



HAL
open science

Modélisation des structures Métal-Oxyde-Semiconducteur (MOS) : Applications aux dispositifs mémoires

Sandrine Bernardini

► **To cite this version:**

Sandrine Bernardini. Modélisation des structures Métal-Oxyde-Semiconducteur (MOS) : Applications aux dispositifs mémoires. Micro et nanotechnologies/Microélectronique. Université de Provence - Aix-Marseille I, 2004. Français. NNT : . tel-00007764

HAL Id: tel-00007764

<https://theses.hal.science/tel-00007764>

Submitted on 15 Dec 2004

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

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

Année 2004

THESE

présentée devant

L'UNIVERSITE D'AIX-MARSEILLE I

par

Sandrine BERNARDINI

Ingénieur INSA LYON

pour obtenir

LE GRADE DE DOCTEUR

Ecole Doctorale : Physique, Modélisation et Sciences pour l'Ingénieur

Spécialité : Physique et modélisation des systèmes complexes

MODELISATION DES STRUCTURES METAL-OXYDE-SEMICONDUCTEUR (MOS) : APPLICATIONS AUX DISPOSITIFS MEMOIRES

Directeur de thèse : Pascal MASSON

Soutenue publiquement le 8 octobre 2004 devant la commission d'examen :

Président	Pierre GENTIL	IMEP, INP Grenoble
Rapporteurs	Carole PLOSSU	LPM, INSA Lyon
	Gérard GHIBAUDDO	IMEP, CNRS, INP Grenoble
Examineurs	Thomas SKOTNICKI	ST Microelectronics, Crolles
	Pascal MASSON	L2MP-Polytech, Univ. Aix-Marseille I
	Rachid BOUCHAKOUR	L2MP-Polytech, Univ. Aix-Marseille I
Invitée	Barbara DE SALVO	CEA/LETI, Grenoble

A ma famille,

A Damien.

Remerciements

Les travaux présentés dans ce manuscrit ont été réalisés au sein de l'équipe Microélectronique du Laboratoire Matériaux et Microélectronique de Provence (L2MP). Je remercie le Directeur, Michel Lannoo, Directeur de recherches au CNRS, de m'avoir accueillie dans son Laboratoire. Je tiens également à remercier Monsieur Rachid Bouchakour, Professeur à l'université de Provence, qui m'a offert l'opportunité d'effectuer cette thèse au sein de son équipe dans des conditions de travail optimales durant toute cette période.

Toute ma gratitude va à mon directeur de thèse, Pascal Masson, Professeur à l'université de Provence avec qui j'ai eu plaisir à travailler durant ces trois années : j'ai beaucoup appris à ses côtés, tant sur le plan scientifique qu'humain. Je tiens à le remercier pour la confiance qu'il a su m'accorder en me proposant ce sujet de recherches.

Je remercie vivement Monsieur Pierre Gentil, Professeur à l'Institut National Polytechnique de Grenoble, qui m'a fait l'honneur de présider le jury de ma thèse. J'exprime toute ma reconnaissance à Madame Carole Plossu, Professeur à l'Institut National des Sciences Appliquées de Lyon et Monsieur Gérard Guibaudo, Directeur de recherches au CNRS, pour l'intérêt qu'ils ont porté à mes travaux en acceptant d'être les rapporteurs de cette thèse ainsi qu'à Monsieur Thomas Skotnicki, Directeur de l'équipe dispositifs avancés à ST Microelectronics Crolles, pour avoir accepté d'examiner ce travail. Je les remercie d'avoir consacré une partie de leur temps à la lecture de ce mémoire et de m'avoir permis de l'améliorer grâce à toutes leurs suggestions.

Je tiens à exprimer ma reconnaissance à toutes celles et à tous ceux qui ont contribué, directement ou indirectement, au bon déroulement de mon travail et en particulier les personnes avec lesquelles j'ai eu la chance d'interagir durant cette thèse : Laurent Raymond et Xavier Cuinet pour l'étude de la rugosité de la capacité, Jean Michel Portal pour son aide dans l'application de mes travaux de recherches au domaine du test, Frédéric Lalande sans qui l'étude expérimentale des charges fixes n'aurait pas été possible, Michel Houssa pour sa disponibilité, nos nombreuses discussions et les renseignements qu'il m'a apportés sur les matériaux high κ et le modèle de l'hydrogène, Barbara Desalvo sans qui l'accomplissement des travaux sur les mémoires à nano-

cristaux de silicium n'aurait pas vue le jour et enfin Romain Laffont et Luca Pernolia pour les études que nous avons effectuées ensemble sur ces mémoires innovantes.

Je souhaiterais également remercier Khalifa Aguir, Professeur à l'université Paul Cézanne pour m'avoir permis de faire mes premiers pas dans le monde de l'enseignement dès mon arrivée en thèse. Je remercie également Lakhdar Zaid et Bruno Imbert, mes collègues de TP de physique des composants.

J'adresse encore mes remerciements à toutes les personnes que j'ai côtoyées pendant ces trois années: les thésards et les permanents, tous ceux qui m'ont aidée à franchir de petites difficultés ou des problèmes plus techniques. Leurs conseils et soutiens sont inestimables.

Mes remerciements ne seraient pas complets si je n'exprimais pas ma profonde gratitude à Sandrine Fisher, Isabelle Gimenez, Cathy Paitel-Haldenwang, Evelyne Santacroce et à Cyril Vidal qui ont toujours répondu présents lors de mes demandes administratives et surtout lors de l'organisation des Journées Nationales du Réseau Doctoral en Microélectronique ; ils ont rendu mon séjour ici plus agréable.

J'ai également une pensée particulière pour les prochains sur la liste : Marc, Hassen, Laurent, Juliano, Bertrand, Valery, Henia, ... courage !

Mes derniers remerciements vont à ma famille et mes amis, et surtout à Damien, qui m'ont tous entourée et m'ont donné la force de passer les moments difficiles.

Merci à ceux que j'ai maladroitement oubliés...

Sommaire

Table des symboles et des abréviations	3
Introduction générale	9
Chapitre I. Rappels sur les structures MOS et les dispositifs mémoires	15
I.1. Introduction	15
I.2. La capacité MOS	15
I.2.1. La structure	15
I.2.2. Principe et régimes de fonctionnement	16
I.2.3. Modélisation de la capacité MOS	17
I.3. Le transistor MOS à enrichissement	25
I.3.1. La structure	25
I.3.2. Principe et régimes de fonctionnement	25
I.3.3. Modélisation du transistor MOS	27
I.4. Les mémoires non volatiles	30
I.4.1. Généralités	30
I.4.2. Les mémoires Flash	33
I.4.3. Les mémoires à nano-cristaux	38
I.5. Conclusion	41
Chapitre II. Modélisation de la capacité MOS	45
II.1. Introduction	45
II.2. La poly-désertion de la grille	45
II.2.1. Détermination des potentiels de surface	45
II.2.2. Tracé des potentiels de surface et des charges	48
II.2.3. Courbes C-V d'une capacité MOS avec poly-désertion	50
II.2.4. Détermination des dopages à partir de la courbe C-V	52
II.3. Non uniformité du dopage du substrat	56
II.3.1. Modélisation	57
II.3.2. Etude d'un dopage de substrat non constant	58
II.4. Effet d'une épaisseur non constante	61
II.4.1. Modélisation pseudo 2D des courbes C-V et I-V	61
II.4.2. Comparaison entre la modélisation pseudo 2D et 2D	66
II.5. Charges fixes latérales non constantes	70
II.5.1. Modélisation des courbes C-V et I-V	70
II.5.2. Impact de la non uniformité de charges	71
II.6. Charges fixes non uniformément réparties dans l'isolant	72
II.6.1. Modélisation de la courbe I-V d'une capacité MOS	72
II.6.2. Modélisation de la courbe I-V d'une capacité MIS	74
II.6.3. Simulation des courbes I-V pour une capacité MOS	76
II.6.4. Simulation des courbes I-V pour une capacité MIS	79
II.6.5. Résolution spatiale de la charge piégée Q_{ox}	80
II.6.6. Détermination de Q_{ox} (Δy)	82
II.6.7. Origine de la charge fixe générée dans l'oxyde	85
II.7. Conclusion	88

Chapitre III. Modélisation du transistor MOS	91
III.1. Introduction	91
III.2. Le modèle segmenté	91
III.3. Les résistances d'accès	93
III.3.1. Impact des résistances d'accès	94
III.3.2. Extraction des résistances d'accès	96
III.3.3. Validité de la méthode d'extraction	97
III.3.4. Résistances d'accès et réduction de la longueur du canal	98
III.3.5. Impact d'une forte résistance d'accès sur la caractéristique $I_{DS}(V_{GS})$	99
III.3.6. Modification apparente de la mobilité	100
III.3.7. Dissymétrie des résistances d'accès	102
III.4. La Poly-désertion de la grille	103
III.4.1. Description du phénomène	103
III.4.2. Modélisation de la poly-désertion du TMOS	105
III.5. Effets d'un dopage latéral de substrat non constant	108
III.5.1. Les structures NUDC (Non Uniformly Doped Channel)	109
III.5.2. Simulations avec le modèle segmenté	110
III.6. Impact de la non uniformité de l'épaisseur de l'oxyde de grille	112
III.6.1. Modèle GLNMOS (Gate Leakage NMOS)	113
III.6.2. Modèle segmenté modifié	114
III.6.3. Simulations de transistors MOS avec fuites de grille	117
III.7. Conclusion	127
Chapitre IV. Etude des structures à nano-cristaux de silicium	131
IV.1. Introduction	131
IV.2. Structures d'étude	131
IV.3. Modélisation d'une charge non uniformément répartie	133
IV.3.1. Modélisation pour un MOSFET	133
IV.3.2. Modélisation d'une mémoire à nodules	135
IV.4. Simulations des structures 1bit	137
IV.4.1. Simulations statiques des mémoires Flash à nodules	138
IV.4.2. Etude de la phase d'écriture des mémoires Flash à nodules	139
IV.5. Caractérisations électriques de structures avec nodules	141
IV.5.1. Etude des capacités avec nodules	142
IV.5.2. Etude des transistors avec nodules	146
IV.6. Conclusion	152
Conclusion générale et perspectives	153
Références bibliographiques	157
Annexe I Les équations du miroir de courant	169

Table des symboles et des abréviations

Symboles	Unités	Définitions
A_{eff}	m^2	Surface du dispositif (= $W \times L$)
a_i	—	Coefficients d'ionisation par impact
BC	—	Bande de conduction
BI	—	Bande interdite
BV	—	Bande de valence
b_i	—	Coefficients d'ionisation par impact
b_{ox}	—	Facteur de dépendance du courant de grille avec le champ électrique dans l'oxyde tunnel
C_D	F m^{-2}	Capacité associée à la zone désertée
C_{D0}	F m^{-2}	Capacité associée à la zone désertée pour $\Psi_s = \Psi_{s0}$
C_{it}	F m^{-2}	Capacité associée aux états d'interface
C_G	F m^{-2}	Capacité associée à la grille
C_{Gd}	F m^{-2}	Capacité associée à la grille en régime de désertion
$C_{G\text{inv}}$	F m^{-2}	Capacité associée à la grille en régime d'inversion
CHEI	—	Channel Hot Electron Injection
C_{ono}	F m^{-2}	Capacité associée à l'empilement Oxyde/Nitruure/Oxyde
C_{ox}	F m^{-2}	Capacité d'oxyde
C_{sc}	F m^{-2}	Capacité associée au semiconducteur
C_{tun}	F m^{-2}	Capacité de recouvrement de la source
D_{dot}	m	Diamètre des nodules
D_{it}	$\text{J}^{-1} \text{m}^{-2}$	Densité d'états d'interface
DIBL	—	Drain Induced Barrier Lowering
d_{tun}	m	Distance tunnel
EPROM	—	Erasable Programmable Read Only Memory
EEPROM	—	Electrically Erasable Programmable Read Only Memory
E	J	Energie

E_C	J	Energie du niveau le plus bas de la bande de conduction
E_{CS}	J	Energie du niveau le plus bas de la bande de conduction à l'interface
E_F	J	Energie du niveau de Fermi dans le semiconducteur loin de l'interface
E_{Fn}	J	Energie du quasi niveau de Fermi pour les électrons
E_{Fp}	J	Energie du quasi niveau de Fermi pour les trous
E_{FM}	J	Energie du niveau de Fermi dans le semiconducteur
E_g	J	Largeur de la bande interdite du semiconducteur
E_i	J	Niveau d'énergie intrinsèque loin de l'interface
E_{iS}	J	Niveau d'énergie intrinsèque à l'interface
E_T	J	Energie d'un niveau piège dans la bande interdite du semiconducteur
E_V	J	Energie du niveau le plus haut de la bande de valence loin de l'interface
E_{VS}	J	Energie du niveau le plus haut de la bande de valence à l'interface
f	Hz	Fréquence du signal de grille
Flash	—	Mémoire non volatile (§ I.4.2)
F_{imp}	Hz	Fréquence d'impact des électrons
FN	—	Fowler-Nordheim
g_m	$A V^{-1}$	Transconductance
I_{cond}	A	Courant de conduction
I_{diff}	A	Courant de diffusion
I_{DS}	A	Courant Drain – Source
I_G	A	Courant de grille
I_{OFF}	A	Courant à $V_{GS} = 0$
I_{sub}	A	Courant de substrat
J_{DT}	$A m^{-2}$	Densités de courants tunnels direct
J_{FN}	$A m^{-2}$	Densités de courants tunnels Fowler-Nordheim
J_n	$A m^{-2}$	Densité de courant d'électrons en chaque point du canal
J_{tunnel}	$A m^{-2}$	Densités de courants tunnels

k	J K^{-1}	Constante de Boltzmann ($k = 1.38 \times 10^{23} \text{ J.K}^{-1}$)
L	m	Longueur de canal dessinée
L_B	m	Longueur de Debye extrinsèque
MIS	—	Metal-Insulator-Semiconductor
MOS	—	Metal-Oxide-Semiconductor
MOSFET	—	Transistor Métal-Oxyde-Semiconducteur à effet de champ
m_e	Kg	Masse effective de l'électron
m_M^*	Kg	Masse effective de l'électron dans le métal
m_{ox}	Kg	Masse effective de l'électron dans l'oxyde (en général on prend $m_{ox} \approx 0.5 m_0$)
N_A	m^{-3}	Concentration en atomes accepteurs
N_D	m^{-3}	Concentration en atomes donneurs
N_{dot}	m^2	Densité de nodules
N_{sub}	m^{-3}	Dopage du substrat
Nodules	—	Semiconducteurs emprisonnés dans un diélectrique (nommés aussi îlots ou dots)
NMOS	—	Transistor MOS ayant un canal de type N (substrat de type P)
n	m^{-3}	Concentration d'électrons libres dans le semiconducteur
n_{GO}	m^{-2}	Densités en porteurs majoritaires de la grille loin de l'interface
n_i	m^{-3}	Concentration intrinsèque d'électrons dans le semiconducteur
N_{pinch}	—	Nombre de transistors élémentaires dont les dots se remplissent par porteurs chauds
n_s	m^{-3}	Concentration d'électrons à l'interface
N_{tot}	m^{-2}	Densité totale de pièges
n_0	m^{-3}	Concentration d'électrons libres dans le substrat loin de l'interface
n_1	m^{-3}	Concentration d'électrons dans le cas où $E_F = E_T$
PMOS	—	Transistor MOS ayant un canal de type P (substrat de type N)
p	m^{-3}	Concentration des trous libres dans le semiconducteur
p_{GO}	m^{-3}	Densités en porteurs minoritaires de la grille loin de l'interface
p_s	m^{-3}	Concentration des trous à l'interface

p_0	m^{-3}	Concentration d'électrons libres dans le substrat loin de l'interface
p_1	m^{-3}	Concentration de trous dans le cas où $E_F = E_T$
Q_{dot}	$C m^{-2}$	Charge totale piégée dans les nodules
Q_G	$C m^{-2}$	Charge dans le poly-silicium de grille
Q_{ox}	$C m^{-2}$	Charge dans l'oxyde
$Q_{ox}(y)$	$C m^{-3}$	Charge volumique dans l'oxyde
Q_{oxeff}	$C m^{-2}$	Charge surfacique dans l'oxyde
Q_D	$C m^{-2}$	Charge dans la zone désertée du semiconducteur
Q_{D0}	$C m^{-2}$	Charge dans la zone désertée du semiconducteur pour $\Psi_s = \Psi_{s0}$
Q_{FG}	$C m^{-2}$	Charge dans la grille flottante
Q_{it}	$C m^{-2}$	Charge due aux états d'interface
Q_{it0d}, Q_{it0a} Q_{it0ad}, Q_{it0}	$C m^{-2}$	Charges constantes dues aux états d'interface selon le type d'état
Q_{SC}	$C m^{-2}$	Charge dans le semiconducteur
Q_n	$C m^{-2}$	Charge de la couche d'inversion
q	C	Valeur absolue de la charge de l'électron ($1.602 \times 10^{-19} C$)
RAM	—	Random Acces Memory
R_{eff}	%	Oxyde de grille occupée par les nodules
ROM	—	Read only Memory
RSCE	—	Reverse Short Channel Effect
R_{SD}	Ω	Résistance série
SCE	—	Short Channel Effect (charge sharing)
t_{ox}	m	Épaisseur d'oxyde
t_{ox1}	m	Épaisseur d'oxyde tunnel (couche d'oxyde entre l'interface Si/SiO ₂ et les nodules)
t_{ox2}	m	Épaisseur d'oxyde de contrôle (couche d'oxyde entre les nodules et la grille de contrôle)
T	K	Température absolue
T(E)	—	Coefficient de transmission des électrons d'énergie E, à travers la barrière énergétique triangulaire

$V(x)$	V	Potentiel le long du canal du à la polarisation Drain - Source
V_{BS}	V	Tension Substrat - Source
V_{CG}	V	Tension appliquée sur la grille de contrôle (Control Gate)
V_{DS}	V	Tension Drain - Source
V_{FB}	V	Tension V_{GS} pour laquelle $\Psi_S = 0$ à la source
V_{FG}	V	Tension aux bornes de la grille flottante
V_{GB}	V	Tension Grille – Bulk
V_{GS}	V	Tension Grille – Source
V_{ox}	V	Tension aux bornes de l'oxyde
V_{mg}	V	Tension V_{GS} pour laquelle $\Psi_S = \Phi_F$ à la source
V_{sat}	V	Potentiel appliqué aux bornes de la zone à saturation
V_T	V	Tension de seuil du transistor
V_{th}	V	Tension V_{GS} pour laquelle $\Psi_S = 2\Phi_F$ à la source
V_{Text}	V	Tension de seuil extrapolée du transistor
W	m	Largeur de canal dessinée
WKB	—	Wentzel-Kramers-Brillouin
x_d	m	Longueur de la zone désertée
X_D	m	Longueur de la région de chargement du transistor à nodules
y_d	m	Largeur de la zone désertée
y_{dM}	m	Extension maximale de la zone désertée (inversion forte)
y_i	m	Épaisseur de la couche d'inversion
ZCE	—	Zone de Charge d'Espace
α_{ox}	—	Facteur de dépendance du courant de grille par rapport au courant substrat
β	V	Potentiel thermique (q/kT)
ΔL	m	Sur-gravure de la longueur du canal
ΔV_T	V	Différence entre les tensions de seuil
ΔW	m	Sur-gravure de la largeur du canal
ϵ_0	F m ⁻¹	Permittivité du vide (8.85×10^{-12} F.m ⁻¹)
ϵ_{ox}	—	Constante diélectrique de l'oxyde (3.82)

ϵ_{SC}	—	Constante diélectrique du semiconducteur (11.9)
ϵ_{Si}	$F m^{-1}$	Permittivité du semiconducteur ($\epsilon_0 \times \epsilon_{SC}$)
θ_1	V^{-1}	Facteur d'atténuation linéaire de la mobilité dans le canal
θ_2	V^{-2}	Facteur d'atténuation quadratique de la mobilité dans le canal
μ_0	$m^2 V^{-1} s^{-1}$	Mobilité des électrons dans le canal à faible champ électrique
μ_{eff}	$m^2 V^{-1} s^{-1}$	Mobilité effective des électrons dans le canal
ρ	$C m^{-3}$	Densité volumique de charge
σ	m^2	Section de capture des pièges
ξ	$V m^{-1}$	Champ électrique
ξ_{ox}	$V m^{-1}$	Champ électrique interne dans l'oxyde
τ	s	Constante de temps de formation de la couche d'inversion
Φ_b	J	Hauteur de barrière
Φ_C	V	Ecart entre les quasi niveaux de Fermi
Φ_F	V	Potentiel de volume du semiconducteur
Φ_{MS}	V	Différence des travaux de sortie de la grille et du semiconducteur
Φ_{poly}	V	Variation de la barrière de potentiel vue par les électrons de la grille
Φ_{si}	V	Variation de la barrière de potentiel vue par les électrons du substrat
Ψ	V	Potentiel dans le semiconducteur
$\Psi(y)$	V	Courbure de bandes du semiconducteur
Ψ_S	V	Potentiel de surface du semiconducteur
Ψ_{S0}	V	Valeur particulière du potentiel de surface du semiconducteur : $\Psi_{S0} = 1.5 \Phi_F - V_{BS}$
Ψ_{SG}	V	Potentiel de surface de la grille
$\hbar = h / 2\pi$	eVs	Constante de Planck réduite ($4.13 \cdot 10^{-15}$ eVs)

Introduction générale

Depuis plus d'un siècle, l'industrie électronique reste surprenante, tant dans le domaine technique qu'économique. Sa croissance repose sur l'apparition incessante de nouveaux marchés, basés sur des produits de plus en plus sophistiqués (télévision, magnétoscope, caméscope, DVD...), et sur la pénétration de bien d'autres secteurs d'activité tels que l'automobile ou le bâtiment. À l'origine et au cœur de cette prodigieuse percée se situe la microélectronique. Celle-ci n'a jamais cessé de répondre à l'exigence de la rapidité et de l'intégration des composants avec leur miniaturisation, tout en recherchant le maintien de leur fiabilité et la réduction des coûts de production. Ainsi, petit à petit, notre quotidien s'est enrichi de nombreux dispositifs issus de l'industrie microélectronique, tels que les téléphones mobiles, les ordinateurs, les appareils photos numériques ou encore les agenda électroniques de poche (PALM). Tous ces produits de service ont en commun la présence de composants mémoires non volatiles, eux même constitués d'éléments électroniques de base : la résistance, le condensateur et le transistor. Par conséquent, une parfaite connaissance et maîtrise des phénomènes physiques intervenant dans le fonctionnement de ces composants élémentaires, qui se miniaturisent de jour en jour, sont nécessaires pour concevoir avec le moins d'empirisme possible les composants de demain.

Alors que Lilienfeld développe le concept du transistor à effet de champ en 1926, c'est le transistor bipolaire en germanium qui fût le premier créé (1947), par les physiciens Bardeen et Brattain. Il faut ensuite attendre l'année 1960, pour que Kahng et Atalla reprennent les travaux de Lilienfeld et aient l'idée du transistor MOS (**M**etal **O**xide **S**emiconductor) avec la silice (SiO_2) comme oxyde [Kahng'60]. Puis en 1963, Hofstein et Heiman proposent le transistor à effet de champ MOSFET (**M**étal **O**xide **S**emiconductor **F**ield **E**ffect **T**ransistor) [Hofstein'63]. De nos jours, celui-ci joue un rôle central dans la technologie silicium : en effet, les circuits à logique CMOS, à base de NMOSFETs (conduction assurée par les électrons) et PMOSFETs (conduction assurée par les trous) constituent la part la plus importante du chiffre d'affaire mondial des circuits intégrés. Bien que la longueur caractéristique des technologies MOS ait été fortement réduite entre le début des années 1970 ($L = 10 \mu\text{m}$) et aujourd'hui ($L = 0.13 \mu\text{m}$), la structure du MOSFET sur silicium et son principe de fonctionnement n'ont pas changé. Cependant, de nouveaux phénomènes physiques sont apparus avec la réduction des

dimensions et l'utilisation de nouveaux procédés de fabrication. Ces derniers dégradent les caractéristiques électriques des transistors et nuisent à leur fiabilité. Les mémoires, suivant la même loi d'échelle que les transistors MOS qui les composent, sont également affectées par ces effets parasites. Actuellement, la modélisation et la simulation électrique s'avèrent être deux outils parfaitement adaptés et peu coûteux pour étudier ces phénomènes et tenter, en les comprenant, de les minimiser au maximum.

Les travaux menés durant cette thèse se sont concentrés sur la modélisation des structures MOS affectées par des défauts qui détériorent leurs propriétés électriques. Le but de ces études était d'acquérir une parfaite connaissance et maîtrise des phénomènes physiques liés à la miniaturisation de la capacité et du transistor MOS, afin de créer les modèles de base permettant de simuler les différents mécanismes de fonctionnement des mémoires. Ainsi, ce manuscrit décrira non seulement nos modèles en tant que contributions supplémentaires aux nombreuses études réalisées sur ces sujets mais aussi en tant que nouveaux outils d'analyse, qui se sont révélés utiles pour une meilleure compréhension des phénomènes étudiés. La plupart de nos modèles ont d'ailleurs été transférés chez nos partenaires industriels dans le cadre de la convention STSI (Service des Technologies et de la Société de l'Information) qui lie notre laboratoire à la société STMicroelectronics. Enfin, nos modélisations des structures MOS ont été appliquées au cas particulier des mémoires Flash à nodules de silicium dans le cadre du projet européen Advanced Memories based on Discrete Traps (ADAMANT) en collaboration avec le CEA-LETI, l'IMEP (Institut de Microélectronique, d'Electromagnétisme et Photonique) et ST Microelectronics Catagne.

Le **premier chapitre** de ce manuscrit est consacré aux rappels des notations et des équations de base utilisées pour les structures étudiées dans le cadre de cette thèse. Le premier paragraphe est dédié à la description du fonctionnement de la capacité MOS ainsi qu'aux principales étapes de calcul de ses caractéristiques électriques. Ces structures sont capables de remplir des fonctions variées telles que la détection d'images, le stockage de données, les opérations logiques, le traitement des signaux. Elles sont également l'élément actif des transistors MOSFET présentés dans un second paragraphe, avec notamment un rappel sur leurs modélisations électriques "classiques". Enfin, dans un dernier paragraphe, nous présentons l'évolution des principales mémoires non volatiles utilisant des structures MOS. Une attention particulière est portée aux mémoires dont le stockage de charges s'effectue dans des nodules de silicium.

Nous proposons dans **le deuxième chapitre** différentes modélisations de la capacité MOS suivant l'effet parasite considéré. Dans un premier temps, la dégénérescence non complète de la grille (poly-désertion de grille) est étudiée en montrant son influence sur le potentiel de surface et l'extraction des paramètres de la capacité. Dans un deuxième temps, nous exposons une modélisation pour simuler l'effet d'un dopage de substrat non constant entre l'interface Si/SiO₂ et le volume du substrat. Les déformations des courbes C-V engendrées par divers profils de dopage sont présentées. L'uniformité de l'épaisseur d'isolant sur toute la surface d'un wafer (plaquette de silicium) est un paramètre clef pour garantir les mêmes caractéristiques électriques d'un composant à un autre. Par suite, dans un troisième paragraphe, nous décrivons un simulateur d'une non uniformité de l'épaisseur d'oxyde d'une capacité MOS afin de mettre en évidence la dégradation de ses caractéristiques électriques. Une comparaison entre notre modèle pseudo 2D et un modèle 2D permet de valider notre approche. Enfin, nous nous sommes intéressés à la présence des charges fixes dans le diélectrique. Une première étude concerne la modélisation de la non uniformité des charges fixes de l'oxyde le long de l'interface Si/SiO₂; puis, une seconde étude est consacrée à la non uniformité des charges fixes dans l'oxyde, entre les deux électrodes de la capacité. A partir de simulations et de mesures électriques, nous expliquons la méthode que nous avons développée afin de déterminer la répartition des charges fixes dans l'oxyde. L'origine physique de ces charges est également analysée.

Le troisième chapitre a pour objectif de présenter notre travail sur la modélisation et la caractérisation du transistor MOS. Dans la première partie, nous décrivons notre approche pseudo 2D pour prendre en compte les non uniformités présentes entre la source et le drain. Une première application de ce modèle est ensuite décrite, avec l'étude de l'influence des résistances séries sur les courbes $I_{DS}(V_{GS}, V_{DS})$. Puis, nous reprenons les études menées dans le chapitre II sur les non uniformités de dopages (dopage de substrat et dopage de grille), afin de les appliquer au cas des transistors. Nous avons également mené un travail dans un cadre plus amont sur le rôle du courant de grille du transistor MOS. Ainsi, le dernier paragraphe de ce chapitre est dédié à la modélisation des transistors à isolants ultra-minces. Nous présentons notamment les modifications de la caractéristique $I_{DS}(V_{GS}, V_{DS})$ induites par une réduction localisée de l'épaisseur d'oxyde du transistor.

Après avoir successivement modélisé le comportement des capacités et des transistors MOS, qui constituent les dispositifs mémoires, dans **le quatrième chapitre** nous appliquons nos modèles aux mémoires Flash à nodules. Jusqu'à présent, les

mémoires Flash ont connu la progression la plus importante observée dans le domaine des mémoires non volatiles. Cependant, la réduction de la taille des cellules Flash traditionnelles est limitée par la fiabilité de leur épaisseur d'oxyde. C'est pourquoi, depuis une dizaine d'années, on assiste à l'émergence de nouveaux concepts dont celui du stockage discret qui a stimulé l'émergence des mémoires Flash à nodules de silicium ainsi que notre travail. Nous commençons le quatrième chapitre par une présentation de la structure, puis, à partir des modèles développés dans le chapitre III, nous proposons une modélisation de la charge localisée dans les nodules proches du drain. Nous détaillons, ensuite, le modèle développé pour simuler l'opération d'écriture de ces mémoires. Enfin, la dernière partie de ce chapitre est consacrée à la caractérisation électrique de structures à nodules de silicium fabriquées par STMicroelectronics Catagne pour valider notre approche.

Une conclusion générale permet de faire la synthèse des résultats obtenus et de décrire les perspectives de ce travail de thèse.

Chapitre I

Rappels sur les structures MOS

&

les dispositifs mémoires

Chapitre I. Rappels sur les structures MOS et les dispositifs mémoires	15
I.1. Introduction	15
I.2. La capacité MOS	15
I.2.1. La structure	15
I.2.2. Principe et régimes de fonctionnement	16
I.2.3. Modélisation de la capacité MOS	17
I.2.3.1. Les équations de bases	17
I.2.3.2. La charge du semiconducteur Q_{sc}	18
I.2.3.3. La charge de la zone désertée Q_D	19
I.2.3.4. La charge de la zone d'inversion Q_n	20
I.2.3.5. La poly-désertion	20
I.2.3.6. Le courant tunnel Fowler-Nordheim	22
I.3. Le transistor MOS à enrichissement	25
I.3.1. La structure	25
I.3.2. Principe et régimes de fonctionnement	25
I.3.3. Modélisation du transistor MOS	27
I.3.3.1. Le modèle de Pao et Sah [Pao'66]	27
I.3.3.2. Le modèle en feuillet [Brews'78]	29
I.4. Les mémoires non volatiles	30
I.4.1. Généralités	30
I.4.2. Les mémoires Flash	33
I.4.2.1. La structure des mémoires Flash	33
I.4.2.2. Architecture des mémoires Flash	34
I.4.2.3. Principe de fonctionnement des mémoires Flash	35
I.4.3. Les mémoires à nano-cristaux	38
I.4.3.1. La structure des mémoires à nano-cristaux de silicium	38
I.4.3.2. Fonctionnement des mémoires à nano-cristaux de silicium	40
I.5. Conclusion	41

Chapitre I. Rappels sur les structures MOS et les dispositifs mémoires

I.1. Introduction

Afin de modéliser les dispositifs composés de transistors de petites dimensions ou des dispositifs plus complexes de type mémoire, il est nécessaire de rappeler le fonctionnement des dispositifs élémentaires tels que les structures MOS (Metal-oxide-semiconducteur) et de définir les paramètres qui serviront au cours de nos études. Nous commencerons ce chapitre par un rappel sur le fonctionnement et la modélisation de la capacité MOS, qui permet une approche simple des phénomènes constituant la physique du transistor MOS. Dans un deuxième temps, nous rappellerons les principales étapes de calculs, les hypothèses et les approximations qui mènent aux modèles couramment utilisés pour le transistor MOS. Enfin, une troisième partie constituera une introduction aux mémoires à grille flottante et plus particulièrement aux mémoires à piégeage discret.

Tout au long de ce document, nous considérerons le cas de composant à substrat de type P. On peut évidemment utiliser le même formalisme pour des dispositifs à substrat de type N (en changeant les N en P et en inversant les polarités).

I.2. La capacité MOS

I.2.1. La structure

Par définition un condensateur est constitué de deux électrodes conductrices séparées par un matériau isolant. Ainsi, on appelle « capacité MIS » la superposition de trois couches de matériaux : le métal ou poly-silicium dégénéré (appelé grille), l'isolant (SiO_2 , HfO_2 , Ta_2O_5 , Si_3N_4 ...), et le semiconducteur (Si, Ge...) de type N ou de type P (appelé bulk ou substrat) (cf. Fig. (I.1)).

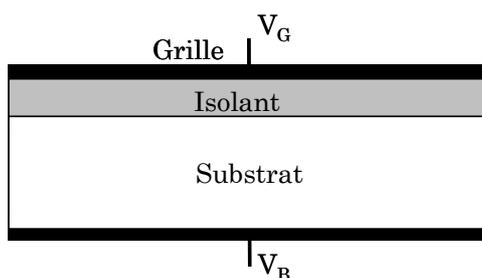


Figure I.1. Schéma en coupe d'une structure MIS.

La dénomination capacité MOS (pour Metal-Oxide-Semiconducteur) résulte de la nature de l'isolant qui est alors un oxyde.

I.2.2. Principe et régimes de fonctionnement

La polarisation de la capacité par une tension V_{GB} , entre la grille métallique et le substrat, implique l'apparition d'une charge Q_G dans la grille et d'une charge opposée Q_{SC} dans le semiconducteur. La variation de la tension V_{GB} modifie la valeur de ces charges, ce qui a pour conséquence les changements de régimes de fonctionnement de la capacité. La figure (I.2) présente les différents diagrammes de bandes du semiconducteur d'une capacité de type P en fonction de la tension V_{GB} .

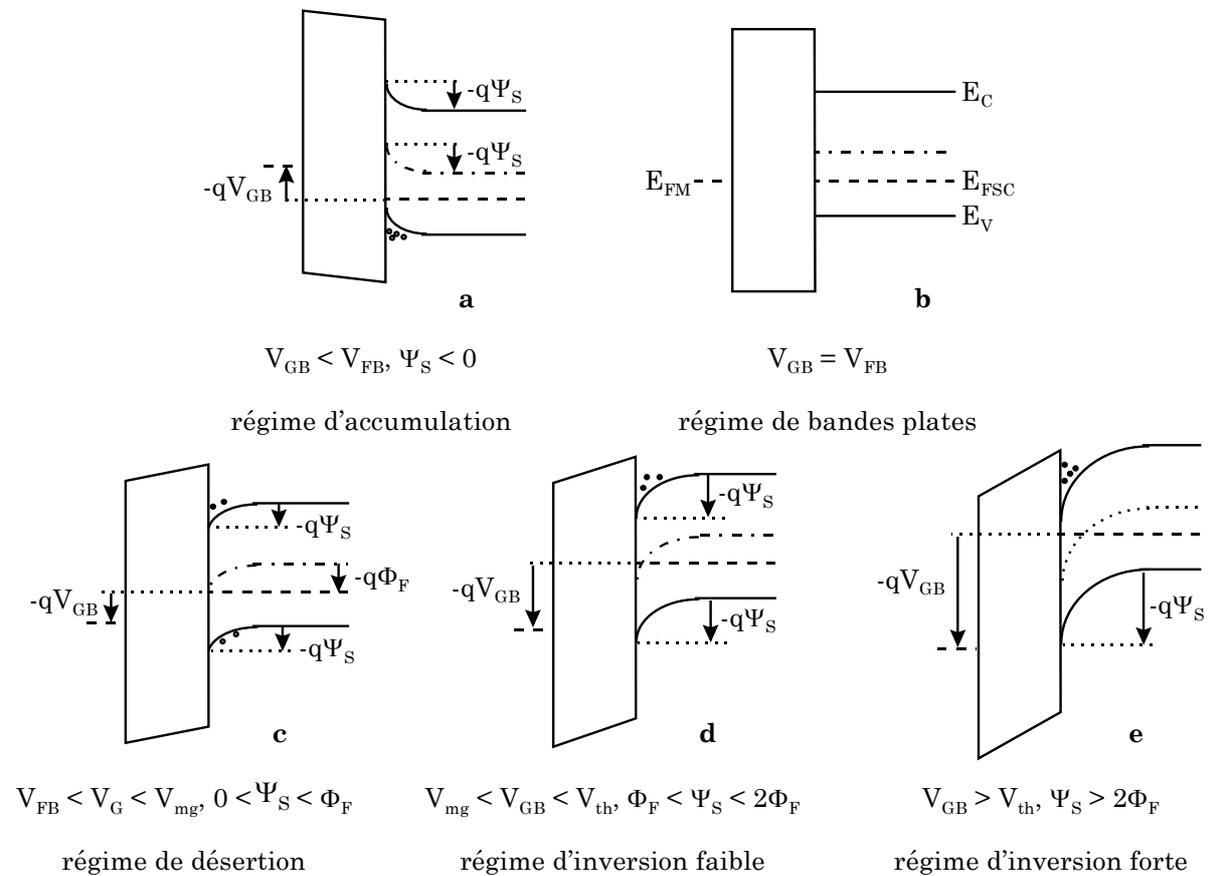


Figure I.2. Diagrammes de bandes représentant les différents régimes du semiconducteur en fonction du potentiel appliqué : le régime d'accumulation (a), le régime de bandes plates (b), le régime de désertion (c), le régime d'inversion faible (d) et le régime d'inversion forte (e).

Ainsi une capacité MOS présente cinq régimes de fonctionnement en fonction de la tension appliquée entre sa grille et son substrat.

I.2.3. Modélisation de la capacité MOS

I.2.3.1. Les équations de bases

La capacité totale d'une capacité MOS, de surface A_{eff} , est composée de la capacité d'oxyde, C_{ox} , en série avec la capacité dynamique du semiconducteur, C_{sc} :

$$\frac{1}{C} = \frac{1}{C_{\text{ox}}} + \frac{1}{C_{\text{sc}}} \quad (\text{I.1})$$

avec :

$$\begin{cases} C_{\text{ox}} = \frac{\epsilon_{\text{ox}} A_{\text{eff}}}{t_{\text{ox}}} \\ C_{\text{sc}} = \frac{dQ_G}{d\Psi_S} = - \frac{dQ_{\text{sc}}}{d\Psi_S} \end{cases} \quad (\text{I.2})$$

où Ψ_S est le potentiel de surface du substrat, et t_{ox} l'épaisseur de la couche d'oxyde. La charge au niveau de la grille, Q_G , est reliée à la tension aux bornes de l'isolant par la relation capacitive :

$$Q_G = C_{\text{ox}} V_{\text{ox}} \quad (\text{I.3})$$

où V_{ox} est la tension appliquée aux bornes de l'oxyde.

Notons que dans les expressions (I.2), les états d'interface et la déplétion de la grille ne sont pas pris en compte.

Pour une capacité MOS, deux équations doivent être respectées : la neutralité de la charge (I.4) et la conservation de l'équation aux potentiels (I.5):

$$Q_G + Q_{\text{ox}} + Q_{\text{sc}} = 0 \quad (\text{I.4})$$

$$V_{\text{GB}} = \Phi_{\text{MS}} + \Psi_S + V_{\text{ox}} \quad (\text{I.5})$$

où Φ_{MS} est la différence entre les travaux de sortie de la grille et du semiconducteur et Q_{ox} la charge fixe dans l'oxyde.

A partir des équations (I.3) à (I.5), l'équation aux potentiels (I.5) s'écrit :

$$V_{\text{GB}} = V_{\text{FB}} + \Psi_S - \frac{Q_{\text{sc}}}{C_{\text{ox}}} \quad (\text{I.6})$$

où la tension de bandes plates, V_{FB} , est définie par :

$$V_{\text{FB}} = \Phi_{\text{MS}} - \frac{Q_{\text{ox}}}{C_{\text{ox}}} \quad (\text{I.7})$$

Notons que dans le cas d'une capacité MOS réelle, les pièges d'interface, Q_{it} , ne sont plus négligeables et la relation donnant la tension de bandes plates, V_{FB} , doit être corrigée pour prendre en compte ces charges :

$$V_{FB} = \Phi_{MS} - \frac{Q_{ox}}{C_{ox}} - \frac{Q_{it}(\Psi_S = 0)}{C_{ox}} \quad (I.8)$$

I.2.3.2. La charge du semiconducteur Q_{sc}

Exprimons, à présent, la charge du semiconducteur Q_{sc} . Celle-ci est déterminée à partir de la résolution de l'équation de Poisson, puis de l'utilisation du théorème de Gauss. Considérons N_A la concentration en atomes accepteurs ionisés (et respectivement N_D la concentration en atomes donneurs ionisés), à une dimension, pour une capacité de type P dont la concentration N_A est uniforme, l'équation de Poisson se résout simplement. Cette équation de Poisson relie la courbure des bandes du semiconducteur, $\Psi(y)$, à la densité de charges, $\rho(y)$:

$$\frac{d^2\Psi(y)}{d^2y} = -\frac{\rho(y)}{\epsilon_{Si}} \quad (I.9)$$

où y correspond à l'axe vertical entre la surface du semiconducteur et le volume de celui-ci et $\epsilon_{Si} = \epsilon_{SC}\epsilon_0$ représente la permittivité du semiconducteur.

La densité de charges dépend à la fois de la densité en porteurs libres et de la charge fixe due aux impuretés dopantes ionisées du substrat :

$$\rho(y) = q[p(y) - n(y) + N_D - N_A] \quad (I.10)$$

où $p(y)$ et $n(y)$ sont respectivement les densités de trous et d'électrons dans le semiconducteur.

$$\begin{cases} n(y) = n_0 \exp(\beta\Psi(y)) \\ p(y) = p_0 \exp(-\beta\Psi(y)) \end{cases} \quad (I.11)$$

où p_0 et $n_0(y)$ sont respectivement les densités de trous et d'électrons libres dans le semiconducteur loin de l'interface et β représente le potentiel thermique (q/kT),.

De plus dans le volume du semiconducteur, la condition de neutralité doit être satisfaite, c'est-à-dire $\rho(y \rightarrow \infty) = p_0 - n_0 + N_D - N_A = 0$ ce qui implique que $p_0 - n_0 = N_A - N_D$. L'équation (I.10) devient alors :

$$\rho(y) = -q N_A \left\{ \left(\frac{n_i}{N_A} \right)^2 [\exp(\beta\Psi(y)) - 1] - [\exp(-\beta\Psi(y)) - 1] \right\} \quad (I.12)$$

avec pour un substrat de type P, $p_0 = N_A$ et $n_0 = (n_i)^2 / N_A$.

A partir de l'équations (I.12) et de l'équation de Poisson (I.9), on obtient le champ électrique, $\xi(y)$:

$$\frac{d\psi(y)}{dy} = -\xi(y) = \pm \sqrt{\frac{2kTN_A}{\epsilon_{Si}}} \left\{ \left(\frac{n_i}{N_A} \right)^2 [\exp(\beta\psi(y)) - \beta\psi(y) - 1] - 1 + \exp(-\beta\psi(y)) + \beta\psi(y) \right\}^{1/2} \quad (I.13)$$

En appliquant le théorème de Gauss au champ électrique à l'interface, $\iint_{(S)} \vec{\xi} \cdot d\vec{S} = \frac{Q_{int}}{\epsilon_{SC} \epsilon_0}$, la densité totale de charges dans le semiconducteur est obtenue :

$$Q_{SC} = \pm \sqrt{2kT\epsilon_{Si}N_A} \left\{ \left(\frac{n_i}{N_A} \right)^2 [\exp(\beta\psi_s) - \beta\psi_s - 1] - 1 + \exp(-\beta\psi_s) + \beta\psi_s \right\}^{1/2} \quad (I.14)$$

avec un signe + si $\Psi_s < 0$ et un signe - si $\Psi_s > 0$ et N_A considéré comme constant.

I.2.3.3. La charge de la zone désertée Q_D

Pour obtenir l'expression de la charge de la zone désertée Q_D , l'équation de Poisson est résolue en omettant le terme ayant pour origine les électrons de la couche d'inversion (quantité n). La densité de charges s'écrit donc à présent :

$$\rho = q [p_0 \exp(-\beta\Psi(y)) + n_0 - p_0] = qp_0 \left[\exp(-\beta\Psi(y)) - 1 + \frac{n_0}{p_0} \right] \quad (I.15)$$

En reportant l'équation (I.15) dans l'équation de Poisson (I.9), il vient :

$$\frac{d^2\Psi}{dy^2} = -\frac{qp_0}{\epsilon_{Si}} \left[\exp(-\beta\Psi(y)) - 1 + \frac{n_0}{p_0} \right] \quad (I.16)$$

En utilisant la même démarche mathématique que celle mise en œuvre pour le calcul de Q_{SC} , on obtient la charge de la zone désertée :

$$Q_D = \sqrt{2kT\epsilon_{Si}p_0} \left[\exp(-\beta\Psi_s) + \beta\Psi_s - \frac{n_0}{p_0} \beta\Psi_s - 1 \right]^{1/2} \quad (I.17)$$

Notons que puisque le substrat est de type P, la zone désertée dans le semiconducteur apparaît uniquement pour $\Psi_s > 0$, c'est pourquoi seule la racine positive de l'équation est considérée. Puisque Ψ_s est positif, il est possible de simplifier l'équation (I.17) en remarquant que :

$$\begin{cases} 1 - \frac{n_0}{p_0} = 1 - \frac{n_i^2}{N_A} \approx 1 \\ \exp(-\beta\Psi_s) \ll -\beta\Psi_s \end{cases} \quad (I.18)$$

La charge de la zone désertée s'exprime alors comme suit :

$$Q_D = -\sqrt{2kT\epsilon_{Si}p_0} [\beta\Psi_S - 1]^{1/2} \quad (I.19)$$

I.2.3.4. La charge de la zone d'inversion Q_n

La charge d'inversion Q_n est définie comme la différence entre la charge du semiconducteur et la charge de la zone désertée :

$$Q_n = Q_{SC} - Q_D \quad (I.20)$$

En faible inversion, puisque $\Psi_S + V_{BS} - 2\Phi_F < 0$, alors $\exp(\beta(\Psi_S - V + V_{BS} - 2\Phi_F)) \ll \beta\Psi_S - 1$ du moins tant que $\Psi_S + V_{BS} \ll 2\Phi_F - kT/q$. Ainsi en développant Q_{SC} au premier ordre, il vient :

$$Q_{SC} \approx -\sqrt{2kT\epsilon_{Si}p_0} \left[1 + \frac{\exp(\beta(\Psi_S - V + V_{BS} - 2\Phi_F))}{2(\beta\Psi_S - 1)} \right] (\beta\Psi_S - 1)^{1/2} \quad (I.21)$$

A partir des relations simplifiées (I.19) et (I.21), on obtient une relation simplifiée de la charge Q_n :

$$Q_n = -\frac{1}{2} \sqrt{\frac{2kT\epsilon_{Si}N_A}{\beta\Psi_S - 1}} \exp[\beta(\Psi_S - V + V_{BS} - 2\Phi_F)] \quad (I.22)$$

On peut également exprimer la charge d'inversion du canal en fonction du potentiel appliqué sur la grille par :

$$Q_n = Q_{SC} - Q_D = C_{ox} \left(V_{FB} - V_{GS} + V_{BS} + \Psi_S + \frac{\sqrt{2kT\epsilon_{Si}N_A}}{C_{ox}} (\beta\Psi_S - 1)^{1/2} \right) \quad (I.23)$$

où $\beta = kT/q$ et N_A est un dopage uniforme du substrat.

I.2.3.5. La poly-désertion

Jusqu'à présent, les capacités modélisées possédaient une grille métallique (ou en poly-silicium dégénéré) ; cependant il existe des capacités dont la grille est constituée de semiconducteur non dégénéré : les capacités SOS (pour Semiconducteur-Oxide-Semiconducteur). Comme le montre la figure (I.3.a), la non dégénérescence de la grille induit une courbure des bandes d'énergie du poly-silicium près de son interface avec l'isolant. Cette courbure varie avec la polarisation de grille rendant ainsi possible l'existence des différents régimes d'un semiconducteur : accumulation, désertion, inversion faible et inversion forte. Cependant en raison des faibles niveaux de dopage de

la grille (mais encore relativement forts par rapport à ceux du substrat), il est plus probable de trouver les régimes d'accumulation et de désertion, ce dernier correspondant à ce que l'on nomme usuellement la poly-désertion (ou poly-déplétion).

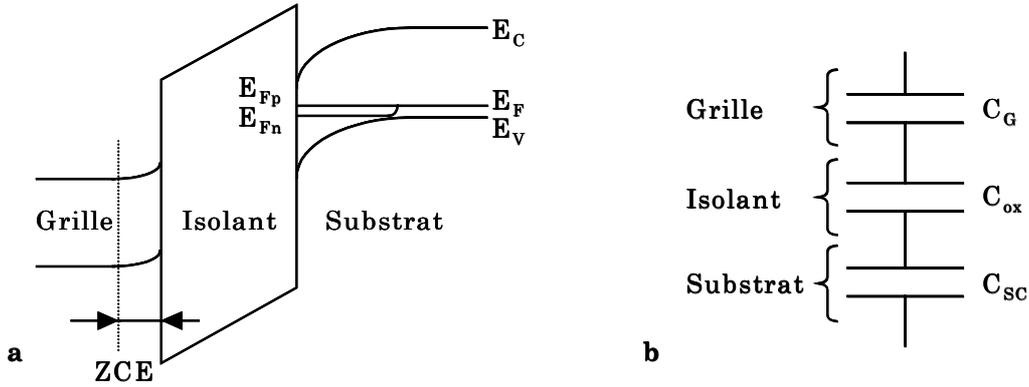


Figure I.3. Courbures des bandes d'énergie de la structure MOS dans le cas d'une dégénérescence non complète du poly-silicium de grille (a) et schéma électrique capacitif équivalent (b).

D'un point de vue capacitif, ce phénomène parasite s'assimile à l'apparition d'une capacité, C_G , en série avec la capacité MOS (avec un vrai métal de grille) comme l'indique la figure (I.3.b). Cela conduit à une extraction imprécise de l'épaisseur de l'oxyde de grille à partir des courbes C-V, puisque la chute de la capacité du dispositif peut être exprimée comme une augmentation de l'épaisseur de l'oxyde de grille [Huang'93].

Considérons une capacité MOS dont le substrat est de type P et la grille est en poly-silicium de type N⁺. En tenant compte du potentiel de surface du poly-silicium Ψ_{SG} , l'équation aux potentiels (I.6) devient :

$$V_{GB} = V_{FB} + \Psi_S - \Psi_{SG} - \frac{Q_{SC}(\Psi_S)}{C_{ox}} \quad (I.24)$$

Les expressions des charges en fonction des potentiels de surface sont :

$$Q_{SC} = \pm \sqrt{2kT\epsilon_{Si}p_0} \left[\frac{n_0}{p_0} (\exp(\beta\Psi_S) - \beta\Psi_S - 1) - 1 + \exp(-\beta\Psi_S) + \beta\Psi_S \right]^{1/2} \quad (I.25)$$

$$Q_G = \pm \sqrt{2kT\epsilon_{Si}n_{G0}} \left[\exp(\beta\Psi_{SG}) - \beta\Psi_{SG} - 1 + \frac{p_{G0}}{n_{G0}} (\exp(-\beta\Psi_{SG}) + \beta\Psi_{SG} - 1) \right]^{1/2} \quad (I.26)$$

avec un signe - lorsque le potentiel de surface considéré (Ψ_S ou Ψ_{SG}) est positif et un signe + lorsqu'il est négatif. n_{G0} et p_{G0} sont les densités en porteurs majoritaires et minoritaires de la grille loin de l'interface.

I.2.3.6. Le courant tunnel Fowler-Nordheim

L'effet tunnel est un mécanisme quantique qui permet à un électron de traverser une barrière énergétique. Le mécanisme de conduction Fowler-Nordheim (FN) a été expliqué pour la première fois par Fowler et Nordheim en 1928 [Fowler-Nordheim'28] dans le cas d'émission d'électrons d'un métal dans le vide. Dans ces conditions, la barrière énergétique vue par les électrons est de forme triangulaire et les électrons peuvent la traverser par effet tunnel en se déplaçant de la bande de conduction de la cathode vers la bande de conduction de l'isolant. Cette conduction apparaît pour des structures soumises à de forts champs électriques. Il faut également préciser que les oscillations observables sur la caractéristique I-V d'une structure MIS sont dues à des effets d'interférences et de réflexions des charges aux frontières de l'isolant. La présence de charges dans l'isolant peut limiter le passage par effet tunnel dans l'isolant de la même façon qu'une zone de déplétion.

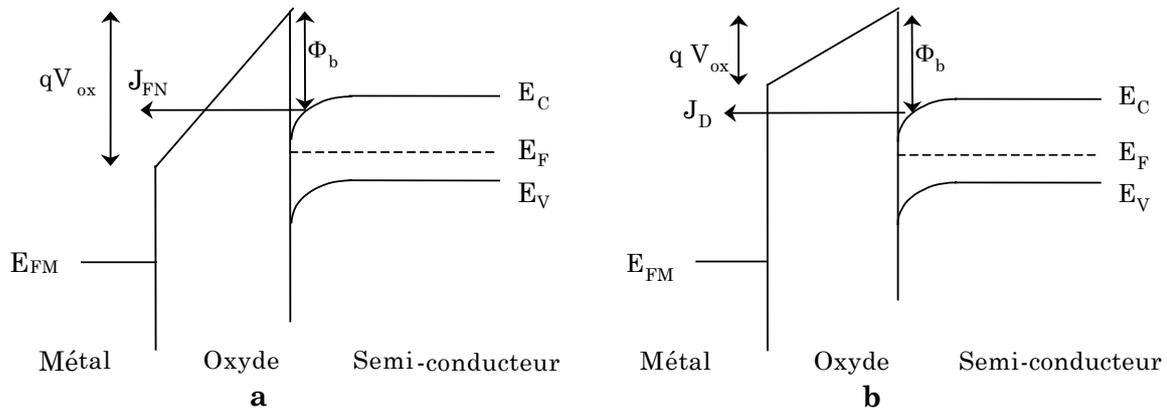


Figure I.4. Diagramme de bandes d'une structure MOS de type P en inversion dans le cas d'un courant tunnel Fowler-Nordheim (a) ou d'un courant tunnel direct (b).

La figure (I.4) met en évidence les deux types de transitions qui apparaissent selon la valeur de la courbure de bande de l'isolant par rapport à la hauteur de barrière, Φ_b , que les électrons voient à l'interface Si / Isolant :

- **La transition Fowler-Nordheim** pour $qV_{ox} > \Phi_b$, (figure (I.4.a)). Le champ électrique appliqué est suffisamment intense pour diminuer la largeur effective de la barrière à traverser. L'électron se retrouve alors dans la bande de conduction de l'isolant, puis il est entraîné vers l'électrode métallique.
- **La transition tunnel directe** pour $qV_{ox} < \Phi_b$, (figure (I.4.b)). Dans ce cas, le courant tunnel est dû aux électrons du semiconducteur qui traverse l'oxyde pour atteindre le métal.

La distance tunnel, d_{tun} , ainsi parcourue dépend de la hauteur de barrière Φ_b que voient les électrons à l'interface Si/Isolant et du champ électrique ξ_{ox} interne au diélectrique :

$$d_{\text{tun}} = \frac{\Phi_b}{\xi_{\text{ox}}} \quad (\text{I.27})$$

où

$$\xi_{\text{ox}} = \frac{V_G - \Phi_{\text{MS}} - \Psi_s}{t_{\text{ox}}} \quad (\text{I.28})$$

Dans le cas d'une structure MOS, en prenant le niveau de Fermi, E_F , comme référence des énergies le courant FN, $I_{\text{FN}}(\xi_{\text{ox}}, T)$, s'exprime de la manière suivante [O'Dweyer'73] :

$$I_{\text{FN}}(\xi_{\text{ox}}, T) = \frac{4q\pi m_M^* kT}{h^3} \int_{-\infty}^{\Phi_b} \ln \left[1 + \exp\left(\frac{-E}{kT}\right) \right] T(E) dE \quad (\text{I.29})$$

où m_M^* est la masse effective de l'électron dans le métal et $T(E)$ est le coefficient de transmission des électrons d'énergie E , à travers la barrière énergétique triangulaire.

En considérant les électrons comme un gaz à 3 dimensions, obéissant à une distribution en énergie de Maxwell-Boltzmann lorsque la longueur d'onde est négligeable devant l'épaisseur du diélectrique, alors la transparence des électrons, $T(E)$, est calculée à partir de l'approximation de Wentzel-Kramers-Brillouin (WKB) [Fromhold'81]:

$$T(E) = \exp \left[-\frac{4\sqrt{2m_{\text{ox}}}}{3q\hbar\xi_{\text{ox}}} (\Phi_b - E)^{3/2} \right] \quad (\text{I.30})$$

Les densités de courants tunnels direct et Fowler-Nordheim ont alors pour expression [Depas'95] :

$$J_{\text{DT}} = \frac{q^3}{16\pi^2 \hbar \Phi_b} \frac{\xi_{\text{ox}}^2}{\left[1 - \left(\frac{\Phi_b - qV_{\text{ox}}}{\Phi_b} \right)^{1/2} \right]^2} \exp \left[-\frac{4}{3} \frac{(2m_{\text{ox}})^{1/2}}{q\hbar} \frac{\Phi_b^{3/2} - (\Phi_b - qV_{\text{ox}})^{3/2}}{\xi_{\text{ox}}} \right] \quad (\text{I.31})$$

$$J_{\text{FN}} = \frac{q^3}{16\pi^2 \hbar \Phi_b} \xi_{\text{ox}}^2 \exp \left[-\frac{4}{3} \frac{(2m_{\text{ox}})^{1/2}}{q\hbar} \frac{\Phi_b^{3/2}}{\xi_{\text{ox}}} \right] \quad (\text{I.32})$$

où m_{ox} (exprimée en kg) est la masse effective des électrons dans l'isolant.

La figure (I.5) présente l'évolution de la densité de courant tunnel en fonction du champ électrique dans l'isolant pour les deux types de courants (direct et Fowler-Nordheim).

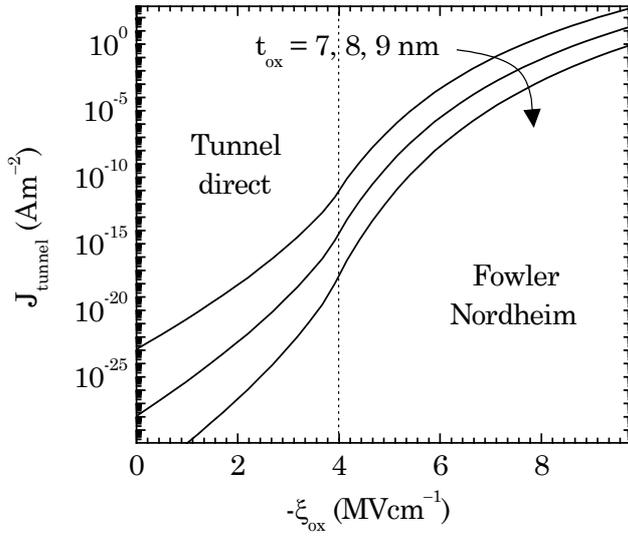


Figure I.5. Densité de courant qui traverse une capacité MOS à substrat N^+ pour des tensions de grille positives en fonction de l'épaisseur d'isolant.

Pour des températures proches de zéro, l'expression du courant FN d'une capacité MOS se simplifie sous la forme [Lenzlinger'69] :

$$I_{\text{FN}}(\xi_{\text{ox}}) = A_{\text{eff}} A \xi_{\text{ox}}^2 \exp\left(-\frac{B}{\xi_{\text{ox}}}\right) \quad (\text{I.33})$$

où A_{eff} correspond à la surface de la capacité et les coefficients Fowler-Nordheim (FN), A et B , dépendent principalement de la hauteur de barrière à l'interface oxyde/semiconducteur et de la masse effective des électrons :

$$A = \frac{q^3}{16\pi^2 \hbar \Phi_b} \frac{m_e}{m_{\text{ox}}} \quad (\text{I.34})$$

$$B = \frac{4}{3} \frac{\sqrt{2m_{\text{ox}}}}{q\hbar} \Phi_b^{3/2} \quad (\text{I.35})$$

où m_{ox} est la masse effective de l'électron dans l'oxyde (en général on prend $m_{\text{ox}} \approx 0.5 m_0$).

Ces paramètres FN sont donc sensibles à la nature des électrodes et à la qualité de l'isolant notamment en terme de charges piégées. Ils se déduisent aisément à partir d'une caractéristique I-V en traçant la courbe $\ln(I_{\text{FN}} / A_{\text{eff}} \xi_{\text{ox}}^2)$ en fonction de $1/\xi_{\text{ox}}$.

I.3. Le transistor MOS à enrichissement

I.3.1. La structure

Avant de présenter les équations permettant le calcul du courant de drain du transistor MOS, il est nécessaire de définir les différentes notations utilisées [Masson'99].

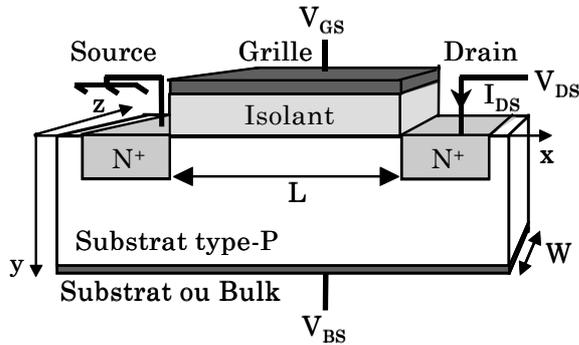


Figure I.6. Vue schématique du transistor MOS de type N [Masson'99].

Le transistor MOS (ou MOSFET pour transistor Métal-Oxyde-Semiconducteur à effet de champ) à canal N est un dispositif quadripolaire constitué d'une électrode de grille (G), de source (S), de drain (D) et de substrat (B) (cf. Fig. (I.6)). La longueur du transistor, notée L , correspond à la longueur de sa grille et sa largeur est notée W . La structure du transistor étant identique selon sa largeur, on le représente communément dans le plan (x,y) . Nous considérerons par la suite un transistor à canal surfacique, c'est-à-dire dont la conduction est assurée par les porteurs minoritaires du substrat (électrons dans le cas d'un NMOSFET), à l'interface entre le diélectrique de grille et le substrat.

Notons que le MOSFET possède deux électrodes supplémentaires par rapport à la capacité MOS, qui sont constituées de deux caissons dopés N^+ pour un NMOS (réservoirs à électrons). Ainsi, de nombreuses propriétés du transistor MOS découlent de celles de la capacité MOS.

I.3.2. Principe et régimes de fonctionnement

Le principe de fonctionnement du transistor MOS (ou MOSFET) repose sur la modulation d'une densité de porteurs d'une zone semi-conductrice par un champ électrique qui lui est perpendiculaire. Ce champ électrique est appliqué par l'électrode de commande (la grille) à travers un isolant (diélectrique de grille). Les porteurs créés sont des charges mobiles : électrons dans le cas d'un transistor NMOS, trous dans le cas d'un transistor PMOS. Lorsque la tension appliquée sur la grille est supérieure à une tension seuil appelée tension de seuil, notée V_T , ces charges mobiles constituent un canal de

conduction entre la source et le drain. Lorsqu'une différence de potentiel, V_{DS} , est appliquée entre la source et le drain, les porteurs affluant (côté source, de façon conventionnelle) sont collectés par le drain sous la forme d'un courant. Ainsi, de façon macroscopique, le transistor MOS se comporte comme un dispositif régulant un courant entre deux électrodes par une commande en tension.

Rappelons qu'il existe trois valeurs particulières de la tension V_{GS} :

- V_{FB} : tension V_{GS} à appliquer pour que $\Psi_S = 0$ au niveau de la source (aussi appelée tension de bandes plates).
- V_{mg} : tension V_{GS} à appliquer pour que $\Psi_S = \Phi_F$ au niveau de la source.
- V_{th} : tension V_{GS} à appliquer pour que $\Psi_S = 2\Phi_F - \Phi_C(0)$ au niveau de la source.

Notons l'apparition de l'écart entre les quasi-niveaux de Fermi, Φ_C , qui dépendent de la tension V_{DS} . En effet, les zones de drain et de source imposent un écart entre les quasi-niveaux de Fermi des électrons, E_{Fn} , et des trous, E_{Fp} , aux bornes du canal. Cet écart, Φ_C , est égal à $(E_{Fp} - E_{Fn})/q$ et prend pour valeur à la source $\Phi_C(0) = V_{SB}$ et au drain $\Phi_C(L) = V_{DB} - V_{SB}$. Le substrat étant de type P, le quasi-niveau de Fermi des trous, E_{Fp} est égal au niveau de Fermi dans le volume du semiconducteur, E_F , et ne varie pas le long du canal : seul le niveau énergétique E_{Fn} varie (cf. Fig. (I.7)).

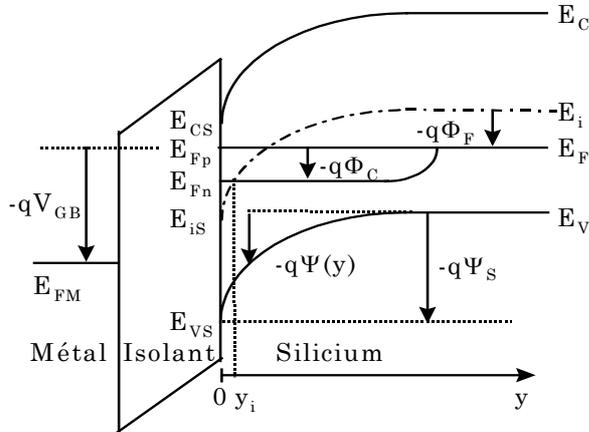


Figure I.7. Diagramme de bandes du transistor MOS en régime d'inversion forte suivant l'axe y en un point quelconque du canal [Masson'99].

La courbure des bandes d'énergie du semiconducteur est notée $\Psi(y)$ et la courbure totale correspond au potentiel de surface, Ψ_S . Le choix du sens des flèches a pour origine la tension que l'on applique entre la grille et le substrat. Cela revient à faire la différence entre les niveaux de Fermi du métal et du semiconducteur.

Le potentiel de volume du semiconducteur Φ_F a pour expression [Sze'81] :

$$\Phi_F = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right) = \frac{1}{\beta} \ln\left(\frac{N_A}{n_i}\right) = -\frac{1}{q}(E_F - E_i) \quad (I.36)$$

I.3.3. Modélisation du transistor MOS

La connaissance des équations de modélisation de la conduction dans le transistor MOS est nécessaire pour l'extraction des paramètres de fonctionnement comme la tension de seuil V_T , la mobilité à faible champ μ_0 ou la transconductance du canal g_m . Parmi les modèles décrivant les propriétés de conduction d'un transistor MOS, les modèles de Pao et Sah [Pao'66] et en feuillet [Brews'78], basés sur le principe de dérive-diffusion, permettent la continuité du courant I_{DS} entre les différents régimes de fonctionnement du transistor MOS (c.a.d. les régimes d'inversion faible, d'inversion forte, ohmique, quadratique et saturé). Ainsi, nous avons choisi d'utiliser ces deux modèles, qui reposent sur le calcul du potentiel de surface (le long du canal ou à ses extrémités).

I.3.3.1. Le modèle de Pao et Sah [Pao'66]

Le modèle de Pao et Sah [Pao'66] décrit le courant de drain en distinguant ou non les termes de conduction et de diffusion :

$$I_{DS} = -\frac{W}{L} \mu_0 \int_{\Psi_s(0)}^{\Psi_s(L)} Q_n d\Psi + \frac{W}{L} \mu_0 \frac{kT}{q} [Q_n(L) - Q_n(0)] = -\frac{W}{L} \mu_0 \int_{\Phi_C(0)}^{\Phi_C(L)} Q_n d\Phi_C \quad (I.37)$$

où Q_n représente la charge de la zone d'inversion (par unité de surface).

D'un point de vue pratique, le calcul du courant de conduction nécessite la connaissance, à V_{GB} donnée, de la variation de la charge d'inversion et du potentiel de surface le long du canal. La relation aux potentiels liant les potentiels aux charges s'écrit :

$$V_{GB} = V_{FB} + \Psi_S - \frac{Q_{SC}(\Psi_S, \Phi_C)}{C_{ox}} - \frac{Q_{it}}{C_{ox}} \quad (I.38)$$

Afin de déterminer Ψ_S le long du canal, à l'aide de l'équation (I.38), on considère par exemple une vingtaine de valeurs de Φ_C entre la source et le drain (c.a.d. $[-V_{BS}, V_{DS} - V_{BS}]$). La charge Q_n est alors calculée pour chaque Ψ_S , puis l'intégrale de l'équation (I.37) est évaluée par la méthode des rectangles ou des trapèzes. Notons que la détermination du potentiel de surface en un point quelconque du canal nécessite la connaissance des charges dans la structure MOS. Tandis que la charge Q_D (charge hors électrons) reste identique à celle d'une capacité MOS (I.17), l'équation de la charge du semiconducteur Q_{SC} (I.14), doit être légèrement modifiée pour tenir compte des quasi-niveaux de Fermi :

$$Q_{SC} = \pm \sqrt{2kT\epsilon_S i} N_A \left\{ \left(\frac{n_i}{N_A} \right)^2 [\exp(\beta\Psi_S - \beta\Phi_c) - \beta\Psi_S - \exp(-\beta\Phi_c)] - 1 + \exp(\beta\Psi_S) + \beta\Psi_S \right\}^{1/2} \quad (\text{I.39})$$

avec un signe + si $\Psi_S < 0$ et un signe - si $\Psi_S > 0$ et N_A considéré comme constant.

Quelle que soit la valeur de la tension de drain, les simulations du courant I_{DS} présentées aux figures (I.8.a) et (I.8.b) montrent qu'en régime d'inversion faible, le courant I_{DS} résulte d'un phénomène de diffusion de porteurs dans le canal, tandis qu'en régime d'inversion forte le courant de drain est presque égal au courant de conduction. De plus, la représentation en échelle semi-logarithmique de la courbe I_{DS} en fonction de V_{GS} , est linéaire en régime d'inversion faible. Cette portion de droite porte le nom de pente sous le seuil.

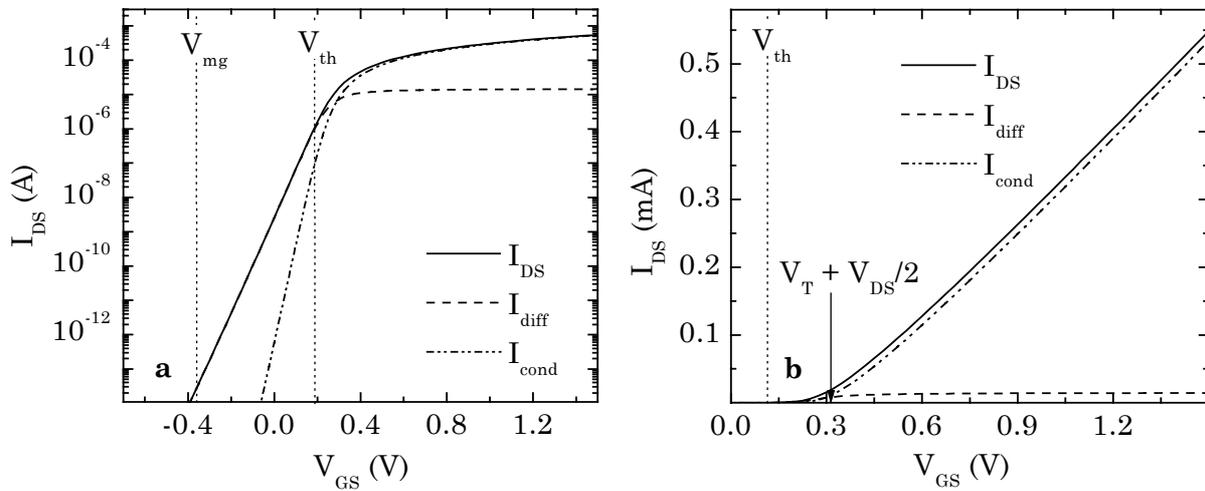


Figure I.8. Evolution des courants de conduction et de diffusion ainsi que du courant total en fonction du potentiel de grille en échelle semi-logarithmique (a) ou linéaire (b). Les paramètres de la simulation sont : $V_{DS} = 0.05$ V, $V_{BS} = 0$ V, $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $\mu_0 = 300 \text{ Vs}^{-1}\text{cm}^{-2}$, $L = 0.5 \text{ }\mu\text{m}$, $W = 1 \text{ }\mu\text{m}$, $V_{FB} = -1$ V, $V_{mg} = -0.36$ V et $V_{th} = 0.18$ V [Masson'99].

De plus, comme le montre la figure (I.8.b), la notion de tension de seuil du transistor MOS, V_T , est différente de celle notée V_{th} . Usuellement, on considère que la tension V_T correspond au déblocage du transistor et donc à la création de la charge d'inversion (ce qui est une approximation). Elle se situe à l'intersection de la partie quasi-linéaire de la courbe avec l'axe V_{GS} . Cette notion de tension de seuil est indispensable pour des mémoires non volatiles puisque c'est elle qui représente l'information stockée. Ainsi, le calcul du courant nécessite le découpage du canal en petits éléments dont on connaît la charge $Q_n(x)$ mais pas la localisation x puisque le découpage a été fait selon Φ_c le long du canal. En supposant que, pour une polarisation donnée, le courant de drain est à flux

conservatif (c.a.d. que le courant est identique en tout point du canal), il est possible de déterminer la localisation (en x) de la charge Q_n et par suite celle du potentiel de surface $\Psi_s(x)$ et de l'écart entre les quasi-niveaux de Fermi $\Phi_C(x)$. Soit x , la distance à partir de la source, le courant de drain peut s'écrire :

$$I_{DS} = -\frac{W}{x} \mu_0 \int_{\Phi_C(0)}^{\Phi_C(x)} Q_n d\Phi_C \quad (\text{I.40})$$

En divisant l'équation (I.37) par l'équation (I.40), on aboutit à l'équation (I.41) :

$$x = L \frac{\int_{\Phi_C(0)}^{\Phi_C(x)} Q_n d\Phi_C}{\int_{\Phi_C(0)}^{\Phi_C(L)} Q_n d\Phi_C} \quad (\text{I.41})$$

Cette approche de type Pao et Sah présente l'avantage d'obtenir une localisation des différentes grandeurs physiques le long du canal (Q_n , Φ_C , Ψ_s). Elle autorise aussi la prise en compte d'un grand nombre d'effets parasites tels que : la présence de pièges dans l'isolant ou à son interface, la poly-déplétion de la grille, etc... La charge Q_n peut aussi être obtenue pour des cas particuliers : comme pour les effets quantiques [Masson'02] ou un dopage (vertical) non uniforme du substrat, comme nous l'expliquerons au chapitre II. Cependant, en raison du découpage en quasi-niveaux de Fermi, le calcul du courant est relativement long et ne prend pas en compte les effets 2D le long du canal. Enfin, la précision du calcul dépend du découpage du canal en quasi-niveaux de Fermi le long du canal notamment en régime de saturation.

I.3.3.2. Le modèle en feuillet [Brews'78]

En 1978, Brews donne également une expression du courant valable de l'inversion faible à l'inversion forte avant saturation en décrivant le courant de drain I_{DS} comme la somme de deux contributions : le courant de conduction et celui de diffusion [Brews'78]. Ce modèle ne nécessite pas la détermination de la charge d'inversion le long du canal puisque le calcul se fait aux frontières du canal (c.a.d. le drain et la source). La résolution de l'équation (I.37) nécessite la connaissance de la primitive de Q_n par rapport à Ψ_s qui peut être calculée à partir de l'expression de Q_n donnée par l'équation (I.23). Ainsi les équations (I.37), et (I.23) nous amènent à écrire l'équation du courant sous la forme suivante :

$$I_{DS} = \frac{W}{L} \mu_0 C_{ox} \left[(V_{GS} - V_{FB} - V_{BS}) \left(\Psi_S - \frac{1}{\beta} \right) - \frac{\Psi_S}{\beta} - \frac{\gamma}{\beta} (\beta \Psi_S - 1)^{1/2} - \frac{1}{2} \Psi_S^2 - \frac{2}{3} \gamma (\beta \Psi_S - 1)^{3/2} \right]_{\Psi_S(0)}^{\Psi_S(L)} \quad (I.42)$$

où

$$\gamma = \frac{\sqrt{2kT\epsilon_{Si}p_0}}{C_{ox}} \quad (I.43)$$

L'intégration de la charge d'inversion le long du canal aboutit à :

$$I_{DS} = \frac{W}{L} \mu_0 C_{ox} [F(L) - F(0)] \quad (I.44)$$

où la fonction F est donnée par :

$$F(x) = (V_{GS} - V_{FB} - V_{BS}) \left(\Psi_S(x) - \frac{1}{\beta} \right) + \frac{\Psi_S}{\beta} + \frac{\gamma}{\beta} (\beta \Psi_S(x) - 1)^{1/2} - \frac{1}{2} \Psi_S(x)^2 - \frac{2}{3} \gamma (\beta \Psi_S(x) - 1)^{3/2} \quad (I.45)$$

Ce modèle, décrit plus précisément dans le livre de Tsividis [Tsividis'99], nécessite la connaissance des potentiels de surface uniquement aux bornes du canal contrairement à l'approche de Pao et Sah. Ainsi le calcul est beaucoup plus rapide. Cependant, ce gain en temps se fait au détriment de la possibilité de prendre en compte la plupart des effets parasites. De plus, ce modèle est pseudo-2D, donc ne prend pas en compte les effets 2D. Comme dans l'approche de Pao et Sah, il est possible de déterminer l'évolution du potentiel de surface et l'écart entre les quasi-niveaux de Fermi le long du canal en considérant que le courant de drain est à flux conservatif avec l'équation suivante :

$$X = L \frac{F(x) - F(0)}{F(L) - F(0)} \quad (I.46)$$

Les modélisations des structures MOS que nous venons de rappeler sont nécessaires pour la modélisation des dispositifs mémoires, basés sur le potentiel des technologies MOS. Pour l'application de nos travaux, nous nous sommes intéressés aux mémoires non volatiles à stockage discret brièvement décrite au paragraphe (§ I.4.3).

I.4. Les mémoires non volatiles

I.4.1. Généralités

Par définition, la mémoire est la propriété de conserver et de restituer des informations. Cependant, en microélectronique, il existe deux moyens pour obtenir cette

propriété. La figure (I.9) donne une classification des principales mémoires MOS qui sont traditionnellement classées en deux grandes familles :

- Les mémoires vives, désignées par le sigle générique RAM (pour Random Access Memory), c'est-à-dire mémoires à accès aléatoire ; ce sont des mémoires dans lesquelles on peut, à tout moment, écrire ou lire des informations, et ce, tant que l'alimentation électrique est présente.
- les mémoires mortes ou ROM (Read-Only Memory) sont des mémoires qui ne peuvent être que lues à partir du moment où les informations y ont été écrites; en revanche, elles possèdent la propriété de garder l'information très longtemps (spécification typique : 10 ans), même en l'absence d'alimentation électrique.

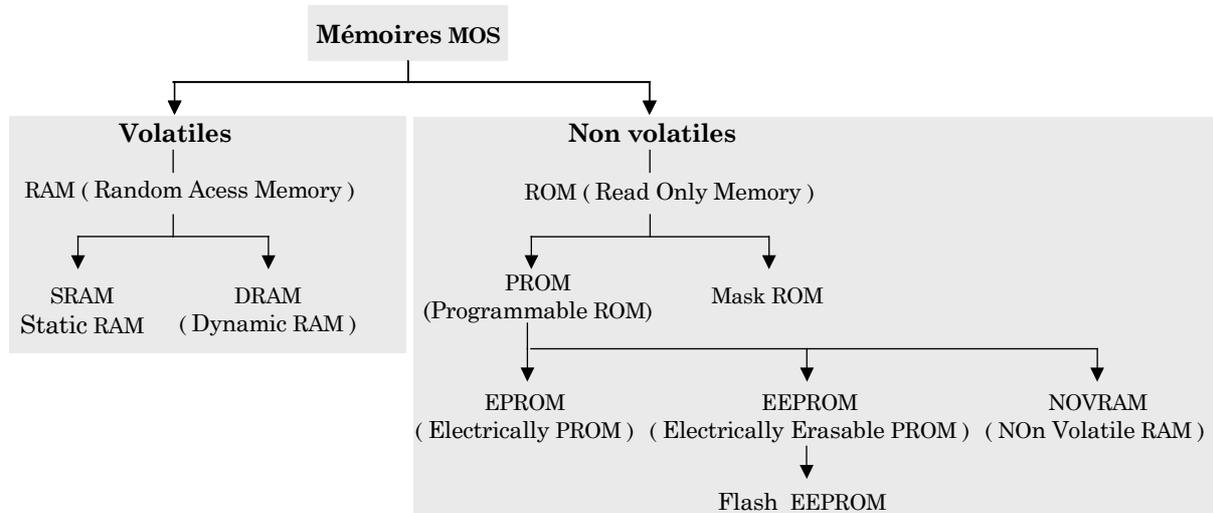


Figure I.9. Classification des principales mémoires MOS.

En 1967, D. Kahng et S.M. Sze [Kahng'67] présentaient la première mémoire MOS non volatile, composée d'un transistor MOS dont la grille était remplacée par un empilement de couches conductrices et non conductrices. De nos jours, les mémoires non volatiles sont quotidiennement présentes dans notre vie avec les cartes bancaires, les téléphones mobiles, les décodeurs de télévision, les ordinateurs personnels, la gestion des moteurs automobiles et beaucoup d'autres applications nécessitant la sauvegarde de l'information de façon permanente même après rupture de l'alimentation. Dans la suite du manuscrit, nous nous intéresserons uniquement aux mémoires non volatiles.

Les mémoires ROM (Read Only Memory) sont destinées uniquement à être lues, et sont essentiellement utilisées pour les jeux vidéo. Elles sont programmées, soit lors de la fabrication (activation ou non d'un transistor par masquage), soit par l'utilisateur avec des structures à base de fusibles. Le fonctionnement d'une ROM est basé sur celui du point mémoire qui est généralement constitué d'un transistor NMOS (ayant une grille

flottante) adressé en lecture par une ligne de bit connectée au drain et une ligne de mot (word line) connectée à la grille (cf. Fig. (I.10)).

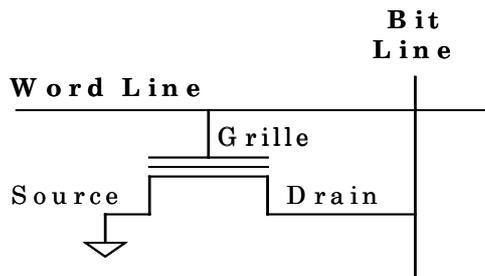


Figure I.10. Schématisation d'un point mémoire.

L'EPROM (Erasable Programmable Read Only Memory) permet d'effacer et de programmer les cellules. Elles sont largement utilisées en bureautique (dans les photocopieurs, les imprimantes lasers, où elles contiennent les différentes polices de caractères, dans les automates programmables, etc ...). L'écriture se fait par stockage d'électrons dans une grille isolée. L'opération d'effacement par rayons Ultra Violets (UV) des EPROMs reste néanmoins lourde à mettre en œuvre : elle suppose un démontage du boîtier de son support et un passage de 15 à 20 minutes sous rayons UV. De plus, les EPROMs utilisent des boîtiers coûteux à fenêtre de quartz pour permettre ce type d'effacement. Le principal problème de fiabilité de ce type de mémoire est la rétention de l'information stockée, car le nombre de cycles d'écriture / effacement reste faible.

Les EEPROMs (Electrically Erasable PROM), développées dans le milieu des années 1970, répondent au problème de l'effacement UV par un effacement bit par bit de type électrique qui évite de retirer le circuit du système électronique pour reprogrammer la mémoire. Les EEPROMs sont en partie dédiées aux applications militaires ou spatiales. Comme le montre la figure (I.11), les EEPROMs utilisent une surface équivalente à deux transistors par cellule mémoire [Yaron'82] : le premier est utilisé comme transistor de sélection et le second est l'élément de stockage. La cellule mémoire EEPROM est traditionnellement réalisée en technologie FLOTOX (« FLOting gate Thin OXide »). Le point critique est l'utilisation d'un oxyde de grille très mince qui sépare le drain de la grille flottante (faible rendement de fabrication) et une surface occupée importante. Sous l'effet d'un champ électrique intense de l'ordre de 10 MVcm^{-1} , des électrons passent par effet tunnel à travers cet oxyde mince, du drain vers la grille flottante ou inversement suivant le sens du champ électrique. Ainsi cette injection d'électrons fait varier la quantité de charges de la grille flottante ce qui modifie la tension de seuil du transistor.

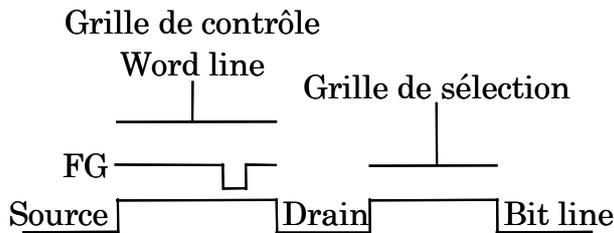


Figure I.11. Schéma équivalent de la cellule EEPROM composée du transistor d'état en série avec le transistor de sélection.

L'apparition des mémoires Flash EEPROMs est issue de la course aux réductions de dimensions. L'utilisation d'un seul transistor par cellule mémoire a permis un gain de place et de rapidité avec la possibilité de re-programmer les mémoires PROM et par conséquent un gain en terme de coût de production. Le terme Flash traduit le fait que les données d'un bloc entier sont effacées d'un seul coup. Actuellement, les mémoires Flash représentent la famille la plus importante des mémoires non-volatiles en raison de leur grande densité d'intégration, de leur rapidité d'écriture et de lecture [Pavan'97].

I.4.2. Les mémoires Flash

La première mémoire flash fut présentée en 1984 par Masuoka [Masuoka'84]. Décrivons la structure et le fonctionnement de ces mémoires Flash.

I.4.2.1. La structure des mémoires Flash

La mémoire Flash est constituée d'un transistor MOS dont la structure de la grille a légèrement été modifiée avec une grille de contrôle et une grille flottante (FG) emprisonnée dans l'isolant.

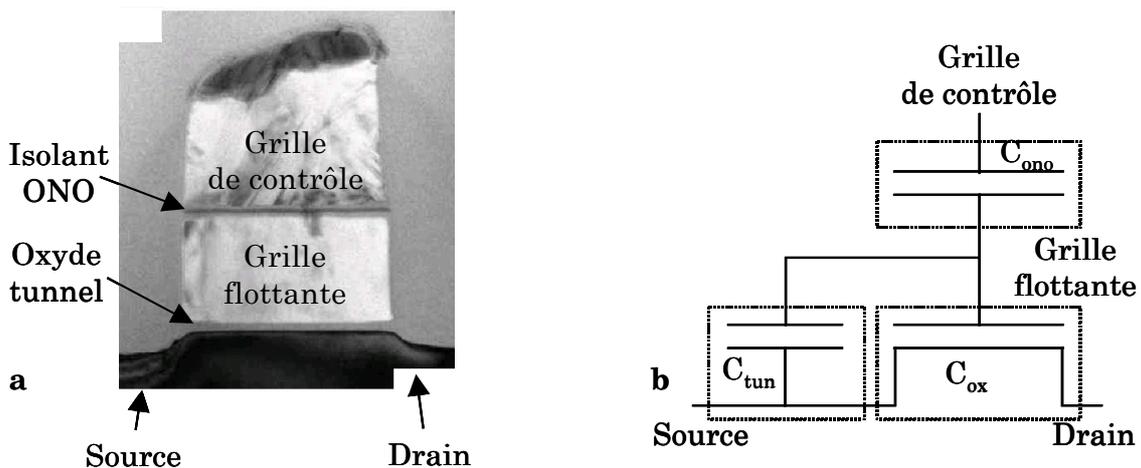


Figure I.12. Coupe SEM (Scanning Electron Microscopy) d'une mémoire de type Flash (a) et schéma électrique équivalent faisant apparaître les différentes capacités (b), [Laffont'03b].

Les figures (I.12.a) et (I.12.b) montrent une coupe SEM (Scanning Electron Microscopy) et le schéma électrique équivalent d'une mémoire Flash. On peut identifier les trois composants principaux que sont : le transistor MOS (avec sa capacité C_{ox}), la capacité inter-poly (C_{ono}) et la capacité de recouvrement de la source (C_{tun}).

I.4.2.2. Architecture des mémoires Flash

Les mémoires Flash peuvent être regroupées en une architecture de type NAND ou de type NOR [Cappelletti'99]. Quelle que soit l'architecture le plan mémoire est constitué d'une matrice de lignes (Word line) et de colonnes (bit line) dont l'intersection correspond à un point mémoire. La figure (I.13.a) présente l'architecture de type NOR. Durant les opérations de lecture, la cellule lue est adressée en polarisant sa word line positivement alors que les autres word lines sont connectées à la masse. Afin d'éviter toute perturbation de la bit line par des courants de fuites, les cellules non sélectionnées doivent avoir une tension de seuil équivalente positive. Le principal inconvénient de ces architectures NOR est leur faible densité d'intégration puisque tous les points mémoires ont leur drain connecté à la bit line et leur source à la ligne commune.

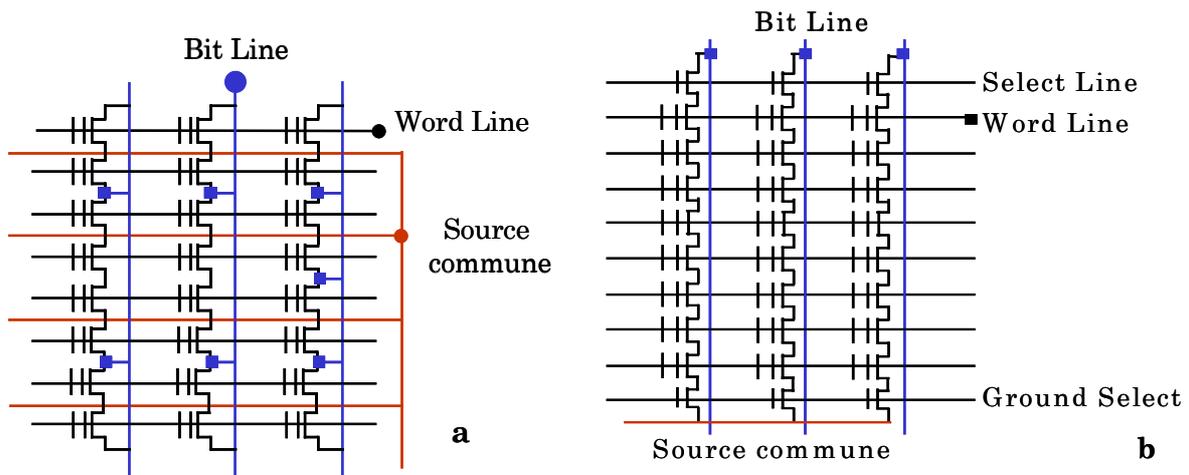


Figure I.13. Architecture de type NOR (a) et de type NAND (b).

Une meilleure densité est obtenue grâce à l'architecture NAND présentée dans la figure (I.13.b). Dans ce cas, les bit lines sont des lignes composées de points mémoires connectés en série. Deux transistors de sélection sont situés sur chaque ligne: le premier, qui sert à sélectionner la bit line, est commandé par le signal SL (« Select Line ») et le second, qui sert à relier les bit lines à la masse, est commandé par le signal GS (« Ground Select »). Afin de choisir une cellule de la ligne, sa word line doit être activée, ainsi que toutes les word lines commandant les autres cellules de la ligne. Pour lire l'état du point mémoire sélectionné, une tension de lecture assez faible est appliquée sur sa grille, alors

qu'une tension supérieure à la tension de seuil équivalente maximale est appliquée aux autres points mémoire. Ainsi la cellule sélectionnée impose le courant de la bit line à lire.

I.4.2.3. Principe de fonctionnement des mémoires Flash

Dans ce paragraphe, nous décrivons uniquement le fonctionnement d'un seul point mémoire. Comme pour tous les dispositifs MOS à grille flottante, le MOSFET fonctionne comme un interrupteur avec une modulation des électrons du canal par la grille de contrôle (GC pour Gate Control). La grille flottante, déconnectée des électrodes où sont appliquées les tensions, joue le rôle d'élément mémoire. Ainsi, la caractéristique $I_{DS}(V_{GC})$ d'une structure à grille flottante dépend de la charge stockée dans celle-ci, Q_{FG} , qui induit une variation de la tension de seuil entre deux valeurs distinctes (cf. Fig. (I.14)). Soit V_{T1} , la tension de seuil initiale du dispositif. L'état écrit de la mémoire résulte du stockage d'électrons dans la grille flottante. La tension de seuil du MOSFET augmente et atteint une valeur V_{T2} , le transistor est alors bloqué. Pour effacer la mémoire, les électrons sont évacués de la grille flottante et la tension de seuil retrouve sa valeur initiale, V_{T1} . Le transistor est alors passant. Notons que la différence, ΔV_T , entre les tensions de seuil de l'état écrit et de l'état effacé correspond à la fenêtre de programmation de la mémoire. L'état de la mémoire est déterminé par une mesure en courant du transistor en polarisant la grille de contrôle par une tension appartenant à la fenêtre de programmation de la mémoire.

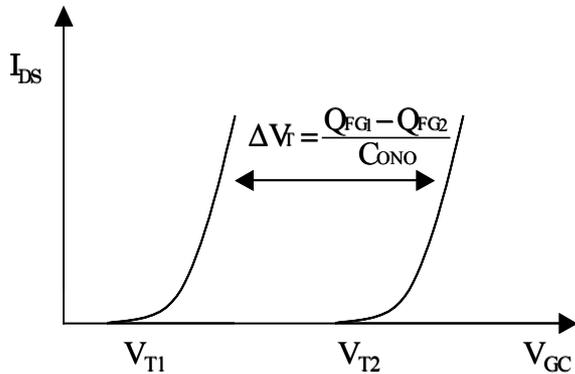


Figure I.14. Caractéristiques $I_{DS}(V_{GC})$ d'une structure à grille flottante pour deux charges différentes sur la grille flottante.

Pour une cellule flash, l'effacement est obtenu par injection d'électrons de la grille flottante vers la source (et le substrat) par le biais d'un courant Fowler-Nordheim (I_{FN}). L'écriture peut être obtenue par injection FN (cas des architectures NAND) ou par injection par porteurs chauds (cas des architectures NOR). Dans ce dernier cas, le transistor est polarisé en régime de saturation. Il existe alors à la jonction canal/drain polarisée en inverse, un champ électrique d'autant plus important que la longueur du canal diminue ($\xi = V / L$). Par conséquent, les électrons qui pénètrent dans la zone de

désertion sont accélérés par ce champ électrique très intense (forte courbure des bandes d'énergie). Ce phénomène est illustré sur la figure (I.15) par le repère (1). Certains électrons acquièrent alors suffisamment d'énergie cinétique (porteurs chauds) pour se comporter comme des particules ionisantes (collision avec le réseau cristallin) et générer des paires électrons-trous : c'est le phénomène d'ionisation par impact. On obtient donc deux électrons dans la bande de conduction et un trou dans la bande de valence. Les paires électrons-trous ainsi créées sont dissociées sous l'effet du champ électrique. Les trous peuvent être attirés par l'électrode de substrat et donner naissance à un important courant de substrat I_{SUB} (repère (2)). Ils peuvent également migrer vers la source et créer un abaissement de la barrière à la jonction source/canal. Il se produit alors une injection d'électrons supplémentaires de la source vers le canal. En toute rigueur, le courant de substrat (I_B) est la somme de I_{SUB} et du courant des jonctions source/substrat et drain/substrat ($I_B = I_{SUB} + I_{diode}$). Pour nos travaux, le courant inverse des diodes sera toujours négligé ($I_B \approx I_{SUB}$). La majeure partie des électrons générés, par ionisation par impact, s'additionne au courant du transistor I_{DS} pour donner le courant noté I_D (repère (3)). Nous en déduisons qu'en présence de porteurs chauds, le courant de drain n'est plus égal au courant de source qui lui, est toujours égal à I_{DS} ($I_S = I_{DS}$). Enfin, lorsque les électrons sont accélérés au niveau de la jonction canal/drain, une polarisation positive de la grille donne la possibilité à ceux qui ont acquis une énergie potentielle suffisante de franchir la barrière de potentiel de l'interface silicium/oxyde, ce qui correspond au courant de grille, I_G (repère (4)).

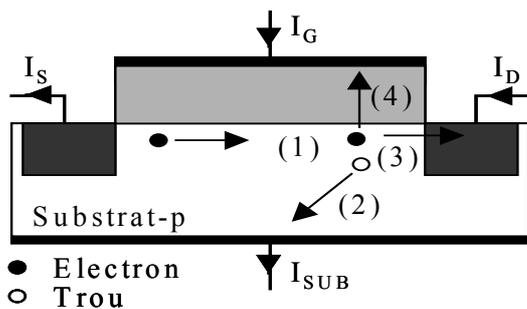


Figure I.15. Localisation de la génération des paires électron-trou due à la présence d'électrons chauds près du drain. Les trous sont collectés par la prise substrat (I_{SUB}). Une partie des électrons générés s'additionne au courant de drain alors que le reste traverse l'isolant de grille (I_G).

Afin de simplifier l'étude, on considèrera seulement trois conditions élémentaires pour l'injection de porteurs chauds vers la grille :

- Les porteurs doivent posséder une énergie suffisante afin de franchir la barrière de potentiel oxyde / semiconducteur.
- Les porteurs doivent avoir une direction perpendiculaire à l'interface Si/SiO₂.

- Lors de leur parcours dans le réseau, après avoir acquis une énergie suffisante, les porteurs ne doivent pas avoir d'interaction avec le réseau et ainsi conserver leur énergie, il en va de même dans l'oxyde.

Pour nos travaux de thèse, nous nous sommes basés sur l'équation du courant d'injection de porteurs chauds (CHEI pour Channel Hot Electron Injection) donnée par Tam *et al.* [Tam'84] :

$$I_G = I_{\text{sub}} \alpha_{\text{ox}} \exp\left(-\frac{b_{\text{ox}}}{\xi_{\text{ox}}}\right) \quad (\text{I.47})$$

où b_{ox} et α_{ox} sont les deux paramètres d'injection.

Dans cette équation b_{ox} est le facteur de dépendance du courant de grille avec le champ électrique dans l'oxyde tunnel (paramètre représentant la probabilité de passage d'un électron à travers l'interface Si/SiO₂ des modèles classiques et de l'électron chanceux) et α_{ox} celui du courant grille par rapport au courant substrat. Le courant substrat a pour expression d'après le modèle de Schokley Read Hall :

$$I_{\text{sub}} = I_{\text{DS}} \frac{a_i}{b_i} (V_D - V_{\text{D sat}}) \exp\left(\frac{-b_i}{V_D - V_{\text{D sat}}}\right) \quad (\text{I.48})$$

où a_i , b_i sont les coefficients d'ionisation par impact et V_{sat} le potentiel appliqué aux bornes de la zone à saturation.

Les coefficients a_i , b_i , α_{ox} et b_{ox} sont obtenus à partir des caractéristiques statiques $I_{\text{SUB}}(V_{\text{GS}}, V_{\text{DS}})$ et $I_G(V_{\text{GS}}, V_{\text{DS}})$ mesurées sur des mémoires dont la grille flottante est reliée à la grille de contrôle (aussi appelées dummy cell).

En raison de la grande intensité de ce dernier type d'injection, l'opération d'écriture d'une mémoire Flash est extrêmement rapide comparée à l'opération d'effacement (courant FN). Cette particularité rend la mémoire Flash très attractive par rapport à la mémoire EEPROM. Cependant, la limitation de la surface consacrée à la mémoire et le volume croissant du stockage souhaité nécessite la réduction de la taille des composants. Suivant la Roadmap International Technologie Roadmap for Semiconductor (ITRS) 2003, la taille limite des mémoires flash serait de 65 nm avec une épaisseur d'oxyde tunnel de l'ordre de 8-9 nm. Cette taille critique est due à l'incompatibilité entre la réduction de l'épaisseur de l'oxyde de grille pour contrôler les phénomènes de canaux courts et la préservation d'une épaisseur de diélectrique minimum pour maintenir sa fiabilité et la rétention de la charge après plusieurs cycles d'écriture et d'effacement. De plus, dès 1990, Bez *et al.* [Bez'90] ont mis en évidence la limitation de la réduction de la longueur

des mémoires flash placées dans une architecture NOR à cause du phénomène appelé « Drain turn on », engendré par le fort couplage entre le drain et la grille flottante. Ce phénomène se traduit par le contrôle du canal par la polarisation de drain lorsque la grille n'est pas (ou peu) polarisée. La discrétisation de la grille supprime le couplage entre le drain et la grille flottante [Lombardo'04] ce qui induit la réduction de l'influence des phénomènes de canaux courts, et permet l'utilisation de tensions de drain plus élevées pour l'opération de lecture.

Ainsi, l'utilisation de stockages discrets en remplacement des traditionnels stockages continus dans la grille flottante est une des solutions envisagées pour surmonter la limitation de dimensions. Il existe plusieurs types de mémoires à piégeages discrets décrits dans la littérature car le matériau utilisé pour le stockage peut être du nitrure présentant de nombreux défauts naturels, ou des matériaux High K (Al_2O_3 ou HfO_2) ou des nano-cristaux de semiconducteur [Tiwari'95] et [Shi'98].

I.4.3. Les mémoires à nano-cristaux

Depuis les années 1990, les nano-cristaux (ou nodules, ou encore dots) de silicium sont très étudiés pour leurs propriétés physiques mais également pour la fabrication de nouveaux dispositifs pour la microélectronique et la photonique. En 1995, Tiwari utilise des nano-cristaux de silicium à la place des traditionnelles grilles flottantes [Tiwari'95]. Ces dispositifs sont très prometteurs en terme de réduction de dimensions car ils présentent l'avantage d'une haute densité d'intégration, d'une basse consommation en puissance (environ 12V contre 18-20V pour les mémoires Flash traditionnelles) et d'un bas coût de fabrication puisque les nodules ne nécessitent pas d'étape de masquage pour les isoler électriquement. Ainsi, le nombre de masques nécessaire à la fabrication des dispositifs décroît de 11 masques pour les Flash à grille flottante traditionnelle à 4 masques pour les mémoires à nano-cristaux [Chang'03].

I.4.3.1. La structure des mémoires à nano-cristaux de silicium

La figure (I.16) présente une coupe TEM (Transmission Electron Microscopy) d'une mémoire à nodules de diamètre 5 nm et de longueur, $L = 0.2$ à 0.3 μm . Cette mémoire est un transistor d'apparence classique si ce n'est la présence de "boules" ou de "demi-boules" de silicium, de tailles nanométriques, réparties dans l'oxyde de grille à une certaine distance de l'interface, recouvrant entièrement la surface du canal. La densité de nodules est de l'ordre de 10^{12} dots. cm^{-2} et l'épaisseur de la couche de diélectrique

séparant les nano-cristaux du substrat est contrôlée afin de diminuer la dispersion de la distribution des tensions de seuils.

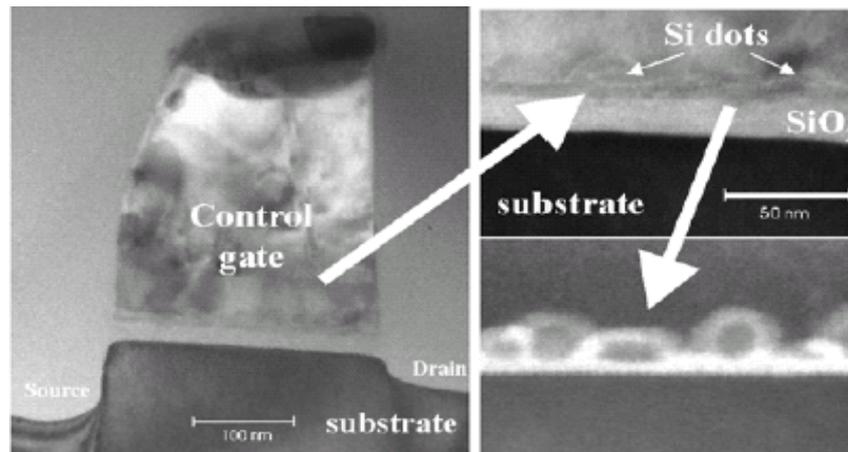


Figure I.16. Coupe TEM (Transmission Electron Microscopy) d'une mémoire à nodules de diamètre 5nm et de longueur $L = 0.2$ à $0.3 \mu\text{m}$ [Corso'03].

D'autres paramètres comme la taille des nano-cristaux, leur forme, l'isolation latérale entre les nano-cristaux, et l'uniformité de la densité surfacique des nano-cristaux doivent également être contrôlés lors du procédé de fabrication pour améliorer les performances de ces mémoires. Ainsi plusieurs techniques de fabrication de nano-cristaux ont successivement été envisagées : la croissance des nano-cristaux par dépôt LPCVD (Low pressure chemical vapor deposition) [Tiwary'95], la précipitation du silicium avec implantation ionique [Hanafi'96], le dépôt par aérosol [Debauwe'00]. Très récemment, une nouvelle technique de dépôt en deux temps, à partir de deux gaz différents (SiH_4 et SiH_2Cl_2), a mené à la séparation des phases de nucléation et de croissance des nano-cristaux, permettant une meilleure maîtrise de leur densité, de leur taille et par conséquent de leur isolement [De Salvo'03].

Hormis les mémoires à nano-cristaux présentées par Tiwary, il existe d'autres types de mémoires à nano-cristaux selon les matériaux utilisés (pour les nodules ou pour le diélectrique de grille) et la disposition des nano-cristaux. En 1998, une alternative aux nano-cristaux de silicium est proposée par l'Université de Berkeley avec les nano-cristaux de germanium. Ces mémoires présentaient de meilleures caractéristiques d'écriture et d'effacement et de meilleurs temps de rétentions que celles des dispositifs à nano-cristaux de silicium [King'98]. En 1998, Kim *et al.* [Kim'98] ont présenté également des mémoires à nano-cristaux de silicium utilisant un diélectrique formé d'oxyde nitrure permettant une meilleure uniformité dans la répartition des dots (à cause de la rugosité de la surface du nitrure). Puis, en 2002, les mémoires à nodules de métal ont été

également proposées [Liu'02]. Ces dispositifs présentent une plus forte densité d'états autour du niveau de Fermi (c.a.d. une plus grande protection contre la fluctuation des niveaux de Fermi causée par des contaminations), une plus grande gamme de valeur de travail de sortie et de plus petites perturbations d'énergie dues au confinement des porteurs.

Dans la suite de ce manuscrit, nous nous limiterons aux dispositifs à nodules de silicium.

I.4.3.2. Fonctionnement des mémoires à nano-cristaux de silicium

La fonction mémoire de ces dispositifs est attribuée à l'échange de charges entre les nano-cristaux de silicium et la couche d'inversion à travers un diélectrique tunnel fin, t_{ox1} . L'isolement des nodules les uns des autres [Chae'99], empêche le mouvement latéral des charges et préserve la mémoire d'une perte totale de l'information lors d'une détérioration locale de l'oxyde. En effet, les mémoires à nano-cristaux de silicium sont des mémoires à stockage discret pour lesquelles quelques électrons sont stockés dans chaque nodule (selon la taille de ces derniers). La charge emmagasinée dans l'ensemble des nodules contrôle la conductivité du canal du transistor mémoire.

L'injection d'un électron à partir de la couche d'inversion s'effectue par effet tunnel lorsque la grille est en polarisation directe par rapport à la source et au drain. La charge stockée écran la charge de la grille et réduit la conduction dans la couche d'inversion, et par suite entraîne une augmentation de la tension de seuil. Le chargement de ces nodules, avec des électrons, peut se faire par injection Fowler-Nordheim en appliquant une tension de grille positive ou par porteurs chauds en appliquant une tension positive sur la grille et sur le drain et/ou sur la source.

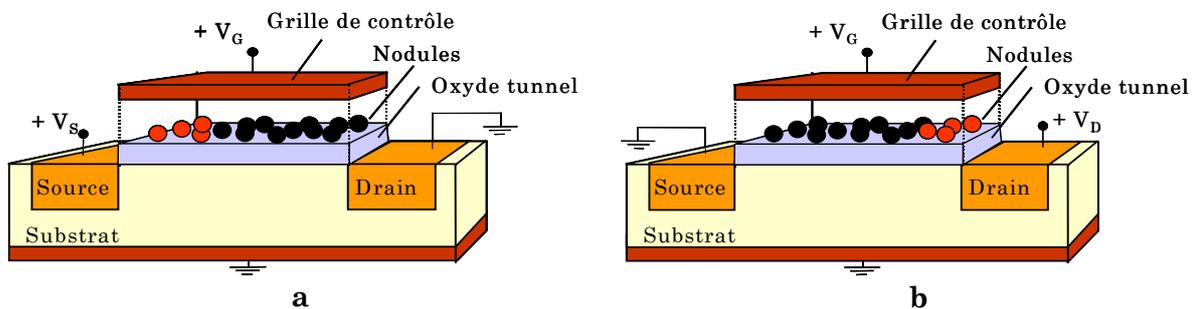


Figure I.17. Schématisation du fonctionnement 2 bits d'une mémoire à nodules avec écriture côté source (a) et côté drain (b).

Par conséquent, l'écriture par porteurs chauds localise l'injection des porteurs dans la région du canal proche du drain et/ou de la source suivant la polarisation choisie (tension

positive sur le drain et/ou la source (cf. Fig. (I.17)). Ainsi, les mémoires à nodules présentent la possibilité d'être utilisées comme des mémoires 2 bits [Eitan'99], [Eitan'00] et [Hradsky'03]. Ce concept de « dual bit » a récemment été amélioré avec des dispositifs à nano-cristaux de métal comportant deux sources et deux drains ce qui permet l'obtention de 4 bits par cellule mémoire [Liu'03].

Le déchargement des nodules (effacement de la mémoire) s'effectue par courant Fowler-Nordheim en appliquant un potentiel de grille négatif. Comme pour la mémoire Flash l'état écrit ou effacé de la mémoire se fera par la détermination de la tension de seuil au cours d'une lecture $I_{DS}(V_{GS})$.

De part leurs petites dimensions, leur fonctionnement à basses tensions, les temps d'écriture et d'effacement (respectivement de l'ordre de la micro-seconde et de la milli-seconde) et leur endurance (10^5 cycles écriture/effacement), les mémoires à nodules sont de bons candidats pour les applications spatiales [Bell'01] et les applications commerciales de type téléphones portables ou ordinateurs portables. Par exemple, grâce à leur fort potentiel de miniaturisation, ces mémoires peuvent être utilisées pour des stockages de photos de caméras digitales. Enfin, notons que les dispositifs à nano-cristaux de silicium ont également des propriétés photoniques. En effet, en 2000, Patch [Patch'00], a montré que des nano-cristaux de silicium emprisonnés dans une couche de SiO_2 pouvaient émettre de la lumière lors de stimulations électriques.

I.5. Conclusion

L'objectif de ce premier chapitre était d'introduire les différentes notations utilisées dans la suite du manuscrit. Nous avons ainsi pu rappeler les principales caractéristiques et le mode de fonctionnement des différentes structures que nous avons étudiées durant nos travaux de thèse, à savoir les capacités MIS, les transistors MOS et les mémoires Flash à nodules de silicium. Les relations de base sur lesquelles reposent nos modèles ont été présentées. En ce qui concerne la modélisation du courant de drain du transistor MOS en inversion faible, et en inversion forte avant saturation, nous utiliserons le modèle de Pao et Sah [Pao'66] ou le modèle en feuillet [Brews'78] basés sur le calcul du potentiel de surface. De plus, les structures étudiées ayant des isolants de grille minces, nous avons également décrit les modes d'injections tunnel (Tunnel Direct et Fowler-Nordheim), ainsi que l'injection par porteurs chauds utilisée pour l'écriture des mémoires Flash à nodules.

Chapitre II

Modélisation de la capacité MOS

Chapitre II. Modélisation de la capacité MOS	45
II.1. Introduction	45
II.2. La poly-désertion de la grille	45
II.2.1. Détermination des potentiels de surface	45
II.2.2. Tracé des potentiels de surface et des charges	48
II.2.2.1. Prise en compte de la couche d'inversion de la grille	48
II.2.2.1. Désertion profonde du poly-silicium	50
II.2.3. Courbes C-V d'une capacité MOS avec poly-désertion	50
II.2.3.1. Prise en compte de la couche d'inversion du poly-silicium	50
II.2.3.2. Désertion profonde du poly-silicium	52
II.2.4. Détermination des dopages à partir de la courbe C-V	52
II.2.4.1. Exemple : cas d'une capacité MOS	54
II.2.4.2. Exemple : cas d'une capacité SOS	55
II.3. Non uniformité du dopage du substrat	56
II.3.1. Modélisation	57
II.3.2. Etude d'un dopage de substrat non constant	58
II.3.2.1. Gaussienne de largeur variable centrée à l'interface	58
II.3.2.2. Gaussienne non centrée à l'interface	59
II.3.2.3. Gaussiennes non centrées à l'interface à largeurs variables	60
II.4. Effet d'une épaisseur non constante	61
II.4.1. Modélisation pseudo 2D des courbes C-V et I-V	61
II.4.1.1. Modélisation des courbes C-V	62
II.4.1.2. Modélisation des courbes I-V	63
II.4.1.3. Simulations des courbes C-V et I-V	64
II.4.2. Comparaison entre la modélisation pseudo 2D et 2D	66
II.5. Charges fixes latérales non constantes	70
II.5.1. Modélisation des courbes C-V et I-V	70
II.5.2. Impact de la non uniformité de charges	71
II.5.2.1. Impact sur les courbes C-V	71
II.5.2.2. Impact sur les courbes I-V	71
II.6. Charges fixes non uniformément réparties dans l'isolant	72
II.6.1. Modélisation de la courbe I-V d'une capacité MOS	72
II.6.2. Modélisation de la courbe I-V d'une capacité MIS	74
II.6.3. Simulation des courbes I-V pour une capacité MOS	76
II.6.4. Simulation des courbes I-V pour une capacité MIS	79
II.6.5. Résolution spatiale de la charge piégée Q_{ox}	80
II.6.5.1. Structure d'étude	81
II.6.5.2. Impact du stress électrique	81
II.6.6. Détermination de Q_{ox} (Δy)	82
II.6.7. Origine de la charge fixe générée dans l'oxyde	85
II.7. Conclusion	88

Chapitre II. Modélisation de la capacité MOS

II.1. Introduction

Ce chapitre est dédié aux codes de simulation développés, lors de ce travail de thèse, pour décrire le fonctionnement de la capacité MOS en présence de phénomènes liés aux variations des paramètres clefs de la capacité. Nous étudierons successivement la diminution du dopage de grille, les variations du dopage de substrat, la variation de l'épaisseur de l'oxyde de grille et les variations latérale et verticale des charges fixes dans l'oxyde.

II.2. La poly-désertion de la grille

Le phénomène de poly-désertion (ou poly-déplétion) de la grille, défini au chapitre I, induit une déformation des courbes C-V ce qui conduit à une extraction imprécise de l'épaisseur de l'oxyde. Or cette épaisseur est un des paramètres clefs de la structure MOS. Dans ce paragraphe, nous présentons la méthode que nous avons développée pour déterminer, à V_{GB} donnée, les deux potentiels de surface (Ψ_S) et (Ψ_{SG}) (respectivement pour le substrat et pour la grille), afin de tracer la caractéristique C-V d'une capacité présentant une désertion du poly-silicium de grille.

II.2.1. Détermination des potentiels de surface

Prenons une capacité MOS dont le substrat est de type P et la grille en poly-silicium de type N^+ non dégénéré. Les équations (I.6) et (I.24) (cf. pages 17 et 21 respectivement) montrent que la détermination des potentiels de surface (c.a.d. Ψ_S et Ψ_{SG}), à partir du potentiel de grille V_{GB} ne peut être obtenue de façon analytique. Afin de résoudre simplement et simultanément ces équations, nous nous sommes basés sur la méthode numérique de Newton et Raphson en remarquant que l'équation (I.24) est une fonction monotone et dérivable par rapport à Ψ_S et Ψ_{SG} .

Considérons les fonctions $h(V_{GB}, \Psi_S, \Psi_{SG})$ et $h_G(V_{GB}, \Psi_S, \Psi_{SG})$ et leurs dérivées respectives $h'(\Psi_S)$ et $h'_G(\Psi_{SG})$ définies à partir des équations suivantes :

$$\begin{aligned}
 h(V_{GB}, \Psi_S, \Psi_{SG}) &= V_{FB} - V_{GB} + \Psi_S - \Psi_{SG} \\
 &\quad \pm \frac{\sqrt{2kT\varepsilon_0\varepsilon_{SC}M_{0S}}}{C_{ox}} \left[\frac{m_{0S}}{M_{0S}} (\exp(t_s\beta\Psi_S) - t_s\beta\Psi_S - 1) - 1 + \exp(-t_s\beta\Psi_S) + t_s\beta\Psi_S \right]^{\frac{1}{2}} \quad (\text{II.1})
 \end{aligned}$$

+ si $\Psi_S < 0$ et - si $\Psi_S > 0$

$$\begin{aligned}
 h_G(V_{GB}, \Psi_S, \Psi_{SG}) &= V_{FB} - V_{GB} + \Psi_S - \Psi_{SG} \\
 &\quad \mp \frac{\sqrt{2kT\varepsilon_0\varepsilon_{SC}M_{0G}}}{C_{ox}} \left[\frac{m_{0G}}{M_{0G}} (\exp(t_G\beta\Psi_{SG}) - t_G\beta\Psi_{SG} - 1) - 1 + \exp(-t_G\beta\Psi_{SG}) + t_G\beta\Psi_{SG} \right]^{\frac{1}{2}} \quad (\text{II.2})
 \end{aligned}$$

+ si $\Psi_{SG} < 0$ et - si $\Psi_{SG} > 0$

$$\begin{aligned}
 h'(\Psi_S) &= \frac{\partial h(V_{GB}, \Psi_S, \Psi_{SG})}{\partial \Psi_S} \\
 &= 1 \pm \frac{1}{2C_{ox}} \frac{\sqrt{2kT\varepsilon_0\varepsilon_{SC}M_{0S}} \left[\frac{m_{0S}}{M_{0S}} (t_s\beta \exp(t_s\beta\Psi_S) - t_s\beta) - t_s\beta \exp(-t_s\beta\Psi_S) + t_s\beta \right]}{\left[\frac{m_{0S}}{M_{0S}} (\exp(t_s\beta\Psi_S) - t_s\beta\Psi_S - 1) - 1 + \exp(-t_s\beta\Psi_S) + t_s\beta\Psi_S \right]^{\frac{1}{2}}} \quad (\text{II.3})
 \end{aligned}$$

$$\begin{aligned}
 h'_G(\Psi_{SG}) &= \frac{\partial h_G(V_{GB}, \Psi_S, \Psi_{SG})}{\partial \Psi_{SG}} \\
 &= 1 \mp \frac{1}{2C_{ox}} \frac{\sqrt{2kT\varepsilon_0\varepsilon_{SC}M_{0G}} \left[\frac{m_{0G}}{M_{0G}} (t_G\beta \exp(t_G\beta\Psi_{SG}) - t_G\beta) - t_G\beta \exp(-t_G\beta\Psi_{SG}) + t_G\beta \right]}{\left[\frac{m_{0G}}{M_{0G}} (\exp(t_G\beta\Psi_{SG}) - t_G\beta\Psi_{SG} - 1) - 1 + \exp(-t_G\beta\Psi_{SG}) + t_G\beta\Psi_{SG} \right]^{\frac{1}{2}}} \quad (\text{II.4})
 \end{aligned}$$

où M_{0S} , m_{0S} , M_{0G} , et m_{0G} , sont les densités de porteurs majoritaires et minoritaires du substrat et de la grille respectivement. t_s est un coefficient égal à (-1) pour un substrat de type N et à $(+1)$ pour un substrat de type P. De même, t_G est un coefficient égal à (-1) pour une grille de type N⁺ et $(+1)$ pour une grille de type P⁺.

Pour une tension V_{GB} donnée, les potentiels de surface Ψ_S et Ψ_{SG} sont obtenus lorsque les fonctions h et h_G sont nulles simultanément. On impose une valeur initiale pour les potentiels de surface Ψ_S et Ψ_{SG} (égale à Φ_F par exemple) que l'on note Ψ_{S0} et Ψ_{SG0} . Supposons que ces valeurs soient les solutions des l'équations (II.1) et (II.2), on obtient par conséquent : $h(V_{GB}, \Psi_{S0}, \Psi_{SG0}) = 0$ et $h_G(V_{GB}, \Psi_{S0}, \Psi_{SG0}) = 0$.

Considérons une valeur Ψ_{S1} proche de Ψ_{S0} et développons $h(V_{GB}, \Psi_{S0}, \Psi_{SG0})$ en série de Taylor autour de Ψ_{S1} et Ψ_{SG1} :

$$h(V_{GB}, \Psi_{S0}, \Psi_{SG0}) = h(V_{GB}, \Psi_{S1}, \Psi_{SG0}) - (\Psi_{S0} - \Psi_{S1})h'(\Psi_{S1}) + \frac{(\Psi_{S0} - \Psi_{S1})^2}{2!}h''(\Psi_{S1}) + \dots \quad (\text{II.5})$$

En utilisant le fait que $h(V_{GB}, \Psi_{S0}, \Psi_{SG0}) = 0$, on peut exprimer Ψ_{S0} :

$$\Psi_{S1} = \Psi_{S0} - \frac{h(V_{GB}, \Psi_{S1}, \Psi_{SG0})}{h'(\Psi_{S1})} + \frac{(\Psi_{S0} - \Psi_{S1})^2}{2!} \frac{h''(\Psi_{S1})}{h'(\Psi_{S1})} + \dots \quad (\text{II.6})$$

Si Ψ_{S1} est suffisamment proche de Ψ_{S0} , les termes d'ordre supérieur, au premier ordre, seront petits. En les négligeant, on obtient ainsi une seconde valeur approchée pour Ψ_{S0} , plus proche de Ψ_{S0} que Ψ_{S1} , donnée par:

$$\Psi_{S2} = \Psi_{S1} - \frac{h(V_{GB}, \Psi_{S1}, \Psi_{SG0})}{h'(\Psi_{S1})} \quad (\text{II.7})$$

En suivant le même raisonnement que précédemment, le développement de $h_G(V_{GB}, \Psi_{S2}, \Psi_{SG0})$ en série de Taylor autour de Ψ_{SG1} conduit à une valeur approchée Ψ_{SG2} , plus proche de Ψ_{SG0} que Ψ_{SG1} :

$$\Psi_{GS2} = \Psi_{SG1} - \frac{h_G(V_{GB}, \Psi_{S2}, \Psi_{SG1})}{h_G'(\Psi_{SG1})} \quad (\text{II.8})$$

En gardant les dernières valeurs des potentiels de surface déterminées et en répétant la même démarche à chaque itération, i , les valeurs Ψ_{Si} et Ψ_{SGi} se rapprochent de la solution exacte de Ψ_{S0} et Ψ_{SG0} (cf. Fig. II.1).

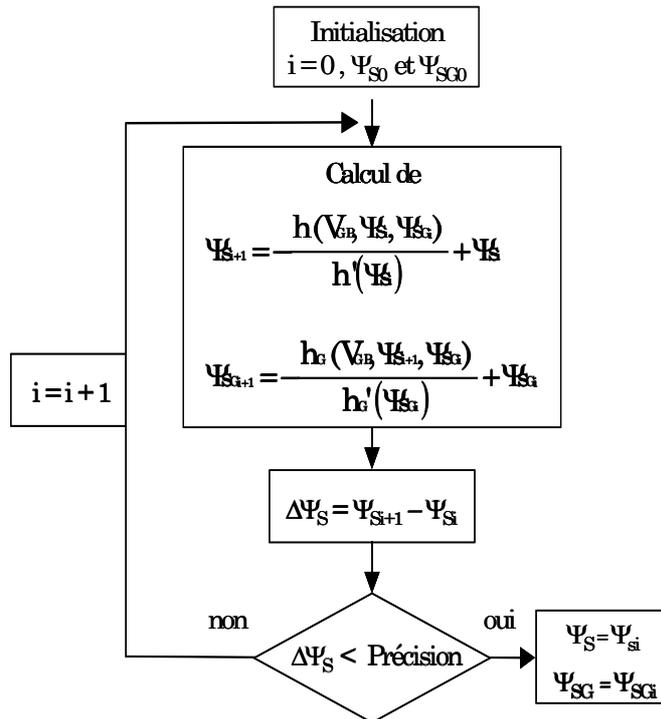


Figure II.1. Algorithme de résolution des potentiels de surface Ψ_S et Ψ_{SG} .

En pratique, après n itérations, il suffit que $h(V_{GB}, \Psi_{Sn}, \Psi_{SGn}) \leq \varepsilon$, avec ε suffisamment petit, pour poser $\Psi_{S0} \approx \Psi_{Sn}$ et $\Psi_{SG0} \approx \Psi_{SGn}$ avec une précision voulue (par exemple 10^{-11}). Si la valeur Ψ_{S1} est bien choisie, (c.a.d. suffisamment proche de Ψ_{S0}), cette méthode présente l'avantage de converger rapidement vers la solution Ψ_{S0} : le nombre n d'itérations nécessaires est alors relativement faible. Ainsi, cette méthode peut être plus

rapide que la méthode de dichotomie pour laquelle l'intervalle de recherche est fixé, empêchant l'optimisation du temps de calcul.

II.2.2. Tracé des potentiels de surface et des charges

A l'aide de l'algorithme décrit au paragraphe précédemment, les valeurs des potentiels de surface peuvent être déterminées pour différentes valeurs de V_{GB} . Dans ce chapitre, nous considérerons que le semiconducteur ne peut jamais être en régime de désertion profonde mais que cela est possible pour le poly-silicium. Dans ce cas l'expression de la charge de la grille (que l'on notera Q_{DG}) se simplifie pour devenir :

$$Q_{DG} = \pm \sqrt{2kT\epsilon_s \sin \theta_0} \left[\exp(\beta \Psi_{SG}) - \left(1 - \frac{p_{G0}}{n_{G0}}\right) \beta \Psi_{SG} - 1 \right]^{1/2} \quad (\text{II.9})$$

avec un signe $-$ pour $\Psi_G > 0$ et un signe $+$ pour $\Psi_G < 0$.

II.2.2.1. Prise en compte de la couche d'inversion de la grille

La figure (II.2.a) présente l'évolution des potentiels de surface du semiconducteur et de grille en fonction de la tension de grille et du dopage du poly-silicium. Notons, que pour un dopage N_G supérieur à 10^{28} m^{-3} (soit un facteur 10^5 par rapport au dopage du substrat), nous pouvons considérer la grille comme un quasi métal (pour la structure considérée), c'est-à-dire que la valeur de Ψ_{SG} reste toujours négligeable devant celles des autres potentiels (voir le point A sur la figure (II.2.a)). Nous constatons également que pour N_G supérieur à 10^{26} m^{-3} , l'évolution de Ψ_s en régime d'inversion faible et forte semble uniquement se décaler, alors que pour N_G inférieur à 10^{28} m^{-3} , elle semble changer de pente. La courbe $\Psi_{SG}(V_{GB})$ est par contre modifiée en régime d'inversion forte dès que N_G est inférieur à 10^{27} m^{-3} (voir le point B sur la figure (II.2.a)).

La figure (II.2.b) donne l'évolution de la charge du semiconducteur (qui est égale à l'opposée de la charge de grille), en fonction de la tension et du dopage de la grille. Pour un rapport (N_G/N_A) supérieur à 100, nous constatons un simple décalage de la courbe $Q_{SC}(V_{GB})$ alors que pour un rapport (N_G/N_A) inférieur à 100 la forme de cette courbe est modifiée en inversion forte. Le déplacement du minimum de la charge Q_{SC} résulte de la dépendance de la tension de bandes plates de la structure avec le dopage N_G :

$$V_{FB} = \Phi_{FG} - \Phi_F = -\frac{kT}{q} \ln\left(\frac{N_G}{n_i}\right) - \frac{kT}{q} \ln\left(\frac{N_{sub}}{n_i}\right) \quad (\text{II.10})$$

La figure (II.2.c) présente l'évolution de la charge d'inversion du semiconducteur en mettant en évidence le régime d'inversion faible (échelle logarithmique). Comme nous l'avons déjà constaté pour la courbe $\Psi_s(V_{GB})$, en raison de la faible valeur de Ψ_{SG} par rapport à celle de Ψ_s , l'évolution de Q_n n'est décalée que pour N_G supérieur 10^{25} m^{-3} . Par contre, la pente de $Q_n(V_{GB})$ diminue pour N_G inférieur 10^{25} m^{-3} . L'évolution de la charge Q_n en régime d'inversion forte (échelle linéaire) est donnée à la figure (II.2.d). A partir de $N_G = 10^{26} \text{ m}^{-3}$, la courbe $Q_n(V_{GB})$ présente une pente plus faible jusqu'à faire apparaître une double pente due au passage en inversion forte du poly-silicium (voir les points C sur la figure (II.2.d)).

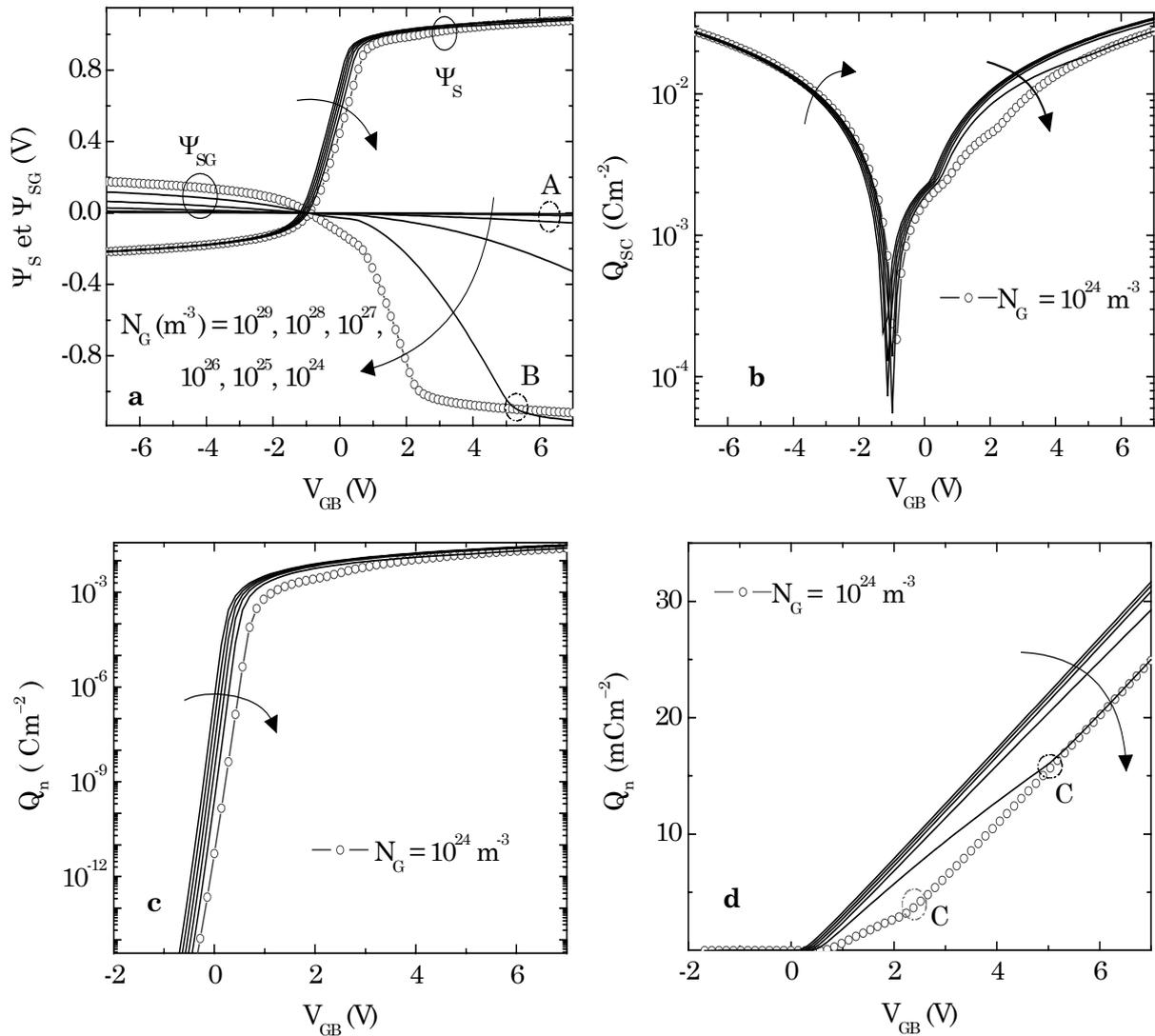


Figure II.2. Evolutions des potentiels de surface en fonction du potentiel de grille et du dopage du poly-silicium (a). Variations de la charge totale du semiconducteur (b) et de la charge d'inversion du semiconducteur en échelle linéaire (c) et logarithmique (d). Les paramètres de la capacité MOS sont : $t_{ox} = 7 \text{ nm}$, $N_A = 2 \times 10^{23} \text{ m}^{-3}$, $V_{FB} \in [-1, -1.2 \text{ V}]$.

II.2.2.1. Désertion profonde du poly-silicium

Comme le montre la figure (II.3), la double pente de la courbe $Q_n(V_{GB})$ (cf. courbe (II.2.d)) disparaît si l'on considère que la couche d'inversion du poly-silicium de grille ne peut jamais être présente (c.a.d. présence du phénomène de désertion profonde de la grille). De plus, la courbe d'évolution de Q_n n'est pas linéaire comme cela est supposé être le cas lorsque le dopage de grille est très important. Cependant cette quasi linéarité apparaît sur la deuxième pente de Q_n lorsque le poly-silicium est en régime d'inversion forte.

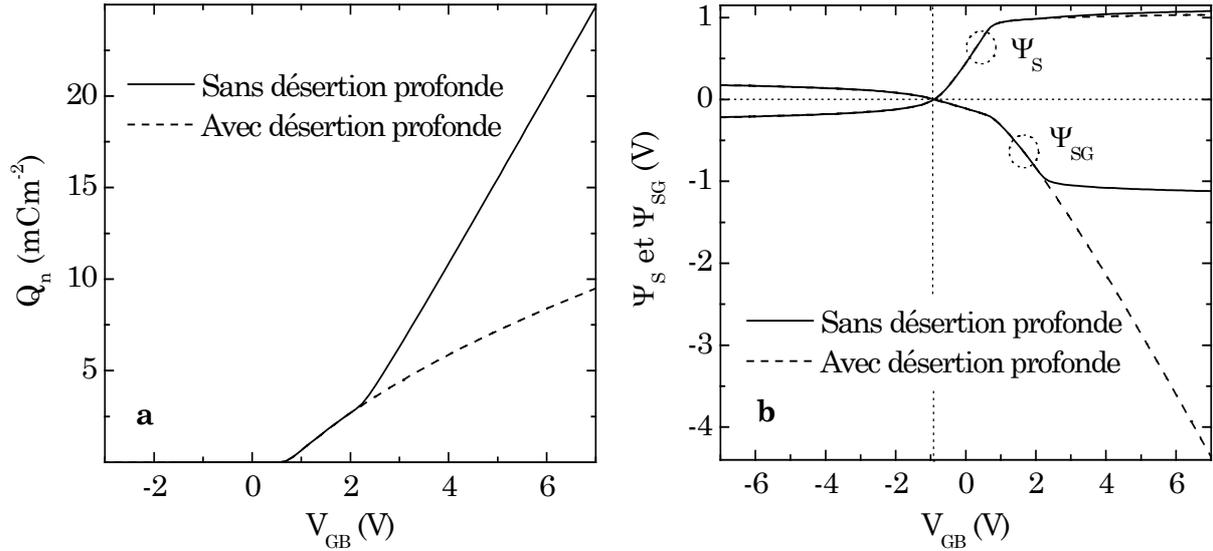


Figure II.3. Evolutions de la charge d'inversion (a) et des potentiels de surface (b) si l'on considère ou non la désertion profonde du poly-silicium. Les paramètres de la capacité MOS sont : $t_{ox} = 7$ nm, $N_A = 2 \times 10^{23} m^{-3}$, $N_G = 1 \times 10^{24} m^{-3}$, $V_{FB} = -1.2$ V.

II.2.3. Courbes C-V d'une capacité MOS avec poly-désertion

II.2.3.1. Prise en compte de la couche d'inversion du poly-silicium

La figure (II.4.a) montre l'évolution de la courbe C-V en fonction du dopage de grille (V_{FB} est fixée à -1 V quel que soit le dopage). On observe une forte diminution de la capacité en régime d'inversion forte du substrat lorsque V_{GB} augmente. La présence d'une telle variation permet la détection du phénomène de désertion de grille. Pour cette simulation, la grille reste dans les régimes d'accumulation et de désertion. Nous pouvons aussi observer, sur cette figure, un abaissement de la courbe en régime accumulation du substrat qui peut conduire à une sur-estimation de l'épaisseur d'isolant [Huang'93]. Pour des dopages de grille plus faibles (cf. Fig. (II.4.b)), le passage en régime d'inversion forte de la grille est mis en évidence par une brusque reprise de l'augmentation de la capacité

de la structure (après la diminution de cette dernière) en régime d'inversion forte du substrat [Schuegraf'93].

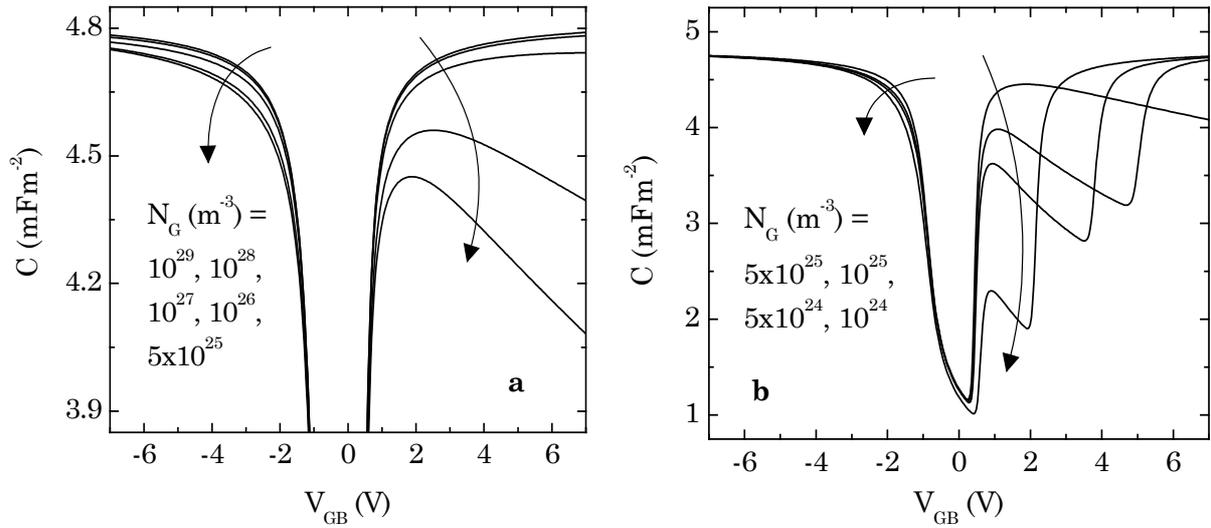


Figure II.4. Evolution de la courbe C-V en BF pour de "fortes" (a) et de "faibles" (b) valeurs du dopage de grille. Les paramètres de la simulation sont : $N_A = 2 \times 10^{23} \text{ m}^{-3}$, $t_{ox} = 7 \text{ nm}$, $V_{FB} = -1 \text{ V}$.

Afin de tracer les courbes de la figure (II.4), nous avons supposé que la couche d'inversion de la grille pouvait suivre les variations du signal de grille (petit signal sinusoïdal utilisé pour les mesures) mais nous pouvons aussi considérer que la formation de cette couche (génération thermique des porteurs) est dominée par une constante de temps τ . La capacité de la structure, C_{inv} , est alors calculée en considérant que la couche d'inversion apparaît au bout d'un temps τ :

$$C_{inv} = \frac{1}{\frac{1}{C_{SC}} + \frac{1}{C_{ox}} + \frac{1}{C_{Gd} + \frac{C_{G_{inv}}}{\sqrt{1 + (2\pi f\tau)^2}}}} \quad (\text{II.11})$$

où f correspond à la fréquence du signal de grille et τ à la constante de temps de formation de la couche d'inversion.

La figure (II.5) montre l'influence de la fréquence du signal de grille sur l'amplitude de la courbe C-V en régime d'inversion forte (de la grille). A haute fréquence (HF), la capacité ne redevient pas égale à C_{ox} à fort V_{GB} .

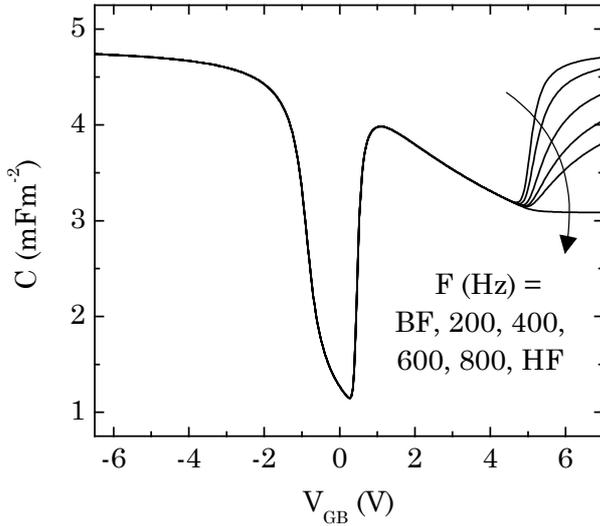


Figure II.5. Evolution de la courbe C-V des basses fréquences (BF) aux fortes fréquences (HF). Les paramètres de la simulation sont : $N_A = 2 \times 10^{23} \text{ m}^{-3}$, $N_G = 5 \times 10^{24} \text{ m}^{-3}$, $t_{ox} = 7 \text{ nm}$, $V_{FB} = -1 \text{ V}$.

II.2.3.2. Désertion profonde du poly-silicium

Nous pouvons aussi considérer le cas où la couche d'inversion de la grille ne peut pas se former : la grille est alors en régime de désertion profonde ce qui signifie que sa zone de charge d'espace continue à s'étendre avec le potentiel de grille et que le potentiel de surface continue à croître (c.a.d. qu'il ne sature pas à environ $2 \times \Phi_{Fg}$). Dans ce cas, comme le montre la figure (II.6), la capacité de la structure continue à chuter lorsque V_{GB} augmente.

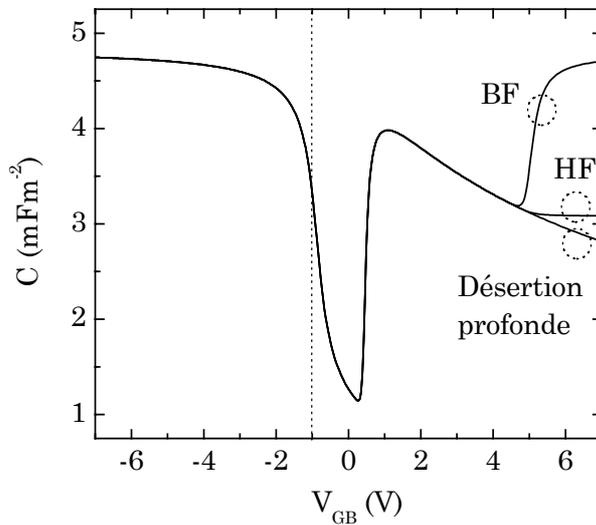


Figure II.6. Courbes C-V en BF, HF et en désertion profonde. Les paramètres de la simulation sont : $N_A = 2 \times 10^{23} \text{ m}^{-3}$, $N_G = 5 \times 10^{25} \text{ m}^{-3}$, $t_{ox} = 7 \text{ nm}$ et $V_{FB} = -1 \text{ V}$.

II.2.4. Détermination des dopages à partir de la courbe C-V

A partir des caractéristiques C-V, il est possible d'extraire le dopage du substrat (N_{Sub}) et celui de la grille (N_G). Dans ce qui suit, les capacités seront données par unité de surface. Les parties de la courbe C-V qui ne dépendent que du dopage du substrat ou du dopage de grille correspondent au régime de désertion et au régime d'inversion faible du

substrat ou de la grille (c.a.d. quand Q_{SC} ou Q_G correspondent à la charge d'une zone désertée).

Prenons le cas du substrat : en régime de désertion ou de faible inversion du substrat, la longueur de la région de désertion, x_d , est donnée par l'équation :

$$x_d = -t_s \sqrt{\frac{2\varepsilon_0 \varepsilon_{SC}}{qN_{sub}} \Psi_S} \quad (II.12)$$

La charge du semiconducteur peut être ré-écrite simplement :

$$Q_{SC} = -(t_s)qN_{sub}x_d = -(t_s)\sqrt{2qN_{sub}\varepsilon_0\varepsilon_{SC}}\Psi_S \quad (II.13)$$

A partir des équations (I.16) et (I.17), l'équation aux potentiels s'écrit :

$$V_{GB} - V_{FB} = \frac{qN_{sub}x_d}{C_{ox}} + \frac{qN_{sub}}{2\varepsilon_{SC}}x_d^2 \quad (II.14)$$

Or

$$C_{SC} = \left| \frac{dQ_{SC}}{d\Psi_S} \right| = \sqrt{\frac{qN_{sub}\varepsilon_{Si}}{2\Psi_S}} \quad (II.15)$$

Des équations (II.12), (II.15) nous déduisons :

$$C_{SC} = \varepsilon_{Si}x_d^{-1} \quad (II.16)$$

En remplaçant x_d dans l'équation (I.18) (cf. page 19) par :

$$x_d = \frac{\varepsilon_{SC}\varepsilon_0}{C_{SC}} \quad (II.17)$$

on obtient :

$$\left(\frac{1}{C_{SC}}\right)^2 + \frac{2}{C_{ox}C_{SC}} - \frac{2}{qN_{sub}\varepsilon_{SC}\varepsilon_0}(V_{GB} - V_{FB}) = 0 \quad (II.18)$$

Cette expression est de la forme : $X^2 + bX + c = 0$ avec $X = 1/C_{SC}$ et a pour solution physique :

$$\frac{1}{C} = \frac{1}{C_{ox}} \left(\sqrt{1 + \frac{2C_{ox}^2(V_{GB} - V_{FB})}{\varepsilon_{sc}qN_{sub}}} \right) \quad (II.19)$$

Dans cette expression, les capacités étant exprimées par unité de surface, pour retrouver la valeur de la capacité, il faut ré-introduire la surface A_{eff} :

$$\left(\frac{A_{eff}}{C}\right)^2 = \left(\frac{A_{eff}}{C_{ox}}\right)^2 + \frac{2}{q\varepsilon_0\varepsilon_{SC}N_{sub}}(V_{GB} - V_{FB}) \quad (II.20)$$

Par conséquent, il n'est pas nécessaire de connaître la valeur de la capacité de l'oxyde (ni celle de son épaisseur) pour déterminer N_{sub} .

Par un raisonnement similaire appliqué au poly-silicium de grille, on peut écrire :

$$\left(\frac{A_{\text{eff}}}{C}\right)^2 = \left(\frac{A_{\text{eff}}}{C_{\text{ox}}}\right)^2 + \frac{2}{q\varepsilon_0\varepsilon_{\text{SC}}N_{\text{G}}}(V_{\text{GB}} - V_{\text{FB}}) \quad (\text{II.21})$$

Ainsi le tracé de $(A_{\text{eff}}/C)^2 = f(V_{\text{GB}})$ permet d'extraire à la fois les dopages N_{sub} et N_{G} de la capacité suivant la portion de la courbe C-V sélectionnée.

II.2.4.1. Exemple : cas d'une capacité MOS

Considérons la mesure C-V d'une capacité MOS représentée par des ronds sur la figure (II.7.a). Le dopage de substrat, N_{sub} est alors extrait à partir de la pente de la droite déterminée en choisissant deux points dans la zone de désertion du substrat sur le graphe $(A_{\text{eff}}/C)^2 = f(V_{\text{GB}})$ (cf. Fig. (II.7.b)). De plus, connaissant le paramètre C_{ox} , l'ordonnée à l'origine permet d'extraire la tension de bandes plates.

$$N_{\text{sub}} = \frac{2}{q\varepsilon_{\text{SC}}\varepsilon_0 \text{ pente1}} \quad (\text{II.22})$$

$$V_{\text{FB}} = \left[\left(\frac{A_{\text{eff}}}{C_{\text{ox}}}\right)^2 - b \right] \left(\frac{q\varepsilon_{\text{SC}}\varepsilon_0 N_{\text{sub}}}{2} \right) \quad (\text{II.23})$$

où b représente l'ordonnée à l'origine du tracé de $(A_{\text{eff}}/C)^2 = f(V_{\text{GB}})$

Notons, que dans ce cas, les charges d'interface Q_{it} , ont été négligées, ce qui permet l'obtention d'une droite. Si ces charges ne sont plus négligeables, la droite de pente « pente 1 » aura tendance à se courber légèrement et l'équation donnant la capacité devra être remplacée par l'expression :

$$\frac{1}{C} = \frac{1}{C_{\text{ox}}} + \frac{1}{C_{\text{D}} + C_{\text{it}}} \quad (\text{II.24})$$

De la même manière, en choisissant les deux points dans la zone correspondant au régime de désertion de la grille (forte valeur de V_{GB} sur le graphe $(A_{\text{eff}}/C)^2 = f(V_{\text{GB}})$), le dopage de la grille, N_{G} , sera déterminé à partir de la pente de la droite (cf. pente 2 sur la figure (II.7.b)) :

$$N_{\text{G}} = \frac{2}{q\varepsilon_{\text{SC}}\varepsilon_0 \text{ pente2}} \quad (\text{II.25})$$

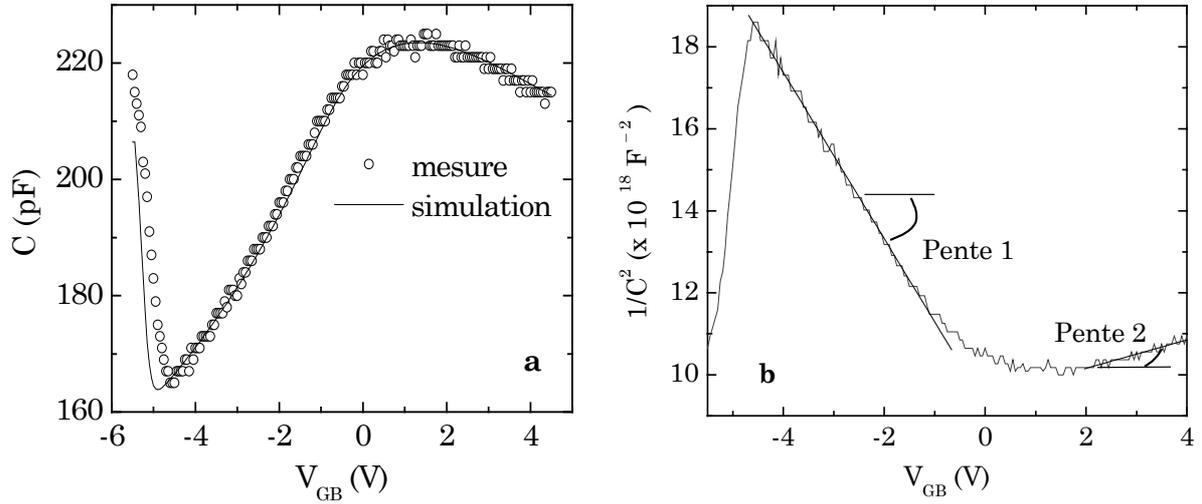


Figure II.7. Comparaison, pour une capacité MOS avec poly-désertion de grille, d'une courbe C-V mesurée et d'une courbe C-V simulée (a). Extraction du dopage de substrat N_{sub} et du dopage de grille N_G à partir du tracé de $1/C^2$ en fonction de V_{GB} (b). Les paramètres de la capacité mesurée sont : $t_{ox} = 11 \text{ nm}$, $V_{FB} = -1 \text{ V}$, $W \times L = 30000 \mu\text{m}^2$. Les dopages extraits sont : $N_{sub} = 2 \times 10^{23} \text{ m}^{-3}$ et $N_G = 1 \times 10^{26} \text{ m}^{-3}$.

II.2.4.2. Exemple : cas d'une capacité SOS

Lorsque la grille et le substrat sont des semiconducteurs dont les dopages sont du même ordre de grandeur, la courbe C-V de la capacité SOS (Semiconducteur-Oxide-Semiconducteur) prend la forme d'une cloche (cf. Fig. (II.8.a)).

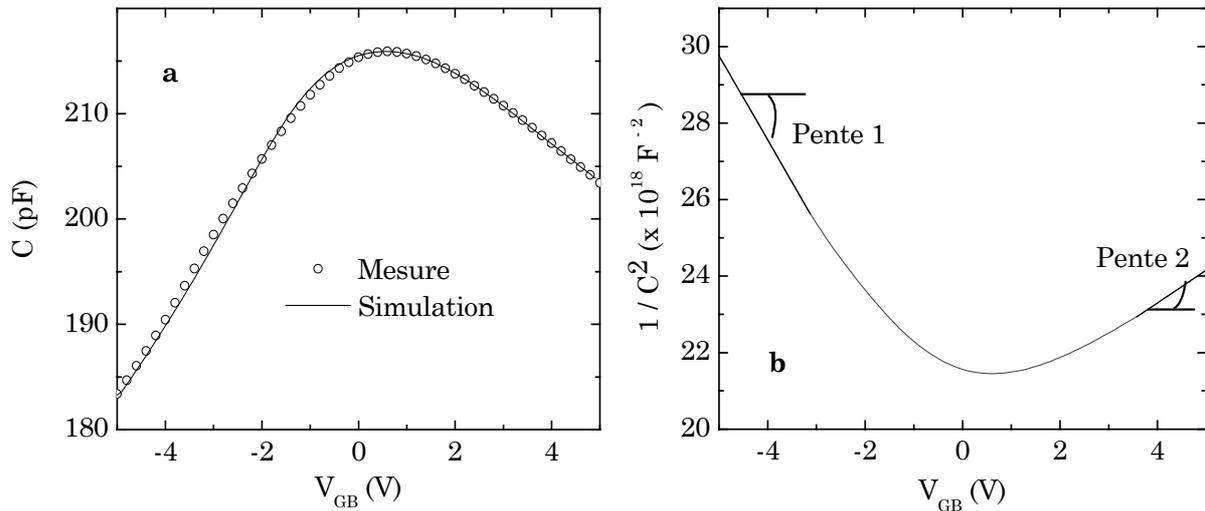


Figure II.8. Comparaison pour une capacité SOS avec poly-désertion de grille d'une courbe C-V mesurée et d'une courbe C-V simulée (a). Extraction du dopage de substrat N_{sub} et du dopage de grille N_G à partir du tracé de $1/C^2$ en fonction de V_{GB} (b). Les dimensions de la capacité mesurée sont $W \times L = 56000 \mu\text{m}^2$. Les paramètres extraits sont : $N_{sub} = 1.721 \times 10^{25} \text{ m}^{-3}$, $N_G = 4.19 \times 10^{25} \text{ m}^{-3}$ et $t_{ox} = 8.3 \text{ nm}$.

En nous basant sur le cas de la capacité MOS, nous avons réalisé un programme semi-automatisé, sous environnement mathcad, pour extraire les dopages du substrat et de la grille de ces capacités SOS. Par exemple, à partir de la courbe $(A_{\text{eff}}/C)^2 = f(V_{\text{GB}})$, tracée sur la figure (II.8.b), et du choix des zones de désertion du substrat et de la grille, les dopages de la capacité SOS mesurée ont pu être extraits : $N_{\text{sub}} = 1.721 \times 10^{25} \text{ m}^{-3}$ et $N_{\text{G}} = 4.19 \times 10^{25} \text{ m}^{-3}$.

Notons, qu'à partir des mesures C-V des capacités MOS (ou SOS) ayant des oxydes minces (30-40Å), il est également possible de déterminer l'épaisseur d'oxyde, t_{ox} , à partir de la méthode de Maserjian [Maserjian'74]. Cette méthode est basée sur la fonction suivante :

$$Y(V_{\text{GB}}) = \frac{1}{C^3} \frac{\partial C}{\partial V_{\text{GB}}} = \frac{1}{C^3} \frac{\partial C_{\text{SC}}}{\partial \Psi_{\text{S}}} \quad (\text{II.26})$$

où C représente la capacité mesurée.

A partir de cette équation en régime d'accumulation, selon Maserjian *et al.*, il est possible d'écrire la relation suivante :

$$\frac{1}{C} = \frac{1}{C_{\text{ox}}} + KY \frac{1}{n+1} \quad (\text{II.27})$$

pour laquelle K est une constante et n un exposant dépendant de la statistique des porteurs en accumulation (n=1 pour la statistique de Boltzmann, n=3 pour la mécanique quantique et n=5 pour la statistique de métal dégénéré).

Ainsi, pour une valeur de n donnée, l'ordonnée à l'origine du tracé de $(1/C)$ en fonction de $(Y^{1/n+1})$ fournit la valeur de C_{ox} et donc de t_{ox} . Dans ce premier paragraphe, nous avons étudié l'impact de la poly-désertion de la grille sur les caractéristiques de la capacité. Nous nous proposons, à présent, d'étudier l'impact des variations du dopage de substrat d'une capacité.

II.3. Non uniformité du dopage du substrat

En raison des procédés de fabrication (croissance de l'oxyde, implantation du substrat, ajustage de la tension de seuil...), le dopage du substrat n'est pas réellement uniforme dans le volume du semiconducteur. Pour un dopage de substrat non constant entre l'interface oxyde/substrat et le volume du substrat (cf. Fig. II.9), il est nécessaire de déterminer la courbure des bandes du semiconducteur afin d'obtenir la charge totale de

la zone désertée. Nous avons donc réalisé un simulateur de courbes C-V (sous environnement Mathcad) pour résoudre l'équation de Poisson en 1 dimension dans le substrat.

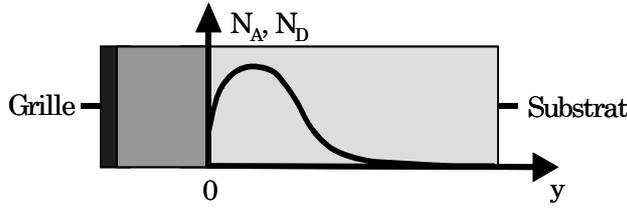


Figure II.9. Capacité MOS à dopage de substrat non constant

II.3.1. Modélisation

Nous développons ici la méthode numérique itérative fondée sur les différences finies à une dimension afin de déterminer $\Psi(y)$ pour un potentiel de surface donné. En conservant la résolution de l'équation de Poisson à une dimension, la densité de charge dans le substrat s'écrit :

$$\rho(y) = q[p(y) - n(y) + N_D(y) - N_A(y)] \quad (\text{II.28})$$

où $N_A(y)$ et $N_D(y)$ représentent les densités de dopants (de types accepteurs et donneurs) de l'interface vers le volume du semiconducteur.

Comme le montre la figure (II.10), la résolution de l'équation de Poisson nécessite le découpage de l'axe y en petits éléments de longueur Δ . L'épaisseur de la couche d'inversion étant très faible par rapport à la longueur de la ZCE, il est indispensable de découper l'axe y de façon non uniforme : le maillage est plus resserré à l'interface.

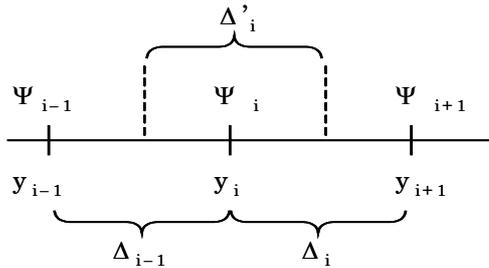


Figure II.10. Maillage du substrat à une dimension.

La dérivée première de la courbure de bandes est donnée par :

$$\Psi_i' = \frac{\Psi_{i+1} - \Psi_{i-1}}{\Delta_{i-1} + \Delta_i} \quad (\text{II.29})$$

Par suite, la dérivée seconde de la courbure de bandes s'écrit de la façon suivante :

$$\Psi_i'' = \frac{\frac{\Psi_{i+1} - \Psi_i}{\Delta_i} - \frac{\Psi_i - \Psi_{i-1}}{\Delta_{i-1}}}{\Delta_i'} = \frac{2}{\Delta_{i-1} + \Delta_i} \left[\frac{\Psi_{i+1} - \Psi_i}{\Delta_i} - \frac{\Psi_i - \Psi_{i-1}}{\Delta_{i-1}} \right] \quad (\text{II.30})$$

car :

$$\Delta_i' = \frac{\Delta_{i-1} + \Delta_i}{2} \quad (\text{II.31})$$

On développe alors l'expression (II.30) :

$$\Psi_i'' = \frac{2}{(\Delta_{i-1} + \Delta_i)\Delta_i\Delta_{i-1}} [\Delta_{i-1}\Psi_{i+1} - \Delta_{i-1}\Psi_i - \Delta_i\Psi_i + \Delta_i\Psi_{i-1}] \quad (\text{II.32})$$

pour exprimer la courbure de bandes au point i :

$$\Psi_i = -\frac{\Delta_i\Delta_{i-1}}{2}\Psi_i'' + \frac{\Delta_{i-1}}{\Delta_{i-1} + \Delta_i}\Psi_{i+1} + \frac{\Delta_i}{\Delta_{i-1} + \Delta_i}\Psi_{i-1} \quad (\text{II.33})$$

D'après l'équation de Poisson (1D), la dérivée seconde de la courbure de bandes au point i dans le semiconducteur correspond à la densité de charge en ce point divisée par la permittivité du substrat, nous pouvons donc écrire :

$$\Psi_i = \frac{\Delta_i\Delta_{i-1}}{2} \frac{\rho_i}{\epsilon_{Si}} + \frac{\Delta_{i-1}}{\Delta_{i-1} + \Delta_i}\Psi_{i+1} + \frac{\Delta_i}{\Delta_{i-1} + \Delta_i}\Psi_{i-1} \quad (\text{II.34})$$

La charge totale du semiconducteur (par unité de surface) s'obtient en intégrant la densité de charge ρ suivant l'axe y :

$$Q_{SC} = \sum_i \frac{\rho_i + \rho_{i+1}}{2} \Delta_i \quad (\text{II.35})$$

II.3.2. Etude d'un dopage de substrat non constant

Dans ce paragraphe, nous considérons un profil de dopage de type gaussien de largeur variable (Δy variable) centré plus ou moins loin de l'interface (y_C variable) décrit par l'expression suivante :

$$N_A(y) = N_{Amax} \exp\left(-\left(\frac{y-y_C}{\Delta y}\right)^2\right) + N_{Amin} \quad (\text{II.36})$$

où N_{Amax} correspond à la valeur maximale de la gaussienne et N_{Amin} au dopage minimum du substrat.

II.3.2.1. Gaussienne de largeur variable centrée à l'interface

Dans cette partie, nous considérons que la gaussienne est centrée à l'interface (c.a.d. $y_c = 0$) et que la largeur de cette gaussienne est variable jusqu'à l'obtention d'un dopage constant (c.a.d. Δy très grand).

Les profils de dopage choisis sont tracés sur la figure (II.11.a). La figure (II.11.b) présente les courbes C-V correspondantes. Soit N_{A0} la courbe de référence (dopage constant), nous remarquons qu'une diminution de la largeur de la gaussienne induit :

- Un décalage de la courbe C-V dans le régime d'inversion forte,
- Une déformation de la courbe dans le régime d'inversion faible puis de désertion.

En revanche, le régime d'accumulation n'est pas affecté par cette variation du profil du dopage et la tension de bandes plates reste constante ($V_{FB} = -1V$) alors que V_T diminue.

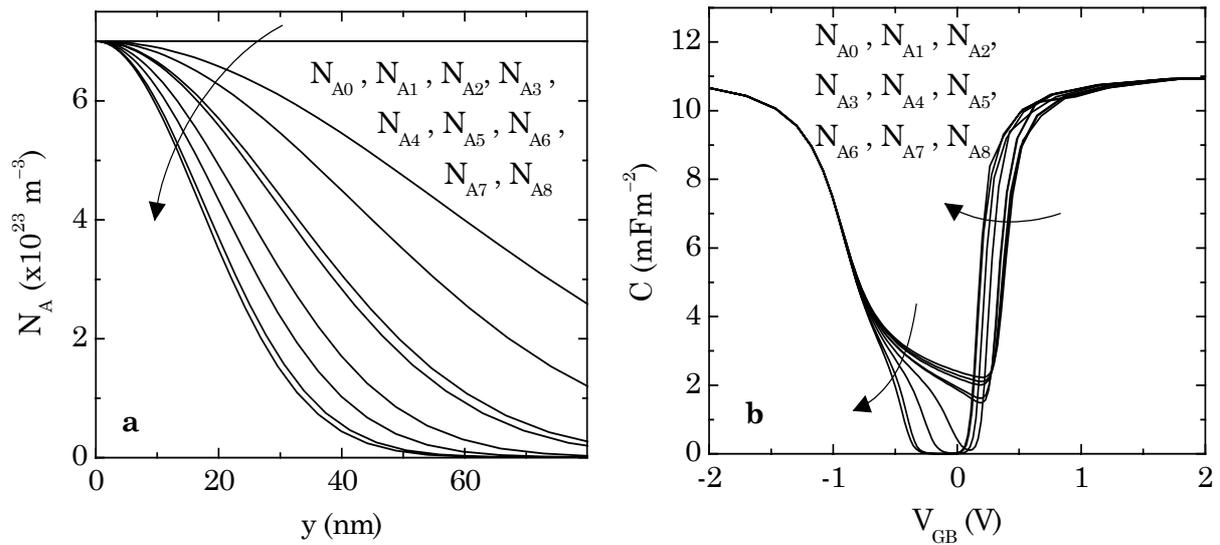


Figure II.11. Profils du dopage (a) et courbes C-V correspondantes (b) si l'on considère un dopage centré à l'interface et de largeur variable. L'épaisseur d'isolant est de 3 nm.

II.3.2.2. Gaussienne non centrée à l'interface

Considérons à présent, à largeur de gaussienne constante, un décalage du maximum de dopage par rapport à l'interface (cf. Fig. (II.12.a)). Soit N_{A1} le dopage de référence de la courbe C-V, la figure (II.12.b) montre qu'un décalage important de la gaussienne déforme assez peu la courbe mais qu'il entraîne un décalage global de celle-ci (variation de V_{FB} et de V_T).

Notons qu'il n'y a pas de déformation caractéristique de la courbe C-V par rapport aux résultats obtenus en diminuant fortement la largeur de la gaussienne.

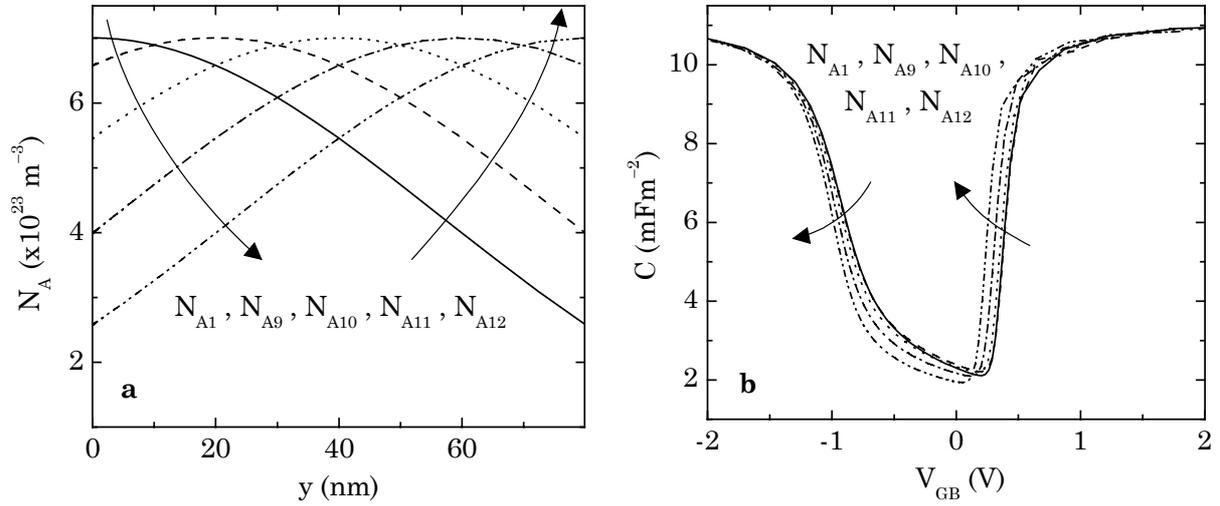


Figure II.12. Profils du dopage (a) et courbes C-V correspondantes (b) si l'on considère une largeur de gaussienne constante et un maximum décalé par rapport à l'interface. L'épaisseur d'isolant est de 3 nm.

II.3.2.3. Gaussiennes non centrées à l'interface à largeurs variables

Finalement, nous considérons un profil de dopage dont le maximum se situe à 20 nm de l'interface avec une largeur variable (cf. Fig. (II.13.a)). En prenant le dopage N_{A9} pour référence, la figure (II.13.b) met en évidence la déformation de la courbe C-V pour une forte diminution de la largeur de la gaussienne ce qui confirme les résultats précédents.

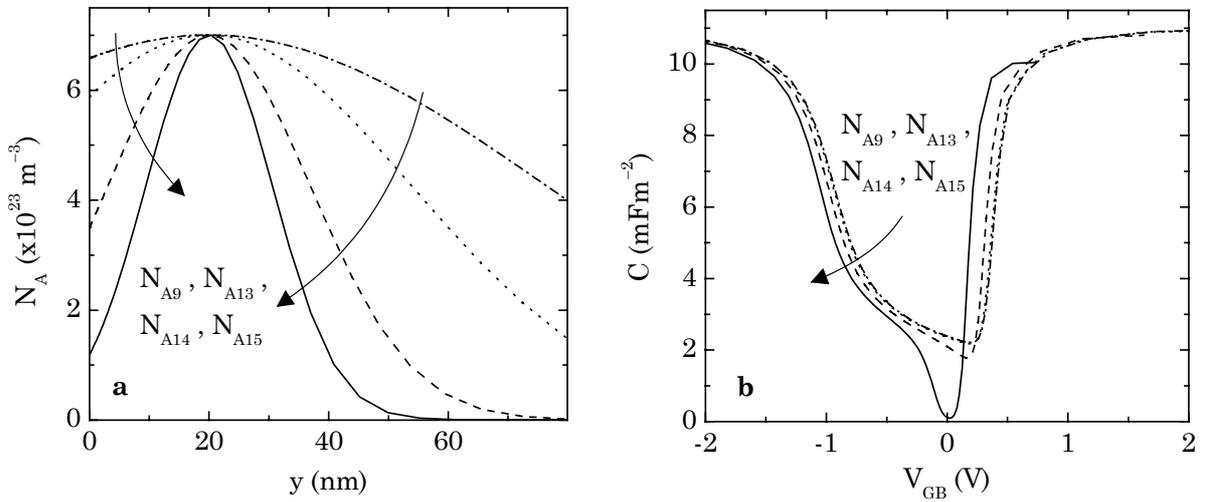


Figure II.13. Profils du dopage (a) et courbes C-V correspondantes (b) si l'on considère une gaussienne centrée à 20 nm de l'interface et de largeur variable. L'épaisseur d'isolant est de 3 nm.

Comme nous l'avons décrit au paragraphe II.2.4, à partir des caractéristiques C-V, il est possible d'extraire le dopage du substrat (N_{Sub}). Si le dopage n'est pas constant, la dérivée de l'équation (II.20) par rapport à V_{GB} permet de retrouver le profil du dopage :

$$\frac{d\left(\frac{A_{\text{eff}}^2}{C^2}\right)}{dV_{\text{GB}}} = \frac{2}{qN_{\text{sub}}\epsilon_{\text{Si}}} \quad (\text{II.37})$$

ou encore :

$$N_{\text{sub}} = \frac{2}{q\epsilon_{\text{Si}} \frac{d\left(\frac{A_{\text{eff}}^2}{C^2}\right)}{dV_{\text{GB}}}} \quad (\text{II.38})$$

Grâce aux deux relations suivantes, nous pouvons déterminer le profil du dopage du substrat de la capacité MOS à partir de sa caractéristique C-V.

$$\begin{cases} N_{\text{sub}}(x) = \frac{2}{q\epsilon_{\text{Si}} \frac{d\left(\frac{A_{\text{eff}}^2}{C(V_{\text{GB}})^2}\right)}{dV_{\text{GB}}}} \\ x(V_{\text{GB}}) = \epsilon_{\text{Si}} \left[\frac{A_{\text{eff}}^2}{C(V_{\text{GB}})} - \frac{A_{\text{eff}}^2}{C_{\text{ox}}} \right] \end{cases} \quad (\text{II.39})$$

Après avoir observé l'impact des dopages sur la capacité MOS, nous nous proposons d'étudier celui de l'épaisseur d'oxyde, t_{ox} .

II.4. Effet d'une épaisseur non constante

Afin de garantir les mêmes caractéristiques électriques des composants appartenant au même Wafer (plaquette de silicium), l'épaisseur d'isolant doit être uniforme. Cependant, durant les étapes technologiques du procédé de fabrication du wafer, des non uniformités d'épaisseur d'oxyde peuvent apparaître. Les origines de ces non uniformités sont multiples avec par exemple : la rugosité de surface du silicium, un défaut du masque, ou encore la position du wafer lors des recuits avec une température non constante dans les fours, etc... Par conséquent, nous nous sommes intéressés à l'impact d'une non uniformité d'épaisseur d'oxyde sur les caractéristiques électriques de la capacité.

II.4.1. Modélisation pseudo 2D des courbes C-V et I-V

Nous avons réalisé un simulateur pseudo 2D prenant en compte d'éventuelles variations d'épaisseur d'oxyde d'une capacité MOS lors de la détermination de ces caractéristiques C-V et I-V. Ce simulateur est basé sur le découpage de la capacité

originelle en N capacités MOS élémentaires d'épaisseur d'isolant constante en supposant qu'elles ne s'influencent pas entre elles (cf. Fig. (II.14)).

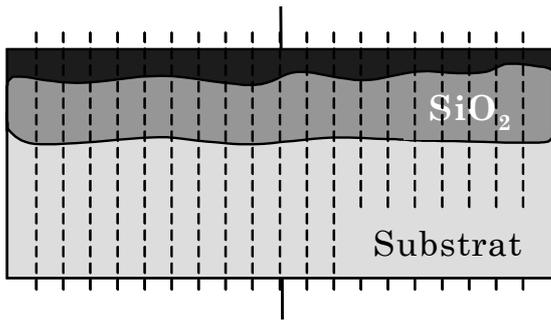


Figure II.14. Capacité MOS à épaisseur d'isolant de grille non constante.

Pour n'étudier que l'influence de l'épaisseur d'oxyde, on suppose :

- qu'il n'y a pas d'état d'interface ($Q_{it} = 0$),
- qu'il n'y a pas de charge fixe ($Q_{ox} = 0$),
- que les dopages sont uniformes.

II.4.1.1. Modélisation des courbes C-V

Soit V_{GB} la tension appliquée sur la grille, le potentiel de surface de chaque capacité élémentaire est déterminé le long de l'interface Si/SiO₂ à l'aide de l'équation liant la polarisation appliquée sur la grille, V_{GB} , et le potentiel de surface, Ψ_S (cf. Eq. (I.6)) :

$$V_{GB} = V_{FB} + \Psi_{Si} - \frac{Q_{SCi}}{C_{oxi}} \quad (\text{II.40})$$

où C_{oxi} et Q_{SCi} correspondent respectivement à la capacité d'oxyde et à la charge du substrat de la $i^{\text{ème}}$ capacité élémentaire (cf. Eq. (I.14)).

Connaissant Q_{SCi} , on détermine la capacité correspondante, C_{SCi} , à partir de l'équation (I.2). Chaque capacité élémentaire, C_i (de surface effective A_{effi}), correspondant à la mise en série des capacités élémentaires de l'isolant, C_{oxi} , et du semiconducteur C_{SCi} , est alors calculée à partir de l'équation (I.1). Ainsi notre simulateur détermine les caractéristiques C-V de chacune des N capacités élémentaires. Pour calculer la capacité totale de la structure, il suffit alors de faire la somme des capacités élémentaires placées en parallèle :

$$C = \sum_i C_i \quad (\text{II.41})$$

I.4.1.2. Modélisation des courbes I-V

En ce qui concerne le calcul du courant de la structure totale, le potentiel (cf. Eq. (II.42)) et le champ électrique (cf. Eq. (II.43)) aux bornes de l'isolant de chaque capacité élémentaire doivent être déterminés :

$$V_{\text{oxi}} = \frac{-Q_{\text{SCi}}}{C_{\text{oxi}}} \quad (\text{II.42})$$

$$\xi_{\text{oxi}} = \frac{-V_{\text{oxi}}}{t_{\text{oxi}}} \quad (\text{II.43})$$

Dans le cas d'un courant tunnel Fowler-Nordheim, le champ élevé à travers l'oxyde abaisse la largeur de la barrière de potentiel et permet aux électrons de passer par effet tunnel. On observe cet effet pour des valeurs, ξ_{oxi} , du champ électrique dans l'oxyde tels que $qV_{\text{oxi}} > \Phi_{\text{bi}}$, (c.a.d. $q\xi_{\text{oxi}} > \Phi_{\text{bi}}/t_{\text{oxi}}$) où Φ_{bi} est la barrière à l'interface oxyde/semiconducteur pour la $i^{\text{ème}}$ capacité élémentaire (différence d'énergie entre le niveau d'énergie considéré du semiconducteur et le bas de la bande de conduction de l'oxyde).

Le courant tunnel de chaque capacité élémentaire est alors donné par la relation:

$$I_{\text{tun}_i} = \begin{cases} A_{\text{effi}} \frac{q^2 m_e \xi_{\text{oxi}}^2}{8\pi h m_{\text{ox}} \phi_{\text{bi}}} \frac{1}{\left(1 - \sqrt{\frac{\Phi_{\text{bi}} \mp V_{\text{oxi}}}{\phi_{\text{bi}}}}\right)^2} \exp\left[\frac{-\gamma}{\pm \xi_{\text{oxi}}} \left[-(\phi_{\text{bi}})^{\frac{3}{2}} + (\phi_{\text{bi}} \mp V_{\text{oxi}})^{\frac{3}{2}}\right]\right] & \text{si } \phi_{\text{bi}} > \pm V_{\text{oxi}} \\ A_{\text{effi}} \frac{q^2 m_e \xi_{\text{oxi}}^2}{8\pi h m_{\text{ox}} \phi_{\text{bi}}} \exp\left[\frac{\gamma (\phi_{\text{bi}})^{\frac{3}{2}}}{\pm \xi_{\text{oxi}}}\right] & \text{si } \phi_{\text{bi}} \leq \mp V_{\text{oxi}} \end{cases} \quad (\text{II.44})$$

où $\gamma = \frac{4}{3\hbar} \sqrt{2q m_{\text{si}} m_0}$, m_e est la masse effective de l'électron, $m_{\text{si}}=0.327$, $m_0 = 9.1 \times 10^{-31} \text{kg}$ et m_{ox} est la masse effective de l'électron dans l'oxyde (en général on prend $m_{\text{ox}} \approx 0.5 \times m_0$ et $\Phi_b \approx 3.05 \text{ eV}$). Les signes \pm et \mp sont dus aux deux passages tunnels possibles des électrons du substrat vers la grille ($V_{\text{oxi}} > 0$) et de la grille vers le substrat ($V_{\text{oxi}} < 0$).

Finalement, la somme des courants tunnel correspondant aux capacités élémentaires est égale au courant tunnel de la structure totale :

$$I = \sum_i I_{\text{tun}_i} \quad (\text{II.45})$$

II.4.1.3. Simulations des courbes C-V et I-V

Considérons une capacité de surface $W \times L$ dont l'épaisseur d'isolant varie linéairement suivant la longueur L (cf. Fig. (II.15.a)). La figure (II.15.b) présente les courbes C-V (par unité de surface) pour les capacités élémentaires ayant les épaisseurs extrêmes et pour la capacité de la structure globale.

A partir de la courbe C-V de la structure globale, nous avons extrait les paramètres caractéristiques (dopages et épaisseur d'oxyde) de cette structure afin de les comparer aux valeurs implémentées pour la simuler. Nous avons observé que :

- Les dopages extraits à l'aide de la courbe C-V de la structure totale, restent inchangés par rapport aux dopages implémentés dans le simulateur.
- L'épaisseur d'isolant de notre structure extraite à partir de la méthode de Maserjean [Maserjean'74] est de 5.46 nm ce qui est légèrement inférieur à la valeur moyenne implémentée.

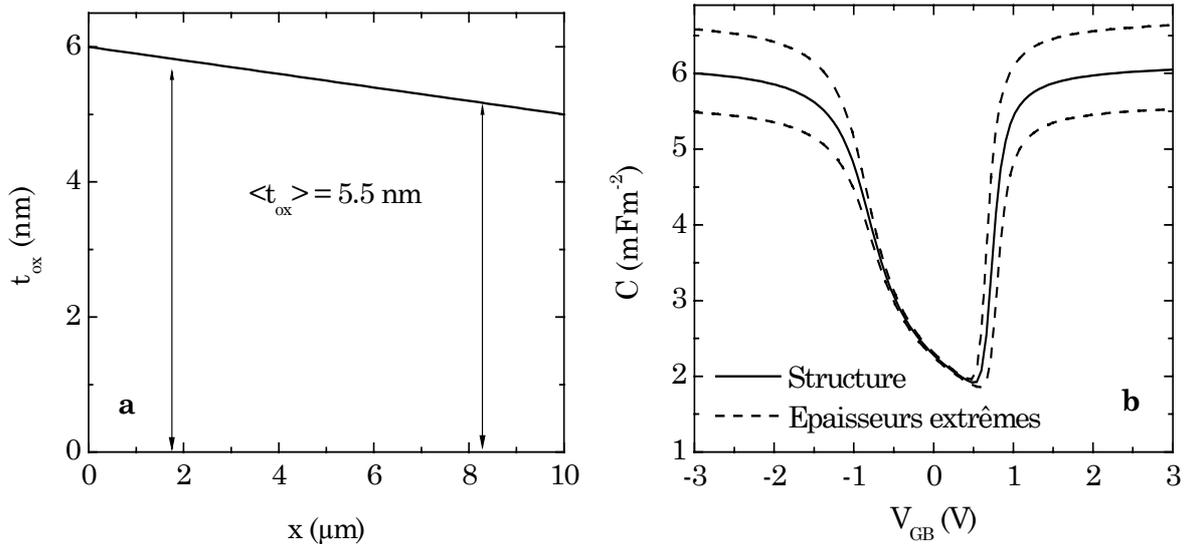


Figure II. 15. Variation de l'épaisseur d'isolant (suivant L) d'une capacité MOS (a) et courbes C-V correspondantes (b). Les paramètres de la simulation sont : $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $V_{FB} = -1 \text{ V}$, $W = L = 10 \text{ } \mu\text{m}$.

Cette comparaison a été réalisée pour différentes valeurs de t_{ox} moyen. La figure (II.16) présente les tracés de t_{ox} moyen et t_{ox} extrait en fonction de t_{ox} moyen. Ce graphe montre que l'épaisseur extraite est toujours inférieure à t_{ox} moyen.

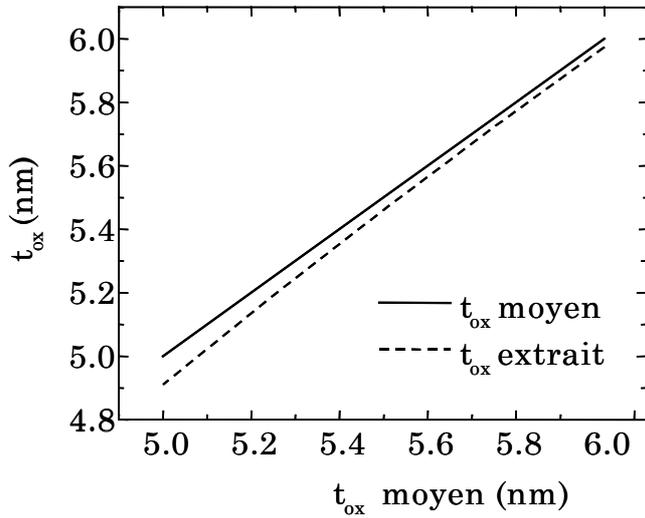


Figure II.16. Tracé de la variation de t_{ox} moyen (trait plein) et de t_{ox} extrait (pointillés) en fonction de t_{ox} moyen.

En ce qui concerne les courbes I-V, la figure (II.17.a) présente les courants d'électrons (par unité de surface) qui traversent les capacités élémentaires ayant des épaisseurs extrêmes et le courant total de la structure. Il est possible de simuler le courant tunnel de la capacité MOS si on suppose que son épaisseur est égale à l'épaisseur extraite. La figure (II.17.b) montre que dans ce cas, le courant est sous-estimé par rapport au courant de la structure.

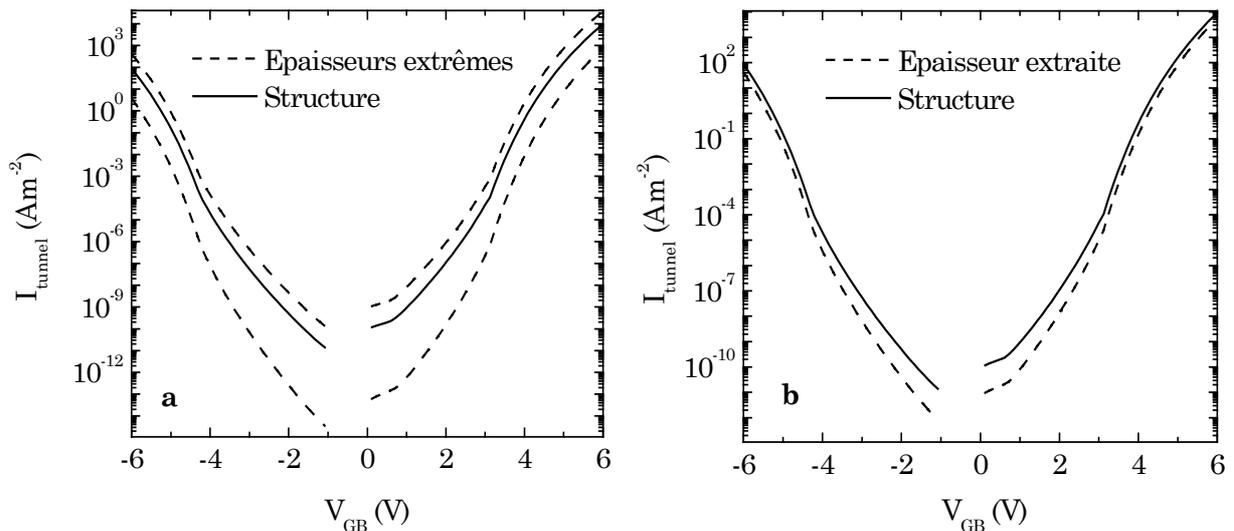


Figure II.17. Courants tunnels (par unité de surface) des capacités élémentaires (pour les épaisseurs extrêmes) et de la structure (a). Comparaison du courant tunnel de la structure avec celui simulé à partir de l'épaisseur d'isolant extraite (b).

En résumé, lorsqu'une capacité présente une non uniformité de son épaisseur d'isolant, l'épaisseur extraite des courbes C-V est toujours inférieure à t_{ox} moyen et les paramètres Fowler-Nordheim extraits des courbes I-V seront erronés. Notons que ce modèle est un modèle pseudo 2D puisque nous supposons que chaque capacité n'influence pas ces voisines. Cependant, ce modèle simplifié n'est plus valable en

présence d'une brusque marche d'escalier qui induit alors un champ distordu. Dans ce cas, l'utilisation d'un simulateur 2D type ISE ou silvaco, est nécessaire. Cependant, la création de son propre simulateur 2D permet une plus grande souplesse pour la modélisation physique et l'implantation des lois de conduction ou certains mécanismes physiques.

II.4.2. Comparaison entre la modélisation pseudo 2D et 2D

L'influence des capacités voisines nécessite la modélisation 2D du potentiel de surface. Certes il existe des simulateurs 2D comme ISE ou silvaco, mais nous avons choisi de faire notre propre simulateur 2D. Ainsi, il nous est possible d'implanter des lois de conduction et certains mécanismes physiques. Ce simulateur réalisé par Laurent Raymond, membre du laboratoire, est basé sur la méthode des éléments finis avec un maillage triangulaire et adaptatif permettant un resserrement à l'interface et aux points anguleux. Ce maillage est obtenu à l'aide d'un logiciel du domaine public (www.geuz.org) avec un code C++ compatible avec notre solveur mathcad.

Soit une capacité avec une épaisseur d'oxyde, t_{ox} , non uniforme, observons la répartition du potentiel du silicium vers la grille (cf. Fig. (II.18)). La grille n'est pas représentée, mais se trouverait en haut de cette figure. La zone supérieure (bleue) représente le diélectrique SiO_2 avec une interface grille/ SiO_2 supposée lisse. La zone noire irrégulière correspond à une interface rugueuse SiO_2 /semiconducteur. Cette zone dense noire est due à l'affinement du maillage. En effet, lorsque l'interface est non uniforme, le maillage est resserré pour tenir compte des changements d'orientation de l'interface. De plus, lorsque la tension appliquée entre la grille et le substrat augmente, la courbure de bande est de plus en plus importante ce qui correspond à une chute de potentiel. L'échelle des potentiels est représentée par une gamme de couleur allant du bleu pour les faibles potentiels au rouge pour les forts potentiels. Les figures (II.18) montrent l'influence de l'augmentation du potentiel appliqué entre la grille et le substrat avec la diminution du niveau de Fermi de la grille et la courbure du diagramme de bandes. Ainsi, la structure simulée fonctionne.

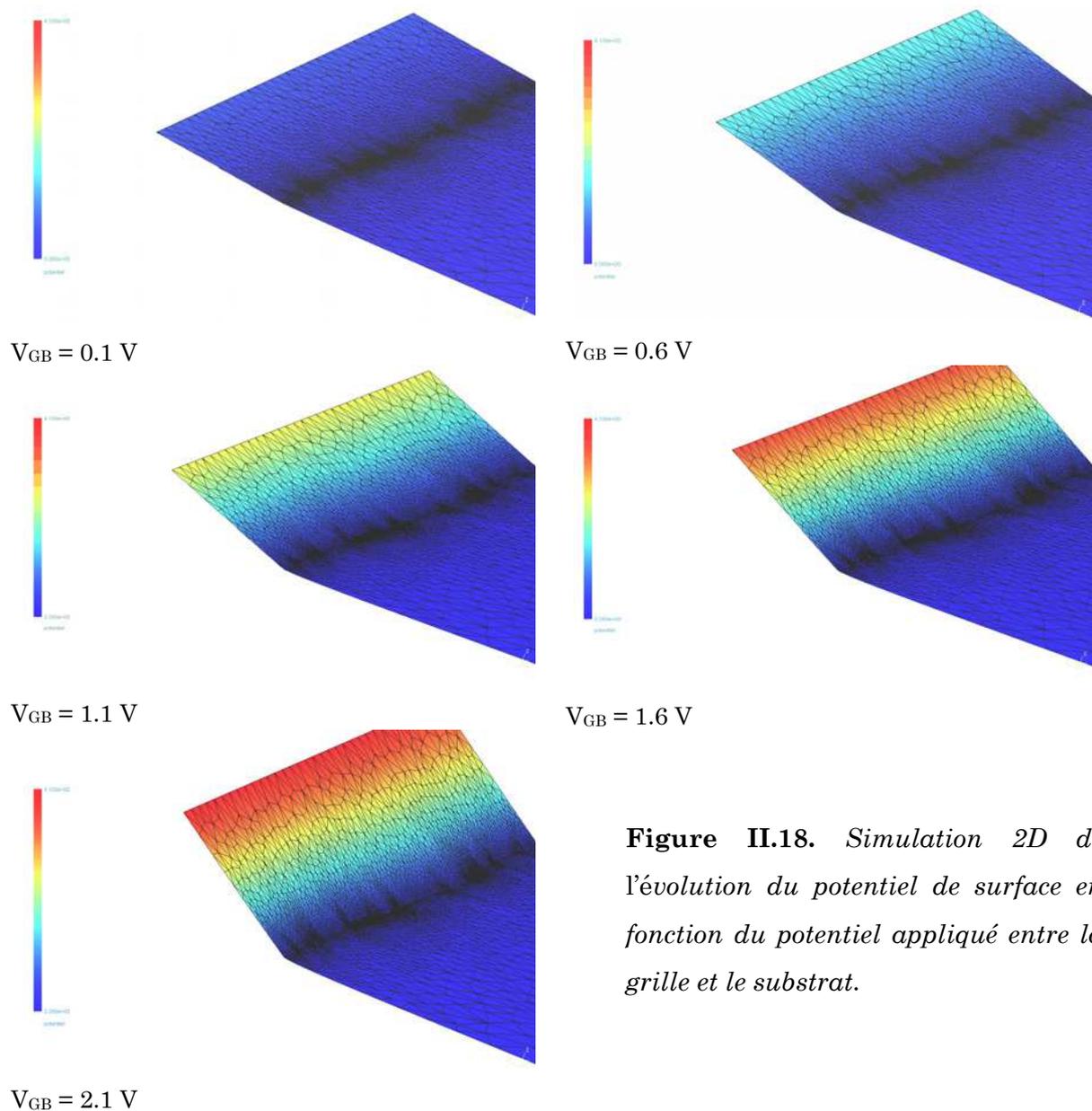


Figure II.18. *Simulation 2D de l'évolution du potentiel de surface en fonction du potentiel appliqué entre la grille et le substrat.*

Ce simulateur permet également la représentation des lignes équipotentielles qui sont perpendiculaires aux lignes de champs. La figure (II.19.a) présente ces équipotentielles avec en noir le maillage; cependant, il est difficile d'observer clairement les lignes de champs sauf au niveau du silicium. La figure (II.19.b) est un zoom de la figure (II.19.a), au niveau de l'interface Si/SiO₂ pour une des non uniformité de t_{ox} en forme de cuvette renversée. Le maillage est resserré au niveau de cette cuvette, surtout au point anguleux (passage de marche). Les équipotentielles sont des droites lorsque l'interface est uniforme, mais sont déformées au niveau de la non uniformité. Ainsi, on peut observer des lignes qui entrent dans le silicium puis en ressortent, ce qui souligne l'influence des capacités voisines au niveau de la non uniformité et donc la nécessité de faire une simulation 2D. De plus, on sait que les électrons suivent les lignes de champ qui sont perpendiculaires aux équipotentielles.

Ainsi, on distinguera deux zones :

- Les zones où il n'y a pas de perturbation c.a.d. avant et après la non uniformité. Dans ce cas, les électrons auront un chemin rectiligne à travers le diélectrique.
- Les zones de non uniformité, où les électrons auront un chemin distordu afin de rester perpendiculaire à l'interface. Dans ce cas, pour le calcul des courants, des tubes de conduction, correspondant aux chemins des électrons du substrat vers la grille, devront être calculés.

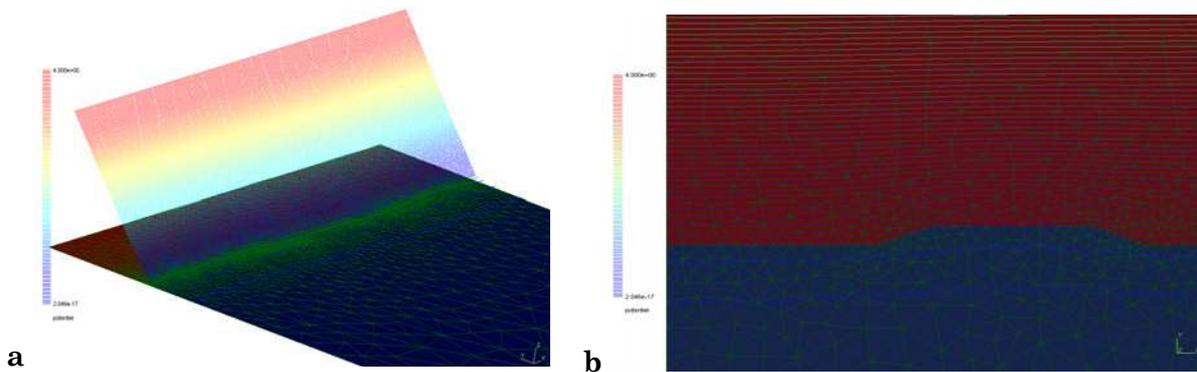


Figure II.19. Représentation des lignes équipotentielles (a) et zoom sur une des non uniformités de l'interface Substrat/SiO₂ (b). Données : $N_A = 10^{24} \text{ m}^{-3}$, $t_{ox} = 7 \text{ nm}$ et la marche constituant la non uniformité est de 2 nm.

Afin de vérifier la validité de notre approche pseudo 2D, nous avons comparé les caractéristiques C-V obtenues avec celles simulées à partir du modèle 2D (c.a.d. du calcul de l'équation de Poisson suivant deux dimensions en chaque nœud du maillage). La figure (II.20) donne un exemple de maillage d'une capacité MOS ayant un oxyde très rugueux d'épaisseur moyenne 3 nm.

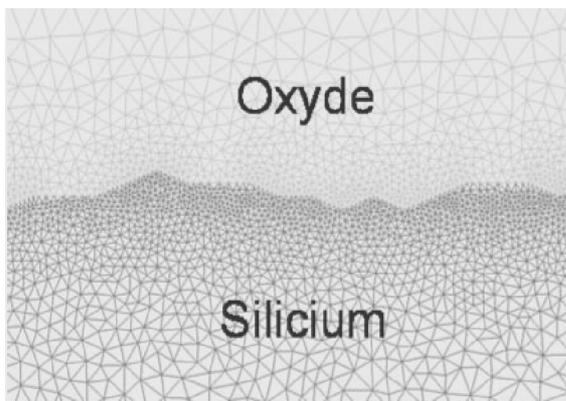


Figure II.20. Exemple d'un maillage 2D de la structure par la méthode des éléments finis. Ce maillage comporte environ 15 000 points [Cuinet'04].

On remarquera la forme particulière de ce maillage qui est triangulaire et de répartition aléatoire. Cette particularité permet d'éviter des effets parasites de symétrie lors du calcul de l'équation de Poisson. On remarquera aussi que le maillage est plus resserré à l'interface oxyde /silicium pour garantir une plus grande précision, et donc une

plus grande fiabilité des résultats. La figure (II.21.a) présente les trois profils de l'épaisseur d'oxyde choisis pour cette comparaison avec une interface SiO_2/Si plate (pointillés), une interface légèrement rugueuse (cercles) et une interface très rugueuse (ligne grise). Ces trois structures d'études ont la même épaisseur d'oxyde moyenne égale à 3 nm. Cependant la figure (II.21.b) met en évidence l'augmentation de la capacité due à la dépendance du chemin des lignes de champ avec la rugosité de l'interface SiO_2/Si .

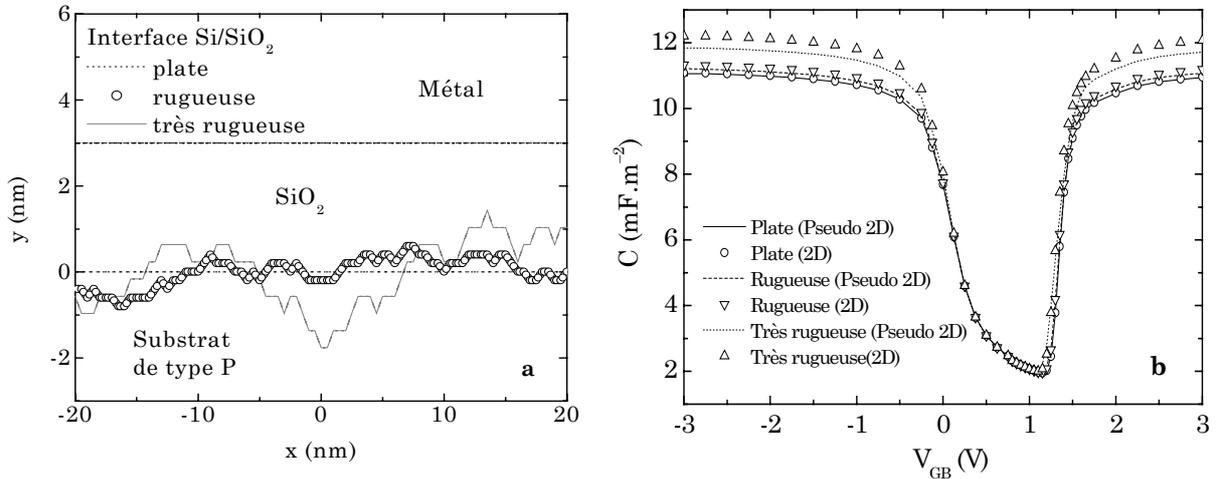


Figure II.21. Profils des interfaces Si/SiO_2 choisis (a) pour étudier l'impact de la rugosité sur les caractéristiques C-V simulées à l'aide de notre modèle pseudo 2D et du modèle 2D (b). Données : $N_A = 5 \cdot 10^{23} \text{ m}^{-3}$, $t_{\text{oxmoy}} = 3 \text{ nm}$.

De surcroît, la comparaison des caractéristiques C-V de la figure (II.21.b) valide notre approche pseudo 2D pour le calcul des capacités dans le cas d'interface non rugueuse ou peu rugueuse. Cette comparaison montre que lorsque la rugosité augmente, les effets 2D ne sont plus négligeables et le modèle pseudo 2D tend à sous estimer la valeur de la capacité. Cependant, notre modèle pseudo 2D permet l'obtention de bonnes caractéristiques C-V au premier ordre et présente un très grand intérêt en terme de rapidité de calcul par rapport au modèle 2D. La même comparaison pour les caractéristiques I-V de ces structures est en cours de réalisation avec le calcul des tubes de conduction correspondant aux chemins des électrons du substrat vers la grille.

Après s'être intéressés aux variations des dopages et de l'épaisseur d'oxyde d'une capacité, nous nous proposons d'étudier l'impact des non uniformités des charges fixes présentes dans l'oxyde d'une capacité. Deux études sont possibles : l'une dans le cas de charges fixes non constantes entre la source et le drain et l'autre dans le cas de charges fixes non uniformément réparties dans le volume de l'isolant.

II.5. Charges fixes latérales non constantes

La présence de défauts dans l'isolant d'une structure Métal/Isolant/Semiconducteur (MIS) entraîne la variation de la tension de seuil V_T ou de la tension de bandes plates V_{FB} , l'augmentation de la pente sous le seuil et du niveau de bruit du courant (courbes $I_{DS}(V_{GS})$), et la déformation de la relation capacité-tension C-V. D'un point de vue électrique, les charges piégées dans les états d'interface ou sur les défauts d'oxyde proches de l'interface sont mesurables car elles évoluent au cours de la mesure, tandis que les charges fixes d'oxyde n'évoluent pas. Nous nous sommes intéressés à ces charges fixes, notées Q_{ox} , dont la présence le long de l'interface isolant / substrat entraîne une translation de la courbe C-V : translation avec $\Delta V_{FB} < 0$ lorsque $Q_{ox} > 0$ et translation avec $\Delta V_{FB} > 0$ lorsque $Q_{ox} < 0$.

II.5.1. Modélisation des courbes C-V et I-V

Dans cette partie, nous nous intéressons à l'effet de la non uniformité des charges fixes, Q_{ox} , vues de l'interface, sur les caractéristiques I-V et C-V d'une capacité MOS. Q_{ox} sera donc une charge surfacique exprimée en Cm^{-2} . Comme pour le cas de la non uniformité d'épaisseur d'isolant, le simulateur que nous avons réalisé, est basé sur le découpage de la capacité MOS en N capacités élémentaires placées en parallèle (cf. Fig. (II.22)).

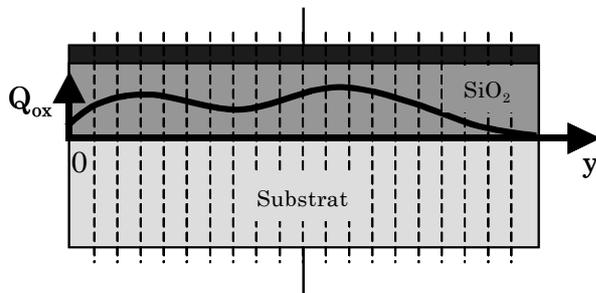


Figure II.22. Capacité MOS à densité de charges fixes non uniforme (vue de l'interface).

La charge fixe Q_{oxi} de chaque capacité élémentaire est considérée comme constante. L'équation (I.7) permet la prise en compte de cette charge Q_{oxi} , dans le calcul du potentiel de surface. Les caractéristiques C-V de chacune de ces capacités sont ensuite évaluées comme précédemment à l'aide de l'équation (I.14). A nouveau, on suppose que les capacités ne s'influencent pas entre elles (modélisation pseudo 2D). Les caractéristiques de la structure globale sont alors déterminées à l'aide de l'équation (II.41). Les courants tunnels sont déterminés de la même manière que dans le cas précédent (§ II.4.1.2.).

II.5.2. Impact de la non uniformité de charges

II.5.2.1. Impact sur les courbes C-V

Pour illustrer l'utilisation de ce simulateur, considérons une capacité MOS de surface $W \times L$ dont la charge fixe varie linéairement suivant sa longueur L comme l'indique la figure (II.23.a). La figure (II.23.b) montre les courbes C-V de la structure et des capacités élémentaires (par unité de surface) présentant les densités de charges fixes extrêmes. Malgré la déformation de la courbe C-V de la structure par rapport à une courbe C-V classique, l'épaisseur d'oxyde extraite correspond à celle implémentée et le dopage du substrat extrait à une valeur de 6 % inférieure à celle implémentée.

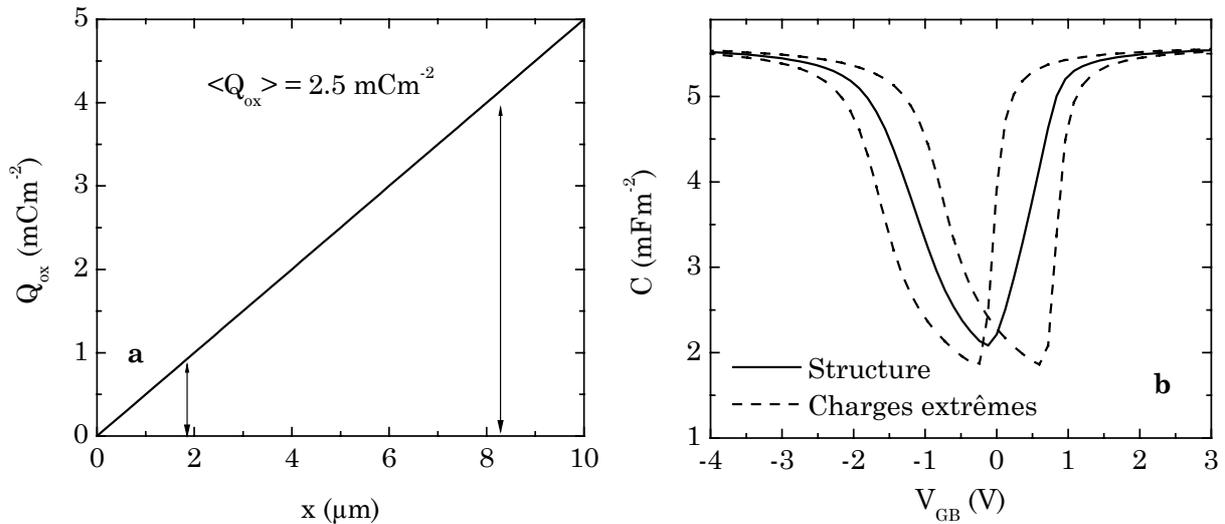


Figure II.23. Variation de la charge fixe (suivant L) d'une capacité MOS (a) et courbes C-V correspondantes (b). Les paramètres de la simulation sont : $t_{ox} = 3 \text{ nm}$, $V_{FB} = -1 \text{ V}$, $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $W = L = 10 \text{ } \mu\text{m}$.

II.5.2.2. Impact sur les courbes I-V

La figure (II.24.a) donne les courbes I-V extrêmes de la structure (par unité de surface) ainsi que le courant de la structure totale. A partir de la détermination de la tension de bandes plates et de la connaissance de la courbe C-V de la capacité ainsi que du dopage du substrat, il est possible de déterminer la charge fixe effective présente dans la structure. On peut alors déterminer le courant qui traverse par effet tunnel ces structures possédant ces charges fixes parasites. La figure (II.24.b) montre que ce courant est inférieur au courant de la structure.

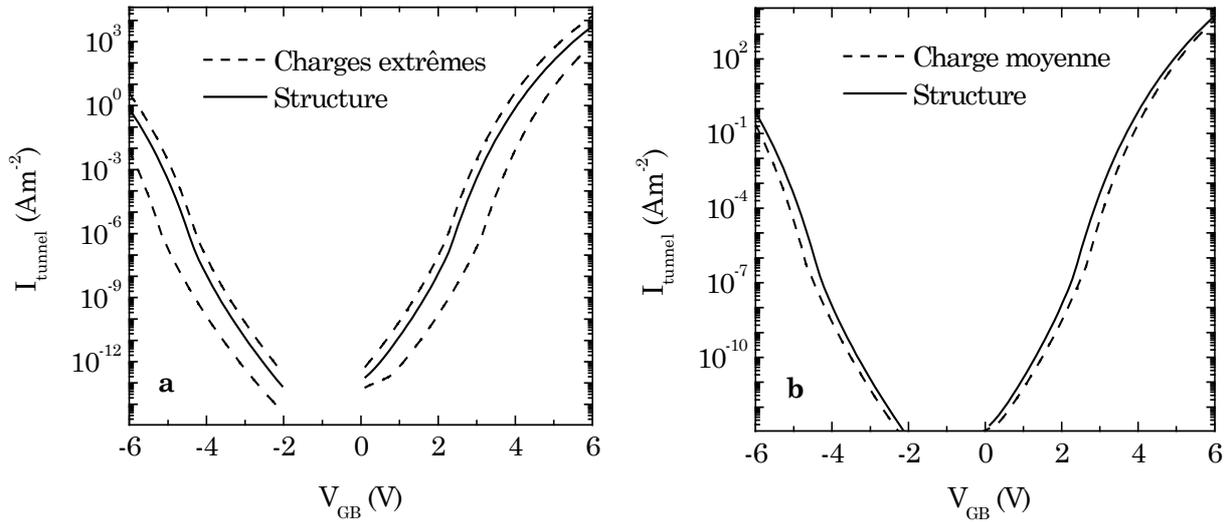


Figure II.24. Courants tunnels des capacités élémentaires (pour les charges fixes extrêmes) et de la structure (a). Comparaison du courant tunnel de la structure avec celui simulé à partir de la charge effective extraite (b).

II.6. Charges fixes non uniformément réparties dans l'isolant

Dans cette partie, nous nous intéressons à l'impact de la non uniformité des charges fixes entre le substrat et la grille, sur les caractéristiques I-V et C-V d'une capacité MIS. L'isolant peut être constitué d'une couche de SiO_2 ou d'un empilement de deux couches formées par du SiO_2 et un matériau à haute constante diélectrique (high κ) comme l'oxyde d'hafnium (HfO_2). Notre approche s'inscrit dans la continuité des travaux de Ku *et al.* [Ku'94] dans lesquels la résolution de Poisson a été réalisée pour le cas d'une charge fixe répartie uniformément dans l'oxyde.

II.6.1. Modélisation de la courbe I-V d'une capacité MOS

Considérons une capacité MOS dont la grille et le substrat sont très fortement dopés (avec des dopages uniformes). Notre simulateur repose sur le découpage de l'oxyde en différentes tranches de même largeur (cf. Fig. (II.25)).

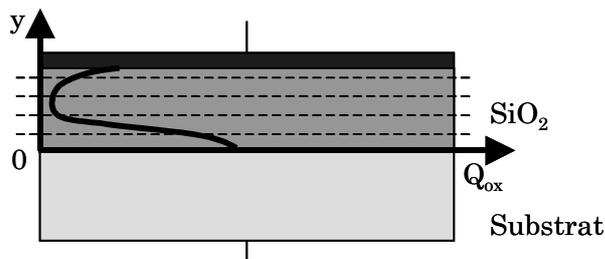


Figure II.25. Capacité MOS présentant une charge non uniforme dans l'isolant de grille, formé par du SiO_2 .

L'équation au potentiel avec ou sans la déplétion de la grille (I.24) peut être écrite en prenant en compte la charge dans l'oxyde :

$$V_{GB} = \Phi_{MS} - \frac{Q_{oxeff}}{C_{ox}} + \Psi_S - \Psi_{SG} - \frac{Q_{SC}(\Psi_S)}{C_{ox}} \quad (II.46)$$

où la charge de l'isolant, Q_{oxeff} , est une charge surfacique. Comme la grille est très fortement dopée, on peut négliger les effets de la charge sur l'interface oxyde/grille. Par suite, la charge volumique, $Q_{ox}(y)$ peut se ramener à une charge surfacique, Q_{oxeff} , vue de l'interface substrat/diélectrique :

$$Q_{oxeff} = \int_0^{t_{ox}} \frac{t_{ox} - y}{t_{ox}} Q_{ox}(y) dy \quad (II.47)$$

Pour un potentiel de grille donné, la résolution de l'équation (II.46) permet la détermination du potentiel de surface et donc du potentiel aux bornes de l'oxyde :

$$V_{ox} = - \frac{Q_{SC} + Q_{oxeff}}{C_{ox}} \quad (II.48)$$

On obtient alors la distribution spatiale du potentiel dans l'oxyde, $V(y)$, à partir de la résolution de l'équation de Poisson (par la méthode des différences finies), en considérant la charge volumique Q_{ox} (et non la charge surfacique Q_{oxeff}) et les conditions aux limites suivantes :

$$\begin{cases} V(y=0) = 0 \\ V(y=t_{ox}) = V_{ox} \end{cases} \quad (II.49)$$

La variation de la barrière de potentiel vue par les électrons du substrat s'exprime alors par la relation suivante :

$$\Phi_{si} = \begin{cases} \Phi_b - V(y) & \text{si } \Phi_b > V(y) \\ 0 & \text{sinon} \end{cases} \quad (II.50)$$

La variation de la barrière de potentiel vue par les électrons de la grille est donnée par :

$$\Phi_{poly} = \begin{cases} \Phi_b - V(y) + V_{ox} & \text{si } \Phi_b > V(y) - V_{ox} \\ 0 & \text{sinon} \end{cases} \quad (II.51)$$

Ainsi, la transparence des électrons en fonction de leur électrode est déterminée à partir de l'expression [Fromhold'81] :

$$T_{Esi,poly} = \exp \left(-2 \sqrt{\frac{2q m_{nox}}{\hbar^2}} \int_0^{t_{ox}} \sqrt{\Phi_{si,poly}} dy \right) \quad (II.52)$$

Par conséquent, on peut déterminer le courant des électrons au niveau du substrat et de la grille :

$$I_{si,poly} = A_{eff} Q_{nsi,poly} F_{imp} T_{Esi,poly} \quad (II.53)$$

où F_{imp} représente la fréquence d'impact des électrons sur la barrière de potentiel et A_{eff} la surface effective de la capacité.

II.6.2. Modélisation de la courbe I-V d'une capacité MIS

Pour modéliser une capacité dont l'isolant est formé par un empilement SiO_2/HfO_2 , nous avons apporté quelques modifications au modèle précédent (§ II.6.1.).

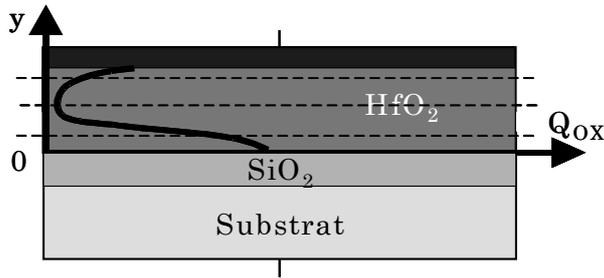


Figure II.26. Schématisation d'une capacité MOS présentant une charge non uniforme dans l'isolant de grille, formé par un empilement SiO_2/HfO_2 .

L'ajout d'une couche de diélectrique entraîne la présence d'une nouvelle capacité (cf. Fig. (II.26)). Ainsi l'expression de la capacité d'oxyde devient :

$$C_{ox} = \frac{C_{ox1} C_{ox2}}{C_{ox1} + C_{ox2}} \quad (II.54)$$

Notons que toute la charge générée dans l'oxyde est supposée être répartie uniquement dans la couche d'hafnium [Garros'03]. Afin de considérer toutes les couches de diélectriques comme équivalentes à une seule couche de SiO_2 , une épaisseur d'isolant équivalente doit être calculée pour la couche de HfO_2 :

$$t_{ox2eq} = \frac{\epsilon_{SiO_2}}{\epsilon_{HfO_2}} t_{ox2} \quad (II.55)$$

$$t_{oxeq} = t_{ox2eq} + t_{ox1} \quad (II.56)$$

En supposant que toutes les charges fixes sont situées à l'intérieur de la couche de HfO_2 , la densité de charge effective vue de l'interface entre le substrat et le diélectrique est donnée par la relation suivante :

$$Q_{oxeff} = \int_0^{t_{ox}} \frac{t_{ox} - t_{ox2eq}}{t_{ox}} Q_{oxeq}(y) dy \quad (II.57)$$

où :

$$Q_{\text{oxeq}} = Q_{\text{ox}} \frac{\varepsilon_{\text{HFO}_2}}{\varepsilon_{\text{SiO}_2}} \quad (\text{II.58})$$

Une dernière modification doit également être apportée pour le calcul du potentiel de surface. En effet, la présence des deux couches d'isolant, implique le calcul du potentiel aux bornes de chaque isolant. Considérons la structure représentée sur la figure (II.27), où le sens positif correspond à l'axe des tensions.

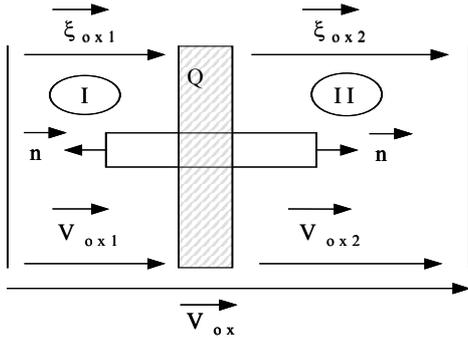


Figure II.27. Schématisation d'une structure avec un isolant constitué de deux matériaux différents (I) et (II), et une charge \$Q\$ piégée à l'interface de ces deux matériaux.

Le théorème de Gauss appliqué à un cylindre perpendiculaire à l'interface entre les deux matériaux permet d'écrire la relation :

$$-\varepsilon_1 \xi_{\text{ox}1} + \varepsilon_2 \xi_{\text{ox}2} = \frac{Q}{\varepsilon_0} \quad (\text{II.59})$$

où \$\varepsilon_1\$, \$\varepsilon_2\$, \$\xi_{\text{ox}1}\$ et \$\xi_{\text{ox}2}\$ correspondent respectivement aux permittivités relatives et aux champs électriques présents dans les matériaux (I) et (II).

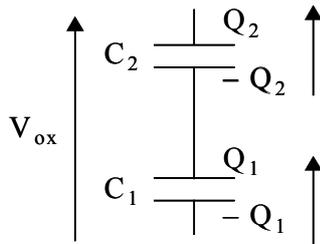


Figure II.28. Schématisation du système capacitif équivalent à la structure présentée sur la figure (II.27).

A partir du système capacitif équivalent présenté sur la figure (II.28), la charge \$Q\$ piégée à l'interface des deux matériaux s'écrit :

$$Q = -Q_2 + Q_1 \quad (\text{II.60})$$

$$\text{avec } \begin{cases} Q_1 = C_1 V_{\text{ox}1} \\ Q_2 = C_2 V_{\text{ox}2} \end{cases} \quad (\text{II.61})$$

La continuité des potentiels implique la relation :

$$V_{\text{ox}1} + V_{\text{ox}2} = V_{\text{ox}} \quad (\text{II.62})$$

A partir des équations (II.59), (II.61) et (II.62), on déduit la relation (II.63) :

$$Q = -C_2 V_{ox2} + C_1 V_{ox1} \quad (\text{II.63})$$

A partir des relations (II.61) à (II.63), on déduit les expressions des potentiels aux bornes des deux matériaux constituant l'isolant :

$$V_{ox1} = \frac{C_{ox2}}{C_{ox1} + C_{ox2}} V_{ox} + \frac{Q_{oxeff12}}{C_{ox1} + C_{ox2}} \quad (\text{II.64})$$

$$V_{ox2} = \frac{C_{ox1}}{C_{ox1} + C_{ox2}} V_{ox} - \frac{Q_{oxeff12}}{C_{ox1} + C_{ox2}} \quad (\text{II.65})$$

où $Q_{oxeff12}$ est la densité de charge effective à l'interface entre les deux diélectriques.

Enfin, la résolution spatiale du potentiel à l'intérieur de l'oxyde, $V(y)$, est obtenue comme précédemment à partir de la résolution de l'équation de Poisson (méthode des différences finies), ce qui permet la détermination de la transparence tunnel et du courant tunnel qui en découle.

II.6.3. Simulation des courbes I-V pour une capacité MOS

Considérons une densité de charges fixes uniformément répartie dans l'isolant SiO_2 d'une capacité MOS et traçons le diagramme de bandes de la structure. La figure (II.29.a) illustre la comparaison entre le diagramme de bandes de la structure obtenu avec et sans la résolution de Poisson dans l'isolant (c'est-à-dire en considérant que le champ électrique y est uniforme). Pour une polarisation positive (cf. Fig. (II.29)), la présence de charges négatives entraîne une déformation du diagramme de bandes de l'isolant visible seulement grâce à la résolution de l'équation de Poisson dans l'isolant. L'aspect bombé de la bande de conduction de l'isolant, dû aux charges fixes négatives, implique une augmentation de la barrière de potentiel à franchir par les électrons provenant du substrat. Par conséquent, la probabilité pour que ces électrons traversent la barrière de potentiel diminue. Cette probabilité de passage est aussi appelée transparence des électrons. Elle est directement reliée au courant tunnel par l'expression (II.64). Nous avons calculé cette transparence dans l'approximation WKB avec ou sans résolution de l'équation de Poisson. La figure (II.29.b) montre clairement l'impact de la déformation des bandes d'énergies qui se traduit par une diminution de la transparence des électrons (cf. Fig. II.29.b) et par suite celle du courant tunnel.

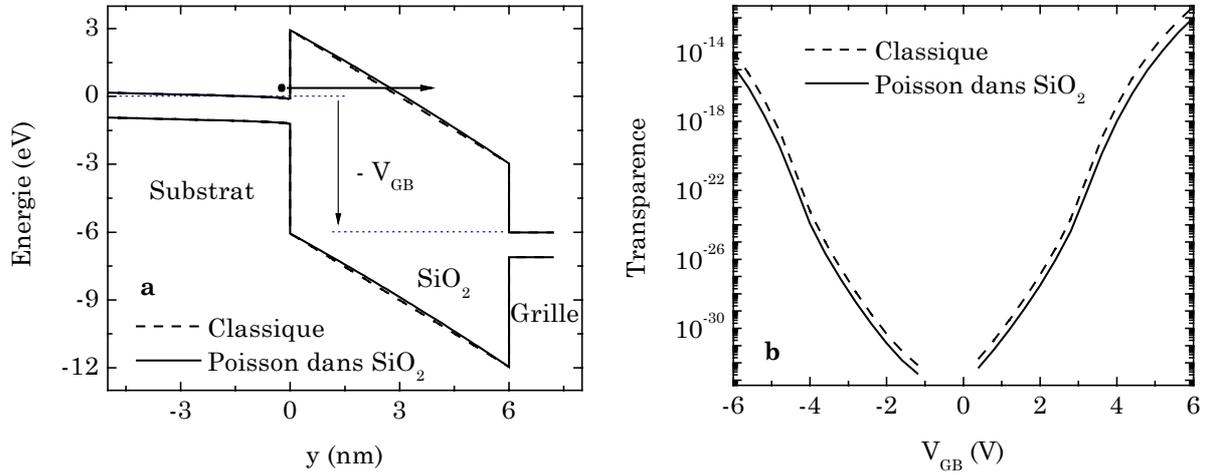


Figure II.29. Diagramme de bandes d'une capacité MOS avec ou sans résolution de l'équation de Poisson dans l'isolant (a), diminution de la transparence tunnel des électrons due aux charges fixes de l'isolant (b). Les paramètres de la simulation sont : $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $t_{ox} = 6 \text{ nm}$, $V_{FB} = -1 \text{ V}$, $W = L = 10 \mu\text{m}$, et $Q_{oxeff} = 2 \cdot 10^{16} \text{ Cm}^{-2}$.

Considérons à présent une densité de charges fixes de type gaussienne centrée dans l'isolant d'une capacité MOS à électrodes fortement dopées (cf. Fig. (II.30.a)). La présence de ces charges induit une déformation du diagramme de bandes de l'isolant que dans le cas où l'équation de Poisson est résolue (cf. Fig. (II.30.b)).

Soit une polarisation positive, la présence de charges fixes négatives bombe la bande de conduction de l'isolant ce qui augmente la largeur de la barrière de potentiel à franchir par les électrons du substrat (cf. Fig. (II.30.b)). La figure (II.30.c) montre la diminution du courant tunnel des électrons en fonction des différentes densités de charges fixes données sur la figure (II.30.a). A priori, on peut supposer que la présence de cette charge négative Q_{ox} est bénéfique puisqu'elle diminue le courant de fuite ; cependant, elle s'accompagne aussi d'une variation de la tension de bandes plates et d'une chute de la mobilité si on considère un transistor MOS et non plus une capacité MOS.

Notons que les effets néfastes de la diminution du courant tunnel ont également été observés pour les mémoires de type EEPROM avec la fermeture de la fenêtre de programmation [Papadas'92]. En effet, le chargement et déchargement de la grille flottante provoque une dégradation de l'oxyde tunnel avec notamment l'apparition de charges fixes chargées négativement. Usuellement, la diminution de l'efficacité d'injection est modélisée avec une modification des paramètres FN au cours de l'utilisation de la mémoire [Plossu'01].

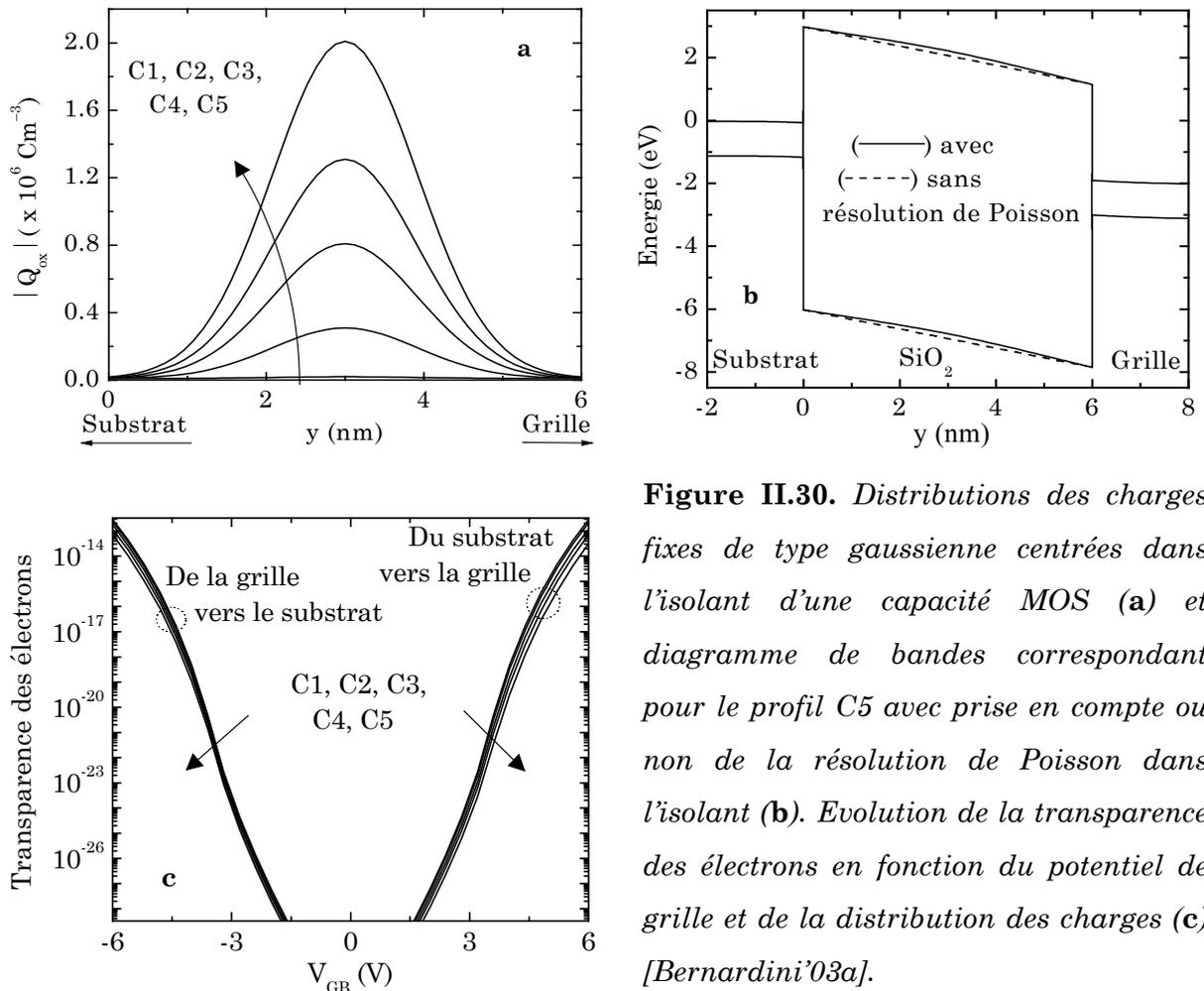


Figure II.30. Distributions des charges fixes de type gaussienne centrées dans l'isolant d'une capacité MOS (a) et diagramme de bandes correspondant pour le profil C5 avec prise en compte ou non de la résolution de Poisson dans l'isolant (b). Evolution de la transparence des électrons en fonction du potentiel de grille et de la distribution des charges (c) [Bernardini'03a].

A partir des courbes I-V et de l'expression classique du courant Fowler-Nordheim, (cf. Eq. (I.33), page 24), il est possible de déterminer les variations des paramètres caractéristiques (c.a.d. A et B) comme le montre la figure (II.31) pour les tensions de grille positives. On observe ainsi une diminution de A et une augmentation de B.

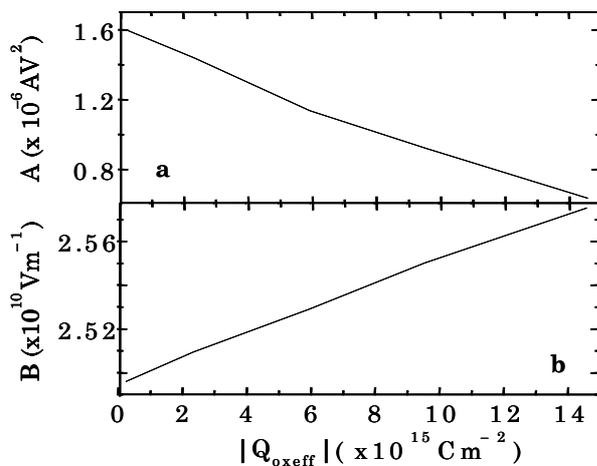


Figure II.31. Diminution du paramètre A (a) et augmentation du paramètre B (b) du courant Fowler-Nordheim (pour les tensions de grille positives) en fonction de la charge vue de l'interface avec le substrat (c.a.d. pour les profils allant de C1 à C5) [Bernardini'04a].

Il est évident que les courbes $A(Q_{\text{oxeff}})$ et $B(Q_{\text{oxeff}})$ données sur la figure (II.31) ne sont valables que pour les profils de charges que nous avons choisis dans ce paragraphe

(gaussiennes centrées) ; celles-ci seront modifiées si l'on suppose un profil uniforme, en exponentiel ou aléatoire.

II.6.4. Simulation des courbes I-V pour une capacité MIS

Dans la suite de ce paragraphe, nous supposons que toute la charge parasite du diélectrique est localisée dans la couches de HfO_2 (cf. Fig. (II.32.a)) [Houssa'00]. Comme dans le cas précédent (isolant formé d'une seule couche de diélectrique), la charge fixe piégée dans le diélectrique induit une déformation du diagramme de bandes. Cette déformation n'est visible que par la résolution de l'équation de Poisson à l'intérieur du diélectrique (cf. Fig. (II.32.b)). Il en résulte une augmentation de la largeur de la barrière à traverser par les électrons (cf. Fig. (II.32.c)) et par conséquent une diminution de la transparence tunnel et du courant tunnel qui en découle.

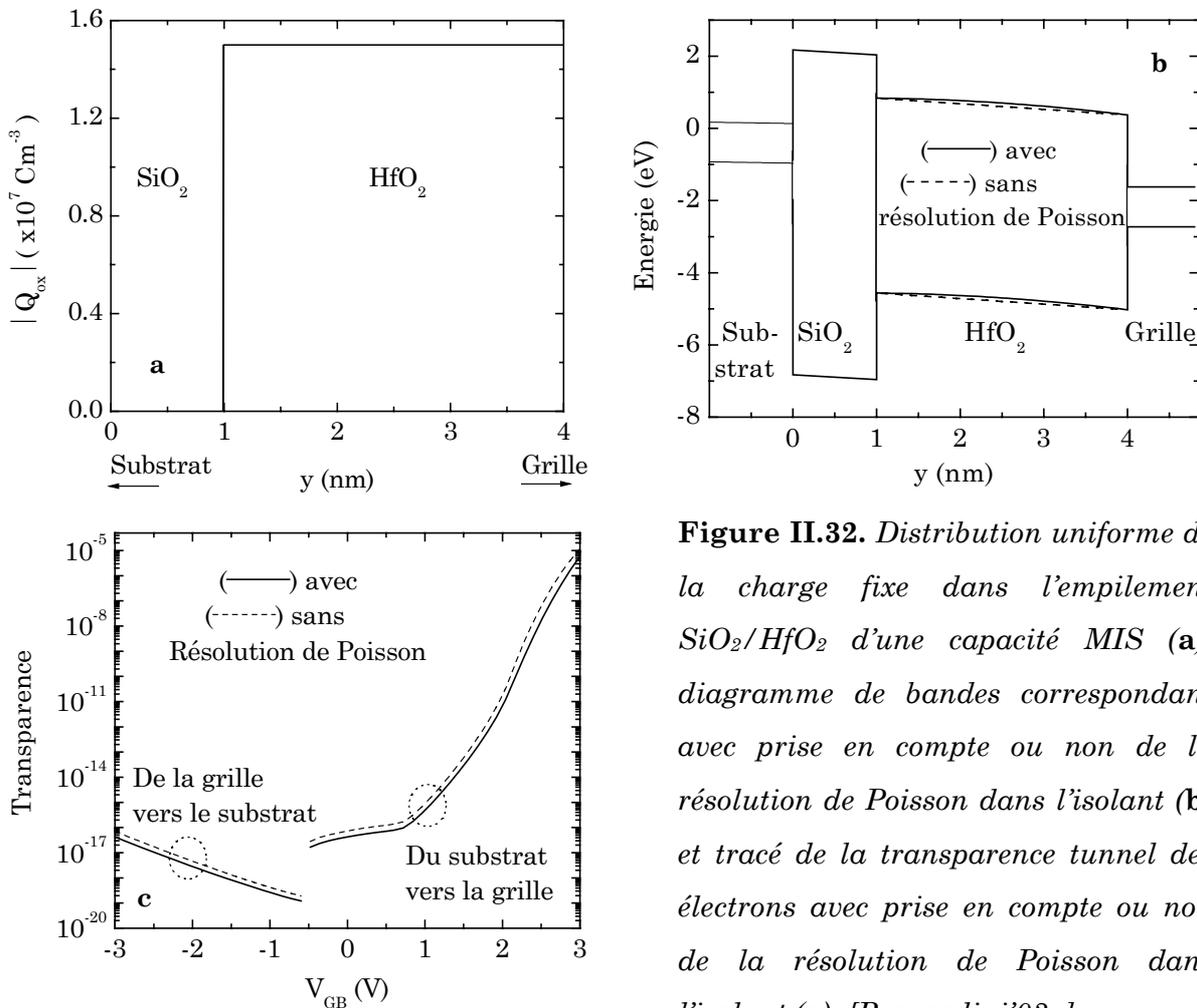


Figure II.32. Distribution uniforme de la charge fixe dans l'empilement $\text{SiO}_2/\text{HfO}_2$ d'une capacité MIS (a), diagramme de bandes correspondant avec prise en compte ou non de la résolution de Poisson dans l'isolant (b) et tracé de la transparence tunnel des électrons avec prise en compte ou non de la résolution de Poisson dans l'isolant (c), [Bernardini'03a].

On peut aussi remarquer que la transparence tunnel est plus élevée pour l'injection à partir du substrat que de la grille. Ce phénomène est dû à la dissymétrie du diagramme de bandes de la structure [Yang'98].

Considérons à présent une situation plus réaliste avec une distribution non uniforme des charges localisées majoritairement à l'interface $\text{SiO}_2/\text{HfO}_2$ (cf. Fig. (II.33.a)). Comme précédemment, la déformation du diagramme de bandes n'est visible que par la résolution de l'équation de Poisson à l'intérieur du diélectrique (cf. Fig. (II.33.b)). La décroissance similaire de la transparence tunnel pour une distribution de charges fixes uniforme ou non uniforme (cf. Fig. (II.32.c) et (II.33.c)) est due aux calculs de celles-ci, calculs qui se font en partie à partir de la charge effective vue de l'interface.

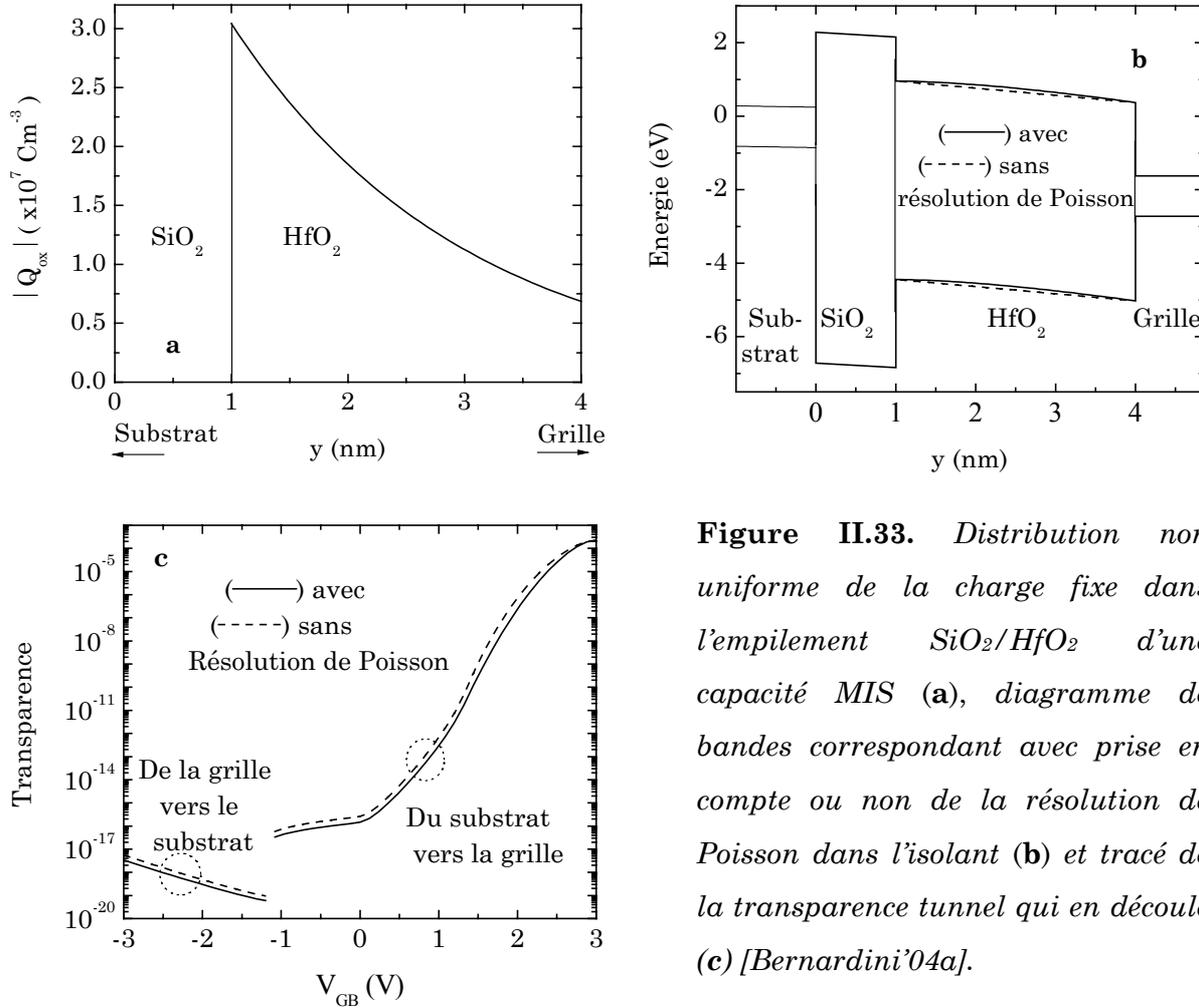


Figure II.33. Distribution non uniforme de la charge fixe dans l'empilement $\text{SiO}_2/\text{HfO}_2$ d'une capacité MIS (a), diagramme de bandes correspondant avec prise en compte ou non de la résolution de Poisson dans l'isolant (b) et tracé de la transparence tunnel qui en découle (c) [Bernardini'04a].

II.6.5. Résolution spatiale de la charge piégée Q_{ox}

A partir de cette étude, nous avons mis au point une nouvelle technique de caractérisation qui permet la détermination de la densité de charges piégées dans l'isolant d'une capacité SOS (Semiconducteur/Oxide/Semiconducteur) après des stress électriques. Cette méthode repose sur le fait que la position de la charge n'a pas le même impact sur les deux courants car la probabilité de passage des électrons est différente suivant l'électrode injectante (grille ou substrat).

II.6.5.1. Structure d'étude

La capacité utilisée pour cette étude a une surface $W \times L = 1000 \times 75 \mu\text{m}^2$. Elle est constituée d'un substrat de type N^+ , sur lequel une couche tunnel SiO_2 a été déposée, puis recouverte par une grille en poly-silicium de type N. Afin d'extraire les paramètres caractéristiques de cette capacité MOS, nous avons réalisé une mesure C-V quasi-statique à l'aide de l'appareil HP4140B piloté par le logiciel ICCAP, Agilent Technologies. L'appareil HP4140B est relié à une platine sous pointes où se trouve le wafer, c'est à dire le dispositif étudié. Lors de la mesure du courant aux bornes de notre dispositif, la tension V_G varie linéairement. Avant de faire la mesure C-V quasi-statique, les capacités parasites des câbles ont été évaluées (en laissant la pointe de la grille en l'air et l'autre pointe posée sur le drain).

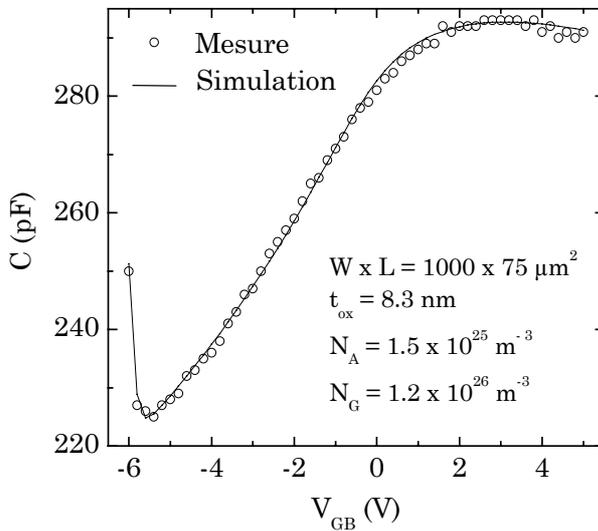


Figure II.34. *Mesure et simulation d'une courbe C-V de notre structure d'étude.*

La figure (II.34) présente deux courbes C-V : l'une simulée (trait plein) et l'autre mesurée (points) pour une vitesse de rampe $\partial V/\partial t = 100 \text{ mV.s}^{-1}$. Les paramètres extraits sont le dopage du substrat $N_{\text{Sub}} = 1.5 \times 10^{25} \text{ m}^{-3}$, celui de la grille $N_G = 1.2 \times 10^{26} \text{ m}^{-3}$ et l'épaisseur d'oxyde $t_{\text{ox}} = 8.3 \text{ nm}$. Notons un léger décalage entre la mesure et la simulation de la courbe C-V qui révèle la présence d'un dopage de substrat non uniforme.

II.6.5.2. Impact du stress électrique

La figure (II.35) correspond aux différentes mesures du courant en fonction de la polarisation appliquée sur la grille et des stress appliqués sur cette électrode. Les mesures ont été réalisées sur différentes capacité SOS ayant toujours les mêmes paramètres caractéristiques. La courbe symbolisée par des ronds correspond à la courbe initiale (dispositif vierge) tandis que les courbes en traits pleins correspondent aux mesures de courant après l'application d'une tension constante soit négative, $V_G = -9.5\text{V}$

(cf. Fig. (II.34.a)), soit positive, $V_G = 8 \text{ V}$ (cf. Fig. (II.35.b)). Ces stress sont dits cumulatifs car une fois qu'ils sont appliqués, la mesure I-V est réalisée puis à nouveau le stress est appliqué et ainsi de suite jusqu'au claquage de l'oxyde.

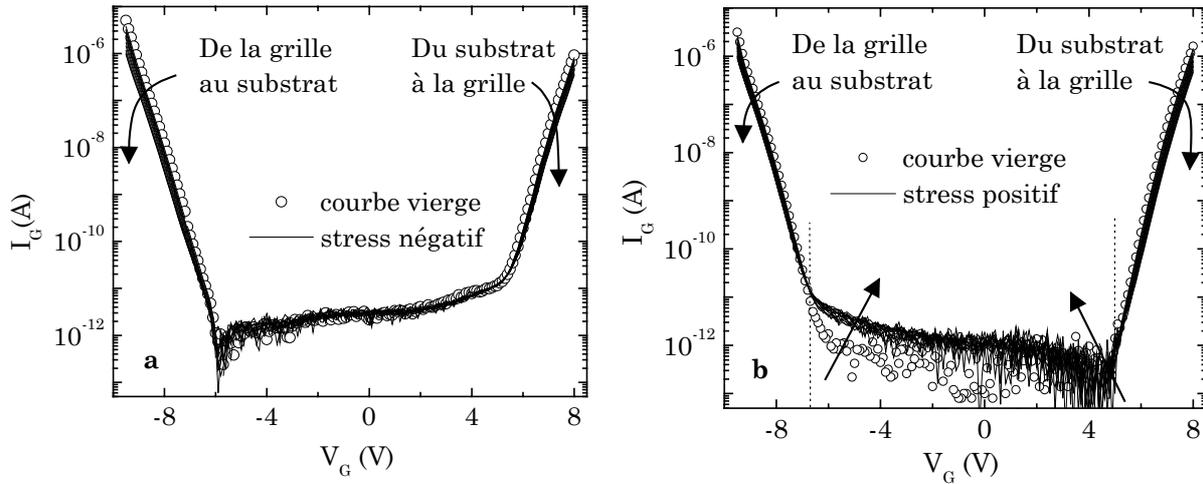


Figure II.35. Evolution de la courbe I-V d'une capacité MOS pour plusieurs stress électriques cumulatifs négatifs ($V_G = -9.5 \text{ V}$) (a) et positifs (b) ($V_G = 8 \text{ V}$) [Bernardini'03b].

Suite aux stress négatifs appliqués, c'est à dire l'injection d'électrons de la grille vers le substrat, on observe une chute du courant plus importante du côté des tensions V_G négatives. De façon symétrique par rapport aux stress négatifs, pour des stress positifs, la diminution du courant Fowler-Nordheim est plus prononcée pour les tensions de grille V_G positives que pour les tensions de grille négatives (cf. Fig. (II.35.b)).

II.6.6. Détermination de Q_{ox} (Δy)

Les dissymétries, observées lors de la diminution du courant, suggèrent une répartition non uniforme des charges piégées dans l'isolant avec par exemple une concentration plus importante de charges près du substrat après des stress positifs. A partir de ces observations, nous nous proposons de comparer trois types de répartition dans l'isolant : uniforme, Gaussienne et exponentielle.

Supposons, tout d'abord, que la distribution spatiale de charges fixes piégées dans l'oxyde soit uniforme. Dans un premier temps, nous calibrons notre simulateur à partir de la caractéristique correspondante à la capacité vierge donnée sur la figure (II.35.b). En effet, en supposant que pour une structure vierge la charge dans l'oxyde soit nulle ($Q_{ox} = Q_0 = 0 \text{ C.m}^{-3}$), comme le courant tunnel est proportionnel à la fréquence d'impact des électrons (F_{imp}) et à la masse de l'électron dans l'oxyde (m_{ox}), nous modifions ces deux paramètres jusqu'à reproduire la courbe mesurée (cf. ronds noirs de la figure (II.36.a)).

Cette figure présente également deux autres mesures, relatives aux caractéristiques $I(V)$ tracées pour des tensions V_{GB} positives après deux stress cumulés à $V_G = 8V$ (symbolisées par des ronds vides et des carrés noirs). Considérons les charges uniformément réparties dans l'oxyde, Q_1 et Q_2 , piégées respectivement durant les deux stress positifs étudiés. En modifiant les valeurs de Q_1 et Q_2 , implémentées dans notre simulateur, nous obtenons les courbes présentées à la figure (II.36.a).

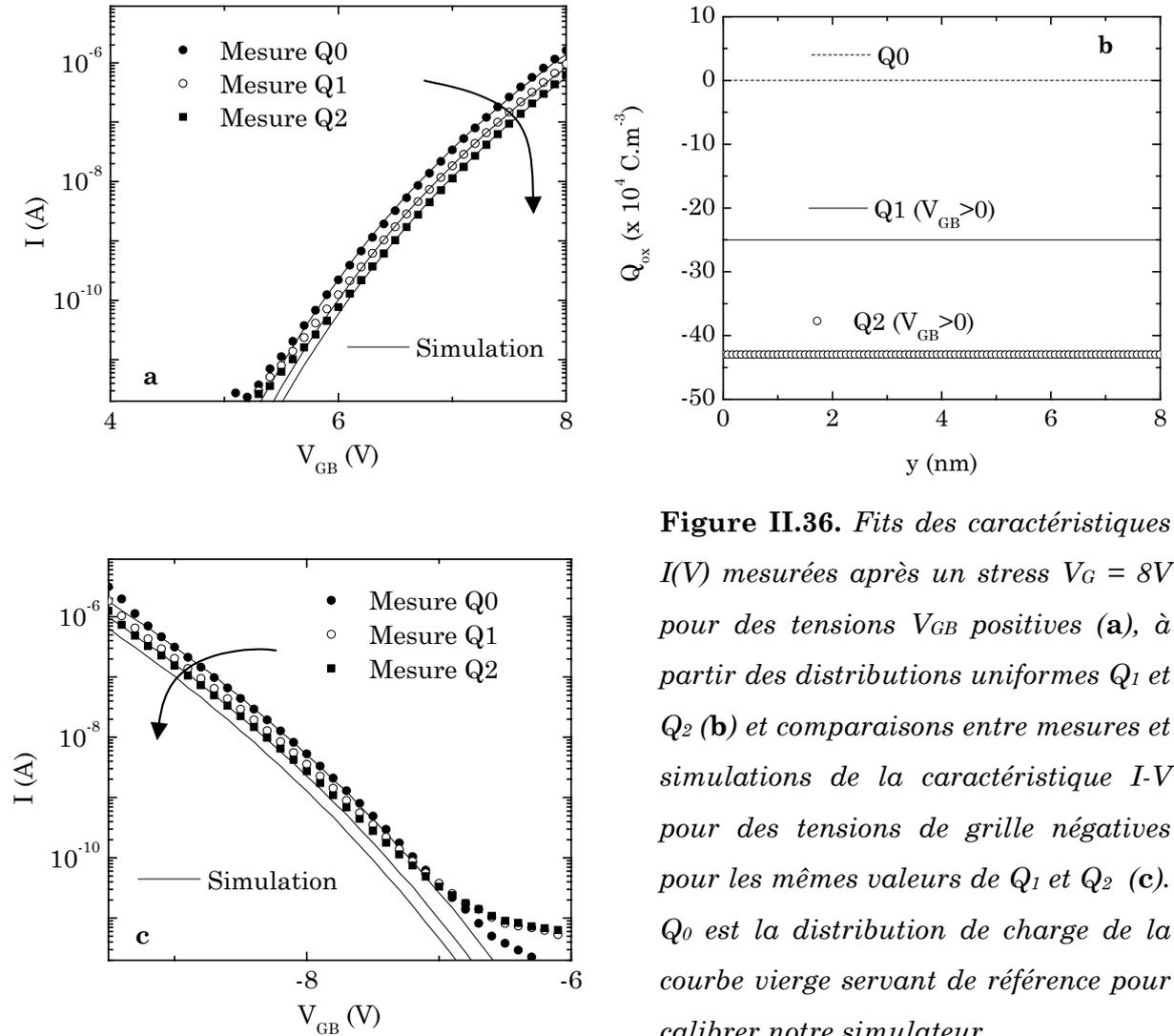


Figure II.36. Fits des caractéristiques $I(V)$ mesurées après un stress $V_G = 8V$ pour des tensions V_{GB} positives (a), à partir des distributions uniformes Q_1 et Q_2 (b) et comparaisons entre mesures et simulations de la caractéristique $I-V$ pour des tensions de grille négatives pour les mêmes valeurs de Q_1 et Q_2 (c). Q_0 est la distribution de charge de la courbe vierge servant de référence pour calibrer notre simulateur.

A partir des valeurs obtenues pour Q_1 et Q_2 , (voir Fig. (II.36.b)), les simulations des caractéristiques $I(V)$ pour des tensions de grille V_{GB} négatives ont aussi été réalisées. Dans ce cas, la figure (II.36.c) témoigne de l'impossibilité de reproduire les résultats expérimentaux mesurés (symbolisés par des ronds vides et des carrés noirs), par des simulations réalisées avec une répartition Q_{ox} uniforme déterminée à partir des caractéristiques $I(V)$ tracées pour des tensions V_{GB} positives.

De la même manière, des résultats similaires ont été obtenus en déterminant la distribution de charges fixes Q_1 et Q_2 à partir du fit des mesures reportés sur la figure

(II.36.c), pour des tensions V_{GB} négatives, puis en simulant les caractéristiques $I(V)$ pour les tensions V_{GB} positives. Par conséquent, il est impossible d'obtenir les résultats expérimentaux par simulation, en utilisant une même répartition Q_{ox} uniforme à la fois pour des tensions positives et négatives.

Supposons, à présent, que la répartition de la charge soit de type Gaussienne. Nous obtenons une bonne corrélation entre mesures et simulations lorsque le maximum de cette gaussienne est très proche de l'interface. Cependant, comme le montrent les figures (II.37), les meilleurs résultats sont obtenus en considérant une répartition de type exponentielle décroissante à partir de l'interface avec le substrat. Les courbes $Q_{ox}(y)$ extraites à partir de la comparaison mesures-simulations sont données sur la figure (II.37.b) pour trois temps de stress cumulatifs.

La loi suivie par la distribution de la charge piégée s'écrit alors sous la forme suivante :

$$Q_{ox} = Q_{max} \exp\left(\frac{-y}{\lambda}\right) + Q_{min} \quad (\text{II. 66})$$

où λ est une longueur caractéristique.

Les principaux paramètres de ces courbes sont regroupés dans le tableau (II.1) ce qui met en évidence l'augmentation de Q_{max} et λ en fonction du temps de stress.

<i>Courbes</i>	<i>Temps de stress (min)</i>	<i>Qmax ($\times 10^4 \text{ Cm}^{-3}$)</i>	<i>$\lambda(\text{nm})$</i>
Q_0	0	0	0
Q_1	16	-300	1.1
Q_2	70	-375	1.325
Q_3	208	-382	1.625

Tableau II.1. Evolution des principaux paramètres de la répartition spatiale de la charge piégée en fonction du temps de stress [Bernardini'03b].

Cette augmentation de λ correspond aussi à un mouvement des charges piégées vers l'intérieur de l'isolant pendant le stress électrique. Ce type de répartitions en exponentielles décroissantes ainsi que leurs augmentations en stress ont déjà été observés pour les pièges proches de l'interface par pompage de charge à deux niveaux (signal carré) [Maneglia'99]. Si on se réfère à l'équation (II.52), on peut aussi remarquer que la probabilité de trouver un électron à une certaine distance de l'interface est donnée

par une exponentielle décroissante (si l'on suppose une barrière trapézoïdale). Cela implique qu'il y a plus d'électrons qui peuvent être piégés près de l'interface que loin de celle-ci.

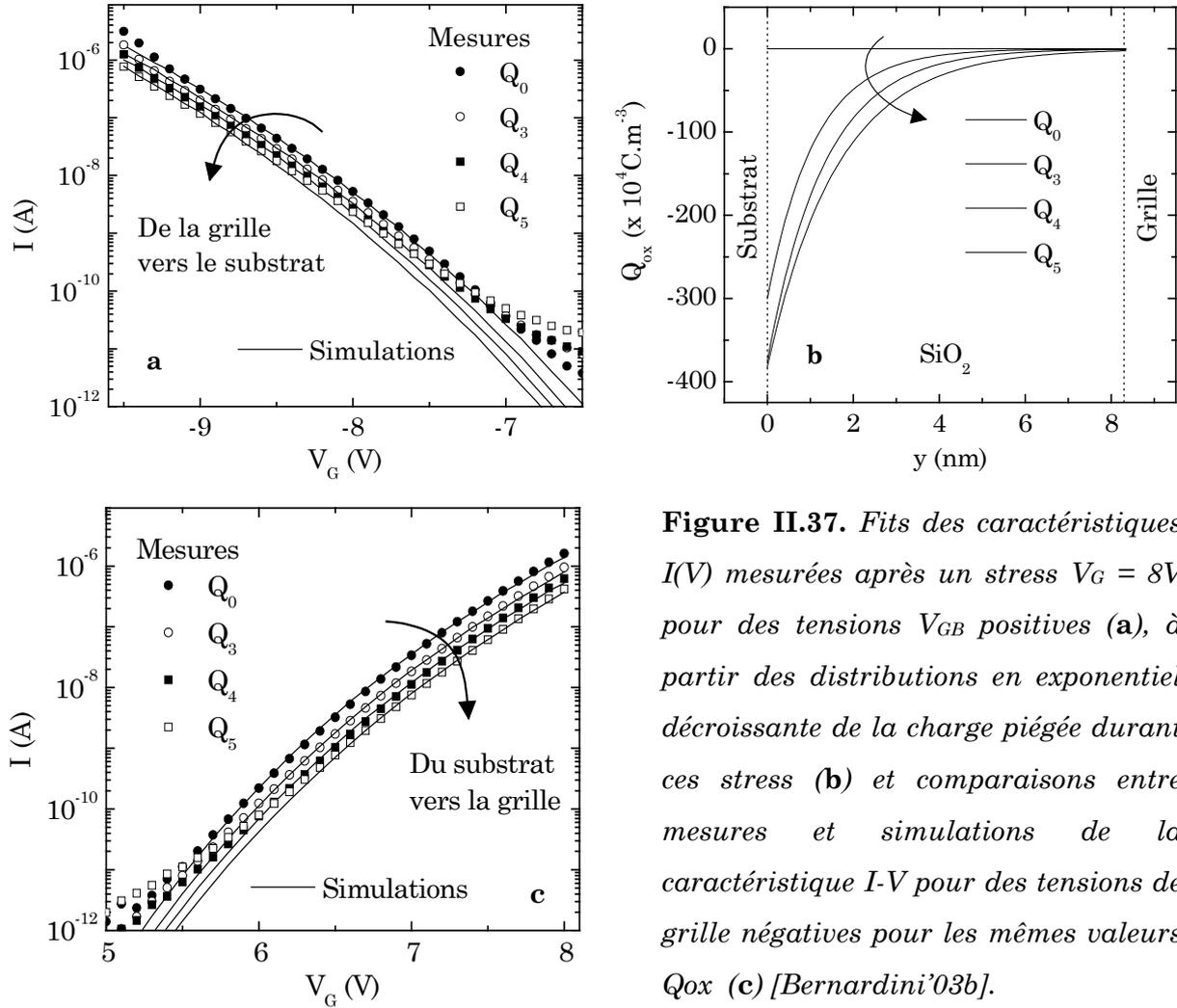


Figure II.37. Fits des caractéristiques $I(V)$ mesurées après un stress $V_G = 8V$ pour des tensions V_{GB} positives (a), à partir des distributions en exponentiel décroissante de la charge piégée durant ces stress (b) et comparaisons entre mesures et simulations de la caractéristique $I-V$ pour des tensions de grille négatives pour les mêmes valeurs Q_{ox} (c) [Bernardini'03b].

Notons, que le décalage entre deux courbes successives de courant augmente avec le temps de stress et la valeur de la tension de stress appliquée.

II.6.7. Origine de la charge fixe générée dans l'oxyde

Afin de trouver une explication possible quant à l'origine des charges Q_{ox} mesurées, nous avons étudié la variation temporelle de Q_{ox} pour les valeurs reportées dans le tableau (II.1). Considérons la cinétique de remplissage des pièges données par l'expression :

$$N_{\text{trap}}(t) = N_{\text{tot}}[1 - \exp(-\sigma(e))] \quad (\text{II.67})$$

où N_{tot} est la densité totale de pièges, e la densité d'électrons injectés et σ la section de capture des pièges.

Le tracé des points obtenus à partir des données du tableau (II.1) est présenté sur la figure (II.38). Les valeurs obtenues à partir de ce fit sont : $N_{\text{tot}} = 2 \times 10^{13} \text{ cm}^{-2}$ et $\sigma = 6 \times 10^{-20} \text{ cm}^2$. Cette valeur de σ est proche de celle d'un centre de piège neutre relié à un groupement hydroxyle, comme un centre Si-O-H.

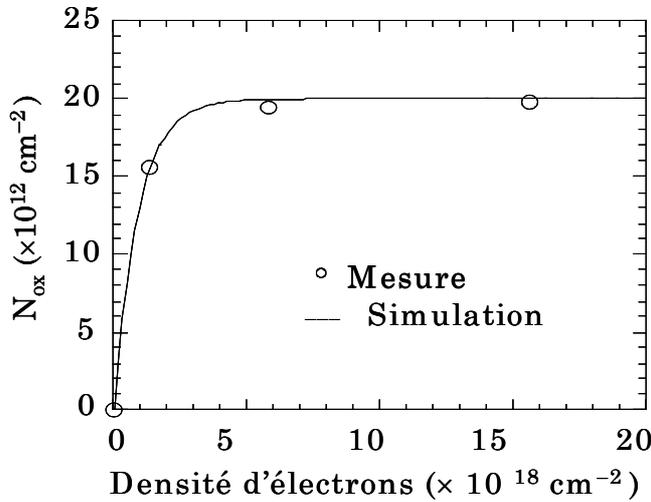


Figure II.38. *Étude cinétique des charges piégées, Q_{ox} à partir des valeurs du tableau (II.1).*

Afin d'affiner notre étude de la cinétique des pièges, de nouvelles courbes $Q_{\text{ox}}(t)$ ont été étudiées (cf. Fig. (II.39)).

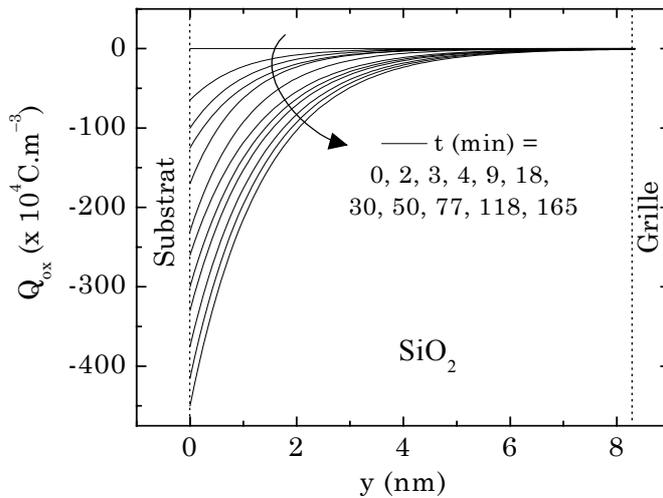


Figure II.39. *Évolution de la résolution spatiale de la charge piégée en fonction du temps de stress [Bernardini'04b].*

Comme le montre la figure (II.40.a), l'utilisation de l'expression (II.67) de la cinétique de remplissage des pièges dans le cas de l'évolution de la charge Q_{ox} présentée sur la figure (II.39), ne permet pas de retrouver les valeurs extraites à partir de la figure (II.38). En effet, la valeur $\sigma = 1.1 \times 10^{-18} \text{ cm}^2$ extraites à partir du fit de la figure (II.40.a) est trop petite (100 fois plus petite que la valeur typiquement obtenue dans le SiO_2). Un autre modèle a donc été utilisé pour simuler la cinétique de ces charges fixes, Q_{ox} . Le modèle est basé sur la libération d'atomes d'hydrogène due aux impacts des électrons pendant le stress [Brown'91]. Nous supposons que la densité de charges générées durant

un stress, ΔN_{ox} , est proportionnelle au décalage de la courbe de courant mesurée avant et après ce stress, ΔI_G . Durant le stress positif, les électrons sont injectés du substrat vers la grille, puis accélérés et permettent la création de paires électrons-trous. Lors des collisions avec l'interface grille/SiO₂, des électrons cassent les liaisons Si-H induites par la passivation de l'interface du poly-silicium. Les trous créés précédemment peuvent alors réagir avec les atomes d'hydrogène pour produire des ions H⁺. Ces derniers circulent à l'intérieur du diélectrique par sauts aléatoires (random hopping mechanism) vers l'interface Si/SiO₂. A l'interface, les ions H⁺ reprennent un électron au substrat et deviennent très réactifs. Ainsi des atomes d'hydrogène vont générer des défauts de type Si-O-H à l'intérieur de l'oxyde. Ces défauts sont connus pour être des pièges à électrons (venant de l'interface Si/SiO₂) en formant des charges SiO⁻ qui peuvent correspondre aux charges fixes que nous détectons durant les stress à V_G constant [Nicollian'82]. Leur présence serait plus importante près de l'interface Si/SiO₂ à cause de la plus forte probabilité de trouver à la fois des défauts Si-O-H et des électrons.

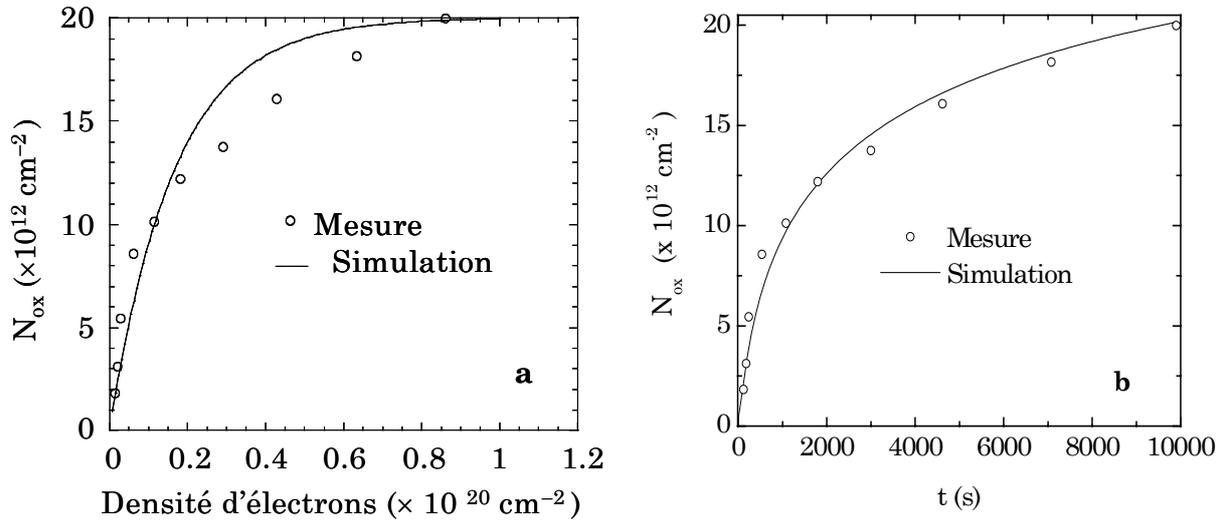


Figure II.40. Nouvelle étude cinétique des charges piégées, variation de la charge générée Q_{ox} en fonction de la densité d'électrons injectés (a) et variation de la charge générée Q_{ox} durant le stress à tension positive constante à partir du modèle de dispersion de l'hydrogène (b) [Bernardini'04b].

La variation ΔN_{ox} est donnée par la relation suivante [Houssa'01] :

$$\Delta N_{ox} = \left(1 - \int_0^{1_{SiO_2}} G_{SiO_2}(y) dy \right) N_{H^+} \quad (\text{II.68})$$

où N_{H^+} est la densité de protons générés suite à l'injection d'électron du substrat vers la grille et $G_{SiO_2}(y)$ est une fonction liée à la probabilité $P(x,t)$ de trouver un ion H⁺ à la distance x de l'anode au temps t .

La fonction $G(y)$ est une approximation de la fonction décrite par McLean and Ausman [McLean'77]. Dans l'équation (II.68), le paramètre $l_{SiO_2} = x_c/l_h t^\alpha$, où x_c correspond à la position du centroïde du défaut, l_h est la distance moyenne du saut des protons et α est un paramètre caractérisant la dispersion du transport des protons dans le matériau. En se basant sur les valeurs utilisées pour le SiO_2 , nous avons fixé, pour notre étude, $l_h=2.5 \text{ \AA}$ et $\alpha=0.3$ [Houssa'01]. La cinétique de génération des charges dans l'oxyde mesurée a été comparée à celle obtenue par le modèle précédemment décrit. La figure (II.40.b) met en évidence le bon accord entre les données extraites des mesures (cercles) et celles obtenues avec le modèle (trait plein). Les valeurs réalistes déterminées par le modèle, $N_{H^+} = 4 \times 10^{13} \text{ cm}^{-2}$ et $x_c = 3 \text{ nm}$, nous laissent penser que les défauts précurseurs de la génération des charges négatives pourraient être les centres Si-O-H.

II.7. Conclusion

Dans ce chapitre, différentes non uniformités présentes dans les capacités MIS ont été envisagées. Diverses méthodes de modélisation ont été développées pour étudier l'impact de ces non uniformités sur les caractéristiques C-V et I-V des capacités MOS. Les simulateurs utilisés ont été développés sous environnement Mathcad. Dans un premier temps, nous avons étudié la dégradation des courbes C-V induite par la poly-désertion de la grille de la capacité. Puis, nous avons mis en évidence la déformation de la courbe C-V pour un dopage de substrat non constant. Dans le troisième paragraphe, nous avons décrit l'impact d'une non uniformité de l'épaisseur d'isolant d'une capacité, avec des paramètres électriques extraits des courbes C-V et I-V erronés. Le dernier paragraphe présente l'impact des non uniformités des charges fixes dans l'isolant. Une méthode de détermination de la répartition des charges fixes a été développée. Enfin, une origine possible de la charge générée au cours des stress électriques a été envisagée avec la création de centres précurseurs Si-O-H.

Chapitre III

Modélisation du transistor MOS

Chapitre III. Modélisation du transistor MOS	91
III.1. Introduction	91
III.2. Le modèle segmenté	91
III.3. Les résistances d'accès	93
III.3.1. Impact des résistances d'accès	94
III.3.2. Extraction des résistances d'accès	96
III.3.3. Validité de la méthode d'extraction	97
III.3.4. Résistances d'accès et réduction de la longueur du canal	98
III.3.5. Impact d'une forte résistance d'accès sur la caractéristique $I_{DS}(V_{GS})$	99
III.3.6. Modification apparente de la mobilité	100
III.3.7. Dissymétrie des résistances d'accès	102
III.4. La Poly-désertion de la grille	103
III.4.1. Description du phénomène	103
III.4.2. Modélisation de la poly-désertion du TMOS	105
III.4.2.1. Etat de l'art	105
III.4.2.2. Modélisation proposée pour la poly-désertion du TMOS	105
III.5. Effets d'un dopage latéral de substrat non constant	108
III.5.1. Les structures NUDC (Non Uniformly Doped Channel)	109
III.5.2. Simulations avec le modèle segmenté	110
III.6. Impact de la non uniformité de l'épaisseur de l'oxyde de grille	112
III.6.1. Modèle GLNMOS (Gate Leakage NMOS)	113
III.6.2. Modèle segmenté modifié	114
III.6.3. Simulations de transistors MOS avec fuites de grille	117
III.6.3.1. Dépolarisation du transistor	117
III.6.3.2. Non uniformité de l'épaisseur de l'oxyde de grille	119
III.6.3.3. Etude de la taille et de la localisation du défaut	120
III.6.3.4. Application au cas d'un miroir de courant	124
III.7. Conclusion	127

Chapitre III. Modélisation du transistor MOS

III.1. Introduction

Lors de la fabrication des transistors MOS, de nombreuses étapes technologiques sont nécessaires et peuvent entraîner la non uniformité de divers paramètres clefs comme l'épaisseur d'oxyde ou le dopage du substrat. De même, l'utilisation de stress électriques peut dégrader les structures ayant des couches d'oxydes minces et entraîner par exemple l'apparition d'un courant de fuite à travers l'isolant d'un transistor. Cependant, les modélisations décrites dans le premier chapitre (modélisation de type Pao et Sah [Pao'66] ou modèle en feuillet [Brews'78]) ne prennent pas en compte ces types de non uniformités dans le développement de l'expression du courant de drain.

Ce chapitre a pour objectif de présenter les modélisations réalisées, lors de ce travail de thèse, afin d'étudier l'impact des non uniformités du transistor MOS sur ses caractéristiques électriques. Dans une première partie, nous décrirons le modèle pseudo 2D que nous avons développé pour prendre en compte les non uniformités présentes entre la source et le drain. Puis dans le deuxième paragraphe, une première application de ce modèle sera décrite avec l'effet des résistances séries. Nous montrerons ensuite, l'impact des non uniformités de dopages (grille et substrat). Enfin, nous présenterons l'étude des courants de fuite de grille engendrés par une réduction localisée de l'épaisseur d'oxyde d'un transistor.

La plupart de nos simulateurs sont développés sous environnement Mathcad, puis nos algorithmes peuvent ensuite être transférés sur des simulateurs commerciaux de circuits comme Eldo pour les rendre utilisables par nos partenaires industriels.

III.2. Le modèle segmenté

Certaines architectures de mémoires sont basées sur le chargement non uniforme de l'isolant du transistor, la charge stockée peut être plus importante près du drain par exemple. Il existe donc une variation de la tension de bandes plates entre le drain et la source du transistor. Les modèles de base du transistor, décrits dans le chapitre I, ne prennent pas en compte les non uniformités latérales du canal (c.a.d. entre la source et le drain) dans le développement de l'expression du courant de drain. Par conséquent, en

s'appuyant sur des modèles existants [Wang'81]-[Charlot'86]-[Bouchakour'95]-[Limbourg'95], nous avons développé une méthode de simulation électrique des transistors affectés par une non uniformité latérale.

Comme le suggère l'équation (I.28), le transistor MOS peut être assimilé à la mise en série de transistors de longueurs plus petites. Considérons alors le découpage d'un transistor de longueur L en N transistors élémentaires de longueur L/N . La figure (III.1) met en évidence la répartition de l'écart entre les quasi-niveaux de Fermi, noté Φ_{C0} au niveau de la source ($= V_{SB}$) et Φ_{CN} au niveau du drain ($= V_{DB}$), sur chaque transistor élémentaire.

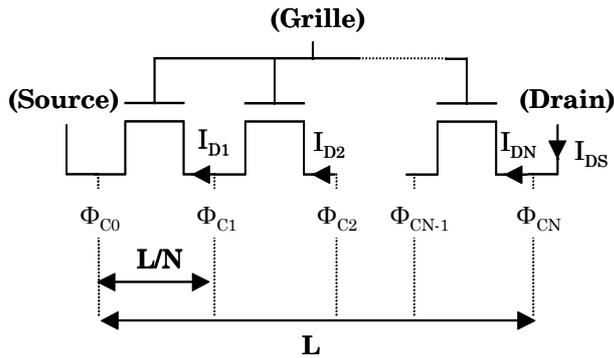


Figure III.1. Segmentation du transistor MOS et variation de l'écart entre les niveaux de Fermi d'un transistor à l'autre [Bernardini'03c].

Afin de connaître le courant de drain du transistor MOS, il faut résoudre un système de N équations à $N-1$ inconnues (les quasi-niveaux de Fermi). Ce problème est usuellement résolu par l'utilisation d'algorithmes relativement complexes faisant intervenir des matrices [Wang'95] ou en utilisant des algorithmes propres au logiciel de simulation utilisé. Pour simplifier la méthode de calcul et rendre transférables nos simulateurs dans n'importe quel langage de simulation (HDLA, C++, ...), nous considérons que chaque transistor élémentaire est assimilable à une résistance et nous utiliserons la loi du pont diviseur de tension pour calculer le courant I_{DS} . Ce calcul, résumé dans la figure (III.2), se déroule de la manière suivante :

1. Le potentiel, appliqué aux bornes de chaque transistor, est initialisé en supposant une variation linéaire de l'écart entre les quasi-niveaux de Fermi de la source au drain. On calcul alors le courant de chaque transistor élémentaire, I_D , ainsi que sa résistance, R , en utilisant la loi d'ohm.
2. A partir de la loi du pont diviseur de tension, le potentiel appliqué à chaque transistor est calculé :

$$V = V_{DS} \frac{R}{\sum R} \quad (\text{III.1})$$

puis, Φ_C , I_D et R sont à nouveau calculés.

3. Le point 2 est répété tant que les fluctuations des potentiels ne sont pas négligeables d'une itération à l'autre. Ainsi, les courants I_{DS} seront identiques pour tous les transistors élémentaires.

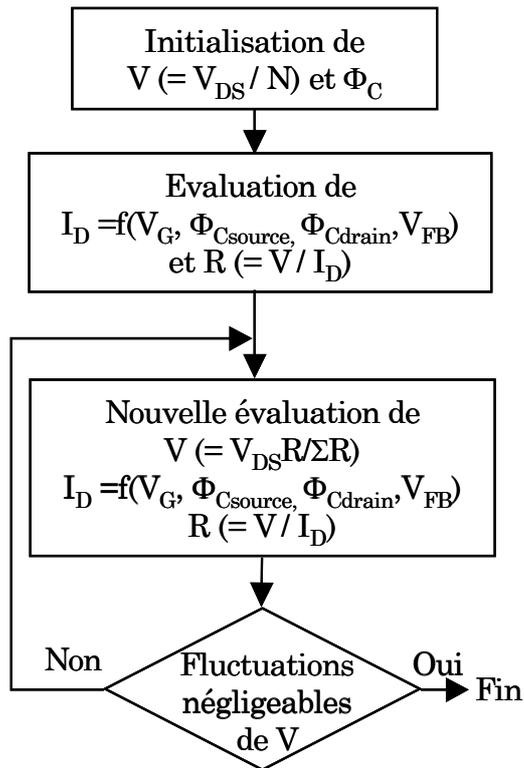


Figure III.2. Algorithme du modèle segmenté permettant le calcul du courant de drain d'un transistor segmenté [Bernardini'03c].

Pour chaque transistor élémentaire, le courant de drain peut être calculé avec le modèle en feuillet ou avec le modèle de Pao et Sah si certains phénomènes parasites (comme les effets quantiques) doivent aussi être pris en compte. Ce modèle segmenté constitue notre modèle de base pour l'étude pseudo 2D des non uniformités (de dopages, d'épaisseur d'oxyde ou de charges) entre la source et le drain des transistors étudiés. Toutefois, l'application la plus intuitive de ce modèle segmenté, où le transistor est assimilé à la mise en série de résistances, est l'étude de l'influence électrique des résistances d'accès du transistor que nous allons décrire dans le paragraphe suivant.

III.3. Les résistances d'accès

Le transistor MOS admet quatre résistances parasites, R_G , R_B , R_S et R_D , localisées sur ses terminaux (cf. Fig. (III.3)). Cependant, ces résistances n'ont pas la même influence sur le fonctionnement du transistor MOS. En considérant un courant à travers l'isolant de grille nul ou extrêmement faible, la résistance de grille, R_G , n'aura pas d'influence sur le comportement statique du transistor et la tension V_G' sera égale à V_G . De plus, si les courants inverses des jonctions et le courant d'ionisation par impact sont faibles alors V_B'

sera égale à V_B . Par conséquent, sous certaines conditions, l'influence des résistances de grille R_G et de substrat R_B peut être négligée.

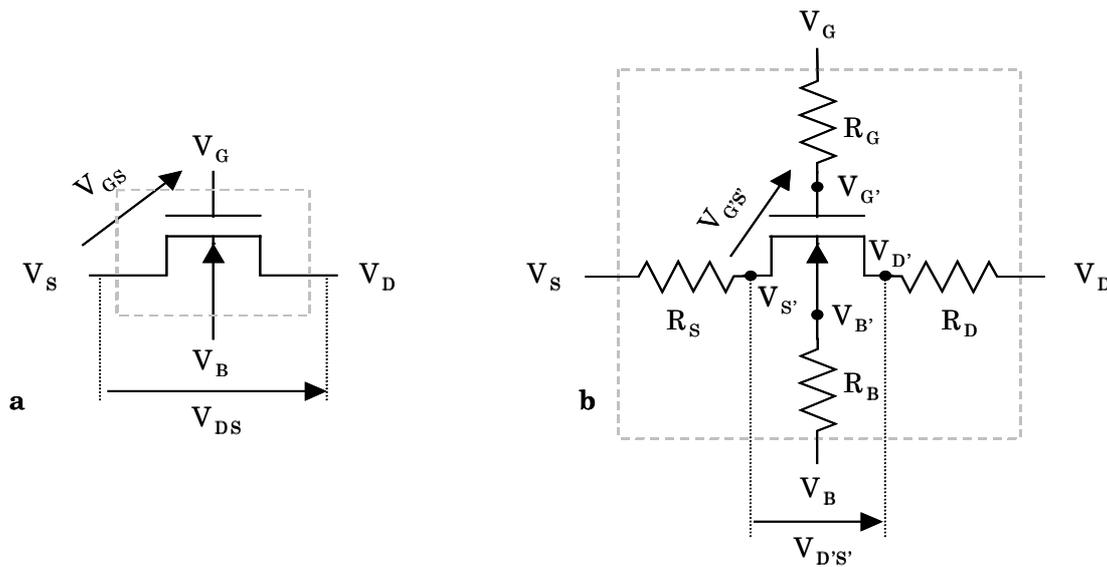


Figure III.3. Schéma électrique d'un transistor idéal (a) et d'un transistor réel faisant apparaître les résistances séries (b).

En revanche, les zones d'accès au canal, où sont appliquées les polarisations de la source et du drain, présentent des résistances, R_S et R_D , appelées résistances d'accès ou résistances séries, dont l'influence peut être néfaste pour l'amplitude du courant de drain. Ces résistances d'accès au canal ne sont pas négligeables, puisqu'elles entraînent une chute de tension à leurs bornes, et par suite une diminution de V_{DS} . Dans la suite de cette étude, nous considérerons que ces résistances R_S et R_D , regroupent à la fois les résistances de contacts, d'interconnexions, de diffusion à l'intérieur des régions source et drain, et d'injection dues au confinement des lignes de courant entre la région du canal et la région source. Enfin, pour des structures dont l'extension des zones désertées, côté source et côté drain, doit être réduite et les forts champs électriques atténués, les résistances des zones faiblement dopées (LDD Lightly Doped Drain [Ogura'80]) seront également prises en compte dans R_S et R_D .

III.3.1. Impact des résistances d'accès

L'utilisation du modèle segmenté permet une prise en compte simple des résistances d'accès, en ajoutant deux résistances, R_S et R_D , à chaque extrémité du canal (cf. Fig. (III.4)). Cet ajout se traduit par l'initialisation de ces deux résistances dans la première étape de l'algorithme de calcul présenté figure (III.2), puis par leur présence lors des sommations des résistances.

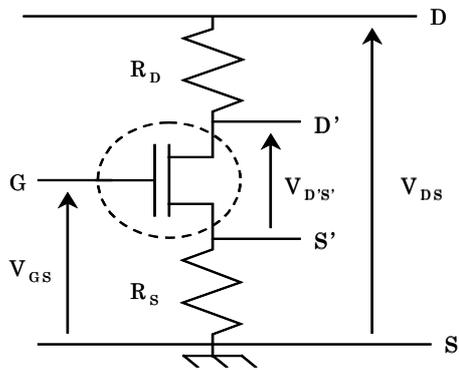


Figure III.4. Représentation schématique d'un transistor MOS (encadré) et de ses résistances d'accès au canal.

Deux régimes de fonctionnement du transistor sont alors dissociables :

- Le régime d'inversion faible où le courant de drain reste relativement faible, et, par conséquent, l'influence des résistances, R_S et R_D , est négligeable (cf. Fig. (III.5.a)).
- Le régime d'inversion forte où le courant de drain devient très important et la tension réellement appliquée au niveau du canal, $V_{D'S'}$, est plus faible que la tension V_{DS} , appliquée au transistor. Ainsi, le courant de drain obtenu est plus faible que celui attendu (cf. Fig. (III.5.b)).

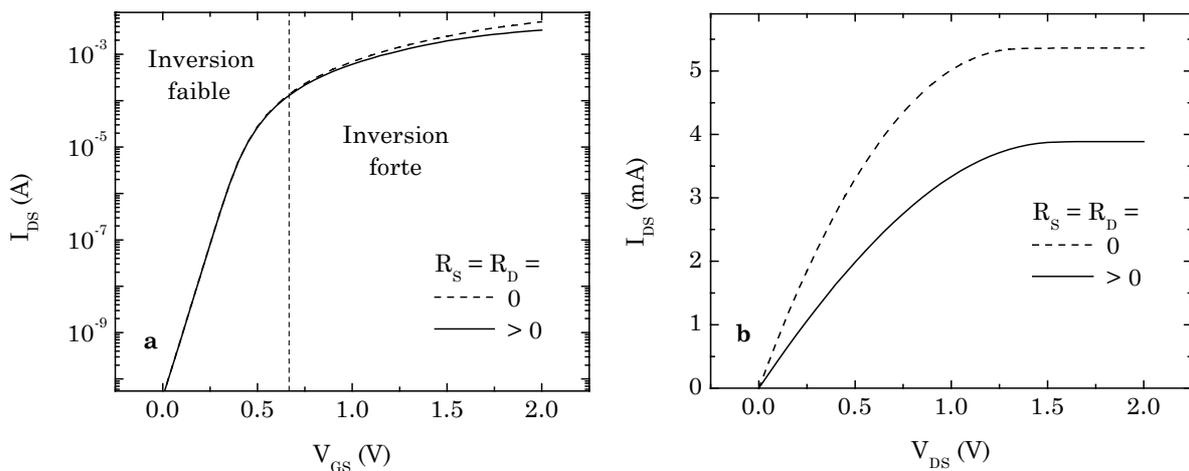


Figure III.5. Mise en évidence, à partir des courbes $I_{DS}(V_{GS})$ et $I_{DS}(V_{DS})$, de l'influence des résistances d'accès au canal en régime d'inversion faible (a) et en régime d'inversion forte (b).

Nous pouvons aussi noter que la chute de tension dans la résistance R_S implique l'apparition d'un effet substrat dont l'amplitude dépend du courant de drain et donc des polarisations. Les courbes présentées figures (III.5.a) et (III.5.b) montrent que l'effet des résistances d'accès au canal peut s'apparenter à une chute de la mobilité des porteurs en régime d'inversion forte.

III.3.2. Extraction des résistances d'accès

La figure (III.4) représente le schéma électrique d'un transistor MOS avec ses résistances d'accès au canal, côté source (R_S) et côté drain (R_D), dont la somme est notée R_{SD} . En considérant qu'en régime d'inversion forte la mobilité des porteurs peut être diminuée par les collisions phoniques et coulombiennes, l'expression classique du courant de drain à faible ($V_{D'S'}$) s'écrit :

$$I_{DS} = \frac{W}{L} \frac{\mu_0 C_{ox} V_{D'S'} (V_{GS'} - V_T)}{1 + \theta_1 (V_{GS'} - V_T) + \theta_2 (V_{GS'} - V_T)^2} \quad (III.2)$$

où μ_0 représente la mobilité à faible champ électrique, et θ_1 et θ_2 , les facteurs linéaire et quadratique de réduction de la mobilité.

La tension de seuil V_T est définie par l'expression:

$$V_T = V_{FB} + 2\Phi_F + \frac{\sqrt{2qN_A \epsilon_{Si}}}{C_{ox}} (-V_{BS'} + 2\Phi_F)^{1/2} \quad (III.3)$$

D'après la figure (III.3), la loi des mailles s'écrit :

$$V_{DS} = (R_S + R_D) I_{DS} + V_{D'S'} \quad (III.4)$$

En posant $R_{SD} = R_S + R_D$ et en remplaçant I_{DS} par son expression (III.2), on obtient :

$$V_{D'S'} = \frac{V_{DS}}{R_{SD} \frac{W}{L} \mu_{eff} C_{ox} (V_{GS'} - V_T) + 1} \quad (III.5)$$

En remplaçant $V_{D'S'}$ par son expression dans l'expression du courant, il vient :

$$I_{DS} = \frac{W}{L} \frac{\mu_0 C_{ox} (V_{GS'} - V_T)}{1 + \theta_1 (V_{GS'} - V_T) + \theta_2 (V_{GS'} - V_T)^2} \times \frac{V_{DS}}{R_{SD} \frac{W}{L} \frac{\mu_0 C_{ox} (V_{GS'} - V_T)}{1 + \theta_1 (V_{GS'} - V_T) + \theta_2 (V_{GS'} - V_T)^2} + 1} \quad (III.6)$$

Pour simplifier l'expression, on pose:

$$\theta_1^* = R_{SD} \frac{W}{L} \mu_0 C_{ox} + \theta_1 \quad (III.7)$$

ce qui permet d'écrire I_{DS} sous sa forme finale :

$$I_{DS} = \frac{W}{L} \mu_0 C_{ox} (V_{GS'} - V_T) \frac{V_{DS}}{1 + \theta_1^* (V_{GS'} - V_T) + \theta_2 (V_{GS'} - V_T)^2} \quad (III.8)$$

D'après l'équation (III.7), nous constatons que la résistance d'accès intervient directement dans l'expression du coefficient θ_1 et donc sur la valeur (apparente) de la

mobilité des porteurs. Cependant, l'expression du courant que nous venons de trouver fait intervenir $V_{GS'}$ au lieu de V_{GS} . Pour retrouver l'expression classique du courant, il faut considérer l'équation suivante :

$$V_{GS} = V_{GS'} + R_S I_{DS} \quad (\text{III.9})$$

dont on déduit que, pour avoir $V_{GS} = V_{GS'}$, il faut et il suffit que le terme $(R_S \times I_{DS})$ soit négligeable devant $V_{GS'}$, c'est à dire V_{DS} faible et R_S pas trop élevée.

L'expression (III.7) indique que le paramètre θ_1^* est une fonction linéaire de l'inverse de la longueur du canal dont la pente est proportionnelle à la résistance d'accès. Il suffit ainsi de déterminer la valeur de θ_1^* pour une série de transistors à L variable et de tracer la fonction $\theta_1^* = f(1/L)$ pour obtenir R_{SD} (et θ_1 en même temps) [McLarty'95]. Cette méthode est illustrée sur la figure (III.6) avec pour résultat $R_{SD} = 70 \Omega$ (et $\theta_1 = -0.13 \text{ V}^{-1}$), ce qui correspond à une résistance d'accès à la source et au drain de 35Ω , si on les considère comme égales.

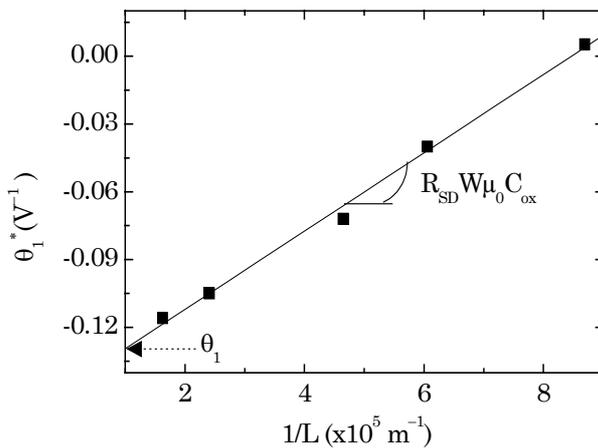


Figure III.6. Evaluation de la résistance d'accès au canal R_{SD} (et du facteur linéaire de réduction de la mobilité θ_1). Les caractéristiques des transistors sont : $t_{ox} = 5.3 \text{ nm}$ et $\mu_0 = 260 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$.

III.3.3. Validité de la méthode d'extraction

Nous avons posé dans les sous-chapitres précédents certaines limites quant à l'utilisation du modèle, notamment au niveau de la valeur de la tension de drain et des résistances permettant ainsi d'égaliser $V_{GS'}$ à V_{GS} . On peut aussi noter que la tension de seuil dépend de la tension $V_{BS'}$ et donc de la résistance d'accès à la source, ce qui rend cette équation (III.7) plus approximative. La variation de la tension $V_{BS'}$ est plus connue sous le nom d'effet substrat.

Afin de vérifier la validité de la méthode, nous avons simulé (avec le modèle en feuillet) des courbes $I_{DS}(V_{GS})$ à L variable en tenant compte des résistances d'accès et en supposant que la mobilité était indépendante de la tension de grille (c.a.d. $\mu_{eff} = \mu_0$).

Notre étude est limitée au cas d'une tension de drain de 50 mV, qui est la valeur couramment utilisée pour déterminer les caractéristiques d'un transistor. Nous avons alors déterminé la résistance $R_{SD\text{-extr}}$ (par la méthode exposée précédemment) que nous comparons à la valeur R_{SD} implémentée dans le simulateur (cf. Fig (III.7)). Nous constatons qu'aux erreurs d'extractions près, la résistance extraite est quasiment égale à la valeur R_{SD} et ce même si les résistances d'accès sont dissymétriques (c.a.d. $R_S \neq R_D$). Ce résultat confirme la pertinence de la méthode d'extraction de la résistance d'accès à faible V_{DS} .

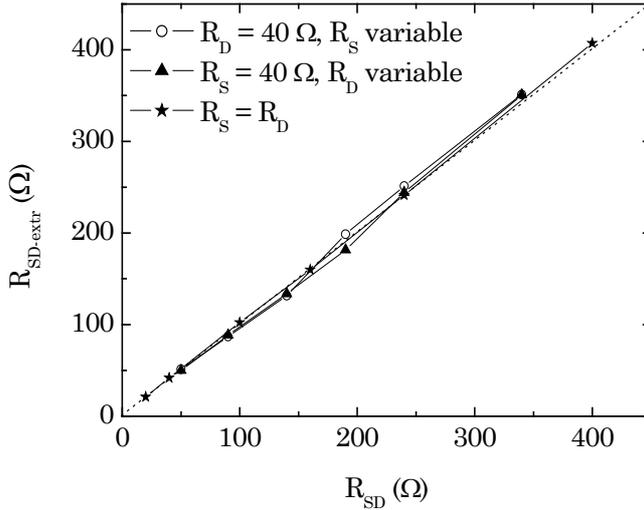


Figure III.7. Valeur de la résistance extraite $R_{SD\text{-extr}}$ en fonction de la résistance R_{SD} implémentée dans le simulateur.

III.3.4. Résistances d'accès et réduction de la longueur du canal

Les figures (III.8.a) et (III.8.b) montrent l'influence des résistances d'accès sur les courbes $I_{DS}(V_{GS})$ en régimes d'inversion faible et forte en fonction de la longueur du canal. D'après la relation (III.8) plus le transistor est court, plus le courant de drain doit être important. Nous constatons que l'impact est négligeable en régime d'inversion faible alors qu'il est très marqué en régime d'inversion forte notamment lorsque la longueur du canal décroît. Pour mettre en évidence ce phénomène, nous introduisons la résistance du canal définie par :

$$R_{\text{canal}} = \frac{V_{D'S'}}{I_{DS}} \quad (\text{III.10})$$

La figure (III.8.c) présente l'évolution de la résistance du canal en fonction de la tension de grille et de la longueur du canal. Plus le transistor est passant plus la résistance du canal diminue. Ce phénomène est amplifié lorsque la longueur du canal diminue puisque le courant I_{DS} est une fonction en $1/L$ (c.a.d. R_{canal} est une fonction de L). De plus, on peut constater que les résistances R_{canal} tendent respectivement vers des asymptotes suivant les longueurs de grille considérées. La valeur de ces asymptotes

semble inversement proportionnelle à la longueur du canal et plus la résistance du canal s'approche de son asymptote, plus les résistances d'accès diminuent l'amplitude du courant de drain.

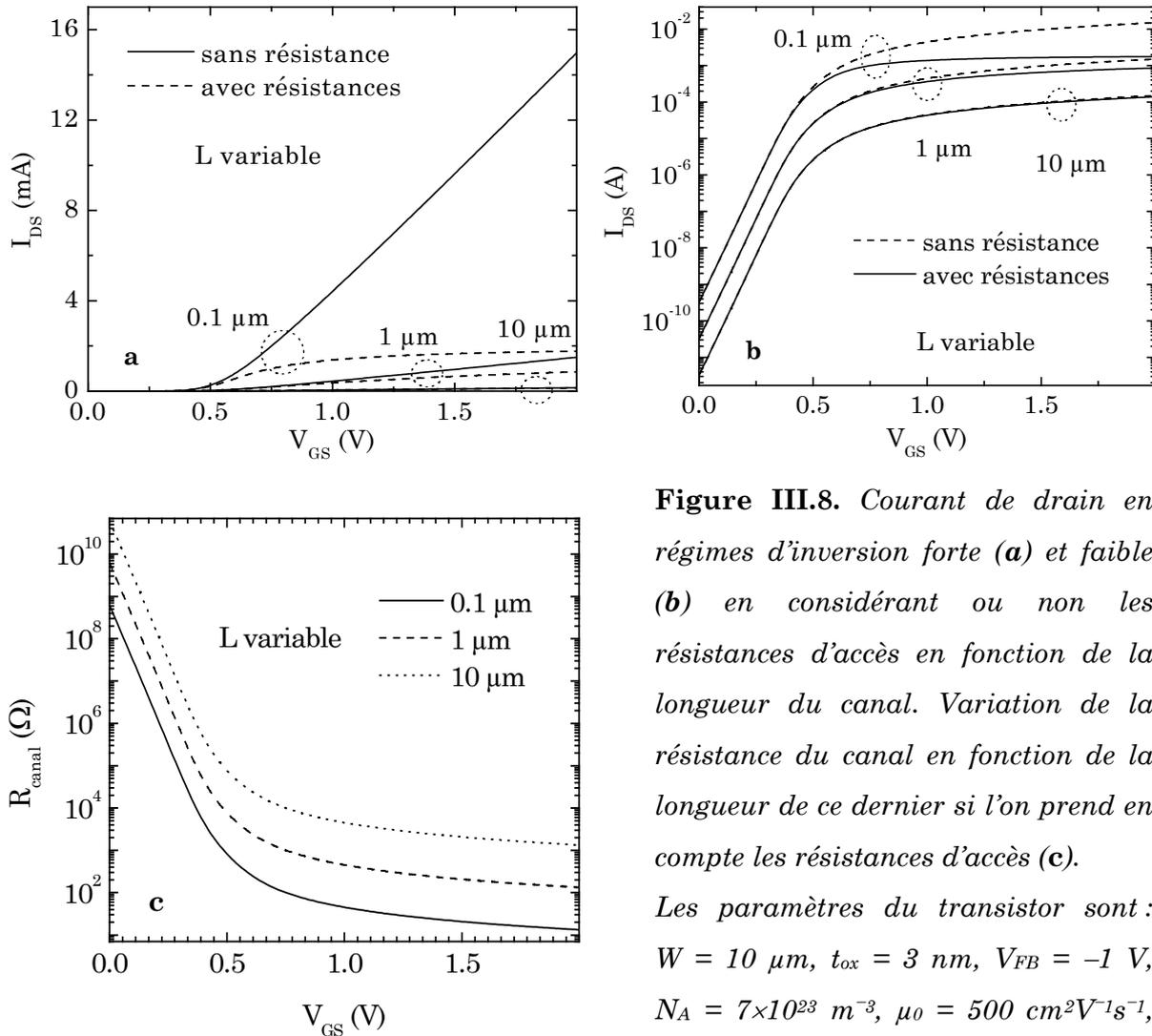


Figure III.8. Courant de drain en régimes d'inversion forte (a) et faible (b) en considérant ou non les résistances d'accès en fonction de la longueur du canal. Variation de la résistance du canal en fonction de la longueur de ce dernier si l'on prend en compte les résistances d'accès (c).

Les paramètres du transistor sont : $W = 10 \mu\text{m}$, $t_{ox} = 3 \text{ nm}$, $V_{FB} = -1 \text{ V}$, $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $\mu_0 = 500 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, $R_S = R_D = 50 \Omega$, $V_{DS} = 0.2 \text{ V}$.

III.3.5. Impact d'une forte résistance d'accès sur la caractéristique $I_{DS}(V_{GS})$

La figure (III.9.a) montre la forte diminution du courant de drain lorsque la résistance d'accès augmente d'un facteur 10. La courbe $I_{DS}(V_{GS})$ normalisée par le courant à $V_{GS} = 2 \text{ V}$ (cf. Fig. (III.9.b)), montre que pour de fortes valeurs de R_{SD} , le courant semble saturer (c.a.d. g_m tend vers 0). Ainsi, l'observation de ce type de comportement du courant de drain lors d'une mesure, peut traduire la présence d'une forte résistance d'accès. La figure (III.9.c) présente l'évolution de la résistance du canal qui, comme évoqué précédemment, tend vers une asymptote égale à R_{SD} . Nous pouvons noter que, ce

qui semble être une saturation du courant de drain, apparaît lorsque la résistance du canal a une valeur très proche de R_{SD} .

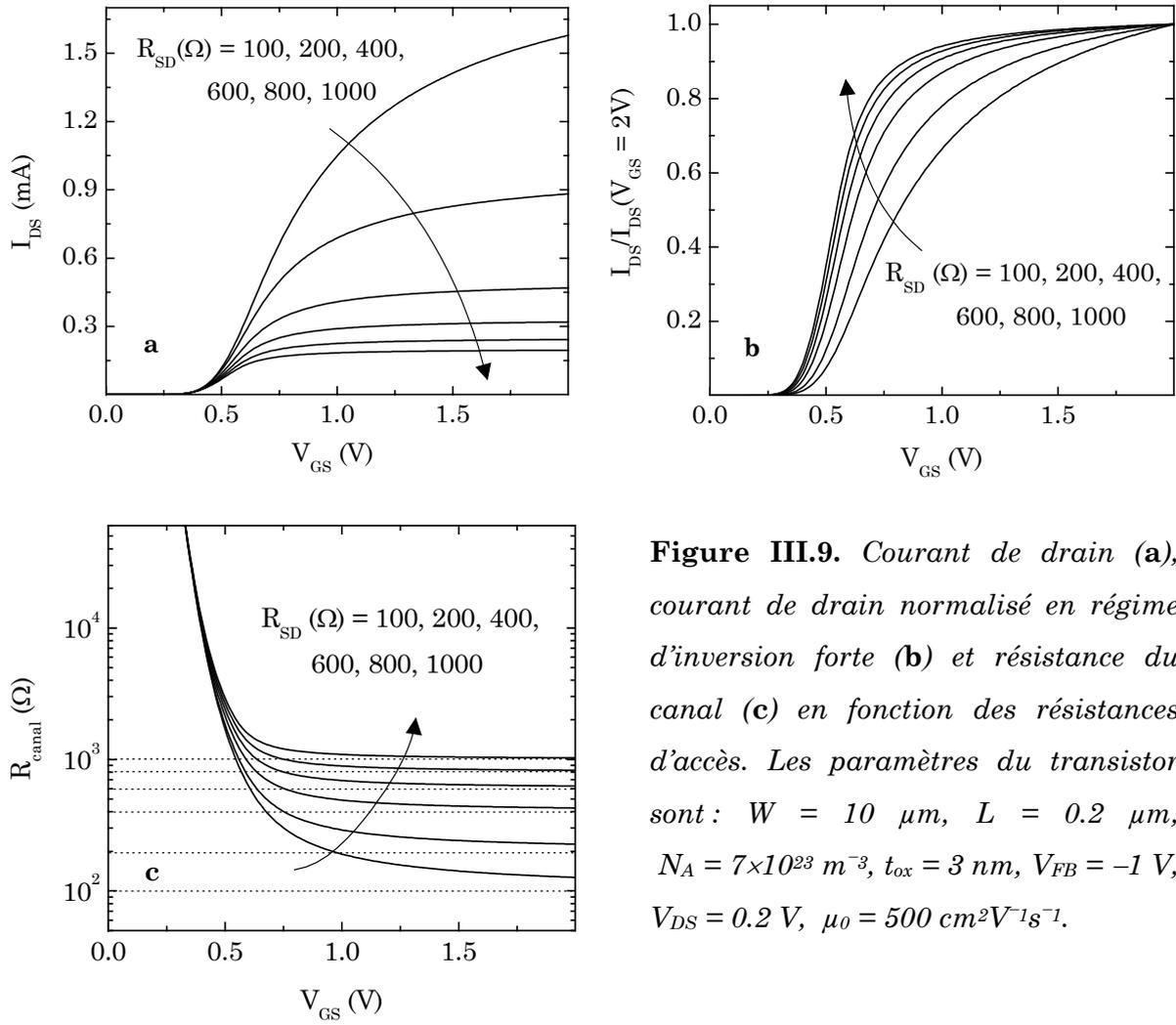


Figure III.9. Courant de drain (a), courant de drain normalisé en régime d'inversion forte (b) et résistance du canal (c) en fonction des résistances d'accès. Les paramètres du transistor sont : $W = 10 \mu\text{m}$, $L = 0.2 \mu\text{m}$, $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $t_{ox} = 3 \text{ nm}$, $V_{FB} = -1 \text{ V}$, $V_{DS} = 0.2 \text{ V}$, $\mu_0 = 500 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

III.3.6. Modification apparente de la mobilité

Considérons à présent, la simulation du courant de drain en fonction de la tension de grille avec (ou sans) résistances d'accès, que nous noterons respectivement I_{DS} et I_{DSs} , tous les autres paramètres étant identiques. Les courbes (III.5.a), (III.8.a) et (III.8.b) montrent que la diminution du courant de drain peut être assimilée à une diminution apparente de la mobilité en inversion forte, puisque pour ce régime $I_{DSs} > I_{DS}$. Comme nous avons considéré une mobilité constante lors des simulations (quel que soit le régime de fonctionnement), en affectant les variations apparentes de la mobilité à une mobilité effective μ_{eff} , nous pouvons donc écrire l'égalité suivante :

$$I_{DSs} \mu_0 = I_{DS} \mu_{\text{eff}} \quad (\text{III.11})$$

Une visualisation de l'inverse de cette mobilité effective est présentée sur la figure (III.10). Nous pouvons observer l'apparition d'un réseau de courbes en régime d'inversion forte, ainsi qu'une droite vers laquelle tendent les courbes μ_0/μ_{eff} en régime de saturation (c.a.d. V_{DS} fort). Cette courbe caractéristique doit pouvoir s'obtenir à partir d'une mesure des caractéristiques $I_{\text{DS}}(V_{\text{GS}}, V_{\text{DS}})$ d'un transistor MOS.

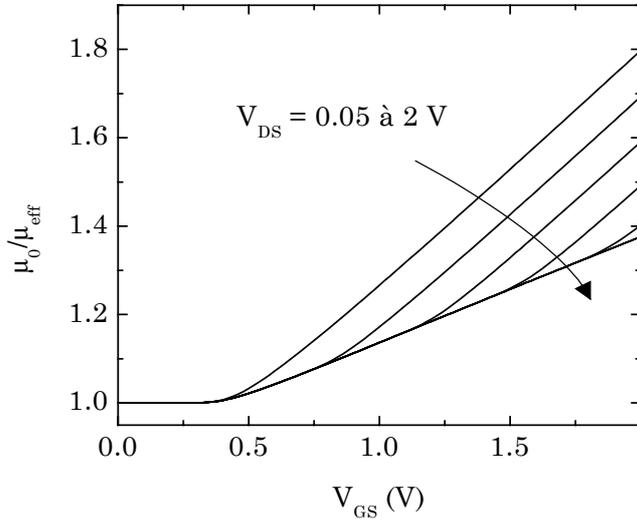


Figure III.10. Tracé de l'inverse de la mobilité effective normalisée par rapport à μ_0 en fonction des tensions de grille et de drain. Les paramètres du transistor sont : $W = 10 \mu\text{m}$, $L = 1 \mu\text{m}$, $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $R_S = R_D = 50 \Omega$, $\mu_0 = 500 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, $t_{\text{ox}} = 3 \text{ nm}$, et $V_{\text{FB}} = -1 \text{ V}$.

Nous avons donc mesuré ces courbes pour un transistor à canal long (afin de s'affranchir des effets de canal court), puis, les paramètres tels que N_A , V_{FB} et μ_0 (hors paramètres de réduction de la mobilité) ont été déterminés. Nous avons alors simulé le courant de drain, puis déterminé les courbes de l'inverse de la mobilité effective normalisée par rapport à μ_0 . Ces courbes sont tracées sur la figure (III.11) et la séparation de celles-ci, en régime d'inversion forte, démontre la présence de résistances d'accès au canal.

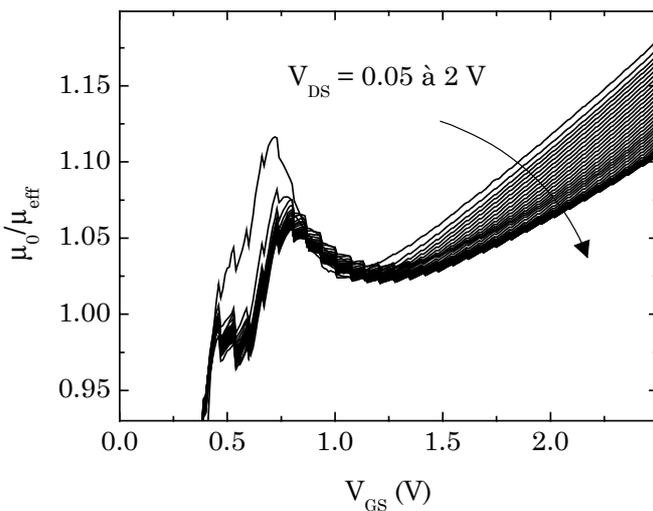


Figure III.11. Tracé de l'inverse de la mobilité effective normalisée par rapport à μ_0 en fonction des tensions de grille et de drain, obtenue après mesures et extraction des paramètres d'un transistor à canal long. Les paramètres du transistor sont : $W = L = 10 \mu\text{m}$, $t_{\text{ox}} = 7 \text{ nm}$ et $N_A = 5.8 \times 10^{23} \text{ m}^{-3}$.

En régime de saturation, les courbes μ_0/μ_{eff} tendent non plus vers une droite mais vers une asymptote, puisque dans l'équation de la mobilité effective classiquement utilisée,

nous devons prendre en compte les facteurs linéaire et quadratique de réduction de la mobilité (respectivement θ_1 et θ_2).

Cette dernière remarque implique qu'il n'est pas nécessaire de mesurer une série de transistors à longueur variable (pas toujours présente sur les wafers) pour déterminer la résistance d'accès. Il suffit, en effet, de modifier la résistance d'accès lors de la simulation du courant de drain I_{DS} , jusqu'à ce que les courbes μ_0/μ_{eff} à V_{DS} variable et en inversion forte se rejoignent.

III.3.7. Dissymétrie des résistances d'accès

Considérons à présent, un transistor ayant une résistance d'accès de source de 100Ω et une résistance de drain nulle. Simulons ensuite les courbes $I_{DS}(V_{GS})$, à faible V_{DS} , en inversant ou non la source et le drain. Ces courbes reportées sur la figure (III.12.a) pour $V_{DS} = 50 \text{ mV}$ montrent des tracés confondus. Par conséquent, la forte dissymétrie des résistances n'influence pas la valeur du courant de drain. Cela explique les résultats reportés sur la figure (III.8), résultats qui indiquaient les mêmes valeurs de résistances extraites (à partir de courbes simulées) avec ou sans dissymétrie, à partir de la méthode classique d'extraction des résistances.

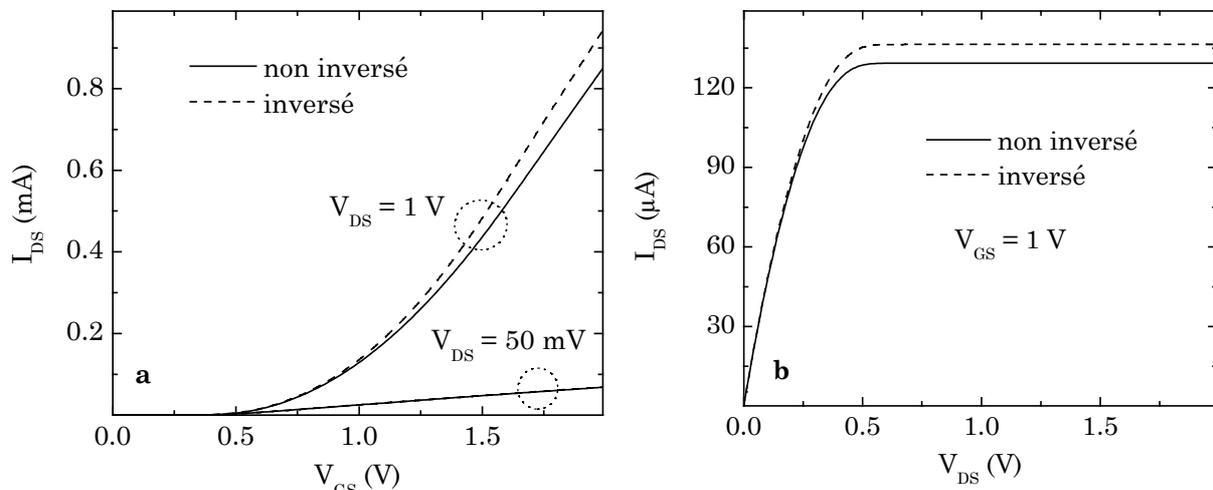


Figure III.12. Courbes $I_{DS}(V_{GS})$ (a) et $I_{DS}(V_{DS})$ (b) pour un transistor MOS à résistances d'accès dissymétriques, $R_S = 100 \Omega$ et $R_D = 0 \Omega$, en inversant ou non source et drain. Les paramètres du TMOS sont : $W \times L = 1 \times 0.5 \mu\text{m}^2$, $t_{ox} = 3 \text{ nm}$, $N_A = 7 \times 10^{23} \text{ m}^{-3}$ et $\mu_0 = 500 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$.

En revanche, une forte augmentation de la valeur de la tension de drain induit un courant plus fort si l'on inverse source et drain (cf. Fig. (III.12.a)). En effet, dans ce cas la résistance de source devient nulle et il n'y a plus d'effet substrat. Par conséquent, la méthode classique d'extraction de la résistance doit être faite à très faible V_{DS} . Ces

résultats sont mis en évidence sur la figure (III.12.b) avec le tracé des courbes $I_{DS}(V_{DS})$ avec ou sans inversion. D'un point de vue pratique, pour vérifier une éventuelle dissymétrie des résistances d'accès du transistor, il faut donc mesurer la courbe $I_{DS}(V_{DS})$ à fort V_{GS} (c.a.d. en régime d'inversion forte) avec et sans inversion de la source et du drain. L'éventuelle différence entre ces deux courbes sera la signature d'une différence entre les résistances R_S et R_D , la résistance la plus forte étant celle pour laquelle le courant est le plus faible. Ainsi, à l'aide du modèle segmenté, nous avons mis en évidence l'influence des résistances d'accès des transistors de petites dimensions, et en particulier la réduction apparente de la mobilité induite par ces résistances.

III.4. La Poly-désertion de la grille

Une réduction apparente de la mobilité peut également apparaître avec d'autres phénomènes comme la poly-désertion de grille que nous nous proposons à présent de décrire.

III.4.1. Description du phénomène

Dans le chapitre II, nous avons mis en évidence la dégradation des courbes C-V due à la poly-désertion des capacités MOS. En ce qui concerne les transistors MOS, les grilles en poly-silicium sont couramment utilisées, et supposées très dégénérées (avec des niveaux de dopage de l'ordre de $5 \cdot 10^{25}$ à 10^{27} m^{-3}). Dans le cas de dopages plus faibles (de l'ordre de 10^{24} à 10^{25} m^{-3} , induits par des procédés destinés à limiter la pénétration des impuretés à travers la couche d'oxyde), le phénomène parasite de poly-désertion se traduit par une chute de tension dans la grille. Celle-ci induit une réduction apparente de la mobilité pour une tension V_{GS} donnée, ce qui entraîne l'augmentation de la tension de seuil. Cet effet est d'autant plus prononcé que la réduction de l'épaisseur de l'oxyde de grille est importante.

A polarisation de substrat nulle, la tension de seuil augmente lorsque le dopage du poly-silicium diminue. Pour une polarisation de substrat non nulle, une augmentation de l'écart de tension de seuil (avec la polarisation du substrat) apparaît. Dans le cas d'une grille faiblement dopée, l'augmentation de la tension de polarisation du substrat contribue à une croissance de la charge de désertion dans le substrat et dans la grille. Celle-ci est d'autant plus forte que la polarisation de grille est faible. Par conséquent, plus le dopage du poly-silicium sera faible, plus la tension de seuil sera élevée pour compenser la perte de potentiel entre la grille et l'oxyde de grille. Notons que cette

dépendance avec la polarisation du substrat se répercute aussi sur l'extraction des profils de dopage dans le canal (de l'interface vers le substrat).

Rappelons que l'augmentation de la tension de seuil lorsque la longueur de grille du transistor diminue, est connue sous le nom d'effet RSCE (pour Reverse Short Channel Effect). Cet effet sera plus longuement étudié dans le paragraphe suivant (§. III.5). Toutefois, en ce qui concerne l'analyse conventionnelle de cet effet RSCE, la concentration en impuretés actives dans la grille est supposée très grande rendant la poly-déplétion négligeable. En réalité, il existe un mécanisme de modulation de l'effet RSCE dû à l'influence du niveau de dopage de la grille. Lors de fortes implantations, celui-ci tend à s'opposer à l'effet TED (Transient Enhanced Diffusion) engendré par les procédés d'implantation source/drain, utilisés lors du dopage du poly-silicium de grille [Sadovnikov'01]. La modulation de l'effet RSCE, donc de la tension de seuil (cf. Eq. (III.12)), par le niveau de dopage du poly-Si se modélise en remplaçant l'épaisseur d'isolant, t_{ox} , par une épaisseur équivalente, $t_{ox,eff}$, qui prend en compte l'isolant de grille et la poly-déplétion :

$$V_T = V_{FB} + 2\Phi_F + \frac{t_{ox,eff}}{\epsilon_{ox}} \sqrt{2\epsilon_{si} q N_A \Phi_F} \quad (III.12)$$

où N_A représente la concentration de dopants dans le canal qui est une fonction de la longueur de la grille.

Le niveau de dopants actifs dans le poly-silicium module la valeur de $t_{ox,eff}$, et par conséquent joue sur l'augmentation de la tension de seuil. Ainsi, dans le cas d'un faible niveau de dopants dans le poly-silicium (c.a.d. une épaisseur d'oxyde équivalente forte) une augmentation plus forte de l'amplitude de l'effet RSCE est observable par rapport au cas d'un fort dopage du poly-silicium. Ce comportement est en contradiction avec la théorie conventionnelle de la RSCE qui prévoyait une augmentation de l'amplitude de ce dernier pour de fortes doses implantées (génération de plus de défauts ponctuels à cause de l'effet TED) [Rafferty'93], [Jacob'93]. En tenant compte de la poly-désertion, on s'aperçoit que pour de fortes doses d'implantation, l'effet TED est concurrencé par le bas niveau de dopage du poly-silicium. C'est pourquoi la poly-désertion peut être considérée comme une des origines de l'effet RSCE.

III.4.2. Modélisation de la poly-désertion du TMOS

III.4.2.1. Etat de l'art

Différentes solutions ont été proposées pour déterminer la concentration de dopants électriquement actifs du poly-silicium de grille. Cependant, elles ne sont pas satisfaisantes : la technique du SIMS par exemple est destructive et ne permet pas de détecter uniquement les impuretés actives et la méthode C_{\min} - C_{\max} nécessite la connaissance préalable de la capacité C_{ox} . En 1994, Rios *et al.* [Rios'94] proposent une modélisation de l'effet de poly-désertion basée sur une correction de l'expression traditionnelle de la charge d'inversion en introduisant une chute du potentiel de la région de désertion du poly-silicium (ϕ_p). L'année suivante, D.N. Arora *et al.* [Arora'95] décrivent l'effet de poly-désertion dégradant la capacité de grille et réduisant le courant de drain, par des modèles AC et DC et l'expression du courant de drain est complétée pour être valable dans tous les régimes de fonctionnement du dispositif. Selon D.N. Arora, la réduction du courant de drain ne serait pas due qu'à la réduction du dopage de grille, mais aussi à une tension de seuil et une résistance série plus importantes. La poly-désertion aurait tendance à être moins marquée pour des dispositifs moins longs si les autres paramètres physiques restent inchangés. En pratique, ceci est peu probable, car il est difficile de garder la valeur des résistances séries constante lorsque le dopage de grille diminue (celles-ci vont avoir tendance à augmenter). Plus récemment, C-H. Choi *et al.* [Choi'02], ont montré que la réduction de la longueur de grille entraîne une dégradation de la capacité de la grille. La chute de potentiel à travers le poly-silicium serait accentuée par la distribution non uniforme des dopants (ΔV_{p1}) et par la réduction de la longueur de grille (ΔV_{p2}).

III.4.2.2. Modélisation proposée pour la poly-désertion du TMOS

Considérons un transistor MOS de type N (substrat type P) dont la grille est en poly-silicium de type N^+ . Dans ce paragraphe, nous nous intéresserons à la prise en compte de l'effet de la poly-déplétion sur le fonctionnement de ce transistor dans l'évolution des potentiels. La grille peut être en régime d'inversion ou de désertion profonde ce qui donne naissance à deux cas que nous allons détailler séparément.

Dans le cas où la grille est en régime d'inversion, les équations donnant la charge dans le poly-silicium de grille et la charge du semiconducteur en fonction des potentiels de surface doivent être modifiées pour tenir compte de l'écart entre les quasi-niveaux de Fermi :

$$Q_{SC} = \pm \sqrt{2kT\epsilon_{Si}p_0} \left[\frac{n_0}{p_0} (\exp(\beta(\Psi_S - \phi_C)) - \beta\Psi_S - \exp(-\beta\phi_C)) - 1 + \exp(-\beta\Psi_S) + \beta\Psi_S \right]^{1/2} \quad (III.13)$$

$$Q_G = \pm \sqrt{2kT\epsilon_{Si}n_{G0}} \left[\frac{p_{G0}}{n_{G0}} (\exp(-\beta\Psi_{SG}) + \beta\Psi_{SG} - 1) - 1 + \exp(+\beta\Psi_{SG}) - \beta\Psi_{SG} \right]^{1/2} \quad (III.14)$$

avec un signe $-$ lorsque le potentiel de surface est positif et un signe $+$ lorsqu'il est négatif. n_{G0} et p_{G0} sont respectivement les densités en porteurs majoritaires et minoritaires de la grille loin de l'interface.

Comme dans le cas de la capacité MOS, à polarisation de grille donnée, il faut résoudre simultanément les équations (III.13) et (III.14) pour déterminer les deux potentiels de surface.

Rappelons que dans notre étude, nous supposons que le semiconducteur ne peut jamais être en régime de désertion profonde mais que cela est possible pour le poly-silicium. Dans ce cas l'expression de la charge de la grille (que l'on notera dans ce cas Q_{DG}) se simplifie pour devenir :

$$Q_{DG} = \pm \sqrt{2kT\epsilon_{Si}n_{G0}} \left[\exp(\beta\Psi_{SG}) - \left(1 - \frac{p_{G0}}{n_{G0}}\right) \beta\Psi_{SG} - 1 \right]^{1/2} \quad (III.15)$$

avec un signe $-$ lorsque le potentiel de surface de grille est positif et un signe $+$ lorsqu'il est négatif.

A partir du calcul des charges, le courant circulant dans le canal peut être calculé, à l'aide du modèle de Pao et Sah [Pao'66], dans le cas où il y a désertion profonde ou inversion du poly-silicium de grille :

$$I_{DS} = -\frac{W}{L} \mu_0 \int_{\Psi_S(0)}^{\Psi_S(L)} Q_n d\Psi + \frac{W}{L} \mu_0 \frac{kT}{q} [Q_n(L) - Q_n(0)] \quad (III.16)$$

où Q_n (charge d'inversion dans le semiconducteur) est donnée par l'expression suivante :

$$Q_n = Q_{SC} - Q_{DSC} = Q_{SC} \pm \sqrt{2kT\epsilon_{Si}p_0} \left[\exp(\beta\Psi_S) - \left(1 - \frac{n_0}{p_0}\right) \beta\Psi_S - 1 \right]^{1/2} \quad (III.17)$$

III.4.2.2.1. Prise en compte de la couche d'inversion

La figure (III.13.a) montre un léger décalage de la pente sous le seuil de la caractéristique $I_{DS}(V_{GS})$ et donc une légère augmentation de la tension de seuil.

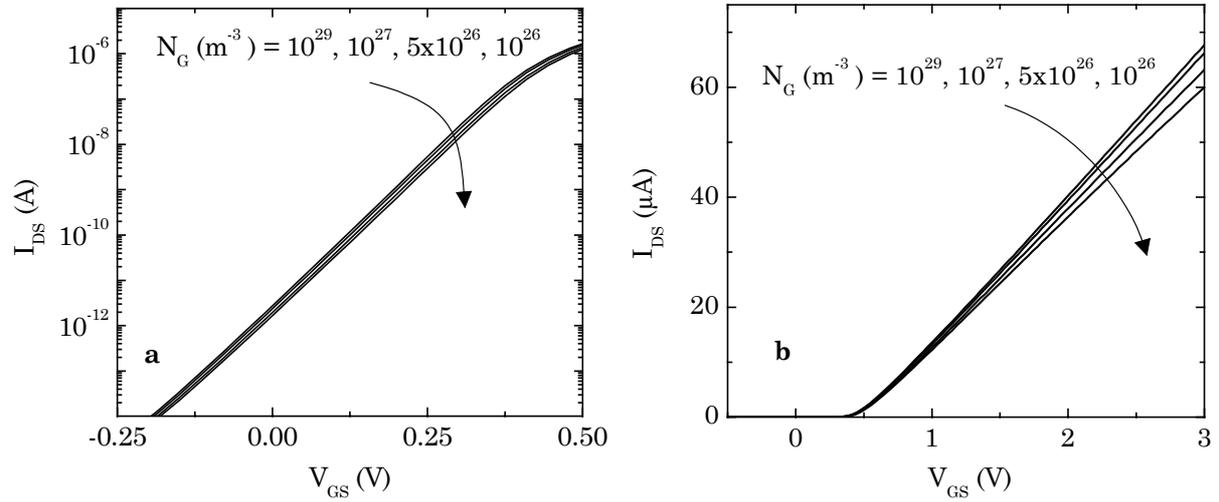


Figure III. 13. Influence de la poly-désertion sur la courbe $I_{DS}(V_{GS})$ sous le seuil (a) ou en inversion forte (b).

La figure (III.13.b) met en évidence une baisse de la transconductance (accroissement moins rapide du courant de drain) en régime d'inversion forte, baisse qui peut s'apparenter à une diminution apparente de la mobilité.

III.4.2.2.2. Désertion profonde du poly-silicium de grille

Les potentiels de surface de grille (Ψ_{SGdp}) et du semiconducteur (Ψ_{Sdp}), en désertion profonde tout au long du canal, sont tracés en fonction de la tension de grille sur la figure (III.14). Ce n'est qu'à partir d'un dopage de grille inférieur à 10^{25} m^{-3} que la séparation du potentiel de surface de grille, suivant l'état de la grille en désertion ou en désertion profonde, apparaît (cf. Fig. (III.14)).

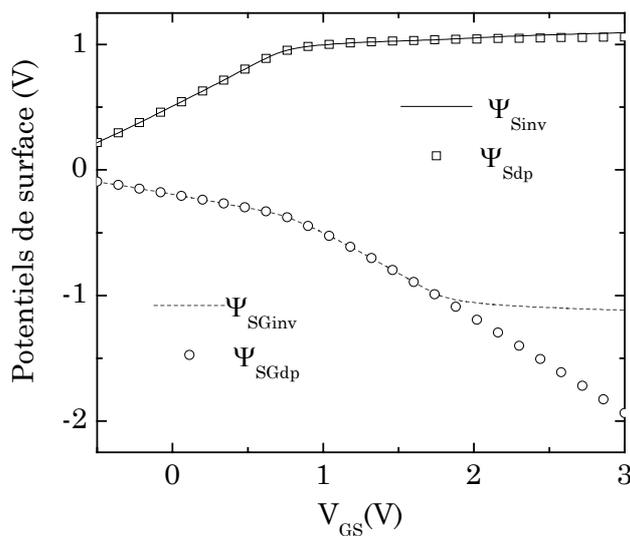


Figure III.14. Evolution des potentiels de surface du substrat et de la grille d'un TMOS en fonction de la tension de grille. Les paramètres de la simulation sont : $N_A = 7 \times 10^{23} \text{ m}^{-3}$, $N_G = 2 \times 10^{24} \text{ m}^{-3}$, $t_{ox} = 3 \text{ nm}$, $V_{FB} = -1 \text{ V}$, 20 points dans le canal.

Les figures (III.15.a) et (III.15.b) illustrent les déformations des caractéristiques du courant de drain d'un transistor ($W = 10 \mu\text{m}$, $L = 10 \mu\text{m}$) qui apparaissent pour un dopage

de grille inférieur à 10^{25}m^{-3} , dans le cas où la grille est en régime de désertion ou en désertion profonde.

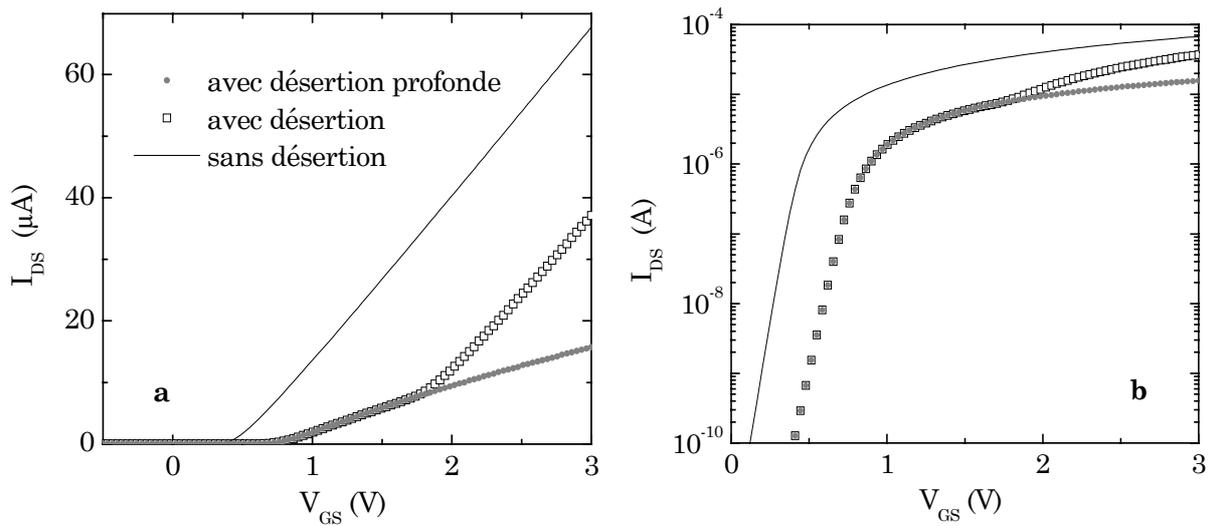


Figure III.15. Influence de la grille en désertion profonde sur la courbe $I_{DS}(V_{GS})$ sous le seuil (a) ou en inversion forte (b), 20 points dans le canal.

Ainsi, nous pouvons dire que l'effet de la poly-désertion pour un transistor MOS est moins marqué que pour la capacité MOS. En effet, pour voir apparaître cet effet, il faut des dopages du poly-silicium de grille très faibles ou travailler avec de fortes tensions de grille.

III.5. Effets d'un dopage latéral de substrat non constant

Le comportement électrique des transistors MOS est fortement conditionné par le profil du dopage. Afin d'ajuster la tension de seuil des nouvelles technologies, dont les dimensions sont de plus en plus petites, le dopage dans le canal est augmenté localement pour limiter les effets de partage de charges (ou SCE pour Short Channel Effect). Par conséquent, le dopage latéral du substrat ne peut plus être considéré comme constant. De surcroît, la concentration moyenne du dopage dans le canal augmente avec la réduction de la longueur de grille ce qui est à l'origine du phénomène d'effets canaux court inverse (RSCE, c'est à dire l'augmentation de la tension de seuil). Le but de notre travail est donc de créer un simulateur permettant d'évaluer les caractéristiques électriques d'un transistor à dopage latéral non constant.

III.5.1. Les structures NUDC (Non Uniformly Doped Channel)

Pour un transistor de type N, la réduction de sa longueur de canal entraîne la chute de l'amplitude de sa tension de seuil. Ce phénomène est connu sous le nom de partage de charge (charge sharing) : la densité de charges de la zone de désertion, contrôlée par la grille, diminue sous l'influence du champ électrique provenant des zones de désertion des deux jonctions PN (substrat/drain et substrat/source). Pour réduire ces zones de désertion, il est indispensable d'augmenter la concentration de dopants du substrat proche du drain et de la source. De plus, pour supprimer la dégradation de la conduction du courant, tout en contrôlant la chute de la tension de seuil, il faut supprimer l'effet « Coulomb scattering » c'est à dire diminuer le champ électrique normal en réduisant la concentration du dopage de substrat au milieu du canal. Ces MOSFET, ayant un canal non uniformément dopé de la source vers le drain, sont appelés structures NUDC (pour Non Uniformly Doped Channel) et présentent trois régions distinctes : les deux régions d'implantation (région proche du drain et de la source), qui ont les mêmes caractéristiques (longueur, dopage, tension de seuil), et la région située au milieu des deux.

En 1992, Okumura *et al.*[Okumura'92] présentent un modèle basé sur le découpage du transistor en trois régions. La tension de seuil du transistor ainsi définie est calculée à partir d'une analyse 2D déterminant le potentiel de surface le long du canal. Puis en 1995, Victory *et al.*[Victory'95] présentent une généralisation du modèle en feuillet du transistor MOS en incorporant la non uniformité du dopage le long du canal et en vérifiant les résultats obtenus avec une simulation 2D SPICE du transistor non uniformément dopé. Cependant, aucune comparaison n'est réalisée entre un transistor à dopage constant et un dopage non uniforme entre source et drain. En 1995, Wang *et al.* [Wang'95] incluent dans leur modèle, un dopage local et une dépendance du champ avec la mobilité, le calcul du potentiel de surface avec l'effet de substrat et un schéma multidimensionnel de relaxation. Cependant, l'utilisation de matrices pour le calcul du courant donne lieu à des algorithmes complexes. Plus récemment, en 2003, Hoewoo *et al.* [Hoewoo'03] se sont basés sur le modèle de Okumura *et al.* pour analyser la diminution et le décalage du courant de drain (simulation 1D) dus aux fortes concentrations de dopants implantés proche du drain et de la source. Dans cette analyse 1D, les variations de tension de seuil et de distribution de champ sont prises en compte en découpant le transistor en trois régions. Cependant, le courant de drain est ensuite modélisé de deux façons différentes suivant les régimes de fonctionnement.

A partir de ces travaux, nous avons utilisé notre modèle segmenté pour développer un simulateur incluant la non uniformité spatiale du dopage de substrat le long du canal du transistor. Cette modélisation offre la possibilité de faire varier le dopage de substrat le long du canal (entre la source et le drain), tout en utilisant une seule modélisation du courant pour tous les régimes de fonctionnement du transistor.

III.5.2. Simulations avec le modèle segmenté

Pour utiliser le modèle segmenté, nous découpons un transistor de longueur L en N transistors élémentaires de longueur L/N . Le dopage de chaque transistor élémentaire est supposé constant mais peut-être différent de celui de son voisin. Par conséquent, on obtient une variation de la tension de bandes plates le long du canal $V_{FB}(x)$, du potentiel de volume $\Phi_F(x)$ et par suite celle de la charge d'inversion $Q_n(x)$.

Soit un dispositif de dimensions $W \times L = 10 \times 1 \mu\text{m}^2$, dont le dopage varie le long du canal entre $N_{Amin} = 5 \times 10^{23} \text{ m}^{-3}$ et $N_{Amax} = 1 \times 10^{24} \text{ m}^{-3}$. La figure (III.16.a) présente un exemple de variation du dopage de substrat le long du canal.

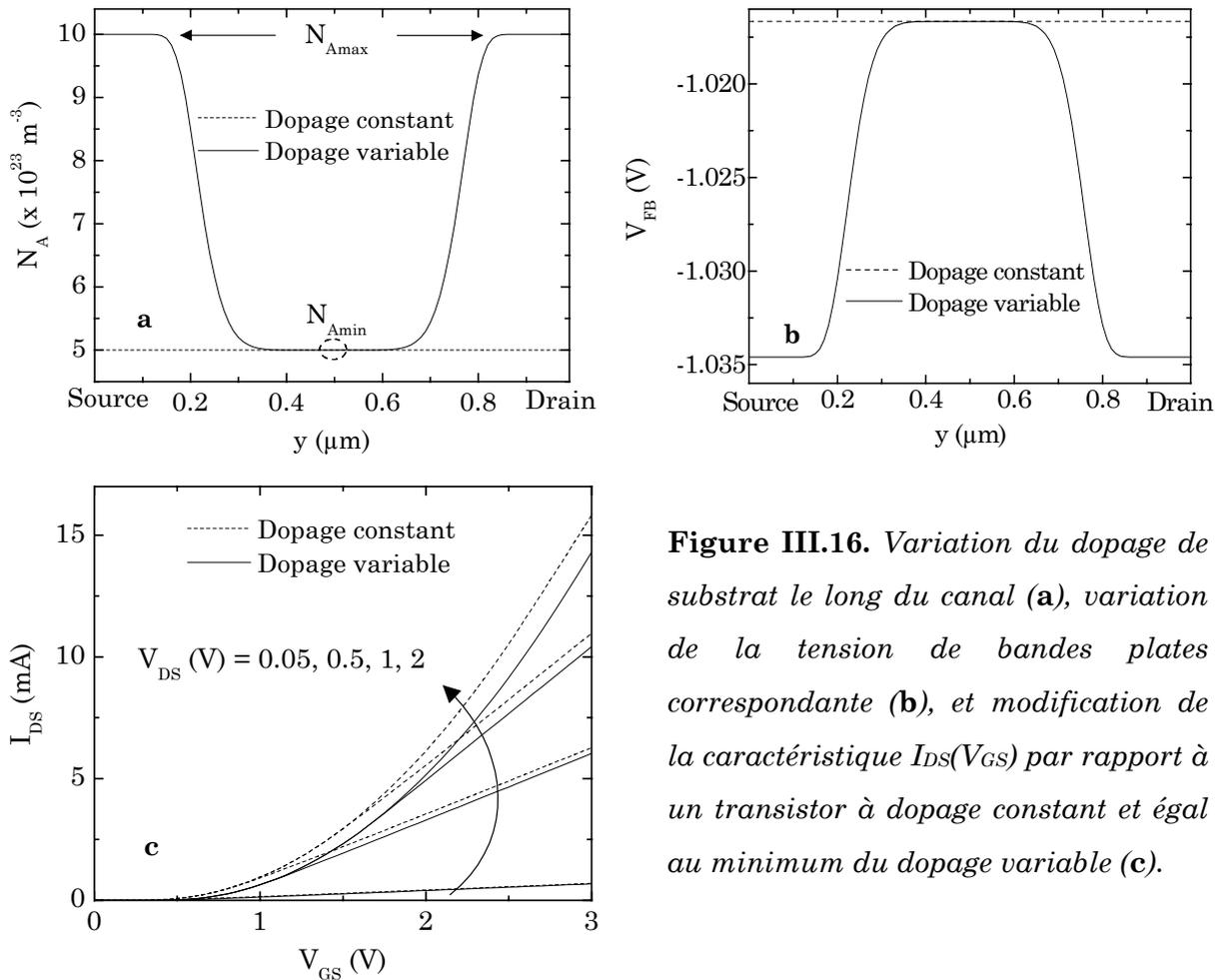


Figure III.16. Variation du dopage de substrat le long du canal (a), variation de la tension de bandes plates correspondante (b), et modification de la caractéristique $I_{DS}(V_{GS})$ par rapport à un transistor à dopage constant et égal au minimum du dopage variable (c).

La figure (III.16.b) montre la variation de la tension de bandes plates correspondante. Nous constatons que la non uniformité du dopage implique une diminution de l'amplitude du courant par rapport à un transistor à dopage uniforme et égal au minimum de la courbe $N_A(y)$ (cf. Fig. (III.16.c)). Nous avons également étudié l'impact du profil du dopage en simulant le courant de drain I_{DS} en fonction de la tension de grille V_{GS} à $V_{DS} = 50$ mV et $V_B = 0$ V, pour différents profils de dopage. Ces derniers sont tous centrés au milieu du canal ($L/2$) avec des zones d'implantation de plus en plus larges (cf. Fig. (III.17.a)). La figure (III.17.b) met en évidence l'augmentation de la tension de seuil avec celle de la longueur des zones d'implantation. En effet, les zones d'implantation ayant un dopage de substrat plus important que celui du centre du canal, plus leur longueur sera importante, plus la valeur du dopage moyen du canal du transistor NUDC sera élevée et plus les caractéristiques $I_{DS}(V_{GS})$ se décaleront vers la caractéristique du transistor à dopage constant maximum.

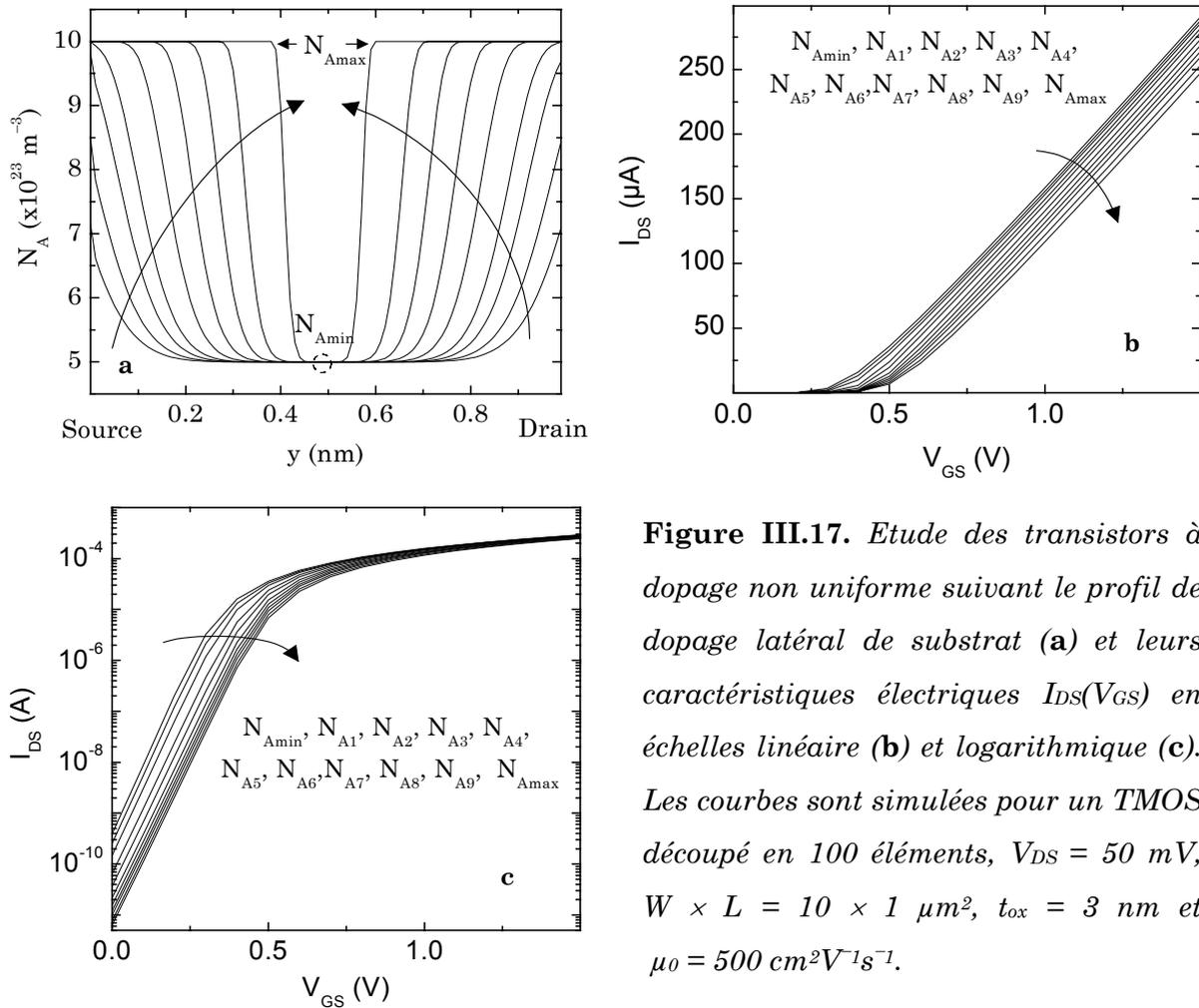


Figure III.17. Etude des transistors à dopage non uniforme suivant le profil de dopage latéral de substrat (a) et leurs caractéristiques électriques $I_{DS}(V_{GS})$ en échelles linéaire (b) et logarithmique (c). Les courbes sont simulées pour un TMOS découpé en 100 éléments, $V_{DS} = 50$ mV, $W \times L = 10 \times 1 \mu\text{m}^2$, $t_{ox} = 3$ nm et $\mu_0 = 500 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$.

Ainsi, pour le régime sous le seuil et celui d'inversion forte, la caractéristique $I_{DS}(V_{GS})$ des transistors NUDC simulés, se situe entre les caractéristiques $I_{DS}(V_{GS})$, tracées pour des transistors à dopage uniforme dont les valeurs sont celles des dopages minimum et

maximum utilisés pour les NUDC. En ce qui concerne le régime sous le seuil (cf. Fig. (III.17.c)), pour une tension V_{GS} donnée, la présence d'une légère non uniformité de dopage induit un décalage important de la région sous le seuil, à cause de la chute de conductivité (c.a.d. augmentation de V_{FB}) ; le décalage est ensuite moins important. Nous avons vu que la réduction des dimensions géométriques ($W \times L$) des transistors MOS sur silicium massif se traduit par une augmentation du dopage dans le canal ; cependant, l'amincissement de l'oxyde de grille a lui aussi un rôle non négligeable. En effet, ces deux conditions permettent de maintenir un certain contrôle des effets canaux courts tout en préservant la valeur de la tension de seuil.

III.6. Impact de la non uniformité de l'épaisseur de l'oxyde de grille

Les dernières générations de transistors MOS ont une épaisseur d'oxyde de grille qui atteint moins de 1.5 nm. Pour de telles dimensions, des erreurs de procédé de fabrication peuvent engendrer une réduction de l'épaisseur de toute la couche d'oxyde que nous appellerons dans la suite défaut uniforme (Fig. (III.18.b)). Néanmoins, une situation plus réaliste résulte d'un défaut propre à l'oxyde qui entraîne une réduction localisée de l'oxyde de grille (Fig. (III.18.c) et (III.18.d)). Cependant, lorsque l'épaisseur de l'oxyde, t_{ox} , diminue, le passage des électrons par effet tunnel à travers l'oxyde de grille du transistor devient possible. Ces électrons sont issus de la déviation vers la grille d'une partie des électrons circulant de la source vers le drain. Ce phénomène donne naissance à des courants de fuite vers la grille d'autant plus importants que l'épaisseur de l'oxyde de grille est fine.

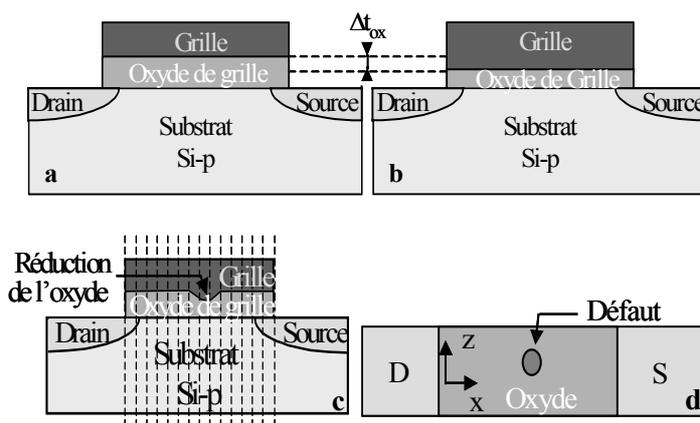


Figure III.18. Description des non uniformités de l'épaisseur d'oxyde étudiées: dispositif de référence (a), réduction uniforme de t_{ox} (b) ou réduction localisée de t_{ox} (c et d).

Ainsi, comme le montre la figure (III.19), l'amplitude du courant de grille mesuré, I_G , augmente lorsque l'épaisseur d'oxyde diminue uniformément de quelques couches atomiques, par exemple de 1.5 nm à 1.3 nm [Gilibert'04]. Par suite, comme le courant le

long du canal n'est plus constant, les courants de drain I_D et de source I_S sont dégradés et les modèles classiques du transistor MOS ne sont plus applicables.

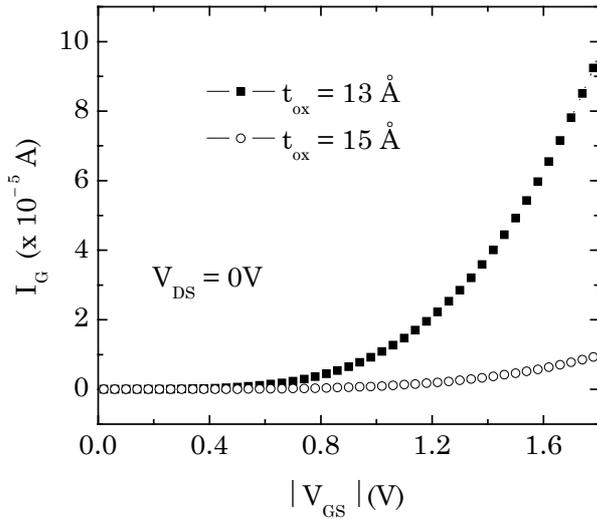


Figure III.19. Courants de grille I_G mesurés en fonction de la tension V_{GS} pour deux transistors PMOS ayant une surface $W \times L = 10 \times 10 \mu\text{m}^2$ et une épaisseur d'oxyde de 1.5 nm et de 1.3 nm [Gilibert'03].

Bien que des modèles compacts comme le BSIM4 [Xi'00] ou le MM11 [Van Langevelde'02] prennent en compte ces courants de fuite, ces modèles compacts ne sont précis que pour de faibles effets tunnel. Nous avons donc utilisé le modèle segmenté pour étudier l'impact de ces courants de grille plus ou moins élevés. Cependant, lorsque le courant de grille devient trop important, la loi du pont diviseur de tension n'est plus valide. Par conséquent, quelques modifications ont dû être apportées à notre modèle segmenté initial, pour obtenir un modèle de transistor avec fuite de grille applicable à chaque transistor élémentaire considéré.

III.6.1. Modèle GLNMOS (Gate Leakage NMOS)

Pour prendre en compte les courants de fuite par la grille, I_G , ce courant est introduit dans la description du transistor MOS. La figure (III.20) montre une vue schématique d'un transistor élémentaire NMOS à oxyde mince, ayant des fuite de grille symbolisées par des sources de courant.

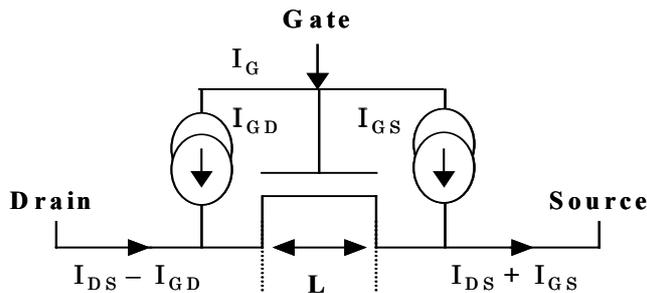


Figure III.20. Vue schématique d'un transistor NMOS avec fuite de grille (GLNMOS).

Pour simplifier les notations, nous appelons ce dispositif un GLNMOS pour « Gate Leakage NMOS ». Afin d'assurer le comportement symétrique de notre modèle (comme dans le cas du modèle en feuillet), le courant de fuite de grille, I_G , est supposé égal à la somme du courant circulant par effet tunnel entre la grille et le drain, I_{GD} , et par effet tunnel entre la grille et la source, I_{GS} .

$$I_G = I_{GS} + I_{GD} \quad (\text{III.18})$$

Malgré l'utilisation symbolique des sources de courant, ce modèle est un modèle physique puisque l'évaluation de ces courants tunnels est réalisée à partir d'une équation provenant de la mécanique quantique :

$$I_G = \frac{WL}{2} (Q_{nS} F_{\text{imp}} T_S + Q_{nD} F_{\text{imp}} T_D) \quad (\text{III.19})$$

où F_{imp} correspond à la fréquence d'impact des électrons sur la barrière de potentiel, Q_{nS} et Q_{nD} correspondent respectivement aux charges d'inversion évaluées côté source et côté drain du transistor. T_S et T_D sont les transparences tunnel des électrons côté source et drain respectivement. Celles-ci sont évaluées à partir de l'approximation de Wentzel, Krammer et Brillouin [Fromhold'81], dont l'expression générale est donnée par :

$$T_{S,D} = \exp \left(-2 \sqrt{\frac{2qm_{\text{nox}}}{\hbar^2}} \int_0^{t_{\text{ox}}} \sqrt{\Phi_{bS,D}} dy \right) \quad (\text{III.20})$$

où m_{nox} est la masse des électrons et $\Phi_{S,D}$ sont les hauteurs des barrières de potentiel vues par les électrons, côté source et côté drain respectivement.

Notons que pour une épaisseur d'oxyde donnée, le courant de grille est proportionnel à $W \times L$ tandis que le courant de drain varie avec W/L . Ainsi pour une épaisseur d'oxyde t_{ox} donnée, plus le transistor sera long, plus l'impact du courant de fuite de grille sera marqué.

III.6.2. Modèle segmenté modifié

Afin de prendre en compte les variations du courant de drain le long du canal d'un transistor à oxyde mince, le calcul des quasi-niveaux de Fermi le long du canal est nécessaire. Lors de nos travaux, le modèle segmenté a été modifié afin de prendre en compte les deux types de non uniformité d'oxyde de grille décrite dans la figure (III.18). Dans le cas d'un défaut uniforme, le modèle segmenté est modifié en considérant N GLNMOS juxtaposés (de longueur L/N) à la place des N transistors MOS élémentaires (cf. Fig. (III.21)).

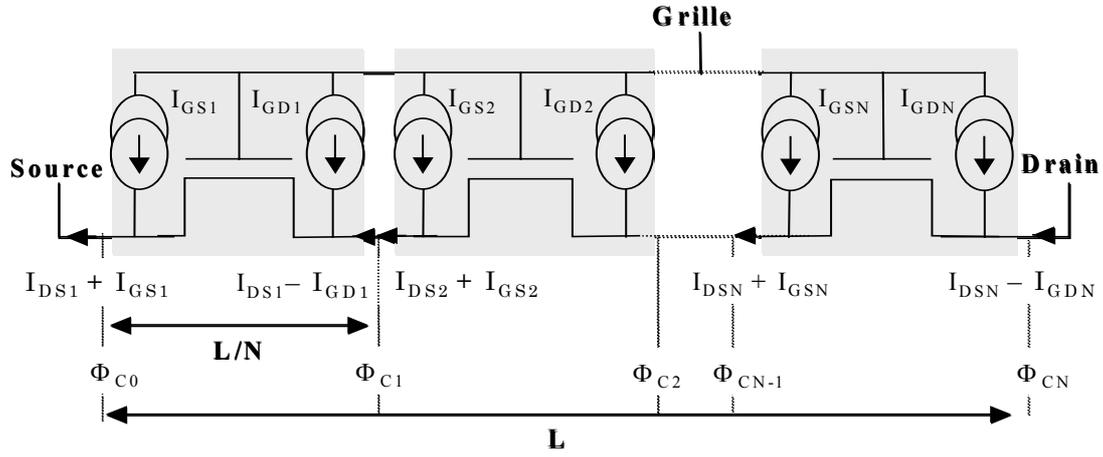


Figure III.21. Partition du transistor à isolant ultra-mince en N transistors élémentaires avec fuite de grille.

Chaque GLNMOS a une épaisseur d'oxyde qui peut être ajustée pour prendre en compte une non uniformité de l'épaisseur de la couche d'oxyde de grille d'un transistor. Pour des tensions V_{GS} et V_D fixées, la détermination de la variation de l'écart entre les quasi-niveaux de Fermi se fait en prenant les transistors élémentaires deux à deux et en imposant :

$$I_{DSN-1} - I_{GDN-1} = I_{DSN} + I_{GSN} \quad (\text{III.21})$$

Dans ce cas, la fluctuation des potentiels en chaque nœud devient négligeable. La précision de ce modèle augmente avec le nombre de GLNMOS utilisés. Par rapport à l'algorithme décrit au paragraphe (§ III.2.), seule la deuxième étape change. Il faut alors effectuer un nouveau calcul du potentiel appliqué à chaque transistor à partir de la loi des nœuds :

$$\Phi_{CN} = \frac{R_N \Phi_{CN+1} + R_{N+1} \Phi_{CN-1} + R_N R_{N+1} I_{GN}}{R_N + R_{N+1}} \quad (\text{III.22})$$

puis V , I_D et R sont à nouveau calculés. Le balayage se fait par exemple de la source vers le drain jusqu'à convergence du système.

Ce modèle est un modèle pseudo bi-dimensionnel utilisable pour une réduction uniforme de l'épaisseur de l'oxyde de grille. On peut également l'utiliser dans le cas d'un défaut local qui s'étendrait sur toute la largeur du transistor. Cependant, un défaut de surface, de part sa taille et sa position, est souvent bidimensionnel car il affecte juste une partie de la surface de l'oxyde de grille du transistor (cf. Fig. (III.18)). Dans ce cas, le modèle segmenté n'est plus utilisable puisque le transistor n'est plus symétrique. Un

maillage 2D du transistor est alors nécessaire. Notre modèle a dû être amélioré pour devenir un modèle bidimensionnel.

Pour étudier l'impact des défauts de surface de l'oxyde de grille en terme de chute d'épaisseur, nous nous sommes basés sur le modèle en matrice publié par Syrsiky [Syrzycki'89]. Le transistor est découpé en une matrice avec $M \times N$ transistors élémentaires. Une illustration de ce découpage est donnée à la figure (III.22) avec un système bidimensionnel constitué de 5×5 transistors élémentaires :

- Les transistors GLNMOS représentés en noir constituent les rangées horizontales de notre système.
- Des transistors classiques (c'est à dire sans aucun courant de fuite de grille) représentés en gris, constituent les rangées verticales de notre système. Ces transistors sont supposés sans courant de fuite pour assurer la continuité du courant de grille en fonction de l'épaisseur d'isolant.

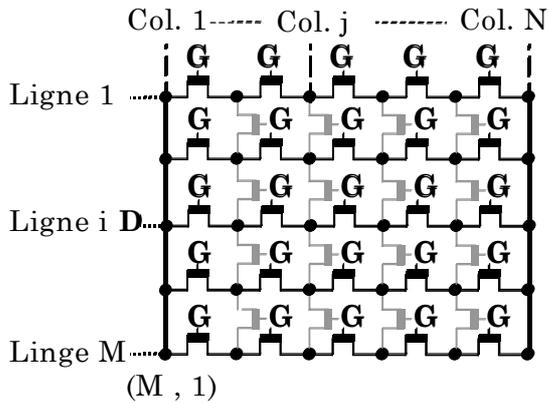


Figure III.22. *Modèle matricé avec un NMOSFET découpé en une matrice 5×5 de transistors élémentaires.*

Cette méthode offre la possibilité d'étudier l'impact de la position du défaut dans la couche d'oxyde de grille selon deux directions. Chaque nœud du tableau est localisé par un couple d'indices (i, j) avec i correspondant aux lignes et j aux colonnes de la matrice. Comme pour le modèle segmenté, les quasi-niveaux de Fermi, Φ_c , sont déterminés aux bornes de chaque GLNMOS et la précision du modèle matricé augmente avec le nombre de GLNMOS utilisés pour définir la matrice. Pour les transistors longs, le nombre de GLNMOS devra donc être grand. Dans ce cas, le calcul étant plus complexe que celui effectué en 1D, il n'est pas envisageable d'utiliser le logiciel Mathcad. Avec l'aide de Jean-Michel Portal, membre du laboratoire, nous avons donc effectué les simulations avec le logiciel Eldo. Les modèles GLNMOS sont décrits en langage HDLA pour être compatibles avec le simulateur ELDO. De plus, certaines fonctions de ces modèles sont écrites en langage C et appelées par la description HDLA, et ce dans le but d'optimiser les temps de calcul.

III.6.3. Simulations de transistors MOS avec fuites de grille

III.6.3.1. Dépolarisation du transistor

Considérons un transistor MOS dont l'isolant de grille est fin ($t_{ox} = 1.5 \text{ nm}$), c'est à dire présentant un courant de fuite.

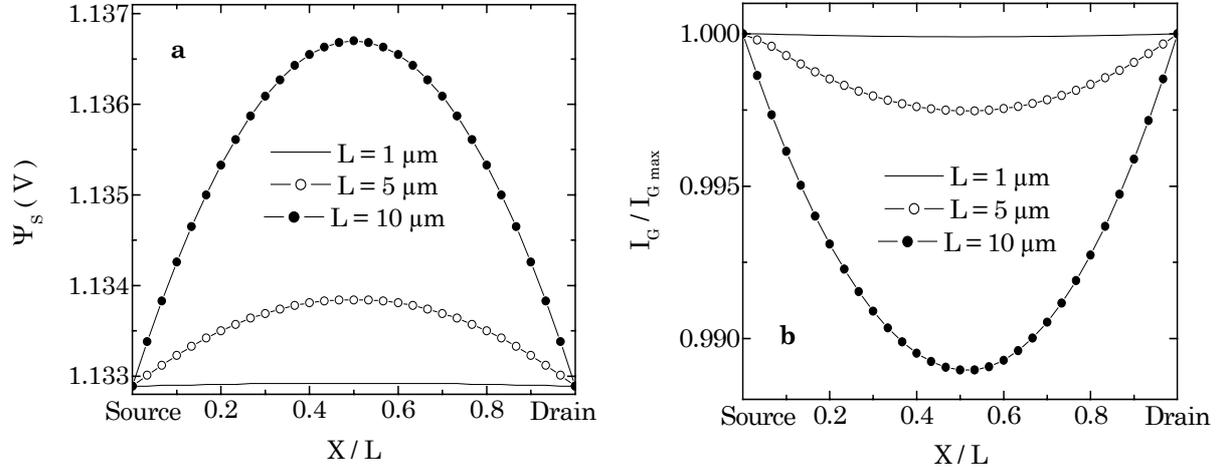


Figure III.23. Evolution du potentiel de surface, Ψ_s (a), et du courant de grille, I_G (b), le long du canal en fonction de la longueur du transistor ($L = 1, 5$ et $10 \mu\text{m}$). Les paramètres du transistor sont : $W = 10 \mu\text{m}$, $V_{DS} = 50 \text{ mV}$, $t_{ox} = 1.5 \text{ nm}$ et $N_A = 7 \cdot 10^{23} \text{ m}^{-3}$.

Le modèle segmenté GLNMOS, décrit au paragraphe (III.4.1.), donne accès à de nombreux paramètres (Ψ_s , Φ_c , I_G , I_s , I_D , ...) en fonction de la position le long du canal. Comme le montre la figure (III.23.a), pour des tensions $V_{DS} = 0 \text{ V}$ et $V_{GS} = 2 \text{ V}$, il apparaît une dépolarisation du canal ce qui entraîne une augmentation du potentiel de surface au milieu du canal. Cette dépolarisation du canal induit la présence d'un courant de fuite de grille non constant le long du canal avec une augmentation localisée dans les régions proches de la source et du drain. Le changement de pente observable sur la figure (III.23.b) témoigne du chemin parcouru par les électrons : du côté de la source, les électrons vont de la source vers la grille ($I_s > 0$), et du côté du drain, les électrons vont du drain vers la grille ($I_D < 0$). De plus, la figure (III.23.b) montre que l'impact du courant de grille sur le potentiel de surface le long du canal est plus important pour des transistors longs que pour des transistors courts, où la charge totale reste constante le long du canal.

La figure (III.24.a) donne la répartition de l'écart entre les quasi-niveaux de Fermi le long du canal en fonction de la longueur du canal pour une polarisation de drain $V_{DS} = 50 \text{ mV}$ et de grille $V_{GS} = 2 \text{ V}$. Pour une telle polarisation de drain, le transistor n'est pas en régime de saturation mais au début du régime non linéaire. Cela signifie que

la variation de $\Phi_C(x)$ doit être quasi-linéaire (avec des valeurs imposées à la source et au drain) ce qui est le cas pour le transistor de longueur très faible (c.a.d. $1 \mu\text{m}$).

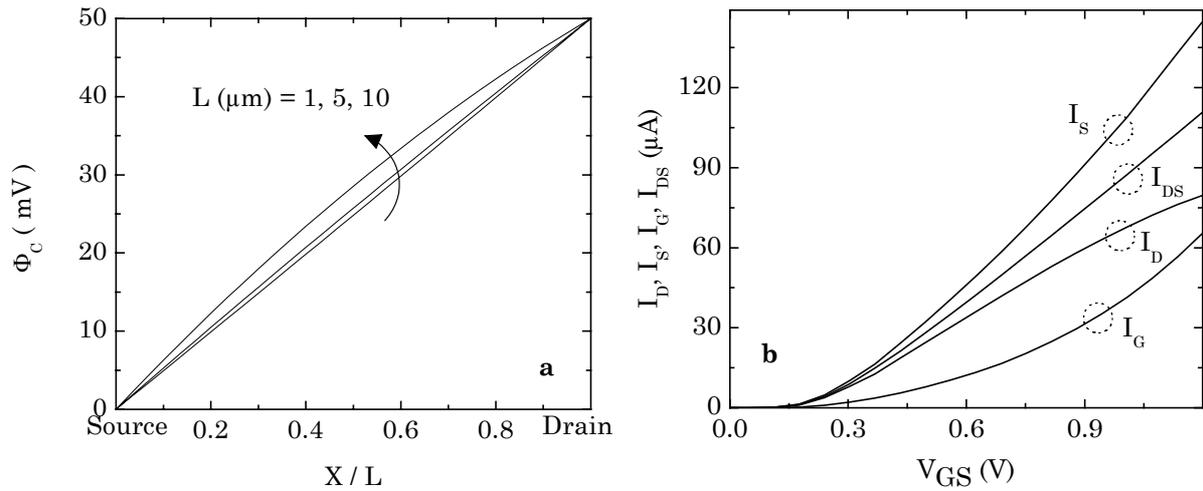


Figure III. 24. Evolution de la répartition de l'écart entre les quasi-niveaux de Fermi le long du canal en fonction de la longueur du transistor (a) et courant de grille, I_G , de drain, I_D , et de source, I_S , d'un transistor de $5 \mu\text{m}$ de long ainsi que le courant de drain, I_{DS} en omettant le courant de fuite (b). Les paramètres du transistor sont : $W = 10 \mu\text{m}$, $V_{DS} = 50 \text{ mV}$, $t_{ox} = 1.5 \text{ nm}$.

Pour les transistors plus longs (ici 5 et $10 \mu\text{m}$), la courbe $\Phi_C(x)$ s'écarte fortement d'une droite indiquant la présence d'un important courant de fuite, qui est dû à la déviation des électrons du canal vers la grille. Ce phénomène s'observe plus clairement sur la figure (III.24.b) qui présente une simulation des courants de drain I_D , de source, I_S , et de grille I_G , d'un transistor MOS de longueur $L = 5 \mu\text{m}$. On donne aussi, en référence, le courant du transistor I_{DS} qui représente le courant dans le cas où il n'y a pas de courant tunnel alors que I_D , I_S et I_G représentent les courants de drain, de source et de grille respectivement. On observe, sur la figure (III.24.b), que les courants I_S et I_D ne sont pas égaux et sont de plus très différents de I_{DS} . Cela signifie que les expressions du courant de drain données par le modèle en feuillet et par le modèle de Pao et Sah ne sont plus applicables pour de tels transistors. Il est alors nécessaire de tenir compte du courant de grille qui devient comparable aux courants de drain et de source. Ainsi, nous constatons que I_D est inférieur à I_S en raison de la fuite de grille très importante. Ce phénomène disparaît lorsque l'on augmente le potentiel de drain. En effet, au premier ordre, le courant de drain dépend linéairement de la tension de drain alors que I_G n'en dépend que faiblement. Cela implique que si on passe d'une tension de 50 mV à 500 mV , le courant de drain est multiplié par un facteur 10. Le courant de grille devient alors négligeable devant le courant de drain qui redevient égal au courant de source (et donc à I_{DS}). Le courant de grille n'a donc aucun impact sur le fonctionnement d'un transistor de

très faible longueur (sauf au niveau circuit en raison du courant total de fuite sur tous les transistors). L'utilisation d'un modèle segmenté n'est pas nécessaire pour ce type de transistor. Par ailleurs, en première approximation, I_G est proportionnel à $W \times L$ alors que I_{DS} est proportionnel à W/L ce qui explique que le courant de grille a un impact de plus en plus fort lorsque la longueur du transistor augmente [Gilibert'03]- [Gilibert'04]. Pour les transistors plus longs, l'utilisation d'un modèle segmenté est donc indispensable sauf à fort V_{DS} .

III.6.3.2. Non uniformité de l'épaisseur de l'oxyde de grille

Nous avons aussi travaillé dans un cadre plus amont sur l'impact d'une non uniformité de l'oxyde de grille du transistor MOS. Pour des isolants ultra-minces (entre 1.5 nm et 1.2 nm), le courant parasite de grille induit une modification de la caractéristique $I_{DS}(V_{GS}, V_{DS})$. Considérons un transistor NMOS de référence de dimensions $W \times L = 10 \times 10 \mu\text{m}^2$, ayant un oxyde de grille fin ($t_{ox} = 1.5 \text{ nm}$). La figure (III.25) présente les caractéristiques en courant $I_G(V_{GS})$ et $I_D(V_{DS})$ de ce transistor de référence, simulées à l'aide du modèle segmenté. On constate que le courant qui traverse la grille par effet tunnel est limité (40 μA) et que son influence sur le courant de drain est donc faible. En effet, les caractéristiques $I_D(V_{DS})$ (Fig. (III.25.b)), montrent deux régions distinctes suivant la polarisation appliquée entre la source et le drain (V_{DS}).

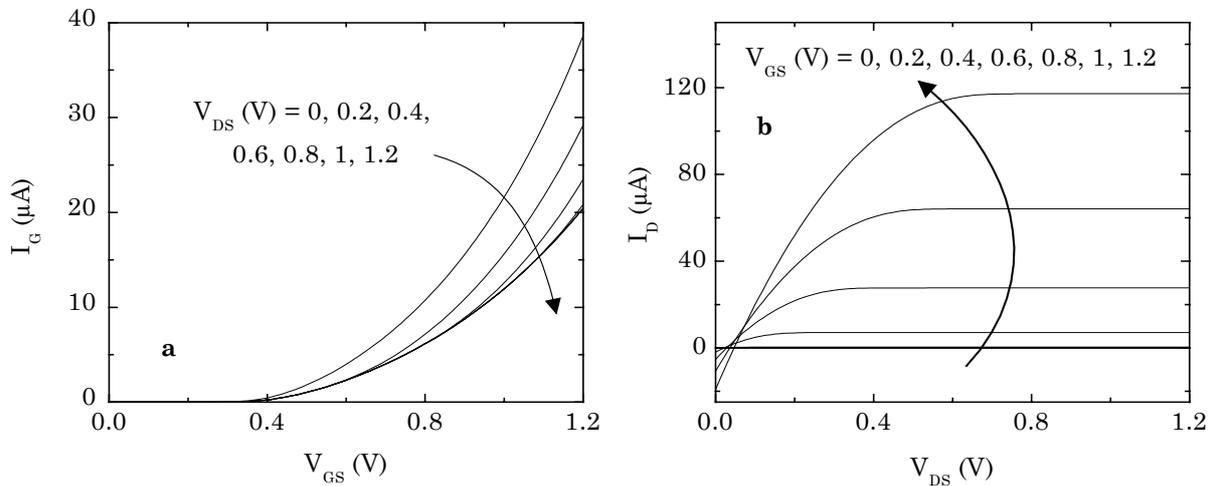


Figure III.25. Simulations des courants de fuite de grille I_G en fonction de la tension appliquée sur la grille (a) et du courant de drain en fonction de la tension appliquée sur le drain (b) pour un transistor de référence ayant $t_{ox} = 1.5 \text{ nm}$ [Bernardini'04c].

Lorsque V_{DS} est petit, le courant de drain est légèrement négatif, tant que V_{DS} reste petit devant la tension V_{GS} . Cette valeur négative du courant de drain signifie que les électrons ne sortent pas par le drain mais y entrent pour sortir par la grille. Lorsque les

valeurs de V_{DS} sont élevées, le courant de drain I_D , devient positif puisque le courant de drain de chaque GLNMOS est proportionnel à la tension V_{DS} , tant que le courant de grille est quasi-indépendant de I_D . Ces caractéristiques ont le même comportement que les mesures montrées dans la figure (III.19), ce qui apporte une validation qualitative à notre modèle. De la même manière, nous avons simulé les mêmes caractéristiques électriques que celles présentées sur la figure précédente mais cette fois-ci dans le cas d'un transistor NMOS défectueux, c'est à dire que localement $t_{ox} = 1.2$ nm (cf. Fig. (III.26)).

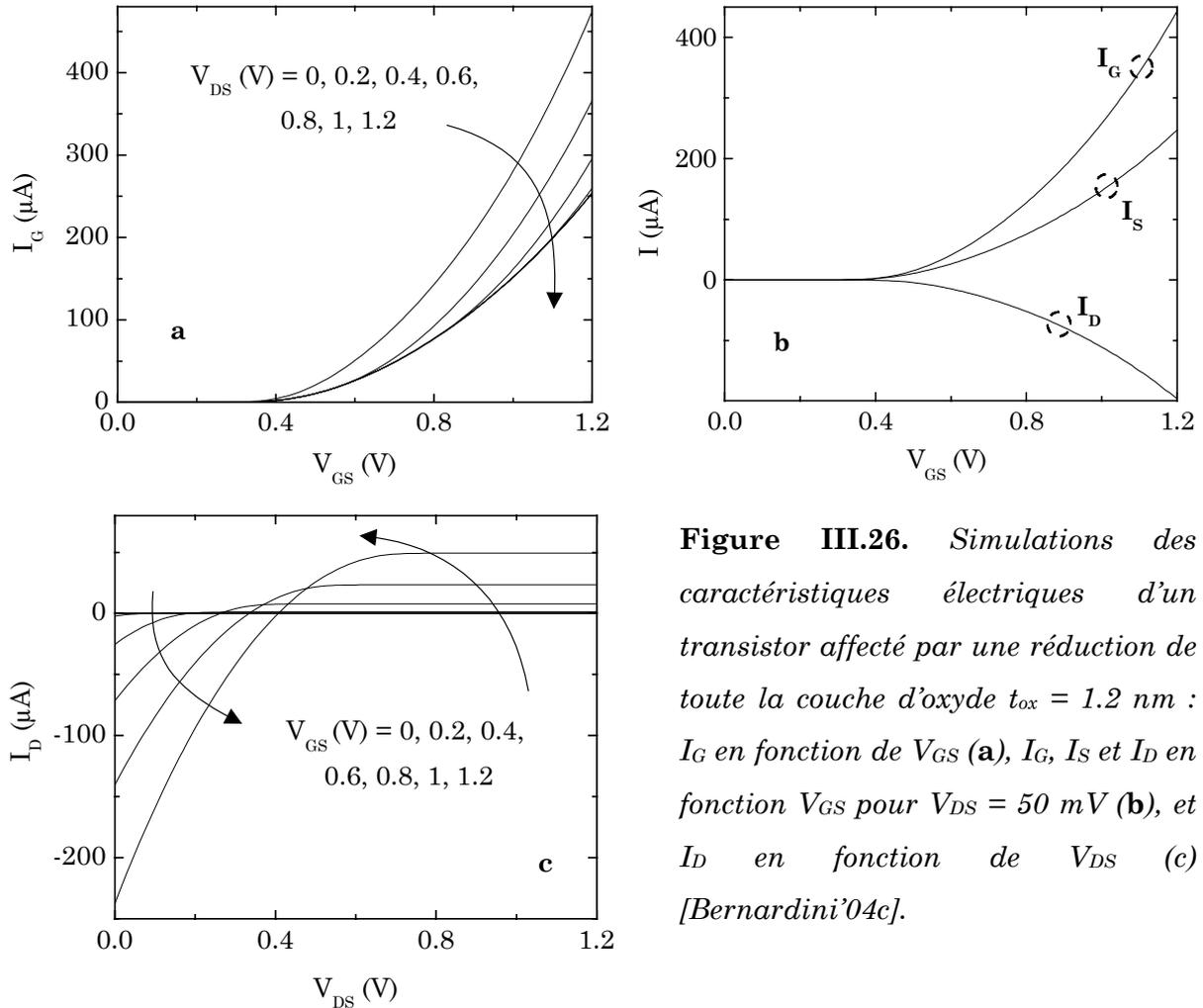


Figure III.26. Simulations des caractéristiques électriques d'un transistor affecté par une réduction de toute la couche d'oxyde $t_{ox} = 1.2$ nm : I_G en fonction de V_{GS} (a), I_G , I_S et I_D en fonction V_{GS} pour $V_{DS} = 50$ mV (b), et I_D en fonction de V_{DS} (c) [Bernardini'04c].

Dans ce cas, le manque de une ou deux couches atomiques d'oxyde de grille, induit un fort courant de grille (Fig. (III.26.a)). Globalement, l'intensité du courant de grille est multiplié par un facteur 10. Les simulations des différents courants, circulant dans le transistor MOS, présentées sur la figure (III.26.b) montrent le même comportement que les mesures reportées sur la figure (III.19), ce qui permet à nouveau une validation qualitative de notre modèle. La différence entre le courant de grille et le courant de source correspond au courant de drain. Comme le montre la figure (III.26.c), le courant de drain est fortement modifié par la présence du courant de grille avec un courant de

drain négatif plus important, dû au nombre croissant d'électrons remontant vers la grille.

III.6.3.3. Etude de la taille et de la localisation du défaut

A présent, nous nous proposons, à l'aide du modèle matricé, d'étudier l'impact de la localisation et de la taille d'une non uniformité de l'oxyde du transistor, en traçant la répartition des courants ainsi que celle du potentiel de surface.

Considérons une matrice de GLNMOS d'épaisseur d'oxyde égale à 1.5 nm (matrice de référence) et supposons qu'un seul GLNMOS ait une épaisseur d'oxyde égale à 1.2 nm (GLNMOS défectueux). Le déplacement de ce GLNMOS défectueux permet l'étude de la localisation d'une diminution de l'épaisseur d'oxyde. Par exemple, les courants tracés sur la figure (III.27) ont été obtenus à partir d'une matrice de référence, d'une matrice avec un GLNMOS défectueux proche du drain (Colonne (1) et Ligne (5)) et une autre avec le GLNMOS défectueux proche de la source (Colonne (5) et Ligne (1)).

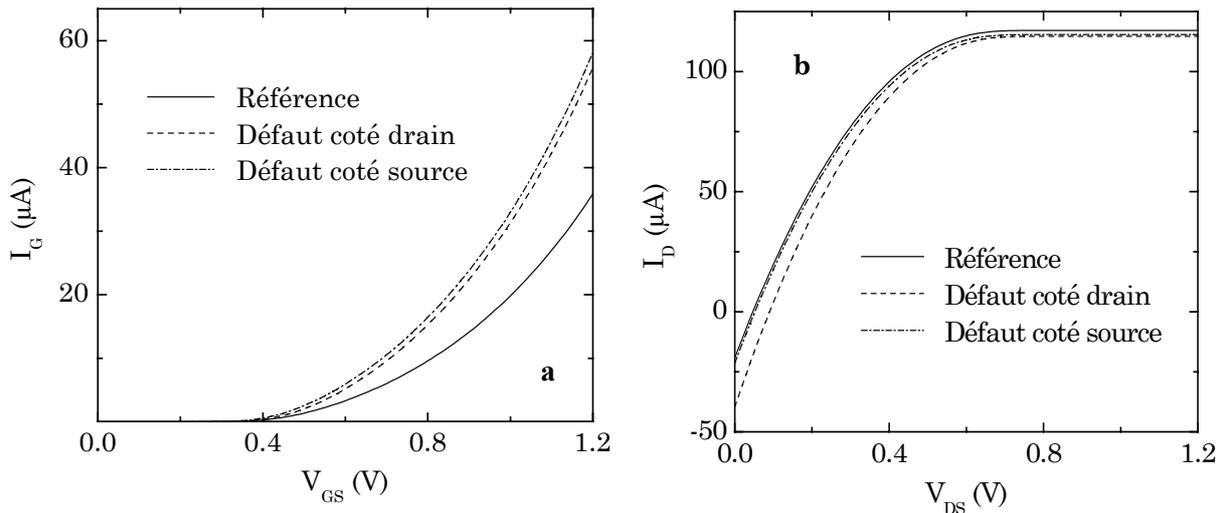


Figure III.27. Simulations des courants de grille I_G en fonction de V_G (avec $V_{DS} = 50 \text{ mV}$) (a) et des courants de drain I_D en fonction de V_D (pour $V_G = 1.2 \text{ V}$) (b) pour un transistor sans défaut ($t_{ox} = 1.5 \text{ nm}$), ou un transistor avec une chute de t_{ox} côté drain ou côté source [Bernardini'04c].

La figure (III.27.a) montre une augmentation du courant de grille de l'ordre de 40% (pour $V_G = 1.2 \text{ V}$) quelle que soit la position de la diminution de t_{ox} , contrairement aux caractéristiques $I_D(V_D)$ qui dépendent de la position du défaut (cf. Fig. (III.27.b)). En effet, pour de faibles valeurs de V_{DS} , le courant I_D est légèrement négatif pour la matrice de référence ou pour celle dont le défaut est proche de la source tandis que, pour celle dont le défaut est proche du drain, le courant I_D devient fortement négatif. Puis pour les

fortes tensions V_{DS} , quel que soit le cas étudié, les valeurs maximales des courants I_D sont légèrement plus faibles. Par conséquent, les caractéristiques $I_D(V_D)$ mettent en évidence l'impact de la position de la diminution de l'épaisseur de l'oxyde. Afin de mieux comprendre le comportement de ces caractéristiques relatives aux courants, nous avons tracé, dans les mêmes conditions, la distribution des quasi-niveaux de Fermi Φ_C , tout au long du canal (cf. Fig.(III.28)) où le GLNMOS défectueux est entouré par un cercle gris.

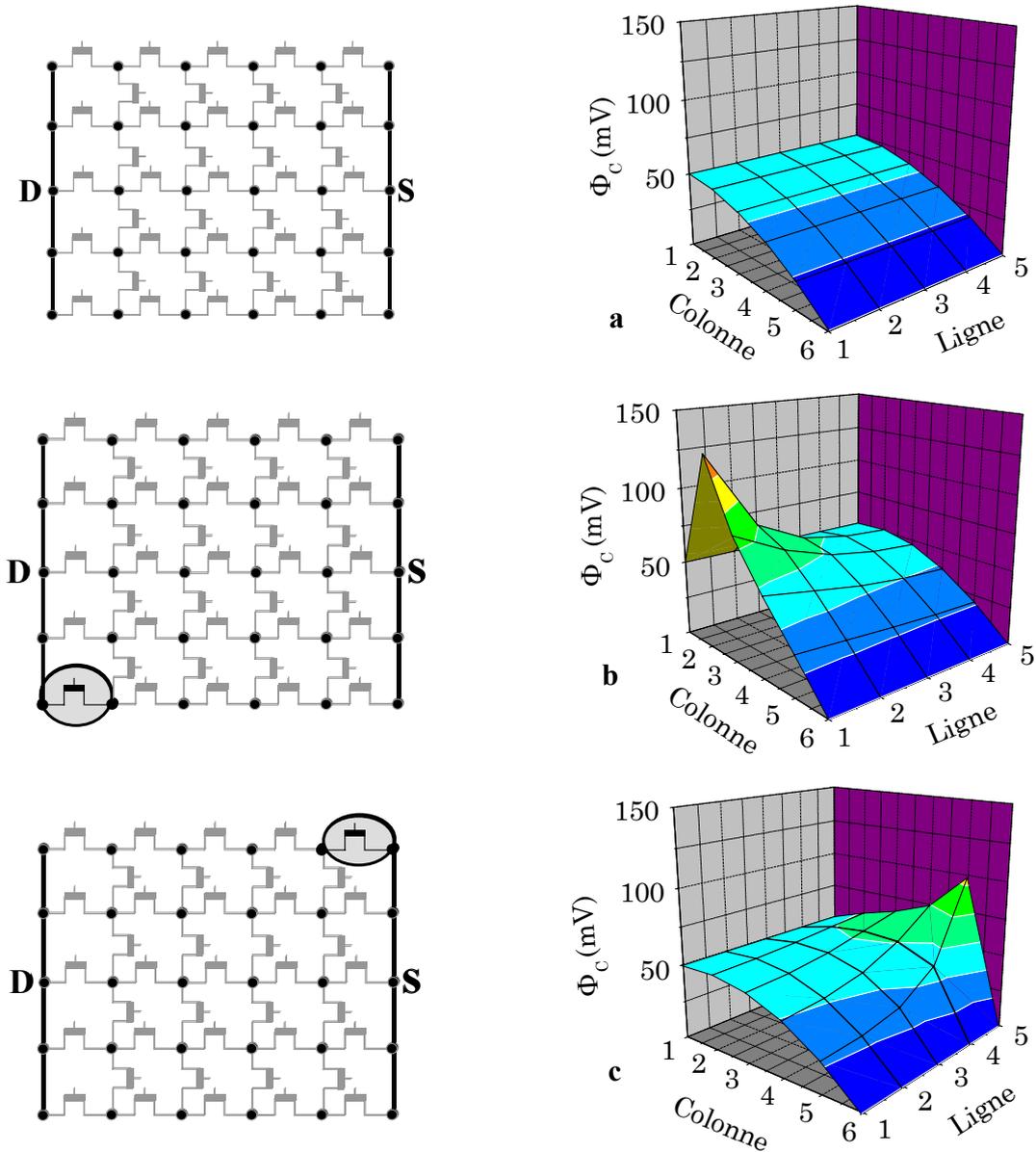


Figure III.28. Simulations de la distribution des quasi-niveaux de Fermi, $\Phi_C(x,y)$ pour un transistor sans défaut avec $t_{ox} = 1.5$ nm (a) et des transistors ayant un défaut soit côté drain (b), soit côté source (c) pour $V_{DS} = 50$ mV et $V_{GS} = 1.2$ V [Bernardini'04c].

Notons que les valeurs du potentiel Φ_C sont imposées au niveau du drain ($\Phi_C(1,y) = V_D$) et au niveau de la source ($\Phi_C(5,y) = V_S$). La figure (III.28.a) montre que la courbe $\Phi_C(x,y)$ est légèrement bombée. Comme nous l'avons décrit précédemment, pour

de faibles tensions V_{DS} ($V_{DS} = 50$ mV, par exemple), la variation des quasi-niveaux de Fermi d'un transistor long avec une couche d'oxyde mince (et par conséquent, des courants de fuite de grille), présente une variation non linéaire. En effet, une partie de la polarisation appliquée sur la grille est appliquée aux bornes de chaque GLNMOS et modifie les variations du potentiel. Notons également que les valeurs de Φ_C restent fixes aux bornes du transistor, car elles sont fixées par la tension V_{DS} . Les figures (III.28.b) et (III.28.c) soulignent la dépendance de l'amplitude du pic de Φ_C suivant la position de la diminution de t_{ox} : avec un pic plus élevé lorsque le défaut est proche du drain. Par conséquent, pour de faibles V_{DS} , le courant de drain sera plus dégradé pour une diminution de t_{ox} plus proche du drain que de la source. Ce phénomène est dû au chemin parcouru par le courant de fuite de grille car, pour un défaut proche du drain, le nombre de transistors ayant un oxyde plus épais sera plus important côté source ; par suite les électrons passant plus facilement à travers une couche d'oxyde plus mince, iront préférentiellement dans le drain.

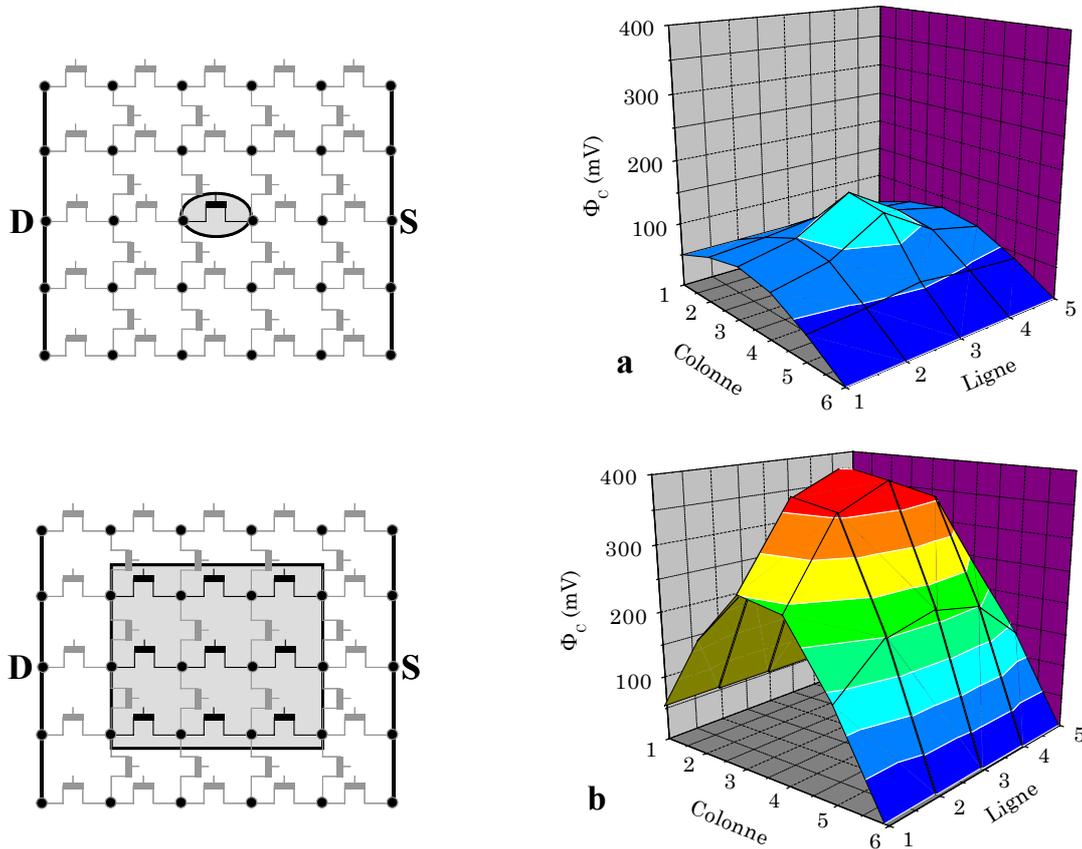


Figure III.29. Simulations, pour $V_{DS} = 50$ mV et $V_{GS} = 1.2$ V, de la distribution des quasi-niveaux de Fermi, $\Phi_C(x,y)$ ayant un seul défaut central (a), ou un défaut neuf fois plus grand (b) [Bernardini'04c].

Lorsque la tension de drain augmente, la variation de $\Phi_C(x,y)$ devient de plus en plus importante le long du canal de la source vers le drain. Par conséquent, un transistor

présentant une épaisseur d'oxyde plus faible côté drain aura un niveau de Fermi $\Phi_C(x,y)$ important côté drain, une faible charge d'inversion Q_n , et par suite un faible courant de fuite, I_G . C'est pourquoi, pour de fortes tensions de drain, la courbe $I_D(V_G)$ d'un transistor présentant une épaisseur d'oxyde plus faible côté drain, tend à rejoindre la courbe $I_D(V_G)$ du transistor de référence (cf. Fig. (III.27)). Finalement, la variation de $\Phi_C(x,y)$ est un bon indicateur des modifications des courants de source et de drain d'un transistor affecté par une non uniformité d'oxyde. Utilisons, à présent, cet indicateur pour étudier l'impact de la taille de cette non uniformité. Le transistor de référence restant le même que pour les simulations de la figure (III.28), nous avons tracé la variation de $\Phi_C(x,y)$ en plaçant un seul (Fig. (III.29.a)) ou neuf (Fig. (III.29.b)) GLNMOS défectueux ($t_{ox} = 1.2$ nm) au centre de la matrice. Ces courbes mettent en évidence à la fois : i) l'impact de la position et de la taille de la réduction d'oxyde, avec une augmentation de $\Phi_C(x,y)$ localisée à l'endroit du défaut (Fig. (III.29.a)) et ii) une augmentation de l'amplitude de $\Phi_C(x,y)$ avec la taille du défaut (Fig. (III.29.b)).

III.6.3.4. Application au cas d'un miroir de courant

Cette étude de la non uniformité de l'épaisseur de l'oxyde de grille a été appliquée au cas simple d'un miroir de courant. A présent, nous plaçons un transistor présentant une réduction d'oxyde dans un miroir de courant (cf. Fig. (III.30)) afin d'étudier l'impact de la non uniformité de t_{ox} sur le comportement du circuit.

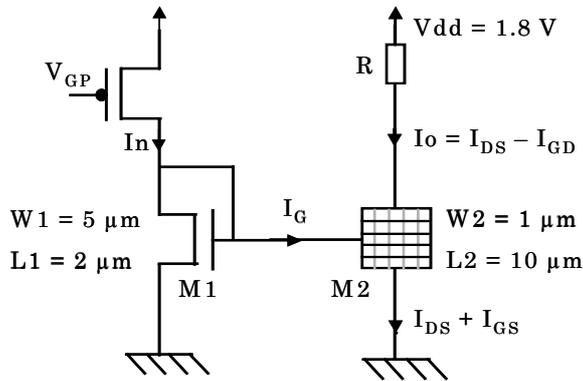


Figure III.30. Schéma électrique du miroir de courant ayant un de ces transistors affecté par une non uniformité de son épaisseur d'oxyde [Bernardini'04d].

Ce miroir de courant est constitué d'un transistor PMOS polarisé par une tension V_{GP} et de deux transistors NMOS M_1 ($W_1 \times L_1 = 5 \times 2 \mu\text{m}^2$) et M_2 ($W_2 \times L_2 = 1 \times 10 \mu\text{m}^2$). Ces dimensions entraînent une réduction du courant de consigne (I_n) par un facteur 25 (cf. Annexe I). Ce circuit peut donc être utilisé pour des applications bas courants. Comme le montre la figure (III.30), seul le transistor NMOS numéro 2 (M_2) présente une réduction localisée de son épaisseur d'oxyde (réduction de 1.5 nm à 1.2 nm). L'utilisation du modèle matricé pour simuler le courant dupliqué, I_o , met en évidence l'impact de la

position et de la taille de la non uniformité de t_{ox} du transistor M_2 . Dans un premier temps, supposons qu'un seul GLNMOS (GLNMOS gris de la figure (III.28)) de notre matrice ait été affecté par cette réduction d'oxyde. Le déplacement de ce GLNMOS permet l'étude de la position de la réduction d'oxyde. La variation relative du courant dupliqué est calculée pour un maximum de courant ($V_{GP} = 0$) à partir du miroir de courant avec un transistor M_2 de référence (I_{0ref}) et un miroir de courant avec un transistor M_2 défectueux ($I_{0defect}$) :

$$S_{I_0} = \frac{I_{0ref} - I_{0defect}}{I_{0ref}} \times 100 \quad (III.23)$$

La figure (III.31) souligne la très faible dépendance du pourcentage d'erreur du courant dupliqué avec la largeur du transistor M_2 (pour $V_{GP} = 0V$). On observe, en revanche, une valeur plus élevée de S_{I_0} pour un défaut plus proche du drain que de la source et un courant dupliqué plus dégradé pour un défaut proche du drain. Comme nous l'avons déjà expliqué, ce phénomène est dû au chemin du courant de fuite de grille.

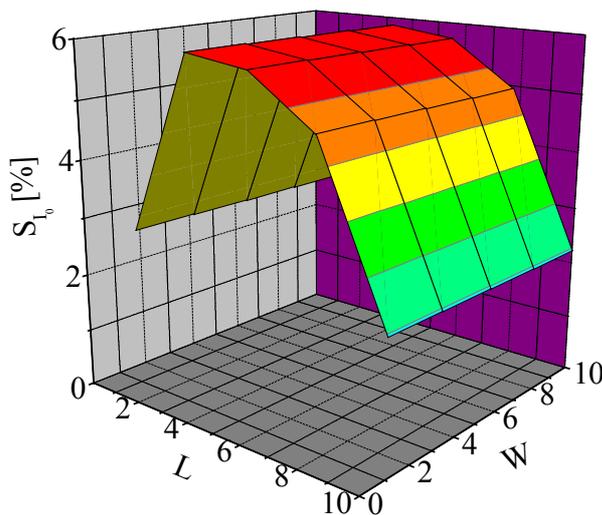


Figure III. 31. *Distribution de la variation relative du pourcentage d'erreur du courant dupliqué en fonction de la position de la réduction de l'épaisseur d'oxyde du transistor M_2 [Bernardini'04d].*

Ainsi, la position de la non uniformité de t_{ox} intervient sur le fonctionnement du miroir de courant. Un autre paramètre non négligeable est l'espace occupé par cette réduction de t_{ox} . Le transistor de référence M_2 de la figure (III.28) est à nouveau utilisé pour étudier l'impact de la taille du défaut.

Prenons un GLNMOS défectueux au centre de la matrice et comparons les résultats à ceux obtenus avec 5 GLNMOS défectueux (cf. Fig. (III.32)).

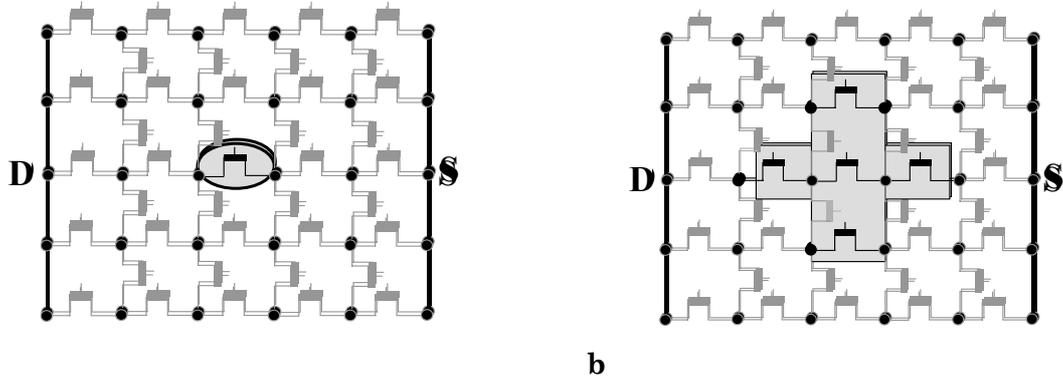


Figure III.32. Schématisation des transistors M2 avec deux distributions spatiales différentes de la réduction de t_{ox} .

Considérons le miroir de courant avec le transistor M2 schématisé sur la figure (III.32.a). Pour $V_{GP} = 0$ V, l'erreur relative sur le courant dupliqué est égale à 5.72% tandis que celle correspondant au transistor M2 de la figure (III.32.b) est plus de 4 fois supérieure. Cette augmentation est liée à l'étendue du défaut qui lorsqu'elle s'accroît entraîne une augmentation des fuites du courant vers la grille.

La figure (III.33.a) montre la variation du courant dupliqué dans le cas du transistor M2 schématisé sur la figure (III.32.b). Cependant le tracé du courant de référence I_n dans le cas du miroir de courant avec un transistor M2 de référence et un miroir de courant avec un transistor M2 défectueux (Fig. (III.33.b)) ne montre pas de différence. En réalité celle-ci existe mais n'est pas visible à cause du facteur 25 qui lie le courant dupliqué et le courant de référence.

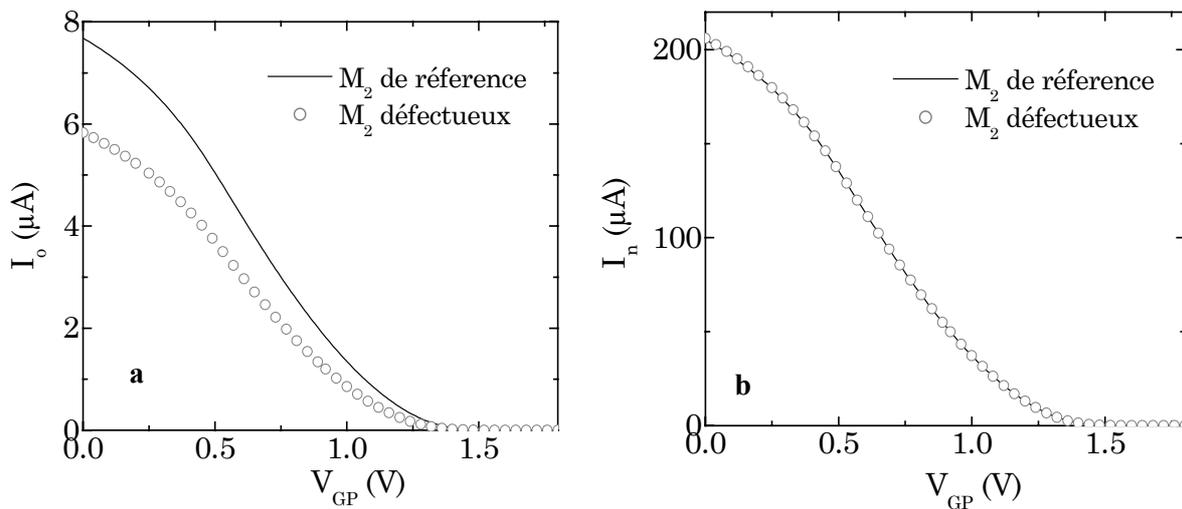


Figure III.33. Comparaison entre les courants dupliqués (a) et des courants de consigne (b) pour des miroirs de courant avec un transistor M2 de référence ou un transistor M2 défectueux (Fig. (III.32.b)) [Bernardini'04d].

Enfin, pour compléter notre étude, nous avons étudié l'impact d'une réduction de t_{ox} de grande taille (6 GLNMOS situés à divers endroits de la surface du transistor). Les résultats obtenus sont regroupés sur la figure (III.34).

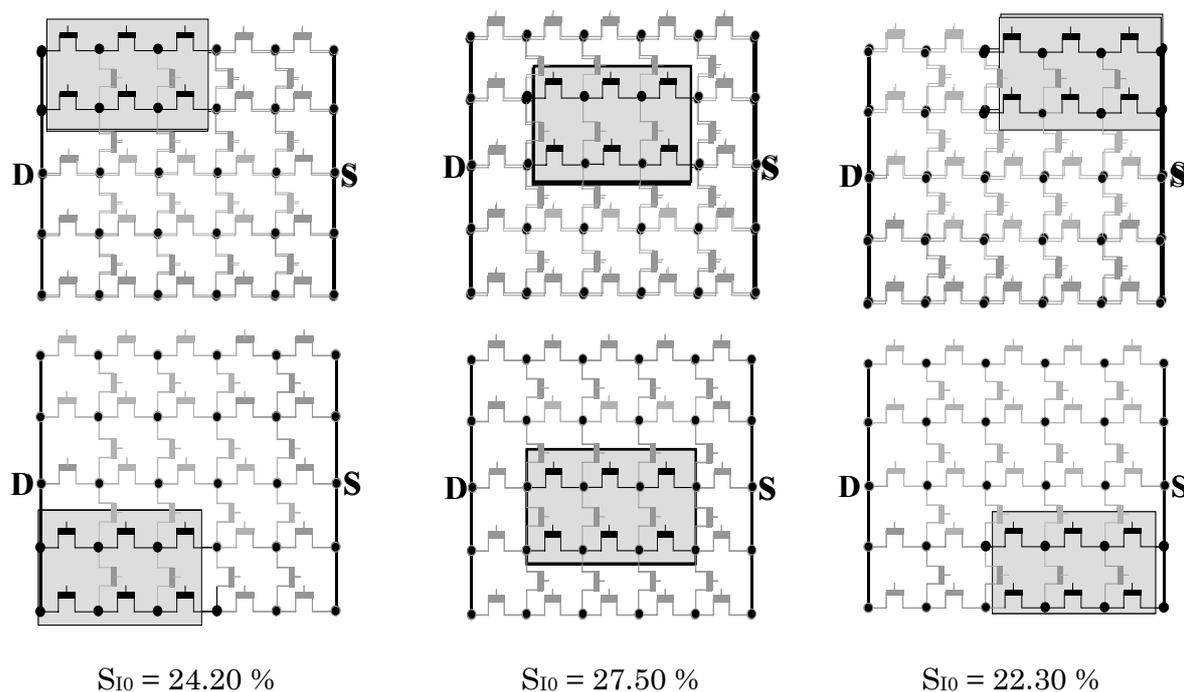


Figure III.34. Schématisation des transistors M2 affectés par une réduction de t_{ox} et pourcentages d'erreur sur le courant dupliqué I_o (pour $V_{GP} = 0V$).

Nous obtenons une combinaison des résultats précédemment décrits, à savoir :

- le coefficient S_{I0} est indépendant de la largeur W du transistor,
- le coefficient S_{I0} augmente avec la taille du défaut,
- le coefficient S_{I0} dépend de la position du défaut (S_{I0} est plus important pour un défaut placé côté drain, par rapport à un défaut placé côté source).

Enfin, nous pouvons constater que la dépolarisation du transistor au centre du canal, induite par le courant de fuite de grille, entraîne l'augmentation du coefficient S_{I0} pour un défaut situé au centre de la couche d'oxyde.

III.7. Conclusion

Dans ce chapitre, différentes applications du modèle segmenté ont montré l'impact des non uniformités des transistors sur leurs caractéristiques I-V. Ce modèle pseudo 2D est valide du régime d'inversion faible (ohmique ou non ohmique) au régime d'inversion forte. Une étude très détaillée de l'impact des résistances séries du transistor a montré d'une part que la réduction de la mobilité apparente du transistor dépendait de la valeur

des résistances séries de celui-ci, et d'autre part, qu'un seul réseau de courbes $I_{DS}(V_{GS}, V_{DS})$ pouvait être suffisant pour déterminer la valeur globale des résistances séries du transistor. A l'aide de notre modèle, nous avons mis en évidence un impact moins important de la poly-désertion de grille sur les transistors par rapport à celui observé sur les capacités MOS. Le modèle segmenté a également été utilisé pour étudier le phénomène inverse de canaux courts (Reverse Short Channel Effect) qui provient principalement d'une non uniformité du dopage de substrat des transistors. Enfin, nous avons montré la dépendance de la dégradation du courant des transistors ultra-minces avec la position et la taille des non uniformités de l'épaisseur d'oxyde.

Chapitre IV

Etude des structures à nano-cristaux de silicium

Chapitre IV. Etude des structures à nano-cristaux de silicium	131
IV.1. Introduction	131
IV.2. Structures d'étude	131
IV.3. Modélisation d'une charge non uniformément répartie	133
IV.3.1. Modélisation pour un MOSFET	133
IV.3.2. Modélisation d'une mémoire à nodules	135
IV.4. Simulations des structures 1bit	137
IV.4.1. Simulations statiques des mémoires Flash à nodules	138
IV.4.2. Etude de la phase d'écriture des mémoires Flash à nodules	139
IV.4.2.1. Variation de la charge stockée	139
IV.4.2.2. Décalage de la tension de seuil	140
IV.5. Caractérisations électriques de structures avec nodules	141
IV.5.1. Etude des capacités avec nodules	142
IV.5.1.1. Etude expérimentale	142
IV.5.1.2. Modélisation des capacités à nodules	142
IV.5.1.3. Impact de la densité, N_{dot} , et du diamètre, D_{dot} , des nodules	143
IV.5.1.4. Extraction des paramètres des capacités ring	144
IV.5.2. Etude des transistors avec nodules	146
IV.5.2.1. Etude des temps d'écriture	146
IV.5.2.2. Etude des tensions d'écriture	147
IV.5.2.2. Etude de la tension de lecture	149
IV.6. Conclusion	152

Chapitre IV. Etude des structures à nano-cristaux de silicium

IV.1. Introduction

Au cours de ces dix dernières années, les nano-cristaux de silicium ont été introduits dans les mémoires Flash afin d'améliorer leurs performances : réduction de dimensions, robustesse par rapport aux défauts de l'oxyde, basse consommation en puissance, bas coût de fabrication, etc... L'objectif de ce chapitre est de présenter une modélisation du comportement électrique des mémoires à nodules de silicium lors de leur écriture en fonction des paramètres des dots. Dans un premier temps, nous avons étudié l'effet de la répartition non uniforme de la charge stockée sur les caractéristiques électriques de ces dispositifs, puis nous avons développé un modèle pseudo 2D de l'opération d'écriture de la mémoire par porteurs chauds. La dernière partie de ce chapitre est consacrée à la caractérisation électrique de structures à nodules fabriquées par la société STMicroelectronics Catagne dans le cadre du projet européen ADAMANT (Advanced Memories based on Discrete Traps).

IV.2. Structures d'étude

Les nodules considérés dans notre étude sont des nano-cristaux de silicium séparés les uns des autres par du SiO_2 . Ils remplacent la grille flottante conventionnelle des mémoires Flash et sont constitués de demi-sphères en silicium offrant leur section la plus large à l'injection tunnel (cf. Fig. (IV.1)).

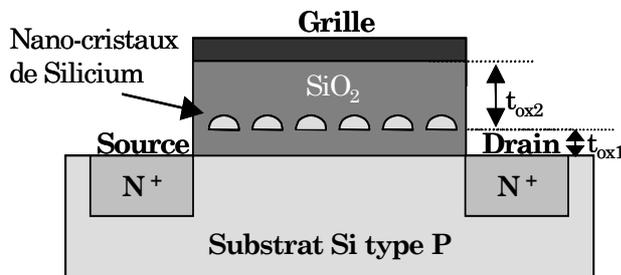


Figure IV.1. Coupe schématique du dispositif mémoire Flash comportant des nodules de silicium.

Ces îlots sont supposés suffisamment espacés les uns des autres, pour empêcher le mouvement latéral des charges et préserver la mémoire d'une perte totale de

l'information lors d'une détérioration locale de l'oxyde. De surcroît, ces nodules sont supposés uniformément répartis dans l'oxyde du transistor entre l'oxyde tunnel d'épaisseur (t_{ox1}) et l'oxyde de contrôle d'épaisseur (t_{ox2}). Comme le montre la figure (IV.1), seule une faible portion R_{eff} de la surface totale de l'oxyde de grille est occupée par les nano-cristaux de silicium. Ce coefficient R_{eff} est proportionnel à la densité (N_{dot}) et au diamètre (D_{dot}) des nodules :

$$R_{eff} = N_{dot} \pi \left(\frac{D_{dot}}{2} \right)^2 \quad (IV.1)$$

Comme nous l'avons rappelé dans le chapitre I, l'opération d'écriture des mémoires Flash à nodules consiste à stocker des électrons dans les nano-cristaux, soit par le biais du mécanisme Fowler–Nordheim (FN), I_{FN} , qui est assez lent (de l'ordre de la micro-seconde) et uniforme (cf. Fig. (IV.2.a)), soit par le mécanisme de porteurs chauds (CHE) [Tam'84], qui est plus rapide (facteur 10) mais plus coûteux en terme d'énergie car le transistor doit être polarisé en régime de saturation (V_{DS} fort) pour créer des paires électrons-trous (par ionisation par impact). Toutefois, l'écriture par porteurs chauds permet un accès sélectif à un seul point mémoire (cf. Fig. (IV.2.b)) puisque la création de paires électrons-trous est localisée dans une zone proche du drain ou de la source suivant la polarisation ($V_D - V_S$) appliquée [Eitan'00]-[Lusky'01]. Les trous ainsi générés se dirigent vers le substrat, tandis que les électrons suivent différents chemins : sous l'influence du champ électrique entre l'interface et la grille, une partie des électrons traverse l'isolant de grille du transistor pour atteindre soit la grille, soit les nodules (qu'ils chargent) ; la majeure partie des électrons générés par ionisation rejoignent le drain (ou la source suivant leur point de départ), sous l'influence du champ électrique entre les extrémités du canal.

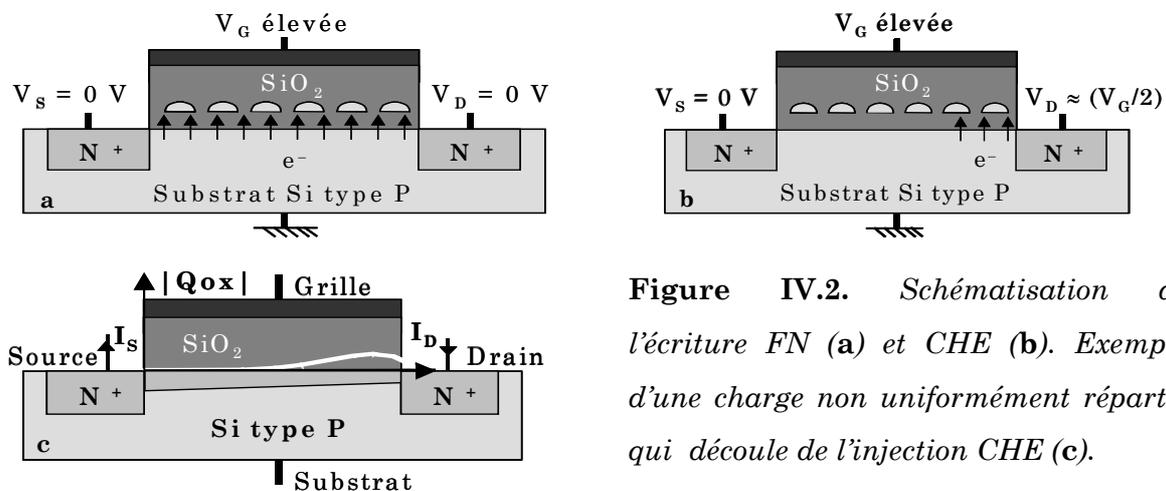


Figure IV.2. Schématisation de l'écriture FN (a) et CHE (b). Exemple d'une charge non uniformément répartie qui découle de l'injection CHE (c).

Par exemple, la figure (IV.2.c) présente une schématisation de la non uniformité de charges qui découlerait de charges stockées uniquement à l'intérieur de quelques nodules proches du drain. Cette possibilité de chargement discret, localisé près d'une jonction (drain et/ou source) offre la possibilité d'une logique 2 bits pour les mémoires à nodules [Bloom'01], c'est à dire quatre états possibles :

- aucun nodule chargé,
- tous les nodules chargés par injection Fowler-Nordheim (charge uniformément répartie dans les nodules),
- des charges présentes dans quelques nodules proches de la source (injection localisée par porteurs chauds).
- des charges présentes dans quelques nodules proches du drain (injection localisée par porteurs chauds).

IV.3. Modélisation d'une charge non uniformément répartie

IV.3.1. Modélisation pour un MOSFET

Considérons les profils de charges en exponentielle décroissante tracés sur la figure (IV.3.a). La charge stockée au dessus du canal modifie la conductivité du canal du transistor par l'intermédiaire de la variation de la tension de bandes plates, V_{FB} (cf. Fig. (IV.3.b)).

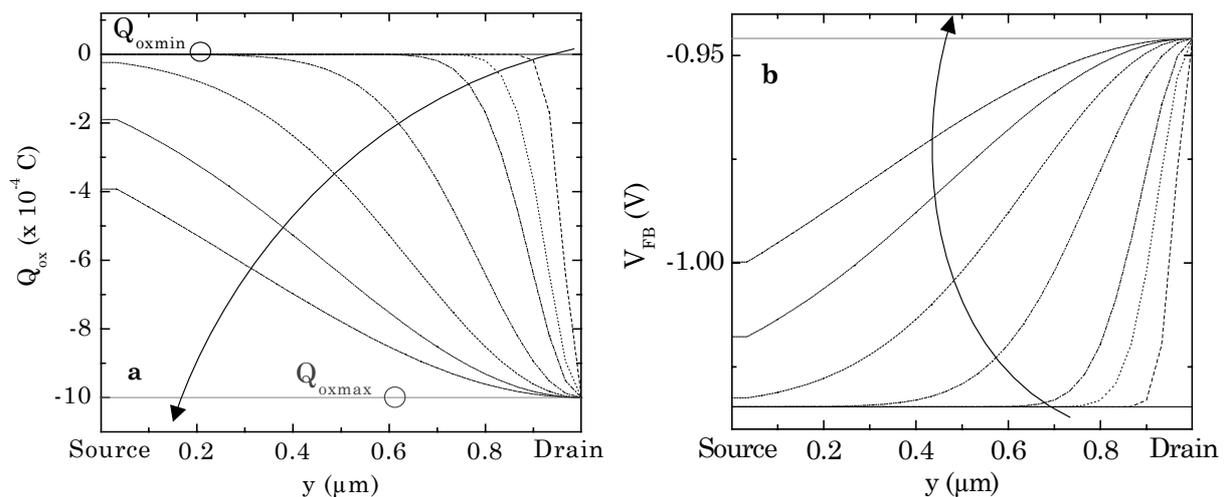


Figure IV.3. Profils de charges non uniformément réparties dans l'oxyde (a) et variations des tensions de bandes plates correspondantes (b).

Comme nous l'avons rappelé au chapitre III, les modélisations de type Pao et Sah et en feuillet ne prennent pas en compte ce type de non uniformité dans le développement de l'expression du courant de drain. Le modèle segmenté, développé lors de nos travaux sur les non uniformités des transistors (voir chapitre III), permet de surmonter cette difficulté. Par conséquent, pour simuler le courant de drain des transistors ayant les profils de charges reportés sur la figure (IV.3.a), nous avons adapté le modèle segmenté en considérant la structure comme équivalente à la juxtaposition de N transistors (de longueur L/N) ayant chacun une charge fixe constante dans l'oxyde mais pouvant être différente d'un segment à l'autre. Le courant de drain des N transistors élémentaires est évalué à l'aide du modèle en feuillet [Brews'78]. A des tensions de grille (V_{GS}) et de drain (V_{DS}) fixées, le potentiel de surface (et par conséquent l'écart entre les quasi-niveaux de Fermi, Φ_C), est calculé pour chaque transistor en supposant un flux de courant conservatif le long du canal. Le système de N équations à $N-1$ inconnues peut alors être résolu par la méthode du pont diviseur. Cependant, comme nous l'avons déjà signalé, l'utilisation du pont diviseur n'est valide que si la somme de tous les courants tunnels est négligeable par rapport au courant de drain I_{DS} . Pour les structures considérées, ces conditions sont respectées puisque le courant d'injection est très petit devant I_{DS} et que les fuites de grille sont considérées comme négligeables.

Les figures (IV.4) montrent le décalage de la courbe $I_{DS}(V_{GS})$ d'un transistor NMOS obtenu pour une tension de drain $V_D = 50$ mV (avec $V_S = 0$ V). Ce décalage résulte de la présence de charges, Q_{ox} , qui correspondent aux profils présentés à la figure (IV.3.a).

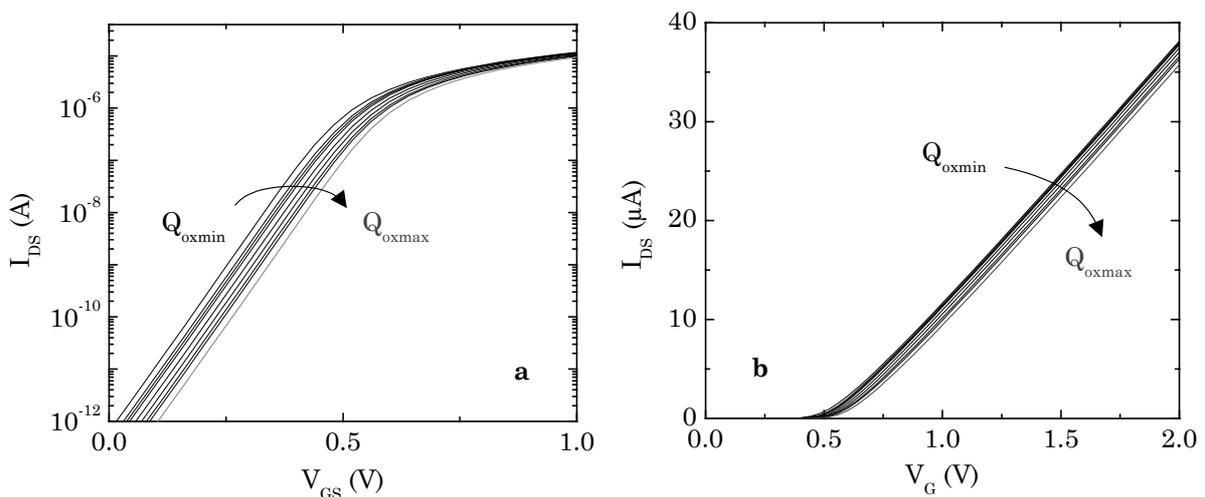


Figure IV.4. Variations du courant de drain en présence de charges non uniformément réparties (cf. Fig. (IV.3.a)) tracées en échelle semi-logarithmique (a) et en échelle linéaire (b). Les paramètres du MOSFET sont : $W = 1 \mu m$, $L = 1 \mu m$, $t_{ox1} = 3$ nm et $V_{DS} = 50$ mV.

Comme pour une variation du dopage de substrat entre le drain et la source, le décalage des caractéristiques $I_{DS}(V_{GS})$ est continu entre la courbe correspondant à une charge nulle et celle correspondant à une charge maximale.

IV.3.2. Modélisation d'une mémoire à nodules

Dans ce paragraphe, nous ne considérerons que l'écriture, par porteurs chauds, des mémoires Flash à nodules, puisque ce mécanisme permet une injection localisée des électrons à partir du canal.

Pour déterminer la charge stockée dans les nodules, la valeur de Φ_C doit être calculée le long du canal. Comme pour un transistor, la mémoire à nodules peut-être supposée équivalente à N transistors juxtaposés de longueur L/N (cf. Fig. (IV.5)).

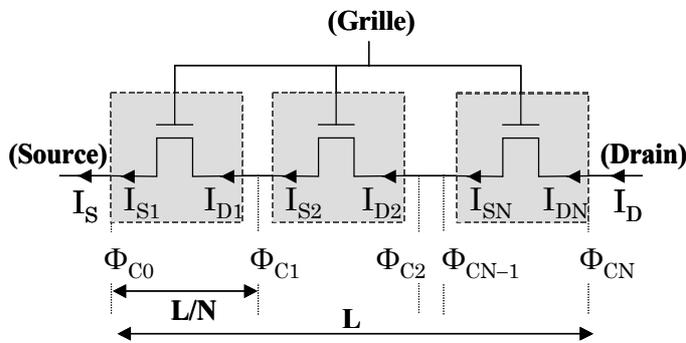


Figure IV.5. Vue schématique de la mémoire à nodules découpée en N transistors.

Ainsi, ce modèle permet la modulation de la longueur de chaque transistor et plus particulièrement sa réduction dans la zone d'injection (près du drain de la mémoire). Notons que cette approche offre d'autres possibilités comme la simulation du courant de drain pour des mémoires ayant des charges stockées à la fois près du drain et de la source (mémoires 2 bits [Eitan'99],[Eitan'00]).

La charge totale Q_{dot} (exprimée en Coulombs) étant stockée dans les nodules enfermés dans la couche d'oxyde d'épaisseur t_{ox2} , le potentiel de surface aux frontières de chaque transistor est alors évaluable à partir de l'équation suivante :

$$V_{GB} = \left(V_{FB} + \Psi_S - \frac{Q_{SC}}{C_{ox}} \right) - \frac{1}{C_{ox}} \frac{Q_{dot}}{W(L/N)} \frac{t_{ox2}}{t_{ox1} + t_{ox2}} \quad (IV.2)$$

où W est la largeur de la mémoire.

Cette équation (IV.2) suggère que l'influence électrostatique des charges stockées dans les nodules est étalée sur la surface totale de chaque transistor élémentaire (comme pour la modélisation des pièges dans un diélectrique).

Le chargement des îlots de silicium s'effectuant par le biais de porteurs chauds, qui apparaissent près du drain lorsque le transistor MOS est polarisé en régime saturé (forte tension V_{DS}), il est nécessaire d'évaluer le champ électrique dans la zone saturée, le courant d'ionisation par impact et par suite, le courant qui traverse l'isolant tunnel.

Usuellement, le canal est dit pincé lorsque la charge d'inversion devient pratiquement nulle au niveau du drain. Cependant, pour une tension de grille donnée, la localisation du point de pincement dépend de la tension V_{DS} . En effet, une augmentation de la polarisation appliquée sur le drain entraîne une augmentation de la largeur de la zone de charge d'espace (ZCE) de la jonction drain/substrat vers le canal (qui est moins dopé que le drain). Par suite, lorsque le potentiel de saturation V_{DSsat} est dépassé l'accroissement de la ZCE conduit à un déplacement du point de pincement vers la source. Dans notre modèle, nous supposons que le point de pincement est atteint lorsque la charge d'inversion dans le canal devient négligeable devant la charge d'inversion de la source [Laffont'03a]. Ainsi, pour des tensions de drain et de grille données, la zone de saturation est localisable le long du canal en recherchant le quasi-niveau de Fermi Φ_{Csat} tel que :

$$Q_n(\Phi_{Csat}, V_{GB}) = \frac{Q_n(V_{SB}, V_{GB})}{FAC} \quad (IV.3)$$

où FAC est un paramètre d'ajustage (typiquement égal à 10).

Par conséquent, le potentiel de saturation V_{DSat} est obtenu en fonction de la résolution spatiale de Φ_C et permet de connaître à la fois la localisation (dans la zone de pincement) et le nombre (N_{pinch}) de transistors élémentaires dont les dots se remplissent par porteurs chauds.

Comme nous l'avons rappelé dans le chapitre I, le courant injecté, I_{CHE} , est fonction du courant de substrat I_{sub} , provenant de l'ionisation par impacts exprimée par l'expression (I.49) dans le modèle de Tam [Tam'84]. Toutefois, contrairement aux mémoires à grille flottante conventionnelle, seule une partie de la charge injectée peut être stockée puisque les nodules ne recouvrent qu'une portion de la surface de l'interface entre les couches d'oxyde t_{ox1} et t_{ox2} . Par suite, le coefficient (R_{eff}) représentant la surface occupée par les nodules a été introduit dans l'expression du courant injecté. En conséquence, en supposant que l'injection soit uniforme dans la zone à saturation pour chaque transistor localisé dans la région de pincement, le courant réellement injecté est donné par la relation suivante :

$$I_W = I_{\text{sub}} \alpha_{\text{ox}} \frac{R_{\text{eff}}}{N_{\text{pinch}}} \exp\left(\frac{b_{\text{ox}}}{\xi_{\text{ox}t1}}\right) \quad (\text{IV.4})$$

où b_{ox} et α_{ox} sont les deux paramètres d'injection et I_{sub} représente le courant substrat décrit par le modèle de Schokley Read Hall :

$$I_{\text{sub}} = I_{\text{DS}} \frac{a_i}{b_i} V_{\text{sat}} \exp\left(-\frac{b_i}{V_{\text{sat}}}\right) \quad (\text{IV.5})$$

où a_i , b_i sont les coefficients d'ionisation par impact et V_{sat} le potentiel appliqué aux bornes de la zone saturée.

Notons que pour des mémoires Flash conventionnelle, les coefficients a_i , b_i , α_{ox} et b_{ox} sont déterminés à partir des caractéristiques statiques $I_{\text{SUB}}(V_{\text{GS}}, V_{\text{DS}})$ et $I_{\text{G}}(V_{\text{GS}}, V_{\text{DS}})$ mesurées sur une cellule dont la grille de contrôle et la grille flottante sont reliées entre elles.

Le paramètre $\xi_{\text{ox}t1}$ correspondant au champ électrique dans le diélectrique tunnel, entre l'interface et les dots, est donné par l'expression suivante :

$$\xi_{\text{ox}t1} = \frac{-V_{\text{ox}}}{t_{\text{ox}1} + t_{\text{ox}2}} - \frac{t_{\text{ox}2}}{t_{\text{ox}1} + t_{\text{ox}2}} \frac{Q_{\text{dot}}}{\epsilon_{\text{ox}} W(L/N) R_{\text{eff}}} \quad (\text{IV.6})$$

où V_{ox} est le potentiel diélectrique total et Q_{dot} résulte de l'intégration du courant d'écriture :

$$\Delta Q_{\text{dot}} = I_W \times \Delta t \quad (\text{IV.7})$$

où Δt , correspond au pas du temps d'écriture.

Remarquons que, pour les mémoires Flash à nodules, à une tension de grille donnée, l'injection dans les nodules est stoppée lorsque le champ électrique, $\xi_{\text{ox}t1}$, tend vers zéro, même si le courant par porteurs chauds I_{CHE} existe toujours. Dans ce cas, la barrière de potentiel n'est plus assez déformée pour laisser passer les électrons par effet tunnel FN. Cette notion est très différente de celle des mémoires Flash traditionnelles pour lesquelles l'opération d'écriture s'arrête lorsque le courant I_{CHE} devient négligeable (le nombre d'électrons injectés entraîne la « dé-saturation » du transistor). De même, lors de l'effacement de la mémoire à nodules, le champ à considérer dans l'expression du courant correspond à $\xi_{\text{ox}t1}$.

IV.4. Simulations des structures 1bit

Afin de simplifier notre modélisation pour une étude 1 Bit, nous avons supposé que la mémoire à nodules était équivalente à un transistor coupé en deux parties : l'une proche

du drain, de longueur X_D , correspondant à la région de chargement et l'autre près de la source de longueur $L - X_D$, correspondant à la partie non chargée du transistor (cf. Fig. (IV.6)).

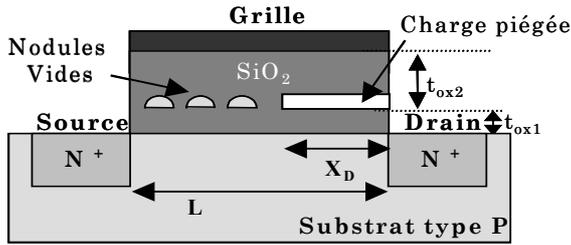


Figure IV.6. Coupe schématique du dispositif mémoire Flash comportant deux parties : une chargée et une non chargée.

Notons que lors de cette étude du chargement des nodules, les effets de canaux courts ne sont pas pris en compte.

IV.4.1. Simulations statiques des mémoires Flash à nodules

La figure (IV.7) montre les courbes $I_{DS}(V_{GS})$ simulées pour un transistor NMOS. Le décalage observé est dû à la présence d'une forte densité de charges, N_Q , localisée dans l'isolant (à 5 nm de l'interface) sur une longueur X_D proche du drain. Pour V_{GS} donnée, cette forte densité de charges induit un fort décalage de la pente sous le seuil pour les transistors chargés sur une petite zone ($X_D < 10^{-2} L$) par rapport au transistor sans charge (c.a.d. $X_D = 0$) (cf. Fig. (IV.7.b)). Pour $0.1 L < X_D < L$, le décalage est faible. En régime d'inversion forte (voir Fig. (IV.7.a)), on constate la disparition de la double pente (due au changement de V_{FB}) lorsque la longueur de la zone chargée augmente.

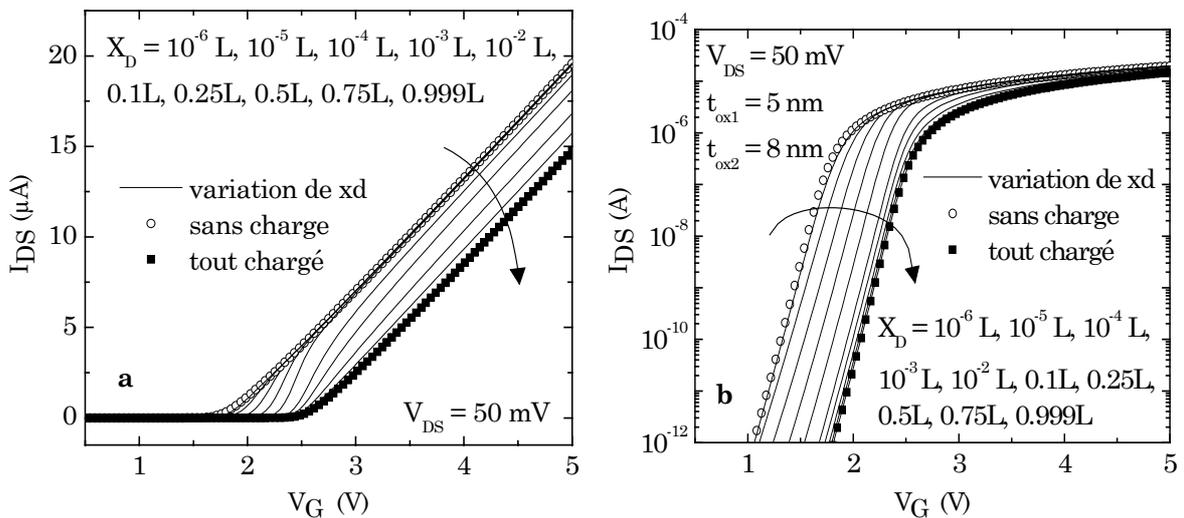


Figure IV.7. Simulation de la courbe $I_{DS}(V_{GS})$ en fonction de la distance X_D proche du drain où sont localisées les charges fixes, en échelle linéaire (a) et en échelle logarithmique (b). Les paramètres de la simulation sont : $W = 1 \mu\text{m}$, $L = 1 \mu\text{m}$, $t_{ox1} = 5 \text{ nm}$, $t_{ox2} = 8 \text{ nm}$, $N_Q = 2 \times 10^{16} \text{ C.m}^{-2}$ et $V_{DS} = 50 \text{ mV}$.

IV.4.2. Etude de la phase d'écriture des mémoires Flash à nodules

IV.4.2.1. Variation de la charge stockée

Dans un souci de simplification, nous montrons ici les résultats obtenus après découpage du transistor en deux transistors élémentaires. Même si pour une tension de grille donnée, le point de pincement se décale le long du canal en fonction de la polarisation appliquée sur le drain, nous supposons que l'injection du courant par porteurs chauds ne s'effectue que dans le second transistor (de longueur $X_D = 0.15 \times L$).

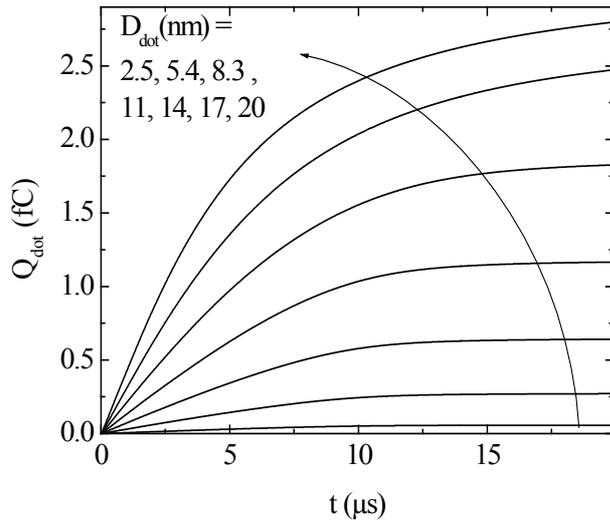


Figure IV.8. Variation de la charge stockée, Q_{dot} , dans les nodules pendant l'écriture de la mémoire en fonction du diamètre des nodules. Les paramètres de la simulation sont : $W = 1 \mu m$, $L = 1 \mu m$, $t_{ox1} = 3 nm$, $t_{ox2} = 5 nm$, $V_{GB} = 5 V$, $V_{DS} = 3.5 V$, $X_D = 0.15 \times L$, $N_{dot} = 2 \times 10^{15} m^{-2}$ [Bernardini'03c].

Nos simulations ont montré que le nombre d'électrons injectés par dot dépend des dimensions de ces derniers et de la durée de l'opération d'écriture. Un exemple de simulation dynamique des charges piégées dans les nodules (Q_{dot}) durant une injection CHE de $20 \mu s$ pour des nodules de différents diamètres est présenté sur la figure (IV.8). Cette figure met en évidence le lien direct entre la taille des nodules, le nombre d'électrons stockés et la durée de l'injection. Pour un temps d'écriture donné, l'augmentation de la taille des nodules induit une augmentation du nombre d'électrons stockés. Avec les paramètres choisis pour nos simulations, en imposant des conditions de polarisation, des temps d'écriture identiques et une densité de nodules faible ($2 \times 10^{11} cm^{-2}$), le nombre d'électrons stockés à la fin du temps d'écriture est respectivement de 1.2 et 58.3 électrons pour des diamètres de nodules de 2.5 nm et 20 nm. Par conséquent, pour conserver la même charge stockée lorsque le diamètre du nodule diminue, leur densité doit être augmentée.

IV.4.2.2. Décalage de la tension de seuil

Pour une tension de drain donnée, la détermination du décalage de la tension de seuil est réalisée à courant constant I_{test} (la valeur est choisie dans la pente sous le seuil) à l'aide d'un algorithme de calcul décrit sur la figure (IV.9). Pour une caractéristique $I_{\text{DS}}(V_{\text{GS}})$ donnée, cet algorithme permet de trouver les points pour lesquels les courants sont respectivement immédiatement supérieur et inférieur à la valeur I_{test} . Celle-ci étant choisie pour le régime de diffusion (pente sous le seuil), la droite qui joint les deux points encadrant la valeur de I_{test} est de la forme :

$$\text{Log } I_{\text{DS}} = a V_{\text{GS}} + b \quad (\text{IV.8})$$

La pente a et l'ordonnée à l'origine b de cette droite sont déterminées à partir des coordonnées des points trouvés autour de I_{test} . Par suite, la valeur de V_{GS} correspondant à I_{test} est obtenue, ainsi que le décalage de tension de seuil, pour des tensions de drain fixées ($\Delta V_{\text{th}} \cong \Delta V_{\text{GS}}$).

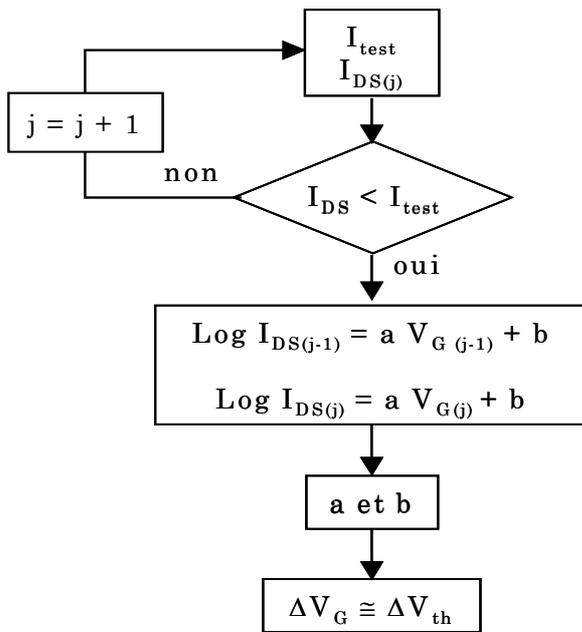


Figure IV.9. Schématisation de l'algorithme de calcul pour la détermination du décalage de la tension de seuil.

Les simulations $I_{\text{DS}}(V_{\text{GS}})$ effectuées pour différentes tensions V_{DS} présentées à la figure (IV.10), mettent en évidence l'influence de la tension appliquée sur le drain par rapport à l'amplitude du décalage de la tension de seuil, ΔV_{T} (déterminée à partir du décalage de la pente sous le seuil pour un courant I_{DS} fixé), pour différentes polarisations de la grille et différents diamètres de dots (D_{dot}). Cette figure (IV.10) montre également que pour une densité de dots fixée, N_{dot} , une petite réduction de leur diamètre entraîne une importante diminution de la tension de seuil d'écriture puisqu'il y a moins d'électrons piégés par nano-cristaux. De plus, la charge piégée augmente avec la polarisation de la grille (tant que le MOSFET est en régime de saturation) puisque la

quantité d'électrons injectés est plus importante. Le mécanisme de chargement s'arrête lorsque le champ électrique dans l'oxyde tunnel devient négligeable : $\Delta V_T(V_{DS})$ tend à saturer pour des valeurs de V_{DS} plus élevées.

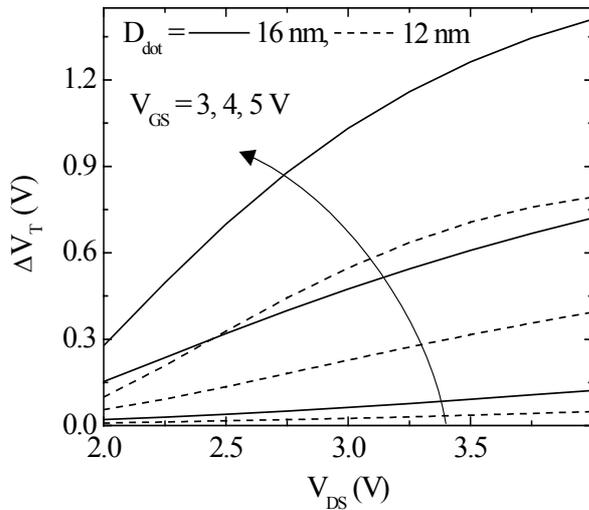


Figure IV.10. Simulation de l'évolution de la tension de seuil en fonction de la tension appliquée sur le drain pour différentes tensions de grille et des dots de différents diamètres. Les paramètres du transistor sont les mêmes que ceux reportés sur la figure (IV.8) avec une période d'écriture de $10\mu s$ [Bernardini'03c].

IV.5. Caractérisations électriques de structures avec nodules

Récemment, de nouveaux procédés de fabrication des mémoires à nodules ont permis une amélioration du contrôle de la densité (N_{dot}) et du diamètre (D_{dot}) des nodules de silicium [De Salvo'03]. Dans la suite de ce chapitre, nous présenterons l'étude menée sur des dispositifs décrits dans cette publication, à savoir, quatre demi-plaques de silicium ayant toutes des nodules de silicium de tailles et de densités différentes permettant de garder cependant un coefficient R_{eff} constant d'environ 25% (voir tableau (IV.1)).

Demi-plaque	N_{dot} ($10^{15} m^{-2}$)	D_{dot} (nm)	R_{eff} calculé (%)
1	16	4.5	25.4
2	9.6	5.5	22.8
3	4	8.5	22.7
4	2.8	10	22

Tableau IV. 1. Récapitulatif des caractéristiques des 4 demi-plaques fabriquées par ST Microelectronics Catagne [De Salvo'03].

Le manque de plaque témoin (sans nodule) de ce lot, nous a orienté vers des études comparatives entre plaques.

IV.5.1. Etude des capacités avec nodules

Dans un premier temps, nous nous sommes intéressés aux capacités avec nodules pour extraire ou vérifier les paramètres caractéristiques de ces structures.

IV.5.1.1. Etude expérimentale

Différentes mesures capacitatives quasi-statiques ont été réalisées sur des capacités à nodules de différentes surfaces, présentes sur les quatre demi-plaques dont nous disposons. La figure (IV.11) présente les mesures quasi-statiques obtenues pour des capacités, de surface $A_{\text{eff}} = 3 \cdot 10^{-3} \text{ cm}^2$, situées dans trois régions différentes des plaques (au milieu, à droite et à gauche). Le faible décalage entre les valeurs maximales des capacités mesurées au milieu, à droite et à gauche des demi-plaques, témoigne de la bonne uniformité des plaques. Pour les tensions de grille positives, on observe un début de chute de la valeur de la capacité puis la courbe tend à saturer.

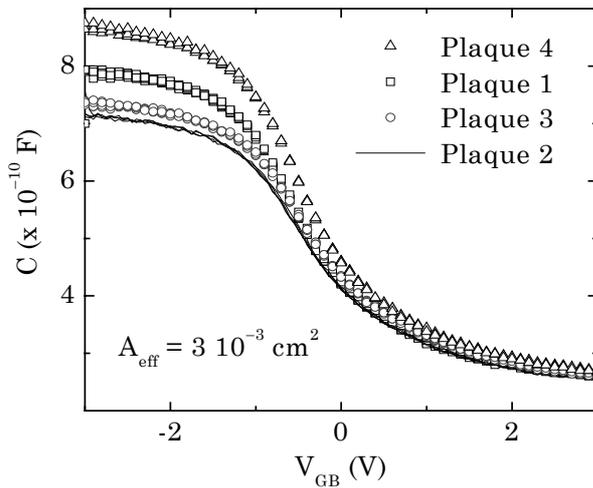


Figure IV.11. Mesures C - V quasi-statiques effectuées sur les capacités à nodules de surface $A_{\text{eff}} = 3 \cdot 10^{-3} \text{ cm}^2$.

On peut également remarquer que l'ordre des courbes tracées ne correspond pas à la réduction de R_{eff} donnée dans le tableau (IV.1), puisque les courbes correspondant à la plaque 1 sont situées entre celles de la plaque 3 et de la plaque 4.

IV.5.1.2. Modélisation des capacités à nodules

Afin d'expliquer l'ordre des courbes C - V mesurées, nous avons modifié les programmes développés lors de l'étude de la capacité MOS pour simuler des capacités à nodules (C_{dot}) en tenant compte du facteur R_{eff} :

$$C_{\text{dot}} = \frac{A_{\text{eff}}}{\frac{1}{C_{\text{SC}}} + \frac{t_{\text{ox1}}}{\epsilon_{\text{ox}} \epsilon_0} + \frac{t_{\text{ox2}}}{\epsilon_{\text{ox}} \epsilon_0} + \frac{\frac{D_{\text{dot}}}{2}}{\epsilon_{\text{Si}} \epsilon_0 R_{\text{eff}}}} \quad (\text{IV.9})$$

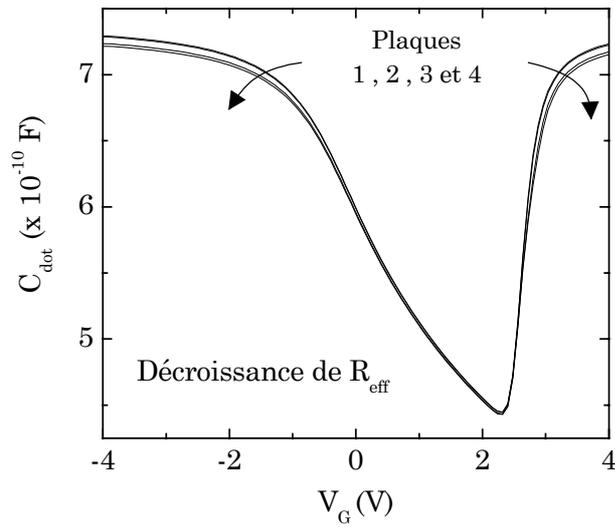


Figure IV.12. Simulations des courbes C-V pour les plaques 1, 2 3 et 4 avec les coefficients R_{eff} donnés dans le tableau (IV.1). Les paramètres utilisés pour les simulations sont : $N_A = 1.3 \cdot 10^{24} \text{ m}^{-3}$, $t_{\text{ox1}} = 5.5 \text{ nm}$, $t_{\text{ox2}} = 8 \text{ nm}$, $A_{\text{eff}} = 3 \cdot 10^{-3} \text{ cm}^2$ et $V_{\text{FB}} = -1 \text{ V}$.

Contrairement aux courbes C-V mesurées (cf. Fig. (IV.11)), les simulations réalisées pour les valeurs de R_{eff} reportées dans le tableau (IV.1), montrent une décroissance continue des valeurs maximales des capacités (voir Fig. (IV.12)). Par conséquent, les coefficients R_{eff} calculés à partir du diamètre et de la densité des nodules ne permettent pas d'expliquer l'ordre des courbes de la figure (IV.11). Nous nous sommes donc intéressés plus particulièrement à l'impact du diamètre des nodules.

IV.5.1.3. Impact de la densité, N_{dot} , et du diamètre, D_{dot} , des nodules

Dans un premier temps, nous avons simulé les courbes C-V des capacités avec et sans couche de semiconducteur, à l'intérieur de la couche d'oxyde (c.a.d. deux ou trois capacités en série). La figure (V.13) met en évidence la diminution de la capacité lorsque l'épaisseur de la couche de silicium augmente à l'intérieur de l'oxyde.

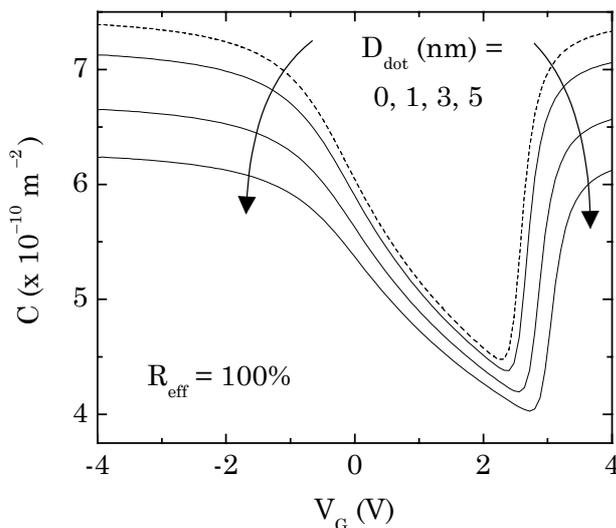


Figure IV.13. Simulation des courbes C-V sans ou avec une couche de silicium à l'intérieur de la couche d'oxyde pour $R_{\text{eff}} = 100\%$. Les paramètres de la simulation sont : $A_{\text{eff}} = 3 \cdot 10^{-7} \text{ m}^2$, $t_{\text{ox1}} = 5.5 \text{ nm}$, $t_{\text{ox2}} = 8 \text{ nm}$, $N_A = 1.3 \cdot 10^{24} \text{ m}^{-3}$, $N_{\text{dot}} = 10^{11} \text{ cm}^{-2}$ et $V_{\text{FB}} = -1 \text{ V}$.

Cependant, les nodules ne recouvrent qu'une partie R_{eff} de la surface de l'oxyde. Par conséquent, les courbes C-V correspondant aux capacités avec nodules (C_{dot}), se situent entre les courbes C-V simulées pour $D_{\text{dot}} = 0$ nm et $D_{\text{dot}} \neq 0$ nm. Néanmoins, à cause des différentes valeurs de R_{eff} (dus à N_{dot} et D_{dot}), les courbes C-V de capacités ayant des nodules de diamètres différents pourront se superposer ou ne pas suivre l'ordre croissant de la taille des nodules. En d'autres termes, les valeurs de N_{dot} et D_{dot} données dans le tableau (IV.1) ne correspondent pas de façon assez précise aux dispositifs mesurés.

De plus, on constate que les valeurs des capacités mesurées (cf. Fig. (IV.11)) sont plus faibles que celles obtenues par simulations (cf. Fig. (IV.13)). Par conséquent, l'épaisseur d'oxyde utilisée pour les simulations ne correspond pas à celle des structures mesurées.

La figure (IV.14) met en évidence le décalage des courbes C-V des capacités à nodules pour différentes valeurs d'épaisseur d'oxyde (t_{ox1} et t_{ox2}). La comparaison entre les mesures, figure (IV.11), et les simulations, figure (IV.14), montrent que lors de nos premières simulations, nous avons surestimé les valeurs des épaisseurs d'oxyde t_{ox1} et/ou t_{ox2} . Cette imprécision sur les épaisseurs peut également expliquer l'ordre des courbes mesurées en considérant que les quatre demi-plaques n'ont pas exactement la même épaisseur d'oxyde (tout en supposant qu'elles aient le même dopage de substrat).

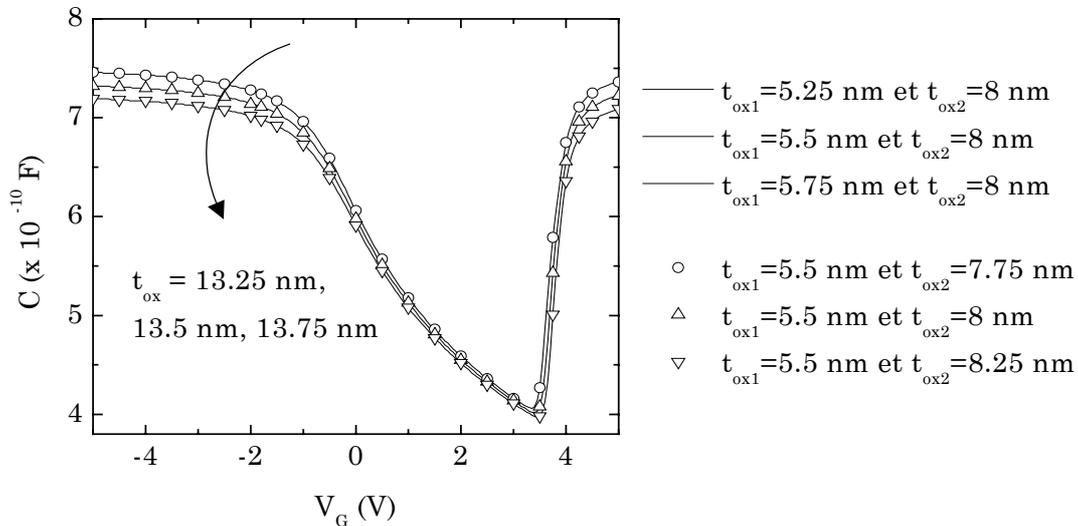


Figure IV.14. Simulation des courbes C-V de capacités à nodules en fonction de la variation des épaisseurs d'oxydes t_{ox1} et t_{ox2} . Les paramètres de simulations sont : $N_A = 1.3 \cdot 10^{24} \text{ m}^{-3}$, $A_{\text{eff}} = 3 \cdot 10^{-7} \text{ m}^2$, $V_{\text{FB}} = -1 \text{ V}$, $N_{\text{dot}} = 16 \cdot 10^{15} \text{ m}^{-2}$, $D_{\text{dot}} = 4.5 \text{ nm}$.

IV.5.1.4. Extraction des paramètres des capacités ring

Les programmes d'extraction de paramètres (dopage de substrat, épaisseur d'oxyde, tension de bandes plates) développés pour nos travaux de recherche sur la capacité MOS

(cf. chapitre II) ont été adaptés au cas des capacités à nodules en tenant compte de l'équation (IV.8) pour simuler la capacité avec dots. La figure (IV.15) donne un exemple de l'extraction des paramètres d'une capacité ring, de surface $A_{\text{eff}} = 5 \cdot 10^{-3} \text{ cm}^2$, située sur la demi-plaque 1 décrite dans le tableau (IV.1). Pour des tensions comprises entre la tension de bandes plates et la tension de seuil, on observe sur la figure (IV.15.a) un croisement de la courbe mesurée et de celle simulée, qui met en évidence la non uniformité du dopage de substrat.

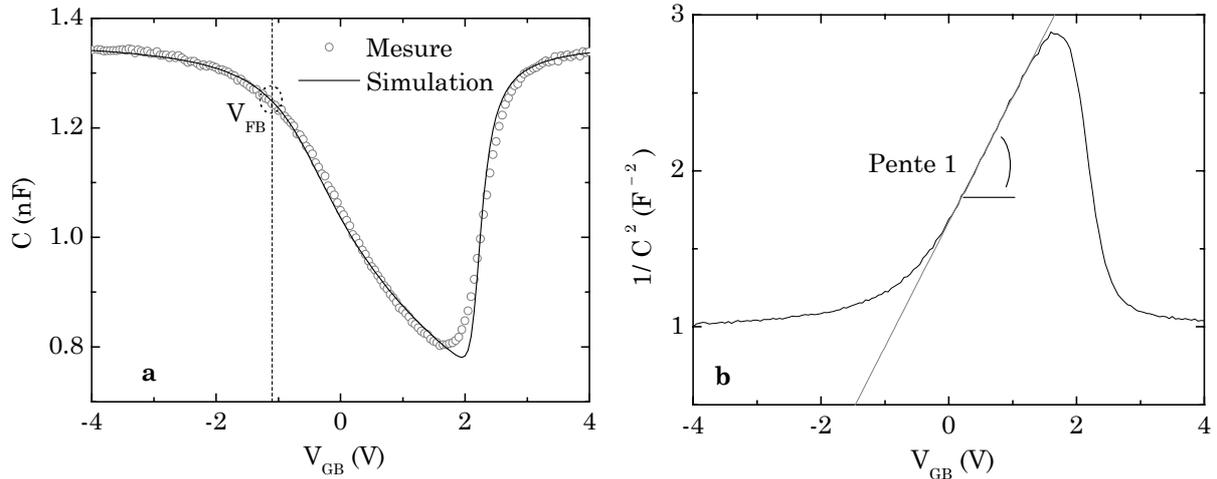


Figure IV.15. Comparaison entre la courbe C-V mesurée et simulée pour une capacité à nodules de la demi-plaque 1 ($A_{\text{eff}} = 5 \cdot 10^{-3} \text{ cm}^2$) (a) et extraction de ses paramètres par exemple Na (b). Les différents paramètres extraits sont : $N_A = 1.29 \cdot 10^{24} \text{ m}^{-3}$, $t_{\text{ox}} = 12.4 \text{ nm}$, $V_{\text{FB}} = -1.1 \text{ V}$.

Le tableau (IV.2) résume les valeurs des paramètres extraits à partir des mesures C-V des capacités ring, de surface $A_{\text{eff}} = 5 \cdot 10^{-3} \text{ cm}^2$, situées sur les quatre demi-plaques étudiées.

Demi-plaques	$N_A (10^{24} \text{ m}^{-3})$	$t_{\text{ox}} (\text{nm})$	$V_{\text{FB}} (\text{V})$
1	1.29	12.4	- 1.1
2	1.32	13.76	
3	1.33	13.32	
4	1.27	11.3	

Tableau IV. 2. Récapitulatif des paramètres extraits à partir des courbes C-V des capacités ring de surface $A_{\text{eff}} = 5 \cdot 10^{-3} \text{ cm}^2$, pour les quatre demi-plaques étudiées.

Ces résultats mettent en évidence la variation des paramètres clefs des capacités entre les quatre demi-plaques.

IV.5.2. Etude des transistors avec nodules

IV.5.2.1. Etude des temps d'écriture

L'écriture par porteurs chauds permet la diminution du temps d'écriture des mémoires. De plus, les simulations, présentées sur la figure (IV.8), ont mis en évidence la saturation de la charge injectée au bout de quelques dizaines de micro-secondes suivant le diamètre des nodules. A partir de ces observations, nous avons procédé à l'opération d'écriture d'un même dispositif en polarisant la grille à 8 V et le drain à 3.5 V avec un ou plusieurs pulses de 500 μ s, 50 μ s et 5 μ s. Avant chaque écriture, nous avons pris soin d'effacer la cellule par injection FN en appliquant une tension nulle sur le drain, la source et le bulk et en appliquant une tension de -12 V sur la grille pendant 100 ms.

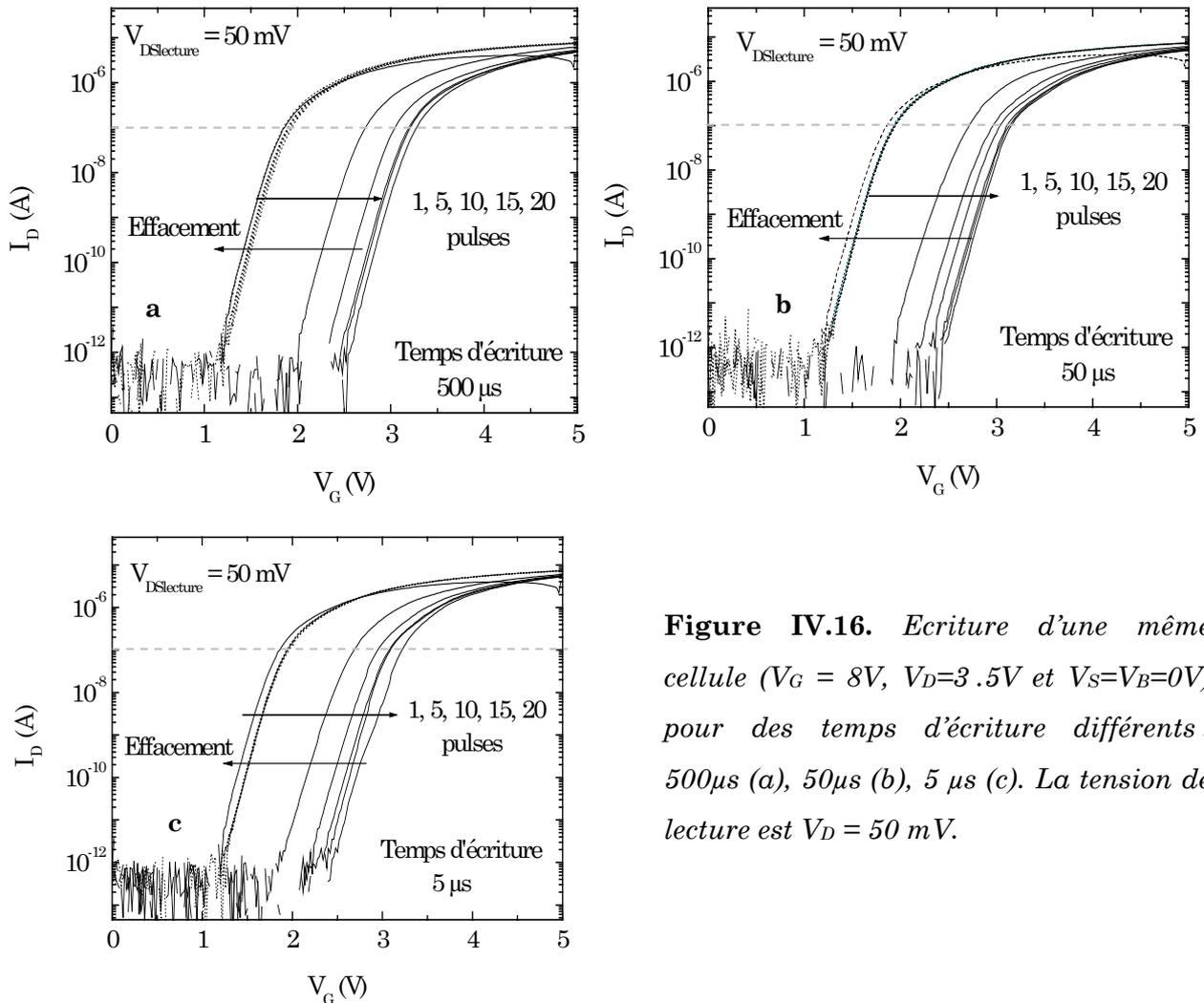


Figure IV.16. *Ecriture d'une même cellule ($V_G = 8V$, $V_D = 3.5V$ et $V_S = V_B = 0V$) pour des temps d'écriture différents : 500 μ s (a), 50 μ s (b), 5 μ s (c). La tension de lecture est $V_D = 50$ mV.*

La figure (IV.16) montre que les courbes correspondant à l'état effacé sont quasiment confondues (un léger décalage peut apparaître suite à la génération d'états d'interface lors de la première écriture de la mémoire) alors qu'après les opérations d'écriture les courbes se décalent de façon similaire. L'extraction des tensions de seuil à courant fixé ($I_{\text{test}} = 10^{-7}\text{A}$) présentée à la figure (IV.17), met en évidence le faible décalage entre les tensions de seuil obtenues pour des temps d'écriture de 500 μs , 50 μs et 5 μs , ce qui témoigne du temps très court nécessaire à l'injection d'électrons dans les nodules.

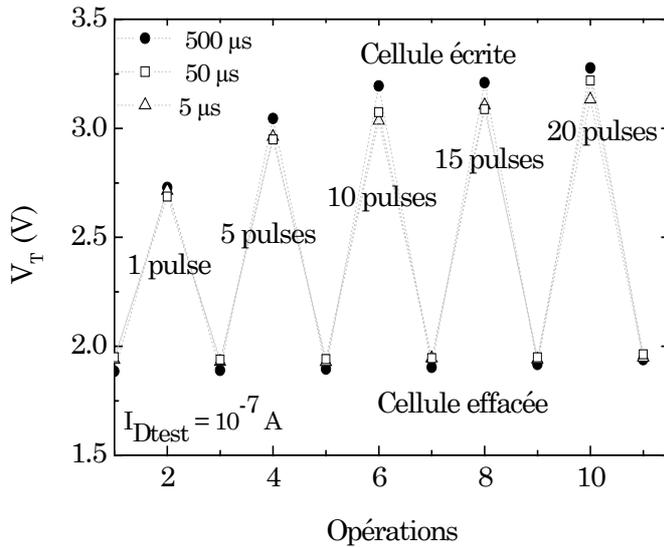


Figure IV.17. Evolution de la tension de seuil des courbes tracées sur la figure (IV.15).

IV.5.2.2. Etude des tensions d'écriture

Pour une tension de grille donnée, l'injection par porteurs chauds se produit pour une tension de drain environ égale à la moitié de cette tension de grille. Cependant dans le but de trouver les conditions d'écriture optimales, à savoir un compromis entre la génération de paires électron-trous et une faible dégradation de l'oxyde, différents couples (V_{GS} , V_{DS}) ont été étudiés. Par exemple, la figure (IV.18) montre que pour une tension de grille d'écriture égale à 8 V, la tension de seuil commence à se décaler à partir d'une tension de drain de lecture de 2 V, puis le décalage est moins prononcé à partir de 5 V avant de revenir en arrière pour des tensions supérieures à 6 V. Ce phénomène peut être dû à la dégradation de l'oxyde ; en effet, lors de ces expériences, nous avons observé le claquage des échantillons testés dès que la tension de seuil extraite commençait à décroître (entre 6 V et 7 V). Afin de vérifier et valider le comportement de notre simulateur lors de l'opération d'écriture, nous avons mesuré les caractéristiques courant-tension des mémoires à nodules après différentes programmations.

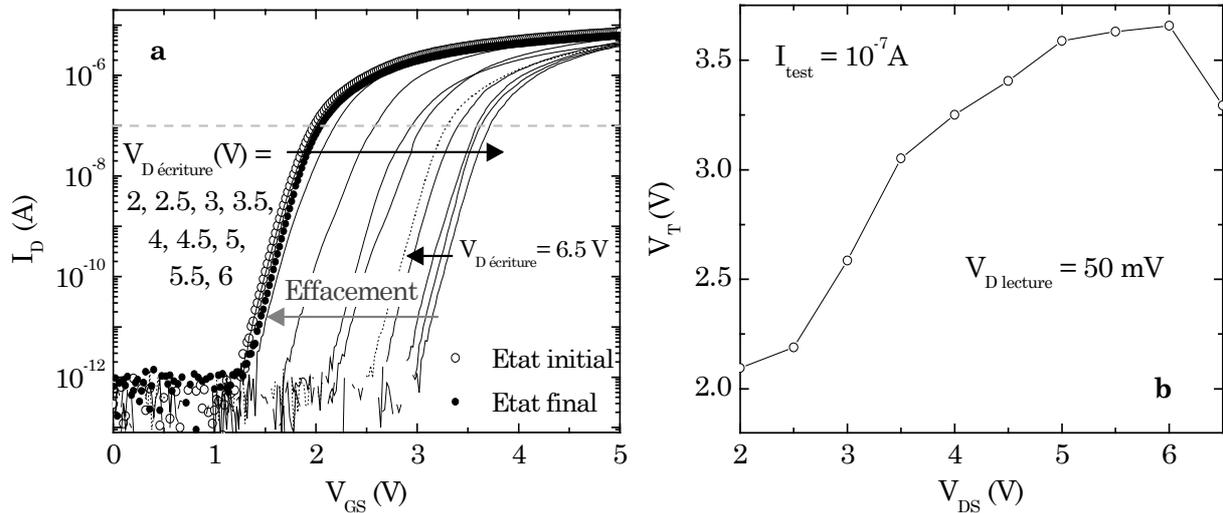


Figure IV.18. Lecture d'un transistor à nodules pour une tension $V_D = 50$ mV après différentes écritures cumulatives avec une tension de grille fixée à 8 V et une tension de drain croissant de 2 V à 6.5 V durant 50 μ s (a) et extraction de la tension de seuil correspondante, à courant fixé ($I_{\text{test}} = 10^{-7}$ A) (b).

La figure (IV.19) présente la variation de la tension de seuil extraite, à courant fixé ($I_{\text{test}} = 10^{-7}$ A), à partir des mesures $I_D(V_{GS})$ effectuées après l'écriture de la mémoire à différentes polarisations de grille et de drain. On observe le même comportement électrique que celui obtenu avec notre simulateur à savoir un grand décalage de la tension de seuil induit par l'augmentation des tensions de grille et de drain (voir Fig. (IV.10)).

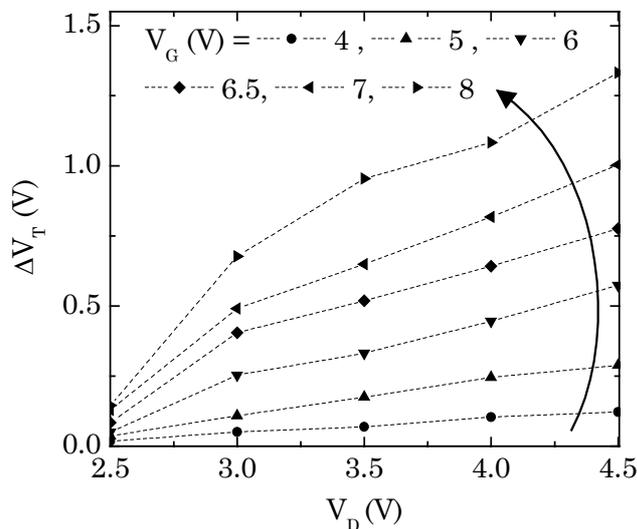


Figure IV.19. Variation de la tension de seuil d'un transistor ($W = 0.16 \mu\text{m}$, $L = 0.28 \mu\text{m}$) situé sur la demi-plaque 2, en fonction du potentiel de drain pour différentes tensions de grille appliquées durant un temps d'écriture égal à 10 μ s (la lecture se fait à $V_D = 50$ mV).

Pour cette gamme de tension, la mémoire reste en régime de saturation et l'injection du courant augmente avec la tension appliquée sur la grille. Pour des temps d'écriture plus longs, la courbe $\Delta V_T(V_{DS})$ aurait tendance à atteindre une valeur constante, c.a.d. une valeur de saturation. Bien qu'il ne nous ait pas été possible de calibrer notre

simulateur sur les dispositifs (en particulier pour les paramètres CHE), par manque de transistors de test (sans nodule), on peut noter, au premier ordre, une bonne concordance entre simulations et mesures ce qui tend à valider notre approche.

IV.5.2.2. Etude de la tension de lecture

Afin de ne pas écrire pendant l'opération de lecture, il est nécessaire de connaître la tension de drain à partir de laquelle l'injection CHE se produit. Différentes séries de mesure $I_D(V_{GS})$ ont donc été effectuées en augmentant la tension de drain de 50 mV à 4 V, tout en gardant les tensions de substrat et de bulk nulles. La figure (IV.20) présente les différentes variations des courants en fonction de la polarisation de grille lorsque la tension de drain V_D augmente.

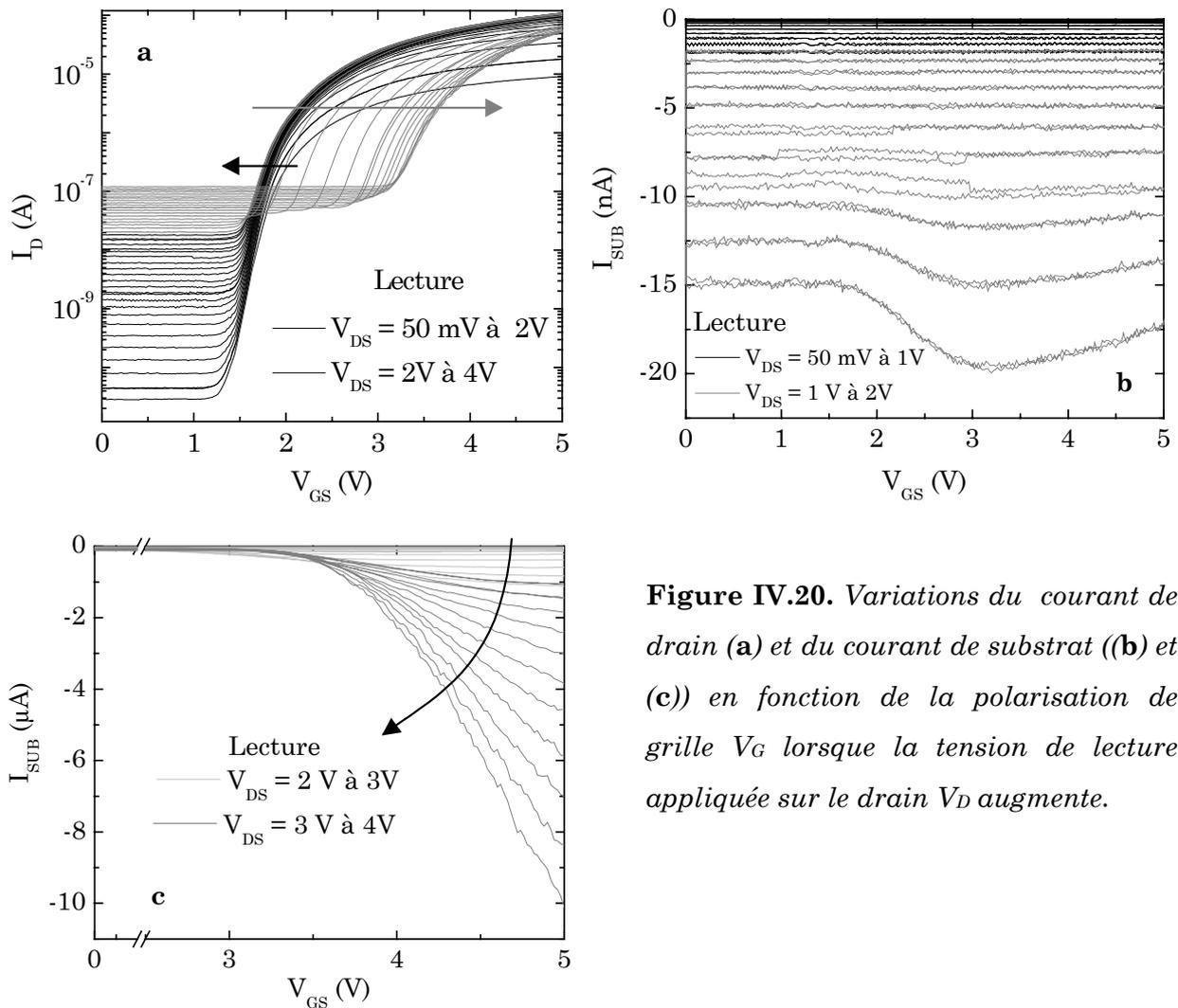


Figure IV.20. Variations du courant de drain (a) et du courant de substrat ((b) et (c)) en fonction de la polarisation de grille V_G lorsque la tension de lecture appliquée sur le drain V_D augmente.

On observe un décalage des courbes $I_D(V_{GS})$ du côté des tensions de grille positives dès que V_D devient supérieur à 2 V, ce qui correspond sur la figure (IV.20.b) au début de l'augmentation du courant de substrat, c'est à dire au début de l'injection par porteurs

chauds. De surcroît, la figure (IV.20.c) met en évidence l'augmentation continue du courant de substrat pour les fortes tensions de drain. Jusqu'à présent, seule la tension de lecture appliquée sur le drain a été présentée. Toutefois, les dispositifs dont nous disposons, présentent un coefficient R_{eff} faible de l'ordre de 25%, ce qui leur confère la possibilité d'un fonctionnement 2 bits. En effet, Mulidhar *et al.* ont montré que tant que le coefficient R_{eff} restait inférieur à la valeur critique du seuil de percolation, les îlots de silicium étaient suffisamment isolés les uns des autres pour éviter le transport latéral des charges entre nodules [Muralidhar'03]. Pour observer ce fonctionnement 2 bits, et vérifier que l'injection est bien localisée d'un seul côté, il est nécessaire de lire la mémoire à la fois en mode direct ($V_D > 0$ et $V_S = 0$) pour déterminer la tension de seuil côté source (V_{thf}) et en mode inverse (c.a.d. en inversant les polarisations source et drain ($V_D=0$ et $V_S>0$)), pour déterminer la tension de seuil du côté drain (V_{thr}). Cependant, les lectures effectuées pour des tensions égales à 50 mV ne permettent pas de faire la distinction entre les deux tensions de seuil. Ce phénomène, observé par Bloom *et al.* [Bloom'02] pour les mémoires à nitrure, est dû à la valeur de la tension de lecture. En effet, pour observer le fonctionnement deux bits des mémoires, il est nécessaire de placer le dispositif en régime de saturation.

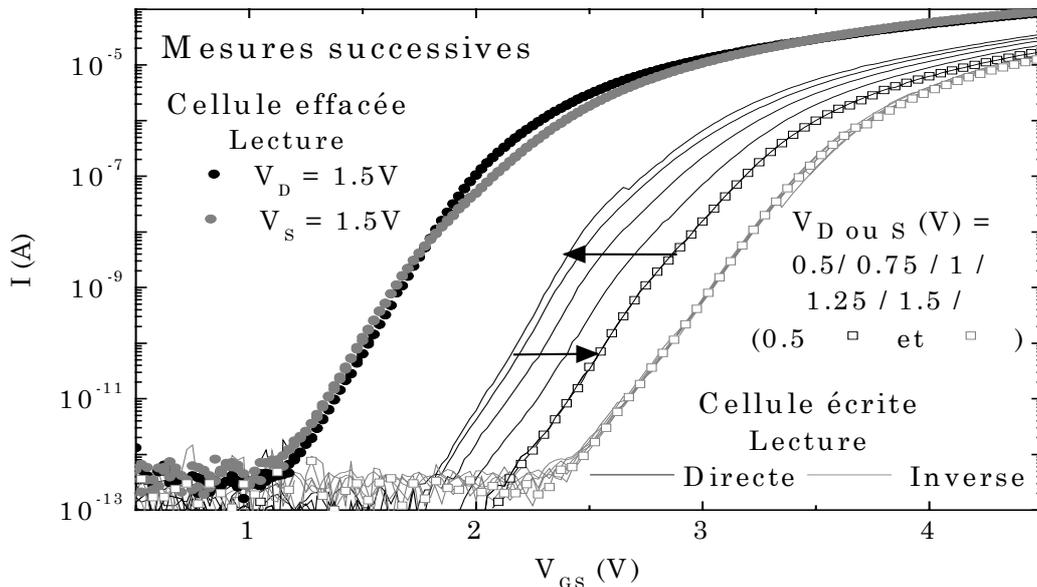


Figure IV.21. Mesures successives des caractéristiques courant-tension d'une cellule (de la demi-plaque 2). Dans un premier temps la cellule a été effacée puis lue en mode direct (points noirs) et en mode inverse (points gris) pour des tensions de lecture de 1.5V; puis dans un deuxième temps la cellule a été programmée ($V_G = 8V$ et $V_D = 3.5V$ durant $500\mu s$) et lue en mode direct (traits noirs) et en mode inverse (trait gris) pour des tensions de lectures variant de 0.5V à 1.5V et à nouveau 0.5V.

Comme le montre la figure (IV.21), à la suite de lectures successives, une fois la cellule écrite (écriture $V_G = 8V$, $V_{DS} = 3.5V$ pendant $500\mu s$), plus la tension de drain augmente, plus les caractéristiques $I_D(V_{GS})$ se décalent vers la caractéristique $I_D(V_{GS})$ de la cellule effacée. Notons, que la cellule mesurée n'a pas été dégradée par la série de lectures puisque les deux caractéristiques $I_D(V_{GS})$ lues en mode direct et inverse pour $V_{lecture} = 0.5V$ avant et après cette étude sont superposables. Le décalage des courbes $I_D(V_{GS})$ dû à la tension de lecture V_D , résulte de la présence des effets canaux courts, et plus particulièrement de l'effet DIBL (Drain Induced Barrier Lowering). En effet, la longueur de la zone d'injection étant très courte [Shappir'03], lorsque la tension appliquée sur le drain augmente, la couche de déplétion s'étend de plus en plus dans le canal vers la source et il se produit un abaissement de la barrière source/canal. L'abaissement de la barrière à la source permet l'injection d'électrons au travers du canal (en surface) et ceci indépendamment de la tension de grille. Par suite, en régime sous seuil, la grille perd le contrôle du courant de drain.

De surcroît, la tension de seuil obtenue par une lecture en mode direct est plus faible que la tension de seuil obtenue par une lecture en mode inverse [Eitan'00], [Bloom'01] et [Larcher'02]. Si la charge est injectée côté drain, elle sera complètement écrantée par la forte valeur appliquée en mode de lecture direct. Par suite, la forte polarisation V_D induit une région de pincement au dessous de la zone chargée (zone où il n'y a pas de couche d'inversion) et la caractéristique $I_D(V_{GS})$ de la cellule écrite reste proche de la caractéristique de la cellule vierge. Par conséquent la tension de seuil reste faible. En revanche, durant le mode de lecture inverse, la forte tension appliquée sur la source n'est pas capable d'écranter l'effet des électrons. Ainsi, pour étudier le fonctionnement 2 bits des cellules mémoires à nodules, des tensions de lecture supérieures à $1V$ doivent être appliquées tout en restant au dessous de la tension où débute l'injection par porteurs chauds (environs $2V$).

Enfin, la dégradation de la pente sous le seuil, surtout visible en mode de lecture inverse, prouve également la présence des effets 2D qui apparaissent à côté de la jonction où s'effectue l'injection. Cette dégradation est due à la courte longueur de la zone chargée par injection CHE plutôt qu'à la génération d'états d'interfaces durant l'injection [Shappir'03].

IV.6. Conclusion

Le travail présenté dans ce chapitre, repose sur les modèles préalablement développés lors de nos travaux de recherches sur les structures MOS. En modifiant ces modèles pour les adapter au cas particulier des mémoires à nodules, nous avons mis en évidence l'impact d'une charge piégée non uniformément dans tous les régimes (faible à forte inversion) de fonctionnement. Puis nous avons développé un algorithme de calcul permettant la modélisation électrique statique et dynamique de ces mémoires. Ce modèle peut également être une aide au design ou pour l'optimisation des signaux d'écriture et d'effacement des mémoires.

Les différentes comparaisons entre les mesures et les simulations des capacités à nodules témoignent de la complexité de la modélisation de ces structures due aux petites variations des valeurs des paramètres clefs (épaisseur d'oxyde, dopage, diamètres des nodules, densité des nodules).

Enfin, bien qu'il ne nous ait pas été possible de calibrer notre simulateur sur les dispositifs par manque de transistors de test (sans nodule), les caractérisations des transistors à nodules ont permis de montrer un bon comportement de notre modèle pour une étude statique 1bit des mémoires à nodules. Cependant, les caractérisations électriques des mémoires à nodules lors d'une étude 2 bits montrent une dégradation de la pente sous le seuil qui selon Lusky *et al.* [Lusky'01] - [Lusky'04] est due aux charges piégées au dessus du drain et du canal proche du drain et aux variations des effets 2D du champ électrique dans cette zone. Ainsi notre modèle, qui ne prend pas en compte les phénomènes 2D tels que les effets canaux courts, n'est plus valide pour ce mode de fonctionnement.

Conclusion générale et perspectives

L'évolution des dispositifs mémoires non volatiles vers le domaine nanométrique nécessite une compréhension très poussée de leurs mécanismes de fonctionnement. Ceci implique la connaissance du fonctionnement des structures Métal-Oxyde-Semiconducteur (MOS) qui composent ces mémoires.

Les travaux reportés dans ce manuscrit ont eu pour objectif la compréhension et la modélisation de nombreux phénomènes physiques liés à la réduction des dimensions de la capacité MOS et du transistor MOS afin de créer des briques élémentaires de la modélisation des mémoires et plus particulièrement des mémoires à nodules de silicium.

Nous avons commencé par réaliser une synthèse des principaux modes de fonctionnement de la capacité MIS, des transistors MOS et des mémoires Flash à nanocristaux de silicium. Un accent particulier a été mis sur le rappel de la modélisation du transistor MOS basée sur le calcul du potentiel de surface, point de départ de toutes nos modélisations. Nous avons également résumé les différents modes d'injections utilisés pour le fonctionnement des mémoires, à savoir l'injection tunnel (tunnel direct et Fowler-Nordheim), ainsi que l'injection par porteurs chauds.

Diverses méthodes de modélisation développées sous environnement Mathcad ont ensuite été présentées pour étudier le rôle des non uniformités de dopages, de charges et d'épaisseur d'oxyde sur les caractéristiques C-V et I-V des capacités MOS. Ainsi, la forte dégradation des courbes C-V générée par la poly-désertion de la grille a été exposée en particulier lorsque le rapport entre le dopage de grille et celui du substrat est inférieur à 100. Nous avons également présenté la déformation de la courbe C-V pour un dopage de substrat non constant entre l'interface et le volume du substrat. La modélisation pseudo 2D d'une non-uniformité de l'épaisseur d'isolant d'une capacité a mis en évidence l'erreur commise sur les paramètres électriques extraits des courbes C-V et I-V. De plus, nous avons comparé nos travaux, portant sur la modélisation pseudo 2D des capacités MOS présentant une rugosité de surface, aux résultats obtenus par ailleurs au laboratoire à l'aide d'une modélisation 2D qui prend en compte l'effet de pointe. Cette comparaison a montré qu'en première approximation, notre approche est correcte et permet un gain en terme de temps de calcul. Nous nous sommes ensuite intéressés aux mécanismes de dégradations électriques des capacités. D'une part, nous avons simulé la chute du courant tunnel des capacités MOS en présence de défauts dans le diélectrique (SiO_2 ou

empilements $\text{SiO}_2/\text{HfO}_2$) grâce à la résolution 1D de l'équation de Poisson dans l'isolant. D'autre part, suite à la comparaison de mesures (effectuées après des stress électriques des capacité MOS) et de simulations, nous avons élaboré une méthode pour déterminer la répartition des charges fixes dans l'oxyde qui se révèle suivre une loi en exponentielle décroissante à partir de l'interface injectante. En extrapolant ce résultat au cas d'une cellule EEPROM, le cyclage de la cellule (écriture par stress positif et effacement par stress négatif) mènerait à deux distributions exponentielles décroissantes de la charge fixe (une côté substrat et l'autre côté poly-silicium), c'est à dire à une courbe en forme de U. Enfin, une explication de l'origine de ces charges, générées au cours de stress électriques, a été proposée avec la création de centres précurseurs Si-O-H.

Le travail sur la capacité, à une dimension, a été réalisé dans le but d'être un pré-développement de la modélisation du transistor MOS (considération pseudo-2D). Par conséquent, nous avons présenté l'impact d'une non-uniformité des paramètres du transistor MOS le long du canal (dopage, épaisseur d'isolant, charges fixes). Pour cela, nous avons exploité les possibilités du modèle segmenté du transistor en commençant par une modélisation de ses résistances séries. Celle-ci a montré d'une part l'augmentation de la réduction de la mobilité apparente du transistor avec la valeur des résistances séries de celui-ci et, d'autre part, la possibilité de déterminer la valeur globale des résistances séries du transistor à partir du tracé d'un seul réseau de courbes $I_{DS}(V_{GS}, V_{DS})$. Nous avons ensuite modélisé la poly-désertion de grille d'un transistor et constaté la faible dégradation des caractéristiques électriques des transistors par rapport à celle observée dans le cas des capacités MOS. Le modèle segmenté a également été adapté pour étudier le phénomène inverse de canaux courts (Reverse Short Channel Effect) induit par une non-uniformité du substrat des transistors. Notre travail a ensuite été étendu au cas des transistors MOS à isolant de grille ultra-mince (< 1.5 nm) présentant un courant de fuite très important. La modélisation d'une non-uniformité d'épaisseur d'isolant le long du canal a été réalisée. Pour compléter cette étude, nous avons aussi matricé le transistor MOS (segmentation suivant la longueur et la largeur du canal) afin d'étudier l'impact d'une non-uniformité locale surfacique du transistor sur ses caractéristiques électriques. Nous avons alors montré la dépendance entre la dégradation des caractéristiques électriques du transistor MOS avec à la fois la taille et la position de la réduction de l'épaisseur d'isolant.

L'approche segmentée offre également la possibilité de simuler le rôle d'une variation locale de la tension de bandes plates sur le fonctionnement du transistor. Ainsi, la non-uniformité de la charge stockée dans l'isolant d'un transistor MOS a été modélisée. Ce

travail a été étendu au cas des mémoires à nodules dont le fonctionnement est basé sur un stockage discret de la charge. Un algorithme de calcul permettant la modélisation électrique de l'opération d'écriture de ces mémoires a ensuite été développé. Afin de valider nos modèles, différentes caractérisations électriques ont été réalisées sur des capacités et des transistors à nodules de silicium. La comparaison entre nos mesures et nos simulations a révélé l'imprécision des paramètres clés des mémoires à nodules (épaisseur d'oxyde, dopage, diamètres des nodules, densité des nodules) qui est source d'erreurs lors des simulations. De surcroît, bien qu'il ne nous ait pas été possible de calibrer notre simulateur par manque de transistors de test (sans nodule), les caractérisations des transistors à nodules ont permis de montrer un bon comportement de notre modèle pour une étude statique 1bit des mémoires à nodules. Cependant, notre modèle, basé actuellement sur une approche pseudo 2D, ne permet pas de simuler le comportement 2 bits de ces cellules mémoires reposant principalement sur les phénomènes 2D qui apparaissent autour des régions de chargement [Shappir'04].

Les différents résultats obtenus durant cette thèse suggèrent quelques perspectives de travail intéressantes :

- L'étude du chargement de l'oxyde de la capacité MOS pourrait être poursuivie notamment pour déterminer la loi de variation de Q_{ox} en fonction du temps de stress et de la polarisation. Cette loi permettrait de déterminer la fermeture de la fenêtre de programmation en fonction des signaux appliqués à la mémoire EEPROM.
- Une étude pourrait également être menée pour simuler la dégradation des propriétés électriques de transistors MOS, plus particulièrement la variation de la tension de seuil et du courant de drain observée au cours de contraintes électriques. Ce modèle pourrait être utilisé pour prédire la durée de vie des composants dans les circuits.
- La comparaison entre le modèle pseudo 2D et 2D réalisée lors de l'étude de la rugosité de surface des capacités MOS pourrait être étendue au cas du transistor MOS. En effet, il nous semble intéressant de vérifier jusqu'à quel point le modèle segmenté est utilisable pour ce type d'application. Comme nous l'avons annoncé, la réalisation d'un simulateur Poisson 2D du transistor MOS

couplé à l'équation du courant (mais sans considérer les diodes d'accès) est actuellement en cours dans un autre travail de thèse au laboratoire.

- Dans la continuité de nos travaux sur les mémoires à nodules, de nouvelles mesures sur des structures test (avec et sans nodules) ne faisant varier qu'un seul paramètre à la fois (épaisseur d'oxyde, dopage, diamètres des nodules, densité des nodules) seraient nécessaires pour effectuer le calibrage des modèles et leur amélioration. Il nous paraît également inévitable d'introduire les effets de canaux courts dans le modèle segmenté.
- Une autre perspective intéressante de travail serait l'étude du comportement des propriétés électriques des mémoires à nodules en fonction de la température du dispositif.

Références bibliographiques

- [Arora'95] N.D. Arora, E. Rios and C.-L. Huang, "Modeling the polysilicon depletion effect and its impact on submicrometer CMOS circuit performance" *IEEE Transactions on Electron Devices*, Vol. **42**, No. 5, p. 935 - 943, 1995.
- [Bell'01] L.D Bell, E Boer, M. Ostraat, and ML Bron- gersma, "A radiation-tolerant, low-power non-volatile memory based on silicon nanocrystal quantum dots", Forum on Innovative Approaches to Outer Planetary Exploration, Lunar Planetary Institute, (Houston), 2001. <www.lpi.usra.edu/meetings/outerplanets2001/pdf/4080.pdf>
- [Bernardini'04a] S. Bernardini, P. Masson and M. Houssa, "Effect of fixed dielectric charges on tunneling transparency in MIM and MIS structures", *Microelectronic Engineering*, Vol. **72**, Issues 1-4, p. 90 - 95, 2004.
- [Bernardini'04b] S. Bernardini, P. Masson, M. Houssa and F. Lalande, "Origin and repartition of the oxide fixed charges generated by electrical stress in memory tunnel oxide", *Applied Physic Letters*, Vol. **84**, Issues 21, p. 4251 - 4253, 2004.
- [Bernardini'04c] S. Bernardini, JM. Portal and P. Masson, "A Tunneling Model for Gate Oxide Failure in Deep Sub-Micron Technology", *Design, Automatisation and Test in Europe (DATE)*, ISBN: 0-7695-2085-5, Vol. **2**, p. 1404 - 1405, 2004.
- [Bernardini'04d] S. Bernardini, P. Masson, JM. Portal, J.M. Gallière and M. Renovell, "Impact of Gate Oxide Reduction Failure on Analog Application: Example of the Current Mirror" *5th Latin-American Test Workshop (LATW'04)*, ISBN: 958-33-5900-9, p. 12-17, 2004.
- [Bernardini'03a] S. Bernardini, P. Masson and M. Houssa, "Effect of fixed dielectric charges on tunneling transparency in MIM and MIS structures", *Insulating Film on Silicon (INFOS), 13th Bi-annual conference, Barcelone*, PS12, 2003.

- [Bernardini'03b] S. Bernardini, P. Masson, M. Houssa and F. Lalande, "Determination of oxide charge repartition in memory tunnel oxide under stress from Fowler-Nordheim current measurements", *33 rd European Solid-State Device Research Conference, (ESSDERC)*, ISBN 0-7803-7999-3, p. 589 - 592, 2003.
- [Bernardini'03c] S. Bernardini, R. Laffont, P. Masson, G. Ghibaudo, S. Lombardo, B. De Salvo and C. Gerardi, "A predictive nano-crystal Flash memory simulator", *4rd European Workshop on Ultimate Integration of Silicon (ULIS), Udine*, p. 143 - 146, 2003.
- [Bez'90] R. Bez, E. Camerlenghi, D. Cantarelli, L. Ravazzi, and G. Crisenza, "A novel method for the experimental determination of the coupling ratios in submicron EPROM and flash EEPROM cells", *Electron Devices Meeting IEDM, Technical Digest. International* , 9-12, p. 99 - 102, 1990.
- [Bloom'02] I. Bloom, P. Pavan and B. Eitan, "NROMTM—a new technology for non-volatile memory products", *Solid-State Electronics*, Vol. **46**, p. 1757 - 1763, 2002.
- [Bloom'01] I. Bloom, P. Pavan and B. Eitan, "NROME — a new non-volatile memory technology : from device to products", *Microelectronic Engineering*, Vol. **59**, p. 213 – 223, 2001.
- [Bouchakour'95] R. Bouchakour, L. Hardy, J.F. Naviner, I. Limbourg, M. Jourdain and M. Jelloul, "Modeling and characterization of the nMOS transistor stressed by hot-carrier injection", *Circuits and Systems, Proceedings of the 38th Midwest Symposium on* , Vol. **1** , p. 61 – 64, 1995.
- [Brews'78] J.R. Brews, "A charge-sheet model of the MOSFET", *Solid-State Electronics*, Vol. **21**, p. 345-355, 1978.
- [Brown'91] D.B. Brown and N.S. Saks, "Time dependence of radiation-induced interface trap formation in metal-oxide-semiconductor devices as a function of oxide thickness and applied field", *J. Appl. Phys*, Vol. **70**, No.7, p. 3734-3747, 1991.

- [Cappelletti'99] P. Cappelletti, C. Golla, P. Olivo, E. Zanoni, "Flash Memories", *Kluwer Academic Publisher*, 1999.
- [Chae'99] D.H. Chae, T.S. Yoon, D.H. Kim, J.Y. Kwon, K.B.Kim, J.D. Lee and B.G. Park, "Programming dynamics of a single electron memory cell with a high density SiGe nanocrystal array at room temperature", *57th Annual Device Research Conf. Dig.*, p. 140-141, 1999.
- [Chang'03] Ko-Min Chang, "Nanocrystals Key to Motorola's Test Memory" *Nanoparticles news, A BCC, Inc. Publication*, Vol. 6, No. 3, 2003.
- [Charlot'86] J.J. Charlot, S. Toutain, "Distributed Charge (Sub)Micron Mos-Transistor Model", *IEE Proceedings-I Communications Speech and Vision* 133 (6), p. 207-213, 1986.
- [Choi'02] C-H.. Choi, P.R. Chidambaram, R. Khamankar, C.F. Machala, Y. Zhiping and R.W. Dutton, "Dopant Profile and Gate geometric effect on polysilicon gate depletion in scaled MOS", *IEEE transaction on Electron Device* , Vol. 49, No. 7, p. 1227 -1231, 2002.
- [Corso'03] D. Corso, I. Crupi, V. Ancarani, G. Ammendola, G. Molas, L. Perniola, S.Lombardo, C. Gerardi, B. De Salvo, "Localized charge storage in nanocrystal memories: feasibility of a multi-bit cell", *33rd Conference on European Solid-State Device Research, ESSDERC*, 16-18 , p. 91 - 94, 2003.
- [Cuinet'04] X. Cuinet, S. Bernardini, P. Masson and L. Raymond, "Simulation of Nanometric Roughness Impact on a MOS Capacitance", *5th symposium SiO₂, Advanced Dielectric & Related Devices*, 2004.
- [De Blauwe'00] J. De Blauwe, M. L. Green, G. Weber, T. Sorch, A. Keber, F. Klemens, R. Cireli, E. Ferry, J.L. Grazul, F. Braumann, Y. Kim, W. Mansfield, J. Bude, J.T.C. Lee, S.J. Hillenius, R.C. Flagan, and H.A. Atwater, "A novel aerosol-nanocrystal floating gate device for non volatile memory applications", *Electron Devices Meeting, IEDM Technical Digest. International*, 10-13, p. 683 - 686, 2000.

- [Depas'95] M. Depas, B. Vermeire, P. W. Mertens, R. L. Van Meirhaeghe and M. M. Heyns, "Determination of tunnelling parameters in ultra-thin oxide layer poly-Si/SiO₂/Si structure", *Solid-State Electronics*, Vol. **38**, No. 8, p. 1465 - 1471, 1995.
- [De Salvo'03] B. De Salvo, C. Gerardi, S. Lombardo, T. Baron, L. Perniola, D. Mariolle, P. Mur, A. Toffoli, M. Gely, M.N. Semeria, S. Deleonibus, G. Ammendola, V. Ancarani, M. Melanotte, R. Bez, L. Baldi, D. Corso, I. Crupi, R.A. Puglisi, G. Nicotra, E. Rimini, F. Mazen, G. Ghibaud, G. Pananakakis, C.M. Compagnoni, D. Ielmini, A. Lacaita, A. Spinelli, Y.M. Wan, and K. V.D. Jeugd, "How far will silicon nanocrystals push the scaling limits of NVMs technologies?", *Electron Devices Meeting, IEDM Technical Digest. IEEE International*, 8-10, p. 26.1.1 - 26.1.4, 2003.
- [Eitan'00] B. Eitan, P.Pavan, I.Bloom, E.Aloni, A.Frommer, and D. Finzi, "NROM: a localized trapping , 2-Bit non volatile Memory cell", *IEEE Elec. Dev. Lett*, Vol. **21**, No. 11, p.543 - 545, 2000.
- [Eitan'99] B. Eitan, P.Pavan, I.Bloom, E.Aloni, A.Frommer, and D. Finzi, "can NROM, a 2 bit, trapping storage NVM cell, give a real challenge to floating gate cells? ", *ext. abs., ICSSDM*, p. 522 - 523, 1999.
- [Fowler-Nordheim'28] R.H Fowler and L. Nordheim, "Electron in Intense Electric Field", *Proc. Soc. London Ser. A*, Vol. **119**, p.173 - 181, 1928.
- [Fromhold'81] A. T. Fromhold, *Quantum mechanical for applied physics and engineering*, Academic Press, New York, 1981.
- [Garros'03] X. Garros¹, C. Leroux, G. Reibold, B. Guillaumot, F.Martin and J.L. Autran, "Investigation of trapping and breakdown in ALD HfO₂ films", *Insulating Film on Silicon (INFOS), 13th Bi-annual conference, Barcelone*, W18, 2003.
- [Gilibert'04] F. Gilbert, D. Rideau, S. Bernardini, P. Scheer, M. Minondo, D. Roy, G. Gouget and A. Juge, "DC and AC MOS transistor modelling in presence of high gate leakage and experimental validation", *Solid-State Electronics*, Vol. **48**, No. 4, p. 597 - 608, 2004.

- [Gilibert'03] F. Gilibert, D. Rideau, S. Bernardini, P. Scheer, M. Minondo, D. Roy, G. Gouget and A. Juge, "Channel Debiasing and Gate Current Modelling in Advanced CMOS Devices", *4rd European Workshop on Ultimate Integration of Silicon (ULISS), Udine*, p. 61-64, 2003.
- [Hanafi'96] H.I.Hanafi, S. Tiwari and I.Khan, "Fast and long retention time nano-cristal memory", *IEEE Trans. Elec. Dev.*, Vol. **43**, p. 1553-1558, 1996.
- [Hoewoo'03] K. Hoewoo, L. Kieyoung, L. Kyungho, T A. Fjeldly and M. S. Shur "Analysis of the anomalous drain current characteristics of halo MOSFETs", *Solid-State Electronics*, Vol. **47**, Issue 1, p. 99-106, 2003.
- [Hofstein'63] S. R. Hofstein and F. P. Heiman, "The Silicon Insulated-Gate Field-Effect Transistor", *Proc. IRE*, Vol. **51**, No.9, p. 1190-1202, 1963
- [Houssa'01] M. Houssa, A. Stesmans, R.J. Carter, and M.M. Heyns, "Stress-induced leakage current in ultrathin SiO₂ layers and the hydrogen dispersive transport model", *Appl. Phys. Lett.*, Vol. **78**, No. 21, p. 3289 - 3291, 2001.
- [Houssa'00] M. Houssa, V.V. Afanas'ev, A. Stesmans and M.M. Heyns, "Variation in the fixed charge density of SiO_x/ZrO₂ gate dielectric stacks during postdeposition oxidation", *Appl. Phys. Lett.*, Vol. **77**, No. 12, p. 1885 - 1887, 2000.
- [Hradsky'03] R.Rao, R.F.Steimle, M.Sadd, B.Hradsky, S.Egley, C.Swift, E.Prinz, J.Yater, S.Bagchi, S.Straub, R.Muralidhar and B.White, "Silicon Nanocrystal Memories", Non volatile memories with discrete storage nodes, Estoril, Portugal, Workshop ESSDERC 2003.
- [Huang'93] C.-L. Huang, N.D. Arora, A.I. Nas and D.A. Bell, " Effect of polysilicon depletion on MOSFET I-V characteristics", *Electronics Letters* , Vol. **29**, No. 13 , p 1208 -1209, 1993.
- [ITRS'03] International Technology Roadmap for Semiconductors, edition 2003. <public.itrs.net/Files/2003ITRS/Home2003.htm>

- [Jacob'93] H. Jacobs, A. von Schwerin, D. Scharfetter, F. Lau, "MOSFET reverse short channel effect due to silicon interstitial capture in gate oxide", *Electron Devices Meeting, Tech. Dig, IEDM*, 5-8, p. 307 - 310, 1993.
- [Kahng'67] D. Kahng, S.M. Sze, "A floating gate and its application to memory devices" *Bell Syst. Tech. J.* 46, p.1288, 1967.
- [Kahng'60] D. Kahng and M. Atalla, "Silicon-Silicon Dioxide Field Induced Devices" *Device Research Conf.* 1960.
- [Kim'98] I. Kim, S. Han, H. Kim, J. Lee, B. Choi, S. Hwang, D. Ahn; H. Shin, "Room temperature single electron effects in Si quantum dot memory with oxide-nitride tunneling dielectrics" *Tech. Dig, IEDM*, 6-9, p. 111 - 114, 1998.
- [King'98] Y-C King, T-J King, C Hu. "MOS memory using germanium nanocrystals formed by thermal oxidation of $\text{Si}_{1-x}\text{Ge}_x$ ", *Tech. Dig (IEDM)*, p. 115 - 118, 1998.
- [Ku'94] P. S. Ku, and D. K. Schroder, "Charges trapped throughout the oxide and their impact on the Fowler-Nordheim current in MOS devices", *IEEE Trans. On Electron Devices*, Vol.41, No.9, p. 1669 - 1672, 1994.
- [Laffont'03a] R. Laffont, P. Masson, S. Bernardini, R. Bouchakour and J.M. Mirabel, "A new floating compact model applied to Flash memory cell", *J. of Non-Cryst. Solids*, Vol. 322, No.1-3, p. 250 - 255, 2003.
- [Laffont'03b] R. Laffont, "Modélisation et contribution à l'amélioration du fonctionnement des mémoires non volatiles", *Thèse de Doctorat, Université de Provence*, 2003.
- [Larcher'02] L. Larcher, G. Verzellesi, P. Pavan, E. Lusky, I. Bloom and B. Eitan, "Impact of programming charge distribution on threshold voltage and subthreshold slope of NROM memory cells", *Electron Devices, IEEE Transactions on*, Vol. 49, No. 11, p. 1939 - 1946, 2002.

- [Lenzlinger'69] M. Lenzlinger and E. H. Snow , “Fowler-Nordheim Tunneling into Thermally Grown SiO₂”, *J. Appl. Phys*, Vol. **40**, No.1, p. 278 - 283, 1969.
- [Limbourg'95] I. Limbourg, L. Hardy, M. Jourdain, R. Bouchakour and J. J. Charlot, “Modeling and simulation of homogeneous degradation for N-channel MOSFETS”, *Journal of Non-Crystalline Solids*, Vol. **187**, No.1, p. 160-164, 1995.
- [Liu'03] Z. Liu, C. Lee, V. Narayanan, G. Pei, and E. C. Kan, “A Novel Quad Source/Drain Metal Nanocrystal Memory Device for Multibit-Per-Cell Storage”, *IEEE Elec. Dev. Lett*, Vol. **24**, No.5, p. 345-347, 2003.
- [Liu'02] Z. Liu, C. Lee, V. Narayanan, G. Pei and E.C. Kan, “Metal nanocrystal memories. I. Device design and fabrication”, *Electron Devices, IEEE Trans. on* , Vol. **49** , No. 9 , p.1606 - 1613, 2002.
- [Lombardo'04] S. Lombardo, B. De Salvo, C. Gerardi and T. Baron, “ Silicon nanocrystal memories ”, *Microelectronic Engineering*, Vol. **72**, No. 1-4, p. 388 - 394 , 2004.
- [Lusky'04] E. Lusky, Y. Shacham-Diamand, G. Mitenberg, A. Shappir, I. Bloom, and B. Eitan, “Investigation of channel hot electron injection by localized charge-trapping nonvolatile memory devices” *Electron Devices, IEEE Transactions on* ,Vol.**51** , No.3, p. 444 - 451, 2004.
- [Lusky'01] E. Lusky, Y. Shacham-Diamand, I. Bloom, and B. Eitan, “Characterization of channel hot electron injection by the subthreshold slope of NROM™ device,” *IEEE Elec. Dev. Lett.*, Vol. **22**, No. 11, p. 556 - 558, 2001.
- [Maneglia'99] Y. Maneglia and D. Bauza, “Evolution of the Si-SiO₂ interface trap characteristics with Fowler-Nordheim“, *Microelectronic Test Structures, ICMTS'99, Göteborg* , p. 117 -120, 1999.

- [Maserjian'74] J. Maserjian, G. Petersson and C. Svensson, " Saturation capacitance of thin oxide MOS structures and the effective surface density of states silicon", *Solid-State Electronics*, Vol. **17**, p. 335 - 339, 1974.
- [Masson'02] P. Masson, J.L. Autran and D. Munteanu, "DYNAMOS : a numerical MOSFET model including quantum-mechanical and near-interface trap transient effects", *Solid-State Electronics*, Vol. **46**, p 1051 - 1059, 2002.
- [Masson'99] P. Masson, "Etude par pompage de charge et par mesures de bruit basse fréquence de transistors MOS à oxynitride de grille ultra-minces" , *Thèse de Doctorat, INSA Lyon*, Grenoble, 1999.
- [Masuoka'84] F. Masuoka, M. Asano, H. Iwahashi, T. Komuto and S. Tanaka, "A new Flash EEPROM cell using triple polysilicon technology", *IEEE, IEDM Tech. Dig*, p.464 - 467, 1984.
- [McLarty'95] P.K. McLarty, S. Cristoloveanu, O. Faynot, V. Mistry, J.R. Hauser and J.J. Wortman "A simple parameter extraction method for ultra-thin oxide MOSFETs", *Solid-State Electronics*, Vol. **38**, No. 6, p. 1175-1177, 1995.
- [McLean'77] F.B. McLean and G.A. Ausman, "Simple approximate solutions to continuous-time random-walk transport", *Jr. Phys. Rev. B*, Vol. **15**, p. 1052-1061, 1977.
- [Muralidhar'03] R. Muralidhar, R.F. Steimle, M. Sadd, R.Rao, C.T. Swift, E.J. Prinz, J. Yater, L. Grieve, K. Harber, B. Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S.G.H. Anderson, M. Rossow, T. Merchant, M. Paransky, T. Huynh, D. Hadad, B.E. Ko-Min Chang and Jr. White, "A 6V Embedded 90nm Silicon Nanocrystal Nonvolatile Memory", *Electron Devices Meeting. IEDM Technical Digest. IEEE International* , 8-10, p. 26.2.1 - 26.2.4, 2003.
- [Nicollian'82] E. H. Nicollian and R. R. Brews, *MOS Physics and Technology*, Wiley, New York, 1982.

- [O'Dwyer'73] J.J. O'Dwyer, *The Theory of Electrical Conduction and Breakdown in Solids Dielectrics*, Carendon, Oxford, 1973.
- [Ogura'80] S. Ogura, P. J. Tsang, W. W. Walker, D. L. Critchlow, and J. F. Shepard, "Design and characteristics of the lightly doped drain-source (LLD) insulated gate field-effect transistor", *IEEE Journal of Solid-State Circuits*, Vol. **SC-15**, No. 4, 1980.
- [Okumura'92] Y. kumur, M. irahata, A. Hachisuka, T. Okudaira, H. Arim, T. Matsukaw, "Source-to-drain nonuniformly doped channel (NUDC) MOSFET structures for high current drivability and threshold voltage controllability", *Electron Devices, IEEE Transactions on*, Vol. **39** , No. 11, p. 2541 - 2552, 1992.
- [Pananakakis'95] G. Pananakakis, G; Ghibaud, R. Kies, C. Papadas, "Temperature dependence of the Fowler–Nordheim current in metal-oxide-degenerate semiconductor structures", *J. Appl. Phys.*, Vol. **78**, No. 4, p. 2635-2641, 1995.
- [Pao'66] H.C Pao and C.T. Sah, "Effects of diffusion current on characteristics of metal oxide(insulator)-semiconductor transistors", *Solid-State Electron.*, Vol. **9**, p. 927, 1966.
- [Papadas'92] C. Papadas, G. Guibaud, G. Pananakakis, C.Riva and P.Ghezzi, "Model for programming window degradation in FLOTOX EEPROM cells", *IEEE Electron Device Letters*, Vol.**13**, No. 12, p. 89- 91, 1992.
- [Patch'00] K. Patch, "Silicon Nanocrystals Glow", TRN News, 2000.
<www.trnmag.com/Stories/121300/Silicon_nanocrystals_glow_121300.html>
- [Pavan'97] P. Pavan, R. Bez, P. Olivo, E. Zanoni, "Flash memory cells-an overview" *IEEE proceedings of the IEEE*, Vol. **85**, No. 8, p. 1248-1271, 1997.
- [Plossu'01] C. Plossu, S. Croci, N. Monti, R. Bouchakour, R. Laffont, Ph. Boivin and J. M. Mirabel, "Conduction properties of electrically erasable read only memory tunnel oxides under dynamic stress", *Journal of Non-Crystalline Solids*, 280, 103, 2001.

- [Rafferty'93] C.S. Rafferty, H.-H. Vuong, S.A. Eshraghi, M.D.Giles, M.R. Pinto, S.J. Hillenius, "Explanation of reverse short channel effect by defect gradients", *Electron Devices Meeting, IEDM, Technical Digest., International* , 5-8 , p. 311-314, 1993.
- [Rios'94] R .Rios, N.D Arora and H. Cheng-Liang, "An analytic polysilicon depletion effect model for MOSFETs", *IEEE Electron Device Letters* , Vol. **15**, No. 4 , p. 129 -131, 1994.
- [Sadovnikov'01] A. Sadovnikov, A. Kalnitsky, A. Bergemont, and P. Hopper, "The effect of polysilicon Doping on the Reverse Short Channel Effect in sub-Quarter Micron NMOS Transistors" *IEEE Trans. Electron Devices*, Vol.**48**, p. 393 - 395, 2001.
- [Schuegraf'93] K.F. Schuegraf, C.C. king and C. Hu, "Impact of polysilicon depletion in thin oxide MOS technology", *VLSI Technology, Systems, and Applications, Proceedings of Technical Papers. International Symposium on* , 12-14, p. 86 - 90, 1993.
- [Shappir'04] A. Shappir , D. Levy, Y. Shacham-Diamand, E. Lusky, I. Bloom, and B. Eitan, "Spatial characterization of localized charge trapping and charge redistribution in the NROM device", *Solid-State Electronics*, Vol. **48**, No. 9, p. 1489 - 1495, 2004.
- [Shappir'03] A. Shappir, Y. Shacham-Diamand, E. Lusky, I. Bloom and B. Eitan, "Subthreshold slope degradation model for localized-charge-trapping based non-volatile memory devices", *Solid-State Electronics*, Vol. **47**, p. 937 - 941, 2003.
- [Shi'98] Y. Shi, K. Saito, H. Ishikuro, T. Hiramoto, "Effets of traps on charge storage characteristics in metal-oxide-semiconductor mememory structures based on silicon nanocrystal", *J. Appl. Phy.*, Vol. **84**, No.4, p. 2358 - 2360, 1998
- [Syrzycki'89] M. Syrzycki, "Modeling of gate oxide shorts in MOS transistors", *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, Vol. **8**, No. 3, p. 193 - 202, 1989.

- [Sze'81] S. M. Sze, "Physics of Semiconductor Devices " Wiley & Sons, New York, 2nd edition, 1981.
- [Tam'84] S Tam, PK KO, and C. Hu, "Lucky electron model of channel hot electron injection in Mosfet's ", *IEEE trans. Electron devices*, Vol. **E-D 31**, p. 111 - 116, 1984.
- [Tiwari'95] S. Tiwari, F. Rana, K. Chan, D. Buchanan, "Volatile and non-volatile memories in silicon with nano-crystal storage" *Electron Devices Meeting, International IEDM*, 10-13, p. 521 - 524, 1995.
- [Tiwari'96] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E.F. Crabbe, and K. Chan, "A silicon nanocrystals based memory", *Appl. Phys. Lett.*, Vol. **68**, No. 10, p. 1377 - 1379, 1996.
- [Tsividis'99] Y. Tsividis, Mc Graw Hill international Edition, 1999.
- [Van Langevelde'02] R. Van Langevelde, AJ Scholten, DBM Klaassen. MOS, Model 11 level 1101 Nat. Lab. unclassi.ed report NL-UR, 802, 2002.
- [Victory'95] J. Victory, J. Sanchez, T. DeMassa and B. Welfert, "Application of the MOS charge-sheet model to nonuniform doping along the channel", *Solid-State Electronics*, Vol. **38**, No. 8, p. 1497-1503, 1995.
- [Wang'95] C-L Wang, "Computation of current and transconductance of a nonuniformly doped channel MOSFET with an arbitrary doping profile", *Solid-State Electronics*, Vol. **38**, No. 8, p. 1423 - 1429, 1995.
- [Wang'81] N.N. Wang, "A simple distributed circuit model for charge-coupled devices", *Circuits and Systems, IEEE Transactions on* , Vol. 28 , No. 1 , p. 2 – 11, 1981.
- [Xi'00] X. Xi, KM. Cao, H. Wan, M. Chan, C. Hu, BSIM4.2.1 MOSFET Model, users' manual, Department of Electrical Engineering and Computer Sciences University of California, Berkeley, 2000. <http://www.eigroup.org/cmc/cmos/b4manual.pdf>

- [Yang'98] H.Y. Yang, H. Niimi, and G. Lucovsky, "Tunneling currents through ultrathin oxide/nitride dual layer gate dielectrics for advanced microelectronic devices", *Journal of Appl. Phys.*, Vol. **83**, No. 4, p. 2327 - 2337, 1998.
- [Yaron'82] G. Yaron, S.J. Prasad, M.S. Ebel and B.M.K. Leong, "A 16K E/SUP 2/PROM employing new array architecture and designed-in reliability features", *IEEE Solid-State Circuits*, , Vol. **17** , No. 5, p. 833 - 840, 1982.

Annexe I

Les équations du miroir de courant

Le miroir de courant est l'élément de base utilisé pour la réplication, l'addition et la soustraction de courants. Il ne fonctionne qu'avec des courants unipolaires. Considérons le miroir de courant constitué de deux transistors MOS représenté sur la figure A1.

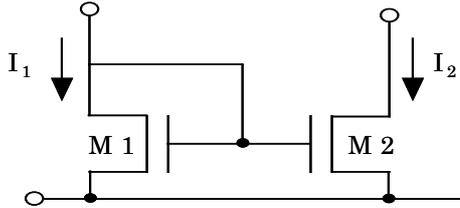


Figure A1: Schématisation d'un miroir de courant constitué de deux transistors MOS M1 et M2.

Le miroir de courant utilise le principe suivant : si les potentiels grille-source de deux transistors MOS, M1 et M2, sont identiques, les courants circulant dans leur canal respectif doivent être égaux pour le régime de saturation (V_{DS} fort).

Soit le courant i_{in} d'entrée et i_o le courant de sortie, appelé aussi l'image de i_{in} . Lorsque le transistor M1 est en saturation on peut écrire :

$$V_{DS1} = V_{GS1} \quad (A.1)$$

Supposons que $V_{DS2} \geq V_{GS} - V_{T2}$ alors V_{DS2} est plus grand que V_{T2} . Cette affirmation permet l'utilisation des équations du transistors MOS dans le régime de saturation. Dans le cas général, le rapport i_o/i_{in} est donné par :

$$\frac{i_o}{i_1} = \left(\frac{L_1 W_2}{W_1 L_2} \right) \left(\frac{V_{GS} - V_{T2}}{V_{GS} - V_{T1}} \right)^2 \left(\frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}} \right) \left(\frac{\mu_{02} C_{ox2}}{\mu_{01} C_{ox1}} \right) \quad (A.2)$$

où i_{in} est représenté par i_1 et λ est le paramètre concernant la modulation de la longueur de canal (V^{-1}) dans la région de saturation, μ_o est la mobilité en surface.

Or, pour les composants d'un miroir de courant fabriqués dans le même circuit intégré, les paramètres physiques tels que V_T , μ_o , C_{ox} , ... sont identiques pour les deux transistors. Nous pouvons donc simplifier la première équation par :

$$\frac{i_o}{i_1} = \left(\frac{L_1 W_2}{W_1 L_2} \right) \left(\frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}} \right) \quad (A.3)$$

Finalement, si $V_{DS2} = V_{DS1}$, le rapport i_o/i_1 devient :

$$\frac{i_o}{i_1} = \left(\frac{L_1 W_2}{W_1 L_2} \right) \quad (A.4)$$

Par conséquent, le rapport i_o/i_1 est une fonction des dimensions des deux transistors.

Modélisation des structures Métal-Oxyde-Semiconducteur (MOS) : Applications aux dispositifs mémoires

Nos travaux concernent la modélisation des structures MOS affectées par des défauts qui détériorent leurs propriétés électriques et par conséquent celles des dispositifs mémoires. Nous avons attaché une grande importance à la compréhension des phénomènes liés à la miniaturisation de la capacité et du transistor MOS qui sont les composants électroniques élémentaires des mémoires. Nos modèles basés sur de nombreuses études réalisées sur ces sujets, représentent de nouveaux outils d'analyses pour créer les modèles de base décrivant le fonctionnement plus complexe des dispositifs mémoires. Après un rappel des notations et des équations de base utilisées pour les capacités MOS et les transistors MOS, nous retraçons l'évolution des dispositifs mémoires jusqu'aux mémoires à nanocristaux. Dans une deuxième partie de notre travail, nous décrivons les différentes modélisations de la capacité MOS développées en fonction de l'effet parasite considéré : la poly-désertion de la grille, la non uniformité du dopage du substrat, de l'épaisseur d'oxyde et des charges fixes présentes dans la couche d'isolant. Nous avons ainsi pu proposer une méthode de détermination de la répartition de la charge générée dans l'oxyde par des stress électriques ainsi qu'une analyse de l'origine de ces charges. La troisième partie est consacrée aux modélisations du transistor MOS basées sur une approche segmentée. Celle-ci a été appliquée à l'étude des résistances séries et aux modélisations des dopages (grille et substrat), puis étendue à la modélisation des transistors à isolants ultra-minces. Nous présentons notamment les modifications de la caractéristique $I_{DS}(V_{GS}, V_{DS})$ du transistor MOS induites par les non uniformités énumérées ci-dessus. Enfin, nous appliquons nos modèles aux mémoires à nanocristaux de silicium. Nous proposons une modélisation de la charge localisée dans les nodules proches du drain, ce qui nous a permis de développer un modèle simulant l'opération d'écriture de ces mémoires. Les caractérisations électriques de ces structures à piègeages discrets sont également analysées à l'aide de nos modèles.

Modeling of Metal-Oxide-Semiconductor (MOS) devices : Applications to memory devices

Our study concerns the modeling of MOS devices affected by defects which deteriorate their electric properties and consequently those of the memory devices. A great importance is given to the knowledge of the phenomena induced by the miniaturization of the capacity and transistor MOS which compose the memories. Our models, based on various studies of these subjects, represent new analysis tools geared to basic models in order to describe the complex operations of the memory devices. After a review of the symbols and basic equations used for MOS capacitors and MOS transistors, we summarize the memory history up to the use of dots. The second part of our work describes various MOS capacitor modeling developed in presence of parasitic effects such as poly-depletion of the gate, non uniformity of the substrate doping, non uniformity of the oxide layer, and the oxide fixed charges. From these models, we come up with a method to determine the repartition of the charges generated within the oxide layer after electrical stress and an analysis of the charge origin. The third part is devoted to MOS transistor modeling based on a segmented approach. This was applied to the series resistance study and the doping (gate and substrate) modeling, then extended to the modeling of ultra-thin insulator transistors. First, we present the modifications of MOS transistor $I_{DS}(V_{GS}, V_{DS})$ characteristics induced by these non uniformities. Then, we use our models for the silicon nanocrystal memories. We propose a model of charge storage in the dots close to the drain which enabled us to develop a model simulating writing operation of these memories. The electric characterizations of these structures with discrete traps are also analyzed by using the same models.

Physique et Modélisation des Systèmes Complexes : Micro et Nanoélectronique

MOTS-CLES

Modélisation, Simulation, Capacité, Transistor MOS, Mémoire, Nanocristaux, High κ , Charges fixes, Non uniformités, Résistances d'accès, Poly-désertion, Courant de fuite de grille.