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 :

- 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 :

$$V = V_{DS} \frac{R}{\sum R}$$
(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, Rs 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 considèrerons que ces résistances Rs 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 Rs 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, Rs 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 (encerclé) 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, Rs et RD, 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* (**a**) *et (***b**).

Nous pouvons aussi noter que la chute de tension dans la résistance Rs 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 (Rs) 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 ($V_{DS'}$) s'écrit :

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

$$V_{T} = V_{FB} + 2\Phi_{F} + \frac{\sqrt{2qN_{A}\epsilon_{Si}}}{C_{ox}} \left(-V_{BS'} + 2\Phi_{F} \right)^{1/2}$$
(III.3)

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

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

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

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

En remplaçant VD'S par son expression dans l'expression du courant, il vient :

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

Pour simplifier l'expression, on pose:

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

ce qui permet d'écrire IDS sous sa forme finale :

$$I_{DS} = \frac{W}{L} \mu_0 C_{ox} \left(V_{GS'} - V_T \right) \frac{V_{DS}}{1 + \theta_1^* \left(V_{GS'} - V_T \right) + \theta_2 \left(V_{GS'} - V_T \right)^2}$$
(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 :

$$V_{GS} = V_{GS'} + R_S I_{DS}$$
(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 Ω (et $\theta_1 = -0.13 \text{ V}^{-1}$), ce qui correspond à une résistance d'accès à la source et au drain de 35 Ω , 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 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'égaler 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 $R_{SD-extr}$ (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 $R_{SD-extr}$ 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 la 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 :

$$R_{canal} = \frac{V_{D'S'}}{I_{DS}}$$
(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 : $W = 10 \ \mu m, \ t_{ox} = 3 \ nm, \ V_{FB} = -1 \ V,$ $N_A = 7 \times 10^{23} \ m^{-3}, \ \mu_0 = 500 \ cm^2 V^{-1} s^{-1},$ $R_S = R_D = 50 \ \Omega, \ V_{DS} = 0.2 \ V.$

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 $V_{GS} = 2 V$ (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 : $W = 10 \ \mu m$, $L = 0.2 \ \mu m$, $N_A = 7 \times 10^{23} \ m^{-3}$, $t_{ox} = 3 \ nm$, $V_{FB} = -1 \ V$, $V_{DS} = 0.2 \ V$, $\mu_0 = 500 \ cm^2 V^{-1} s^{-1}$.

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 :

$$I_{\text{DSs}} \mu_0 = I_{\text{DS}} \mu_{\text{eff}} \tag{III.11}$$

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_{DS}(V_{GS},V_{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 : $W = 10 \ \mu m, \ L = 1 \ \mu m,$ $N_A = 7 \times 10^{23} \ m^{-3}, \ R_S = R_D = 50 \ \Omega,$ $\mu_0 = 500 \ cm^2 V^{-1} s^{-1}, \ t_{ox} = 3 \ nm, \ et$ $V_{FB} = -1V.$

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 : $W = L = 10 \ \mu m, t_{ox} = 7 \ nm$ et $N_A = 5.8 \times 10^{23} \ m^{-3}$.

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 IDS, jusqu'à ce que les courbes μ_0/μ_{eff} à VDS 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 100 Ω 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 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: $W \times L = 1 \times 0.5 \ \mu m^2$, $t_{ox} = 3 \ nm$, $N_A = 7 \times 10^{23} \ m^{-3}$ et $\mu_0 = 500 \ cm^2 V^{-1} s^{-1}$.

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 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, tox_{eff}, qui prend en compte l'isolant de grille et la poly-déplétion :

$$V_{T} = V_{FB} + 2\Phi_{F} + \frac{tox_{eff}}{\varepsilon_{ox}} \sqrt{2\varepsilon_{si}qN_{A}\Phi_{F}}$$
(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 tox_{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 Cmin-Cmax nécessite la connaissance préalable de la capacité Cox. 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 ($\Delta Vp1$) et par la réduction de la longueur de grille (Δ Vp2).

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 polysilicium 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 :

$$Q_{SC} = \pm \sqrt{2kT\varepsilon_{Sip0}} \left[\frac{n_0}{p_0} \left(\exp(\beta(\Psi_S - \phi_C)) - \beta\Psi_S - \exp(-\beta\phi_C) \right) - 1 + \exp(-\beta\Psi_S) + \beta\Psi_S \right]^{\frac{1}{2}}$$
(III.13)

$$Q_{G} = \pm \sqrt{2kT\varepsilon_{SinG0}} \left[\frac{p_{G0}}{n_{G0}} \left(\exp(-\beta\Psi_{SG}) + \beta\Psi_{SG} - 1 \right) - 1 + \exp(+\beta\Psi_{SG}) - \beta\Psi_{SG} \right]^{\frac{1}{2}}$$
(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 polysilicium. Dans ce cas l'expression de la charge de la grille (que l'on notera dans ce cas Q_{DG}) se simplifie pour devenir :

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

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

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

$$Q_{n} = Q_{SC} - Q_{DSC} = Q_{SC} \pm \sqrt{2kT\varepsilon_{Si}p_{0}} \left[\exp(\beta\Psi_{S}) - \left(1 - \frac{n_{0}}{p_{0}}\right)\beta\Psi_{S} - 1 \right]^{\frac{1}{2}}$$
(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 : $N_A = 7 \times 10^{23} \text{ m}^{-3}, N_G = 2 \times 10^{24} \text{ m}^{-3},$ $t_{ox} = 3 \text{ nm}, V_{FB} = -1 \text{ V}, 20 \text{ points}$ 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 = $10\mu m$, L = $10\mu m$) qui apparaissent pour un dopage

de grille inférieur à 10²⁵m⁻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 $W \times L = 10 \times 1 \ \mu m^2$, dont le dopage varie le long du canal entre $N_{Amin} = 5 \times 10^{23} \ m^{-3}$ et $N_{Amax} = 1 \times 10^{24} \ m^{-3}$. 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, $W \times L = 10 \times 1 \ \mu m^2$, $t_{ox} = 3$ nm et $\mu_0 = 500 \ cm^2 V^{-1} s^{-1}$.

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 × 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 $W \times L = 10 \times 10 \ \mu m^2$ 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}.

$$I_{G} = I_{GS} + I_{GD}$$
(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 :

$$I_{G} = \frac{WL}{2} (Q_{nS} F_{imp} T_{S} + Q_{nD} F_{imp} T_{D})$$
(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 :

$$T_{s,D} = \exp\left(-2\sqrt{\frac{2qm_{nox}}{\hbar^2}} \int_0^{t_{ox}} \sqrt{\Phi_{bS,D}} \, dy\right)$$
(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×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 :

$$I_{DSN-1} - I_{GDN-1} = I_{DSN} + I_{GSN}$$
(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 :

$$\Phi_{\rm CN} = \frac{R_{\rm N} \Phi_{\rm CN+1} + R_{\rm N+1} \Phi_{\rm CN-1} + R_{\rm N} R_{\rm N+1} I_{\rm GN}}{R_{\rm N} + R_{\rm N+1}}$$
(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×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 ($t_{ox} = 1.5$ nm), c'est à dire présentant un courant de fuite.



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

Le modèle segmenté GLNMOS, décrit au paragraphe (III.4.1.), donne accès à de nombreux paramètres (Ψ s, Φ c, I_G, Is, 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 V et V_{GS} = 2 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 (Is > 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 $V_{DS} = 50$ mV et de grille $V_{GS} = 2$ 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 µm).



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 µm de long ainsi que le courant de drain, I_{DS} en omettant le courant de fuite (**b**). Les paramètres du transistor sont : $W = 10 \mu m$, $V_{DS} = 50 mV$, $t_{ox} = 1.5 nm$.

Pour les transistors plus longs (ici 5 et 10 μ 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, Is, et de grille IG, d'un transistor MOS de longueur L = 5 μ 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 Is et ID 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 IG 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 IDS). 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×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 $W\times L = 10\times 10 \ \mu m^2$, ayant un oxyde de grille fin ($t_{ox} = 1.5 \ nm$). 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 μ A) 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 $t_{ox} = 1.5$ 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 descaractéristiques 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 fonction de V_{DS} I_D en(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 $V_G = 1.2$ 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 $\Phi_{\rm C}$ sont imposées au niveau du drain $(\Phi_{\rm C}(1,y) = V_{\rm D})$ et au niveau de la source $(\Phi_{\rm C}(5,y) = V_{\rm S})$. La figure (III.28.a) montre que la courbe $\Phi_{\rm C}(x,y)$ 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 $V_{DS} = 50 \text{ mV}$ et $V_{GS} = 1.2 \text{ V}$, 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 (tox = 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 M₁ (W₁ × L₁ = 5 × 2 µm²) et M2 (W₂ × L₂ = 1 × 10 µm²). 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 (M2) 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, I₀, met en évidence l'impact de la position et de la taille de la non uniformité de t_{ox} du transistor M₂. 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 M2 de référence (I_{0ref}) et un miroir de courant avec un transistor M₂ défectueux (I_{0defect}) :

$$S_{I_0} = \frac{I_{0ref} - I_{0defect}}{I_{0ref}} \times 100$$
(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_{10} 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 M2 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 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 Io (pour $V_{GP} = 0V$).

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

- le coefficient S10 est indépendant de la largeur W du transistor,
- le coefficient S₁₀ 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.