_{s2}(V_g, V_{ch}) = \phi_{s[\mathrm{qm}]}(V_g, V_{ch}) \tag{4.138}$$

La combinaison de ces équations fournit alors une expression très pratique et précise du potentiel de surface, valide quel que soit le régime de fonctionnement du TMOS, et complètement dépendante de toutes les tensions appliquées au dispositif. Cette description analytique complète du potentiel de surface quantique est finalement donnée par :

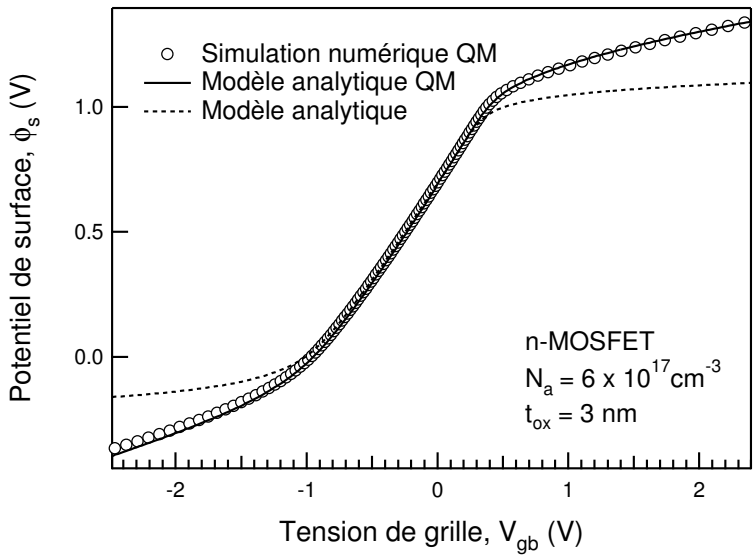
$$\phi_s(V_g, V_{ch}) = \begin{cases} \phi_{s1}(V_g) & \text{pour } V_g < V_0, \\ \phi_{s2}(V_g, V_{ch}) & \text{pour } V_g \geqslant V_0. \end{cases} \tag{4.139}$$

La tension de transition V_0 dans (4.139) est choisie comme étant égale à zéro. En fait, il est possible de choisir n'importe quelle valeur comprise entre -0.1 et 0.1 V pour la détermination de V_0. Suite à de nombreux essais, nous avons trouvé que la valeur $V_0 = 0$ était un bon choix, non seulement pour la précision du modèle, mais aussi pour sa simplicité. Nous avons vérifié la validité de ce choix pour un large éventail de dopages substrat allant de 2×10^{17} à 3×10^{18} cm^{-3}, couplés à différentes épaisseurs d'oxyde de grille comprises entre 1.7 et 5 nm. Dans tous les cas, l'erreur relative $\frac{|\phi_{s2}(V_0)-\phi_{s1}(V_0)|}{\min(\phi_{s1},\phi_{s2})}$ est inférieure à 0.6 %, ce qui confirme la cohérence de notre approche. Pour prouver plus concrètement la validité et l'intérêt de cette modélisation par parties, nous allons présenter les résultats obtenus avec (4.139) dans la section suivante.

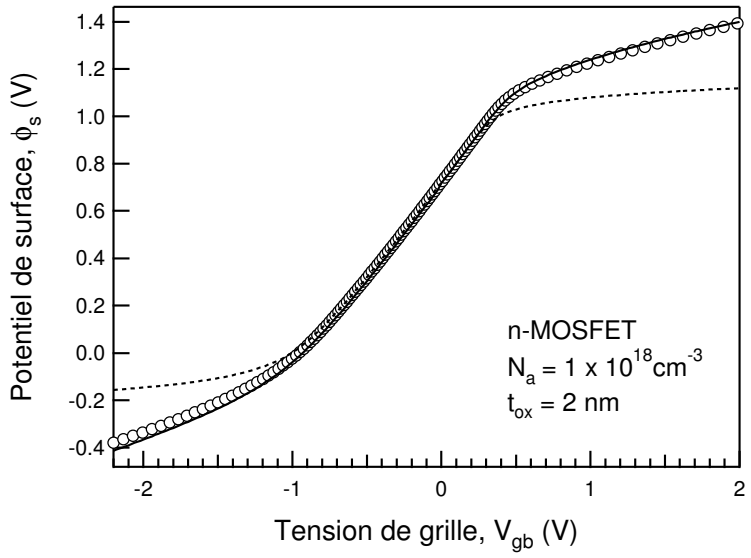
4.10 Validation du modèle complet

Une première validation du modèle final est réalisée par une série de comparaisons avec des simulations numériques Schrödinger–Poisson. Ces comparaisons portent tout d'abord sur le potentiel de surface lui-même. Les Figs. 4.26a et 4.26b montrent le potentiel de surface ϕ_s calculé pour des transistors n-MOS ayant des caractéristiques technologiques typiques des générations actuelles et futures. Les effets quantiques ont un impact fort sur les deux régions d'accumulation et d'inversion, et sont d'autant plus importants que le dopage substrat augmente et que l'épaisseur d'oxyde diminue. Comme le montrent clairement ces figures, notre modèle décrit avec une bonne précision le potentiel de surface sur toute la dynamique de tensions de grille. Notons quand même que la région d'accumulation est légèrement moins bien prédite que celle d'inversion, ce qui peut s'expliquer par le caractère plus empirique de l'approche développée dans cette partie du modèle. Cette précision est néanmoins réellement suffisante

dans le contexte d'un modèle compact, ce que nous allons démontrer tout de suite.



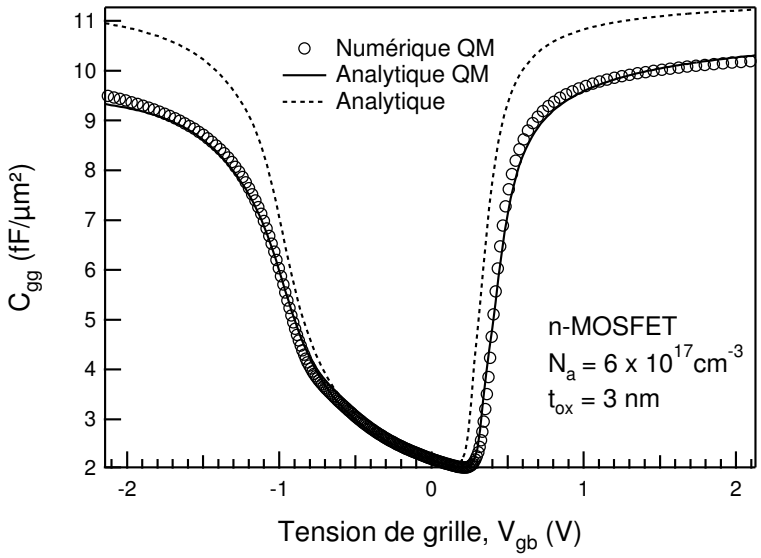
(a) Technologie CMOS actuelle.



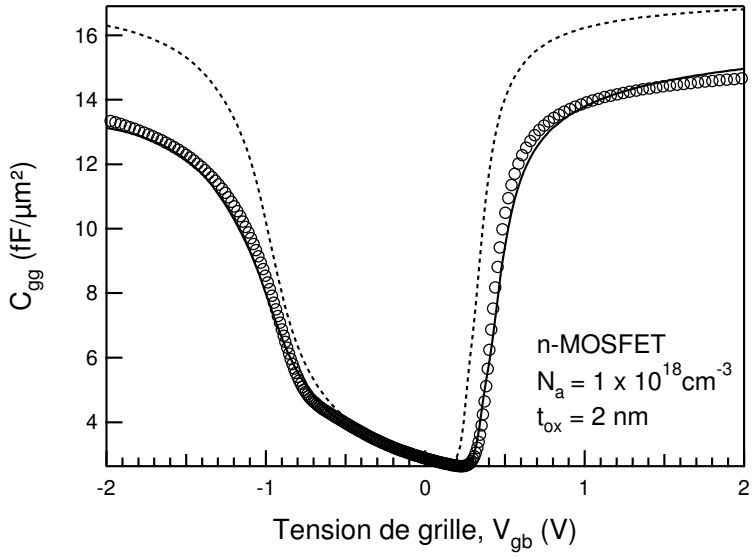
(b) Technologie CMOS avancée.

FIG. 4.26 : *Comparaison entre le calcul numérique (cercles), analytique quantique (ligne continue) et analytique classique (pointillés) du potentiel de surface ϕ_s.*

Les Figs. 4.27a et 4.27b présentent des simulations de la transcapacité de grille C_{gg} en fonction de la tension de grille. Il apparaît que cette capacité est modélisée précisément dans tous les régimes de fonctionnement, et que l'importance des effets quantiques en accumulation est aussi importante qu'en inversion. Le décalage de la tension de seuil est aussi clairement mis en évidence sur les deux figures, et correctement pris en compte par notre modèle. Il



(a) Technologie CMOS actuelle.



(b) Technologie CMOS avancée.

FIG. 4.27 : *Comparaison entre le calcul numérique (cercles), analytique quantique (ligne continue) et analytique classique (pointillés) de la transcapacité de grille C_{gg} pour des TMOS de différentes technologies. Dans ces simulations, l'effet de polydéplétion n'est pas pris en compte, i.e. $\phi_p = O$ dans l'expression de la densité de charge de grille Q_g.*

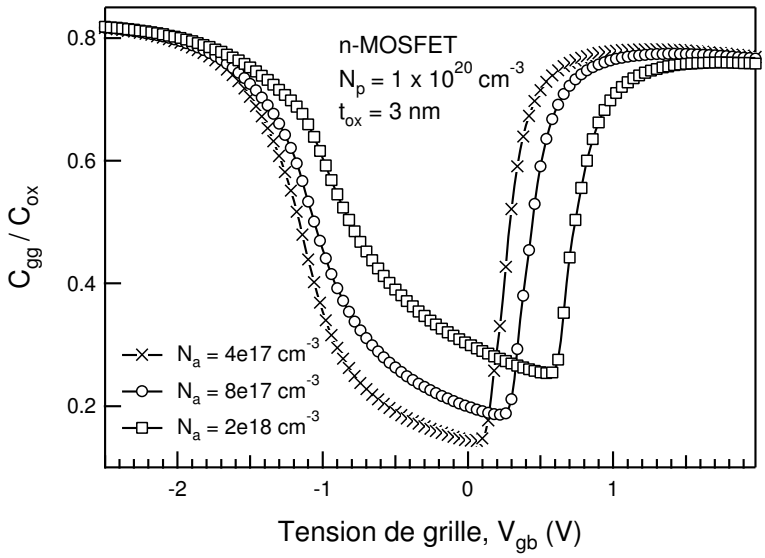
est important de préciser que tous les paramètres d'entrée du modèle ($t_{ox}, N_a, V_{fb} \ldots$) sont rigoureusement les mêmes que ceux utilisés lors des simulations auto-cohérentes Schrödinger–Poisson. De plus, les différentes constantes de lissage du modèle (ε, κ) n'ont en aucun cas servi de paramètres d'ajustement ; quelles que soient les simulations effectuées, leurs valeurs restent respectivement fixées à 0.15 et 0.1.

D'un point de vue efficacité de calcul, il est intéressant de constater que la résolution numérique de la transcapacité de grille requiert plusieurs minutes, tandis que la même simulation réalisée avec notre modèle analytique — à nombre de points simulés identique — ne dure que quelques secondes.

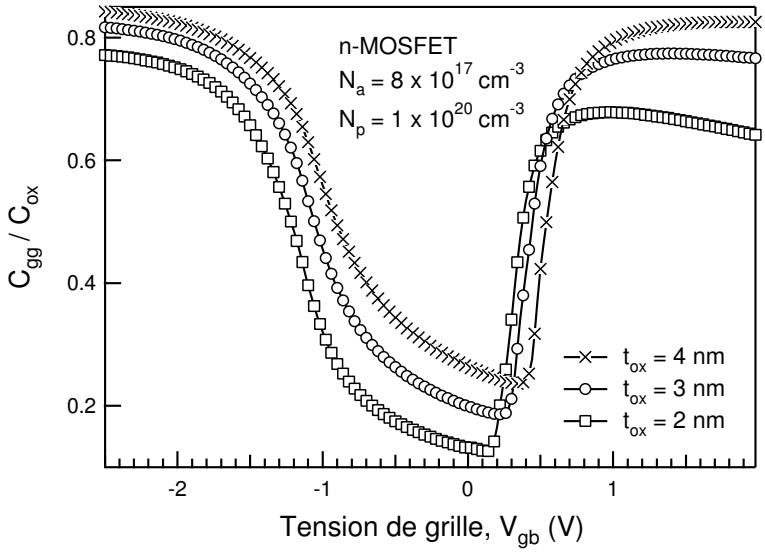
Avant de présenter notre seconde validation, nous allons nous intéresser au comportement général du modèle en tant que fonction des principaux paramètres technologiques responsables des effets quantiques. L'objectif des simulations présentées à la Fig. 4.28 est de mettre en évidence le bon comportement du modèle capacitif autour des zones critiques que sont la tension de bandes plates et la région d'inversion modérée. En particulier la zone correspondant à la tension de bandes plates est bien décrite (sans discontinuité), ce qui n'est pas toujours le cas dans les modèles quantiques globaux (i.e. valable de l'accumulation à l'inversion) [54,55].

En dépit des bons résultats présentés jusqu'ici, notre nouveau modèle n'est cependant pas exempt de toute critique. Bien qu'il soit proche de la physique du semi-conducteur, ce n'est pas au sens propre un modèle physique, et en conséquence il présente quelques limitations. En fait, notre modèle est capable de prédire correctement toutes les caractéristiques électriques du TMOS pour une large gamme de valeurs de dopages substrat (de 10^{17} à 3×10^{18} cm^{-3}) et d'épaisseurs d'oxyde (de 1.7 à 6 nm), *mais à condition de respecter la logique associée aux lois de réduction d'échelle*. Concrètement cela signifie que si l'on considère une épaisseur d'oxyde de 3 nm, le modèle permet d'obtenir des simulations très précises si $4 \times 10^{17} \leqslant N_a \leqslant 8 \times 10^{17}$ cm^{-3}, précises si $2 \times 10^{17} \leqslant N_a \leqslant 10^{18}$ cm^{-3} mais pas au-delà. Par exemple, le modèle ne fournira pas des résultats corrects si l'on souhaite simuler un TMOS avec une épaisseur d'oxyde de 2 nm et un dopage substrat de 5×10^{16} cm^{-3}. Dans ce cas, le potentiel de surface sera surestimé en inversion, ceci résultant directement des hypothèses liées au concept d'approximation de l'inversion modérée. Une solution envisageable pour améliorer l'approximation de l'inversion modérée pourrait être de faire intervenir explicitement le dopage substrat dans son expression. Il serait alors possible d'obtenir un véritable modèle physique, du moins pour la modélisation de l'inversion.

Cette parenthèse sur les limitations du modèle étant refermée, nous allons maintenant achever sa validation. Les comparaisons de la Fig. 4.27 ont précédemment illustré le bon accord entre notre modèle et les calculs numériques. Une telle validation n'est cependant pas suffisante, en particulier pour la région d'accumulation. Dans une récente publication, RICHTER et al. ont



(a) N_a variable, t_{ox} fixée.



(b) N_a fixé, t_{ox} variable.

FIG. 4.28 : *Simulations de la transcapacité de grille C_{gg} en fonction de la tension de grille V_{gb} pour (a) différents dopages substrat N_a et (b) différentes épaisseurs d'oxyde t_{ox}. La polydéplétion est aussi incluse dans ces simulations.*

comparé les courbes *C–V* obtenues avec un lot représentatif de différents simulateurs quantiques numériques [19]. Les conclusions de cet article sont très instructives :

- En inversion, tous les simulateurs convergent vers les mêmes valeurs de transcapacité de grille et de niveau de seuil d'inversion.
- En accumulation, des écarts très significatifs (20 % de la valeur maximale de la transcapacité de grille) sont observés entre les différents résolveurs Schrödinger–Poisson.

En conséquence, une validation définitive du modèle impose de le comparer à des résultats expérimentaux. Ayant déjà validé la partie du modèle dédiée à l'inversion (comparaisons avec plusieurs jeux de mesures *I–V* et *C–V*), nous souhaitons maintenant vérifier les résultats du modèle complet. Pour ce faire il faut forcément procéder à une comparaison avec des mesures *C–V*, et plus précisément avec des mesures de la transcapacité de grille.

La Fig. 4.29 montre la comparaison entre notre modèle et des résultats expérimentaux provenant d'une technologie Philips 0.18 μm. Sans utiliser aucun paramètre d'ajustement, le modèle de potentiel de surface quantique mène à d'excellents résultats vis à vis de cette technologie. En particulier et à l'instar de toutes les autres simulations réalisées avec notre modèle, l'épaisseur d'oxyde de grille n'a pas été considérée comme un paramètre d'ajustement, uniquement l'épaisseur technologique (réelle) a été employée dans nos simulations. En conclusion, le bon accord observé à la Fig. 4.29 apporte la preuve — en complément des validations précédentes — de la validité des concepts développés au cours de la réalisation de ce nouveau modèle compact.

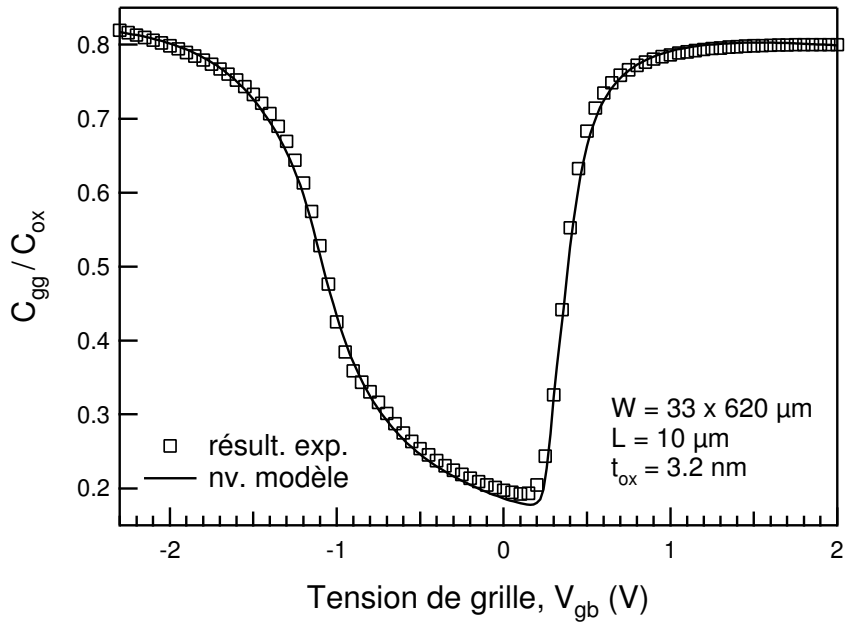


FIG. 4.29 : *Transcapacité de grille normalisée par rapport à la capacité d'oxyde de grille, en fonction de la tension de grille. Les paramètres physiques de cette technologie Philips 0.18 μm sont les suivants : épaisseur d'oxyde $t_{ox} = 3.2$ nm, dopage substrat $N_a = 6 \times 10^{17}$ cm^{-3} et dopage polysilicium $N_p = 1.2 \times 10^{20}$ cm^{-3}.*

4.11 Conclusion

Tout au long de ce chapitre, il a clairement été démontré que les effets quantiques perturbent de façon significative le comportement électrique des transistors MOS. Plus problématique encore est que leur influence va fortement s'accroître avec l'arrivée des prochaines générations de TMOS. En effet, la constante réduction d'échelle des dispositifs impose une réduction drastique des épaisseurs d'oxyde de grille, tandis qu'en parallèle les niveaux de dopage substrat augmentent. À moyen terme, une telle situation risque de devenir préoccupante, car dans le cadre de cette loi de réduction d'échelle, les tensions d'alimentation sont aussi réduites. Or, une des conséquences des effets quantiques étant l'augmentation de la tension de seuil, cela signifie que si la tension d'alimentation diminue, la plage d'utilisation du transistor sera forcément réduite. Dès lors, l'intérêt de tels dispositifs dans le domaine de la conception analogique risque de s'en trouver quelque peu limité. Cependant, étant données les contraintes industrielles et commerciales, c'est l'utilisation du TMOS en conception numérique qui dicte les évolutions futures. En résumé, ce sera au monde de l'analogique de s'adapter à cette nouvelle donne technologique.

C'est dans ce contexte précis que réside l'intérêt de ce chapitre majeur. En effet, qui dit transfert de technologie de la conception numérique à la conception analogique, dit aussi adaptation des modèles compacts à ces nouvelles exigences. En conséquence, un modèle compact performant devra non seulement être capable de simuler des circuits numériques mais aussi des circuits analogiques, en d'autres termes il devra pouvoir simuler correctement des circuits mixtes. La simulation analogique est cependant plus complexe que la simulation digitale, puisque dans ce cas, les caractéristiques électriques du TMOS doivent être correctement prédites sur toute la dynamique usuelle de fonctionnement, c'est-à-dire de l'inversion faible à l'inversion forte. En particulier la modélisation correcte de la région d'inversion modérée devient très importante, en raison de la baisse des tensions d'alimentation. En outre, au-delà des impératifs liés aux tensions d'alimentation, la région d'inversion modérée est très intéressante d'un point de vue conception analogique, car elle permet d'obtenir un compromis idéal entre une valeur élevée de transconductance g_m et un niveau de courant de drain raisonnable. Un modèle compact permettant de modéliser précisément cette région tend donc à devenir indispensable, ce qui explique l'intérêt croissant des concepteurs analogiciens pour des modèles de dernières générations tels que EKV3.0, MM11 ou SP, au détriment des modèles dits en tension de seuil, tels que MM9 ou BSIM4.

L'inclusion des effets quantiques en modélisation compacte a originellement consisté en la définition de corrections quantiques. Les travaux de VAN DORT ont tout d'abord mis en évidence le fort impact de ces effets sur la tension de seuil du TMOS. Différents modèles prenant en compte l'augmentation quantique de la tension de seuil ont été présentés et testés dans ce chapitre. Finalement, l'analyse critique des résultats obtenus avec ces modèles nous a conduit à reformuler un modèle à la fois plus simple, et plus efficace. Dans le but de vérifier la validité des différents modèles, nous avons développé une méthode d'extraction de la tension de seuil, basée sur la connaissance des caractéristiques C–V du dispositif[1]. Cette méthode d'extraction a été validée par comparaison avec des résultats expérimentaux de VAN DORT [9].

La seule considération de l'augmentation de la tension de seuil ne permet cependant qu'une prise en compte partielle des effets quantiques sur les caractéristiques électriques du TMOS. Si la région d'inversion faible est effectivement bien modélisée par le décalage du niveau du seuil, celles d'inversion modérée et forte ne le sont pas du tout. Pour résoudre ce problème, une seconde correction quantique, la capacité d'oxyde effective, est généralement utilisée. Cette capacité est définie comme une fonction d'une épaisseur d'oxyde effective, elle-même dépendante de la polarisation de grille du TMOS. L'emploi conjoint de ces deux corrections quantiques peut alors permettre — au sein d'un modèle compact formulé en tension de seuil — une prise en compte complète des effets quantiques, de l'inversion faible à l'inversion forte. Cependant, à notre sens l'intérêt d'un tel modèle est limité, puisque dans ce type d'approche la région d'inversion modérée est décrite de façon purement empirique. C'est pourquoi nous nous sommes intéressés au développement d'un nouveau modèle, quantique et analytique, formulé intégralement en potentiel de surface.

Notre objectif initial était de proposer un modèle complètement analytique, et prenant en compte les effets quantiques de manière intrinsèque. En d'autres mots, nous souhaitions obtenir un modèle de potentiel de surface dans lequel les effets quantiques ne seraient pas vus comme une simple correction se greffant sur le modèle classique. C'est par exemple ce qui est fait dans MM11, où les effets quantiques sont inclus sous la forme d'une capacité d'oxyde effective, intervenant au niveau du calcul des charges. Ainsi, bien que ce modèle soit entièrement écrit en potentiel de surface, il ne permet pas d'obtenir la valeur réelle — tenant compte des effets quantiques — du potentiel de surface. Aussi, en dépit de l'efficacité certaine de MM11, la cohérence de cette approche est in fine discutable.

[1] Cette méthode est présentée en Annexe, page 175.

Le modèle développé dans ce manuscrit permet quant à lui de calculer la valeur réelle du potentiel de surface. Il repose sur une approximation analytique classique du potentiel de surface, qui correspond en fait au noyau du modèle MM11. Nous avons modifié ce modèle classique de l'intérieur, dans le but d'y inclure les effets quantiques de façon transparente, c'est-à-dire en ne modifiant aucun paramètre physique[1] du transistor MOS.

Ce nouveau modèle est en fait constitué de deux sous-modèles, l'un destiné à modéliser les effets quantiques en inversion, l'autre en accumulation. La modélisation de l'inversion est basée sur un nouveau concept (l'approximation de l'inversion modérée) lié à la définition d'une densité équivalente de porteurs minoritaires. Ce modèle autorise une modélisation physique de l'augmentation quantique du potentiel de surface en inversion ; il ne dépend d'aucune constante empirique, ni d'aucun paramètre d'ajustement. En accumulation, en raison de la structure complexe de la bande de valence, il est nettement plus difficile de réaliser un modèle à la fois simple, analytique et physiquement rigoureux. Nous avons contourné ce problème en définissant une densité équivalente de porteurs majoritaires, modélisant de façon semi-empirique le nombre de porteurs libres se trouvant dans la couche d'accumulation quantifiée. Grâce à l'utilisation conjointe de cette densité équivalente de porteurs majoritaires et de l'approximation du puits de potentiel triangulaire, nous pouvons décrire avec une précision satisfaisante les effets quantiques dans ce régime de fonctionnement. La réunion de ces deux sous-modèles en un modèle unique est immédiate, et ne nécessite aucune manipulation mathématique particulière.

En conclusion, un modèle analytique de potentiel de surface incluant de manière explicite les effets quantiques vient d'être présenté. Ce modèle est continu de l'accumulation à l'inversion et ne dépend d'aucun paramètre d'ajustement. Il permet une évaluation précise du potentiel de surface et de ses dérivées, pour une large gamme de dopages substrat et d'épaisseurs d'oxyde de grille. Intégré à un modèle en feuille de charge, il entraîne une simulation précise des différentes caractéristiques électriques du TMOS, telles que les charges, le courant de drain et les transcapacités. Enfin, les nombreuses comparaisons entre ce nouveau modèle, les simulations Schrödinger–Poisson et les résultats expérimentaux ont objectivement démontré la validité de notre approche de modélisation.

[1] Épaisseur d'oxyde de grille, dopage substrat, concentration intrinsèque de porteurs, etc.

Bibliographie

[1] F. Stern and W. E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit," *Phys. Rev.*, vol. 163, pp. 816–835, 1967.

[2] F. Stern, "Quantum properties of surface space-charge layers," *CRC Crti. Rev. Solid State Sci.*, vol. 4, pp. 499–514, 1974.

[3] T. Ando, A. B. Fowler, and F. Stern, "Electronic proporties of two-dimensional systems," *Rev. Mod. Phys.*, vol. 54, pp. 437–672, 1982.

[4] R. Rios, N. Arora, C.-L. Huang, *et al.*, "A physical compact MOSFET model, including quantum mechanical effects, for statistical circuit design applications," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, 1995, pp. 937–940.

[5] T. Janik and B. Majkusiak, "Analysis of the MOS transistor based on the self-consistent solution to the Schrödinger and poisson equations and on the local mobility model," *IEEE Trans. Electron Devices*, vol. 45, no. 6, pp. 1263–1271, June 1998.

[6] G. Gildenblat, T. L. Chen, and P. Bendix, "Closed-form approximation for the perturbation of MOSFET surface potential by quantum-mechanical-effects," *Electronics Letters*, vol. 36, no. 12, pp. 1072–1073, June 2000.

[7] Y. Ma, L. Liu, W. Deng, L. Tian, Z. Li, and Z. Yu, "A new charge model including quantum mechanical effects in MOS structure inversion layer," *Solid-State Electron.*, vol. 44, pp. 1697–1702, 2000.

[8] R. Clerc, P. O'Sullivan, K. G. McCarthy, G. Ghibaudo, *et al.*, "A physical compact model for direct tunneling from NMOS inversion layers," *Solid-State Electron.*, vol. 45, pp. 1705–1716, 2001.

[9] M. J. van Dort, P. H. Woerlee, A. J. Walker, *et al.*, "Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 39, no. 4, pp. 932–938, Apr. 1992.

[10] M. J. van Dort, P. H. Woerlee, and A. J. Walker, "A simple model for quantisation effects in heavily-doped silicon MOSFETs at inversion conditions," *Solid-State Electron.*, vol. 37, no. 3, pp. 411–414, 1994.

[11] J. Suñé, P. Olivo, and B. Riccò, "Quantum-mechanical modeling of accumulation layers in MOS structure," *IEEE Trans. Electron Devices*, vol. 39, no. 7, pp. 1732–1738, July 1992.

[12] G. Ghibaudo, S. Bruyère, T. Devoivre, B. DeSalvo, and E. Vincent, "Improved method for the oxide tickness extraction in MOS structures with ultrathin gate dielectrics," *IEEE Trans. Semiconductor Manufacturing*, vol. 13, no. 2, pp. 152–158, May 2000.

[13] A. S. Spinelli, A. Benvenuti, and A. Pacelli, "Self-consistent 2-D model for quantum effects in n-MOS transistors," *IEEE Trans. Electron Devices*, vol. 45, no. 6, pp. 1342–1349, June 1998.

[14] R. Clerc, "Etudes des effets quantiques dans les composants CMOS à oxydes de grille ultra minces – modélisation et caractérisation," Ph.D. dissertation, INPG, nov 2001.

[15] J. A. López-Villanueva, P. Cartujo-Casinello, J. Banqueri, *et al.*, "Effects on the inversion layer centroid on MOSFET behavior," *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1915–1922, Nov. 1997.

[16] *Quantum-Mechanical CV Simulator*, UC Berkeley Device Group, Apr. 2000. [Online]. Available : http ://www-device.eecs.berkeley.edu/research/qmcv/qmcv.html

[17] D. K. Blanks *et al.*, "NEMO : General release of a new comprehensive quantum device simulator," in *Proc. IEEE 24th Int. Symp. Compound Semiconductors*, 1998, pp. 639–642.

[18] W. K. Henson, K. Z. Ahmed, E. M. Vogel, J. R. Hauser, J. J. Wortman, R. D. Venables, M. Xu, and D. Venables, "Estimating oxide thickness of tunnel oxides down to 1.4 nm using conventional capacitance-voltage measurements on MOS capacitors," *IEEE Electron Device Lett.*, vol. 20, pp. 179–181, 1999.

[19] C. A. Richter, A. R. Hefner, and E. M. Vogel, "A comparison of quantum-mechanical capacitance-voltage simulators," *IEEE Electron Device Lett.*, vol. 22, no. 1, pp. 35–37, Jan. 2001.

[20] F. Stern, "Self-consistent results for n-type Si inversion layers," *Phys. Rev. B*, vol. 5, no. 12, pp. 4891–4899, June 1972.

[21] S. A. Hareland, M. Manassian, W. K. Shih, S. Jallepalli, *et al.*, "Computationally efficient models for quantization effects in MOS electron and hole accumulation layers," *IEEE Trans. Electron Devices*, vol. 45, no. 7, pp. 1487–1493, July 1998.

[22] H. Mathieu, *Physique des semiconducteurs et des composants électroniques*. Masson, 1998, ISBN 2-225-83151-3.

[23] Y. Ma, Z. Li, L. Liu, L. Tian, and Z. Yu, "Effective density-of-states approach to QM correction in MOS structures," *Solid-State Electron.*, vol. 44, pp. 401–407, 2000.

[24] Y. Ma, Z. Li, L. Liu, and Z. Yu, "Comprehensive analytical physical model of quantized inversion layer in MOS structure," *Solid-State Electron.*, vol. 45, pp. 267–273, 2001.

[25] Y. Ma, L. Liu, Z. Yu, and Z. Li, "Analytical charge-control and I–V model for submicrometer and deep-submicrometer MOSFETs fully comprising quantum mechanical effects," *IEEE Trans. Computer-Aided Design*, vol. 20, no. 4, pp. 495–502, Apr. 2001.

[26] X. Liu, J. Kang, X. Guan, R. Han, and Y. Wang, "The influence of tunneling effect and inversion layer quantization effect on threshold voltage of deep submicron MOSFETs," *Solid-State Electron.*, vol. 44, pp. 1435–1439, 2000.

[27] C. Lallement, J.-M. Sallese, M. Bucher, W. Grabinski, and P. Fazan, "Accounting for quantum effects and polysilicon depletion from weak to strong inversion, in a charge based design-oriented MOSFET model," *IEEE Trans. Electron Devices*, vol. 50, no. 2, pp. 406–417, Feb. 2003.

[28] J. W. Slotboom and H. C. de Graaf, "Measurements of bandgap narrowing in Si bipolar transistors," *Solid-State Electron.*, vol. 19, pp. 857–862, 1976.

[29] S. A. Hareland, S. Krishnamurthy, S. Jallepalli, *et al.*, "A computationally efficient model for inversion layer quantization effects in deep-submicron n-channel MOSFETs," *IEEE Trans. Electron Devices*, vol. 43, no. 1, pp. 90–96, Jan. 1996.

[30] M. Bucher, J.-M. Sallese, and C. Lallement, *Proc. IEEE Int. Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*. New York : D. Tsoukalas, C. Tsamis, Springer Vienna, pp. 296–299, 2001, ISBN 3-211-83708-6.

[31] D. B. M. Klaassen, R. van Langevelde, A. J. Scholten, *et al.*, *The MOS model, level 1101*, Philips Research Laboratories, 2002. [Online]. Available : http ://www.semiconductors.philips.com/philips_models/mos_models/model11

[32] *BSIM 4.2 Manual*, Univ. of California, Berkeley, Apr. 2001. [Online]. Available : http ://www-device.eecs.berkeley.edu/˜bsim

[33] G. Gildenblat and T. L. Chen, "Overview of an advanced surface-potential-based MOSFET model (SP)," in *Proc. 5th Int. Conf. on Modeling and Simulation of Microsystems (MSM)*, 2002.

[34] R. van Langevelde, A. J. Scholten, and R. J. Havens, "Advanced compact MOS modelling," in *Proc. ESSDERC*, 2001.

[35] T. L. Chen and G. Gildenblat, "Analytical approximation for the MOSFET surface potential," *Solid-State Electron.*, vol. 45, pp. 335–339, 2001.

[36] F. Prégaldiny, C. Lallement, and D. Mathiot, "Quantum surface potential model suitable for advanced MOSFETs simulation," in *Proc. IEEE Int. Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sept. 2003, pp. 227–230.

[37] F. Prégaldiny, C. Lallement, R. van Langevelde, and D. Mathiot, "An advanced explicit surface potential model physically accounting for the quantization effects in deep-submicron MOSFETs," *Solid-State Electron.*, vol. 48, no. 3, pp. 427–435, Mar. 2004.

[38] R. van Langevelde *et al.*, "Physical background of MOS model 11," unclassified Report, Philips Electronics N.V.

[39] J.-M. Sallese, M. Bucher, F. Krummenacher, and P. Fazan, "Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model," *Solid-State Electron.*, vol. 47, no. 4, pp. 677–683, Apr. 2003.

[40] R. Clerc, T. Devoivre, G. Ghibaudo, *et al.*, "Capacitance-voltage (C–V) characterization of 20 å thick gate oxide : parameter extraction and modeling," *Microelectronics Reliability*, vol. 40, pp. 571–575, 2000.

[41] R. van Langevelde and F. M. Klaassen, "An explicit surface-potential-based MOSFET model for circuit simulation," *Solid-State Electron.*, vol. 44, pp. 409–418, 2000.

[42] W. Liu *et al.*, *BSIM3v3.2 MOSFET model – Users' manual*, Univ. of California, Berkeley, 1998. [Online]. Available : http ://www-device.eecs.berkeley.edu/˜bsim

[43] *MM9 Model*, Philips Semiconductors, 2001. [Online]. Available : http ://www.semiconductors.philips.com/philips_models/mos_models

[44] N. Arora, *MOSFET Models for VLSI Circuit Simulation. Theory and Practice*. New York : Springer-Verlag, 1993, ISBN 3-211-82395-6.

[45] R. van Langevelde and F. M. Klaassen, "Effect of gate-field fependent mobility degradation on distortion analysis in MOSFET's," *IEEE Trans. Electron Devices*, vol. 44, pp. 2044–2052, 1997.

[46] K. Remashan, N. A. Wong, K. Chan, S. P. Sim, and C. Y. Yang, "Modeling inversion-layer carrier mobilities in all regions of MOSFET operation," *Solid-State Electron.*, vol. 46, pp. 153–156, 2002.

[47] M. Bucher, C. Lallement, C. C. Enz, F. Théodoloz, and F. Krummenacher, "The EPFL–EKV MOSFET model, version 2.6," EPFL, Tech. Rep., 1999. [Online]. Available : http ://legwww.epfl.ch/ekv/

[48] M. Bucher, "Analytical MOS transistor modelling for analog circuit simulation," Ph.D. dissertation, EPFL, 1999, thèse no. 2114.

[49] C. Caillat, "Etude des dispositifs MOS de longueur de grille inférieure à 0.1 μm, vers les limites de l'intégration du CMOS sur silicium," Ph.D. dissertation, INPG, nov 1999.

[50] D. Foty, *MOSFET Modeling with Spice. Principles and Practice*. Prentice Hall, 1997, ISBN 0-13-227935-5.

[51] F. Prégaldiny, C. Lallement, W. Grabinski, J.-B. Kammerer, and D. Mathiot, "An analytical quantum model for the surface potential of deep-submicron MOSFETs," in *Proc. 10th International Conference on Mixed Design of Integrated Circuits and Systems (MIXDES)*, June 2003, pp. 98–104.

[52] M. I. Vexler, "A simple quantum model for the MOS structure in accumulation mode," *Solid-State Electron.*, vol. 47, no. 8, pp. 1283–1287, Aug. 2003.

[53] F. Prégaldiny, C. Lallement, and D. Mathiot, "Accounting for quantum mechanical effects from accumulation to inversion, in a fully analytical surface-potential-based MOSFET model," accepted in Solid-State Electronics (Dec. 2003).

[54] V. I. Kol'dyaev, G. van den Bosch, and L. Deferm, "A comprehensive closed-form model for the quantized accumulation layer in MOS structures," *Solid-State Electron.*, vol. 42, no. 1, pp. 49–56, 1998.

[55] S. i. Saito, K. Torii, M. Hiratani, and T. Onai, "Analytical quantum mechanical model for accumulation capacitance of MOS structures," *IEEE Electron Device Lett.*, vol. 23, no. 6, pp. 348–350, June 2002.

Chapitre 5

Conclusion et perspectives

5.1 Conclusion

Ce manuscrit fait état d'une étude approfondie de différents effets perturbant le fonctionnement conventionnel des dispositifs MOSFET des générations actuelles et futures. En particulier nous nous sommes intéressés à la description et à la modélisation, d'une part, des capacités parasites en régime extrinsèque et, d'autre part, des effets quantiques au sein des TMOS de technologies avancées. L'objectif majeur de ce travail était de réaliser des modèles analytiques ayant des liens forts avec la physique du dispositif, dans le but d'améliorer les modèles compacts existants.

Une synthèse des principaux modèles compacts destinés à la simulation de circuits est présentée au début de ce manuscrit. Tout d'abord, le principe d'une modélisation physique idéale (pour un TMOS de grande géométrie) est explicité. Ce modèle physique repose sur une formulation implicite du potentiel de surface et n'est donc pas adapté à la simulation des circuits VLSI–ULSI. Son étude est cependant très instructive car elle définit les principes d'un modèle en feuille de charge et le principe de l'approche drift–diffusion pour modéliser le courant de drain. En outre, la connaissance et la compréhension d'un modèle physique rigoureux permet de vérifier la validité des hypothèses et approximations utilisées dans les modèles compacts usuels. Ce type de modèle est aussi très utile dans le domaine de la validation, en particulier lorsque des résultats expérimentaux ne sont pas disponibles.

Suite à un bref historique de la modélisation compacte, les caractéristiques majeures des modèles de troisième génération et de quelques approches alternatives ont été discutées. Les modèles *classiques* de troisième génération regroupent BSIM3, BSIM4 et MM9 entre autres.

Leur caractéristique commune repose sur la formulation d'une tension de seuil qui sépare le fonctionnement du TMOS à l'état 'on' en deux régions : l'inversion faible et l'inversion forte. L'inconvénient d'une telle modélisation est que la zone de transition, dite région d'inversion modérée, est uniquement décrite par une fonction de lissage. Le caractère prédictif d'un tel modèle est alors incertain. Comme nous l'avons vu à la conclusion du chapitre précédent, la région d'inversion modérée est une zone d'un grand intérêt, en particulier pour la conception analogique. De plus, elle tend à devenir importante aussi pour la conception digitale, en raison de la baisse drastique des tensions d'alimentation. En conséquence, un modèle décrivant de façon uniquement empirique cette région, n'est en tout état de cause, plus adapté à la modélisation rigoureuse des TMOS actuels. Si de tels modèles parviennent quand même à obtenir des résultats acceptables, c'est uniquement grâce à l'utilisation d'innombrables paramètres (~ 80 pour le noyau de BSIM4) qui in fine font définitivement perdre au modèle tout caractère physique.

Dans le but de renverser cette tendance à augmenter la complexité des modèles, de nouvelles approches tendant à accroître le contenu physique des modèles ont été développées. Pour ne citer que les modèles les plus connus, nous avons par ordre chronologique d'apparition : EKV, MM11 et SP. Le modèle EKV est basé sur le concept de linéarisation de la charge d'inversion tandis que les modèles MM11 et SP reposent sur une formulation explicite du potentiel de surface. Ces trois modèles sont par nature proches de la physique du TMOS, et de ce fait nécessitent beaucoup moins de paramètres que BSIM4 (~ 40 pour le noyau de MM11). Ils apparaissent donc naturellement comme les successeurs des modèles classiques de troisième génération, et à ce titre, ils méritent très certainement le qualificatif de modèles de quatrième génération tant leurs approches diffèrent de celles de BSIM3/4 ou MM9.

Le troisième chapitre est consacré à la réalisation d'un nouveau modèle de capacité extrinsèque, et cela indépendamment de toute considération relative au modèle de base (modèle capacitif intrinsèque). L'avantage d'un tel modèle est d'être très facilement implémentable dans n'importe quel type de modèle compact : en potentiel de surface, en linéarisation de la charge ou en tension de seuil. Le développement de notre modèle repose sur une étude approfondie des résultats obtenus lors des simulations 2D de TMOS d'une technologie 0.25 μm. L'influence de divers paramètres technologiques (longueur de grille, niveau de dopage des régions LDD) sur le comportement capacitif extrinsèque a été clarifiée. La compréhension physique de ces résultats de simulation nous a permis de développer un modèle basé sur la physique du TMOS,

et ayant un faible degré d'empirisme. En effet, un unique paramètre additionnel est nécessaire par rapport à un modèle intrinsèque conventionnel. Les caractéristiques essentielles de notre modèle extrinsèque consistent en une nouvelle formulation de la capacité d'overlap et de la capacité de bord interne. En particulier, pour la première fois en modélisation compacte, la dépendance de la capacité de bord interne à la polarisation de la grille est explicitée. Le modèle extrinsèque global (capacité d'overlap + capacité de bord interne + capacité de bord externe) est continu de l'accumulation à l'inversion, et ne présente aucun problème de convergence. Enfin, l'implémentation du modèle extrinsèque au sein d'un modèle compact quelconque étant immédiate, nous l'avons intégré au modèle EKV2.6. La formulation du modèle complet est relativement simple, et mène à d'excellents résultats en comparaison aux simulations 2D. Pour conclure, disons que l'intérêt d'avoir un modèle précis de capacités parasites est directement lié à la politique de réduction d'échelle du TMOS. En effet, plus les dimensions des dispositifs diminuent, plus le caractère extrinsèque devient important : dans le cas d'une technologie CMOS 0.18 μm, en inversion, la valeur de la capacité extrinsèque est quasiment équivalente à celle de la capacité intrinsèque. Les applications potentielles d'un tel modèle sont donc multiples et intéressent tout particulièrement la simulation des circuits digitaux (TMOS utilisé en commutation) et RF.

Le quatrième chapitre de ce manuscrit est consacré à l'étude et à la modélisation des effets quantiques au sein des transistors n-MOS. L'influence croissante des effets quantiques dans les composants CMOS fortement submicroniques constitue l'une des préoccupations majeures de la microélectronique d'aujourd'hui. L'impact du confinement quantique des porteurs (électrons et trous) se répercute sur l'ensemble du comportement électrique du TMOS et sur toutes ses grandeurs électriques fondamentales telles que le potentiel de surface, les charges, les transcapacités, le courant de drain, la tension de seuil, etc. Développer un modèle compact adapté à la description des nouvelles générations de TMOS est donc une réelle nécessité. S'il est vrai que les principaux modèles compacts existants (BSIM4, EKV3.0, MM11 et SP) prennent en compte les effets quantiques (du moins en inversion), l'intérêt de proposer un nouveau modèle est réel, comme cela a clairement été démontré au chapitre 4. Aussi, nous avons réalisé un nouveau modèle intégralement formulé en potentiel de surface qui est à la fois, analytique, quantique, simple et proche de la physique ; compromis qu'aucun modèle actuel ne réalisait jusqu'alors. Contrairement aux modèles BSIM4, EKV3.0 ou MM11, notre modèle ne fait intervenir aucune notion de capacité d'oxyde effective (BSIM4, MM11), de capacité d'oxyde équivalente (EKV3.0) ou de coefficient de substrat équivalent (EKV3.0). En

fait, aucun des paramètres du noyau du modèle classique n'est modifié avec notre approche, seul le potentiel de surface est traité de façon quantique. Sa connaissance rigoureuse permettant de calculer toutes les autres grandeurs électriques du TMOS, ce type d'approche est a priori prometteur. Le problème est que prendre en compte les effets quantiques directement en termes de potentiel de surface est très complexe, en particulier d'un point de vue mathématique. Le modèle SP était l'unique modèle — avant celui réalisé dans le cadre de cette thèse — à décrire analytiquement la valeur réelle du potentiel de surface sans utiliser les habituelles corrections quantiques citées précédemment. Cependant, son inconvénient majeur réside dans une transformation quantique extrêmement lourde et mathématiquement complexe du modèle classique, si bien que les expressions obtenues ne sont plus du tout parlantes en termes de paramètres technologiques et de tensions appliquées au dispositif. De plus, le conditionnement mathématique du modèle a nécessité de nombreuses approximations, ce qui in fine occulte l'indéniable base physique du modèle classique.

Le modèle quantique développé dans ce manuscrit repose sur une approximation analytique classique du potentiel de surface correspondant au noyau du modèle MM11. Nous avons modifié ce modèle en y incluant les effets quantiques de façon transparente, c'est-à-dire en ne modifiant aucun paramètre physique du TMOS. Contrairement à MM11, notre modèle permet de calculer la valeur réelle du potentiel de surface — i.e. tenant compte des effets quantiques — et ne requiert pas la définition d'une capacité d'oxyde effective. Dans notre approche, les effets quantiques ne sont pas vus comme une simple correction se greffant sur le modèle classique, mais au contraire font partie intégrante d'un modèle global, unifié. Une approche originale, développée dans le cadre de cette thèse consiste en une prise en compte séparée des effets quantiques en inversion et en accumulation. Les meilleures approximations analytiques ont été utilisées pour chaque région, à savoir l'approximation variationnelle pour la modélisation de l'inversion et l'approximation du puits de potentiel triangulaire pour la modélisation de l'accumulation. L'originalité réelle de ce travail est alors liée au développement de nouveaux concepts permettant d'exprimer simplement les variations quantiques (augmentation en inversion, réduction en accumulation) du potentiel de surface en fonction des paramètres technologiques et des tensions appliquées au TMOS. Ces concepts permettent de modéliser analytiquement l'élargissement quantique du gap du semi-conducteur en termes de potentiel de surface, et cela pour une polarisation externe donnée. Ce dernier point est fondamental ; en effet, s'il est très simple de transcrire l'élargissement du gap en termes de potentiel de surface pour une charge d'inversion donnée, le problème se complique sérieusement lorsque l'on se

place à une polarisation de grille fixée. L'intérêt des nouveaux concepts développés ici est justement d'apporter une solution à ce problème, en proposant un modèle simple, physique et compréhensible en termes de paramètres technologiques. Enfin, un point important est que notre modèle de potentiel de surface quantique ne nécessite d'une part, aucun paramètre d'ajustement et d'autre part, aucun paramètre additionnel par rapport à un modèle classique.

L'inclusion de ce nouveau modèle dans un modèle en feuille de charge permet une description précise des caractéristiques électriques essentielles du TMOS telles que les charges, les transcapacités, le courant de drain, etc. En particulier, nous avons expliqué comment reformuler de façon rigoureuse l'approche drift–diffusion (calcul du courant de drain) dans le but de déterminer précisément les deux courants de diffusion et de conduction, en tenant compte des effets quantiques simultanément sur le potentiel de surface et sur la densité de charge d'inversion. Le modèle quantique mis au point au cours de ce travail a finalement été validé par comparaison avec des simulations numériques (résolutions couplées Schrödinger–Poisson) et des résultats expérimentaux (I–V, C–V) de différentes technologies CMOS avancées. Précisons aussi que l'implémentation de ce modèle dans des langages de haut niveau tels que Verilog-AMS ou VHDL-AMS est immédiate et permet une simulation rapide et précise des caractéristiques électriques du TMOS.

En conclusion, l'intention réelle de cette thèse est de démontrer qu'il est possible de réaliser des modèles analytiques de TMOS à la fois simples, physiquement cohérents et ayant un degré d'empirisme minimum. Naturellement, le développement de tels modèles suppose au préalable une compréhension précise de la physique sous-jacente aux phénomènes étudiés. En effet, seule une compréhension claire des causes et conséquences des effets perturbant le fonctionnement classique du TMOS autorise la définition d'approximations analytiques viables. Augmenter le contenu physique des modèles compacts apparaît donc comme une solution efficace, aux avantages multiples tels que l'amélioration de la précision du modèle, la réduction du nombre de paramètres, la facilité d'extraction des paramètres, etc. En résumé, une approche pragmatique de la modélisation compacte est tout à fait envisageable et sera à coup sûr le maître mot des futures générations de modèles compacts destinés à la simulation de circuits.

5.2 Perspectives

Les modèles décrits dans cette thèse s'attachent à prendre en compte certains effets perturbant de façon significative le comportement idéal du transistor MOS. En aucun cas leur prétention n'est de se substituer aux modèles compacts existants, réaliser un tel modèle étant une tâche titanesque, nécessitant des moyens qui ne sont pas les nôtres. C'est pourquoi notre but consiste plutôt à ouvrir de nouvelles voies et à développer de nouveaux concepts permettant d'améliorer les modèles actuels, tels que EKV3.0 et MM11 par exemple.

En ce qui concerne notre modèle de potentiel de surface quantique, un certain nombre d'extensions devraient être réalisées. Dans sa forme actuelle, le modèle n'est valide que pour des transistors de grande géométrie, les effets canaux courts n'y étant pas inclus. De plus, le modèle drift–diffusion de courant de drain n'est pour l'instant valable qu'en régime linéaire. Une première évolution de ce modèle consistera donc à étendre sa zone de validité pour les fortes tensions de drain (source), i.e. en régime de saturation. L'inclusion d'effets canaux courts tels que la modulation de la longueur du canal, le DIBL, la réduction de la mobilité et la saturation de la vitesse des porteurs sera à moyen terme effectuée. En outre, l'étude de l'impact des effets quantiques sur la mobilité des porteurs fera l'objet d'une attention toute particulière.

À court terme, l'étude d'une reformulation complète de l'approximation de l'inversion modérée — i.e. le concept utilisé pour modéliser les effets quantiques en inversion — serait très intéressante. En particulier la redéfinition d'une nouvelle densité équivalente de porteurs prenant explicitement en compte la densité de charge fixe de déplétion, permettrait peut-être de réaliser un modèle encore plus proche de la physique. Dans ce cas, le modèle pourrait être valable pour tous les couples dopage substrat/épaisseur d'oxyde, sans que sa précision n'en soit altérée ; cela pourrait éventuellement étendre son utilité à d'autres domaines que la modélisation du TMOS. Enfin, la transposition du modèle de potentiel de surface quantique au cas des transistors p-MOS fera l'objet d'une prochaine étude.

En dernier lieu, l'utilisation conjointe de notre modèle de capacités parasites avec le modèle de potentiel de surface quantique constitue aussi une perspective intéressante. Mais plus encore, c'est le développement d'un nouveau modèle de capacité extrinsèque, en charge cette fois-ci, qui serait le plus utile. Au sein d'un tel modèle, la charge extrinsèque serait définie comme une fonction du potentiel de surface des zones d'extensions des source/drain (zones LDD), de telle sorte que sa modélisation ne nécessiterait alors plus aucun paramètre d'ajustement.

Annexe A

Extraction de la tension de seuil à partir des caractéristiques *C–V*

Une méthode très simple permettant de définir précisément la tension de seuil (quantique ou non) consiste à l'extraire d'une caractéristique *C–V*.

Pour cela, il suffit de définir la fonction suivante :

$$P(V_g) = \frac{dC_{gg}}{dV_g} \tag{A.1}$$

où C_{gg} représente la transcapacité de grille basse fréquence (quasi-statique) et V_g la tension de grille appliquée au transistor MOS. La tension de seuil V_{th} est alors définie comme la valeur de V_g pour laquelle la fonction P est maximale (cf. Figs. A.1 et A.2).

L'expression de la tension de seuil est donc donnée par la bijection suivante :

$$V_{th} = (\max P)^{-1} \tag{A.2}$$

En d'autres termes, la tension de seuil est simplement définie comme l'abscisse du point d'inflexion de la courbe $C_{gg}(V_g)$ en régime d'inversion faible. Bien que cette définition puisse sembler simpliste, elle est cependant physiquement cohérente : en effet, quand le TMOS entre en régime de fonctionnement 'on', le canal d'inversion est juste formé, ce qui correspond au changement d'état représenté par le maximum de la fonction $P(V_g)$.

La Fig. A.3 illustre la précision et la validité de notre procédure d'extraction. L'excellent accord entre les mesures expérimentales réalisées par VAN DORT [1] et les résultats obtenus avec notre méthode est clairement montré.

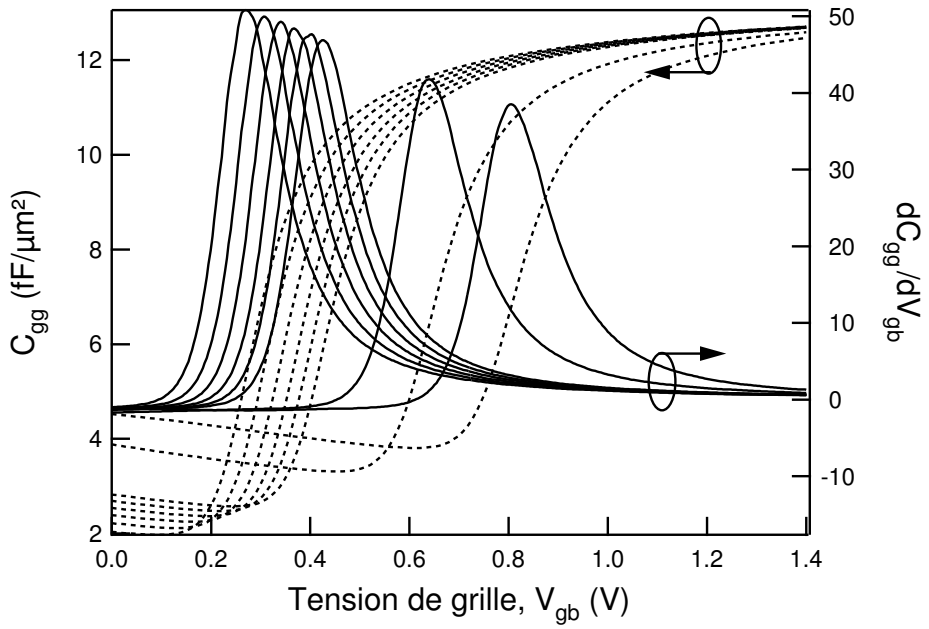


FIG. A.1 : *Principe de notre méthode d'extraction de la tension de seuil à partir d'un réseau de courbes C–V. Les différentes courbes C_{gg} (pointillés) correspondent à différents dopages substrat, allant de 5×10^{17} à 3×10^{18} cm^{-3} (de gauche à droite). Elles ont été obtenues par résolutions auto-cohérentes des équations de Schrödinger et de Poisson.*

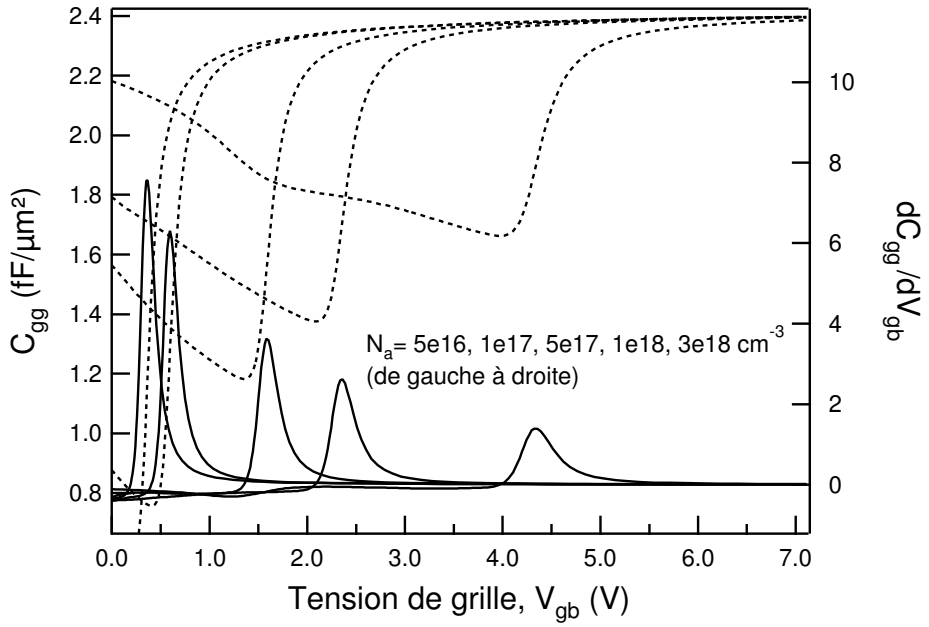


FIG. A.2 : *Caractéristiques C–V correspondant aux paramètres de l'expérience de van Dort [1]. En particulier, l'épaisseur d'oxyde de grille est égale à 14 nm.*

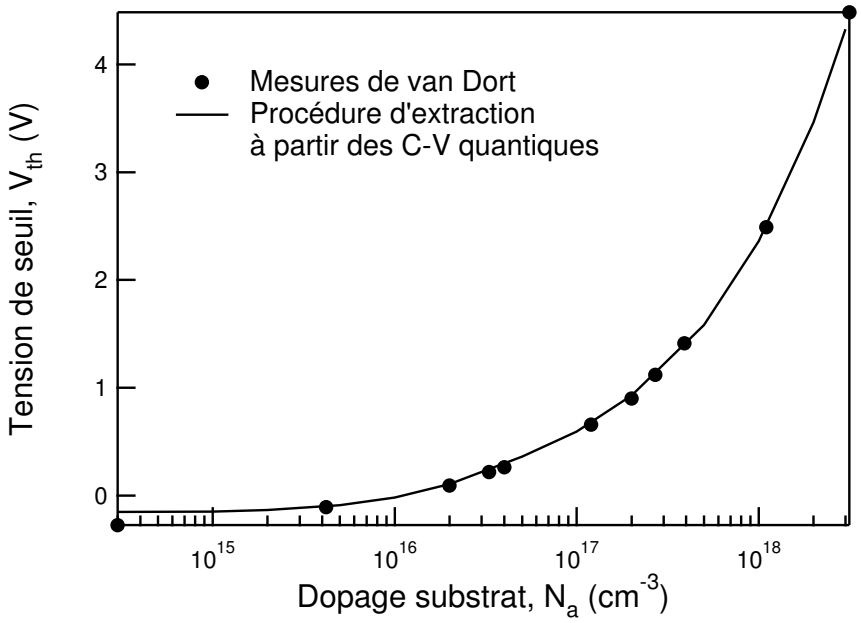


FIG. A.3 : *Tension de seuil en fonction du dopage substrat. Les symboles représentent les résultats expérimentaux de van Dort [1] et la ligne continue les résultats obtenus avec notre procédure d'extraction réalisée à partir des courbes C–V montrées à la Fig. A.2.*

Bibliographie

[1] M. J. van Dort, P. H. Woerlee, A. J. Walker, *et al.*, "Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 39, no. 4, pp. 932–938, Apr. 1992.

Annexe B

Code VHDL-AMS du modèle de potentiel de surface quantique – modélisation de l'inversion

```
-- All the functions
-- *****************
library ieee;
library disciplines;

use disciplines.ELECTROMAGNETIC_SYSTEM.all;
use ieee.math_real.all;

-- functions declaration
-- **********************
package the_functions is
    pure function nall(Cox,Vg,Vt,Vch:real) return real;
    pure function bq(Cox,Vg,Vt,Vch:real) return real;
    pure function Ew(Cox,Vg,Vt,Vch:real) return real;
    pure function Vg_hyp(Vg,Vt,Vch,eps:real) return real;
    pure function dPhis(E:real) return real;
    pure function Phiswi(Vg,Vfb,Gamma:real) return real;
    pure function PsiB(Phif,Phit,Gamma,Vg,Vfb,Vch:real) return real;
    pure function f(Vg,Vch,Vfb,Gamma,Phif,delta:real) return real;
    pure function fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps:real) return real;
    pure function Phis(Vg,Vch,Vfb,Gamma,Phif,Phit,delta:real) return real;
    pure function Phisq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,Phit,delta,eps:real) return real;
end;

-- functions definition
-- ********************
package body the_functions is

    -- Equivalent carrier density (moderate inversion approximation)
    -- ************************************************************
    pure function nall(Cox,Vg,Vt,Vch:real) return real is
        constant q      :real       :=  1.602e-19;
        variable ret    :real;
    begin
        ret :=  2.0*Cox*(Vg-Vt-Vch)/q;
        return ret;
    end nall;

    -- Quantum effects 1
    -- ****************
    pure function bq(Cox,Vg,Vt,Vch:real) return real is
        constant q      :real   :=  1.602e-19;
```

```
    constant eps0   :real   :=  8.86e-12;
    constant eps_si :real   :=  11.7*eps0;
    constant h      :real   :=  6.626e-34;
    constant hb     :real   :=  h/(2.0*math_pi);
    constant me     :real   :=  9.1e-31;
    constant m      :real   :=  0.98*me;
    variable ret    :real;
begin
    ret :=  (nall(Cox,Vg,Vt,Vch)*(4.0*m*q**2)/(eps_si*hb**2))**(1.0/3.0);
    return ret;
end bq;

-- Quantum effects 2
-- ****************
pure function Ew(Cox,Vg,Vt,Vch:real) return real is
    constant h      :real   :=  6.626e-34;
    constant hb     :real   :=  h/(2.0*math_pi);
    constant me     :real   :=  9.1e-31;
    constant m      :real   :=  0.98*me;
    variable ret    :real;
begin
    ret :=  (3.0 * hb**2 * bq(Cox,Vg,Vt,Vch)**2)/(8.0*m);
    return ret;
end Ew;

-- Smooth Vg
-- *********
pure function Vg_hyp(Vg,Vt,Vch,eps:real) return real is
    variable ret    :real;
begin
    ret :=  0.5*(Vg + sqrt((Vg-Vt-Vch)**2+4.0*eps**2) + sqrt((Vt+Vch)**2+4.0*eps**2));
    return ret;
end Vg_hyp;

-- New delta-Phis
-- **************
pure function dPhis(E:real) return real is
    constant q      :real   :=  1.602e-19;
    variable ret    :real;
begin
    ret :=  E/q;
    return ret;
end dPhis;

-- Weak inversion surface potential
-- ********************************
pure function Phiswi(Vg,Vfb,Gamma:real) return real is
    variable ret    :real;
begin
    ret :=  (sqrt(Vg-Vfb+Gamma**2/4.0)-Gamma/2.0)**2;
    return ret;
end Phiswi;

-- PsiB
-- ****
pure function PsiB(Phif,Phit,Gamma,Vg,Vfb,Vch:real) return real is
    variable ret    :real;
begin
    ret :=  2.0*Phif + Vch + (Phiswi(Vg,Vfb,Gamma) - 2.0*Phif -Vch)/
            (sqrt(1.0+((Phiswi(Vg,Vfb,Gamma)-2.0*Phif-Vch)/(4.0*Phit))**2));
    return ret;
end PsiB;

-- Classical description
-- *********************
pure function f(Vg,Vch,Vfb,Gamma,Phif,delta:real) return real is
    variable ret    :real;
begin
```

```
        ret :=  0.5*(2.0*Phif + Phiswi(Vg,Vfb,Gamma) + Vch) -
                0.5*sqrt((Phiswi(Vg,Vfb,Gamma)-2.0*Phif-Vch)**2+4.0*delta**2);
        return ret;
    end f;

    -- Quantum description
    -- *******************
    pure function fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps:real) return real is
        variable ret    :real;
    begin
        ret :=  0.5*(2.0*Phif + Phiswi(Vg,Vfb,Gamma) + Vch +
                dPhis(Ew(Cox,Vg_hyp(Vg,Vt,Vch,eps),Vt,Vch))) -
                0.5*sqrt((Phiswi(Vg,Vfb,Gamma)-2.0*Phif-Vch-
                dPhis(Ew(Cox,Vg_hyp(Vg,Vt,Vch,eps),Vt,Vch)))**2+4.0*delta**2);
        return ret;
    end fq;

    -- Classical surface potential
    -- ***************************
    pure function Phis(Vg,Vch,Vfb,Gamma,Phif,Phit,delta:real) return real is
        variable ret    :real;
    begin
        ret :=  f(Vg,Vch,Vfb,Gamma,Phif,delta) +
                Phit*log(( (Vg-Vfb-f(Vg,Vch,Vfb,Gamma,Phif,delta)-
                (Phiswi(Vg,Vfb,Gamma)-f(Vg,Vch,Vfb,Gamma,Phif,delta))/
                (sqrt(1.0+((Phiswi(Vg,Vfb,Gamma)-f(Vg,Vch,Vfb,Gamma,Phif,delta))/
                (4.0*Phit))**2)))/(Gamma*sqrt(Phit)) )**2 -
                f(Vg,Vch,Vfb,Gamma,Phif,delta)/Phit + 1.0 );
        return ret;
    end Phis;

    -- Quantum surface potential
    -- *************************
    pure function Phisq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,Phit,delta,eps:real) return real is
        variable ret    :real;
    begin
        ret :=  fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps) +
                Phit*log(((Vg-Vfb-fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps)-
                (Phiswi(Vg,Vfb,Gamma)-fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps))/
                (sqrt(1.0+((Phiswi(Vg,Vfb,Gamma)-fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps))/
                (4.0*Phit))**2))) / (Gamma*sqrt(Phit)) )**2 -
                fq(Cox,Vg,Vt,Vch,Vfb,Gamma,Phif,delta,eps)/Phit + 1.0 );
        return ret;
    end Phisq;

end the_functions;


-- *******
-- *******
-- *******

library ieee;
library disciplines;

use disciplines.ELECTROMAGNETIC_SYSTEM.all;
use ieee.math_real.all;
use work.all;
use work.the_functions.all;

-- Entity MOS
-- **********

entity MOS is

    generic (Na         :real   :=  2.4e23;     -- Substrate doping level
        Np              :real   :=  1.0e26;     -- Poly doping level
        Tox             :real   :=  3.5e-9;     -- Oxide thickness
```

```
        Mu0             :real   :=  0.066;      -- Low-field mobility
        Vfb_I           :real   :=  -0.87;      -- Flatband voltage for I-V curves
        Vfb_C0          :real   :=  -0.56;      -- Flatband voltage reference for C-V curves
        Theta1          :real   :=  0.19;       -- Mobility parameter 1
        Theta2          :real   :=  0.33;       -- Mobility parameter 2
        eps             :real   :=  0.15;       -- Smoothing constant
        delta           :real   :=  0.02;       -- Smoothing constant
        W               :real   :=  1.0e-6;     -- Width
        L               :real   :=  1.0e-6);    -- Length
    port    (terminal G,D,S,B   :electrical);

end MOS;

-- architecture Quantum_Polydep
-- ****************************

architecture Quantum_Polydep of MOS is

    constant T          :real   :=  300.0;
    constant q          :real   :=  1.602e-19;
    constant k          :real   :=  1.38066e-23;
    constant eps0       :real   :=  8.86e-12;
    constant eps_si     :real   :=  11.7*eps0;
    constant epsox      :real   :=  3.9*eps0;
    constant Eg         :real   :=  1.79422e-19;
    constant Nc         :real   :=  2.8e25;
    constant Nv         :real   :=  1.04e25;
    constant Phit       :real   :=  k*T/q;
    constant Cox        :real   :=  epsox/Tox;
    constant ni         :real   :=  sqrt(Nc)*sqrt(Nv)*exp(-Eg/(2.0*k*T));
    constant Phif       :real   :=  Phit*log(Na/ni);
    constant Vfb_C      :real   :=  Vfb_C0-Phif;
    constant Gamma      :real   :=  sqrt(2.0*q*eps_si*Na)/Cox;
    constant Vt0_C      :real   :=  Vfb_C + 2.0*Phif + Gamma*sqrt(2.0*Phif);
    constant Vt0_I      :real   :=  Vfb_I + 2.0*Phif + Gamma*sqrt(2.0*Phif);
    constant Gammap     :real   :=  sqrt(2.0*q*eps_si*Np)/Cox;

    quantity Vgeff          :real;
    quantity Mus            :real;
    quantity Idrift,Idiff   :real;
    quantity Phip           :real;
    quantity Qg             :real;
    quantity Qbulk          :real;
    quantity Qinv           :real;
    quantity iqinv          :real;

    quantity vdb across D to B;
    quantity vsb across S to B;
    quantity vgb across G to B;
    quantity ids through D to S;
    quantity igs through G to S;
    quantity igd through G to D;
    quantity isb through S to B;
    quantity idb through D to B;

begin

    -- Polydepletion effect (used only for current evaluation)
    Vgeff   ==  Vfb_I + Phis(Vgb,Vsb,Vfb_I,Gamma,Phif,Phit,delta) +
                (q*eps_si*Np/(Cox**2))*(sqrt(1.0+2.0*(Cox**2)*(Vgb-Vfb_I-
                Phis(Vgb_I,Vsb,Vfb_I,Gamma,Phif,Phit,delta))/(q*eps_si*Np))-1.0);

    -- Mobility model
    Mus     ==  Mu0/(1.0+Theta1*(Vgeff-Vt0_I)+Theta2*(Vgeff-Vt0_I)**2);

    -- Drift current
    Idrift  ==  (Mus*Cox*W/L)*( (Vgeff-Vfb_I)*
                (Phisq(Cox,Vgeff,Vt0_I,Vdb,Vfb_I,Gamma,Phif,Phit,delta,eps)-
```

```
                Phisq(Cox,Vgeff,Vt0_I,Vsb,Vfb_I,Gamma,Phif,Phit,delta,eps)) -
                (Phisq(Cox,Vgeff,Vt0_I,Vdb,Vfb_I,Gamma,Phif,Phit,delta,eps)**2-
                Phisq(Cox,Vgeff,Vt0_I,Vsb,Vfb_I,Gamma,Phif,Phit,delta,eps)**2)/2.0 -
                (2.0*Gamma/3.0)*
                (sqrt(Phisq(Cox,Vgeff,Vt0_I,Vdb,Vfb_I,Gamma,Phif,Phit,delta,eps)**3)-
                sqrt(Phisq(Cox,Vgeff,Vt0_I,Vsb,Vfb_I,Gamma,Phif,Phit,delta,eps)**3)) );

    -- Diffusion current
    Idiff   ==  (Mus*Cox*Phit*Gamma*W/L)*( sqrt(Phisq(Vgeff,Vsb,Vfb_I,Gamma,Phif,Phit,delta)+
                Phit*exp((Phisq(Vgeff,Vsb,Vfb_I,Gamma,Phif,Phit,delta)-Vsb-2.0*Phif-
                dPhis(Ew(Cox,Vg_hyp(Vgeff,Vt0_I,Vsb,eps),Vt0_I,Vsb))))/Phit)) -
                sqrt(Phisq(Vgeff,Vdb,Vfb_I,Gamma,Phif,Phit,delta)+
                Phit*exp((Phisq(Vgeff,Vdb,Vfb_I,Gamma,Phif,Phit,delta)-Vdb-2.0*Phif-
                dPhis(Ew(Cox,Vg_hyp(Vgeff,Vt0_I,Vdb,eps),Vt0_I,Vdb))))/Phit)) -
                sqrt(Phisq(Vgeff,Vsb,Vfb_I,Gamma,Phif,Phit,delta)) +
                sqrt(Phisq(Vgeff,Vdb,Vfb_I,Gamma,Phif,Phit,delta)) );

    -- Poly surface potential
    Phip    ==  (sqrt(Vgb-Vfb_C-Phis(Vgb,Vsb,Vfb_C,Gamma,Phif,Phit,delta)+Gammap**2/4.0)
                -Gammap/2.0)**2;

    -- Gate charge
    Qg      ==  W*L*Cox*(Vgb-Vfb_C-Phisq(Cox,Vgb,Vt0_C,Vsb,Vfb_C,Gamma,Phif,Phit,delta,eps)-Phip);

    -- Bulk charge
    Qbulk   ==  W*L*Cox*Gamma*sqrt(Phisq(Cox,Vgb,Vt0_C,Vsb,Vfb_C,Gamma,Phif,Phit,delta,eps));

    -- Inversion charge
    Qinv    ==  Qg-Qbulk;
    iqinv   ==  Qinv'dot;

    -- Currents
    ids     ==  Idrift+Idiff;
    igs     ==  0.5*Qg'dot;
    igd     ==  0.5*Qg'dot;
    isb     ==  0.5*Qbulk'dot;
    idb     ==  0.5*Qbulk'dot;

end Quantum_Polydep;
```

Publications associées à ce travail

Revues spécialisées

1. *F. Prégaldiny, C. Lallement et D. Mathiot*
   "A simple efficient model of parasitic capacitances of deep-submicron LDD MOSFETs," *Solid-State Electronics*, vol. 46, no. 12, pp. 2191–2198, Décembre 2002.

2. *F. Prégaldiny, C. Lallement, R. van Langevelde et D. Mathiot*
   "An advanced explicit surface potential model physically accounting for the quantization effects in deep-submicron MOSFETs," *Solid-State Electronics*, vol. 48, no. 3, pp. 427–435, Mars 2004.

3. *F. Prégaldiny, C. Lallement et D. Mathiot*
   "Accounting for quantum mechanical effects from accumulation to inversion, in a fully analytical surface-potential-based MOSFET model," accepté dans *Solid-State Electronics* (Décembre 2003).

Conférences internationales avec actes

1. *F. Prégaldiny, C. Lallement, W. Grabinski, J.-B. Kammerer et D. Mathiot*
   "An analytical quantum model for the surface potential of deep-submicron MOSFETs," *10th International Conference on Mixed Design of Integrated Circuits and Systems (MIXDES)*, Lodz, Pologne, pp. 98–104, Juin 2003 (communication invitée).

2. *F. Prégaldiny, C. Lallement et D. Mathiot*
   "Quantum surface potential model suitable for advanced MOSFETs simulation," *IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Boston, USA, pp. 227–230, Septembre 2003.

Séminaire international

1. *F. Prégaldiny, C. Lallement et D. Mathiot*
   "Extrinsic capacitance model for advanced MOSFET design," *MOS Modeling and Parameter Extraction Group Meeting*, Erfurt, Allemagne, Octobre 2002 (communication invitée).

Conférence nationale avec actes

1. *F. Prégaldiny, C. Lallement et D. Mathiot*
   "Modélisation analytique avancée des capacités parasites du transistor MOS fortement submicronique," *VIèmes Journées Nationales du Réseau Doctoral Microélectronique, Micro et Nano Technologies (JNRDM)*, Toulouse, pp. 311–313, Mai 2003.

Étude et modélisation du comportement électrique des transistors MOS fortement submicroniques

La modélisation précise des transistors MOS pour la conception et la simulation de circuits est un défi constant en raison de la nature évolutive de la technologie CMOS. L'objectif de cette thèse est d'une part d'étudier les principaux effets résultant de la miniaturisation des TMOS et d'autre part de proposer des modèles analytiques simples et originaux permettant de les prendre en compte. Les bases physiques nécessaires à la formulation d'un modèle idéal sont présentées au chapitre 2, de même qu'un état de l'art des principaux modèles compacts de TMOS (modèles destinés à la simulation de circuits) actuellement utilisés. Le troisième chapitre est consacré à une étude détaillée du comportement capacitif extrinsèque du TMOS fortement submicronique. Un nouveau modèle de capacités parasites est également proposé puis validé à partir de simulations numériques à deux dimensions. Le quatrième chapitre fait état d'une étude approfondie des effets quantiques au sein des transistors n-MOS. L'influence des effets quantiques sur les différentes caractéristiques électriques (I–V, C–V) du TMOS est discutée. Un nouveau modèle quantique, formulé intégralement en potentiel de surface, est alors développé. Ce modèle est complètement analytique, valable de l'accumulation à l'inversion, et ne nécessite aucun paramètre d'ajustement. Utilisé conjointement à un modèle en feuille de charge, il autorise une description précise et continue des caractéristiques électriques majeures du TMOS telles que les charges, les capacités, le courant de drain, la transconductance, etc. Le nouveau modèle est finalement validé par comparaison avec des résultats expérimentaux de différentes technologies CMOS avancées. En conclusion, cette thèse démontre qu'une approche pragmatique de la modélisation compacte permet de réaliser des modèles simples, efficaces et physiquement cohérents.

Mots clés : MOSFET, modélisation compacte, effets quantiques, modèle en feuille de charge, comportement extrinsèque.

Study and modeling of the electrical behavior of deep-submicron MOSFETs

Accurate MOS transistor modeling for circuit design and simulation is a constant challenge due to the continuously evolving of CMOS technology. The objective of this thesis is on the one hand to study the main effects resulting from MOSFET miniaturization and on the other hand to propose simple and original analytical models accounting for them. The physical basis necessary to the formulation of an ideal MOSFET model is presented in chapter 2. In addition, a state of the art of the most widely used compact MOSFET models (models for circuit simulation) is also discussed. Chapter 3 is devoted to a detailed study of the extrinsic capacitive behavior of deep-submicron MOSFETs. A new model of parasitic capacitances is developed and then validated by two-dimensional numerical simulations. Chapter 4 introduces a depth study of the quantization effects in both accumulation and inversion layers of n-MOS transistors. The impact of quantum effects on the various electrical characteristics (I–V, C–V) is discussed. A new fully analytical surface-potential-based MOSFET model accounting for the quantum effects is then derived in full detail. This model is valid from accumulation to inversion and does not need any fitting parameter. Within the context of a charge sheet model, it leads to an accurate and continuous description of major MOSFET electrical characteristics such as charges, capacitances, drain current, transconductance, etc. The new model is finally validated by comparison with experimental results from various advanced CMOS technologies. In conclusion, this thesis demonstrates that a pragmatic approach of compact modeling enables the development of simple, efficient and physically coherent models.

Keywords: MOSFET, compact modeling, quantum effects, charge sheet model, extrinsic behavior.