

TABLE DES MATIÈRES

INTRODUCTION	1
1 LE TRANSISTOR MOS NANOMETRIQUE : ARCHITECTURE ET NOUVEAUX EFFETS PHYSIQUES	7
1.1 Introduction	7
1.2 Evolution de la structure des transistors MOS	10
1.2.1 Structure de grille	11
1.2.1.1 Evolution de la structure de grille	11
1.2.1.2 Limites de la grille en polysilicium	12
1.2.1.3 Phénomène de déplétion de grille	12
1.2.2 Diélectrique de grille	14
1.2.3 Profils de concentration de dopants dans le canal	16
1.2.3.1 Substrat	16
1.2.3.2 Dopage conventionnel du canal : <i>Diffused-Well</i>	17
1.2.3.3 Canal à profil de dopage rétrograde (<i>Retrograde Well</i>)	18
1.2.3.4 Canal à profil de dopage Super Steep retrograde	19
1.2.4 Structure de Source/Drain	20
1.2.4.1 Structures anti-perçage (Anti Punchthrough Structures)	22
1.2.4.2 Résistances d'accès des transistors MOS fortement submicroniques	24
1.2.4.3 Zones de recouvrement des extensions de source et de drain	26
1.3 Limites de la technologie CMOS Silicium	27
1.3.1 Les limites de la réduction conventionnelles des dimensions	27
1.3.2 Technologie SOI (Silicon on Insulator)	28
1.4 Nouveaux effets physiques dans les transistors MOS nanométriques	30
1.4.1 Effets quantiques	30
1.4.1.1 Quantification de l'énergie des porteurs	30
1.4.1.2 Influence des effets quantiques	32
1.4.1.3 Structure à une dimension : Capacité MOS	33
1.4.1.4 Structure à deux dimensions : Transistor MOS	34
1.4.2 Courant de grille dans les oxydes ultra minces	35
1.4.2.1 Classification des mécanismes tunnel	35
1.4.2.1.1 L'effet tunnel Fowler-Nordheim	35
1.4.2.1.2 L'effet tunnel direct	37
1.4.2.1.3 L'effet tunnel indirect	40
1.4.3 Impact des courants de fuite sur le fonctionnement du transistor MOS	41
1.4.3.1 Transistor nMOS	42
1.4.3.1.1 Régime de forte inversion $V_{GS} > V_T$	42

1.4.3.1.2	Régime de faible inversion et de déplétion ($V_{FB} < V_{GS} < V_T$)	43
1.4.3.1.3	Régime d'accumulation ($V_{GS} < V_{FB}$)	44
1.4.3.2	Transistor pMOS	45
1.4.3.2.1	Régime de forte inversion $V_{GS} < V_T$	45
1.4.3.2.2	Régime de faible inversion et de déplétion ($V_T < V_{GS} < V_{FB}$)	46
1.4.3.2.3	Régime d'accumulation ($V_{GS} > V_{FB}$)	47
1.4.3.3	Synthèse des composantes tunnels	47
1.5	Conclusion	48
2	MODELISATION DES EFFETS QUANTIQUES ET DU COURANT TUNNEL DANS LES CAPACITES MOS	51
2.1	Introduction	51
2.2	Les régimes de fonctionnement	52
2.2.1	Le régime d'accumulation ($V_{GB} < V_{FB}$)	53
2.2.2	Le régime de déplétion et faible inversion ($V_{FB} < V_{GB} < V_T$)	53
2.2.3	Le régime de forte inversion ($V_{GB} > V_T$)	54
2.3	Modélisation de la capacité MOS basée sur le potentiel de surface	55
2.4	Modélisation du phénomène de déplétion dans le polysilicium de grille.	57
2.5	Modélisation explicite du potentiel de surface	62
2.5.1	Régimes de déplétion et d'inversion	62
2.5.2	Régime d'accumulation	65
2.5.3	Modèle explicite complet	67
2.6	Modélisation des effets Quantiques	69
2.6.1	Structure de sous bandes d'énergie	69
2.6.2	Calcul auto-cohérent de l'équation de Poisson et de l'équation de Schrödinger	73
2.6.3	Approche Analytique des effets Quantiques	77
2.6.3.1	L'approximation du puits de potentiel triangulaire.	77
2.6.3.2	L'approximation variationnelle	78
2.6.4	Modélisation implicite du potentiel de surface avec prise en compte des effets quantiques.	80
2.6.4.1	Modèle de van Dort	81
2.6.4.2	Modèle implicite quantique basé sur le modèle de van Dort	84
2.6.4.3	Modèle implicite quantique basé sur l'approche variationnelle : Régime de déplétion et d'inversion	87
2.6.4.4	Modèle implicite quantique basé sur l'approche variationnelle : Régime d'accumulation	94
2.6.4.5	Modèle implicite quantique complet pour les structures pMOS	101
2.6.5	Modélisation explicite quantique du potentiel de surface en régime de déplétion et d'inversion	103
2.6.5.1	Modèle Compact MM11	103

2.6.5.2	Modèle explicite basé sur l'élargissement du gap	104
2.6.5.3	Modèle quantique explicite basé sur l'approche variationnelle	106
2.7	Modélisation du courant de grille	112
2.7.1	Expression générale du courant tunnel direct	112
2.7.2	Calcul analytique de la probabilité tunnel	114
2.7.3	Formulation explicite du courant tunnel direct en fonction du potentiel de surface	121
2.8	Conclusion	123
3	MODELISATION DU TRANSISTOR MOS AVEC FUITE DE GRILLE	127
3.1	Introduction	127
3.2	Modélisation classique en feuille de charge du transistor MOS	127
3.3	Mobilité de la couche d'inversion	132
3.3.1	Réduction de la mobilité	132
3.3.2	Saturation de la vitesse des porteurs	135
3.4	Modélisation du potentiel de surface du transistor MOS	136
3.4.1	Modélisation quasi-statique sans courant de grille	136
3.4.2	Influence des effets quantiques et du phénomène de polydépétion de grille sur le courant drain.	139
3.5	Modélisation du potentiel de surface avec fuite de grille	141
3.5.1	Formulation du modèle quasi statique	141
3.5.2	Influence du courant de grille à $V_{DS} = 0$ V	143
3.5.2.1	Dépolarisation du canal	143
3.5.2.2	Impact sur la densité de courant de grille	144
3.5.2.3	Impact sur la capacité quasi statique	146
3.5.3	Impact du courant de grille sur le fonctionnement du transistor MOS en régime linéaire	147
3.5.3.1	Courant de grille total	147
3.5.3.2	Courants drain et source	148
3.6	Etude expérimentale et validation du modèle	149
3.6.1	Présentation des technologies étudiées	149
3.6.2	Caractéristiques courant – tension I - V	151
3.6.2.1	Lot RTN14	151
3.6.2.2	Lot RTN12	153
3.6.2.3	Lot DPN12	155
3.6.3	Mesure de capacités à diélectriques ultrafins	157
3.6.4	Validation du modèle statique du transistor MOS avec fuite de grille.	160
3.7	Caractérisation de capacités MOS de grande surface	165
3.7.1	Caractérisation statique	165
3.7.2	Modélisation non quasi statique de la capacité MOS.	166

3.7.2.1	Etude théorique non quasi statique de l'impact du courant tunnel	167
3.7.2.2	Validation du modèle non quasi statique	168
3.8	Conclusion	171
4	MODELISATION DU BRUIT BASSE FREQUENCE DANS LES STRUCTURES MOS A DIELECTRIQUES ULTRAFINS	177
4.1	Introduction	177
4.2	Les sources de bruit	177
4.3	Modélisation du phénomène de piégeage/dépiégeage	179
4.3.1	Modélisation du piégeage en forte inversion	180
4.3.2	Modélisation du piégeage en faible inversion	186
4.4	Formulation du bruit dans les transistors MOS	194
4.4.1	Densité spectrale des fluctuations de la tension de bandes plates	194
4.4.2	Densité spectrale des fluctuations du courant de drain.	195
4.4.2.1	Modèle ΔN	195
4.4.2.2	Modèle corrélé $\Delta N - \Delta \mu$	199
4.4.2.3	Lien avec les formulations existantes	201
4.5	Caractérisation en bruit du courant de drain	202
4.6	Bruit basse fréquence associé au courant de grille	206
4.6.1	Modélisation du bruit associé au courant tunnel direct	206
4.6.2	Résultats expérimentaux et validation du modèle	207
4.6.3	Influence du courant de grille sur le bruit du canal dans les transistors longs	211
4.6.4	Bruit RTS sur la grille [MARTINEZ'05b]	214
4.7	Conclusion	219
	CONCLUSIONS ET PERSPECTIVES	223
	ANNEXE A : APPROXIMATION DE LA FEUILLE DE CHARGE	231
	ANNEXE B : TECHNIQUE DU TRANSFERT MATRICIEL [CASSAN'00]	233
	ANNEXE C : RESOLUTION NUMERIQUE DE L'EQUATION DIFFERENTIELLE REGISSANT LE POTENTIEL DE SURFACE	237
	ANNEXE D : MODELE PETIT SIGNAL NON QUASI STATIQUE	239
	PUBLICATIONS	253

Introduction

La technologie CMOS a connu, durant ces quarante dernières années, un essor spectaculaire. Les circuits électroniques basés sur cette technologie sont le coeur de l'électronique moderne. Les très fortes densités d'intégration et la faible consommation des transistors MOS (Metal Oxide Semiconductor) expliquent en grande partie le fait que la technologie CMOS représente près de 90 % du marché des semiconducteurs. La course à la réduction des dimensions s'accompagne d'un challenge aussi bien technologique que physique, entraînant l'industrie du semiconducteur à se surpasser et à se projeter en permanence vers le futur.

La recherche de performances toujours accrues et de densités d'intégration toujours plus élevées se heurte à de nombreuses difficultés. Ces problèmes sont liés notamment à la difficulté d'obtenir les performances attendues par une simple loi de réduction d'échelle. Un nombre considérable de nouveaux effets physiques liés à la miniaturisation deviennent prépondérants. Les modèles permettant de prédire le comportement des transistors MOS font apparaître leurs limites quand ils sont appliqués à ces structures nanométriques. Parmi celles-ci, la réduction de l'épaisseur d'oxyde et les forts niveaux de dopage entraînent l'apparition de forts courants de fuites au travers du film diélectrique et le confinement quantique des porteurs de la couche d'inversion.

La compréhension de ces nouveaux phénomènes nécessite à la fois une modélisation avancée et des outils de caractérisation adaptés à ces dispositifs. Les modèles classiques, qui sont à la base des méthodes de caractérisation, souffrent de sévères limitations et l'utilisation des modèles physiques existants adaptés aux technologies nanométriques comme outil de caractérisation est souvent impossible, essentiellement dû aux temps de calculs trop importants. Il est donc primordial de disposer de modèles physiques prenant en compte ces nouveaux phénomènes et nécessitant de faibles temps de calculs pour caractériser les structures MOS nanométriques. Cette thèse se positionne dans le cadre de la modélisation analytique ou semi-analytique des effets liés à la miniaturisation du transistor MOS, dans le but de développer des modèles physiques utilisables pour la caractérisation des dispositifs.

Ce mémoire est composé de quatre chapitres.

Dans le premier chapitre, nous détaillons l'architecture des transistors MOS nanométriques. L'accent est mis sur les différents points clé de l'optimisation des dispositifs ainsi que sur les techniques de fabrication. Nous abordons également les limites des technologies CMOS sur substrat massif, et les potentialités des technologies SOI pour la

course à la miniaturisation. Enfin, nous présentons qualitativement les nouveaux effets physiques à prendre en compte dans ces technologies nanométriques, à savoir les effets quantiques induisant le confinement des porteurs de la couche d'inversion, et les mécanismes de fuites par effet tunnel au travers du diélectrique de grille ultrafin.

Le deuxième chapitre est consacré à la modélisation du potentiel de surface et du courant tunnel direct dans une structure MOS unidimensionnelle à diélectrique ultrafin. Dans un premier temps, nous rappelons la formulation classique (i.e. sans effets quantiques) permettant de définir les relations implicites et explicites liant le potentiel de surface, la tension de grille et le quasi niveau de Fermi, en tenant compte de la déplétion dans le matériau de grille (polydéplétion). Nous présentons ensuite les effets quantiques et leur modélisation dans les capacités MOS. Après un rappel des bases de mécanique quantique nécessaires pour appréhender l'impact des effets quantiques, nous présentons une approche analytique permettant d'établir une relation implicite permettant de lier le potentiel de surface, la tension de grille et le quasi niveau de Fermi, en tenant compte de la polydéplétion, dans tous les régimes de fonctionnement. Nous présentons également un nouveau modèle explicite quantique du potentiel de surface, valable de la faible à la forte inversion. Enfin, après avoir rappelé les bases des mécanismes de fuite de grille par effet tunnel direct, nous reformulons le modèle de courant tunnel direct pour le coupler à notre modèle explicite quantique de potentiel de surface. Tous ces modèles servent de base pour le modèle de transistors MOS développé au chapitre trois.

Le troisième chapitre est consacré à la modélisation des transistors MOS à canal long basé sur le modèle en feuille de charge, en tenant compte des effets quantiques, de polydéplétion, et de courant tunnel. Nous présentons également dans ce chapitre les résultats expérimentaux issus de trois lots de transistors de technologies avancées. A partir des mesures statiques ($I-V$) et dynamiques ($C-V$), nous avons validé notre modèle sur les transistors à canal long et nous avons extrait les paramètres technologiques des transistors des trois lots. Enfin, l'accent est mis sur la modélisation de structures MOS de grande surface ($100 \times 100 \mu\text{m}^2$), pour lesquelles nous développons un modèle petit signal non quasi statique pour la simulation de la capacité.

Le quatrième et dernier chapitre est consacré à la modélisation et la caractérisation du bruit basse fréquence associé au courant de drain et de grille. Nous réexaminons les modèles de piégeage - dépiégeage de porteurs libres par des pièges d'oxyde, dans le but de déterminer la validité des modèles classiques dans le cas des oxydes ultrafins. Nous détaillons ensuite les expressions du modèle de fluctuation du nombre de porteurs et du modèle de fluctuation de la

mobilité corrélée à la fluctuation du nombre de porteurs, dans le but de coupler ces modèles au modèle de transistor MOS établi au chapitre trois. Nous présentons ensuite les résultats de la caractérisation en bruit basse fréquence des transistors des trois lots, qui serviront de support à la comparaison des procédés de nitruration de l'oxyde grille. Nous abordons ensuite la modélisation du bruit basse fréquence associé au courant de grille. Notre approche est validée avec les mesures effectuées sur diverses géométries. L'impact du bruit du courant de grille sur le bruit du courant drain est également abordé. Finalement, nous présentons la modélisation et la caractérisation d'un défaut singulier responsable de bruit de type RTS sur le courant de grille.

Nous concluons cette étude par une synthèse des résultats obtenus et quelques perspectives de travail sur les futures générations de composants. A la suite de la conclusion, ce manuscrit est complété par quatre annexes.

L'annexe A décrit la formulation de la densité de charges de la couche d'inversion sous l'hypothèse de la feuille de charge. Cette formulation est primordiale pour les modèles formulés dans les chapitres deux et trois.

L'annexe B décrit la méthode du transfert matriciel sur laquelle repose le calcul de la probabilité tunnel exacte en résolvant par morceau l'équation de Schrödinger.

L'annexe C présente les détails de la méthode de calcul numérique utilisée pour déterminer le potentiel de surface le long du canal. Elle décrit la résolution numérique d'une équation différentielle non linéaire dont les coefficients dépendent de l'inconnue par méthode itérative.

Enfin l'annexe D décrit la linéarisation et la résolution numérique de l'équation présentée dans l'annexe C dans le but d'établir un modèle petit signal non quasi statique du potentiel de surface.

CHAPITRE 1

LE TRANSISTOR MOS NANOMÉTRIQUE :

ARCHITECTURE ET NOUVEAUX EFFETS

PHYSIQUES

1 Le transistor MOS nanométrique : Architecture et nouveaux effets physiques

1.1 Introduction

Le transistor MOSFET est le composant le plus utilisé de l'industrie des circuits intégrés ULSI. Depuis les trente dernières années, les dimensions des transistors MOS ont été réduites de façon considérable ; les longueurs de grilles des transistors MOS sont passées de 10 μm au début des années 70 à moins de 0.13 μm de nos jours. Cette course à la réduction des dimensions va se poursuivre et il est probable qu'on atteigne des longueurs de grilles de 20 nm vers 2012 [ITRS'03]. La figure 1.1 montre l'évolution des longueurs de grilles des transistors MOS telles que le prédit l'ITRS en 2003. Cette figure donne les longueurs de grille (dessinées et physiques) à atteindre pour que les microprocesseurs (MPU) réalisés avec de tels transistors soient conformes aux spécifications du nœud technologique.

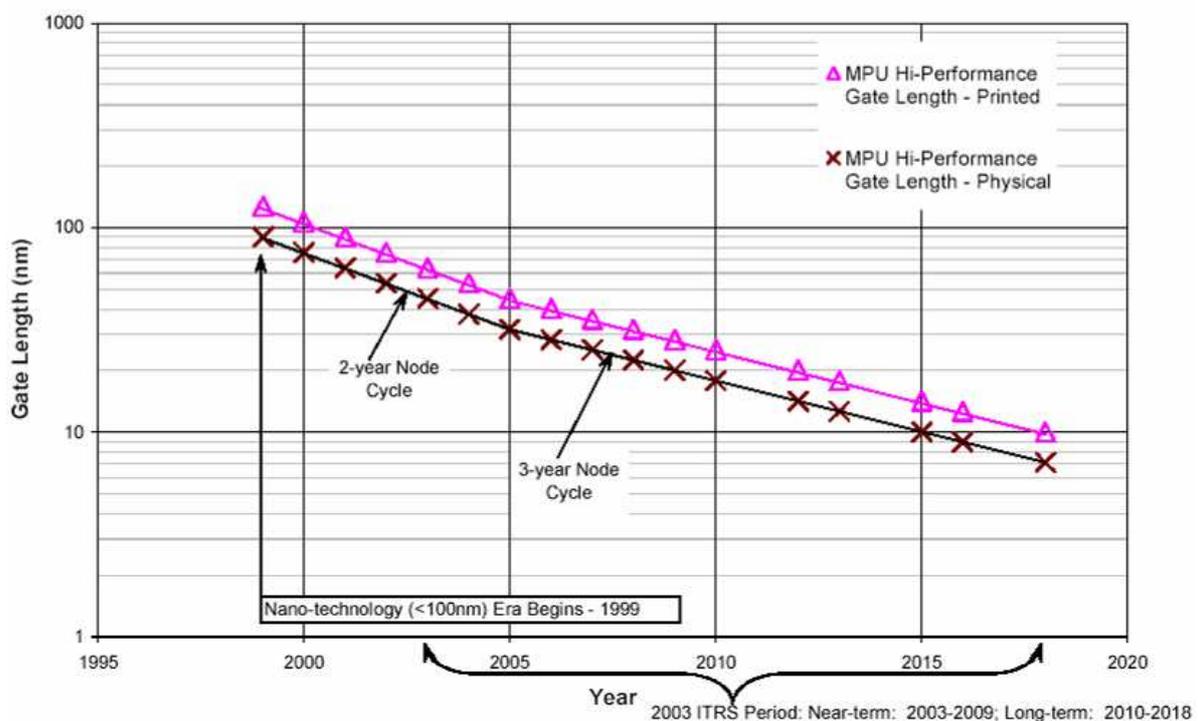


Figure 1.1 : Evolution des longueurs de grilles dessinées et physiques [ITRS'03]

La notion de nœud technologique est liée à l'introduction des mémoires DRAM (*Dynamic Random Access Memory*). Le fait de passer d'un nœud technologique au suivant correspond à une multiplication par quatre du nombre de bits par puce. Le nœud technologique porte alors le nom lié à la distance entre deux lignes de métallisation, désignée

par *Metal Pitch* sur la figure 1.2. Cependant, compte tenu de la grande diversité des produits (mémoires, microprocesseurs, etc.), il est difficile de prédire l'évolution d'un nœud technologique vers le suivant en se basant uniquement sur les mémoires DRAM. En fait, l'amélioration des mémoires va nécessiter une diminution du *half pitch*, alors que pour les microprocesseurs (*MPU*), une diminution de la longueur de grille des transistors est nécessaire. Le DRAM *half-Pitch* et le *MPU Gate length* sont donc les deux grands indicateurs permettant de prédire l'évolution d'une technologie. On définit également *MPU/ASIC M1 half Pitch*, correspondant à la distance entre deux lignes de métallisation dans les microprocesseurs (*MPU*) ou des circuits ASIC (*Application-Specific Integrated Circuit*). La figure 1.3 donne l'évolution du *half pitch* pour les DRAM et les microprocesseurs (*MPU*), ainsi que les nœuds technologiques 90 nm (hp90), 65 nm (hp65), etc.

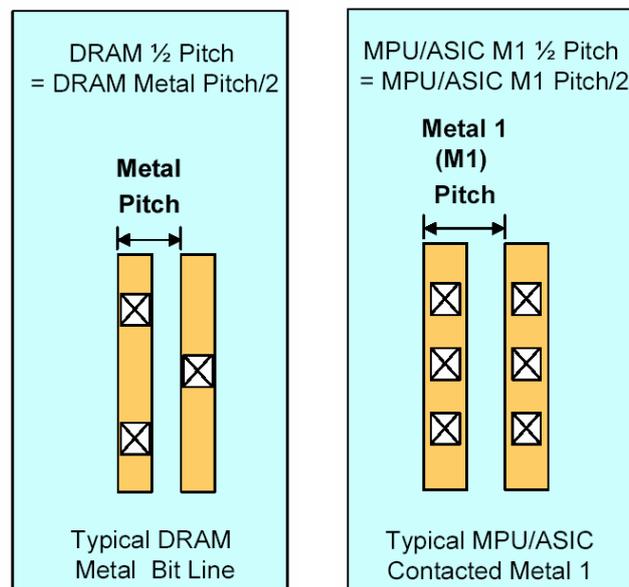


Figure 1.2 : Définition du DRAM half Pitch et du MPU/ASIC Half Pitch [ITRS'03]

La distance *half pitch* est multipliée par 0.7 d'un nœud au suivant. Elle est donc réduite de moitié tous les deux nœuds technologiques. Depuis 2001, l'évolution d'un nœud au suivant prend trois années, alors qu'elle se faisait en deux ans auparavant.

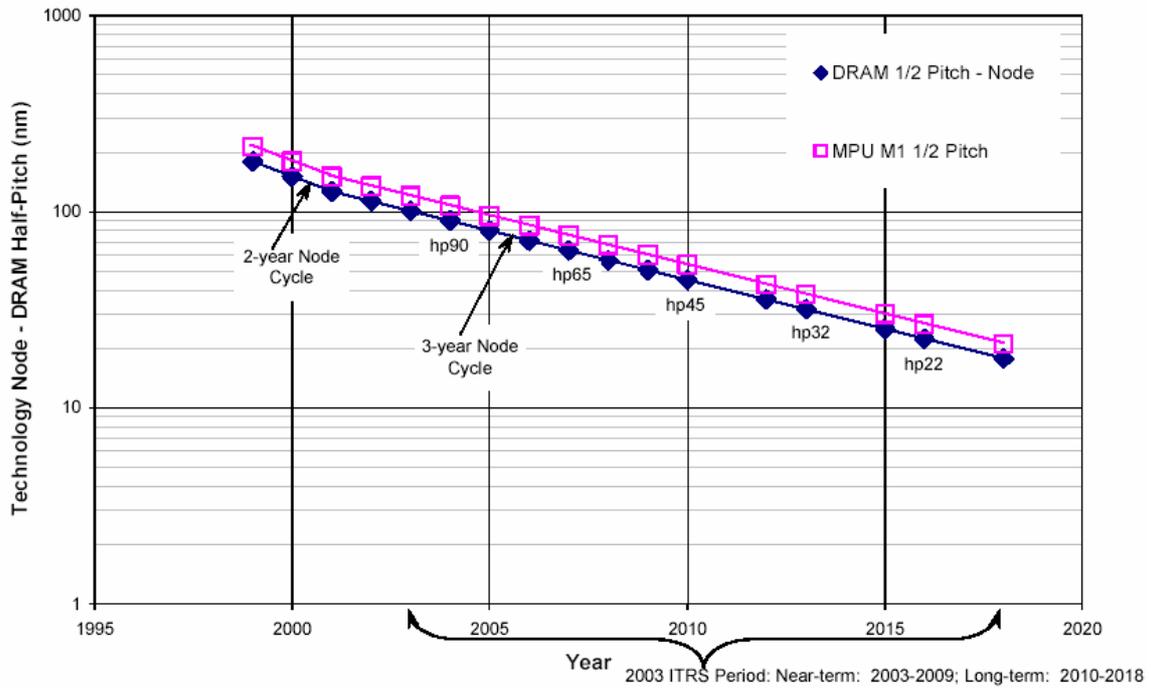


Figure 1.3 : Evolution des noeuds technologiques, DRAM et MPU half Pitch [ITRS'03]

Cependant ces spécifications ne reflètent pas la réalité industrielle en terme de production. La figure 1.4 illustre cette remarque. Par exemple, en 2003, la production de wafers de technologie inférieure à $0,16\ \mu\text{m}$ ne constitue environ que 27% de la production totale.

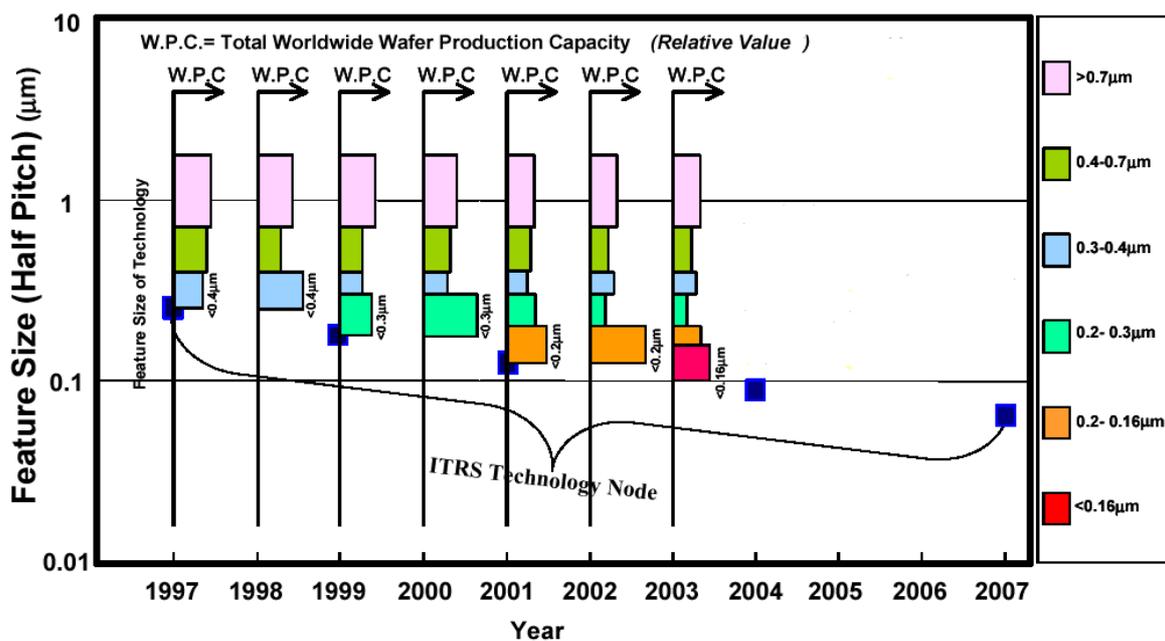


Figure 1.4 : Comparaison entre les noeuds technologiques et la production réelle [ITRS'03]

1.2 Evolution de la structure des transistors MOS

Les transistors MOS sont les composants les plus utilisés dans les circuits intégrés silicium car leurs dimensions peuvent être extrêmement réduites. La réduction de la longueur de grille implique l'emploi des règles de réduction d'échelle. Ces lois prévoient la réduction des différentes dimensions (L_L , W , t_{ox}) par un facteur K tout en conservant l'intégrité électrostatique des transistors. Jusqu'au nœud technologique $0.5 \mu\text{m}$, la tension d'alimentation est restée constante (5 Volts). Pour des raisons de fiabilité des couches diélectriques, une réduction de la tension d'alimentation est désormais nécessaire avec la réduction de l'épaisseur de l'oxyde de grille.

La figure 1.5 montre une image SEM (*Scanning Electron Microscopy*) d'un transistor avec une longueur de grille $2 \mu\text{m}$ (Figure 1.5a) et un transistor avec une longueur de grille $0.2 \mu\text{m}$ (Figure 1.5b). La structure générale n'a pas fondamentalement changé, mais une étude détaillée montre que de nouveaux éléments ont été ajoutés pour un bon fonctionnement (Figure 1.6).

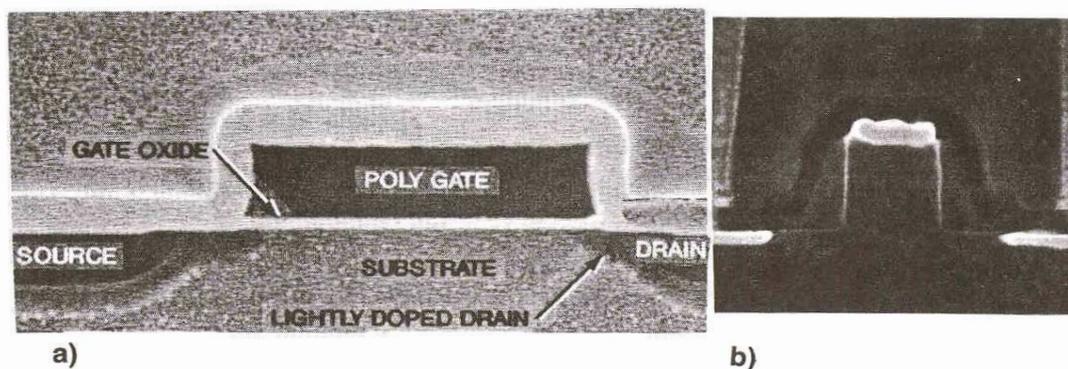


Figure 1.5 : Image SEM de transistors MOS avec différentes longueurs de grille : a) $2 \mu\text{m}$ b) $0,2 \mu\text{m}$

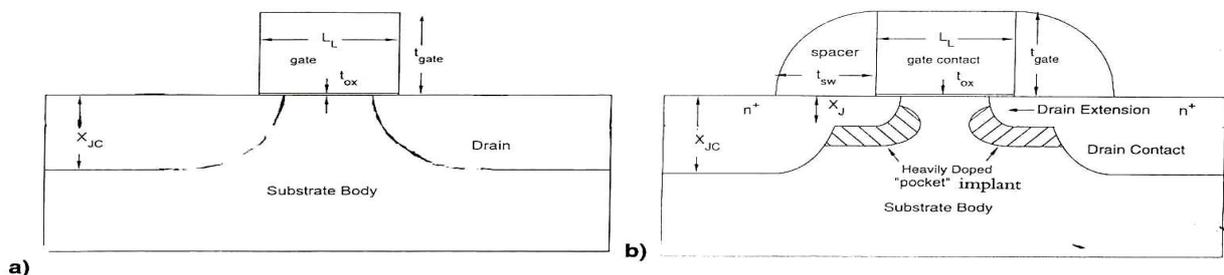


Figure 1.6 : Vue schématique du transistor MOS avec différentes longueurs de grille a) $2 \mu\text{m}$ b) $0,2 \mu\text{m}$

Dans cette partie, les différentes parties de la structure d'un transistor MOS vont être examinées. Les modifications nécessaires au bon fonctionnement des transistors fortement submicroniques sont abordées, ainsi que l'impact sur les caractéristiques de fonctionnement.

1.2.1 Structure de grille

1.2.1.1 Evolution de la structure de grille

Le matériau de métallisation de la grille utilisé est le polysilicium depuis le milieu des années 1970, remplaçant ainsi les grilles aluminium. Jusqu'à la technologie $0.35\mu\text{m}$, le polysilicium était de type n+, que ce soit pour les nMOS ou les pMOS. Avec la réduction des longueurs de grille à $1.25\mu\text{m}$, la résistivité des lignes de polysilicium est devenue si grande qu'elle ne permettait plus de jouer le rôle de ligne d'interconnexion. La solution a consisté à déposer une couche de Métal/Silicium (Siliciure ou *Silicide* en anglais) au dessus du polysilicium qui permet de diminuer la résistivité des lignes polysilicium. Historiquement, le WSi_2 (siliciure de tungstène) a été utilisé, divisant par quatre la résistance des lignes. Actuellement, le WSi_2 a été remplacé par le TiSi_2 , le CoSi_2 , ou le NiSi , qui ont des résistivités plus faibles. De plus, une couche de siliciure peut être formée simultanément sur les contacts de source et drain en même temps que sur le polysilicium. On parle alors de *self-aligned silicide* (ou *salicide*).

Les avantages sont donc une plus faible résistivité des lignes polysilicium, et une résistance de contact source et drain plus faible que sans salicidation [SKOTNICKI'00].

La figure 1.7 donne la variation de la résistance carrée de lignes siliciurées à base de titane (TiSi_2) et de cobalt (CoSi_2).

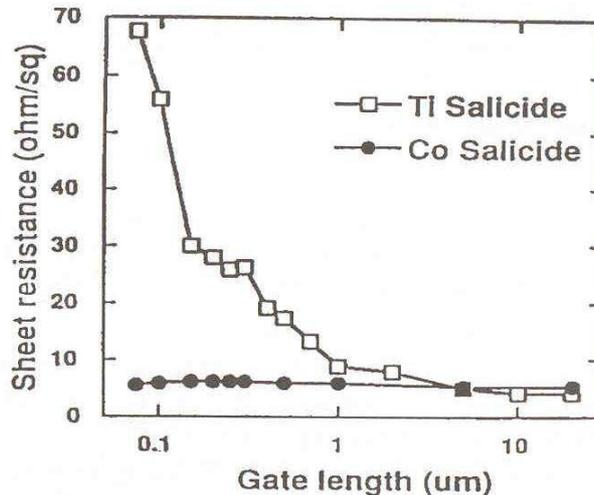


Figure 1.7 : variation de la résistance carrée de lignes siliciurée à base de titane (TiSi_2) et de cobalt (CoSi_2) [WOLF'02]

1.2.1.2 Limites de la grille en polysilicium

Le polysilicium fortement dopé est un matériau connu et qui a fait ses preuves depuis plusieurs générations technologiques. Grâce à sa nature semiconductrice, il offre la possibilité de modifier le travail de sortie de la grille, sur l'amplitude d'un gap de silicium, via l'ajustement de son niveau de Fermi (figure 1.8). En pratique, le polysilicium est dopé jusqu'à la dégénérescence, de façon que le niveau de Fermi soit dans la bande de valence (pMOS), ou dans la bande de conduction (nMOS). C'est l'architecture à grilles duales, permettant d'utiliser un canal surfacique pour les deux types de transistors. Ce procédé a permis de symétriser les dispositifs dans le substrat pour les transistors nMOS et pMOS.

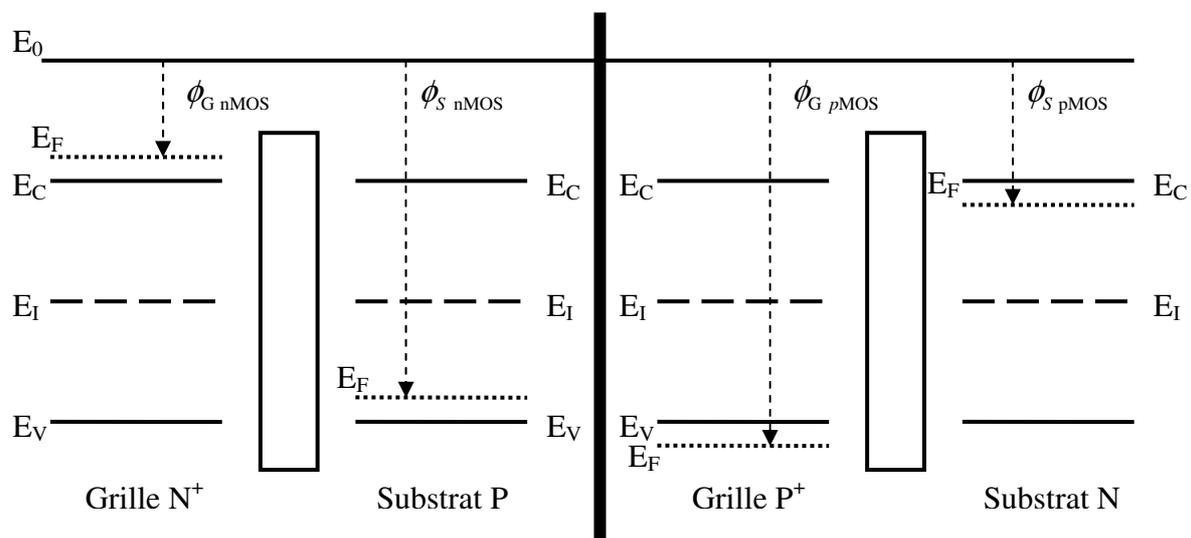


Figure 1.8 : Bandes d'énergie dans la grille et dans le substrat pour une grille N⁺ sur un substrat P (nMOS) et une grille P⁺ sur un substrat N (pMOS)

1.2.1.3 Phénomène de déplétion de grille

Le dopage de la grille est effectué en même temps que les zones de source et de drain (implantation auto alignée). Traditionnellement, l'arsenic est utilisé pour les nMOS et le bore pour les pMOS. On peut renforcer le dopage en ayant recours à une étape supplémentaire, intervenant avant la gravure de la grille : c'est le prédopage de la grille, réalisé en général avec du phosphore coté nMOS et du bore coté pMOS. Le but est de dégénérer complètement la grille sur toute son interface avec l'oxyde, afin de réduire le phénomène de déplétion dans le polysilicium à l'interface polysilicium/oxyde.

Si on considère une capacité MOS avec une grille en polysilicium dopée n⁺ et un substrat dopé p, en régime de déplétion et d'inversion, la charge négative coté substrat est

compensée par une charge positive dans le polysilicium de la grille. Les atomes donneurs ionisés forment cette charge positive. Dans ces régimes de fonctionnement, une zone de déplétion se forme dans le polysilicium. Cette zone de déplétion sera d'autant plus grande que le dopage du polysilicium est faible. D'un point de vue électrique, la zone de désertion joue le rôle d'une capacité se connectant en série avec la capacité d'oxyde, diminuant la capacité totale. Cet effet est clairement mis en évidence sur les mesures capacité-tension pour différentes valeurs de dopage de la grille (figure 1.9). Pour de faible valeur de dopage du polysilicium ($N_p = 5 \cdot 10^{18} \text{ cm}^{-3}$), la remontée de la capacité pour une tension de grille supérieure à 4 volts correspond au début de l'inversion de l'interface polysilicium/oxyde (charge constituée de trous libres), car la charge négative d'inversion du canal à compenser ne peut plus se faire uniquement par une charge positive de déplétion dans le polysilicium.

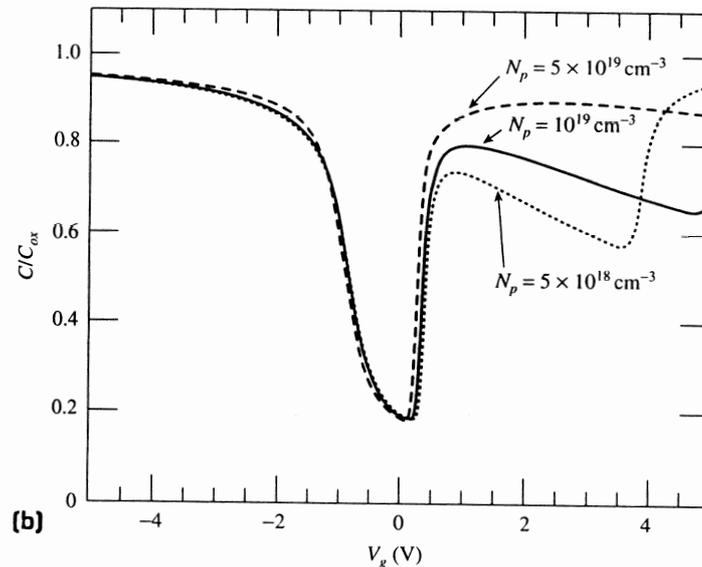


Figure 1.9 : impact de la déplétion de grille sur la capacité en fonction du dopage de grille (Capacité nMOS, $T_{ox} = 7 \text{ nm}$) [RIOS'94]

Pour la même polarisation de grille, la capacité et la charge d'inversion sont plus faibles. Ainsi le courant de drain diminue si le dopage de la grille n'est pas élevé à l'interface polysilicium/oxyde. D'un point de vue performance, cet effet augmente par exemple le temps de propagation dans les oscillateurs en anneaux (*ring oscillator*) [CHEN'96].

Pour limiter ces effets (ou les supprimer), il faut donc répartir une dose de dopant la plus forte possible, afin d'y obtenir une concentration forte et régulière. Les dopants sont introduits dans la grille, puis conduits à l'interface par diffusion préférentiellement le long des joints de grain lors des recuits. Cette démarche est limitée essentiellement pour deux raisons.

D'une part, la pénétration du bore dans l'oxyde de grille laisse des charges négatives dans la grille. Les dopants peuvent même traverser l'oxyde et venir compenser localement le substrat. On doit donc trouver un compromis entre augmenter le budget thermique pour augmenter l'activation des dopants près de l'interface, mais aussi le réduire pour limiter la pénétration du bore à travers l'oxyde. Cette limitation est propre aux grilles dopées p, car il n'a pas été recensé dans la littérature un phénomène similaire pour l'arsenic ou le phosphore [JOSSE'00].

D'autre part, la limite de solubilité des dopants (Bore, Arsenic, Phosphore) dans le polysilicium est de l'ordre de 10^{20} atome/cm³. Cela signifie que le polysilicium même fortement dopé ne peut pas atteindre la dégénérescence complète et qu'une déplétion de grille sera toujours présente.

Une alternative aux grilles polysilicium est d'utiliser des grilles métalliques. L'intérêt essentiel réside dans le fait que quel que soit le métal, aucune déplétion ne pourra s'y produire à son interface avec l'oxyde. De plus, l'utilisation de grilles métalliques permet l'intégration de nombreux diélectriques à haute permittivité, donc la réactivité avec le polysilicium nécessite un empilement de grille spécifique. Cependant, l'intégration d'une grille métal dans un procédé CMOS conventionnel est très délicate, du fait des budgets thermiques de l'ordre de 1000 °C, nécessaire pour activer les dopants. A ces températures, on risque de dégrader le métal et de l'injecter massivement dans le diélectrique de grille, et ceci d'autant plus sensiblement que l'oxyde est fin.

1.2.2 Diélectrique de grille

L'oxyde de grille SiO₂ a été utilisé comme diélectrique de grille pendant plus de 30 ans. L'épaisseur de l'oxyde n'a cessé d'être diminué avec la diminution de la longueur du canal pour augmenter le courant de drain du transistor et pour diminuer les effets de canal court. L'épaisseur de la couche diélectrique décroît linéairement avec la longueur du canal. Empiriquement on vérifie $L_e/T_{ox} \approx 45$ [THOMPSON'98] (Figure 1.10). Pour des longueurs de canal inférieures à 0.1 μm, et si le rapport L_e/T_{ox} est respecté, l'épaisseur d'oxyde devrait être inférieure à 20 Å.

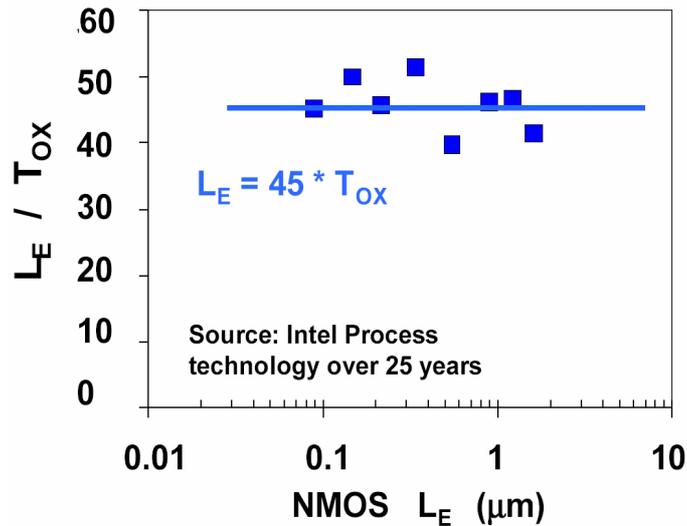


Figure 1.10 : Rapport de la longueur du canal par l'épaisseur d'oxyde en fonction de la longueur du canal pour des technologies Intel depuis les 25 dernières années [THOMPSON'98]

Pour ces épaisseurs, l'apparition d'un courant au travers de la couche diélectrique vient perturber le fonctionnement du transistor. Le courant de fuite grille - canal peut devenir aussi grand que le courant de fuite du transistor à l'état bloqué. Il devient donc difficile d'utiliser ces transistors pour des circuits logiques [KIRKLEN'00]. Le dioxyde de silicium pur n'est donc plus utilisable pour les technologies sub 0,18 μm . L'utilisation de nouveaux matériaux diélectriques à haute permittivité (matériaux *high K*) permettra de diminuer la longueur des transistors. Avec ces matériaux, tels que l'oxyde tantale Ta_2O_5 ($\epsilon_r = 26$) ou l'oxyde d'Hafnium ($\epsilon_r = 24$), des couches plus épaisses peuvent être utilisées tout en conservant les caractéristiques électriques requises pour un bon fonctionnement. Le courant de fuite au travers de l'oxyde étant lié à l'épaisseur physique du film, ces matériaux constituent les diélectriques des technologies MOS ultimes.

Une solution intermédiaire consiste à nitrurer le dioxyde de silicium. L'incorporation d'atomes d'azote dans le film diélectrique permet d'augmenter la constante diélectrique. Elle peut donc varier entre $\epsilon_r = 3,9$ (SiO_2) et $\epsilon_r = 7,8$ (Si_3N_4). De plus ce type d'oxyde permet d'obtenir une meilleure barrière à la diffusion des atomes de bore qu'un oxyde pur lors du dopage du polysilicium des transistors pMOS. Dans les nombreux procédés de nitruration utilisés, on distingue des nitrurations thermiques, où les atomes d'azote sont introduits dans un oxyde thermique par recuit, et les nitrurations utilisant des procédés de déposition chimique, tel que le procédé RPN (*Remote Plasma Nitridation*) [WOLF'02].

1.2.3 Profils de concentration de dopants dans le canal

1.2.3.1 Substrat

Le matériau de départ est du silicium monocristallin obtenu par tirage Czochralski (*Czochralski-grown single-crystal silicon*). Il existe deux types de profil de dopage du substrat. La plaque de silicium est soit dopée uniformément, soit présentant deux niveaux de dopage.

Dans le cas d'un dopage uniforme, la concentration des dopants (de type n ou p) est relativement faible. Ces wafers sont appelés *bulk wafers* ou *polished wafers*.

Une autre approche consiste à fortement doper (de type p, de l'ordre de 10^{19} atomes/cm³) pendant le processus de croissance du silicium, puis à épitaxier une couche de silicium peu dopée (de type p, de l'ordre de 10^{15} à 10^{16} atomes/cm³) sur la surface du wafer. Ces wafers sont appelés *epitaxial* ou *epi* ou *p-epi-on-p+* wafers. La figure 1.11 présente la structure d'un inverseur CMOS sur un substrat épitaxié.

Les *Bulk Wafers* étaient utilisés dans les technologies NMOS et dans quelques technologies CMOS bon marché car les wafers épitaxiés coûtent une fois et demi plus cher à réaliser qu'un *bulk wafer*.

Les wafers épitaxiés permettent un meilleur contrôle de l'effet de latch up dans les circuits logiques CMOS qui nécessitent une forte intégration. Ils permettent également une amélioration de la fiabilité de l'oxyde de grille.

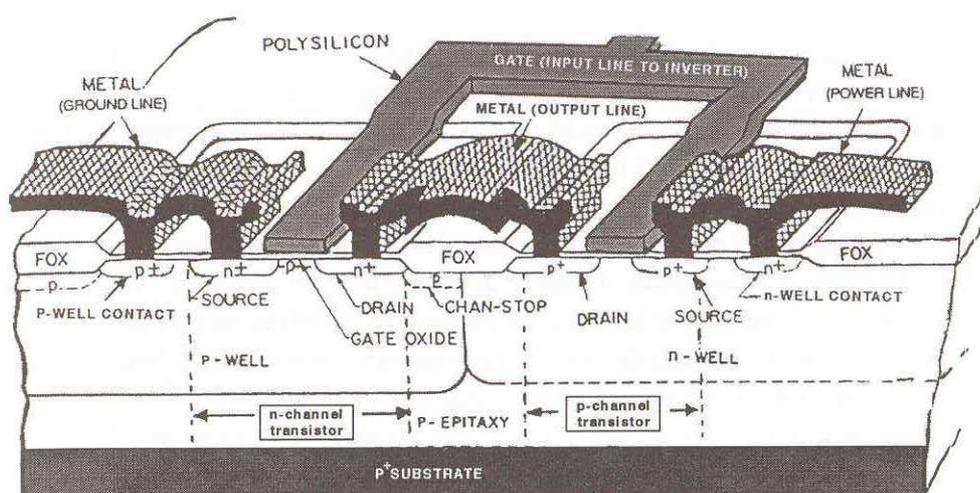


Figure 1.11 : Vue schématique d'une structure CMOS Twin-well sur substrat épitaxié [WOLF'02]

1.2.3.2 Dopage conventionnel du canal : *Diffused-Well*

La figure 1.11 présente la structure d'inverseurs CMOS où les transistors nMOS et pMOS sont réalisés dans des caissons (*well*) de type p et n, respectivement. Cette technologie CMOS appelée *Twin-well CMOS* est la plus utilisée actuellement. Ces caissons sont formés indépendamment en implantant les dopants dans une zone du substrat faiblement dopée (zone épitaxiée de type p du wafer). Ainsi on a la possibilité de réaliser une optimisation des transistors nMOS et pMOS séparément. Initialement, ces caissons étaient réalisés en implantant avec une énergie modérée, de façon à laisser les dopants près de la surface. Par un long processus de diffusion (haute température > 1100°C, longue durée 12-24h), les dopants pénétraient dans le substrat. Le profil net de concentration résultant est donné par la figure 1.12. Cette étape permettant une pénétration profonde des dopants est nécessaire pour éviter le phénomène de perçage (*punchthrough*). Cependant ce type de dopage du canal ne permet pas de réaliser des technologies CMOS en dessous de 0,5 μm , car des problèmes limitent fortement son utilisation.

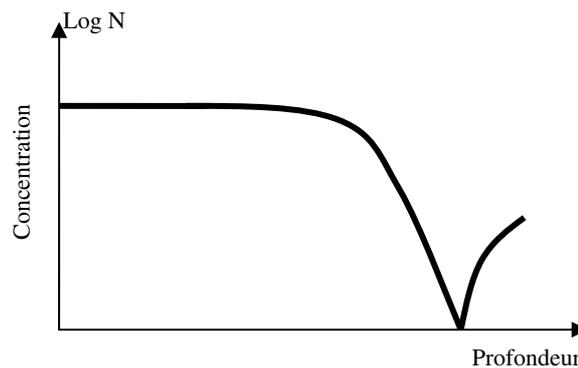


Figure 1.12 : Profil de dopage conventionnel

D'une part, une diffusion latérale intervient pendant l'étape de diffusion, ce qui réduit la densité d'intégration. Par exemple, une diffusion de 3-4 μm de profondeur va s'accompagner d'une diffusion latérale de 2-3 μm . Il y a donc une possibilité de recouvrement des caissons dans le cas de caissons étroits.

D'autre part ce type de profil (relativement constant) ne permet d'obtenir qu'un compromis entre des phénomènes tels que : le contrôle de la tension de seuil, la suppression des effets de porteurs chauds, le contrôle de l'effet de perçage, etc. Par exemple, un fort niveau de dopage est nécessaire en profondeur pour diminuer les risques de latchup. Ce critère est incompatible avec la nécessité d'avoir un dopage en surface faible pour obtenir une grande mobilité des porteurs.

Les technologies 0.5 μm et en deçà ont donc nécessité l'introduction d'une nouvelle méthode de dopage du canal. La méthode adoptée est un profil rétrograde (*retrograde well*) avec des étapes d'implantation d'ions à haute énergie.

1.2.3.3 Canal à profil de dopage rétrograde (*Retrograde Well*)

L'implantation à énergie moyenne (50-100 keV) utilisée précédemment est remplacée par une implantation à haute énergie (300 keV à 1MeV). Du fait de cette haute énergie, le pic de concentration est beaucoup plus profond dans le substrat (environ 1 μm). La concentration des impuretés diminue donc quand on s'approche de la surface du wafer. La figure 1.13 montre l'implantation de bore avec une énergie de 400 keV.

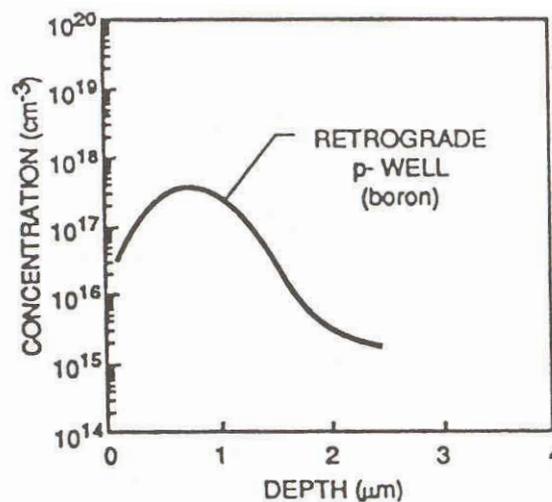


Figure 1.13 : Profil de dopage rétrograde

Le profil de concentration rétrograde a de nombreux avantages sur le profil conventionnel. La diffusion latérale est plus faible, et n'excède pas 1 μm après tous les traitements thermiques (à comparer aux 2 μm pour un procédé conventionnel). De plus, les implantateurs haute énergie peuvent placer les dopants sous les structures d'isolation (STI – *Shallow Trench Isolation* – ou LOCOS – *Local Oxidation of Silicon*) permettant ainsi une implantation après avoir formé les structures d'isolation.

En combinant ces étapes d'implantation avec les autres étapes à faible énergie d'implantation (ajustement de la tension de seuil et diminution du perçage électrostatique), le contrôle très précis des niveaux de dopage devient possible. Cela permet d'optimiser les régions critiques du transistor MOS, et donc d'améliorer les performances caractéristiques : rapidité du transistor (mobilité des porteurs et capacité drain/source), courants de fuites entre

composants, contrôle de la tension de seuil, et suppression des phénomènes de perçage. En effet, un faible niveau de dopage à l'interface Si/SiO₂ permet de maintenir une forte mobilité. Des niveaux plus forts en volume (dopage rétrograde) permet de réduire les zones de déplétion. Cette liberté d'optimisation n'est pas possible avec un dopage conventionnel du canal (*diffused wells*).

1.2.3.4 Canal à profil de dopage Super Steep retrograde

Avec la réduction des dimensions, le niveau de dopage entre source et drain doit augmenter, pour éviter les phénomènes de perçage. Par exemple, pour une longueur de grille de 0,13 μm, le contrôle de l'effet de perçage nécessite un niveau de dopage supérieur à $5 \cdot 10^{17} \text{ cm}^{-3}$. Pour éviter une trop forte réduction de la mobilité, et conformément à la courbe de mobilité universelle, le niveau de dopage surfacique doit être maintenu inférieur à $5 \cdot 10^{17} \text{ cm}^{-3}$. La transition entre la surface faiblement dopée et le volume fortement dopé doit être réalisée sur une distance très faible, et le gradient de concentration qui en résulte est très fort. On parle alors de *super-steep retrograde profile (SSR Profiles)*.

La formation de profil SSR se fait par implantation ionique. On utilise des ions plus lourds, ainsi l'indium (In) est utilisé à la place bore (B) pour les transistors nMOS, l'arsenic (As) ou l'antimoine (Sb) à la place du phosphore (P) pour les transistors pMOS. Les espèces lourdes permettent d'implanter de forts gradients de concentration, et diffusent moins pendant les phases de recuit. Le profil de dopage après recuit reste donc très proche du profil implanté.

La figure 1.14 montre les profils de dopage après recuit, pour plusieurs températures. Les profils de dopage comparant profil SSR et profil conventionnel sont présentés sur la figure 1.15. Les profils rétrogrades pour nMOS (figure 1.15a) et pMOS (figure 1.15b) font apparaître un dopage surfacique de l'ordre de $5 \cdot 10^{16} \text{ cm}^{-3}$ et un pic de dopage de $1 \cdot 10^{18} \text{ cm}^{-3}$. Le maximum de dopage pour l'indium ne peut pas dépasser 10^{18} cm^{-3} , car ce niveau de dopage constitue la limite de solubilité de l'indium dans le silicium.

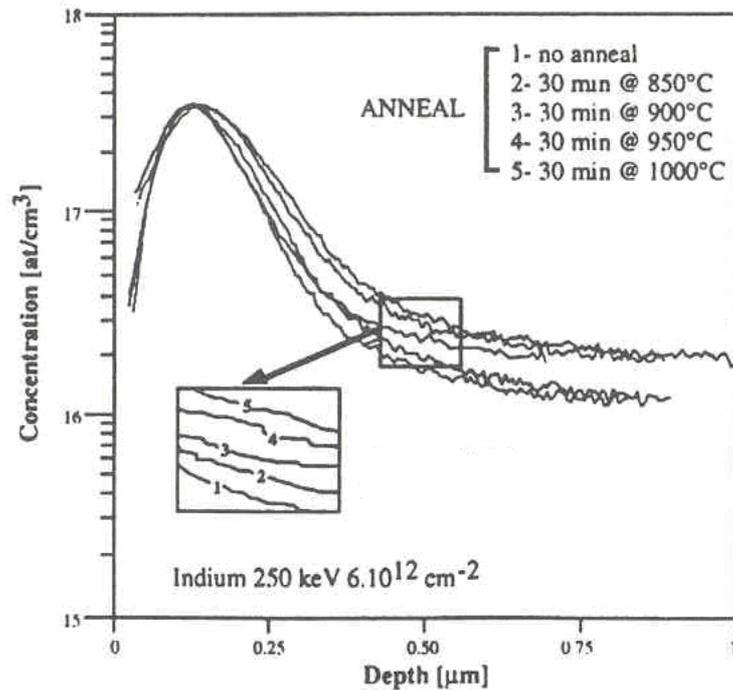


Figure 1.14 : Profil de concentration SIMS de l'Indium pour différentes températures de recuits
[BOULLION'95]

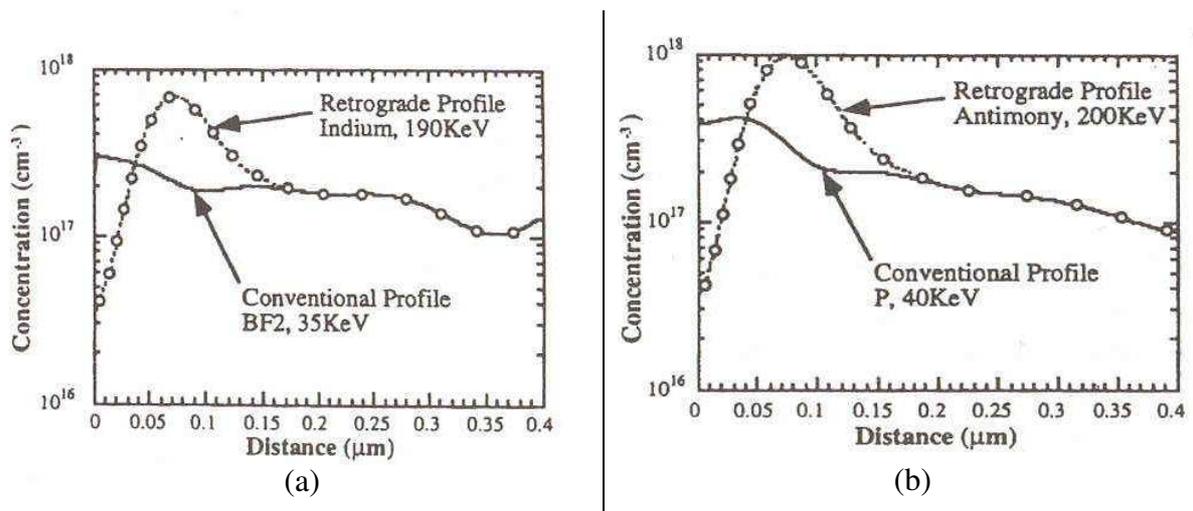


Figure 1.15 : Profils conventionnels et rétrogrades pour nMOS (a) et pMOS (b) (simulation)
[BOULLION'95]

1.2.4 Structure de Source/Drain

Dans les transistors fortement submicroniques (figure 1.16), les structures de contact source (drain) se décomposent en deux parties : une partie peu profonde appelée extension de source (drain), avec une profondeur de jonction X_J , et une zone plus profonde appelée zone de contact source (drain), avec une profondeur de jonction X_{JC} .

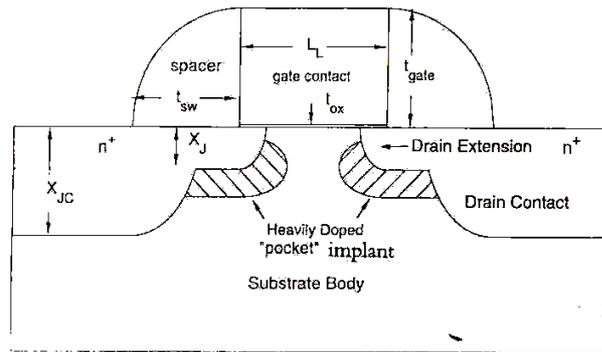


Figure 1.16 : Transistor MOS avec poches pour le contrôle des effets de perçage

Ce type de structure a été introduit quand la longueur du canal est devenue inférieure à $1,5 \mu\text{m}$. L'extension source/drain est plus faiblement dopée que la zone de contact, et ce type de structure est appelé *LDD* (*Lightly-doped drain*). La fonction première de ce type de structure était de réduire le maximum (pic) de champ électrique, et donc de limiter le phénomène d'injection de porteurs chauds. La réduction des dimensions a également nécessité la réduction des tensions d'alimentations, et le pic de champ électrique au niveau du drain s'est fortement atténué. Les structures LDD ne sont donc plus nécessaires pour supprimer les effets d'injection de porteurs chauds.

Cependant, cette architecture du drain et de la source a été conservée dans les transistors fortement submicroniques. Leur rôle est désormais de réduire les effets de canal court (SCE, *Short Channel Effects*). La région de contact est nécessaire pour pouvoir déposer une couche de salicide et permet de réduire les courants de fuite des jonctions. Cependant, la profondeur de jonction X_{JC} ne doit pas être plus grande que ce qui est nécessaire, car il en résulterait une largeur importante des espaceurs (*spacers*). Les régions d'extension source/drain (LDD) seraient alors longues et la résistance parasite d'accès R_{SDE} augmenterait, dégradant ainsi les performances du transistor.

Les dimensions des régions source/drain sont donc à choisir en essayant d'avoir peu de courant de fuite, des résistances parasites faibles, et un bon contrôle des effets de canal court.

La figure 1.17 présente les différentes longueurs caractéristiques pour des longueurs de grilles inférieures à 200 nm . Ces données issues de l'ITRS 1999 donnent les profondeurs de jonction pour les zones de contact, pour les zones d'extension, ainsi que la profondeur (SSRD) de la zone peu dopée du profil SSR (dopage canal).

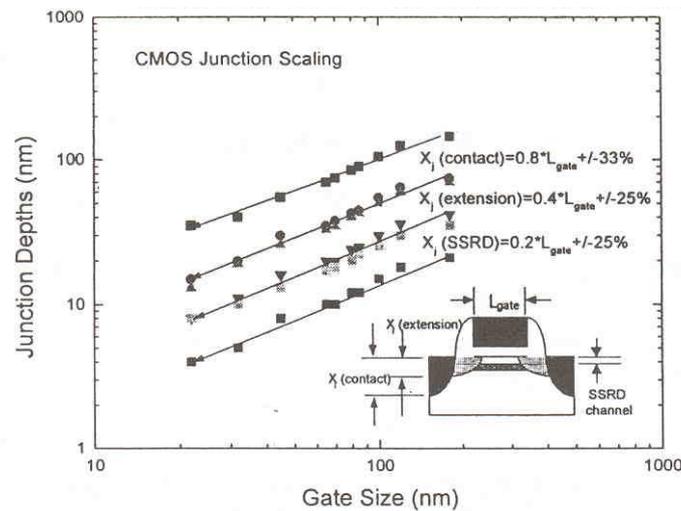


Figure 1.17 : Profondeurs de jonctions Source/Drain et profil SSR pour les technologies CMOS [ITRS'99]

1.2.4.1 Structures anti-perçage (Anti Punchthrough Structures)

La réduction des dimensions doit suivre un certain nombre de règles permettant de conserver l'intégrité électrostatique du dispositif. La pente sous le seuil S (*Subthreshold Swing*) est un bon indicateur de l'intégrité électrostatique du transistor après sa réduction géométrique. Dans le cas idéal, la pente sous le seuil d'un transistor à faible longueur de grille devrait être indépendante de la tension V_{DS} et pratiquement égale à celle du transistor à canal long correspondant [TSIVIDIS'99].

Les premières structures réalisées sont décrites sur la figure 1.16. Des zones surdopées avec le même type de dopant que le substrat sont implantées sous les zones d'extension de source et de drain. La zone où se crée le canal reste donc faiblement dopée, permettant une tension de seuil faible et une mobilité élevée, et les zones de déplétions sont réduites dans les zones surdopées. Ces régions sont appelées halo (*halo anti punchthrough structures*). Les halos sont réalisés en implantant avec un faisceau normal à la surface du wafer, après l'implantation des LDD, avec une énergie plus grande que pour l'implantation des extensions.

Dans les technologies plus avancées, cette seconde implantation est réalisée avec une implantation ionique avec un grand angle par rapport à la surface ($20^\circ - 45^\circ$). Cette technique permet une implantation plus précise des niveaux de dopage et de la profondeur. On parle alors de LATIPS (*Large Angle Tilted Implanted Punchthrough Structure*) ou de poches (*pockets*). La figure 1.18 montre les niveaux de dopages résultant d'une implantation de poches pour un transistor nMOSFET.

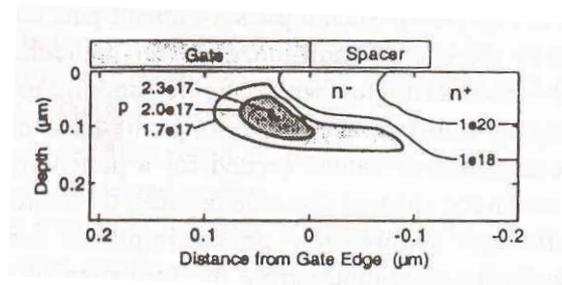


Figure 1.18 : Niveaux de dopage des poches pour un transistor nMOS [HORI'94]

L'implantation de poches améliore les performances des transistors. Les capacités des jonctions sont réduites jusqu'à 65% et la pente sous le seuil augmente (suppression des effets de canaux courts). D'autres effets liés à l'introduction de poches sont reportés dans la littérature, notamment la diminution de la mobilité effective avec la réduction de la longueur du canal pour une même technologie.

La figure 1.19 donne les niveaux de dopages nécessaires pour un bon fonctionnement d'un transistor nMOS avec une longueur de grille de 25 nm. Les niveaux de dopages très élevés de source et de drain côtoient les poches également très dopées. La difficulté dans la réalisation de structures avancées est donc la nécessité d'avoir non seulement de forts niveaux de dopages mais également de forts gradients de concentration.

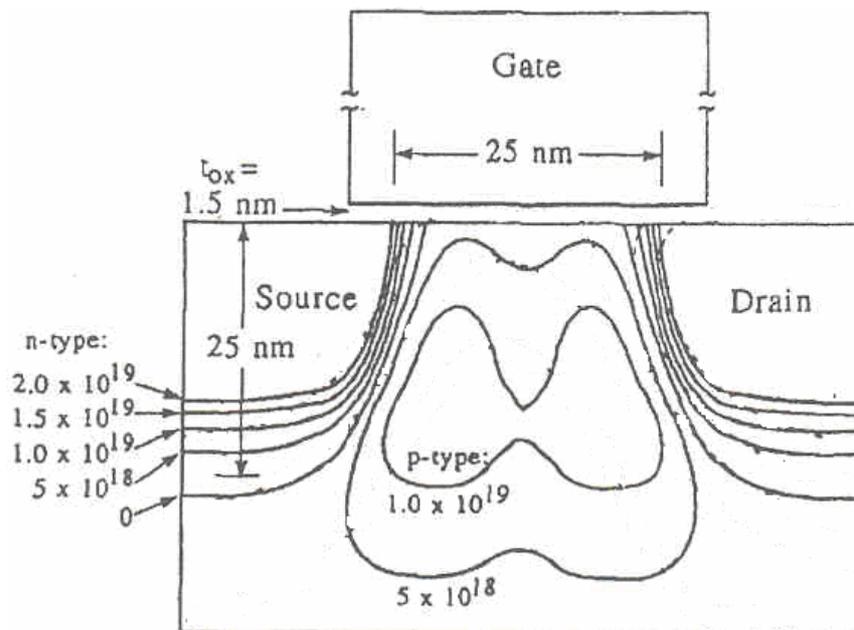


Figure 1.19 : Niveaux de dopage pour un transistor pMOS avec une longueur de grille de 25 nm [TAUR'98]

1.2.4.2 Résistances d'accès des transistors MOS fortement submicroniques

La zone active du transistor (canal) étant située sous la grille, le courant provient des lignes d'interconnexions métalliques et passe par le drain et la source. La figure 1.20 montre les lignes de courants dans une structure LDD. On peut séparer la résistance totale R_S en plusieurs résistances correspondant à des effets différents.

R_{C0} est la résistance de contact entre le métal et la région de source et de drain, R_{sh} la résistance de diffusion dans les régions de source et de drain, R_{sp} la résistance due au confinement des lignes de courants à proximité du début du canal, et R_{idd} la résistance de la couche d'accumulation dans la zone faiblement dopée (LDD).

Les résistances d'accès au canal réduisent le courant I_{DSAT} du transistor. L'architecture des zones de source et drain est déterminée de façon à limiter cette dégradation à 10% de I_{DSAT} . On peut également formuler ce critère en comparant les valeur de R_{SAT} (résistance du canal quand le transistor est dans l'état passant) avec la résistance d'accès. Par exemple si la résistance d'accès atteint 10% de R_{SAT} , I_{DSAT} est réduit de 8% [WOLF'02]. Avec la réduction des dimensions, R_{SAT} va diminuer fortement. Cela implique donc une réduction des résistances d'accès pour pouvoir maintenir une atténuation acceptable de I_{DSAT} .

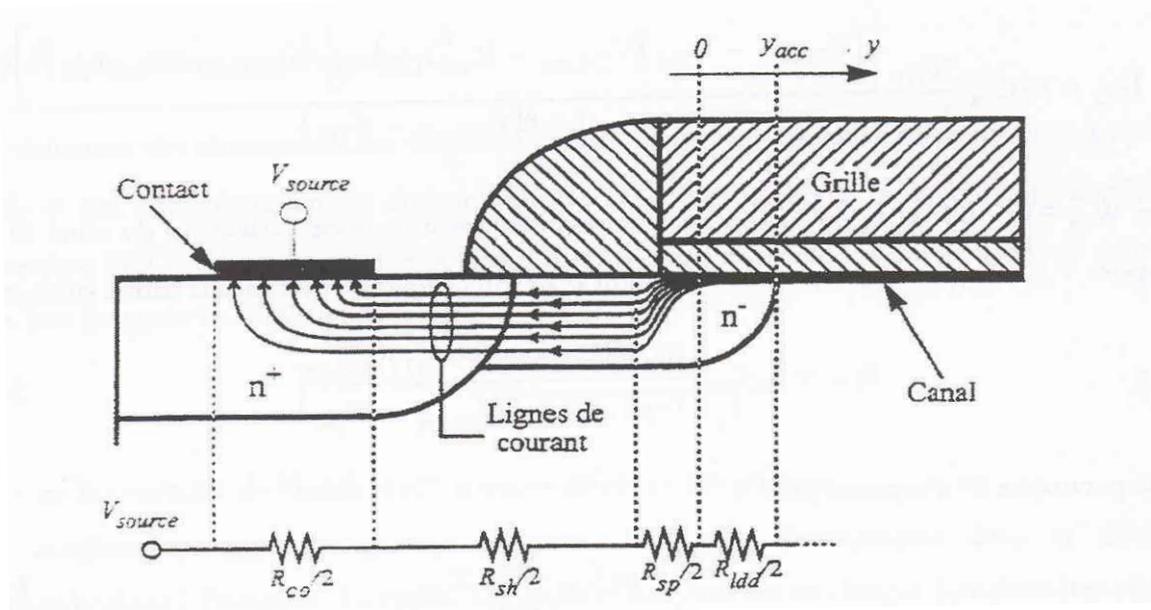


Figure 1.20 : Représentation des différentes composantes des résistances d'accès au canal du transistor [BRUT'96]

La valeur de la résistance d'accès dépend bien sûr de l'architecture de la source et du drain. Ces zones doivent avoir une profondeur de jonction correcte, mais il a été montré qu'un

fort gradient de niveau de dopage permet de limiter la résistance R_{sp} associée au confinement des lignes de courant aux extrémités de la couche d'inversion. Les mesures effectuées pour caractériser l'impact du gradient de dopage latéral des extensions de source et drain permettent de dégager une règle pour dimensionner ces zones de transition. Le niveau de dopage doit être une décade plus faible pour une distance d'environ 3% de la longueur de la grille [CURRENT'00]. Cette règle permet d'obtenir une résistance R_{sp} plus faible que 5% de R_{SAT} . Pour une technologie MOS 100 nm, le niveau de dopage doit donc diminuer d'une décade sur 3 nm. Cette jonction abrupte est obtenue par l'implantation de Bore à 0.2 keV, sans diffusion thermique.

Avec la réduction des dimensions, il est clair que de nouvelles techniques devront être mises en œuvre pour obtenir des profils de jonctions plus abruptes. Les figures 1.21 et 1.22 montrent des résultats de simulation pour plusieurs profils de dopage latéral [VILLANUEVA'03]. On peut remarquer que la longueur électrique du canal est également affectée par le profil de la jonction.

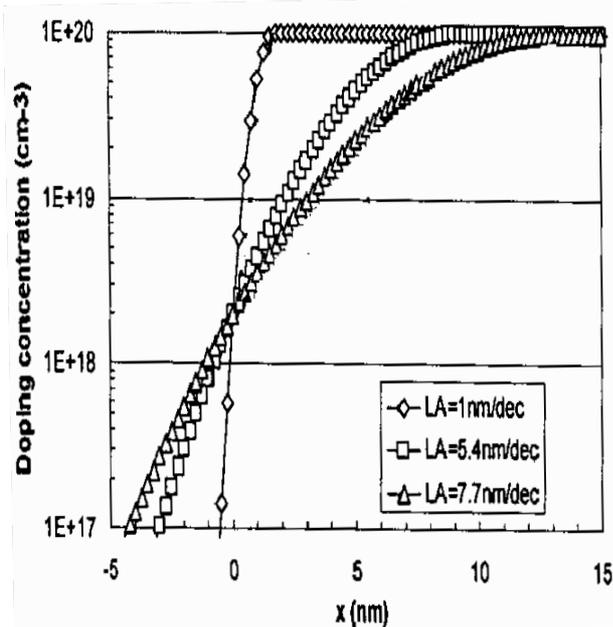


Figure 1.21 : Différents profils de dopages des extensions Source/Drain

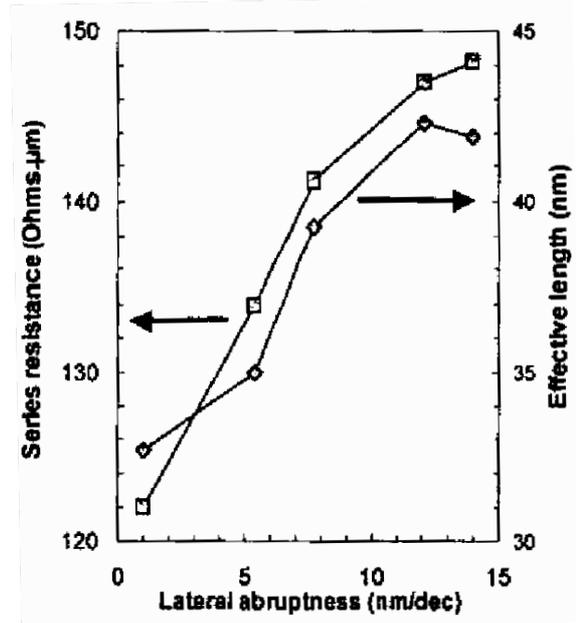


Figure 1.22 : Résistance d'accès et longueur effective du canal en fonction du gradient de concentration

1.2.4.3 Zones de recouvrement des extensions de source et de drain

L'architecture des zones de source et drain sont d'une importance capitale dans la réduction des dimensions. Les performances des dispositifs sont également très sensibles aux zones de recouvrement de source et de drain. Des zones de recouvrement trop grandes entraînent des fuites excessives du transistor à l'état bloqué, et la capacité liée aux zones de recouvrement va augmenter (Capacité Miller). Une capacité trop élevée va limiter les temps de commutations et les courants à l'état bloqué ne doivent pas excéder une valeur de $1\text{nA}/\mu\text{m}$ pour des raisons liées à la conception des circuits.

D'un autre coté, si le recouvrement est trop faible ou voire absent, le courant de saturation I_{DSAT} va être fortement diminué. On notera que la capacité Miller tend vers une capacité minimale non nulle (C_{fringe}), correspondant aux effets de bord de la grille. La figure 1.23 montre l'évolution du courant de saturation pour trois transistors nMOS issus d'une technologie $0.25\ \mu\text{m}$ avec plusieurs longueurs de recouvrement [THOMPSON'98]. Il apparaît clairement que la longueur des zones de recouvrement ne doit pas être inférieure à $20\ \text{nm}$ pour conserver un courant de saturation maximum. En l'absence de recouvrement, la fin du canal d'inversion n'est pas couplée électriquement avec les extensions de source et de drain. Dans ce cas, elles sont séparées par une zone de haute résistivité, donc la résistance parasite de source et de drain augmente, faisant de ce fait diminuer le courant de saturation.

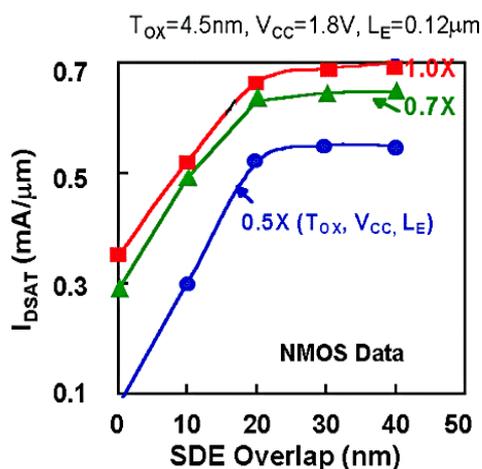


Figure 1.23 : I_{DSAT} en fonction de la longueur des zones de recouvrement pour un nMOS [THOMPSON'98]

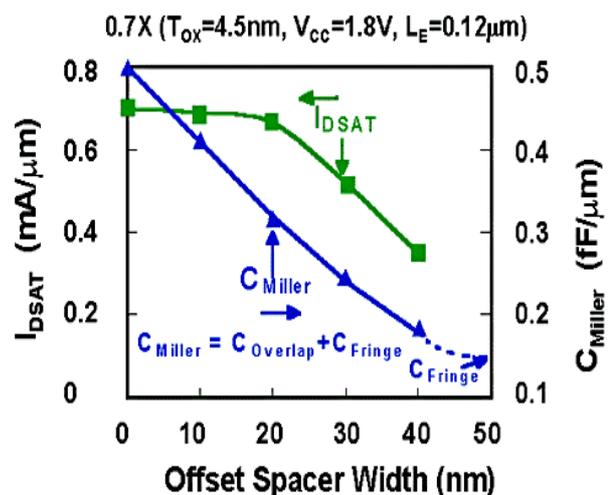


Figure 1.24 : I_{DSAT} et C_{MILLER} en fonction de la longueur de l'espaceur latéral (*offset spacer*) [THOMPSON'98]

Le contrôle de la longueur des zones de recouvrement de source et de drain lors de la fabrication se fait grâce à la présence ou non d'espaceurs latéraux (*offset-sidewall spacer*).

La figure 1.25 montre la structure faisant apparaître ces espaceurs. L'implantation des extensions de source et drain ne se fait qu'après avoir formé ces espaceurs, et leur épaisseur varie de zéro à près de 50 nm. La largeur optimum est déterminée de façon à trouver le meilleur compromis entre courant de saturation maximum et courant de fuite/capacité Miller minimum. La figure 1.24 donne un exemple de dimensionnement optimal de la largeur des espaceurs. Dans le cas du transistor étudié, le meilleur compromis est pour une largeur des espaceurs égale à 20 nm.

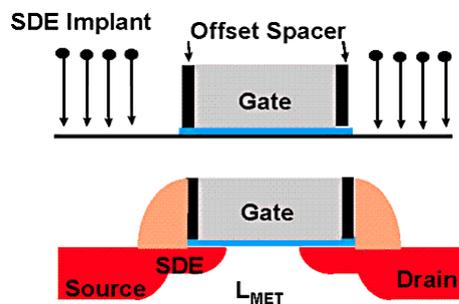


Figure 1.25 : structure de test pour déterminer la longueur optimale des zones de recouvrement [THOMPSON'98]

1.3 Limites de la technologie CMOS Silicium

1.3.1 Les limites de la réduction conventionnelles des dimensions

L'ITRS 2001 prévoyait que la technologie CMOS planaire conventionnelle devait atteindre ses limites pour une longueur de grille de 50 nm (nœud technologique 90 nm). La course à la miniaturisation au delà du nœud 45 nm accélère l'introduction de nouvelles technologies, de manière à « pousser » encore plus loin l'utilisation du silicium et donc continuer avec la loi de Moore.

Pour ces dimensions, le transistor MOS conventionnel se heurte à des limites physiques fondamentales. Les plus citées sont [HUTCHBY'02] :

- L'effet tunnel au travers du film diélectrique
- L'effet tunnel des porteurs de la source vers le drain (à l'état bloqué), et du drain vers le substrat

- Le contrôle de la densité et de la position des dopants dans le canal et dans les régions source/drain (optimisation du rapport I_{ON}/I_{OFF})
- La valeur limite de la pente sous le seuil.

Ces limitations ont donné lieu à des prédictions pessimistes sur l'avenir du transistor MOS conventionnel. Cependant, on constate que les technologies actuelles ont réussi à aller au delà de ces limitations, permettant ainsi la réduction des dimensions des transistors MOS conventionnels.

Le problème vient des matériaux utilisés, à savoir le silicium, l'oxyde de silicium et le polysilicium. L'introduction de nouveaux matériaux est donc nécessaire pour le futur, et la technologie CMOS planaire conventionnelle sera remplacée par de nouvelles architectures de composants, tel que la technologie SOI, multi grilles, grille enrobante, etc. [ITRS'03b]

1.3.2 Technologie SOI (Silicon on Insulator)

De nombreux problèmes liés à la réduction des dimensions des transistors MOS conventionnels sont supprimés par l'introduction de wafers SOI. De plus, ces derniers offrent de nouvelles possibilités en termes de conception de circuits. Les transistors CMOS SOI permettent un meilleur contrôle du courant à l'état bloqué (I_{OFF}), et leur vitesse de fonctionnement est augmentée de 20-50% par rapport aux technologies conventionnelles.

L'utilisation de wafers SOI permet le prolongement de la *roadmap* des technologies CMOS planaires d'environ une décade, et de dépasser la limite associée aux grilles de longueur 50 nm des transistors CMOS conventionnels. La figure 1.26 présente les longueurs de grille et les profondeurs de jonction pour des dimensions sub-100 nm et montre que les dimensions les plus petites seront atteintes avec des transistors CMOS SOI.

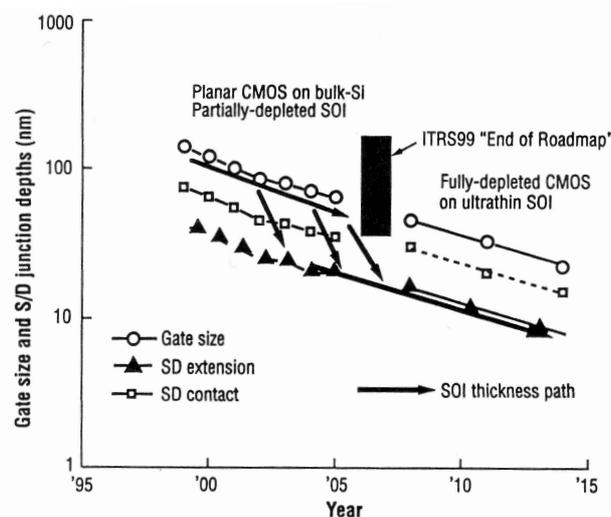


Figure 1.26 : Extension de la Roadmap grâce à la technologie SOI [ITRS'01].

Les avantages des technologies CMOS sur substrat SOI sont nombreux :

- Réduction de la consommation des circuits
- Réduction des capacités de jonction
- Elimination du phénomène de latchup
- Pente sous le seuil plus importante (caractéristique de mise à l'état bloqué amélioré)
- Compatibilité avec les précédés CMOS conventionnels

Les technologies SOI actuellement commercialisées utilisent des transistors partiellement déplétés *PD-SOI MOSFETs (Partially-depleted SOI MOSFETs)*. Les vitesses de commutations des circuits réalisés avec de tels transistors sont de 20-30% plus rapides comparées aux mêmes circuits réalisés en technologie MOS conventionnelle. Les transistors SOI totalement déplétés *FD-SOI MOSFETs (Fully-depleted SOI MOSFETs)* présentent des avantages encore plus intéressants, notamment dans le contrôle des effets de canal court [COLINGE'98]. Pour ces technologies, la difficulté réside dans la réalisation d'une couche de silicium fine avec une densité de défauts très faible, ce qui limite fortement son utilisation pour les circuits intégrés commerciaux. Néanmoins, Intel a annoncé l'utilisation de FD-SOI MOSFETs pour ces circuits intégrés à partir de 2005. A l'heure actuelle, Intel n'utilise toujours pas de technologie SOI, préférant les transistors sur silicium contraint (augmentation de la mobilité, et aucun changement dans l'architecture du composant), alors que AMD l'utilise pour la totalité de ces microprocesseurs.

1.4 Nouveaux effets physiques dans les transistors MOS nanométriques

1.4.1 Effets quantiques

1.4.1.1 Quantification de l'énergie des porteurs

Dans le cas d'une structure MOS à une dimension (Capacité MOS idéale sans courant de fuite), l'équation de Poisson permet de calculer la nature et la densité de charges en fonction de la polarisation. Le résultat est présenté figure 1.27, pour du silicium de type p dopé avec une concentration d'accepteurs $N_A=5.10^{17} \text{ cm}^{-3}$. Le calcul classique est développé en considérant d'une part, que tous les donneurs et les accepteurs sont ionisés, et d'autre part, que les porteurs libres obéissent à la statique de Boltzmann. La première hypothèse est peu justifiée en régime d'accumulation et à basse température, car les impuretés majoritaires sont alors en partie neutralisées par les porteurs libres. La deuxième hypothèse est mise en défaut en régimes d'accumulation et d'inversion, en raison du fait que le semi-conducteur est dégénéré.

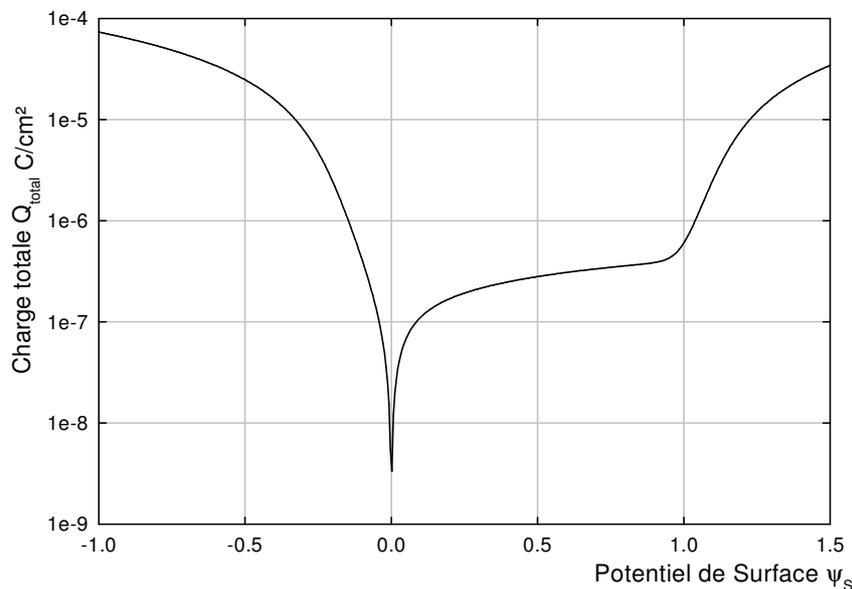


Figure 1.27 : Charge d'espace dans le semiconducteur en fonction du potentiel de surface ($T_{\text{ox}}=1.4 \text{ nm}$, $N_A=5.10^{17} \text{ cm}^{-3}$)

Les résultats donnent néanmoins une bonne approximation de l'ordre de grandeur des densités de porteurs. La figure 1.28 montre le potentiel et la densité de porteurs dans une capacité MOS pour une polarisation de grille $V_{GB}=2 \text{ V}$ (régime de forte inversion). La densité

de charge de déplétion est de l'ordre de 10^{11} cm^{-2} , et la densité superficielle de charge d'inversion est de l'ordre de 10^{12} cm^{-2} . Si on suppose le dopage homogène ($N_A=5 \cdot 10^{17} \text{ cm}^{-3}$), la largeur de la zone de déplétion est de l'ordre de 60 nm. La largeur moyenne de la couche d'inversion y_{inv} est de l'ordre de la dizaine d'Angstrom.

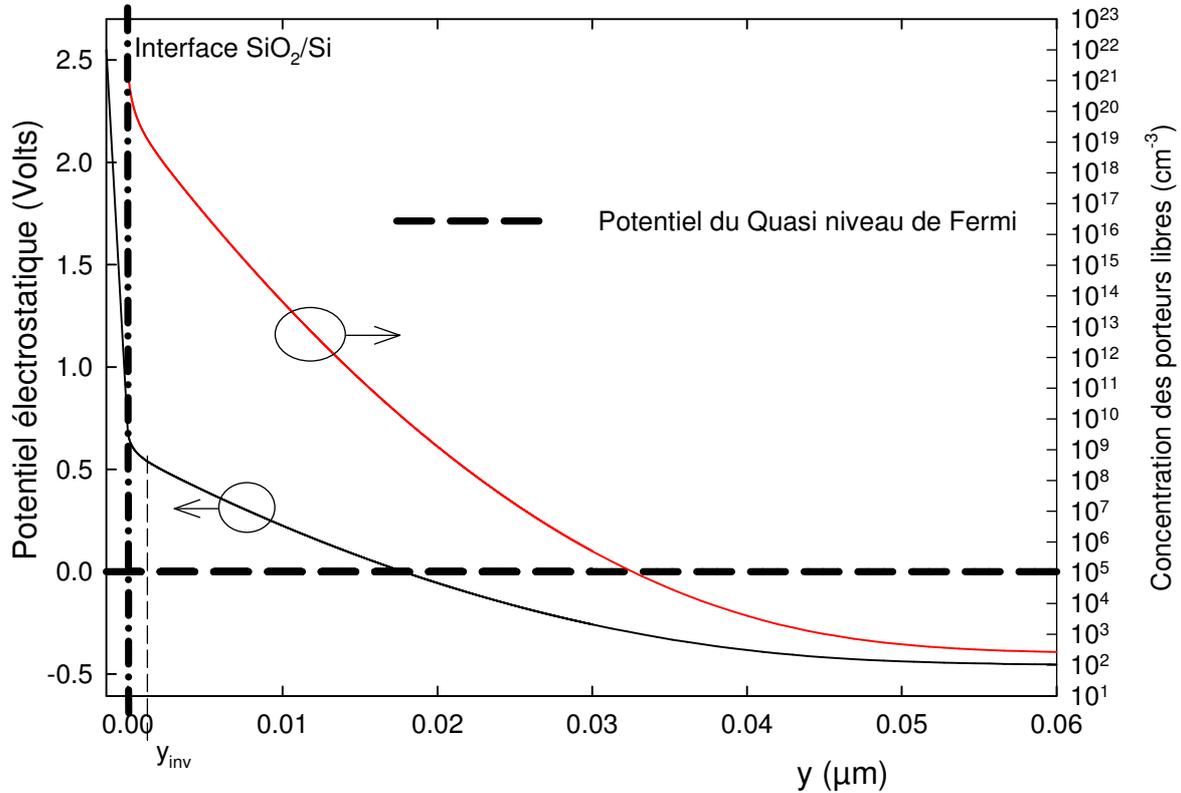


Figure 1.28 : Potentiel électrostatique et concentration des porteurs libres en régime de forte inversion ($N_A=5 \cdot 10^{17} \text{ cm}^{-3}$, $T_{\text{ox}}=1.4 \text{ nm}$, $V_{\text{GB}}=2 \text{ Volts}$, $V_{\text{FB}}=-1 \text{ Volt}$).

Les électrons de la couche d'inversion sont donc confinés dans un puits de potentiel de quelques dizaines d'Angstrom d'épaisseur. Ils sont animés d'un mouvement brownien caractérisé par un libre parcours moyen $l=v_{\text{th}}\tau_c$ où v_{th} est la vitesse thermique des porteurs et τ_c leur temps de collision. Un calcul rapide donne $l=100$ à 1000 \AA ; la largeur du puits de potentiel est donc inférieure au libre parcours moyen des électrons. Ces électrons se comportent comme des particules libres dans un puits de potentiel, leur mouvement est conditionné par les réflexions sur les faces du puits, et leurs états sont quantifiés.

Dans les technologies CMOS avancées, la combinaison de forts dopages substrat et la très faible épaisseur du diélectrique de grille induisent un fort champ électrique à l'interface oxyde/semi-conducteur. La courbure des bandes d'énergie est très importante, le puits de potentiel ainsi créé est donc suffisamment étroit pour quantifier l'énergie des porteurs. Les électrons (ou les trous) confinés dans une couche d'inversion ou d'accumulation se

comportent comme un gaz d'électrons (ou de trous). Leur mouvement est libre dans le plan de la structure (interface Si/SiO₂) et quantifié dans la direction perpendiculaire. Le calcul classique (résolution de l'équation de Poisson associée à la statistique de Boltzmann) de la distribution des porteurs libres n'est qu'approché, une étude détaillée passe par un traitement quantique du problème.

1.4.1.2 Influence des effets quantiques

Le spectre d'énergie des porteurs à la surface du semi-conducteur devient discret. Les figures 1.29 et 1.30 illustrent les principaux effets liés au confinement quantique en régime de forte inversion. Compte tenu de la nature multi-vallée du silicium, deux spectres d'énergies sont à calculer, un pour les deux vallées normales et l'autre pour les quatre vallées parallèles. L'éclatement de la bande de conduction en sous bandes discrètes ($E_0, E_1, E_2 \dots$) met en évidence le phénomène d'élargissement du gap du semi-conducteur (ΔE_g), ainsi que le phénomène de déplacement de la distribution des porteurs par rapport à l'interface Si/SiO₂ (Δy). La densité de porteurs libres s'annule à l'interface, ce qui augmente la distance moyenne $\overline{y_{qm}}$ du centroïde d'inversion par rapport à une solution classique (résolution de l'équation de Poisson couplée à la statistique de Boltzmann) d'une quantité Δy .

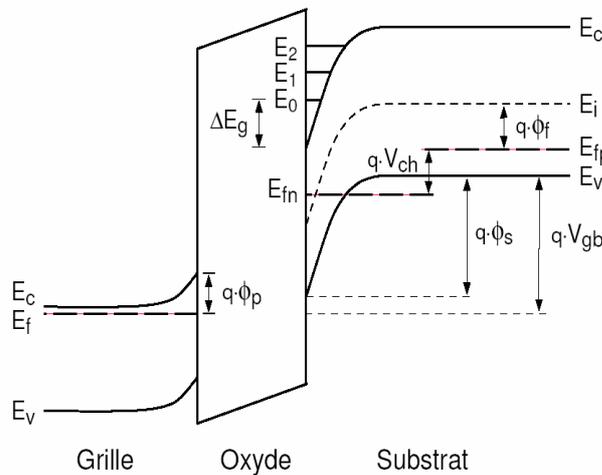


Figure 1.29 : Diagramme de bandes d'une capacité n-MOS en régime de forte inversion

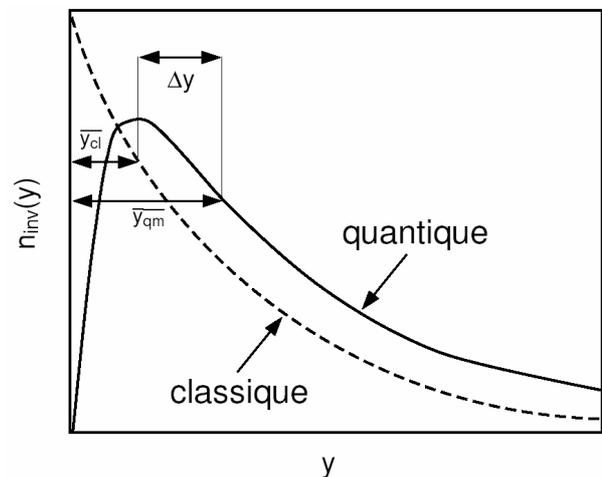


Figure 1.30 : Densité d'électrons n_{inv} en fonction de la distance transversale y

1.4.1.3 Structure à une dimension : Capacité MOS

La figure 1.31 illustre le principal impact des effets quantiques : une forte atténuation de la capacité en forte inversion et en accumulation. En effet, en absence de phénomène de déplétion dans le polysilicium, la capacité par unité de surface dans ces régimes de fonctionnement est égale à $C_{ox} = \frac{\epsilon_{ox}}{T_{ox}}$.

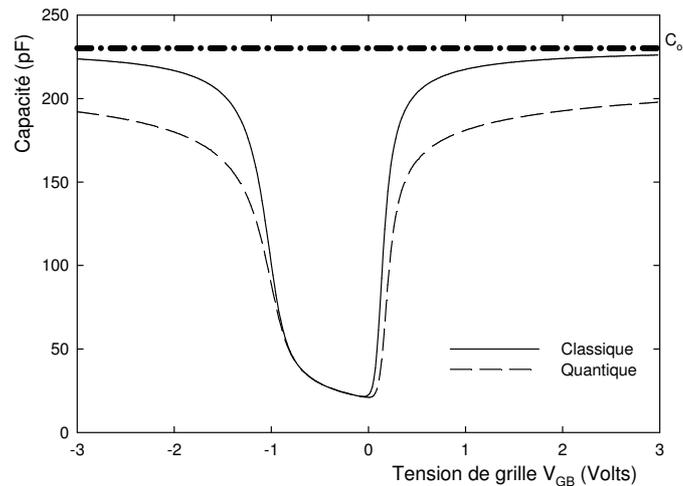


Figure 1.31 : Capacité d'une structure MOS à une dimension calculée en prenant ou en ne prenant pas en compte les effets quantiques. ($N_A=5 \cdot 10^{17} \text{ cm}^{-3}$, $T_{ox}=1.4 \text{ nm}$, $V_{FB}=-1 \text{ Volt}$)

L'évolution classique de la charge d'inversion avec la tension appliquée sur la grille peut être évaluée par la relation $Q_{inv} \approx C_{ox} (V_{GB} - V_{th})$. Lorsque les effets quantiques sont pris en compte, la charge d'inversion est modifiée. D'une part, la quantification de l'énergie des porteurs provoque une augmentation de la tension de seuil. D'autre part, la capacité du diélectrique C_{ox} est plus faible ; ce phénomène est lié à l'apparition d'une deuxième capacité en série avec la capacité du diélectrique située dans le substrat (*dark space*) (voir figure 1.30). De plus, cette nouvelle valeur de capacité d'oxyde (C_{oxeff}) est non constante et dépend de la tension de polarisation V_{GB} .

L'approche quantique nécessite de coupler l'équation de Poisson et l'équation de Schrödinger. Cette approche souvent appelée calcul auto cohérent permet d'obtenir la simulation « référence ». Un calcul auto cohérent Poisson/Schrödinger requiert des temps de calculs très importants et n'est possible que pour des structures simples à une dimension. Ce type d'approche devient très complexe dans le cas d'un transistor MOS (structure à deux dimensions), et n'est applicable que pour des transistors à canal long. [JANIK'98]

Compte tenu de l'importance des mesures capacité-tension dans la caractérisation des transistors MOS fortement submicroniques, il semble évident qu'on ne peut plus négliger les effets quantiques dans la modélisation des capacités MOS.

1.4.1.4 Structure à deux dimensions : Transistor MOS

Les effets quantiques sur le comportement d'un transistor MOS sont directement liés à ceux décrits au paragraphe précédent. Le premier effet discuté dans la littérature est la modification de la tension de seuil. La figure 1.32 met en évidence l'augmentation de la tension de seuil, la pente sous le seuil (*subthreshold swing*) restant inchangée. Les travaux de Van Dort [VAN'92] ont ouvert la voie à de nombreux modèles de tension de seuil basés sur l'élargissement du gap du semiconducteur et le déplacement du centroïde de la couche d'inversion. Le potentiel de surface est également affecté par les effets quantiques, et le champ électrique à l'interface est modifié. La figure 1.33 montre l'impact sur le maximum de transconductance, qui se trouve diminué.

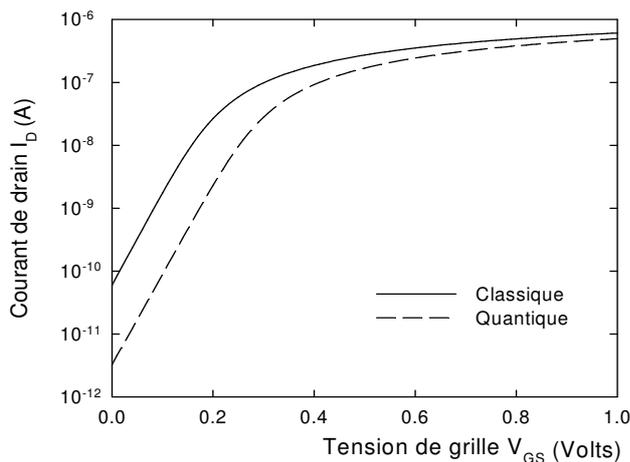


Figure 1.32 : Impact des effets quantiques sur la tension de seuil

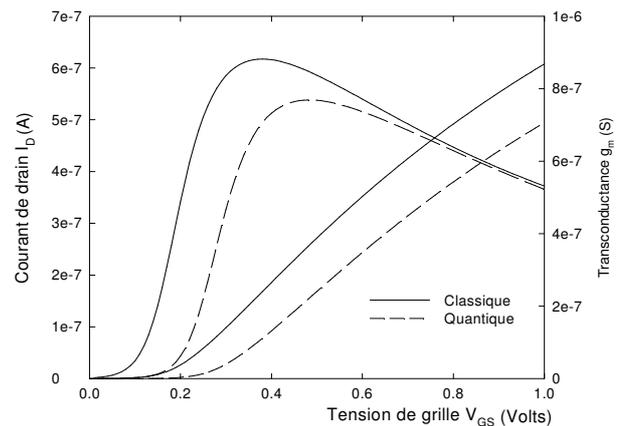


Figure 1.33 : impact des effets quantiques sur la transconductance

Les effets quantiques sont donc un élément très important dans la modélisation électrique des transistors MOS actuels, puisque l'augmentation de la tension de seuil peut atteindre 200 mV pour les technologies MOS les plus récentes. Pour la modélisation électrique compacte (BSIM4, EKV 3, MM11), la tendance est plutôt d'utiliser des paramètres effectifs permettant d'ajuster au mieux les résultats de simulation avec les mesures expérimentales. Cette approche, même si elle est basée sur des modèles physiques et des approximations justifiées, ne permet pas toujours de conserver un sens physique aux paramètres ainsi extraits. Le gros avantage réside dans les faibles temps de calculs puisque les équations implantées sont des équations explicites ne nécessitant pas d'itérations numériques.

1.4.2 Courant de grille dans les oxydes ultra minces

L'oxyde de silicium pur (SiO_2) possède de très bonnes propriétés en tant que diélectrique de grille dans les structures MOS grâce à la grande barrière de potentiel et la faible densité de pièges à l'interface Si/SiO_2 . Les phénomènes de courants de fuite ont fait l'objet de nombreuses études. Pour les couches d'épaisseur supérieure à 6 nm, le mécanisme de courant de fuite s'explique par l'effet tunnel Fowler-Nordheim. A partir de ce modèle, la masse effective des électrons traversant l'oxyde a pu être déterminée. De plus, les phénomènes de courant tunnel dans les couches ultra minces ont été examinés dès 1974. L'effet tunnel direct domine dans les couches d'épaisseur inférieure à 2 nm.

1.4.2.1 Classification des mécanismes tunnel

L'injection de charges à travers la couche d'oxyde de grille des composants MOS peut être liée à plusieurs mécanismes. On distingue ainsi essentiellement 3 mécanismes tunnel.

1.4.2.1.1 L'effet tunnel Fowler-Nordheim

Cet effet est caractérisé par l'injection de porteurs dans la bande de conduction de l'oxyde. Il se manifeste essentiellement lorsque la tension d'alimentation du dispositif est suffisante pour courber de manière notable le profil de bande de conduction du dispositif MOS (figure 1.34). En effet, les porteurs « voient » alors une épaisseur d'oxyde plus faible, et peuvent donc traverser la barrière de potentiel plus facilement. Une fois injectés dans la bande de conduction de l'oxyde, les électrons sont accélérés par le champ électrique présent dans l'oxyde vers la grille, provoquant ainsi un courant de grille. Dans le cas d'une structure Si/SiO_2 , la différence de potentiel aux bornes de l'oxyde doit être supérieure à 3 Volts pour être dans le cas d'un tel mécanisme. Cette injection est appelée injection Fowler-Nordheim (FN).

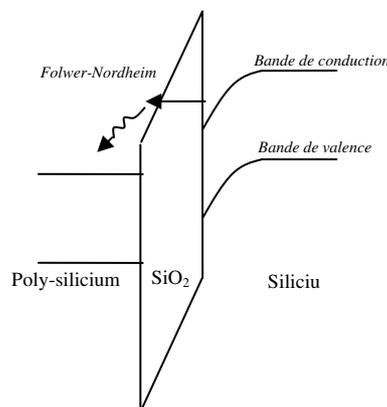


Figure 1.34 : Effet tunnel Fowler-Nordheim

Ce type d'injection est très connu car c'est un des mécanismes utilisés pour la programmation des mémoires non volatiles (EEPROMs). C'est aussi une méthode pour réaliser des contraintes électriques sur les structures MOS afin de déterminer la fiabilité du composant.

Pour déterminer la probabilité qu'un électron traverse la barrière de potentiel par injection Fowler-Nordheim, l'équation de Schrödinger doit être résolue pour une barrière triangulaire. L'approximation WKB est utilisée pour obtenir l'expression la densité de courant Folwer-Nordheim [DEPAS'95] :

$$J_{FN} = \alpha E_{ox}^2 \exp\left(-\frac{\beta}{E_{ox}}\right) \quad 1.1$$

$$\text{avec } \alpha = \frac{q^3 m}{8\pi \hbar m^*}, \quad \beta = \frac{4(2m^*)^{1/2} \phi^{3/2}}{3\hbar q},$$

q la charge de l'électron, k la constante de Boltzmann, T la température absolue, E_{ox} le champ électrique dans l'oxyde, m la masse de l'électron libre, m^* la masse effective de l'électron dans l'oxyde, et ϕ la hauteur de barrière énergétique à l'interface Si/SiO₂.

En traçant $\ln(J_{FN}/E_{ox}^2)$ en fonction de $1/E_{ox}$ (diagramme connu sous le nom de *Fowler-Nordheim plot*), on obtient une droite si le phénomène de conduction est du FN pur. On peut donc extraire la hauteur de la barrière de potentiel au niveau de l'interface Si/SiO₂. Cependant, cette approche n'est valable qu'en l'absence de charges électriques dans l'oxyde. Les charges présentes dans l'oxyde modifient la barrière de potentiel au niveau de l'interface Si/SiO₂.

La figure 1.35 présente l'évolution du champ électrique dans l'oxyde depuis les 30 dernières années [DEGRAEVE'99]. Il apparaît que le champ électrique augmente car les tensions d'alimentation ne diminuent pas linéairement avec la réduction de l'épaisseur d'oxyde. De ce fait, les tensions à appliquer sur les technologies MOS récentes (oxyde ultra mince) pour être dans un régime FN provoquent un claquage de l'oxyde. Par exemple, pour un oxyde de 2 nm d'épaisseur, la tension à appliquer sur la grille pour être dans le régime FN est de l'ordre de 4 Volts, correspondant à un champ électrique de l'ordre de 13MV/cm. Ce régime n'interviendra donc pas dans cette étude.

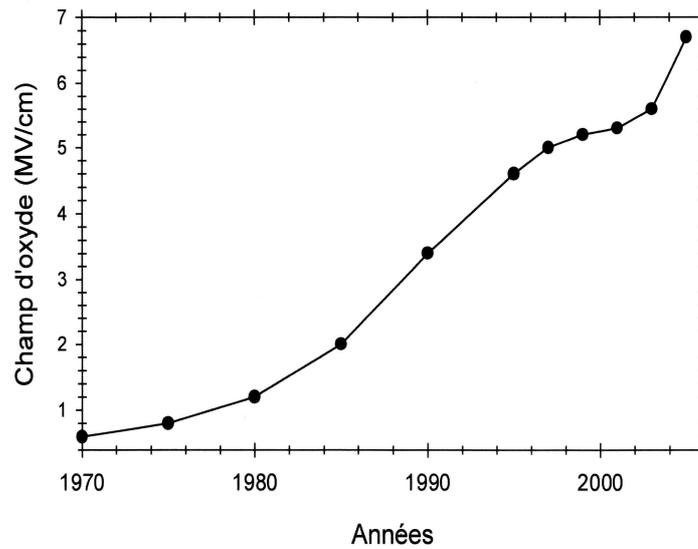


Figure 1.35 : Champ électrique dans l'oxyde de grille dans les technologies CMOS en fonction des années de fabrication [DEGRAEVE'99]

1.4.2.1.2 L'effet tunnel direct

Ce phénomène est lié à la pénétration de la fonction d'onde des porteurs dans la couche diélectrique contre laquelle ils sont confinés (onde évanescente). Lorsque l'épaisseur de cette couche est suffisamment fine, la probabilité par unité de temps pour qu'une fraction des porteurs la traverse par effet tunnel n'est pas négligeable, ce qui se traduit par un flux de charges non nul (figure 1.36).

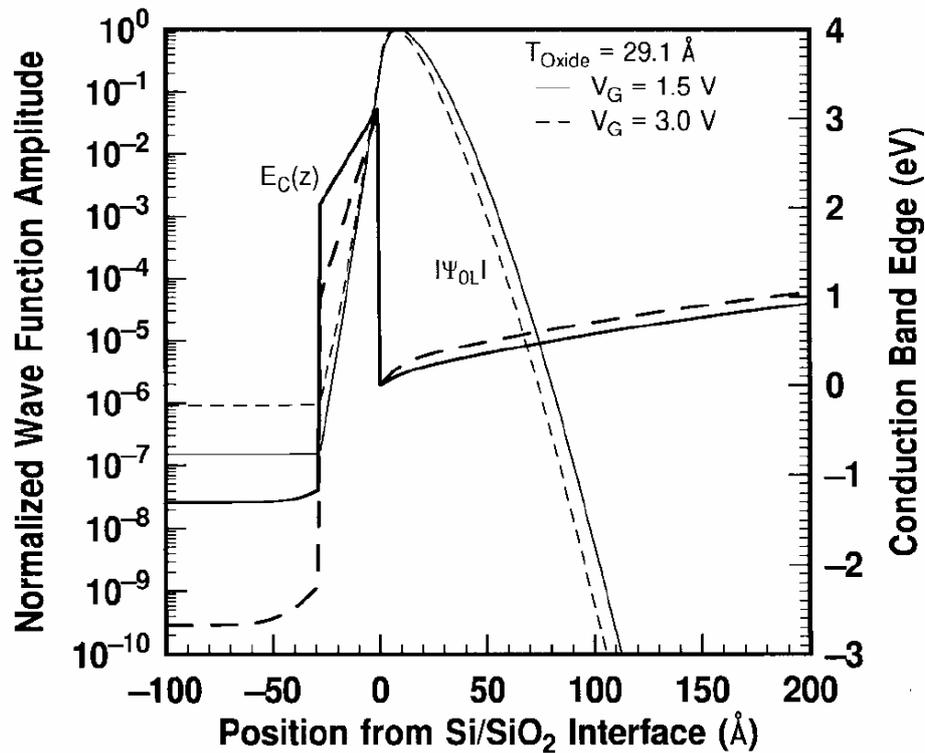


Figure 1.36 : Pénétration de la fonction d'onde dans l'oxyde pour deux polarisations de grille en forte inversion [LO'97]

Dans une structure métal - oxyde - semi-conducteur, trois mécanismes tunnel peuvent être distingués (voir figure 37):

- le passage par effet tunnel d'électrons libres de la bande de conduction (ECB)
- le passage par effet tunnel d'électrons liés de la bande de valence (EVB)
- le passage par effet tunnel de trous libres de la bande de valence (HVB)

Le mécanisme de passage par effet tunnel ECB est dominant pour un transistor nMOS en régime de forte inversion, alors que le mécanisme HVB est le plus important pour les pMOS.

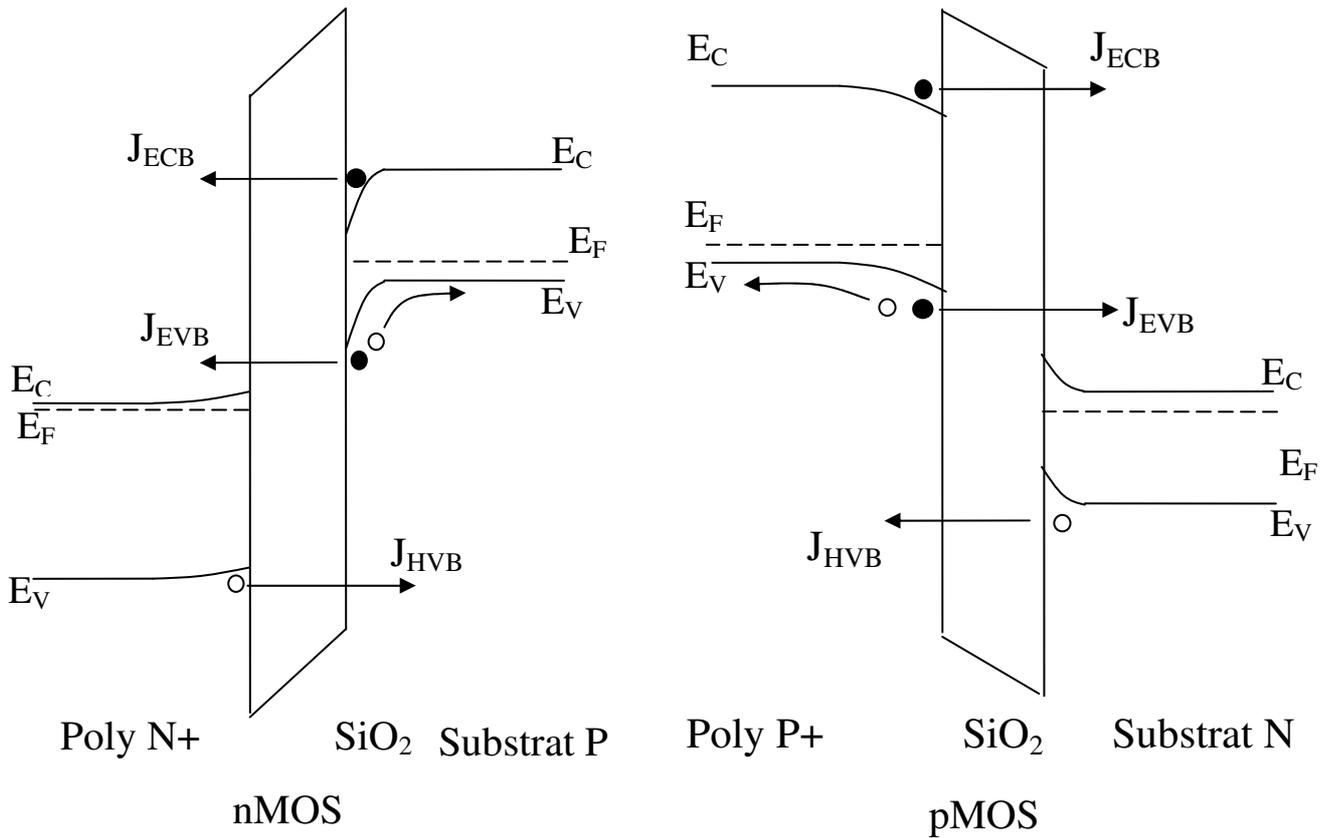


Figure 1.37 : Diagramme des bandes d'énergie d'une structure MOS polarisée en forte inversion et différents mécanismes d'effet tunnel direct.

Le passage par effet tunnel EVB ne peut intervenir qu'en cas de forte courbure des bandes d'énergie, de sorte que le bas de la bande de valence dans le substrat corresponde à la bande de conduction du polysilicium, ou vice-versa. La chute de potentiel dans l'oxyde V_{ox} doit donc être supérieure à largeur de la bande interdite ($E_g/q = 1.15$ V). Puisque les tensions d'alimentations sont de l'ordre du volt, ce type de conduction n'intervient que très peu dans les régimes de fonctionnement normaux des transistors nanométriques. Cependant, on peut être amené lors de la caractérisation de ces composants à polariser au delà des tensions nominales. Les transports de type EVB ne sont alors plus négligeables, notamment pour le courant substrat.

Plus précisément, les phénomènes de transport tunnel sont décrits sur la figure 1.37, pour des nMOS et pMOS polarisés en régime de forte inversion. La valeur $|V_{GB}|$ est suffisamment grande pour avoir un transport de type EVB. Le passage d'un électron lié de la bande de valence va donc générer un trou dans le substrat pour le nMOS ou dans le

polysilicium pour le pMOS. Le courant substrat associé au transport EVB est donc un courant de trous pour les nMOS et un courant d'électron pour les pMOS.

1.4.2.1.3 L'effet tunnel indirect

On rassemble sous ce terme l'ensemble des mécanismes faisant intervenir un défaut du réseau cristallin de l'oxyde (silice) se comportant comme un piège. Ce type de mécanisme est en général assisté thermiquement. Compte tenu de la variété des défauts ponctuels de la silice et de leur activité électrique en tant que piège, il faut noter que le nombre de mécanismes de piégeages et de réémissions rassemblés dans cette catégorie est très important.

Le courant tunnel assisté par pièges d'oxyde apparaît en régime d'inversion ou d'accumulation. Les électrons passent de la bande de conduction de la grille vers la bande de conduction du substrat via des pièges situés dans l'oxyde chargés positivement ou neutre. Le passage peut être élastique ou inélastique. Il semble plus ou moins admis que le passage par effet tunnel inélastique est le plus probable dans les oxydes minces [TAKAGI'99].

1.4.3 Impact des courants de fuite sur le fonctionnement du transistor MOS

La figure 1.38 décrit les différents courants entre les électrodes de grille, de source et de drain. Les courants de fuites liés au passage des électrons de la bande de valence sont considérés comme négligeables pour les tensions nominales de fonctionnement. On distingue le courant de fuite grille – canal ($I_{GC} = I_{GCS} + I_{GCD}$), les courants de fuites liés aux zones de recouvrements (I_{GS_0} , I_{GD_0}), et le courant de fuite grille – substrat (I_{GB}).

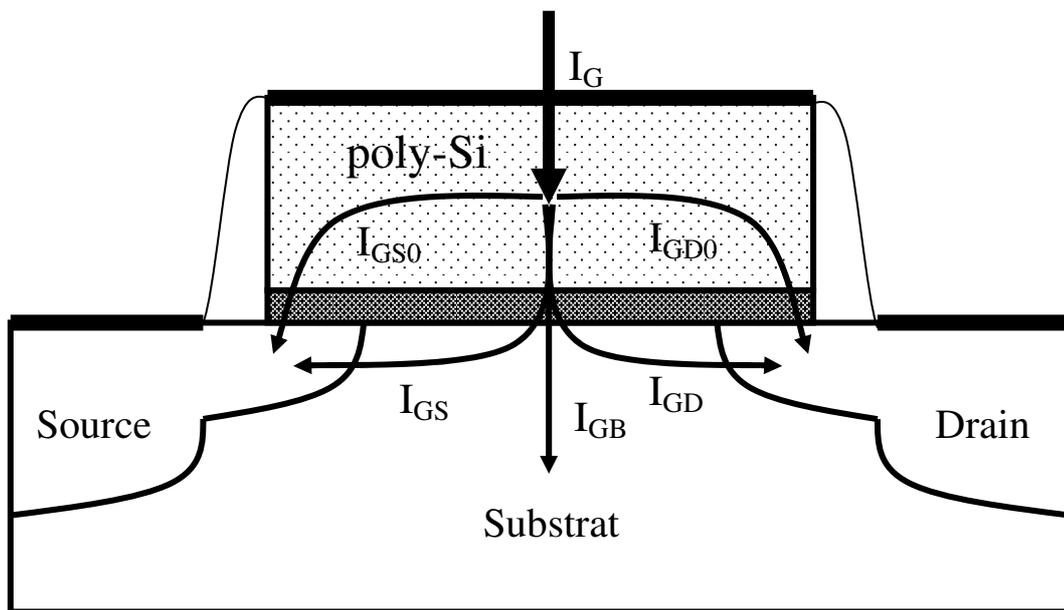


Figure 1.38 : Illustration des différentes composantes du courant tunnel de grille.

1.4.3.1 Transistor nMOS

La figure 1.39 présente les courants pour un transistor nMOS ($T_{ox} = 1.4$ nm, $W / L = 10 \mu\text{m} / 10 \mu\text{m}$), avec les électrodes de source, de drain et de substrat reliées à la masse, la tension de grille variant de -3 Volts à +3 Volts. La tension de bandes plates V_{FB} est égale à -1 Volt, la tension de seuil est de 400 mV.

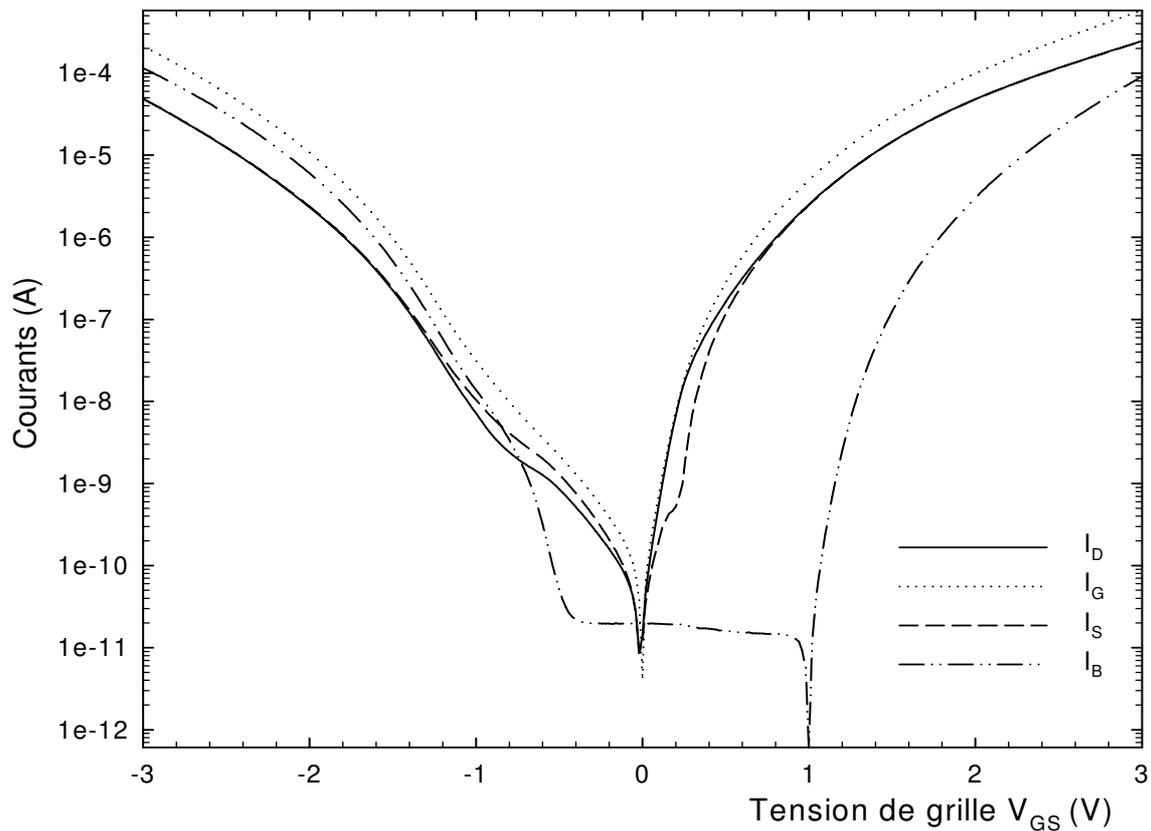


Figure 1.39 : Caractéristique I-V d'un transistor nMOS ($W/L = 10/10 \mu\text{m}$; $T_{ox} = 1.4$ nm)

1.4.3.1.1 Régime de forte inversion $V_{GS} > V_T$

La figure 1.40 présente les différents courants dans la structure pour ce régime ainsi que les diagrammes de bandes associés aux zones de recouvrement et du canal.

En régime de forte inversion ($V_{GS} > V_T$) le courant de grille est dominé par le passage par effet tunnel des électrons de la couche d'inversion (ECB). Les porteurs libres sont fournis par la source et le drain, le courant grille est égal à la somme du courant de source et de drain.

Les régions de recouvrement sont en accumulation et ne contribuent que très faiblement au courant de grille total.

Pour $V_{GS} > 1$ Volt, le courant substrat augmente, dû au passage d'électrons de la bande de valence (EVB). Comme le montre la figure 1.39, un courant supplémentaire d'électrons vient s'ajouter au courant de grille, et le courant substrat est un courant de trous.

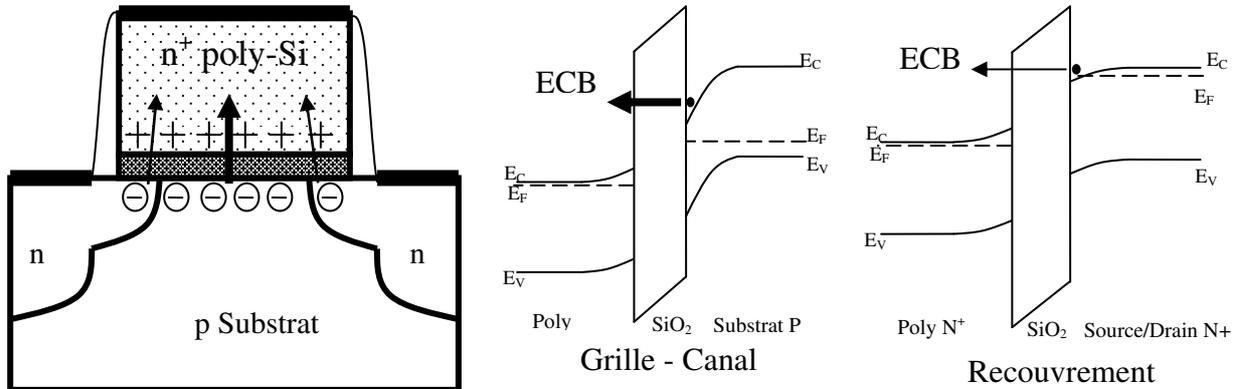


Figure 1.40 : Courants de fuites d'un transistor nMOS en régime de forte inversion ($V_{GS} > V_T$)

1.4.3.1.2 Régime de faible inversion et de déplétion ($V_{FB} < V_{GS} < V_T$)

La figure 1.41 présente les différents courants dans la structure pour ce régime ainsi que les diagrammes de bandes associés aux zones de recouvrement et du canal.

Le courant de grille est dominé par les courants de fuites des zones de recouvrement. Le courant I_{GC} est négligeable car le substrat de type p est en déplétion. Quelques électrons peuvent traverser par effet tunnel assisté par pièges (effet tunnel indirect). Les électrons accumulés dans le polysilicium au voisinage des zones de recouvrement contribuent au courant de grille total.

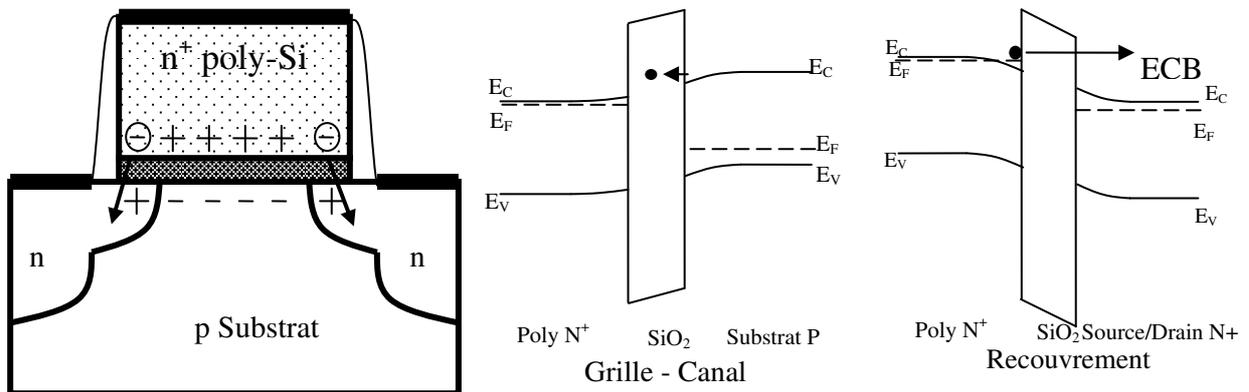


Figure 1.41 : Courants de fuites d'un transistor nMOS en régime de faible inversion et déplétion ($V_{FB} < V_{GS} < V_T$)

1.4.3.1.3 Régime d'accumulation ($V_{GS} < V_{FB}$)

La figure 1.42 présente les différents courants dans la structure pour ce régime ainsi que les diagrammes de bande associés aux zones de recouvrement et du canal.

Une couche d'accumulation de trous est formée dans le substrat de type p et une accumulation d'électrons apparaît dans le polysilicium. La barrière d'énergie pour les électrons ($\chi_{B_n} = 3.1$ eV) étant bien plus faible que pour les trous ($\chi_{B_p} = 4.5$ eV), le courant de grille est majoritairement constitué d'électrons traversant par effet tunnel de la grille vers le substrat.

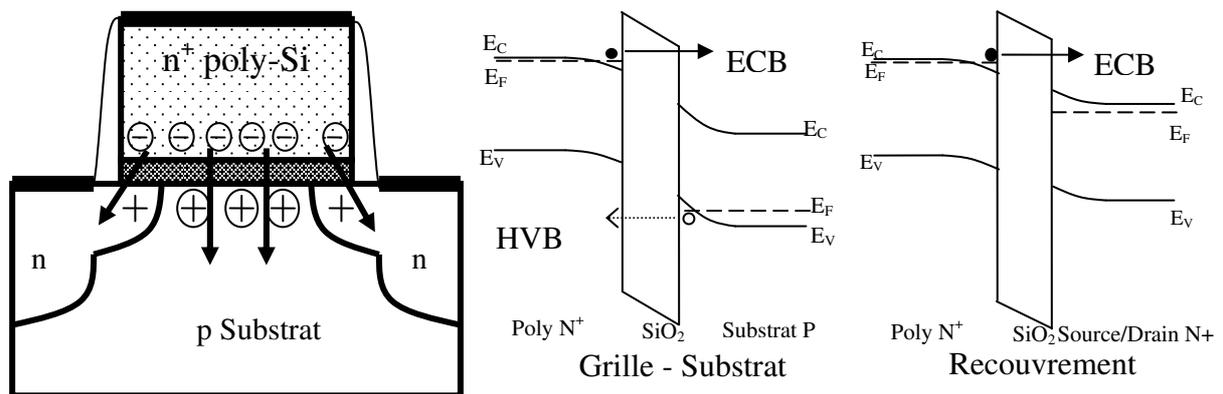


Figure 1.42 : Courants de fuites d'un transistor nMOS en régime d'accumulation ($V_{GS} < V_{FB}$)

1.4.3.2 Transistor pMOS

La figure 1.43 présente les courants pour un transistor pMOS ($T_{ox} = 1.4$ nm, $W/L = 10 \mu\text{m}/10 \mu\text{m}$), avec les électrodes de source, de drain et de substrat reliées à la masse, la tension de grille variant de -3 Volts à +3 Volts. La tension de bandes plates V_{FB} est égale à 1 Volt, la tension de seuil est de -400 mV.

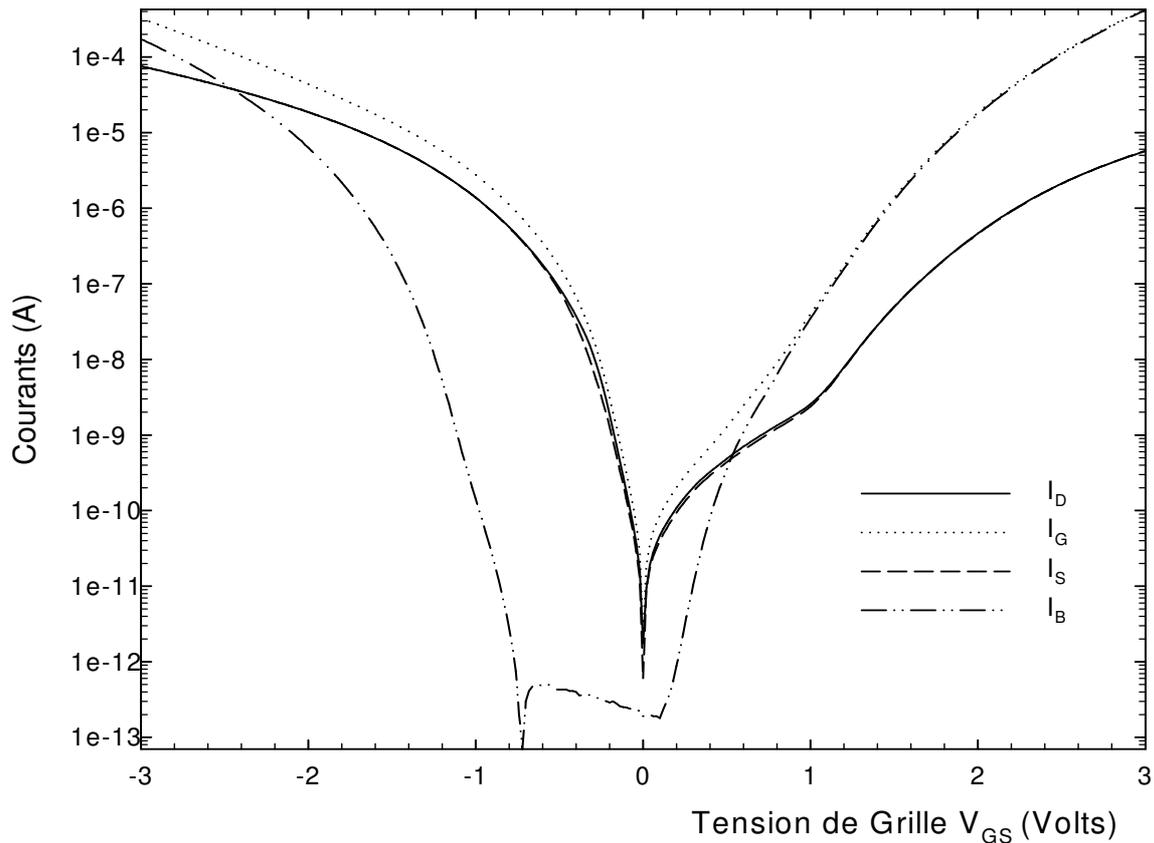


Figure 1.43 : Caractéristique I-V d'un transistor pMOS ($W/L = 10/10 \mu\text{m}$; $T_{ox} = 1.4$ nm)

1.4.3.2.1 Régime de forte inversion $V_{GS} < V_T$

La figure 1.44 présente les différents courants dans la structure pour ce régime ainsi que les diagrammes de bandes associés aux zones de recouvrement et du canal.

En régime d'inversion ($V_{GS} < V_T$) le courant de grille est dominé par le passage par effet tunnel des trous de la couche d'inversion (HVB). Les porteurs libres sont fournis par la source et le drain, le courant de grille est égal à la somme du courant de source et de drain.

Les régions de recouvrement sont en accumulation et ne contribuent que très faiblement au courant de grille total.

Pour $V_{GS} < -1$ Volt, le courant substrat augmente, dû au passage d'électrons de la bande de valence (EVB). Comme le montre la figure 1.43, un courant supplémentaire d'électrons vient s'ajouter au courant de grille, et le courant substrat est un courant de trous.

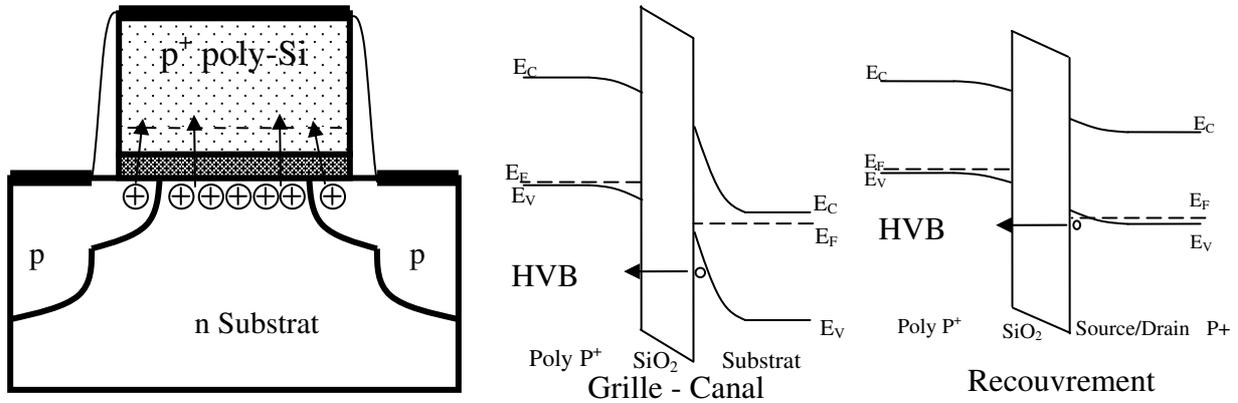


Figure 1.44 : Courants de fuites d'un transistor pMOS en régime de forte inversion ($V_{GS} < V_T$)

1.4.3.2 Régime de faible inversion et de déplétion ($V_T < V_{GS} < V_{FB}$)

La figure 1.45 présente les différents courants dans la structure pour ce régime ainsi que les diagrammes de bandes associés aux zones de recouvrement et du canal.

Le courant de grille est dominé par les courants de fuites des zones de recouvrement. Le courant I_{GC} est négligeable car le substrat de type N est en déplétion ou faible inversion. Les trous accumulés dans le polysilicium des zones de recouvrement contribuent au courant de grille total.

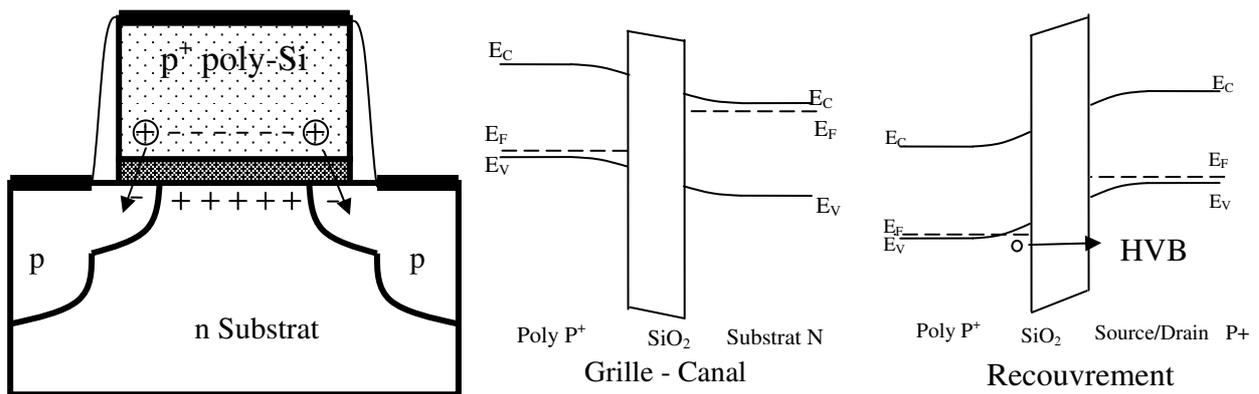


Figure 1.45 : Courants de fuites d'un transistor pMOS en régime de faible inversion et déplétion ($V_T < V_{GS} < V_{FB}$)

1.4.3.2.3 Régime d'accumulation ($V_{GS} > V_{FB}$)

La figure 1.46 présente les différents courants dans la structure pour ce régime ainsi que les diagrammes de bandes associés aux zones de recouvrement et du canal. Une couche d'accumulation de trous est formée dans le substrat de type n et une accumulation d'électrons est présente dans le polysilicium. La barrière d'énergie pour les électrons ($\chi_{B_n} = 3.1$ eV) étant bien plus faible que pour les trous ($\chi_{B_p} = 4.5$ eV), le courant de grille est majoritairement constitué d'électrons traversant par effet tunnel de la grille vers le substrat.

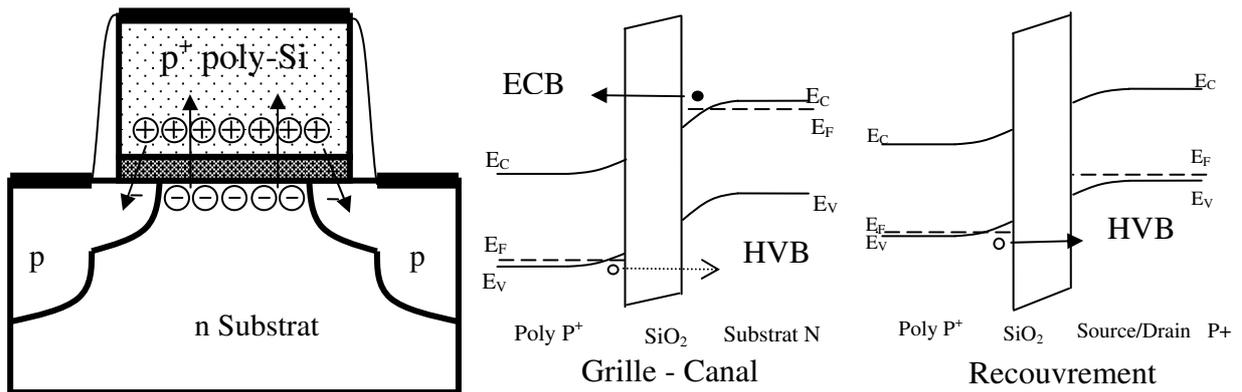


Figure 1.46 : Courants de fuites d'un transistor pMOS en régime d'accumulation ($V_{GS} < V_{FB}$)

1.4.3.3 Synthèse des composantes tunnels

Pour des tensions d'utilisation normale (typiquement 1 Volt de tension de grille pour une épaisseur de diélectrique de 1.4 nm), le type de porteur contribuant aux courants de fuites par effet tunnel à travers la grille est résumé dans le tableau 1.

Transistor	Transistor intrinsèque		Zone de recouvrement
	Accumulation	Inversion	
nMOS	Electrons	Electrons	Electrons
pMOS	Electrons	Trous	Trous
	I_{GB}	$I_{GS}+I_{GD}$	I_{GS_o}, I_{GD_o}

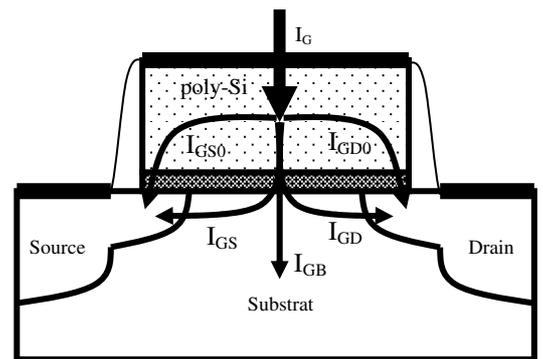


Tableau 1 : Type de porteurs participant aux courants de fuites par effet tunnel au travers de l'oxyde de grille

1.5 Conclusion

Nous avons présenté dans ce chapitre l'architecture des transistors nanométriques. Tous les points clés de l'optimisation du transistor MOS nanométrique ont été abordés, ainsi que les techniques de fabrication associées. Les limites de cette technologie CMOS conventionnelle ont été détaillées et les potentialités de la technologie SOI ont été abordées. Une large part a été faite aux nouveaux phénomènes physiques à prendre en compte dans ces technologies MOS nanométriques. D'une part l'impact des effets quantiques a été clairement identifié, et d'autre part, les mécanismes de fuites par effet tunnel au travers de la grille ultra mince ont été détaillés.

CHAPITRE 2

MODÉLISATION DES EFFETS

QUANTIQUES ET DU COURANT TUNNEL

DANS LES CAPACITÉS MOS

2 Modélisation des effets quantiques et du courant tunnel dans les capacités MOS

2.1 Introduction

La miniaturisation poussée des structures MOS fait apparaître plusieurs phénomènes limitant les performances des transistors MOS avancés. La diminution du courant tunnel de grille reste un challenge primordial pour rendre utilisables ces nouvelles technologies. Il est donc très important de comprendre les mécanismes physiques mis en jeu dans ces courants, d'où l'importance croissante de la modélisation statique et dynamique du fonctionnement des structures MOS (capacités et transistors).

Compte tenu des dimensions nanométriques des couches utilisées, les effets physiques liés au phénomène de polydéplétion de grille, des effets de confinement quantique de la couche d'inversion, et de courant tunnel au travers du diélectrique doivent être pris en compte.

Dans la littérature, plusieurs auteurs ont proposé des modèles pour prendre en compte ces phénomènes. Dans ce travail, nous avons abouti à la conclusion que pour une modélisation efficace, il est nécessaire d'avoir une formulation la plus fine possible du potentiel de surface, qui servira de point de départ pour le calcul des courants de grille et de drain dans les transistors MOS à oxyde ultrafins.

Ainsi dans ce chapitre, nous présentons l'état de l'art des méthodes de calcul et de prises en compte des différents phénomènes. Nous montrons les limites de ces méthodes qui utilisent souvent des paramètres de lissage, non physiques, pour modéliser le fonctionnement des structures MOS, et nous proposons de nouvelles méthodes explicites pour prendre en compte ces phénomènes.

Dans la première partie de ce chapitre, nous rappelons de façon critique la formulation implicite et explicite liant le potentiel de surface, la tension de grille et le quasi niveau de fermi, en tenant compte du phénomène de polydéplétion.

Nous abordons ensuite les effets liés au confinement quantiques des porteurs de la couche d'inversion, dans le but d'établir une équation implicite quantique. Nous proposons dans cette partie une nouvelle approche explicite quantique que nous comparons à celles déjà publiées.

Enfin, ce chapitre se termine par la mise en œuvre d'un modèle de courant de grille entièrement basé sur notre modèle explicite du potentiel de surface.

2.2 Les régimes de fonctionnement

Dans ce paragraphe, nous allons rappeler les principes de fonctionnement d'une structure MOS à une dimension, qui est la base du transistor MOS. La figure 2.1 présente l'empilement d'une couche de métal, d'isolant, et de semi-conducteur de type p. C'est donc la structure de base d'un transistor nMOS qui est présenté ici, et un changement de signe approprié permet aisément d'obtenir le même raisonnement pour un transistor pMOS. En appliquant une tension V_{GB} entre la grille et le substrat, la structure de bande au voisinage de l'interface Si/SiO₂ est modifiée. Trois régimes de fonctionnement sont à distinguer :

- le régime d'accumulation
- le régime de déplétion et de faible inversion
- le régime de forte inversion

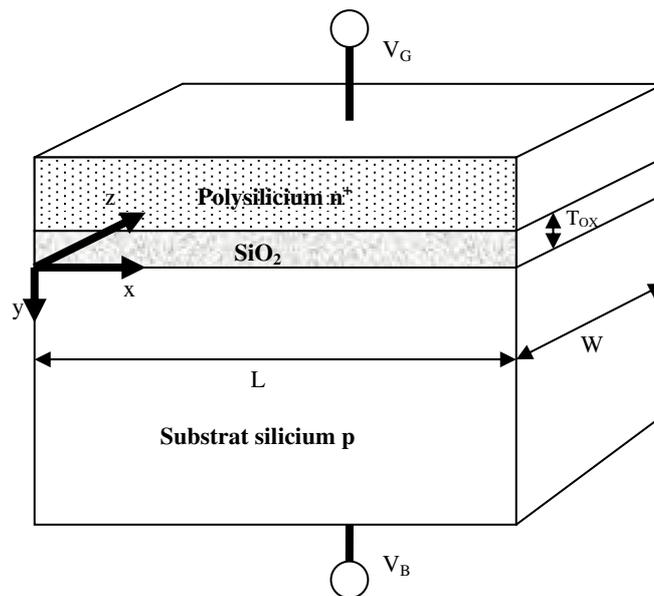


Figure 2.1 : Structure de base d'une capacité nMOS avec un substrat de type p, et une grille en polysilicium de type n+.

2.2.1 Le régime d'accumulation ($V_{GB} < V_{FB}$)

Pour des tensions inférieures à la tension de bandes plates V_{FB} , qui est la tension pour laquelle les niveaux de Fermi de la grille E_{Fm} et du substrat E_F s'alignent, les trous situés dans le substrat p sont attirés vers la surface (interface Si/SiO₂). La concentration des porteurs majoritaires près de la surface devient supérieure à celle du substrat. Une fine couche de charges positives est formée, c'est une couche dite *d'accumulation*. Un schéma représentatif et le diagramme de bandes d'énergies associé sont représentés dans sur la figure 2.2.

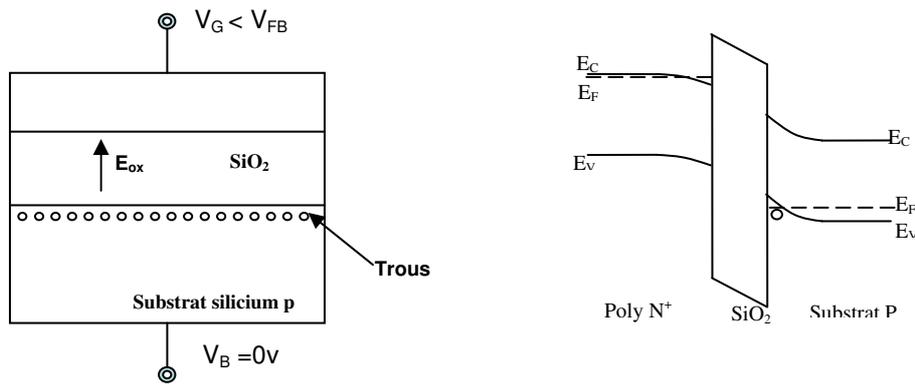


Figure 2.2 : Vue en coupe et diagramme de bandes d'énergies d'une capacité nMOS en régime d'accumulation

2.2.2 Le régime de déplétion et faible inversion ($V_{FB} < V_{GB} < V_T$)

Dans ce régime de fonctionnement, le champ électrique E_{ox} se retrouve dirigé vers le substrat. La courbure des bandes s'inverse et s'accroît vers le bas. La charge positive de la grille repousse les trous majoritaires dans le substrat et laisse une charge négative fixe due aux accepteurs ionisés. Ainsi une zone de déplétion est créée près de la surface. Dans ces conditions de polarisation, la région au voisinage de la surface est vidée de ses porteurs libres.

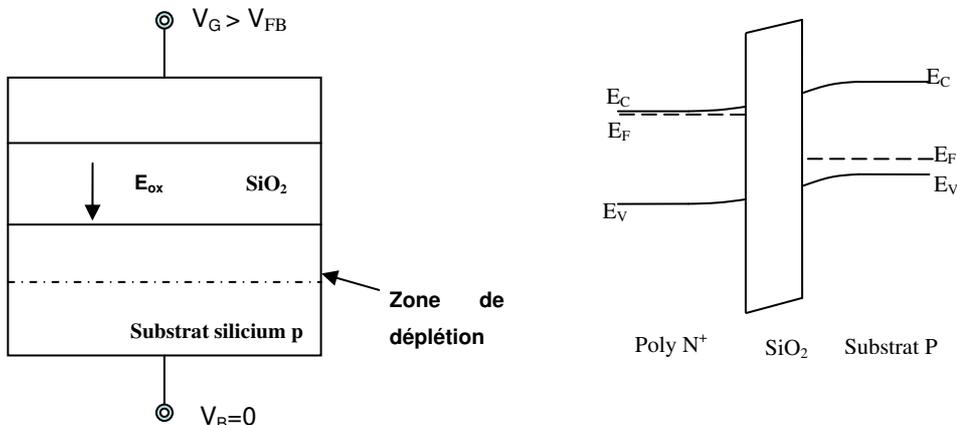


Figure 2.3 : Vue en coupe et diagramme de bandes d'énergie d'une capacité nMOS en régime de déplétion

Lorsque la courbure de bande est égale à ϕ_F (potentiel de Fermi), la structure est polarisée dans le régime de faible inversion. La densité d'électrons au voisinage de la surface augmente mais reste très inférieure à celle des trous majoritaires. La faible inversion est essentiellement conditionnée par le phénomène de déplétion.

2.2.3 Le régime de forte inversion ($V_{GB} > V_T$)

Lorsque la tension de grille augmente et atteint une polarisation supérieure à la tension de seuil V_T , pour laquelle le potentiel de surface Ψ_s est égale à $2\phi_F$, la structure fonctionne dans le régime dit de forte inversion. Le niveau de Fermi intrinsèque E_i devient inférieur au niveau de Fermi E_f et la densité d'électrons y est supérieure à celle des trous majoritaires, ce qui signifie que le semiconducteur est de type n dans cette région. Cette région est appelée *couche d'inversion*. La création de cette surface d'inversion conductrice à l'aide d'une tension de grille est le phénomène essentiel pour la conduction du courant dans un transistor MOS. Une fois la surface inversée, tout accroissement de la tension de grille conduit à une augmentation de la concentration d'électrons libres.

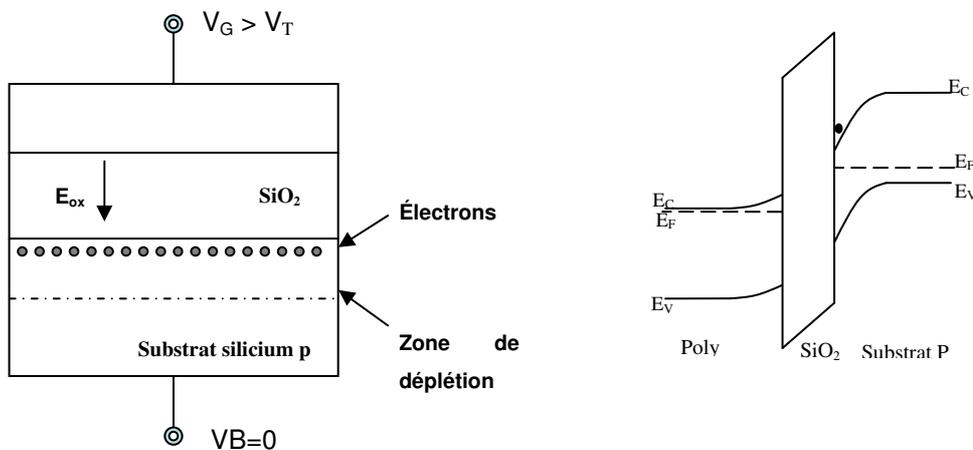


Figure 2.4 : Vue en coupe et diagramme de bandes d'énergies d'une capacité nMOS en régime d'inversion

Lorsque la forte inversion est atteinte ($\Psi_s = 2\phi_F$), la densité de charges mobiles dans la couche d'inversion est supérieure à la densité de charges fixes dans la couche de déplétion.

2.3 Modélisation de la capacité MOS basée sur le potentiel de surface

La figure 2.5 introduit la grandeur de base utilisée dans tout ce manuscrit, à savoir le potentiel de surface Ψ_s (Volts). Il traduit la courbure de bande totale entre la région neutre du substrat et l'interface Substrat/Oxyde. Cette figure fait également apparaître la chute de potentiel dans le polysilicium de grille Ψ_p , le potentiel du niveau de Fermi ϕ_F , V le potentiel du quasi niveau de Fermi des électrons et E_{Fn} le quasi niveau de Fermi des électrons.

Dans un premier temps, nous allons calculer la relation liant le potentiel de surface Ψ_s , la tension grille substrat V_{GB} , et le potentiel du quasi niveau de Fermi V en utilisant les hypothèses suivantes :

- L'oxyde est un isolant parfait, aucun courant tunnel ne traverse la structure. De ce fait, le quasi niveau de Fermi des électrons et le quasi niveau de Fermi des trous sont égaux au niveau de Fermi¹. On parle de quasi-équilibre thermodynamique.
- La grille est considérée comme idéale et il n'y a aucun phénomène de polydépétion dans le polysilicium. On considère donc $\Psi_p = 0$.
- La structure est à une dimension.

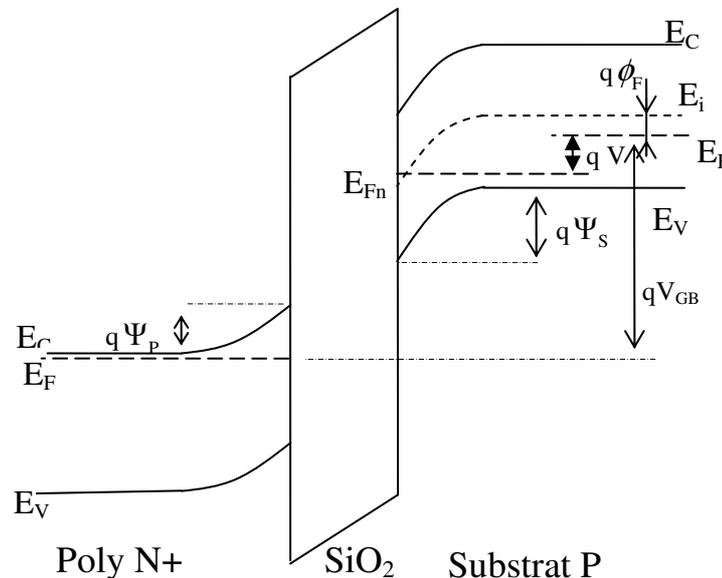


Figure 2.5 : Définition des grandeurs physiques sur une structure nMOS.

¹ En partant de la formulation du courant de dérive-diffusion $\vec{J}_n = n\mu_n\vec{\nabla}E_{Fn}$ et $\vec{J}_p = p\mu_p\vec{\nabla}E_{Fp}$, et puisque il n'y pas de courant dans la structure, le gradient des quasi niveaux Fermi est nul.

Dans le cas d'un substrat de type p, la charge d'espace $\rho(y)$ s'exprime par :

$$\rho(y) = q \cdot [p(y) - n(y) - N_A] \quad (2.1)$$

où N_A est la concentration nette des dopants accepteurs. Les densités d'électrons $n(y)$ et de trous $p(y)$ sont données, pour des semiconducteurs non dégénérés par la statistique de Maxwell-Boltzmann :

$$n(y) \approx N_A \cdot \exp\left[\frac{\Psi(y) - V - 2\phi_F}{\phi_T}\right] \quad (2.2)$$

$$p(y) = N_A \cdot \exp\left(-\frac{\Psi(y)}{\phi_T}\right) \quad (2.3)$$

où $\Psi(y)$ est le potentiel électrostatique au point y par rapport à la zone neutre du substrat et ϕ_T la tension thermodynamique définie par kT/q . Le potentiel de Fermi ϕ_F est défini comme $\phi_T \ln(N_A/n_i)$, avec n_i la concentration intrinsèque des porteurs.

L'équation de Poisson dans le substrat s'écrit :

$$\nabla^2 \Psi = -\frac{\rho(y)}{\epsilon_{Si}} \quad (2.4)$$

Puisque la structure étudiée se réduit à une dimension, l'équation de Poisson s'écrit donc sous la forme :

$$\frac{\partial^2 \Psi(y)}{\partial y^2} = \frac{qN_A}{\epsilon_{Si}} \cdot \left\{ -\exp\left(-\frac{\Psi(y)}{\phi_T}\right) + \exp\left(\frac{\Psi(y) - V - \phi_B}{\phi_T}\right) + 1 \right\} \quad (2.5)$$

avec $\phi_B = 2\phi_F$. Dans la zone neutre du substrat, les quantités Ψ et $\partial\Psi/\partial y$ sont nulles.

De plus, en utilisant la relation $\partial^2 \Psi / \partial y^2 = \frac{1}{2} \cdot \partial(\partial\Psi/\partial y)^2 / \partial\Psi$, on peut déterminer la charge totale par unité de surface Q_{sc} dans le semiconducteur en appliquant la loi de Gauss :

$$Q_{sc} = \epsilon_{Si} \cdot \left. \frac{\partial \Psi}{\partial y} \right|_{y=0} = \pm \gamma C_{ox} \left\{ \Psi_s + \phi_T \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right] + \phi_T \cdot \exp\left(\frac{-V - \phi_B}{\phi_T}\right) \left[\exp\left(\frac{\Psi_s}{\phi_T}\right) - 1 \right] \right\}^{1/2} \quad (2.6)$$

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 $\sqrt{2q\epsilon_{Si}N_A}/C_{ox}$.

La charge Q_{sc} est négative pour $V_{GB} > V_{FB}$ (régime d'inversion), et positive quand $V_{GB} < V_{FB}$ (régime d'accumulation).

En appliquant le théorème de Gauss à l'interface Oxyde/Semiconducteur, la charge Q_{sc} peut être reliée à la tension de grille appliquée à la grille :

$$Q_{sc} = -C_{ox} (V_{GB} - V_{FB} - \Psi_s) \quad (2.7)$$

En combinant les équations (2.6) et (2.7), on obtient une relation implicite liant le potentiel de surface Ψ_s , la tension grille substrat V_{GB} , et le potentiel du quasi niveau de Fermi V :

$$\left(\frac{V_{GB} - V_{FB} - \Psi_s}{\gamma} \right)^2 = \Psi_s + \phi_T \cdot \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right] + \phi_T \cdot \exp\left(\frac{-V - \phi_B}{\phi_T}\right) \left[\exp\left(\frac{\Psi_s}{\phi_T}\right) - 1 \right] \quad (2.8)$$

Cette relation très importante constitue l'élément essentiel de tout modèle basé sur le potentiel de surface. Elle permet évidemment de déterminer les différentes charges et capacités d'une structure MOS à une dimension dans tous les régimes de fonctionnement, mais elle peut également être utilisée pour la modélisation d'un transistor MOS.

Cette équation comporte de nombreux avantages, puisqu'elle est valable dans tous les régimes de fonctionnement. Cette équation est dite équation implicite, car elle ne peut être résolue que par des techniques numériques nécessitant des itérations. Plusieurs auteurs [LANGEVELDE'00, CHEN'01, PREGALDINY'04] ont développé une expression explicite destinée avant tout aux modèles compacts basés sur le potentiel de surface. Avant de développer le principe de ces modèles explicites, nous allons compléter l'analyse d'une structure à une dimension, en incluant les effets de polydépétion qui ne peuvent plus être négligés dans les technologies actuelles.

2.4 Modélisation du phénomène de dépétion dans le polysilicium de grille.

Pour augmenter la conductivité des lignes d'interconnexion, le polysilicium a remplacé l'aluminium dans le milieu des années 1970. L'utilisation d'une grille en polysilicium provoque la formation d'une couche de dépétion à l'interface polysilicium-SiO₂, lorsque le transistor est en régime d'inversion. Cette charge fixe dans le matériau de la grille vient équilibrer la charge dans le semiconducteur, assurant ainsi l'électro-neutralité.

Une chute de potentiel Ψ_p s'établit dans le polysilicium de grille au travers de la couche de déplétion, réduisant ainsi la différence de potentiel appliquée à l'oxyde. La figure 2.5 schématise cette chute de potentiel dans le polysilicium.

L'application du théorème de Gauss à l'interface Si/SiO₂ sans tenir compte du phénomène de polydéplétion permet d'obtenir la charge totale dans le substrat formulée suivant (2.7). En incluant le phénomène de polydéplétion, l'équation (2.7) devient :

$$Q_{SC} = -C_{ox} \cdot (V_{GB} - V_{FB} - \Psi_s - \Psi_p) \quad (2.9)$$

Le but de cette partie est d'établir une relation explicite entre la chute de potentiel dans le polysilicium Ψ_p et le potentiel de surface Ψ_s .

Considérons une grille en polysilicium fortement dopée n⁺ (avec un niveau de dopage compris entre 10¹⁹ et 10²⁰ cm⁻³). La densité de trous dans ce matériau est considérée comme négligeable, car la grille fonctionne uniquement dans les régimes d'accumulation ($V_{GB} < V_{FB}$) ou de déplétion ($V_{GB} > V_{FB}$), et ne sera jamais en régime d'inversion dans les conditions normales d'utilisation du transistor MOS².

La charge d'espace $\rho'(y)$ dans la grille de polysilicium est donnée par :

$$\rho'(y) = q \cdot (N_p - n'(y)) \quad (2.10)$$

Où N_p représente la concentration nette d'atomes donneurs du polysilicium et $n'(y)$ la densité d'électrons dans le polysilicium. En utilisant la statistique de Maxwell-Boltzmann, cette densité d'électron s'exprime par :

$$n'(y) \simeq N_p \cdot \exp\left(-\frac{\Psi(y)}{\phi_T}\right) \quad (2.11)$$

La grandeur Ψ est le potentiel électrostatique qui s'annule dans la zone neutre de la grille. L'équation de Poisson dans la grille peut donc s'écrire :

$$\frac{\partial^2 \Psi}{\partial y^2} = \frac{q N_p}{\epsilon_{SI}} \cdot \left[1 - \exp\left(-\frac{\Psi}{\phi_T}\right) \right] \quad (2.12)$$

² A titre indicatif, pour une capacité MOS à substrat de type p dopée à 10¹⁸cm⁻³, avec un oxyde d'épaisseur 1.2 nm et une grille en polysilicium de type n dopée à 5.10¹⁹ cm⁻³, la tension V_{GB} à appliquer pour inverser le matériau de grille est de l'ordre de 3 Volts, ce qui est bien au delà des tensions nominales d'utilisation.

Les conditions aux limites sur Ψ et $\frac{\partial\Psi}{\partial y}$ sont choisies de façon à avoir ces deux quantités nulles dans la zone neutre de la grille. Comme précédemment, en utilisant la relation $\partial^2\Psi/\partial y^2 = \frac{1}{2} \cdot \partial(\partial\Psi/\partial y)^2/\partial\Psi$, on peut déterminer la charge totale par unité de surface Q_g dans le semiconducteur en appliquant la loi de Gauss :

$$Q_g = \varepsilon_{SI} \cdot \left. \frac{\partial\Psi}{\partial y} \right|_{y=-T_{ox}} = \pm\gamma_p C_{ox} \left\{ \Psi_p + \phi_T \left[\exp\left(-\frac{\Psi_p}{\phi_T}\right) - 1 \right] \right\}^{1/2} \quad (2.13)$$

où γ_p est le facteur de corps du polysilicium, égal à $\sqrt{2q\varepsilon_{SI}N_p}/C_{ox}$. La charge par unité de surface Q_g est calculée en $y = -T_{ox}$ car l'interface SiO₂/Si correspond à $y = 0$.

En raison de la neutralité électrique de la charge de part et d'autre du diélectrique, l'opposée de la charge totale dans le substrat est égale à la charge totale dans le polysilicium ($Q_{sc} = -Q_g$). La charge totale dans le substrat est explicitement reliée au potentiel de surface Ψ_s et à la chute de potentiel dans le polysilicium Ψ_p (cf. équation (2.9)). En égalant les deux équations (2.9) et (2.13), on a donc :

$$C_{ox} (V_{GB} - V_{FB} - \Psi_s - \Psi_p) = \pm\gamma_p C_{ox} \left\{ \Psi_p + \phi_T \left[\exp\left(-\frac{\Psi_p}{\phi_T}\right) - 1 \right] \right\}^{1/2} \quad (2.14)$$

On distingue deux cas en fonction de V_{GB} :

Pour le régime d'accumulation ($V_{GB} < V_{FB}$), une couche d'accumulation de trous dans le substrat et une couche d'accumulation d'électrons dans le polysilicium sont formées. Comme Ψ_s et Ψ_p sont négatifs, le terme en $\exp\left(\frac{\Psi_s}{\phi_T}\right)$ dans (2.6) devient négligeable, et (2.14) peut s'écrire de la manière suivante :

$$\gamma C_{ox} \left\{ \Psi_s + \phi_T \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right] \right\}^{1/2} = \gamma_p C_{ox} \left\{ \Psi_p + \phi_T \left[\exp\left(-\frac{\Psi_p}{\phi_T}\right) - 1 \right] \right\}^{1/2} \quad (2.15)$$

Compte tenu des différences entre les niveaux de dopages ($N_A \ll N_p$), on a $\gamma \ll \gamma_p$. De ce fait, on obtient $|\Psi_s| \gg |\Psi_p|$. Il est donc fréquent de considérer que Ψ_p est nul en régime d'accumulation, et de ne considérer aucun phénomène de polydéplétion dans ce régime de fonctionnement.

En régime d'inversion ($V_{GB} > V_{FB}$), la charge totale dans le substrat est constituée d'une couche d'inversion d'électrons et d'une couche de déplétion d'atomes accepteurs

ionisés. Une couche de déplétion d'atomes donneurs ionisés est aussi présente dans le polysilicium (polydéplétion). Comme Ψ_S et Ψ_P sont positifs, l'équation (2.13) peut être approximée par :

$$Q_g \approx \gamma_p C_{ox} \cdot \sqrt{\Psi_P} \quad (2.16)$$

Compte tenu de la neutralité électrique, on peut donc déterminer une relation explicite entre Ψ_S et Ψ_P , en égalant (2.16) et (2.9) :

$$\Psi_P = \left[\sqrt{V_{GB} - V_{FB} - \Psi_S + \frac{\gamma_p^2}{4}} - \frac{\gamma_p}{2} \right]^2 \quad (2.17)$$

Cette formulation reste valable également dans le cas d'une grille idéale métallique ($N_p \rightarrow \infty$). Dans ce cas $\gamma_p \rightarrow \infty$ et Ψ_P est nul (aucun phénomène de polydéplétion).

L'équation implicite (2.8) peut donc rendre compte de l'effet de la déplétion dans le polysilicium en incluant l'équation (2.17) pour $V_{GB} > V_{FB}$:

$$\left(\frac{V_{GB} - V_{FB} - \Psi_S - \Psi_P}{\gamma} \right)^2 = \Psi_S + \phi_T \cdot \left[\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1 \right] + \phi_T \cdot \exp\left(\frac{-V - \phi_B}{\phi_T}\right) \left[\exp\left(\frac{\Psi_S}{\phi_T}\right) - 1 \right] \quad (2.18)$$

avec :

$$\Psi_P = \begin{cases} 0 & \text{pour } V_{GB} < V_{FB} \\ \left[\sqrt{V_{GB} - V_{FB} - \Psi_S + \frac{\gamma_p^2}{4}} - \frac{\gamma_p}{2} \right]^2 & \text{pour } V_{GB} > V_{FB} \end{cases} \quad (2.19)$$

Pour rendre compte de l'impact du phénomène de polydéplétion dans le matériau de grille, la figure 2.6 présente la simulation du potentiel de surface, de la charge totale par unité de surface et de la capacité d'une structure MOS avec un substrat de type n dopé uniformément à 10^{18} cm^{-3} , un polysilicium dopé de type n à $5 \cdot 10^{19} \text{ cm}^{-3}$, et une épaisseur d'oxyde de 1.2 nm. Nous avons également extrait la tension du dispositif³.

On peut constater un faible impact sur le potentiel de surface (figure 2.6a). Cependant, la figure 2.6b présentant l'évolution de la charge totale montre l'importance de la prise en compte de l'effet de polydéplétion. Enfin la capacité de la structure (figure 2.6c) confirme que

³ La méthode d'extraction de la tension de seuil à partir de la mesure de la capacité totale C_{gg} utilisée ici est basée sur l'étude de la fonction $P(V_{GB}) = \partial C_{gg} / \partial V_{GB}$. La tension V_{GB} correspondant au maximum de la fonction P est la tension de seuil V_{th} .

la polydéplétion ne peut être ignorée dans les techniques de caractérisation, puisque elle fait chuter la capacité maximale de près de 35 %, ainsi qu'une augmentation de 3 % de la tension de seuil (figure 2.6d).

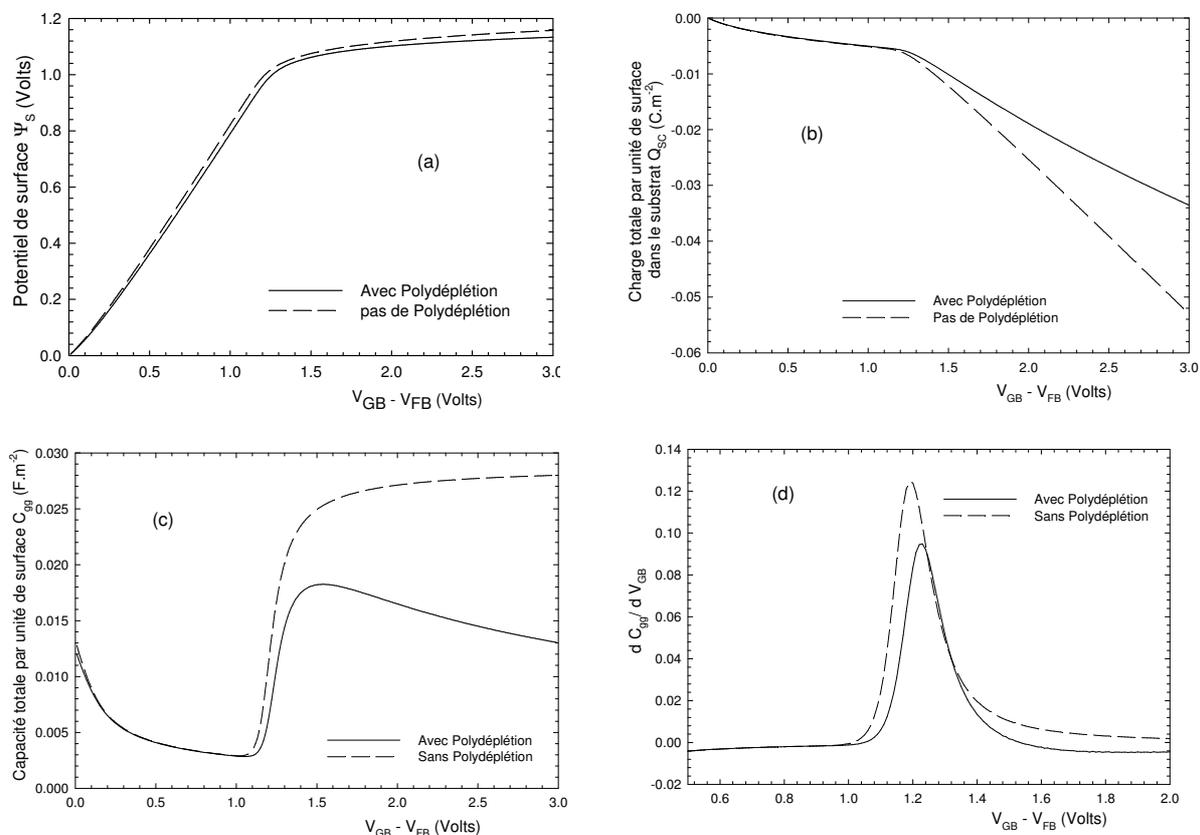


Figure 2.6 : Comparaison entre le potentiel de surface (a), entre la charge totale par unité de surface dans le substrat (b), entre la capacité totale C_{gg} (c) et : Extraction de la tension de seuil (d) calculés en prenant en compte ou non le phénomène de polydéplétion, pour une capacité nMOS, avec un dopage substrat de type p $N_A=10^{18} \text{ cm}^{-3}$, dopage de type n du polysilicium $N_P= 5.10^{19} \text{ cm}^{-3}$ et $T_{ox}=1.2 \text{ nm}$.

2.5 Modélisation explicite du potentiel de surface

Nous allons rappeler dans cette partie le traitement mathématique de l'équation (2.8) permettant d'obtenir un calcul explicite du potentiel de surface par la seule connaissance de V_{GB} et V . Dans un premier temps, nous allons établir une équation explicite valable depuis le régime de déplétion au régime d'inversion forte. Ces régimes de fonctionnement sont destinés à la modélisation compacte. Enfin, nous compléterons le modèle en détaillant la formulation valable pour le régime d'accumulation.

2.5.1 Régimes de déplétion et d'inversion

Dans ces conditions, l'équation (2.8) se réduit à :

$$\left(\frac{V_{GB} - V_{FB} - \Psi_S}{\gamma} \right)^2 = \Psi_S + \phi_T \cdot \exp\left(\frac{-V - \phi_B}{\phi_T} \right) \left[\exp\left(\frac{\Psi_S}{\phi_T} \right) - 1 \right] \quad (2.20)$$

Cette équation garde évidemment son caractère implicite. Le principe de l'approche explicite est de déterminer une première approximation valable dans le régime de déplétion, puis une deuxième dans le régime d'inversion forte. L'unification des deux approximations est ensuite assurée par une fonction de lissage de type hyperbolique.

Dans le régime de déplétion et de faible inversion, le potentiel de surface est défini comme $0 < \Psi_S < \phi_B + V$. Le terme exponentiel dans le second membre de l'équation (2.20) est négligeable, et le potentiel de surface s'exprime selon :

$$\Psi_{S_{wi}} = \left[\sqrt{V_{GB} - V_{FB} + \frac{\gamma^2}{4}} - \frac{\gamma}{2} \right]^2 \quad (2.21)$$

Dans ce régime de fonctionnement, le potentiel de surface est quasiment linéaire avec la tension de grille V_{GB} .

Dans le régime de forte inversion, le potentiel de surface est défini comme $\Psi_S > \phi_B + V$. Le terme exponentiel ne peut plus être négligé. Langevelde [LANGEVELDE'00] a proposé de remplacer dans l'équation (2.20) le terme Ψ_S par $\Phi_B + V$ sauf dans le terme exponentiel. Cette approche permet d'obtenir l'approximation suivante, valable pour le régime de forte inversion :

$$\Psi_{S_{si}} = \Phi_B + V + \phi_T \ln \left\{ \frac{1}{\phi_T} \left[\left(\frac{V_{GB} - V_{FB} - \phi_B - V}{\gamma^2} \right)^2 - \phi_B - V + \phi_T \right] \right\} \quad (2.22)$$

La figure 2.7 représente les résultats de l'approximation valable en régime de déplétion et de faible inversion (équation (2.21)), et régime de forte inversion (équation (2.22)). Le modèle d'approximation en déplétion et faible inversion donne de bons résultats, mais en régime de forte inversion, l'accord n'est pas bon, ce qui montre que le modèle doit être amélioré. L'introduction de fonctions empiriques pour améliorer la précision de l'approximation dans le régime d'inversion modérée est nécessaire. En observant la solution issue de l'équation implicite, on remarque que le potentiel de surface varie de $\phi_B + V$ jusqu'à $4\phi_T$ pour des tensions de grilles élevées. Cette observation a conduit Langevelde à remplacer ϕ_B par la fonction empirique suivante :

$$\phi_B^* = \phi_B + V + \frac{\Psi_{S_{wi}} - \phi_B - V}{\sqrt{1 + \left[\frac{\Psi_{S_{wi}} - \phi_B - V}{4\phi_T} \right]^2}} \quad (2.23)$$

En considérant que le terme quadratique est dominant, la reformulation du potentiel de surface en forte inversion utilisant cette fonction empirique donne :

$$\Psi_{S_{si}} = \Phi_B + V + \phi_T \ln \left\{ \frac{1}{\phi_T} \left[\left(\frac{V_{GB} - V_{FB} - \phi_B^* - V}{\gamma^2} \right)^2 - \phi_B - V + \phi_T \right] \right\} \quad (2.24)$$

La dernière étape est de lier les deux expressions valables en faible et forte inversion par une fonction de lissage de type hyperbolique. Prégaldiny [PREGALDINY'03] a introduit l'utilisation d'une fonction f , qui permet le passage progressif de la fonction $\Psi_{S_{wi}}$ en $\phi_B + V$:

$$f(V_{GB}, V) = \frac{\Phi_B + V + \Psi_{S_{wi}}}{2} - \frac{1}{2} \sqrt{(\Psi_{S_{wi}} - \phi_B - V)^2 + 4\delta^2} \quad (2.25)$$

avec δ une constante de lissage, fixée à 0,02.

La modèle explicite valable de la déplétion à la forte inversion s'exprime donc par [PREGALDINY'03]:

$$\Psi_S = f + \phi_T \ln \left\{ \frac{1}{\gamma^2 \phi_T} \left[V_{GB} - V_{FB} - f - \frac{\Psi_{S_{wi}} - f}{\sqrt{1 + \left[\frac{\Psi_{S_{wi}} - f}{4\phi_T} \right]^2}} \right]^2 - \frac{f}{\phi_T} + 1 \right\} \quad (2.26)$$

Les termes $\phi_B + V$ de l'équation (2.24) ont été remplacés par f . Cela n'affecte en aucun cas l'expression du potentiel de surface en forte inversion, puisque f sature à la valeur $\phi_B + V$ dans ce régime de fonctionnement. Par contre, dans le régime de déplétion et de faible

inversion, l'expression (2.26) est dominée par la fonction f et le terme logarithmique est négligeable. Les résultats obtenus sont présentés sur la figure 2.7. Les résultats intermédiaires (f et $\Psi_{S_{wi}}$) sont également présentés. On observe un très bon accord entre la solution implicite (calcul itératif) et cette expression explicite.

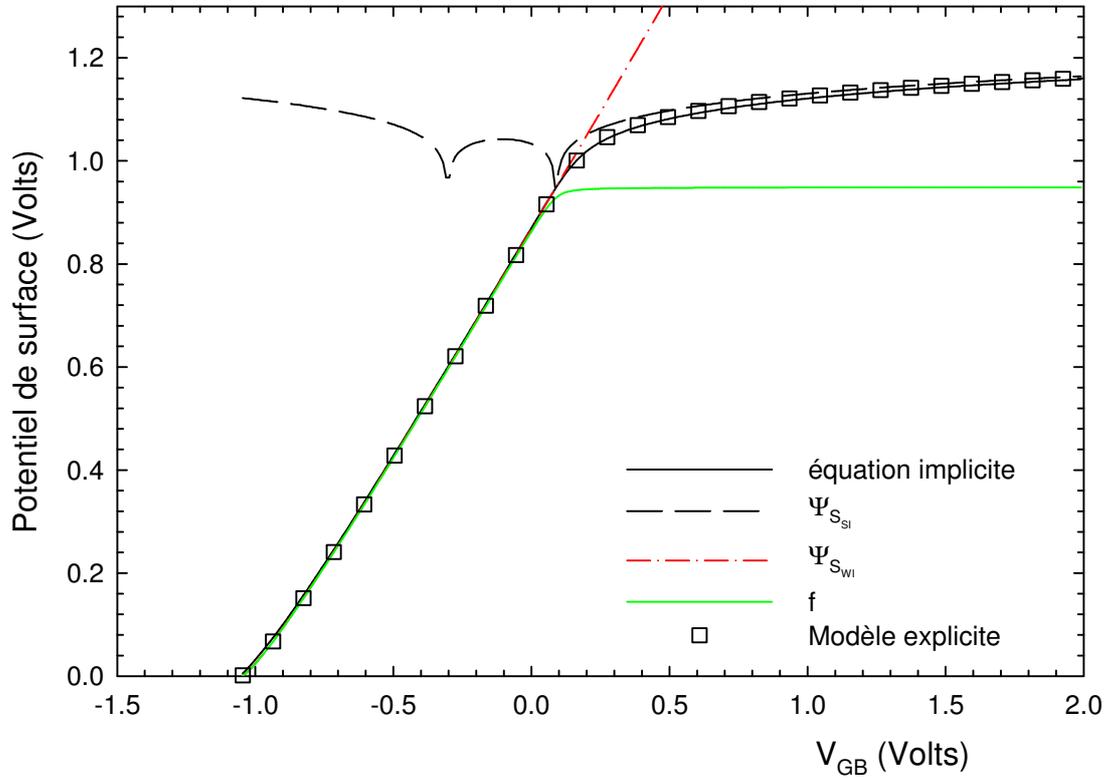


Figure 2.7 : Modélisation explicite du potentiel de surface, dans les régimes de déplétion et d'inversion, pour une structure nMOS avec $N_A=10^{18} \text{ cm}^{-3}$, $T_{ox}=1.2 \text{ nm}$, $N_p=9 \cdot 10^{19} \text{ cm}^{-3}$.

Le phénomène de polydéplétion n'est pas inclus de façon intrinsèque dans ce modèle. La prise en compte de la polydéplétion est réalisée par l'utilisation d'une tension effective de grille, définie par :

$$V_{GB_{eff}} = V_{GB} - \Psi_p \quad (2.27)$$

La chute de potentiel dans le polysilicium est explicitement reliée à Ψ_s par la relation (2.19). Cependant, en remplaçant dans l'expression Ψ_p le potentiel de surface par la fonction f , on obtient une approximation satisfaisante de la chute de potentiel dans le polysilicium. Au final, il suffit donc de calculer la tension de grille effective correspondant à une tension de grille donnée, puis de calculer le potentiel de surface avec l'équation (2.26) par rapport à cette tension de grille effective. La figure 2.8 présente le résultat du modèle complet classique explicite tenant compte de la polydéplétion de grille.

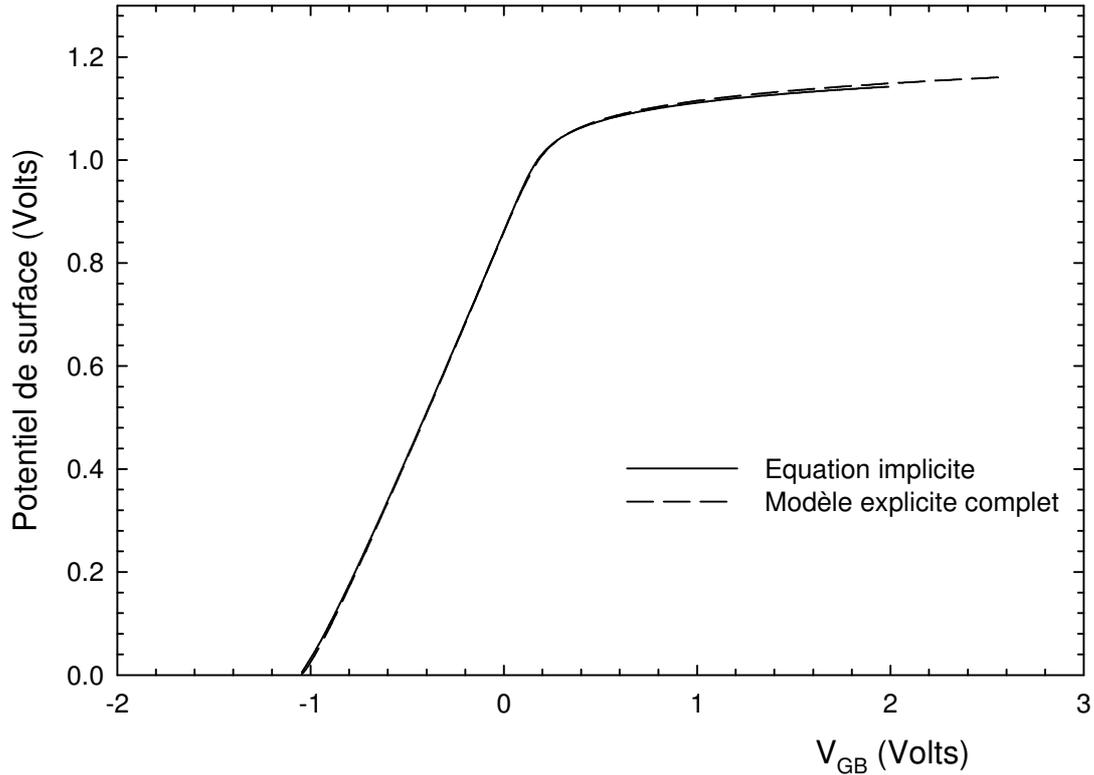


Figure 2.8 : Modélisation explicite du potentiel de surface, dans les régimes de déplétion et d'inversion, pour une structure nMOS avec $N_A=10^{18} \text{ cm}^{-3}$, $T_{ox}=1.2 \text{ nm}$, $N_p=9 \cdot 10^{19} \text{ cm}^{-3}$, avec prise en compte de la polydéplétion de grille.

2.5.2 Régime d'accumulation

Le régime d'accumulation est relativement plus simple à traiter que les régimes de déplétion et d'inversion. De plus, il n'y a pas de phénomène de polydéplétion dans le régime d'accumulation. L'influence des électrons étant négligeable, l'équation implicite (2.18) se réduit donc à la forme suivante :

$$\left(\frac{V_{GB} - V_{FB} - \Psi_S}{\gamma} \right)^2 = \Psi_S + \phi_T \left[\exp\left(-\frac{\Psi_S}{\phi_T} \right) - 1 \right] \quad (2.28)$$

Rappelons que dans ce régime de fonctionnement, V_{GB} et Ψ_S sont négatifs. Le terme exponentiel est donc dominant et le potentiel de surface s'écrit :

$$\Psi_S = -\phi_T \ln \left(\frac{(V_{GB} - V_{FB} - \Psi_S)^2 / \gamma^2 - \Psi_S + \phi_T}{\phi_T} \right) \quad (2.29)$$

De la même façon que dans le régime d'inversion, les termes Ψ_S présents dans le logarithme vont être remplacés par une fonction empirique de V_{GB} .

Le potentiel de surface varie de 0 à la tension de bandes plates, jusqu'à une valeur proche de $-4\phi_T$ à forte accumulation. L'approche proposée par Langevelde est d'utiliser la fonction [LANGEVELDE'03]:

$$\Psi_B^* = \frac{(V_{GB} - V_{FB}) / (1 + \gamma\sqrt{2\phi_T})}{\sqrt{1 + \left[\frac{(V_{GB} - V_{FB}) / (1 + \gamma\sqrt{2\phi_T})}{4\phi_T} \right]^2}} \quad (2.30)$$

Cette fonction empirique, de la même forme que la fonction (2.23), permet un passage progressif de 0 à $-4\phi_T$.

Le potentiel de surface en accumulation s'exprime donc par :

$$\Psi_S = -\phi_T \ln \left(\frac{(V_{GB} - V_{FB} - \Psi_B^*)^2 / \gamma^2 - \Psi_B^* + \phi_T}{\phi_T} \right) \quad (2.31)$$

Le modèle explicite valable dans le régime d'accumulation est présenté sur la figure 2.9. La simulation a été réalisée pour deux épaisseurs d'oxyde différentes. Cette comparaison met en évidence que cette approche donne des résultats satisfaisants pour des oxydes épais. Cependant l'écart entre la solution numérique de l'équation implicite et l'approximation explicite s'accroît avec la diminution de l'épaisseur d'oxyde. Ce phénomène est lié au fait que le potentiel de surface, dans ce cas, ne tend pas vers $-4\phi_T$. Ce paramètre de la fonction empirique Ψ_B^* a été ajusté pour des oxydes relativement épais (supérieurs à 3.5 nm). De ce fait, il serait nécessaire de modifier cette valeur $-4\phi_T$ en fonction de l'épaisseur de l'oxyde, ce qui va à l'encontre d'une approche physique.

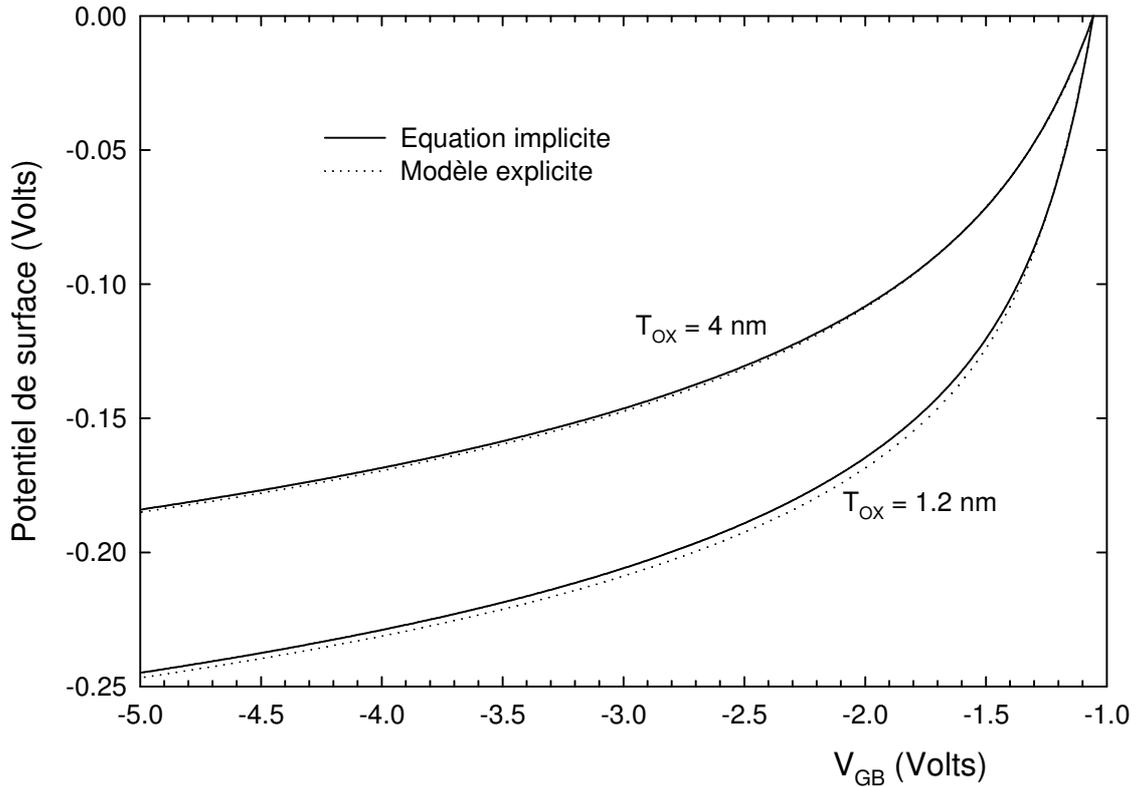


Figure 2.9 : Potentiel de surface en régime d'accumulation pour deux épaisseurs d'oxyde différentes.

2.5.3 Modèle explicite complet

On dispose de deux équations explicites, l'une valable en accumulation et l'autre en régime de déplétion et d'inversion. La figure 2.10 montre le résultat de l'ensemble de ces deux équations, laissant apparaître un problème de « liaison » au niveau de la tension de bandes plates. En effet, l'équation (2.26) a été établie en négligeant la contribution des trous majoritaires. L'approximation explicite (2.26) est donc peu précise pour des tensions de grilles proches de la tension de bandes plates, et la continuité avec l'équation explicite valable en régime d'accumulation n'est pas assurée. Ce problème a été corrigé par Langevelde en reformulant l'équation (2.21) de la façon suivante [LANGEVELDE'03]:

$$\Psi_{S_{wi}} = \left(\sqrt{V_{GB} - V_{FB} + \Delta_{acc} + \frac{\gamma^2}{4}} - \frac{\gamma}{2} \right)^2 - \Delta_{acc} \quad (2.32)$$

$$\text{où } \Delta_{acc} = \phi_T \left[\exp \left(- \frac{1}{1 + \frac{\gamma}{\sqrt{2\phi_T}}} \frac{V_{GB} - V_{FB}}{\phi_T} \right) - 1 \right]. \text{ Ce terme permet de prendre en}$$

compte de façon explicite l'influence des trous en régime de déplétion pour des tensions de

grille légèrement supérieures à la tension de bandes plates. La continuité entre l'accumulation et la déplétion est ainsi assurée avec l'utilisation du terme Δ_{acc} .

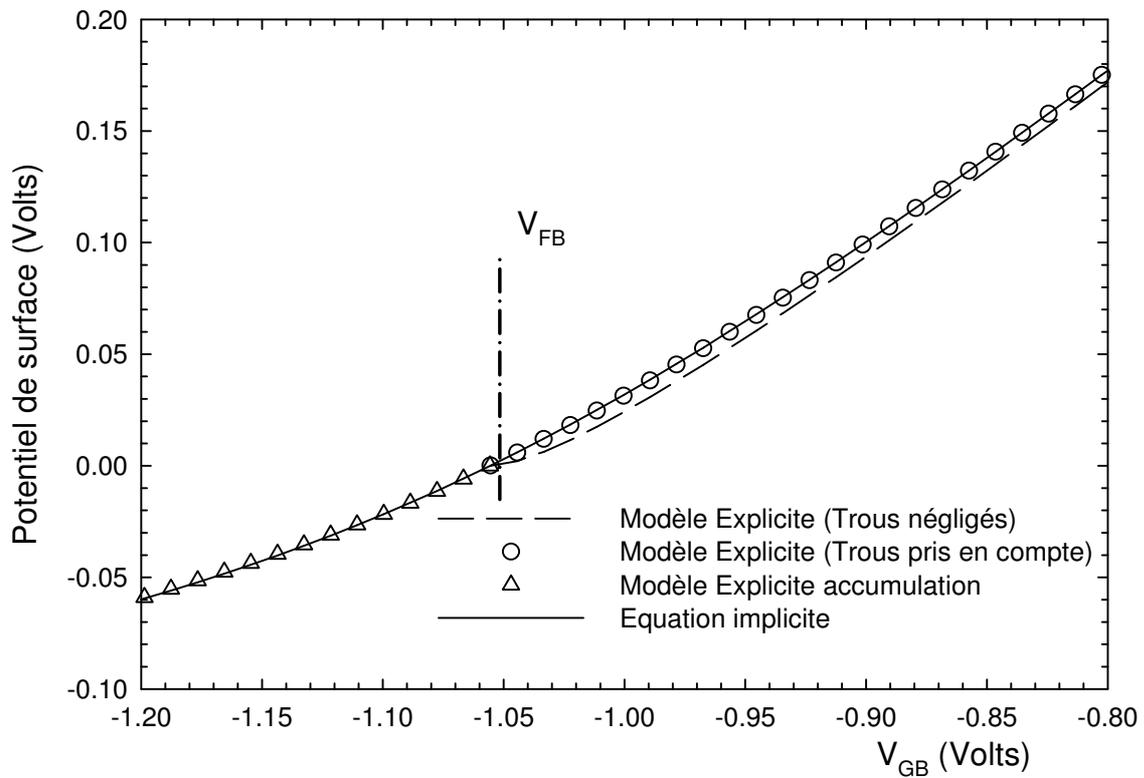


Figure 2.10 : Influence de la prise en compte des trous libres en régime de déplétion au voisinage de la tension de bandes plates.

2.6 Modélisation des effets Quantiques

Nous avons vu au chapitre I une étude qualitative de l'impact du confinement de la couche d'inversion, nécessitant ainsi la prise en compte des effets quantiques dans la modélisation des capacités MOS. Le but de cette partie est de rappeler dans un premier temps la méthode de calcul numérique résolvant l'équation de Poisson et l'équation de Schrödinger couplées. Les résultats ainsi obtenus constituent le résultat de simulation « référence ». Un état de l'art des méthodes d'approximation de la résolution de ces équations sera ensuite détaillée et enfin une nouvelle méthode sera présentée et largement utilisée comme outil de caractérisation des diélectriques des technologies CMOS nanométriques.

2.6.1 Structure de sous bandes d'énergie

Pour calculer les différents niveaux d'énergie et les fonctions d'ondes des électrons confinés dans une couche d'inversion ou d'accumulation au voisinage de l'interface silicium-isolant, il suffit d'intégrer l'équation de Schrödinger

$$\left(T + V(\vec{r})\right) \xi_i(\vec{r}) = E_i \xi_i(\vec{r}) \quad (2.33)$$

où T représente l'opérateur énergie cinétique des électrons, et $V(\vec{r})$ l'opérateur énergie potentielle. ξ_i et E_i représentent respectivement la fonction d'onde et l'énergie associées à l'état propre i . Dans la couche d'inversion (ou d'accumulation), les électrons voient un potentiel relativement complexe. La résolution de l'équation (2.33) implique de calculer le terme d'énergie potentielle $V(\vec{r})$. La discontinuité de la bande de conduction à l'interface silicium-isolant, qui résulte de la différence des affinités électroniques de l'isolant et du semiconducteur, est de l'ordre de 3 eV pour une structure Si-SiO₂. Pour les structures MOS avec un isolant relativement épais, on peut considérer que cette barrière de potentiel est infinie et donc que la fonction d'onde s'annule à l'interface. C'est dans cette hypothèse qu'ont été menés les travaux les plus importants relatifs aux effets quantiques dans les structures MOS [ANDO'82]. Il a été montré que pour les oxydes ultrafins, la présence d'un courant de fuite au travers de l'oxyde perturbe peu les fonctions d'ondes dans le semiconducteur [CASSAN'00]. Pour toute notre étude, nous annulerons la fonction d'onde à l'interface, même pour des oxydes de grille ultrafins faisant apparaître de forts courants de grille.

Parmi les différentes contributions au potentiel $V(\vec{r})$ que voient les électrons, le potentiel périodique lié aux ions du cristal (potentiel cristallin), joue un rôle important. Les

états propres (énergies de la bande de conduction) et les fonctions propres (fonctions de Bloch) sont connus, mais on peut s'affranchir d'une prise en compte de toutes ces contributions en utilisant l'approximation de la masse effective. On considèrera donc un électron se déplaçant dans un réseau périodique vide d'ions, affecté d'une masse effective différente de la masse m_0 de l'électron dans le vide. Cette hypothèse est largement utilisée dans la modélisation de tous les composants électroniques, mais dans le cas des structures MOS, on atteint la limite de validité dans la mesure où la couche d'inversion est confinée dans quelques couches atomiques. Cependant, la méconnaissance de nombreux paramètres physiques tel que la structure réelle de la zone de transition entre silicium et isolant, perturbe bien plus la modélisation du fonctionnement de la structure. L'hypothèse de la masse effective sera donc employée dans toute l'étude.

Dans l'hypothèse de la masse effective, l'opérateur énergie cinétique pour un électron de masse m^* dans la bande de conduction supposée isotrope et parabolique s'écrit :

$$T = -\frac{\hbar^2}{2m^*} \Delta \quad (2.34)$$

L'opérateur énergie potentielle s'écrit :

$$V(\vec{r}) = -q\Psi(\vec{r}) \quad (2.35)$$

où $\Psi(\vec{r})$ est le potentiel électrostatique. L'étude est menée sur une structure à une dimension, donc le vecteur position \vec{r} se réduit à une seule dimension (y). Le mouvement de l'électron n'est donc perturbé que dans la direction y , et reste libre dans le plan (x - z). La fonction d'onde peut s'écrire sous la forme d'un produit de la fonction d'enveloppe par une onde plane :

$$\xi(y) = \varphi_i(y) \exp(ik_1x + ik_2z) \zeta(\vec{r}) \quad (2.36)$$

où $\zeta(\vec{r})$ est la fonction de Bloch, et $\varphi_i(y)$ est une solution de l'équation de la masse effective :

$$\frac{\hbar^2}{2m^*} \frac{d^2\varphi_i(y)}{dy^2} + [E_i - V(z)]\varphi_i(y) = 0 \quad (2.37)$$

avec comme conditions aux limites :

$$\varphi_i(y=0) = 0 \quad \varphi_i(y \rightarrow \infty) = 0 \quad (2.38)$$

Les énergies E_i sont les énergies des électrons dans leur mouvement dans la direction y . Elles décrivent la quantification des états électroniques dans la direction perpendiculaire à l'interface, le mouvement dans le plan (x - z) n'étant pas modifié. Ce système à deux

dimensions est caractérisé par une densité d'état constante en fonction de l'énergie [ANDO'82], et s'exprime par :

$$g(E) = \frac{m^*}{\pi\hbar^2} \sum_i H(E - E_i) \quad (2.39)$$

où $H(x)$ est la fonction de Heaviside. On peut donc définir les densités d'électrons sur chaque sous bande :

$$N_i = \int_{E_i}^{\infty} g_i(E) f(E) dE = \frac{m^*}{\pi\hbar^2} \int_0^{\infty} \frac{dE}{1 + e^{(E-E_i)/kT}} = \frac{m^*}{\pi\hbar^2} kT \ln \left[1 + \exp\left(\frac{E_F - E_i}{kT}\right) \right] \quad (2.40)$$

où $f(E)$ représente l'intégrale de Fermi.

Le carré du module de la fonction d'enveloppe $|\varphi_i(y)|^2$ représente la densité de probabilité de présence d'un électron à la position y . Cette densité d'électrons s'exprime donc simplement par :

$$n(y) = \sum_i N_i |\varphi_i(y)|^2 \quad (2.41)$$

Le résultat donné ici concerne les électrons, et il est facile de l'étendre au cas des trous. La charge totale dans le semiconducteur s'exprime par :

$$Q = -q \int_0^{+\infty} n(y) dy = -q \sum_i N_i \quad (2.42)$$

Dans un semiconducteur à gap direct, le minimum de la bande de conduction se situe en $k=0$. Ce minimum est unique et le semiconducteur est dit univallée. Dans ce cas, les surfaces d'énergie constante au voisinage du minimum sont des sphères centrées en $k_m=0$ et l'énergie d'un électron de la bande de conduction est donnée par [MATHIEU'99] :

$$E = E_C + \frac{\hbar^2 k^2}{2m^*} \quad \text{avec } k^2 = k_x^2 + k_y^2 + k_z^2 \quad (2.43)$$

Le silicium est un semiconducteur possédant un gap indirect avec 6 minima de la bande de conduction. Il en résulte l'existence de 6 vallées. Les surfaces d'énergie constante au voisinage de chacun des minima sont des ellipsoïdes de révolution autour de l'axe considéré, centrés en \bar{k}_m . Dans le cas de l'orientation $\langle 100 \rangle$, les six vallées Δ ellipsoïdales du bas de la bande de conduction du silicium sont à considérer en deux groupes (voir figure 2.11), conduisant à deux séries de niveaux :

- Les deux vallées Δ pour lesquelles la masse longitudinale des électrons m_l^* correspond à la direction perpendiculaire à l'interface sont caractérisées par $m^* = m_l^*$ et $m_{||}^* = m_t^*$. (vallées normales).

- Les quatre vallées Δ pour lesquelles l'axe longitudinal de la surface ellipsoïdale équi-énergie est parallèle à l'interface sont caractérisées par $m^* = m_t^*$ et $m_{//}^* = \sqrt{m_l^* m_t^*}$ (vallées parallèles)

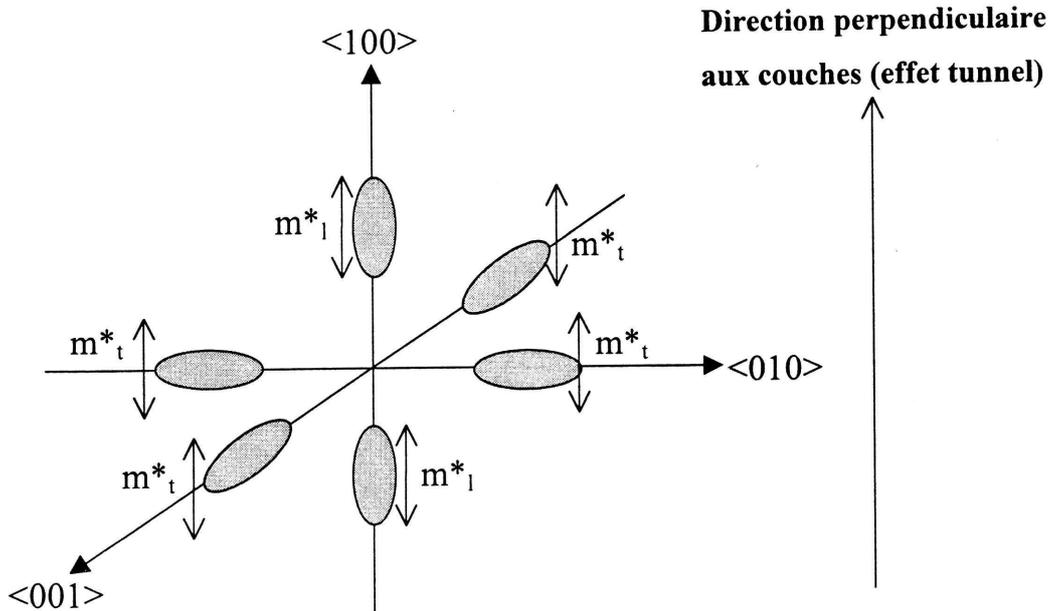


Figure 2.11 : Effet de l'orientation du cristal de silicium par rapport à l'interface Si/SiO₂ sur le nombre de types de vallées à prendre en compte pour la résolution de l'équation de Schrödinger (orientation <100>) [CASSAN'00]

Les densités de charges de chaque niveau pour chacune des vallées s'écrivent donc :

$$N_{v,i} = g_v \frac{m_{//}^{*v}}{\pi \hbar^2} kT \ln \left[1 + \exp \left(\frac{E_F - E_{v,i}}{kT} \right) \right] \quad (2.44)$$

où g_v représente le degré de dégénérescence du type de vallée 'v' ($g_{parallele} = 4$ et $g_{normal} = 2$). Les états d'énergie discrets E_i sont reparties en deux ensembles et sont notés $E_{v,i}$.

Il faut noter que la masse à prendre en compte dans l'évaluation des états discrets (équation (2.37)) est la masse m^* des électrons dans la direction perpendiculaire aux couches, tandis que la masse à prendre en compte dans l'évaluation des densités de charges des niveaux discrets est la masse $m_{//}^*$ dans le plan de l'interface.

Pour les trous, il y a également deux familles, correspondant aux trous lourds de masse effective m_{hh}^* et aux trous légers de masse m_{lh}^* . Le tableau 2 présente les valeurs numériques des différents paramètres utilisés pour le silicium.

Masse longitudinale des électrons	m_l^*	0.92 m_0
Masse transverse des électrons	m_t^*	0.19 m_0
Masse des trous légers	m_{lh}^*	0.16 m_0
Masse des trous lourds	m_{hh}^*	0.49 m_0

Tableau 2 : Paramètres physiques du silicium.

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

$$n(y) = \sum_i 2 \frac{m_t^*}{\pi \hbar^2} kT \ln \left[1 + \exp \left(\frac{E_F - E_{1,i}}{kT} \right) \right] |\varphi_{1,i}(y)|^2 + \sum_j 4 \sqrt{\frac{m_t^* m_l^*}{\pi \hbar^2}} kT \ln \left[1 + \exp \left(\frac{E_F - E_{2,j}}{kT} \right) \right] |\varphi_{2,j}(y)|^2 \quad (2.45)$$

$$p(y) = \sum_i \frac{m_{lh}^*}{\pi \hbar^2} kT \ln \left[1 + \exp \left(\frac{E_F - E_{3,i}}{kT} \right) \right] |\varphi_{3,i}(y)|^2 + \sum_j \frac{m_{hh}^*}{\pi \hbar^2} kT \ln \left[1 + \exp \left(\frac{E_F - E_{4,j}}{kT} \right) \right] |\varphi_{4,j}(y)|^2 \quad (2.46)$$

où les indices {1, 2, 3, 4} pour les énergies propres et les fonctions d'enveloppe représentent les électrons des vallées normales, les électrons des vallées parallèles, les trous légers et les trous lourds, respectivement.

2.6.2 Calcul auto-cohérent de l'équation de Poisson et de l'équation de Schrödinger

Connaissant $n(y)$ et $p(y)$, il est possible de calculer la densité de charges donnée par :

$$\rho(y) = q \cdot (p(y) - n(y) - N_A + N_D) \quad (2.47)$$

où N_A et N_D sont les concentrations des dopants accepteurs et donneurs.

Dans l'approximation de Hartree, en négligeant les énergies d'échange et la polarisation de l'oxyde (force image), l'énergie potentielle $V(y)$ se réduit à une énergie potentielle électrostatique, qui est donnée en résolvant l'équation de Poisson :

$$\frac{\partial^2 V(z)}{\partial y^2} = q \frac{\rho(y)}{\epsilon_{SI}} \quad (2.48)$$

A partir des expressions des densités de porteurs données par les équations (2.45) et (2.46), on peut donc calculer la densité de charges (2.47), et l'intégration de l'équation (2.48) permet d'obtenir l'énergie potentielle électrostatique. Enfin l'équation de Schrödinger peut être intégrée pour déterminer les fonctions d'enveloppe et les niveaux discrets d'énergie. Ce

processus illustre bien le caractère auto-cohérent d'une résolution Schrödinger-Poisson couplée. La technique de résolution est itérative et les temps de calculs sont très longs.

Un potentiel d'essai V_{essai}^i est choisi, permettant l'intégration de l'équation de Schrödinger. Les densités de porteurs ainsi obtenues permettent, après intégration de l'équation de Poisson, de calculer un nouveau potentiel $V_{calculé}^i$, dit potentiel « calculé ». Tant qu'il existe une différence (à une tolérance près fixée par l'utilisateur) entre le potentiel d'essai et le potentiel calculé, on recommence en prenant comme nouveau potentiel d'essai V_{essai}^{i+1} [ANDO'82] :

$$V_{essai}^{i+1} = V_{essai}^i + k(V_{calculé}^i - V_{essai}^i) \quad (2.49)$$

La résolution numérique de l'équation de Schrödinger et de l'équation de Poisson couplées a été longuement étudiée dans la littérature. L'ensemble de ces travaux ont été synthétisés dans l'article de Ando, Folwer et Stern publié en 1982 [ANDO82]. Il décrit toutes les propriétés physiques d'un gaz d'électrons bidimensionnel et fait office de référence dans le domaine. Actuellement, il existe de nombreux simulateurs auto-cohérents Schrödinger-Poisson. L'article de Richter [RICHTER'01] compare les résultats obtenus par les simulateurs les plus répandus. Il met en évidence que tous les simulateurs donnent des résultats similaires en forte inversion, alors qu'en régime d'accumulation, les disparités sur les résultats peuvent donner une erreur significative sur l'extraction de l'épaisseur du film diélectrique (jusqu'à 20%). Nous avons utilisé le simulateur de Berkeley, QMCV [QMCV'00], comme simulateur Schrödinger Poisson de référence dans toute notre étude, d'une part parce qu'il est basé sur la résolution des équations Schrödinger Poisson suivant l'algorithme décrit précédemment, et d'autre part, parce que le code source est accessible librement, permettant ainsi de connaître tous les paramètres physiques utilisés (notamment les masses effectives ou le nombre de sous bandes calculées).

Les résultats de calculs auto cohérents sont présentés sur les figures 2.12 et 2.13. Ces résultats ont été obtenus en négligeant le potentiel image et les interactions électrons-électrons. La figure 2.12 représente le diagramme énergétique et la distribution spatiale des électrons, pour une capacité nMOS avec un substrat p dopé à $N_A=10^{18} \text{ cm}^{-3}$, une grille polysilicium dopée à $N_P=5 \cdot 10^{19} \text{ cm}^{-3}$, où la polydéplétion n'a pas été prise en compte, dans le but de montrer l'impact des effets quantiques seuls. L'origine des énergies est prise dans la zone neutre du substrat, loin de l'interface. Les deux premiers niveaux d'énergie de chaque vallée ont été représentés, ainsi que les courbes $|\varphi_{i,j}|^2$, représentant la distribution spatiale des

électrons. La figure 2.13 présente les densités d'électrons de chaque niveau d'énergie, ainsi que la densité totale d'électrons (cf. équation (2.45)). Le maximum de la densité n'est plus à l'interface comme le prévoit un modèle classique. Il est décalé dans le substrat de l'ordre de 1.5 nm. On constate que la majorité des porteurs est situé sur le premier niveau d'énergie des vallées normales. Ce résultat très important constitue la base de tous les modèles approchés de calcul quantique sur les couches d'inversion sans utiliser un algorithme auto cohérent trop lourd en terme de temps de calcul.

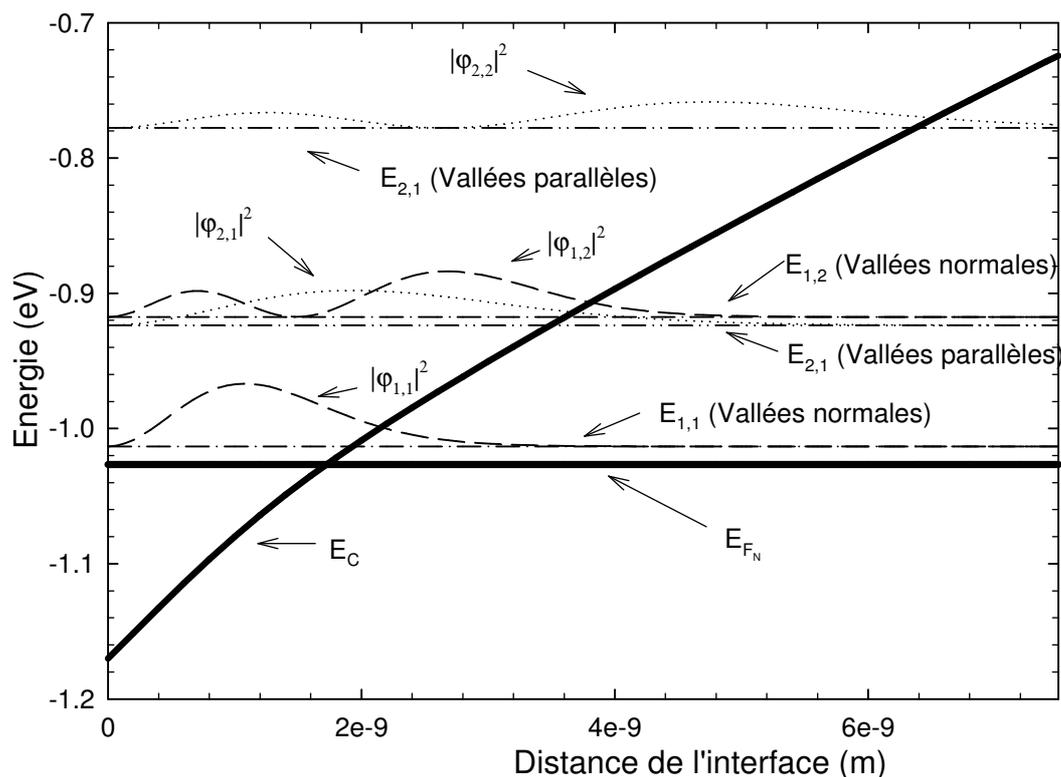


Figure 2.12 : Résultats d'un calcul auto-cohérent dans une capacité nMOS à substrat p dopé à 10^{18} cm^{-3} , avec une grille polysilicium dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$ polarisée en forte inversion avec $V_{GB} = 0.5$ Volts. La polydéplétion n'a pas été prise en compte. Les deux premiers niveaux d'énergie de chaque vallée sont représentés ainsi que les carrés des modules des fonctions d'enveloppe (densité de probabilité de présence) associées.

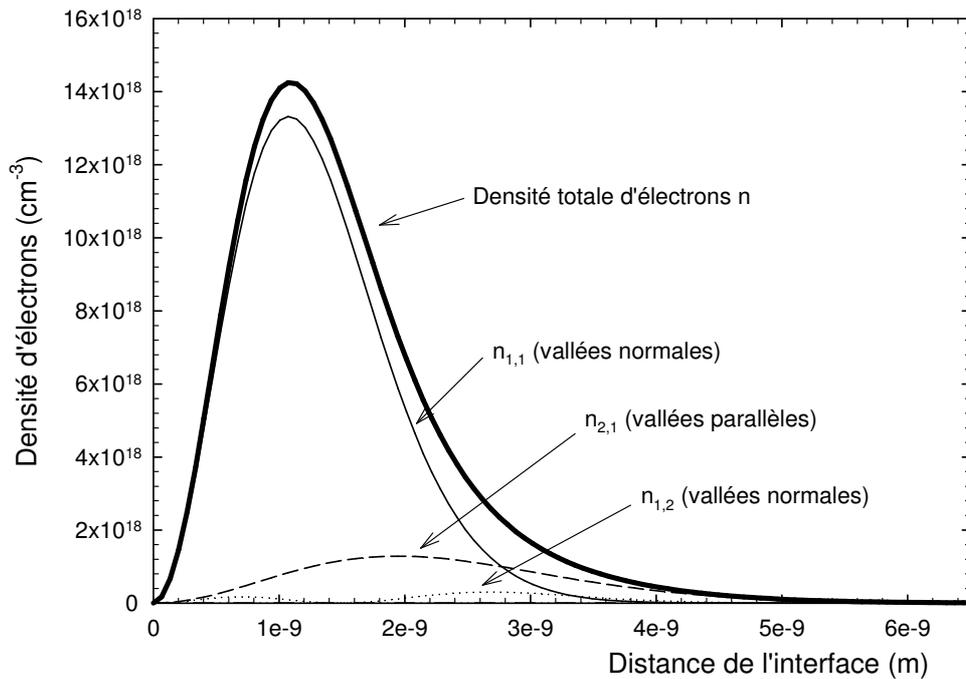


Figure 2.13 : Résultat d'un calcul auto-cohérent dans une capacité nMOS à substrat p dopé à 10^{18} cm^{-3} , avec une grille polysilicium dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$ polarisée en forte inversion avec $V_{GB} = 0.5$ Volts. La polydépétion n'a pas été prise en compte. La densité totale d'électrons est représentée, ainsi que la contribution à cette densité totale des deux premiers niveaux d'énergie des vallées parallèles (vallées de basse énergie) ainsi que le premier niveau d'énergie des vallées normales (vallées de haute énergie). On remarque que la majorité des porteurs se trouve sur le premier niveau d'énergie des vallées normales .

Les simulateurs numériques physiques permettent de modéliser précisément l'impact des effets quantiques sur les caractéristiques électriques d'une capacité MOS, tel que le potentiel électrostatique, la charge d'inversion, et par suite la capacité de la structure. Néanmoins, les temps de calculs très importants rendent leur utilisation impossible dans les modèles destinés à la simulation de circuits (BSIM, EKV, MM11). Leur utilisation reste difficile comme outil de caractérisation : les techniques itératives d'optimisation de paramètres sont elles mêmes déjà très longues, et l'implantation d'un algorithme auto cohérent à l'intérieur d'une itération d'optimisation ne permet pas d'obtenir des paramètres dans des délais raisonnables. D'autre part, on verra dans la suite de ce chapitre que la simulation d'un transistor MOS incluant les effets quantiques reste difficile avec la résolution des équations Schrödinger Poisson couplées. De nombreuses méthodes d'approximation de la solution exacte Schrödinger Poisson couplée a été largement étudiée, donnant parfois des résultats très différents [YU'00]. Dans la partie suivante, nous allons présenter les principaux modèles analytiques.

2.6.3 Approche Analytique des effets Quantiques

Dans le but de résoudre de manière approchée les équations couplées de Poisson et de Schrödinger, deux principales méthodes d'approximation ont été développées. Les formulations ainsi obtenues ne requièrent que de faibles temps de calculs et sont facilement intégrables dans des algorithmes de caractérisation ou simulation de circuits.

2.6.3.1 L'approximation du puits de potentiel triangulaire.

Cette méthode consiste à approximer l'énergie potentielle par une droite, et par suite de considérer le puits de potentiel où sont confinés les électrons comme triangulaire (figure 2.14)

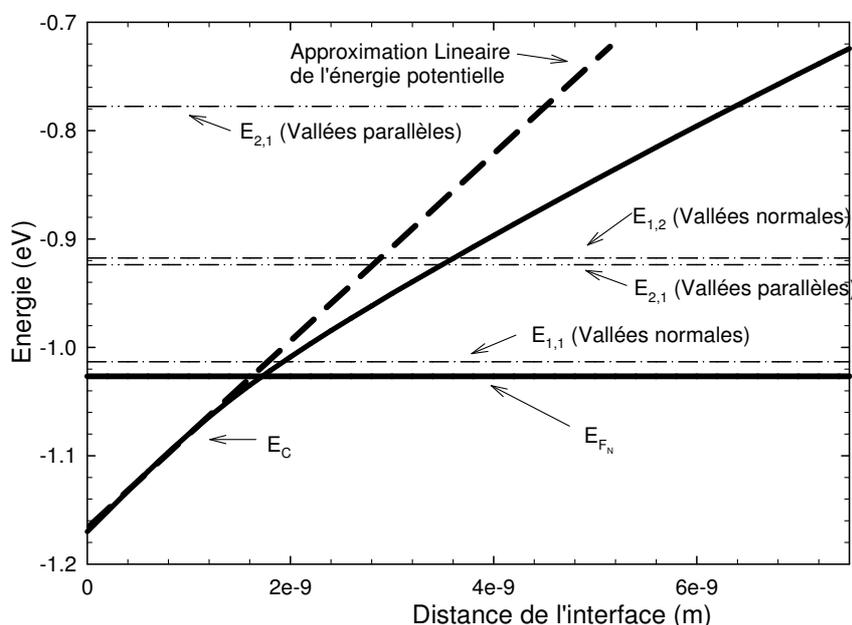


Figure 2.14 : Approximation du potentiel triangulaire.

L'énergie potentielle est approximée par la relation [STERN'72]:

$$V(y) = q E_{eff} y = \frac{q^2 (N_{dep} + f N_{inv})}{\epsilon_{Si}} y \quad (2.50)$$

où N_{dep} représente la charge de déplétion, N_{inv} la charge d'inversion, f représente un coefficient pondérant la participation moyenne des charges d'inversion au champ électrique effectif présent au voisinage de l'interface coté semiconducteur. $f=0$ correspond à la seule contribution de la charge de déplétion, $f=1$ donne à E_{eff} la valeur du champ à l'interface, et $f=1/2$ donne à E_{eff} une valeur moyenne à l'intérieur de la zone d'inversion.

Dans ces conditions, l'équation de Schrödinger (2.36) admet des solutions analytiques, et s'exprime en utilisant les fonctions de Airy. Les fonctions enveloppes des fonctions d'onde s'exprime par :

$$\varphi_i(y) = A_i \left(\frac{2m^* q E_{eff}}{\hbar^2} \right)^{1/3} \left[y - \frac{E_i}{q E_{eff}} \right] \quad (2.51)$$

Et les énergies correspondantes sont données par :

$$E_i = \left(\frac{\hbar^2}{2m^*} \right)^{1/3} \left[\frac{3}{2} \pi q E_{eff} \left(i + \frac{3}{4} \right) \right]^{2/3} \quad (2.52)$$

Les niveaux d'énergies ainsi calculés sont référencés par rapport au bas de la bande de conduction en $y=0$.

Cette méthode présente l'avantage de pouvoir calculer toutes les énergies et fonctions d'ondes de toutes les sous bandes. Même si l'on a vu que la majorité des porteurs libres se trouve sur le premier niveau d'énergie des vallées normales en forte inversion, ce n'est pas forcément le cas en déplétion et en accumulation. Cette méthode est cependant limitée par le terme E_{eff} , car si en déplétion le facteur f doit être pris à zéro, il faut l'augmenter avec l'accroissement de la charge d'inversion pour obtenir des résultats proches d'un calcul auto cohérent Schrödinger Poisson.

Le calcul par cette méthode de la densité de porteurs N_{inv} passe donc par un calcul auto cohérent. Les équations (2.51) et (2.52) permettent de calculer la population de chaque sous bande (équation (2.45)). La charge de déplétion et d'inversion permettent de calculer le champ effectif, et on itère jusqu'à convergence. Une utilisation de l'approximation de potentiel triangulaire de cette manière ne présente que peu d'intérêt. Dans le cadre de la recherche d'un modèle quantique liant le potentiel de surface et la tension de grille, l'équation (2.52) a été utilisée, en introduisant un pseudo élargissement du gap dans la zone de confinement des porteurs. Nous détaillerons cette méthode au paragraphe 2.6.4.

2.6.3.2 L'approximation variationnelle

L'approximation triangulaire est intéressante car elle fournit l'ensemble des fonctions d'ondes et des niveaux d'énergie, mais leur utilisation est peu pratique pour le calcul des propriétés de la couche d'inversion, et plus particulièrement leur réponse à une perturbation extérieure. Il est utile de disposer d'une fonction d'onde analytique simple. De plus, dans les régimes d'inversion, le fait de ne considérer que les électrons de la première sous bande des

vallées normales constitue une très bonne approximation. Une fonction d'onde analytique simple pour ces électrons a été proposé par Fang et Howard [FANG'66] :

$$\varphi_{1,1}(y) = \left(\frac{b^3}{2}\right)^{1/2} y e^{-\frac{by}{2}} \quad (2.53)$$

Le paramètre variationnel b est obtenu en minimisant l'énergie totale du système pour des valeurs données de densités de charge de déplétion et d'inversion.

Nous allons présenter ici les points importants du calcul du paramètre b , puisque l'approximation variationnelle constitue la base de notre modèle de prise en compte des effets quantiques dans la relation entre le potentiel de surface, la tension de grille et le quasi niveau de Fermi, qui sera détaillé par la suite dans ce mémoire.

La pénétration moyenne des électrons est donnée par :

$$y_m = \int_0^\infty y |\varphi_{1,1}(y)|^2 dy = \frac{b^3}{2} \int_0^\infty y^3 e^{-by} dy = \frac{3}{b} \quad (2.54)$$

Le paramètre $\frac{3}{b}$ représente la position du centroïde de la couche d'inversion. Compte tenu de la simplicité de la fonction d'onde, on peut facilement calculer l'énergie totale des électrons, qui est donnée par :

$$\begin{aligned} E &= \langle T \rangle + \langle V \rangle \\ \text{avec } \langle T \rangle &= -\frac{\hbar^2}{2m^*} \int_0^\infty \varphi_{1,1}(y) \varphi_{1,1}''(y) dy \\ \langle V \rangle &= \int_0^\infty \left(V_{dep}(y) + \frac{1}{2} V_{inv}(y) \right) |\varphi_{1,1}(y)|^2 dy \end{aligned} \quad (2.55)$$

Compte tenu de l'expression analytique de la fonction d'onde, la valeur propre de l'énergie cinétique s'exprime par :

$$\langle T \rangle = \frac{\hbar^2 b^2}{8m^*} \quad (2.56)$$

L'énergie potentielle liée à la charge de déplétion se calcule facilement en intégrant l'équation de Poisson et en considérant que la zone de déplétion s'étend sur une distance y_d . L'énergie potentielle liée à la couche d'inversion se calcule dans l'hypothèse du potentiel de Hartree, c'est-à-dire en remplaçant l'énergie d'interaction electron-electron par une énergie moyenne, représentant le potentiel créé au point y par l'ensemble des autres électrons. Dans ces conditions, l'énergie potentielle des électrons est donnée par :

$$\langle V \rangle = \langle V_{dep} \rangle + \frac{1}{2} \langle V_{inv} \rangle = \frac{3q^2 N_{dep}}{\epsilon_{Si} b} + \frac{33}{32} \frac{q^2 N_{inv}}{\epsilon_{Si} b} \quad (2.57)$$

L'énergie totale du système est donnée par :

$$E = \langle T \rangle + \langle V \rangle = \frac{\hbar^2 b^2}{8m^*} + \frac{3q^2 N_{dep}}{\epsilon_{Si} b} + \frac{33}{32} \frac{q^2 N_{inv}}{\epsilon_{Si} b} \quad (2.58)$$

Pour minimiser l'énergie du système, il suffit de prendre la valeur du paramètre variationnel b tel que $\frac{dE}{db} = 0$. Il faut donc prendre :

$$b = \left[\frac{12m^* q^2}{\epsilon_{Si} \hbar^2} \left(N_{dep} + \frac{11}{32} N_{inv} \right) \right]^{1/3} \quad (2.59)$$

Avec l'aide de la fonction d'onde variationnelle, l'équation de Schrödinger s'écrit :

$$H\varphi \simeq E(b)\varphi \quad (2.60)$$

où H est l'Hamiltonien du système et E son énergie. L'énergie s'exprime donc en fonction de b :

$$E(b) = \frac{3}{8} \frac{\hbar^2}{m^*} b^2 \quad (2.61)$$

L'énergie E représente l'écart entre le bas de la bande de conduction et le premier niveau d'énergie des vallées normales (niveau le plus peuplé). On a donc, grâce à l'approximation variationnelle, une relation entre les charges de déplétion et d'inversion, et le pseudo élargissement du gap à l'interface oxyde/semiconducteur.

Les deux techniques d'approximations sont à la base de nombreux modèles analytiques permettant d'inclure les effets quantiques dans les équations liant le potentiel de surface à la tension de grille et au quasi niveau de Fermi. Dans la partie suivante, nous allons détailler les principales approches existantes, puis nous présenterons le modèle développé spécifiquement dans le cadre de cette thèse.

2.6.4 Modélisation implicite du potentiel de surface avec prise en compte des effets quantiques.

La modélisation du potentiel de surface est une étape cruciale dans l'étude d'une structure MOS. Elle permet en effet d'établir le lien entre les différentes variables physiques, telles que les densités de porteurs, le champ électrique, le courant de grille, capacité de la structure, et les tensions externes appliquées. La caractérisation du diélectrique passe par la connaissance de ces paramètres. De plus, les modèles de transistor, de courant de fuite et de bruit basse fréquence lié au courant de grille développés dans ce mémoire reposent sur la connaissance précise du potentiel de surface. Dans un premier temps nous allons détailler les

techniques mises en œuvre pour modéliser les régimes de déplétion et d'inversion. Le régime d'accumulation sera ensuite abordé.

2.6.4.1 Modèle de van Dort

Parmi les différents impacts des effets quantiques sur le fonctionnement d'une structure MOS, le premier phénomène étudié fut l'augmentation de la tension de seuil. Le modèle de Van Dort est essentiellement destiné à corriger la tension de seuil des modèles compacts électriques, en ajoutant à la valeur de tension de seuil classique un terme correctif. Naturellement, au regard de ce qui a été développé dans les parties précédentes, ce terme doit être étroitement lié à la tension de grille externe appliquée. De plus, ces modèles sont orientés pour donner de bons résultats en terme de simulation compacte, avec la possibilité d'utiliser un nombre de paramètres empiriques, ce qui est incompatible avec la finalité de notre étude, c'est-à-dire élaborer un modèle précis pour la caractérisation et non pour la simulation de circuits. Le caractère non itératif des modèles compacts est une obligation pour la simulation de circuits, et ceci au détriment de la signification physique des paramètres.

L'hypothèse de base de l'approche de Van Dort est de conserver les équations classiques des structures MOS, c'est-à-dire la résolution de l'équation de Poisson et des densités de porteurs décrites par la statistique de Boltzmann, tout en augmentant la valeur du gap à l'interface oxyde/semiconducteur. Cette approche intuitive est motivée par le fait que la majorité des électrons se situent sur le premier niveau d'énergie des vallées normales situé au dessus du bas de la bande de conduction.

La valeur du gap intervient dans la définition de la concentration intrinsèque n_i , dont l'expression est donnée par :

$$n_i = \sqrt{N_c N_v} \exp\left[-\frac{E_g}{2kT}\right] \quad (2.62)$$

où N_c désigne la densité d'états dans la bande de conduction ($N_c = 1.04 \cdot 10^{19} \text{ cm}^{-3}$), N_v la densité d'états dans la bande de valence ($N_v = 2.8 \cdot 10^{19} \text{ cm}^{-3}$) et E_g la largeur du gap du silicium. En modélisant l'impact des effets quantiques par une augmentation du gap, on introduit alors une concentration intrinsèque « quantique » n_i^{QM} , relative à un gap de largeur égale à $E_g + \Delta E_g$, qui s'exprime par :

$$n_i^{QM} = n_i \exp\left[-\frac{\Delta E_g}{2kT}\right] \quad (2.63)$$

où n_i est la concentration intrinsèque classique. Van Dort a publié deux articles relatifs à la modélisation de l'augmentation de la tension de seuil [VAN'92, VAN'94]. Nous ne détaillerons pas les résultats du premier article [VAN'92], car ce modèle contient des paramètres empiriques et les choix effectués par l'auteur sont discutables. De plus, la finalité de ce modèle concerne uniquement la modélisation de la tension de seuil, et non la modélisation de la caractéristique $\Psi_s - V_{GB}$ complète. La comparaison de ce seul paramètre peut donner des résultats corrects, alors qu'il y a un très mauvais accord pour la caractéristique $\Psi_s - V_{GB}$. Dans le deuxième article de Van Dort [VAN'94], l'augmentation de la largeur du gap est donnée par la relation suivante :

$$\Delta E_g = E_0 + qE_s \Delta y = \beta \left[\frac{\epsilon_{Si}}{4kT} \right]^{1/3} \max(E_s, 0)^{2/3} + qE_s \Delta y \quad (2.64)$$

où β est une constante physique fixée par l'auteur à $4.1 \cdot 10^{-8}$ eV cm⁴. Le premier terme correspond à l'énergie du premier niveau d'énergie calculé dans l'approximation triangulaire et dépend de la tension de polarisation par le terme E_s représentant le champ électrique à la surface, coté semiconducteur. Le second terme représente l'effet de déplacement dans le volume de la couche d'inversion (centroïde d'inversion). L'auteur a choisi d'utiliser le résultat de l'approche variationnelle (2.54) pour modéliser le terme Δy . Après simplifications, Δy s'exprime par :

$$\Delta y = \frac{4}{9} \frac{E_0}{qE_s} \quad (2.65)$$

Au final, le terme d'élargissement du gap pris en compte s'exprime donc par :

$$\Delta E_g = \frac{13}{9} \beta \left[\frac{\epsilon_{Si}}{4kT} \right]^{1/3} \max(E_s, 0)^{2/3} \quad (2.66)$$

La tension de seuil correspond à la tension de grille à appliquer pour obtenir un potentiel de surface égal à $2\phi_F$. En appliquant le théorème de Gauss (cf. équation(2.9)) , la tension de seuil s'exprime par :

$$V_{th} = V_{FB} + 2\phi_F + \frac{Q_{SC}(2\phi_F)}{C_{ox}} \approx V_{FB} + 2\phi_F + \frac{Q_{dep}(2\phi_F)}{C_{ox}} \approx V_{FB} + 2\phi_F + \gamma \sqrt{2\phi_F} \quad (2.67)$$

⁴ Conformément à l'approximation du potentiel triangulaire, la valeur théorique de β est $4.4 \cdot 10^{-8}$ eV

Pour établir cette relation, la charge totale dans le semiconducteur est approximée à la seule charge de déplétion, qui est uniquement liée au potentiel de surface⁵.

La courbure de bande nécessaire pour arriver à l'état d'inversion induit une augmentation du potentiel de surface, s'exprimant par :

$$\Delta\Psi_s = \frac{\Delta E_g}{q} \quad (2.68)$$

La tension tenant compte des effets quantiques s'exprime alors :

$$V_{th}^{QM} = V_{FB} + 2\phi_F + \Delta\Psi_s + \gamma\sqrt{2\phi_F + \Delta\Psi_s} \quad (2.69)$$

En considérant que cet incrément de potentiel de surface est petit devant $2\phi_F$, on obtient :

$$\Delta V_{th} = V_{th}^{QM} - V_{th}^{CL} = \Delta\Psi_s \left(1 + \gamma \frac{1}{2\sqrt{2\phi_F}} \right) = \Delta\Psi_s \left(1 + \frac{1}{2C_{ox}} \sqrt{\frac{qN_A \epsilon_{Si}}{\phi_F}} \right) \quad (2.70)$$

Cette formulation n'est pas celle exactement employée par l'auteur, puisque qu'il a introduit une capacité effective, en considérant la zone près de l'interface peu peuplée de porteurs (dark space) comme une capacité en série avec la capacité d'oxyde C_{ox} . Comme l'ont souligné certains travaux focalisés sur la modélisation des effets quantiques pour les modèles compacts [PREGALDINY'03], l'utilisation d'une valeur effective de capacité est surprenante, car le déplacement du centroïde d'inversion a déjà été pris en compte dans l'expression de l'élargissement du gap.

Cette modélisation de l'augmentation de la tension de seuil est évidemment destinée aux modèles compacts basés sur la tension de seuil, dont le plus utilisés est BSIM [BSIM'01]. Cependant, l'impact des effets quantiques ne se limite pas à l'augmentation de la tension de seuil, ils affectent tout le régime d'inversion modérée et de forte inversion. Le régime d'inversion modérée concerne la zone de transition entre la faible inversion, où la charge totale développée dans le substrat est dominée par la charge de déplétion, et le régime de forte inversion où la charge totale est dominée par la charge d'inversion. Dans les modèles basés sur la tension de seuil, cette transition est assurée par une fonction de lissage liant un modèle valable pour la faible inversion et une autre valable pour la forte inversion.

⁵ On utilise ici le concept de feuille de charge pour modéliser les propriétés de la charge de déplétion. Ce concept sera développé en détail plus tard dans ce chapitre lors de la formulation du modèle pour le transistor MOS.

2.6.4.2 Modèle implicite quantique basé sur le modèle de van Dort

La modélisation d'un transistor MOS basée sur le potentiel de surface permet de s'affranchir de toutes les limitations liées à la définition d'une tension de seuil. Il n'y a plus à considérer de décalage de la tension de seuil, ni à utiliser des paramètres effectifs tels que la capacité d'oxyde effective. Tous les paramètres gardent leur sens physique. De plus, le modèle est formulé dans une seule équation⁶, garantissant ainsi une approche unifiée et cohérente du régime de déplétion jusqu'à la forte inversion.

L'approche de van Dort introduit une densité intrinsèque quantique prenant en compte l'élargissement du gap ΔE_g . L'équation implicite classique valable dans tous les régimes de fonctionnement peut se réécrire en faisant apparaître explicitement la densité de porteurs intrinsèque :

$$\begin{aligned} & \left(\frac{V_{GB} - V_{FB} - \Psi_S - \Psi_P}{\gamma} \right)^2 \\ &= \Psi_S + \phi_T \frac{n_i}{N_A} \exp\left(\frac{\phi_F}{\phi_T}\right) \cdot \left[\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1 \right] + \phi_T \cdot \frac{n_i}{N_A} \exp\left(\frac{-V - \phi_F}{\phi_T}\right) \left[\exp\left(\frac{\Psi_S}{\phi_T}\right) - 1 \right] \end{aligned} \quad (2.71)$$

En utilisant la densité intrinsèque de porteurs n_i^{QM} à la place de n_i , on obtient la relation explicite suivante [RIOS'95]⁷

$$\begin{aligned} & (V_{GB} - V_{FB} - \Psi_S - \Psi_P)^2 \\ &= \gamma^2 \left[\Psi_S + \phi_T \exp\left(-\frac{\Delta E_g}{2kT}\right) \left(\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1 \right) + \phi_T \exp\left(-\frac{V + \phi_B}{\phi_T}\right) \exp\left(-\frac{\Delta E_g}{2kT}\right) \left(\exp\left(\frac{\Psi_S}{\phi_T}\right) - 1 \right) \right] \end{aligned} \quad (2.72)$$

La modélisation de l'élargissement du gap ΔE_g est conforme à l'expression proposée par van Dort, à savoir :

$$\begin{aligned} \Delta E_g &= \frac{13}{9} \beta \left[\frac{\epsilon_{Si}}{4kT} \right]^{1/3} \max(E_S, 0)^{2/3} \\ \text{avec} \quad E_S &= \frac{V_{GB} - V_{FB} - \Psi_S - \Psi_P}{T_{ox}} \end{aligned} \quad (2.73)$$

⁶ Pour l'instant, cette équation est implicite, mais nous verrons par la suite le conditionnement mathématique permettant d'obtenir une expression explicite.

⁷ Cette approche a été proposée par Rios et al., mais la publication comportait une erreur de typographie dans l'ordre des parenthèses.

En injectant l'équation (2.73) dans (2.72), on obtient donc une équation implicite liant V_{GB} , Ψ_s et V . Le terme Ψ_p est modélisé en utilisant l'équation (2.19). Ce modèle tient compte à la fois des effets quantiques et de la polydéplétion de grille.

La figure 2.15 présente la comparaison entre la solution obtenue par calcul auto cohérent (simulateur QMCV) et le modèle décrit par les équations (2.72) et (2.73), pour une capacité nMOS avec $T_{ox}=1.2$ nm, $N_A=10^{18}$ cm⁻³, $N_P=9 \cdot 10^{19}$ cm⁻³. Deux valeurs du paramètre β ont été utilisées. La valeur $4.1 \cdot 10^8$ eV cm correspond à la valeur proposée par van Dort. Le paramètre β a une valeur physique qui provient directement de l'approximation du puits de potentiel triangulaire⁸. Cependant, la valeur permettant d'ajuster au mieux le résultat auto cohérent est $3.6 \cdot 10^8$ eV cm. La figure 2.16 présente les mêmes résultats pour un oxyde plus fin ($T_{ox}=1$ nm) et un substrat plus dopé ($N_A=5 \cdot 10^{18}$ cm⁻³), avec les mêmes valeurs du paramètre β .

Il apparaît clairement que ce modèle peut donner des résultats proches de la solution auto cohérente, mais sous réserve d'ajuster le paramètre β , ce qui n'est pas, d'un point de vue physique, satisfaisant. L'idée d'utiliser la notion de gap élargie est intéressante, mais cela sous entend de disposer d'un modèle d'élargissement du gap performant. Le modèle de van Dort a été initialement prévu pour corriger la tension de seuil, et son extension à la simulation de la caractéristique complète $\Psi_s(V_{GB})$ ne donne pas de bons résultats. Le problème principal vient de l'approximation du puits triangulaire elle-même : ce modèle est dépendant d'un champ effectif, qui doit dériver plutôt du potentiel de déplétion en faible inversion, et plutôt du potentiel d'inversion en forte inversion. D'autre part, la zone de transition (inversion modérée) doit aussi être prise en considération. Il donc impossible de concilier tous ces éléments en utilisant uniquement le champ électrique à l'interface E_s comme l'a fait van Dort.

D'un point de vue physique, cette approche est donc très discutable. Ce modèle n'est donc pas satisfaisant en tant qu'outil de caractérisation électrique, puisque un maximum de précision est recherché. D'un point de vue de la modélisation compacte, une approche empirique n'est pas rédhibitoire ; un article [TANG'02] propose d'ailleurs une loi empirique pour le paramètre β . Une relation empirique liant β à T_{ox} et N_A permet d'obtenir une bonne

⁸ Les travaux de van Dort sur les effets quantiques ont été publiés en deux parties. Dans sa première formulation l'auteur a effectivement proposé la valeur de $4.4 \cdot 10^8$ eV cm (valeur théorique qui découle directement du puits de potentiel triangulaire), puis a utilisé la valeur de $4.1 \cdot 10^8$ eV cm dans son deuxième article.

concordance pour différentes épaisseurs d'oxydes et niveaux de dopages. Cette approche est évidemment totalement inadéquate avec l'approche physique que nous recherchons.

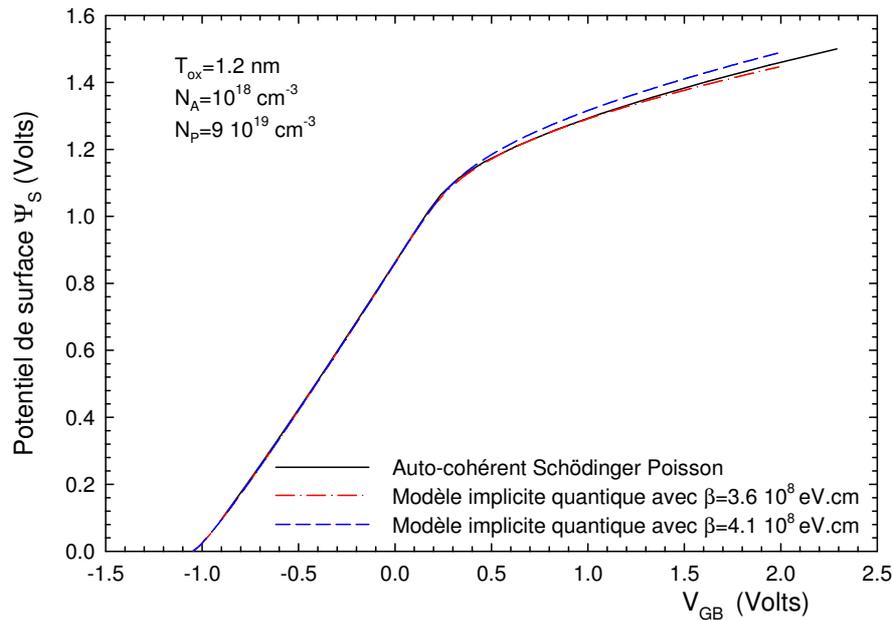


Figure 2.15 : Résultats de simulation du potentiel de surface en fonction de la tension de grille pour une capacité nMOS avec $T_{ox} = 1.2 \text{ nm}$, $N_A = 10^{18} \text{ cm}^{-3}$, $N_P = 9 \cdot 10^{19} \text{ cm}^{-3}$.

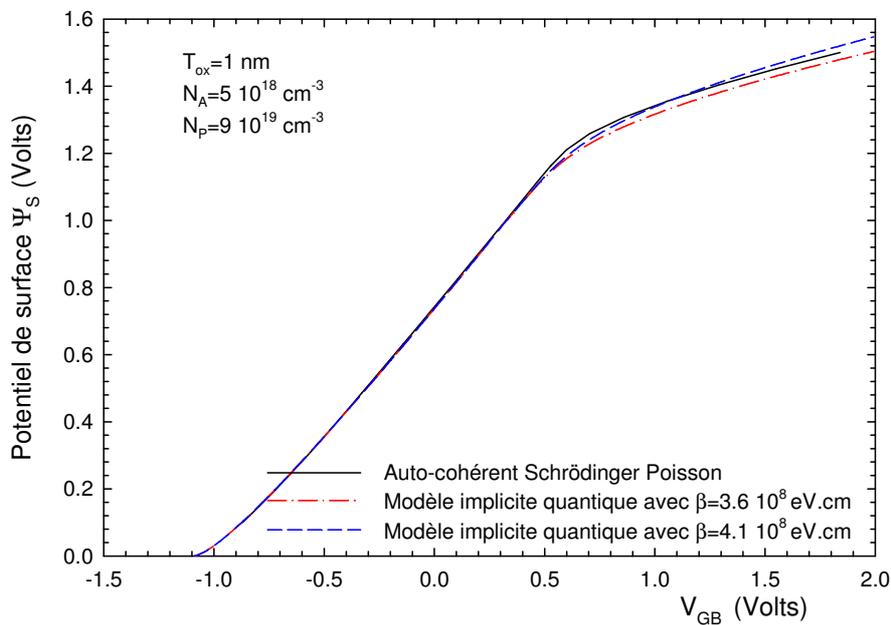


Figure 2.16 : Résultats de simulation du potentiel de surface en fonction la tension de grille pour une capacité nMOS avec $T_{ox} = 1 \text{ nm}$, $N_A = 5 \cdot 10^{18} \text{ cm}^{-3}$, $N_P = 9 \cdot 10^{19} \text{ cm}^{-3}$.

2.6.4.3 Modèle implicite quantique basé sur l'approche variationnelle : Régime de déplétion et d'inversion

La prise en compte des effets quantiques par un pseudo élargissement du gap mène à des résultats peu satisfaisants d'un point de vue physique, puisque le paramètre principal du modèle β doit être ajusté suivant l'épaisseur d'oxyde ou le dopage du substrat. Nous avons décidé d'abandonner l'approximation du puits triangulaire comme approche analytique des effets quantiques, et de ne plus utiliser le modèle de pseudo élargissement du gap à l'interface.

En régime de déplétion et d'inversion, La majorité des porteurs libres est située sur le premier niveau d'énergie des vallées de basse énergie. L'approche variationnelle repose sur la connaissance du paramètre b (cf § 2.6.3.2). La connaissance de ce paramètre permet le calcul du premier niveau d'énergie et donc de la densité de charge d'inversion (équation (2.45)).

Le paramètre b est fonction de la densité de charges de la couche d'inversion et de la densité de charges de déplétion. Le point clé de cette approche est donc d'exprimer ces densités de charges en fonction du potentiel de surface. L'approximation de la feuille de charge [BREWS'78] permet d'obtenir, d'une part, une expression ne dépendant que du potentiel de surface pour la charge de déplétion, et d'autre part, une expression ne dépendant que du potentiel de surface et de la tension de grille pour la charge d'inversion. Ces résultats sont détaillés en annexe A. Suivant l'approche proposée par van Dort, nous avons, dans un premier temps, reformulé les expressions de densité de charges en utilisant la concentration intrinsèque « quantique ». Les densités de charges s'expriment alors par :

$$Q_{dep}(\Psi_S) = -C_{ox}\gamma\sqrt{\Psi_S + \phi_T \exp\left(-\frac{\Delta E_g}{2kT}\right)\left(\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1\right)} \quad (2.74)$$

$$Q_{inv}(V_{GB}, \Psi_S) = -C_{ox}\left[V_{GB} - V_{FB} - \Psi_S - \Psi_P - \gamma\sqrt{\Psi_S + \phi_T \exp\left(-\frac{\Delta E_g}{2kT}\right)\left(\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1\right)}\right] \quad (2.75)$$

Ces deux expressions modifiées permettent de mieux appréhender l'impact des effets quantiques sur les charges de déplétion et d'inversion. Il apparaît que le calcul des charges de façon classique est valable. En effet, le terme exponentiel dans l'équation (2.74) tend vers zéro pour le régime de forte inversion, c'est-à-dire dans le régime où le terme ΔE_g est le plus grand. De plus, dans le régime de déplétion (région où le terme exponentiel n'est pas négligeable), c'est le terme ΔE_g qui est proche de zéro. On peut donc conclure que le calcul

de la charge de déplétion sous l'hypothèse de la feuille de charge reste valable. La charge d'inversion est obtenue en enlevant à la charge totale (issue du théorème de Gauss, donc toujours valable même en présence d'effets quantiques) la charge de déplétion calculée précédemment. Le calcul de la charge d'inversion classique reste donc valable.

Pour valider ces conclusions, nous avons simulé le potentiel de surface, la densité de charges de déplétion et la densité de charges libres avec le simulateur auto cohérent QMCV. Sur la figure 2.17 sont reportées les densités de porteurs simulées de cette façon, ainsi que les charges calculées à partir des expressions suivantes :

$$Q_{dep}(\Psi_s) = -C_{ox} \gamma \sqrt{\Psi_s + \phi_T \left(\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right)} \quad (2.76)$$

$$Q_{inv}(V_{GB}, \Psi_s) = -C_{ox} \left(V_{GB} - V_{FB} - \Psi_s - \Psi_p - \gamma \sqrt{\Psi_s + \phi_T \left(\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right)} \right) \quad (2.77)$$

Le très bon accord entre les deux approches valide ainsi l'hypothèse que la formulation en feuille de charges pour les densités de charges de déplétion et d'inversion reste valide en présence d'effets quantiques.

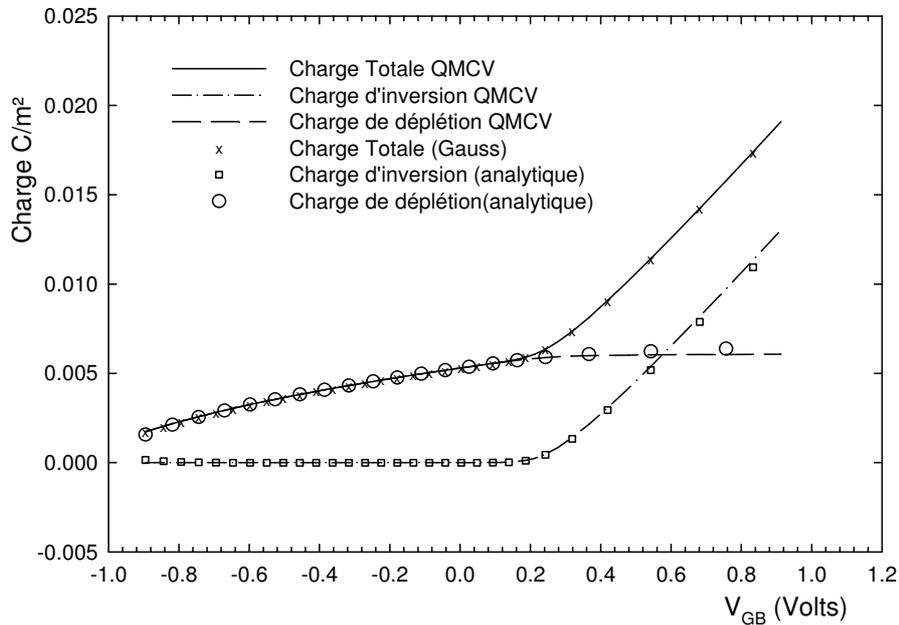


Figure 2.17 : Densités de charges issues d'un calcul auto cohérent Poisson Schrödinger et de l'approximation de la feuille de charge.

Le modèle implicite quantique repose donc d'une part sur l'approche variationnelle, et d'autre part sur le modèle en feuille de charges des densités de charges. Pour mémoire, la fonction d'onde des électrons présents sur le premier niveau d'énergie des vallées de plus basse énergie s'écrit sous la forme :

$$\zeta(y) = \sqrt{\frac{b^3}{2}} y \exp\left(-\frac{b \cdot y}{2}\right) \quad (2.78)$$

En minimisant l'énergie du système, le paramètre variationnel b s'écrit donc :

$$b(\Psi_s) = \left[\frac{12}{\epsilon_{Si} \hbar^2} m^* q^2 \left(\frac{11 N_{inv}(\Psi_s)}{32} + N_{dep}(\Psi_s) \right) \right]^{1/3} \quad (2.79)$$

Dans cette expression, le paramètre b est donc explicitement relié au potentiel de surface, puisque les charges d'inversion N_{inv} et de déplétion N_{dep} sont des fonctions explicites de Ψ_s sous l'hypothèse de la feuille de charges.

La formulation complète des densités de charge utilisés dans le calcul du paramètres b , prenant en compte l'effet de polydéplétion dans le matériau de grille, s'écrit :

$$N_{dep}(\Psi_s) = \gamma \frac{C_{OX}}{q} \sqrt{\Psi_s + \phi_T \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right]} \quad (2.80)$$

$$N_{inv}(\Psi_s) = \frac{C_{OX}}{q} (V_{GB} - V_{FB} - \Psi_s - \Psi_P) - \gamma \frac{C_{OX}}{q} \sqrt{\Psi_s + \phi_T \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right]} \quad (2.81)$$

$$\text{avec } \Psi_P = \left(\sqrt{V_{GB} - V_{FB} - \Psi_s + \gamma_P^2/4} - \gamma_P/2 \right)^2.$$

Le paramètre b étant désormais clairement exprimé en fonction du potentiel de surface, on peut déterminer le premier niveau d'énergie des vallées de plus basse énergie :

$$E(\Psi_s) = \frac{3\hbar^2 b(\Psi_s)^2}{8m^*} \quad (2.82)$$

En considérant que la majorité des porteurs libres sont situés sur le premier niveau d'énergie des vallées de plus basse énergie, la statistique de Fermi-Dirac permet donc de calculer la densité de charges d'inversion, en fonction du potentiel de surface :

$$N_{inv}^{QM}(\Psi_s) = \frac{2m_i^*}{\pi \hbar^2} kT \ln \left[1 + \exp\left(\frac{E_{F_n}(\Psi_s) - E(\Psi_s)}{kT}\right) \right] \quad (2.83)$$

L'énergie de la première sous bande d'énergie $E(\Psi_s)$ est référencée par rapport au bas de la bande de conduction à l'interface. Dans ces conditions, le quasi niveau de Fermi s'exprime par :

$$E_{F_n}(\Psi_s) = -kT \ln\left(\frac{N_A}{n_i}\right) - \frac{E_g}{2} + \Psi_s - V \quad (2.84)$$

L'expression (2.83) est donc totalement explicite du potentiel de surface. On peut donc définir la relation implicite quantique en égalant les expressions (2.83) et (2.81) :

$$\frac{2m_t^*}{\pi\hbar^2} kT \ln \left[1 + \exp \left(\frac{E_{F_n}(\Psi_S) - E(\Psi_S)}{kT} \right) \right] = \frac{C_{OX}}{q} \left\{ (V_{GB} - V_{FB} - \Psi_S - \Psi_P) - \gamma \sqrt{\Psi_S + \phi_T \left[\exp \left(-\frac{\Psi_S}{\phi_T} \right) - 1 \right]} \right\} \quad (2.85)$$

Dans le but d'alléger les calculs, nous utiliserons, dans ce manuscrit, une forme compacte pour l'équation (2.85), qui peut être écrite sous la forme :

$$F(V_{GB} - V_{FB}, \Psi_S) = 0 \quad (2.86)$$

La figure 2.18 présente les résultats obtenus avec ce modèle, pour diverses épaisseurs d'oxydes et divers niveaux de dopages. On constate un très bon accord avec les résultats obtenus avec le solveur auto cohérent Poisson Schrödinger QMCV. L'avantage majeur avec cette approche est la réduction significative des temps de calculs. Ils sont réduits d'un facteur 150, faisant de ce modèle un outil idéal pour la caractérisation. Compte tenu des faibles temps de calculs, il peut être aisément implémenté dans une boucle d'optimisation, pour un ajustement optimal des paramètres [CHOI'00].

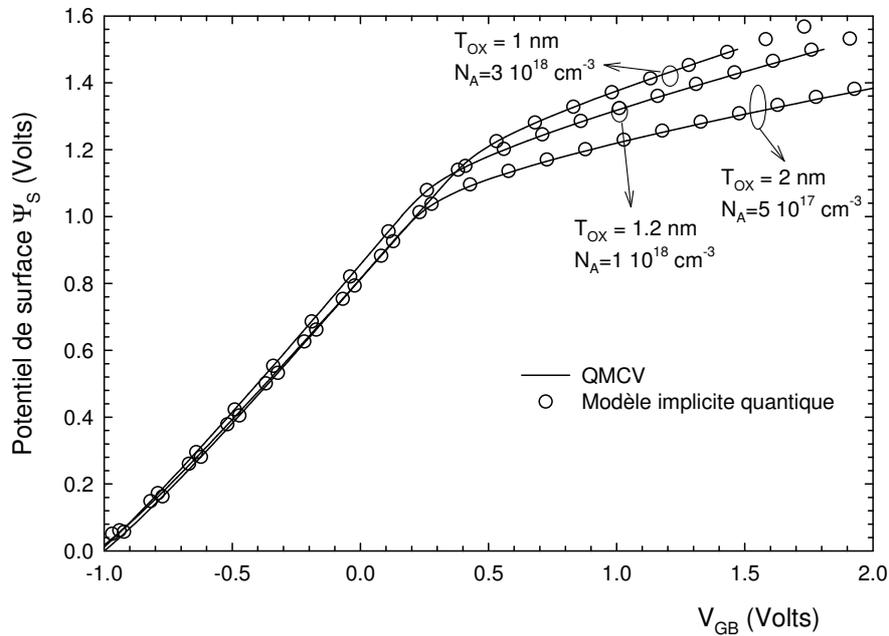


Figure 2.18 : Potentiel de surface calculé par QMCV (lignes) et le modèle implicite quantique (symboles), pour des structures nMOS avec plusieurs jeux de paramètres technologiques. Le phénomène de polydépétion n'a pas été pris en compte.

La modélisation de la dérivée du potentiel de surface doit être également très précise. Tous les calculs de capacités de la structure MOS font intervenir la dérivée du potentiel de surface. Dans le simulateur QMCV, la capacité totale de la structure MOS est obtenue en calculant dans un premier temps la charge totale pour chaque tension de grille, puis en dérivant numériquement cette charge par rapport à la tension de grille⁹. Cette technique souffre d'un gros défaut : il est nécessaire de calculer la charge pour un grand nombre de points de polarisation avec un faible intervalle entre les points pour obtenir une précision suffisante sur la capacité. En d'autres termes, il faut calculer plus de points que nécessaire, induisant donc des temps de calculs trop longs. L'approche analytique (2.86) permet de contourner ce problème, et nous proposons donc une méthode de calcul permettant d'évaluer précisément la capacité de façon explicite à partir de la connaissance du potentiel de surface au point de polarisation considéré.

Nous allons nous intéresser à la capacité totale d'une structure MOS, habituellement notée C_{gg} ¹⁰, définie simplement par la relation :

$$C_{gg} = \frac{\partial Q_g}{\partial V_{GB}} = C_{ox} \left(1 - \frac{\partial \Psi_s}{\partial V_{GB}} - \frac{\partial \Psi_p}{\partial V_{GB}} \right) \quad (2.87)$$

La dérivée de la chute de potentiel dans le polysilicium par rapport à la tension de grille se calcule aisément en utilisant comme définition de Ψ_p l'équation (2.19). La capacité C_{gg} , en tenant compte de la polydéplétion, s'exprime donc suivant la relation valable pour $V_{GB} - V_{FB} > 0$:

$$C_{gg} = C_{ox} \left\{ 1 - \frac{\partial \Psi_s}{\partial V_{GB}} - \frac{\left(\sqrt{4(V_{GB} - V_{FB} - \Psi_s) + \gamma_p^2} - \gamma_p \right) \left(1 - \frac{\partial \Psi_s}{\partial V_{GB}} \right)}{\sqrt{4(V_{GB} - V_{FB} - \Psi_s) + \gamma_p^2}} \right\} \quad (2.88)$$

La connaissance de la dérivée du potentiel de surface est donc essentielle pour déterminer la capacité pour une tension de grille donnée.

Une petite fluctuation de la tension de grille $\widetilde{V_{GB}}$ autour de la valeur moyenne V_{GB_0} induit une petite fluctuation du potentiel de surface $\widetilde{\Psi_s}$ autour de la valeur moyenne

⁹ Cette opération est réalisée simplement en utilisant la dérivée à gauche d'ordre 1 : $C = \frac{Q_n - Q_{n-1}}{V_{GB_n} - V_{GB_{n-1}}}$

¹⁰ Cette notation est liée à une modélisation de la structure par les paramètres Y, détaillée dans l'ouvrage de référence [TSIVIDIS'99]

Ψ_{s_0} . Dans ces conditions, la fonction implicite (2.86) reste valable. On peut développer au premier ordre la fonction F , de sorte que l'on obtienne la relation suivante :

$$F(V_{GB0} + \widetilde{V}_{GB}, \Psi_{s_0} + \widetilde{\Psi}_s) = F(V_{GB0}, \Psi_{s_0}) + \frac{\partial F}{\partial V_{GB}} \widetilde{V}_{GB} + \frac{\partial F}{\partial \Psi_s} \widetilde{\Psi}_s = 0 \quad (2.89)$$

Compte tenu que $F(V_{GB0}, \Psi_{s_0}) = 0$, on obtient donc une relation donnant la dérivée du potentiel de surface par rapport à la tension de grille en fonction des dérivées partielles de la fonction F :

$$\frac{\partial \Psi_s}{\partial V_{GB}} = \frac{\widetilde{\Psi}_s}{\widetilde{V}_{GB}} = - \frac{\partial F / \partial V_{GB}}{\partial F / \partial \Psi_s} \quad (2.90)$$

Les dérivées partielles de la fonction F sont des fonctions explicites de la tension de grille et du potentiel de surface. La connaissance du potentiel de surface pour une tension de grille permet donc de calculer la capacité pour cette même tension de grille, sans pour autant avoir besoin de connaître le potentiel de surface pour des tensions de grille voisines. La simulation de référence est donnée par QMCV, où 170 points ont été calculés pour des tensions de grille variant de V_{FB} à 1.6 Volts. Ce nombre élevé de points permet d'obtenir la valeur de capacité de la structure, puisqu'une augmentation du nombre de points ne change pas le résultat. En d'autres termes, le calcul de la dérivée de la charge ne change plus à partir de 170 points simulés. Les résultats obtenus par notre modèle avec le même nombre de points donnent des résultats similaires, que la capacité soit calculée suivant l'équation (2.90) ou tout simplement par une différentiation de la charge totale comme le fait QMCV. Ces résultats sont présentés sur la figure 2.19a. L'écart entre notre modèle et QMCV pour des tensions de grilles proche de la tension de bandes plates sera discuté dans la partie traitant du régime d'accumulation.

La figure 2.19b présente l'intérêt de la formulation (2.90) pour le calcul de la dérivée. En effet, il n'est pas nécessaire de simuler autant de points pour caractériser une structure MOS. Nous avons reporté sur la figure 2.19b la courbe de référence QMCV de la figure 2.19a (170 points), la capacité calculée sur 20 points avec QMCV, ainsi que notre modèle avec 170 et 20 points. A la vue de tous ces résultats, il est clair que la valeur de capacité mesurée dépend fortement du nombre de points calculés quand on différencie la charge, alors que la formulation (2.90) donne un résultat indépendant du nombre de points choisis.

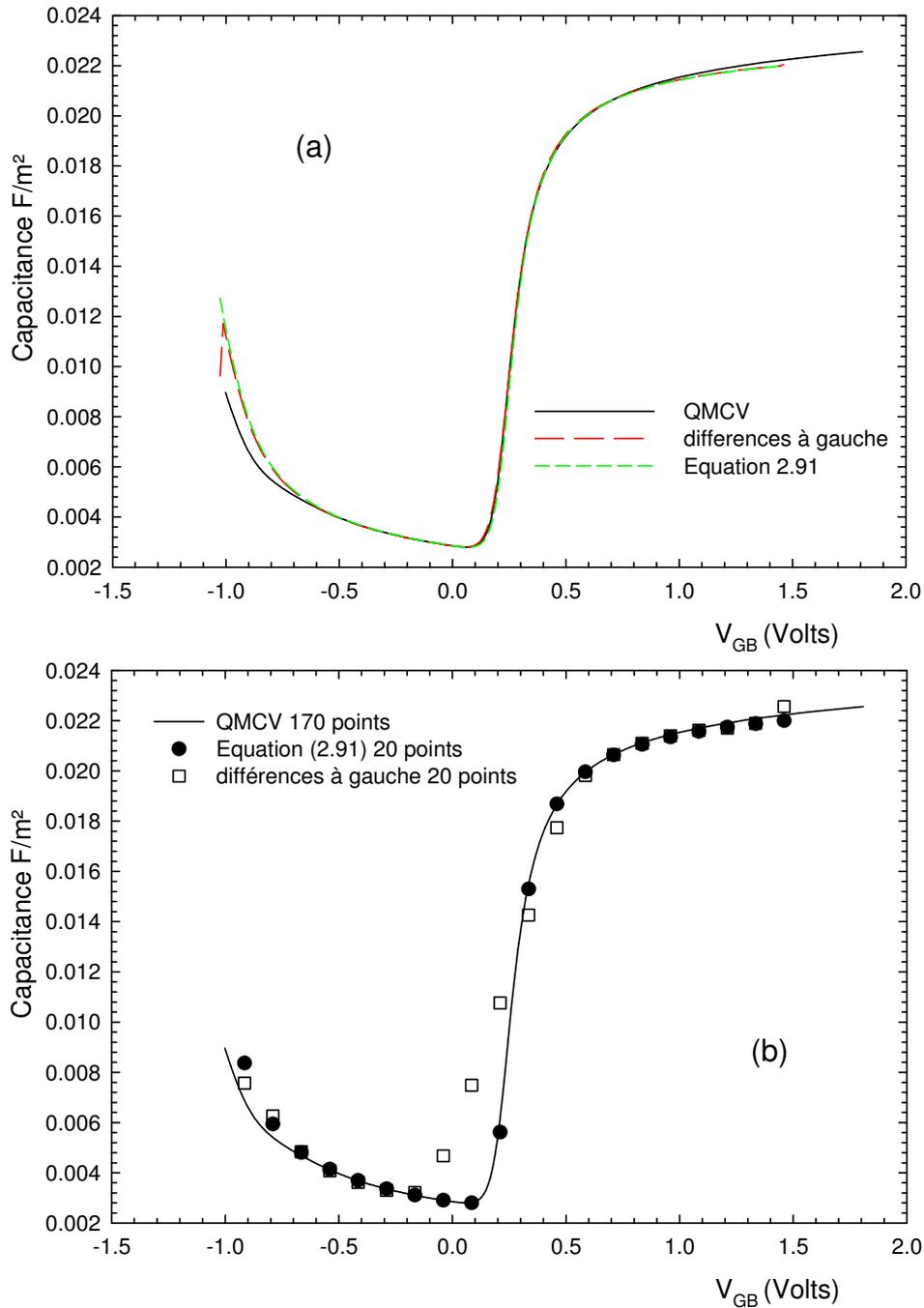


Figure 2.19 : Capacité C_{gg} d'une structure nMOS avec une épaisseur d'oxyde de 1.2 nm, un substrat de type p dopé à 10^{18} cm^{-3} , et une grille polysilicium de type n dopé à $5 \cdot 10^{19} \text{ cm}^{-3}$. Le phénomène de polydépétion n'a pas été pris en compte. La courbe notée « différences à gauche » réfère au calcul de la capacité via la différentiation de la charge totale par rapport à la tension de grille. Le nombre de points pour toutes les simulations de figure (a) est 170.

2.6.4.4 Modèle implicite quantique basé sur l'approche variationnelle : Régime d'accumulation

Le traitement quantique de la couche d'accumulation est nettement plus compliqué que le traitement quantique de la couche d'inversion, en raison, notamment, de la complexité de la structure de la bande de valence. Dans le cas d'une structure nMOS, la couche d'accumulation est constituée de trous.

Dans un premier temps, nous allons nous intéresser aux résultats de simulation auto cohérente en régime de forte accumulation. La figure 2.20 donne les densités pour les deux premières sous bandes des trous légers et des trous lourds. La structure est de type nMOS, avec un substrat de type p dopé à 10^{18} cm^{-3} , une épaisseur d'oxyde de 1.2 nm, et une grille polysilicium dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$. Comme expliqué au paragraphe 1.1, les phénomènes de polydépuration sont négligeables en régime d'accumulation. La tension de grille est de -1.4 volts, de façon à être dans un régime de forte accumulation.

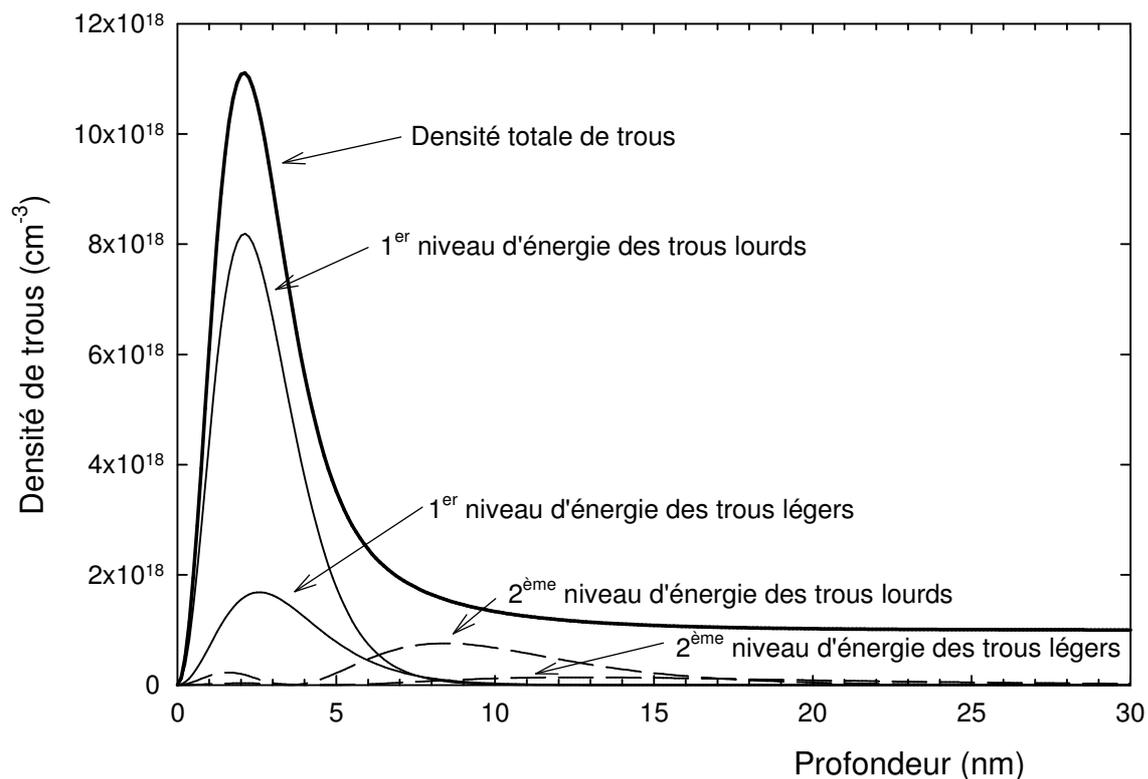


Figure 2.20 : Résultat d'un calcul auto-cohérent dans une capacité nMOS à substrat p dopé à 10^{18} cm^{-3} , avec une grille polysilicium dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$ et une épaisseur d'oxyde de 1.2 nm, polarisée en régime d'accumulation avec $V_{GB} = -1.4$ Volts.

En régime d'accumulation, les porteurs libres majoritaires (trous) compensent la charge fixe des accepteurs ionisés dans la zone neutre. Les résultats de simulation reportés sur la figure 2.20 montre en effet que la densité totale de trous tend vers la valeur du dopage, fixée à 10^{18} cm^{-3} dans notre cas. La charge totale des trous libres peut donc se décomposer en une partie compensant la charge fixe des accepteurs ionisés, et une autre partie constituant la charge excédentaire (charge d'accumulation). La figure 2.21 présente la charge d'espace dans le semiconducteur pour la même tension $V_{GB} = -1.4$ Volts, ainsi que la densité de charges contenues sur le premier niveau d'énergie des trous lourds. D'un point de vue strict, plusieurs niveaux d'énergies doivent être pris en compte pour décrire la charge d'espace. La prise en compte des deux premiers niveaux d'énergie des trous lourds et légers permet d'obtenir une bonne approximation. Même si la prise en compte de plusieurs niveaux d'énergies dans une approche variationnelle est possible [ANDO'82], elle reste néanmoins difficile et nécessitera forcément une résolution numérique, notamment dans la minimisation du paramètre variationnel. Nous avons choisi de formuler l'hypothèse que la charge d'accumulation est contenue sur ce premier niveau d'énergie, permettant ainsi une approche variationnelle analytique du problème de l'accumulation.

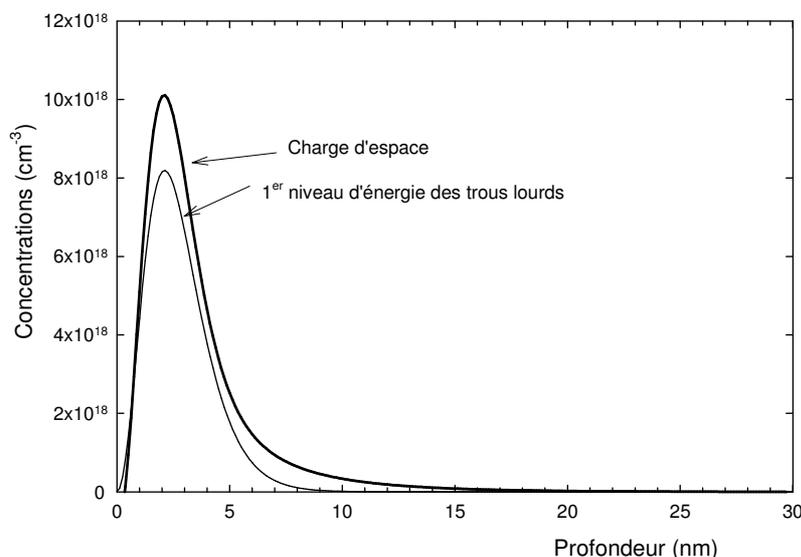


Figure 2.21 : Résultat d'un calcul auto-cohérent dans une capacité nMOS à substrat p dopé à 10^{18} cm^{-3} , avec une grille polysilicium dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$ et une épaisseur d'oxyde de 1.2 nm, polarisée en régime d'accumulation avec $V_{GB} = -1.4$ Volts.

Avec cette hypothèse, la formulation de l'approche variationnelle en accumulation est similaire au raisonnement développé au paragraphe 2.6.3.2. L'énergie potentielle des trous accumulés à la surface est calculée en intégrant l'équation de Poisson où seul les trous du premier niveau des trous lourds sont pris en compte :

$$\frac{\partial^2 \phi}{\partial y^2} = -\frac{q}{\epsilon_{Si}} P_{acc}(y) \quad (2.91)$$

où P_{acc} représente la densité des trous lourds du premier niveau d'énergie.

Les calculs mènent à des résultats similaires à ceux établis précédemment. La charge d'accumulation se substitue à la charge d'inversion, et la charge de déplétion est supprimée. Finalement, le paramètre variationnel s'écrit sous la forme :

$$b(\Psi_s) = \left[\frac{12}{\epsilon_{Si} \hbar^2} m^* q^2 \left(\frac{11 N_{acc}(\Psi_s)}{32} \right) \right]^{1/3} \quad (2.92)$$

où m^* représente la masse effective des trous lourds ($m^* = 0.291 m_0$).

La densité de charges d'accumulation N_{acc} est simplement formulée en utilisant le théorème de Gauss et s'exprime donc par :

$$N_{acc}(\Psi_s) = -\frac{C_{ox}}{q} (V_{GB} - V_{FB} - \Psi_s) \quad (2.93)$$

Il faut noter que dans ce régime de fonctionnement, la tension V_{GB} et le potentiel de surface Ψ_s sont négatifs.

L'étape suivante consiste à calculer l'énergie de la première sous bande $E(\Psi_s)$. Dans un premier temps, nous avons utilisé la formulation (2.82), et les résultats obtenus au final n'étaient pas satisfaisants. En fait, l'équation (2.82) est une forme simplifiée qui est d'autant plus valable que l'extension spatiale de la fonction d'onde est faible. Dans le cas d'une couche d'accumulation, l'extension spatiale de la fonction d'onde est nettement plus grande que celle d'électrons d'une couche d'inversion. De ce fait, nous avons utilisé l'expression exacte pour calculer l'énergie de la première sous bande des trous lourds [STERN'72]:

$$E(\Psi_s) = \left(\frac{3}{2} \right)^{5/3} \left(\frac{q^2 \hbar}{\sqrt{m^*} \epsilon_{Si}} \right)^{2/3} \frac{\left(\frac{55}{96} N_{acc} \right)}{\left(\frac{11}{96} N_{acc} \right)^{1/3}} \quad (2.94)$$

Nous pouvons donc calculer la population de cette sous bande en utilisant la statistique de Fermi-Dirac :

$$N_{acc}^{QM}(\Psi_s) = \frac{m_{hh}^*}{\pi \hbar^2} kT \ln \left[1 + \exp \left(\frac{E_{F_p}(\Psi_s) - E(\Psi_s)}{kT} \right) \right] \quad (2.95)$$

où m_{hh}^* est la masse de la densité d'état des trous lourds ($m_{hh}^* = 0.645 m_0$).

Au final, en égalant la charge d'accumulation calculée de façon quantique à la charge d'accumulation calculée par le théorème de Gauss, on obtient la relation implicite valable pour le régime d'accumulation suivante :

$$\frac{m_{hh}^*}{\pi\hbar^2} kT \ln \left[1 + \exp \left(\frac{E_{F_p}(\Psi_S) - E(\Psi_S)}{kT} \right) \right] = -\frac{C_{ox}}{q} (V_{GB} - V_{FB} - \Psi_S) \quad (2.96)$$

La Figure 2.22 présente les résultats obtenus avec ce modèle, pour diverses épaisseurs d'oxydes et divers niveaux de dopages. On constate un très bon accord avec les résultats obtenus avec le solveur auto cohérent Poisson Schrödinger QMCV. De la même façon qu'en régime de déplétion et d'inversion, les temps de calculs sont significativement réduits avec cette approche.

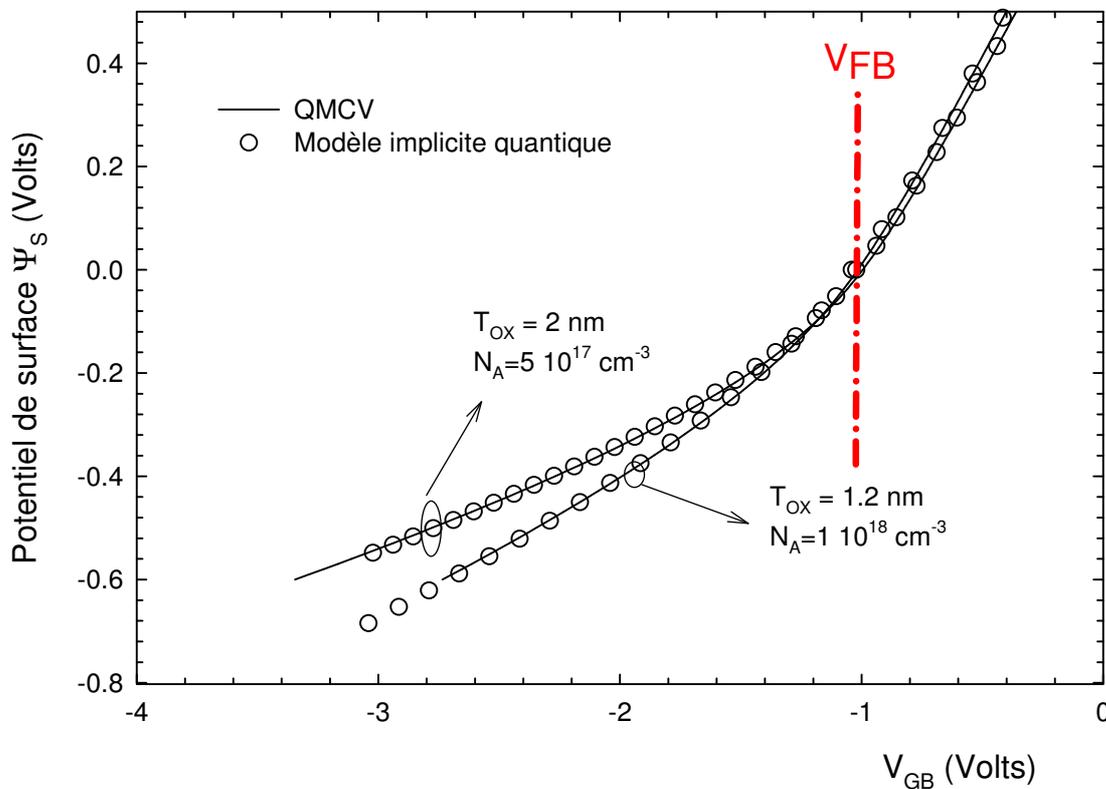


Figure 2.22 : Potentiel de surface calculé par QMCV (lignes) et le modèle implicite quantique (symboles), pour des structures nMOS avec plusieurs jeux de paramètres technologiques.

Toute cette approche repose sur l'hypothèse que la charge d'accumulation est contenue sur le premier niveau d'énergie des trous lourds. A très faible accumulation (typiquement pour $V_{GB} - V_{FB} < 0.1$ Volts pour une épaisseur d'oxyde de 1.2 nm et un dopage substrat de 10^{18} cm^{-3}), cette hypothèse n'est plus vérifiée. La figure 2.23 présente un agrandissement de cette zone. On constate que le potentiel de surface calculé par QMCV ne

passer par zéro pour le régime de bandes plates ($V_{GB}=V_{FB}$). Ceci est dû au fait qu'à très faible polarisation supérieure ou inférieure à la tension de bandes plates, les bandes d'énergies sont très peu discrétisées. De ce fait, il est nécessaire de calculer un grand nombre de sous bandes d'énergie pour se rapprocher de la statistique de Maxwell Boltzmann qui est valide pour ces faibles tensions de grille. De plus, en régime d'accumulation, les fonctions d'ondes associées aux niveaux supérieurs (ceux hors du puits de potentiel) ne doivent pas tendre vers zéro. Une solution consiste à les annuler le plus loin possible de l'interface. Ceci revient évidemment à résoudre l'équation de Schrödinger dans une zone encore plus grande, et donc à augmenter le temps de calcul. La convergence des solveurs Poisson Schrödinger autour de la tension de seuil a été peu traitée dans la littérature. On peut cependant trouver dans [PACELLI'99] une étude détaillée de l'impact de tous les paramètres de simulation (longueur de la zone où est résolue de l'équation de Schrödinger, nombre de niveaux calculés) sur le résultat final.

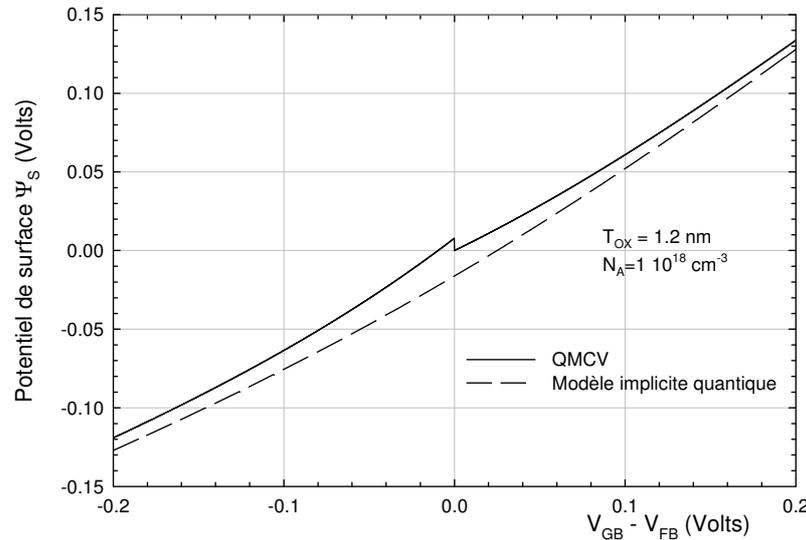


Figure 2.23 : Discontinuité du potentiel de surface à la tension de bandes plates par l'approche variationnelle.

Le modèle implicite converge mal vers la solution car le nombre de niveaux d'énergies pris en compte est insuffisant. Nous avons adopté une approche empirique pour contourner ce problème. Une fonction de lissage faisant passer doucement la charge calculée de façon quantique de zéro pour la tension de bandes plates à la formulation (2.95) permet d'améliorer le calcul du potentiel de surface. La nouvelle formulation de l'équation (2.95) est donc la suivante :

$$N_{acc}^{QM}(\Psi_s) = \frac{m_{hh}^*}{\pi \hbar^2} kT \ln \left[1 + \exp \left(\frac{E_{F_p}(\Psi_s) - E(\Psi_s)}{kT} \right) \right] \times \left(1 - \exp \left(-\frac{V_{GB}}{\lambda} \right) \right) \quad (2.97)$$

où λ est une constante empirique fixée à 0.04. Cette valeur est fixe quels que soient les valeurs des paramètres technologiques.

La figure 2.24 présente les résultats obtenus avec le modèle complet, permettant donc de calculer le potentiel de surface dans tous les régimes de fonctionnement, et tenant compte des effets quantiques et de la polydéplétion.

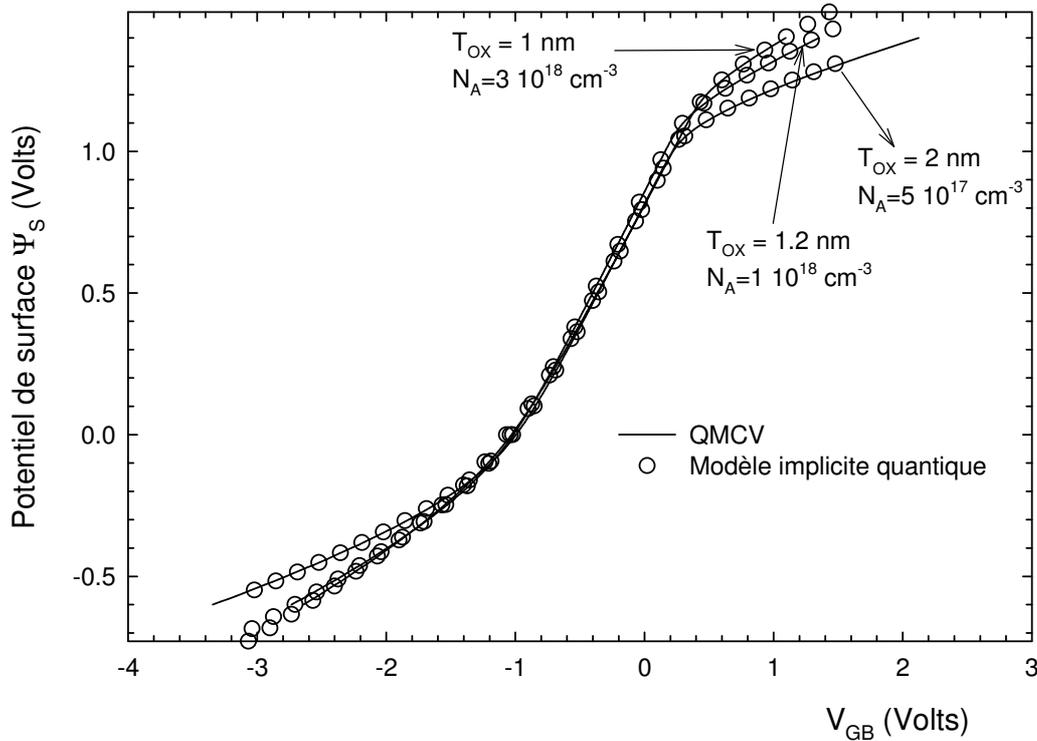


Figure 2.24 : Potentiel de surface pour des structures nMOS avec plusieurs jeux de paramètres technologiques, du régime d'accumulation à la forte inversion.

L'utilisation de cette fonction de lissage ne parvient cependant pas à assurer une bonne continuité du potentiel de surface au voisinage de la tension de bandes plates. Une forte discontinuité apparaît lors du calcul de la dérivée du potentiel de surface pour calculer la capacité de la structure. Ce problème est récurrent dans de nombreux simulateurs ou modèle quantiques de capacité MOS [YU'00, LIME'03]. La solution consiste souvent à l'emploi de fonction de lissage assurant la continuité du potentiel de surface lors du passage du régime d'accumulation au régime de déplétion. Dans le but d'obtenir une caractéristique $C-V$ quantique utilisable pour la caractérisation, nous avons utilisé une fonction réalisant une extrapolation cubique entre le potentiel de surface calculé pour $V_{GB}-V_{FB}=-0.1$ Volts et $V_{GB}-V_{FB}=0.1$ Volts. La charge totale calculée à partir de ce potentiel de surface donne des résultats très satisfaisants dans tous les régimes de fonctionnement. Le résultat est présenté sur la figure 2.25 pour plusieurs épaisseurs d'oxyde et niveaux de dopages substrat. Pour toutes ces

simulations, le polysilicium est de type n, dopé à $5 \cdot 10^{19} \text{ cm}^{-3}$, et le phénomène de polydéplétion n'a pas été prise en compte (dans le but de bien séparer l'impact de chaque effet intervenant dans la capacité de la structure).

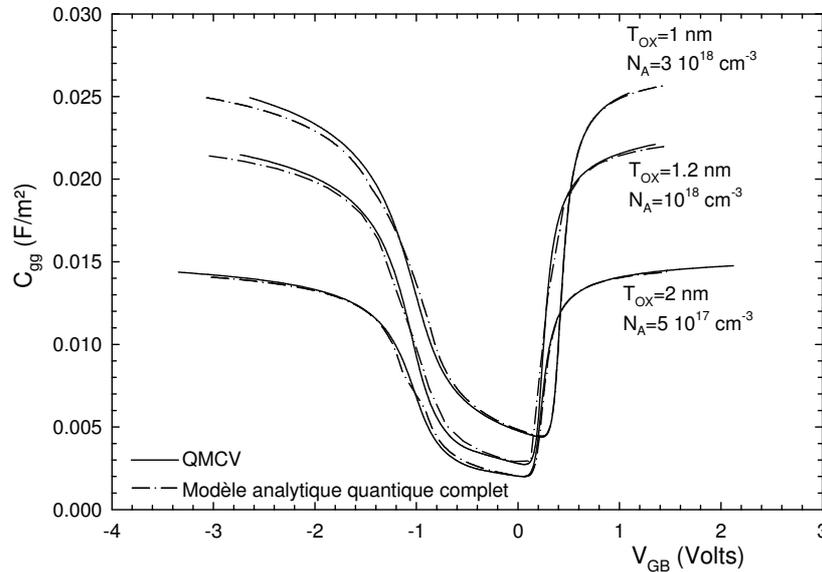


Figure 2.25 : Capacité C_{gg} pour des structures nMOS avec plusieurs jeux de paramètres technologiques, du régime d'accumulation à la forte inversion.

La figure 2.26 présente les résultats du modèle complet pour des structures nMOS en tenant compte du phénomène de polydéplétion. La structure est un transistor nMOS, avec une épaisseur d'oxyde de 1.2 nm et un substrat de type p dopé à 10^{18} cm^{-3} . Le dopage du polysilicium varie de 10^{19} cm^{-3} à 10^{20} cm^{-3} . Un très bon accord est obtenu avec le résultat du calcul auto cohérent.

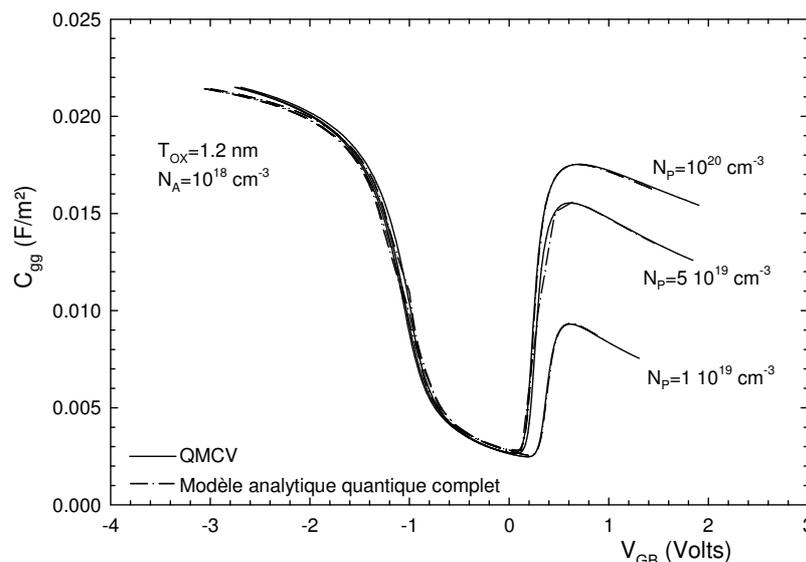


Figure 2.26 : Capacité C_{gg} d'une structure nMOS avec une épaisseur d'oxyde de 1.2 nm, un substrat de type p dopé à 10^{18} cm^{-3} , et une grille polysilicium de type n dopée de 10^{19} cm^{-3} à 10^{20} cm^{-3} .

2.6.4.5 Modèle implicite quantique complet pour les structures pMOS

Le modèle complet décrit précédemment valable pour les structures nMOS peut être aisément étendu aux structures pMOS. En régime de déplétion, la charge est constituée de donneurs ionisés, et l'équation (2.76) reste valable. La charge d'inversion, constituée de trous libres, est supposée être majoritairement sur le premier niveau d'énergie des trous lourds. Sous cette hypothèse, l'équation (2.83) se reformule sous la forme suivante :

$$N_{inv}^{QM}(\Psi_s) = \frac{m_{hh}^*}{\pi\hbar^2} kT \ln \left[1 + \exp \left(\frac{E_{F_p}(\Psi_s) - E(\Psi_s)}{kT} \right) \right] \quad (2.98)$$

On obtient donc une équation implicite en suivant le même processus que celui décrit au paragraphe 2.6.4.3.

En régime d'accumulation, la charge est constituée d'électrons majoritaires, et l'essentiel de cette charge est contenue sur le premier niveau d'énergie des vallées normales. L'équation (2.95) s'écrit donc sous la forme :

$$N_{acc}^{QM}(\Psi_s) = \frac{2m_t^*}{\pi\hbar^2} kT \ln \left[1 + \exp \left(\frac{E_{F_n}(\Psi_s) - E(\Psi_s)}{kT} \right) \right] \quad (2.99)$$

Pour les structures pMOS, la discontinuité du potentiel de surface par une approche variationnelle au niveau de la tension de bandes plates est moins importante que pour les structures nMOS. En régime de faible accumulation d'électrons, l'hypothèse que la charge est contenue sur un seul niveau d'énergie est mieux vérifiée que pour une couche d'accumulation de trous. Enfin, le phénomène de polydéplétion est modélisé de la même façon que pour les structures nMOS, suivant l'équation (2.19).

Les figures 2.27 et 2.28 présentent le potentiel de surface ainsi que la capacité totale de structures pMOS pour plusieurs jeux de paramètres technologiques. La comparaison avec les résultats de calculs auto cohérents montre la très grande précision obtenue avec l'approche variationnelle sur les structures pMOS.

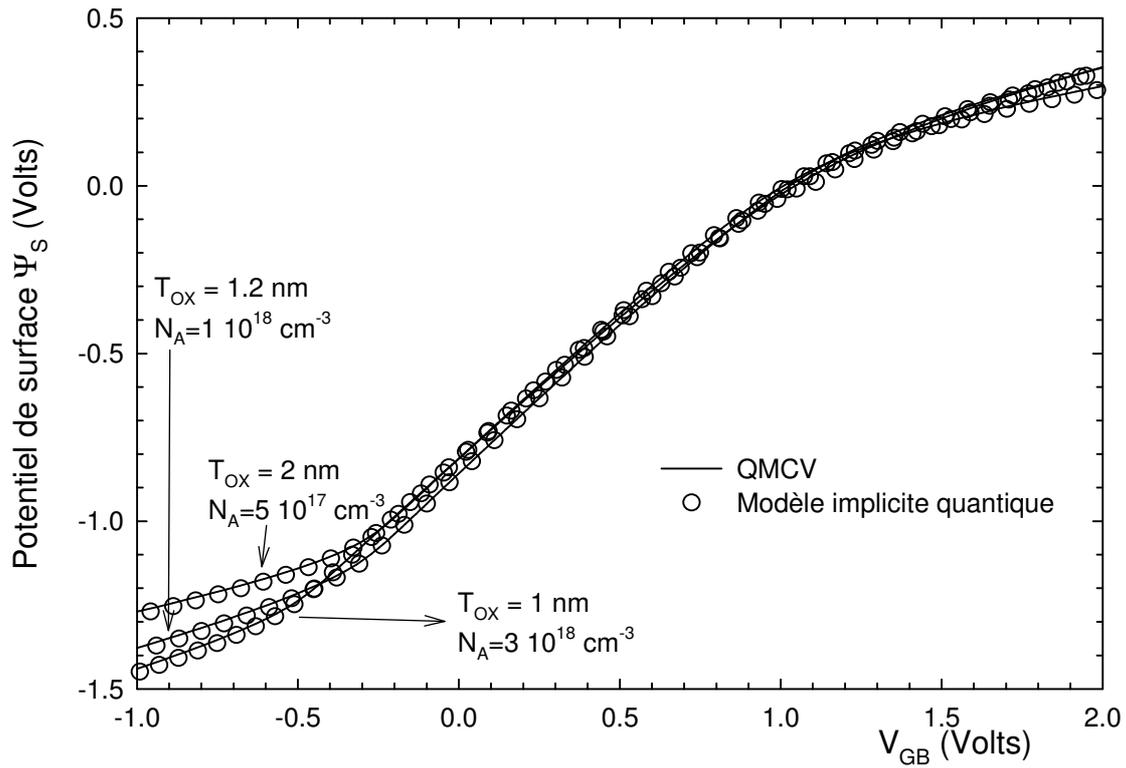


Figure 2.27 : Potentiel de surface pour des structures pMOS avec plusieurs jeux de paramètres technologiques, du régime d'accumulation à la forte inversion.

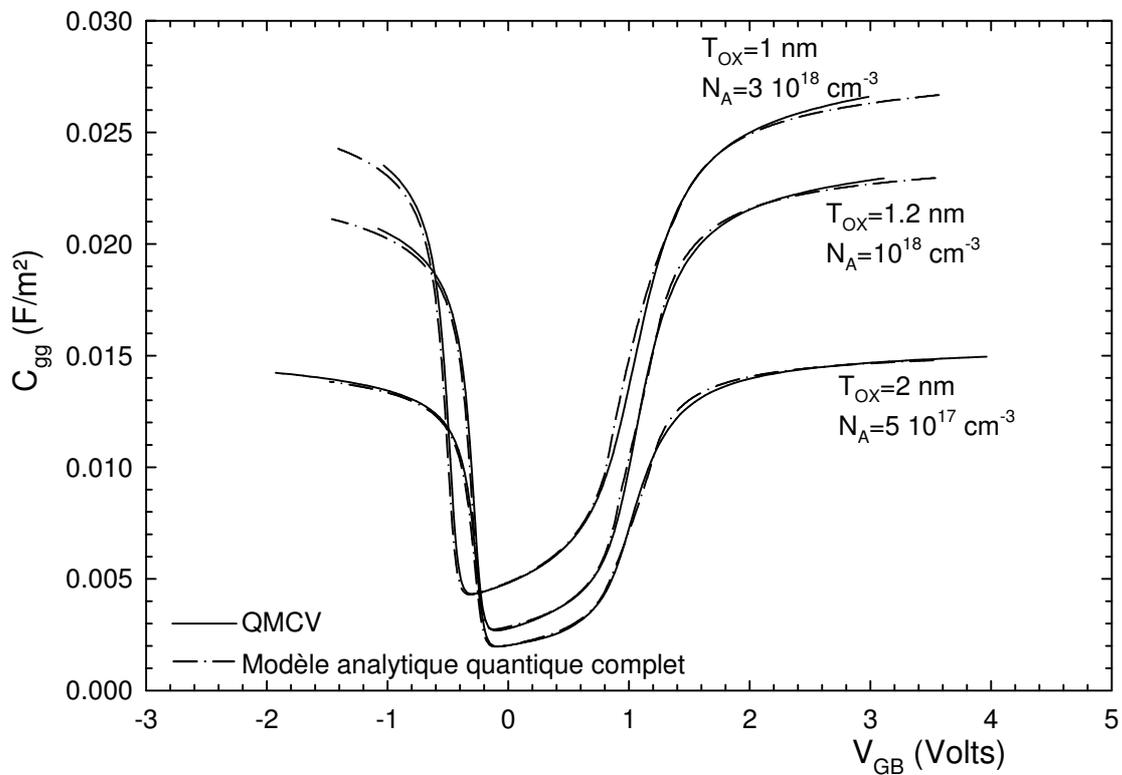


Figure 2.28 : Capacité C_{gg} pour des structures pMOS avec plusieurs jeux de paramètres technologiques, du régime d'accumulation à la forte inversion.

2.6.5 Modélisation explicite quantique du potentiel de surface en régime de déplétion et d'inversion

2.6.5.1 Modèle Compact MM11

Les excellents résultats obtenus avec l'approche variationnelle en terme de temps de calculs par rapport au calcul auto cohérent permettent d'utiliser ce modèle en tant qu'outil de caractérisation. La connaissance du potentiel de surface permet d'accéder à un maximum d'information sur la structure MOS. Nous verrons par la suite que les modèles de transistors MOS basés sur le potentiel de surface sont également très intéressants, puisqu'ils utilisent au final moins de paramètres que les modèles basés sur la tension de seuil, et gardent de ce fait tout leur sens physique.

Cependant, un modèle compact ne peut pas être basé sur un calcul numérique itératif. Il est impératif que toutes les équations soient explicites. Dans la famille des modèles compacts basés sur le potentiel de surface, le modèle MM11 (Mos Model 11) [LANGEVELDE'03] est certainement le plus abouti. Il est implanté dans des simulateurs de type Spice non commerciaux (Spice3f4) et commerciaux (Eldo, Spectre, ADS). Les équations explicites utilisées dans MM11 sont basées sur celles développées au paragraphe 2.5. Les effets quantiques sont pris en compte par l'utilisation d'une épaisseur d'oxyde effective :

$$T_{ox_{eff}} = T_{ox} + \frac{\epsilon_{ox}}{\epsilon_{Si}} \Delta y \quad (2.100)$$

Le terme Δy est calculé à partir de l'élargissement du gap à l'interface, modélisé suivant l'approche variationnelle :

$$\Delta E_g = \frac{3}{2} \left(\frac{3q\hbar}{2\epsilon_{Si}\sqrt{m^*}} \right)^{2/3} \frac{Q_{dep} + \frac{55}{96}Q_{inv}}{\left(Q_{dep} + \frac{11}{32}Q_{inv} \right)^{1/3}} \approx q \frac{3}{5} QM (\epsilon_{Si} F_{eff})^{2/3} \quad (2.101)$$

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 normal à l'interface Si/SiO₂, donné par :

$$F_{eff} = - \frac{Q_{dep} + Q_{inv}/3}{\epsilon_{Si}} \quad (2.102)$$

Dans le modèle MM11, le terme prenant en compte la modification de la distribution des porteurs libres s'écrit donc, conformément aux résultats établis pour l'approximation du puits triangulaire¹¹ :

$$\Delta y \approx \frac{2}{3} \frac{\Delta E_g}{q F_{eff}} \quad (2.103)$$

En définissant l'augmentation effective de l'épaisseur d'oxyde par $\Delta T_{ox} = T_{ox,eff} - T_{ox}$, on obtient l'expression suivante :

$$\frac{\Delta T_{ox}}{T_{ox}} = \frac{2}{3} \frac{\Delta E_g}{q \epsilon_{Si} F_{eff}} \frac{C_{ox}}{C_{ox}} = QM_{tox} \left[\frac{C_{ox}}{\epsilon_{Si} F_{eff}} \right]^{1/3} \quad (2.104)$$

Cette approche est donc très simple et se limite à la modification de la capacité d'oxyde effective dépendante de la polarisation de grille. Puisqu'elle est basée sur un élargissement du gap, la modélisation des effets quantiques dans MM11 souffre des problèmes évoqués au paragraphe 2.6.4.2. Un ajustement du paramètre QM_{tox} est nécessaire pour simuler au mieux le potentiel de surface.

2.6.5.2 Modèle explicite basé sur l'élargissement du gap

Une approche plus physique a été proposée par Prégaldiny [PREGALDINY'04]. Elle est également basée sur un élargissement du gap, formulé à partir des résultats de l'approche variationnelle.

Le potentiel de surface quantique étant relativement proche du potentiel de surface classique, il s'exprime simplement par :

$$\Psi_s^{QM} = \Psi_s + \delta\Psi_s \quad (2.105)$$

Dans cette formulation, l'élargissement du gap se répercute donc directement sur une augmentation du potentiel de surface. L'augmentation du potentiel de surface due aux effets quantiques est donnée par les relations suivantes :

¹¹ Il est surprenant que la position du centroïde d'inversion Δy n'est pas été formulée suivant l'approximation variationnelle, où elle s'exprime simplement par $\frac{3}{b}$.

$$\begin{aligned}\delta\Psi_s &= \frac{1}{q} \frac{3}{8} \frac{\hbar^2}{m^*} b(V_{GB}, V)^2 = \frac{1}{q} \frac{3}{8} \frac{\hbar^2}{m^*} \left[\frac{12 m^* q^2 n_{all}(V_{GB}', V)}{\epsilon_{Si} \hbar^2} \right]^{2/3} \\ n_{all}(V_{GB}', V) &= \frac{2C_{ox}}{q} (V_{GB}' - V_{to} - V) \\ V_{GB}' &= \frac{1}{2} \left[V_{GB} + \sqrt{(V_{GB} - V_{to} - V)^2 + 4\epsilon^2} + \sqrt{(V_{to} + V)^2 + 4\epsilon^2} \right]\end{aligned}\quad (2.106)$$

Le paramètre variationnel b a été modifié de façon à pouvoir le calculer explicitement en fonction de V_{GB} et V . Il fait intervenir une fonction empirique n_{all} pour le calcul du terme $\left(N_{dep} + \frac{11}{32} N_{inv}\right)$ de l'équation (2.59). Le paramètre ϵ est un paramètre de lissage fixé par l'auteur à 0.15 et V_{to} est la tension de seuil canal long classique donnée par $V_{FB} + \Phi_B + \gamma\sqrt{\Phi_B}$.

Au final, l'auteur a intégré toutes ces modifications au modèle explicite décrit au paragraphe 2.5. Le modèle explicite complet s'exprime donc par :

$$\begin{aligned}\Psi_s &= f + \phi_T \ln \left\{ \frac{1}{\gamma^2 \phi_T} \left[V_{GB} - V_{FB} - f^{QM} - \frac{\Psi_{S_{wl}} - f^{QM}}{\sqrt{1 + \left[\frac{\Psi_{S_{wl}} - f^{QM}}{4\phi_T} \right]^2}} \right]^2 - \frac{f^{QM}}{\phi_T} + 1 \right\} \\ \text{avec } f^{QM} &= \frac{\phi_B + V + \Psi_{S_{wl}} + \delta\Psi_s}{2} - \frac{1}{2} \sqrt{(\Psi_{S_{wl}} - \phi_B - V - \delta\Psi_s)^2 + 4\delta^2} \\ \text{et } \Psi_{S_{wl}} &= \left[\sqrt{V_{GB} - V_{FB} + \gamma^2/4} - \gamma/2 \right]^2\end{aligned}\quad (2.107)$$

La constante de lissage δ est fixée à 0.02. Les résultats obtenus avec ce modèle sont présentés sur la figure 2.29. L'écart entre ce modèle explicite et les résultats du calcul auto cohérent n'est pas négligeable, et tend à augmenter avec une augmentation du niveau de dopage substrat et une diminution de l'épaisseur d'oxyde. D'après l'auteur, ce modèle ne permet pas de simuler avec précision le potentiel de surface pour des dopages substrat supérieures à 10^{18} cm^{-3} et des épaisseurs d'oxyde inférieures à 1.7 nm. Ces limitations compromettent fortement l'utilisation de ce modèle pour des modèles des nœuds technologiques avancés, tel que le nœud 65 nm [ITRS'03] où le dopage substrat est de l'ordre de $3 \cdot 10^{18} \text{ cm}^{-3}$ et une épaisseur d'oxyde équivalente de l'ordre de 1 nm.

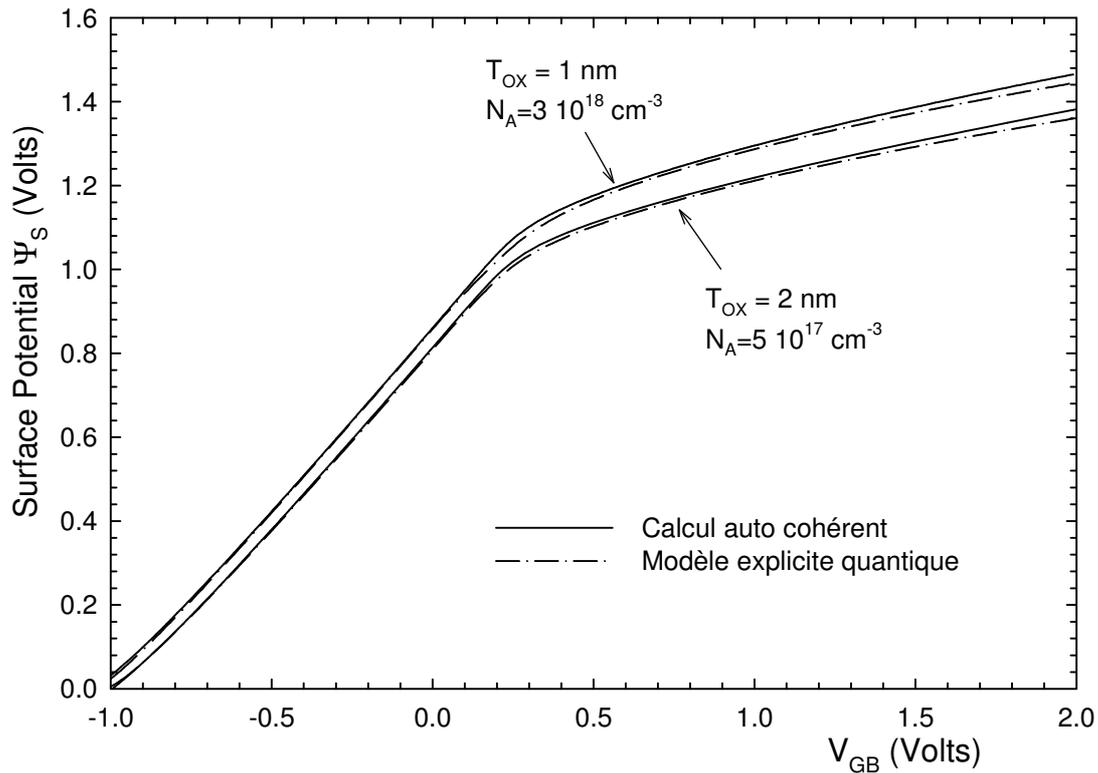


Figure 2.29 : Potentiel de surface issu du modèle explicite quantique [PREGALDINY'04] et d'un calcul auto cohérent, pour des structures nMOS avec un polysilicium de type n dopé à $5 \cdot 10^{19} \text{ cm}^{-3}$. Le phénomène de polydéplétion n'a pas été pris en compte.

2.6.5.3 Modèle quantique explicite basé sur l'approche variationnelle

Nous allons exposer dans ce paragraphe l'approche explicite que nous avons développée pour prendre en compte à la fois les effets quantiques et le phénomène de polydéplétion dans les structures MOS.

Dans le paragraphe 2.6.4.3, la fonction implicite F a été établie grâce aux résultats de l'approche variationnelle. D'une manière générale, on peut constater que le potentiel de surface calculé en prenant en compte les effets quantiques est légèrement plus grand que celui calculé de façon classique pour une même tension de grille. Ce comportement est d'autant plus vrai que la tension de grille est grande. On peut donc poser :

$$\Psi_s = \Psi_{s_0} + \Delta\Psi_s \quad (2.108)$$

Dans ces conditions, on peut aisément exprimer la quantité $\Delta\Psi_s$ en fonction de la fonction F et de Ψ_{s_0} en utilisant un développement en série de Taylor d'ordre 2 :

$$F(\Psi_{s_0} + \Delta\Psi_s) = F(\Psi_{s_0}) + \left. \frac{\partial F}{\partial \Psi_s} \right|_{\Psi_{s_0}} \Delta\Psi_s + \frac{1}{2} \left. \frac{\partial^2 F}{\partial \Psi_s^2} \right|_{\Psi_{s_0}} (\Delta\Psi_s)^2 = 0 \quad (2.109)$$

$$\Delta\Psi_s = \frac{-\left. \frac{\partial F}{\partial \Psi_s} \right|_{\Psi_{s_0}} + \sqrt{\left(\left. \frac{\partial F}{\partial \Psi_s} \right|_{\Psi_{s_0}} \right)^2 - 2F(\Psi_{s_0}) \left. \frac{\partial^2 F}{\partial \Psi_s^2} \right|_{\Psi_{s_0}}}}{\left. \frac{\partial^2 F}{\partial \Psi_s^2} \right|_{\Psi_{s_0}}} \quad (2.110)$$

Cette approche nécessite donc un potentiel Ψ_{s_0} de départ autour duquel on développe la fonction F . Malgré un développement d'ordre 2, le terme $\Delta\Psi_s$ doit rester très faible pour que l'équation (2.109) reste valable. Pour le calcul de Ψ_{s_0} , une première approche a été d'utiliser la solution classique (2.26). La figure 2.30 montre que cette approche ne donne pas de bons résultats. La non convergence de la solution provient simplement du fait que la solution Ψ_{s_0} est trop éloignée de la solution quantique Ψ_s , ou, en d'autres termes, implique $\Delta\Psi_s$ trop grand.

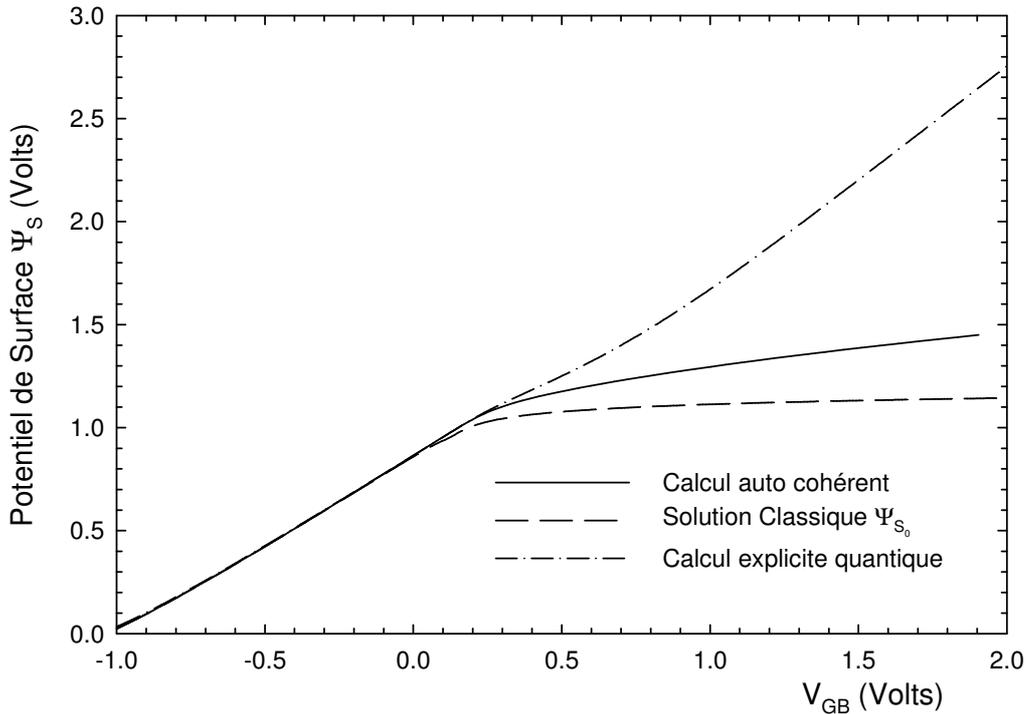


Figure 2.30 : Potentiel de surface calculé suivant l'équation (2.110) diverge si la solution de départ est trop éloignée de la solution finale quantique.

Nous devons donc choisir comme première approximation Ψ_{s_0} une expression analytique explicite, tenant compte approximativement des effets quantiques, pour ne pas

avoir de terme $\Delta\Psi_s$ trop important. Nous avons présenté au paragraphe 2.6.4.2 l'équation implicite permettant de décrire le potentiel de surface quantique basée sur l'élargissement du gap :

$$\begin{aligned} & (V_{GB} - V_{FB} - \Psi_s - \Psi_p)^2 \\ &= \gamma^2 \left[\Psi_s + \phi_T \exp\left(-\frac{\Delta E_g}{2kT}\right) \left(\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right) + \phi_T \exp\left(-\frac{V + \phi_B}{\phi_T}\right) \exp\left(-\frac{\Delta E_g}{2kT}\right) \left(\exp\left(\frac{\Psi_s}{\phi_T}\right) - 1 \right) \right] \end{aligned} \quad (2.111)$$

En simplifiant cette expression par le même processus qu'au paragraphe 2.5.1, on obtient l'expression suivante pour le potentiel de surface Ψ_{s_0} :

$$\Psi_{s_0} = f + \phi_T \ln \left\{ \frac{1}{\gamma^2 \phi_T \exp\left(-\frac{\Delta E_g}{2kT}\right)} \left[V_{GB} - V_{FB} - f - \frac{\Psi_{s_{wi}} - f}{\sqrt{1 + \left[\frac{\Psi_{s_{wi}} - f}{4\phi_T}\right]^2}} \right]^2 - \frac{f}{\phi_T} + 1 \right\} \quad (2.112)$$

Les fonctions f et $\Psi_{s_{wi}}$ restent définies de la même façon qu'au paragraphe 2.5.1. Le terme ΔE_g est défini suivant la formulation de van Dort :

$$\begin{aligned} \Delta E_g &= \frac{13}{9} \beta \left[\frac{\epsilon_{Si}}{4kT} \right]^{1/3} \max(E_S, 0)^{2/3} \\ &\text{avec } E_S = \frac{C_{ox}}{\epsilon_{Si}} (V_{GB} - \phi_B - \gamma \sqrt{\phi_B}) \end{aligned} \quad (2.113)$$

Le potentiel de surface Ψ_{s_0} est donc explicitement défini, puisque ΔE_g est directement exprimé à V_{GB} . Une fois ce potentiel calculé, on calcule le potentiel final quantique à l'aide de l'équation (2.110).

La figure 2.31 présente les résultats obtenus avec ce nouveau modèle quantique explicite. Les résultats issus de calculs auto cohérents y sont également reportés, ainsi que le modèle décrit par l'équation (2.107). Notre modèle offre une très grande précision dans les régimes de déplétion, faible et forte inversion. De plus, la précision par rapport à des résultats issus de calculs auto cohérents est présentée sur la figure 2.32. L'approche que nous proposons est donc nettement plus proche du calcul de référence que le modèle proposé dans [PREGALDINY'04], et ce quelque soit le régime de fonctionnement.

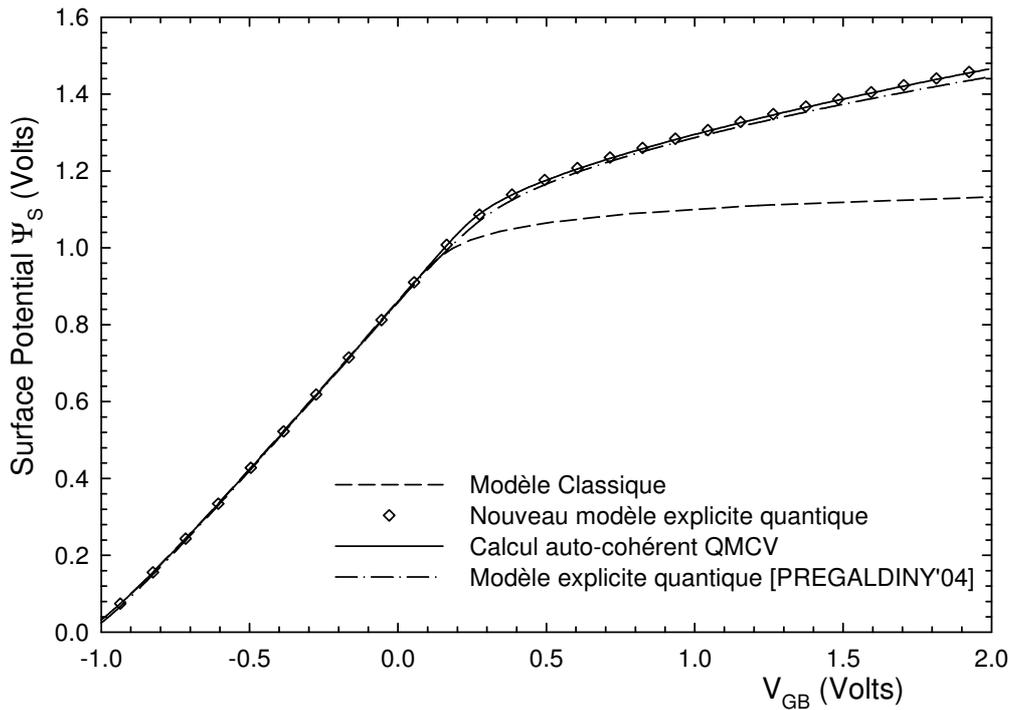


Figure 2.31 : Potentiel de surface pour une structure nMOS avec une épaisseur d'oxyde de 1 nm, un substrat de type p dopé à $3 \cdot 10^{18} \text{ cm}^{-3}$, et une grille polysilicium de type n dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$.

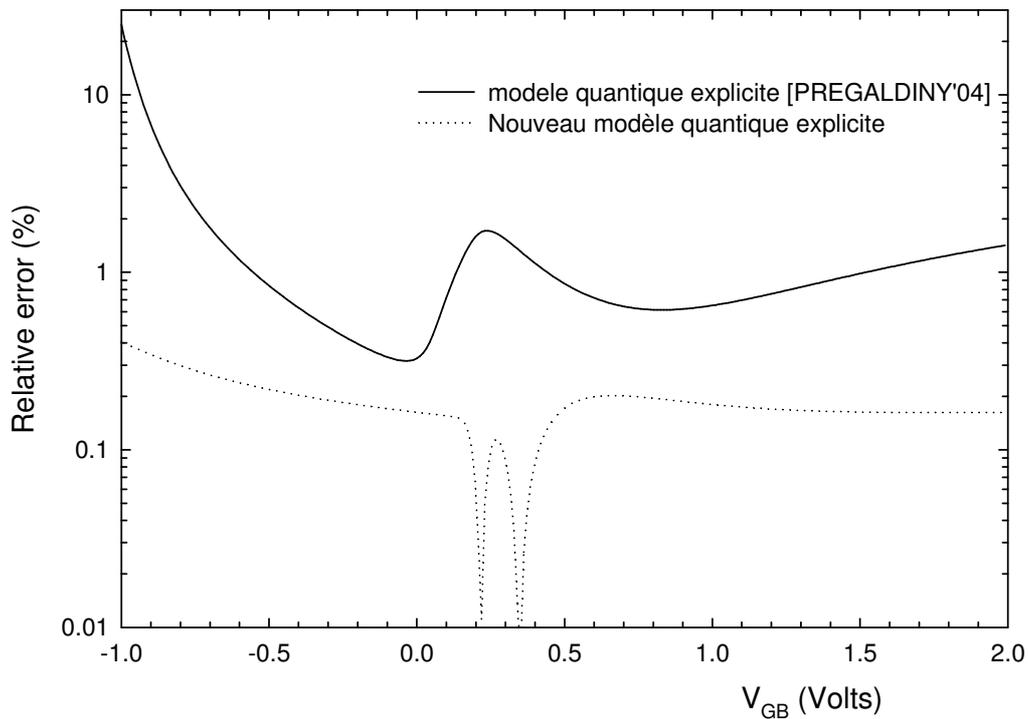


Figure 2.32 : Erreur relative par rapport aux simulations auto cohérente pour notre modèle explicite et le modèle explicite proposé dans [PREGALDINY'04]. Structure nMOS avec une épaisseur d'oxyde de 1 nm, un substrat de type p dopé à $3 \cdot 10^{18} \text{ cm}^{-3}$, et une grille polysilicium de type n dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$.

La figure 2.33 présente les résultats obtenus sur des structures nMOS correspondant à plusieurs technologies. Un très bon accord est observé quel que soit les paramètres technologiques, de la déplétion à la forte inversion.

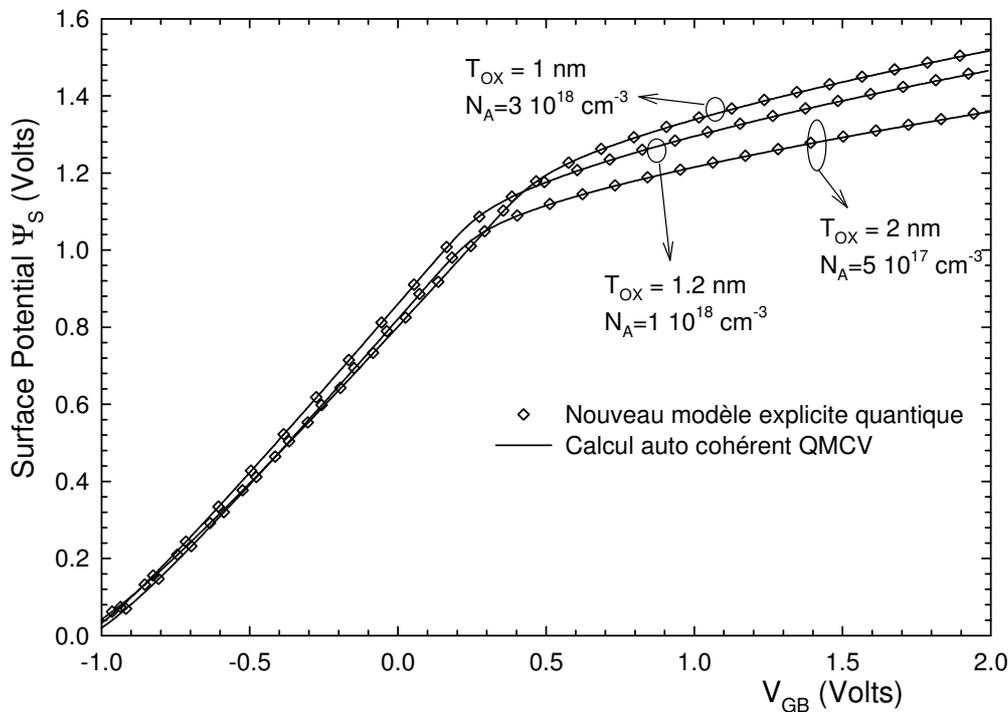


Figure 2.33 : Potentiel de surface calculé par QMCV (lignes) et le modèle explicite quantique (symboles), pour des structures nMOS avec plusieurs jeux de paramètres technologiques. Le phénomène de polydéplétion n'a pas été pris en compte.

En modélisation compacte, le modèle de potentiel de surface doit non seulement être précis en terme de potentiel de surface, mais également en terme de dérivé. Ce critère est essentiel pour le calcul des capacités de la structure, ou plus généralement dans le cadre d'une d'étude petit signal. La figure 2.34 montre l'excellente précision de la dérivée du potentiel de surface comparée à celle obtenue par calcul auto cohérent.

Nous avons donc développé un modèle explicite de potentiel de surface précis tenant compte des effets quantiques et du phénomène de polydéplétion. A la différence des modèles existants, il n'utilise aucune fonction de lissage, et, de ce fait, les paramètres gardent leur sens physique. En particulier, les très bons résultats obtenus avec des paramètres technologiques correspondant à des technologies très avancées montre qu'il peut être utilisé comme base des futurs modèles compacts. D'un point de vue caractérisation, nous verrons par la suite qu'il permet l'extraction de paramètres sur des oxydes ultrafins et sert de base à la modélisation du bruit basse fréquence sur le courant de fuite de la grille.

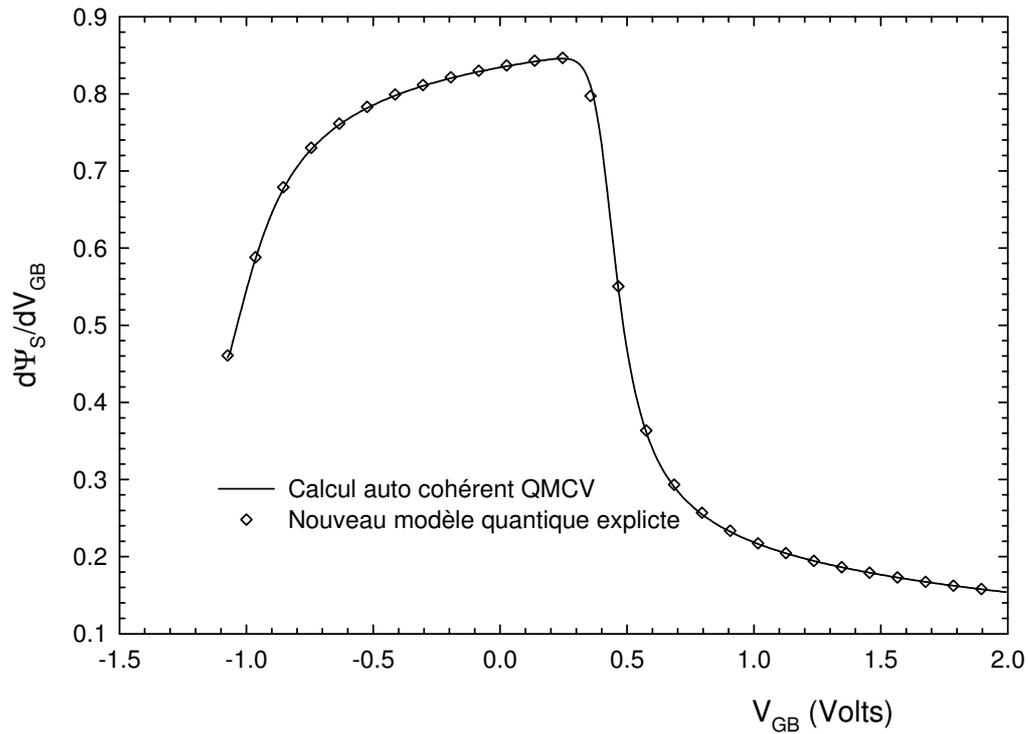


Figure 2.34 : Dérivée du potentiel de surface issu du modèle explicite (symboles) et auto cohérent (ligne) pour une structure nMOS avec une épaisseur d'oxyde de 1.2 nm, un substrat de type p dopé à 10^{18} cm^{-3} , et une grille polysilicium de type n dopée à $5 \cdot 10^{19}$ cm^{-3} . Le phénomène de polydéplétion n'a pas été pris en compte.

2.7 Modélisation du courant de grille

Cette partie est consacrée à l'établissement du modèle classique du courant de grille. Nous allons nous limiter à l'étude du courant tunnel direct qui domine le courant de fuite des oxydes ultraminces d'épaisseur inférieure à 2 nm.

Dans un premier temps, nous allons rappeler la formulation générale du courant tunnel direct, puis nous détaillerons les expressions utilisées dans ce mémoire, et notamment le couplage avec le modèle de potentiel de surface quantique.

2.7.1 Expression générale du courant tunnel direct

Le raisonnement présenté ici est basé sur l'hypothèse que la structure considérée peut être assimilée à l'empilement de trois matériaux cristallins, ne possédant chacun qu'une seule bande d'énergie parabolique. L'étude qualitative des courants de grille directs présentés au premier chapitre fait apparaître quatre mécanismes distincts : l'injection d'électrons ou de trous, depuis le substrat ou le matériau de grille. Par soucis de simplification, nous détaillerons uniquement l'injection d'électrons depuis le substrat vers le matériau de grille, c'est-à-dire dans les régimes de faible et forte inversion pour des structures nMOS.

La densité de courant tunnel est obtenue en pondérant le flux élémentaire d'électrons incident au point (x,z) $d^3\Phi(y=0, x, z, \vec{k})$ par la probabilité $P(x, z, \vec{k})$ de chaque électron, de vecteur d'onde \vec{k} frappant l'interface en (x,z) , de traverser la couche d'oxyde par effet tunnel direct [Cassan'00]:

$$d^3 J_G(y=0, x, z, \vec{k}) = q \cdot d^3\Phi(y=0, x, z, \vec{k}) \cdot P(x, z, \vec{k}) \quad (2.114)$$

Le flux élémentaire d'électrons $d^3\Phi(y=0, x, z, \vec{k})$ est calculé grâce à la fonction de distribution f des électrons appliquée au point $(y=0, x, z)$:

$$d^3\Phi(y=0, x, z, \vec{k}) = d^3n \times v_y = \frac{1}{4\pi^3} f(y=0, x, z, \vec{k}) d^3\vec{k} \times v_y \quad (2.115)$$

Dans cette expression, v_y désigne la vitesse des électrons selon la direction y perpendiculaire à l'interface. Elle est liée au diagramme des bandes du semiconducteur par l'intermédiaire de la relation de dispersion $\varepsilon = \varepsilon(\vec{k})$, où ε désigne l'énergie totale des électrons. La vitesse v_y s'exprime en fonction du vecteur d'onde des électrons dans la direction y :

$$v_y = \frac{1}{\hbar} \frac{\partial \mathcal{E}}{\partial k_y} \quad (2.116)$$

La densité de courant tunnel au point (x, z) de l'interface semiconducteur oxyde est finalement donnée par :

$$J_G(x, z) = q \iiint \frac{1}{\hbar} \frac{\partial \mathcal{E}}{\partial k_y} f(y=0, x, z, \vec{k}) P(x, z, \vec{k}) d^3 \vec{k} \quad (2.117)$$

Cette approche générale permet donc de calculer la densité de courant tunnel connaissant la fonction de distribution des électrons et la probabilité tunnel en tout point de l'interface.

Le calcul de la probabilité tunnel nécessite la résolution de l'équation de Schrödinger indépendante du temps, dans le but d'évaluer l'atténuation de la fonction d'onde des électrons incidents. Dans l'hypothèse où les directions de l'espace sont découplées, la fonction enveloppe des électrons $\varphi(x, y, z)$ se décompose comme le produit de deux fonctions enveloppe $\chi(x, z)$ et $\psi(y)$ correspondant respectivement à la direction parallèle x - z et à celle perpendiculaire à l'interface, chacune de ces fonctions obéissant à une équation de Schrödinger.

L'étude de la transition des fonctions d'ondes du semiconducteur à l'oxyde, et notamment la continuité de la fonction d'onde et du courant de probabilité [COHEN'73] permettent d'établir que :

- Le vecteur d'onde parallèle se conserve
- $\psi(y)$ est continu
- $\frac{1}{m_y^*} \frac{d\psi(y)}{dy}$ est continu

L'équation (2.117) et les hypothèses énoncées ci-dessus reposent sur la nature cristalline de tous les matériaux du dispositif. Cependant, on montre que la continuité du vecteur d'onde parallèle induit une variation de la hauteur de barrière suivant l'orientation cristalline. Cette propriété n'a pas été vérifiée expérimentalement [WEINBERG'83]. La nature non cristalline de l'oxyde conduit donc à une violation de la conservation du vecteur d'onde parallèle. Cette hypothèse interdit en toute rigueur l'application de l'équation de Schrödinger dans la direction perpendiculaire à l'interface. Cependant, compte tenu de la très faible épaisseur de la couche d'oxyde (1-2 nm), on peut supposer que la couche diélectrique possède selon la direction perpendiculaire à l'interface des propriétés proches d'un matériau

cristallin. L'approche la plus communément adoptée pour calculer l'atténuation de la fonction d'onde dans la direction perpendiculaire est donc de résoudre l'équation de Schrödinger dans la direction perpendiculaire en supposant que l'injection des porteurs par effet tunnel s'effectue à énergie perpendiculaire ε_y constante.

Dans ces conditions, la probabilité tunnel ne dépend que de la forme de la barrière de potentiel et de leur énergie perpendiculaire ε_y . Le calcul devient purement unidimensionnel.

2.7.2 Calcul analytique de la probabilité tunnel

L'évaluation de la probabilité tunnel $P(\varepsilon_y)$ passe par la résolution de l'équation de Schrödinger dans l'approximation de la masse effective. La résolution exacte de l'équation de Schrödinger est basée sur une description de la barrière de potentiel réelle par un profil constant par morceaux. La probabilité tunnel est calculée selon la méthode du transfert matriciel (Annexe B). Cette méthode requiert des temps de calculs importants et constitue notre simulation de référence pour évaluer la précision des approximations analytiques détaillées dans cette partie.

La figure 2.35 présente un exemple de barrière de potentiel utilisée dans la résolution par transfert matriciel. La probabilité tunnel dépend de la forme de la barrière de potentiel (de la largeur T_{ox} , de la hauteur χ_B de la barrière et du champ électrique dans l'oxyde F_{ox}), de l'énergie perpendiculaire ε_y et des masses des porteurs dans les trois matériaux.

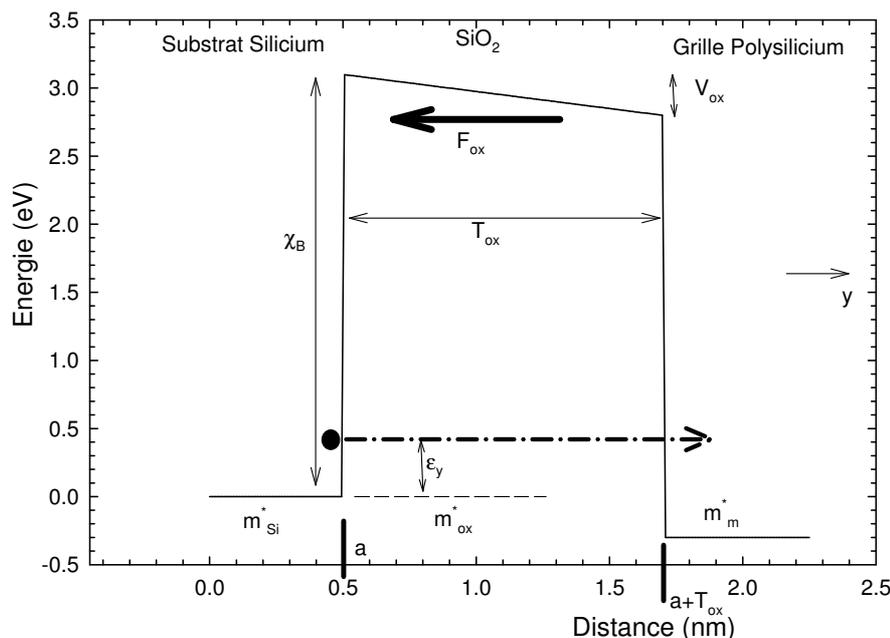


Figure 2.35 : Représentation de la barrière de potentiel vue par les électrons du canal pour une structure nMOS avec une épaisseur d'oxyde de 1 nm, polarisé en forte inversion avec $V_{ox} = 0.3$ Volts

La figure 2.36 illustre l'impact de la masse effective des électrons dans l'oxyde. L'influence de la masse effective des porteurs dans la bande interdite de l'oxyde est considérable. D'un point de vue théorique pour un matériau SiO_2 parfait, la masse effective à prendre en compte est $m_{ox}^* = 0,5 m_0$, m_0 étant la masse de l'électron dans le vide. La figure 2.37 illustre l'influence de la hauteur de barrière sur la probabilité tunnel pour deux énergies d'injection distinctes. On observe également une très forte dépendance du résultat à ce paramètre. La hauteur de barrière à prendre en compte pour une orientation cristalline (100) est $\chi_B = 3.1\text{eV}$.

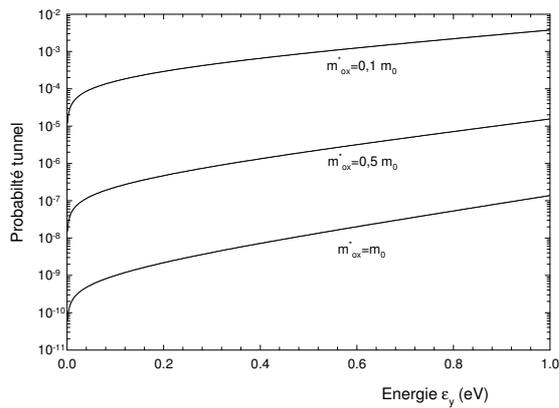


Figure 2.36 : Influence de la masse effective des électrons dans l'oxyde sur la probabilité tunnel.

$T_{ox} = 1.2 \text{ nm}$, $V_{ox} = 0.3 \text{ Volts}$, $\chi_B = 3.1 \text{ eV}$.

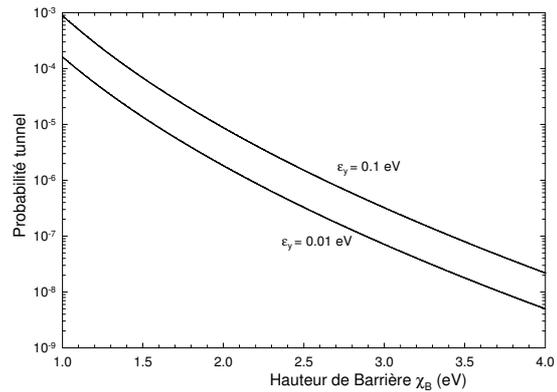


Figure 2.37 : Influence de la hauteur de barrière sur la probabilité tunnel. $T_{ox} = 1.2 \text{ nm}$, $V_{ox} = 0.3$

Volts, $m_{ox}^* = 0.5 m_0$

La figure 2.38 présente la probabilité tunnel en fonction de l'épaisseur d'oxyde, pour deux valeurs d'énergie perpendiculaire des électrons incidents. On constate une très forte dépendance de la probabilité tunnel et la diminution de la probabilité varie quasiment exponentiellement avec l'épaisseur.

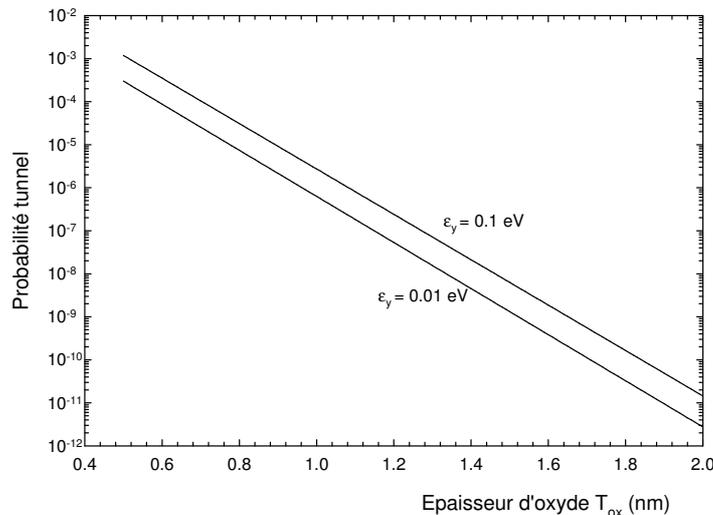


Figure 2.38 : Influence de l'épaisseur d'oxyde sur probabilité tunnel. $V_{ox} = 0.3 \text{ Volts}$, $\chi_B = 3.1 \text{ eV}$, $m_{ox}^* = 0.5$

m_0 .

Les masses effectives des électrons dans le silicium et dans le polysilicium influencent peu la probabilité tunnel. Compte tenu de la faible dépendance de la hauteur de barrière vue par les électrons à l'interface Si-SiO₂ en fonction de l'orientation cristalline, il n'est pas nécessaire de faire une distinction entre les différentes vallées ellipsoïdales du silicium lors du calcul de la probabilité tunnel des électrons [CASSAN'00]. La masse m_{Si}^* qui est communément utilisée dans le calcul de la probabilité tunnel est la masse effective de conduction du silicium $m_{\text{Si}}^* = 0.258 m_0$. La masse effective dans le matériau de grille, dont l'influence sur la probabilité tunnel est faible, sera prise égale à m_0 .

La probabilité tunnel est donc très sensible à la hauteur de barrière, l'épaisseur d'oxyde et la masse effective des porteurs dans l'oxyde. De plus, la constante diélectrique de l'oxyde n'intervient pas dans la probabilité tunnel. Cette propriété est très intéressante, puisqu'elle justifie l'utilisation d'oxydes nitrurés pour les technologies MOS avancées, voire des oxydes à haute permittivité (*high K*). En effet, pour travailler à capacité d'oxyde équivalente avec une constante diélectrique supérieure à celle du SiO₂, on doit augmenter l'épaisseur physique du film diélectrique. Le courant de grille étant sensible à l'épaisseur physique, le courant tunnel se trouve fortement diminué. Nous verrons par la suite la nécessité de la prise en compte de l'épaisseur physique du film tant pour la modélisation du courant de grille que pour le bruit basse fréquence associé au courant tunnel.

L'approche analytique la plus utilisée pour calculer la probabilité tunnel est basée sur la méthode WKB (Wentzel Kramers Brillouin). Soit la fonction d'onde $\varphi(y)$ satisfaisant à l'équation de Schrödinger :

$$\frac{d^2\varphi(y)}{dy^2} + \frac{2m^*}{\hbar^2} [E - V(y)] \varphi(y) = 0 \quad (2.118)$$

En posant $\varphi = e^{i\omega/\hbar}$ et $\omega = S + \frac{\hbar}{i} \ln(A)$, où S et $\ln(A)$ sont des fonctions paires de \hbar , on obtient une équation différentielle du 3^{ème} ordre en S rigoureusement équivalente à l'équation de Schrödinger (2.118) [MESSIAH'59]:

$$S'^2 = 2m^* [E - V(y)] + \hbar^2 \left[\frac{3}{4} \left(\frac{S''}{S'} \right)^2 - \frac{1}{2} \frac{S'''}{S'} \right] \quad (2.119)$$

L'approximation WKB consiste à développer S en série de puissance de \hbar^2 :

$$S = S_0 + \hbar^2 S_1 + \dots \quad (2.120)$$

En substituant le développement (2.120) dans l'équation (2.119) et en ne gardant que les termes d'ordre zéro, on obtient une équation différentielle qui s'intègre sans difficulté :

$$S'^2 \simeq S_0'^2 = 2m^* [E - V(y)] \quad (2.121)$$

Suivant le signe de la quantité $E - V(y)$, on obtient les solutions WKB de la forme :

$$\begin{aligned} \varphi(y) &= \alpha \cos\left(\int \frac{dy}{\lambda} + \phi\right) & \text{avec } \lambda(y) &= \frac{\hbar}{\sqrt{2m^* [E - V(y)]}} \text{ pour } E > V(y) \\ \varphi(y) &= \left[\gamma \exp\left(+\int \frac{dy}{l}\right) + \delta \exp\left(-\int \frac{dy}{l}\right) \right] & \text{avec } l(y) &= \frac{\hbar}{\sqrt{2m^* [V(y) - E]}} \text{ pour } E < V(y) \end{aligned} \quad (2.122)$$

où $\alpha, \phi, \gamma, \delta$ sont des constantes arbitraires.

L'approche classique du courant de grille par la méthode WKB [DEPAS'95] consiste à calculer la transmission au travers de la barrière de potentiel sans tenir compte des discontinuités de la bande de conduction et des masses effectives. Conformément aux équations (2.122) et en suivant les notations de la figure 2.35, la fonction d'onde dans l'oxyde s'écrit :

$$\varphi(y) = \exp\left(-\int_0^{T_{ox}} \frac{\sqrt{2m_{ox}^*}}{\hbar} \sqrt{q\chi_B - qF_{ox}y - \varepsilon_y} dy\right) \quad (2.123)$$

La probabilité tunnel est donnée par la probabilité de présence de l'électron en $y = T_{ox}$ et s'écrit donc :

$$P(\varepsilon_y) = \varphi^* \varphi = \exp\left(-2\sqrt{\frac{2m_{ox}^*}{\hbar^2}} \int_0^{T_{ox}} \sqrt{q\chi_B - qF_{ox}y - \varepsilon_y} dy\right) \quad (2.124)$$

En intégrant l'équation (2.124) et en utilisant la relation $F_{ox}T_{ox} = V_{ox}$, on obtient la relation classique donnant la probabilité tunnel WKB en fonction de l'énergie et de la chute de potentiel dans l'oxyde :

$$P_{wkb}(\varepsilon_y, V_{ox}) = \exp\left(-4\frac{\sqrt{2m_{ox}^*}}{3q\hbar} \frac{T_{ox}}{V_{ox}} \left[(q\chi_B - \varepsilon_y)^{3/2} - (q\chi_B - qV_{ox} - \varepsilon_y)^{3/2} \right]\right) \quad (2.125)$$

La Figure 2.39 compare les probabilités calculées par la méthode du transfert matriciel (annexe B) et celles calculées à l'aide de l'équation (2.125) pour une épaisseur d'oxyde de 1.2 nm. On constate que les résultats sont loin d'être satisfaisants, puisque l'approximation WKB décrite précédemment surestime la probabilité à faible énergie alors qu'elle la sous-estime à forte énergie. La non prise en compte des réflexions aux interfaces Si-SiO₂ et SiO₂-PolySi ne permet pas d'obtenir une approximation analytique correcte de la probabilité tunnel.

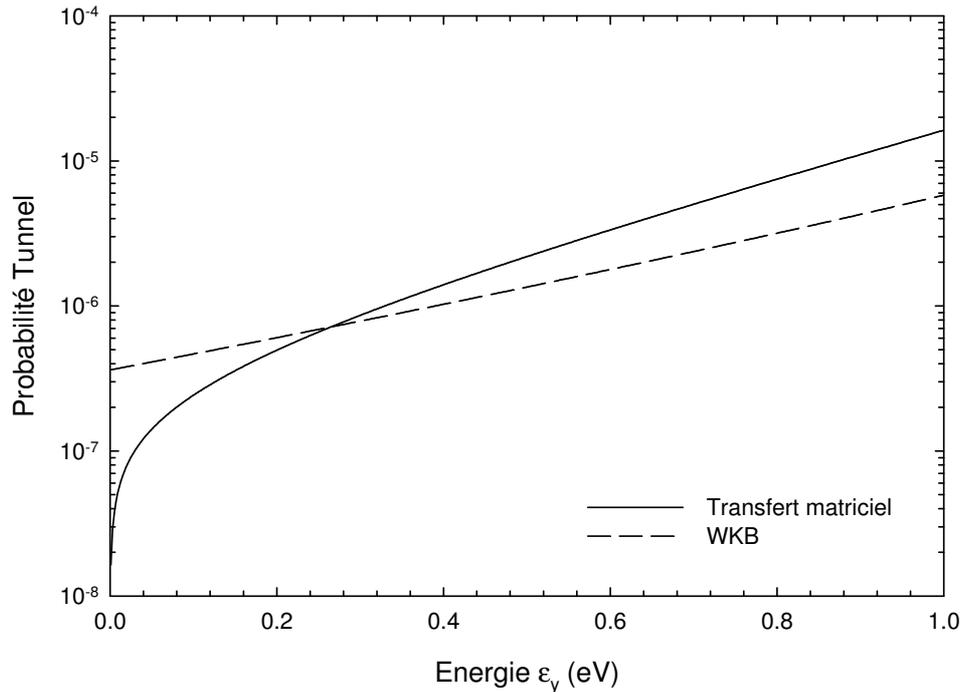


Figure 2.39 : Probabilité tunnel calculée par la méthode du transfert matriciel et par la méthode WKB pour une épaisseur d'oxyde de 1.2 nm.

La modélisation analytique des réflexions liées aux interfaces Si-SiO₂ et SiO₂-Polysilicium a été proposée par Register [REGISTER'99]. L'auteur propose de multiplier la probabilité tunnel (2.125) par un facteur correctif donné par :

$$C(\varepsilon_y) = \frac{4v(\varepsilon_y)v_{ox}(\varepsilon_y)}{v(\varepsilon_y)^2 + v_{ox}(\varepsilon_y)^2} \times \frac{4v(\varepsilon_y + qV_{ox})v_{ox}(\varepsilon_y + qV_{ox})}{v(\varepsilon_y + qV_{ox})^2 + v_{ox}(\varepsilon_y + qV_{ox})^2} \quad (2.126)$$

$$\text{où } v(\varepsilon_y) = \sqrt{\varepsilon_y / 2m_{si}^*} \text{ et } v_{ox}(\varepsilon_y) = \sqrt{(q\chi_B - \varepsilon_y) / 2m_{ox}^*}.$$

Ce résultat est systématiquement utilisé dans les publications récentes relatives aux courants tunnel dans les oxydes. Cependant, la référence dans [REGISTER'99] associée à ce résultat reste mystérieuse¹². De notre point de vue, ce terme ne constitue pas une correction de l'approche WKB (2.125), mais plutôt d'un traitement global du calcul de la probabilité par la méthode WKB. Nous proposons donc de développer les calculs aboutissant à une nouvelle formulation WKB de la probabilité tunnel.

Nous allons tout d'abord définir 3 zones sur la figure 2.35. La zone 1 correspond au substrat silicium qui s'étend depuis $y \rightarrow -\infty$ jusqu'à l'abscisse notée $y=a$. Dans cette zone, le potentiel est considéré comme constant, et l'équation de Schrödinger s'intègre

¹² L'auteur se cite lui même pour des travaux non publiés.

analytiquement. La fonction d'onde résultante est la somme d'une onde incidente et d'une onde réfléchie, définie par :

$$\varphi_1(y) = A \sin\left(\frac{y-a}{\lambda_a} + \phi\right) \quad (2.127)$$

avec $\lambda_a = \frac{\hbar}{\sqrt{2m_{Si}^* \varepsilon_y}}$ en prenant $V=0$ dans cette zone. A et ϕ sont deux constantes

d'intégration qui seront déterminées par la suite.

La zone 2 correspond à l'oxyde tel que $a < y < a + T_{ox}$. En utilisant les résultats de la méthode WKB (2.122), la fonction d'onde dans cette zone s'écrit :

$$\varphi_2(y) = \sqrt{l} \left[\gamma \exp\left(+\int \frac{dy}{l}\right) + \delta \exp\left(-\int \frac{dy}{l}\right) \right] \quad (2.128)$$

avec $l(y) = \frac{\hbar}{\sqrt{2m_{ox}^* [q\chi_B - qF_{ox}(y-a) - \varepsilon_y]}}$. γ et δ sont deux constantes

d'intégration qui seront déterminées par la suite.

Enfin, la zone 3 correspond au polysilicium, dans laquelle il n'y a qu'une onde transmise :

$$\varphi_3(y) = \sqrt{\lambda_b} \exp\left(j \frac{y-(a+T_{ox})}{\lambda_b}\right) \quad (2.129)$$

avec $\lambda_b = \frac{\hbar}{\sqrt{2m_{Si}^* (\varepsilon_y + V_{ox})}}$ et $j^2 = -1$.

Posons $l(a) = l_a$, $l(b = a + T_{ox}) = l_b$ et définissons les vitesses de groupe suivantes :

$$\begin{aligned} v(\varepsilon_y) &= \frac{\hbar}{\lambda_a} \frac{1}{2m_{Si}^*} & v_{ox}(\varepsilon_y) &= \frac{\hbar}{l_a} \frac{1}{2m_{ox}^*} \\ v(\varepsilon_y + qV_{ox}) &= \frac{\hbar}{\lambda_b} \frac{1}{2m_{Si}^*} & v_{ox}(\varepsilon_y + qV_{ox}) &= \frac{\hbar}{l_b} \frac{1}{2m_{ox}^*} \end{aligned} \quad (2.130)$$

Les règles de continuité des fonctions d'onde entre les zones 2 et 3 permettent d'établir le système suivant :

$$\begin{cases} e^\tau \gamma + e^{-\tau} \delta = \sqrt{\frac{v_{ox}(\varepsilon_y + qV_{ox}) m_{ox}^*}{v(\varepsilon_y + qV_{ox}) m_{Si}^*}} \\ e^\tau \gamma - e^{-\tau} \delta = j \sqrt{\frac{v(\varepsilon_y + qV_{ox}) m_{ox}^*}{v_{ox}(\varepsilon_y + qV_{ox}) m_{Si}^*}} \end{cases} \quad \tau = \int_a^{a+T_{ox}} \frac{\sqrt{2m_{ox}^*}}{\hbar} \sqrt{q\chi_B - qF_{ox}(y-a) - \varepsilon_y} dy \quad (2.131)$$

En appliquant les règles de continuité entre les zones 1 et 2, on obtient le système d'équations permettant de calculer l'amplitude A de l'onde incidente de la zone 1:

$$\begin{cases} A \sin \phi = \sqrt{l_a} (\gamma + \delta) \\ \frac{A}{\lambda_a m_{Si}^*} \cos \phi = \frac{1}{m_{ox}^*} \frac{1}{\sqrt{l_a}} (\gamma - \delta) \end{cases} \quad (2.132)$$

Le module carré de l'amplitude de l'onde incidente s'écrit donc :

$$A^2 = l_a (\gamma + \delta)^2 + \frac{\lambda_a^2}{l_a} \left(\frac{m_{Si}^*}{m_{ox}^*} \right)^2 (\gamma - \delta)^2 \quad (2.133)$$

La constante d'intégration γ est associée à la réflexion de l'interface SiO₂-Poly. On fait l'hypothèse que cette composante de la fonction d'onde est négligeable à l'interface Si-

SiO₂¹³, permettant ainsi d'exprimer simplement le flux incident $\frac{|A|^2}{4} \frac{\hbar}{\lambda_a m_{Si}^*}$ par :

$$\frac{|A|^2}{4} \frac{\hbar}{\lambda_a m_{Si}^*} = \left\{ \frac{l_a}{\lambda_a} + \frac{\lambda_a}{l_a} \left(\frac{m_{Si}^*}{m_{ox}^*} \right)^2 \right\} |\delta|^2 = \frac{1}{4} \frac{m_{Si}^*}{m_{ox}^*} \left\{ \frac{v(\varepsilon_y)}{v_{ox}(\varepsilon_y)} + \frac{v_{ox}(\varepsilon_y)}{v(\varepsilon_y)} \right\} |\delta|^2 \quad (2.134)$$

En injectant la valeur de δ calculée à partir du système (2.131), le flux total incident s'exprime par :

$$\frac{|A|^2}{4} \frac{\hbar}{\lambda_a m_{Si}^*} = \frac{1}{4} \cdot \frac{v(\varepsilon_y)^2 + v_{ox}(\varepsilon_y)^2}{v(\varepsilon_y)v_{ox}(\varepsilon_y)} \cdot \frac{1}{4} \cdot \frac{v_{ox}(\varepsilon_y + qV_{ox})^2 + v(\varepsilon_y + qV_{ox})^2}{v(\varepsilon_y + qV_{ox})v_{ox}(\varepsilon_y + qV_{ox})} e^{2\tau} \quad (2.135)$$

Comme le flux de l'onde transmise $\varphi_3(y)$ est égal à $\frac{\hbar}{m_{Si}^*}$, la probabilité tunnel s'exprime donc par :

$$P(\varepsilon_y, V_{ox}) = 4 \cdot \frac{v(\varepsilon_y)v_{ox}(\varepsilon_y)}{v(\varepsilon_y)^2 + v_{ox}(\varepsilon_y)^2} \cdot 4 \cdot \frac{v(\varepsilon_y + qV_{ox})v_{ox}(\varepsilon_y + qV_{ox})}{v_{ox}(\varepsilon_y + qV_{ox})^2 + v(\varepsilon_y + qV_{ox})^2} \cdot P_{wkb}(\varepsilon_y, V_{ox}) \quad (2.136)$$

La méthode WKB et les conditions usuelles de passage d'une zone à une autre nous ont donc permis de formuler le facteur de correction proposé par Register. La figure 2.40 présente la probabilité tunnel calculée par transfert matriciel et l'équation (2.136). Les résultats obtenus avec cette approche analytique donnent d'excellents résultats (erreur relative inférieure au pourcent) et sa formulation explicite de l'énergie de l'électron et de la chute de potentiel dans l'oxyde permet de la coupler facilement aux modèles de potentiel de surface.

¹³ Pour justifier cette approximation, il suffit de factoriser l'expression (2.133) par δ^2 . Comme le terme $\gamma/\delta \propto e^{-2\tau}$, on peut le négliger devant 1 et obtenir l'équation (2.134)

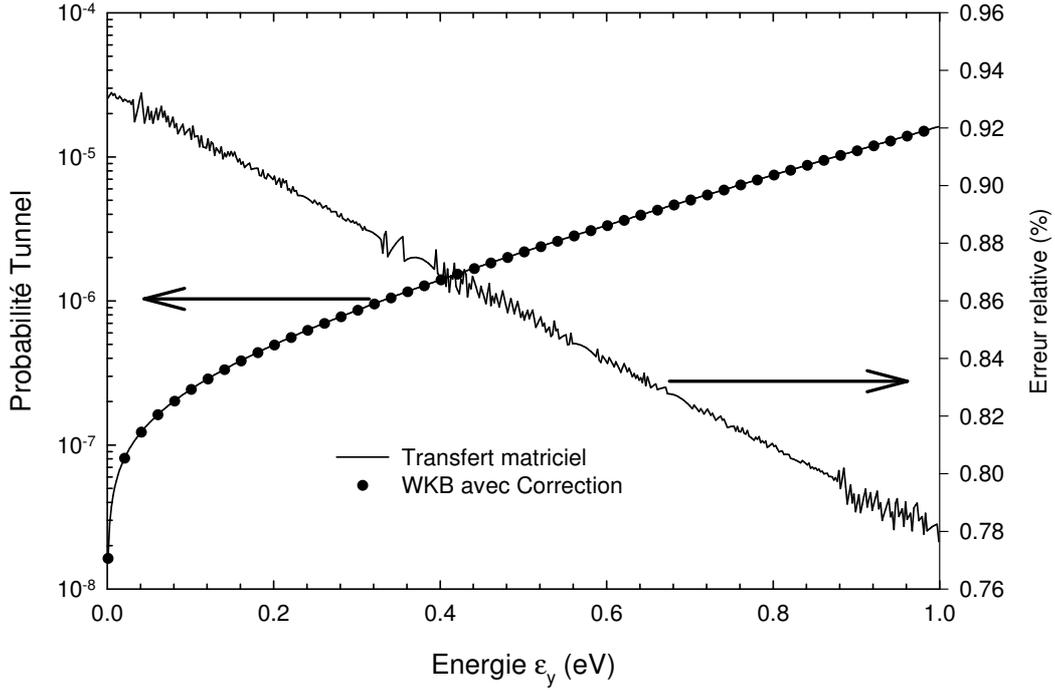


Figure 2.40 : Probabilité tunnel calculée par la méthode du transfert matriciel et par la méthode WKB avec la prise en compte des réflexions des interfaces, pour une épaisseur d'oxyde de 1.2 nm et $V_{ox}=0.3$ Volts.

2.7.3 Formulation explicite du courant tunnel direct en fonction du potentiel de surface

Nous disposons donc d'une expression analytique de la probabilité tunnel. L'expression générale du courant (2.117) requiert la connaissance du flux incident d'électrons. Dans le cas des structures fortement dopées où les niveaux d'énergies sont discrétisés, on fait l'approximation que seuls les électrons de la sous bande la plus peuplée contribuent au courant tunnel. Sous cette hypothèse, la charge susceptible de traverser par effet tunnel est donc définie par l'équation (2.81) ou (2.83), et l'énergie des porteurs injectés est :

$$\varepsilon_y(\Psi_s) = E(\Psi_s) = \frac{3\hbar^2 b(\Psi_s)^2}{8m_{Si}^*} \quad (2.137)$$

La fréquence d'impact f est définie comme étant le flux de porteurs frappant la barrière de potentiel par unité de temps. Pour le premier niveau d'énergie, la fréquence d'impact est donnée par[CLERC'02] :

$$f = \frac{\hbar b(\Psi_s)^2}{4\pi m_{Si}^*} = \frac{2}{3} \frac{2E(\Psi_s)}{h} \quad (2.138)$$

La densité de courant tunnel s'exprime simplement comme le produit de la charge d'inversion, de la fréquence d'impact et de la probabilité tunnel :

$$J_G(V_{GB}, \Psi_S) = q \cdot f(\Psi_S) \cdot N_{inv}(V_{GB}, \Psi_S) P(V_{GB}, \Psi_S) \quad (2.139)$$

Cette expression est donc totalement compatible avec le modèle quantique de potentiel de surface établi dans la partie 2.6. La figure 2.41 présente la simulation du courant de grille pour une capacité nMOS en régime de faible et forte inversion, avec une épaisseur d'oxyde de 1.2 nm, un substrat de type p dopé à 10^{18} cm^{-3} , une grille polysilicium de type n dopée à $5 \cdot 10^{19} \text{ cm}^{-3}$. La masse effective des électrons dans l'oxyde est égale à $0.5 m_0$ et la barrière de potentiel est égale à 3.1 eV.

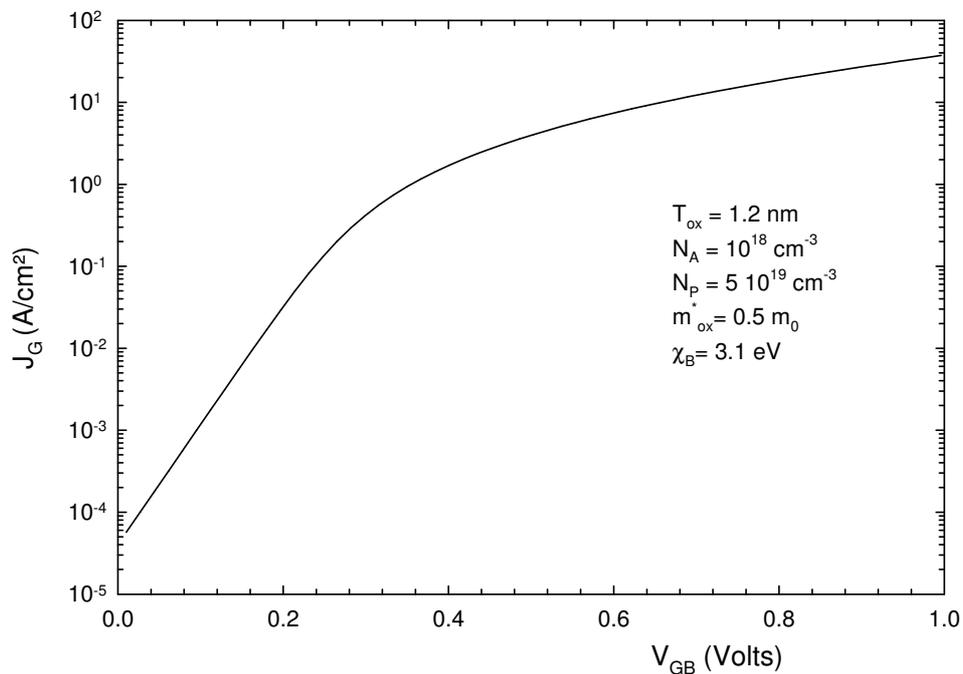


Figure 2.41 : Simulation de la densité de courant de grille à partir du modèle de potentiel de surface quantique explicite et de la probabilité tunnel WKB tenant compte des réflexions aux interfaces.

2.8 Conclusion

Tous le long de ce chapitre, nous avons présenté la modélisation des nouveaux effets physiques présents dans les structures MOS avancées. Notre étude s'est plus précisément focalisée sur le phénomène de déplétion dans le polysilicium de grille, des effets de confinement de la couche d'inversion et enfin de courant tunnel direct au travers du film diélectrique ultrafin. Pour la modélisation de tous ces phénomènes, nous avons choisi d'utiliser comme point central le potentiel de surface. En effet, la connaissance précise du potentiel de surface permet d'accéder par la suite à toutes les grandeurs physiques d'une capacité MOS, et ouvre la voie de la modélisation du transistor MOS basée sur le potentiel de surface.

Dans un premier temps, nous avons rappelé la formulation classique de l'équation implicite liant le potentiel de surface, la tension de grille, et le quasi-niveau de Fermi dans une structure MOS.

Nous avons ensuite présenté la modélisation du phénomène de polydéplétion dans le matériau de grille. Ce modèle est formulé en fonction de la tension de grille et du potentiel de surface, ce qui permet de l'inclure facilement dans l'équation implicite définie au paragraphe 2.3. Au travers des simulations de la capacité de la structure en tenant compte des effets de polydéplétion, nous avons montré qu'il n'est plus possible de négliger ce phénomène dans une procédure de caractérisation de diélectrique de grille, puisque la capacité maximum se trouve atténuée de près de 35 %.

Le paragraphe 2.5 a présenté l'état de l'art de la modélisation explicite du potentiel de surface. Les résultats fournis par ces modèles sont très bons comparés à ceux des modèles implicites, et les temps de calculs sont quasi-nuls. Cependant, nous avons mis en évidence la nécessité d'utiliser de nombreuses équations de lissage et donc de paramètres empiriques pour obtenir d'une part un potentiel de surface satisfaisant, et d'autre part, une dérivée du potentiel de surface satisfaisante.

Le paragraphe 2.6 traite des effets quantiques et de leur modélisation dans les capacités MOS. Dans un premier temps, nous avons rappelé les bases de mécanique quantique nécessaires pour appréhender l'impact des effets quantiques, et notamment la structure de sous bande d'énergie du puits de potentiel où sont confinés les porteurs libres. Nous avons ensuite présenté les principaux résultats de simulation obtenus avec le solveur auto cohérent QMCV, et établi les hypothèses simplificatrices permettant de formuler par une approche analytique les effets quantiques dans les structures MOS.

Les deux approches analytiques communément utilisées ont été présentées au paragraphe 2.6.3. A partir de ces méthodes de calcul, et notamment à partir de l'approximation variationnelle, nous avons formulé dans le paragraphe 2.6.4 un modèle implicite quantique de potentiel de surface, tenant compte de la polydépétion. Ce modèle est valable dans tous les régimes de fonctionnement, pour des structures de type n- ou p - MOS.

Dans le paragraphe 2.6.5, nous nous sommes intéressés à un traitement mathématique du modèle implicite précédent dans le but de le rendre explicite. Nous avons établi un nouveau modèle explicite valable en régime de dépétion et d'inversion, ne nécessitant aucune fonction de lissage, et tenant compte de façon intrinsèque du phénomène de polydépétion. La comparaison avec les modèles explicites quantiques déjà publiés dans la littérature a montré une diminution importante de l'erreur relative. De plus, la dérivée du potentiel de surface issue de ce nouveau modèle présente un excellent accord avec les résultats issus de calculs auto cohérents. Les très bons résultats obtenus avec des paramètres technologiques correspondant à des technologies très avancées montrent qu'il peut être utilisé comme base des futurs modèles compacts.

Dans le dernier paragraphe, nous avons rappelé les propriétés physiques du courant tunnel d'une part au travers d'une méthode numérique (transfert matriciel), et d'autre part avec une formulation analytique (méthode WKB). L'étude par le transfert matriciel a permis de déterminer l'influence des paramètres physiques sur la probabilité tunnel, à savoir que les paramètres essentiels sont l'épaisseur physique, la masse effective des porteurs, et la hauteur de barrière de potentiel à l'interface Si/SiO₂. Nous avons rappelé l'approche classique WKB du calcul de la probabilité tunnel, et nous avons montré que les réflexions aux deux interfaces ne pouvaient pas être négligés. Le facteur de correction proposé par Register a donc été redémontré dans cette partie.

Au final, nous disposons donc d'une part d'un modèle explicite de potentiel de surface tenant compte à la fois des effets quantiques et du phénomène de polydépétion, et d'autre part d'un modèle de courant tunnel de grille totalement lié au potentiel de surface. Ces modèles sont donc la base du modèle de transistor MOS qui sera développé au chapitre suivant.

CHAPITRE 3
MODÉLISATION DU TRANSISTOR MOS
AVEC FUITE DE GRILLE

3 Modélisation du transistor MOS avec fuite de grille

3.1 Introduction

Le modèle de transistor MOS que nous présenterons dans ce chapitre est un modèle destiné à la caractérisation de structures MOS à diélectrique ultrafin. Toutefois, nous ne nous intéresserons pas aux problèmes de transports dans les canaux ultracourts. Le but n'est pas de définir un modèle compact pour la simulation de circuits, mais d'élaborer simplement un modèle analytique ou quasi analytique nécessitant de faibles temps de calculs comparés à des simulateurs physiques.

Dans un premier temps, nous définissons le modèle de transistor MOS permettant de simuler la caractéristique statique I_D-V_G en régime linéaire, sur des structures à canal long (typiquement avec une longueur de canal supérieure à $1\ \mu\text{m}$). Nous verrons ensuite l'impact de la polydéplétion et des effets quantiques sur la caractéristique statique.

Nous présenterons ensuite les structures que nous avons eu à disposition durant cette thèse. Nous nous focaliserons sur les transistors à canal long pour valider le modèle de transistors prenant en compte le courant tunnel de grille.

Enfin nous nous intéressons aux structures MOS de grande surface. Nous montrons que ces géométries nécessitent l'utilisation d'un modèle non quasi statique pour la simulation de la capacité de la structure.

3.2 Modélisation classique en feuille de charge du transistor MOS

Les équations et les modèles développés dans cette partie concernent un transistor nMOS avec un dopage substrat de type p uniforme N_A , d'une grille en polysilicium avec un dopage de type n uniforme N_p . La figure 3.1 présente une vue schématique de la structure étudiée.

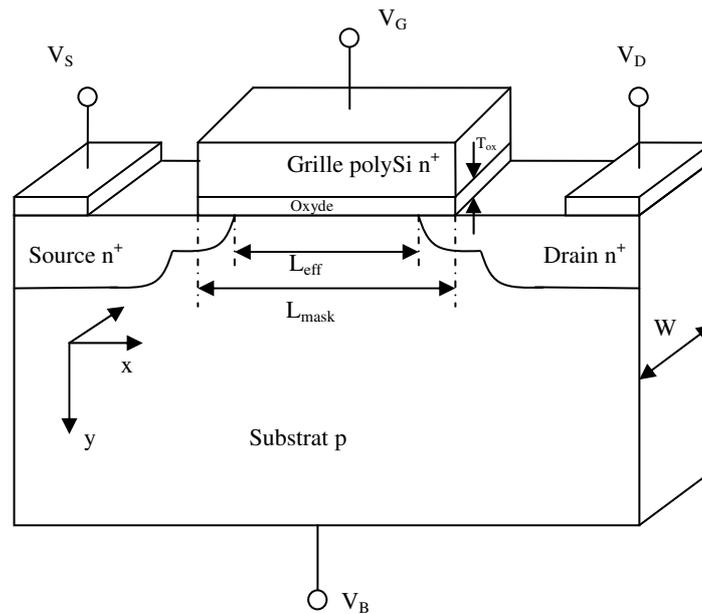


Figure 3.1 . Structure d'un transistor MOS de type n

Dans le cas où la structure est longue et large, les effets de canal court et étroit sont négligés. Les équations à prendre en compte pour déterminer les caractéristiques statiques et dynamiques sont l'équation de Poisson, l'équation de transport et les équations de continuité pour les électrons et les trous.

La résolution couplée de ces équations est un problème à trois dimensions, voire quatre dimensions si on inclut la variable temporelle¹⁴. Dans un premier temps, nous n'étudierons que l'aspect statique ou quasi statique du problème. La réduction de trois à deux dimensions se fait aisément en ne considérant uniquement que les dimensions x et y . Cependant, la résolution numérique des équations citées précédemment est nécessaire car il n'existe pas de solution analytique du problème. Mathématiquement, cela revient souvent à résoudre un système d'équations aux dérivées partielles. La méthode la plus communément utilisée est celle des éléments finis, comme par exemple dans le logiciel ISE TCAD. Le problème majeur réside dans les temps de calculs souvent extrêmement longs quand les structures deviennent complexes. On peut néanmoins simplifier analytiquement les équations dans le cas qui nous intéresse, à savoir les transistors à canal long. Les hypothèses pour réduire l'étude à un problème unidimensionnel sont les suivantes :

- i) La variation du champ électrique longitudinal E_x dans la direction x est négligeable devant la variation du champ électrique transverse E_y dans la

¹⁴ On parle alors de modèle non quasi statique.

direction y . Cette hypothèse s'appelle l'approximation du canal graduel et permet de réduire le Laplacien de l'équation de Poisson à une simple dérivée seconde suivant l'axe des y . Grâce à cette hypothèse, modèle de potentiel de surface établi au chapitre précédent reste donc valable le long du canal.

- ii) Le courant de trous (d'électrons) est négligé pour les transistors nMOS (pMOS), et le courant total s'exprime suivant le modèle de dérive-diffusion :

$$J_n = qn \frac{dE_{F_{n(p)}}}{dx} \quad (3.1)$$

- iii) Les phénomènes de génération et recombinaison sont négligés, de sorte que l'équation de continuité des porteurs libres s'exprime simplement par :

$$\frac{dI(x)}{dx} = 0 \quad (3.2)$$

- iv) Le courant ne circule que dans la direction longitudinale. Le quasi niveau de Fermi des électrons est donc constant dans la direction y . Cette hypothèse est également compatible avec le modèle de potentiel de surface du chapitre 2.

En suivant le modèle dérive diffusion, le courant de drain s'exprime sous la forme :

$$I(x) = -\mu(x)W Q_{inv}(x) \frac{dV}{dx} \quad (3.3)$$

où $\mu(x)$ est la mobilité des porteurs libres au point x du canal.

Le principe du modèle en feuille de charge [BREWS'78] réside dans la formulation de la charge d'inversion uniquement en fonction du potentiel de surface et de la tension de grille. L'annexe A décrit le calcul pour parvenir aux formulations suivantes :

$$Q_{dep}(\Psi_S) = -C_{ox} \gamma \sqrt{\Psi_S + \phi_T \left(\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1 \right)} \quad (3.4)$$

$$Q_{inv}(V_{GB}, \Psi_S) = -C_{ox} \left(V_{GB} - V_{FB} - \Psi_S - \Psi_P - \gamma \sqrt{\Psi_S + \phi_T \left(\exp\left(-\frac{\Psi_S}{\phi_T}\right) - 1 \right)} \right) \quad (3.5)$$

En partant de l'équation implicite classique (2.8), et en différentiant le potentiel de surface par rapport à V , on obtient :

$$\frac{\partial \Psi_S}{\partial V} = \frac{\gamma^2 \exp\left(\frac{\Psi_S - V - \Phi_B}{\phi_T}\right)}{\gamma^2 \left[1 + \exp\left(\frac{\Psi_S - V - \Phi_B}{\phi_T}\right) - \exp\left(-\frac{\Psi_S}{\phi_T}\right) \right] - 2V_{ox} \frac{\partial V_{ox}}{\partial \Psi_S}} \quad (3.6)$$

où $V_{ox} = V_{GB} - V_{FB} - \Psi_S - \Psi_P$.

Cette relation peut se mettre sous la forme simplifiée [BREWS'78] :

$$\frac{\partial V}{\partial \Psi_s} \approx 1 + \phi_T \frac{\gamma^2 \left[1 - \exp\left(-\frac{\Psi_s}{\phi_T}\right) \right] - 2V_{ox} \frac{\partial V_{ox}}{\partial \Psi_s}}{V_{ox}^2 - \gamma^2 \left(\Psi_s + \phi_T \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right] \right)} \quad (3.7)$$

En utilisant les expressions (3.4) et (3.5), on obtient l'expression suivante :

$$\frac{\partial V}{\partial \Psi_s} \approx 1 + \phi_T \frac{C_{ox}}{Q_{inv}} \left(\frac{\partial V_{ox}}{\partial \Psi_s} + \frac{Q_{inv} \frac{\partial V_{ox}}{\partial \Psi_s} + \gamma^2 C_{ox} \left[1 - \exp\left(-\frac{\Psi_s}{\phi_T}\right) \right]}{Q_{inv} + 2Q_{dep}} \right) \quad (3.8)$$

Le terme entre parenthèse devient important uniquement quand la charge d'inversion devient petite. En négligeant Q_{inv} dans le terme entre parenthèses, la relation (3.8) se met sous la forme :

$$\begin{aligned} \frac{\partial V}{\partial \Psi_s} &\approx 1 + \phi_T \frac{C_{ox}}{Q_{inv}} \left(\frac{\partial V_{ox}}{\partial \Psi_s} + \frac{\gamma^2 C_{ox} \left[1 - \exp\left(-\frac{\Psi_s}{\phi_T}\right) \right]}{2Q_{dep}} \right) \\ &\approx 1 + \phi_T \frac{C_{ox}}{Q_{inv}} \left(\frac{\partial V_{ox}}{\partial \Psi_s} - \frac{1}{C_{ox}} \frac{\partial Q_{dep}}{\partial \Psi_s} \right) = 1 - \frac{\phi_T}{Q_{inv}} \frac{\partial Q_{inv}}{\partial \Psi_s} \end{aligned} \quad (3.9)$$

Au final, en remplaçant le terme $\frac{dV}{dx}$ par $\frac{dV}{d\Psi_s} \frac{d\Psi_s}{dx}$ dans (3.3), on obtient l'équation

de dérive diffusion sous l'approximation de la feuille de charge :

$$I(x) = -\mu_n W Q_{inv}(x) \frac{\partial \Psi_s}{\partial x} + \mu_n \phi_T \frac{\partial Q_{inv}}{\partial x} = I_{dérive} + I_{diffusion} \quad (3.10)$$

Cette équation classique montre que le courant drain est la somme de deux composantes, $I_{dérive}$ (courant de conduction) et $I_{diffusion}$ (courant de diffusion). Dans le cas où une seule composante domine le courant total, on peut établir les formulation analytiques classiques du courant de drain.

En régime de faible inversion, le courant drain est dominé par le courant de diffusion. La mobilité étant constante dans ce régime de fonctionnement, le courant drain est obtenu par l'intégration de l'équation (3.10) :

$$\int_0^L I_{diffusion} dx = \mu_n W \phi_T \int_{\Psi_{S0}}^{\Psi_{SL}} dQ_{inv} \quad (3.11)$$

où Ψ_{s0} et Ψ_{sL} représentent la valeur du potentiel de surface aux extrémités du canal, respectivement en $x=0$ et $x=L$.

Compte tenu de l'équation de continuité du courant le long du canal, l'utilisation de la forme (3.5)¹⁵ pour exprimer la charge d'inversion permet d'obtenir :

$$I_{diffusion} = \mu_n \frac{W}{L} \phi_T [Q_{inv}(x=L) - Q_{inv}(x=0)] \quad (3.12)$$

Le courant de conduction est calculé suivant l'intégrale :

$$\int_0^L I_{dérive} dx = -W \int_{\Psi_{s0}}^{\Psi_{sL}} \mu_n(x) Q_{inv}(x) d\Psi_s \quad (3.13)$$

En considérant que la mobilité ne varie pas fortement le long du canal, le courant drain de conduction s'exprime suivant la relation :

$$I_{dérive} = \mu_n C_{ox} \frac{W}{L} \left[(V_{GB} - V_{FB})(\Psi_{sL} - \Psi_{s0}) - \frac{1}{2}(\Psi_{sL}^2 - \Psi_{s0}^2) - \frac{2}{3} \gamma \left(\Psi_{sL}^{\frac{3}{2}} - \Psi_{s0}^{\frac{3}{2}} \right) \right] \quad (3.14)$$

L'expression totale du courant drain est donc la somme des composantes de dérive et de diffusion définies par les équations (3.12) et (3.14). La figure 3.2 présente le courant drain calculé par ce modèle pour un transistor nMOS, avec $W/L = 10 \mu\text{m}/10 \mu\text{m}$, $T_{ox} = 2 \text{ nm}$, $N_A = 5 \cdot 10^{17} \text{ cm}^{-3}$ et $N_P = 5 \cdot 10^{19} \text{ cm}^{-3}$.

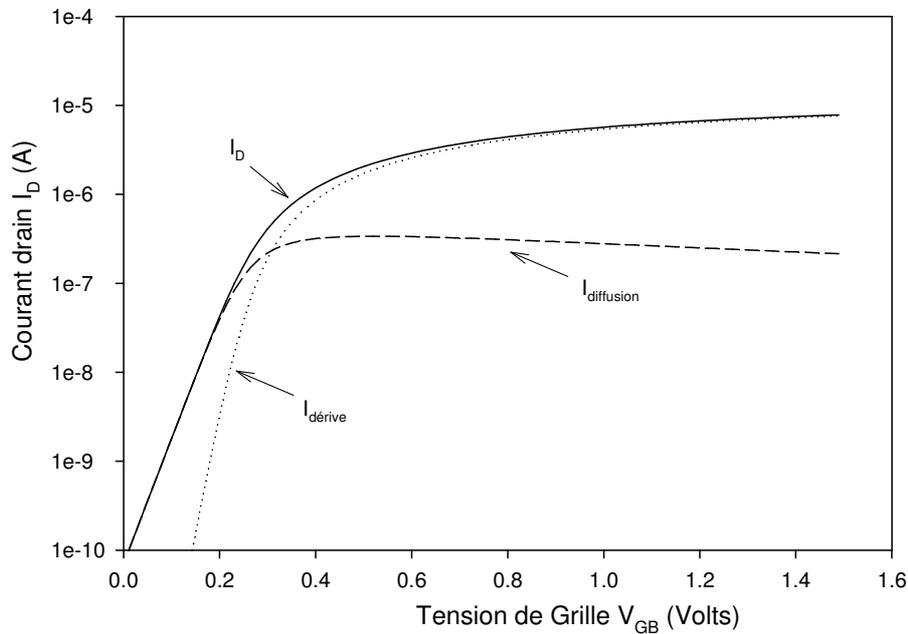


Figure 3.2 : Courant drain I_D calculé par l'approximation de la feuille de charge et du modèle de transport dérive-diffusion

¹⁵ On néglige dans cette approche le terme Ψ_p dans l'intégration de la charge d'inversion.

L'intégration analytique du courant drain n'a pu être possible qu'en considérant que la mobilité ne dépend pas de la position du canal. En réalité, elle dépend fortement des champs électriques E_x et E_y . Nous allons détailler dans le paragraphe suivant l'impact du champ électrique sur la mobilité et sur le courant drain.

3.3 Mobilité de la couche d'inversion

La mobilité des porteurs libres dans la couche d'inversion dépend fortement du champ électrique transverse E_x et du champ électrique longitudinal E_y . Le phénomène de dépendance de la mobilité au champ transverse est connu sous le nom de « réduction de mobilité », alors que le phénomène de dépendance de la mobilité au champ longitudinal est appelé « saturation de la vitesse des porteurs ».

3.3.1 Réduction de la mobilité

Dans les structures MOS, le champ électrique transverse confine les porteurs sur une fine épaisseur, et des phénomènes de dispersion provoquent une dépendance de la mobilité aux conditions de polarisation. La littérature est très abondante sur le sujet, et il apparaît expérimentalement que la mobilité des couches d'inversion suit une loi universelle indépendante de la polarisation substrat V_{SB} , de la concentration N_A ou de l'épaisseur d'oxyde T_{ox} quand elle est tracée en fonction du champ effectif E_{eff} , défini par :

$$E_{eff} = -\frac{Q_{dep} + \eta Q_{inv}}{\epsilon_{Si}} \quad (3.15)$$

où le terme empirique η est égal à $1/2$ pour les électrons et $1/3$ pour les trous dans le cas d'une surface orientée (100). Plus généralement, ce paramètre empirique est lié au processus de fabrication [KRUTSICK'88], à la température [HUANG'90], et à l'orientation cristalline [TAKAGI'94b].

Nous utiliserons pour notre modèle, une approche semi empirique de la mobilité, basée sur 3 mécanismes majeurs intervenant dans la dépendance de la mobilité au champ transverse.

L'interaction des porteurs avec les charges fixes (du même type que les porteurs libres) près de l'interface Si-SiO₂ engendre une répulsion coulombienne. Cet effet est d'autant plus important que la charge d'inversion est faible, et que la densité surfacique de charges fixes ou le dopage substrat sont forts. Cet effet devient négligeable en régime de forte inversion à cause de l'effet d'écran des porteurs libres. Empiriquement, l'interaction Coulombienne avec les impuretés s'exprime sous la forme [KONDO'99] :

$$\mu_C \propto -\frac{Q_{inv}}{N_A} \quad (3.16)$$

Le modèle de mobilité universel ne tient pas compte de l'interaction coulombienne. De plus, ce terme étant prépondérant uniquement en faible inversion, nous le négligerons pour la modélisation du courant drain.

Les phonons issus des vibrations du réseau cristallin interagissent avec les porteurs libres. Dans l'hypothèse que les porteurs de la couche d'inversion occupent la sous bande de plus basse énergie, la mobilité déterminée par les interactions avec les phonons acoustiques est donnée par [LEE'91] :

$$\mu_{ph} \propto \left(-\frac{11/32 Q_{inv} + Q_{dep}}{\epsilon_{Si}} \right)^{-\frac{1}{3}} \quad (3.17)$$

D'un point de vu expérimental, les électrons et les trous suivent la loi empirique suivante :

$$\mu_{ph} \propto E_{eff}^{-\frac{1}{3}} \quad (3.18)$$

Le dernier effet important à prendre en compte est l'interaction avec la rugosité de surface. Cet effet est prépondérant dans le régime de forte inversion. Pour les électrons, il a été montré expérimentalement que la mobilité limitée par l'interaction avec la rugosité de surface a la forme suivante [TAKAGI'94b] :

$$\mu_{SR} \propto E_{eff}^{-2} \quad (3.19)$$

Pour les trous, la mobilité prend la forme suivante :

$$\mu_{SR} \propto E_{eff}^{-1} \quad (3.20)$$

La différence de dépendance de la mobilité des électrons et des trous s'explique par le fait que le confinement des trous est moins fort que celui des électrons. La charge de trous est donc globalement plus loin de l'interface que celle constitué d'électrons, et la dépendance au champ transverse de la mobilité des trous est plus faible que celle des électrons. L'influence des différents mécanismes sont représentés sur la figure 3.3.

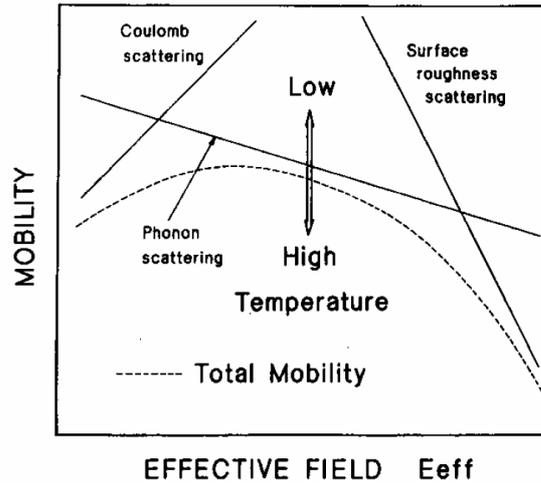


Figure 3.3 : Variation de la mobilité en fonction du champ électrique effectif d'après [TAKAGI'94a]

Les mobilités liées à chaque mécanisme sont incorporées dans un terme de mobilité globale suivant la loi de Matthiessen :

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_0} + \frac{1}{\mu_{SR}} + \frac{1}{\mu_{ph}} \quad (3.21)$$

Où μ_0 est la mobilité limitée par les impuretés ionisées. On obtient donc, au final, la mobilité effective de la couche d'inversion en fonction du champ effectif transverse [LANGEVELDE'97]:

$$\mu_{eff} = \frac{\mu_0}{1 + \left[(\theta_{ph}^* E_{eff})^{v/3} + (\theta_{SR}^* E_{eff})^{2v} \right]^{1/v}} \quad \text{pour les électrons} \quad (3.22)$$

$$\text{et } \mu_{eff} = \frac{\mu_0}{\left[1 + (\theta_{ph}^* E_{eff})^{v/3} + (\theta_{SR}^* E_{eff})^v \right]^{1/v}} \quad \text{pour les trous} \quad (3.23)$$

Pour ces deux équations, $v = 2$ à température ambiante.

L'implantation de ce modèle au sein de l'équation du courant drain (3.13) ne conduit pas à un résultat analytique. La charge d'inversion et de déplétion ne sont pas constantes le long du canal, donc la mobilité varie le long du canal. Cependant, en prenant la mobilité moyenne correspondant au potentiel de surface moyen $\overline{\Psi_s} = \frac{\Psi_{s0} + \Psi_{sL}}{2}$, on obtient des résultats satisfaisants. Nous montrerons par la suite l'impact réel de cette approximation quand nous développerons un modèle fournissant le potentiel de surface le long du canal.

3.3.2 Saturation de la vitesse des porteurs

La vitesse de dérive des porteurs n'est proportionnelle au champ électrique longitudinal qu'en dessous d'une valeur critique E_{sat} au delà de laquelle celle-ci atteint une valeur limite de l'ordre de $v_{sat}=10^7$ cm/s [MULLER'86]. Ce mécanisme peut provoquer la saturation du courant avant le pincement du canal pour des transistors suffisamment courts. La figure 3.4 illustre le mécanisme de saturation de la vitesse des porteurs dans le silicium.

Le modèle de vitesse des porteurs le plus simple, couramment utilisé en modélisation compacte, est le suivant [JACOBONI'77] :

$$v = \frac{\mu_{eff} \frac{\partial \Psi_s}{\partial x}}{1 + \frac{\mu_{eff}}{v_{sat}} \left| \frac{\partial \Psi_s}{\partial x} \right|} \quad (3.24)$$

Dans le cas des transistors avec un canal ultracourt, la longueur du canal est comparable au libre parcours moyen des porteurs. Le transport est dit quasi balistique et la vitesse des porteurs peut être supérieure à la vitesse de saturation v_{sat} . On parle alors de phénomène de survitesse (*velocity overshoot*) [SIM'95].

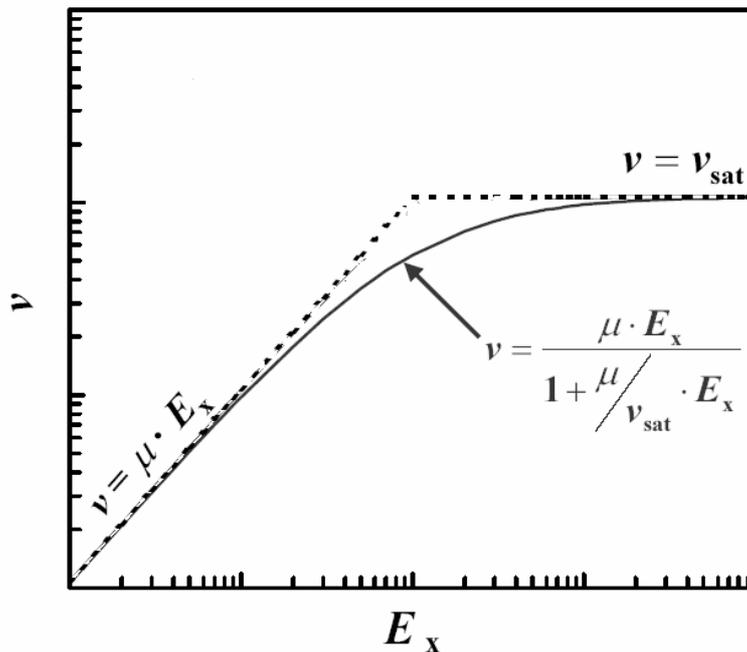


Figure 3.4 : Illustration de la saturation de la vitesse de dérive des porteurs dans le silicium d'après [LANGEVELDE'03]

3.4 Modélisation du potentiel de surface du transistor MOS

3.4.1 Modélisation quasi-statique sans courant de grille

Dans le cadre de l'approximation du canal graduel, nous allons définir une équation différentielle définissant le potentiel de surface le long du canal. En combinant l'équation de dérive diffusion (3.10) et l'équation de continuité $\frac{dI(x)}{dx} = 0$, on obtient l'équation suivante :

$$\frac{d}{dx} \left[-\mu_{eff}(x)W Q_{inv}(x) \frac{d\Psi_s(x)}{dx} + \mu_{eff}(x)\phi_T W \frac{dQ_{inv}(x)}{dx} \right] = 0 \quad (3.25)$$

Dans le but de simplifier les calculs, posons la variable Q_{inv}^* définie par :

$$Q_{inv}^* = Q_{inv} - \phi_T \frac{\partial Q_{inv}}{\partial \Psi_s} \quad (3.26)$$

L'équation (3.25) s'écrit donc :

$$\frac{d}{dx} \left[\mu_{eff}(x)W Q_{inv}^*(x) \frac{d\Psi_s(x)}{dx} \right] = 0 \quad (3.27)$$

Le développement de cette équation permet d'obtenir une équation différentielle du deuxième ordre non linéaire où les coefficients dépendent de l'inconnue $\Psi_s(x)$:

$$\left[\mu_{eff}(\Psi_s) \frac{\partial Q_{inv}^*}{\partial \Psi_s} + \frac{\partial \mu_{eff}}{\partial \Psi_s} Q_{inv}^*(\Psi_s) \right] \left(\frac{\partial \Psi_s}{\partial x} \right)^2 + \mu_{eff}(\Psi_s) Q_{inv}^* \frac{\partial^2 \Psi_s}{\partial x^2} = 0 \quad (3.28)$$

Tous les termes intervenant dans les coefficients de cette équation différentielle sont explicites du potentiel de surface. Le modèle de mobilité effective est lié au champ effectif, qui s'exprime explicitement en fonction du potentiel de surface. La dérivée de la mobilité en fonction du potentiel de surface s'exprime également en fonction du potentiel de surface. Puisque Q_{inv}^* est une fonction explicite du potentiel de surface, sa dérivée l'est aussi.

Les conditions aux limites de l'équation différentielle (3.28) sont données par :

$$\begin{cases} \Psi_s(x=0) = \Psi_{s0} \\ \Psi_s(x=L) = \Psi_{sL} \end{cases} \quad (3.29)$$

Les potentiels de surface Ψ_{s0} et Ψ_{sL} sont calculés en utilisant le modèle développé au chapitre 2, en prenant $V=V_S$ pour $x=0$ (Source) et $V=V_D$ pour $x=L$ (Drain).

L'équation (3.28) n'ayant pas de solution analytique, nous avons opté pour une résolution numérique matricielle [NOUGIER'87]. La résolution numérique de cette équation est présentée en annexe C.

La figure 3.5 présente le potentiel de surface le long du canal pour un transistor nMOS, avec $W/L = 10\mu m/10\mu m$, $T_{ox}=2\text{ nm}$, $N_A = 5 \cdot 10^{17}\text{ cm}^{-3}$ et $N_p = 5 \cdot 10^{19}\text{ cm}^{-3}$. La mobilité a été modélisée suivant le modèle de mobilité universel [SCHRODER'98] :

$$\mu_{eff} = \frac{638}{1 + \left(\frac{E_{eff}}{710^5}\right)^{1.69}} \quad [\text{cm}^2 \text{V}^{-1} \text{s}^{-1}] \quad (3.30)$$

Le courant drain est obtenu en calculant $I(x=L)$ suivant l'équation :

$$I_D = I(x=L) = -\mu_{eff}(\Psi_{SL}) Q_{inv}^*(\Psi_{SL}) \frac{\partial \Psi_s}{\partial x} \Big|_{x=L} \quad (3.31)$$

La figure 3.6 présente le courant drain calculé par les équations (3.14) et (3.28). La différence réside dans la prise en compte de la dégradation de mobilité dans le modèle de courant. L'utilisation de l'équation (3.14) impose l'emploi d'une mobilité moyenne pour tout le canal, alors que l'approche par l'équation (3.28) intègre une mobilité dépendant du potentiel de surface et de la variation de mobilité le long du canal.

On peut donc constater que cette approximation ne change absolument pas la valeur du courant drain en régime linéaire. Cela est dû au fait que la variation de mobilité le long du canal est quasiment linéaire et que l'on peut utiliser la valeur correspondant au potentiel de surface moyen.

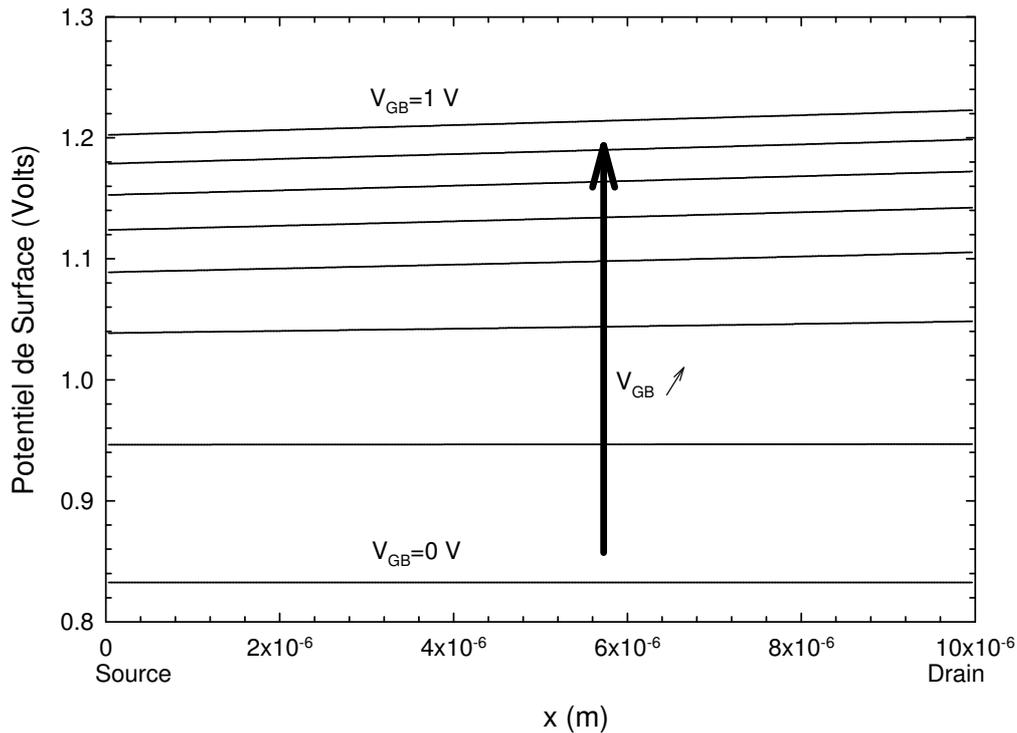


Figure 3.5 : Evolution du potentiel de surface le long du canal pour plusieurs polarisations de grille de 0 V à 1 V pour un transistor nMOS, avec $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, $T_{\text{ox}}=2 \text{ nm}$, $N_A= 5 \cdot 10^{17} \text{ cm}^{-3}$ et $N_p=5 \cdot 10^{19} \text{ cm}^{-3}$.

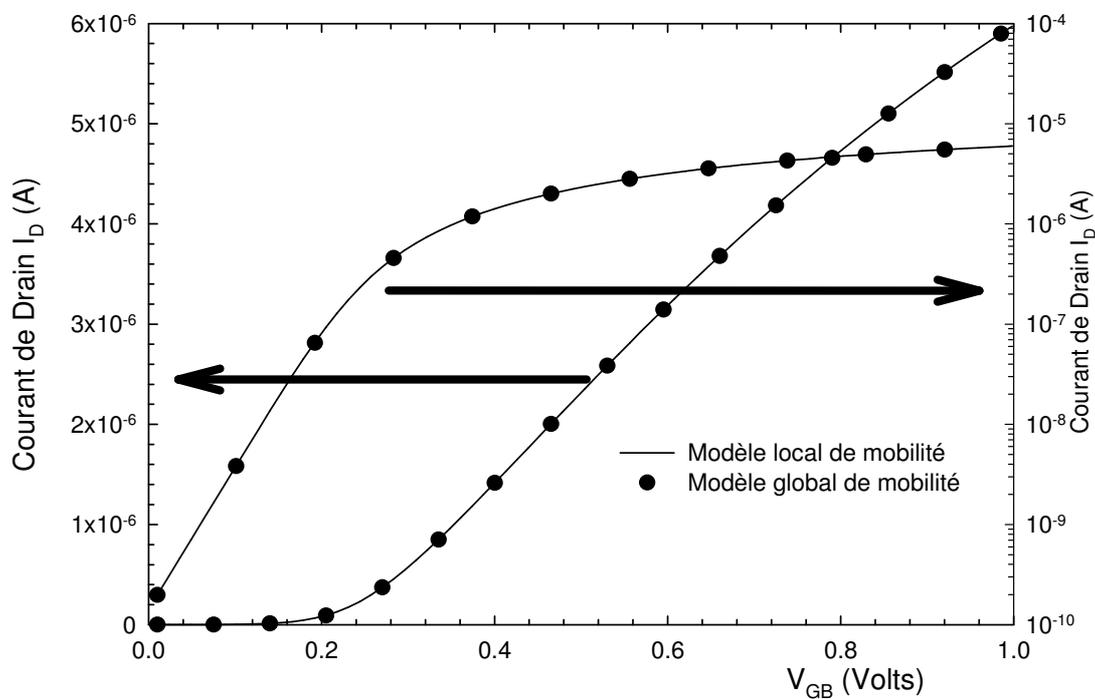


Figure 3.6 : Courant drain simulé avec une mobilité calculée pour le potentiel de surface moyen (modèle global) et le modèle local en fonction du champ transverse le long du canal pour un transistor nMOS, avec $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, $T_{\text{ox}}=2 \text{ nm}$, $N_A= 5 \cdot 10^{17} \text{ cm}^{-3}$ et $N_p=5 \cdot 10^{19} \text{ cm}^{-3}$.

3.4.2 Influence des effets quantiques et du phénomène de polydépétion de grille sur le courant drain.

Dans le chapitre 2, nous avons présenté la modélisation unidimensionnelle des effets quantiques et de la polydépétion sur le potentiel de surface. Le modèle de transistor MOS en feuille de charge et l'approximation du canal graduel restent valables en présence d'effets quantiques et de polydépétion. En effet, nous avons montré que l'expression de la charge d'inversion sous l'hypothèse de la feuille de charge reste valable (voir paragraphe 2.6.4.3). De plus, la polydépétion est incluse dans cette même expression. Au final, les équations (3.28) et (3.31) restent valables.

La figure 3.7 montre sur une échelle linéaire trois simulations pour une même structure nMOS, avec une épaisseur d'oxyde de 1.2 nm, un substrat de type p dopé uniformément à 10^{18} cm^{-3} et une grille polysilicium dopée à 10^{20} cm^{-3} . La mobilité est modélisée suivant le modèle de mobilité universel (3.30).

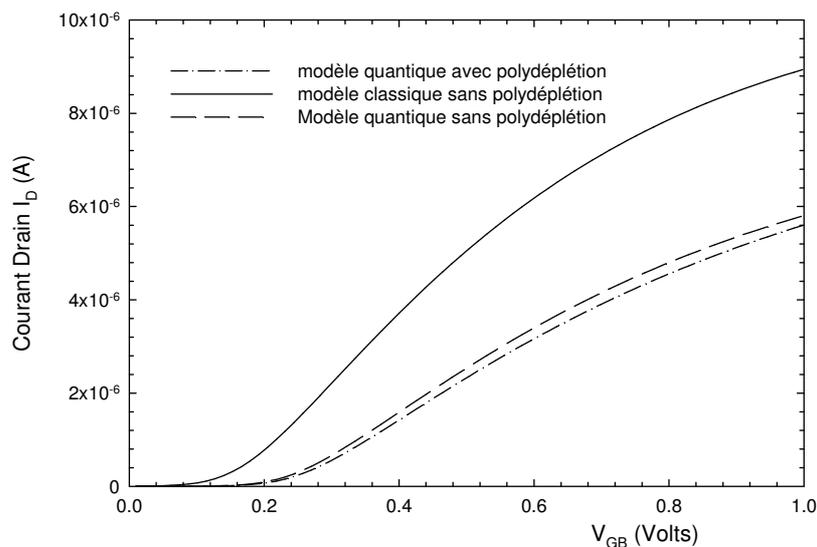


Figure 3.7 : Influence des effets quantiques et de la polydépétion sur le courant drain, pour un transistor nMOS, avec $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, $T_{\text{ox}} = 2 \text{ nm}$, $N_A = 5 \cdot 10^{17} \text{ cm}^{-3}$ et $N_p = 5 \cdot 10^{19} \text{ cm}^{-3}$.

La première simulation utilise un modèle classique pour le calcul du potentiel de surface et le phénomène de polydépétion n'est pas pris en compte. La deuxième simulation inclut les effets quantiques sans polydépétion. Enfin, la troisième simulation prend en compte les effets quantiques et la polydépétion. La comparaison de ces trois modèles permet de confirmer la très forte influence de tous ces effets liés à la miniaturisation. Les effets quantiques augmentent la tension de seuil de façon significative. De plus, le champ électrique transverse moyen dans le canal est globalement plus faible en présence d'effets quantiques

(voir figure 3.8). La mobilité est donc moins dégradée avec l'augmentation de la tension de grille.

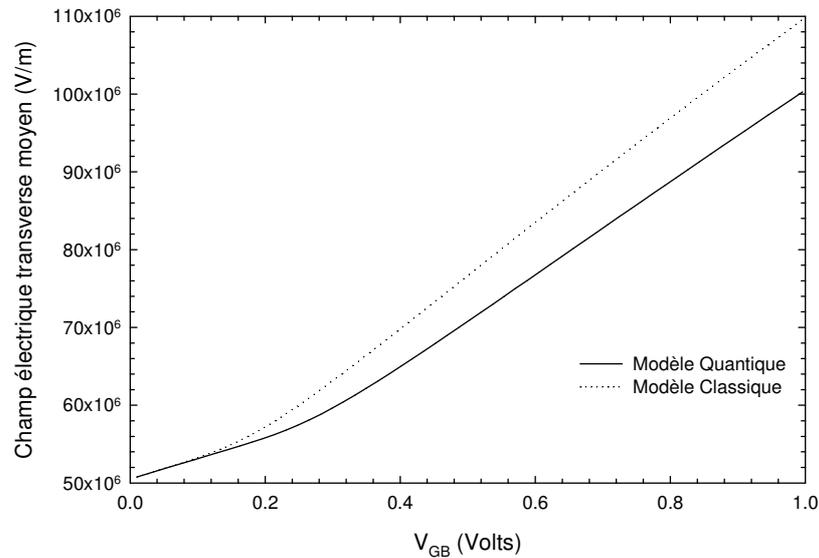


Figure 3.8 : Influence des effets quantiques sur le champ électrique transverse, pour un transistor nMOS, avec $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, $T_{\text{ox}}=2 \text{ nm}$, $N_A=5 \cdot 10^{17} \text{ cm}^{-3}$ et $N_p=5 \cdot 10^{19} \text{ cm}^{-3}$.

La polydépétion affecte légèrement la tension de seuil ; cet effet joue un rôle important uniquement si le dopage du polysilicium de grille est très faible. La figure 3.9 montre l'impact pour différents dopages. L'impact est donc nettement moins important que pour la simulation de la capacité de la structure pour laquelle leur prise en compte est indispensable.

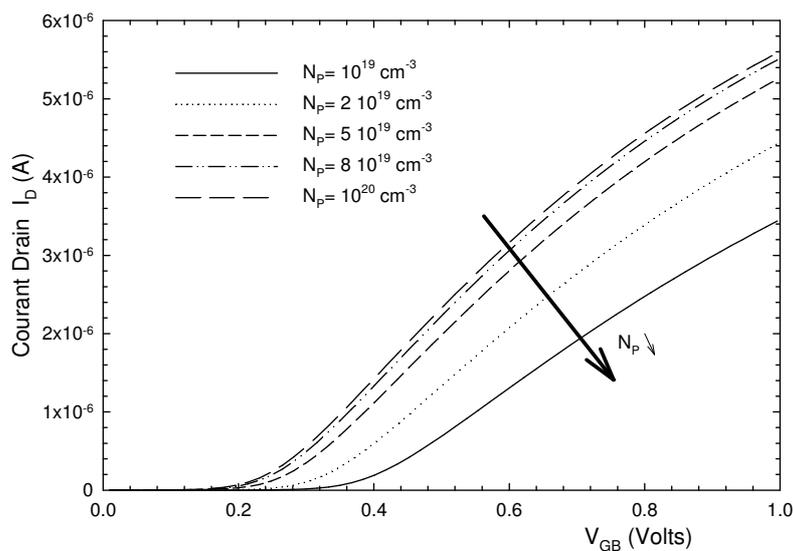


Figure 3.9 : Influence du dopage du polysilicium sur le courant de drain, pour un transistor nMOS, avec $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, $T_{\text{ox}}=2 \text{ nm}$, $N_A=5 \cdot 10^{17} \text{ cm}^{-3}$ et $N_p=5 \cdot 10^{19} \text{ cm}^{-3}$.

3.5 Modélisation du potentiel de surface avec fuite de grille

3.5.1 Formulation du modèle quasi statique

Dans le but de coupler l'équation du courant tunnel direct à l'équation différentielle (3.28) régissant le potentiel de surface le long du canal, nous devons reformuler l'équation de continuité $\frac{dI(x)}{dx} = 0$. Un élément du canal est représenté par la figure 3.10. La charge entrant dans cet élément de volume pendant un temps dt est égale à :

$$q_{in} = -i dt \quad (3.32)$$

La charge sortant du volume s'exprime par :

$$q_{out} = -(i + di) dt - (J_G W dx) dt \quad (3.33)$$

La charge d'inversion présente dans le volume élémentaire correspond à la variation de charge $q_{in} - q_{out}$. On obtient donc :

$$dQ_{inv} = \frac{dq}{W dx} = \frac{di dt}{W dx} + \frac{J_G W dx dt}{W dx} \quad (3.34)$$

L'équation de continuité est une équation aux dérivées partielles donnée par :

$$\frac{di}{dx} = -W J_G(x) + W \frac{dQ_{inv}}{dt} \quad (3.35)$$

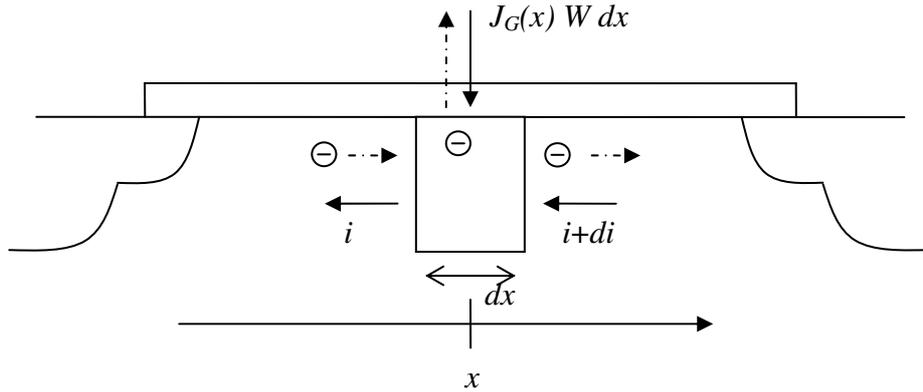


Figure 3.10 : Prise en compte de la fuite de la grille pour l'équation de continuité. Les flèches en trait plein représentent les courants et les flèches en trait pointillé le trajet réel des électrons.

Dans un premier temps, nous nous plaçons dans le cadre d'un calcul quasi statique, de sorte que l'équation de continuité se réduit à :

$$\frac{di}{dx} = -W J_G(x) \quad (3.36)$$

L'équation (3.28) se reformule sous la forme :

$$\left[\mu_{eff}(\Psi_S) \frac{\partial Q_{inv}^*}{\partial \Psi_S} + \frac{\partial \mu_{eff}}{\partial \Psi_S} Q_{inv}^*(\Psi_S) \right] \left(\frac{\partial \Psi_S}{\partial x} \right)^2 + \mu_{eff}(\Psi_S) Q_{inv}^* \frac{\partial^2 \Psi_S}{\partial x^2} = W J_G(\Psi_S) \quad (3.37)$$

Il faut noter que pour cette équation, la tension de grille V_{GB} est constante. Toutes les grandeurs intervenant dans cette équation dépendent de V_{GB} et Ψ_S . Dans le but d'alléger l'écriture, nous n'avons pas fait apparaître la tension V_{GB} .

Cette équation constitue la base de notre approche des structures MOS à deux dimensions. Elle tient compte à la fois des effets quantiques et de la polydéplétion de grille.

La résolution numérique est identique à celle de l'équation (3.28). Pour étudier l'impact du courant de grille sur le transistor MOS, nous allons considérer une structure nMOS avec les paramètres technologiques d'une génération 65 nm [ITRS'03]. Les paramètres technologiques utilisés pour les simulations dans toute cette partie sont reportés dans le tableau 3.1. Dans un premier temps, nous allons détailler l'impact du courant de grille pour une polarisation $V_{DS} = 0$ Volt. Nous présenterons ensuite l'impact du courant de grille sur les caractéristiques statiques du transistor MOS polarisé en régime linéaire.

T_{ox}	1.2 nm
N_A	10^{18} cm^{-3}
N_P	10^{20} cm^{-3}
L	$1 \mu\text{m} < L < 100 \mu\text{m}$
W	10 μm
χ_B	3.1 eV
m_{ox}^*	0.5 m_0
μ_{eff}	Mobilité universelle (3.30)

Tableau 3.1 : Paramètres des transistors nMOS utilisés pour l'évaluation de l'impact du courant tunnel sur les différentes grandeurs physiques du transistor.

3.5.2 Influence du courant de grille à $V_{DS} = 0$ V

3.5.2.1 Dépolarisation du canal

La figure 3.11 présente le potentiel de surface le long du canal pour plusieurs longueurs de grille à $V_{GB}=1$ V, la source et le drain étant polarisés à 0 V. On constate que le potentiel de surface est quasiment constant pour une longueur de $1 \mu\text{m}$, et une augmentation de la longueur induit une courbure du potentiel de surface. Cet effet a déjà été mis en évidence par plusieurs auteurs [GILIBERT'04, LANGEVELDE'01]. La modélisation de la dépolarisation du canal a été réalisée par la segmentation du canal, chaque segment étant simulé par un modèle compact. A notre connaissance, une seule publication a adopté une modélisation de type analytique [PALESTRI'03]; cependant seul le régime de forte inversion a été pris en compte¹⁶.

La polarisation $V_{DS} = 0$ V est traditionnellement dédiée à la mesure de capacité. Dans ces conditions, la mobilité des porteurs libres de la couche d'inversion n'est jamais prise en compte. Dans le cas où le courant de grille est pris en compte, l'équation (3.37) laisse supposer que la mobilité joue un rôle dans la courbure du potentiel de surface. La figure 3.12 présente le potentiel de surface le long du canal pour un transistor nMOS avec $L = 50 \mu\text{m}$ et différents modèles de mobilité (représentés sur l'axe de droite). La première courbe a été obtenue en prenant une mobilité constante égale à $500 \text{ cm}^2/\text{Vs}$ (courbe en trait plein). Pour la deuxième simulation, le modèle de mobilité universel a été utilisé (courbe en trait pointillé).

¹⁶ L'auteur a couplé numériquement l'équation de continuité avec l'équation du courant de dérive uniquement. Le modèle de courant de grille est totalement empirique et les effets quantiques sont pris en compte par une simple augmentation de la tension de seuil.

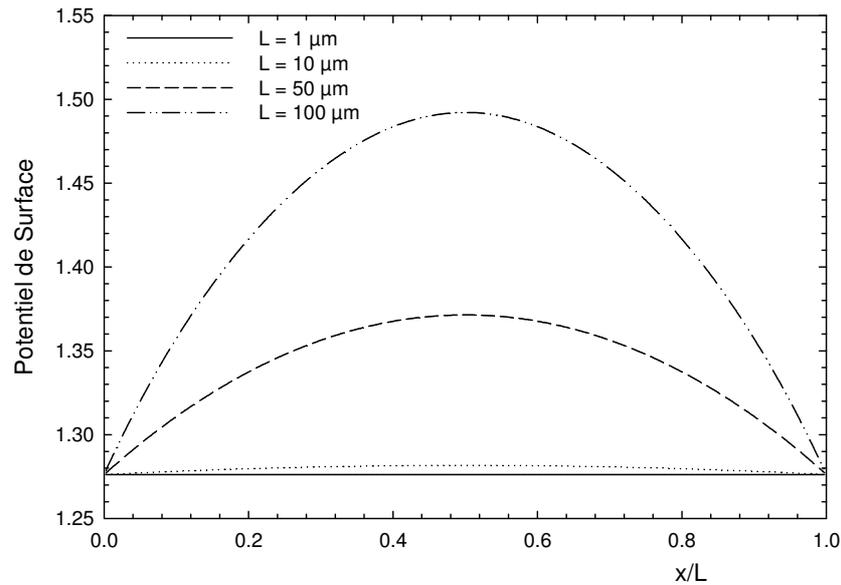


Figure 3.11 : Evolution du potentiel de surface avec la prise en compte du courant de grille pour différentes longueurs de transistor nMOS à $V_{GB}=1$ V, $T_{ox}=1.2$ nm.

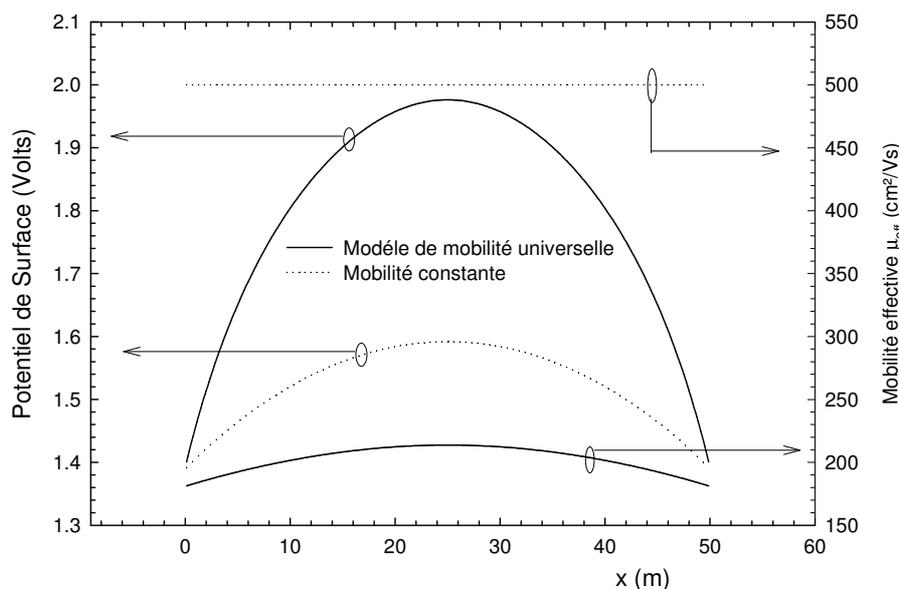


Figure 3.12 : Influence de la mobilité sur la courbure du potentiel de surface en présence de courant de fuite de grille pour un transistor nMOS de longueur 50 µm et $T_{ox}=1.2$ nm.

3.5.2.2 Impact sur la densité de courant de grille

La courbure du potentiel de surface est d'autant plus grande que la densité de courant de fuite est forte. Avec les paramètres technologiques définis dans le tableau 3.1, les transistors MOS de longueurs supérieures à 10 µm polarisés à tension de drain nulle ne peuvent plus être considérés comme des structures unidimensionnelles. En effet, comme toutes les grandeurs physiques du transistor (charge d'inversion, de déplétion, courant tunnel, cf. chapitre 2) dépendent du potentiel de surface, la prise en compte de cette courbure est

indispensable lors de la caractérisation de structures avec de grandes longueurs. La figure 3.13 présente la densité de courant le long du canal pour différentes longueurs. Les densités de courant de grille ont donc tendance à diminuer fortement au centre du canal pour les transistors très longs ($L > 10 \mu\text{m}$). La figure 3.14 présente l'évolution du courant de grille total normé à la surface en fonction de la longueur du canal. On observe une chute importante liée à la réduction de la densité de courant de grille au centre du canal avec une augmentation de la longueur.

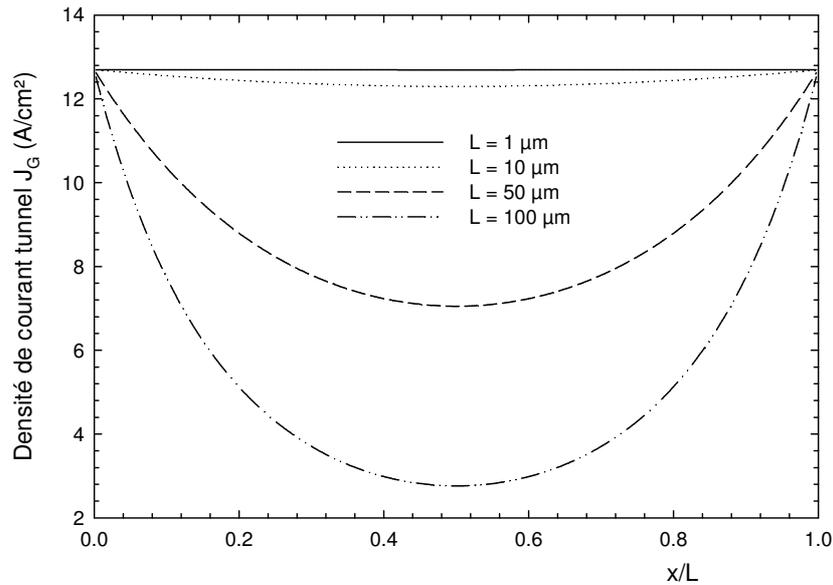


Figure 3.13 : Evolution de la densité de courant de grille le long du canal en fonction de la longueur du transistor à $V_{GB} = 1 \text{ V}$ pour des transistors nMOS de largeur $10 \mu\text{m}$ et $T_{\text{ox}} = 1.2 \text{ nm}$.

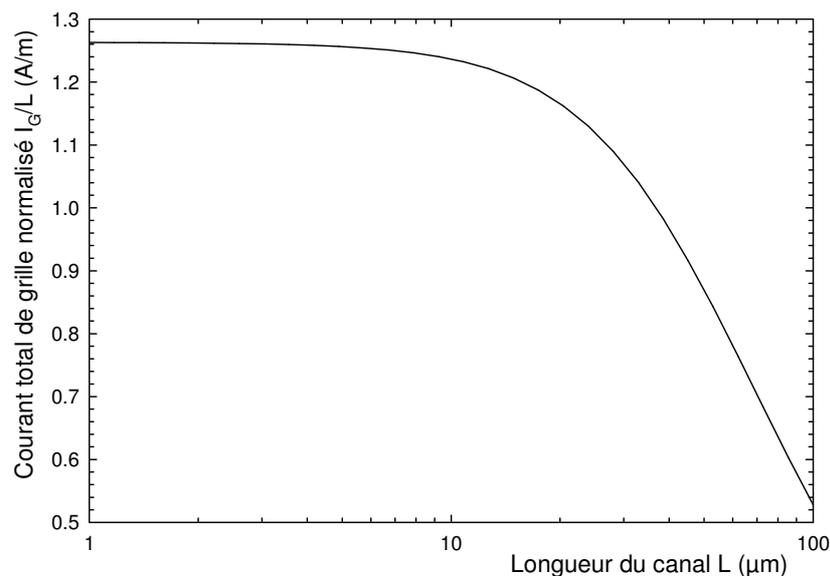


Figure 3.14 : Diminution du courant de grille total en fonction de la longueur du transistor à $V_{GB} = 1 \text{ V}$ pour des transistors nMOS de largeur $10 \mu\text{m}$ et $T_{\text{ox}} = 1.2 \text{ nm}$.

3.5.2.3 Impact sur la capacité quasi statique

La capacité quasi statique est présentée dans ce paragraphe dans le but mettre en évidence l'impact théorique du courant de grille sur la charge totale dans les transistors longs. En effet, il est absolument impossible de réaliser une mesure de capacité quasi statique avec un courant de fuite aussi fort. La mesure de capacité quasi statique est basée sur l'évaluation de la charge en fonction de la tension de grille. La méthode est basée sur l'intégration du courant transitoire à une rampe de tension V_{GB} ; le courant obtenu se calcule simplement par :

$$C_{gg} = \frac{I}{\frac{dV_{GB}}{dt}} \quad (3.38)$$

Malheureusement, en présence de courant tunnel direct, le courant I est la somme de ce courant transitoire (lié à la variation de charge dq/dt) et du courant tunnel direct. Le courant issu de la variation de charge est négligeable devant le courant tunnel direct. Les techniques de compensation de courant tunnel de fuite [AGILENT'01] dans les mesures quasi statiques ne permettent pas de surmonter ce problème avec des oxydes aussi fins.

La charge totale de la grille dépend fortement du potentiel de surface, et de ce fait, elle va être fortement perturbée par la courbure de celui ci. La figure 3.15 présente la capacité C_{gg} normée à la surface de la structure nMOS en régime de déplétion et d'inversion ($V_{GB} > V_{FB}$) décrit par les paramètres technologiques du tableau 3.1.

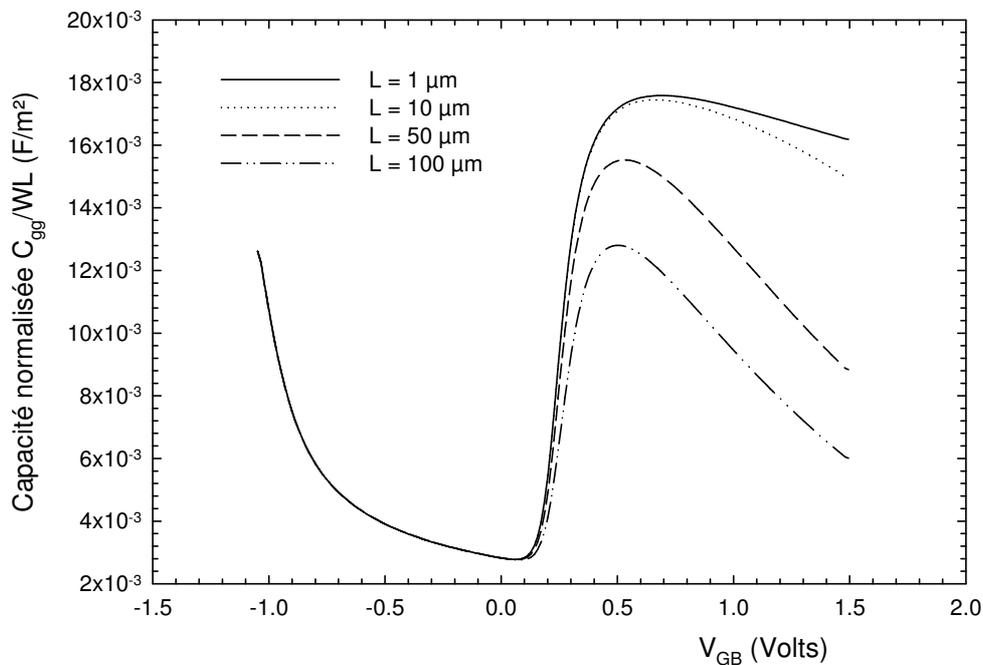


Figure 3.15 : Impact du courant de grille sur la capacité quasi statique C_{gg} en fonction de la longueur de grille pour des transistors nMOS de largeur 10 μm et $T_{ox}=1.2$ nm.

3.5.3 Impact du courant de grille sur le fonctionnement du transistor MOS en régime linéaire

3.5.3.1 Courant de grille total

Dans le but d'évaluer l'impact du couplage de l'équation de drift diffusion et l'équation de courant tunnel, nous avons réalisé deux simulations (figure 3.16) sur une structure nMOS ($L=10\ \mu\text{m}$) décrite dans le tableau 3.1 pour plusieurs valeurs de V_{DS} . La première simulation a été obtenue en calculant le potentiel de surface avec l'équation (3.28), puis en calculant le courant de grille de façon non couplée. La deuxième simulation utilise l'équation (3.37).

La différence entre les deux approches est très faible, et elle décroît quand V_{DS} augmente. La différence serait plus marquée avec des longueurs de transistors plus importantes ; cependant, nous allons voir au paragraphe suivant que les courants drain et source sont fortement perturbés par le courant de grille, rendant impossible la caractérisation en régime linéaire de composants très long. L'intérêt d'utiliser le modèle (3.37) réside donc dans la simulation des courants de drain et de source, qui sont fortement perturbés par la présence d'un courant de grille.

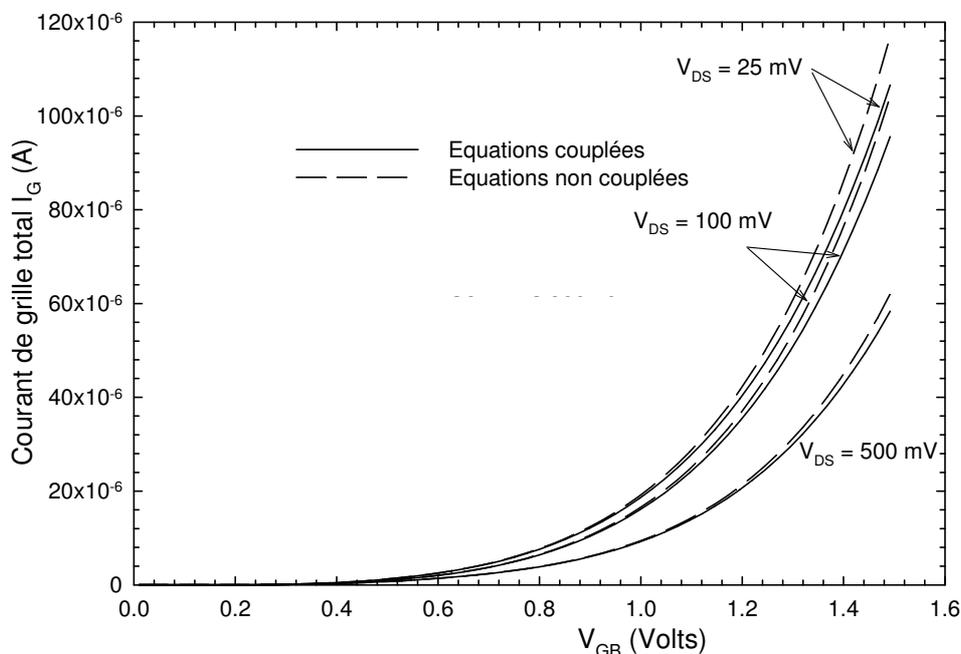


Figure 3.16 : Influence du modèle sur le courant de grille total en régime linéaire pour un transistor nMOS avec $W \times L = 10\ \mu\text{m} \times 10\ \mu\text{m}$ et $T_{ox}=1.2\ \text{nm}$.

3.5.3.2 Courants drain et source

Le dernier point de notre modèle statique à examiner est l'impact du courant de grille sur les courants de source et de drain.

Les courants de drain et de source sont fortement perturbés par le courant de grille. Sur la figure 3.17 sont reportées les simulations d'un transistor nMOS de longueur 10 μm , avec et sans la prise en compte du courant de grille, c'est-à-dire à l'aide respectivement de l'équation (3.37) et (3.28),.

Dans le cas où l'équation du courant de grille n'est pas couplée à l'équation de dérive diffusion, on obtient évidemment $I_S=I_D$ puisque le courant est le même dans tout le canal. On peut toutefois obtenir des résultats similaires aux courants calculés par l'équation (3.37) en calculant la part de courant de grille totale présente sur le courant de source et drain. Cette méthode est employée en modélisation compacte, en particulier dans le modèle MM11. En effet, nous avons pu voir que le potentiel de surface est faiblement courbé pour des longueurs de canal de 10 μm . Il en résulte que le potentiel de surface peut être calculé avec l'équation (3.28) sans faire trop d'erreur. Le courant de drain et de source s'expriment alors par[LANGEVELDE'01] :

$$\begin{aligned} I_D &= I_0 - W \int_0^L \frac{x}{L} J_G(x) dx \\ I_S &= I_0 + W \int_0^L \left[1 - \frac{x}{L}\right] J_G(x) dx \end{aligned} \quad (3.39)$$

où I_0 est le courant drain (et source) calculé par l'équation (3.28).

Le modèle défini par l'équation (3.37) prend en compte de façon intrinsèque le phénomène de partitionnement du courant de grille sur le courant de source et de drain. Le fait de pouvoir simuler tous les phénomènes physiques à partir d'une seule équation est un avantage majeur vis-à-vis de l'approche employée en modélisation compacte, où chaque phénomène nécessite une équation supplémentaire.

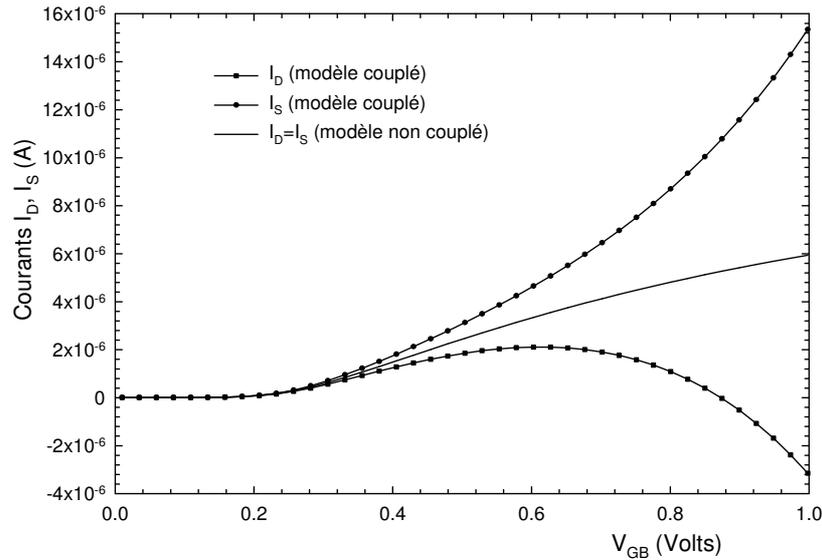


Figure 3.17 : Simulation des courants drain et source avec l'équation (3.37) (modèle couplé) et l'équation (3.28) (modèle non couplé), pour un transistor nMOS avec $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$ et $T_{\text{ox}}=1.2 \text{ nm}$.

On peut donc extraire le courant intrinsèque I_0 à partir d'une mesure de courant de drain fortement perturbée par le courant de fuite. Ce point est très important pour la caractérisation, notamment pour l'extraction de la mobilité par Split - CV [LIME'03].

Nous avons donc à notre disposition un modèle complet statique permettant de simuler les caractéristiques I_G-V_{GB} , I_D-V_{GB} , et I_S-V_{GB} en régime linéaire pour des transistors longs de technologies avancées. Dans la partie suivante, nous allons valider ce modèle en le confrontant aux mesures réalisées sur des transistors MOS issues de technologies avancées.

3.6 Etude expérimentale et validation du modèle

3.6.1 Présentation des technologies étudiées

Au cours de cette thèse, nous avons étudiés dans le cadre du projet européen ARTEMIS des transistors MOS sur silicium massif sub- $0.1 \mu\text{m}$ fabriqués par ST-microelectronics. L'étude concerne 3 lots différents de transistors n et p de longueur de grille ultra courtes et à oxydes ultrafins.

Le premier et le deuxième lot sont des technologies où l'oxyde a été réalisé par le procédé RTN (Rapid Thermal Nitridation), c'est-à-dire une oxydation faite sous atmosphère d'azote et d'oxygène à environ $900 \text{ }^\circ\text{C}$. Nous avons eu à notre disposition deux épaisseurs d'oxyde, à savoir 1.4 nm et 1.2 nm . Les transistors issus de cette technologie seront nommés par la suite respectivement RTN14 et RTN12.

Le troisième lot est différent des deux premiers car l'oxyde de grille, d'épaisseur 1.2 nm, a été réalisé par un procédé DPN (Decoupled Plasma Nitridation), c'est-à-dire un oxyde où la nitruration est assistée par plasma. Les transistors issus de cette technologie seront nommés par la suite DPN12.

Les transistors montés en batteries ont les électrodes de grille, de source et de substrat communes. Seule l'électrode de drain est indépendante. Le problème majeur avec ce type de structure est lié au courant de fuite de grille mesuré qui correspond à l'ensemble des fuites de tous les transistors. Le courant de grille mesuré est donc quasiment indépendant du transistor sélectionné. Seul le courant drain est représentatif du transistor que l'on souhaite mesurer. Les longueurs de grille physiques (mesurées par TEM) des transistors en batterie sont données dans le tableau 3.2. La largeur des transistors est de 10 μm .

<i>RTN12 & RTN14</i>		<i>DPN12</i>		
<i>0.055</i>	<i>1</i>	<i>0.03</i>	<i>0.12</i>	<i>0.34</i>
<i>0.085</i>	<i>5</i>	<i>0.04</i>	<i>0.14</i>	<i>0.54</i>
<i>0.105</i>	<i>10</i>	<i>0.05</i>	<i>0.16</i>	<i>0.74</i>
<i>0.125</i>		<i>0.06</i>	<i>0.18</i>	<i>1</i>
<i>0.145</i>		<i>0.07</i>	<i>0.2</i>	<i>2</i>
<i>0.185</i>		<i>0.08</i>	<i>0.22</i>	<i>5</i>
<i>0.285</i>		<i>0.09</i>	<i>0.26</i>	<i>10</i>
<i>0.465</i>		<i>0.1</i>	<i>0.3</i>	

Tableau 3.2 : Longueurs de grilles physiques (μm) pour les batteries de transistors à source et grille communes pour les lots RTN14, RTN12 et DPN12.

Des transistors totalement isolés sont également disponibles. Les électrodes de source, de drain, de grille et de substrat sont indépendantes. Les mesures de capacités sont effectuées sur ces structures, et les longueurs de grilles physiques disponibles sont reportées sur le tableau 3.3.

<i>RTN12 & RTN14</i>	<i>DPN12</i>
<i>0.055</i>	<i>0.04</i>
<i>0.15</i>	<i>0.06</i>
<i>0.285</i>	<i>0.15</i>
<i>10</i>	<i>1</i>
<i>-</i>	<i>10</i>

Tableau 3.3 : Longueurs de grilles physiques (μm) pour les transistors isolés pour les lots RTN14, RTN12 et DPN12.

Dans un premier temps, nous allons présenter les mesures statiques réalisées sur ces différents lots pour avoir une vue d'ensemble de ces différentes technologies afin d'illustrer certains points développés au chapitre 1, notamment l'influence des poches (*pockets*) sur les effets de canal court et l'effet canal court inverse. Nous nous focaliserons ensuite sur les transistors ayant des longueurs de grille supérieures à $1\ \mu\text{m}$ dans le but de valider les modèles développés au chapitre 2 et dans la première partie du chapitre 3.

3.6.2 Caractéristiques courant – tension I - V

3.6.2.1 Lot RTN14

Les caractéristiques I_G - V_{GS} à $V_{DS}=0\ \text{V}$ de transistors n et p MOS sont présentées respectivement sur la figure 3.18. La partie commune des courants de grille pour une tension de grille comprise entre 0 et V_{FB} correspond à la contribution des zones de recouvrement (*overlap*) de la grille sur la source et le drain. Dans cette zone, le courant est proportionnel à la largeur du transistor, qui est égale à $10\ \mu\text{m}$ avec cette série de transistors isolés.

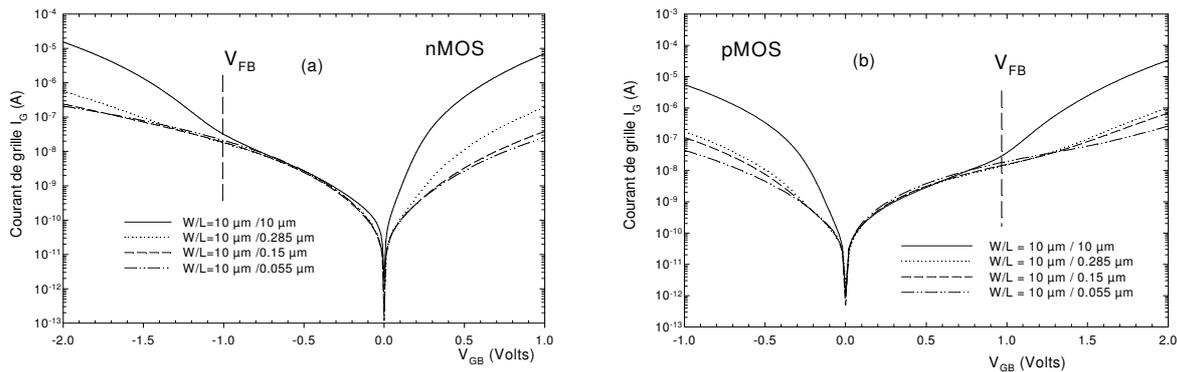


Figure 3.18: Courants de grille des transistors nMOS (a) et pMOS (b) isolés issus du lot RTN14.

Les caractéristiques I_D - V_{GS} à $V_{DS}=25\ \text{mV}$ de transistors n et p MOS sont présentées respectivement sur la figure 3.19 et la figure 3.20. Le transistor de longueur $55\ \text{nm}$ souffre de problèmes importants liés d'une part, aux effets de canal court mal maîtrisés, et d'autre part, des résistances d'accès limitant l'augmentation du courant avec la diminution de la longueur de grille. L'impact des effets de canal court induit une diminution de la tension de seuil et une dégradation de la pente sous le seuil. Cet effet est d'ailleurs plus marqué pour les transistors pMOS que les transistors nMOS. Les tensions de seuil de ces dispositifs sont reportées sur la figure 3.21. L'effet de canal court inverse, présent sur les deux types de transistors, est essentiellement dû au surdopage apporté par les poches qui devient prépondérant dans les transistors de faibles longueurs. Un paramètre à observer pour juger de la qualité des

transistors à canal court est la pente sous seuil. La figure 3.22 présente l'évolution de l'inverse de la pente sous le seuil (*Subthreshold swing*) en fonction de la longueur. Pour les transistors nMOS, ce paramètre augmente peu, alors que les transistors pMOS souffrent d'une augmentation importante. La répercussion directe de cette dégradation est une augmentation du courant drain à l'état bloqué I_{off} , augmentant ainsi la consommation statique.

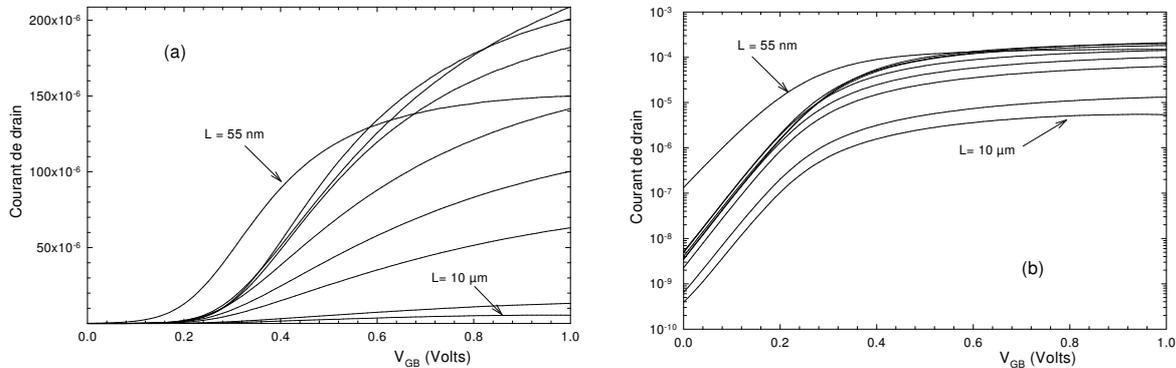


Figure 3.19 : Courant de drain en fonction de la tension de grille à $V_{DS}=25$ mV pour une batterie de transistors nMOS à source et grille communes, en échelle linéaire (a) et logarithmique (b).

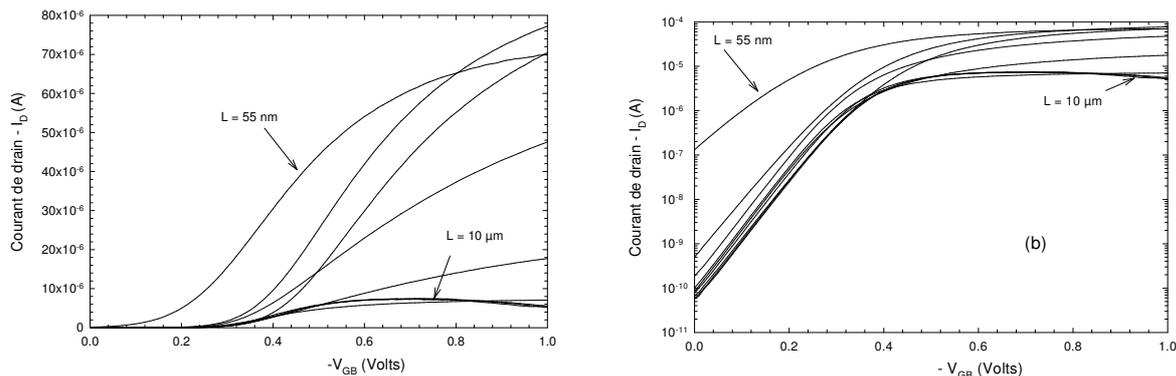


Figure 3.20 : Courant de drain en fonction de la tension de grille à $V_{DS}= - 25$ mV pour une batterie de transistors pMOS à source et grille communes, en échelle linéaire (a) et logarithmique (b).

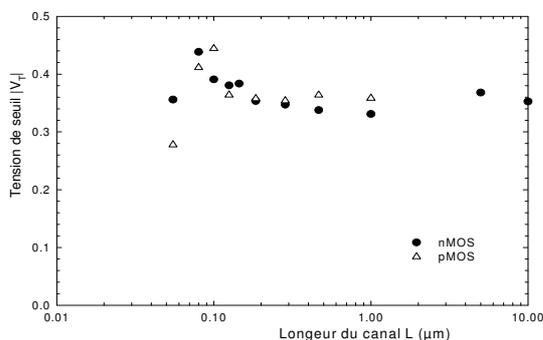


Figure 3.21 : Tension de seuil en fonction de la longueur du canal pour les transistors n et p MOS RTN14.

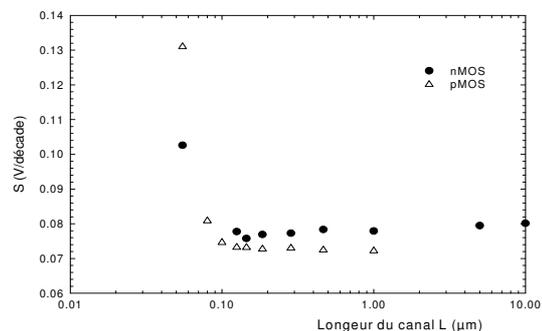


Figure 3.22 : Inverse de la pente sous le seuil S en fonction de la longueur du canal pour des transistors n et pMOS RTN14.

3.6.2.2 Lot RTN12

Les caractéristiques I_G - V_{GS} à $V_{DS}=0$ V de transistors n et p MOS sont présentées respectivement sur la Figure 3.23.

La partie commune des courants de grille pour une tension V_{GB} comprise entre 0 et V_{FB} correspond à la contribution des zones de recouvrement (*overlap*) de la grille sur la source et le drain. Dans cette zone, le courant est proportionnel à la largeur du transistor, qui est égale à $10\ \mu\text{m}$ avec cette série de transistors isolés.

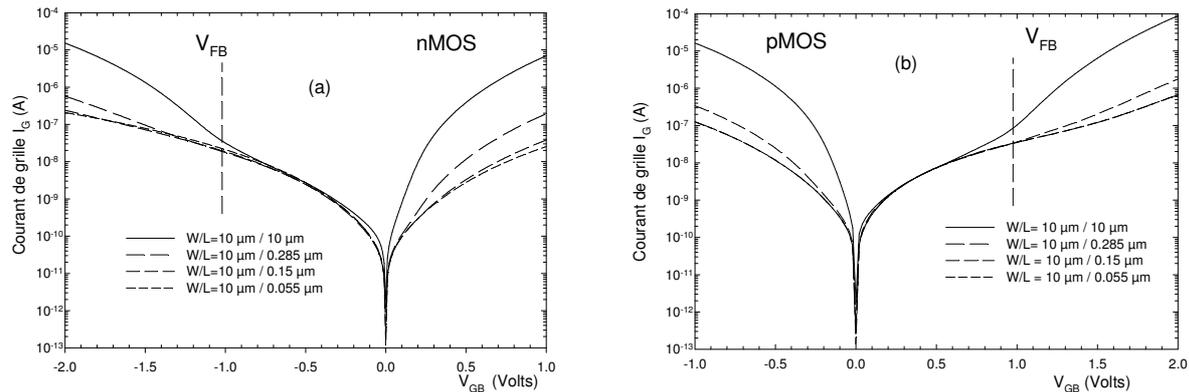


Figure 3.23: Courants de grille des transistors nMOS (a) et pMOS (b) isolés issus du lot RTN12.

Les caractéristiques I_D - V_{GS} à $V_{DS}=25$ mV de transistors n et p MOS sont présentées respectivement sur la figure 3.24 et figure 3.25. Les tensions de seuil de ces dispositifs sont reportées sur la figure 3.26. La figure 3.27 présente l'évolution de l'inverse de la pente sous le seuil en fonction de la longueur du canal.

Par rapport aux transistors RTN14, ce lot ne diffère que par l'épaisseur d'oxyde. Pour les transistors nMOS, on observe une nette amélioration pour le transistor de longueur 55 nm. La loi empirique liant la longueur du transistor à l'épaisseur du film diélectrique formulée par Intel [THOMPSON'98] se trouve donc vérifiée¹⁷.

Les effets de canal court restent mal maîtrisés pour le pMOS de longueur 55 nm, puisque la tension de seuil et la pente sous le seuil se dégradent fortement. En effet, il est technologiquement plus difficile d'optimiser un pMOS qu'un nMOS pour une longueur de grille donnée.

La technologie RTN12 donne donc des résultats satisfaisants pour les transistors nMOS jusqu'à une longueur de grille de 55 nm. L'implantation des poches de surdopage leur confère une bonne tenue aux effets de canal court, car la diminution de la tension de seuil est

¹⁷ Cette loi propose que $L = 45 \times T_{ox}$, soit $T_{ox} = 1.2$ nm pour $L = 55$ nm.

très faible et la pente sous le seuil reste proche de celle des transistors à canal long. Par contre, les transistors pMOS de longueur 55 nm ne sont pas satisfaisants en terme de tenue aux effets de canal court.

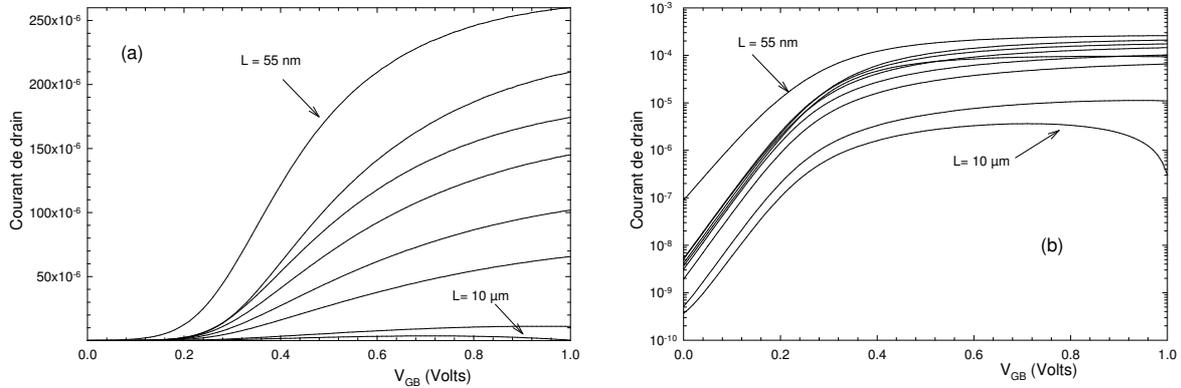


Figure 3.24 : Courants drain en fonction de la tension de grille à $V_{DS}=25$ mV pour une batterie de transistors nMOS RTN12 à source et grille communes, en échelle linéaire (a) et logarithmique (b).

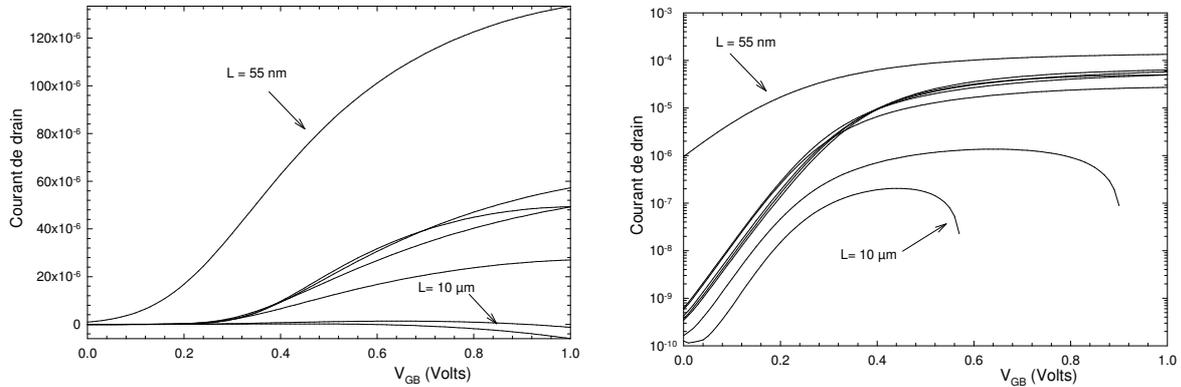


Figure 3.25 : Courants drain en fonction de la tension de grille à $V_{DS} = -25$ mV pour une batterie de transistors pMOS RTN12 à source et grille communes, en échelle linéaire (a) et logarithmique (b).

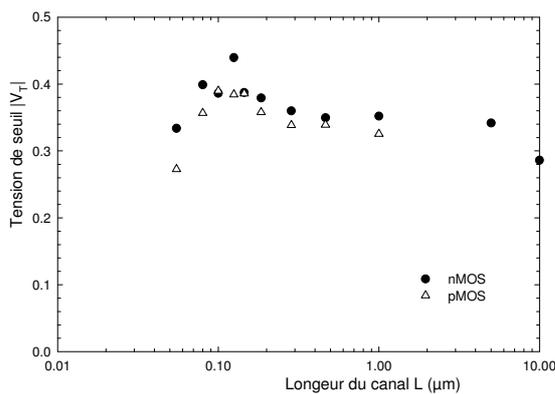


Figure 3.26 : Tensions de seuil en fonction de la longueur du canal pour les transistors n et p MOS RTN12.

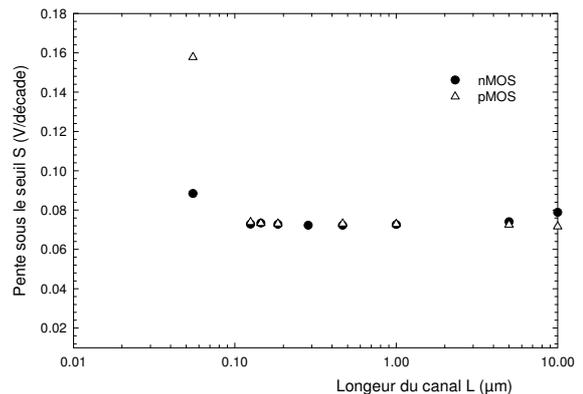


Figure 3.27 : Inverse de la pente sous le seuil S en fonction de la longueur du canal pour des transistors n et pMOS RTN12.

Les transistors du lot RTN12 à canal long souffrent du problème lié aux forts courants de fuite évoqués au paragraphe 3.5.3.2. Le fort courant de grille présent sur les transistors de longueur $10\ \mu\text{m}$ diminue fortement le courant de drain, au point qu'il peut devenir négatif pour les fortes polarisations de grille. Nous nous focaliserons sur la modélisation des courants des transistors longs au paragraphe 3.6.4.

3.6.2.3 Lot DPN12

Les caractéristiques I_G - V_{GS} à $V_{DS}=0\ \text{V}$ de transistors n et p MOS sont présentées sur la Figure 3.28. La partie commune des courants de grille pour une tension V_{GB} comprise entre 0 et V_{FB} correspond à la contribution des zones de recouvrement (*overlap*) de la grille sur la source et le drain. Dans cette zone, le courant est proportionnel à la largeur du transistor, qui est égale à $10\ \mu\text{m}$ avec cette série de transistors isolés.

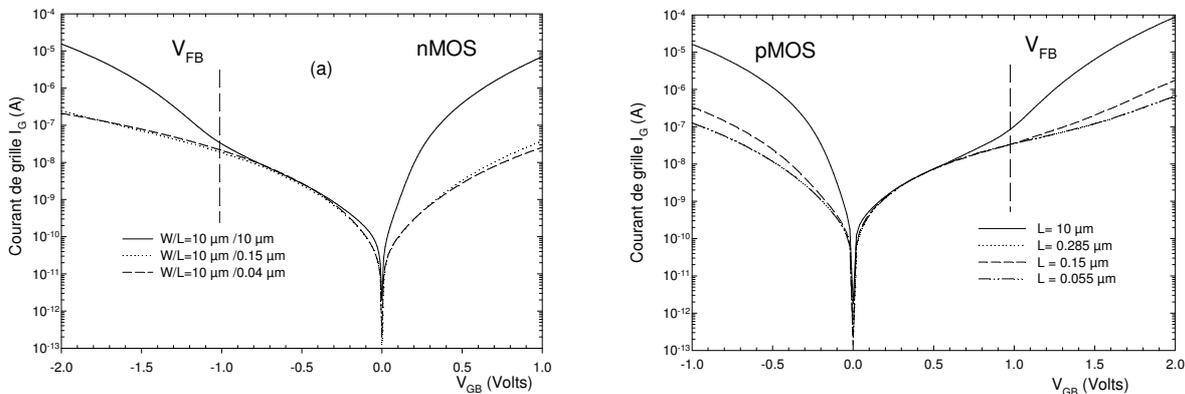


Figure 3.28: Courants de grille des transistors nMOS (a) et pMOS (b) isolés issus du lot DPN12.

Les caractéristiques I_D - V_{GS} à $V_{DS}=25\ \text{mV}$ de transistors n et p MOS sont présentées respectivement sur la figure 3.24 et figure 3.25. Les tensions de seuil de ces dispositifs sont reportées sur la figure 3.26, et la figure 3.27 présente l'évolution de l'inverse de la pente sous le seuil en fonction de la longueur.

D'une manière générale, on constate que cette technologie de fabrication permet d'obtenir des transistors n et pMOS de longueur de grille de $30\ \text{nm}$ performants. Les effets de canal court sont maîtrisés pour les deux types de transistors. La pente sous le seuil ne se dégrade que très faiblement, et de façon similaire pour les transistors n et pMOS. Cependant on observe un fort effet de canal court inverse pour les pMOS. Ceci est lié à l'utilisation de poches fortement dopées. L'examen du processus de fabrication montre que celles-ci sont réalisées avec de l'arsenic alors que le lot RTN12 utilise du phosphore. Les doses étant

identiques, il semblerait que le changement d'espèce du dopant influence fortement l'effet de canal court inverse.

La technologie DPN12 donne donc des résultats satisfaisants pour les transistors nMOS jusqu'à une longueur de grille de 30 nm. L'implantation des poches de surdopage produit un effet de canal court inverse sur les transistors pMOS, provoquant ainsi une forte augmentation de la tension de seuil pour les transistors de longueur 30 nm. Malgré ce point, cette technologie permet d'obtenir des transistors performants jusqu'à une longueur de grille de 30 nm.

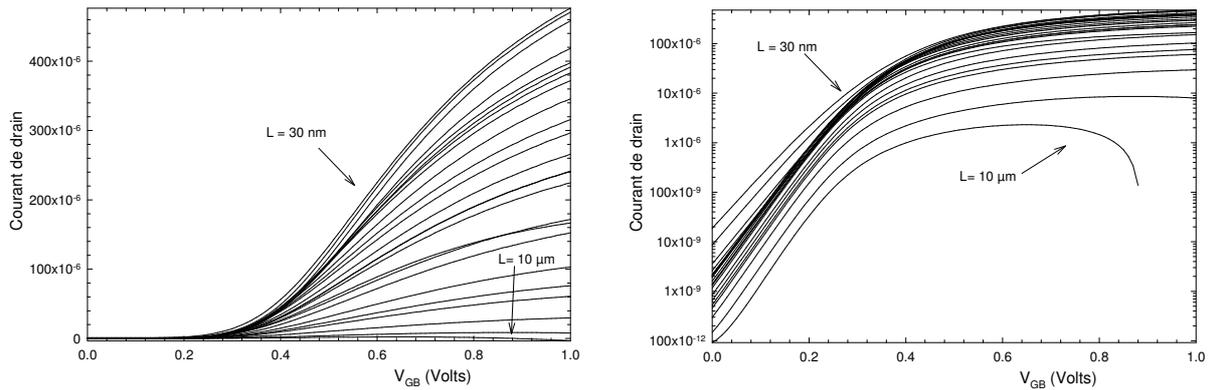


Figure 3.29 : Courants drain en fonction la tension de grille à $V_{DS}=25 \text{ mV}$ pour une batterie de transistors nMOS DPN12 à source et grille communes, en échelle linéaire (a) et logarithmique (b).

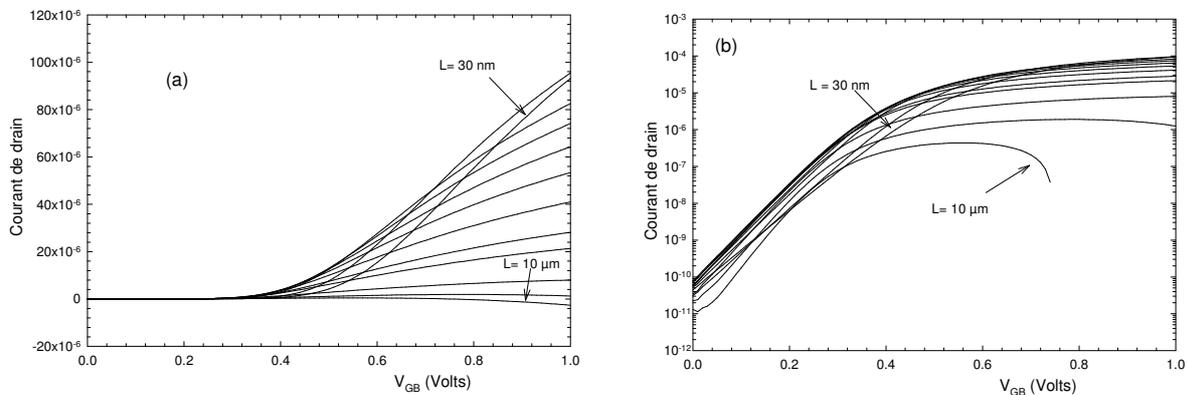


Figure 3.30 : Courants drain en fonction de la tension de grille à $V_{DS} = -25 \text{ mV}$ pour une batterie de transistors pMOS DPN12 à source et grille communes, en échelle linéaire (a) et logarithmique (b).

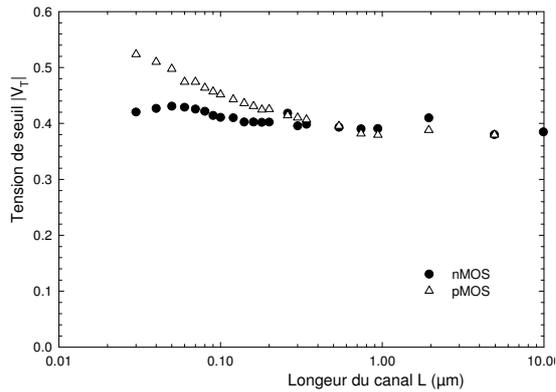


Figure 3.31 : Tensions de seuil en fonction de la longueur du canal pour les transistors n et p MOS DPN12.

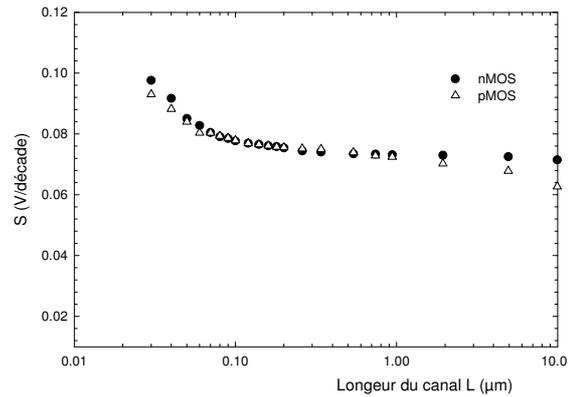


Figure 3.32 : Inverse de la pente sous le seuil S en fonction de la longueur du canal pour des transistors n et pMOS DPN12.

3.6.3 Mesure de capacités à diélectriques ultrafins

Le modèle que nous avons développé ne requiert que très peu de paramètres. Dans le but de déterminer les dopages du substrat, du polysilicium ainsi que l'épaisseur d'oxyde, nous avons réalisé des mesures de capacités sur les lots présentés précédemment. Pour chaque lot, nous avons mesuré sur un transistor isolé la capacité totale de grille C_{gg} , la capacité grille-canal C_{gc} et la capacité grille-bulk C_{gb} . Les figures 3.33 à 3.38 présentent l'ensemble des résultats. Les simulations pour chaque transistor ont été réalisées avec le modèle quantique décrit au chapitre 2.

Une capacité parasite est présente sur les capacités de type nMOS. En effet, la capacité C_{gb} doit tendre vers zéro en régime d'inversion, ce qui n'est pas le cas. La figure 3.39 est une vue en coupe de la structure et met en évidence la capacité parasite entre le plot de grille et le substrat. Pour les transistors pMOS, le caisson de type n fait office d'isolation ; cette capacité parasite n'existe donc pas pour les transistors pMOS. De plus, la capacité C_{gc} n'est pas affectée par cette capacité parasite. La capacité notée « C_{gg} corrigée » pour les transistors nMOS correspond donc à la capacité C_{gg} à laquelle nous avons soustrait la valeur de la capacité C_{gb} en forte inversion (typiquement $V_{GB} = 1$ V).

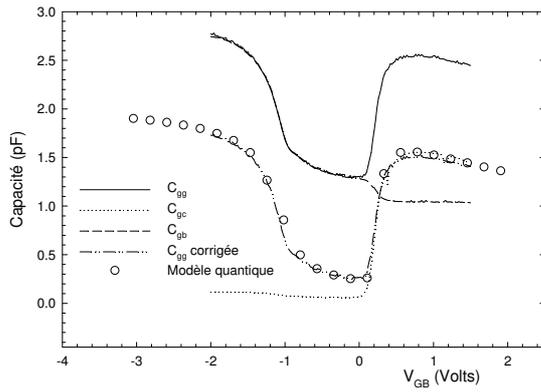


Figure 3.33: Capacités C_{gg} , C_{gc} , C_{gb} et C_{gg} corrigée pour un transistor nMOS RTN14 $10\ \mu\text{m} \times 10\ \mu\text{m}$.

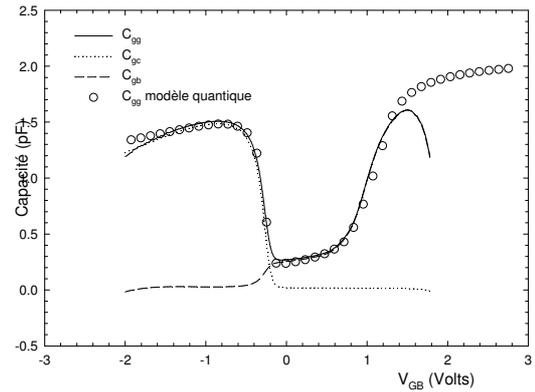


Figure 3.34 : Capacités C_{gg} , C_{gc} et C_{gb} pour une transistor pMOS RTN14 $10\ \mu\text{m} \times 10\ \mu\text{m}$.

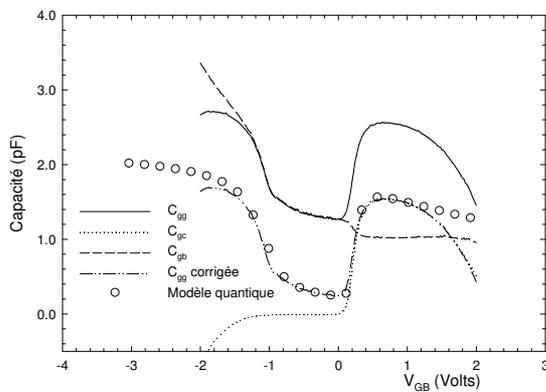


Figure 3.35 : Capacités C_{gg} , C_{gc} , C_{gb} et C_{gg} corrigée pour une transistor nMOS RTN12 $10\ \mu\text{m} \times 10\ \mu\text{m}$.

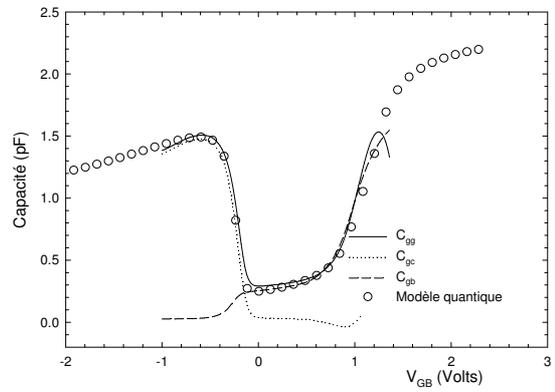


Figure 3.36 : Capacités C_{gg} , C_{gc} , et C_{gb} pour une transistor pMOS RTN12 $10\ \mu\text{m} \times 10\ \mu\text{m}$.

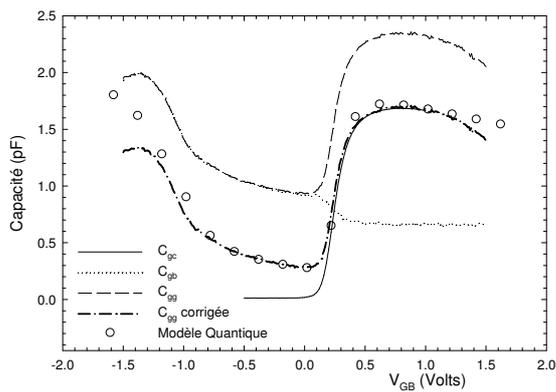


Figure 3.37 : Capacités C_{gg} , C_{gc} , C_{gb} et C_{gg} corrigée pour un transistor nMOS DPN12 $10\ \mu\text{m} \times 10\ \mu\text{m}$.

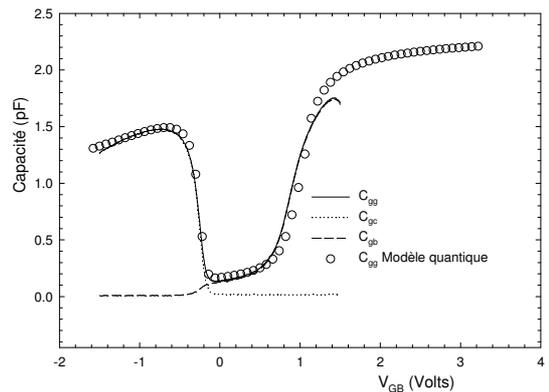


Figure 3.38 : Capacités C_{gg} , C_{gc} et C_{gb} pour un transistor pMOS DPN12 $10\ \mu\text{m} \times 10\ \mu\text{m}$.

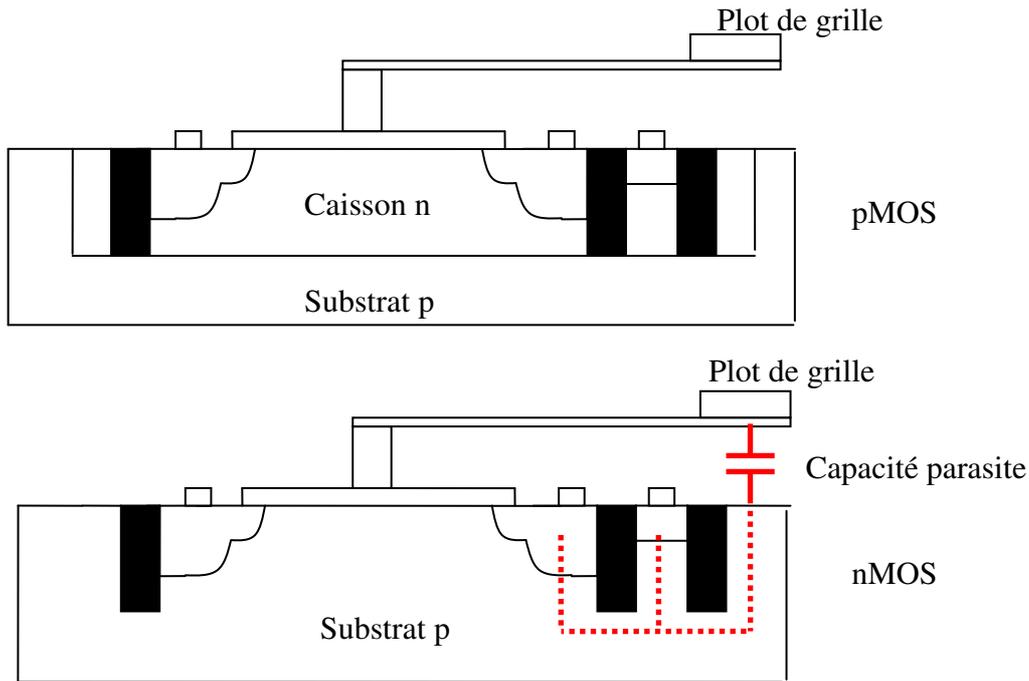


Figure 3.39 : Définition de la capacité parasite dans les transistor nMOS

Un autre problème important est lié au fort courant de fuite statique présent dans ces structures. Les mesures sont réalisées à l'aide d'un LCR mètre, dans notre cas le HP 4284A, qui, pour une polarisation DC et une fréquence de travail donnée, mesure l'impédance de la structure. Cependant, la précision avec laquelle le LCR mètre mesure une capacité décroît en présence d'une conductance parallèle trop importante. En effet, l'erreur relative est donnée par [HP'94]:

$$\frac{\Delta C}{C} (\%) = 0.1\sqrt{1+D^2} \quad (3.40)$$

où le facteur de dissipation D s'exprime en fonction de la conductance G_m , la capacité C_m et de la fréquence de mesure :

$$D = \frac{G_m}{2\pi f C_m} \quad (3.41)$$

La capacité se dégrade fortement dans les zones où le courant de grille devient très fort. Nous reviendrons en détail sur ce point dans la partie consacrée à la mesure et à la modélisation des capacités de grandes surfaces. Dans ce paragraphe, pour l'extraction des paramètres N_A , N_P , et T_{ox} , nous ne considérerons que la capacité mesurée dans des régimes de polarisations correspondant à l'utilisation normale du transistor, à savoir $|V_{GB}| < 1.5$ V

Les paramètres extraits sont regroupés dans le tableau 3.4. Les meilleurs résultats pour la simulation de la capacité du transistor pMOS DPN12 ont été obtenus en ajoutant une charge fixe d'interface $N_{SS}=2 \cdot 10^{12} \text{ cm}^{-2}$. Cette charge décale la tension de bandes plates de

0.1V. Nous disposons donc d'un ensemble de jeux de paramètres permettant la simulation des caractéristiques statiques des transistors. Le seul paramètre manquant est la mobilité ; de ce fait, l'ajustement des courbes simulées sur les mesures va nous permettre d'extraire une valeur de mobilité. Le paragraphe suivant décrit cette méthode d'extraction sur les transistors longs pour les transistors des lots RTN14, RTN12 et DPN12.

3.6.4 Validation du modèle statique du transistor MOS avec fuite de grille.

La première étape consiste à modéliser le courant de grille à $V_{DS}=0$ pour un transistor ayant une longueur de canal comprise entre 1 μm et 10 μm . En effet, nous avons pu voir qu'une trop forte longueur du canal provoque une courbure du potentiel de surface et une non homogénéité des densités de courant le long du canal. En revanche, une trop faible longueur implique des effets de canal court que notre modèle ne prend pas en compte. De plus, nous ne tenons pas compte du courant de fuite tunnel issu des zones de recouvrement, donc nos simulations sous-estiment légèrement le courant de grille en faible inversion. Nous avons donc réalisé une extraction des paramètres de courant tunnel sur des structures isolées avec une géométrie $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$. La figure 3.40 présente les mesures et les simulations du courant de grille à $V_{DS}=0$ V pour tous les transistors n et pMOS des trois lots.

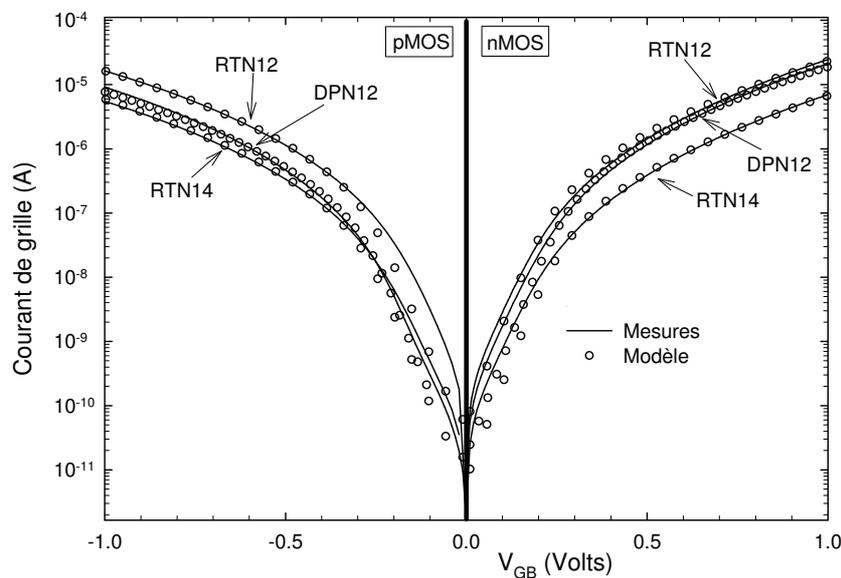


Figure 3.40 : Courant de grille en régime de faible et forte inversion pour les transistors n et pMOS des lots RTN14, RTN12 et DPN12. Les symboles représentent les résultats de simulation utilisant les paramètres reportés sur le tableau 3.4.

L'équation du courant tunnel développée dans le chapitre deux a été établie pour un oxyde pur SiO₂. Pour la comparaison avec les mesures des trois lots, nous avons obtenus de très bons résultats en prenant en compte le caractère nitruré des oxydes. Pour un matériau SiON, les hauteurs de barrières vues par les électrons et les trous sont modifiées ainsi que leurs masses effectives. L'épaisseur à considérer pour le courant tunnel est donc très proche de l'épaisseur physique, noté T_{ox}^{SiON} dans le tableau 3.4. Nous utilisons donc une épaisseur équivalente à un oxyde pur SiO₂ T_{ox}^{EOT} pour la simulation du potentiel de surface et du courant de drain, et l'épaisseur physique du diélectrique T_{ox}^{SiON} pour la simulation du courant de grille. Tous les paramètres ainsi extraits sont regroupés dans le tableau 3.4.

	RTN14		RTN12		DPN12	
	<i>nMOS</i>	<i>pMOS</i>	<i>nMOS</i>	<i>pMOS</i>	<i>nMOS</i>	<i>pMOS</i>
N_A (cm ⁻³)	8 10 ¹⁷	7 10 ¹⁷	7 10 ¹⁷	8 10 ¹⁷	10 ¹⁸	3 10 ¹⁷
N_P (cm ⁻³)	8 10 ¹⁹	8 10 ¹⁹	6 10 ¹⁹	5 10 ¹⁷	9 10 ¹⁹	4 10 ¹⁹
T_{ox}^{EOT} (nm)	1.4	1.4	1.2	1.2	1.2	1.2
T_{ox}^{SiON} (nm)	2.23	2.23	2.11	2.11	2.1	2.2
χ_B (eV)	2.15	2	2.14	2	2.15	1.85
m_{ox}^*/m_0	0.3	0.4	0.29	0.4	0.3	0.4

Tableau 3.4 : Paramètres extraits à partir de mesures de capacités et de courants de grille à $V_{DS}=0$ V pour les trois lots de transistors.

Avec le jeu de paramètres extraits à partir des mesures de capacités et du courant de grille, nous pouvons simuler les caractéristiques courant tension (I_D-V_{GB} , I_S-V_{GB}) des transistors long en régime linéaire soumis à de fortes fuites de grille. Le seul paramètre inconnu est la mobilité. Nous avons donc dans un premier temps utilisé le modèle de mobilité universel décrit par l'équation empirique (3.30) que nous avons par la suite ajusté pour que les simulations soient en accord avec les courbes expérimentales, et cela sans modifier les autres paramètres. Les résultats pour les transistors nMOS du lot RTN14, RTN12 et DPN12 de géométrie $W \times L = 10 \times 10 \mu\text{m}$ sont présentés sur la figure 3.41. Le modèle de mobilité que nous avons utilisé est présenté sur la figure 3.42. De plus nous avons reporté sur cette figure la mobilité extraite par la méthode Split CV adaptée aux transistors à oxyde ultramince [LIME'03], ainsi que le modèle de mobilité universel (3.30).

D'une part, les mobilités obtenues sont supérieures à celles prévues par le modèle de mobilité universel. D'autre part, on constate que la mobilité obtenue par la méthode Split CV est inférieure à la mobilité obtenue par ajustement avec notre modèle. Quelle que soit la méthode employée, la même tendance est obtenue, à savoir une réduction de la mobilité avec la diminution de l'épaisseur d'oxyde (RTN14 \rightarrow RTN12) et une réduction de la mobilité avec une augmentation du dopage du substrat (RTN12 \rightarrow DPN12). Il est possible que l'extrême simplicité de l'expression du courant de drain exploitée dans la méthode Split CV soit à l'origine de l'écart entre la mobilité extraite par la méthode split CV celle extraite par ajustement avec notre modèle. Il faut rappeler que le modèle de courant de drain dans la méthode Split CV ne tient pas compte de la présence de forts courants de grille, des effets quantiques et du phénomène de polydéplétion.

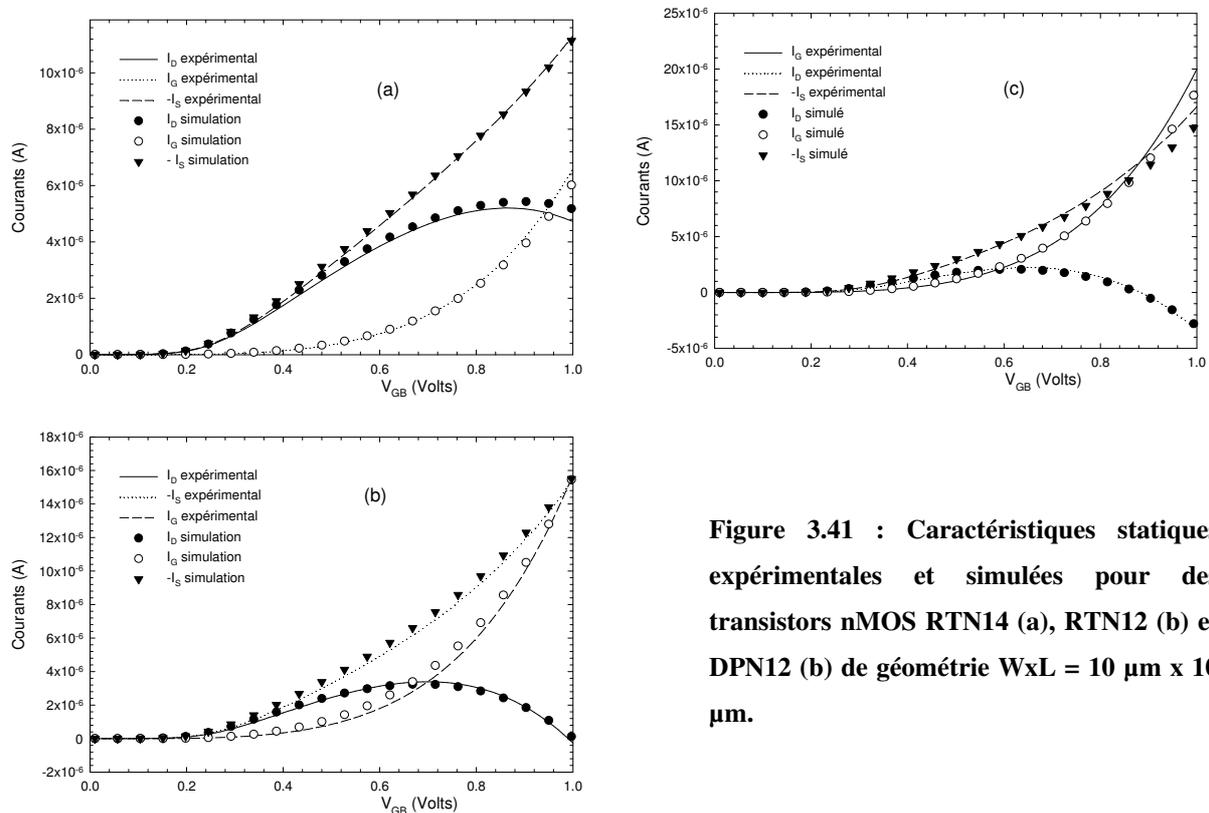


Figure 3.41 : Caractéristiques statiques expérimentales et simulées pour des transistors nMOS RTN14 (a), RTN12 (b) et DPN12 (b) de géométrie $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$.

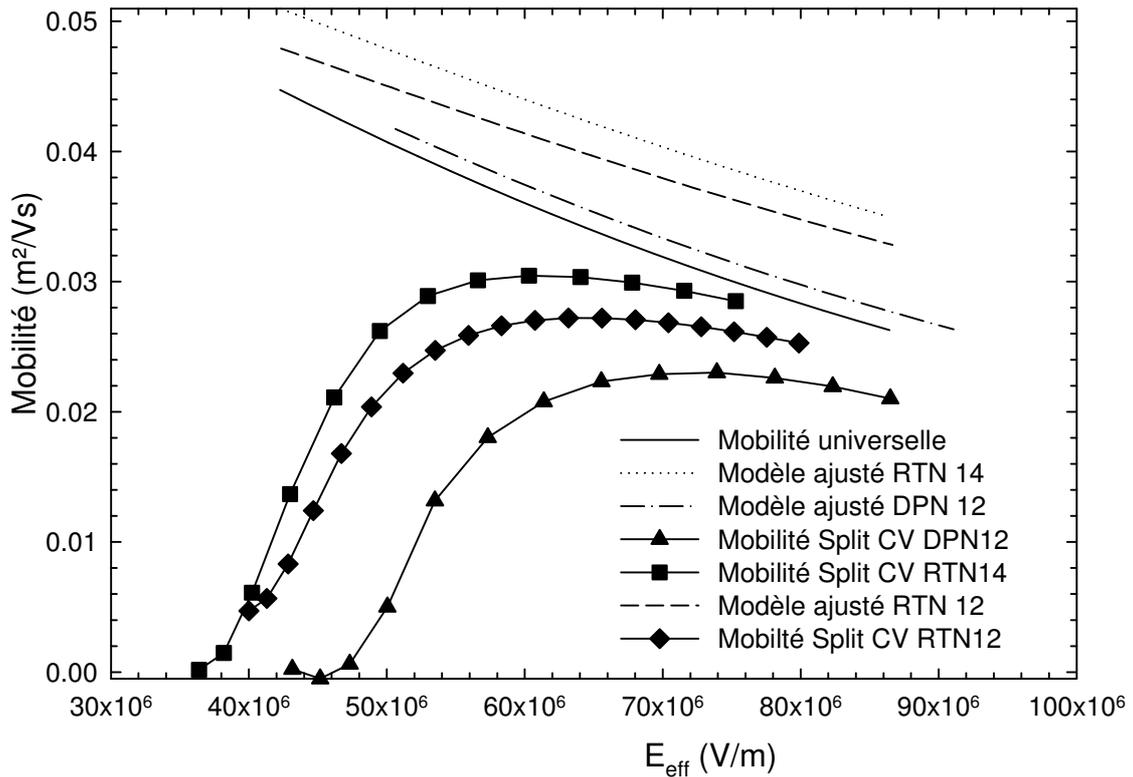


Figure 3.42 : Comparaison des mobilités extraites par Split CV et par ajustement du modèle pour les transistors nMOS RTN14, RTN 12, et DPN12 de géométrie $W \times L = 10 \mu \times 10 \mu \text{m}$.

Les résultats pour les transistors pMOS du lot RTN14, RTN12 et DPN12 de géométrie $W \times L = 10 \mu \text{m} \times 10 \mu \text{m}$ sont présentés sur la figure 3.43. Les mobilités extraites par Split CV et par ajustement du modèle pour les transistors pMOS sont présentées sur la figure 3.44. De la même façon que pour les nMOS, la mobilité extraite par Split CV est inférieure à celle extraite par ajustement du modèle. Pour la simulation des caractéristiques statiques, nous avons obtenu de très bons résultats en conservant le même modèle de mobilité pour les trois lots. On obtient donc trois courbes confondues sur la figure 3.44. Ainsi, il semble que la mobilité soit peu affectée par le changement d'épaisseur d'oxyde où la modification du dopage du substrat. Cette remarque ne s'applique que dans le cas où les transistors sont à canal long, car il a été montré, que ce soit pour les transistors n ou pMOS, que la diminution de la longueur de canal diminue la mobilité à faible champ [ROMANJEK'05]. Cet effet s'explique par la présence des poches de surdopage qui deviennent prépondérante lorsque le canal devient très court.

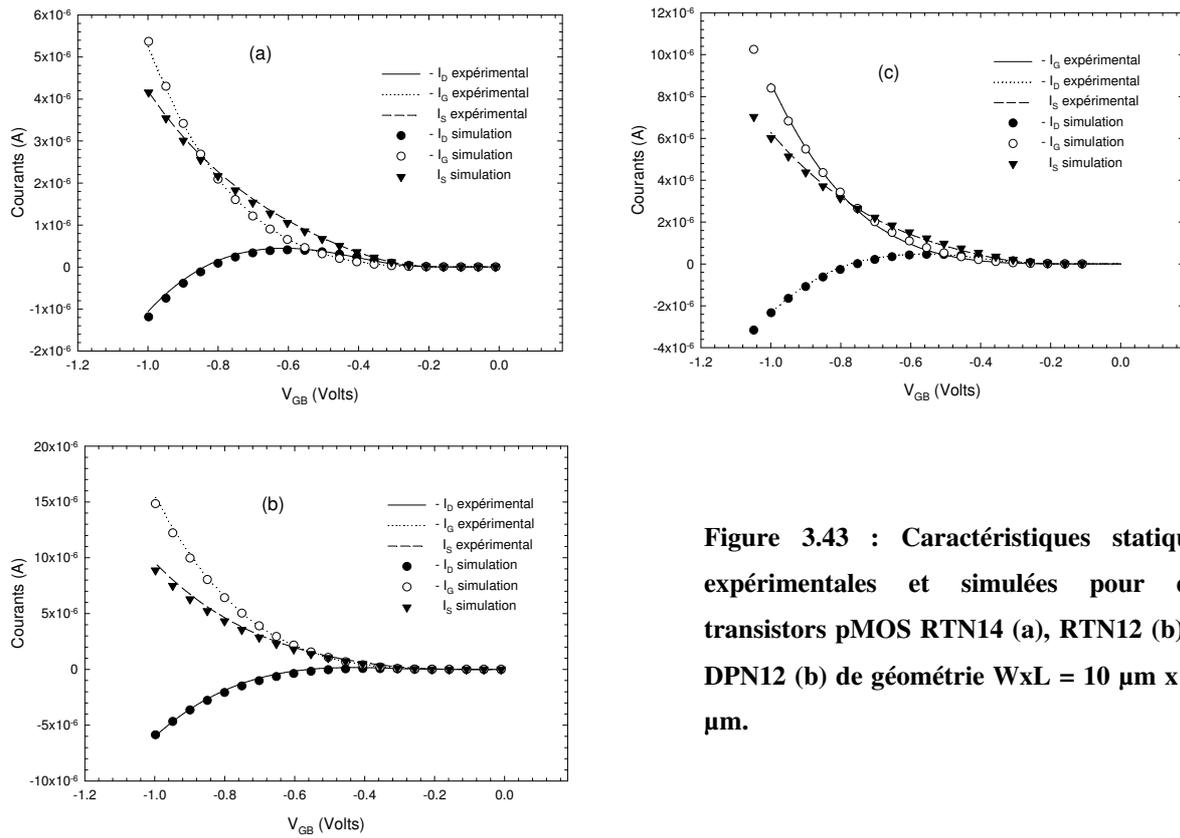


Figure 3.43 : Caractéristiques statiques expérimentales et simulées pour des transistors pMOS RTN14 (a), RTN12 (b) et DPN12 (c) de géométrie $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$.

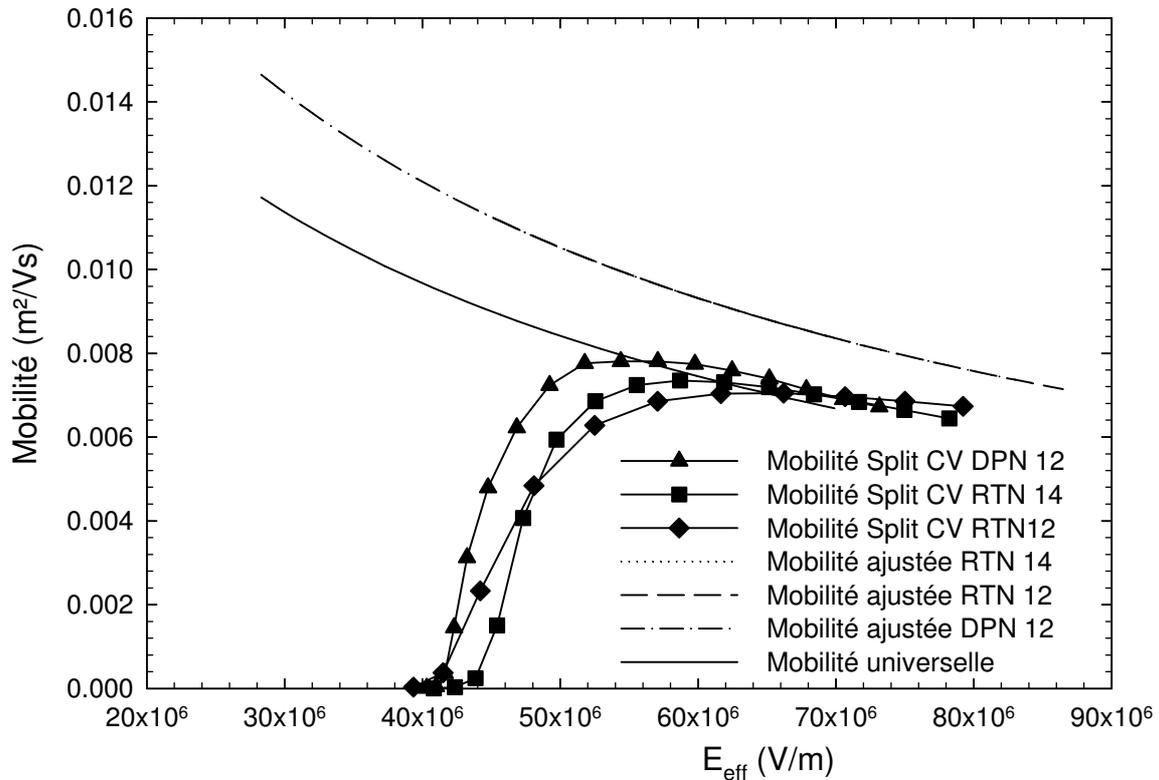


Figure 3.44 : Comparaison des mobilités extraites par Split CV et par ajustement du modèle pour les transistors pMOS RTN14, RTN 12, et DPN12 de géométrie $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$.

En résumé, à partir d'une caractérisation des capacités, nous avons donc extrait les niveaux de dopage du substrat et du polysilicium. Nous avons ensuite extrait les paramètres physiques relatifs au courant tunnel de grille. L'ajustement du modèle de mobilité pour simuler au mieux les caractéristiques courant - tension en régime linéaire ($|V_{DS}| = 25 \text{ mV}$) nous a permis d'extraire des valeurs de mobilité. Nous avons mis en évidence les différences entre la mobilité extraite de cette façon et celle obtenue avec la méthode Split CV. L'utilisation d'un modèle ne prenant pas en compte tous les phénomènes physiques mis en jeu dans ces technologies pourrait expliquer la sous estimation de la mobilité par la méthode Split CV.

3.7 Caractérisation de capacités MOS de grande surface

Dans cette partie, nous allons nous focaliser sur les capacités MOS de grande surface. Nous avons à notre disposition dans le lot DPN12 des structures de type n et p, où la source et le drain sont court circuités par un niveau de métallisation. Les surfaces disponibles sont $W \times L = 10 \times 10 \mu\text{m}^2$, $100 \times 10 \mu\text{m}^2$ et $100 \times 100 \mu\text{m}^2$.

3.7.1 Caractérisation statique

La figure 3.45 présente les densités de courants pour les structures de type nMOS, ainsi les simulations réalisées avec les paramètres extraits précédemment. L'effet de la longueur sur le potentiel de surface apparaît clairement sur ces structures. On observe un rapport 10 quand la largeur passe de $10 \mu\text{m}$ à $100 \mu\text{m}$, la longueur restant égale à $10 \mu\text{m}$. En revanche, le courant de grille total n'est pas proportionnel à la surface pour la structure de longueur et de largeur $100 \mu\text{m}$. Comme expliqué au paragraphe 3.5.2, la mobilité influence la courbure du potentiel de surface, et donc le courant de grille total. Nous avons ajusté la mobilité pour simuler au mieux le courant de grille de la structure $100 \times 100 \mu\text{m}^2$; Avec une mobilité légèrement supérieure à celle extraite sur les transistors MOS précédemment caractérisés, nous obtenons un très bon accord pour le courant de grille pour les trois structures simulées avec le même jeu de paramètres du tableau 3.4. Compte tenu que ces structures de test n'ont pas une architecture de transistors (absence de zone de recouvrement, poches, etc...) nous pouvons faire l'hypothèse que la mobilité pouvait être plus grande dans ces structures que dans les transistors MOS précédemment caractérisés.

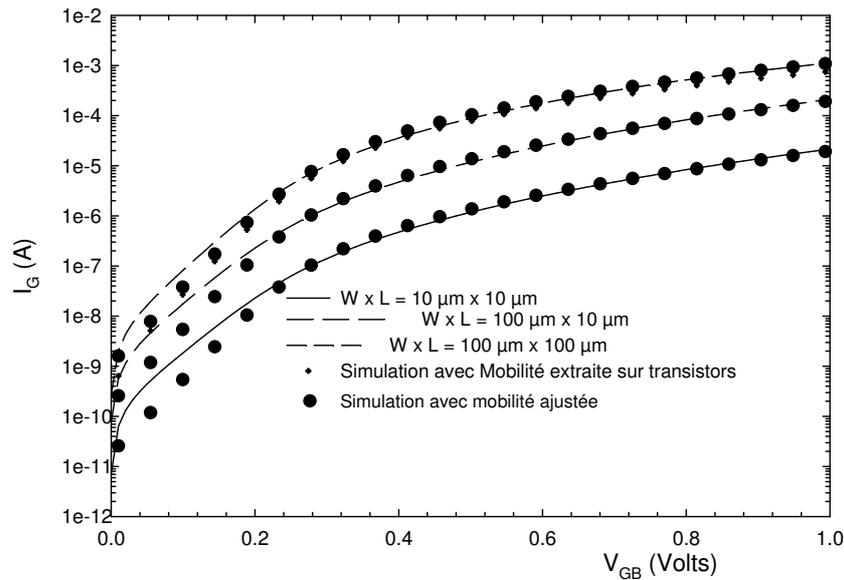


Figure 3.45 : Courants de grilles mesurés et simulés pour les structures nMOS DPN12 de grande surface.

3.7.2 Modélisation non quasi statique de la capacité MOS.

Nous avons réalisé des mesures de capacité sur ces structures de grandes surfaces. Pour les transistors de longueur $10\ \mu\text{m}$, la fréquence à laquelle est réalisée la mesure n'influe pas le résultat dans la gamme $100\ \text{kHz} - 1\ \text{MHz}$. La figure 3.46 présente les résultats sur les structures $100 \times 100\ \mu\text{m}^2$. Pour les structures de longueurs $100\ \mu\text{m}$, les mesures montrent une dépendance de la capacité à la fréquence. Il est donc nécessaire de prendre en compte le terme de la dérivée de la charge d'inversion par rapport au temps dans l'équation de continuité (3.35).

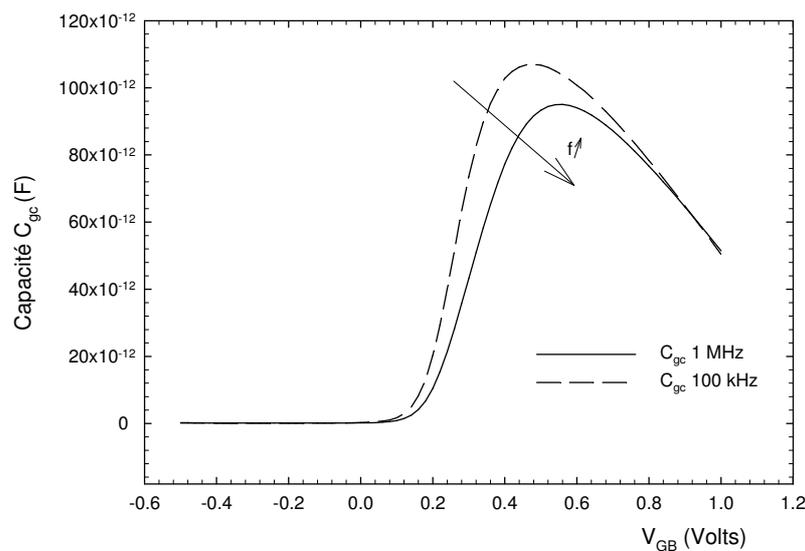


Figure 3.46 : Capacité C_{gc} d'une structure nMOS $100 \times 100\ \mu\text{m}^2$ à 1 MHz et 100 kHz.

D'un point de vue mathématique, on se trouve confronté à la résolution numérique d'une équation aux dérivées partielles. Nous avons donc développé un modèle non quasi statique en régime de petit signal, tenant compte des effets quantiques, du phénomène de polydéplétion et du courant tunnel de grille. Les détails de ce modèle sont donnés en annexe D. Cette approche nous fournit pour chaque point du canal l'amplitude complexe (module et phase) du potentiel de surface. A partir de cette valeur, on peut donc accéder à toutes les amplitudes complexes des grandeurs physiques, et cela en fonction de la fréquence.

3.7.2.1 Etude théorique non quasi statique de l'impact du courant tunnel

Dans un premier temps, nous avons comparé les résultats fournis par notre modèle aux résultats d'ISE TCAD en l'absence de courant tunnel au travers de l'oxyde. La figure 3.47 montre un excellent accord, que ce soit pour la capacité ou la conductance. A titre d'exemple, le résultat ISE TCAD a nécessité environ deux heures de calculs contre quelques dizaines de seconde pour notre modèle.

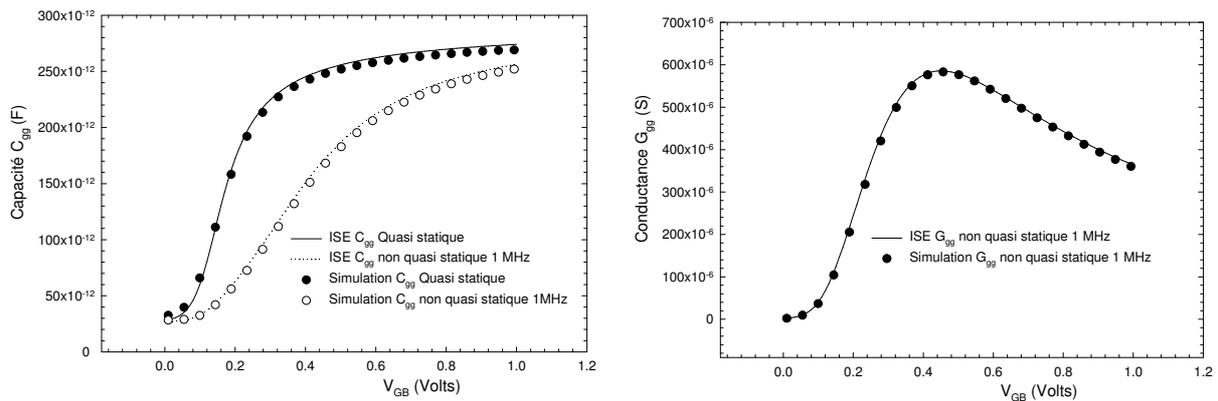


Figure 3.47 : Capacité C_{gg} et conductance G_{gg} simulées avec ISE TCAD et avec notre modèle sur une structure nMOS sans courant tunnel de grille.

En l'absence de courant de grille, la capacité dynamique tend vers la capacité quasi statique quand la fréquence tend vers zéro. En présence de courant tunnel direct, ce n'est plus le cas comme le montre la figure 3.48a. En effet, la fluctuation du potentiel de surface, de part la grande longueur du composant, n'est pas en phase avec la fluctuation de la tension de grille (figure 3.48b). Ceci implique donc que le courant tunnel n'est pas en phase avec la tension de grille. Une partie du courant tunnel direct alternatif a donc une composante imaginaire contribuant au final à la capacité mesurée.

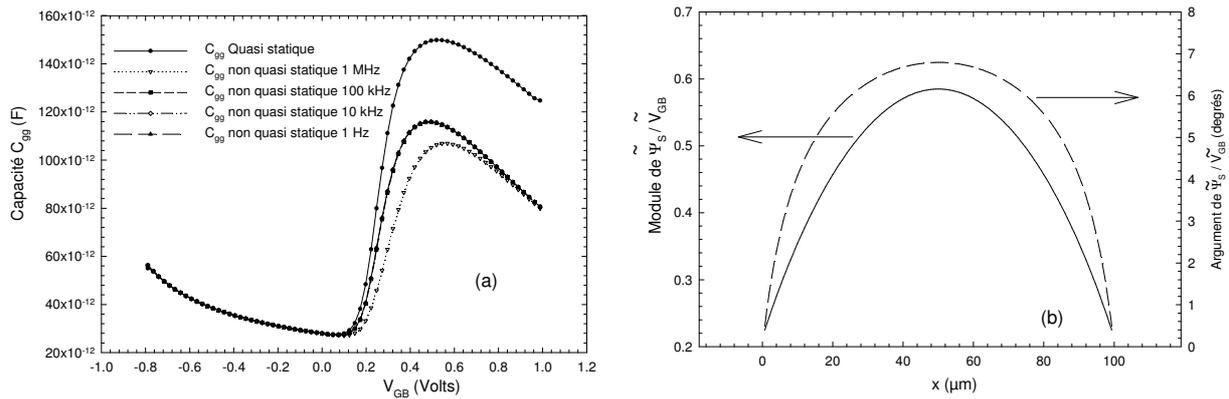


Figure 3.48 : Evolution de la capacité C_{gg} pour une structure nMOS $100 \times 100 \mu\text{m}^2$ en fonction de la fréquence en présence de courant tunnel de grille (a) et module et phase de la fluctuation du potentiel de surface à $V_{GB}=1\text{V}$ (b).

3.7.2.2 Validation du modèle non quasi statique

Dans un premier temps, comparons les capacités des structures $W \times L=10 \mu\text{m} \times 10 \mu\text{m}$ et $W \times L=100\mu\text{m} \times 10 \mu\text{m}$ (figure 3.50). On remarque que la capacité normée à la surface chute à partir de $V_{GB} = 0.5$ Volts pour la capacité de largeur $100 \mu\text{m}$. Cet effet peut être modélisé simplement en prenant en compte le caractère inductif de l'accès au composant, en particulier l'ensemble des câbles et pointes, qui d'après la documentation constructeur, présentent un schéma équivalent suivant la figure 3.49.

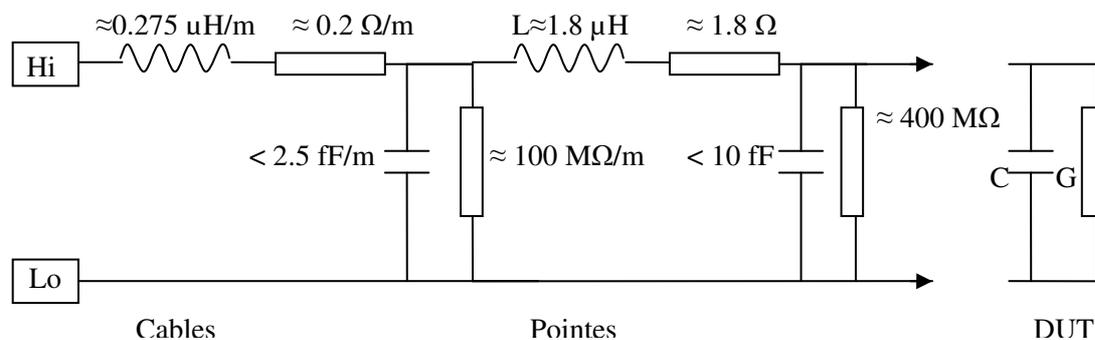


Figure 3.49 : Modélisation petit signal des câbles et des pointes. C et G représente la capacité et la conductance du dispositif mesuré.

Compte tenu de la faible longueur des câbles (inférieur au mètre) que nous avons utilisés, nous considérons que l'impédance mesurée est en fait l'impédance du dispositif en série avec l'inductance de $1.8 \mu\text{H}$. Les fuites modélisées par la capacité de 10 fF et la résistance de $400 \text{ M}\Omega$ sont considérées comme négligeables. La capacité mesurée s'exprime donc simplement par :

$$C_{mesuré} = \frac{C - LG^2 - LC^2\omega^2}{1 + L^2C^2\omega^4 - (2C - LG^2)L\omega^2} \quad (3.42)$$

où C et G sont la capacité et la conductance calculées par le modèle petit signal, et L l'inductance série due aux pointes. On obtient un très bon accord entre ce modèle de prise en compte des accès et les mesures avec la valeur $L=1.8 \mu\text{m}$ (figure 3.50).

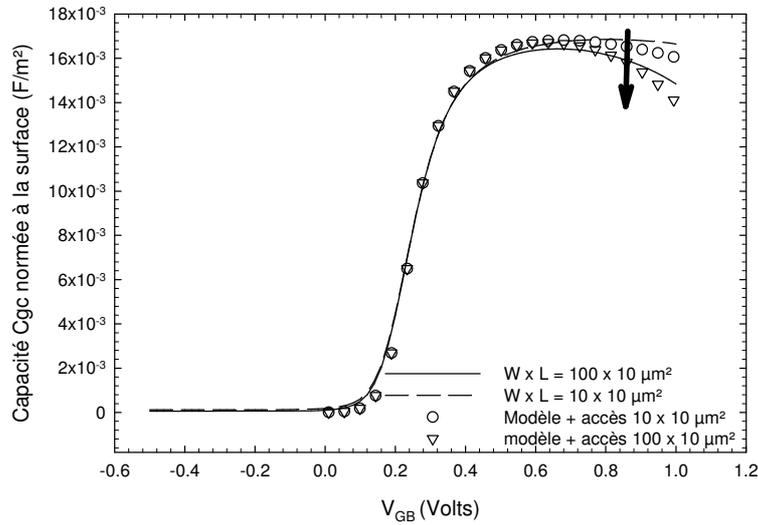


Figure 3.50 : Chute de la capacité pour une augmentation de la largeur sur les structures nMOS DPN 12.

La figure 3.51 présente les résultats obtenus avec notre modèle non quasi statique pour différentes géométries et aux fréquences de 100 kHz et 1 MHz. On obtient de très bons résultats en utilisant la mobilité précédemment utilisée pour ajuster au mieux le courant de grille statique.

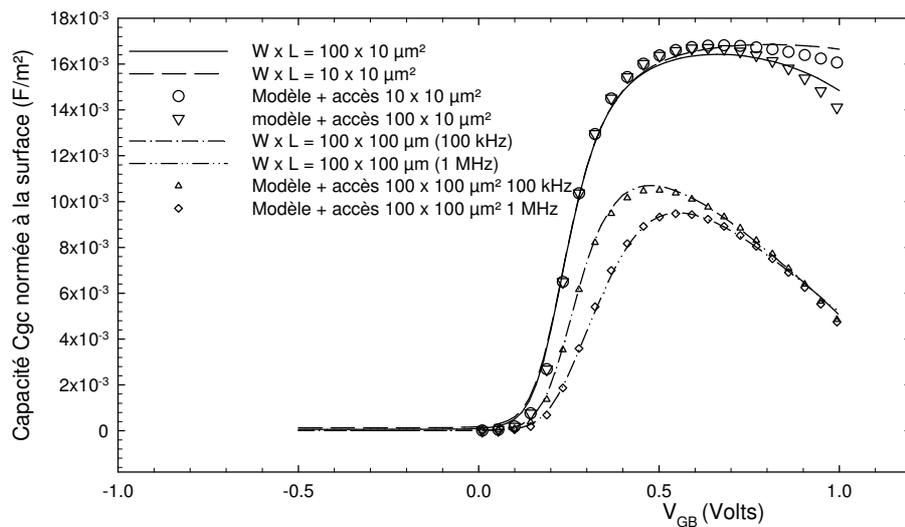


Figure 3.51 : Capacité C_{gc} normée à la surface pour les structures nMOS DPN12.

Nous présentons sur la figure 3.52 la comparaison des conductances mesurées sur la structure $100 \times 100 \mu\text{m}^2$ avec notre modèle petit signal. Avec les paramètres extraits précédemment, on observe une légère surévaluation de la conductance avec notre modèle. Toutefois, la dépendance à la fréquence est très bien prise en compte par notre modélisation. La structure de grande surface $100 \times 100 \mu\text{m}$ est donc simulée avec précision d'un point de vue statique (I_G - V_{GB}) et dynamique (capacité et conductance à fréquence variable) avec un faible nombre de paramètres. Il est certain que la prise en compte des non homogénéités de la structure, tel que la variation de l'épaisseur d'oxyde ou les profils réels de dopage, permettrait d'obtenir une simulation encore plus proche des données expérimentales.

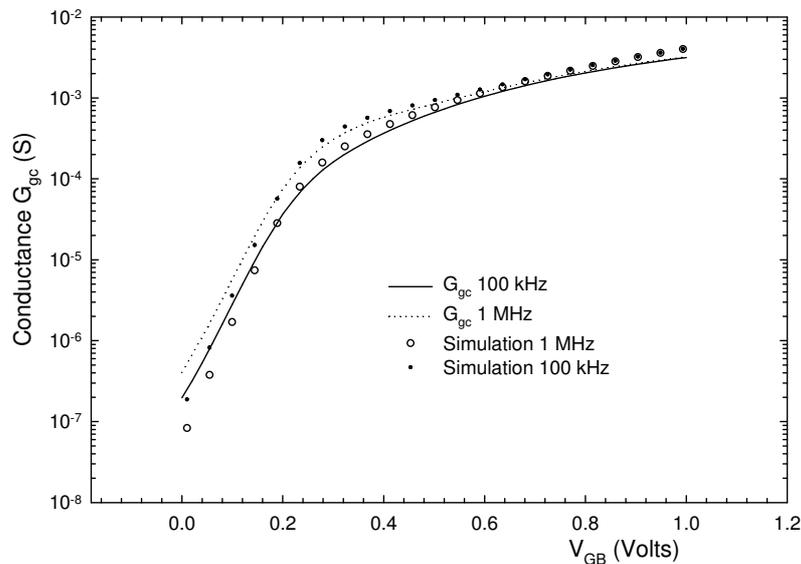


Figure 3.52 : Conductance G_{gc} mesurée et simulée pour une structure nMOS $100 \times 100 \mu\text{m}^2$ à 100 kHz et 1 MHz.

Nous avons réalisé une étude similaire sur les structures pMOS issues du lot DPN 12. La figure 3.53 présente les courants de grille statiques et la figure 3.54 présente les capacités C_{gc} normées à la surface. La mobilité des trous étant nettement plus faible que celle des électrons, le potentiel de surface va se courber nettement plus que dans le cas des structures nMOS de grande surface. Il en résulte une plus grande dégradation du courant total de grille ou de la capacité normés à la surface du composant. De la même façon que pour les nMOS, la mobilité permettant d'ajuster au mieux la simulation par rapport aux mesures statiques et dynamique est légèrement plus grande que celle extraite sur les transistors du paragraphe 3.6.4.

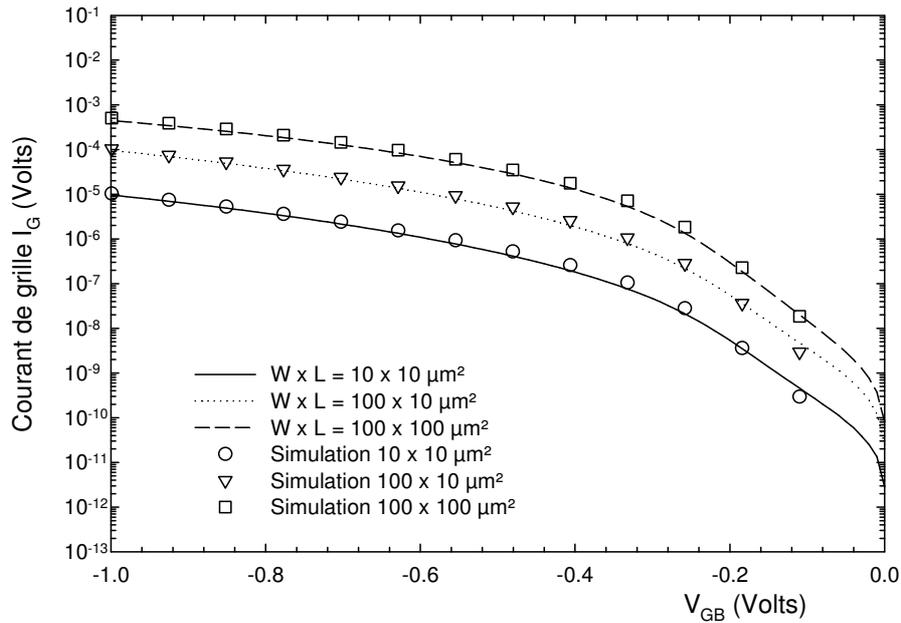


Figure 3.53 : Courants de grilles mesurés et simulés pour les structures pMOS DPN12 de grande surface.

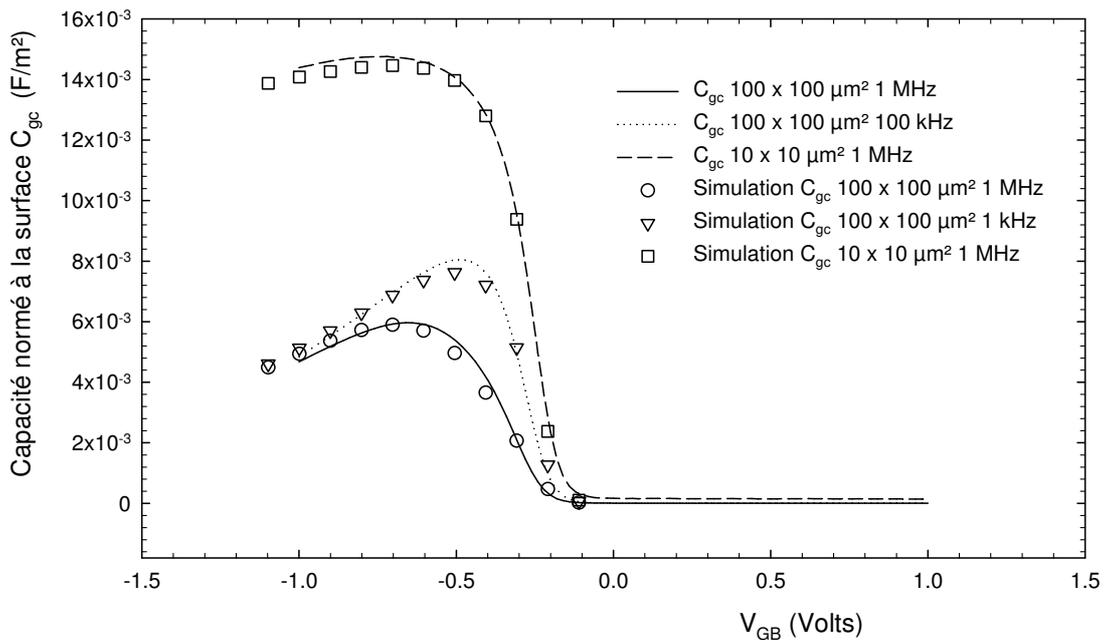


Figure 3.54 : Capacité C_{gc} normée à la surface pour les structures pMOS DPN12.

3.8 Conclusion

Dans les trois premiers paragraphes de ce chapitre, nous avons présenté la modélisation de transistors MOS soumis à de fortes fuites de grille. Ce modèle est valable pour les transistors à canal long, puisqu'il est basé sur le modèle en feuille de charge originalement proposé par Brews en 1978. En nous appuyant sur les résultats établis au

chapitre 2, nous avons modifié le modèle de façon à tenir compte de la polydéplétion, des effets quantiques et du courant tunnel direct au travers du diélectrique de grille.

Nous avons montré au paragraphe 3.5 que la présence du courant tunnel induit une courbure sur le potentiel de surface. Cette courbure est fortement liée à la densité de courant tunnel, à la longueur du transistor et à la mobilité des porteurs. Contrairement aux travaux déjà publiés précédemment sur ce phénomène, nous n'avons pas utilisé de modèle segmenté ; nous avons développé une approche analytique basée sur une seule équation différentielle nécessitant de faibles temps de calculs. La courbure du potentiel de surface induit une diminution du courant de grille total et de la capacité des structures normées à la surface avec l'augmentation de la longueur du transistor.

Nous avons également exposé l'effet du courant de grille sur le fonctionnement du transistor en régime linéaire. Notre modèle basé sur l'équation différentielle (3.37) prend en compte le partitionnement du courant de grille sur le courant de drain et de source.

Nous avons présenté dans le paragraphe 3.6 les différents lots de transistors que nous avons eu à disposition au cours de cette thèse. Les différentes mesures statiques ont permis de montrer les excellentes performances des transistors nMOS de longueur de grille 55 nm pour le lot RTN12 et 30 nm pour le lot DPN12. Les transistors pMOS souffrent de problèmes de canal court mal maîtrisé (dégradation de la pente sous le seuil pour les transistors pMOS RTN12) ou d'effet de canal court inverse trop important (augmentation importante de la tension de seuil des transistors pMOS DPN12).

Dans le paragraphe 3.6.4, nous nous sommes focalisés sur les transistors longs pour valider notre modèle basé sur l'équation (3.37). Dans un premier temps, nous avons extrait les niveaux de dopage du substrat et du polysilicium à partir de mesures de la capacité des structures, en utilisant le modèle quantique décrit au chapitre 2. Nous avons ensuite extrait les paramètres relatifs au courant de grille des différentes structures à $V_{DS} = 0V$. Enfin, nous avons ajusté la mobilité de façon à simuler au mieux le courant de drain et de source en régime linéaire. Cette approche nous a donc permis d'extraire la mobilité effective nécessaire à l'ajustement du modèle. La mobilité extraite suivant la méthode Split CV est inférieure à celle extraite pour obtenir une simulation en accord avec les mesures. Cet écart est probablement dû au modèle de conduction utilisé dans la méthode Split CV, qui ne prend pas en compte ni les effets quantiques, ni les phénomènes de polydéplétion.

Dans le dernier paragraphe de ce chapitre, nous avons présenté l'impact du courant tunnel de grille sur les caractéristiques statiques et dynamiques de structures MOS de grande surface. D'un point de vue statique, nous avons mis en évidence expérimentalement la

nécessité de prendre en compte le phénomène de courbure du potentiel de surface pour simuler avec le même jeu de paramètres le courant de grille des transistors MOS de $10 \times 10 \mu\text{m}^2$ à $100 \times 100 \mu\text{m}^2$. Nous avons montré expérimentalement que la mobilité des porteurs influait sur le courant de grille total. Les mesures de capacité sur les transistors très longs ($L=100\mu\text{m}$) ont montré la nécessité de prendre en compte les effets non quasi statiques dans le modèle. Nous avons développé un modèle petit signal non quasi statique prenant en compte les effets quantiques, la polydéplétion et le courant tunnel de grille. Avec le même jeu de paramètres utilisés pour la simulation statique du courant de grille, nous avons obtenu un très bon accord avec les mesures de capacités et de conductances effectuées à plusieurs fréquences.

Les modèles développés dans ce chapitre ont donc permis de caractériser les structures MOS de technologies avancées soumises aux forts courants de fuites de grille. L'extraction précise des paramètres technologiques est une étape indispensable avant la modélisation et la caractérisation de ces structures en terme de bruit basse fréquence. Dans le chapitre suivant, portant sur l'étude du bruit basse fréquence, nous nous appuyerons donc, d'une part, sur les modèles établis dans les chapitres 2 et 3, et d'autre part, sur tous les paramètres extraits dans ce chapitre.

CHAPITRE 4

MODÉLISATION DU BRUIT BASSE
FRÉQUENCE DANS LES STRUCTURES
MOS À DIÉLECTRIQUES ULTRAFINS

4 Modélisation du bruit Basse fréquence dans les structures MOS à diélectriques ultrafins

4.1 Introduction

Les modèles développés au cours des chapitres deux et trois nous ont permis d'extraire les paramètres technologiques et de conduction de dispositifs MOS à diélectriques ultrafins à partir des caractéristiques statiques ($I-V$) et dynamiques ($C-V$). Cependant, les performances d'un composant ne se limitent pas à ces seuls paramètres. Parmi les paramètres dynamiques à étudier, le bruit basse fréquence est un facteur important pour certains circuits électroniques, notamment en radio fréquence avec le bruit de phase des oscillateurs. Le bruit basse fréquence constitue également, de part sa sensibilité, un bon indicateur de qualité des composants.

Ce chapitre est consacré au bruit basse fréquence dans les transistors MOS à diélectriques ultrafins. Dans un premier temps, nous nous intéresserons à la formulation du phénomène de piégeage - dépiégeage d'un porteur libre par un piège d'oxyde. Le but de cette partie est d'une part de vérifier la validité des modèles de la littérature dans le cas de structures à diélectriques ultrafins, et d'autre part de coupler le modèle de piégeage - dépiégeage avec notre modèle de potentiel de surface.

Nous développerons ensuite la formulation du bruit basse fréquence associé au courant de drain et de source. Cette partie sera illustrée par la caractérisation en bruit basse fréquence des lots de transistors étudiés au chapitre trois.

Nous aborderons enfin la modélisation et la caractérisation du bruit basse fréquence associé au courant de fuite de grille. Nous détaillerons également l'impact d'un fort courant de grille sur le bruit du courant drain. La modélisation et la caractérisation d'un défaut singulier responsable de bruit de type RTS (*Random Telegraph Signal*) sur le courant de grille seront finalement présentées.

4.2 Les sources de bruit

Le bruit est un signal aléatoire ergotique et stationnaire du second ordre. Les deux premiers moments du signal de bruit, sa valeur moyenne et sa variance, sont invariants dans le temps. L'analyse temporelle des fluctuations ne permet pas de séparer les différentes sources de bruit car l'amplitude des fluctuations se répartie généralement de manière gaussienne

autour de la valeur moyenne du signal. Toutefois, son utilisation se prête bien à l'étude du bruit RTS.

Dans ce paragraphe, nous allons rappeler brièvement les principales sources de bruit que l'on rencontre dans les composants à semiconducteur :

i) Le bruit thermique

Ce bruit est dû à l'agitation thermique des porteurs. C'est un bruit blanc (indépendant de la fréquence) sur une large bande de fréquence ($f < 10^{12}$ Hz). Pour une résistance R , son niveau est donné par :

$$S_v(f) = 4kTR \quad (4.1)$$

où k est la constante de Boltzmann et T la température. Le bruit thermique est un bruit fondamental qui ne peut être réduit que par le biais de la température.

ii) Le bruit de grenaille

Ce bruit est dû à l'émission aléatoire de porteurs à travers une barrière de potentiel. Sa densité spectrale est donnée par :

$$S_i(f) = 2q\langle I \rangle \quad (4.2)$$

où $\langle I \rangle$ est la valeur moyenne du courant.

iii) Bruit de génération recombinaison

Le bruit de génération recombinaison provient de la fluctuation des charges piégés dans la zone de charge d'espace du transistor. Le bruit associé à un seul piège présente un spectre dit Lorentzien :

$$S_i(f) = \frac{A}{1 + \omega^2 \tau^2} \quad (4.3)$$

où τ est la constante de temps caractéristique du piège et A une constante.

iv) Le bruit en 1/f

Ce type de bruit, qui varie comme l'inverse de la fréquence, est généralement prépondérant à basse fréquence. Ce bruit est dû à une fluctuation de la conductivité du matériau. L'origine des fluctuations peut être liée à la fluctuation de la mobilité [HOOGE'69], ou à la fluctuation du nombre de porteurs liée au phénomène de piégeage dynamique [FU'72] ou à la fluctuation du nombre de porteurs et de la mobilité corrélée [HUNG'90a,HUNG'90b]. Dans les transistors MOS, l'origine du bruit basse fréquence n'a pas été clairement établie, puisque la quantité observable est la fluctuation du courant du canal, qui est proportionnelle au produit de la mobilité par le nombre de porteurs. La séparation des effets de la fluctuation du nombre de porteurs ou de la mobilité est donc difficile. Le modèle de bruit basé sur la

fluctuation de la mobilité pose un problème en faible inversion, puisqu'il prévoit une diminution du bruit normalisé au carré du courant S_{I_D}/I_D^2 avec l'augmentation du courant, phénomène qui n'est pas vérifié expérimentalement. Un autre problème avec ce modèle est lié à l'explication physique de la fluctuation de la mobilité. La constante de bruit α_H intervenant dans le modèle de Hooge, considérée par le passé comme une constante fondamentale égale à 10^{-3} pour tous les matériaux [SIMOEN'99], est désormais considérée comme empirique, puisqu'elle dépend du matériau et des données technologiques. Dans ce mémoire, nous avons considéré que l'origine du bruit est liée à la fluctuation du nombre de porteurs et à la fluctuation de leur mobilité corrélée, d'une part, parce qu'elle permet de modéliser le bruit dans les transistors n et pMOS de la faible à la forte inversion, et d'autre part, elle permet d'expliquer l'origine du bruit basse fréquence associé au courant de grille.

v) **Le bruit RTS (Random Telegraph signal)**

Dans les dispositifs MOS, ce bruit provient de l'activité d'un piège unique à l'interface Si-SiO₂. La modulation de la conductance du canal induit par la capture d'un porteur se traduit donc par des signaux rectangulaires directement observable dans le domaine temporel. Dans le domaine fréquentiel, le spectre est constitué d'une ou plusieurs Lorentziennes, chacune d'entre elles étant caractéristique d'un piège actif pour des conditions de polarisation bien définies de la structure.

Dans ce chapitre, nous allons focaliser notre étude sur le bruit en $1/f$ associé au courant du drain, de la source et de la grille. Nous présenterons également la caractérisation et la modélisation du bruit de type RTS observé sur le courant de grille.

4.3 Modélisation du phénomène de piégeage/dépiégeage

Dans le cas où on considère que le bruit en $1/f$ dans les transistors MOS est dû à la fluctuation du nombre de porteurs et de la mobilité corrélée, la densité spectrale de bruit des courants dans le transistor MOS dépend essentiellement du modèle de piégeage – dépiégeage d'un porteur libre par un piège d'oxyde. Dans cette partie nous considérerons une structure MOS unidimensionnelle avec un substrat de type p polarisée de la faible à la forte inversion, dans le but de vérifier la validité du modèle couramment utilisé de piégeage - dépiégeage des porteurs libres de la couche d'inversion par les pièges d'oxyde pour les structures à oxydes ultrafins. Dans un premier temps, nous décrirons le processus de capture d'un électron de la couche d'inversion par un piège d'oxyde en régime de forte inversion, puis nous détaillerons le processus de capture en faible inversion.

4.3.1 Modélisation du piégeage en forte inversion

Le modèle de fluctuation du nombre de porteurs a été proposé pour expliquer le bruit en $1/f$ dans le germanium par McWhorter [McWHORTER'57]. L'origine des fluctuations est due au mécanisme de piégeage dépiégeage par des états lents situés dans l'oxyde de grille à proximité de l'interface Si-SiO₂.

Soient $N_t(E, x, y, z)$ le nombre de pièges dans l'oxyde par unité de volume et par unité d'énergie et $n_t(E, x, y, z, t)$ le nombre de pièges occupés. Par la suite, du fait que la structure est considérée comme unidimensionnelle, nous utiliserons les notations $N_t(E, y)$ et $n_t(E, y, t)$.

La Figure 4.1 illustre le mécanisme de piégeage d'un électron de la couche d'inversion par une densité de piège $N_t(E, y)$ située à la profondeur y et d'énergie E .

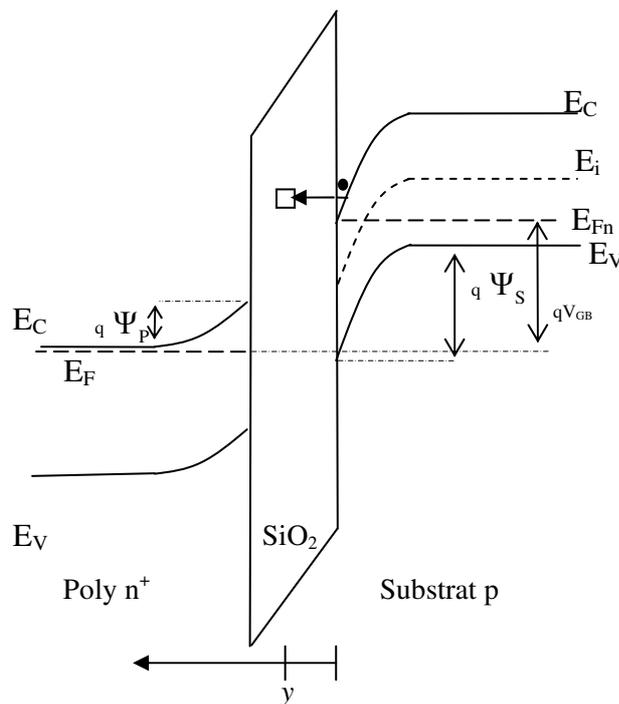


Figure 4.1 : Diagramme de bande d'une structure nMOS, illustrant le mécanisme de piégeage - dépiégeage d'un piège d'oxyde en régime de forte inversion

La probabilité élémentaire qu'un piège soit occupé par un électron dépend de sa position énergétique par rapport au quasi niveau de Fermi E_{F_n} . Sa probabilité d'occupation est donnée par la statistique de Fermi Dirac :

$$f_i(E) = \frac{1}{1 + \exp\left(\frac{E - E_{F_n}}{kT}\right)} \quad (4.4)$$

Si les pièges sont indépendants les uns des autres, on peut supposer que le nombre d'électrons piégés suit une loi binomiale. On a alors :

$$\langle n_i(E, x, y, z, t) \rangle = n_{i_0}(E, x, y, z) = N_i(E, x, y, z) f_i(E) \quad (4.5)$$

$$\langle n_i(E, x, y, z, t)^2 \rangle = N_i(E, x, y, z) f_i(E) [1 - f_i(E)] \quad (4.6)$$

Le mécanisme de capture et d'émission d'un porteur suit la statistique Shockley-Read-Hall (SRH) [SHOCKLEY'52]. La probabilité par unité de temps et par unité de surface qu'un électron soit capturé est égale au produit de la probabilité de capture $c_n(E, y)$, du nombre de pièges libres $N_i(E, y) - n_i(E, y, t)$ et du nombre de porteurs libres n . Dans le cas où la couche d'inversion est quantifiée, le flux incident contre l'interface Si-SiO₂ est définie comme le produit de la charge présente sur le premier niveau d'énergie ε et la fréquence d'impact $f(\varepsilon)$. Dans ces conditions, le taux de capture par unité de temps et par unité de surface est donc :

$$g(n_i) = \sigma_n f(\varepsilon) N_{inv}(\varepsilon) T(\varepsilon, y) (N_i(E, y) - n_i(E, y, t)) \quad (4.7)$$

où σ_n est la section de capture du piège, $N_{inv}(\varepsilon)$ le nombre de porteurs libres situés sur le premier niveau d'énergie E et $T(E, y)$ la probabilité tunnel d'accéder au piège situé en y :

$$T(\varepsilon, y) = \frac{4v(\varepsilon)v_{ox}(\varepsilon)}{v(\varepsilon)^2 + v_{ox}(\varepsilon)^2} \times \exp\left(-4 \frac{\sqrt{2m_{ox}^*} T_{ox}}{3q\hbar V_{ox}} \left[(q\chi_B - \varepsilon)^{3/2} - \left(q\chi_B - qV_{ox} \frac{y}{T_{ox}} - \varepsilon \right)^{3/2} \right]\right) \quad (4.8)$$

Dans l'expression (4.8), nous avons tenu compte de la réflexion à l'interface Si-SiO₂, en prenant pour $v(\varepsilon)$ et $v_{ox}(\varepsilon)$ les mêmes expressions que celles définies au chapitre 2.

Le taux d'émission par unité de temps et par unité de surface est donné par :

$$r(n_i) = e_n n_i(E, y, t) \quad (4.9)$$

Le nombre de pièges occupés $n_i(E, y, t)$ vérifie donc l'équation différentielle suivante :

$$\frac{dn_i(E, y, t)}{dt} = g(n_i) - r(n_i) + h(t) \quad (4.10)$$

où $h(t)$ représente une source aléatoire de bruit de Langevin.

En régime d'équilibre, on a la relation $g(n_{t_0}) = r(n_{t_0})$. L'équation différentielle relative aux fluctuations $\delta n_t(E, y, t)$ autour de la valeur moyenne $n_{t_0}(E, y)$ s'écrit :

$$\frac{d\delta n_t(E, y, t)}{dt} = - \left[\left. \frac{\partial r}{\partial n_t} \right|_{n_{t_0}} - \left. \frac{\partial g}{\partial n_t} \right|_{n_{t_0}} \right] \delta n_t(E, y, t) + \gamma(t) \quad (4.11)$$

Cette équation peut être écrite sous la forme classique :

$$\begin{aligned} \frac{d\delta n_t(E, y, t)}{dt} + \frac{\delta n_t(E, y, t)}{\tau(E, y)} &= \gamma(t) \\ \tau &= \left[\left. \frac{\partial r}{\partial n_t} \right|_{n_{t_0}} - \left. \frac{\partial g}{\partial n_t} \right|_{n_{t_0}} \right]^{-1} = \left[\frac{\sigma_n f(\varepsilon) N_{inv}(\varepsilon) T(\varepsilon, y)}{f_t(E)} \right]^{-1} \end{aligned} \quad (4.12)$$

En combinant (4.12) et (4.8) constante de temps τ s'exprime donc sous la forme :

$$\begin{aligned} \tau(E, \varepsilon, y) &= \left[\frac{f_t(E)}{\sigma_n f(\varepsilon) N_{inv}(\varepsilon)} \frac{v(\varepsilon)^2 + v_{ox}(\varepsilon)^2}{4v(\varepsilon)v_{ox}(\varepsilon)} \right] \\ &\times \exp \left(4 \frac{\sqrt{2m_{ox}^*} T_{ox}}{3q\hbar} \frac{V_{ox}}{V_{ox}} \left[(q\chi_B - \varepsilon)^{3/2} - \left(q\chi_B - qV_{ox} \frac{y}{T_{ox}} - \varepsilon \right)^{3/2} \right] \right) \end{aligned} \quad (4.13)$$

A partir de l'équation (4.12), la densité spectrale de la fluctuation du nombre de pièges occupés à la profondeur y est donnée par :

$$\Delta S_{n_t}(f, y, E) = 4 \left\langle \delta n_t(E, y)^2 \right\rangle \frac{\tau(E, \varepsilon, y)}{1 + \omega^2 \tau(E, \varepsilon, y)^2} = 4 f_t(E) [1 - f_t(E)] N_t(E, y) \frac{\tau(E, \varepsilon, y)}{1 + \omega^2 \tau(E, \varepsilon, y)^2} \quad (4.14)$$

Pour obtenir la densité spectrale de la fluctuation du nombre total de pièges occupés, il suffit d'intégrer (4.14) sur l'énergie et sur l'épaisseur de l'oxyde. Dans un premier temps, nous allons nous intéresser à l'intégration sur l'énergie. Le produit $f_t(E)[1 - f_t(E)]$ se comporte comme une fonction qui n'est différente de zéro que pour $E_{F_n} - kT/2 < E < E_{F_n} + kT/2$, et son intégrale est égale à :

$$\int_{-\infty}^{\infty} f_t(E) [1 - f_t(E)] dE = kT \quad (4.15)$$

Le densité spectrale de la fluctuation du nombre de pièges à la profondeur y occupés intégrée sur l'énergie est donc donnée par :

$$\begin{aligned}
S_n(f, y) &= \int 4f_t(E)[1-f_t(E)]N_t(E, y) \frac{\tau(E, \varepsilon, y)}{1+\omega^2\tau(E, \varepsilon, y)^2} dE \\
&\simeq 4kTN_t(E_{F_n}, y) \frac{\tau(E_{F_n}, \varepsilon, y)}{1+\omega^2\tau(E_{F_n}, \varepsilon, y)^2}
\end{aligned} \tag{4.16}$$

Pour établir cette expression, nous n'avons pas pris en compte la dépendance à l'énergie E de la constante de temps τ . L'intégration numérique de (4.16) et l'approximation en prenant $E = E_{F_n}$ pour τ donnent des résultats similaires ; cette hypothèse permet donc de simplifier les calculs sans faire trop d'erreur.

Ce résultat montre que seuls les pièges situés quelques kT autour du quasi niveau de Fermi ont un impact sur la fluctuation du nombre de pièges occupés. On peut donc simplifier l'approche du phénomène de piégeage - dépiégeage en définissant une densité effective de pièges $N_t'(E_{F_n}, y) = 4kTN_t(E_{F_n}, y)$, et en remplaçant E par E_{F_n} dans les expressions précédemment établies[HOU'03] ; ainsi il n'y a plus à considérer une intégration suivant l'énergie :

$$\begin{aligned}
S_n(f, y) &= 4f_t(E_{F_n})[1-f_t(E_{F_n})]N_t'(E_{F_n}, y) \frac{\tau(E_{F_n}, \varepsilon, y)}{1+\omega^2\tau(E_{F_n}, \varepsilon, y)^2} \\
&= 4\left(\frac{1}{2}\right)\left[1-\frac{1}{2}\right]4kTN_t(E_{F_n}, y) \frac{\tau(E_{F_n}, \varepsilon, y)}{1+\omega^2\tau(E_{F_n}, \varepsilon, y)^2} \\
&= 4kTN_t(E_{F_n}, y) \frac{\tau(E_{F_n}, \varepsilon, y)}{1+\omega^2\tau(E_{F_n}, \varepsilon, y)^2}
\end{aligned} \tag{4.17}$$

A présent, intéressons nous à l'intégration sur la profondeur. Le modèle classiquement utilisé pour décrire la constante de temps τ est une fonction exponentiellement croissante avec la profondeur. Cette formulation est issue d'un développement de Taylor à l'ordre 1 de la probabilité :

$$\begin{aligned}
\tau(E, \varepsilon, y) &= \tau_0 \times \exp\left(4\frac{\sqrt{2m_{ox}^*} T_{ox}}{3q\hbar} \frac{T_{ox}}{V_{ox}} \left[(q\chi_B - \varepsilon)^{3/2} - \left(q\chi_B - qV_{ox} \frac{y}{T_{ox}} - \varepsilon \right)^{3/2} \right] \right) \\
&\simeq \tau_0 \exp(\gamma y) \quad \text{avec } \gamma = 2\frac{\sqrt{2m_{ox}^*} (q\chi_B - \varepsilon)}{\hbar}
\end{aligned} \tag{4.18}$$

Cette approche mène à la valeur classique de $\gamma \simeq 10^{10} \text{ m}^{-1}$ pour une barrière de potentiel de 3 eV ($m_{ox}=0.5 m_0$, $\varepsilon = 0$).

En toute rigueur, nous voyons que τ_0 et γ dépendent de la polarisation. Etudions cette dépendance pour déterminer la validité de l'approximation $\tau \approx \tau_0 \exp(\gamma y)$.

La figure 4.2 montre les constantes de temps en fonction de la profondeur, calculées avec l'équation (4.13) et l'équation simplifiée (4.18), pour plusieurs valeurs de la tension V_{GB} .

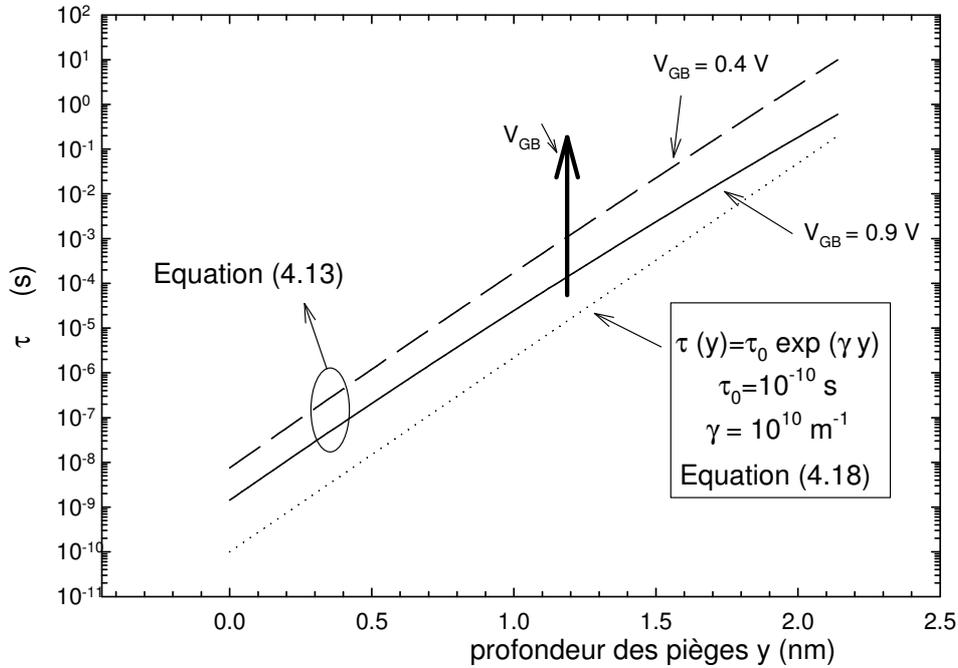


Figure 4.2 : Evolution de la constante de temps τ avec la profondeur des pièges et la polarisation de grille, pour une structure nMOS avec $T_{\text{ox}}^{\text{SiON}} = 2.14 \text{ nm}$, $N_A = 10^{18} \text{ cm}^{-3}$, $N_P = 10^{20} \text{ cm}^{-3}$.

On observe donc que l'évolution de la constante de temps calculée avec l'expression complète (4.13) est de type exponentielle, ce qui permet de justifier l'approximation $\tau(y) = \tau_0 \exp(\gamma y)$. Cependant, la Figure 4.2 montre que les valeurs de τ_0 et de γ ne doivent pas être prises constantes et qu'elles dépendent fortement de la polarisation de grille. Pour déterminer quelles valeurs donner à τ_0 et γ , nous devons établir l'expression de la densité spectrale de la fluctuation du nombre de pièges d'oxyde occupés. Dans le cas où on considère une distribution spatiale de pièges uniforme, l'approximation (4.18) permet de déterminer analytiquement l'expression de la densité spectrale du nombre de pièges occupés :

$$S_{N_t} = \int_0^{y_1} 4kTN_t(E_{Fn}) \frac{\tau(y)}{1 + \omega^2 \tau(y)^2} dy = \frac{4kTN_t(E_{Fn})}{\gamma \omega} [\arctan(\omega \tau_1) - \arctan(\omega \tau_0)] \quad (4.19)$$

où y_1 représente la profondeur maximale des pièges et $\tau_1 = \tau_0 \exp(\gamma y_1)$.

Pour $\frac{1}{\tau_1} < \omega < \frac{1}{\tau_0}$, la différence des arctangentes est constante et égale à $\pi/2$, et on

obtient une densité spectrale inversement proportionnelle à la fréquence :

$$S_{N_t}(f) = \frac{kTN_t(E_{F_n})}{\gamma f} \quad (4.20)$$

Dans cette formulation, la constante τ_0 n'intervient plus. La première remarque importante concerne la profondeur y_1 nécessaire pour obtenir une densité spectrale en $1/f$ à 1 Hz. Nous avons reporté sur la figure 4.3a la densité spectrale calculée par l'équation (4.19) pour y_1 égal à l'épaisseur équivalente (EOT) 1.2 nm, et pour y_1 égal à l'épaisseur physique 2.1 nm, c'est-à-dire pour les paramètres extraits du lot DPN12. Dans le cas où $y_1 = EOT$, la fréquence $1/2\pi\tau_1$ est supérieure à 1 Hz. Pour obtenir un spectre en $1/f$ à la fréquence de 1Hz, il faut considérer l'épaisseur physique du diélectrique et non son épaisseur électrique équivalente (EOT). Nous avons donc utilisé pour nos simulations l'épaisseur physique, ainsi que la barrière de potentiel χ_B et la masse effective de l'électron dans l'oxyde déterminées au chapitre 3.

La figure 4.3b présente la densité spectrale calculée par l'équation (4.20) et par intégration numérique des équations (4.17) et (4.13). La courbe notée « modèle analytique » a été obtenue avec $\gamma = 10^{10} \text{ m}^{-1}$. L'observation de l'évolution de la constante de temps avec la tension V_{GB} (figure 4.2), nous a montré qu'en toute rigueur, il fallait considérer une valeur de γ dépendante de V_{GB} . Sur la figure 4.3b, on observe une augmentation de $1/2\pi\tau_1$ quand on augmente V_{GB} , mais l'augmentation de γ avec V_{GB} a très peu d'impact sur la partie en $1/f$ de la densité spectrale. De plus, les mesures de bruit ne sont réalisées que pour une gamme de fréquence de 1 Hz à 100 kHz ; l'observation du plateau pour des fréquences inférieures à $1/2\pi\tau_1$ est donc difficile. Nous pouvons donc considérer, pour les oxydes ultrafins, que l'approximation (4.20) est valable pour décrire la densité spectrale de l'occupation des pièges d'oxyde dans la gamme 1 Hz – 100 kHz, avec $\gamma = 10^{10} \text{ m}^{-1}$.

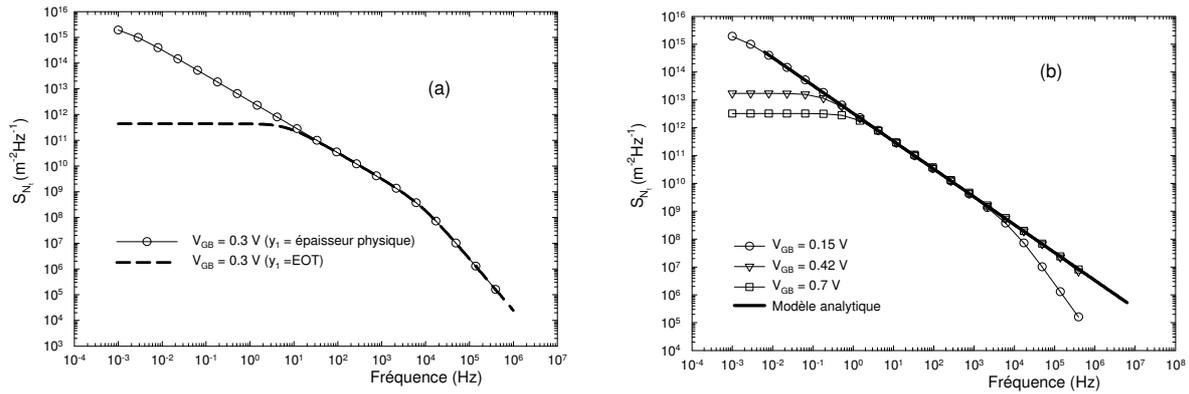


Figure 4.3 : Simulation de la densité spectrale de la fluctuation du nombre de pièges occupés pour une structure nMOS avec les paramètres technologiques du lot DPN12. La figure (a) présente l'impact de l'épaisseur dans l'expression (4.17), et la figure (b) montre l'impact de la prise en compte de la courbure de bande pour le calcul de la constante de temps τ .

4.3.2 Modélisation du piégeage en faible inversion

Nous nous plaçons dans le cas où la surface n'est pas dégénérée, c'est-à-dire quand l'énergie du quasi niveau de Fermi des électrons est inférieure au bas de la bande de conduction à l'interface Si/SiO₂. Le diagramme de bande représenté sur la figure 4.4 illustre le mécanisme de piégeage - dépiégeage des porteurs libres de la couche d'inversion dans cette configuration. Deux processus de capture peuvent être mis en jeu. Christensson [CHRISTENSSON'68] et al ont proposé un modèle où les porteurs traversent par effet tunnel à énergie constante jusqu'à la distance y_t du piège (processus (i)), puis sont capturés par le piège situé à l'énergie E_t , proche de E_{Fn} (processus (ii)). Ce modèle suppose qu'une dissipation d'énergie se produit dans l'oxyde. Les données expérimentales n'ont pas confirmé cette hypothèse [BURSTEIN'69]. Un autre modèle, proposé par Fu et Sah [FU'72], fait l'hypothèse que les porteurs sont tout d'abord piégés par les états d'interfaces rapides suivant un processus Shockley-Read-Hall (SRH), comme indiqué par les transitions (a) et (b) sur la figure 4.4. A partir de ces pièges d'interface, les porteurs peuvent alors accéder au piège d'oxyde par effet tunnel (processus (c) et (d)). Avec ce modèle, les porteurs ne dissipent pas d'énergie dans l'oxyde. Contrairement au modèle exposé au paragraphe 4.3.1, le processus aléatoire met en jeu deux variables aléatoires qui sont couplées. Dans ce paragraphe, nous allons détailler la modélisation de ce processus bi-variable pour déterminer si la prise en compte des états d'interface est nécessaire pour modéliser la fluctuation du nombre de pièges d'oxyde occupés.

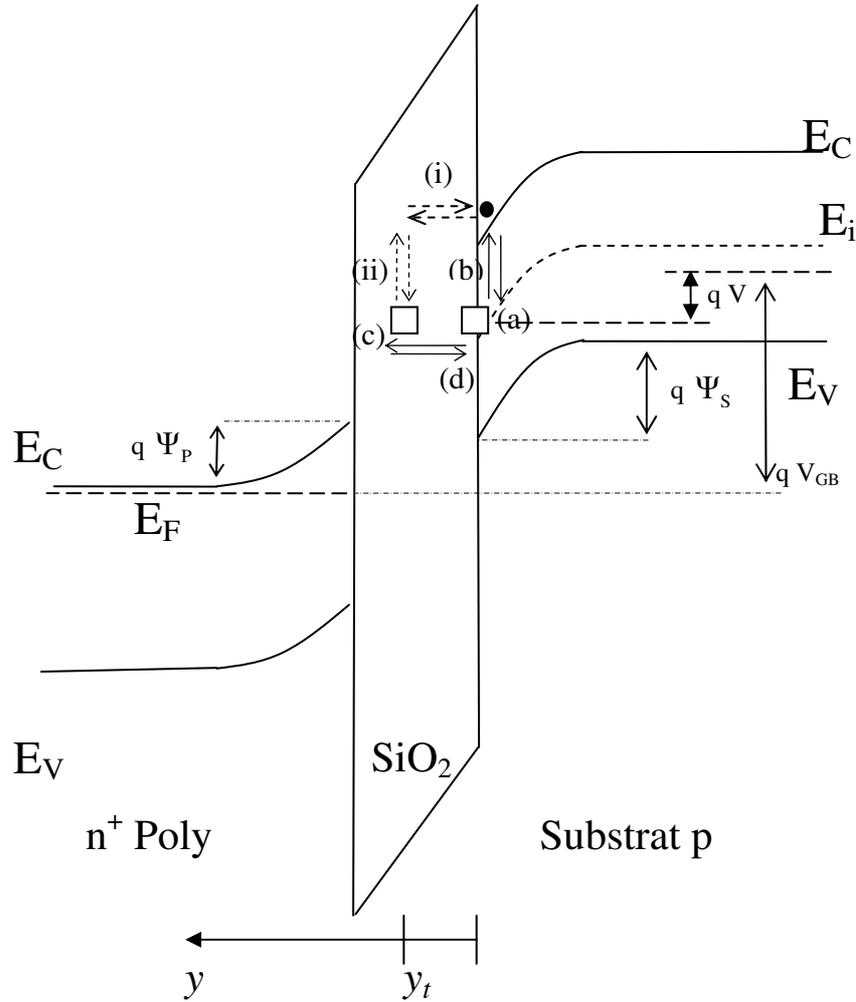


Figure 4.4 : Diagramme de bande d'une structure nMOS, illustrant les différents mécanismes de piégeage –dépiégeage d'un piège d'oxyde en régime de faible inversion.

Considérons $N_{it}(E)$ la densité de pièges d'interface par unité de surface et d'énergie, $n_{it}(E, t)$ le nombre de pièges d'interface occupés à l'instant t par unité de surface et d'énergie, $N_t(E, y)$ la densité de pièges d'oxyde à une profondeur y par unité de volume et d'énergie et $n_t(E, y, t)$ le nombre de pièges d'oxyde occupés à une profondeur y à l'instant t par unité de volume d'énergie. De la même façon que dans le paragraphe 4.3.1, nous définissons également $\delta n_t(E, y, t)$ et $\delta n_{it}(E, t)$ comme étant les fluctuations de l'occupation des pièges autour de la valeur moyenne :

$$\begin{aligned} n_{it}(E, t) &= f_t(E) N_{it}(E) + \delta n_{it}(E, t) \\ n_t(E, y, t) &= f_t(E) N_t(E, y) + \delta n_t(E, y, t) \end{aligned} \quad (4.21)$$

où $f_i(E)$ est la fonction d'occupation du piège. L'occupation moyenne du piège est définie par la statistique de Fermi, et s'exprime par :

$$f_i(E) = \frac{1}{1 + \exp\left(\frac{E - E_{F_n}}{kT}\right)} \quad (4.22)$$

Nous allons utiliser dans cette partie le formalisme des processus stochastiques multi-variables [VANVLIET'65]. On considère que le processus est markovien, c'est-à-dire que l'état futur du système (occupation des pièges d'interface et d'oxyde) ne dépend que de l'état présent. Tout le système est donc conditionné par la connaissance des probabilités de transition d'un état à un autre. De plus, on considère que pendant un temps $\Delta t \rightarrow 0$, les variables n_{it} ou n_t ne changent que d'une unité.

Nous devons formuler les taux de capture et d'émission des processus (a), (b), (c) et (d). De la même façon que dans le paragraphe 4.3.1, nous utiliserons les densités effectives de pièges (notées $N'_t, n'_t, N'_{it}, n'_{it}$), évitant ainsi une intégration suivant l'énergie.

Considérons maintenant le processus de capture d'électrons situés à l'interface dans la bande de conduction. La probabilité de capture s'exprime sous la forme $c_n = \sigma_{n,it} v_{th}$, avec σ_n la section de capture du piège et v_{th} la vitesse thermique de l'électron. Dans l'approche classique d'un processus de capture SRH, la charge libre susceptible d'être piégée est égale à la densité d'électrons libres à l'interface. A cause des effets quantiques, la charge n'est plus située à l'interface mais légèrement dans le substrat ; de ce fait, on peut définir une charge équivalente, tenant compte de la distribution spatiale des électrons à l'interface Si-SiO₂ [SIERGIEJ'92, CELIK'00]. Le processus de capture (a) est caractérisé par le taux de capture défini par :

$$r_a(E, t) = \sigma_{n,it} v_{th} \left(N'_{it}(E_{F_n}) - n'_{it}(E_{F_n}, t) \right) \bar{n} \quad (4.23)$$

avec

$$\begin{aligned} \bar{n} &= N_{inv} \frac{1}{y} \int_0^{\bar{y}} |\zeta(y)|^2 dy = N_{inv} \frac{b}{3} \int_0^{3/b} \frac{b^3}{2} y^2 \exp(-b \cdot y) \\ &= \frac{b \cdot N_{inv}}{3} \left(1 - \frac{17}{2} e^{-3} \right) \approx 0.19 b N_{inv} \end{aligned} \quad (4.24)$$

En définissant e_n comme étant le coefficient d'émission, le taux d'émission (b) s'exprime par la relation suivante :

$$r_b(E, t) = e_n n_{it}'(E_{F_n}, t) \quad (4.25)$$

A présent, complétons l'étude du processus de piégeage - dépiégeage en ajoutant les pièges d'oxyde. Le taux de capture associé au processus (c) $r_c(E, y, t)$ des pièges d'oxyde inclut la probabilité que l'électron situé dans un piège d'interface traverse par effet tunnel jusqu'à l'abscisse y .

$$r_c(E_{F_n}, y, t) = T(y) \left[N_{it}'(E_{F_n}, y) - n_{it}'(E_{F_n}, y, t) \right] n_{it}(E_{F_n}, t) \Delta y \quad (4.26)$$

La probabilité $T_{wkb}(E, y)$ est calculée suivant la formulation de Fu et Sah [FU'72], en prenant en compte la courbure de bande dans l'oxyde [HOU'03] :

$$T(E_{F_n}, y) = \frac{1}{\tau_0 N_{it}(E_{F_n})} \exp \left(-4 \frac{\sqrt{2qm_{ox}^* (q\chi_B - E_{F_n})}}{\hbar} y \right) \approx T_0 \exp(-\gamma y) \quad (4.27)$$

avec $\tau_0 \approx 10^{-10}$ s [CELIK'85]. Les énergies sont référencées par rapport au bas de la bande conduction à l'interface ; le quasi niveau de Fermi s'écrit donc :

$$E_{F_n} = +q\Psi_s - \frac{E_g}{2} - q\phi_F - qV \quad (4.28)$$

En utilisant la probabilité d'émission $e_{n_{ox}}$, le taux d'émission associé au processus (d) $r_d(E_{F_n}, y, t)$ des pièges d'oxyde s'exprime par :

$$r_d(E_{F_n}, y, t) = e_{n_{ox}} n_t'(E, y, t) \left(N_{it}'(E_{F_n}) - n_{it}'(E_{F_n}, t) \right) \Delta y \quad (4.29)$$

Les fluctuations du nombre de porteurs piégés par les pièges d'interface et d'oxyde satisfont le système d'équations suivant :

$$\begin{aligned} \frac{dn_{it}'(E_{F_n}, t)}{dt} &= r_a(E_{F_n}, t) - r_b(E_{F_n}, t) - \left[r_c(E_{F_n}, y, t) - r_d(E_{F_n}, y, t) \right] + \gamma_{it}(t) + \gamma_{ox}(t) \\ \frac{dn_t'(E_{F_n}, y, t)}{dt} dy &= r_c(E_{F_n}, y, t) - r_d(E_{F_n}, y, t) + \gamma_{ox}(t) \end{aligned} \quad (4.30)$$

Où $\gamma_{it}(t)$ et $\gamma_{ox}(t)$ sont les sources de bruit de Langevin décrivant respectivement les transitions aléatoires pour les pièges d'interface et les pièges d'oxyde. Avant d'établir les équations régissant les fluctuations $\delta n_{it}'(E_{F_n}, t)$ et $\delta n_t'(E_{F_n}, y, t)$, Nous pouvons simplifier (4.30) en explicitant le système d'équation relatif à la composante continu de n_{it}' et n_t' :

$$\begin{aligned} \frac{dn_{i_0}'(E, y_t)}{dt} = 0 &\Rightarrow r_{c_0}(E, y_t) = r_{d_0}(E, y_t) \\ \frac{dn_{i_0}'(E)}{dt} = 0 &\Rightarrow r_{a_0}(E) - r_{b_0}(E) = [r_{c_0}(E, y_t) - r_{d_0}(E, y_t)] = 0 \end{aligned} \quad (4.31)$$

La résolution de (4.31) permet d'éliminer e_n et $e_{n_{ox}}$:

$$\begin{aligned} c_n [N_{it}(E) - f_t(E)N_{it}(E)] \bar{n} = e_n f_t(E) N_{it}(E) &\Rightarrow e_n = \bar{n} c_n \frac{1 - f_t(E)}{f_t(E)} \\ T(y) [N_t(E, y_t) - f_t(E)N_t(E, y_t)] f_t(E) N_{it}(E) & \\ = e_{n_{ox}}(y) [N_{it}(E) - f_t(E)N_{it}(E)] f_t(E) N_t(E, y_t) &\Rightarrow T(y) = e_{n_{ox}}(y) \end{aligned} \quad (4.32)$$

Le système d'équations régissant les fluctuations $\delta n_{it}(E, t)$ et $\delta n_t(E, y, t)$ est obtenu en développant au premier ordre le système (4.30) :

$$\begin{aligned} \frac{d\delta n_{it}'(t)}{dt} &= \left[-\frac{c_n \bar{n}}{f_t} - T(y) N_t'(y) \Delta y \right] \delta n_{it}'(t) + \left[T(y) N_{it}' \Delta y \right] \delta n_t'(y, t) + \gamma_{it}(t) + \gamma_{ox}(t) \\ \frac{d\delta n_t'(y, t)}{dt} &= \left[T(y) N_t'(y) \right] \delta n_{it}'(t) + \left[-T(y) N_{it}' \right] \delta n_t'(y, t) + \gamma_{ox}(t) \end{aligned} \quad (4.33)$$

Le système (4.33) décrit avec le formalisme de Langevin un processus aléatoire bidimensionnel. Dans le but de calculer la densité spectrale d'occupation des pièges d'oxyde, nous reformulons le système (4.33) sous une forme matricielle :

$$\frac{d}{dt} \begin{bmatrix} \delta n_t(E_{F_n}, y, t) \Delta y \\ \delta n_{it}(E_{F_n}, t) \end{bmatrix} = -[M] \begin{bmatrix} \delta n_t(E_{F_n}, y, t) \Delta y \\ \delta n_{it}(E_{F_n}, t) \end{bmatrix} + \begin{bmatrix} \gamma_1(t) \\ \gamma_2(t) \end{bmatrix} \quad (4.34)$$

Cette formulation matricielle a été obtenue de façon directe en détaillant les probabilités par unité de temps liées à chaque transition, puis en linéarisant le système d'équations donnant l'évolution temporelle de l'occupation des pièges. Il existe une méthode systématique pour obtenir d'une part, le système (4.34) et d'autre part, aboutir au calcul des densités spectrales d'occupations des pièges [VanVliet'65]. Nous ne développerons pas dans cette thèse les détails mathématiques de cette formulation générale des problèmes de génération recombinaison pour un nombre quelconque de variables aléatoires. Nous rappellerons simplement les résultats qui permettent d'obtenir les densités spectrales.

La matrice M est appelée matrice phénoménologique de relaxation. La matrice de corrélation est définie par :

$$\Phi(s) = \langle \delta_n(t) \delta_n(t+s) \rangle$$

$$\text{avec } \delta_n(t) = \begin{bmatrix} \delta n_i(E, y_i, t) \Delta y \\ \delta n_{ii}(E, t) \end{bmatrix} \quad (4.35)$$

Le système d'équations (4.34) s'écrit sous la forme :

$$\frac{d\delta_n(t)}{dt} = -M\delta_n(t) + \Upsilon(t) \quad (4.36)$$

avec $\Upsilon(t)$ représentant l'ensemble des sources aléatoires. Cette relation est également vérifiée par la matrice de corrélation, de sorte que :

$$\frac{d\Phi(t)}{dt} = -M\Phi(t) \quad (4.37)$$

La matrice de corrélation s'exprime simplement par :

$$\Phi(t) = \Phi(0) \exp(-Mt) \quad (4.38)$$

En appliquant le théorème de Wiener-Khintchine , on obtient la matrice des densités spectrales S :

$$S(\omega) = 2 \int_{-\infty}^{\infty} \Phi(s) e^{j\omega s} ds \quad (4.39)$$

Les densités spectrales $S_{\delta n_i \delta n_i}$ et $S_{\delta n_{ii} \delta n_{ii}}$ sont réelles alors que les densités spectrales croisées $S_{\delta n_i \delta n_{ii}}$ et $S_{\delta n_{ii} \delta n_i}$ sont complexes conjuguées. En posant que G est la partie réelle de S, on obtient :

$$G(\omega) = 4 \operatorname{Re} \int_0^{\infty} \Phi(s) e^{j\omega s} ds = 4 \operatorname{Re} (M - j\omega I)^{-1} \Phi(0)$$

$$\text{avec } \Phi(0) = \begin{bmatrix} \langle \delta n_i^2 \rangle & \langle \delta n_i \delta n_{ii} \rangle \\ \langle \delta n_{ii} \delta n_i \rangle & \langle \delta n_{ii}^2 \rangle \end{bmatrix} \quad (4.40)$$

Le calcul des variances $\Phi(0)$ peut se faire à partir de la connaissance des différentes probabilités au moyen de la « *master equation* », en postulant que toutes les transitions pendant un temps dt ne font augmenter ou diminuer le nombre de pièges occupés que d'une seule unité. Dans ces conditions, on définit la matrice B à partir des probabilités de transitions :

$$B = \begin{bmatrix} r_c + r_d & -(r_c + r_d) \\ -(r_c + r_d) & r_c + r_d + r_a + r_b \end{bmatrix} \quad (4.41)$$

La relation entre la matrice des variances et la matrice B est donnée par la relation $\Phi(0) = \frac{1}{2} M^{-1} B$. On obtient donc les variances à partir uniquement des probabilités de transition ;

$$\begin{aligned} \langle \delta n_t^2 \rangle &= N_t(E, y_t) f_t(E) (1 - f_t(E)) \\ \langle \delta n_{it}^2 \rangle &= N_{it}(E) f_t(E) (1 - f_t(E)) \\ \langle \delta n_{it} \delta n_t \rangle &= \langle \delta n_t \delta n_{it} \rangle = 0 \end{aligned} \quad (4.42)$$

Les spectres se calculent donc en suivant la formulation (4.40). En utilisant les valeurs propres $1/\tau_1$ et $1/\tau_2$ de la matrice M, on obtient les spectres $G_{\delta n_t, \delta n_t}$, $G_{\delta n_t, \delta n_{it}}$ et $G_{\delta n_{it}, \delta n_{it}}$. Une formulation analytique complète mène à des expressions lourdes, mais cette approche matricielle s'utilise facilement dans le cadre de techniques numériques de calculs. Dans le cas où nous nous intéressons à la densité spectrale de la fluctuation de l'occupation des pièges d'oxyde à une profondeur y, on obtient :

$$\begin{aligned} S_{n_t} = G_{\delta n_t, \delta n_t} &= 2 \left[\frac{\tau_1 \tau_2}{\tau_2 - \tau_1} \frac{\tau_1^2}{1 + \omega^2 \tau_1^2} \left[\left(\frac{1}{\tau_1} - M_{22} \right) B_{11} + M_{12} B_{12} \right] + \frac{\tau_1 \tau_2}{\tau_1 - \tau_2} \frac{\tau_2^2}{1 + \omega^2 \tau_2^2} \left[\left(\frac{1}{\tau_2} - M_{22} \right) B_{11} + M_{12} B_{12} \right] \right] \\ &= 4 \left[\frac{\tau_1 \tau_2}{\tau_2 - \tau_1} \frac{\tau_1^2}{1 + \omega^2 \tau_1^2} \left[r_d \left(\frac{1}{\tau_1} - M_{22} - M_{12} \right) \right] + \frac{\tau_1 \tau_2}{\tau_1 - \tau_2} \frac{\tau_2^2}{1 + \omega^2 \tau_2^2} \left[r_d \left(\frac{1}{\tau_2} - M_{22} - M_{12} \right) \right] \right] \end{aligned} \quad (4.43)$$

L'équation (4.43) représente la même quantité que l'équation (4.17), mais en tenant compte des états d'interfaces dans le processus de capture. Il apparaît que pour des pièges d'oxyde de profondeur y, on obtient une lorentzienne liée aux états d'interface, et une autre liée aux pièges d'oxyde. La figure 4.5 présente l'évolution de l'inverse des valeurs propres $1/\tau_1$ et $1/\tau_2$ en fonction de la profondeur. Il apparaît que les deux constantes de temps sont très différentes. La constante de temps liée aux états d'interface est constante, alors que celle liée aux pièges d'oxyde augmente exponentiellement avec la profondeur. Le résultat important réside dans la faible interdépendance des constantes de temps. En effet, la modification des paramètres propres aux états d'interface (section de capture, densité d'états d'interface) ne modifie pas les constantes de temps de la lorentzienne associée aux pièges d'oxyde, et inversement.

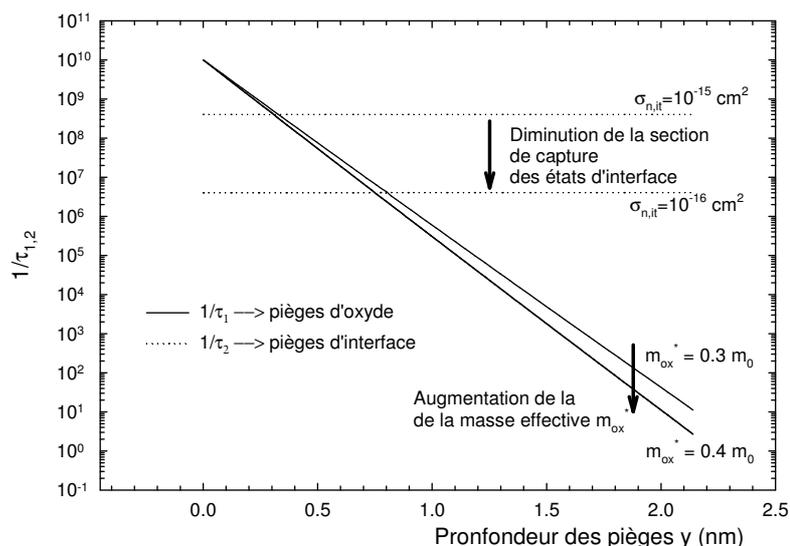


Figure 4.5 : Constantes de temps des Lorentziennes associées aux pièges d'oxyde et d'interface, en fonction de la profondeur du piège d'oxyde, pour une structure nMOS polarisée en faible inversion ($V_{GB} = 0.15$ Volts) avec $T_{ox}^{SiON} = 2.14$ nm, $N_A = 10^{18}$ cm $^{-3}$, $N_P = 10^{20}$ cm $^{-3}$, $N_t = 10^{18}$ cm $^{-3}$ eV $^{-1}$ et $N_{it} = 10^{12}$ cm $^{-2}$ eV $^{-1}$

La figure 4.6 montre la densité spectrale de l'occupation des pièges d'oxydes calculés par l'intégration suivant la profondeur de l'équation (4.43). Nous avons également reporté le modèle analytique formulé pour la forte inversion (4.20). On observe un très bon accord entre ces deux calculs de la densité spectrale, confirmant ainsi que l'activité des pièges d'interfaces est négligeable dans la modélisation de la fluctuation de l'occupation des pièges d'oxyde.

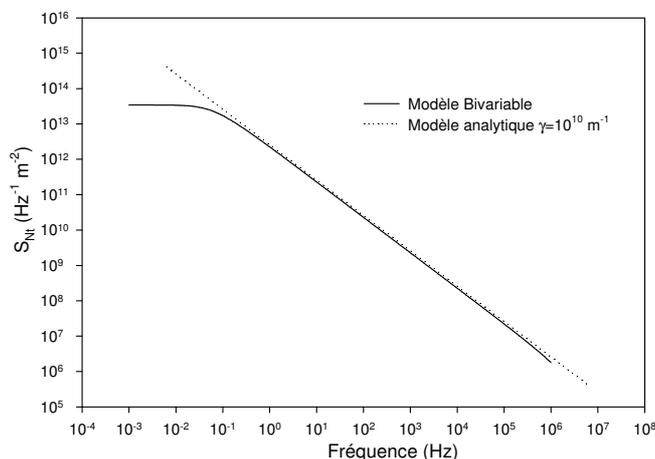


Figure 4.6 : Densité spectrale des fluctuations d'occupation des pièges d'oxyde, en tenant compte du des états d'interface (modèle bivariable) comparé au modèle analytique (4.20) avec $\gamma = 10^{10}$ m $^{-1}$, pour une structure nMOS polarisée en faible inversion ($V_{GB} = 0.15$ Volts) avec $T_{ox}^{SiON} = 2.14$ nm, $N_A = 10^{18}$ cm $^{-3}$, $N_P = 10^{20}$ cm $^{-3}$, $N_t = 10^{18}$ cm $^{-3}$ eV $^{-1}$, $N_{it} = 10^{12}$ cm $^{-2}$ eV $^{-1}$ et $\sigma_{n,it} = 10^{-15}$ cm 2 .

Le modèle de piégeage - dépiégeage en régime de faible inversion peut donc se modéliser avec une équation de la forme (4.20) sans tenir compte des états d'interface par lesquels passent les électrons vers les pièges d'oxyde. Pour la suite de ce mémoire, nous utiliserons, pour la densité spectrale (par unité de surface) d'occupation des pièges d'oxyde,

$$\text{l'équation } S_{N_t}(f) = \frac{kTN_t(E_{F_n})}{\gamma f}$$

4.4 Formulation du bruit dans les transistors MOS

4.4.1 Densité spectrale des fluctuations de la tension de bandes plates

Nous avons donc établi la densité spectrale de la fluctuation du nombre de pièges occupés dans l'oxyde. La fluctuation du nombre de porteurs piégés dans l'oxyde est équivalente à une fluctuation de la tension de bandes plates V_{FB} de la structure :

$$V_{FB} = \Phi_{ms} - \frac{Q_{ox}}{C_{ox}} \quad (4.44)$$

Pour une surface élémentaire $\Delta x \Delta z$, les fluctuations de la tension de bandes plates sont liées à la fluctuation du nombre total de pièges occupés ΔN_t , et s'expriment par :

$$\delta V_{FB} = -\frac{q \Delta N_t}{\Delta x \Delta z C_{ox}} \quad (4.45)$$

La densité spectrale des fluctuations du nombre total de pièges occupés s'exprime par :

$$S_{\Delta N_t}(f) = S_{N_t}(f) \Delta x \Delta z \quad (4.46)$$

La densité spectrale de la fluctuation de bandes plates pour une surface élémentaire $\Delta x \Delta z$ s'écrit donc :

$$S_{V_{FB}}(f) = \frac{q^2}{C_{ox}^2} \frac{kTN_t(E_F)}{\Delta x \Delta z \gamma f} \quad (4.47)$$

De plus, on peut ramener les fluctuations de bandes plates à la fluctuation du potentiel de surface. Dans la formulation du potentiel de surface, la tension de grille V_{GB} n'intervient jamais seule, mais toujours sous la forme $V_{GB} - V_{FB}$. Des fluctuations de la tension de bandes plates sont, d'un point de vue mathématique, équivalentes à une fluctuation de la tension de grille V_{GB} . On obtient donc :

$$S_{\Psi_s}(f) = \left(\frac{\partial \Psi_s}{\partial V_{FB}} \right)^2 S_{V_{FB}}(f) = \left(\frac{\partial \Psi_s}{\partial V_{GB}} \right)^2 S_{V_{FB}}(f) = \left(\frac{\partial F / \partial V_{GB}}{\partial F / \partial \Psi_s} \right)^2 S_{V_{FB}}(f) \quad (4.48)$$

où F est la fonction implicite liant Ψ_s et V_{GB} (cf. chapitre 2).

4.4.2 Densité spectrale des fluctuations du courant de drain.

4.4.2.1 Modèle ΔN

Le modèle physique qui attribue l'origine du bruit en $1/f$ aux fluctuations du nombre de porteurs est appelé couramment modèle ΔN . Considérons un transistor nMOS sans fuite de grille, mais en tenant compte des effets quantiques et de la polydépétion. En chaque point du canal, la densité de pièges occupés fluctue ; nous avons montré au paragraphe 4.4.1 que la fluctuation du nombre de pièges d'oxyde occupés est équivalente à une fluctuation de la tension de bandes plates. Dans la formulation de la charge d'inversion dans le cadre de l'approximation de la feuille de charge, nous avons obtenu une expression fonction de la tension $V_{GB}-V_{FB}$ et du potentiel de surface Ψ_s . Grâce à notre modèle petit signal du transistor établi au chapitre 3, nous avons également à notre disposition l'amplitude de la variation de tous les paramètres par rapport à une variation de la tension de grille. Compte tenu que la tension de grille est toujours présente dans les équations sous la forme $V_{GB}-V_{FB}$, la dérivée d'une grandeur physique par rapport à la tension de grille V_{GB} équivaut à l'opposé de cette même grandeur par rapport à V_{FB} .

La fluctuation de la charge des porteurs libres est donc liée à la fluctuation de la charge du nombre de pièges occupés par la relation :

$$\delta Q_{inv} = \frac{dQ_{inv}}{dV_{FB}} \delta V_{FB} = - \frac{dQ_{inv}}{dV_{GB}} \times \left(- \frac{1}{C_{ox}} \delta Q_t \right) \quad (4.49)$$

On définit le coefficient $R(x)$ en tout point du canal comme étant le rapport entre la fluctuation des charges libres et la fluctuation des pièges d'oxyde occupés :

$$R(x) = \frac{\delta Q_{inv}(x)}{\delta Q_t(x)} = \frac{dQ_{inv}}{dV_{GB}} \Big|_x \frac{1}{C_{ox}} \quad (4.50)$$

La figure 4.7 présente l'évolution du coefficient R dans le cas d'une structure MOS unidimensionnelle pour évaluer l'impact des effets quantiques et de la polydépétion sur le coefficient R . Ce coefficient, introduit par Jindal et Van der Ziel [JINDAL'78] est habituellement considéré comme très petit en faible inversion ($\delta Q_{inv} \ll \delta Q_t$) et proche de -1

en forte inversion ($\delta Q_{inv} \approx -\delta Q_i$). Ces approximations sont valables dans le cadre d'une approche classique, c'est-à-dire sans effets quantiques ni polydéplétion. Lorsque ces effets sont pris en compte, l'approximation en forte inversion $R \approx -1$ n'est plus valable.

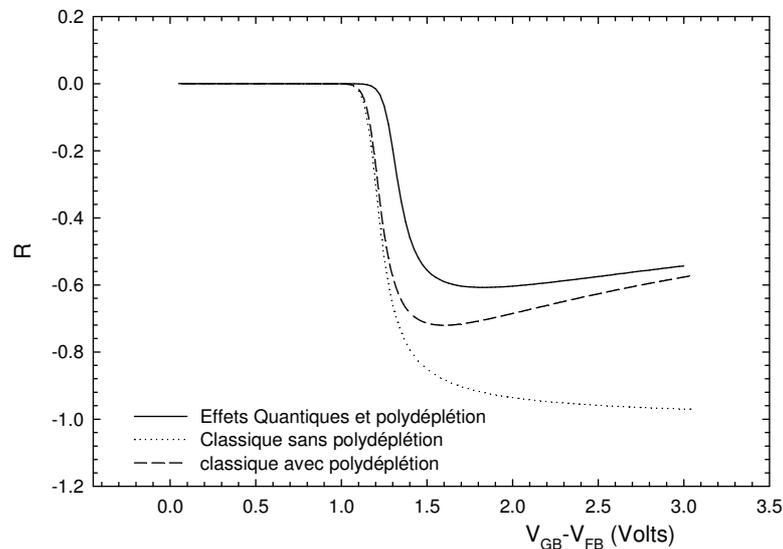


Figure 4.7 : Evolution du coefficient R avec la tension de grille avec la prise en compte des effets quantiques et de la polydéplétion, pour une structure nMOS d'épaisseur d'oxyde 1.2 nm, $N_A=10^{18} \text{ cm}^{-3}$ et $N_P=10^{20} \text{ cm}^{-3}$.

La figure 4.8 présente le coefficient R le long du canal pour plusieurs polarisations de grille à $V_{DS} = 25 \text{ mV}$, pour un transistor nMOS avec une épaisseur d'oxyde de 1.2 nm, $W/L=10 \mu\text{m} / 10 \mu\text{m}$, $N_A=10^{18} \text{ cm}^{-3}$, $N_P=10^{20} \text{ cm}^{-3}$.

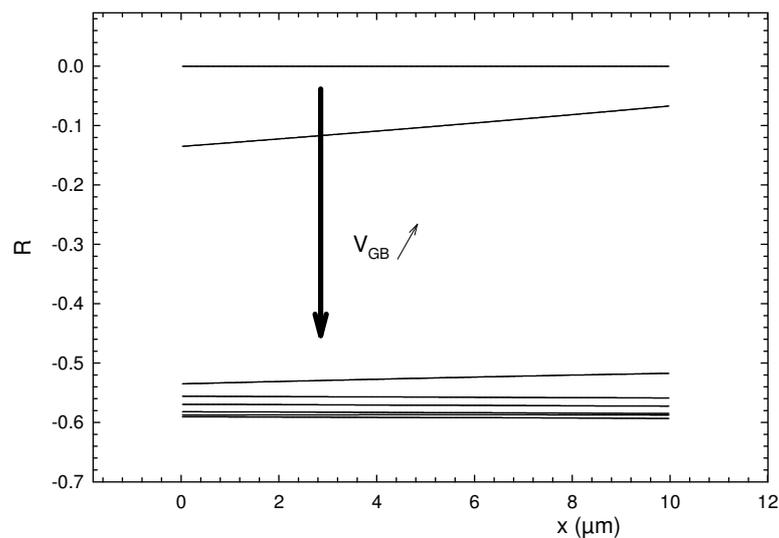


Figure 4.8 : Evolution du coefficient R le long du canal pour plusieurs tensions de grille V_{GB} à $V_{DS} = 25 \text{ mV}$, pour un transistor nMOS d'épaisseur d'oxyde 1.2 nm, $N_A=10^{18} \text{ cm}^{-3}$ et $N_P=10^{20} \text{ cm}^{-3}$.

Cette simulation a été effectuée en prenant en compte les effets de polydépétion et le confinement quantique des porteurs. Il apparaît donc clairement que le coefficient R ne doit plus être considéré comme égale à -1 en forte inversion sur les transistors de technologies avancées.

Dans le cas où la fluctuation de courant de drain est due à la fluctuation du nombre de porteurs, la fluctuation locale du courant à la position x dans une tranche dx s'écrit :

$$\frac{\delta I(x,t)}{I(x)} = \frac{\delta Q_{inv}}{Q_{inv}} = \frac{1}{Q_{inv}} \frac{\delta Q_{inv}}{\delta Q_t} \delta Q_t = \frac{R(x)}{Q_{inv}(x)} \delta Q_t \quad (4.51)$$

La densité spectrale associée à la tranche Δx du canal s'écrit donc :

$$\Delta S_i(x, f) = I^2(x) \left(\frac{R(x)}{Q_{inv}(x)} \right)^2 S_{\delta Q_t} = I^2(x) \left(\frac{R(x)}{Q_{inv}(x)} \right)^2 \frac{q^2 k T N_t (E_{F_n})}{\gamma f W \Delta x} \quad (4.52)$$

Pour déterminer la densité spectrale du courant de drain, nous avons suivi l'approche développée par Klaassen et Prins [KLAASSEN'67], qui ont montré que l'équation de Langevin pouvait s'écrire dans le cas d'un transistor à effet de champ comme :

$$I_D(x,t) = g(V(x,t)) \frac{dV(x,t)}{dx} + h(x,t) \quad (4.53)$$

où $h(x,t)$ est la source de bruit induisant les fluctuations suivantes :

$$\begin{aligned} I_D(x,t) &= I_{D_0}(x) + \delta I_D(x,t) \\ V(x,t) &= V_0(x) + \delta V(x,t) \\ g(V(x,t)) &= g(V_0(x)) + \frac{dg}{dV} \delta V(x,t) \end{aligned} \quad (4.54)$$

La linéarisation de l'équation (4.53) permet d'obtenir l'expression de la fluctuation du courant de drain :

$$\delta I(x,t) = \frac{d}{dx} \left\{ g(V_0(x)) \delta V(x,t) \right\} + h(x,t) \quad (4.55)$$

En considérant que le courant est constant dans le tout canal, et que l'on calcule le bruit du canal en sortie court circuitée ($\delta V(0,t) = \delta V(L,t)$), on obtient :

$$\int_0^L \delta I(x,t) dx = \delta I_D(t) L = \int_0^L h(x,t) dx \quad (4.56)$$

La fonction d'auto-corrélation s'écrit :

$$\langle \delta I_D(t) \delta I_D(t+s) \rangle = \frac{1}{L^2} \int_0^L \int_0^L \langle h(x,t) h(x',t+s) \rangle dx dx' \quad (4.57)$$

La densité spectrale de $I_D(t)$ s'exprime par :

$$S_{I_D}(f) = \frac{1}{L^2} \int_0^L \int_0^L S_h(x, x', f) dx dx' \quad (4.58)$$

La densité spectrale $S_h(x, x', f)$ est de la forme $F(x, f)\delta(x-x')$. En appliquant (4.58) à une tranche du canal comprise entre x et $x+\Delta x$, on obtient la relation entre $\Delta S_I(x, f)$ et $F(x, f)$:

$$\Delta S_I(x, f) = \frac{F(x, f)}{\Delta x} \quad (4.59)$$

La densité spectrale de la fluctuation du courant de drain s'exprime donc par [KLAASSEN'67]:

$$S_{I_D}(f) = \frac{1}{L^2} \int_0^L [\Delta S_I(x, f) \Delta x] dx \quad (4.60)$$

En utilisant (4.52), on obtient donc la formulation de la densité spectrale de bruit associé au courant de drain :

$$S_{I_D}(f) = \frac{I_D^2}{WL^2} \frac{q^2 k T N_t(E_{F_n})}{\gamma f} \int_0^L \left(\frac{R(x)}{Q_{inv}(x)} \right)^2 dx \quad (4.61)$$

La figure 4.9 présente la simulation de la densité spectrale pour un transistor nMOS avec une épaisseur d'oxyde de 1.2 nm, $W/L=10 \mu\text{m} / 10 \mu\text{m}$, $N_A=10^{18} \text{cm}^{-3}$, $N_P=10^{20} \text{cm}^{-3}$. Dans le but d'éviter une intégration numérique, nous avons également simulé la densité

spectrale de bruit en considérant $R(x)/Q_{inv}(x) \approx \left. \frac{R}{Q_{inv}} \right|_{x=L/2}$. On remarque que cette

approximation donne de très bons résultats, car l'évolution de R/Q_{inv} est pratiquement linéaire de la source au drain pour de faibles polarisations du drain.

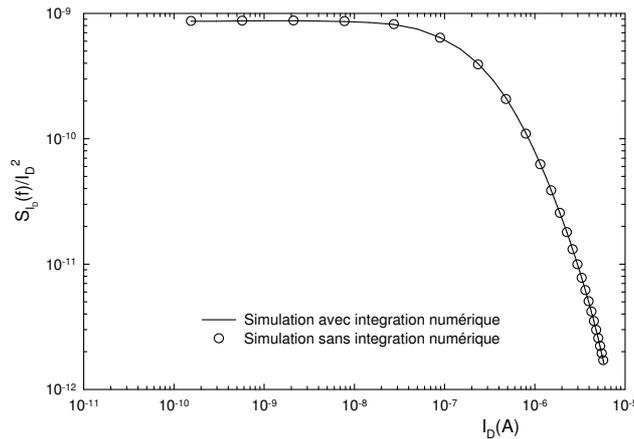


Figure 4.9 : Simulation de la densité spectrale de bruit du courant de drain à $V_{DS} = 25$ mV par intégration numérique de l'équation (4.61) et par approximation analytique (sans intégration), pour un transistor nMOS $W/L=10/10$ μm , d'épaisseur d'oxyde 1.2 nm, $N_A=10^{18}$ cm^{-3} et $N_P=10^{20}$ cm^{-3} .

En nous basant sur l'hypothèse qu'une fluctuation du nombre de porteurs dans le canal est responsable de la fluctuation du courant de drain, nous avons donc formulé une expression donnant la densité spectrale de bruit associé au courant de drain prenant en compte les effets quantiques et la polydéplétion.

4.4.2.2 Modèle corrélé $\Delta N - \Delta\mu$

Le modèle corrélé postule que la fluctuation du nombre de porteurs dans le canal induit une fluctuation de leur mobilité. En effet, la charge piégée à proximité de l'interface Si-SiO₂ est une source potentielle de diffusion coulombienne pour les porteurs libres du canal. Les fluctuations de la charge piégée dans l'oxyde peuvent donc induire des fluctuations du temps de relaxation associé à ce mécanisme d'interaction et donc de la mobilité globale des porteurs. La figure 4.10 illustre la diffusion coulombienne des porteurs de la charge piégée.

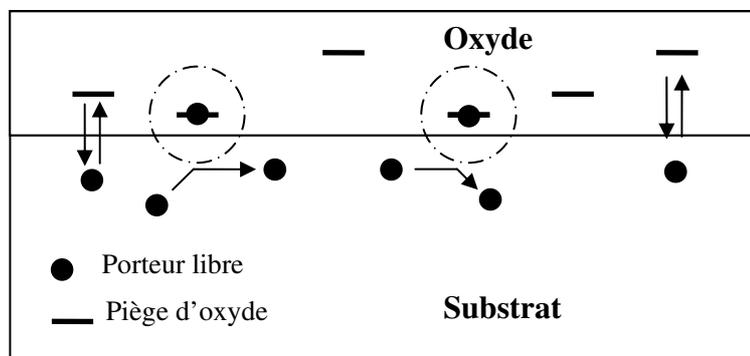


Figure 4.10 : Illustration de la diffusion des porteurs par la charge d'oxyde.

Les fluctuations locales du courant s'écrivent donc sous la forme :

$$\frac{\delta I}{I} = \frac{\delta Q_{inv}}{Q_{inv}} \pm \frac{\delta \mu_{eff}}{\mu_{eff}} = \left[\frac{1}{Q_{inv}} \frac{\delta Q_{inv}}{\delta Q_t} \pm \frac{1}{\mu_{eff}} \frac{\delta \mu_{eff}}{\delta Q_t} \right] \delta Q_t \quad (4.62)$$

En introduisant dans cette relation le coefficient R défini au paragraphe 4.4.2.1, on obtient :

$$\frac{\delta I}{I} = \left[\frac{R}{Q_{inv}} \pm \frac{1}{\mu_{eff}} \frac{\delta \mu_{eff}}{\delta Q_t} \right] \delta Q_t \quad (4.63)$$

Le terme $\delta \mu_{eff} / \delta Q_t$ s'écrit en utilisant la loi de Matthiessen [SUN'80] pour traduire la dépendance entre la mobilité effective des porteurs et la densité de pièges occupés dans l'isolant :

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_{eff_0}} + \frac{1}{\mu_{ox}} \quad (4.64)$$

Où μ_{eff_0} représente la mobilité sans l'interaction coulombienne des pièges d'oxyde définie au chapitre 3 et μ_{ox} la mobilité limitée par l'interaction avec la charge d'oxyde, liée empiriquement au nombre de pièges occupés par unité de surface, par la relation :

$$\frac{1}{\mu_{ox}} = \alpha_S N_t \quad (4.65)$$

D'après les relations (4.64) et (4.65), on obtient donc :

$$\frac{\delta \mu_{eff}}{\mu_{eff}} \frac{1}{\delta Q_t} = -\mu_{eff} \alpha_S \quad (4.66)$$

La fluctuation relative de courant s'exprime donc sous la forme :

$$\frac{\delta I}{I} = \left[\frac{R}{Q_{inv}} \pm \frac{1}{q} \alpha_S \mu_{eff} \right] \delta Q_t = - \left[\frac{R}{Q_{inv}} \pm \frac{1}{q} \alpha_S \mu_{eff} \right] C_{ox} \delta V_{FB} \quad (4.67)$$

La densité spectrale des fluctuations du courant dans une tranche comprise entre x et $x + \Delta x$ s'écrit donc :

$$\Delta S_I(x, f) = I^2(x) \left(\frac{R(x)}{Q_{inv}(x)} \pm \frac{1}{q} \alpha_S \mu_{eff} \right)^2 \frac{q^2 k T N_t(E_{F_n})}{\gamma f W \Delta x} \quad (4.68)$$

En utilisant (4.60), on obtient donc la formulation de la densité spectrale de bruit associé au courant de drain :

$$S_{I_D}(f) = \frac{I_D^2}{W L^2} \frac{q^2 k T N_t(E_{F_n})}{\gamma f} \int_0^L \left(\frac{R(x)}{Q_{inv}(x)} \pm \frac{1}{q} \alpha_S \mu_{eff} \right)^2 dx \quad (4.69)$$

De la même façon qu'au paragraphe 4.4.2.1, l'intégration numérique le long du canal peut être évitée en régime linéaire. On obtient donc :

$$S_{I_D}(f) = \frac{I_D^2}{WL} \frac{q^2 k T N_t(E_{F_n})}{\gamma f} \left(\frac{R}{Q_{inv} \Big|_{x=L/2}} \pm \frac{1}{q} \alpha_s \mu_{eff} \right)^2 \quad (4.70)$$

La figure 4.11 présente les simulations de la densité spectrale de bruit associé au courant de drain pour un transistor nMOS avec une épaisseur d'oxyde de 1.2 nm, $W/L=10 \mu\text{m} / 10 \mu\text{m}$, $N_A=10^{18} \text{ cm}^{-3}$, $N_P=10^{20} \text{ cm}^{-3}$, obtenues à l'aide des relations (4.69) et (4.70). On obtient un très bon accord.

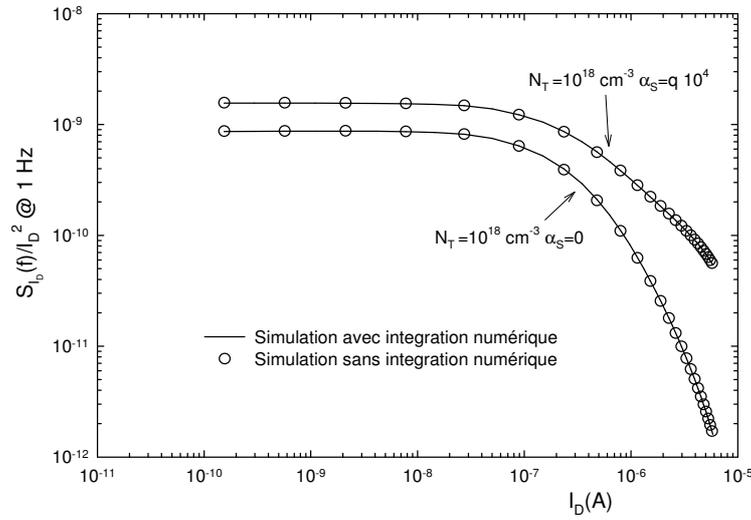


Figure 4.11 : Simulation de la densité spectrale de bruit du courant de drain à $V_{DS} = 25 \text{ mV}$ par intégration numérique de l'équation (4.69) et par approximation analytique (4.70) sans intégration, pour un transistor nMOS $W/L=10 / 10 \mu\text{m}$, d'épaisseur d'oxyde 1.2 nm, $N_A=10^{18} \text{ cm}^{-3}$ et $N_P=10^{20} \text{ cm}^{-3}$.

4.4.2.3 Lien avec les formulations existantes

Nous avons montré que les formulations des modèles ΔN ou $\Delta N - \Delta \mu$ faisant intervenir une intégration numérique sur la longueur du canal peuvent se réduire à une forme analytique ne faisant intervenir que le rapport R/Q_{inv} moyen. Compte tenu que le modèle ΔN n'est rien d'autre que le modèle $\Delta N - \Delta \mu$ avec le paramètre α_s égal à zéro, nous allons montrer le lien entre la formulation développée au paragraphe 4.4.2.2 et les modèles existants.

Dans le cas d'une structure unidimensionnelle, le rapport R/Q_{inv} s'écrit :

$$\frac{R}{Q_{inv}} = \frac{1}{Q_{inv} C_{ox}} \frac{dQ_{inv}}{dV_{GB}} \approx \frac{1}{C_{ox}} \frac{dI_D}{dV_{GB}} \frac{1}{I_D} = g_m \frac{1}{C_{ox} I_D} \quad (4.71)$$

En reportant ce résultat dans (4.70), on obtient :

$$\begin{aligned}
S_{i_b}(f) &= \frac{I_D^2}{WL} \frac{q^2 kTN_t(E_{F_n})}{\gamma f} \left(g_m \frac{1}{C_{ox} I_D} \pm \frac{1}{q} \alpha_s \mu_{eff} \right)^2 \\
&= \frac{1}{WLC_{ox}^2} \frac{q^2 kTN_t(E_{F_n})}{\gamma f} g_m^2 \left(1 \pm \frac{1}{q} \alpha_s \mu_{eff} C_{ox} \frac{I_D}{g_m} \right)^2
\end{aligned} \tag{4.72}$$

En utilisant la densité spectrale des fluctuations de bandes plates, la densité spectrale du courant de drain s'écrit :

$$S_{i_b}(f) = \frac{q^2 kTN_t(E_{F_n})}{\gamma f WLC_{ox}^2} g_m^2 \left(1 \pm \frac{1}{q} \alpha_s \mu_{eff} \frac{I_D}{g_m} \right)^2 = g_m^2 \left(1 \pm \frac{1}{q} \alpha_s \mu_{eff} \frac{I_D}{g_m} \right)^2 S_{V_{FB}}(f) \tag{4.73}$$

Cette formulation a été proposée par Ghibaudo [GHIBAUDO'91]. Elle permet de simuler le bruit à partir de la connaissance de la caractéristique courant tension expérimentale. L'expression (4.73) reste valable en présence d'effets quantiques et de polydépétion. Concernant la densité spectrale des fluctuations de la tension de bandes plates, il faut noter que la valeur de C_{ox} à considérer doit toujours être ϵ_{ox}/T_{ox} . En fait, nous pensons qu'il est erroné de remplacer C_{ox} par C_{oxeff} (cf. chapitre 2) pour tenir compte des effets quantiques [MERCHA'02], puisque ce terme est directement issu de l'application du théorème de Gauss (cf. équation (4.44)). Les effets quantiques sont simplement pris en compte dans la transconductance g_m .

Une autre publication [AHSAN'05] relative à l'impact des effets quantiques sur la modélisation du bruit basse fréquence propose de tenir compte du déplacement de la charge dans le substrat pour la constante α_s . D'un point de vue théorique, il est évident que l'interaction coulombienne diminue quand on s'éloigne de l'interface Si-SiO₂. Cependant, ce modèle donne un coefficient α_s deux fois plus petit à une distance de 1.5 nm de l'interface. Compte tenu qu'il est impossible d'obtenir une telle précision expérimentale sur le paramètre α_s , nous n'avons pas pris en compte ce phénomène dans notre modélisation de la fluctuation de la mobilité due à l'interaction coulombienne.

4.5 Caractérisation en bruit du courant de drain

Le but de cette partie est d'estimer la densité de pièges lents des différents lots présentés au chapitre 3. Nous allons nous focaliser dans un premier temps sur les transistors ayant une longueur de grille inférieure à 1 μm pour ne pas être perturbé par le courant de grille. Dans ces conditions, la formulation du bruit du courant de drain (4.73) reste valable, puisqu'elle est équivalente à notre formulation tenant compte des effets quantiques et de la

polydépétion. De plus, elle permet d'extraire simplement la densité de piège en connaissant les caractéristiques expérimentales statiques du transistor.

Nous avons donc réalisé des extractions de la densité de pièges d'oxyde sur les transistors nMOS et pMOS des lots RTN14, RTN12, et DPN12. La figure 4.12 présente l'évolution typique de la densité spectrale du courant drain normée au carré du courant de drain pour une géométrie (n- et pMOS) de chaque lot. Les extractions ont été réalisées sur l'ensemble des géométries disponibles pour chaque lot par ajustement des mesures à l'équation (4.73). Les gammes de valeurs de densité de pièges lents d'oxyde et le coefficient de corrélation α_S sont reportées dans le tableau 4.1.

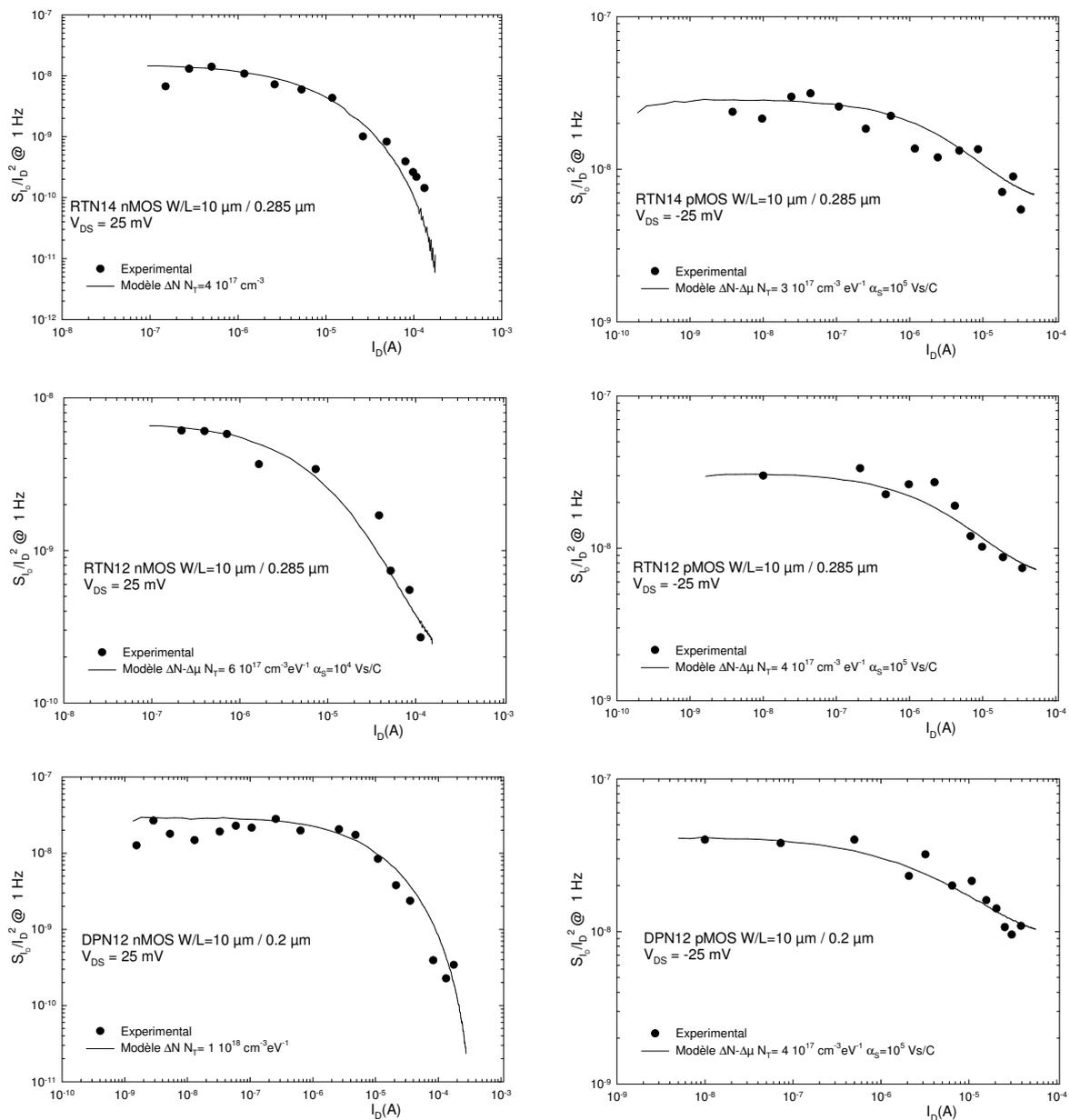


Figure 4.12 : Densités spectrales des fluctuations basse fréquence du courant drain pour les lots RTN14, RTN12 et DPN12.

	$N_t(E_F)$ ($eV^{-1}cm^{-3}$)		α_S (Vs/C)	
	nMOS	pMOS	nMOS	pMOS
RTN14	$1 \cdot 10^{17} - 3 \cdot 10^{17}$	$2 \cdot 10^{17} - 5 \cdot 10^{17}$	0	10^4
RTN12	$4 \cdot 10^{17} - 6 \cdot 10^{17}$	$3 \cdot 10^{17} - 4 \cdot 10^{17}$	10^3	10^4
DPN12	$7 \cdot 10^{17} - 1 \cdot 10^{18}$	$4 \cdot 10^{17} - 6 \cdot 10^{17}$	0	10^4

Tableau 4.1 : Paramètres de bruit extraits sur les lots RTN14, RTN12, DPN12 pour les transistors n- et pMOS

Pour les nMOS, on observe une augmentation de la densité de pièges avec la réduction de l'épaisseur d'oxyde (RTN14 \rightarrow RTN12), et avec le changement de processus de nitruration de la grille (RTN12 \rightarrow DPN12). Le modèle de fluctuation du nombre de porteurs ΔN seul a permis de simuler avec une bonne précision les lots RTN14 et DPN12. L'augmentation de la densité d'états lents avec la réduction de l'épaisseur d'oxyde nitruré a déjà été mise en évidence à de nombreuses reprises dans la littérature. Cette dégradation de la qualité de l'oxyde est liée essentiellement à la réduction des budgets thermiques [SIMOEN'99]. Il a été montré que l'introduction d'azote par le procédé DPN améliore les performances en bruit des transistors comparé au procédé RTN [MARIN'05]. Cependant, cette étude est relative à des composants possédant une épaisseur d'oxyde supérieures à celle de nos transistors. De plus, le lot RTN12 et DPN12 ne diffèrent pas uniquement par le procédé de nitruration ; le canal ainsi que les poches sont nettement plus dopés dans le cas du lot DPN12 pour permettre le bon fonctionnement des transistors de longueur de grille de 30 nm. Il est donc probable que la faible épaisseur d'oxyde combinée à des budgets thermiques différents masque totalement l'amélioration potentielle liée au procédé de nitruration DPN. Nous avons également constaté que la simulation du bruit du canal pour les transistors nMOS RTN12 a nécessité l'utilisation du modèle corrélé $\Delta N - \Delta \mu$ avec un coefficient de corrélation égal à 10^4 Vs/C. Compte tenu que le bruit des transistors nMOS DPN12 vérifie le modèle ΔN , on peut supposer que le procédé DPN sur les transistors nMOS a permis de diminuer l'impact de la charge piégée sur la mobilité des porteurs.

Dans le cas des pMOS, nous avons extrait sensiblement les mêmes valeurs de densités de pièges lents et de coefficients de corrélation pour les trois lots. La densité de pièges est de l'ordre de $3-4 \cdot 10^{17} cm^{-3}$, et le coefficient de corrélation est de l'ordre de 10^5 Vs/C. Dans le cas du lot RTN12, le coefficient de corrélation de mobilité des transistors pMOS est supérieur à celui des nMOS d'une décade, ce qui est conforme à la littérature. Ce phénomène s'explique

par le fait que le coefficient d'interaction coulombien avec les pièges d'oxyde pour les trous est plus grand que celui pour les électrons. De la même façon que pour les nMOS, les améliorations en terme de bruit basse fréquence potentiellement amené par le procédé de nitruration DPN est masqué par l'épaisseur d'oxyde très faible de nos composants [MARTINEZ'05a]. Dans le cas d'un oxyde plus épais, le procédé DPN préserve l'interface Si-SiO₂ et les performances en bruit basse fréquence sont donc améliorées [MARIN'05]. Avec une épaisseur d'oxyde équivalente de l'ordre de 1.2 nm, on peut considérer que tous les pièges présents dans l'oxyde participent au bruit BF.

Sur la figure 4.13 nous comparons nos valeurs de $N_t(E_F)$ avec celles trouvées dans la littérature en fonction de l'épaisseur d'oxyde équivalente. Avec l'introduction d'oxyde nitruré, on observe donc une augmentation de la densité de pièges d'oxyde avec une diminution de l'épaisseur d'oxyde. Cette tendance est nettement plus marquée pour les transistors de type pMOS que nMOS.

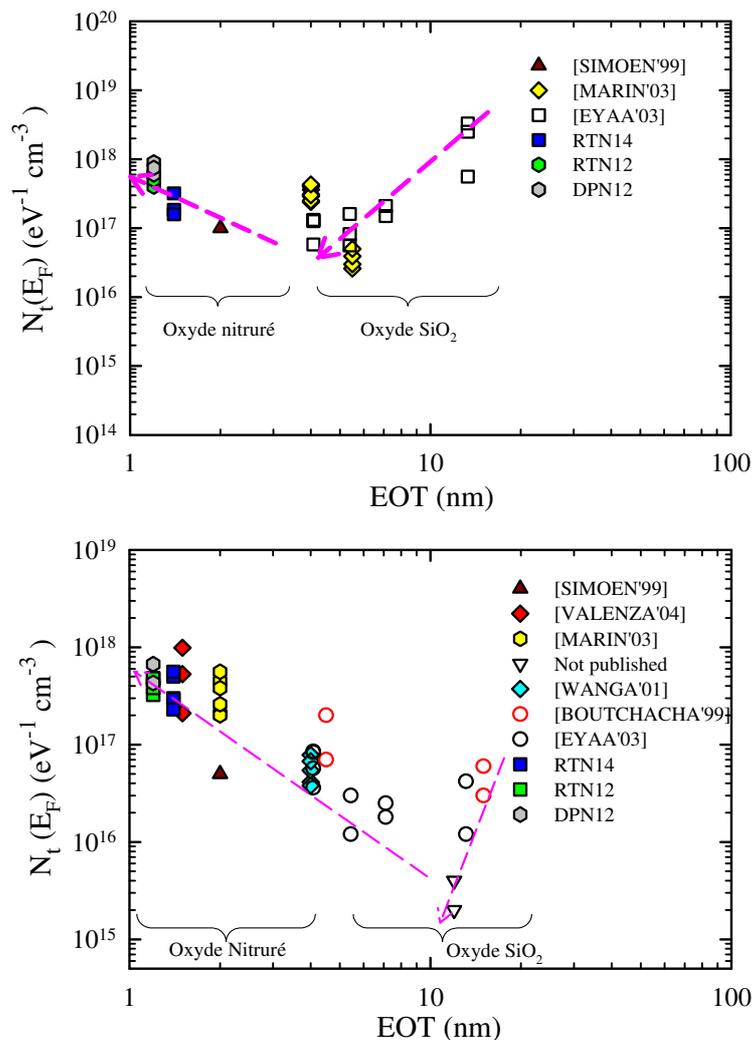


Figure 4.13 : Variations de la densité de pièges $N_t(E_F)$ en fonction de l'épaisseur équivalente (EOT) pour les transistors nMOS (a) et pMOS (b).

4.6 Bruit basse fréquence associé au courant de grille

4.6.1 Modélisation du bruit associé au courant tunnel direct

Le bruit basse fréquence associé à la traversée par effet tunnel d'une barrière de potentiel à été étudié dans le cadre des diodes MOS dans les années 70 [KLEINPENNING'78, KLEINPENNING'82, KUMAR'77]. Ce n'est que très récemment qu'un modèle dédié au courant tunnel de grille de transistors MOS avec un diélectrique ultrafin a été publié [LEE'03]. Ce modèle est basé sur une fluctuation de la probabilité tunnel liée au phénomène de piégeage - dépiégeage des états lents de l'oxyde. L'approche choisie par l'auteur est basée sur la modélisation électrique des pièges d'oxyde (modélisation des pièges par des circuits R-C). Cependant, la formulation proposée par Lee et al ne fait pas intervenir une densité de pièges par unité de volume et d'énergie, mais une densité surfacique de pièges d'oxyde. Tout lien avec la densité de pièges lents extrait par les mesures sur le courant de drain est donc difficile. De ce fait, nous proposons donc dans cette partie une formulation du bruit basse fréquence associé au courant de grille basée sur le phénomène de piégeage - dépiégeage des pièges d'oxyde faisant intervenir explicitement la densité de pièges par unité de volume et d'énergie.

Nous avons montré que la fluctuation du nombre de pièges occupés dans l'oxyde induit une fluctuation de la tension de bandes plates. La formulation du courant tunnel détaillée au chapitre 2 montre que le courant tunnel direct dépend de la tension de grille et du potentiel de surface. Sur une tranche du canal située entre x et $x + \Delta x$, la fluctuation de courant de grille s'exprime par :

$$\Delta \delta I_G(x, t) = \delta J_G(x, t) W \Delta x \quad (4.74)$$

La fluctuation du courant de grille est donc liée à la fluctuation de la tension de grille par la relation :

$$\Delta \delta I_G(x, t) = \frac{dJ_G}{dV_{GB}} \delta V_{GB}(x, t) W \Delta x \quad (4.75)$$

La densité spectrale de la fluctuation du courant de grille sur une tranche comprise entre x et $x + \Delta x$ du canal est donc donnée par :

$$\Delta S_{I_G}(x, f) = W^2 \Delta x^2 \left(\frac{dJ_G}{dV_{GB}|_x} \right)^2 S_{V_{FB}}(f) = \left(\frac{dJ_G}{dV_{GB}|_x} \right)^2 \frac{q^2 kT N_t}{C_{ox}^2 \gamma f} W \Delta x \quad (4.76)$$

Puisque le courant de grille total s'exprime suivant la relation :

$$I_G(t) = W \int_0^L J_G(x,t) dx \quad (4.77)$$

Les fluctuations du courant de grille total sont donc égale à :

$$\delta I_G(t) = W \int_0^L \delta J_G(x,t) dx \quad (4.78)$$

La fonction d'auto-corrélation de la fluctuation totale du courant de grille s'écrit :

$$\langle \delta I_G(t) \delta I_G(t+s) \rangle = W^2 \int_0^L \int_0^L \langle \delta J_G(x,t) \delta J_G(x',t+s) \rangle dx dx' \quad (4.79)$$

La densité spectrale des fluctuations du courant de grille s'exprime donc par :

$$S_{I_G} = W^2 \int_0^L \int_0^L S_{J_G}(x,x',f) dx dx' = W^2 \int_0^L F(x,f) dx \quad (4.80)$$

où $S_{J_G}(x,x',f) = F(x,f) \delta(x-x')$.

En appliquant la relation (4.80) à une tranche comprise en x et $x+\Delta x$, on montre facilement que $F(x,f) = \frac{\Delta S_{I_G}}{W^2 \Delta x}$. La densité spectrale des fluctuations du courant de grille peut donc se mettre sous la forme :

$$S_{I_G}(f) = \int_0^L \frac{\Delta S_{I_G}(x,f)}{\Delta x} dx \quad (4.81)$$

avec $\Delta S_{I_G}(x,f) = \left(\left. \frac{dJ_G}{dV_{GB}} \right|_x \right)^2 \frac{q^2 kT N_t}{C_{ox}^2 \gamma f} W \Delta x$

Dans le cas où on considère une structure homogène de surface $W \times L$ (capacité MOS), l'expression (4.81) se met sous la forme :

$$S_{I_G}(f) = \left(\frac{dJ_G}{dV_{GB}} \right)^2 \frac{q^2 kT N_t}{C_{ox}^2 \gamma f} WL = \left(\frac{dI_G}{dV_{GB}} \right)^2 \frac{q^2 kT N_t}{WLC_{ox}^2 \gamma f} \quad (4.82)$$

4.6.2 Résultats expérimentaux et validation du modèle

Pour étudier le bruit associé au courant de grille, nous devons nécessairement disposer de structures à grilles isolées. Compte tenu du nombre de géométries disponibles, nous avons concentré notre étude expérimentale sur les transistors isolés nMOS du lot DPN12.

La figure 4.14 présente un spectre de bruit du courant de grille typiquement observé sur un transistor nMOS polarisé en régime de forte inversion. Le plateau de bruit blanc est attribué au bruit de grenaille $2qI_G$. Plusieurs auteurs ont publié que le bruit de grenaille associé au courant tunnel pouvait être supérieur à $2qI_G$ (*enhanced shot noise*) [REKLAITIS

'99, IANNACCONE'98] ou inférieur (*suppressed shot noise*) [IANNACCONE'00]. Ce phénomène est expliqué à partir de la nature inélastique du courant de grille, c'est-à-dire dans le cas où le courant de grille est un courant tunnel indirect assisté par pièges. Selon les constantes de temps des pièges mis en jeu, on observe un bruit de grenaille $2qI_G$ modulé par des lorentziennes liées aux processus de capture/émission. Cependant, nous n'avons pas observé ce phénomène sur les composants que nous avons mesurés puisque nous avons toujours observé un plateau de bruit blanc égal à $2qI_G$.

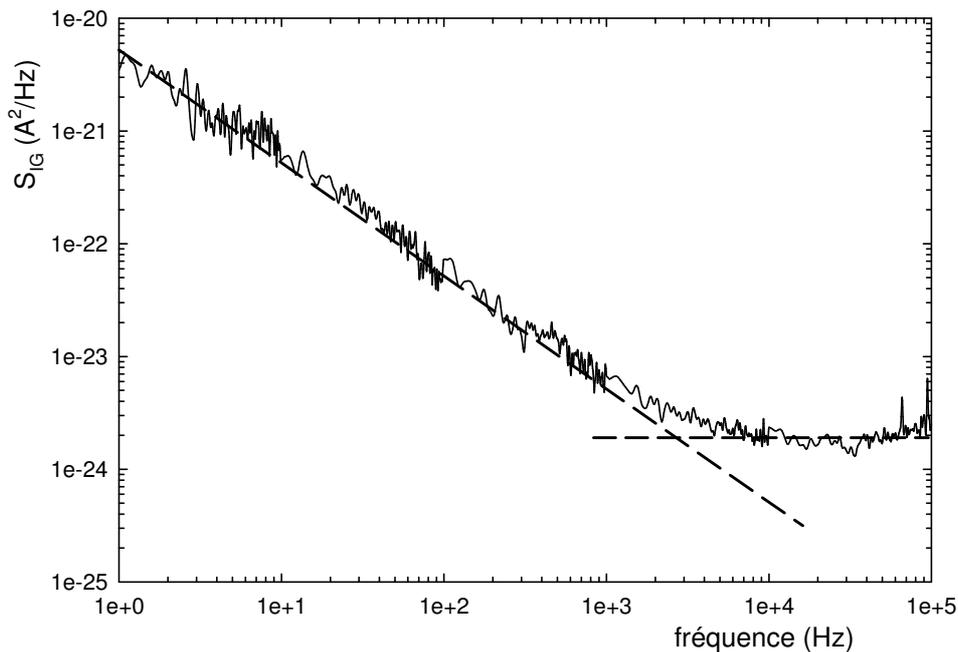


Figure 4.14 : Spectre de bruit typique du courant de grille pour un transistor nMOS du lot DPN12 polarisé en forte inversion.

Nous avons également observé du bruit de type RTS sur le courant de grille, induisant une composante lorentzienne sur le spectre de bruit. Nous détaillerons dans la partie suivante l'étude de ce phénomène ; dans cette partie, nous nous focalisons sur le niveau de bruit en $1/f$ du courant de grille.

La figure 4.15 présente le niveau de bruit à 1 Hz du courant de grille pour un transistor nMOS avec $W/L = 10 \mu\text{m} / 10 \mu\text{m}$ du lot DPN12. Cette mesure a été effectuée avec les électrodes de source, de drain et de substrat reliées à la masse. La simulation utilise le modèle défini par l'équation (4.82), et le meilleur ajustement nous permet d'extraire la densité de

pièges d'oxyde $N_t=2 \cdot 10^{18} \text{ cm}^{-3} \text{ eV}^{-1}$. Cette simulation utilise les modèles ainsi que les paramètres extraits dans les chapitres 2 et 3 pour évaluer les quantités $\frac{dI_G}{dV_{GB}}$.

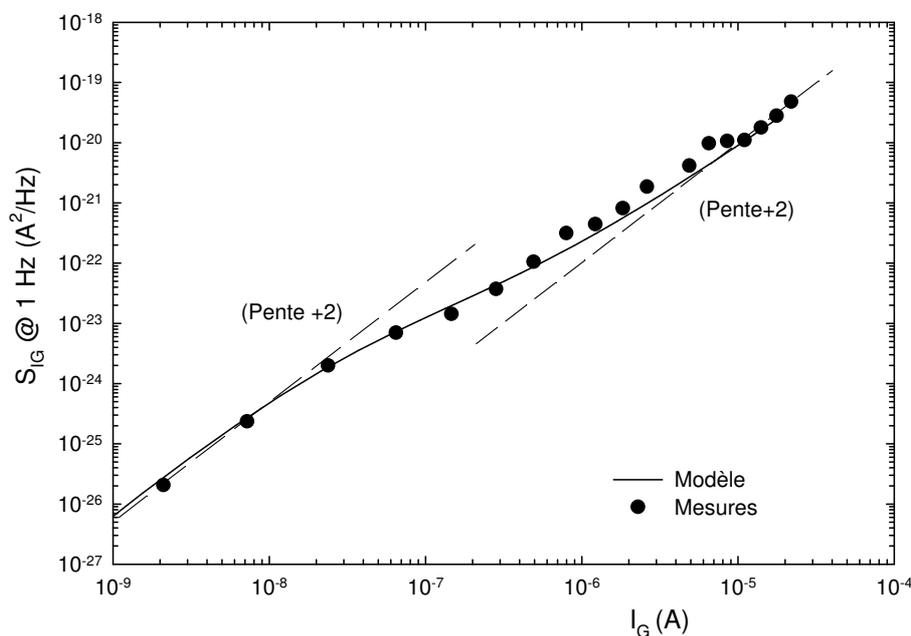


Figure 4.15 : Niveau de bruit en 1/f du courant de grille normé à la surface pour un transistor nMOS du lot DPN12 avec $W/L = 10 \mu\text{m} / 10 \mu\text{m}$. La source, le drain et le substrat sont reliés à la masse.

La densité de pièges extraite est donc en accord avec les mesures effectuées sur le courant de drain des transistors nMOS du lot DPN12. De plus, on note que notre modèle prévoit le changement de courbure de la caractéristique $S_{IG}-I_G$ au voisinage de la tension de seuil du composant. On peut noter la dépendance quadratique au courant statique avant et après cette zone de transition. La modélisation empirique [VALENZA'03,VALENZA'04] du bruit basse fréquence associé au courant de grille comme une simple loi quadratique du courant statique peut donc induire une erreur suivant que l'on se place dans la partie supérieure ou inférieure de la caractéristique $S_{IG}-I_G$

Nous avons mesuré un grand nombre de géométries différentes du lot DPN12, et reporté l'ensemble des densités spectrales de bruit de grille normées à la surface sur la figure 4.16. Les mesures sont toujours effectuées avec la source, le drain et le substrat à la masse. Les simulations ont été réalisées en utilisant la même densité de pièges d'oxyde $N_t = 2 \cdot 10^{18} \text{ cm}^{-3} \text{ eV}^{-1}$ pour toutes les géométries. On observe un très bon accord entre les mesures et le modèle décrit par l'équation (4.82).

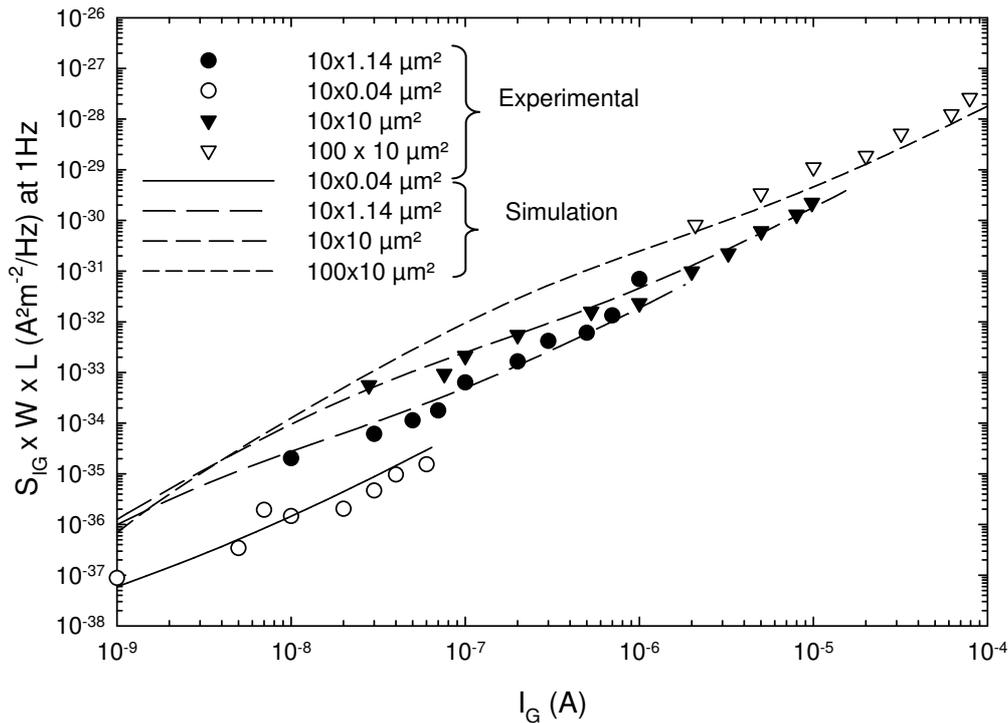


Figure 4.16 : Niveau de bruit en 1/f du courant de grille normé à la surface pour des transistor nMOS du lot DPN12. La source, le drain et le substrat sont reliés à la masse.

La figure 4.17 présente le niveau de bruit à 1 Hz normé à la surface pour différentes surfaces à courant de grille constant $I_G = 1 \mu A$. Il apparaît clairement sur cette figure que la densité spectrale de bruit de grille normée à la surface n'est pas indépendante de la surface, comme le prévoit le modèle empirique [VALENZA'04]. La simulation avec le modèle (4.82) présente un excellent accord avec les données expérimentales.

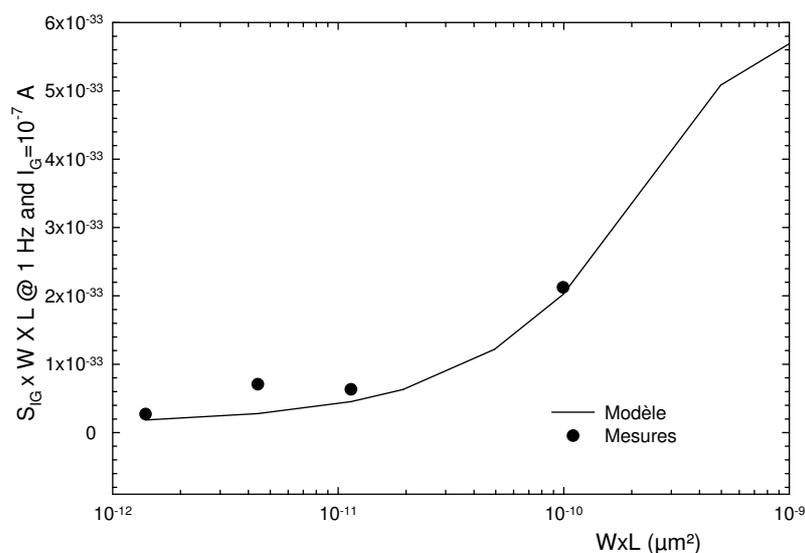


Figure 4.17 : Niveau de bruit en 1/f normé à la surface pour un courant de grille constant $I_G = 10^{-7} A$ en fonction de la surface du transistor, pour les transistors nMOS du lot DPN12.

4.6.3 Influence du courant de grille sur le bruit du canal dans les transistors longs

Nous nous intéressons dans cette partie à la modélisation du bruit basse fréquence associé au courant drain dans les transistors fortement perturbés par le courant de grille. Dans cette étude, nous avons mesuré le bruit sur des transistors de géométrie $10\ \mu\text{m} \times 10\ \mu\text{m}$ du lot DPN12. Les caractéristiques statiques mettant en évidence la courbure du potentiel de surface le long du canal en présence de forte fuite de la grille ont été détaillées au chapitre 3. Le premier point important est de quantifier l'impact de la courbure du potentiel de surface sur le coefficient R. La figure 4.18 présente la simulation pour le transistor nMOS à $V_{DS} = 25\ \text{mV}$ et $V_{GB} = 1\ \text{V}$, avec la prise en compte ou non du courant de grille dans l'équation de continuité. Nous avons également reporté le potentiel de surface sur cette figure.

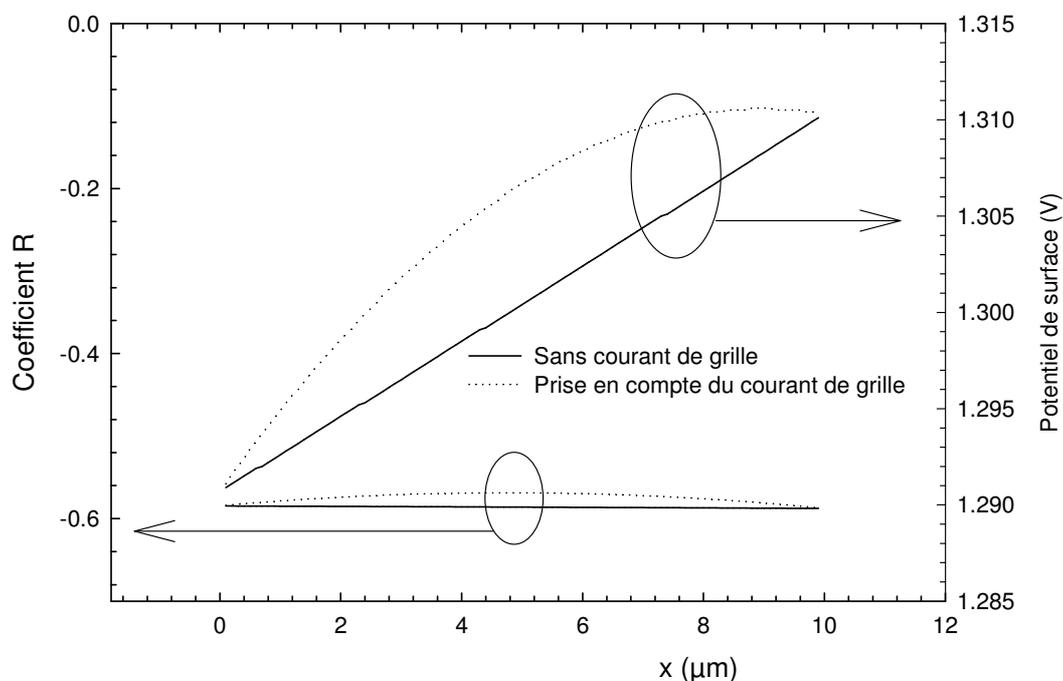


Figure 4.18 : Impact de la courbure du potentiel de surface sur le coefficient R, pour un transistor nMOS simulé avec les paramètres du lot DPN12, à $V_{DS} = 1\ \text{V}$ et $V_{GB} = 1\ \text{V}$

Le coefficient R est donc faiblement perturbé par la courbure du potentiel de surface. Nous faisons donc l'hypothèse que le bruit total présent sur le drain est la somme du bruit du drain formulé dans le paragraphe 4.4.2.2 et d'une partie du bruit de grille, que nous allons évaluer en utilisant l'équation de continuité du courant dans le canal du transistor.

L'équation de continuité dans le canal du transistor pour les fluctuations de courant s'écrit :

$$\frac{d\delta i(x,t)}{dx} = -W \delta J_G(x,t) \quad (4.83)$$

En suivant les équations du partitionnement du courant de grille, la fluctuation de courant de drain due au courant de grille sur une tranche comprise entre x et $x + \Delta x$ s'écrit [RANUAREZ'05] :

$$\delta i_{GD}(x,t) = \delta i_G(x,t) \frac{x}{L} \quad (4.84)$$

On a donc la relation :

$$\Delta S_{I_{GD}}(x,f) = \frac{x^2}{L^2} \Delta S_{I_G}(x,f) \quad (4.85)$$

La fluctuation totale au niveau du drain $\delta I_{GD}(t)$ s'exprime sous la forme :

$$\delta I_{GD}(t) = W \int_0^L \delta J_G(x,t) \frac{x}{L} dx \quad (4.86)$$

La densité spectrale de I_{GD} s'exprime donc :

$$S_{I_{GD}} = \int_0^L \frac{x^2}{L^2} \frac{\Delta S_{I_G}(x,f)}{\Delta x} dx \quad (4.87)$$

En utilisant (4.81), la densité spectrale au niveau du drain est donnée par :

$$S_{I_D}(f) = \frac{I_D^2}{WL^2} \frac{q^2 kTN_t}{\gamma f} \int_0^L \left(\frac{R}{Q_{inv}|_x} \pm \frac{1}{q} \alpha_S \mu_{eff} \right)^2 dx + \frac{q^2 kT N_t}{C_{ox}^2 \gamma f} W \int_0^L \frac{x^2}{L^2} \left(\frac{dJ_G}{dV_{GB}} \Big|_x \right)^2 dx \quad (4.88)$$

La formulation pour le bruit total mesuré sur la source est analogue aux expressions précédentes. La fluctuation totale au niveau de la source $\delta I_{GS}(t)$ s'exprime sous la forme :

$$\delta I_{GS}(t) = W \int_0^L \delta J_G(x,t) \left(1 - \frac{x}{L} \right) dx \quad (4.89)$$

La densité spectrale de I_{GS} s'exprime donc :

$$S_{I_{GS}} = \int_0^L \left(1 - \frac{x}{L} \right)^2 \frac{\Delta S_{I_G}(x,f)}{\Delta x} dx \quad (4.90)$$

La densité spectrale au niveau de la source est donnée par :

$$S_{I_S}(f) = \frac{I_S^2}{WL^2} \frac{q^2 kTN_t}{\gamma f} \int_0^L \left(\frac{R}{Q_{inv}|_x} \pm \frac{1}{q} \alpha_S \mu_{eff} \right)^2 dx + \frac{q^2 kT N_t}{C_{ox}^2 \gamma f} W \int_0^L \left(1 - \frac{x}{L} \right)^2 \left(\frac{dJ_G}{dV_{GB}} \Big|_x \right)^2 dx \quad (4.91)$$

La figure 4.19 présente l'évolution des densités spectrales de courant de bruit du courant source normées au carré du courant source pour un transistor nMOS du lot DPN12, $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, pour des polarisations de drain de 25 mV et 100 mV. Nous avons également reporté les simulations correspondantes en utilisant l'équation (4.91). Ces simulations ont été réalisées avec $N_t = 2 \cdot 10^{18} \text{ cm}^{-3} \text{ eV}^{-1}$, $\alpha_s = 0$, et les paramètres technologiques et de conduction extraits au chapitre 3. On obtient un excellent accord entre l'équation (4.91) et données expérimentales, en particulier la dépendance de l'excès de bruit lié au courant de grille mesuré sur le courant source avec la tension V_{DS} .

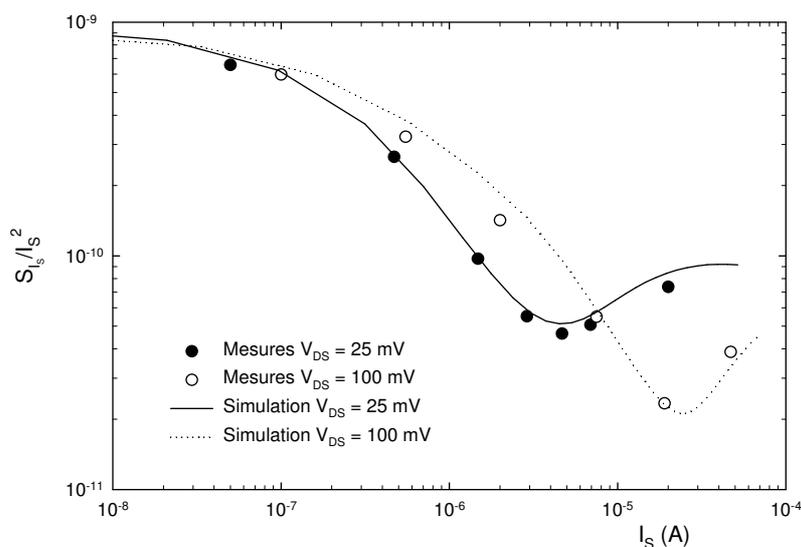


Figure 4.19 : Densité spectrale des fluctuations du courant source expérimentales et simulées, pour un transistor nMOS du lot DPN12, $W/L = 10 \mu\text{m} / 10 \mu\text{m}$.

Le banc de mesure ne permettant de ne mesurer que le bruit au niveau de la source (amplificateur trans-impédance), il n'a pas été possible de compléter l'étude expérimentale par des mesures sur le courant drain. La figure 4.20 présente néanmoins la simulation des densités spectrales de bruit du courant drain normées au carré du courant drain pour un transistor nMOS du lot DPN12, $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$, pour des polarisations de drain de 25 mV et 100 mV, simulées avec les paramètres précédemment extraits.

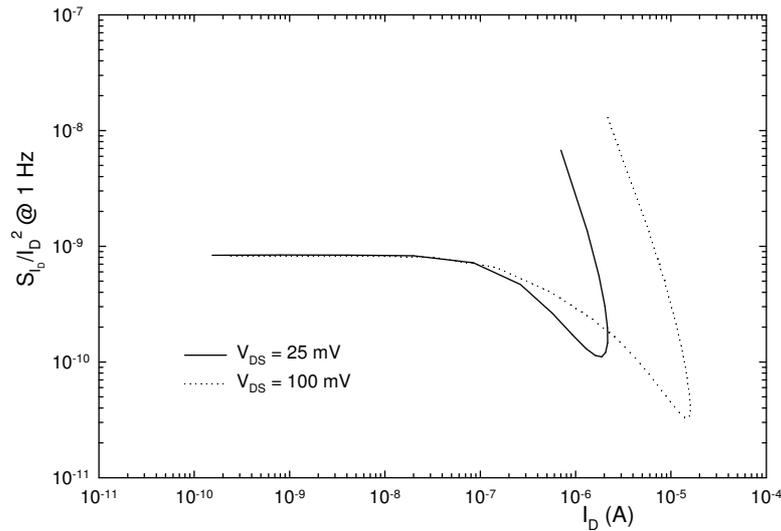


Figure 4.20 : Densité spectrale des fluctuations du courant drain simulées, pour un transistor nMOS du lot DPN12, $W/L = 10 \mu\text{m} / 10 \mu\text{m}$.

On peut donc s'attendre à une augmentation significative du niveau de bruit du courant drain du au courant de grille dès que le courant drain atteint son maximum.

4.6.4 Bruit RTS sur la grille [MARTINEZ'05b]

Les fluctuations de type RTS (*Random Telegraph Signal*) ont été observées sur le courant de grille. Ces fluctuations, sont caractérisées par une signature spectrale (spectre lorentzien) et temporelle (signaux rectangulaires aléatoires). Un exemple de mesures dans le domaine temporel et fréquentiel typiquement observé de bruit RTS sur le courant de grille est représenté sur la figure 4.21.

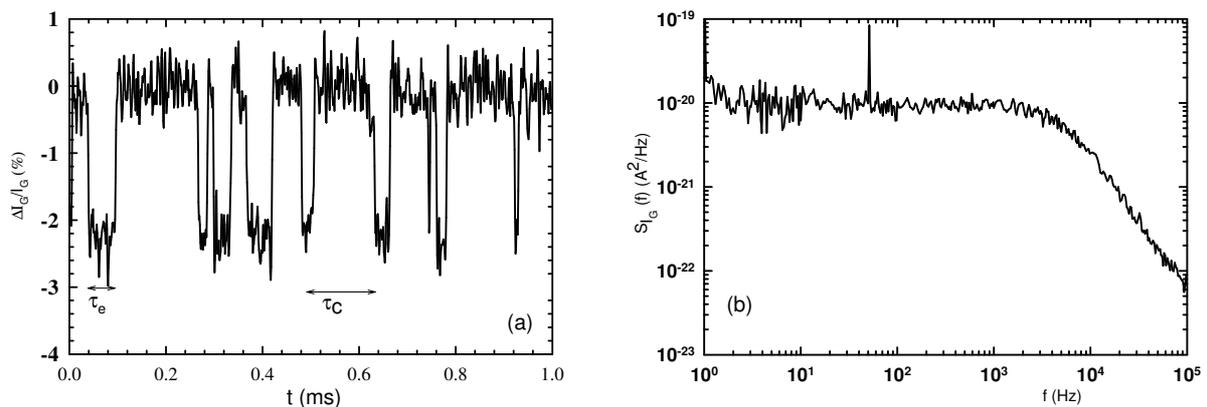


Figure 4.21 : Observation du bruit de type RTS sur le courant de grille des transistors nMOS du lot DPN12 dans le domaine temporel (a) et fréquentiel (b)

Ce type de bruit associé au courant de drain a été largement traité dans la littérature. L'origine physique de ce bruit est liée à l'activité d'un piège unique à l'interface Si-SiO₂ [ROUX'93, OHATA'90]. La modulation de la conductance du canal induit par la capture d'un porteur se traduit donc par des signaux rectangulaires directement observables dans le domaine temporel. La caractérisation de ces défauts singuliers par les mesures de bruit RTS permet d'extraire la position du piège, son énergie d'activation, et sa section de capture. Néanmoins, les fluctuations de type RTS ne sont observables sur le courant drain que pour des transistors de surface inférieure au μm^2 .

Nous avons observé du bruit de type RTS sur le courant de grille de transistors nMOS de géométrie $10 \times 10 \mu\text{m}^2$, la source, le drain, et le substrat étant reliés à la masse. Dans ce cas, on peut donc considérer que cette structure est unidimensionnelle, et que le potentiel de surface est constant sur toute la surface du composant.

Pour modéliser le temps de capture et d'émission des porteurs, nous utilisons les relations établies au paragraphe 4.3.1. Toutes les définitions des grandeurs physiques utilisées dans cette partie sont rappelées sur la figure 4.22.

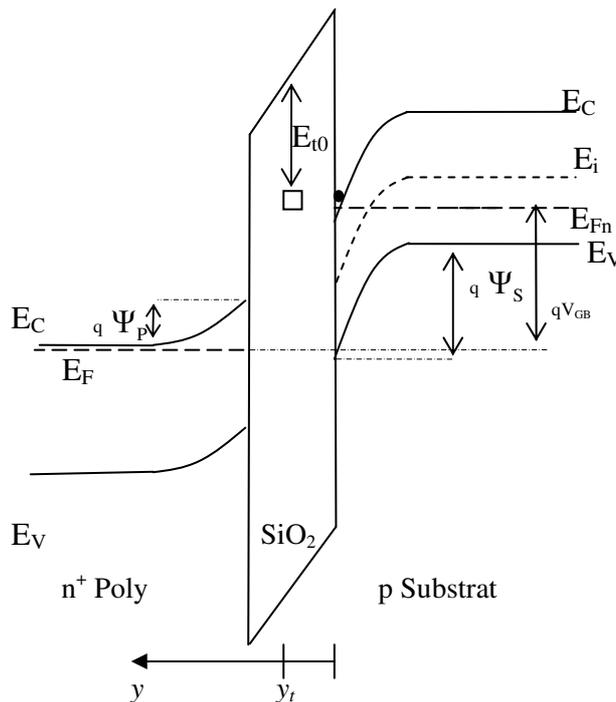


Figure 4.22 : Diagramme de bandes illustrant le piégeage d'un porteur responsable du bruit de type RTS sur le courant de grille.

Conformément à la statistique de Shockley-Read-Hall, la constante de temps de capture d'un piège d'énergie E_t à la profondeur y_t est donnée par :

$$\tau_c = \frac{1}{\sigma_n f(\varepsilon) N_{inv}(\varepsilon) T(\varepsilon, y_t)} \quad (4.92)$$

La section de capture se met sous la forme $\sigma_n = \sigma_0 \exp(-\Delta E_B/kT)$, où σ_0 est le pré-facteur de la section de capture, et ΔE_B la barrière d'énergie pour la capture [AMARASINGHE'03, KIRTON'89].

La probabilité moyenne d'occupation du piège est donnée par [KIRTON'89] :

$$f_t(y) = \frac{\tau_e}{\tau_e + \tau_c} = \frac{1}{1 + \exp\left[\frac{E_t - E_{F_n}}{kT}\right]} \quad (4.93)$$

La constante de temps d'émission est donc donnée par :

$$\tau_e = \frac{\exp\left[\frac{E_t - E_{F_n}}{kT}\right]}{\sigma_n f(\varepsilon) N_{inv}(\varepsilon) T(\varepsilon, y_t)} \quad (4.94)$$

Conformément aux notations utilisées dans ce mémoire, la quantité $E_{F_n} - E_T$ est calculée en prenant comme référence en énergie le bas de la bande de conduction à l'interface :

$$E_{F_n} - E_T \equiv q\Psi_s - \frac{E_g}{2} - q\phi_F - (q\chi_B - qF_{ox}y_t - qE_{T_0}) \quad (4.95)$$

où E_{T_0} est l'énergie du piège par rapport à la bande de conduction de l'oxyde.

A partir des mesures temporelles, nous avons extrait les constantes de temps des signaux RTS en fonction de la polarisation de grille pour un transistor nMOS W/L=10 μm / 10 μm du lot DPN12. Sur la figure 4.23a sont reportés les résultats expérimentaux ainsi que la simulation en utilisant la modélisation décrite par (4.92) et (4.94). Pour ajuster au mieux la simulation, nous avons utilisé les paramètres extraits au chapitre 3 pour simuler le potentiel de surface, puis nous avons recherché la profondeur y_t , la section de capture σ_n , et l'énergie E_{T_0} optimum. Les meilleurs résultats ont été obtenus pour $y_t = 0.1$ nm, $E_{T_0} = 2.95$ eV, et $\sigma_0 = 5 \cdot 10^{-20}$ cm². Ces valeurs sont en accord avec les paramètres extraits pour des pièges d'oxydes nitrurés à partir de mesures RTS sur le courant de drain [LEYRIS'05].

La Figure 4.23b présente l'évolution de la fonction d'occupation du piège mesurée et simulée. Un très bon accord est observé ; l'activité du piège suit donc un processus SRH.

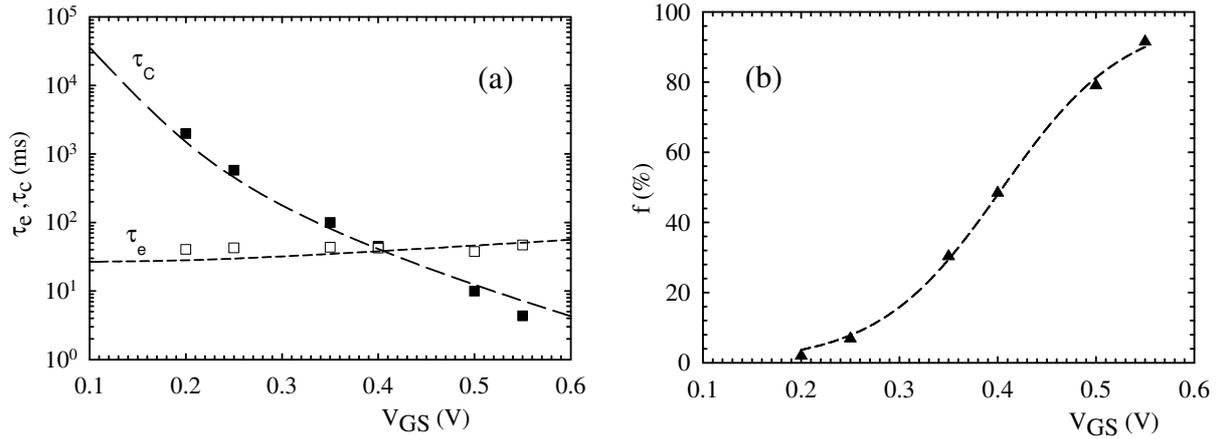


Figure 4.23 : Constantes de temps (a) d'émission et de capture mesurées (symboles) et simulées (lignes), et fonction d'occupation du piège (b) mesurée (symboles) et simulée (ligne), pour un transistor nMOS du lot DPN12 WxL = 10 μ m x 10 μ m à partir de mesure de bruit RTS sur le courant de grille à $V_{DS} = 0$ V.

Nous avons reporté sur la figure 4.24 l'amplitude des fluctuations du courant en fonction du courant de polarisation. On observe une évolution quadratique entre ΔI_G et le courant de grille statique I_G . L'amplitude des signaux RTS sur le courant de drain fait intervenir également le concept de fluctuations de tension de bandes plates. Dans le cas du piégeage d'une charge élémentaire, la fluctuation de bandes plates pour une surface $\Delta x \Delta z$ s'écrit [GHIBAUDO'91] :

$$\delta V_{FB} = \frac{q}{\Delta x \Delta z C_{ox}} \left(1 - \frac{y_t}{T_{ox}} \right) \quad (4.96)$$

La fluctuation du courant total de grille va donc s'écrire :

$$\Delta I_G = \left(\frac{dI_G}{dV_{GB}} \right) \frac{q}{WLC_{ox}} \left(1 - \frac{y_t}{T_{ox}} \right) \quad (4.97)$$

Cette fluctuation du courant correspond à la capture d'un porteur. La simulation réalisée avec ce modèle donne une amplitude largement inférieure à celle que nous avons mesurée.

Les phénomènes de RTS sur le courant de grille ont été étudiés dans l'article de Avellan et al [AVELLAN'01]. Dans cette étude, l'auteur modélise uniquement l'amplitude du signal RTS, et ne propose pas de modèles pour les temps de capture et d'émission. L'approche est également basée sur une modulation de la conductance tunnel, celle-ci étant liée à la barrière de potentiel susceptible de varier par la présence d'une charge dans l'oxyde. Cette approche est dérivée du modèle d'atténuation de la barrière des diodes Schottky développé par Sze [SZE'81]. Cette approche donne des résultats similaires à ceux obtenus

avec le modèle de fluctuation équivalente de la tension de bandes plates, à savoir qu'elle sous estime l'amplitude du signal RTS. Cependant, l'auteur propose, non pas de considérer la capture d'une charge, mais d'un ensemble de charges, qui peut être interprété comme un « chemin de pièges » (*conductive path*)[CESTER'00]. Un facteur multiplicatif η doit donc être appliqué à l'équation (4.97) pour prendre en compte le nombre de charges piégées. L'amplitude du bruit RTS du courant de grille s'écrit donc :

$$\Delta I_G = \eta \left(\frac{dI_G}{dV_{GB}} \right) \frac{q}{WLC_{ox}} \left(1 - \frac{y_t}{T_{ox}} \right) \quad (4.98)$$

La valeur typique du facteur η est comprise entre 10 et 1000. Sur la figure 4.24, on observe un excellent accord entre les mesures et le modèle (4.98) avec $\eta = 200$.

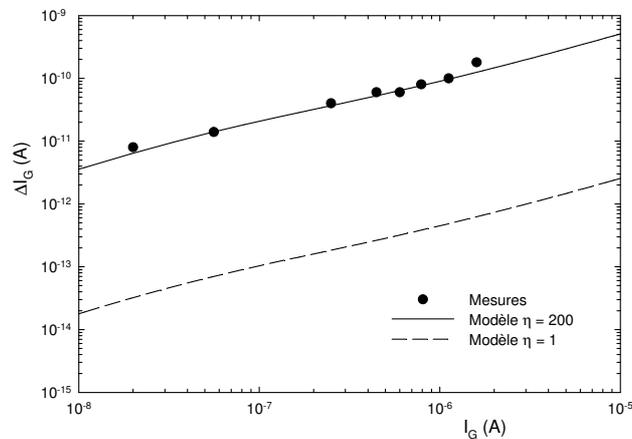


Figure 4.24 : Amplitude du bruit RTS associé au courant de grille sur un transistor nMOS du lot DPN12 W x L = 10 μ m x 10 μ m à $V_{DS} = 0$ V.

4.7 Conclusion

Dans le paragraphe 4.3, nous avons détaillé la formulation du phénomène de piégeage-dépiégeage d'un porteur libre dans le régime de forte inversion, puis dans le régime de faible inversion. Dans le cas du régime de forte inversion, nous avons reformulé les expressions classiques du processus aléatoire par rapport au modèle de potentiel de surface défini au chapitre 2. Nous avons notamment introduit dans l'expression de la constante de temps les phénomènes de réflexion à l'interface Si-SiO₂. Nous avons montré que la formulation classique de la densité spectrale des fluctuations de l'occupation des pièges d'oxyde reste valable pour les diélectriques ultrafins, et notamment que, comparé à un calcul numérique de la densité spectrale, la valeur $\gamma = 10^{10} \text{ m}^{-1}$ permet d'obtenir de bons résultats dans la gamme 1 Hz - 100 kHz.

Nous nous sommes ensuite intéressés dans le paragraphe 4.3.2 au mécanisme de piégeage via les états d'interface. Compte tenu du peu de littérature à ce sujet, nous avons appliqué le formalisme des processus aléatoires bivariable développé par vanVliet à notre structure MOS en régime de faible inversion. Nous avons montré que d'une part, les constantes de temps liées aux états d'interface sont négligeables devant celles liées aux pièges d'oxyde, et d'autre part, que la formulation analytique de la densité spectrale d'occupation des pièges d'oxyde avec $\gamma = 10^{10} \text{ m}^{-1}$ permettait de modéliser analytiquement le processus de piégeage dépiégeage en faible inversion.

Dans le paragraphe 4.4, nous avons détaillé la formulation du bruit associé au courant du canal, dans le cadre du modèle de fluctuation du nombre de porteurs ΔN , et du modèle de fluctuation du nombre de porteurs et de mobilité corrélée $\Delta N - \Delta \mu$. Nous avons adapté la formulation classique aux variables utilisées dans le modèle développé au chapitre trois. Nous avons notamment montré que le rapport entre la fluctuation des charges libres et la fluctuation des charges piégées dans l'oxyde n'était plus égal à -1 en régime de forte inversion, à cause des effets quantiques et de la polydéplétion.

Nous avons présenté dans le paragraphe 4.5 les résultats expérimentaux relatifs à la caractérisation en bruit des transistors des lots RTN14, RTN12 et DPN12. Pour chacun des trois lots, nous avons extrait les densités moyennes de pièges d'oxyde et le coefficient moyen de corrélation de mobilité. Pour les transistors nMOS, la réduction de l'épaisseur d'oxyde s'accompagne d'une augmentation de la densité de pièges d'oxyde. Nous avons également observé que le procédé de nitruration DPN n'amenait aucune amélioration en terme de bruit.

D'après les études de la littérature menées sur des oxydes plus épais, les mesures de bruit permettent d'établir que le processus de nitruration DPN diminue le nombre pièges d'oxyde proche de l'interface Si-SiO₂. Avec nos composants à oxyde ultramine, nous n'avons pas observé d'amélioration. Nous avons attribué ce phénomène d'une part à la faible épaisseur du film diélectrique, et d'autre part à des défauts occasionnés par le fort dopage du canal nécessaire au bon fonctionnement des transistors de longueur 30 nm. Pour les transistors pMOS, il n'a pas été mesuré d'évolution significative des paramètres de bruit entre les trois lots de transistors.

Le paragraphe 4.6 est consacré à l'étude du bruit basse fréquence associé au courant de grille. Le modèle que nous avons proposé est basé sur la modulation de la conductance tunnel induite par la fluctuation de l'occupation des pièges d'oxyde. Les simulations avec ce modèle présentent un excellent accord avec les données expérimentales. Les densités de pièges d'oxyde extraites avec ce modèle sont en accord avec les densités de pièges extraites à partir mesures de bruit sur le courant drain.

Nous avons également modélisé la forte augmentation de bruit du drain et source sur les transistors nMOS de longueur 10 µm. Cette augmentation de bruit est directement liée au partitionnement du courant de grille vers le drain et la source. Les simulations pour plusieurs valeurs de la tension de drain présentent un excellent accord avec les mesures.

Enfin le paragraphe 4.6.4 présente la caractérisation et la modélisation de bruit de type RTS sur le courant de grille. Nous avons mis en évidence que le défaut singulier vérifiait la statistique de Shockley Read Hall. En utilisant notre modèle de potentiel de surface, nous avons modélisé les constantes de temps de capture et d'émission, permettant ainsi d'extraire la profondeur du piège, sa position en énergie et sa section de capture. L'amplitude du signal RTS a également été étudiée. Conformément à la littérature, nous avons fait l'hypothèse que le défaut responsable du bruit RTS n'était pas unique, mais que l'on était en présence d'un ensemble de pièges se comportant de façon similaire. L'hypothèse du « chemin de pièges » nous a permis de modéliser correctement l'amplitude du signal RTS en fonction de la polarisation de grille, et d'extraire un nombre de charges piégées simultanément.

CONCLUSIONS
ET
PERSPECTIVES

Conclusions et Perspectives

Ce travail de thèse porte sur la modélisation et la caractérisation de technologies CMOS nanométriques. Pendant cette étude, nous nous sommes plus particulièrement focalisés sur la modélisation et la caractérisation des effets du confinement des porteurs de la couche d'inversion et du courant de grille, au travers d'une étude statique et dynamique. L'objectif principal de ce travail était de mettre en place des modèles de simulations prenant en compte les effets liés à la miniaturisation et nécessitant de faibles temps de calculs. Les modèles ainsi élaborés nous ont permis de caractériser les dispositifs MOS à diélectriques ultrafins.

Nous avons présenté, au chapitre 1, l'architecture générale des transistors MOS nanométriques. Au travers des différents points clé de l'optimisation des dispositifs, nous avons justifié l'introduction de modifications majeures dans leur fabrication permettant d'obtenir des dispositifs sub 100 nm. Tous les critères de performances des transistors MOS ont été rappelés, ainsi que les limites des technologies MOS sur substrat massif. Un bref aperçu des potentialités de la technologie SOI a permis de donner quelques pistes sur le futur de la course à la miniaturisation. La dernière partie de ce chapitre est consacrée à une description qualitative des nouveaux effets physiques dans les technologies MOS nanométriques. D'une part, l'impact des effets quantiques a été clairement identifié, et d'autre part, les mécanismes de fuites par effet tunnel au travers de la grille ultra mince ont été détaillés.

Le chapitre 2 est consacré à la modélisation du potentiel de surface et du courant tunnel direct dans les structures MOS unidimensionnelles à diélectrique ultrafin. Le potentiel de surface est le point central de tous les modèles décrits dans ce mémoire. Dans une première partie, nous avons rappelé la formulation des modèles implicites et explicites liant le potentiel de surface, la tension de grille, et le quasi niveau de Fermi, sans tenir compte du confinement quantique. Nous avons montré que la déplétion dans le matériau de grille ne pouvait plus être ignoré dans la modélisation de la capacité de la structure, puisqu'elle est responsable d'une diminution du couplage capacitif de près de 35 % pour les technologies nanométriques. Nous avons ensuite détaillé les résultats de mécanique quantique nécessaires à la compréhension du confinement des porteurs de couche d'inversion ou d'accumulation. Nous avons montré que les effets quantiques sont également responsables d'une diminution importante de la capacité de la structure, et qu'il est donc indispensable d'en tenir compte dans la caractérisation des oxydes ultrafins. Sur la base des approximations analytiques de la résolution de l'équation de Schrödinger issues de la littérature, nous avons présenté un modèle implicite permettant

d'obtenir le potentiel de surface en tenant compte du phénomène de polydépétion. Ce modèle, valable dans tous les régimes de fonctionnement, a été, au chapitre 3, un élément clé des caractérisations expérimentales de transistors MOS issus de technologies avancées. Nous avons modifié le modèle implicite de façon à obtenir un modèle explicite valable en régime de faible et forte inversion. Ce modèle, qui n'utilise aucune fonction de lissage, a permis d'obtenir une précision supérieure aux modèles publiés dans la littérature. Sa nature explicite lui permet également de servir de base aux modèles compacts basés sur le potentiel de surface du type MM11. Enfin, dans une dernière partie, nous avons détaillé la formulation du modèle de courant tunnel direct. Nous avons rappelé les bases de l'approximation WKB largement utilisée pour la modélisation des phénomènes tunnels. Nous avons également détaillé le calcul permettant de tenir compte des réflexions aux interfaces polySi-SiO₂ et Si-SiO₂ dans l'expression de la probabilité tunnel. Nous avons ensuite présenté le modèle de courant tunnel direct totalement couplé au modèle de potentiel de surface, permettant ainsi de simuler une caractéristique I_G-V_{GB} à partir des paramètres technologiques de la structure MOS.

Le chapitre 3 est consacré à la modélisation de transistors MOS soumis à de fortes fuites de grille. Sur la base du modèle en feuille de charge développé par Brews, nous avons intégré les modèles décrits au chapitre 2 pour obtenir un modèle prenant en compte les phénomènes de polydépétion, de confinement de la couche d'inversion, et de courant tunnel direct au travers du diélectrique de grille. Nous avons formulé une équation différentielle générale prenant en compte tous ces phénomènes. Nous avons notamment montré l'impact de la courbure du potentiel de surface induite par les fuites de la grille. La validation du modèle développé a été réalisée sur trois lots de transistors de technologie avancée. Au travers des mesures statiques courant – tension, nous avons montré les excellentes performances de ces transistors jusqu'à des longueurs de grille de 30 nm pour les transistors du lot DPN12. Dans le but de simuler les caractéristiques courant tension des transistors longs, nous avons tout d'abord extrait à partir des mesures $C-V$ les niveaux de dopages du canal et du polysilicium de grille, ainsi que l'épaisseur équivalente d'oxyde (EOT). Les mesures des courants de grille nous ont permis d'extraire les paramètres intervenant dans le courant tunnel direct, à savoir l'épaisseur physique du film diélectrique, la masse effective des porteurs dans l'oxyde, et les hauteurs de barrière de potentiel. L'ajustement de notre modèle aux mesures de courant drain, source et grille en régime linéaire nous a permis d'extraire une valeur de mobilité effective. Nous avons observé que la mobilité extraite de cette façon est supérieure à la mobilité extraite par la méthode Split CV. Cet écart, observé à la fois pour les transistors nMOS et pMOS des trois lots, est probablement dû au modèle de conduction utilisé dans la

méthode Split CV, qui ne prend pas en compte ni les effets quantiques, ni les effets de polydépuration. La dernière partie de ce chapitre présente l'impact du courant de grille sur les structures MOS de grande surface, d'un point de vue statique (I_G-V_{GB}) et dynamique ($C-V$). Cette partie nous a permis de mettre en évidence le phénomène de courbure du potentiel de surface sur le courant de grille, à savoir une décroissance du courant de grille normé à la surface avec une augmentation de la longueur du transistor. Les mesures de capacités sur les transistors de longueur 100 μm ont montré la nécessité de prendre en compte les effets non quasi statiques. Nous avons développé un modèle non quasi statique petit signal prenant en compte les effets quantiques, la polydépuration et le courant tunnel de grille. Que ce soit pour les transistors nMOS ou pMOS, nous avons obtenus un très bon accord entre les mesures et les simulations pour plusieurs fréquences, avec de faibles temps de calculs. Les modèles et méthodes de caractérisations présentées dans ce chapitre nous ont donc permis d'extraire des jeux de paramètres technologiques et de conduction pour les transistors à diélectriques ultrafins des trois lots.

Le quatrième et dernier chapitre est consacré à la modélisation et caractérisation en bruit basse fréquence des transistors MOS à diélectriques ultrafins. Dans un premier temps, nous avons vérifié la validité des modèles couramment utilisés dans la modélisation du bruit en $1/f$ sur le courant drain. Nous avons montré que le modèle classique donnant la constante de temps $\tau = \tau_0 \exp(\gamma y)$ reste valable pour les diélectriques ultrafins avec $\gamma = 10^{10} \text{ m}^{-1}$, en régime de faible et forte inversion. Nous avons reformulé le modèle de fluctuation du nombre de porteurs et de la mobilité corrélée par rapport aux variables introduites dans le modèle du chapitre 3. Nous avons également présenté le lien avec les formulations existantes, et notamment montré que notre formulation était équivalente à celle introduite par G. Ghibaudo. Nous avons ensuite présenté les résultats expérimentaux relatifs aux mesures de bruit en $1/f$ sur tous les transistors des différents lots. Cette étude nous a permis de mettre en évidence l'impact du changement du procédé de nitruration du diélectrique de grille. Nous n'avons pas observé d'amélioration de la densité de pièges d'oxyde pour les oxydes de grille ultrafins. Une étude équivalente avait été réalisée sur des oxydes plus épais, montrant une sensible amélioration en passant d'un procédé RTN à un procédé DPN. Avec nos composants à oxyde ultrafin, nous n'avons pas observé d'amélioration. Nous avons attribué ce résultat d'une part, à la faible épaisseur du film, et d'autre part, à des défauts supplémentaires occasionnés par les forts niveaux de dopages lors du passage du lot RTN12 au lot DPN12, dans le but de mieux maîtriser les effets de canal court. La dernière partie du chapitre est consacrée au bruit en $1/f$

associé au courant de grille, et à l'impact de ce courant de grille sur le bruit associé au courant drain et source. Nous avons présenté un modèle de bruit en $1/f$ associé au courant tunnel direct basé sur la modulation de la conductance tunnel par un phénomène de piégeage - dépiégeage des pièges d'oxyde. Nous avons donc pu extraire une densité de pièges d'oxyde par des mesures de bruit sur le courant de grille. Nous avons observé un très bon accord entre les mesures et les simulations pour plusieurs géométries différentes. La densité de pièges ainsi extraite est proche de la densité de pièges extraite par les mesures de bruit en $1/f$ sur le courant drain. Nous avons également modélisé la forte augmentation de niveau de bruit sur le courant drain et source observée expérimentalement sur les transistors de géométrie $W \times L = 10 \mu\text{m} \times 10 \mu\text{m}$. Les mesures et les simulations pour plusieurs valeurs de tension de drain ont présenté un très bon accord. Enfin, nous avons modélisé et caractérisé le bruit de type RTS observé sur le courant de grille. En utilisant le modèle de potentiel de surface du chapitre 2, nous avons modélisé les constantes de temps de capture et d'émission d'un défaut singulier responsable du bruit de type RTS. Les paramètres du piège ainsi extraits, à savoir sa section de capture, sa profondeur et son énergie sont du même ordre de grandeur que ceux trouvés dans la littérature relative aux défauts singuliers d'oxydes nitrurés, caractérisés par des mesures RTS sur le courant drain. L'étude de l'amplitude du bruit RTS sur le courant de grille a mis en évidence l'hypothèse que le défaut responsable du signal RTS n'est pas unique, mais que l'on est en présence d'un ensemble de pièges se comportant de façon similaire. En nous basant sur la littérature, nous avons fait l'hypothèse d'un « chemin de pièges ». Cette approche nous a permis de modéliser avec un très bon accord l'amplitude du bruit RTS observé expérimentalement.

En conclusion, le but de cette thèse était de développer des modèles simples permettant de prendre en compte, lors de la caractérisation statique et dynamique, les phénomènes liés à la miniaturisation des dispositifs. Devant l'étendue des phénomènes à traiter, nous avons concentré notre étude sur l'impact sur les caractéristiques courant - tension, capacité - tension et bruit basse fréquence du confinement quantique, de la polydéplétion de grille et du courant tunnel au travers du diélectrique, pour les dispositifs à canal long.

Cette étude permet de dégager quelques perspectives de travail futures en matière de modélisation en vue de la caractérisation de technologies MOS avancées. D'une part, la modélisation des effets de canal court permettrait de prendre en compte les phénomènes traités dans cette thèse sur l'ensemble des géométries disponibles. Le développement de modèles nécessitant de faibles temps de calculs pour la caractérisation des transistors courts

permettrait de compléter l'ensemble des outils présentés dans cette thèse. De plus, il serait intéressant d'adapter ces modèles à des dispositifs autres que des transistors MOS sur silicium massif, tels que les transistors MOS SOI ou double grille. D'autre part, la modélisation du bruit basse fréquence du courant de grille que nous avons présenté n'est qu'un point de départ pour une étude plus poussée des phénomènes de piégeage - dépiégeage dans les oxydes ultrafins. Nous pensons qu'il est nécessaire de considérer des phénomènes plus complexes que ceux développés dans cette thèse en matière d'interaction entre les pièges eux mêmes. Une étude en température pourrait nous renseigner sur la nature exacte des phénomènes de transport dans les oxydes ultrafins, et nous permettrait d'identifier clairement les phénomènes tunnels directs (processus élastiques) et indirects (processus inélastiques). La modélisation de ces processus pourrait ce faire grâce aux outils mathématiques des processus stochastiques multi-variables abordés dans cette thèse.

ANNEXES

Annexe A : Approximation de la feuille de charge

La densité de charge totale Q_{sc} développée dans le semiconducteur, donnée par les relations (2.6) et (2.7), est le résultat de la densité de charges de la couche d'inversion Q_{inv} constituée d'électrons et de la densité de charges du substrat Q_{dep} constituée des atomes d'impuretés ionisés. Afin de donner une expression exacte de la charge d'inversion Q_{inv} , il est nécessaire d'intégrer la densité d'électrons n dans la direction transversale de l'interface Si/SiO₂ jusqu'à la zone neutre du substrat :

$$Q_{inv} = -q \cdot \int n \cdot dy \quad (A.1)$$

En utilisant l'expression du champ transversal $E_y = -\partial\Psi/\partial y$ et la relation (II.4) pour déduire n , nous pouvons écrire que :

$$Q_{inv} = -q \cdot N_A \cdot \int_0^{\Psi_s} \frac{\exp\left[\frac{(\Psi - \Phi_b - V)}{\phi_T}\right]}{E_y(\Psi, V)} d\Psi \quad (A.2)$$

où selon (II.7) et pour $V_{GB} > V_{FB}$, E_y est donné par :

$$E_y = \frac{\gamma \cdot C_{ox}}{\epsilon_{Si}} \sqrt{\Psi + \phi_T \cdot \left[\exp\left(-\frac{\Psi}{\phi_T}\right) - 1 \right] + \phi_T \cdot \exp\left(-\frac{V + \Phi_B}{\phi_T}\right) \cdot \left[\exp\left(\frac{\Psi}{\phi_T}\right) - 1 \right]} \quad (A.3)$$

L'équation (A.3) ne peut cependant être résolue explicitement. Afin d'obtenir une expression explicite de Q_{inv} , l'approximation de la feuille de charge est utilisée. Dans cette approximation, l'épaisseur de la couche d'inversion est supposée nulle (feuille de charge). En considérant cette hypothèse valide, la densité de charges du substrat peut être calculé en utilisant l'expression (2.5) pour laquelle l'influence des électrons est négligée :

$$\frac{\partial^2 \Psi}{\partial y^2} \approx \frac{q \cdot N_A}{\epsilon_{Si}} \cdot \left[1 - \exp\left(-\frac{\Psi}{\phi_T}\right) \right] \quad (A.4)$$

En intégrant cette expression, Q_{dep} s'obtient en appliquant la loi de Gauss :

$$Q_{dep} = \epsilon_{Si} \cdot \frac{\partial \Psi}{\partial y} \Big|_{y=0^+} = \pm \gamma \cdot C_{ox} \cdot \sqrt{\Psi_s + \phi_T \cdot \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right]} \quad (A.5)$$

Q_{dep} est négative pour $V_{GB} > V_{FB}$ (inversion) et positive pour $V_{GB} < V_{FB}$ (accumulation). La densité de charges de la couche d'inversion Q_{inv} est maintenant et simplement donnée par (pour $V_{GB} > V_{FB}$) :

$$Q_{inv} = Q_{sc} - Q_{dep} = -C_{ox} \cdot \left(V_{GB} - V_{FB} - \Psi_s - \gamma \sqrt{\Psi_s + \phi_T \cdot \left[\exp\left(-\frac{\Psi_s}{\phi_T}\right) - 1 \right]} \right) \quad (A.6)$$

Annexe B : Technique du transfert matriciel [CASSAN'00]

La technique du transfert matriciel présentée dans cette annexe est dédiée à la résolution de l'équation de Schrödinger indépendante du temps à une dimension dans le cadre de l'approximation de la masse effective. L'idée générale de la méthode consiste à approcher un profil d'énergie potentielle de forme quelconque par un profil d'énergie potentielle formé par la juxtaposition de la même forme élémentaire.

La figure B.1 précise l'essentiel des hypothèses et des notations adoptées. La structure globale décrite est formée de $N+2$ zones, où N est un entier, numérotées de 0 à $N+1$, à l'intérieure desquelles l'énergie potentielle est constante :

- Les zones extrêmes 0 et $N+1$ sont supposées infinies
- Chaque zone i , où $0 \leq i \leq N+1$, est caractérisée par une énergie potentielle E_{pi}
- La masse effective des porteurs dans la zone i est m^*_i

Dans chacune des zones élémentaires i , l'équation de Schrödinger indépendante du temps prend la forme suivante :

$$\frac{\partial^2 \Psi_i(x)}{\partial x^2} + \frac{2m^*_i(E - E_{pi})}{\hbar^2} \Psi_i(x) = 0 \quad (\text{B.1})$$

La fonction enveloppe Ψ_i d'énergie E se décompose donc en une somme de deux ondes planes de vecteur d'onde $\pm k_i$:

$$\Psi_i(x) = A_i \exp(j k_i x) + B_i \exp(-j k_i x) \quad (\text{B.2})$$

$$\text{Avec } k_i = \sqrt{\frac{2m^*_i(E - E_{pi})}{\hbar^2}} \quad (\text{B.3})$$

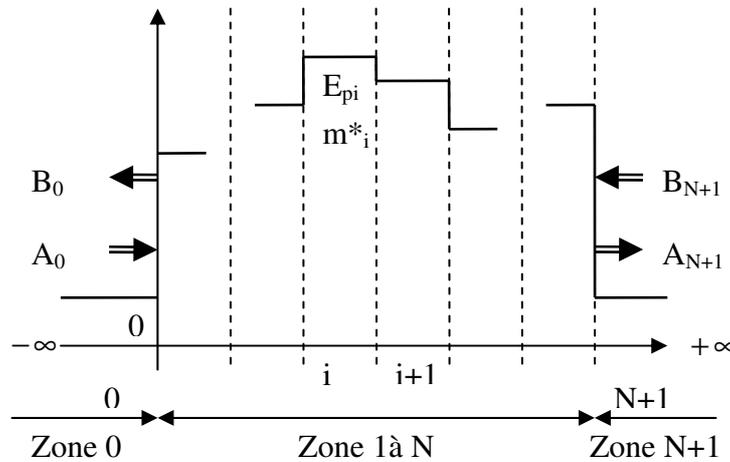


Figure B.1 : Profil d'énergie potentielle constant par morceaux à $N+2$ zones

La connaissance de la fonction enveloppe $\Psi(x)$ dans tout l'espace à une dimension est donc subordonnée à celles des $2(N+2)$ coefficients $((A_0, A_1, \dots, A_{N+1}), (B_0, B_1, \dots, B_{N+1}))$. Les relations de continuités portant sur la fonction d'onde et le courant de probabilité permettent de réduire le nombre de ces inconnues. A la frontière x_i entre les zones i et $i+1$, elles s'expriment par :

$$\Psi_i(x_i) = \Psi_{i+1}(x_i) \quad (\text{B.4})$$

$$\frac{1}{m^*_i} \left(\frac{\partial \Psi_i}{\partial x} \right)_{x_i} = \frac{1}{m^*_{i+1}} \left(\frac{\partial \Psi_{i+1}}{\partial x} \right)_{x_i} \quad (\text{B.5})$$

De ces deux expressions, un couple de relations liant $(A_i, A_{i+1}, B_i, B_{i+1})$ peut être tiré et se mettre sous la forme d'une matrice 2×2 notée M_i .

$$\begin{bmatrix} A_{i+1} \\ B_{i+1} \end{bmatrix} = M_i \begin{bmatrix} A_i \\ B_i \end{bmatrix} \quad (\text{B.6})$$

M_i étant définie par les deux relations suivantes :

$$M_i = \frac{1}{2} \begin{bmatrix} (1+S_i) \exp(-j(k_{i+1}-k_i)x_i) & (1-S_i) \exp(-j(k_{i+1}+k_i)x_i) \\ (1-S_i) \exp(j(k_{i+1}+k_i)x_i) & (1+S_i) \exp(j(k_{i+1}-k_i)x_i) \end{bmatrix} \quad (\text{B.7})$$

$$S_i = \frac{m^*_{i+1} k_i}{m^*_i k_{i+1}} \quad (\text{B.8})$$

La répétition de cette procédure pour i compris entre 0 et N fournit ainsi l'ensemble des $2(N+1)$ relations qui peut se mettre de manière synthétique sous la forme d'une matrice M_A :

$$\begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} = M_A \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} \quad (\text{B.9})$$

La matrice M_A qui contient une description de l'ensemble du profil d'énergie potentielle de la structure, a pour expression :

$$M_A = \prod_{i=N}^0 M_i = \begin{bmatrix} m_{11}(E) & m_{12}(E) \\ m_{21}(E) & m_{22}(E) \end{bmatrix} \quad (\text{B.10})$$

La détermination complète de la fonction enveloppe obéissant au profil d'énergie potentielle constant par morceaux nécessite la connaissance supplémentaire de deux des quatre coefficients $(A_0, A_{N+1}, B_0, B_{N+1})$. Cette dernière information est fournie par la situation physique traitée.

Exemple d'une barrière de potentiel de largeur finie :

Considérons le cas d'une onde plane d'énergie E lancée depuis $-\infty$ sur une barrière de potentiel de largeur finie (figure B.2).

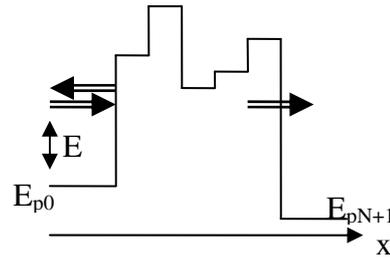


Figure B.2 : Onde plane lancée sur une barrière de potentiel

Les deux dernières relations permettant d'expliciter la fonction enveloppe dans tout l'espace sont obtenues dans cette situation en écrivant que l'amplitude de l'onde incidente est unitaire ($A_0=1$), et que le milieu après transmission ne possède pas d'onde réfléchie se propageant vers $-\infty$ ($B_{N+1}=0$). Les relations (B.9) et (B.10) permettent de déduire le coefficient de transmission complexe t et la probabilité tunnel P , respectivement définis comme le rapport complexe des ondes incidentes et transmises, et le rapport des courants de probabilité incident et transmis :

$$t(E) = \frac{\frac{\hbar k_0}{m^*_0}}{\frac{\hbar k_{N+1}}{m^*_{N+1}} m_{22}(E)} \quad (B.11)$$

$$P(E) = \frac{\frac{\hbar k_{N+1}}{m^*_{N+1}}}{\frac{\hbar k_0}{m^*_0}} |t(E)|^2 \quad (B.12)$$

Annexe C : Résolution Numérique de l'équation différentielle régissant le potentiel de surface

L'équation différentielle à résoudre est la suivante :

$$\left[\mu_{eff}(\Psi_s) \frac{\partial Q_{inv}^*}{\partial \Psi_s} + \frac{\partial \mu_{eff}}{\partial \Psi_s} Q_{inv}^*(\Psi_s) \right] \left(\frac{\partial \Psi_s}{\partial x} \right)^2 + \mu_{eff}(\Psi_s) Q_{inv}^* \frac{\partial^2 \Psi_s}{\partial x^2} = 0 \quad (C.1)$$

L'axe x est discrétisé suivant la figure C.1. Les potentiels de surface Ψ_{s0} et Ψ_{sL} aux points d'abscisse x_0 et x_L sont donnés par les conditions initiales.

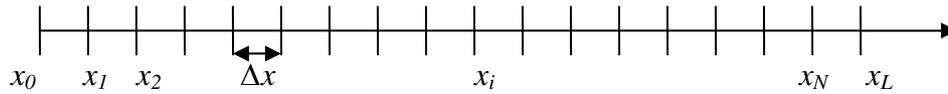


Figure C.4.25 : Discrétisation de l'axe x .

Les dérivées du potentiel de surface s'expriment à l'aide des différences finies par :

$$\begin{aligned} \frac{\partial \Psi_s}{\partial x}(x_i) &= \Psi_{s_i}' = \frac{\Psi_{s_{i+1}} - \Psi_{s_{i-1}}}{2\Delta x} \\ \frac{\partial^2 \Psi_s}{\partial x^2}(x_i) &= \Psi_{s_i}'' = \frac{\Psi_{s_{i-1}} - 2\Psi_{s_i} + \Psi_{s_{i+1}}}{\Delta x^2} \end{aligned} \quad (C.2)$$

Tous les termes intervenant dans les coefficients sont explicites de Ψ_s ; ils s'expriment donc en fonction de Ψ_{s_i} au point x_i .

L'équation (C.1) se met sous la forme discrétisée suivante :

$$A_i \Psi_{s_i}'' + B_i \Psi_{s_i}' = 0 \quad (C.3)$$

où les coefficients A_i et B_i s'expriment en fonction de Ψ_{s_i} , $\Psi_{s_{i-1}}$ et $\Psi_{s_{i+1}}$:

$$\begin{aligned} A_i &= \mu_{eff}(\Psi_{s_i}) Q_{inv}^*(\Psi_{s_i}) \\ B_i &= \left[\mu_{eff}(\Psi_{s_i}) \frac{\partial Q_{inv}^*}{\partial \Psi_s} \Big|_{\Psi_s = \Psi_{s_i}} + \frac{\partial \mu_{eff}}{\partial \Psi_s} \Big|_{\Psi_s = \Psi_{s_i}} Q_{inv}^*(\Psi_{s_i}) \right] \left(\frac{\Psi_{s_{i+1}} - \Psi_{s_{i-1}}}{2\Delta x} \right) \end{aligned} \quad (C.4)$$

L'équation différentielle discrétisée se ramène donc à une relation de récurrence entre Ψ_{s_i} , $\Psi_{s_{i-1}}$ et $\Psi_{s_{i+1}}$:

$$\begin{aligned} a_i \Psi_{s_{i-1}} + b_i \Psi_{s_i} + c_i \Psi_{s_{i+1}} &= 0 \\ a_i &= \frac{A_i}{\Delta x^2} - \frac{B_i}{2\Delta x} \quad b_i = -\frac{2A_i}{\Delta x^2} \quad c_i = \frac{A_i}{\Delta x^2} + \frac{B_i}{2\Delta x} \end{aligned} \quad (C.5)$$

Cette relation de récurrence, valable en tout point i , se met sous la forme matricielle suivante :

Annexe D : Modèle petit signal non quasi statique

L'équation différentielle à résoudre est la suivante :

$$\begin{aligned} & \left[\mu_{eff}(\Psi_S(x,t)) \frac{\partial Q_{inv}^*(x,t)}{\partial \Psi_S} + \frac{\partial \mu_{eff}(x,t)}{\partial \Psi_S} Q_{inv}^*(\Psi_S(x,t)) \right] \left(\frac{\partial \Psi_S}{\partial x} \right)^2 \\ & + \mu_{eff}(\Psi_S(x,t)) Q_{inv}^*(x,t) \frac{\partial^2 \Psi_S(x,t)}{\partial x^2} \\ & = J_G(\Psi_S(x,t)) - \frac{\partial Q_{inv}(\Psi_S(x,t))}{\partial t} \end{aligned} \quad (D.1)$$

Dans le cadre de l'approche petit signal, la tension de grille est un signal sinusoïdal de faible amplitude centré autour d'une valeur de polarisation. Le potentiel de surface est la somme du potentiel de surface $\Psi_{S_0}(x)$ et d'une fluctuation de faible amplitude $\widetilde{\Psi}_S(x,t)$. La discrétisation de l'équation (D.1) suivant la variable x est similaire à celle décrite dans l'annexe C. L'équation D.1 se met sous la forme d'une relation de récurrence :

$$F_i(\Psi_{S_{i-1}}(t), \Psi_{S_i}(t), \Psi_{S_{i+1}}(t), V_{GB}(t)) = \frac{\partial Q_{inv}(\Psi_{S_i}(t), V_{GB}(t))}{\partial t} \quad (D.2)$$

Posons les notations complexes pour les variables $\Psi_S(x,t)$ et $V_{GB}(t)$:

$$\begin{aligned} V_{GB} &= V_{GB_0} + v_g e^{j\omega t} \\ \Psi_{S_i}(t) &= \Psi_{S_{0i}} + \widetilde{\Psi}_{S_i} e^{j\omega t} \end{aligned} \quad (D.3)$$

$\Psi_{S_{0i}}$ (indépendant du temps) correspond à la solution de l'équation C.1 correspondant à un potentiel de grille constant V_{GB_0} . v_g est l'amplitude du signal appliqué sur la grille. Il constitue notre référence de phase (v_g est réel). $\widetilde{\Psi}_{S_i}$ est l'amplitude complexe de la variation du potentiel de surface autour de $\Psi_{S_{0i}}$.

Q_{inv} dépend du potentiel de surface et de la tension de grille. On peut donc écrire la relation suivante pour la fluctuation de la charge d'inversion :

$$\frac{\partial Q_{inv}(\Psi_{S_i}(t), V_{GB}(t))}{\partial t} = j\omega \left[\frac{\partial Q_{inv}}{\partial \Psi_S} \widetilde{\Psi}_{S_i} + \frac{\partial Q_{inv}}{\partial V_{GB}} v_g \right] \quad (D.4)$$

En développant l'équation D.1 à l'ordre 1, on obtient :

$$\frac{\partial F_i}{\partial \Psi_{S_{i-1}}} \widetilde{\Psi}_{S_{i-1}} + \frac{\partial F_i}{\partial \Psi_{S_i}} \widetilde{\Psi}_{S_i} + \frac{\partial F_i}{\partial \Psi_{S_{i+1}}} \widetilde{\Psi}_{S_{i+1}} + \frac{\partial F_i}{\partial V_{GB}} v_g = j\omega \frac{\partial Q_{inv}}{\partial \Psi_{S_i}} \widetilde{\Psi}_{S_i} + j\omega \frac{\partial Q_{inv}}{\partial V_{GB}} v_g \quad (D.5)$$

Références bibliographiques

[A]

- [AGILENT'01] Agilent, “**Evaluation of Gate Oxides Using a Voltage Step Quasi-Static CV Method**”, note d’application 4156-10, 2001
- [AHSAN'05] A. K. M. Ahsan, D. K. Schroder, “**Impact of channel carrier displacement and barrier height lowering on the low-frequency noise characteristics of surface-channel n-MOSFETs**”, *Solid-State Electronics*, vol. 49, pp. 654–662, 2005.
- [AMARASINGHE'03] N. Amarasinghe, Z. Celik-Butler, A. Zlotnicka. And F. Wang, “**Model for random telegraph signals in sub-micron MOSFETS**”, *Solid-State Electronics*, vol. 47, pp.1443–1449, 2003.
- [ANDO'82] T Ando, A. B. Fowler and F Stern, “**Electronic properties of two dimensional systems**”, *Rev. Mod. Phys.* 54, pp 437-672, 1982.
- [AVELLAN'01] A. Avellan, W. Krautschneider and S. Schwantes, “**Observation and modelling of random telegraph signals in the gate and drain currents of tunnelling metal-oxide-semiconductor field-effect transistors**”, *Applied Physics letters*, Vol. 78(18), pp 2790-2792, 2001.

[B]

- [BOULLION'95] P. Boullion et al., “**Re-Examination of Indium Implantation for low power 0.1 μm technology**”, *Tech. Dig. IEDM*, 1995 p 897.
- [BOUTCHACHA'99] T. Boutchacha, G. Ghibaud, B. Belmeki, “**Study of low frequency noise in 0.18 μm silicon CMOS transistors**”, *proc IEEE ICMTS*, Goteberg, Sweden, 1999, Vol 12, pp 84-88.
- [BREWS'78] J. R. Brews, “**A charge sheet model of the MOSFET**”, *Solid-state Electronics*, vol 21, pp 345-355, 1978
- [BRUT'96] H. Brut, “**Contribution à la modélisation et à l’extraction des paramètres de tension de seuil, de résistance série et de**

réduction de longueur dans les transistors MOS submicroniques”, Thèse, INPG Grenoble, 1996.

[BSIM'01]

BSIM 4.2 Manual, Univ. of California, Berkeley, Apr 2001.
<http://www.device.eecs.berkeley.edu/~bsim>

[BURSTEIN'69]

E. Burstein and S Lunquist, **“Tunnelling phenomena in solids”**, New York: Plenum, 1969.

[C]

[CASSAN'00]

E. Cassan., **“Etude théorique du courant tunnel de grille dans les transistors MOSFET sub-0,1 μm ”**”, *PhD Dissertation*, IEF, 2000.

[CESTER'00]

MA. Cester, A. Paccagnella, L. Bandiera and G. Ghidini, **“Switching Behaviour and Noise of Soft Breakdown Current in Ultra-Thin Gate Oxide ”**, *Proc. ESSDERC 2000*, pp. 500-504, 2000.

[CHEN'01]

T. L. Chen and G. Gildenblat, **“Analytical approximation for the MOSFET surface potential”**, *Solid-State Electronics*, 45, pp 335-339, 2001.

[CHEN'96]

K. Chen et al., **“The impact of device scaling and power supply change on CMOS gate performance”**, *IEEE Electron Device Letters*, 17, May 1996 p 202.

[CHOI'00]

C.-H. Choi et al., **“Capacitance Reconstruction from Measured C-V in High Leakagen Nitride/Oxyde MOS”**, *IEEE Trans. Elec. Dev.*, 47, October 2000 pp 1843-1850.

[CHRISTENSSON'68]

S. Christensson, I. Lundström and C. Svensson, **“Low frequency noise in MOS Transistors –I theory”**, *Solid-state Electronics*, vol 11, pp 797-812, 1968.

[CLERC'02]

R. Clerc et al., **“Electrical characterization and modelling of MOS structures with ultra-thin oxide”**, *Solid-state Electronics*, 46, March 2002 pp 407-416.

[COHEN'73]

C. Cohen-Tanoudji, **“Mécanique quantique”**, Hermann, 1973

[COLINGE'98]

J.-P. Colinge, **“Fully-Depleted SOI CMOS for Analog Applications”**, *IEEE trans. Electron Dev.*, 45, May 1998, p 1010

- [CURRENT'00] M. I. Curent et al., « **Process Integration Issues for Doping of Ultra-Shallow Doping Junctions** », *J. Vac. Sci. Technol.*, B-18, 2000, p 468.

[D]

- [DEGRAEVE99] R. Degraeve et al., “**Degradation and breakdown in thin xide layers : mechanism, models and reliability prediction**”, *Mircoelectronics Reliability*, 39, p 1445.

- [DEPAS'95] M. Depas et al., “**Determination of tunneling parameters in ultra-thin oxide layer poly-Si/SiO₂/Si structures**”, *Solid-state electronics*, 38, August 1995, pp. 1465-1471

[E]

- [EYAA'03] A. Eya'a Monvgbote., “**Caractérisation et modélisation du bruit en 1/f de transistors MOS de technologies CMOS à 0.25 et 0.18 µm**”, Thèse de Doctorat, Université Montpellier 2, 2003

[F]

- [FANG'66] F. F. Fang and W. E. Howard, “**Negative Field-Effect Mobility on (100) Si Surfaces**”, *Phys Rev. Let.*, 16 1966, p797

- [FU'72] S. T. Fu and C. T. Sah, “**Theory and experiments on surface 1/f noise**”, *IEEE trans. Electron. Devices*, vol 19, pp 273-285, Feb 1972.

[G]

- [GHIBAUDO'91] G. Ghibauda and al., “**Improved Analysis of Low Frequency noise in Field effect MOS transistor**”, *Phys. Stat. Sol (a)*, vol. 124, pp.571, 1991.

- [GILIBERT'04] F. Gilibert, D. Rideau, S Bernardini, P. Scheer, M Minondo et al., “**DC and AC MOS transistor modelling in presence of high leakage and experimental validation**”, *Solid-State Electronics*, Volume 48, Pages 721-726, 2004.

[H]

- [HOOGE'69] F. N. Hooge, "**1/f noise is no surface effect**", *Physics letters*, Vol 29A,3, p 139, 1969.
- [HOU'03] V. Kumar and W. E. Dahlke, "**Simulation of oxide trapping noise in submicron n-channel MOSFET**" *IEEE Trans. Electron Devices*, vol. 50, pp. 846–852, Mar.2003.
- [HORI'94] T. Hori, "**A 0.1 μm CMOS Technology with Tilt Implanted Punchthrough Stopper (TIPS)**", *Tech. Dig. IEDM*, 1994, p. 75.
- [HP'94] HP 4284A Operating manual, Hewlett-Packard, Japan, 1994.
- [HUANG'90] C.-L. Huang and G.S. Gildenblat, "**Measurements and Modeling of the n-Channel MOSFET Inversion Layer Mobility and Device Characteristics in the Temperature Range 60-300 K**," *IEEE Trans. Electron Devices*, Vol. ED-37, No. 5, pp. 1289-1300, 1990.
- [HUNG'90a] K. K. Hung et al., "**A unified model for flicker noise in Metal-Oxide-Semiconductor**", *IEEE transaction on Electron Devices*, Vol 37, 3, pp 654, 1990
- [HUNG'90b] K. K. Hung et al., "**A physics-Based MOSFET noise model for circuit simulator**", *IEEE transaction on Electron Devices*, Vol 37, 5, pp 1323, 1990
- [HUTCHBY'02] J. A. Hutchby et al., "**Extending the road beyond CMOS**", *IEEE circuits & devices magazine*, March 2002, p 28.
- [I]**
- [IANNACCONI'98] G. Iannaccone, G. Lombardi, M. Macucci, and B. Pellegrini, "**Enhanced shot noise in resonant tunneling : Theory and experiment**," *Phys. Rev. Lett.*, vol. 80, no. 5, pp. 1054–1057, 1998.
- [IANNACCONI'00] G. Iannaccone, F. Crupi, B. Neri, and S. Lombardo, "**Suppressed shot noise in trap-assisted tunneling of metal-oxide-semiconductor capacitors**", *Appl. Phys. Lett.*, vol. 77, no. 18, pp. 2876–2878, 2000.
- [ITRS'99] **The international Roadmap for Semiconductor**, Semiconductor Industry Association, 1999 Edition
- [ITRS'01] **The international Roadmap for Semiconductor**, Semiconductor Industry Association, 2001 Edition

- [ITRS'03] **Executive Summary Chapter. The international Roadmap for Semiconductor**, *Semiconductor Industry Association*, 2003 Edition
- [ITRS'03b] **Emerging Research Devices Chapter. The international Roadmap for Semiconductor**, *Semiconductor Industry Association*, 2003 Edition

[J]

- [JACOBONI'77] C. Jacoboni, C. Canalo, G. Ottaviani and A. Quaranta, "A Review of Some Charge Transport Properties of Silicon," *Solid-State Electron.*, Vol. 20, No. 1, pp. 77-89, 1977.
- [JANIK'98] T. Janik, "Analysis of the MOS Transistor based on the Self Consistent solution to the Schrödinger and Poisson equation and on the Local Mobility Model", *IEEE Trans. Electron Dev.*, 45, June 1998, p. 1263.
- [JINDAL'78] R. P. Jindal and A. van der Ziel, "Carrier fluctuation noise in a MOSFET channel due to traps in the oxide", *Solid-State Electronics*, Volume 21, Issue 6, June 1978, Pages 901-903.
- [JOSSE'00] E. Josse, "Nouvelles architectures de grilles pour les générations CMOS 0.1 μm et en deçà", Thèse, Institut Polytechnique de Grenoble, 2000.

[K]

- [KIRKLEN'00] W.Kirklen Henson et al., "Analysis of leakage Currents and Impact on Off-State Power Consumption for CMOS Technology in the 100-nm Regime", *IEEE Trans. Electron Dev.*, 47, July 2000, p. 1393.
- [KIRTON'89] A. M.J Kirton. M.J Uren Ohata , "Noise in solid-state microstructures: A new perspective on individual defects, interface states, and low-frequency noise", *J Advances Physics*, 38 n°4, 1989,pp 367-468.
- [KLAASSEN'67] F. M. Klaassen and J. Prins, "Thermal noise of MOS transistors" *Philips Res. Repts.*, vol. 22, pp. 504–514, 1967.
- [KLEINPENNING'78] T. G. M. Kleinpenning, "On low-

- frequency noise in tunnel diodes**", *Solid-State Electronics*, vol. 21, pp. 927–931, 1978.
- [KLEINPENNING'82] T. G. M. Kleinpenning, "**On low-frequency noise in tunnel junctions**", *Solid-State Electronics*, vol. 25, pp. 78–79, 1982.
- [KONDO'99] M. Kondo and H. Tanimoto, "**An Accurate Coulomb Mobility Model for MOS Inversion Layer and Its Application to NO-Oxynitride Devices**," *IEEE Trans. Electron Devices*, Vol. ED-48, No. 2, pp. 265-270, 1999.
- [KUMAR'77] V. Kumar and W. E. Dahlke, "**Low-frequency noise in Cr-SiO₂-n-Si tunnel diodes**" *IEEE Trans. Electron Devices*, vol. 24, pp. 146–153, Feb.1977.
- [KRUTSICK'88] T.J. Krutsick and M.H. White, "**Consideration of Doping Profiles in MOSFET Mobility Modeling**," *IEEE Trans. Electron Devices*, Vol. ED-35, No. 7, pp. 1153-1155, 1988.
- [L]**
- [LANGEVELDE'97] R. van Langevelde and F.M. Klaassen, "**Effect of Gate-Field Dependent Mobility Degradation on Distortion Analysis in MOSFET's**," *IEEE Trans. Electron Devices*, Vol. ED-44, No. 11, pp. 2044-2052, 1997.
- [LANGEVELDE'00] R. van Langevelde et al., "**An explicit surface-potential based MOSFET model for circuit simulation**", *Solid-State Electronics*, 44, pp 409-418, 2000.
- [LANGEVELDE'01] R van Langevelde, et al., "**Gate current:Modeling, DL extraction and impact on RF performance**", IEDM 2001 pp 289-292.
- [LANGEVELDE'03] R van Langevelde, et al., "**Physical background of MOS model 11**", unclassified Report, Philips Electronics N. V. 2003.
- [LEE'91] K. Lee, J.-S. Choi, S.-P. Sim and C.-K. Kim, "**Physical Understanding of Low-Field Carrier Mobility in Silicon MOSFET Inversion Layer**," *IEEE Trans. Electron Devices*, Vol. ED-38, No. 8, pp. 1905-1911, 1991.
- [LEYRIS'05] C. Leyris

- [LIME'03] F. Lime et al., **“Characterization of effective mobility by split C(V) technique in N-MOSFETs with ultra-thin gate oxides”**, *Solid-State Electronics*, 47,2003, pp. 1147-1153.
- [LO'97] S. -H. Lo et al., **“Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's”**, *IEEE Electron. Dev. Letters*, 18, May 1997, p. 209.
- [M]**
- [MARIN'03] M. Marin, **“Etude expérimentale du bruit en 1/f dans les composants CMOS issus de technologies sub-0.18 μ m”**, Thèse de doctorat, Université Montpellier 2, 2003
- [MARIN'05] M. Marin, J.C. Vildeuil, B. Tavel, B. Duriez, F. Arnaud, P. Stolk and M. Woo, **“Can 1/f noise in MOSFETs be reduced by gate oxide and channel optimization”**, proceedings of INCF 2005, Salamanca, Spain.
- [MARTINEZ'05a] F. Martinez, C. Leyris, M. Valenza, A. Hoffmann, F. Boeuf, T. Skotnicki, M. Bidaud, D. Barge, B. Tavel, **“Effect of oxide thickness and nitridation process on pMOS Gate and drain Low frequency noise”**, proceedings of INCF 2005, Salamanca, Spain.
- [MARTINEZ'05b] F. Martinez, C. Leyris, G. Neau, M. Valenza, A. Hoffmann, J.C. Vildeuil, E. Vincent, F. Boeuf, T. Skotnicki, M. Bidaud et al., **“Oxide trap characterization of 45 nm MOS transistors by gate current R.T.S. noise measurements”**, *Microelectronic Engineering*, Volume 80, June 2005, Pages 54-57.
- [MATHIEU'99] H. Mathieu, **“Physique des semiconducteurs et des composants électroniques”**, Masson, 1999
- [McWHORTER'57] A. L. Mc Whorter, **“1/f noise and germanium surface properties”**, Semiconductor surface physics, edited by R. H. Kingston, University of Pennsylvania Press, Philadelphia, 1957.
- [MERCHA'02] A. Mercha, E. Simoen, G. Richardson and C. Claeys, **“Inversion layer quantization impact on the interpretation of 1/f Noise in**

deep submicron CMOS transistors", *Proc. ESSDERC 2002*, pp. 79–82, 2002.

[MESSIAH'59] A. Messiah, "Mécanique Quantique", Dunod 1959, tome 1 pp194-203

[MULLER'86] R. Muller and T. Kamins, "**Device electronics for integrated circuit**", J. Wiley & Sons 2nd edition 1986.

[N]

[NOUGIER'87] J.P.Nougier, "**Méthodes de calcul numérique**", Masson, 1987.

[O]

[OHATA'90] A. Ohata et al., "**Observation of random telegraph signals: Anomalous nature of defects at the Si-SiO₂ interface**", *Journal of Applied Physics*, vol. 68(1), p 200, 1990.

[P]

[PACELLI'99] A. Pacelli et al., "**Carrier quantization at flat bands in MOS devices**", *IEEE Trans. Elec. Dev.*, 46, February 1999 pp 383-387.

[PALESTRI'03] Palestri, P.; Esseni, D.; Selmi, L.; Guegan, G.; Sangiorgi, E. , "**A methodology to extract the channel current of permeable gate oxide MOSFETs**", *Electron Devices, IEEE Transactions on*, Vol.50, Iss.5, May 2003.Pages: 1314- 1321.

[PREGALDINY'03] F. Prégaldiny, "**Etude et Modélisation du comportement électrique des transistors MOS fortement submicroniques**", PhD dissertation, 2003.

[PREGALDINY'04] F. Prégaldiny et al., "**An advanced explicit surface potential model physically accounting for the quantization effects in deep-submicron MOSFETs**", *Solid-State Electronics*, 48, pp 427-435, March 2004.

[Q]

[QMCV'00] Quantum-mechanical CV Simulator, UC Berkeley Device Group, April 2000, <http://www.device.eecs.berkeley.edu/research/qmcv/>

[R]

- [RANUAREZ'05]** Ranuarez, J.C.; Deen, M.J.; Chih-Hung Chen “**Modeling the partition of noise from the gate-tunneling current in MOSFETs**”, *Electron Device Letters, IEEE*, Vol.26, Iss.8, Pages: 550- 552, Aug. 2005
- [REGISTER'99]** L.F. Register, E. Rosenbaum and K. Yang, “**Analytic model for direct tunneling current in polycrystalline silicon-gate metal-oxide-semiconductor devices**”, *Appl Phys Lett.*, 74, 1999, pp. 457-459.
- [REKLAITIS'99]** A. Reklaitis and L. Reggiani, “**Prediction of enhanced shot-noise in tunneling-controlled transport,**” *Phys. B*, vol. 272, pp. 279–281, 1999.
- [RICHTER'01]** C. A. Richter et al., “**A comparison of quantum-mechanical capacitance-voltage simulators**”, *IEEE Electron. Dev. Letters*, 22, January 2001, pp.35-37.
- [RIOS'94]** R. Rios et al., “**Determination of ultra-thin gate oxide thickness for CMOS structures using quantum effects**”, *Tech. Dig. IEDM*, 1994 p 613.
- [RIOS'95]** R. Rios and al., “**A physical compact MOSFET model including quantum mechanical effects, for statistical circuit design application**”, *proc. IEEE Int. Electron Devices Meeting (IEDM)*, 1995, pages 937-940.
- [ROMANJEK'05]** K. Romanjek, F. Andrieu, T. Ernst and G. Ghibaudo, “**Characterization of the effective mobility by split C(V) technique in sub 0.1 μm Si and SiGe PMOSFETs**”, *Solid-State Electronics*, Volume 49, Issue 5, May 2005, Pages 721-726.
- [ROUX'93]** O. Roux-dit-Buisson, “**Étude du bruit électrique en 1/f et des fluctuation RTS aux basses frequencias dans le transistors MOS submicronique**”, Thèse, INPG Grenoble, 1993

[S]

- [SCHRODER'98]** D. Schroder, “**Semiconductor Material and Device Characterization**”, J. Wiley & Sons 2nd edition 1998

- [SHOCKLEY'52] W. Shockley and W. T. Read, "**Statistics of Recombinations of Holes and Electrons**", *Physical review*, Vol. 87(5), pp. 835-842, 1952
- [SIM'95] J. Sim, "**An Analytical Deep Submicron MOS Device Model Considering Velocity Overshoot Behavior Using Energy Balance Equation**," *IEEE Trans. Electron Devices*, Vol. ED-42, No. 5, pp. 864-869, 1995.
- [SIMOEN'99] E. Simoen and C. Claeys, "**On the flicker noise in submicron silicon MOSFET's**", *Solid-State Electronics*, vol. 43, pp.865–882, 1999.
- [SKOTNICKI'00] T Skotnicki, "**Transistor MOS et sa technologie de fabrication**" *Encyclopédie des sciences et techniques de l'ingénieur, Traité d'Electronique*, Article E 2430, 2000.
- [STERN'72] F Stern, "**Self-consistent results for n-type Si inversion layers**", *Phys. Rev B*, 5, pp 4491-4899, June 1972.
- [SZE'81] S. M. Sze, "**Physics of semiconductor Devices**", 2nd Edition, Wiley, New York, 1981
- [T]**
- [TAKAGI'94a] S. Takagi, A. Toriumi, M. Iwase and H. Tango, "**On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I - Effects of Substrate Impurity Concentration**," *IEEE Trans. Electron Devices*, Vol. ED-41, No. 12, pp. 2357-2362, 1994.
- [TAKAGI'94b] S. Takagi, A. Toriumi, M. Iwase and H. Tango, "**On the Universality of Inversion Layer Mobility in Si MOSFET's: Part II - Effects of Surface Orientation**," *IEEE Trans. Electron*
- [TAKAGI'99] S. Takagi et al., "**Experimental evidence of inelastic tunneling in stress-induced leakage current**", *IEEE Trans. Electron Device*, 46, p 335. *Devices*, Vol. ED-41, No. 12, pp. 2363-2368, 1994.
- [TANG'02] Ting-wei Tang; Yiming Li, "**A SPICE-compatible model for nanoscale MOSFET capacitor simulation under the inversion condition**", *Nanotechnology, IEEE Transactions on*, Vol.1, Iss.4, Dec 2002,Pages: 243- 246

- [TAUR'98] Y. Taur, "**25 nm CMOS Design Considerations**", *Tech. Dig. IEDM*, 1998, p. 789.
- [THOMPSON'98] S. Thompson et al., "**MOS Scaling: Transistor Challenges for the 21st Century**", *Intel Technology Journal*, 3rd Quarter 1998.
- [TSIVIDIS'99] Y. P. Tsividis, "**Operation and modelling of the MOS transistor**", *Mc Graw Hill* 1999..

[V]

- [VALENZA'03] M. Valenza, A. Laigle, F. Martinez, A. Hoffmann and D. Rigaud, "**Impact of gate current noise on drain current noise in 90 nm CMOS technology**", *European Solid-State Device Research Conference ESSDERC 2003 proceedings*, pp. 287-290, 2003
- [VALENZA'04] M.Valenza and al., "**Overview of the impact of downscaling technology of 1/f noise in p-MOSFETs to 90nm**", *IEE Proc-Circuits Devices Syst.*, Vol 151, No 2, pp 102-110, April 2004
- [VAN'92] M. J. Van Dort et al., "**Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFET's**", *IEEE Trans. Electron Devices*, 39, April 1992, p. 932.
- [VAN'94] M. J. Van Dort et al., "**A simple model for quantisation effects in heavily-doped silicon MOSFET at inversion conditions**" *Solid-State Electronics*, 37 March 1994 pp. 411-4114
- [VANVLIET'65] K. M. VanVliet and J. R. Fasset, "**Fluctuations due to electronic transitions and transport in solids**", *Fluctuations phenomena in solids*, Edited by R. E. Burgess, Academic Press, Chapter 7, 1965
- [VILLANUEVA'03] D. Villanueva et al., "**Impact of the lateral Source / Drain Abruptness on MOSFET Characteristics and Transport Properties**", *Tech. Dig. IEDM*, 2003, p 9.4.1.

[W]

- [WANGA'01] F. Wanga, Z. Celik-Butler, "**An improved physics-based 1/f noise model for deep sub-micron MOSFETs**", *Solid state Electronics*, Vol 45 , pp 351-357, 2001

[WEINBERG'83] Z. A. Weinberg et al., **“Effect of silicon orientation and hydrogen anneal on tunneling from Si on SiO₂”**, *J Appl. Phys.*, 54, 1983, pp 2517-2521.

[WOLF'02] S. Wolf, **“Silicon Processing for the VLSI Era – Volume 4 – Deep Submicron Process Technology”**, Lattice Press, 2002.

[YU]

[YU'00] Z. YU et al., **“Circuit/Device modelling at quantum level”**, *IEEE trans. Electron Dev.*, 47, October 2000, p 1819

Publications

REVUES INTERNATIONALES AVEC COMITÉ DE LECTURE

- P1 M. Valenza, A. Hoffmann, D. Sodini, A. Laigle, F. Martinez, D. Rigaud, **“An overview of the impact of downscaling technology on 1/f noise in P MOSFETs to 90 nm”**, IEE Proc.-Circuits Devices Syst. Vol. 151, N°2, pp. 102-110, 2004.
- P2 F. Martinez, C. Leyris, G. Neau, M. Valenza, A. Hoffmann, J.C. Vildeuil, E. Vincent, F. Boeuf, T. Skotnicki, M. Bidaud et al, **“Oxide trap characterization of 45 nm MOS transistors by gate current R.T.S. noise measurements”**, Microelectronic Engineering, Volume 80, 17 June 2005, Pages 54-57.

CONFÉRENCES INTERNATIONALES AVEC COMITÉ DE LECTURE ET ACTES

- CI1 M. Valenza, A. Hoffmann, F. Martinez, A. Laigle, J. Rhayem, R. Gillon, M. Tack, **“Discussion on standard 1/f noise models in software packages: SPICE, HSPICE and BSIM3v3. Comparison to MOSFET noise data on commercial c025”**, SPIE's First International Symposium on Fluctuations and Noise, Santa Fe, New Mexico, USA, June 1-4, 2003. p. 494-502, ISBN 0-8194-4973-3 (poster).
- CI2 M. Valenza, A. Hoffmann, A. Eya'a Mvongbote, F. Martinez, A. Laigle, D. Rigaud **“Impact of process scaling on BSIM noise parameters 17th International Conference on Noise in Physical Systems and 1/f Fluctuations”**, Prague, République Tchèque, August 18-22, 2003.
- CI3 A. Hoffmann, M. Valenza, D. Rigaud, A. Laigle, F. Martinez, **“Correlated carrier number-mobility fluctuation model or access resistance noise in p-MOSFETs”**, 17th International Conference on Noise in Physical Systems and 1/f Fluctuations, Prague, République Tchèque, August 18-22, 2003.
- CI4 M. Valenza, A. Laigle, F. Martinez, A. Hoffmann and D. Rigaud, **“Impact of gate current noise on drain current noise in 90 nm CMOS technology”**, 33rd European Solid-State Device Research Conference, ESSDERC 2003, pp 287-290, Estoril, Portugal, September 16-18, 2003.
- CI6 T. Nguyen, M. Valenza, F. Martinez, G. Neau, J.C. Vildeuil, G. Ribes, M. Muller, V. Cosnier, T. Skotnicki, **“Gate and drain low frequency noise in HfO2 NMOSFETs”**, 18th International Conference on Noise in Physical Systems and 1/f Fluctuations, Salamanca, Espagne, September 19-23, 2005.

- CI7 F. Martinez, A. Laigle, A. Hoffmann, M. Valenza, A. Veloso, M. Jurczak, **“Contributions of channel gate and overlap gate currents on 1/f gate current noise for thin oxide gate P-MOSFETs”**, 18th International Conference on Noise in Physical Systems and 1/f Fluctuations, Salamanca, Espagne, September 19-23, 2005.
- CI8 F. Martinez, C. Leyris, M. Valenza, A. Hoffmann, F. Boeuf, T. Skotnicki, M. Bidaud, D. Barge, B. Tavel, **“Effect of oxide thickness and nitridation process on PMOS gate and drain low frequency noise”**, 18th International Conference on Noise in Physical Systems and 1/f Fluctuations, Salamanca, Espagne, September 19-23, 2005.

CONFÉRENCES INVITÉES

- II M. Valenza, A. Hoffmann, F. Martinez and A. Laigle **“Noise Modelling for circuits simulation”**, Advanced Compact Modeling Workshop, Estoril, 16 – 18 September, 2003.(invité)

CONFÉRENCES NATIONALES AVEC ACTES

- CN1 A. Laigle, F. Martinez, M. Valenza, **“Bruit BF associé au courant de grille dans les composants nanométriques, Workshop "AS Bruit" ”**, La Grande Motte 7-8 juin 2004.
- CN2 F. Martinez, A. Laigle, C. Leyris, S. Soliveres, P. Benoit, J. Raoult, C. Delseny, A. Hoffmann, F. Pascal, M. Valenza, **“Le bruit basse fréquence : outil de caractérisation des nanocomposants”**, 5èmes Journées Nationales RMNT 11-13 Octobre 2004, Cassis.

RAPPORTS INTERMÉDIAIRES:

- RI1 A. Hoffmann, A. Laigle, F. Martinez, M. Valenza, **“Noise characterisation for 45 nm devices”**, Contrat Europeen "ARTEMIS" (Advanced Research and Technology for Microelectronics Integrated Silicon devices).
- RI2 A. Hoffmann, A. Laigle, F. Martinez, M. Valenza, **“Noise characterisation for 37 nm devices”**, Contrat Europeen "ARTEMIS" (Advanced Research and Technology for Microelectronics Integrated Silicon devices).

RESUME

Ce travail de thèse porte sur la caractérisation et la modélisation analytique ou semi-analytique des effets liés à la miniaturisation du transistor MOS dans le but de développer des modèles physiques utilisables pour la caractérisation des dispositifs. Dans une première partie, les différents points clés de l'optimisation des MOSFETs nanométriques sont abordés, présentant ainsi les limites des technologies CMOS. Notre étude se focalise ensuite sur les effets du confinement quantique et des forts courants de grille présents dans les structures MOS à diélectriques ultrafins. Nous proposons un modèle de potentiel de surface prenant en compte les effets quantiques et la polydépletion de grille dans tous les régimes de fonctionnement. Un modèle de transistors MOS à canal long basé sur l'hypothèse de la feuille de charge et utilisant le modèle de potentiel de surface a été développé pour l'extraction des paramètres de conduction et technologiques des transistors d'une technologie 65 nm avec de forts courants de grille. De plus, un modèle non quasi statique petit signal a été développé pour la caractérisation des structures MOS de grande surface. La dernière partie de ce mémoire est consacrée à la modélisation du bruit basse fréquence de ces technologies avancées, et notamment le bruit basse fréquence associé au courant de grille. Les sources de bruits BF associées aux courants de grille et de drain ont été mesurées et interprétées à partir des modèles développés.

TITLE

Modeling and characterization of MOS structures with ultrathin dielectric: Application to the low frequency gate current noise.

ABSTRACT

This thesis deals with the characterization and analytical or semi-analytical modeling of the effects related to miniaturization of MOS transistor in order to develop physical models for device characterization. In a first part, various key points of the optimization of nanometric MOSFETs are reviewed, pointing out the limits of CMOS technologies. Then, our study focuses on the effects of quantum confinement and on high gate currents in MOS structures with ultrathin dielectrics. We propose a model of surface potential taking into account of the quantum and polydepletion effects in all regions of operation. A long channel MOS transistor model based on the charge sheet approximation and using the quantum surface potential model has been developed for the extraction of the technological and conduction parameters of 65 nm technology node transistors under high gate leakage conditions. Moreover, a small signal non quasi static model has been developed for the characterization of large area MOS structures. The last part of this thesis is devoted to modeling the low frequency noise related to ultrathin gate dielectrics, in particular to low frequency noise associated to the gate current. These models are used for the characterization of the ultrathin dielectric film by low frequency gate current noise measurements.

DISCIPLINE

Microélectronique

MOTS-CLES

MOSFET, modélisation, effets quantiques, polydéplétion, modèle Non quasi statique petit signal, potentiel de surface, bruit basse fréquence, courant de grille.

INTITULE ET ADRESSE DU LABORATOIRE :

Centre d'Electronique et de Micro-optoélectronique de Montpellier (CEM2)
Unité Mixte de Recherche 5507 CNRS - MEN
UNIVERSITE MONTPELLIER II - Sciences et Techniques du Languedoc
Place Eugène Bataillon - 34095 Montpellier Cedex 5 (France)
