

THESE

En vue de l'obtention du titre de

Docteur de l'Université des Sciences et Technologies de Lille

Spécialité : Electronique

Clément Charbuillet

Recherche et étude de dispositifs à commutation abrupte

Soutenue devant la commission d'examen le 9 Novembre 2007.

Membres du jury :

M.	G. Dambrine	Président
M.	A. Ionescu	Rapporteur
M.	D. Flandre	Rapporteur
M.	F. Balestra	Examineur
M.	T. Hoffman	Examineur
M.	T. Skotnicki	Co-Directeur de thèse
M.	E. Dubois	Directeur de thèse

Thèse réalisée au sein de STMicroelectronics et entrant dans le cadre du
laboratoire commun IEMN - STMicroelectronics.

L'électronique grand public est entraînée par les applications numériques, dont le composant fondamental est le transistor MOS. Ses performances sont au premier ordre données par le courant de conduction I_{ON} , élevé pour des circuits rapides, et le courant de fuite I_{OFF} , faible pour les circuits basse consommation. Ces deux grandeurs sont intimement liées par la pente sous le seuil S du transistor, traduisant sa capacité à passer de l'état bloqué à l'état passant.

Nous proposons une modélisation du courant sous le seuil, notamment en décrivant l'évolution de la pente sous le seuil avec les paramètres technologiques, et soulignons sa limite fondamentale à kT/q (60mV/dec @ 300K).

Puis diverses architectures permettant de passer cette limitation sont évaluées.

La plus prometteuse est le transistor I-MOS (*Impact Ionization MOS*), basé sur l'effet d'avalanche. Il est étudié d'un point de vue théorique, en s'appuyant sur des simulations TCAD et des données de littérature.

Les performances de circuits réalisées en technologie I-MOS complémentaire sont évaluées. Ces circuits montrent un gain en terme de vitesse mais une consommation accrue par rapport à la technologie CMOS conventionnelle.

Enfin, nous poussons les limites de l'intégration planaire "conventionnelle" de dispositifs I-MOS en réalisant les dispositifs les plus courts de la littérature (17nm de longueur de grille). Nous montrons et expliquons qu'il existe une taille minimale sous laquelle le transistor I-MOS n'est plus fonctionnel. Nous proposons également une intégration verticale permettant de s'affranchir de cette limite en terme de densité d'intégration.

Customer electronics is drained by digital applications, which is essentially based on the MOS transistor. Its performance at the first order is given by the conduction current I_{ON} , high for high speed circuits, and the leakage current I_{OFF} , low for low power applications. Both values are tightly linked by the subthreshold slope S , reflecting the efficiency of the transistor to switch from the OFF state to the ON state.

We propose a modeling for the subthreshold current, especially by describing the impact of the technological parameters on the subthreshold slope. We emphasize its fundamental limit of kT/q (60mV/dec @ 300K).

Then different architectures that may be able to overcome this limit are reviewed.

The most promising one is the Impact Ionization MOS transistor. It is studied from a theoretic point of view, relying on TCAD simulations and literature data.

The performance of circuits realized in complementary I-MOS technology are evaluated. These circuits present a gain in term of speed but an increased consumption with respect to the conventional CMOS technology.

Finally, we push the limits of the "conventional" planar integration of I-MOS devices by realizing the shortest of the literature (17nm gate length). We show and explain that a minimal size exists, under which the I-MOS transistor is not functional any more. We also propose a vertical integration that allow to overcome this limitation in terms of integration density.

Remerciements

Je tiens à remercier particulièrement:

Emmanuel, pour les nombreuses discussions qui ont fait avancer le schmilblick, pour les moyens mis à ma disposition, pour m'avoir permis de terminer dans de bonnes conditions.

Thomasz, pour m'avoir confié ce travail de recherche, pour n'avoir jamais considéré que "c'est comme ça" était une réponse satisfaisante, et pour avoir aplani un grand nombre de difficultés administratives.

Stéphane, sans qui l'intégration technologique de l'I-MOS n'aurait pu être si avancée.

Philippe, Pam pour leur soutien dans le déroulement la thèse.

Fabrice, pour les discussions où d'avantage que d'habitude j'avais l'impression d'être compris.

Fabienne, Marie-Pierre, Yves, Campi, Alexandre, Nicolas, Pierre pour avoir transformé des fantasmes Power Point en réalité technologique.

Arnaud, pour les échanges sur la modélisation du courant sous le seuil du MOS.

Alexandre pour avoir embarqué 3 ans avant moi des structures permettant de tenter l'intégration verticale.

Fred, Claire, Claire (si, 2 fois), Stéphane (le second), Ben, Markus, Jessy, Alexandre, Russ, Daniel, Robin, Romain et les premiers pour leurs discussions actives, en particulier pour leurs apports dans la mise au point du procédé d'intégration.

Je sers la science et c'est ma joie.

Discipulus Simplex

Fort peu de choses sont impossibles.

Remerciements.....	3
Introduction générale.....	9
Chapitre 1. Présentation du transistor MOS.....	11
Introduction.....	12
1.1. Introduction au transistor MOS.....	13
1.1.1. Principe de fonctionnement.....	14
1.1.2. Performances.....	14
1.1.3. Description morphologique du transistor MOS.....	16
1.1.4. Principales étapes du procédé de fabrication du transistor MOS.....	20
1.2. Eléments de modélisation.....	24
1.2.1. Potentiel de Fermi, potentiel de surface et courbure de bandes.....	25
1.2.2. Tension de seuil V_{th}	26
1.2.3. Courant de conduction I_{on}	28
1.2.4. Pente sous le seuil S	30
1.2.5. Courant de fuite I_{off}	31
1.3. Réduction des dimensions: les effets canaux courts.....	31
1.3.1. Noeuds technologiques.....	31
1.3.2. Les effets canaux courts.....	33
1.3.3. Empilement de grille.....	34
1.3.4. Optimisation du dopage canal.....	37
1.3.5. Jonctions.....	37
1.4. Amélioration des performances: le silicium contraint.....	38
1.4.1. Principe.....	38
1.4.2. Contrainte uniaxiale, biaxiale, tensile, compressive.....	38
1.4.3. Méthodes technologiques pour obtenir la contrainte.....	38
1.4.4. Quelle méthode pour quel porteur?.....	42
1.5. Architectures de rupture.....	42
1.5.1. Nouvelles architectures de jonctions.....	42
1.5.2. Silicium sur isolant.....	44
1.5.3. Transistors multigrilles.....	47
Conclusion du chapitre.....	50

Chapitre 2. Le transistor MOS sous le seuil.....	51
Introduction	52
2.1. Pente sous le seuil des dispositifs semi-conducteurs	53
2.1.1. Limite fondamentale de la pente sous le seuil.....	53
2.1.2. Pente des dispositifs semi-conducteurs élémentaires.....	53
2.2. Modélisation de la pente sous le seuil du TMOS.....	55
2.2.1. La transformation tension dopage (VDT)	55
2.2.2. Pente sous le seuil en technologie substrat massif.....	56
2.2.3. Amélioration du modèle de pente	61
2.2.4. Extension au transistor SOI complètement déplété.....	63
2.2.5. Discussion sur les paramètres d'ajustement	65
2.3. Modélisation du courant sous le seuil	67
2.3.1. Principe de modélisation	67
2.3.2. Approche capacité MOS	69
2.3.3. Approche statistique.....	69
2.3.4. Comparaison des modèles.....	72
2.3.5. Relation avec la tension de grille	73
2.3.6. Discussion et interprétations	73
Conclusion du chapitre.....	76
Chapitre 3. Dispositifs présentant une pente meilleure que 60mV/dec	77
Introduction.....	78
3.1. Les structures de rupture	79
3.1.1. Matériaux à transition de Mott.....	79
3.1.2. Résultats morphologiques	80
3.1.3. Transistor à grille mobile	83
3.2. Structures utilisant l'effet tunnel.....	84
3.2.1. Transistor à jonctions Schottky	84
3.2.2. Structure tunnel résonnant.....	85
3.2.3. Transistor à effet tunnel (TFET)	86
3.3. Structures mettant en jeu un phénomène d'amplification.....	87
3.3.1. Thyristor	87
3.3.2. Hybride bipolaire/MOS.....	88
3.3.3. SOI partiellement déplété.....	89

3.3.4. MOS à Ionisation par Impact (I-MOS)	90
Conclusion du chapitre.....	91
Chapitre 4. Le transistor de type I-MOS.....	92
Introduction.....	93
4.1. Introduction au transport dans les semi-conducteurs	94
4.1.1. Equations élémentaires de base de la physique du transport	94
4.1.2. Modèle dérive-diffusion.....	94
4.1.3. Modèle thermodynamique.....	95
4.1.4. Modèle hydrodynamique.....	95
4.2. Le phénomène d'ionisation par impact.....	96
4.2.1. Description – phénomène d'avalanche	96
4.2.2. Modélisation du phénomène d'ionisation par impact.....	97
4.2.3. Modélisation des coefficients d'ionisation par impact	98
4.2.4. Choix d'un modèle.....	102
4.3. La diode p-i-n.....	105
4.3.1. Description et grandeurs caractéristiques.....	105
4.3.2. Modélisation d'une diode p-i-n.....	107
4.3.3. Tension d'avalanche d'une diode p-i-n	109
4.3.4. Modélisation du courant d'une diode p-i-n.....	112
4.4. Le transistor I-MOS	113
4.4.1. Présentation	113
4.4.2. Polarisation.....	116
4.4.3. Performances	121
Conclusion du chapitre.....	127
Chapitre 5. Les circuits en I-MOS	128
Introduction.....	129
5.1. Conception d'un inverseur CI-MOS.....	130
5.1.1. L'inverseur CMOS	130
5.1.2. L'inverseur CI-MOS.....	133
5.1.3. Conception d'un inverseur CI-MOS.....	135
5.1.4. Spécificités de l'inverseur CI-MOS.....	142
5.2. Performances d'un inverseur CI-MOS	143
5.2.1. Structures simulées – stratégie de comparaison.....	143

5.2.2. Consommation statique	144
5.2.3. Consommation dynamique.....	144
5.2.4. Puissance de court-circuit.....	148
5.2.5. Fréquence maximum de fonctionnement	150
5.3. Circuits élémentaires	155
5.3.1. La porte NAND.....	155
5.3.2. La porte NOR.....	158
5.4. La SRAM	160
5.4.1. Présentation	160
5.4.2. Calcul de la marge de bruit statique.....	161
5.4.3. Marge de bruit d'une cellule SRAM réalisée en technologie I-MOS.....	163
Conclusion du chapitre.....	164
Chapitre 6. Intégration de l'I-MOS	165
Introduction.....	166
6.1. Etat de l'art.....	167
6.1.1. Double lithographie.....	167
6.1.2. Double espaceur	168
6.1.3. I-MOS en forme de L	169
6.1.4. Utilisation du masque S/D décalé	170
6.1.5. Comparaison des intégrations	170
6.2. Intégration horizontale	173
6.2.1. Procédé de fabrication.....	173
6.2.2. Résultats électriques – observations générales.....	175
6.2.3. Résultats électriques – impact de la géométrie des dispositifs.....	177
6.2.4. Extraction de la figure de mérite I_{on}/I_{off}	182
6.3. Intégration verticale.....	183
6.3.1. Procédé de fabrication.....	186
6.3.2. Mesures sur l'intégration verticale.....	191
Conclusion du chapitre.....	194
Conclusion générale et perspectives	195
Bibliographie.....	199

Introduction générale

Le numérique envahit peu à peu notre quotidien. Le développement de l'informatique personnelle est flagrant, mais le multimédia également profite du numérique: ventes records de baladeurs numériques et de lecteurs DVD, explosion des sites de vente de musique en ligne, introduction de la Télévision Numérique Terrestre et de la télévision haute définition, offres groupées téléphonie/Internet/télévision des fournisseurs d'accès Internet en sont les aspects les plus visibles. L'électronique prend également une place de plus en plus importante dans l'automobile (commandes d'airbag, assistance au freinage...).

Un exemple moins connu est l'identification par radiofréquence ou RFID (*Radio Frequency Identification*). Il s'agit d'une méthode pour stocker et récupérer des données à distance en utilisant des marqueurs. Ces marqueurs sont de petits objets, tels que des étiquettes autoadhésives, qui comprennent une antenne et une puce électronique. Ils sont utilisés pour les contrôles d'accès par badge, l'ouverture sans serrure de certaines voitures, et sont destinés à remplacer les étiquettes des marchandises.

"The economic future is digital" [Barett'02]. Tel est le message délivré le 3 octobre 2002 par Graig Barett, ex-PDG d'Intel et aujourd'hui chargé du développement stratégique du fondateur, même si le marché du semi-conducteur est encore loin de celui de l'automobile: le chiffre d'affaire s'élève à près de 235 milliards de dollars en 2005 [Gartner'06] pour les semi-conducteurs contre près de 1600 milliards de dollars pour l'automobile en 2003 [Oica'03].

Le marché du semi-conducteur est tiré par le composant fondamental de ces applications numériques, le transistor MOS. Celui-ci joue le rôle d'interrupteur permettant de différencier le 0 du 1 logique, représentés par des niveaux de courant. Il est la cible des principaux investissements de recherche et développement. Pour exemple, le site de l'Alliance Crolles 2 représente un investissement de trois milliards d'euros en 5 ans. C'est le plus important investissement industriel européen de ces dix dernières années [C2Alliance'03].

Le premier chapitre de ce manuscrit présente le transistor MOS, associant principes généraux et intégration technologique. En affinant la modélisation de ce dispositif, il devient possible mettre en évidence les leviers technologiques disponibles pour améliorer les performances. Ce chapitre passe également en revue les architectures MOS de rupture, telles les intégrations sur substrat SOI ou SON ou les transistors double grille.

Le second chapitre traite de la modélisation sous le seuil du courant du transistor MOS (planaire et SOI) en utilisant la transformation tension/dopage (VDT). Cette méthode permet

d'assimiler les effets de champ électrique latéral à une modification apparente du dopage canal, tout en gardant la modélisation usuelle du transistor. L'une des conclusions essentielles, bien que commune, de ce chapitre, est que la pente sous le seuil du transistor MOS est limitée à kT/q , quelle que soit la technologie employée.

C'est pourquoi le troisième chapitre passe en revue différents dispositifs présentant des caractéristiques électriques proches de celles du MOS, mais pouvant permettre de passer cette limite thermodynamique de 60mV/dec à température ambiante. Ils peuvent être complètement innovants, utiliser un effet physique différent, ou bien intégrer un système d'amplification. Parmi ces dispositifs, le transistor I-MOS, basé sur l'effet d'ionisation par impact, semble le plus prometteur. C'est donc sur lui que porte l'essentiel des travaux de recherche présentés dans ce manuscrit.

Le quatrième chapitre introduit le transistor I-MOS d'un point de vue théorique, en commençant par présenter le phénomène d'ionisation par impact. Son fonctionnement est expliqué en détail, sur la base de simulations TCAD et de modélisation analytique.

Le cinquième chapitre illustre de quelle manière les avantages de l'I-MOS peuvent être utilisés dans les circuits. Les performances d'un inverseur réalisé en technologie I-MOS complémentaire sont comparées à celles d'un inverseur CMOS. Des circuits plus complexes comme un point mémoire SRAM, une porte NAND et une porte NOR sont également abordés.

Le sixième et dernier chapitre traite de l'intégration technologique, dont le principal enjeu est l'asymétrie n/p de l'I-MOS. Une approche planaire s'inspirant de celles présentes dans la littérature et une intégration verticale innovante sont comparées.

Chapitre 1. Présentation du transistor MOS

Dans la première partie de ce chapitre, le transistor MOS est présenté d'un point de vue fonctionnel et technologique. Puis, ses performances statiques sont abordées.

La modélisation au premier ordre du courant du transistor est abordée dans la seconde partie. Elle permet de montrer que les performances augmentent lorsque la longueur de grille du transistor diminue. C'est pourquoi les évolutions technologiques ont longtemps eu pour but principal de réaliser les grilles les plus courtes possibles.

La baisse continue des dimensions entraîne des effets parasites de plus en plus nombreux: effets canaux courts, poly-déplétion, ou encore effets quantiques tels les fuites par effet tunnel. Ces phénomènes physiques ainsi que les moyens technologiques mis en oeuvre pour contrecarrer ces effets parasites sont présentés dans la troisième partie.

La quatrième partie de chapitre explique comment améliorer les performances grâce à de la contrainte mécanique, et présente des architectures de ruptures: silicium sur isolant (SOI, SON) et multigrilles.

Introduction

Le marché du semi-conducteur est drainé par le transistor MOS sur silicium. Il présente les avantages d'une grande densité d'intégration et d'une quasi universalité: les applications des autres composants semi-conducteurs sont limitées à des "niches technologiques": applications opto-électroniques (télécommunication) pour les composés III-V, analogiques à haute fréquence (radar, GSM) pour les transistors bipolaires.

En outre, le transistor MOS est le composant essentiel des applications numériques, actuellement en plein essor. En effet, les circuits associant des transistors MOS de type n (courant d'électrons) avec des transistors MOS de type p (courant de trous) permettent de réaliser des circuits dont la consommation statique est extrêmement limitée. On parle alors de technologie MOS complémentaire (CMOS pour *Complementary MOS*).

Enfin, les performances du transistor MOS s'améliorent à mesure que sa dimension caractéristique, la longueur de grille, diminue. Cela implique que réduire les dimensions permet à la fois d'améliorer les performances (fréquence d'horloge d'un micro processeur par exemple) et d'augmenter la quantité de dispositifs produits par plaque, et donc la rentabilité.

Ce chapitre présente le transistor MOS, d'un point de vue fonctionnel et technologique. Il introduit les expressions de modélisation élémentaire du transistor MOS, qui permettent de déterminer les leviers disponibles pour l'amélioration des performances.

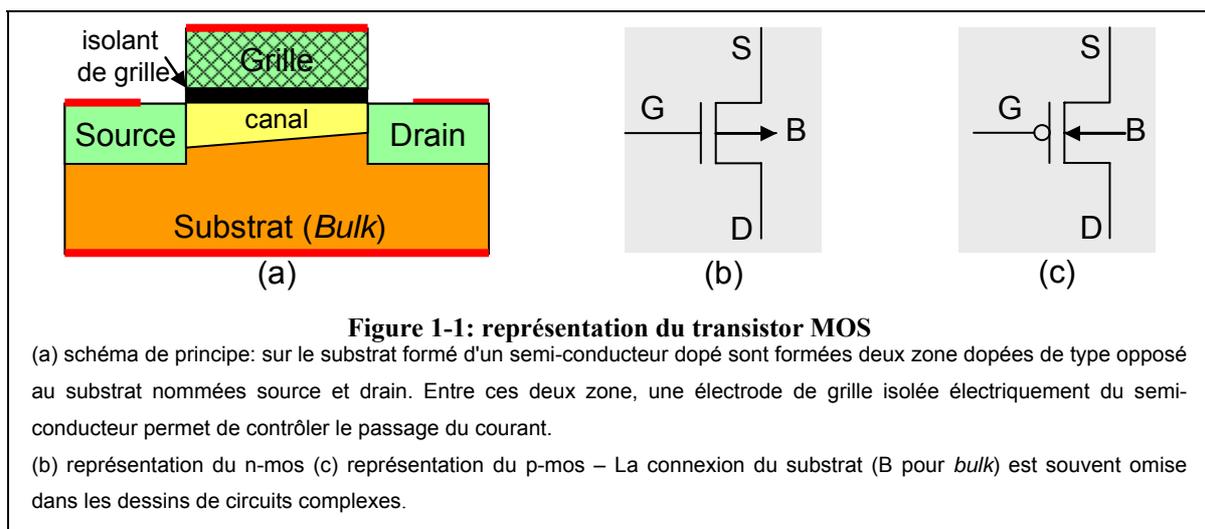
1.1. Introduction au transistor MOS

Les **transistors à effet de champ**, ou FET pour *Field Effect Transistor*, se comportent comme une résistance dont la valeur est contrôlée par un champ électrique, c'est-à-dire une tension. Il s'agit généralement d'un semi-conducteur séparé de l'électrode de commande par un oxyde. On parlera donc de **MOSFET** pour *Metal Oxide Semiconductor FET*. Par la suite, nous utiliserons indifféremment l'abréviation MOSFET ou **TMOS** (*Transistor Métal Oxyde Semi-conducteur*).

Dans le transistor MOS, le phénomène de conduction implique majoritairement un seul type de porteurs (électrons ou trous), par opposition au transistor bipolaire, où les deux types de porteurs sont impliqués. Un transistor MOS où le courant de conduction est un courant d'électrons est appelé **n-mos** ou **transistor de type n**. De même, un transistor où le courant de conduction est un courant de trous est appelé **p-mos** ou **transistor de type p**. Par défaut, c'est un transistor de type n qui servira d'étude dans ce manuscrit.

Le transistor MOS compte quatre électrodes (**figure 1-1**):

- la **grille** permet de contrôler la résistivité du semi-conducteur sur une zone appelée **canal**
- la **source** et le **drain** viennent contacter le canal de part et d'autre. Par extension, on appellera zones de source et de drain les zones dopées du semi-conducteur où sont pris ces contacts.
- le **substrat** sert de tension de référence. Il s'agit généralement de la masse pour les transistors de type n, de la tension d'alimentation V_{DD} pour les transistors de type p.

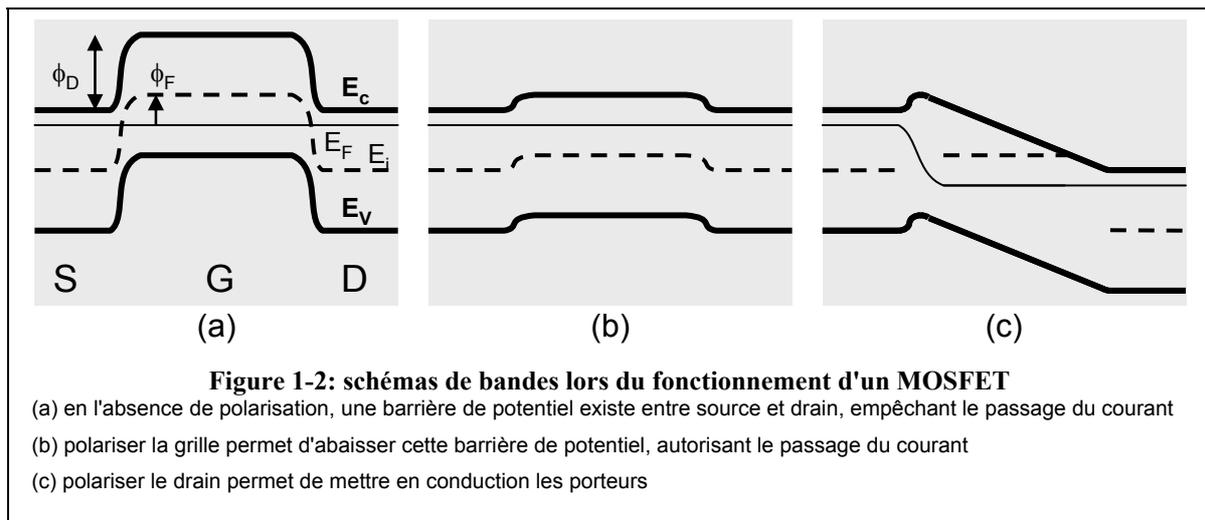


Il existe différentes architectures de MOSFETs: à canal enterré, à canal surfacique. Actuellement, seul ce dernier est utilisé et sera donc présenté ici. Plus que la résistivité du canal, c'est le **courant** que fournit le dispositif qui est étudié.

1.1.1. Principe de fonctionnement

Le canal et les zones source/drain ont des dopages de nature opposée. Pour un transistor n-mos, les zones source/drain sont de type n et le canal de type p. C'est l'inverse pour un p-mos. Lorsque source, drain, et grille sont polarisés à la masse, les porteurs à la source voient une barrière de potentiel (**figure 1-2a**). Polariser la grille permet de moduler cette barrière par effet électrostatique (**figure 1-2b**). C'est le **régime de blocage**.

Une fois la barrière de conduction supprimée, l'électrode de grille permet de moduler la quantité de charges dans le canal. Ces charges sont mises en conduction par le champ latéral source/drain **figure 1-2c**). C'est le **régime de conduction**.



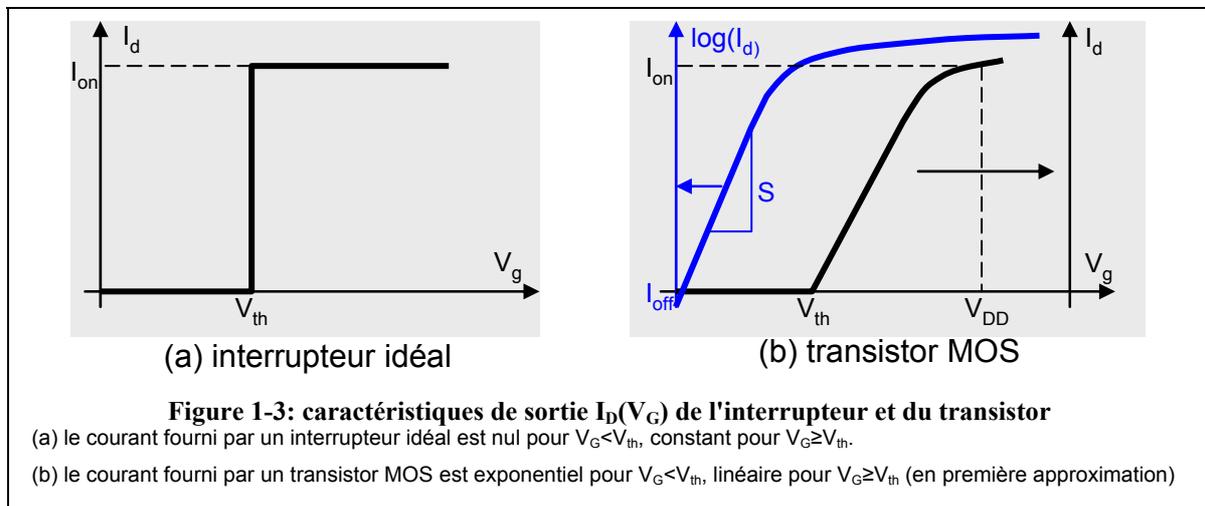
1.1.2. Performances

Il a été signalé en introduction que le TMOS est principalement utilisé pour des applications logiques. Dans ce genre d'applications, les deux points de fonctionnement essentiels sont ceux qui traduisent le "0" et le "1" logiques. Le dispositif idéal est un interrupteur (*switch*) qui permet de basculer d'un état bas (courant nul) à un état haut (courant important) en fonction de la tension de grille (**figure 1-3a**). La commutation se fait de façon abrupte à une valeur remarquable appelée **tension de seuil** (V_{th}).

Même s'il remplit sa fonction d'interrupteur, le TMOS est loin d'être idéal (**figure 1-3b**):

- à l'état bas, un faible courant circule. C'est le **courant de fuite** I_{off} . Ce courant est responsable de consommation d'énergie et doit donc être le plus faible possible.

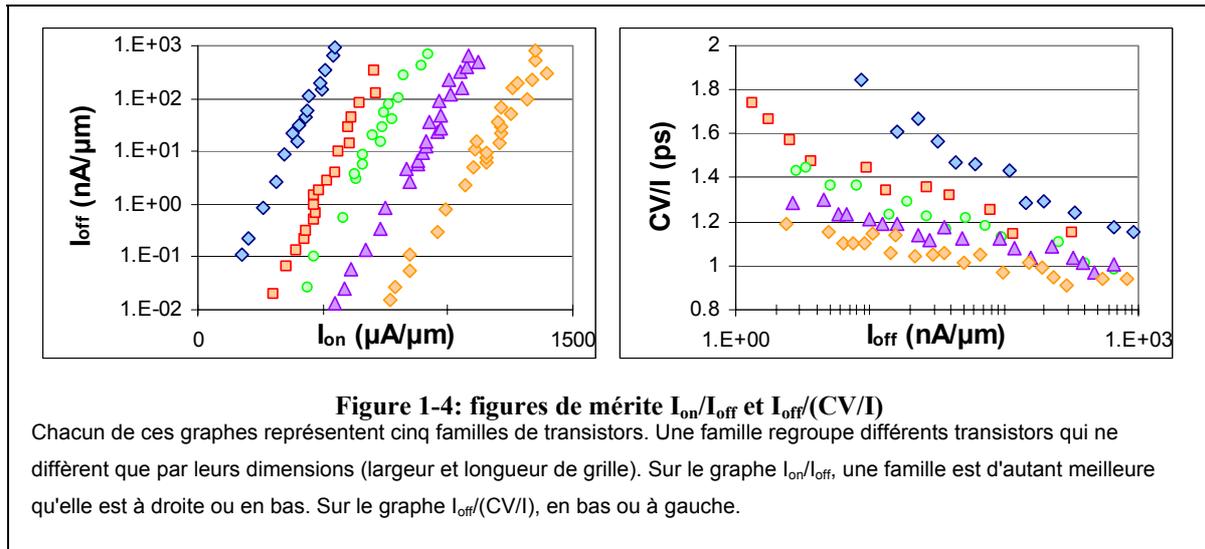
- à l'état haut, le **courant de conduction** I_{on} est limité. Ce courant sert à charger et décharger les capacités des circuits. Il doit donc être le plus élevé possible pour avoir des dispositifs rapides.
- la commutation n'est pas abrupte. Le courant passe de l'état haut à l'état passe de façon exponentielle. Cette partie est caractérisée par la **pente sous le seuil** S , qui est la différence de tension à appliquer sur la grille pour obtenir une variation de courant de drain d'une décade. Son unité courante est le millivolt par décade (mV/dec). Elle est égale à l'inverse de la pente mathématique de la courbe $\log(I_D)(V_G)$.
- de manière similaire à l'interrupteur, la **tension de seuil** V_{th} est définie comme étant la tension de grille à partir de laquelle le transistor commence à conduire.



Ces quatre grandeurs I_{on} , I_{off} , V_{th} et S sont représentatives des performances statiques d'un transistor. Une façon rapide d'évaluer les performances d'un transistor est d'utiliser les figures de mérites que sont les graphes I_{on}/I_{off} et $I_{off}/(CV/I)$ (**figure 1-4**).

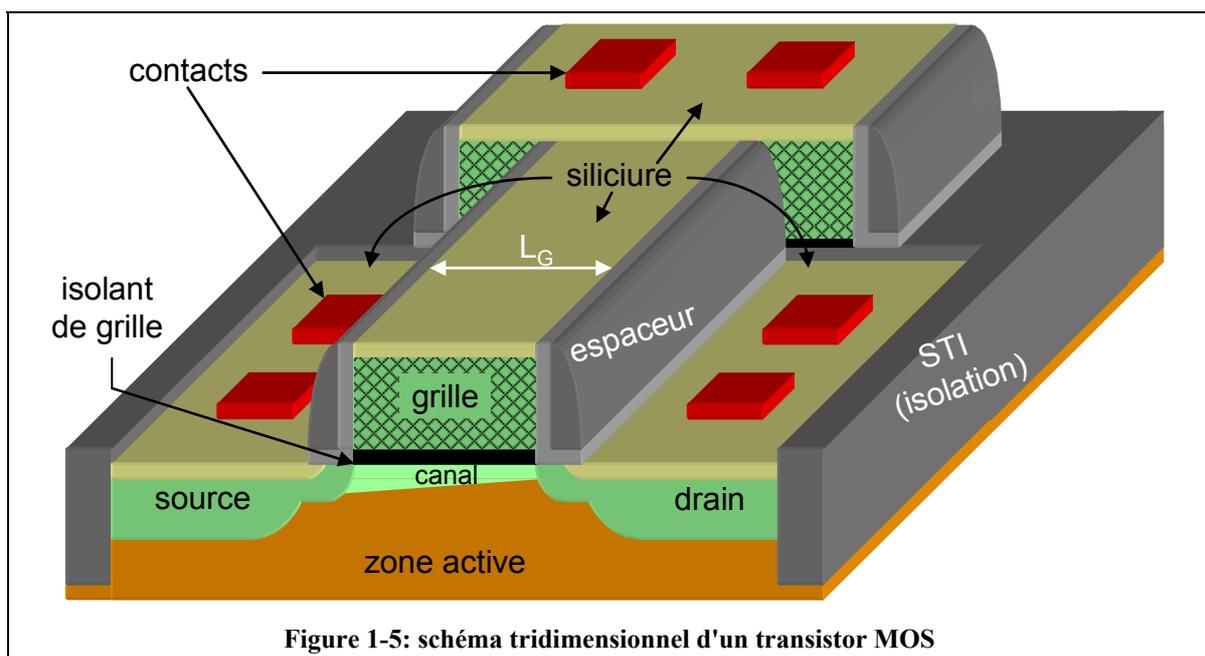
Le graphe I_{on}/I_{off} représente le logarithme du courant de fuite I_{off} en fonction du courant de conduction I_{on} . Par conséquent, un dispositif est d'autant meilleur que le point le représentant est situé en bas à droite du graphique. Une famille de dispositifs ne différant que par la longueur de grille forme une droite sur ce graphe. Une famille est donc meilleure si sa représentation est plus basse ou plus à droite.

Le graphe $I_{off}/(CV/I)$ représente le temps de délai CV/I en fonction du courant de fuite I_{off} . C représente les capacités intrinsèques du transistor, de couplage et d'interconnexions. La tension V est la tension d'alimentation V_{DD} et I le courant de conduction I_{on} . La grandeur CV/I est donc homogène à un temps, et est représentative du délai intrinsèque d'une porte logique, et peut donc être reliée à la vitesse de commutation des circuits. Sur ce graphe, un point sera d'autant meilleur qu'il est situé en bas à gauche.



1.1.3. Description morphologique du transistor MOS

Les différentes parties du transistor et leur influence sur les performances statiques sont décrites. Dans un souci de clarté, seuls les aspects nécessaires à la compréhension du transistor sont présentés.



1.1.3.1. Grille

C'est l'électrode de commande. Elle est réalisée en polysilicium dopé jusqu'à dégénérescence et est souvent considérée comme métallique. La longueur de grille L est la plus petite dimension que l'on obtient sur la plaque. En effet, on montrera plus loin dans ce chapitre que le courant de conduction est inversement proportionnel à la longueur de grille. Réduire la longueur de grille a donc longtemps été le moteur de l'amélioration des performances du transistor MOS.

1.1.3.2. Isolant de grille

C'est la zone qui isole électriquement la grille du canal. Il s'agit généralement d'un oxynitride de silicium dont la stoechiométrie varie selon la position: il est plus riche en azote près du semi-conducteur que de la grille. Cet isolant doit avoir deux qualités:

- être un bon isolant électrique. Dans le cas contraire, un courant de fuite entre la grille et le drain apparaît, augmentant ainsi I_{off} et donc la consommation du transistor.

- la capacité de grille doit être la plus élevée possible pour que la grille soit la plus efficace possible. C'est pourquoi l'épaisseur de l'oxyde de grille t_{ox} diminue progressivement ($C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$). Cette réduction s'accompagne d'une baisse de la tension d'alimentation. En effet,

l'oxyde de silicium soumis à un champ électrique supérieur à 10^6 V/cm se dégrade et n'isole plus correctement. Ainsi, la tension d'alimentation est adaptée à l'épaisseur d'oxyde (**figure 1-6**).

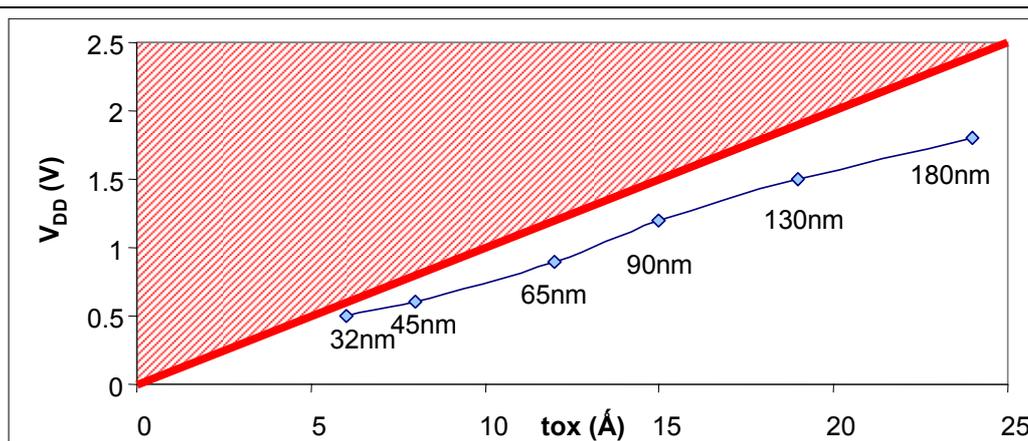


Figure 1-6: évolution de la tension d'alimentation avec l'épaisseur d'oxyde

La zone hachurée correspond à un champ électrique dans l'oxyde de grille supérieur à son champ de claquage. Au fur et à mesure que les générations technologiques progressent, le champ électrique dans l'isolant de grille est de plus en plus proche du champ de claquage.

Données issues de [ITRS'05] pour la catégorie *High Performance*.

1.1.3.3. Dopage canal

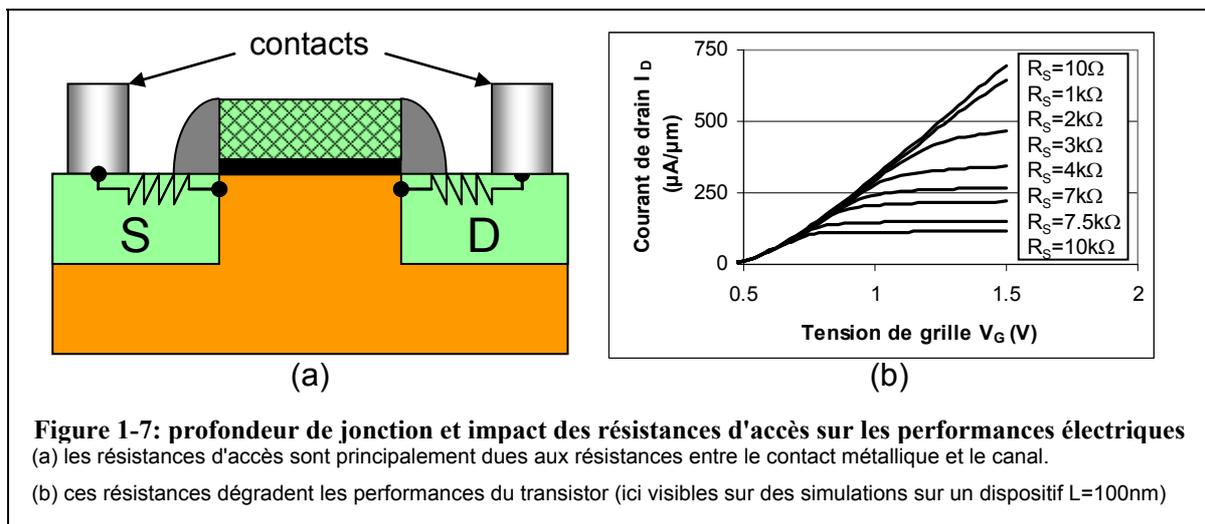
C'est la quantité de dopants ionisés introduits en surface sous la grille. Comme dans tout semi-conducteur, la position du niveau de Fermi dans le silicium dépend du dopage. Notamment, cela implique que la hauteur de barrière du TMOS dépend du dopage canal. C'est principalement en ajustant que le dopage canal qu'est fixée la tension de seuil des transistors.

1.1.3.4. Jonctions

Il s'agit des zones dopées du semi-conducteur qui forment les zones source et drain. Elles servent à conduire le courant du transistor depuis le canal jusqu'aux contacts métalliques.

Résistance d'accès

Les jonctions n'étant pas parfaitement métalliques, elles sont à l'origine de résistances qui viennent dégrader le fonctionnement du transistor (**figure 1-7**).

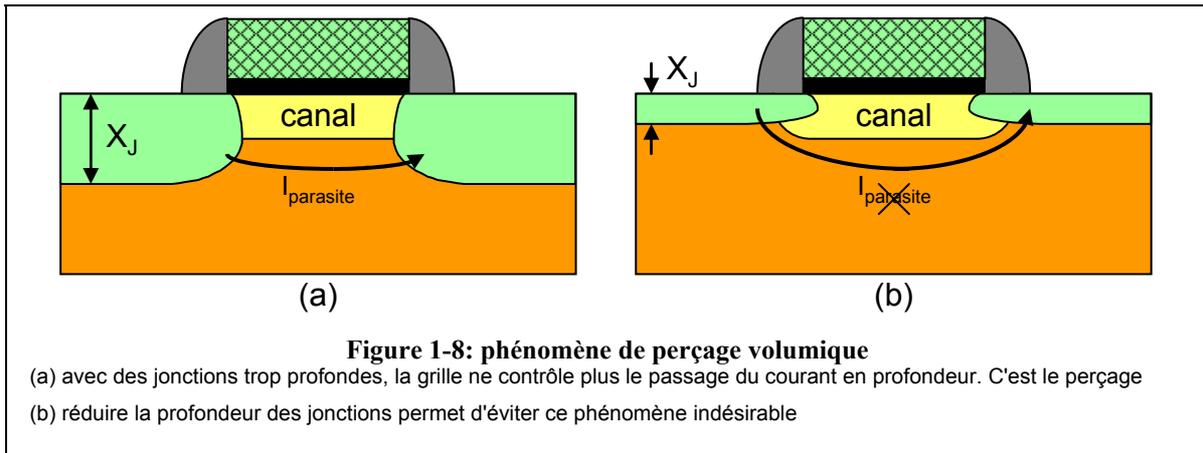


Profondeur de jonction

Les jonctions sont d'autant plus résistives qu'elles sont minces par simple effet géométrique:

$$R = \rho \cdot \frac{L}{S} \quad (1.1)$$

Mais des jonctions profondes posent un problème plus important: la grille ne contrôle les porteurs qu'en surface. Si les jonctions sont trop profondes, alors le courant peut circuler de la source au drain en profondeur, hors de contrôle de l'effet de champ induit par la grille (**figure 1-8a**). C'est le perçage volumique. Pour éviter ce phénomène indésirable, il faut donc réduire la **profondeur de jonction X_J** (**figure 1-8b**). D'autres solutions seront présentées plus loin dans ce chapitre.



Isolation

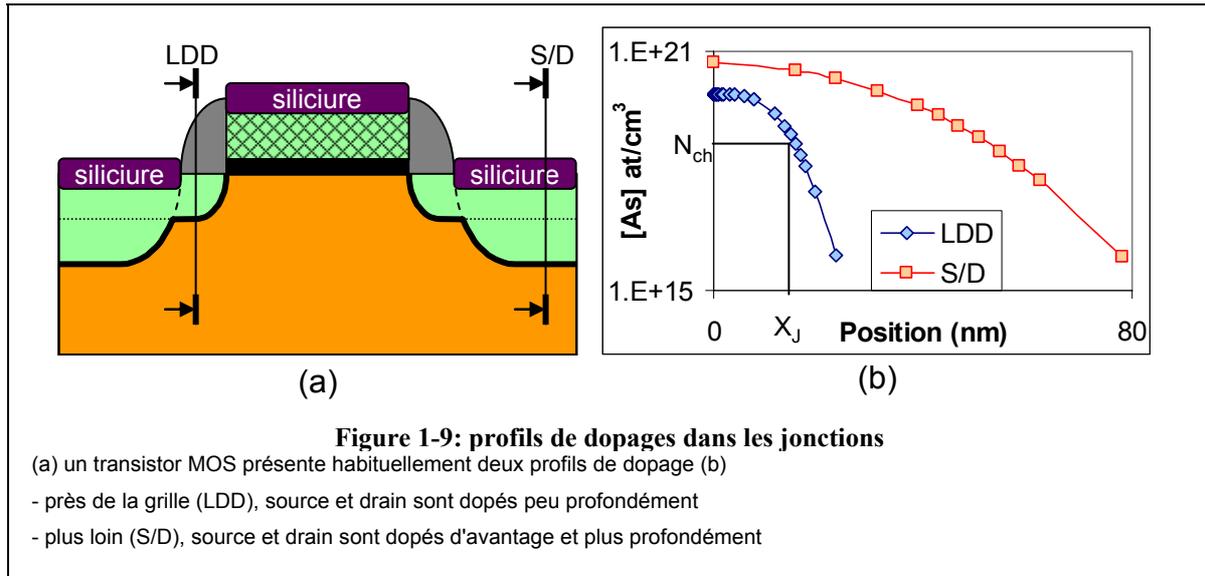
D'autre part, les jonctions étant des diodes, elles servent à isoler électriquement les électrodes source et drain du substrat. Les fuites de jonctions peuvent dégrader les performances, en particulier le courant I_{off} . Ce courant de fuite est du:

- à faible champ, aux fuites de jonction en inverse, liées à la génération de porteurs en zones de déplétion (effet *Shockley-Reed-Hall* ou *SRH*)
- en cas de forte polarisation, au courant tunnel assisté par défauts ou au courant tunnel direct

Architecture des jonctions

Pour optimiser tous ces phénomènes, les jonctions sont aujourd'hui réalisées en plusieurs parties (**figure 1-9**):

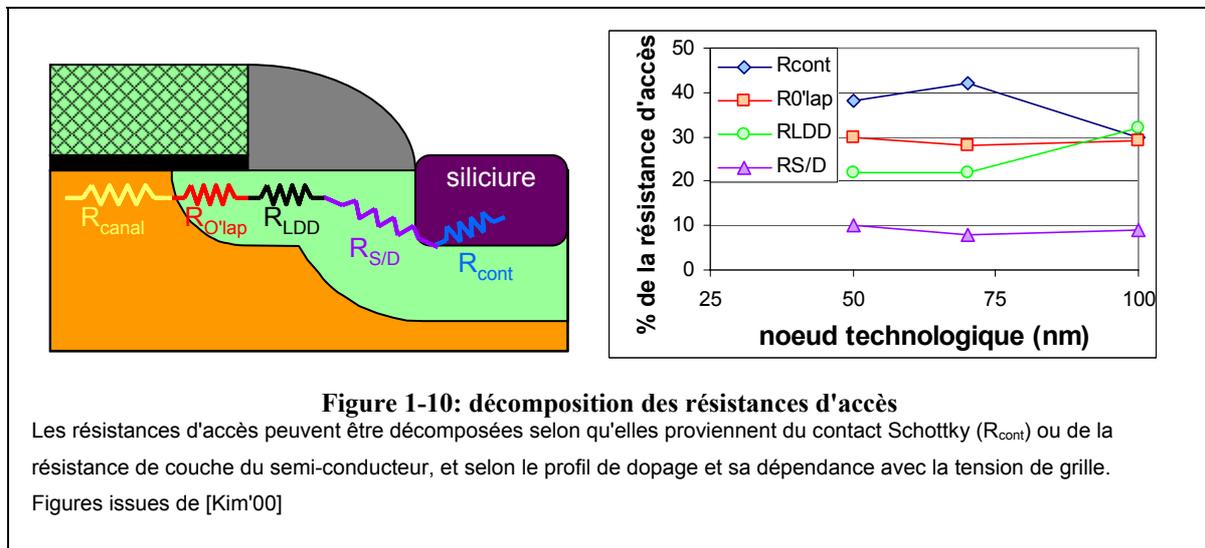
- les **LDD** (*Lightly Doped Drain*) sont peu profondes et localisées près du canal. Elles permettent d'éviter le perçage volumique
- les **source/drain** (S/D1) sont éloignées du canal. Plus profondes, elles permettent de diminuer la résistivité.
- les **source/drain 2** (S/D2) sont éloignées du canal. Plus profondes que les S/D1, mais au profil moins abrupt, elles servent à réduire les fuites de jonctions.
- pour limiter la résistance Schottky entre le contact métallique et le silicium dopé, les zones source/drain sont siliciurées. Les siliciures sont des alliages métalliques de silicium. Les deux alliages les plus utilisés sont le siliciure de cobalt $CoSi_2$ et le siliciure de nickel $NiSi$. Pour les réaliser, une couche de métal (cobalt ou nickel) est déposée sur la plaque. La réaction est activée thermiquement. Les couches les plus minces sont obtenues avec le nickel: de l'ordre de 30nm contre 50nm pour le cobalt.



Répartition des composantes des résistances d'accès

Les résistances d'accès peuvent être séparées en quatre composantes [Kim'00] (**figure 1-10**):

- la résistance de contact est due à la jonction Schottky entre le siliciure métallique et le semi-conducteur.
- les résistances source/drain et LDD sont dues à la résistance de feuille du semi-conducteur
- la résistance de recouvrement (*overlap*) diffère de la résistance de LDD par le fait qu'elle dépend de la tension de grille.



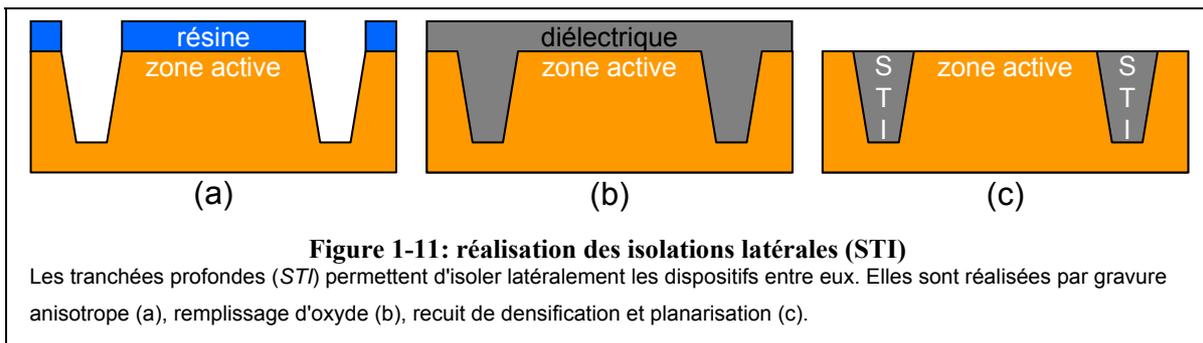
1.1.4. Principales étapes du procédé de fabrication du transistor MOS

Substrat

Le substrat est une plaque de silicium intrinsèque, c'est-à-dire non intentionnellement dopé. La densité résiduelle d'impuretés est de l'ordre de 10^{14}cm^{-3} . Ces plaques sont ensuite généralement dopées à 10^{15}cm^{-3} en bore en profondeur, et environ 10^{17}cm^{-3} en surface.

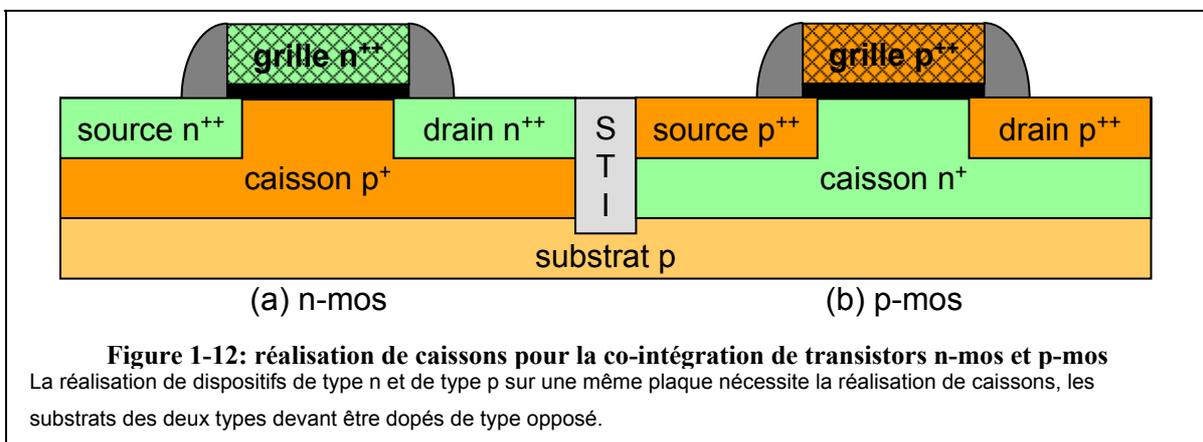
Définition des zones actives – isolation latérale

Dans les circuits, il est important d'isoler électriquement les n-mos des p-mos. L'isolation latérale se fait sous forme de tranchées peu profondes remplies d'oxyde, les **STI** (*Shallow Trench Isolation*). Les zones qui seront dites **zones actives** sont masquées par lithographie, puis les tranchées (**figure 1-11a**) sont creusées par gravure plasma. Une fois la résine retirée, un oxyde est déposé sur toute la plaque (**figure 1-11b**). Le dépôt doit remplir correctement les tranchées malgré un important facteur d'aspect: 300nm de profondeur pour une largeur pouvant descendre à 50nm. Enfin, le surplus d'oxyde est éliminé par polissage mécano-chimique (**figure 1-11c**) ou **CMP** (*Chemical Mechanical Polishing / Chemical Mechanical Planarization*) recuit de densification.



Implantation: caissons, dopage canal

Le dopage de type p du substrat est adapté aux transistors de type n. Pour implémenter des transistors de type p, il faut un substrat de type n. Celui-ci est réalisé par l'implantation d'un **caisson**. Un caisson symétrique est réalisé pour les transistors de type n lorsque le dopage du



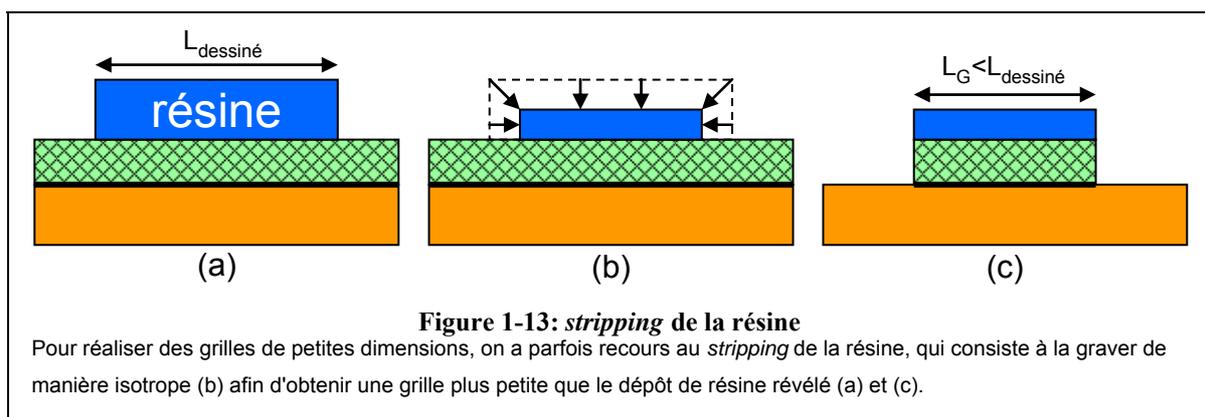
substrat n'est pas suffisant. En même temps sont faites les implantations du canal qui permettent de fixer la tension de seuil des dispositifs.

Empilement de grille

L'oxyde de grille est généralement réalisé par oxydation thermique sous atmosphère contrôlée. Puis le polysilicium de grille est déposé. Il est important de noter que ces deux étapes doivent être réalisées avec le plus court délai possible entre elles. En effet, en raison de sa petite dimension, l'oxyde de grille est très sensible aux impuretés [Bidaud'01]. Le polysilicium est ensuite dopé par implantation, en différenciant les zones p et n. Les grilles sont ensuite définies par lithographie et gravure anisotrope. Cette gravure est une étape délicate du procédé de fabrication. Elle est en effet soumise à de nombreuses contraintes technologiques:

- la longueur de grille est la plus petite dimension sur la plaque. C'est pourquoi elle est décrite comme **dimension critique ou CD** (*Critical Dimension*).
- les flancs doivent être verticaux
- le facteur d'aspect est important et variable. Le polysilicium de grille mesure de l'ordre de 120nm d'épaisseur, et les longueurs de grille font de 25nm à 10 μ m.
- la sélectivité de gravure doit être excellente: il faut graver les 120nm de polysilicium et s'arrêter sur un oxyde pouvant mesurer 1.2nm.

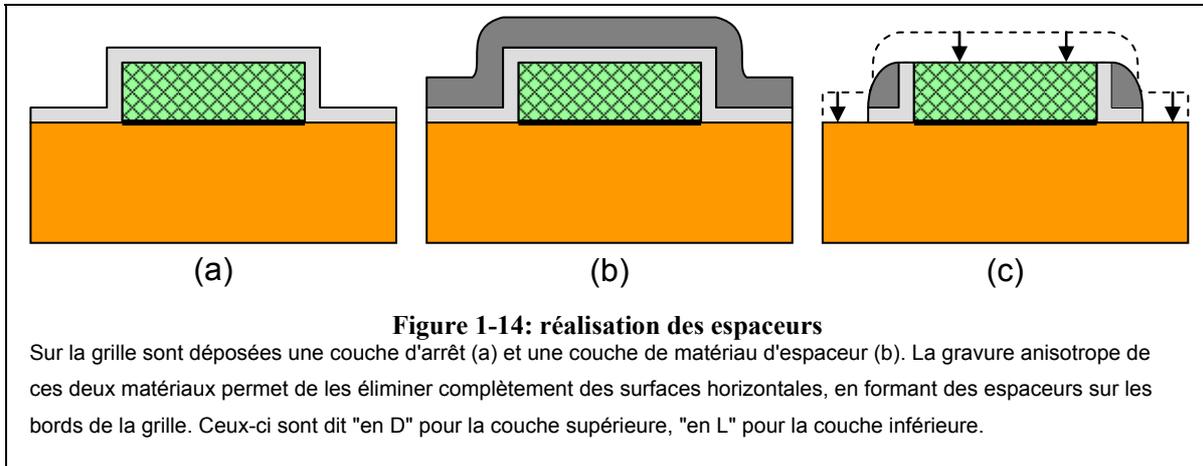
Il est parfois nécessaire de réaliser des grilles plus petites que le motif dessiné. Pour cela, on procède à une légère gravure isotrope de la résine. C'est le *stripping* (**figure 1-13**).



Réalisation des jonctions

En premier lieu, les extensions LDD sont implantées en utilisant la grille comme masque d'implantation. C'est pourquoi cette intégration est qualifiée d'**auto-alignée**. Pour écarter les zones S/D, des espaceurs sont réalisés: une couche d'arrêt (**figure 1-14a**) et le matériau

d'espaceur (**figure 1-14b**) sont successivement déposées. Il s'agit généralement de silice SiO_2 et de nitrure de silicium Si_3N_4 . Les dépôts doivent être conformes, c'est-à-dire que la topologie doit être respectée. Par conséquent, la hauteur de nitrure est plus importante sur les bords de la grille. Puis le tout est rabaissé par gravure anisotrope (**figure 1-14c**).

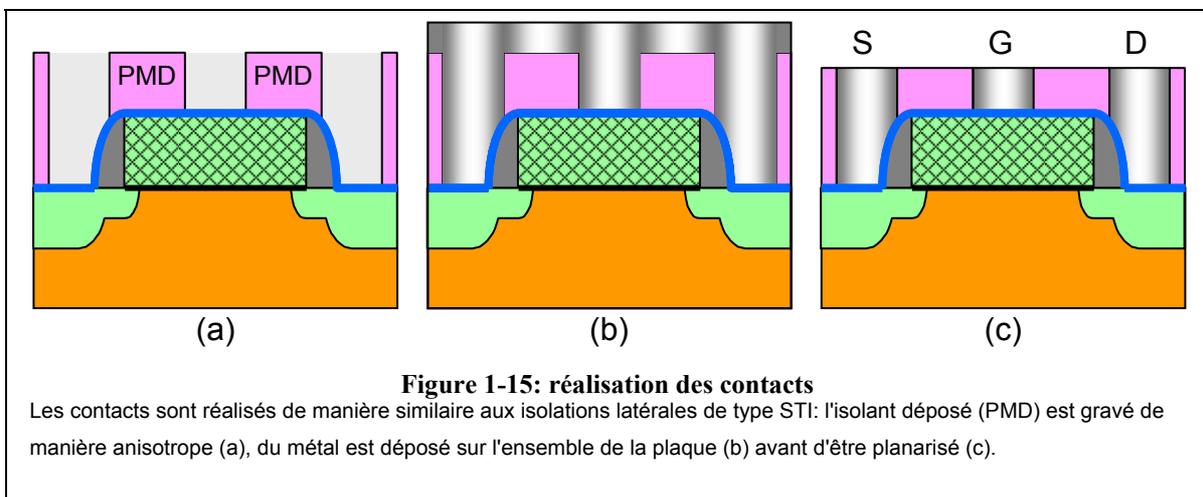


Les dopants implantés doivent être rendus électriquement actifs par recuit thermique. Celui-ci est court et intense, environ 1000°C quelques secondes. Enfin, les zones S/D sont siliciurées en même temps que le haut de la grille.

Contacts

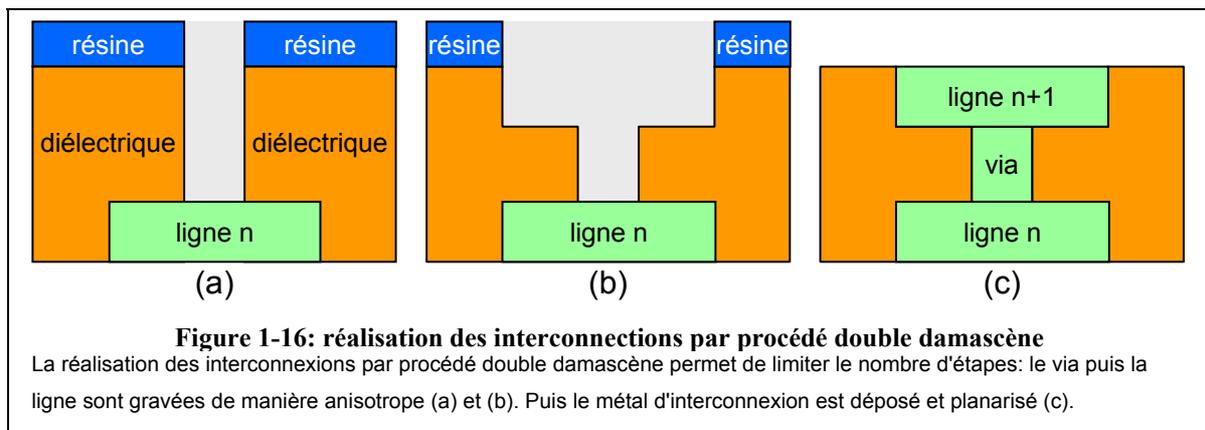
Une couche d'arrêt pour la gravure des contacts est déposée. Elle est appelée **CESL** pour *Contact Etch Stop Layer*. Par-dessus est déposée une épaisse couche de diélectrique, le **PMD** (*Pre Metal Dielectric*). Puis les trous de contact sont gravés (**figure 1-15a**). Dans les technologies avancées, il peut arriver que le contact débouche en partie sur l'espaceur.

Puis, de façon similaire aux STI, le métal de contact est déposé (**figure 1-15b**) puis planarisé (**figure 1-15c**). Les métaux utilisés pour les contacts sont le TiN et le tungstène.



Lignes de métal – interconnexions

Les interconnexions se composent de deux parties: les **vias** verticaux assurent la connexion entre deux niveaux de lignes. Les **lignes** horizontales assurent les connexions d'une partie à l'autre de la puce. Niveau après niveau, vias et lignes sont réalisés simultanément en technologie double damascène: dépôt du diélectrique, photolithographie puis gravure via (**figure 1-16a**). Puis la ligne est définie par photolithographie avant d'être gravé (**figure 1-16b**). Enfin, une fois la résine retirée, la cavité ainsi créée est remplie par du métal qui est ensuite planarisé (**figure 1-16c**).



Depuis le noeud 120nm, l'aluminium a été remplacé par le cuivre comme métal d'interconnexion. Le diélectrique entre les lignes doit avoir la permittivité la plus faible possible, pour limiter les capacités parasites inter-lignes. C'est pourquoi des matériaux poreux comme certains SiOC sont intégrés.

1.2. Éléments de modélisation

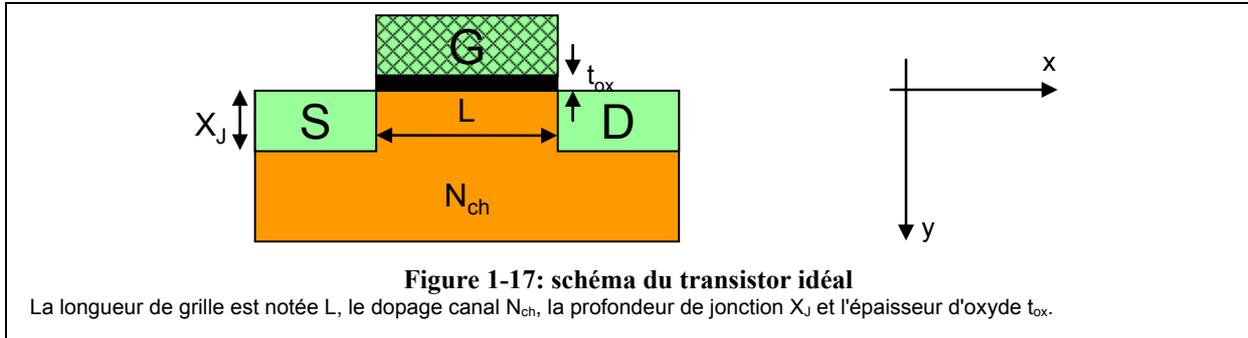
Les notions fondamentales nécessaires à la modélisation du transistor MOS sont introduites, telles qu'utilisées dans le modèle MASTAR ([Skotnicki'88], [Skotnicki'03]). Il est important de souligner que seuls les effets du premier ordre sont considérés ici. En particulier, la longueur de grille est supposée supérieure à $1\mu\text{m}$ (transistor "à canal long").

L'axe horizontal (source => drain) est l'axe x, et y celui de la profondeur. Les grandeurs suivantes sont utilisées:

- **L**: longueur de grille.
- **t_{ox}** : épaisseur de l'oxyde de grille (*t* pour *thickness*)
- **N_{ch}** : dopage canal (*ch* pour *channel*)
- **X_j** : profondeur des jonctions

- $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$: capacité d'oxyde par unité de surface

La grille, bien que n'étant pas rigoureusement métallique, sera considérée comme telle.

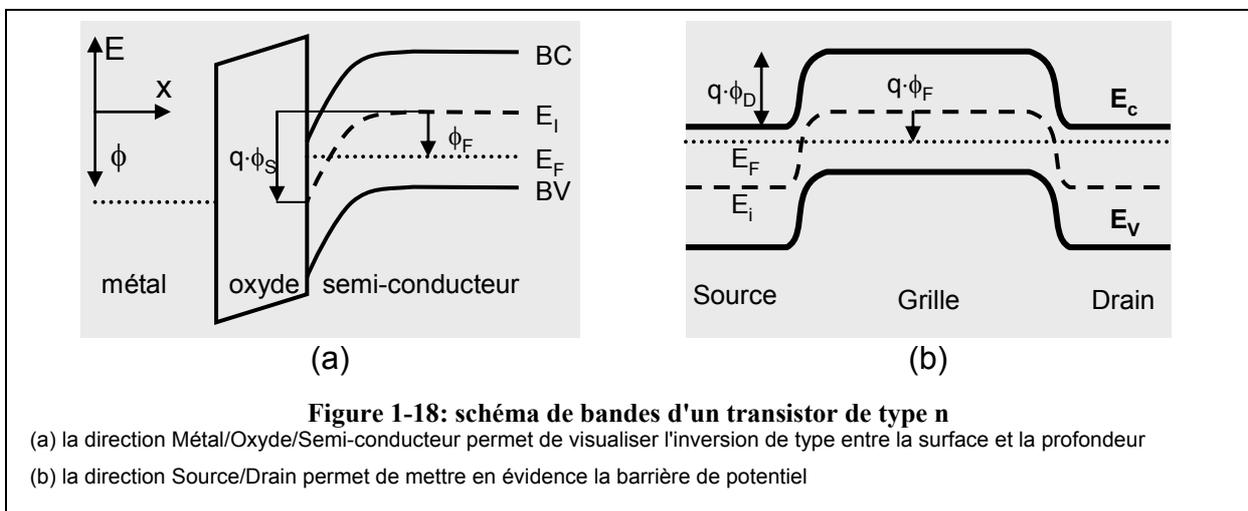


1.2.1. Potentiel de Fermi, potentiel de surface et courbure de bandes

La **figure 1-18** représente le schéma de bandes de la structure MOS en régime d'inversion. La position du niveau de Fermi est donnée par le dopage:

$$\phi_F = \frac{k_b \cdot T}{q} \cdot \ln\left(\frac{N_{ch}}{n_i}\right) \quad (1.2)$$

Le potentiel de surface ϕ_S est la déviation de potentiel à l'interface oxyde/semi-conducteur ($x=0$) par rapport à la profondeur ($x=\infty$)



La barrière de potentiel entre la source et le canal en l'absence de polarisation est:

$$\phi_D = \frac{k \cdot T}{q} \cdot \ln\left(\frac{N_D \cdot N_{ch}}{n_i^2}\right) \quad (1.3)$$

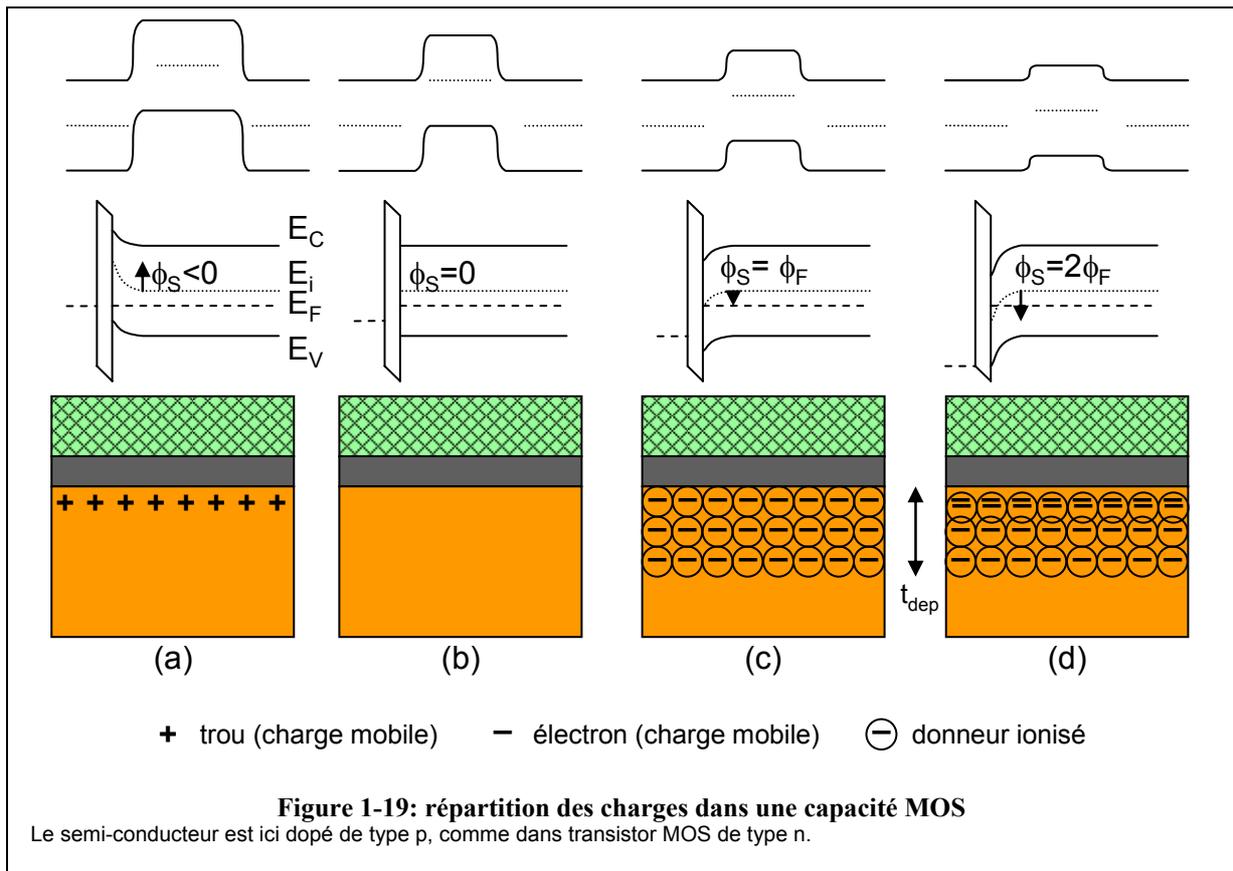
Sous l'effet de la polarisation de grille, le potentiel vient se courber selon l'axe y. Quatre situations (**figure 1-19**) peuvent se présenter:

(a) $\phi_S < 0$: régime d'accumulation: il y a d'avantage de trous en surface qu'en profondeur.

(b) $\phi_S=0$: tension de bandes plates: c'est la tension de grille pour laquelle la courbure de bande est nulle dans tout le semi-conducteur.

(c) $0 < \phi_S < \phi_F$: déplétion et inversion faible: les trous sont repoussés de la surface, laissant les donneurs ionisés.

(d) $\phi_F < \phi_S$: inversion: les porteurs majoritaires sont maintenant les électrons, comme dans un semi-conducteur de type n.



1.2.2. Tension de seuil V_{th}

La tension de seuil V_{th} (*THreshold Voltage*) est la tension de grille nécessaire à la suppression de la barrière de potentiel. Il faut donc amener la structure en régime de bandes plates, puis augmenter la tension de grille jusqu'à obtenir $\phi_S = \phi_D$. Cet incrément de tension peut être calculé à l'aide du schéma capacitif équivalent de la structure MOS [Skotnicki] (**figure 1-20a**) où:

- C_{dep} est la capacité associée aux charges de déplétion.
- C_{ss} est la capacité associée aux charges d'interface (*Surface State*).
- C_{inv} est la capacité associée aux charges d'inversion.

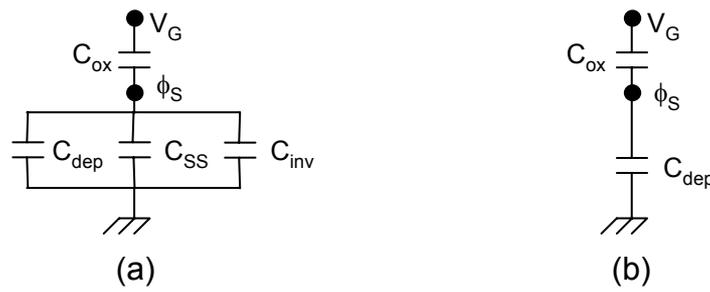


Figure 1-20: schémas capacitifs équivalents d'une structure MOS

Le schéma (a) est complet. Il peut être réduit au schéma (b) sous deux conditions:

- dans une technologie bien maîtrisée, les charges de surfaces et la capacité associée C_{SS} peuvent être négligées.
- en régime de déplétion et d'inversion faible, les charges d'inversion et la capacité associée peuvent être négligées.

Dans une technologie bien maîtrisée, les charges de surface peuvent être négligées, et par conséquent la capacité C_{SS} qui leur est associée. De plus, lorsque $V_G \sim V_{th}$, la charge d'inversion est négligeable par rapport à la charge de déplétion. L'ensemble des capacités peut donc être réduit à la seule capacité de déplétion C_{dep} (**figure 1-20b**).

La tension de seuil est donc la somme de la tension de bandes plates V_{FB} , de la tension correspondant à la hauteur de barrière ϕ_D , et de la chute de potentiel dans l'oxyde Q_{dep}/C_{ox} .

$$V_{th} = V_{FB} + \phi_D + \frac{Q_{dep}}{C_{ox}} \quad (1.4)$$

La charge de déplétion est simplement exprimée à partir du dopage canal N_{ch} et de la profondeur de déplétion t_{dep} :

$$Q_{dep} = t_{dep} \cdot N_{ch} \quad (1.5)$$

La profondeur de déplétion est calculée en intégrant deux fois l'équation de Poisson:

$$\frac{d^2\phi}{dy^2} = -\frac{q \cdot N_{ch}}{\epsilon_{Si}} \quad (1.6)$$

et en appliquant les conditions limites:

$$\begin{aligned} E(x = t_{dep}) &= -\frac{d\phi}{dy} = 0 \\ \phi(x = t_{dep}) &= V_{BS} \\ \phi(x = 0) &= \phi_S \end{aligned} \quad (1.7)$$

ce qui donne:

$$t_{dep} = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_{ch}} (\phi_S - V_{BS})} \quad (1.8)$$

En combinant les équations (1.4), (1.5) et (1.8):

$$V_{th} = V_{FB} + \phi_D + \frac{1}{C_{ox}} \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{ch} \cdot (\phi_D - V_{BS})} \quad (1.9)$$

La tension de seuil a été définie à l'aide de la relation $\phi_S = \phi_D$. Cependant, il existe d'autres moyens de définir la tension de seuil. Notamment, il est possible de décrire le seuil comme étant le moment où la densité de charges d'inversion est égale à la densité de charges libres en régime de bandes plates. La relation $\phi_S = \phi_D$ devient alors $\phi_S = 2\phi_F$. Pour une architecture de type substrat massif (*bulk*), où le canal est fortement dopé, la différence est minime. Mais la définition par $\phi_S = \phi_D$ est plus appropriée pour des canaux faiblement dopés.

1.2.3. Courant de conduction I_{on}

Une fois la barrière de potentiel supprimée, la grille permet d'accumuler des porteurs dans le canal. Il devient possible de mettre en mouvement cette **charge d'inversion** à l'aide de la tension de drain V_{DS} . Le courant de conduction est le quotient de la charge d'inversion par le temps t mis par les porteurs pour aller de la source au drain:

$$I_{DS} = \frac{Q_{inv}}{t} \quad (1.10)$$

Ce temps dépend de la vitesse v des porteurs:

$$t = \frac{L_G}{v} \quad (1.11)$$

Celle-ci dépend de la mobilité μ des porteurs et du champ électrique source/drain:

$$v = \mu \cdot E \quad (1.12)$$

Le champ électrique étant simplement:

$$E = \frac{V_{DS}}{L_G} \quad (1.13)$$

En groupant ces quatre équations, on obtient:

$$I_{DS} = \frac{Q_{inv} \cdot \mu \cdot V_{DS}}{L_G^2} \quad (1.14)$$

La mobilité des porteurs est connue [Sze'81]. Pour exprimer le courant, il faut donc exprimer la charge d'inversion. Il est important de souligner que la grille est équipotentielle alors que le potentiel du canal varie entre source et drain. Cela implique que la chute de potentiel à travers la capacité C_{ox} varie également. Il s'ensuit nécessairement que la charge d'inversion n'est pas constante le long du canal. Notamment, elle vaut à la source:

$$Q_{inv,source} = C_{ox} \cdot (V_G - V_{th}) \quad (1.15)$$

alors qu'au drain elle devient:

$$Q_{inv,drain} = C_{ox} \cdot (V_G - V_{th} - V_{DS}) \quad (1.16)$$

En première approximation, on peut considérer que la charge varie linéairement de la source au drain. Par conséquent, la charge d'inversion vaut:

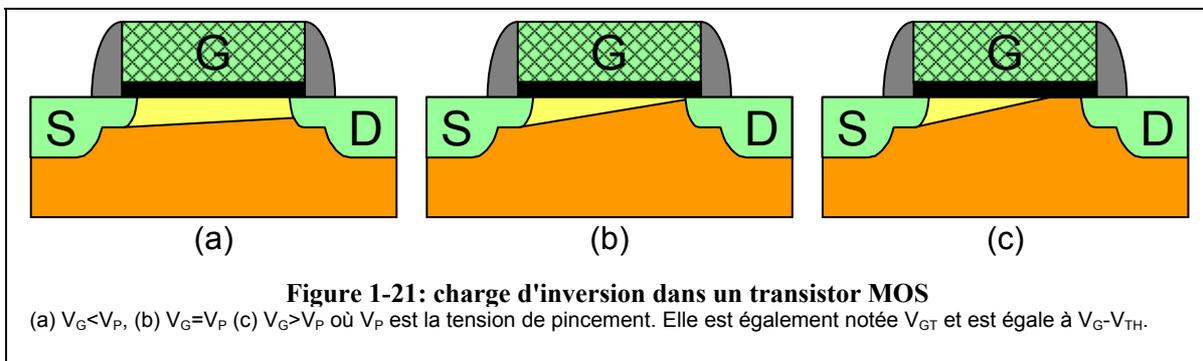
$$Q_{inv} = \left(\frac{Q_{inv,drain} + Q_{inv,source}}{2} \right) \cdot W \cdot L = W \cdot L \cdot C_{ox} \cdot \left(V_G - V_{th} - \frac{V_{DS}}{2} \right) \quad (1.17)$$

Note: les quantités $Q_{inv,source}$ et $Q_{inv,drain}$ sont des charges surfaciques, alors que la quantité Q_{inv} est une charge. Il s'ensuit l'expression du courant en régime linéaire:

$$I_{DS} = \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot \left(V_G - V_{th} - \frac{V_{DS}}{2} \right) \cdot V_{DS} \quad (1.18)$$

Cette expression n'est valable que lorsque la tension de grille est suffisamment élevée pour obtenir une inversion tout le long du canal. Ceci n'est pas vérifié pour une tension de drain supérieure à $V_G - V_{th}$. De fait, il importe de décrire l'évolution de la charge d'inversion avec le courant de drain.

A tension de drain faible, l'ensemble du canal est inversé (**figure 1-21a**). La relation (1.18) est donc valable. La charge d'inversion devient nulle au drain lorsque la chute de potentiel s'annule au drain (**figure 1-21b**), pour $V_{DS} = V_G - V_{th}$. Cette tension est appelée tension de pincement et est notée V_{GT} ou V_P .



Toute augmentation de la tension de drain au dessus de la tension de pincement ne crée plus de charge d'inversion. Le point de pincement (point où la charge d'inversion devient nulle) se déplace vers la source (**figure 1-21c**). La charge d'inversion n'évoluant plus, le courant sature. La valeur du courant de saturation est obtenue en donnant la valeur V_{GT} à la tension V_{DS} dans la relation (1.18):

$$I_{DSat} = \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot \frac{V_{GT}^2}{2} \quad (1.19)$$

En régime de fonctionnement normal, le drain et la grille sont polarisés à la tension d'alimentation V_{DD} . Par conséquent, c'est la relation (1.19) qui est utilisée pour le calcul du courant de conduction:

$$I_{on} = \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot \frac{V_{DD}^2}{2} \quad (1.20)$$

1.2.4. Pente sous le seuil S

Elle traduit le comportement du courant en régime sous le seuil, c'est-à-dire pour des tensions de grille inférieures à la tension de seuil. Les mesures montrent que le courant sous le seuil varie exponentiellement avec la tension de grille. La pente sous le seuil est donc définie comme étant l'incrément de tension de grille nécessaire à la variation de courant de drain d'une décade:

$$S = \frac{\partial V_G}{\partial \log(I_{DS})} \quad (1.21)$$

C'est l'inverse de la pente mathématique sur une caractéristique $\log(I_D)(V_G)$. Cette expression peut s'écrire sous la forme:

$$S = \frac{1}{\frac{\partial \log(I_{DS})}{\partial V_G}} \quad (1.22)$$

La charge d'inversion obéit à une loi de Fermi-Dirac. En régime sous le seuil, elle peut être approximée par une loi de Boltzmann:

$$Q_{inv}(V_G) = Q_0 \cdot \exp\left(\frac{k_B \cdot T}{q \cdot \phi_s(V_G)}\right) \quad (1.23)$$

L'hypothèse classique illustrée par la l'équation (1.10) est reprise pour le calcul du courant:

$$I_{DS,ss}(V_G) = \frac{Q_0}{t} \cdot \exp\left(\frac{k_B \cdot T}{q \cdot \phi_s(V_G)}\right) \quad (1.24)$$

La variation de mobilité avec la tension de grille est négligeable devant la variation de charge d'inversion avec la tension de grille [Sze'81]:

$$Q_{inv} \cdot \frac{d\mu}{dV_G} \ll \mu \cdot \frac{dQ_{inv}}{dV_G} \quad (1.25)$$

Le temps t étant indépendant des polarisations (équations (1.11) à (1.13)). Nous pouvons donc injecter la relation (1.24) dans la relation (1.22):

$$S = \frac{k_B \cdot T}{q} \ln(10) \frac{\partial V_G}{\partial \phi_s} \quad (1.26)$$

La quantité $\frac{\partial \phi_s}{\partial V_G}$ représente le couplage entre la tension de grille et le potentiel de surface.

Elle est égale à l'unité pour un couplage parfait, inférieure sinon. Par conséquent, la pente sous le seuil est limitée:

$$S \geq 60 \text{mV/dec} @ 300\text{K} \quad (1.27)$$

Dans une technologie bien maîtrisée, $S=80\text{mV/dec}$ [Skotnicki'00]. En technologie SOI, des valeurs de $S=65\text{mV/dec}$ ont été mesurées, soit presque la valeur idéale.

1.2.5. Courant de fuite I_{off}

Le courant varie exponentiellement avec la tension de grille entre 0V et V_{th} . Par conséquent, le courant de fuite peut être calculé par:

$$\log(I_{\text{off}}) = \log(I_{\text{th}}) - \frac{V_{\text{th}}}{S} \quad (1.28)$$

où I_{th} est l'extrapolation du courant sous le seuil à $V_G=V_{\text{th}}$. Il peut s'exprimer sous la forme [Mastar'04]:

$$I_{\text{th}} = 5 \cdot 10^{-7} [\text{A}] \frac{W}{L} 8 \cdot 10^8 N_{\text{ch}}^{0.4865} [\text{cm}^{-3}] \quad (1.29)$$

Cependant, le courant I_{off} peut être dégradé par le courant tunnel à travers l'oxyde de grille entre la grille et le drain. Ce sont les "fuites de grille".

1.3. Réduction des dimensions: les effets canaux courts

1.3.1. Noeuds technologiques

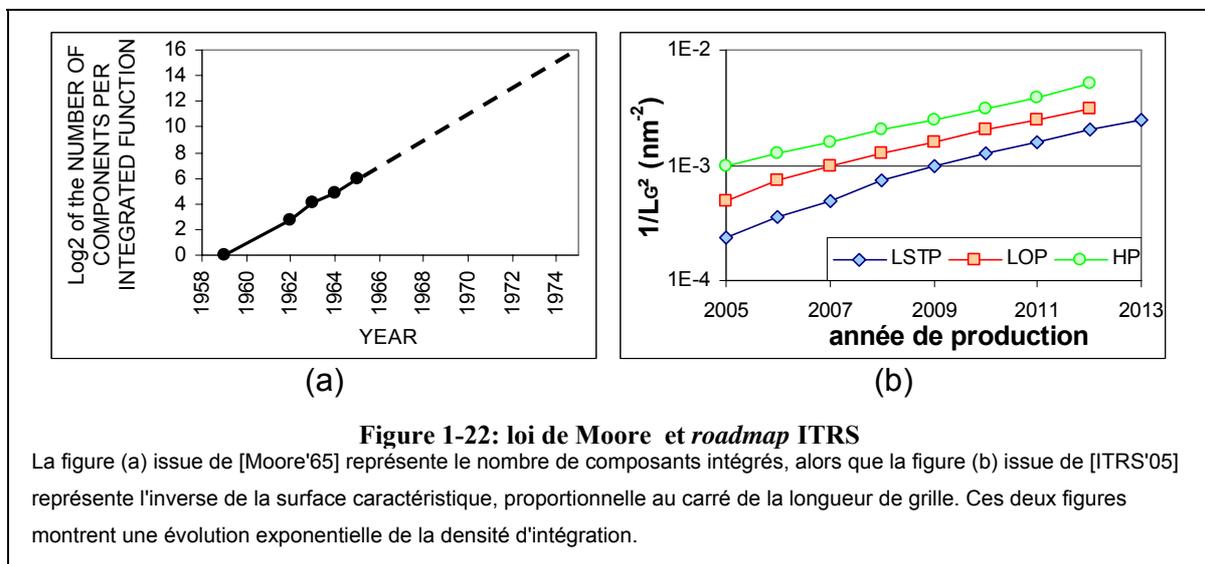
L'équation (1.20) fournit une information fondamentale: le courant de conduction est inversement proportionnel à la longueur de grille. Par conséquent, pour améliorer les performances, il "suffit" de diminuer la longueur de grille. Cette réduction des dimensions a longtemps été le moteur principal de l'évolution des transistors MOS. D'autre part, la réduction des dimensions permet d'augmenter la densité d'intégration, et donc d'augmenter le nombre de puces par plaque, ce qui présente un bénéfice économique évident.

Dans un article désormais célèbre, Gordon Moore fait remarquer que la densité d'intégration augmente régulièrement avec le temps [Moore'65]. Cet article a donné naissance à la fameuse

loi de Moore, selon laquelle la puissance des processeurs à l'instar de la densité d'intégration double tous les dix-huit mois.

Voici posée la règle qui a gouverné le monde du semi-conducteur, bien qu'à présent la tendance aille plutôt à la parallélisation (multiplication des coeurs). Afin de conserver cette vitesse d'évolution, les transistors sont regroupés par générations technologiques. A chaque génération correspond une longueur de grille nominale et des performances à atteindre. La génération porte le nom du demi-*pitch* de la DRAM, motif répétitif d'une cellule mémoire élémentaire.

Le descriptif des générations à venir, ainsi que les moyens technologiques mis oeuvre pour y parvenir, sont publiés tous les deux ans sous forme de *roadmaps* par le comité ITRS (*International Technology Roadmap for Semiconductors*) [ITRS'05].



Cependant, courants de fuites et courant de conduction résultent de contraintes technologiques opposées. Il convient donc de trouver un bon compromis I_{on}/I_{off} . Ce compromis dépend de l'application commerciale à laquelle est destinée le transistor. Augmenter I_{on} revient à améliorer la vitesse, et donc la puissance de calcul, alors que diminuer I_{off} revient à limiter la consommation, c'est-à-dire augmenter l'autonomie de dispositifs embarqués. Pour répondre à cette logique de différenciation, ce sont donc trois *roadmaps* distinctes qui sont proposées par l'ITRS (**figure 1-22b**):

- *High Performance* (HP): haute vitesse. Il s'agit typiquement des processeurs et de mémoires rapides.
- *Low Operating Power* (LOP): basse consommation en fonctionnement. Ce sont les produits embarqués qui tournent de façon continue. Les baladeurs MP3 en sont un bon exemple.

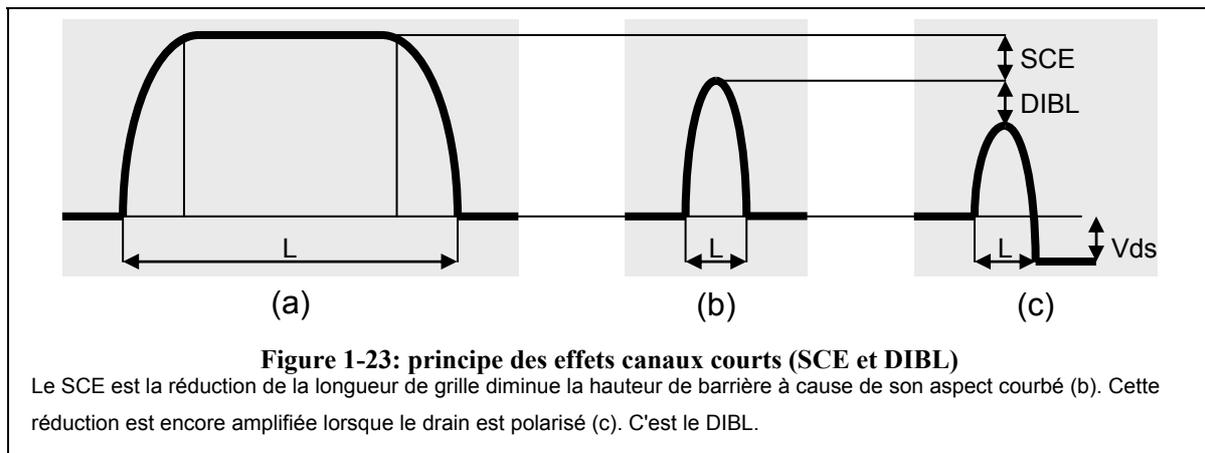
- *Low Standby Power* (LSTP): faible consommation en veille. Le produit phare de cette catégorie est le téléphone portable.

1.3.2. Les effets canaux courts

Lorsque les dimensions réduisent, de nouveaux phénomènes physiques viennent perturber le comportement du transistor. Notamment, les **effets canaux courts** apparaissent lorsque la longueur de grille est réduite.

Les jonctions ne sont pas infiniment abruptes. La barrière de potentiel passe progressivement de l'état haut vers l'état bas (**figure 1-23a**). Lorsque la longueur de grille diminue, ce passage progressif donne naissance à un abaissement de la barrière (**figure 1-23b**): c'est le *Short Channel Effect* (SCE). Cet effet est encore accentué lorsque le drain est polarisé (**figure 1-23c**): c'est le *Drain Induced Barrier Lowering* (DIBL).

Cet abaissement se traduit par une chute de la tension de seuil lorsque la longueur de grille diminue (**figure 1-24a**). Sur la caractéristique $I_D(V_G)$, cela se traduit par un décalage sur l'axe des tensions (**figure 1-24b**).



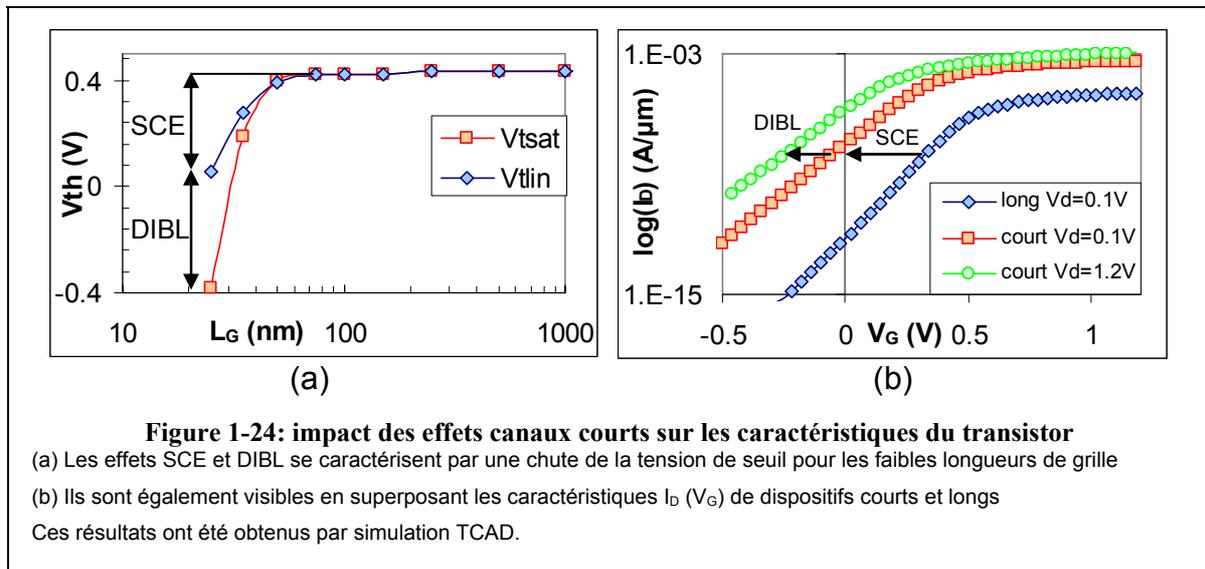
Ces effets sont quantifiés par:

$$\begin{aligned} \text{SCE}(L) &= V_{\text{th,lin}}(L = \infty) - V_{\text{th,lin}}(L) \\ \text{DIBL}(L) &= V_{\text{th,lin}}(L) - V_{\text{th,sat}}(L) \end{aligned} \quad (1.30)$$

Il est possible [Skotnicki'88] de modéliser ces quantités:

$$\begin{aligned} \text{SCE}(L) &= 0.64 \cdot \frac{\epsilon_{\text{Si}}}{\epsilon_{\text{ox}}} \cdot \left(1 + \frac{X_J^2}{L_G^2}\right) \cdot \frac{t_{\text{ox_el}}}{L_G} \cdot \frac{t_{\text{dep}}}{L_G} \cdot \phi_d \\ \text{DIBL}(L) &= 0.8 \cdot \frac{\epsilon_{\text{Si}}}{\epsilon_{\text{ox}}} \cdot \left(1 + \frac{X_J^2}{L_G^2}\right) \cdot \frac{t_{\text{ox_el}}}{L_G} \cdot \frac{t_{\text{dep}}}{L_G} \cdot V_{\text{DS}} \end{aligned} \quad (1.31)$$

où X_J est la profondeur de jonction et t_{ox_el} l'épaisseur de diélectrique. Ces expressions mettent en évidence les moyens technologiques disponibles pour lutter contre les effets canaux courts.



1.3.3. Empilement de grille

La quantité t_{ox_el} est l'épaisseur équivalente en silice, ou **EOT** (*Equivalent Oxide Thickness*). Elle correspond à l'épaisseur qu'aurait la couche de SiO_2 pour former une capacité de même valeur. Cette capacité provient de l'isolant employé, ainsi que de la poly-déplétion et des effets quantiques qui apparaissent dans le canal. En effet, les porteurs sont délocalisés: leur barycentre s'éloigne de l'interface. C'est le **dark space** (à l'interface oxyde/semi-conducteur) et la **poly déplétion** (à l'interface oxyde/polysilicium de grille). L'expression de t_{ox_el} devient donc:

$$t_{ox_el} = t_{phys} \cdot \frac{\epsilon_{SiO_2}}{\epsilon_{isolant}} + D_{ark_space} + P_{oly_depl} \quad (1.32)$$

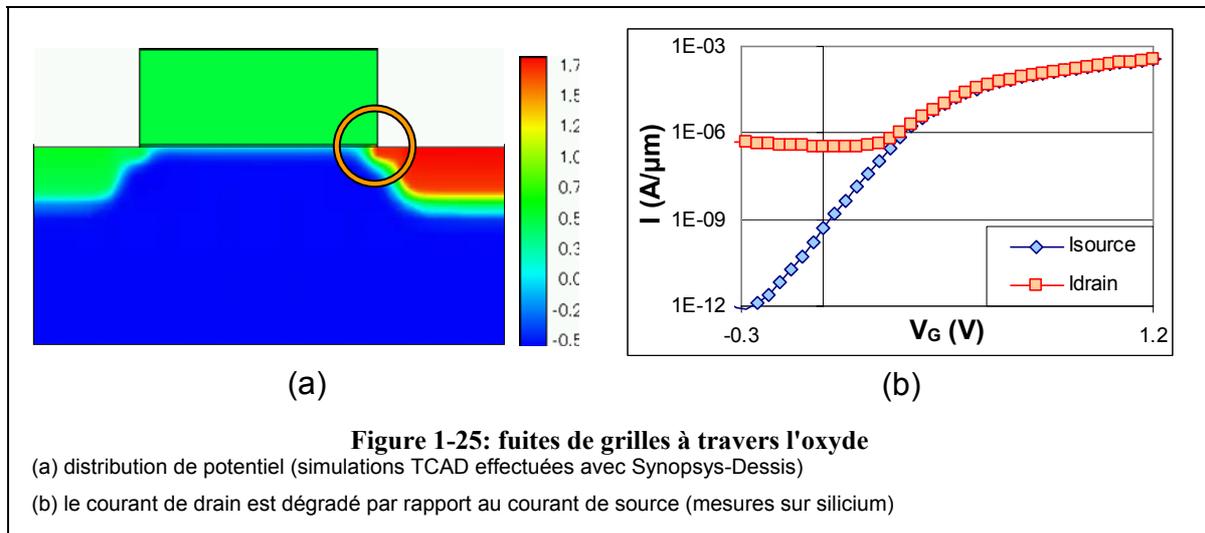
Le *dark space* peut difficilement être supprimé. Par contre, il est possible de:

- diminuer l'épaisseur physique d'oxyde
- changer de matériau pour une plus grande permittivité
- intégrer une grille métallique pour supprimer la poly-déplétion

Diminution de l'épaisseur d'isolant

Jusqu'à présent, la réduction de l'épaisseur d'oxyde était efficace. Cependant, les oxydes les plus minces ($\sim 12\text{\AA}$) ne sont plus totalement isolants. A ces dimensions, les porteurs peuvent passer à travers l'oxyde par effet tunnel. La fuite de grille se fait là où la différence de potentiel est la plus importante, c'est-à-dire entre le drain et la grille à faible tension de grille

et forte tension de drain (**figure 1-25a**). Ces fuites peuvent fortement dégrader le courant I_{off} du transistor (**figure 1-25b**).



Matériau d'isolant de grille

Il s'agit historiquement de silice SiO_2 , qui présente de nombreux avantages:

- intégration aisée
- compatibilité exemplaire avec le silicium
- stabilité (contrairement à l'oxyde de germanium par exemple)
- possibilité d'obtenir des oxydes très minces ($\sim 12\text{Å}$) de bonne qualité

La recherche d'un nouveau matériau de grille est motivée par l'amélioration de la constante

diélectrique de ce matériau. Cela permet d'augmenter la capacité de grille $C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}}$ et donc

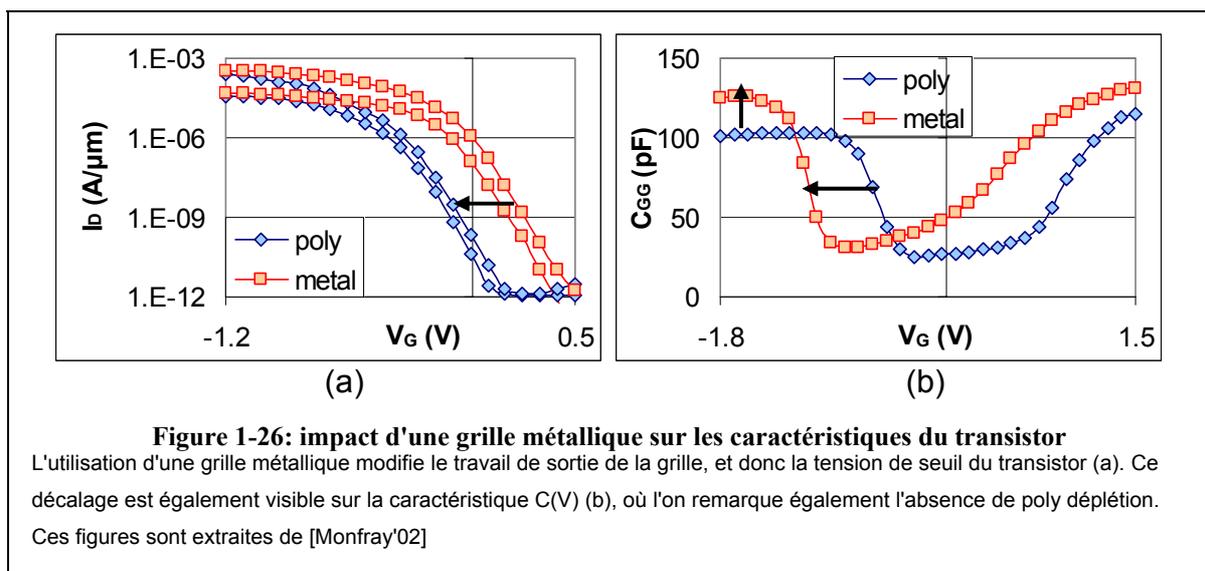
d'améliorer les performances, en évitant d'utiliser des épaisseurs d'oxyde trop minces conduisant à des fuites de grille par effet tunnel. Une solution simple consiste à incorporer de l'azote dans la silice. Cela permet d'élever la permittivité relative de l'oxyde de 3.9 à 4.2 environ.

La solution de rupture consiste à remplacer le SiO_2 par un matériau à haute permittivité, dit high- κ . Cependant, l'intégration de ces nouveaux matériaux reste difficile. Les matériaux actuellement étudiés sont le dioxyde d'hafnium HfO_2 et ses dérivés, les silicates d'hafnium HfSiO_x éventuellement nitrurés. Signalons aussi d'autres études sur des alliages de zirconium.

Matériau de grille

Le matériau de grille utilisé est du polysilicium dopé jusqu'à dégénérescence. N'étant pas rigoureusement métallique, une zone de déplétion vient se créer à l'interface grille/oxyde. Cette zone de l'ordre de 3-4 Å joue le rôle d'isolant électrique et est à rajouter à l'épaisseur de l'oxyde de grille. Ceci a donc pour effet de diminuer le couplage entre la grille et le canal.

Pour pallier ce problème, le polysilicium de grille doit être remplacé par un métal. Pour reproduire les travaux de sortie différents du silicium dopé p et du silicium dopé n, ce sont donc deux métaux dits de type n+ (*n+ like*) et de type p+ (*p+ like*) qui doivent être intégrés. En premier lieu, il est plus aisé de commencer par intégrer un unique matériau *midgap*, dont le travail de sortie est intermédiaire. Par conséquent, cela se traduit par un décalage de tension de seuil sur la caractéristique $I_D(V_G)$ (figure 1-26a) ainsi que sur la caractéristique $C(V)$ (figure 1-26b). Sur cette même figure, on peut noter l'absence de poly-déplétion. Il est intéressant de noter que les architectures à film mince n'étant pas ou peu dopées, le travail de sortie d'un matériau *midgap* est parfaitement adapté.



La grille métallique pose des difficultés d'intégration accrues par rapport que la grille en polysilicium, notamment pour la gravure de la grille. C'est pourquoi d'autres méthodes sont explorées:

- la siliciuration totale de la grille: la grille est réalisée en polysilicium dopé, puis siliciurée comme les source/drain. La différence tient au fait qu'il faut siliciurer la grille jusqu'à l'oxyde, soit 120nm, alors qu'on cherche à réaliser des siliciures fins aux jonctions. On parlera de grille ToSi (*Totally Silicied*) ou FuSi (*Fully Silicided*) ([Muller'05] et références incluses).

- le procédé PRETECH [Harrison'04]: la grille est réalisée en polysilicium, puis elle est gravée par voie humide en passant par les trous de contacts. La cavité est ensuite remplie par du métal. L'isolant de grille peut également être remplacé de cette manière.

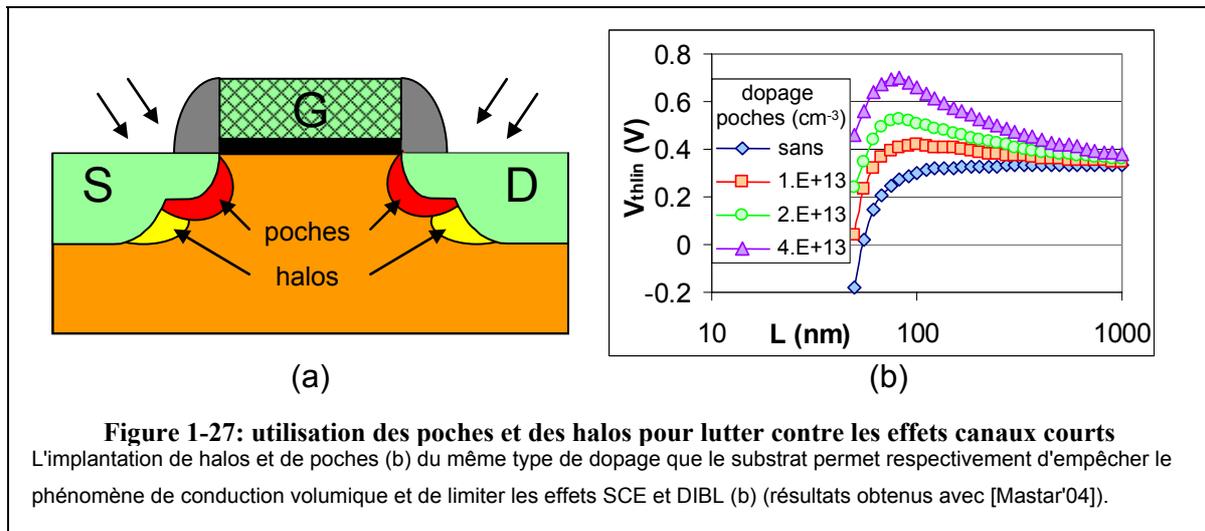
1.3.4. Optimisation du dopage canal

Les effets SCE et DIBL sont proportionnels à la profondeur de déplétion t_{dep} , qui est directement dépendante du dopage canal (équation (1.8)):

$$t_{dep} = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_{ch}} (\phi_s - V_{BS})}$$

Par conséquent, un moyen simple de lutter contre les effets canaux courts est d'augmenter le dopage canal. L'inconvénient majeur est que cela décale la tension de seuil. L'idéal serait de pouvoir doper le canal d'avantage pour les transistors courts que pour les longs.

C'est pourquoi une implantation inclinée (*tiltée*) est ajoutée. Cela permet d'augmenter le dopage canal sur les bords de grilles (**poches**). Plus profonds que les poches, plus dopés, les **halos** viennent empêcher le phénomène de conduction volumique.



1.3.5. Jonctions

Pour limiter les effets canaux courts, les expressions (1.31) montrent qu'il faut réduire la profondeur de jonction. Pour cela, le moyen le plus élémentaire est de limiter la diffusion des dopants. De nombreux moyens technologiques sont mis en oeuvre dans cette optique:

- abandon des procédés à budget thermique élevé (haute température / longue durée)
- utilisation de procédés "froids": les nitrures utilisés pour les espaceurs ne sont plus déposés thermiquement mais avec un procédé de CVD.

- utilisation de recuits d'activation rapides (type *spike* ou recuit laser)
- les implantations sont modifiées:
 - implantation à faible énergie (ULE pour *Ultra Low Energy*) ou plasma
 - préamorphisation pour éviter les effets de canalisation (*channeling*)

1.4. Amélioration des performances: le silicium contraint

1.4.1. Principe

L'équation (1.20) montre que le courant de conduction I_{on} est proportionnel à la mobilité des porteurs:

$$I_{on} = \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot \frac{V_{DD}^2}{2}$$

Par conséquent, l'augmentation de la mobilité se traduit par un gain en performances. Or, la mobilité comme toutes les propriétés électroniques des semi-conducteurs dépend de la périodicité des atomes [Lundstrom'00]. En modifiant celle-ci, la mobilité est impactée. Le principe est d'appliquer une contrainte (*stress*) sur le canal. La maille est alors déformée (*strain*), ce qui modifie la mobilité.

1.4.2. Contrainte uniaxiale, biaxiale, tensile, compressive

Les contraintes et les déformations s'expriment sous forme de tenseurs. Grâce aux symétries du cristal de silicium, il est possible de réduire le tenseur des contraintes à deux cas utiles principaux:

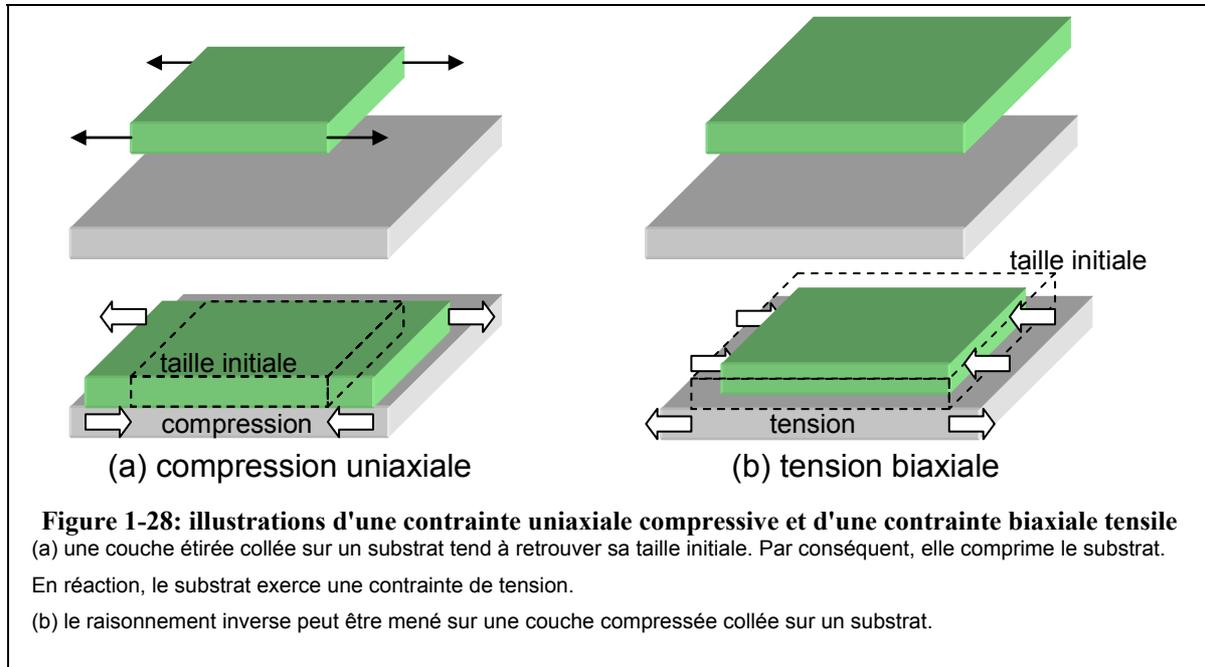
- la contrainte est dite uniaxiale lorsque elle est appliquée uniquement dans le sens source/drain
- la contrainte est dite biaxiale lorsque elle appliquée dans le plan du transistor.

De plus, la contrainte sera compressive si elle tend à réduire le paramètre de maille du cristal de silicium, tensile dans le cas contraire.

1.4.3. Méthodes technologiques pour obtenir la contrainte

L'effet de la contrainte n'est pas toujours bénéfique. C'est pourquoi il est important de comprendre de quelle manière introduire la contrainte voulue par procédé de fabrication. L'analogie à un élément collé à une planche permet de comprendre ce principe:

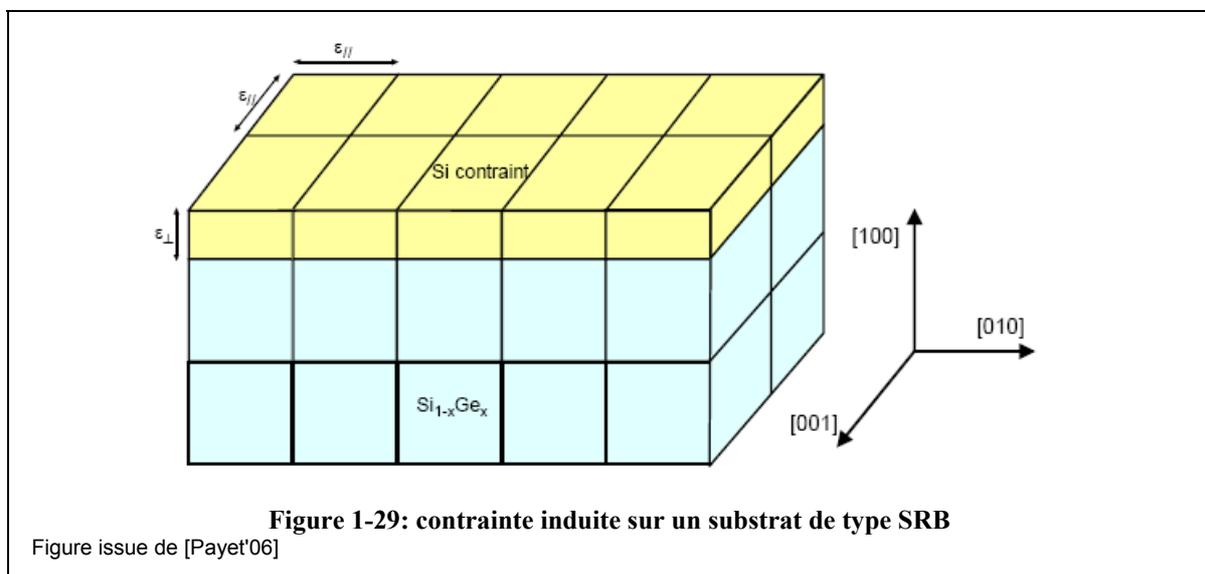
Un élastique étiré puis collé sur une planche tend à retrouver sa forme initiale. Cela va dans le sens de réduire la dimension de la plaque (**figure 1-28a**). Il s'agit donc d'une contrainte uniaxiale compressive.



Un morceau de caoutchoucs comprimé puis collé sur une planche tend également à retrouver sa taille initiale. Cette fois, cela va dans le sens d'augmenter les dimensions de la plaque dans les deux directions (**figure 1-28b**). Il s'agit donc d'une contrainte biaxiale tensile.

Ces deux exemples sont simples et permettent d'évaluer la contrainte résultante de l'action donnée. Il est cependant beaucoup plus difficile de déterminer l'état de contraintes qui résulte d'un procédé de fabrication réel.

Substrat contraint de type SRB (Strain-Relaxed Buffer)



Sur une plaque de silicium, sont épitaxiées (**figure 1-30**):

- une zone tampon de SiGe dont la concentration en germanium croît progressivement
- du SiGe pur
- une fine couche de silicium.

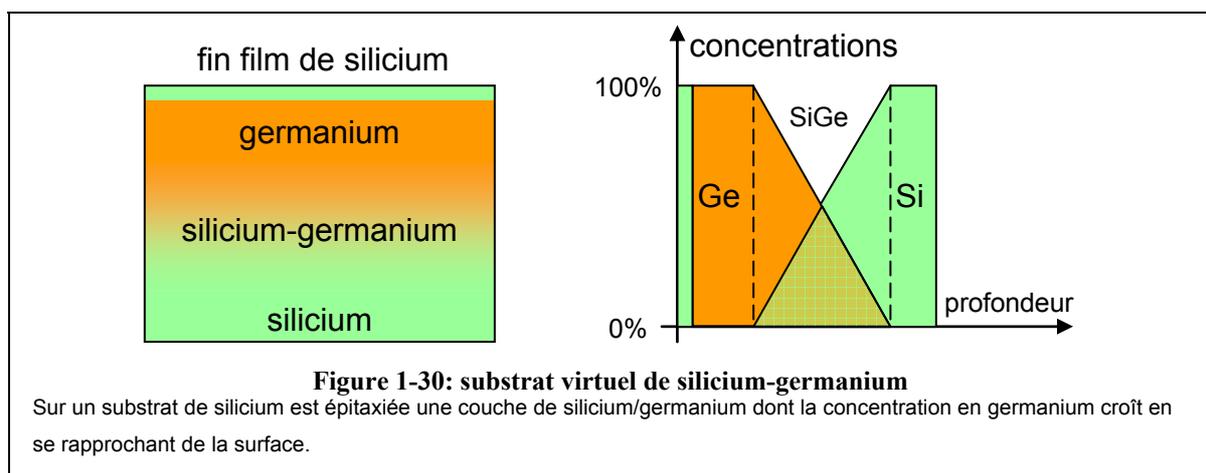
Dans ces conditions, le paramètre de maille dans le substrat est celui du silicium. En se rapprochant de la surface, celui-ci croît progressivement grâce la concentration croissante en germanium. En effet, le germanium possède un paramètre de maille plus élevé que celui du silicium [Sze'81], et les alliages SiGe ont un paramètre intermédiaire [Pakfar'03].

La fine couche de silicium étant déposé sur un substrat à paramètre de maille différent, les atomes viennent s'adapter et sont donc contraints (**figure 1-29**).

Cette méthode permet d'obtenir un substrat dont la zone active est contrainte. Le procédé de fabrication standard est donc utilisable, ce qui est extrêmement avantageux.

Cependant, quelques inconvénients pratiques doivent être mentionnés:

- le substrat est délicat à fabriquer. En effet, des contraintes apparaissent à cause du désaccord de maille, même si la zone tampon de SiGe permet de limiter cet effet.
- le procédé de fabrication complet doit éviter les recuits trop importants. Sinon, les contraintes disparaissent (relâchement des contraintes), et le substrat perd tout son intérêt.
- le fin film de silicium risque d'être consommé au cours des différents nettoyages. Il y a donc lieu d'adapter effectivement le procédé de fabrication.



Contrainte induite lors de la mise en boîtier

La contrainte peut également être induire après la réalisation du dispositif, lors de la mise en boîtier. Cette méthode a le mérite d'être simple, et de ne pas influencer le procédé de

fabrication. Par contre, la contrainte n'est pas maîtrisée, ce qui fait que cette technique n'est pas utilisée industriellement.

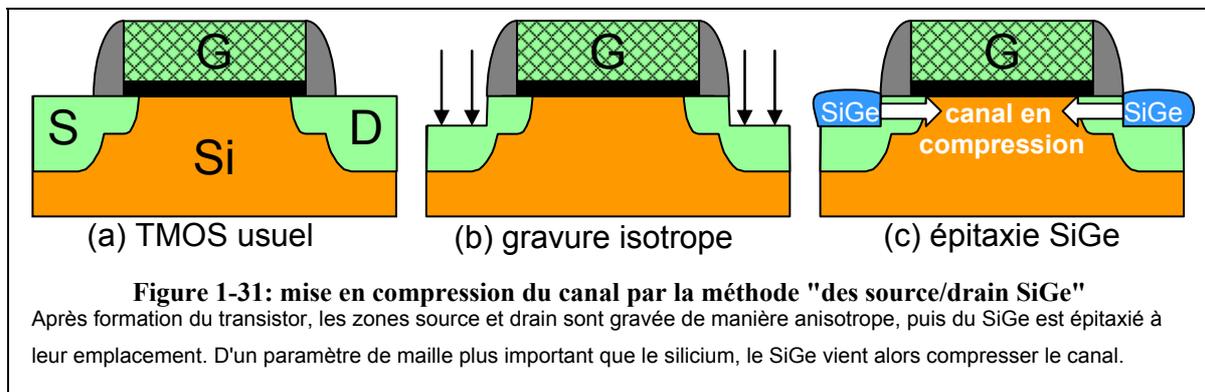
Contrainte induite par CESL

Lors du procédé de fabrication, une couche de nitrure est déposée afin de servir de couche d'arrêt pour la gravure des contacts. C'est le CESL (*Contact Etch Stop Layer*). Or, il est possible de déposer ce matériau de façon contrainte. Celle-ci se transmet au canal et permet d'améliorer la mobilité [Ortolland'04, Shimizu'01]. Cette méthode est utilisée industriellement et permet d'obtenir un gain de 5 à 10% sur le courant de conduction. Contrairement aux cas précédents, la contrainte dépend de la longueur de grille.

Il est possible de discriminer les zones contraintes des zones non contraintes en bombardant ces dernières avec des atomes de germanium, ce qui a pour effet d'amorphiser la couche de nitrure et de faire disparaître la contrainte [Shimizu'01].

Contrainte induite par source/drain SiGe

Une fois le transistor réalisé, les jonctions sont gravées (**figure 1-31b**). Puis une couche de silicium/germanium est épitaxiée. Celui-ci présentant une maille plus grande le silicium, le canal de conduction est donc mis en compression [Ghani'03]. Cette méthode est actuellement utilisée pour la réalisation de microprocesseurs.



Contrainte induite par modification du matériau de grille

Il est possible de générer des contraintes en modifiant le matériau de grille: en le métallisant [Mondot'05], ou en le recristallisant [Duriez'04]. Le matériau de départ (polysilicium) et le matériau final n'ayant pas les mêmes densités, ils sont donc déformés pour pouvoir occuper le même volume. La contrainte issue de cette déformation est transmise au canal de conduction.

Contraintes involontaires

Outres ces méthodes utilisées pour introduire volontairement une contrainte, les étapes traditionnelles du procédé de fabrication peuvent générer également des contraintes. L'exemple le plus marqué est celui des isolations latérales de type STI.

1.4.4. Quelle méthode pour quel porteur?

Il s'agit d'une part d'identifier la contrainte résultant d'une étape technologique, et d'autre part d'évaluer quel impact a cette contrainte sur la mobilité des porteurs. Ce dernier point peut être effectué à l'aide de l'étude de la structure de bandes du silicium [Payet'06].

Le **tableau 1** résume:

- l'impact positif ou négatif de la contrainte
- la contrainte que l'on peut obtenir avec les différentes méthodes technologiques

	contrainte uniaxiale		contrainte biaxiale	
	compressive	tensile	compressive	tensile
électrons	-	+	-	+
trous	+	-	+	+
SRB			✓	
SSOI			✓	✓
mise en boîtier			✓	✓
CESL	✓	✓	✓	✓
S/D SiGe	✓			
matériau de grille				
STI	✓			

Tableau 1: impact de la contrainte sur la mobilité

Le lecteur cherchant plus d'informations sur l'utilisation de la contrainte pourra se reporter au mémoire de thèse de Fabrice Payet [Payet'06].

1.5. Architectures de rupture

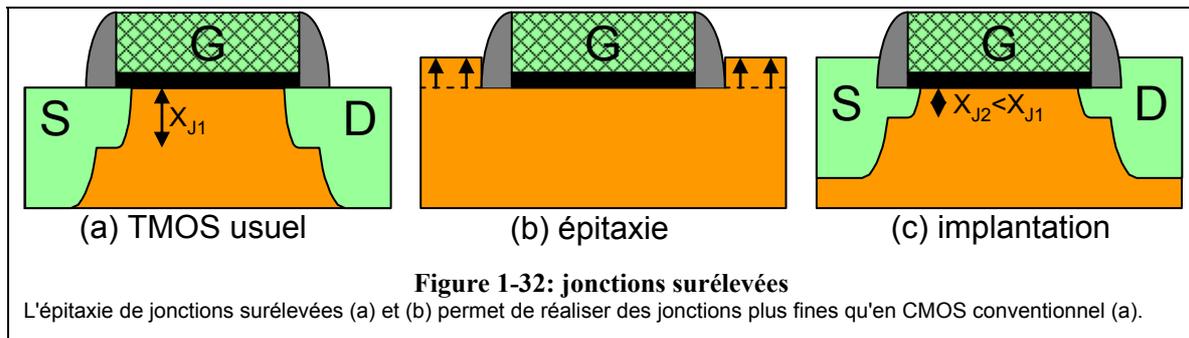
Jusqu'à présent, le transistor a pu être considéré comme un ensemble de modules: module de grille, module de jonctions. Les améliorations évoquées n'affectent qu'un unique module, et ne perturbent que peu les autres. D'autres architectures néanmoins apportent une rupture plus franche.

1.5.1. Nouvelles architectures de jonctions

Jonctions épitaxiées

De la même manière que sont créées les source/drain SiGe, le semi-conducteur peut être remplacé au niveau des jonctions par du SiGe ou du SiGeC. La diffusion des dopants dans ces

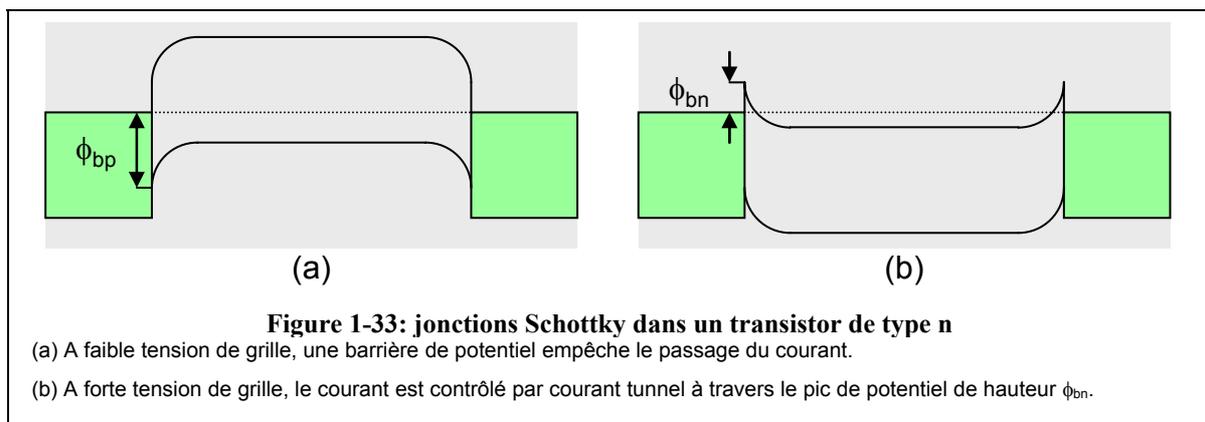
matériaux est moindre que dans le silicium [Pakfar'03], cela permet de limiter la diffusion des jonctions. En outre, ce procédé permet de réaliser des jonctions surélevées (*raised junctions*): l'épétaxie est prolongée de manière à obtenir des zones source/drain plus élevées que l'interface initiale (**figure 1-32a**, **figure 1-32b**). Les implantations sont donc surélevées d'autant, ce qui permet de diminuer la profondeur de jonction (**figure 1-32c**).



Jonctions Schottky

Dans une architecture à jonctions Schottky [Sze'81], les zones source/drain ne sont plus formées de semi-conducteur dopé mais de métal, généralement du siliciure. Par conséquent, la jonction entre source/drain et canal est une jonction métallique ou jonction Schottky. Ce type de jonction présente une barrière de potentiel sous forme de pic, ce qui modifie fortement les mécanismes d'injections.

A faible tension de grille, une barrière de potentiel empêche le passage du courant (**figure 1-33a**). A forte tension de grille, le courant est contrôlé par courant tunnel à travers le pic de potentiel de hauteur ϕ_{bn} (**figure 1-33b**).



L'avantage évident de telles jonctions est de supprimer les principales résistances d'accès: résistance de contact, résistances source/drain et extensions. Cependant, la difficulté majeure

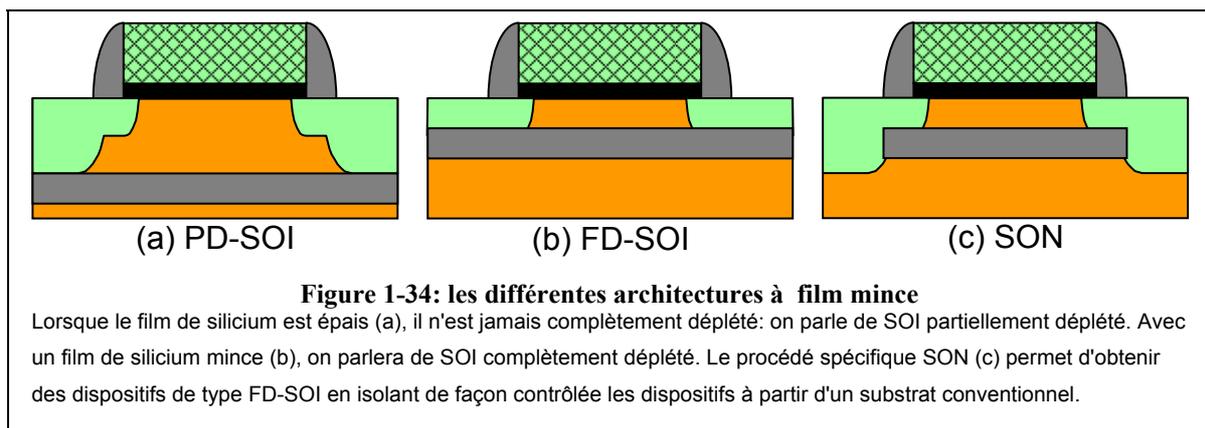
consiste à obtenir un métal formant un pic ϕ_{bn} suffisamment faible pour que l'injection par effet tunnel soit efficace.

1.5.2. Silicium sur isolant

Dans les architectures de type SOI (*Silicon On Insulator*), un isolant enterré, le *BOX* (*Burried Oxide*), sépare électriquement le substrat du film de conduction. L'oxyde enterré doit son nom à l'isolant majoritairement utilisé, la silice SiO_2 . Tout autre isolant (le nitrure Si_3N_4 par exemple) conviendrait également.

Deux paramètres supplémentaires viennent caractériser ces architectures: l'épaisseur du film de silicium et l'épaisseur de l'oxyde enterré. Lorsque le film de silicium est épais ($>150\text{nm}$), la déplétion située sous le canal ne s'étend pas jusqu'à l'oxyde. On parlera alors de SOI partiellement déplété (**figure 1-34a**) ou PD-SOI (*Partially-Depleted SOI*). Dans le cas contraire, on parlera de SOI totalement déplété (**figure 1-34b**) ou FD-SOI (*Fully-Depleted SOI*).

Un cas particulier de FD-SOI est l'architecture SON (*Silicon On Nothing*) [Jurczak'00]. Dans l'architecture FD-SOI, le transistor est réalisé sur un substrat spécifique où l'oxyde enterré est présent sous toute la surface de la plaque. En technologie SON, un substrat standard est utilisé. L'oxyde enterré est réalisé en même temps que le transistor grâce à un système de couche sacrificielle. Ceci permet de limiter la présence de l'oxyde aux zones actives (**figure 1-34c**). On parle également de SOI localisé.



Il existe plusieurs avantages à utiliser une technologie SOI:

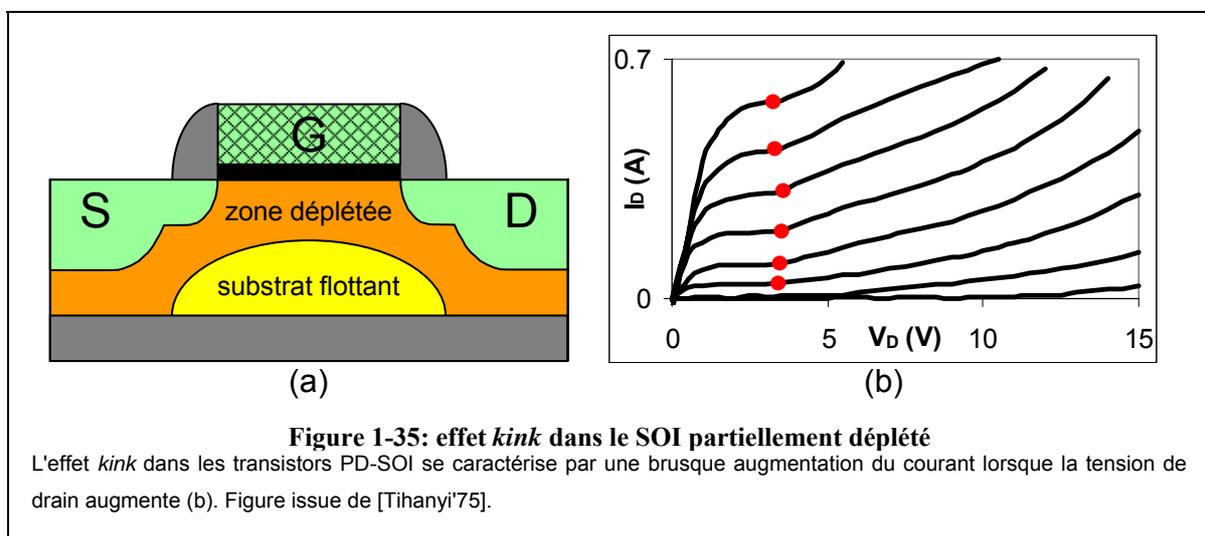
- profondeur de déplétion et profondeur de jonction sont définies par l'épaisseur du film de silicium. Si ces profondeurs sont de l'ordre de 20 à 40nm, des films de silicium d'une dizaine de nanomètres en technologie FD-SOI et SON ont été obtenus. Cette diminution rend ces architectures intrinsèquement plus résistantes aux effets canaux courts.

- l'oxyde enterré constitue une isolation entre les transistors et le substrat. Après réalisation du STI, les transistors sont donc parfaitement isolés les uns par rapport aux autres.
- cette isolation entre transistor et substrat permet également de s'affranchir des effets des radiations ionisantes. Celles-ci induisent des charges en profondeur, qui grâce à cette isolation n'ont aucun impact sur le transistor. Le SOI constitue donc un candidat idéal pour les applications spatiale et avionique.

1.5.2.1. Les transistors partiellement déplétés

Le principal avantage de ce type de transistor est qu'il utilise une technologie similaire à celle des transistors conventionnels. Il peut donc facilement s'insérer dans un procédé de fabrication standard. Le principal inconvénient de la technologie PD-SOI est la présence d'une zone neutre isolée du substrat sous le canal de conduction (**figure 1-35a**). Ce substrat flottant (*body*) est à l'origine de différents effets parasites:

- des effets transitoires: lorsque la polarisation de grille évolue de 0V à V_{DD} , les porteurs majoritaires sont repoussés dans le substrat, abaissant la tension de seuil et augmentant le courant de drain. Lorsque la jonction source/substrat devient passante, ces charges sont évacuées et le système retrouve un état stationnaire. De façon similaire, lorsque la polarisation de grille passe de V_{DD} à 0V, les porteurs majoritaires devraient revenir sous la grille. Or, ces porteurs ne sont pas accessibles car il n'y a pas de connexion avec le substrat. Il faut donc attendre qu'ils soient recréés par mécanisme de génération, relativement lent.
- l'effet *kink*: lorsque le drain est fortement polarisé, des paires électron-trou se créent à la frontière canal/drain [Tihanyi'75]. Les porteurs majoritaires sont évacués par le drain, mais les porteurs minoritaires sont stockés dans le substrat flottant, abaissant ainsi la tension de seuil. Ceci entraîne une hausse du courant à la forme de coude caractéristique (**figure 1-35b**).



- l'effet bipolaire parasite: lorsque le substrat flottant est chargé et que la diode source/substrat est mise en conduction, le système se retrouve dans une configuration de type npn (ou p pour le P-MOS) entre la source, le substrat flottant et le drain. Ce système rentre en conduction (le bipolaire est activé) lorsque la base (substrat) est suffisamment polarisée. Le drain joue alors le rôle de collecteur.

- effet d'auto-échauffement: la couche d'oxyde sous le canal de conduction constitue un obstacle pour la dissipation de la chaleur générée dans le dispositif. En effet, la conductivité thermique de l'oxyde est environ 100 fois plus faible que celle du silicium ($K_{Si}=150W/m.K$ comparé à $K_{SiO_2}=1,4W/m.K$) [Sze'81]. Cette élévation de la température entraîne une chute des performances: le courant de fuite I_{off} augmente à cause d'une pente sous le seuil élevée (équations (1.26) et (1.28)). Le courant de conduction I_{on} chute car la hausse de température augmente les interactions porteur/phonon, diminuant ainsi la mobilité.

Il est possible de supprimer ces effets parasites en utilisant un contact de substrat. Ceci se fait au détriment de la densité d'intégration.

1.5.2.2. Les transistors totalement déplétés

Dans un transistor totalement déplété, l'épaisseur du film de silicium est plus faible que la profondeur de déplétion t_{dep} :

$$t_{dep} = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_{ch}} (\phi_s - V_{BS})}$$

Or, celle-ci dépend du dopage canal et de la tension de grille par l'intermédiaire du potentiel de surface ϕ_s . Notamment, cette relation doit se vérifier pour $V_G=V_{th}$, là où la déplétion est minimale et le transistor en conduction. Cette condition détermine la relation qui lie l'épaisseur maximale du film et le dopage canal:

$$t_{Si,max} = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_{ch}} (V_{bi} - V_{BS})} \quad (1.33)$$

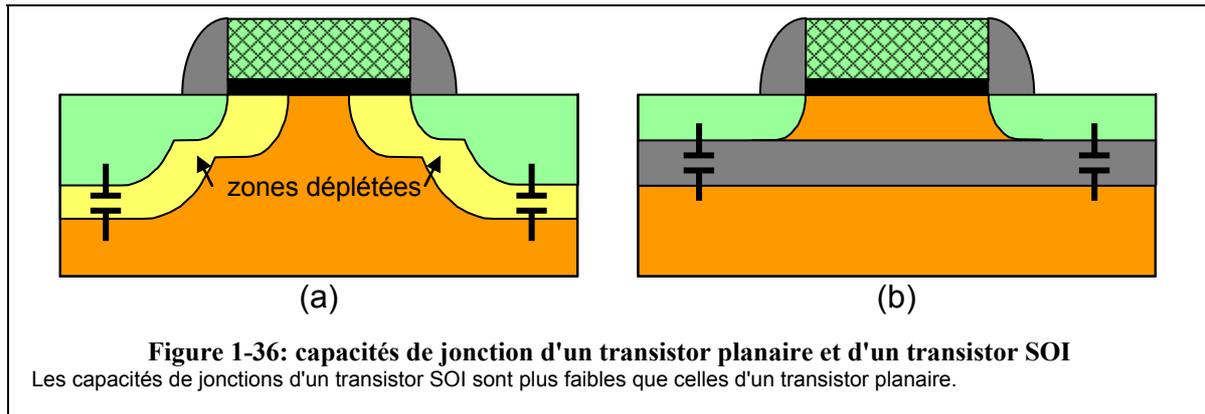
Les transistors totalement déplétés présentent de nombreux avantages par rapport aux transistors partiellement déplétés et conventionnels:

- la profondeur des jonctions est limitée par l'épaisseur du film de silicium. Cela permet donc un meilleur contrôle des effets canaux courts, sans recours à un dopage canal élevé.
- l'effet de substrat disparaît. Par conséquent, le contact de substrat devient inutile.
- les capacités de jonctions sont réduites. En effet, les zones de déplétions sous les zones source/drain d'un transistor planaire forment une capacité parasite (**figure 1-36a**). En

technologie FD-SOI, les source/drain descendent jusqu'à l'oxyde enterré. Cette capacité est donc remplacée par la capacité d'oxyde enterré:

$$\frac{\epsilon_{\text{Si}}}{t_{\text{dep}}} \Rightarrow \frac{\epsilon_{\text{SiO}_2}}{t_{\text{box}}} \quad (1.34)$$

Or, la permittivité de la silice (3.9) est quatre fois moindre que celle du silicium (11.9), et l'épaisseur de l'oxyde enterré est contrôlée. Ceci entraîne donc une diminution des capacités de jonction.



- les fuites de jonction sont supprimées, ce qui est un avantage majeur pour les dispositifs à rétention de charges.

- la résistance aux effets canaux courts permet d'utiliser des films non dopés. Ceci se traduit par un gain en mobilité.

Cependant, les transistors FD-SOI présentent quelques difficultés d'intégration:

- l'utilisation de films minces et donc de jonctions minces entraîne une hausse des résistances série. Les extensions sont alors fortement dopées, et deviennent alors plus sensibles à la diffusion latérale.

- l'épaisseur du film détermine la profondeur de déplétion et donc la tension de seuil. Toute fluctuation technologique se voit donc directement répercutée sur les performances du transistor.

- l'utilisation de film non dopé modifie le travail de sortie. Le polysilicium dopé n'est donc plus adapté pour une tension de seuil acceptable. L'intégration d'un autre matériau devient nécessaire.

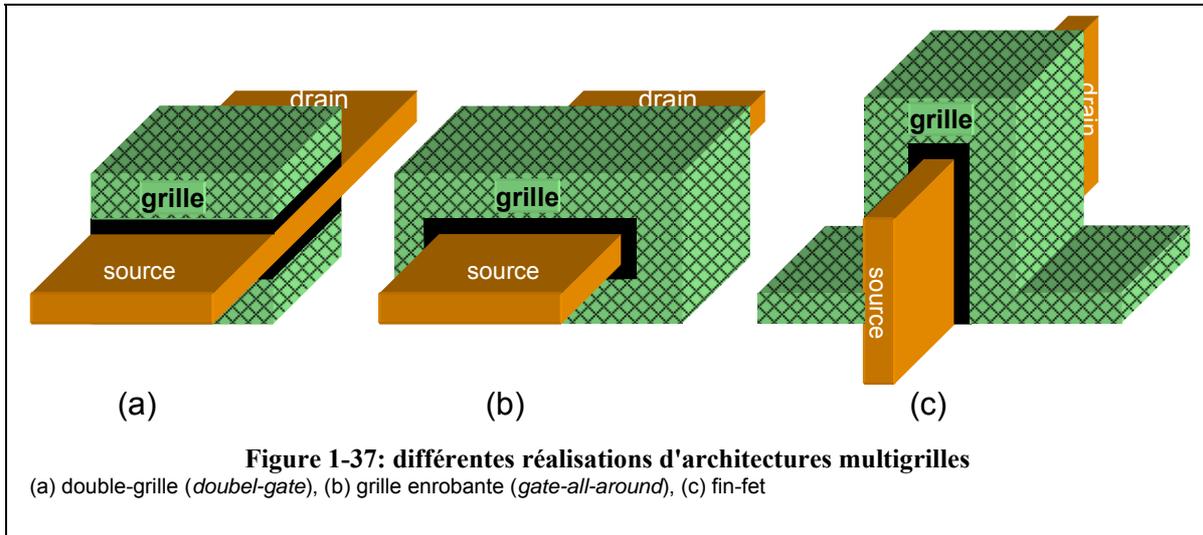
- isolation latérale: les STI ne peuvent pas être réalisées de façon standard

- la siliciuration des films minces est délicate

1.5.3. Transistors multigrilles

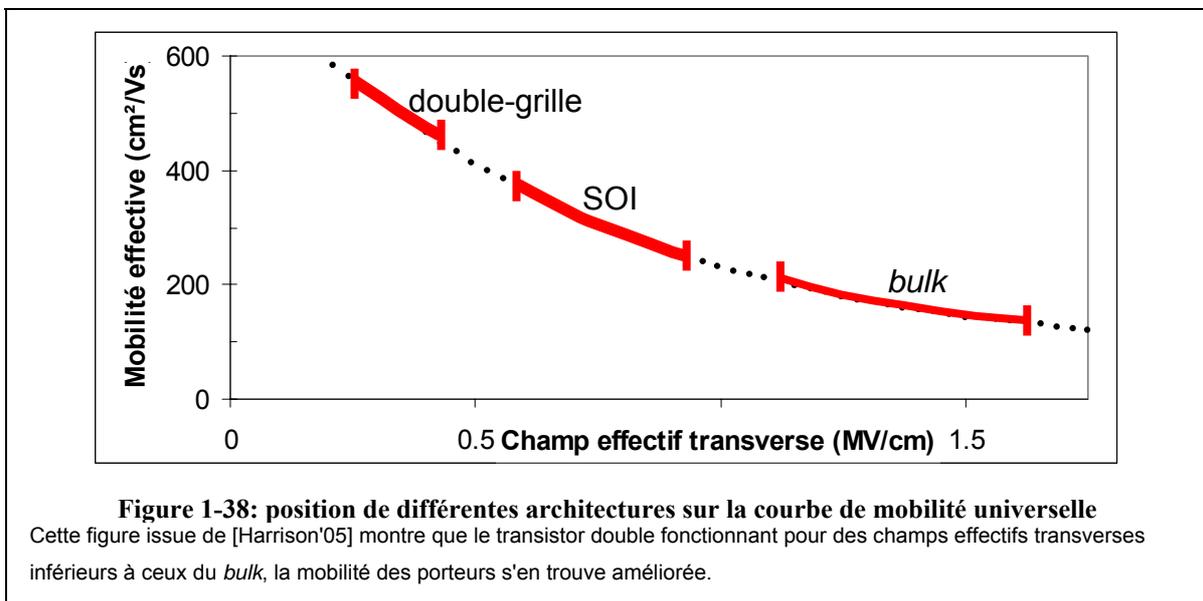
Une autre approche consiste à utiliser plusieurs grilles pour commander plusieurs interfaces de conduction [Harrison'04]. Plusieurs intégrations sont proposées:

- le double-grille (**figure 1-37a**) permet d'avoir deux grilles indépendantes
- la grille enrobante (**figure 1-37b**) permet d'avoir quatre interfaces de conduction
- le *fin-FET* (**figure 1-37c**) doit permettre une plus grande densité d'intégration



Les architectures multigrilles présentent de nombreux avantages:

- présentant plusieurs interfaces, elles multiplient théoriquement d'autant le courant de conduction.
- sur film mince, elles offrent un meilleur contrôle du potentiel. Ceci entraîne un meilleur contrôle des effets canaux courts.
- en considérant des grilles symétriques, il s'ensuit que le champ électrique vertical est nul au milieu du film. Par conséquent, le champ vertical moyen est plus faible que dans une architecture planaire ou SOI. Cette baisse de champ électrique permet de limiter les



interactions de surface, et donc d'augmenter la mobilité. En effet, celle-ci décroît avec le champ effectif (**figure 1-38**).

- en réduisant l'épaisseur du film de conduction, le minimum de potentiel au centre du film augmente [Josse'00]. Par conséquent, la densité de porteurs libres augmente également et la conduction ne se fait plus uniquement aux interfaces (**conduction volumique**), ce qui augmente encore le courant de conduction.

- enfin, utiliser deux grilles indépendantes permet d'implémenter des fonctions logiques sur un seul dispositif, et donc d'augmenter la densité d'intégration.

Conclusion du chapitre

La complexité croissante d'intégration du transistor MOS a toujours été accompagnée de pronostics sur les limitations dans la réduction des dimensions. Les limites annoncées furent une longueur de grille de $1\mu\text{m}$, puis 100nm , puis 10nm . La perpétuelle réussite à passer les premières barrières rend prudent quant à l'existence d'une réelle limitation. En outre, des transistors planaire d'une longueur de grille de 6nm ont été présentés.

Cependant, la difficulté croissante d'intégration est bien réelle. C'est pourquoi les architectures de rupture semblent devoir à terme supplanter l'intégration classique. Bien que les premiers produits grand publics ont déjà initialisé ce processus (les processeurs des consoles *next gen* de Microsoft (XBox 360) et Sony (PlayStation 3) sont réalisés sur substrat SOI partiellement déplétés), l'utilisation de ces technologies est onéreuse. C'est pourquoi leur introduction tend à être repoussée autant que possible.

La modélisation du transistor MOS a permis dans ce chapitre de mettre en évidence les moyens technologiques utiles à l'amélioration des performances. Cependant, une modélisation plus fine devient nécessaire pour prendre en compte correctement les effets canaux courts.

Chapitre 2. Le transistor MOS sous le seuil

La pente sous le seuil du transistor MOS est intrinsèquement limitée à 60mV/décade à température ambiante, quelle que soit la technologie employée (*bulk*, SOI, double-grille) ou les améliorations technologiques utilisées (stress, oxyde de grille à haute constante diélectrique...). L'origine physique de cette limitation est rappelée dans la première partie, en constatant qu'elle s'applique aux dispositifs élémentaires sur semi-conducteur: jonction p-n, transistor bipolaire.

Puis, la pente sous la seuil est modélisée, et en particulier son évolution avec la réduction de la longueur de grille. Cette modélisation se base sur la transformation tension dopage (VDT), qui permet de prendre en compte le couplage latéral en tant que diminution du dopage apparent du substrat. Ce modèle sera validé par des simulations 2D pour les technologies bulk et FD-SOI.

Enfin, le courant du transistor en régime sous le seuil est modélisé. Ceci permet en particulier d'exprimer le courant de fuite I_{off} .

Introduction

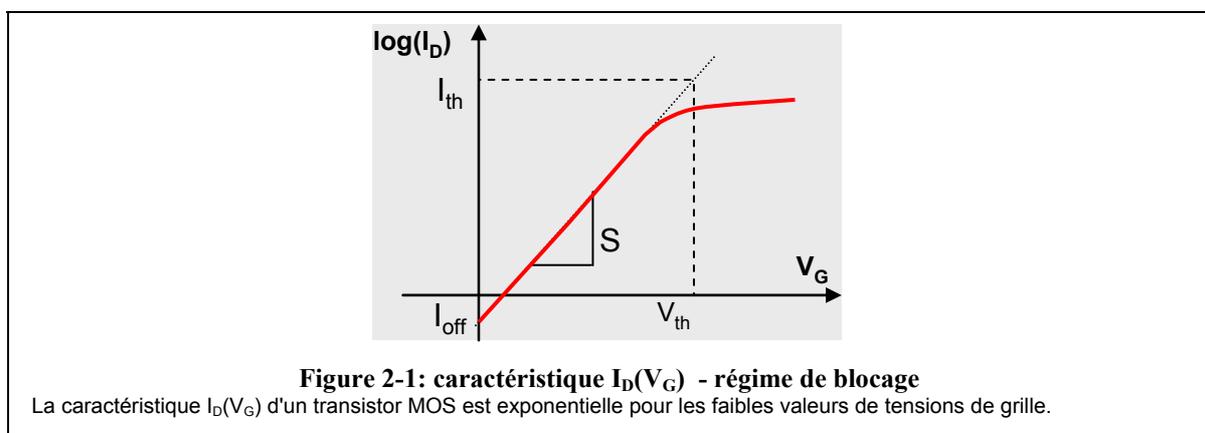
Le régime de blocage du transistor MOS correspond aux tensions de grille inférieures à la tension de seuil. C'est pourquoi ce régime est également appelé régime **sous le seuil**, le courant correspondant **courant sous le seuil** I_{sth} (*SubTHreshold*) et la pente de la courbe $\log(I_D)$ en fonction de V_G **pente sous le seuil**.

Il apparaît que lorsque les fuites de grille sont négligeables, le courant de drain à tension de grille nulle I_{off} est un cas particulier de courant sous le seuil. Or, ce paramètre est essentiel pour l'évaluation des performances d'un dispositif, car il est directement relié à la consommation statique des circuits. Pour les systèmes embarqués, cela joue sur leur autonomie. Pour des applications haute performance, cela permet de réduire l'échauffement des puces, et donc d'améliorer la fiabilité.

Au chapitre précédent, les résultats de la modélisation analytique des effets canaux courts ont mis en évidence les moyens technologiques pour les combattre. De manière similaire, il serait donc intéressant de disposer d'un modèle analytique simple permettant de calculer I_{off} en fonction de paramètres technologiques. Actuellement, I_{off} est calculé dans le modèle Mastar [Mastar'04] par l'expression:

$$\log(I_{off}) = \log(I_{th}) - \frac{V_{th}}{S} \quad (2.1)$$

où I_{th} est la prolongation de la partie sous le seuil pour $V_G = V_{th}$ (**figure 2-1a**). Par conséquent, la modélisation rigoureuse du courant sous le seuil par l'intermédiaire des grandeurs S et I_{th} est nécessaire pour la détermination précise du courant I_{off} .



2.1. Pente sous le seuil des dispositifs semi-conducteurs

La notion de pente sous le seuil n'est pas spécifique au transistor MOS. Elle existe pour tout dispositif semi-conducteur (diode, transistor bipolaire...) lorsque le courant est attribué aux porteurs minoritaires. De façon générale, la pente sous le seuil peut être définie comme l'incrément de tension de commande nécessaire à la variation de courant de sortie d'une décade:

$$S = \frac{\partial V_{in}}{\partial \log(I_{out})} \quad (2.2)$$

2.1.1. Limite fondamentale de la pente sous le seuil

Le courant sous le seuil est un déplacement de porteurs minoritaires. Ils obéissent à une loi statistique de Fermi-Dirac, qui peut être approximée par une statistique de Boltzman:

$$Q = Q_0 \cdot \exp\left(\frac{k_b \cdot T}{q \cdot \phi}\right) \quad (2.3)$$

où Q est la charge de déplacement, ϕ son potentiel moyen et Q_0 une constante homogène à une charge. Le courant est le quotient de cette charge par le temps de parcours des porteurs:

$$I_{out} = \frac{Q}{t} = \frac{Q_0}{t} \cdot \exp\left(\frac{k_b \cdot T}{q \cdot \phi}\right) \quad (2.4)$$

Il est habituel de considérer [Skotnicki] que le temps de parcours des porteurs ne dépend pas du potentiel ϕ . Par conséquent, l'expression (2.2) devient:

$$S = \frac{k_b \cdot T}{q} \cdot \ln(10) \cdot \frac{\partial V_{in}}{\partial \phi} \quad (2.5)$$

La quantité $\frac{\partial V_{in}}{\partial \phi}$ représente l'incrément de tension de commande nécessaire à l'augmentation du potentiel de surface. Cette quantité est au mieux égale à l'unité. Par conséquent, la pente de tout dispositif semi-conducteur est intrinsèquement limitée:

$$S \geq \frac{k_b \cdot T}{q} \cdot \ln(10) = 60\text{mV} / \text{dec} @ 300\text{K} \quad (2.6)$$

2.1.2. Pente des dispositifs semi-conducteurs élémentaires

2.1.2.1. La jonction p-n

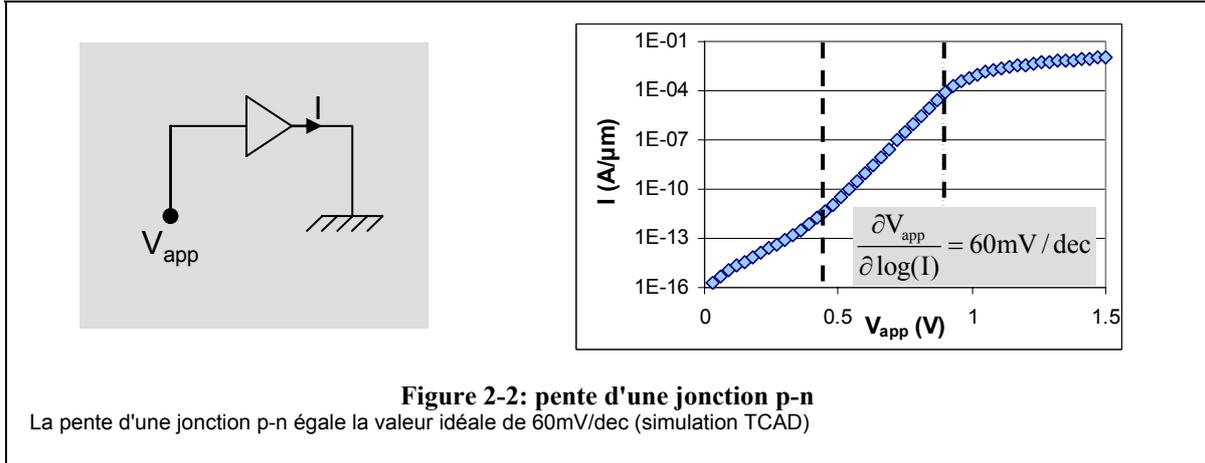
Le courant en inverse d'une diode polarisée s'écrit [Sze'81]:

$$I_{pn} = I_S \cdot \left[\exp\left(\frac{q \cdot V_{app}}{k_b \cdot T}\right) - 1 \right] \quad (2.7)$$

Cette relation permet de déduire l'expression de la pente:

$$S_{p-n} = \frac{k_b \cdot T}{q} \cdot \ln(10) \quad (2.8)$$

Dans ce cas, la pente de la caractéristique $I(V_{app})$ atteint la valeur minimale de 60mV/dec.



2.1.2.2. Le transistor bipolaire

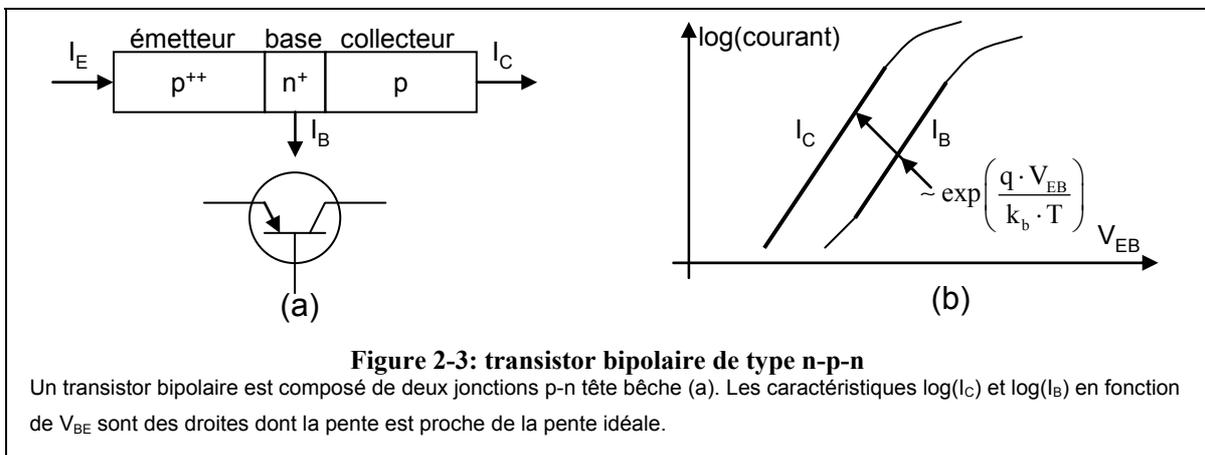
Un transistor bipolaire est la réunion de deux jonctions p-n tête-bêche. Le courant de collecteur est le courant inverse de la jonction base-collecteur. Celui est proportionnel à la densité de porteurs minoritaires injectés par la jonction émetteur-base polarisée en direct.

On peut déduire du modèle de diode les équations de courant du transistor bipolaire [Sze'81]:

$$I_E = A \cdot q \cdot \frac{D_B \cdot p_B}{L_B} \coth\left(\frac{W}{L_B}\right) \cdot \left[\left(e^{q \cdot V_{EB}/k \cdot T} - 1 \right) - \frac{1}{\cosh(W/L_B)} \cdot \left(e^{q \cdot V_{CB}/k \cdot T} - 1 \right) \right] + A \cdot q \cdot \frac{D_E \cdot n_E}{L_E} \cdot \left(e^{q \cdot V_{EB}/k \cdot T} - 1 \right)$$

$$I_E = A \cdot q \cdot \frac{D_B \cdot p_B}{L_B} \frac{1}{\sinh(W/L_B)} \cdot \left[\left(e^{q \cdot V_{EB}/k \cdot T} - 1 \right) - \cosh\left(\frac{W}{L_B}\right) \cdot \left(e^{q \cdot V_{CB}/k \cdot T} - 1 \right) \right] + A \cdot q \cdot \frac{D_C \cdot n_C}{L_C} \cdot \left(e^{q \cdot V_{CB}/k \cdot T} - 1 \right) \quad (2.9)$$

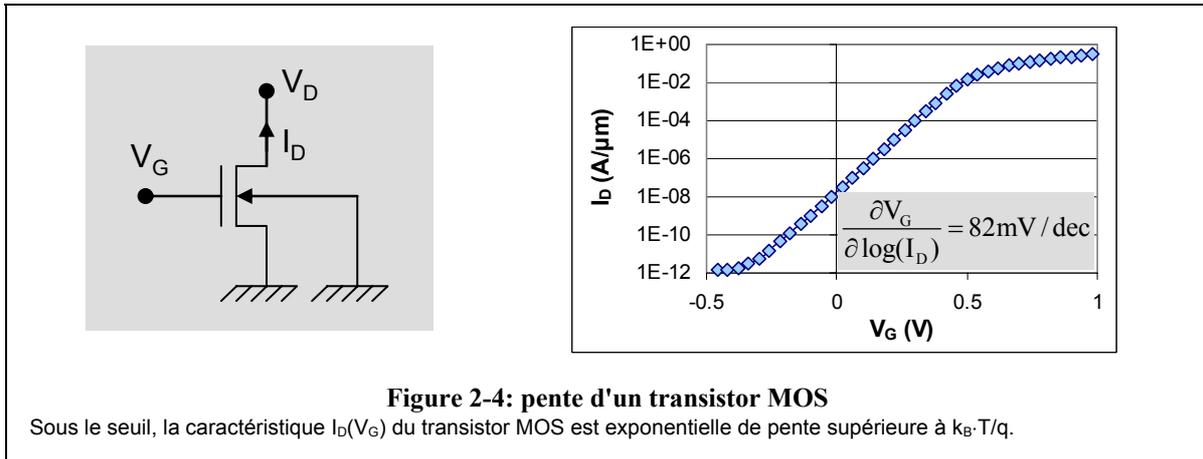
$$I_B = I_E - I_C$$



2.1.2.3. Le transistor MOS

Dans le cas du transistor MOS, la tension de grille V_G module le potentiel de surface ϕ_s . Le courant de sortie est le courant de drain:

$$S_{\text{T MOS}} = \frac{\partial V_G}{\partial \log(I_D)} = \frac{k_b \cdot T}{q} \cdot \ln(10) \cdot \frac{\partial V_G}{\partial \phi_s} \quad (2.10)$$



Cette limite fondamentale résulte de la distribution des porteurs. Elle ne dépend donc pas de l'architecture (SOI, double-grille) ou des moyens technologiques utilisés pour améliorer les performances (grille métallique, silicium contraint).

2.2. Modélisation de la pente sous le seuil du TMOS

Le courant de fuite I_{of} dépend principalement de la tension de seuil V_{th} et de la pente sous le seuil S . C'est pourquoi la dépendance de ces grandeurs avec les paramètres technologiques a été étudiée [Eitan'82, Fu'84, Fu'97, VanDamme'97]. Cependant, aucune de ces études ne permet de rendre compte de la dégradation de la pente avec la réduction de la longueur de grille. Un moyen élégant de modéliser cet effet est tenir compte du champ électrique source/drain [Gwoziecki'02] en utilisant la transformation tension dopage (VDT pour *Voltage Doping Transformation*) [Skotnicki'88].

2.2.1. La transformation tension dopage (VDT)

Cette transformation repose sur l'expression de l'équation de Poisson:

$$\vec{\nabla}(\epsilon \cdot \vec{E}) = \rho \Leftrightarrow \vec{\nabla}(-\epsilon \cdot \vec{\nabla}(\phi)) = \rho \quad (2.11)$$

Dans un transistor en déplétion ou en inversion faible, la charge peut être réduite à la charge déplétion $q \cdot N_b$. Dans cette expression et par la suite, nous supposons étudier un transistor de

type n, avec un substrat de type p. En considérant l'approche usuelle qui consiste effectuer une étude uni dimensionnelle du transistor dans l'axe vertical, l'équation (2.11) peut s'écrire:

$$\Delta\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = \frac{q \cdot N_{ch}}{\epsilon} \quad (2.12)$$

où y est l'axe des profondeurs, x l'axe source/drain et z l'axe de la largeur et N_{ch} le dopage canal. En canal large et étant données les symétries, la composante en z de $\Delta\phi$ est nulle. En canal long, les variations de champ dans l'axe source/drain peuvent être négligées. L'équation de Poisson se ramène alors à un problème unidimensionnel:

$$\frac{\partial^2\phi}{\partial x^2} = \frac{q \cdot N_{ch}}{\epsilon} \quad (2.13)$$

Lorsque la longueur de grille réduit, la composante y de $\Delta\phi$ ne peut plus être négligée. L'équation de Poisson s'écrit alors:

$$\Delta\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = \frac{q \cdot N_{ch}}{\epsilon} \quad (2.14)$$

On peut réécrire cette équation sous la forme:

$$\frac{\partial^2\phi}{\partial x^2} = \frac{q \cdot N_{ch}}{\epsilon} - \frac{q}{\epsilon} \cdot \left(\frac{\epsilon \partial^2\phi}{q \partial y^2} \right) \quad (2.15)$$

Le terme de droite peut être assimilée à une modification du dopage. L'équation (2.15) devient donc:

$$\frac{\partial^2\phi}{\partial x^2} = \frac{q \cdot N_{ch}^*}{\epsilon} \quad (2.16)$$

où N_{ch}^* est un **dopage apparent**. On montre [Skotnicki'88] que la valeur de ce dopage peut être exprimée en fonction des paramètres de la structure:

$$N_{ch}^*(x) = N_{ch} - \frac{2 \cdot \epsilon_{Si}}{q \cdot L_{el}(x)^2} \left[V_{DS} + 2(V_{bi} - \phi(x)) + 2\sqrt{(V_{bi} - \phi(x))(V_{DS} + V_{bi} - \phi(x))} \right] \quad (2.17)$$

où L_{el} est la distance électrique source/drain, ϕ_s le potentiel de surface et V_{bi} la barrière de potentiel du TMOS.

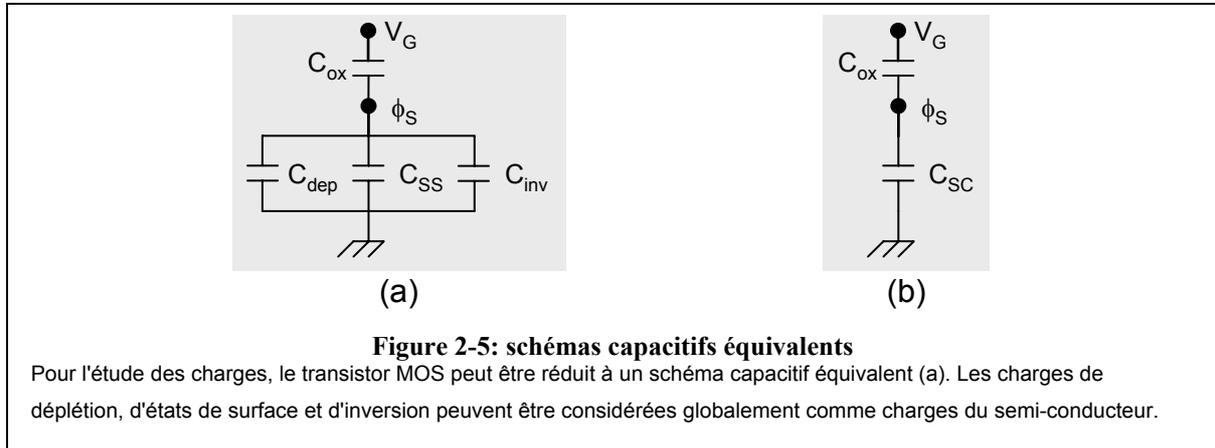
2.2.2. Pente sous le seuil en technologie substrat massif

2.2.2.1. Expression de la pente

La pente sous le seuil du transistor MOS est:

$$S = \frac{k_b \cdot T}{q} \cdot \ln(10) \cdot \frac{\partial V_G}{\partial \phi_s} \quad (2.18)$$

Calculer la pente sous le seuil du TMOS revient donc à exprimer la relation entre tension de grille V_G et potentiel de surface ϕ_s . Pour cela, il faut considérer le schéma capacitif équivalent de la structure MOS (**figure 2-5a**) introduit au chapitre 1. Les capacités dues aux charges de déplétion, d'inversion et d'interface peuvent être considérées globalement comme charges du



semi conducteur Q_{SC} (**figure 2-5b**). Elles sont l'image des charges sur la grille Q_G :

$$Q_G = -Q_{SC} \quad (2.19)$$

De plus, le schéma capacitif nous permet d'écrire que la charge dans la grille est proportionnelle à la chute de potentiel dans l'oxyde grille $V_G - \phi_s$:

$$Q_G = (V_G - \phi_s) \cdot C_{OX} \quad (2.20)$$

Les équations (2.19) et (2.20) permettent donc de déduire:

$$V_G = \phi_s - \frac{Q_{SC}}{C_{OX}} \quad (2.21)$$

L'injection de cette relation dans l'équation (2.10) fournit la formulation de la pente sous le seuil en fonction de la charge du semi-conducteur:

$$S = \frac{k \cdot T}{q} \cdot \ln(10) \cdot \left(1 - \frac{1}{C_{OX}} \cdot \frac{\partial Q_{SC}}{\partial \phi_s} \right) \quad (2.22)$$

La modélisation de la pente sous le seuil du TMOS revient donc à exprimer la charge du semi-conducteur Q_{SC} en fonction du potentiel de surface ϕ_s . Pour cela, la solution courante consiste résoudre l'équation de Poisson en 1D dans le sens de la profondeur du semi-conducteur.

En outre, l'expression analytique de la dérivée $\frac{\partial Q_{SC}}{\partial \phi_s}$ peut être simplifiée par discrétisation:

$$\frac{\partial Q_{SC}}{\partial \phi_s} \sim \frac{Q_{SC}(\phi_{s1}) - Q_{SC}(\phi_{s2})}{\phi_{s1} - \phi_{s2}} \quad (2.23)$$

La charge sera calculée pour $\phi_{s1} = \phi_D$ et $\phi_{s2} = \phi_D / 2$, où ϕ_D (également noté V_{bi}) est la différence de potentiel interne entre la source et le canal en l'absence de polarisation:

$$\phi_D = \frac{k \cdot T}{q} \cdot \ln\left(\frac{N_D \cdot N_{ch}}{n_i^2}\right) \quad (2.24)$$

où N_D est le dopage de la source. L'expression (2.22) devient alors:

$$S = \frac{k \cdot T}{q} \ln(10) \left(1 - \frac{1}{C_{ox}} \frac{Q_{SC}(\phi_s = \phi_D) - Q_{SC}(\phi_s = \phi_D / 2)}{\phi_D / 2} \right) \quad (2.25)$$

Il est intéressant de remarquer que le calcul de la charge à $\phi_s = \phi_D$ permet d'accéder à la valeur de la tension de seuil:

$$V_{th} = \phi_{MS} + \phi_D + \frac{Q_{SC}(\phi_s = \phi_D)}{C_{ox}} \quad (2.26)$$

2.2.2.2. Approximation du profil de dopage apparent

L'expression du dopage apparent dans la VDT fait intervenir le potentiel. Par conséquent, l'utilisation rigoureuse de la VDT implique une résolution auto-cohérente des équations fondamentales du semi-conducteur (équation de Poisson, conservation des trous et des électrons) en même temps que de la VDT. C'est l'approche utilisée en simulation TCAD, mais nécessite d'importantes ressources numériques.

Or, l'intérêt majeur d'un modèle analytique réside dans la simplification judicieuse des équations de la physique, qui permet à la fois de rendre compte des effets physiques avec acuité, tout en ne nécessitant qu'une unique itération de calcul. Il est donc nécessaire de simplifier l'utilisation conjointe de l'équation de Poisson et de la VDT.

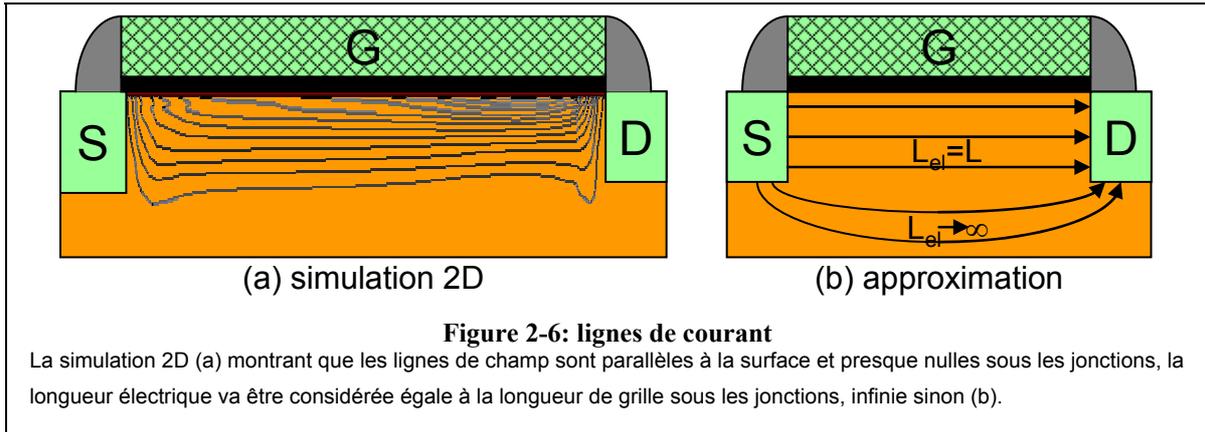
Dans une technologie bien maîtrisée, les charges d'interface peuvent être négligées [Skotnicki'00]. De plus, en régime sous le seuil, la charge d'inversion peut être négligée par rapport à la charge de déplétion. La charge dans le semi-conducteur se réduit donc à la charge de déplétion:

$$Q_{SC} \sim Q_{dep} \quad (2.27)$$

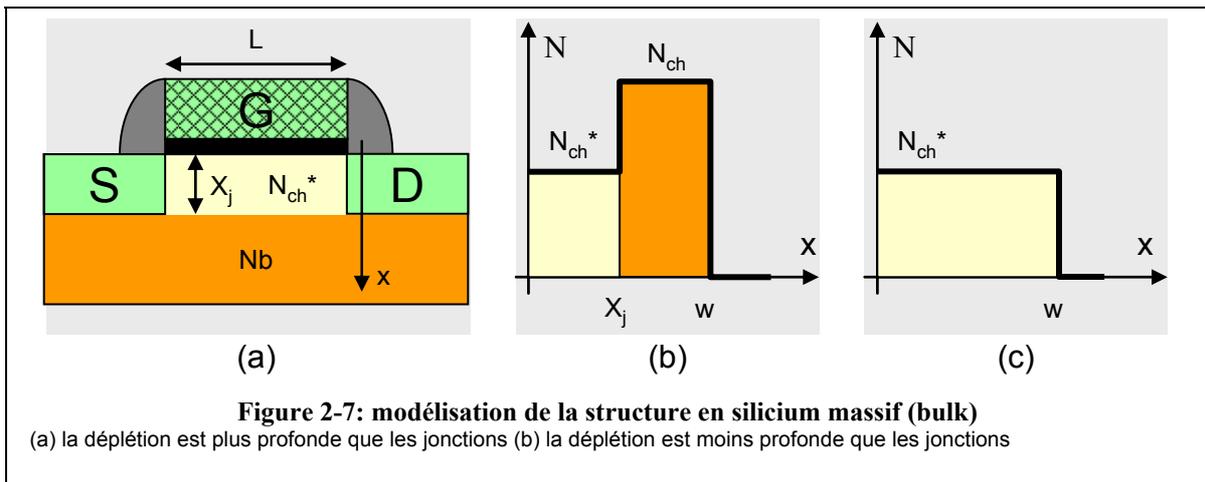
Il suffit donc de connaître le profil de dopage pour résoudre l'équation de Poisson de manière conventionnelle. Pour opérer de manière similaire, la VDT va d'abord être utilisée pour déterminer le profil de dopage apparent.

Les simulations numériques montrent (**figure 2-6a**) que les lignes de courant sont parallèles à la direction de conduction au niveau des jonctions. Ces lignes s'allongent rapidement dès que l'on considère des profondeurs supérieures à celles des jonctions. Ceci se traduit dans la VDT par la longueur électrique $L(x)$ (**figure 2-6b**):

$$\begin{aligned} L_{el}(x) &= L \text{ si } x \leq X_j \\ L_{el}(x) &\rightarrow \infty \text{ si } x > X_j \end{aligned} \quad (2.28)$$



Par conséquent, le dopage va être supposé constant. Il sera modifié par la VDT pour des profondeurs inférieures à la profondeur de jonction, là où le champ électrique est important, et laissé identique au dopage canal pour des profondeurs plus importantes (**figure 2-7a** et **figure 2-7b**). Cependant, lorsque la profondeur déplétion est inférieure à la profondeur de jonction, le profil de dopage se résume à un unique dopage uniforme (**figure 2-7c**).



Pour rendre compte de la variation de potentiel au long de la profondeur, la VDT est appliquée pour un potentiel égal à une fraction du potentiel de surface:

$$N_{ch}^*(x) = N_{ch} - \frac{2 \cdot \epsilon_{Si}}{q \cdot L_{el}(x)^2} \left[V_{DS} + 2(V_{bi} - \alpha \cdot \phi_s) + 2\sqrt{(V_{bi} - \alpha \cdot \phi_s)(V_{DS} + V_{bi} - \alpha \cdot \phi_s)} \right] \quad (2.29)$$

où α servira de paramètre d'ajustement.

2.2.2.3. Résolution de l'équation de Poisson

L'équation de Poisson peut désormais être intégrée avec les conditions limites:

$$\begin{aligned} E(x = t_{\text{dep}}) &= 0 \\ \phi(x = 0) &= \phi_s \\ \phi(x = t_{\text{dep}}) &= 0 \end{aligned} \quad (2.30)$$

Ces limites permettent de déterminer les profondeurs de déplétion:

$$\begin{aligned} \text{(a) } t_{\text{dep1}} &= \sqrt{\frac{2 \cdot \epsilon_{\text{Si}} \cdot \phi_s}{q \cdot N_{\text{ch}}^*}} \\ \text{(b) } t_{\text{dep2}} &= \sqrt{\frac{q \cdot (N_{\text{ch}} - N_{\text{ch}}^*) X_j^2 + 2 \cdot \epsilon_{\text{Si}} \cdot \phi_s}{q \cdot N_{\text{ch}}}} \end{aligned} \quad (2.31)$$

La première expression sera utilisée si la déplétion est inférieure aux jonctions (**figure 2-7c**):

$$X_j \geq \sqrt{\frac{2 \cdot \epsilon_{\text{Si}} \cdot \phi_s}{q \cdot N_{\text{ch}}^*}} \quad (2.32)$$

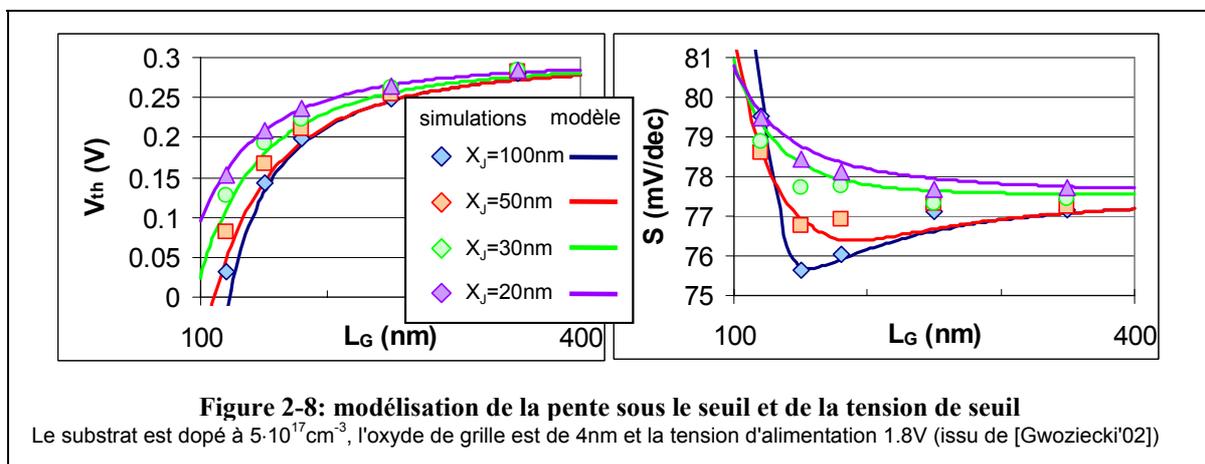
La seconde expression sera utilisée dans le cas contraire (**figure 2-7b**).

La charge dans le semi-conducteur est obtenue par simple intégration du profil de dopage:

$$\begin{aligned} \text{(a) } Q &= q \cdot t_{\text{dep}} \cdot N_{\text{ch}}^* \\ \text{(b) } Q &= q \cdot X_j \cdot N_{\text{ch}}^* + q \cdot (t_{\text{dep}} - X_j) \cdot N_{\text{ch}} \end{aligned} \quad (2.33)$$

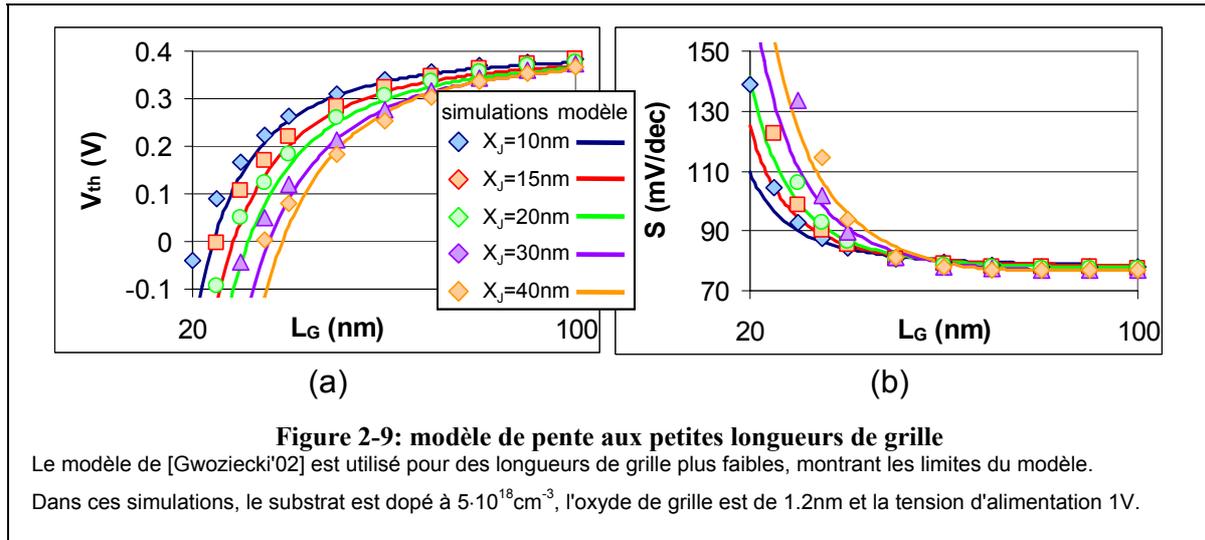
2.2.2.4. Validation du modèle

L'ensemble des équations (2.25) à (2.33) fournit les expressions analytiques de la pente sous le seuil et de la tension de seuil du transistor MOS [Gwoziecki'02]. La confrontation de ce modèle avec des résultats issus de simulations TCAD (**figure 2-8**) montre qu'il reproduit parfaitement le profil de tension de seuil, la légère amélioration puis très nette dégradation de la pente sous le seuil aux faibles longueurs de grille, ainsi que l'influence de la profondeur de jonction.

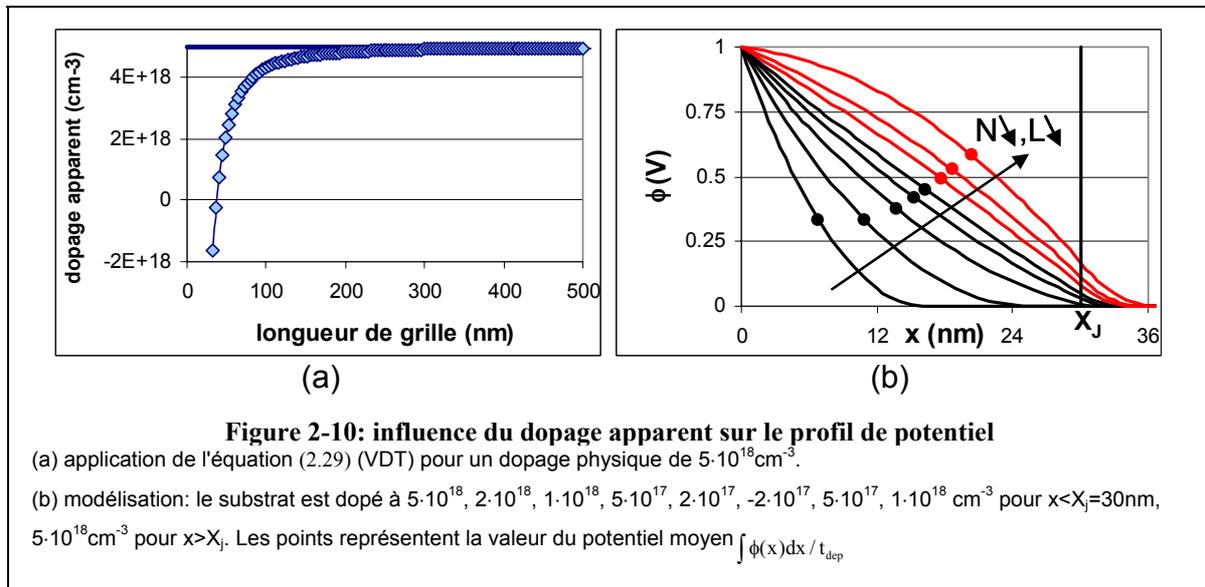


2.2.3. Amélioration du modèle de pente

Ce modèle a été validé pour des longueurs grilles descendant à 100nm. Or, les dispositifs actuels et futurs sont bien en deçà [ITRS'05]. Or, il apparaît que ce modèle n'est plus adéquat fiable dès lors que la longueur de grille devient inférieure à 50nm (**figure 2-9**). La chute de tension de seuil (SCE) est surestimée alors que l'augmentation de la pente est sous-estimée. Il est donc nécessaire d'améliorer le modèle pour les petites longueurs de grille.



La VDT montre que le dopage apparent diminue lorsque la longueur de grille diminue. Il peut même changer de signe (**figure 2-10a**).



Cette chute du dopage entraîne globalement une hausse du potentiel dans la structure (**figure 2-10b**). En particulier, le potentiel moyen défini par:

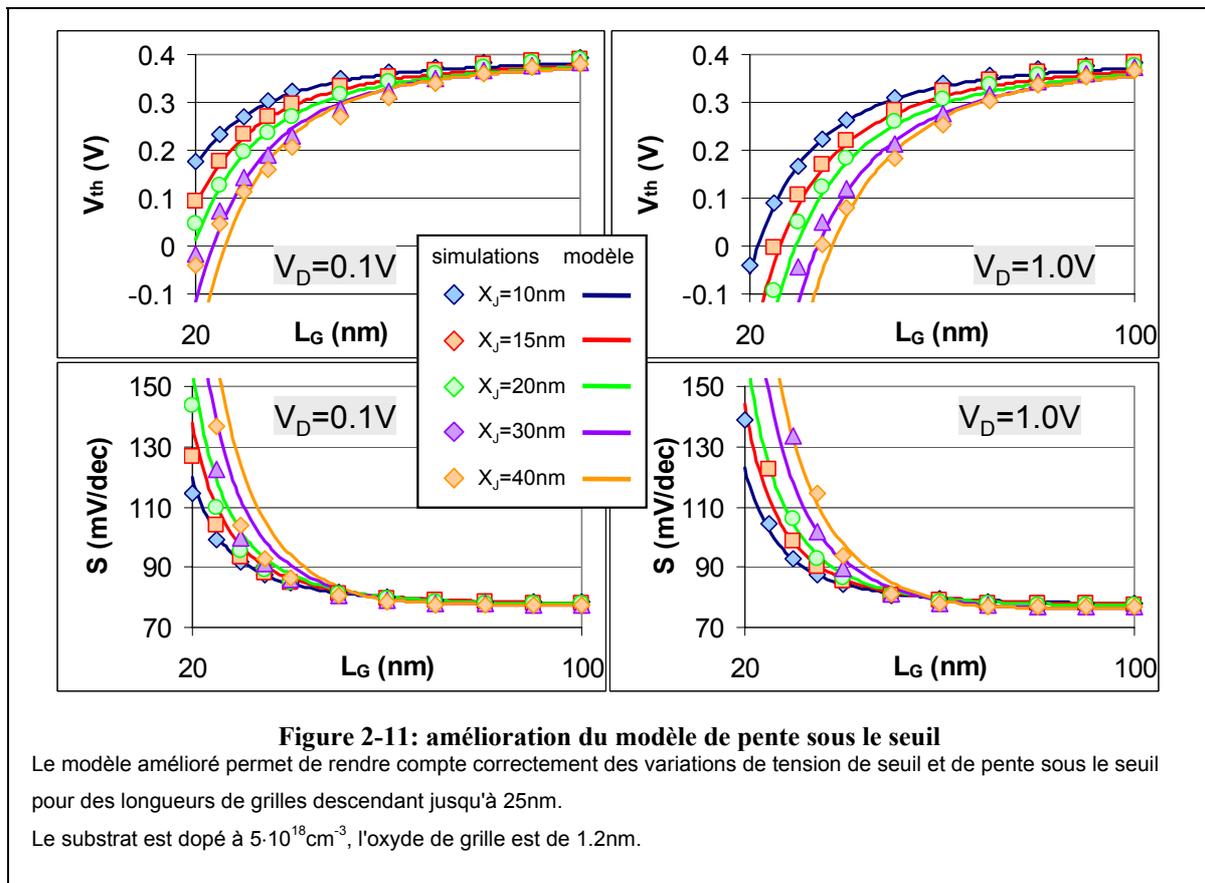
$$\phi_{\text{moyen}} = \frac{\int \phi(x) \cdot dx}{t_{\text{dep}}} \quad (2.34)$$

augmente également. Bien que $\alpha \cdot \phi_s$ et ϕ_{moyen} soient des quantités différentes, elles ont le même comportement: elles augmentent lorsque la longueur de grille diminue. Ceci peut s'exprimer pour le coefficient α par la relation empirique:

$$\alpha(L) = \alpha_{\infty} \cdot \left(1 + \frac{L_c}{L}\right) \quad (2.35)$$

où α_{∞} est la valeur de α pour les grandes longueurs de grille L_c est une longueur critique. Il est important de noter qu'en canal long, le dopage apparent est égal au dopage physique car le champ latéral est faible. La valeur α_{∞} ne doit donc pas être extraite en canal long.

Ce raffinement du modèle permet de rendre compte des variations de tension de seuil et de pente sous le seuil pour des longueurs de grilles descendant jusqu'à 25nm, que ce soit en régime linéaire ou en régime de saturation (**figure 2-11**).



2.2.4. Extension au transistor SOI complètement déplété

2.2.4.1. Approximation du profil de dopage apparent

Comme pour le transistor planaire, le dopage du film de silicium est remplacé par un dopage apparent N_{Si}^* constant dans le film de silicium:

$$N_{Si}^*(x) = N_{ch} - \frac{2 \cdot \epsilon_{Si}}{q \cdot L_{el}(x)^2} \left[V_{DS} + 2(V_{bi} - \alpha \cdot \phi_s) + 2\sqrt{(V_{bi} - \alpha \cdot \phi_s)(V_{DS} + V_{bi} - \alpha \cdot \phi_s)} \right] \quad (2.36)$$

L'oxyde ne contient pas de charges. Mais comme pour le silicium, il est possible d'introduire une charge apparente issue de la VDT. Contrairement au film de silicium, la longueur électrique dans l'oxyde enterré $L_{box}(x)$ n'est pas constante. Elle dépend de la profondeur. Cette différence peut être primordiale, notamment lorsque l'on considère un oxyde enterré épais (145nm) pour les longueurs de grille actuelles ($L < 50nm$). Deux cas particuliers peuvent être relevés:

- les lignes de champ sont circulaires (**figure 2-12a**):

$$\Delta L = \frac{\pi}{2} \cdot (x - t_{Si} - t_{box}) \quad (2.37)$$

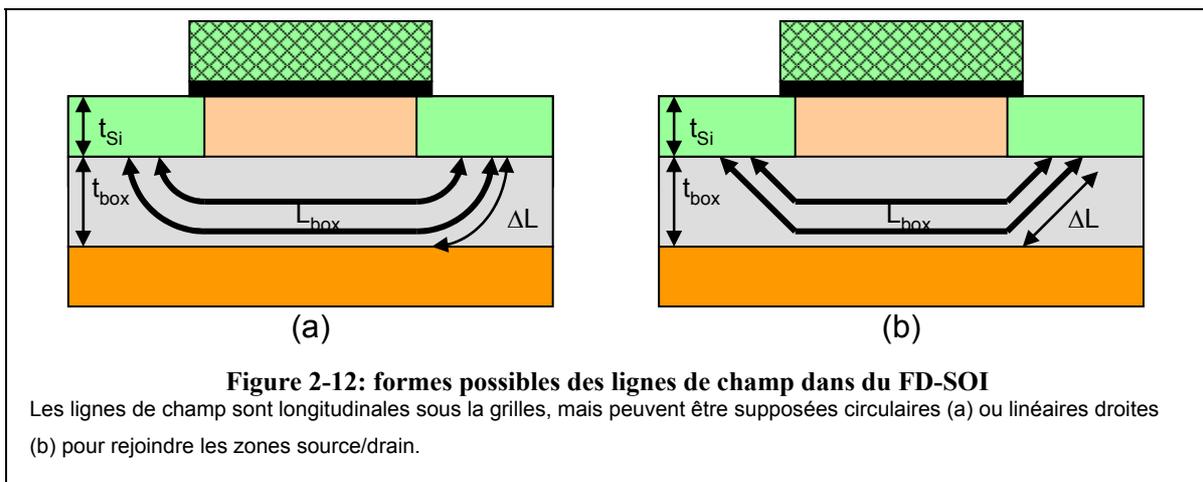
- les lignes de champ sont des droites à 45° (**figure 2-12b**):

$$\Delta L = \sqrt{2} \cdot (x - t_{Si} - t_{box}) \quad (2.38)$$

Dans ces deux cas la quantité ΔL est proportionnelle à la profondeur. Ceci peut être exprimée de façon générique par:

$$L_{box} = L + k_{box} \cdot t_{box} \quad (2.39)$$

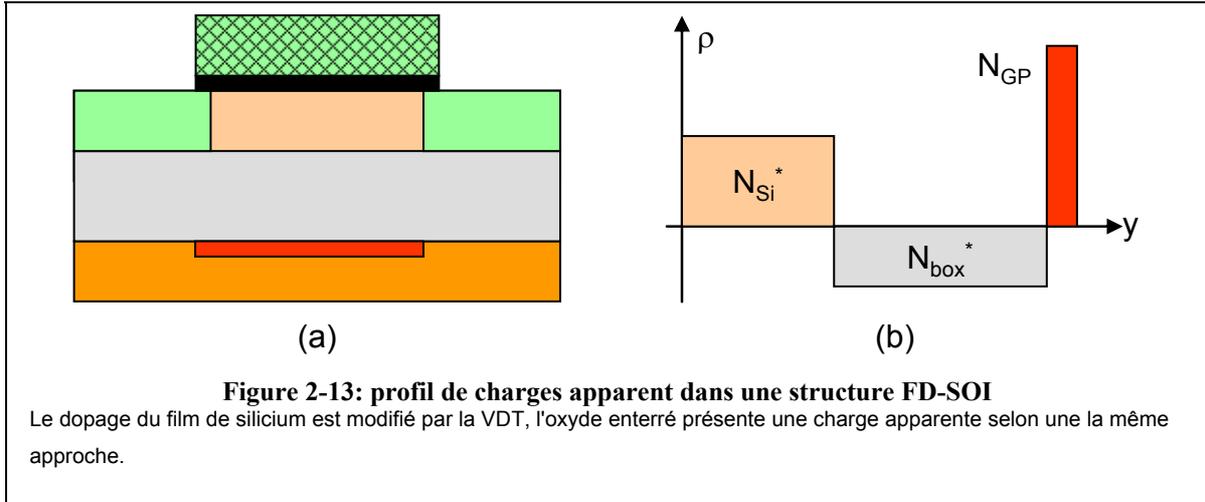
où k_{box} servira de paramètre d'ajustement.



La valeur du dopage apparent dans le box sera donc:

$$N_{box}^*(x) = 0 - \frac{2 \cdot \epsilon_{box}}{q \cdot L_{box}^2} \left[V_{DS} + 2(V_{bi} - \alpha \cdot \phi_s) + 2\sqrt{(V_{bi} - \alpha \cdot \phi_s)(V_{DS} + V_{bi} - \alpha \cdot \phi_s)} \right] \quad (2.40)$$

Ces considérations amènent à considérer le profil de charges présenté sur la **figure 2-13**.



Comme pour le transistor planaire, le paramètre α utilisé dans les expressions de N_{Si}^* et N_{box}^* s'exprime sous la forme:

$$\alpha(L) = \alpha_{\infty} \cdot \left(1 + \frac{L_c}{L} \right)$$

Les paramètres d'ajustement α_{∞} et L_c sont *a priori* différents pour le silicium et pour l'oxyde, et également différents des valeurs trouvées pour le transistor planaire.

2.2.4.2. Résolution de l'équation de Poisson

L'équation de Poisson peut être intégrée avec les conditions limites:

$$\begin{aligned} E(x = t_{dep}) &= 0 \\ \phi(x = 0) &= \phi_s \\ \phi(x = t_{dep}) &= 0 \end{aligned} \quad (2.41)$$

Les relations de continuité du déplacement électrique doivent également être utilisées:

$$\begin{aligned} \epsilon_{Si} \cdot \frac{\partial \phi}{\partial x} \Big|_{x \leq t_{Si}} &= \epsilon_{ox} \cdot \frac{\partial \phi}{\partial x} \Big|_{x \geq t_{Si}} \\ \epsilon_{ox} \cdot \frac{\partial \phi}{\partial x} \Big|_{x \leq t_{Si} + t_{box}} &= \epsilon_{Si} \cdot \frac{\partial \phi}{\partial x} \Big|_{x \geq t_{Si} + t_{box}} \end{aligned} \quad (2.42)$$

Ces limites permettent de déterminer la profondeur de déplétion:

$$t_{dep} = t_{box} \cdot \left(1 - \frac{\epsilon_{Si}}{\epsilon_{ox}} \right) + \sqrt{\left(t_{ox} \cdot \frac{\epsilon_{Si}}{\epsilon_{ox}} + t_{Si} \right) - \frac{N_{Si}^*}{N_{GP}} \cdot t_{Si}^2 - 2 \cdot \frac{N_{box}^*}{N_{GP}} \cdot t_{box} \cdot t_{Si} - \frac{N_{box}^*}{N_{GP}} \cdot t_{box}^2 \cdot \frac{\epsilon_{Si}}{\epsilon_{ox}} + \frac{2 \cdot \phi_s \cdot \epsilon_{Si}}{q \cdot N_{GP}}} \quad (2.43)$$

La charge dans le semi-conducteur vaut simplement:

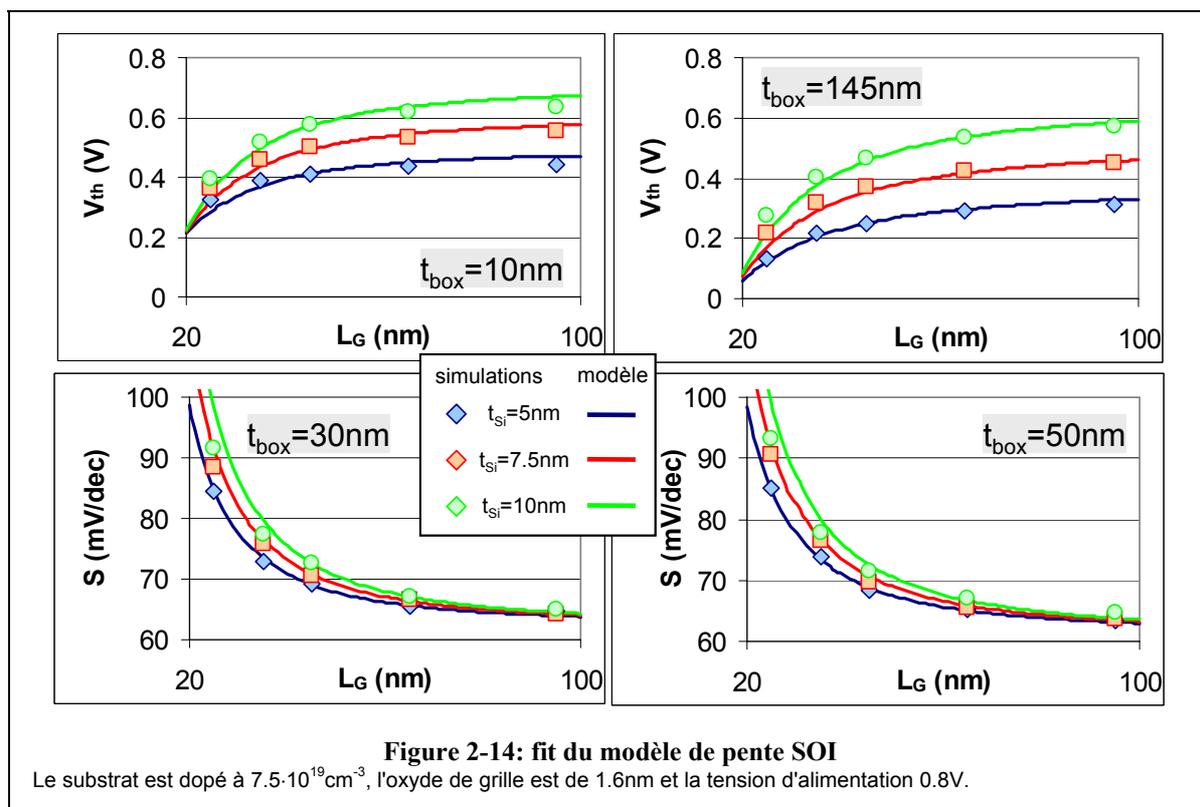
$$Q_{SC} = q \cdot t_{Si} \cdot N_{Si}^* + q \cdot t_{box} \cdot N_{box}^* + q \cdot N_{GP} \cdot (t_{dep} - t_{Si} - t_{box}) \quad (2.44)$$

2.2.4.3. Validation du modèle

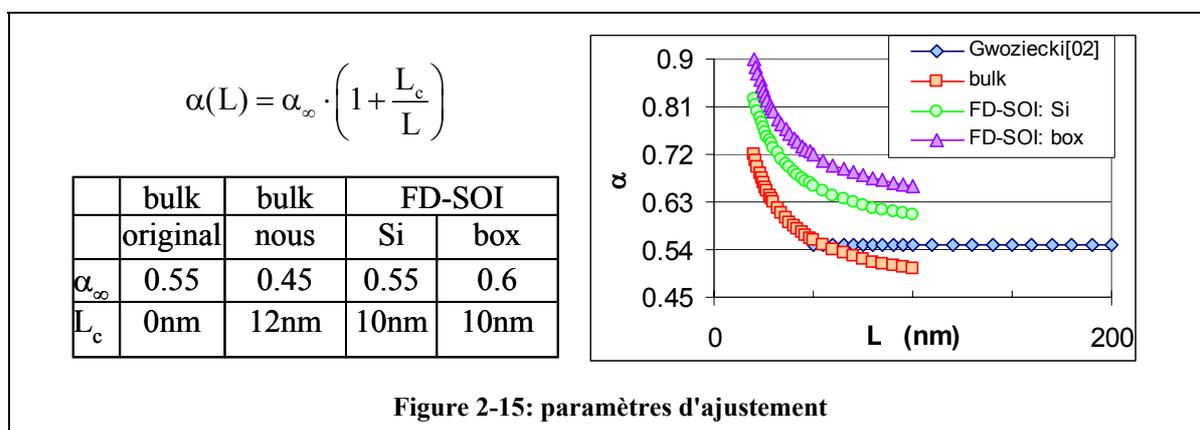
Le modèle est confronté aux simulations TCAD pour un dispositif dont le canal est dopé à $7.5 \cdot 10^{19} \text{cm}^{-3}$. Pour être complètement déplété, le film de silicium doit être inférieur à :

$$\sqrt{\frac{2 \cdot \phi_s \cdot \epsilon_{\text{Si}}}{q \cdot N_{\text{ch}}}} = 12 \text{nm} . \text{ L'oxyde de grille fait } 16 \text{\AA} \text{ et la tension d'alimentation } 0.8 \text{V} .$$

Le modèle arrive à reproduire tous les points de simulation pour une large plage d'épaisseurs d'oxyde enterré (10nm à 145nm) pour différentes épaisseurs de silicium (5 à 10nm). Ceci est vrai pour la tension de seuil, la pente, en régime linéaire ($V_D=0.1\text{V}$) ou saturé ($V_D=0.8\text{V}$).

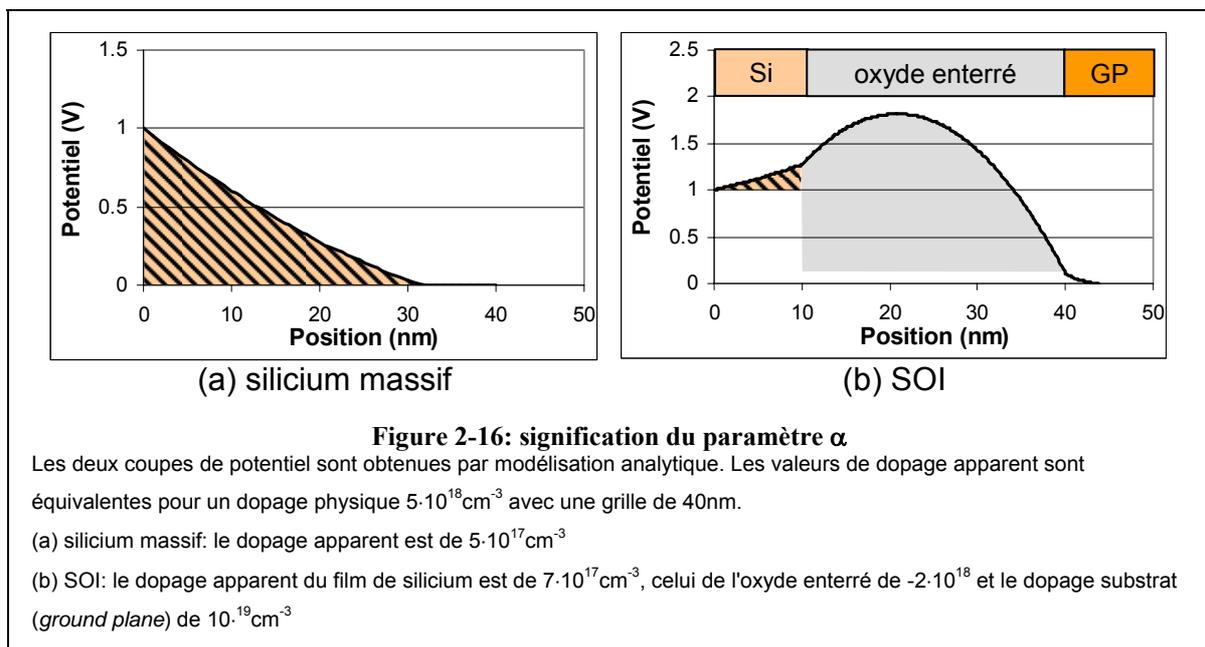


2.2.5. Discussion sur les paramètres d'ajustement



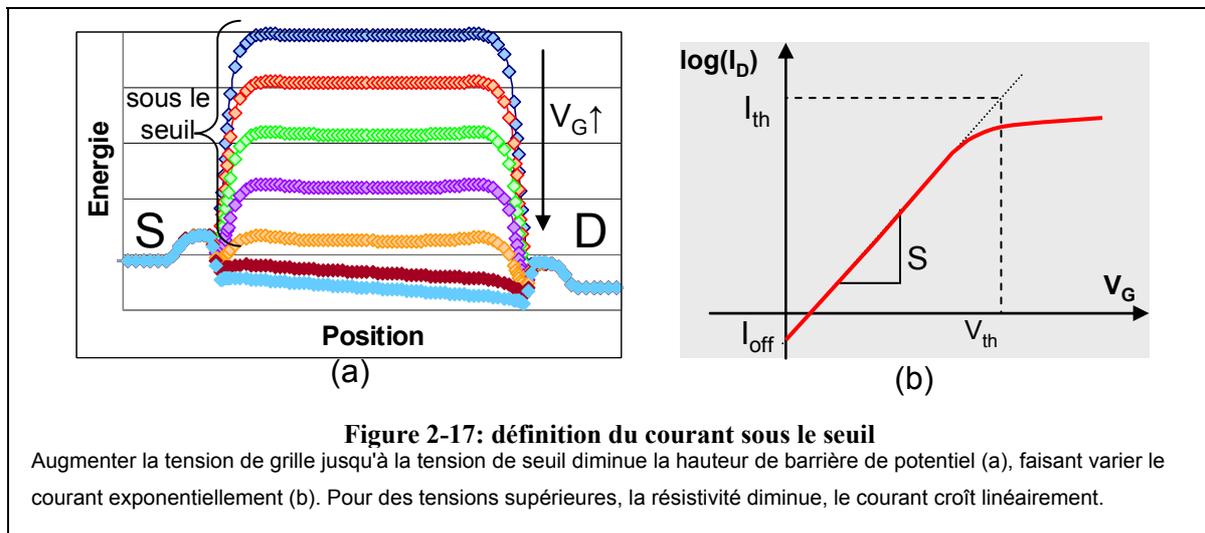
Si la même approche a été utilisée dans deux architectures complètement différentes, ce sont au total trois jeux de paramètres différents (α_∞ et L_c) qui sont nécessaires, tous différents du modèle de départ. La **figure 2-15** permet de relever un certain nombre de points intéressants pour les valeurs de α :

- la valeur utilisée dans [Gwoziecki'02] est une valeur moyennée sur les différentes longueurs. Elle permet d'être globalement bonne, mais pas de retranscrire les effets des plus courtes longueurs de grilles.
- les valeurs dans le *bulk* sont inférieures à celles dans le film de silicium de FD-SOI, elles même inférieures à celles du box dans le FD-SOI. En effet, le coefficient α représente une globalisation du potentiel. En substrat silicium massif, cela concerne l'ensemble du silicium où le potentiel varie de 0 à ϕ_s (surface hachurée sur la **figure 2-16a**). Avec un substrat SOI, dans le film de silicium, la variation de potentiel est bien moindre (surface hachurée sur la **figure 2-16b**).
- la longueur critique pour le FD-SOI est plus petite que pour le *bulk*, ce qui traduit la meilleure immunité aux effets canaux courts.



2.3. Modélisation du courant sous le seuil

Le courant en régime sous le seuil est dû aux porteurs qui peuvent passer par injection thermoélectronique par-dessus la barrière de potentiel créée par la grille. La **figure 2-17a** présente le profil de la bande de conduction en fonction de la tension de grille pour un transistor de type n. A tension de grille faible, (les courbes les plus hautes), le profil de la bande de conduction est horizontal, ce qui caractérise une conduction se fait par diffusion. L'augmentation de la tension de grille diminue la barrière de potentiel. Aux tensions de grille les plus élevées, la barrière est supprimée. La bande de conduction est linéairement décroissante de la source vers le drain, ce qui caractérise une conduction par dérivation.



On trouve dans la littérature [Skotnicki'00] l'expression du courant sous le seuil:

$$I_{sth} = I_{th} \cdot \exp\left(\frac{V_{GS} - V_{th}}{S} \cdot \ln(10)\right) \cdot \left[1 - \exp\left(\frac{-V_{DS}}{u_t}\right)\right] \quad (2.45)$$

qui traduit le fait que le courant de drain varie exponentiellement avec la tension de grille en régime sous le seuil (**figure 2-17b**). Une étude par simulation a montré que le courant I_{th} dépend essentiellement du dopage canal [Skotnicki'03]:

$$I_{th_new} = 5 \cdot 10^{-7} [A] \frac{W}{L} 8 \cdot 10^8 N_{ch}^{-0.4865} [cm^{-3}] \quad (2.46)$$

Cette partie vise à démontrer les relations les relations (2.45) et (2.46) en utilisant une approche physique analytique similaire à celle utilisée dans Mastar.

2.3.1. Principe de modélisation

Dans un transistor MOS de type n, les porteurs à l'origine du courant sont des électrons. En régime sous le seuil, une barrière de potentiel existe entre source et drain. Cette barrière

présente un potentiel constant, c'est-à-dire que le champ électrique est nul. Par conséquent, la composante de dérive du courant entre source et drain est nulle:

$$J_{\text{dérive}} = q \cdot \mu \cdot n \cdot E = 0 \quad (2.47)$$

où q est la charge élémentaire, μ est la mobilité des électrons, n leur densité et E le champ électrique. Le courant sous le seuil est donc un courant de diffusion:

$$J_{\text{diff}} = q \cdot D_n \cdot \nabla n \quad (2.48)$$

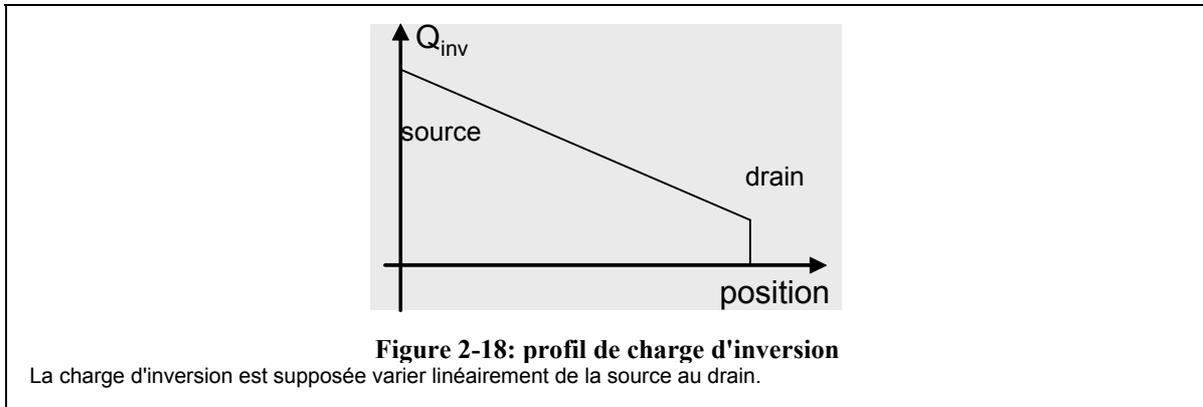
La constante de diffusion des électrons D_n peut être exprimée en fonction de la mobilité par la relation d'Einstein:

$$\frac{D_n}{\mu_n} = \frac{k_b \cdot T}{q} \quad (2.49)$$

Nous supposons que la densité de porteurs varie linéairement de la source au drain. Par conséquent, le gradient de densité de charges est constant et vaut:

$$\nabla n = \frac{Q_{\text{inv}}(\phi_{S,\text{source}}) - Q_{\text{inv}}(\phi_{S,\text{drain}})}{L} \quad (2.50)$$

où $Q_{\text{inv}}(\phi_{S,\text{source}})$ est la charge d'inversion à la source au potentiel de surface ϕ_S et $Q_{\text{inv}}(\phi_{S,\text{drain}})$ est la charge d'inversion au drain correspondante.



Les quantités de porteurs $Q_{\text{inv}}(\phi_{S,\text{source}})$ et $Q_{\text{inv}}(\phi_{S,\text{drain}})$ peuvent être calculées de deux façons différentes:

- en utilisant les expressions issues de l'étude de la capacité MOS
- en utilisant le produit densités de porteurs·fonction de distribution

Le courant sous le seuil I_{sth} est l'intégrale de la densité de courant sur la largeur W et la profondeur t :

$$I_{\text{sth}} = W \cdot t \cdot J_{\text{diff}} \quad (2.51)$$

t sera définie plus loin dans ce chapitre.

2.3.2. Approche capacité MOS

2.3.2.1. Calcul de la charge d'inversion

L'étude de la capacité MOS [Skotnicki] est rappelée en annexe. On y montre que la charge dans le semi-conducteur vaut:

$$Q_{sc} = \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{ch} \cdot \phi_s} \sqrt{\frac{u_t}{\phi_s} \cdot \exp\left(-\frac{\phi_s}{u_t}\right) + 1 - \frac{u_t}{\phi_s} + \frac{u_t}{\phi_s} \cdot \exp\left(\frac{\phi_s - 2 \cdot \phi_f}{u_t}\right) - \exp\left(\frac{-2 \cdot \phi_f}{u_t}\right) \cdot \left(1 + \frac{u_t}{\phi_s}\right)} \quad (2.52)$$

Cette expression peut être simplifiée sous la forme:

$$Q_{sc} \sim \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{ch} \cdot \phi_s} \sqrt{1 + \frac{u_t}{\phi_s} \cdot \exp\left(\frac{\phi_s - 2 \cdot \phi_f}{u_t}\right)} \quad (2.53)$$

et linéarisée:

$$Q_{sc} \sim \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{ch} \cdot \phi_s} \left[1 + \frac{\frac{u_t}{\phi_s} \cdot \exp\left(\frac{\phi_s - 2 \cdot \phi_f}{u_t}\right)}{2} \right] \quad (2.54)$$

Le premier terme $\sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot N_{ch} \cdot \phi_s}$ peut être identifié à la charge de déplétion. Il s'ensuit que la charge d'inversion vaut:

$$Q_{inv} \sim \sqrt{\frac{\epsilon_{Si} \cdot q \cdot N_{ch}}{2 \cdot \phi_s}} \cdot u_t \cdot \exp\left(\frac{\phi_s - 2 \cdot \phi_f}{u_t}\right) \quad (2.55)$$

2.3.2.2. Calcul du courant sous le seuil

Le potentiel de surface vaut ϕ_s à la source $\phi_s + V_{DD}$ au drain. Par conséquent, en utilisant les relations (2.51) et (2.48) à (2.50), le courant sous le seuil vaut:

$$I_{sth, capa} = \frac{W}{L} \cdot \mu_n \cdot u_t \cdot \sqrt{\frac{\epsilon_{Si} \cdot q \cdot N_{ch}}{2 \cdot \phi_s}} \cdot u_t \cdot \frac{n_i^2}{N_{ch}^2} \cdot \exp\left(\frac{\phi_s}{u_t}\right) \cdot \left(1 - \exp\left(\frac{-V_{DD}}{u_t}\right)\right) \quad (2.56)$$

2.3.3. Approche statistique

Pour estimer la quantité de porteurs en un point du canal, il suffit d'utiliser le produit densité d'états-fonction de distribution, puis d'intégrer.

2.3.3.1. Calcul de la densité de charges

Fonction de distribution

Les porteurs considérés ont une énergie supérieure à $\frac{k_b \cdot T}{q}$. Leur distribution peut donc être décrite par la fonction de Boltzmann:

$$f(E) = \exp\left(-\frac{E - E_{fe}}{k_b \cdot T}\right) \quad (2.57)$$

D'autre part, les électrons sous la grille ont une énergie [Mathieu'98]:

$$E = E_{fe} + q \cdot \phi + \frac{\hbar^2 \cdot k^2}{2 \cdot m^*} \quad (2.58)$$

Cette expression peut être développée:

$$E = E_{fe} + q \cdot \phi + \frac{\hbar^2}{2 \cdot m^*} (k_x^2 + k_y^2 + k_z^2) \quad (2.59)$$

Densité d'états

Dans un système à trois dimensions, la densité d'états dans l'espace des k s'écrit:

$$g(k) = \frac{2}{(2 \cdot \pi)^3} \quad (2.60)$$

Densité de charges

Il suffit de multiplier la densité d'états par la fonction de distribution et d'intégrer sur les k:

$$n_{\text{phys}}(k) = \iiint g(k) \cdot f(k) \cdot dk_x \cdot dk_y \cdot dk_z \quad (2.61)$$

En utilisant les expressions (2.61) et (2.60), on aboutit à:

$$n_{\text{phys}}(E) = \frac{2}{(2\pi)^3} \cdot \exp\left(-\frac{q \cdot \phi}{k_b \cdot T}\right) \cdot \int_{-\infty}^{+\infty} \exp\left(-\frac{\hbar^2 \cdot k_x^2}{2 \cdot m^* \cdot k_b \cdot T}\right) \cdot dk_x \cdot \int_{-\infty}^{+\infty} \exp\left(-\frac{\hbar^2 \cdot k_y^2}{2 \cdot m^* \cdot k_b \cdot T}\right) \cdot dk_y \cdot \int_{-\infty}^{+\infty} \exp\left(-\frac{\hbar^2 \cdot k_z^2}{2 \cdot m^* \cdot k_b \cdot T}\right) \cdot dk_z \quad (2.62)$$

Les trois intégrales sont de la forme $\int_{-\infty}^{+\infty} \exp(-a^2 \cdot x^2) \cdot dx = \frac{\sqrt{\pi}}{a}$. Donc:

$$n_{\text{phys}}(\phi) = \frac{2}{(2 \cdot \pi)^3} \exp\left(-\frac{\phi}{u_t}\right) \left(\frac{\pi \cdot 2 \cdot m \cdot k_b \cdot T}{\hbar^2}\right)^{\frac{3}{2}} \quad (2.63)$$

Cette expression peut être simplifiée et réécrite en fonction de $h = 2\pi\hbar$:

$$n_{\text{phys}}(E) = 2 \left(\frac{\pi \cdot 2 \cdot m^* \cdot k \cdot T}{h^2}\right)^{\frac{3}{2}} \exp\left(\frac{-\phi}{u_t}\right) \quad (2.64)$$

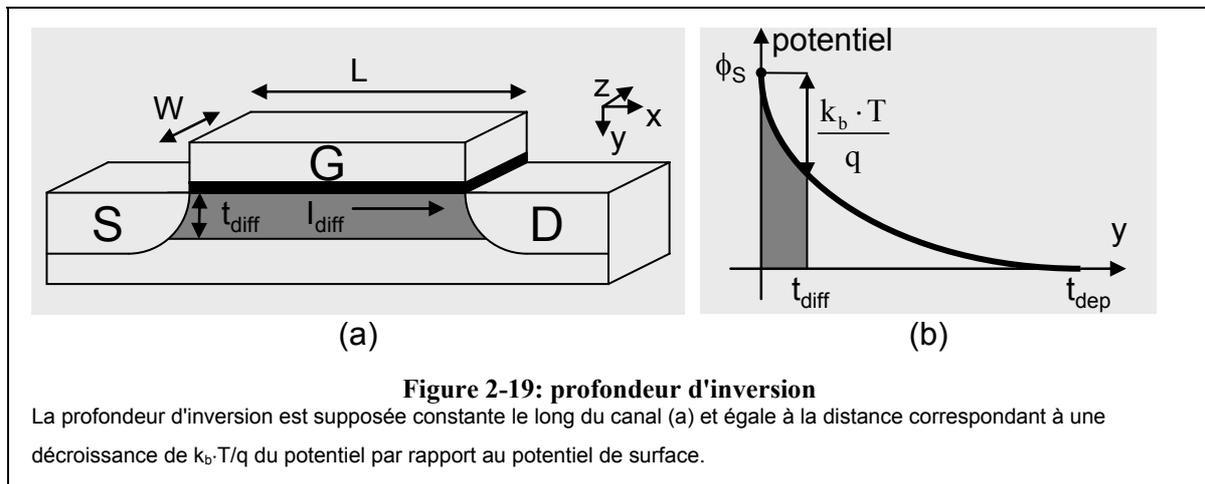
Cette quantité représente une densité de porteurs (cm^{-3}). Or, nous cherchons une charge d'inversion, c'est-à-dire une densité de porteurs intégrée sur la profondeur.

2.3.3.2. Estimation de la profondeur d'inversion

Nous allons considérer que la densité de porteurs n_{stat} est constante sur une profondeur t_{inv} constante (**figure 2-19a**). Cette dimension peut être estimée comme la profondeur sur laquelle le potentiel chute de $\frac{k_b \cdot T}{q}$ (**figure 2-19b**).

Au chapitre 1, nous avons utilisé l'équation de Poisson pour calculer la profondeur de déplétion. Cette méthode nous fournit la variation du potentiel:

$$\phi(x) = \frac{q \cdot N_{\text{ch}}}{2 \cdot \epsilon_s} \left(\sqrt{\frac{2 \cdot \epsilon_s \cdot \phi_s}{q \cdot N_{\text{ch}}}} - x \right)^2 \quad (2.65)$$



La profondeur d'inversion est donc solution de l'équation:

$$\phi_s - u_t = \frac{q \cdot N_{\text{ch}}}{2 \cdot \epsilon_s} \left(\sqrt{\frac{2 \cdot \epsilon_s \cdot \phi_s}{q \cdot N_{\text{ch}}}} - t_{\text{diff}} \right)^2 \quad (2.66)$$

soit:

$$t_{\text{diff}}(\phi_s) = \sqrt{\frac{2 \cdot \epsilon_s \cdot \phi_s}{q \cdot N_{\text{ch}}}} \left(1 - \sqrt{1 - \frac{u_t}{\phi_s}} \right) \quad (2.67)$$

en considérant $u_t \ll \phi_s$, on peut simplifier cette expression:

$$t_{\text{diff}}(\phi_s) = u_t \cdot \sqrt{\frac{\epsilon_s}{2 \cdot q \cdot N_{\text{ch}} \cdot \phi_s}} \quad (2.68)$$

2.3.3.3. Calcul du courant sous le seuil

Charge d'inversion

La charge d'inversion est le produit de la densité de charge par la profondeur de la couche d'inversion:

$$Q_{\text{inv_stat}}(\phi, \phi_s) = t_{\text{inv}}(\phi_s) \cdot n_{\text{stat}}(\phi) \quad (2.69)$$

Gradient de charges

La **figure 2-20** nous permet de déterminer l'énergie des porteurs à la source et au drain:

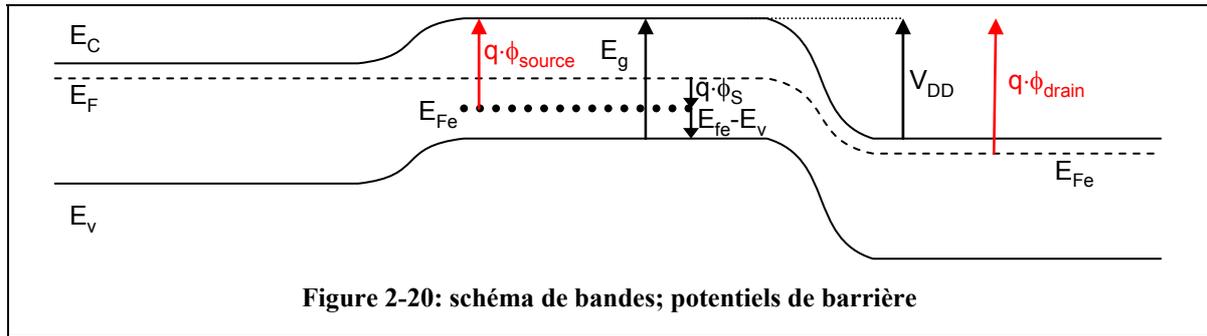


Figure 2-20: schéma de bandes; potentiels de barrière

$$q \cdot \phi_{\text{source}} = E_g - q \cdot \phi_s - (E_{\text{fe}} - E_v) \quad (2.70)$$

$$q \cdot \phi_{\text{drain}} = E_g - q \cdot \phi_s - (E_{\text{fe}} - E_v) + V_{\text{DD}}$$

Les valeurs du gap et la position du quasi-niveau de Fermi sont données par:

$$E_g = u_t \cdot \ln\left(\frac{N_c \cdot N_v}{n_i^2}\right) \quad (2.71)$$

$$E_{\text{fe}} - E_v = u_t \cdot \ln\left(\frac{N_v}{N_a}\right) \quad (2.72)$$

Le courant sous le seuil peut être calculé à partir des relations (2.51), (2.48), (2.49), (2.50), (2.69), (2.68), (2.64), (2.70), (2.71) et (2.57):

$$I_{\text{sth,phys}} = \frac{W}{L} \cdot q \cdot \mu_n \cdot u_t \cdot u_t \cdot \sqrt{\frac{\epsilon_s}{2 \cdot q \cdot N_{\text{ch}} \cdot \phi_s}} \cdot 2 \cdot \left(\frac{\pi \cdot 2 \cdot m \cdot k_b \cdot T}{h^2}\right)^{\frac{3}{2}} \cdot \frac{n_i^2}{N_{\text{ch}} \cdot N_c} \cdot \exp\left(\frac{\phi_s}{u_t}\right) \left(1 - \exp\left(\frac{-V_{\text{DD}}}{u_t}\right)\right) \quad (2.73)$$

2.3.4. Comparaison des modèles

Les deux approches utilisées (capacité MOS et statistique) aboutissent à des expressions similaires:

$$I_{\text{sth}} = A(N_{\text{ch}}) \cdot \frac{W}{L} \cdot \frac{1}{\sqrt{\phi_s}} \cdot \exp\left(\frac{\phi_s}{u_t}\right) \cdot \left(1 - \exp\left(\frac{-V_{\text{dd}}}{u_t}\right)\right) \quad (2.74)$$

où $A(N_{ch})$ est une valeur qui dépend du dopage substrat N_{ch} . Les deux approches donnent des valeurs de A différentes de 12% dues aux différentes approximations, quelle que soit la valeur de N_{ch} .

2.3.5. Relation avec la tension de grille

Ces expressions relient le courant sous le seuil au potentiel de surface. Il reste donc à faire le lien avec la tension de grille.

L'expression de la tension de seuil est:

$$S = \frac{k_b \cdot T}{q} \cdot \ln(10) \cdot \frac{dV_G}{d\phi_s} \quad (2.75)$$

Par conséquent, le potentiel de surface vaut:

$$\phi_s = \frac{k_b \cdot T}{q} \cdot \frac{\ln(10)}{S} \cdot V_G + V_0 \quad (2.76)$$

La constante V_0 peut être déterminée grâce au cas particulier $V_G = V_{th}$. Dans ce cas, $\phi_s = \phi_{s,th}$:

$$\phi_s = \frac{k_b \cdot T}{q} \cdot \frac{\ln(10)}{S} \cdot (V_G - V_{th}) + \phi_{s,th} \quad (2.77)$$

En remarquant que $V_G - V_{th} < \phi_{s,th}$, on peut raisonnablement faire l'approximation:

$$\frac{1}{\sqrt{\phi_s}} \sim \frac{1}{\sqrt{\phi_{s,th}}} \quad (2.78)$$

En injectant ces relations dans les équations (2.56) et (2.73), on aboutit à:

$$I_{sth} = I_{th} \cdot \exp\left(\frac{V_{GS} - V_{th}}{S} \cdot \ln(10)\right) \cdot \left[1 - \exp\left(\frac{-V_{DS}}{u_t}\right)\right] \quad (2.79)$$

L'expression de I_{th} dépend de l'approche utilisée. Pour l'approche capacité MOS, elle vaut:

$$I_{th, \text{capa}} = \frac{W}{L} \cdot \mu_n \cdot u_t \cdot \sqrt{\frac{\epsilon_{Si} \cdot q \cdot N_a}{2 \cdot \phi_{Sth}}} \cdot u_t \cdot \frac{n_i^2}{N_a} \cdot \exp\left(\frac{\phi_{Sth}}{u_t}\right) \quad (2.80)$$

Pour l'approche statistique, elle vaut:

$$I_{th, \text{stat}} = \frac{W}{L} \cdot q \cdot \mu_n \cdot u_t \cdot u_t \cdot \sqrt{\frac{\epsilon_S}{2 \cdot q \cdot N_{ch} \cdot \phi_{Sth}}} \cdot 2 \cdot \left(\frac{\pi \cdot 2 \cdot m \cdot k_b \cdot T}{h^2}\right)^{\frac{3}{2}} \cdot \frac{n_i^2}{N_{ch} \cdot N_c} \cdot \exp\left(\frac{\phi_{s,th}}{u_t}\right) \quad (2.81)$$

2.3.6. Discussion et interprétations

La relation (2.79) issue de cette modélisation est identique à la relation (2.45) trouvée dans la littérature. Les relations (2.80) et (2.81) peuvent être calibrées sur les données de simulations données par la relation (2.46):

$$I_{th_new} = 5.10^{-7} [A] \frac{W}{L} 8.10^8 N_{ch}^{-0.4865} [cm^{-3}]$$

En fait, le paramètre sur lequel nous pouvons jouer est le potentiel de surface au seuil ϕ_{Sth} . Nous avons signalé au chapitre 1 que la définition usuelle de la tension de seuil est $\phi_S = 2 \cdot \phi_F$. Or, compte tenu de l'expression:

$$\phi_F = \frac{k_b \cdot T}{q} \cdot \ln \left(\frac{N_{ch}}{n_i} \right) \quad (2.82)$$

les relations (2.81) peuvent s'écrire sous la forme:

$$I_{th} = A \cdot \sqrt{N_a} \quad (2.83)$$

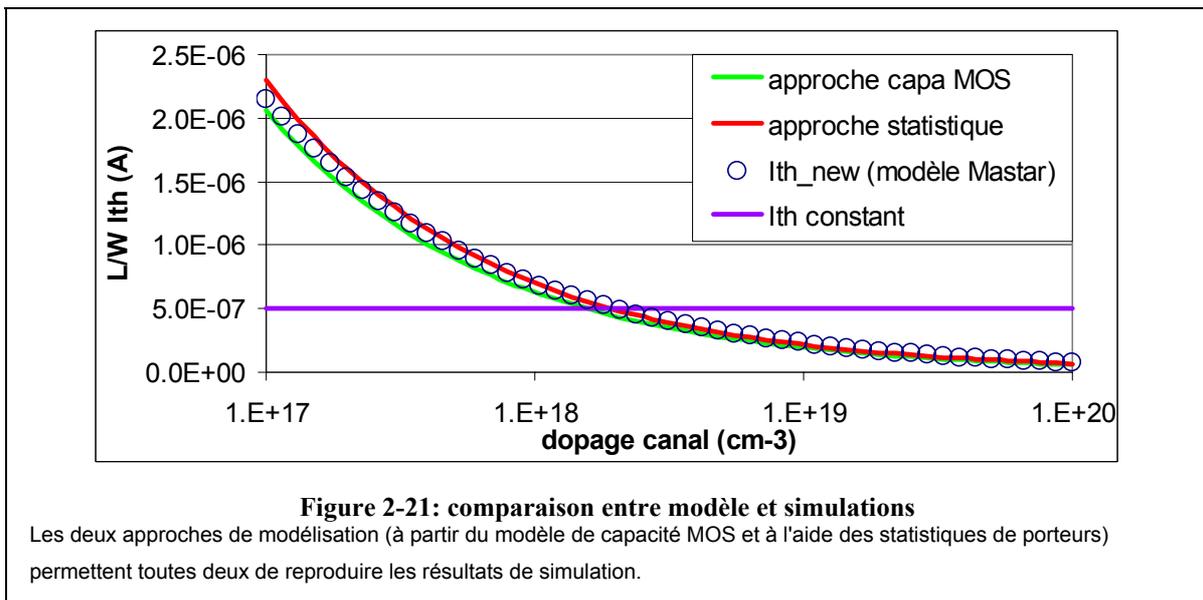
ce qui ne semble pas correspondre aux simulations numériques.

Nous avons également rappelé que la définition de la tension de seuil $\phi_s = V_{bi}$ est de plus en plus utilisée, notamment parce qu'elle est vérifiée pour les technologies planaire, SOI et DG. C'est cette définition qui est utilisée dans le modèle Mastar. En utilisant cette définition, l'expression du courant de seuil devient:

$$I_{th} = \frac{A}{\sqrt{N_{ch}}} \quad (2.84)$$

ce qui est très proche des résultats de simulations numériques.

La relation qui donne la meilleure adéquation entre modèle et simulations (**figure 2-21**) est:



$$\phi_{s,th} = V_{bi} - \frac{k_b \cdot T}{q} \quad (2.85)$$

Le niveau de dopage canal en technologie planaire est de l'ordre de quelques 10^{18}cm^{-3} . La valeur approchée $5 \cdot 10^{-7} \text{A}/\mu\text{m}$ est donc tout à fait utilisable.

Cette relation est parfaitement cohérente avec notre définition de la tension de seuil. Elle indique que l'on peut définir le seuil comme le moment où la barrière de potentiel devient égale à l'énergie thermique des porteurs.

Notons que la définition de la tension de seuil utilisée dans Mastar est $\phi_{Sth}=V_{bi}$. Cette différence de $k_b \cdot T/q$ est acceptable, compte tenu des différentes approximations faites sur le modèle.

Conclusion du chapitre

La modélisation du transistor MOS au chapitre 1 a montré que la pente sous le seuil est intrinsèquement limitée. Cette limitation est due à la statistique de porteurs en $\exp(k_b \cdot T/q)$. La première partie de ce chapitre a montré que cette limitation de la quantité $dV/d\log(I)$ est commune aux autres dispositifs semi-conducteurs élémentaires: la diode et le transistor bipolaire.

Dans la seconde partie de ce chapitre, la pente sous le seuil du transistor MOS conventionnel est modélisée. Cette modélisation est basée sur la transformation tension dopage (VDT), qui permet d'intégrer les effets de champ latéral dans la modélisation 1D usuelle. Cette approche déjà présentée dans la littérature est améliorée pour reproduire plus précisément le comportement des transistors les plus courts. Cette modélisation est également étendue au transistor SOI complètement déplété.

Dans la troisième et dernière partie de ce chapitre, le courant sous le seuil du transistor MOS est modélisé. Deux approches sont proposées: la première utilise le calcul de la capacité MOS planaire, alors que la seconde reprend les équations de statistique des porteurs. Les deux approches donnent des résultats identiques, reproduisant les résultats de simulations TCAD.

Ce chapitre a développé la modélisation du transistor MOS en régime sous le seuil. L'établissement d'expressions analytiques simples permet de déterminer les leviers efficaces pour l'amélioration des performances: optimisation des profondeurs de jonctions en technologie planaire, épaisseur d'oxyde enterré en technologie SOI.

Cependant, la limitation fondamentale de la pente sous le seuil ne peut pas être passée avec des dispositifs de type MOSFET. Le chapitre suivant se propose de passer en revue un certain nombre de dispositifs pouvant passer sous cette limite.

Chapitre 3. Dispositifs présentant une pente meilleure que 60mV/dec

Le chapitre précédent a montré que la pente sous le seuil des dispositifs électroniques usuels (diode, transistors MOS et bipolaire) est intrinsèquement limitée à 60mV/décade à température ambiante. Afin de s'affranchir de cette limite, il est par conséquent nécessaire d'envisager un autre dispositif basé sur un autre principe de fonctionnement. Dans ce chapitre, les structures de la littérature présentant une caractéristique électrique proche de celle du MOS, mais avec une pente inférieure à 60mV/dec, sont passées en revue. Elles peuvent être divisées en trois catégories:

- les structures de rupture, qui utilisent de nouveaux matériaux (à transition de Mott par exemple), ou d'autres effets (mécanique par exemple)
- les structures utilisant l'effet tunnel: en effet, la conduction par effet tunnel est différente de la conduction par diffusion, ce qui permet de s'affranchir de la limite usuelle de pente
- les structures utilisant un phénomène d'amplification, comme le gain d'une structure bipolaire ou le phénomène d'ionisation par impact.

Introduction

Le courant électrique résulte d'un déplacement de charges. Dans les dispositifs usuels, il est proportionnel à la charge présente dans le dispositif. Celle-ci obéit à une loi de Fermi-Dirac, qui à température ambiante peut être approximée par une statistique de Boltzmann. Cela implique la limite fondamentale de 60mV/dec pour tous les dispositifs courants (diode, transistor bipolaire, transistor MOS), et ce quel que soit le matériau employé.

La littérature présente d'autres dispositifs présentant des caractéristiques électriques comparables à celles du transistor MOS traditionnel, mais avec une pente sous le seuil inférieure à 60mV/dec. Différents moyens sont utilisés:

- changer le phénomène physique à l'origine de la commutation
- utiliser un phénomène d'amplification
- utiliser un autre phénomène de transport, comme l'effet tunnel.

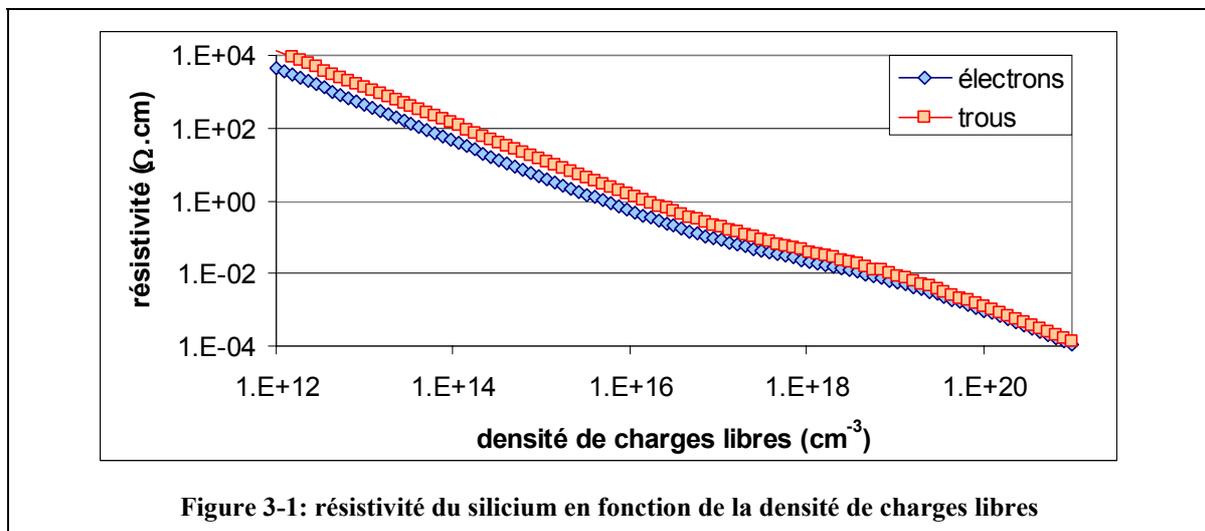
Ce chapitre propose d'évaluer leur potentiel, à la fois en terme de performances et de simplicité d'intégration. En effet, une compatibilité, voire une possible co-intégration avec la logique CMOS actuelle, est un avantage prépondérant.

3.1. Les structures de rupture

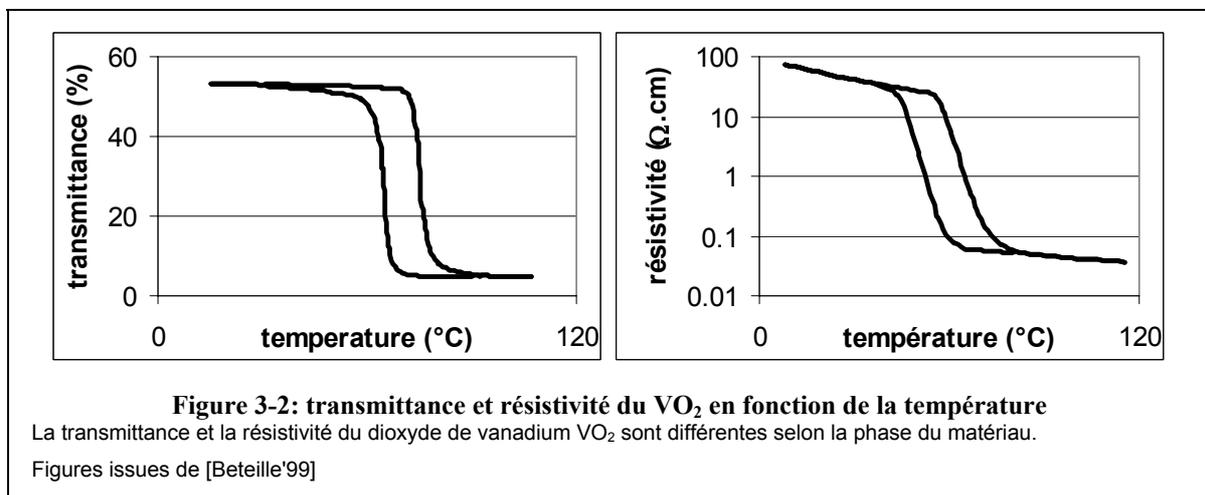
3.1.1. Matériaux à transition de Mott

3.1.1.1. Introduction aux matériaux

On peut considérer l'effet transistor comme la modulation de la conductivité d'un matériau. Cette modulation peut être obtenue par une tension (transistor à effet de champ) ou un courant (transistor bipolaire). Le matériau présente donc les propriétés d'un isolant (régime de blocage) ou d'un métal (régime de conduction). Dans le cas des semi-conducteurs, c'est la densité de charges libres qui est à l'origine de la différence métal/isolant (**figure 3-1**).



Cependant, il existe d'autres types de matériaux dont la résistivité peut varier sur une large plage. On parlera alors de transition métal-isolant [Mott'90]. En particulier, les matériaux à transition de Mott (du nom du chercheur qui obtiendra le prix Nobel pour ses travaux sur ce sujet) présentent cette caractéristique. Pour exemple, le dioxyde de vanadium VO_2 dispose de phases dont les propriétés électriques et optiques sont très différentes (**figure 3-2**). Cette



propriété de disposer d'une phase isolante et d'une phase conductrice permet de concevoir un dispositif de type transistor. Le phénomène physique à l'origine de la modulation de la conductivité étant différent de celui du MOS conventionnel, la pente sous le seuil d'un tel dispositif n'est pas soumise à la limite de 60mV/dec.

3.1.1.2. Procédé d'intégration

La première étape consiste à réaliser un film de matériau à transition de Mott, un oxyde de vanadium. Selon les recuits, il peut se présenter sous différentes stochiométries, dont les plus courantes sont le pentoxyde de vanadium V_2O_5 et le dioxyde de vanadium VO_2 . En outre, dans un souci de compatibilité avec la microélectronique actuelle, il est intéressant de conserver un substrat silicium. L'oxyde de vanadium est isolé du substrat par une couche de silice réalisée par oxydation thermique du silicium.

Le matériau est déposé par procédé sol-gel [Livage'00]. Ce procédé a été choisi pour sa simplicité et son efficacité. Un alcoxyde (du vanadium oxy isopropoxide) contenant des ions vanadium est déposé à la surface de la plaque (la couche supérieure étant de la silice). Il est ensuite réparti de façon plus homogène par rotation du substrat (*spin-coating*).

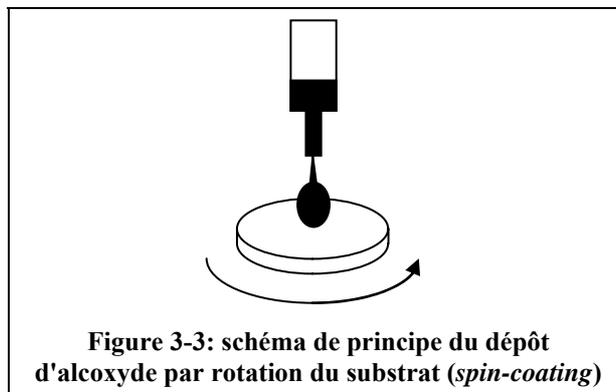


Figure 3-3: schéma de principe du dépôt d'alcoxyde par rotation du substrat (*spin-coating*)

3.1.2. Résultats morphologiques

Un premier recuit (650°C 10min) aboutit à la formation de cristaux (**figure 3-5**). Une vue plus détaillée au microscope électronique à balayage (MEB) permet d'évaluer la taille de ces cristaux: une centaine de nanomètres environ (**figure 3-4**). Cependant, ce dépôt n'est pas de suffisamment bonne qualité pour pouvoir être testé électriquement. Il convient donc de l'améliorer.

Un recuit à température moindre (500°C 12min) sur un autre échantillon a été réalisé. On observe un début de coalescence des cristaux, indiquant une organisation de surface plus

propice à la continuité électrique. Cependant, le matériau déposé est discontinu, ce qui rend impossible une caractérisation électrique.

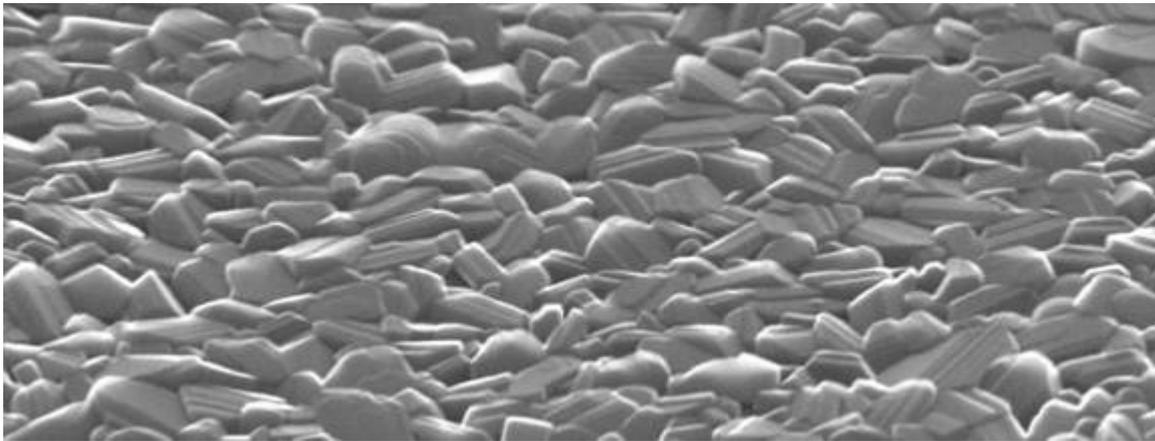


Figure 3-5: image MEB d'un dépôt après un recuit à 650°C

Des cristaux discontinus se forment à la surface de l'échantillon.

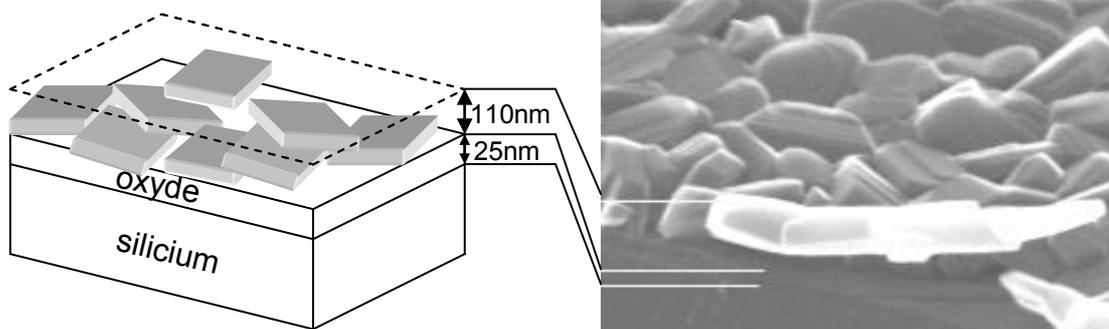


Figure 3-4: dimensions des cristaux

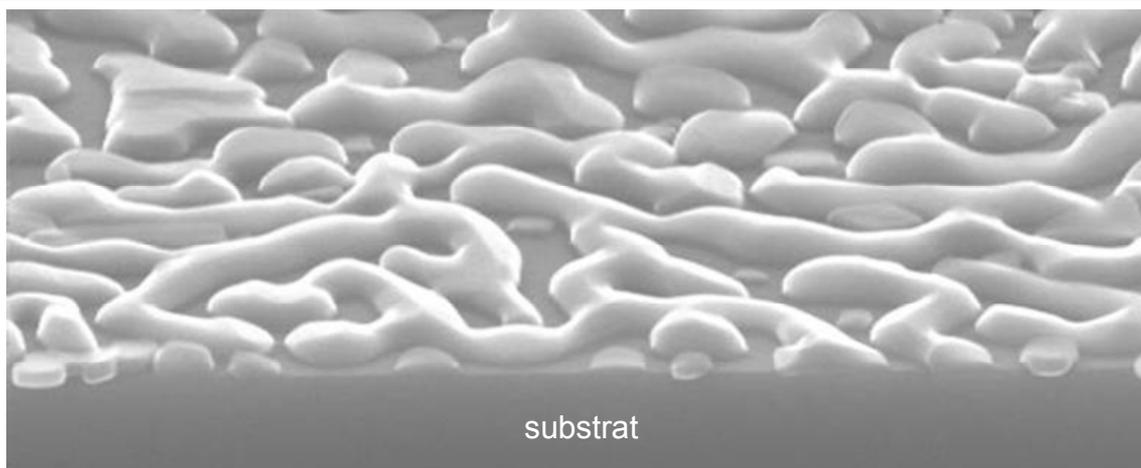


Figure 3-6: image MEB du dépôt après un recuit à 500°C

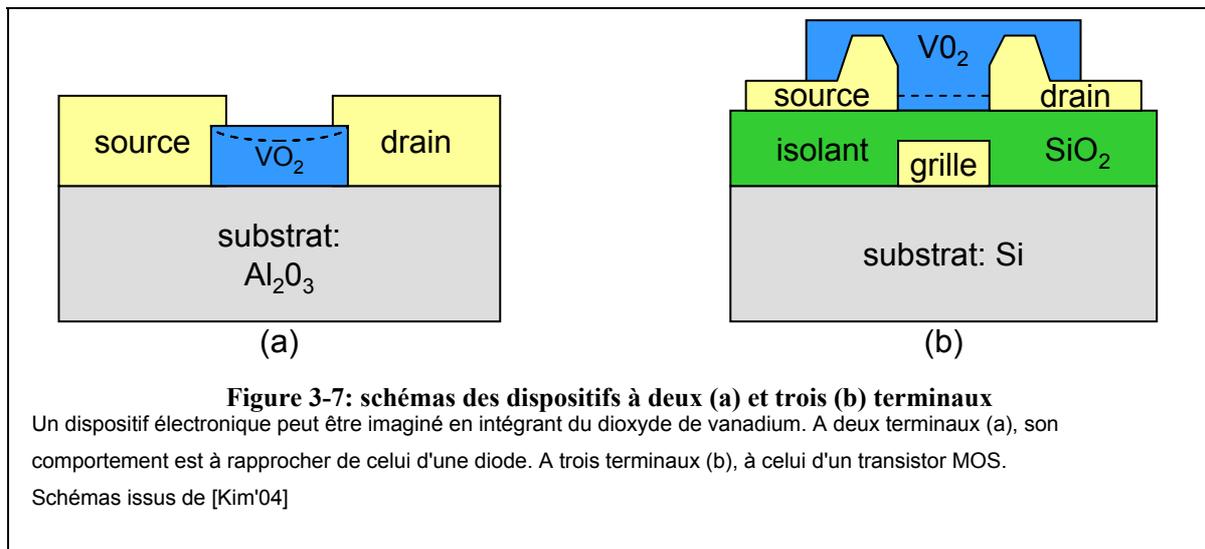
On peut observer un début de coalescence des cristaux.

Le meilleur dépôt est obtenu pour une température de recuit plus basse (500°C, **figure 3-6**), ce qui est en accord avec les résultats que l'on peut trouver dans la littérature

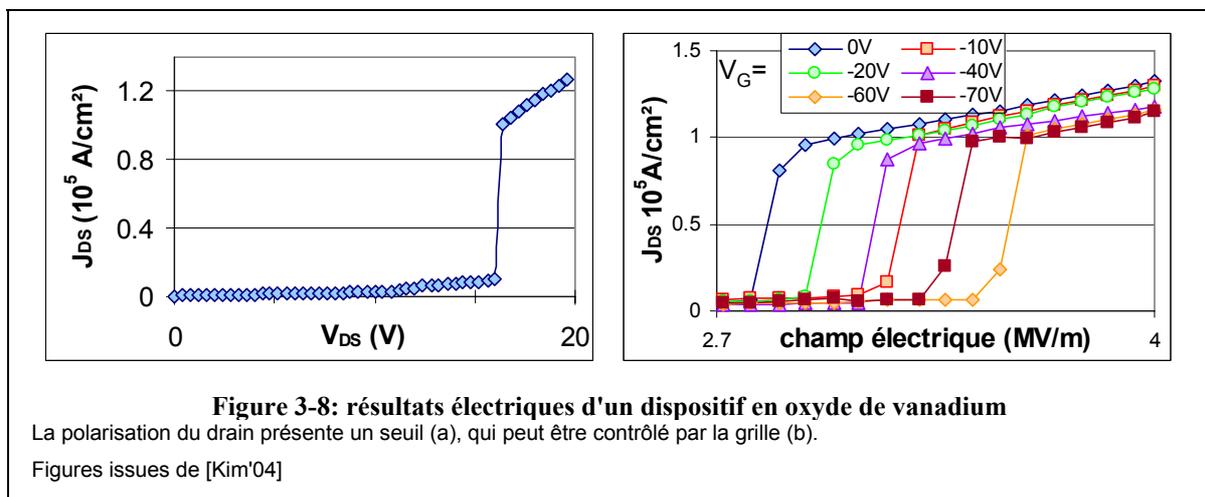
[Beteille'99,Livage'01]. Bien qu'une température de recuit plus faible associée à un temps plus long ait donné un meilleur résultat, baisser encore la température ou augmenter le temps de recuit n'a pas permis d'obtenir un film continu. Il est probable que la coalescence des îlots a vidé le seuil de percolation.

3.1.2.1. Un dispositif à commutation

Des dispositifs à deux ou trois terminaux ont été réalisés [Kim'04] (**figure 3-7**) afin de servir d'outil de caractérisation des matériaux à transition de Mott. Cependant, le dispositif à trois terminaux peut être assimilé à un transistor.

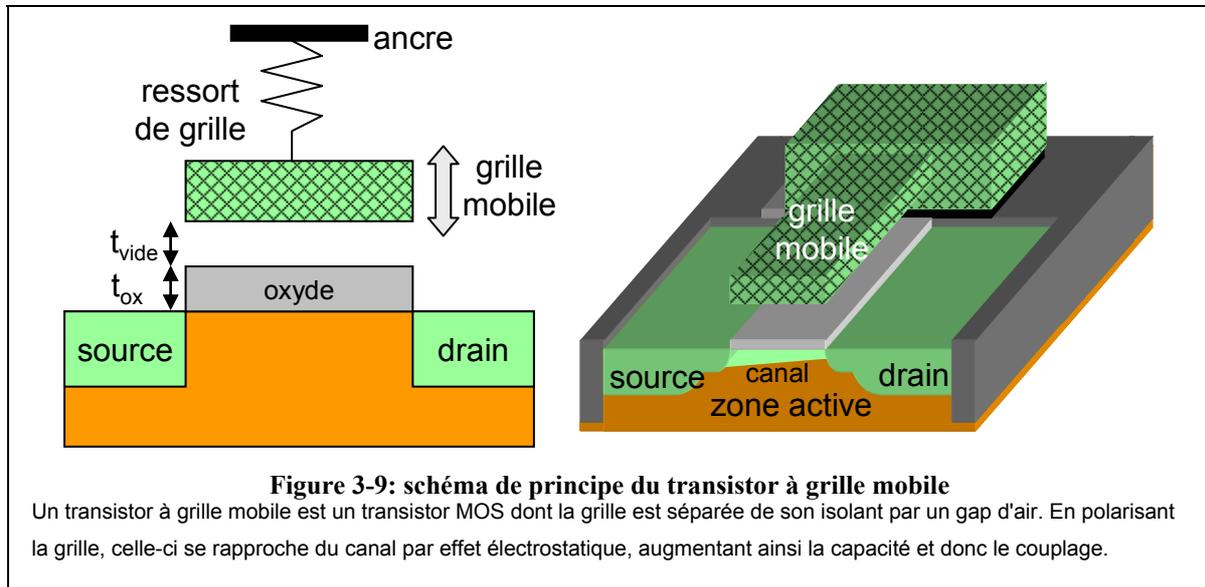


Les dispositifs obtenus montrent bien que le courant électrique est modulé par la grille (**figure 3-8**), signe d'un effet de transition. Cependant, le saut en courant est de l'ordre d'une décade, ce qui ne constitue pas encore une performance à la hauteur des transistors MOS.



3.1.3. Transistor à grille mobile

3.1.3.1. Principe



Le principe, qui date d'une quarantaine d'années [Newell'66], est d'intercaler une couche de vide entre la grille en poly silicium et l'oxyde de grille. La grille se présente donc comme une poutre suspendue. La capacité de grille se décrit alors en deux composantes: la capacité associée à l'oxyde C_{ox} , et la capacité liée à l'air C_{vide} . La capacité de grille est donc:

$$C_G = \frac{1}{\frac{1}{C_{ox}} + \frac{1}{C_{vide}}} \quad (4.1)$$

Si la poutre est suffisamment flexible pour être mobile, ses déplacements entraînent une variation de la capacité $C_{vide} = \epsilon_0 / t_{vide}$, et donc de la capacité de grille. Par conséquent, la formule de la pente sous le seuil doit prendre en compte cette nouvelle dépendance et devient donc:

$$S = \frac{k_b \cdot T}{q} \cdot \ln(10) \cdot \left(1 - \frac{\partial}{\partial \phi_s} \frac{Q_{sc}}{C_G} \right) \quad (4.2)$$

Lorsque l'on polarise la poutre (grille), celle-ci est attirée vers le canal par force électrostatique. Cet effet contribue à faire diminuer l'épaisseur de l'isolant de grille et à renforcer l'efficacité du contrôle de la grille sur le canal. Des études basées sur des simulations numériques et du modèle analytique [Abele'05b, Ionescu'02] ont permis de montrer que sous certaines conditions, un tel dispositif pourrait permettre d'obtenir une pente inférieure à 60mV/dec. Ceci fut également démontré expérimentalement [Abele'05a], les dispositifs obtenus montrant une pente de 2mV/dec.

3.1.3.2. Intégration

Ce dispositif appartient à la famille des MEMS (*Micro Electrical Mechanical Systems*). Cette famille compte notamment des résonateurs mécaniques permettant de filtrer les fréquences dans les applications de type téléphonie mobile. Ils sont alors réalisés *above IC*, c'est-à-dire au-dessus des niveaux d'interconnexions des circuits. Cependant, on peut imaginer implémenter ce genre de structure au niveau des zones actives du silicium, notamment avec une technologie de type SON [Segueni'07].

Cependant, le fait que la poutre soit mobile a quelques conséquences:

- la poutre doit être longue (transistor large) pour être plus facilement défléchie
- la capacité de grille est beaucoup moins importante que pour une TMOS classique, du faite de la composante due au vide.

3.1.3.3. Performances

En approximation canal long, le courant de conduction s'écrit:

$$I_{on} = \mu \cdot C_g \cdot \frac{W}{L} \cdot \frac{V_{DD}^2}{2}$$

Le fait de disposer d'une capacité de grille faible entraîne un niveau de courant faible.

Par contre, le courant tunnel à travers la grille est extrêmement limité, du fait la grande épaisseur physique de l'isolant de grille.

En terme de vitesse de commutation, il convient de prendre en compte la partie mécanique du système. Avec des dimensions suffisamment faibles, telles qu'on les obtient en technologie SON, la fréquence de résonance d'une telle poutre est le d'ordre de la centaine de GHz. La vitesse n'est donc pas un élément bloquant.

3.1.3.4. Applications envisagées

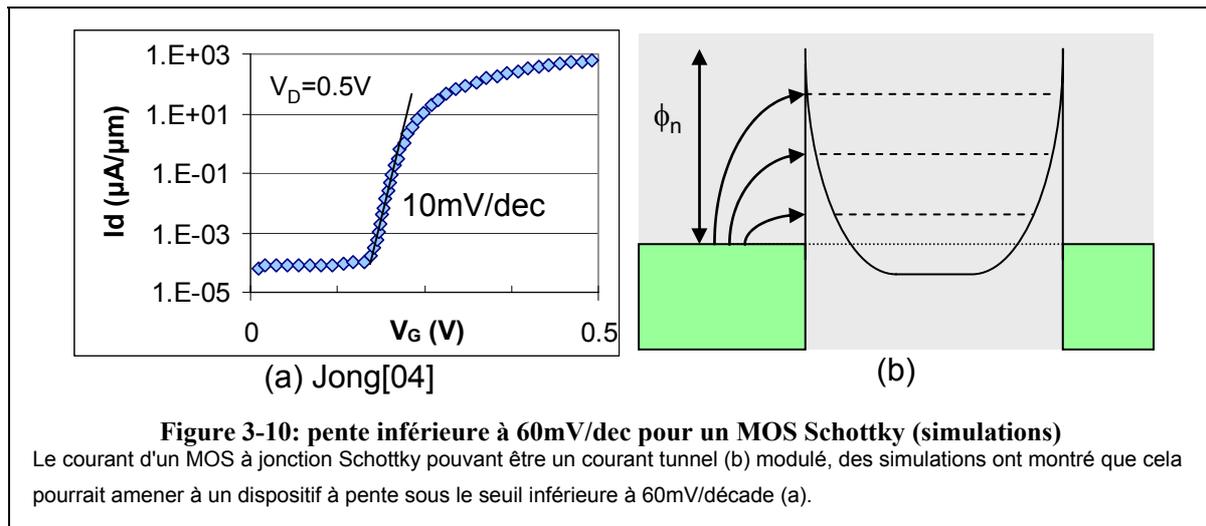
Malgré la faible pente, le rapport I_{on}/I_{off} ne permet pas de l'utiliser comme interrupteur. Les mesures ont montré un hystérésis, permettant de l'utilisant comme une mémoire [Abele'06].

Cette structure peut également être utilisée comme résonateur, dont l'application principale est le filtrage de signal. L'utilisation d'un matériau cristallin et les faibles dimensions utilisées permettent d'envisager un facteur de qualité bien supérieur aux résonateurs actuels.

3.2. Structures utilisant l'effet tunnel

3.2.1. Transistor à jonctions Schottky

Les jonctions métalliques de type Schottky ont été présentées au le chapitre 1. Dans ce cas, la courant est alors régi par l'injection par effet tunnel. Des simulations ont montré [Jang'04] que de tels dispositifs sont susceptibles de présenter une pente inférieure à 60mV/dec. En effet, le mécanisme responsable du passage du courant n'est plus la diffusion comme dans le cas du MOS classique, mais l'effet tunnel. Par conséquent, la pente n'est plus constante. Dans la simulation (**figure 3-10a**), la pente descend à 10mV/dec.

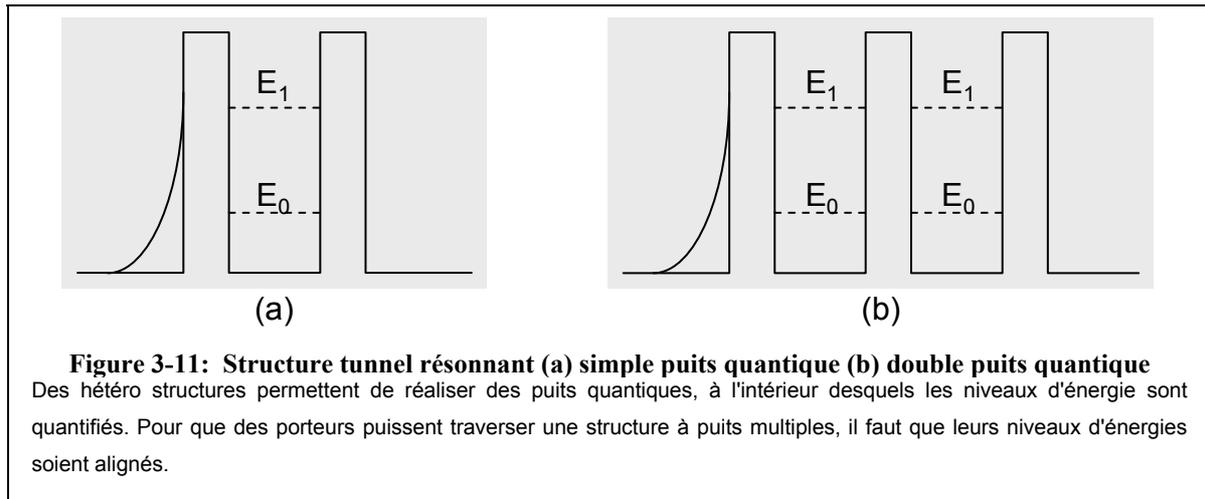


Cependant, ces résultats sont à prendre avec précaution. L'effet tunnel actuellement implémenté dans les simulateurs électriques usuels ne résulte pas d'une modélisation physique, trop complexe. Par conséquent, on lui préfère des expressions analytiques simples, qui ne sont valables que dans le domaine d'application où elles ont été définies. Or, ces équations ont été définies pour des épaisseurs de barrière tunnel bien supérieures à celle résultant d'une jonction Schottky.

D'autre part, il a été montré ailleurs [Sodamos] que l'effet tunnel ne pouvait pas améliorer la pente sous le seuil d'un transistor MOS à jonctions Schottky. Notons également qu'à ce jour, aucun résultat expérimental n'est venu illustrer cette hypothèse.

3.2.2. Structure tunnel résonnant

L'effet tunnel résonnant [Blanks'98] est une application du courant tunnel direct ou de l'effet Fowler Nordheim utilisant un niveau intermédiaire quantifié. Celui-ci est usuellement formé par une région semi-conductrice, car la résonance nécessite la conservation du vecteur d'onde, ce qui équivaut à avoir peu d'interactions. Le niveau quantifié est encadré par deux barrières (de l'oxyde de silicium par exemple), formant ainsi un puits quantique (**figure 3-11a**).



Le courant peut passer par effet tunnel de la source vers le puits quantique, puis du puits quantique au drain à nouveau par effet tunnel. Ce mécanisme de transport conduit à une variation très abrupte du courant en fonction de la position du niveau de quantification. Cet effet peut être amplifié en utilisant des barrières multiples (**figure 3-11b**), ce qui entraîne les porteurs à passer d'un état quantifié à un autre.

En pratique, les niveaux d'énergie quantifiés ne sont pas abrupts à cause de différents mécanismes d'élargissement tels que la relaxation d'énergie, les vibrations thermiques, etc. qui entraîne une distribution non abrupte en énergie. Cependant, la densité d'états a une résolution énergétique inférieure à ' kT/q ', spécialement pour des barrières très fines (les interactions sont minimisées).

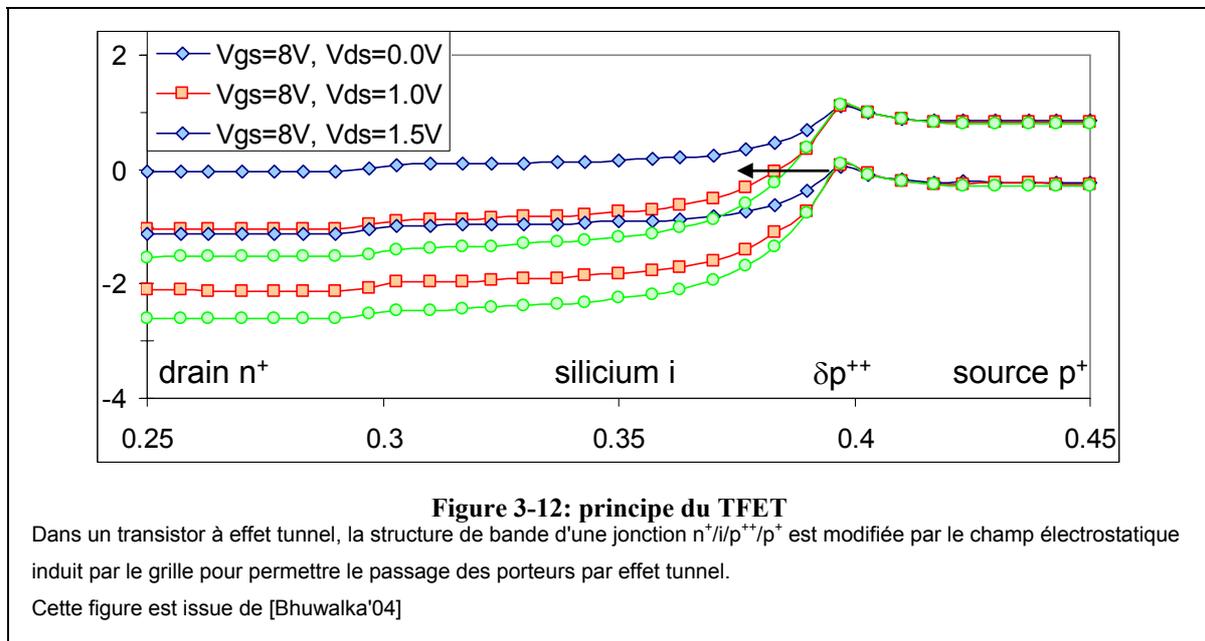
Pour des oxydes minces (<3nm), l'effet tunnel résonnant ne présente pas de transition abrupte ON/OFF à cause des courants tunnel direct et de leur impact sur l'élargissement des niveaux d'énergie dans le puits de potentiel. De plus, il n'est pas évident que des systèmes à 3 terminaux puissent être réalisés et qu'ils montrent des propriétés similaires à celles des dipôles. Pour des barrières plus épaisses, la densité de courant est faible, ce qui signifie une faible vitesse de commutation et affecte la fiabilité. Nous devons également mentionner qu'aucune structure tunnel résonnant n'a encore été observée dans des systèmes Si-SiO₂-Si-SiO₂-Si. Ceci a été attribué aux interactions dans le poly silicium qui forme le puits.

3.2.3. Transistor à effet tunnel (TFET)

Le transistor à effet tunnel fonctionne sur le principe du passage du courant par effet tunnel bande à bande. Contrairement au transistor MOS, source et drain sont de dopages opposés. Ils sont séparés par une zone intrinsèque, c'est-à-dire non dopée. Ce dispositif est donc une diode p-i-n. En polarisant suffisamment la grille et le drain, les bandes du semi-conducteur se

courbent jusqu'à faire passer la bande de conduction dans la zone i sous la bande de valence de la source (**figure 3-12**). Ceci entraîne du courant par effet tunnel bande à bande.

Le phénomène de conduction n'est plus régi par la statistique des porteurs. Des simulations de cette structure ont montré une pente sous le seuil inférieure à 60mV/dec [Bhuwalka'05]. Cependant, comme pour le transistor à effet Schottky, ces considérations reposent sur un modèle phénoménologique de l'effet tunnel extrapolé hors de sa zone d'utilisation. Aucune intégration n'a démontré la possibilité de passer la limite de 60mV/dec. Les résultats qui semblaient les plus prometteurs [Nirschl'05a] se sont avérés inexacts. Les auteurs avaient réalisé un transistor MOS au lieu d'un transistor à effet tunnel en implantant des LDD avec un mauvais type de dopants.



3.3. Structures mettant en jeu un phénomène d'amplification

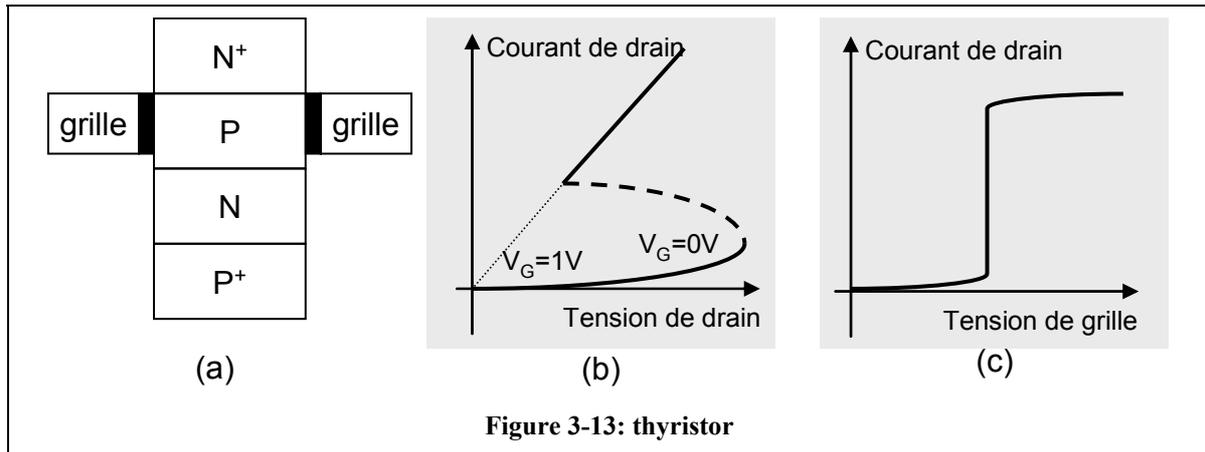
3.3.1. Thyristor

Les thyristors utilisent des variantes de structure p-n-p-n (**figure 3-13a**). Ils présentent une résistance dynamique négative ou **NDR** (*Negative Dynamic Resistance*). A l'état bloqué, le thyristor peut supporter une forte tension directe. A l'état de conduction, le thyristor se comporte comme jonction p-n polarisée en directe, ce qui lui permet de fournir une grande densité de courant.

Les structures p-n-p-n commutent de façon très abrupte de l'état bloqué à l'état de conduction et inversement (**figure 3-13b**). C'est parce que le courant à l'état "OFF" est limité par le

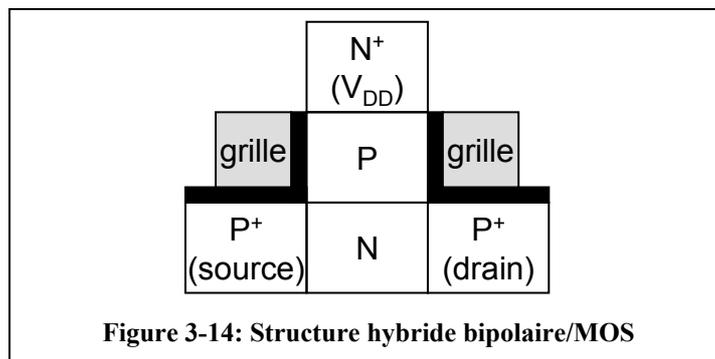
courant en inverse de la jonction p-n alors que le courant "ON" est limité par le courant direct des jonctions polarisées.

L'inconvénient du thyristor est sa faible vitesse de commutation. Bien que quelques avancées aient été observées en utilisant les propriétés films minces [Nemati'99], les plus rapides montrent des temps de commutation de quelques nanosecondes, ce qui est incompatible avec les Thz nécessaires pour les futurs noeud CMOS.



3.3.2. Hybride bipolaire/MOS

Le dispositif hybride bipolaire/MOS [Bhat'88] exploite un mécanisme de gain pour améliorer la pente sous le seuil du MOSFET. La structure utilise un NMOS qui émet vers la base d'un transistor bipolaire à hétérojonction (HBT) de type PNP. Le courant de sortie est donc le produit du courant du TMOS par le gain β du HBT. Or, le gain β du transistor bipolaire augmente de façon monotone avec le courant de base. Ce gain intrinsèque permet une amplification qui réduit la pente sous le seuil.



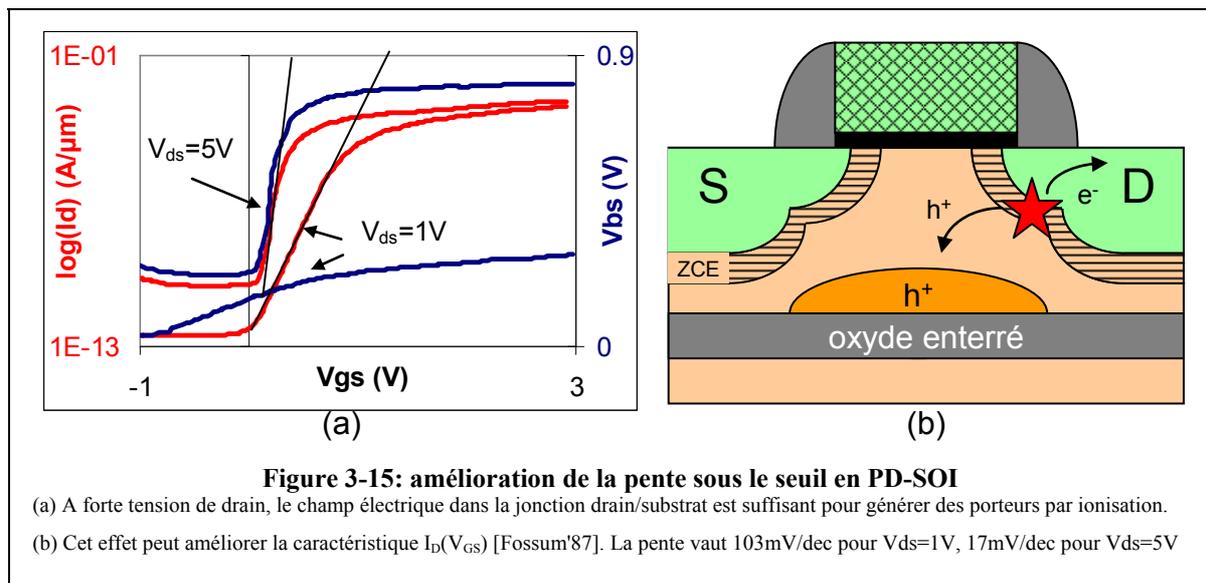
Cependant, ce dispositif présente deux sérieux inconvénients. D'abord, il bascule très lentement de l'état "ON" à l'état "OFF" à cause des porteurs minoritaires en excès dans la base du HBT. Ces porteurs ne peuvent en effet être évacués que par recombinaison. Ensuite, d'un

point de vue encombrement, la réduction des dimensions du HBT est plus difficile que celle du MOS, et cela fait deux transistors pour une cellule logique.

3.3.3. SOI partiellement déplété

Certains dispositifs réalisés sur SOI partiellement déplété ont présenté une pente sous le seuil améliorées [Davis'86]. Cette amélioration a même pu conduire à une pente inférieure à 60mV/décade [Fossum'87], comme illustré sur la **figure 3-15a**. L'auteur explique ce phénomène de manière similaire à l'effet "kink" parfois observé en régime de saturation:

A forte tension de drain **figure 3-15b**, le champ électrique entre canal et drain est suffisamment élevé pour créer des paires électrons-trous par ionisation par impact. Les électrons sont évacués par le drain, mais les trous sont repoussés à l'interface inférieure du film de silicium. Le film étant partiellement déplété, cette interface peut être assimilée au substrat (body). Ces trous vont alors modifier le potentiel de substrat s'il n'est pas polarisée (*floating body* ou substrat flottant). Ce changement de potentiel de substrat entraîne une chute de la tension de seuil et donc une augmentation de courant.

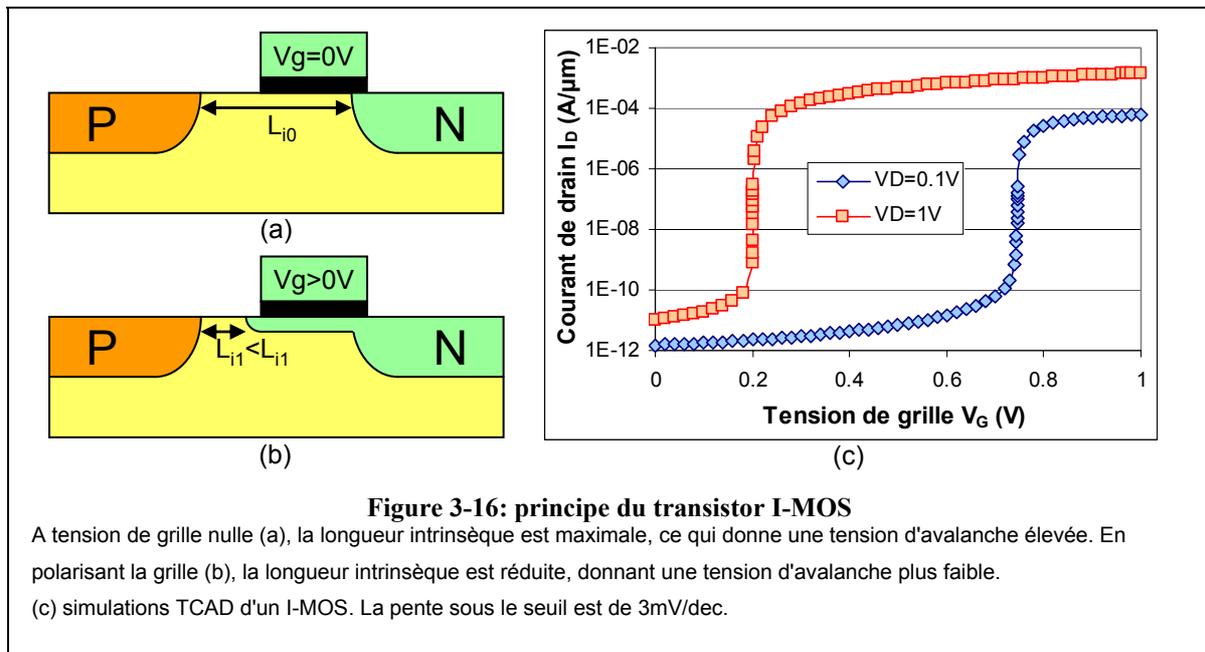


Cependant, l'effet de substrat flottant est associé à d'autres phénomènes parasites, notamment des effets transitoires (voir chapitre 1). Ces effets sont supprimés par l'ajout d'une prise de terre permettant d'évacuer l'excès de porteurs majoritaires, empêchant donc d'utiliser cette pente anormalement faible.

3.3.4. MOS à Ionisation par Impact (I-MOS)

Le principe du transistor I-MOS est de contrôler la tension de claquage par avalanche d'une jonction p-i-n par une électrode de grille. Une jonction p-i-n est composée de trois zones: une zone dopée n, une zone non dopée (intrinsèque) et une zone p. Lorsque cette jonction est polarisée en inverse, le champ électrique peut être suffisant pour générer des porteurs par ionisation par impact. La tension à appliquer à ce dipôle est d'autant plus faible que la zone intrinsèque est réduite.

Dans un I-MOS, l'électrode de grille recouvre une partie seulement de la zone intrinsèque. En polarisant cette électrode, une accumulation de charges est créée, ce qui équivaut à réduire la longueur intrinsèque de la jonction p-i-n (**figure 3-16a et b**). Le phénomène d'avalanche étant très abrupt et n'étant pas assimilable à un processus activé thermodynamiquement, des pentes sous-le seuil très faibles peuvent être atteintes (**figure 3-16c**).



Sur les premiers dispositifs réalisés sur germanium [Gopalakrishnan'02], la pente sous le seuil a été mesurée à 10mV/dec. Depuis, des réalisations sur silicium ont montré des pentes jusqu'à 2mV/dec [Choi'04a,Choi'04b,Choi'05b,Mayer'06,Toh'05].

Conclusion du chapitre

Ce chapitre a passé en revue un certain nombre de structures présentant ou supposant pouvoir présenter une caractéristique semblable à celle du TMOS, mais avec une pente sous le seuil meilleure que 60mV/dec. La structure ayant un potentiel avenir industriel à moyen terme, c'est-à-dire susceptible de remplacer le transistor MOS conventionnel dans certaines applications, implique certaines contraintes:

- l'intégration doit être compatible avec celle du TMOS actuel. Une structure en silicium est donc obligatoire. Par conséquent, l'étude sur les matériaux à transition de Mott, trop exotiques, est abandonnée.
- les concepts de structures basées sur l'effet tunnel reposent sur des simulations dont on peut mettre en doute la pertinence: le modèle de l'effet tunnel est utilisé en dehors de son domaine de validité. En outre, nulle réalisation technologique ne vient étayer ces concepts.
- un atout majeur du transistor MOS est sa densité d'intégration. Il ne serait pas raisonnable d'étudier une structure, si intéressante soit-elle, dès lors qu'elle nécessite une surface d'intégration trop importante. C'est ceci qui en partie rend délicat l'intégration de transistor à grille mobile, la poutre devant être suffisamment longue pour être flexible. D'autre part, l'encapsulation de ce genre de structures reste délicat.
- la rapidité de commutation doit être également considérée. La structure hybride TMOS/bipolaire étant limitée par la recombinaison par les porteurs minoritaires, sa vitesse de commutation est limitée.

De toutes les structures passées en revue, la plus prometteuse est donc le transistor de type I-MOS. C'est une structure parfaitement compatible silicium, co-intégrable avec des TMOS [Mayer'06]. Ses caractéristiques ont été démontrées expérimentalement, il n'y a pas de limitation particulière à la réduction des dimensions. Le procédé d'avalanche est un phénomène rapide (les structures IMPATT reposant sur l'effet d'avalanche fonctionnent jusqu'à 30GHz [Sze'81]), la relaxation diélectrique se fait par courant de dérive, ce qui fait des commutations rapides. C'est donc sur cette structure que nous allons nous focaliser.

Chapitre 4. Le transistor de type I-MOS

Le transistor de type I-MOS est un dispositif basé sur le phénomène de multiplication de porteurs par avalanche. Il se présente comme une diode p-i-n contrôlée par une grille.

Dans ce chapitre, le phénomène physique de multiplication par avalanche est présenté. Les différentes approches de modélisation de ce phénomène sont passées en revue. Il en résulte une sélection du modèle le plus approprié à utiliser dans les simulations et les modélisations analytiques.

Puis, la physique du composant est détaillée. A cette fin, la modélisation unidimensionnelle de la diode p-i-n est abordée en premier lieu. Elle sert de base pour appréhender le phénomène d'avalanche.

Enfin, une étude détaillée du transistor I-MOS est proposée: principe de fonctionnement, phénomènes physiques mis en jeu, similitudes et différences par rapport à un transistor MOS conventionnel.

Les discussions et démonstrations de ce chapitre sont largement documentées et étayées par la simulation électrique.

Introduction

De tous les dispositifs recensés au chapitre 3, le plus prometteur est indéniablement le dispositif de type I-MOS (*Impact Ionization MOS*). Ce dispositif basé sur le procédé de multiplication de porteurs par ionisation par impact est le seul dont les réalisations technologiques ont montré une pente inférieure à 60mV/dec.

Initialement intégré sur germanium [Gopalakrishnan'02], il a rapidement été porté sur silicium [Choi'04a]. Diverses publications [Choi'04b,Choi'05a] ont montré qu'il est possible de réduire les dimensions, et que plusieurs schémas d'intégrations pour y parvenir sont envisageables [Toh'05].

Comme le transistor MOS conventionnel, ce dispositif peut être décliné de type n ou de type p. Ceci permet d'imaginer des circuits en technologie I-MOS complémentaire. Des circuits élémentaires (inverseur, SRAM) ont pu être simulés à partir des caractéristiques individuelles de dispositifs I-MOS [Choi'05a], montrant ainsi leur utilité.

Ce chapitre présente les dispositifs I-MOS d'un point de vue théorique, et est largement soutenue par des simulations TCAD. Le chapitre suivant traite de l'intégration technologique de dispositifs I-MOS, et un dernier chapitre traite plus spécifiquement des circuits réalisés en technologie I-MOS complémentaire.

4.1. Introduction au transport dans les semi-conducteurs

Les phénomènes physiques dans les dispositifs semi-conducteurs sont nombreux et variés. Ils peuvent être décrits par un jeu d'équations différentielles, plus ou moins complexe selon les phénomènes considérés. Ces équations permettent notamment de décrire le transport des charges, c'est-à-dire la position des porteurs et leur vitesse, à tout instant et à tout endroit. Cette première partie introduit les notions élémentaires de la physique du transport, telles que présentées dans [ISE'04].

4.1.1. Equations élémentaires de base de la physique du transport

Trois équations fondamentales gouvernent la physique du transport. La première est l'équation de Poisson, qui relie le potentiel électrostatique ψ à la densité spatiale de charges ρ :

$$\text{div}(\varepsilon \nabla \psi) = -\rho \quad (4.3)$$

où ε est la permittivité électrique.

Les deux autres sont les équations de continuité pour les trous et les électrons:

$$\nabla \cdot \vec{J}_N = q \cdot R_n - q \cdot G_n + q \frac{\partial n}{\partial t} \quad (4.4)$$

$$-\nabla \cdot \vec{J}_p = q \cdot R_p - G_p + p \frac{\partial p}{\partial t} \quad (4.5)$$

où R_n et R_p sont les taux de recombinaison des électrons et des trous respectivement, où G_n et G_p sont les taux de génération des électrons et des trous respectivement, \vec{J}_N la densité de courant d'électrons et \vec{J}_p la densité de courant de trous. La description d'autres phénomènes, les effets quantiques par exemple, nécessite l'ajout de nouvelles équations.

Pour une description du transport dans un dispositif semi-conducteur, ces équations doivent être résolues de manière auto cohérente. Selon le degré de précision dans la description des phénomènes physiques, des expressions plus ou moins complexes des densités de courant sont utilisables, aboutissant à différents modèles.

4.1.2. Modèle dérive-diffusion

Les densités de courant sont dérivées à l'ordre 1. Les expressions des densités de courant sont:

$$\vec{J}_N = -n \cdot q \cdot \mu_n \cdot \nabla \phi_n \quad (4.6)$$

$$\vec{J}_p = -p \cdot q \cdot \mu_p \cdot \nabla \phi_p \quad (4.7)$$

où μ_n et μ_p sont les mobilités des électrons et des trous, et ϕ_n et ϕ_p sont les quasi-niveaux de Fermi des électrons et des trous.

4.1.3. Modèle thermodynamique

Le modèle thermodynamique (ou non isothermique) étend l'approche dérive-diffusion pour prendre en compte les effets électrothermiques. Il suppose que les porteurs de charge sont en équilibre thermique avec le réseau. Les expressions des densités de courant deviennent alors:

$$\bar{J}_N = -n \cdot q \cdot \mu_n \cdot (\nabla \phi_n + P_n \cdot \nabla T) \quad (4.8)$$

$$\bar{J}_p = -p \cdot q \cdot \mu_p \cdot (\nabla \phi_p + P_p \cdot \nabla T) \quad (4.9)$$

où P_n et P_p sont les puissances thermoélectriques absolues [Callen'85].

Pour calculer la distribution de température dans le dispositif due à l'auto échauffement, l'équation suivante peut être utilisée [ISE'04]:

$$c \frac{\partial T}{\partial t} - \nabla \cdot \kappa \nabla T = -\nabla \cdot \left[(P_n \cdot T + \phi_n) \bar{J}_N + (P_p \cdot T + \phi_p) \bar{J}_p \right] - \left(E_C + \frac{3}{2} \cdot k_B \cdot T \right) \nabla \bar{J}_N - \left(E_V - \frac{3}{2} \cdot k_B \cdot T \right) \nabla \bar{J}_p + q \cdot R \cdot (E_C - E_V - 3 \cdot k_B \cdot T) \quad (4.10)$$

4.1.4. Modèle hydrodynamique

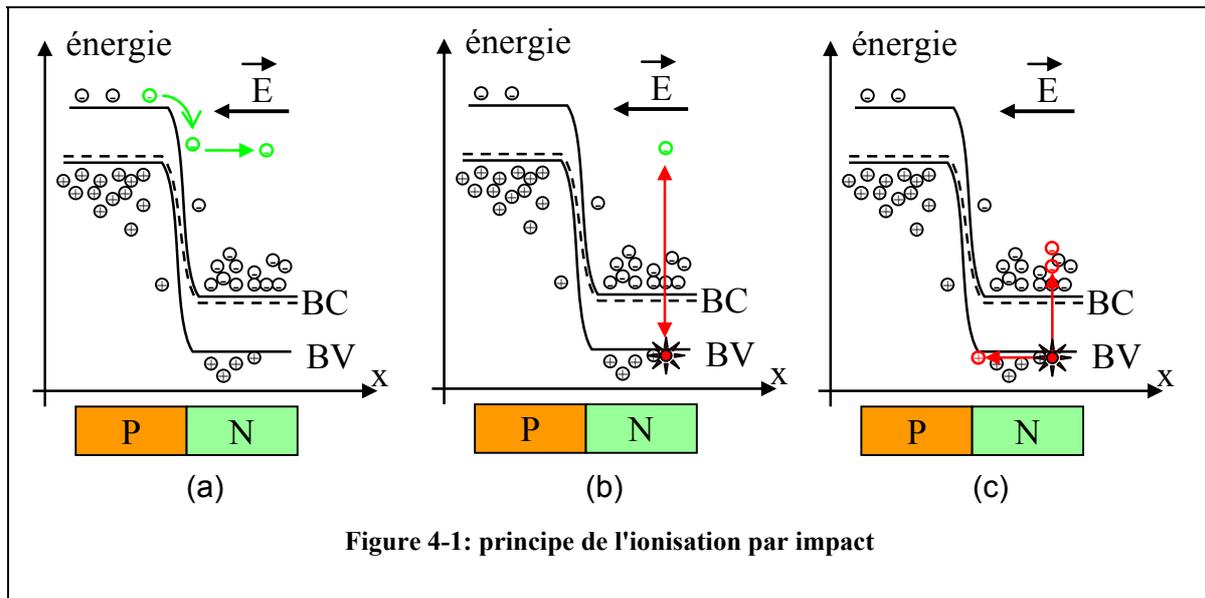
Le modèle hydrodynamique étend le modèle thermodynamique, en incluant le fait que les porteurs ne sont pas forcément en équilibre thermodynamique avec le réseau. De nombreuses variations du modèle existent dans la littérature. La formulation complète, incluant les termes dits convectifs, consiste en un jeu de huit équations différentielles partielles [Benvenuti'93a, Benvenuti'93b, Benvenuti'92]. La forme simplifiée comprend six équations différentielles partielles [Chen'93, ISE'04].

A mesure que les dimensions des dispositifs réduisent, de nouveaux phénomènes physiques apparaissent ou prennent une importance croissante. Ceux-ci ne peuvent pas toujours être décrits par le modèle de dérive-diffusion. Par exemple, ce modèle ne peut pas reproduire le phénomène de survitesse et a tendance à surestimer les taux de génération d'ionisation par impact. Le modèle hydrodynamique, plus précis, permet de décrire ces phénomènes.

4.2. Le phénomène d'ionisation par impact

4.2.1. Description – phénomène d'avalanche

Considérons un électron de la bande de conduction dans une jonction p-n. Celui-ci est accéléré par le champ électrique. Il acquiert ainsi de l'énergie cinétique (**figure 4-1a**). Lorsque cette énergie est très supérieure à l'énergie thermique, le porteur est appelé porteur chaud. En entrant en collision avec un élément du réseau (**figure 4-1b**), il arrache un électron de la bande de valence pour (**figure 4-1c**).

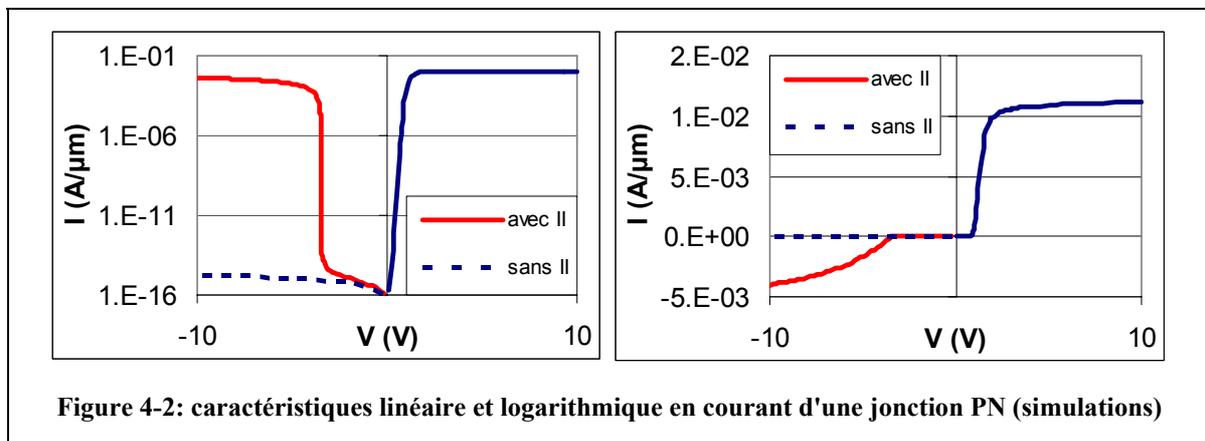


Le bilan de cette collision est donc :



L'ionisation par impact est donc un phénomène de génération de porteurs. Il est d'autant plus important que le champ électrique est élevé.

Si le champ électrique est suffisant, alors les porteurs créés peuvent à leur tour générer d'autres porteurs par ionisation par impact. C'est le phénomène d'avalanche, qui se traduit par



exemple par une brusque augmentation du courant dans les caractéristiques électriques de jonctions p-n polarisées en inverse (**figure 4-2**).

4.2.2. Modélisation du phénomène d'ionisation par impact

4.2.2.1. Les coefficients d'ionisation

Le coefficient d'ionisation α est défini comme étant la quantité de porteurs créés par unité de longueur et par porteur. L'unité courante est le cm^{-1} . Ce coefficient est différent pour les électrons et les trous, et est donc noté différemment: α pour les électrons et β pour les trous, ou bien α_n et α_p . Dans ce manuscrit, c'est la seconde notation qui sera utilisée.

Il y a deux façons de prendre en compte ce phénomène:

La première est d'inclure la modélisation de ce phénomène dans les équations de continuité des densités de porteurs (4.4) et (4.5). Elle apparaît dans le terme de génération/recombinaison R . Les équations différentielles servant à la modélisation sont alors résolues de manière auto cohérente. C'est ce qui est fait dans un simulateur TCAD. La modélisation des coefficients d'ionisation va alors dépendre du modèle utilisé (dérive-diffusion, thermodynamique ou hydrodynamique). La quantité de paires électron-trou générées par unité de temps et par unité de volume à la position x s'exprime alors sous la forme:

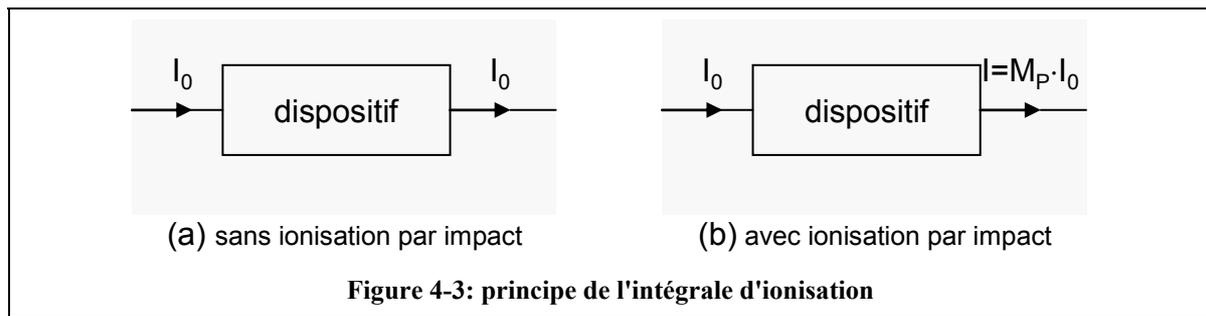
$$\begin{aligned} g_n(x) &= \alpha_n(x) \cdot n(x) \cdot |v_n(x)| \\ g_p(x) &= \alpha_p(x) \cdot p(x) \cdot |v_p(x)| \end{aligned} \quad (4.12)$$

La seconde façon de prendre en compte l'ionisation par impact est de la traiter comme une perturbation: la distribution de porteurs et de potentiel est résolue sans tenir compte de l'ionisation par impact. Puis, l'effet sur le courant est ajouté en considérant que cela **influe de façon négligeable la distribution de porteurs et de potentiel**. Cette approche sera également utilisée dans ce chapitre, pour la modélisation analytique du phénomène.

4.2.2.2. Intégrale d'ionisation

Cette notion n'a de sens que lorsque l'on traite l'ionisation par impact comme une perturbation du transport. Sans ionisation, le courant qui traverse la structure est alors accru par un facteur multiplicatif M_p qui dépend de la structure et des polarisations (**figure 4-3**). Le calcul de ce coefficient est présenté en annexe. L'avalanche apparaît lorsque $M_p \rightarrow \infty$; c'est l'intégrale d'ionisation:

$$\int_0^w \alpha_p \exp\left(-\int_0^x (\alpha_p - \alpha_n) dx'\right) dx = 1 \quad (4.13)$$



4.2.3. Modélisation des coefficients d'ionisation par impact

Il n'existe pas de modèle physique théorique de l'ionisation par impact dans un semi-conducteur. Il n'existe que des modèles empiriques. ceux-ci peuvent être séparés en deux catégories:

- les **modèles locaux** sont utilisés en résolution auto cohérente dans le modèle dérive-diffusion et dans le traitement en perturbation. Ils considèrent que les coefficients d'ionisation au point x ne dépendent que des conditions à ce point.
- les **modèles non locaux** sont utilisés en résolution auto cohérente dans les thermodynamique et hydrodynamique. Ils permettent de prendre en compte l'"historique" de la particule.

4.2.3.1. Modèles locaux

Relativement simples, ils ont été les premiers proposés. Les coefficients dépendent:

- du **matériau**: l'avalanche est plus efficace dans les matériaux à faible gap (germanium par rapport à silicium par exemple). Cela peut s'expliquer par la moindre énergie à apporter pour exciter un porteur de la bande de valence vers la bande de conduction.
- de l'**orientation cristallographique**
- du **type de porteur**: l'avalanche est généralement plus facile à obtenir avec des électrons qu'avec des trous. Ceci peut être dû à leur masse plus faible.
- de la **température**: les interactions avec les phonons augmentent avec la température. Cela réduit la probabilité qu'un porteur atteigne une énergie suffisante pour générer une paire par avalanche. Par conséquent, l'avalanche décroît avec la température.
- du **champ électrique E**

Pour un substrat donné, on cherche donc des relations du type $\alpha_n(E,T)$ et $\alpha_p(E,T)$.

Similitude à la décharge dans un gaz

L'ionisation par impact dans un semi-conducteur est modélisée de manière similaire à la décharge dans un gaz [Chynoweth'58]:

$$\alpha_{n,p}(E) = A_{n,p} \cdot \exp\left(-\frac{B_{n,p}}{E}\right) \quad (4.14)$$

où E est la composante du champ électrique parallèle au déplacement de charges, A et B des coefficients d'ajustement différents pour les électrons (indice n) et les trous (indice p).

C'est le modèle le plus ancien de la littérature et le plus simple. Il a été étendu pour prendre en compte l'effet de la température [Van Overstraeten'70]:

$$\alpha_{n,p} = \gamma \cdot A_{n,p} \cdot \exp\left(-\gamma \cdot \frac{B_{n,p}}{E}\right) \quad (4.15)$$

$$\gamma = \frac{\tanh\left(\frac{h \cdot \omega_{op}}{2 \cdot k_b \cdot T_0}\right)}{\tanh\left(\frac{h \cdot \omega_{op}}{2 \cdot k_b \cdot T}\right)}$$

où k_b est la constante de Boltzmann, T_0 une température de référence, T la température du dispositif et ω_{op} l'énergie des phonons optiques.

L'extraction de ces paramètres se fait à partir de mesures expérimentales de tension d'avalanche sur des structures de type p-i-n. Le profil de charges est postulés, puis le profil de champ électrique résultant est calculé analytiquement. Enfin, l'intégrale d'ionisation fournit l'équation qui permet de déterminer α et β . De nombreuses publications traitent de ce modèle, dont les hypothèses quant au profil de charges diffèrent, aboutissant donc à de nombreuses valeurs de coefficients:

		faible champ		champ limite (V/cm)	fort champ	
		A (cm ⁻¹)	B (V/cm)		A (cm ⁻¹)	B (V/cm)
[Chynoweth'58]	électrons					
	trous					
[Lee'64]	électrons	3,37·10 ⁶	1,74·10 ⁶			
	trous	1,12·10 ⁷	3,02·10 ⁶			
[Sze'66]	électrons	3.8·10 ⁶	1,75·10 ⁶			
	trous	2,25·10 ⁶	3.26·10 ⁶			
[Van Overstraeten'70]	électrons	7,03·10 ⁵	1,12·10 ⁶			
	trous	1,58·10 ⁶	2,04·10 ⁶	4·10 ⁵	6,71·10 ⁶	1,69·10 ⁶
[Grant'73]	électrons	6,2·10 ⁵	1,08·10 ⁶	5,3·10 ⁵	5·10 ⁵	0,99·10 ⁶
	trous	2·10 ⁶	1,97·10 ⁶	5,3·10 ⁶	5,6·10 ⁶	1,32·10 ⁶

Tableau 2: coefficients des modèles d'ionisations par impact pour le modèle de Chynoweth

Le modèle le plus utilisé est celui de Van Overtraeten. En effet, les extractions ont été faites sur des jonctions implantées en utilisant des fonctions exponentielles pour représenter les

profils de dopage, ce qui est correct. Les autres publications font l'hypothèse d'un profil de charges constant ou linéaire.

Intégration d'une énergie seuil

D'autres articles [Lackner'91, Okuto'75] prennent en compte le fait qu'une particule doit atteindre une énergie suffisante pour générer de l'ionisation par impact. Le calcul de l'intégrale d'ionisation intègre donc une énergie seuil ainsi que les libres parcours moyens des porteurs.

Ainsi, Thomas Lackner propose à partir de ces considérations une amélioration de la loi de Chynoweth, qui se trouve alors être un cas particulier d'une théorie plus générale:

$$\alpha_{n,p} = \frac{\gamma \cdot A_{n,p}}{Z} \cdot \exp\left(-\gamma \cdot \frac{B_{n,p}}{E}\right)^m \quad \text{avec} \quad \gamma = \frac{\tanh\left(\frac{h \cdot \omega_{op}}{2 \cdot k_b \cdot T_0}\right)}{\tanh\left(\frac{h \cdot \omega_{op}}{2 \cdot k_b \cdot T}\right)} \quad (4.16)$$

$$Z = 1 + \frac{\gamma \cdot B_n}{E} \cdot \exp\left(-\gamma \cdot \frac{B_p}{E}\right)^m + \frac{\gamma \cdot B_p}{E} \cdot \exp\left(-\gamma \cdot \frac{B_p}{E}\right)^m$$

Utilisation des outils numériques

Tous les modèles présentés jusqu'ici reposent sur la résolution analytique de l'intégrale d'ionisation. Plus récemment [Valdinoci'99], l'université de Bologne a proposé un modèle extrait à partir d'un plan d'expériences et une extraction numérique des coefficients d'ionisation. La loi extraite possède 12 paramètres d'ajustement, ce qui la rend peu facile à utiliser en dehors d'un outil de simulation TCAD.

$$\alpha = \frac{F}{a(T) + b(T) \exp\left[\frac{d(T)}{F + c(T)}\right]}$$

$$\begin{aligned} a(T) &= a_0 + a_1 T^{a_2} \\ b(T) &= b_0 \exp(b_1 T) \\ c(T) &= c_0 + c_1 T^{c_2} + c_3 T^2 \\ d(T) &= d_0 + d_1 T + d_2 T^2 \end{aligned} \quad (4.17)$$

4.2.3.2. Modèles non locaux

Il s'agit là d'exploiter pleinement l'hypothèse selon laquelle les coefficients d'ionisation dépendent de l'énergie des porteurs. Il n'est alors plus possible de les déterminer à partir des conditions locales, il faut tenir compte de l'historique de la particule. Cela impose donc de calculer l'énergie des porteurs, et exclue donc l'approche de dérive diffusion.

4.2.3.3. Comparaison

Localisation

Dans une diode p-i-n, le champ électrique est constant dans la zone i. L'utilisation d'un modèle local (**figure 4-4**) montre donc un profil de génération presque constant sur cette plage.

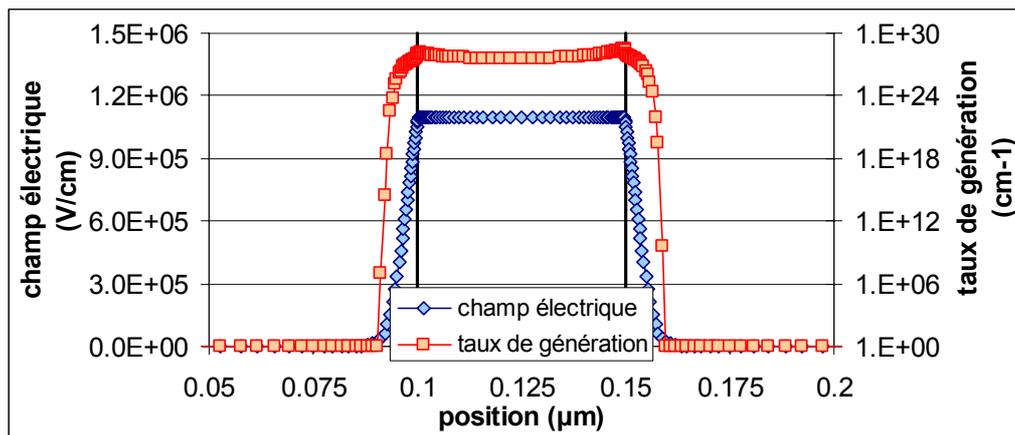


Figure 4-4: profil de génération par impact pour un modèle local

Simulation d'une diode p-i-n de longueur intrinsèque 50nm. Les zones n et p sont dopées à 10^{19}cm^{-3} . La coupe est faite à la tension de claquage (5.3V) à $I=10^{-6}\text{A}/\mu\text{m}$.

Par contre, utiliser un modèle local montre une plus génération croissante avec la position (**figure 4-5**). Cela est du au fait que, sous l'effet du champ électrique, l'énergie des porteurs augmente avec la position et le taux de génération avec elle.

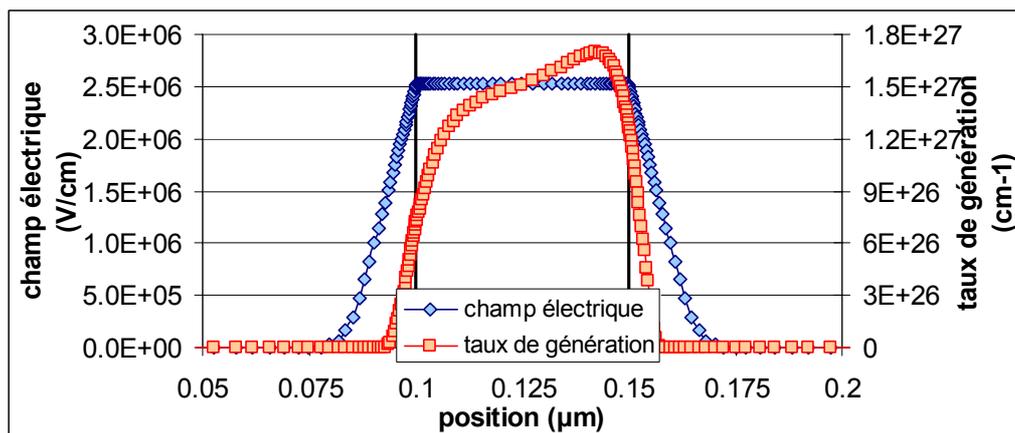


Figure 4-5: profil de génération par impact pour un modèle non local

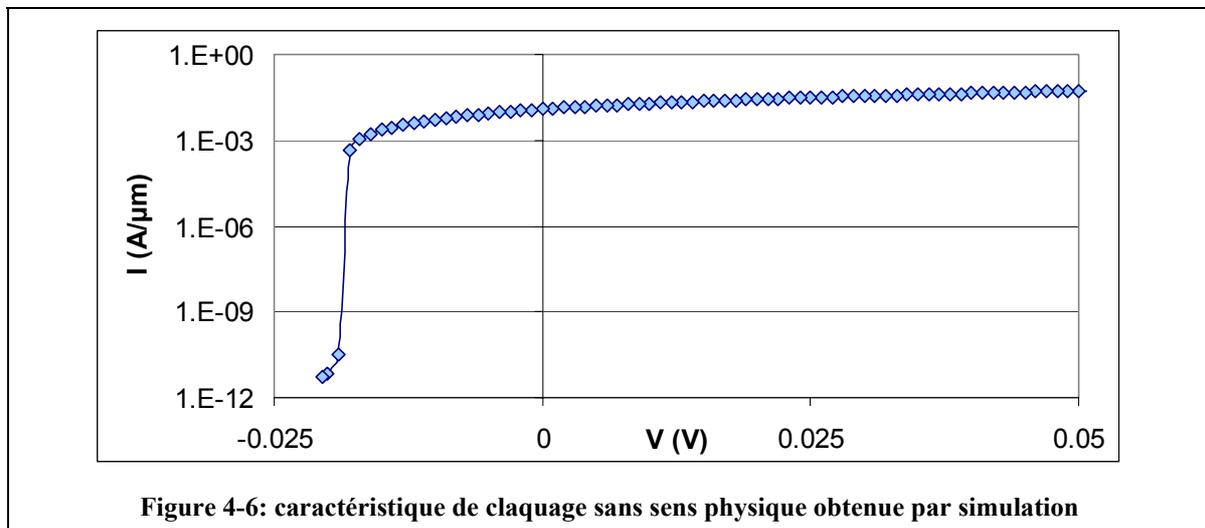
Simulation d'une diode p-i-n de longueur intrinsèque 50nm. Les zones n et p sont dopées à 10^{19}cm^{-3} . La coupe est faite à la tension de claquage (15.7V) à $I=10^{-6}\text{A}/\mu\text{m}$.

Ces figures ont été obtenues par simulations TCAD de la même structure, à densités de courant égales ($10^{-6}\text{A}/\mu\text{m}$).

On peut remarquer sur les **figures Figure 4-4 et Figure 4-5** que la génération a lieu pour des positions plus éloignées dans le cas d'un modèle non local. Cet effet est appelé effet d'espace mort (*deadspace effect*), et a d'autant plus d'impact que les dimensions du dispositif étudié sont réduites.

Limite des modèles locaux

La prise en compte de l'énergie permet d'éviter certaines situations aberrantes que l'on peut obtenir avec un modèle local. Ainsi, il est possible de simuler une jonction p-n- dont le champ électrique interne serait suffisant pour générer des porteurs par ionisation par impact (**figure 4-6**), ce qui n'a pas le moindre sens physique.



4.2.4. Choix d'un modèle

4.2.4.1. Critères de choix

Outre la précision, le choix d'un ou plusieurs modèle(s) va se fonder sur l'utilisation qui en est faite. L'étude de dispositifs I-MOS sera illustrée à la fois par de la simulation TCAD et de la modélisation analytique. Dans ce dernier, l'ionisation par impact sera traitée en tant que perturbation, et un modèle local sera donc utilisé. Le même modèle sera utilisé en simulation TCAD pour des raisons de cohérence dans le travail, ce qui implique l'utilisation de dérive-diffusion.

Ce choix d'utiliser un modèle local peut étonner, les modèles locaux ayant la réputation de ne pas être précis pour des dispositifs de dimensions réduites, en particulier à cause de l'effet d'espace mort (*deadspace effect*). s'avérer problématique. Cependant, la méthodologie d'extraction des coefficients d'ionisation prend intrinsèquement en compte cet effet. On peut

alors les considérer comme des coefficients effectifs [Plimmer'00]. Cependant, les résultats de simulation devront être confrontés à des résultats expérimentaux afin de valider ce choix.

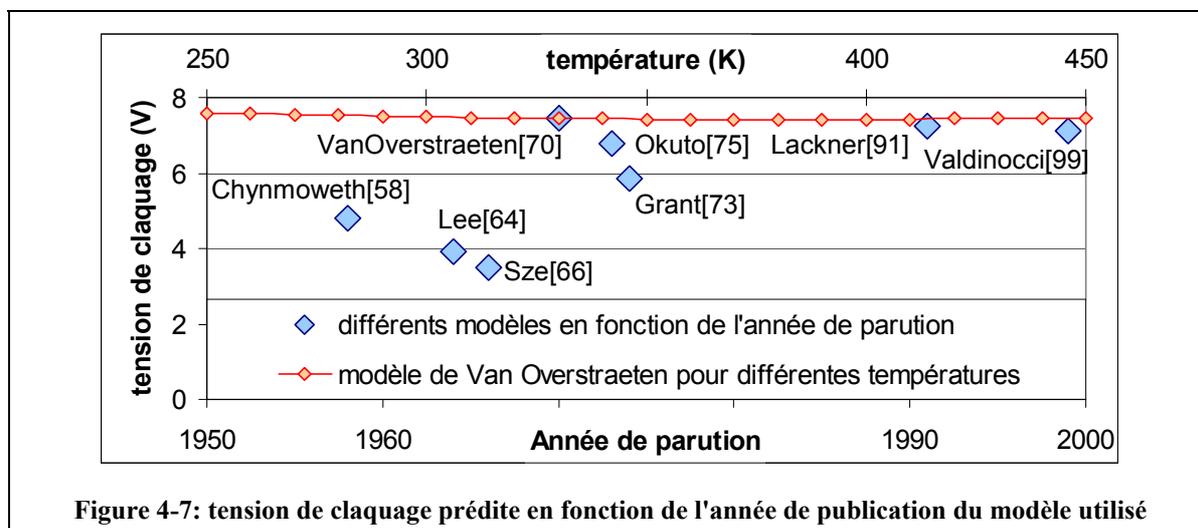
4.2.4.2. Choix d'un modèle local

Parmi les modèles locaux présentés aux paragraphes 4.2.3.1, les plus réalistes physiquement sont ceux de [Van Overstraeten'70], [Okuto'75], [Lackner'91] et [Valdinoci'99]. Cependant, il reste nécessaire d'évaluer la précision des autres modèles. D'autres aspects plus pragmatiques tels que la simplicité ou la convergence numérique dans un simulateur TCAD participent également au choix du modèle.

Précision

La **figure 4-7** représente la tension de claquage d'une jonction p-n pour différents modèles, en fonction de leur année de publication. L'influence de la température pour le modèle de Van Overstraeten [Van Overstraeten'70] y figure également.

L'influence de la température de la température apparaît suffisamment faible pour être négligée. Par conséquent, dans la suite de ce manuscrit, toutes les simulations et modélisations seront faites à température ambiante.



Les trois modèles les plus anciens ([Chynoweth'58], [Lee'64] et [Sze'66]) présentent des valeurs très en deçà de celles obtenues avec des modèles plus récents. Pour ces "vieux" modèles, l'extraction des paramètres de la loi de Chynoweth se fonde sur la mesure et la modélisation simple des profils de dopage. Cependant, on peut légitimement s'interroger quant à l'adéquation entre cette modélisation, et les dispositifs expérimentaux utilisés. Etant donnée le caractère abrupt de l'ionisation par impact, même de faibles erreurs sur la

caractérisation des échantillons peuvent conduire à des écarts importants dans les coefficients d'avalanche extraits.

En outre, les modèles estimés les plus fiables *a priori* donnent des valeurs de tension de claquage homogènes. Ceci peut être attribué à la sensible amélioration des méthodes de caractérisation des échantillons, et conforte dans l'idée qu'ils sont plus représentatifs de la réalité que les modèles les plus anciens.

Ainsi, le modèle de [Grant'73] dont les résultats diffèrent de ceux des modèles les plus récents est également écarté.

Problème de convergence numérique

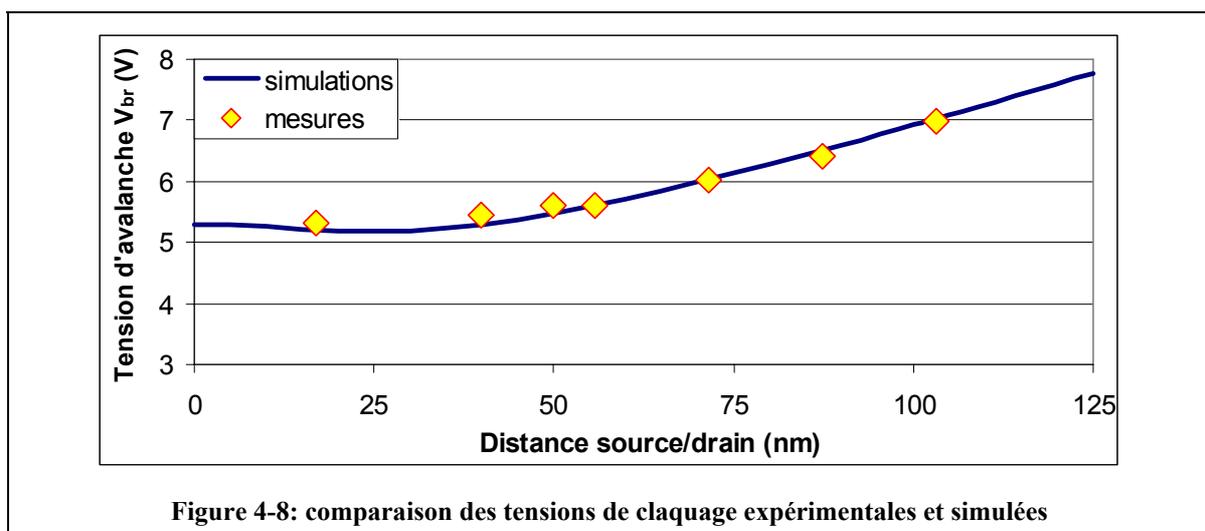
Le modèle d'Okuto [Okuto'75] a présenté des instabilités numériques lors de simulations TCAD. Il a donc été écarté pour des raisons pratiques.

Problème de complexité

Le modèle de l'université de Bologne [Valdinoci'99] utilise une douzaine de paramètres d'ajustement, ce qui est rédhibitoire pour son utilisation en modélisation analytique

4.2.4.3. Validation du choix du modèle

Deux modèles demeurent pertinents: [Lackner'91] et [Van Overstraeten'70]. Ce dernier est couramment utilisé, notamment en simulation industrielle, et a prouvé sa pertinence vis-à-vis des résultats expérimentaux. La **figure 4-8** montre que cette pertinence est valable également pour des dispositifs aux dimensions réduites, pour les raisons rappelées au paragraphe 4.2.3.2.



Dans la suite de ce manuscrit, c'est donc le modèle de [Van Overstraeten'70] qui sera utilisé, sans tenir compte de la température:

$$\alpha_{n,p}(E) = A_{n,p} \cdot \exp\left(-\frac{b_{n,p}}{E}\right) \quad (4.18)$$

avec:

		électrons	trous
1,75.10 ⁵ V/cm < E < 4.10 ⁵ V/cm	A (cm ⁻¹)	7.03E+05	1.58E+06
	b (V/cm)	1.23E+06	2.04E+06
4.10 ⁵ V/cm < E < 6.10 ⁵ V/cm	A (cm ⁻¹)	7.03E+05	6.71E+05
	b (V/cm)	1.23E+06	1.69E+06

Tableau 3: paramètres du modèle de Van Overstraeten

4.3. La diode p-i-n

4.3.1. Description et grandeurs caractéristiques

Il s'agit simplement de la succession d'une zone dopée de type p (anode), d'une zone non dopée dite intrinsèque (i) et d'une zone dopée de type n (cathode). Elle se comporte de façon similaire à une jonction p-n, en considérant que la zone de déplétion est augmentée de la taille de la zone intrinsèque.

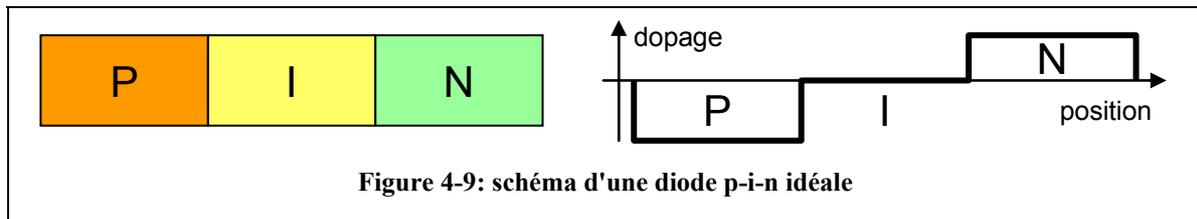
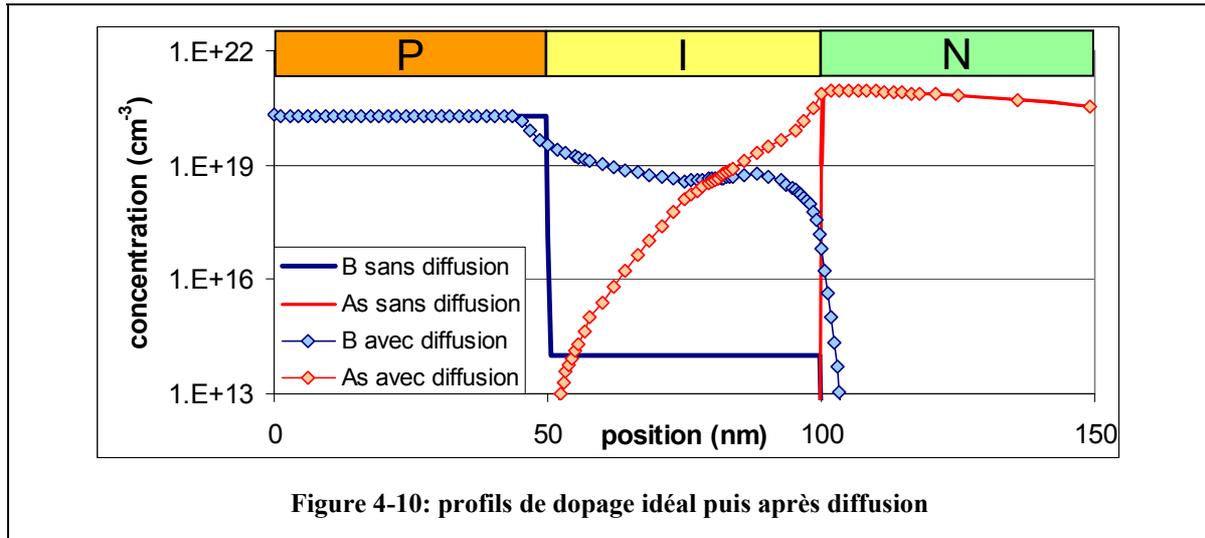


Figure 4-9: schéma d'une diode p-i-n idéale

Une diode p-i-n est caractérisée par son profil de dopants. Idéalement, les zones n et p sont uniformément dopées et la zone intrinsèque non dopée. Il y a donc discontinuité du profil de dopage aux jonctions p-i et i-n (**figure 4-9**).

Bien évidemment, ce cas n'a aucune existence réelle:

- on ne sait pas réaliser de semi-conducteur non dopé. Le silicium "intrinsèque" utilisé dans l'industrie compte au mieux 10^{15} impuretés au centimètre cube (10^{15}cm^{-3})
- la discontinuité de dopage équivaut à un gradient infini: les dopants des zones n et p vont donc naturellement diffuser vers la zone i.
- les traitements thermiques d'un procédé de fabrication vont favoriser les diffusions et donc adoucir les profils de dopage.



La **figure 4-10** montre une simulation d'empilement p-i-n. Les courbes en traits pleins montrent les profils de dopants si on néglige les diffusions. Or, les étapes de réalisation se font à relativement haute température, de l'ordre de 700°C. Ce budget thermique non négligeable entraîne une diffusion importante des dopants: la zone "intrinsèque" est dopée ici au minimum à $4 \cdot 10^{18} \text{ cm}^{-3}$.

4.3.1.1. Influence des paramètres sur la tension de claquage

Si l'on considère une diode p-i-n idéale, c'est-à-dire que les profils de dopage sont infiniment abrupts, les paramètres qui la caractérisent sont:

- la valeur du dopage dans les zones n et p
- la longueur intrinsèque, distance entre les zones n et p

Comment ces paramètres influent-ils sur la tension d'avalanche du dispositif?

L'ionisation par impact est un phénomène physique qui dépend essentiellement du champ électrique. Plus celui-ci est élevé, plus l'ionisation est efficace. Au premier ordre, le champ électrique dans une diode p-i-n va être le rapport entre la tension appliquée et la distance n/p:

$$E \approx \frac{V}{L_{n,p}} \quad (4.19)$$

Par conséquent, la tension d'avalanche doit être d'autant plus faible que la dimension $L_{n,p}$ est faible. Cette distance entre les zones n et p se décompose entre la longueur intrinsèque et les zones de charges d'espace aux jonctions n/i et i/n:

$$L_{n,p} = L_i + L_{ZCE,n/i} + L_{ZCE,i/p} \quad (4.20)$$

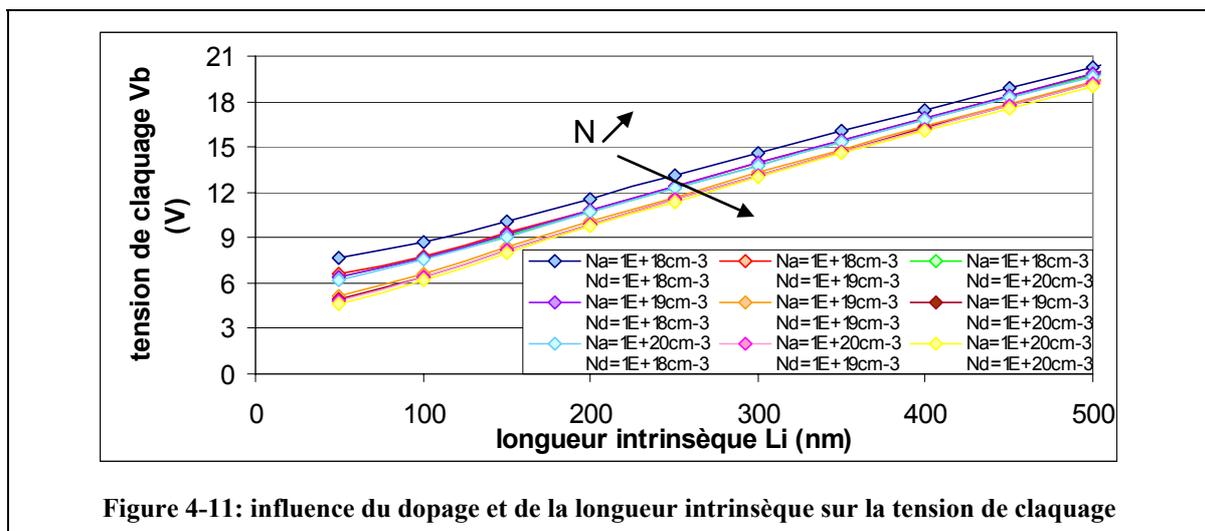
Par conséquent, nous pouvons déduire que:

- la tension de claquage diminue avec la longueur intrinsèque L_i .

- la tension de claquage diminue lorsque les zones de charges d'espace réduisent, c'est-à-dire que les zones n et p sont fortement dopées.

Ces remarques se vérifient par simulations TCAD (**figure 4-11**). Quatre séries de courbes peuvent être identifiées:

- la courbe $N_a=N_d=10^{18}\text{cm}^{-3}$
- en dessous, les courbes pour lesquelles $\min(N_a,N_d)=10^{18}\text{cm}^{-3}$.
- encore en dessous, les courbes pour lesquelles $\min(N_a,N_d)=10^{19}\text{cm}^{-3}$.
- la courbe $N_a=N_d=10^{20}\text{cm}^{-3}$ est légèrement plus basse



Autrement dit, à dimensions fixées, c'est la zone la moins dopée qui va imposer la tension de claquage. Ainsi, il est suffisant de travailler avec les zones n et p identiquement dopées.

4.3.2. Modélisation d'une diode p-i-n

Il s'agit de déterminer les profils de charges, de champ électrique et de potentiel à l'intérieur de la structure polarisée.

4.3.2.1. Profil de charges

Le profil de charges est supposé ne dépendre que des profils de dopage. En particulier l'influence de la tension appliquée est négligée et la génération de porteurs par ionisation par impact est supposée ne pas perturber ce profil de charges.

Nous considérons une structure idéale:

- la zone i est faiblement dopée de type p. En effet, les substrats les plus purs présentent 10^{15} défauts par cm^3 . Ce niveau de dopage est électriquement équivalent à du silicium intrinsèque. Cependant, lors des simulations numériques, l'utilisation d'un dopage à 10^{15}cm^{-3} améliore la convergence.

- les zones n et p sont uniformément dopées
- la transition entre zone dopée et zone non dopée est infiniment abrupte. Cette hypothèse de jonction abrupte est tout à fait valable dans le cas de jonctions implantées ou épitaxiées.

Des zones de charge d'espace (ZCE) se créent aux jonctions p-i et i-n. Nous utiliserons les hypothèses usuelles:

- la charge est nulle hors de la zone intrinsèque et des ZCE (neutralité électrique)
- la charge est égale au dopage dans les ZCE et la zone intrinsèque

La **figure 4-12** schématise la structure et présente le profil de densité de charges considéré.

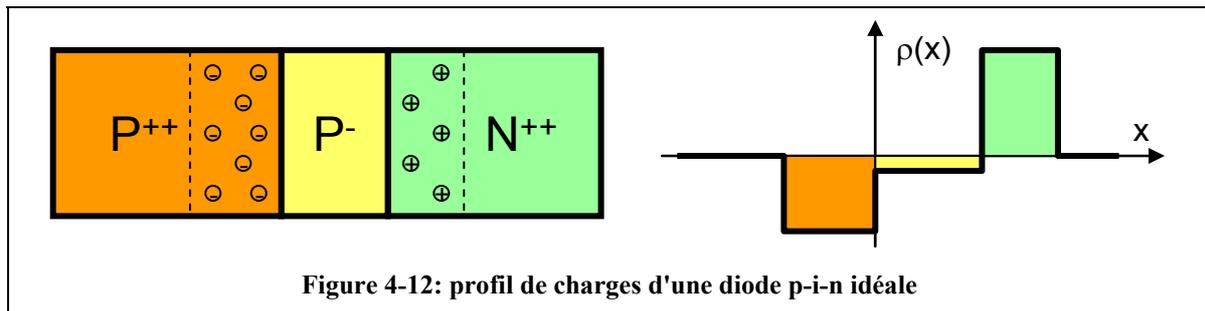


Figure 4-12: profil de charges d'une diode p-i-n idéale

4.3.2.2. Résolution de l'équation de Poisson

Le profil de charges donne la densité de porteurs. L'équation de Poisson:

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{\rho(x)}{\epsilon_{Si}} \quad (4.21)$$

peut alors être résolue pour déterminer le champ électrique et le potentiel électrique. Il est suffisant de se limiter au domaine où la densité de charges est non nulle.

Le champ électrique est obtenu par intégration de la densité de charges:

$$E(x) = \int \frac{\rho(x)}{\epsilon_{Si}} \cdot dx \quad (4.22)$$

Le potentiel est obtenu par intégration du champ électrique:

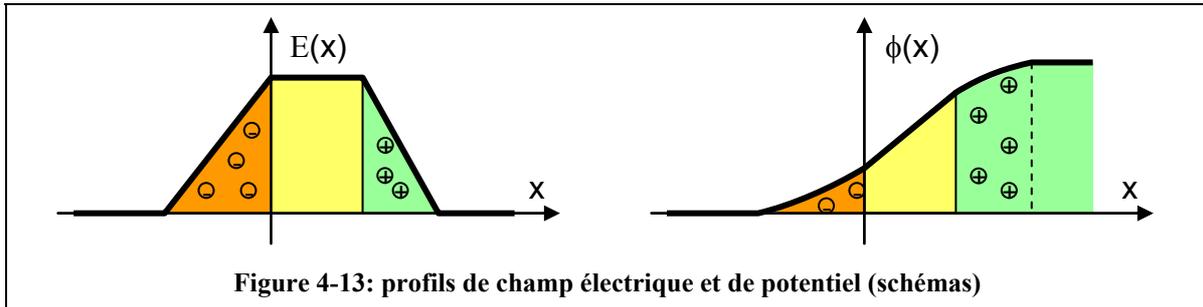
$$\phi(x) = \int -E(x) \cdot dx \quad (4.23)$$

Pour obtenir les solutions des équations (4.22) et (4.23), nous devons utiliser les équations de continuité du champ électrique et de potentiel, ainsi que les conditions limites suivantes:

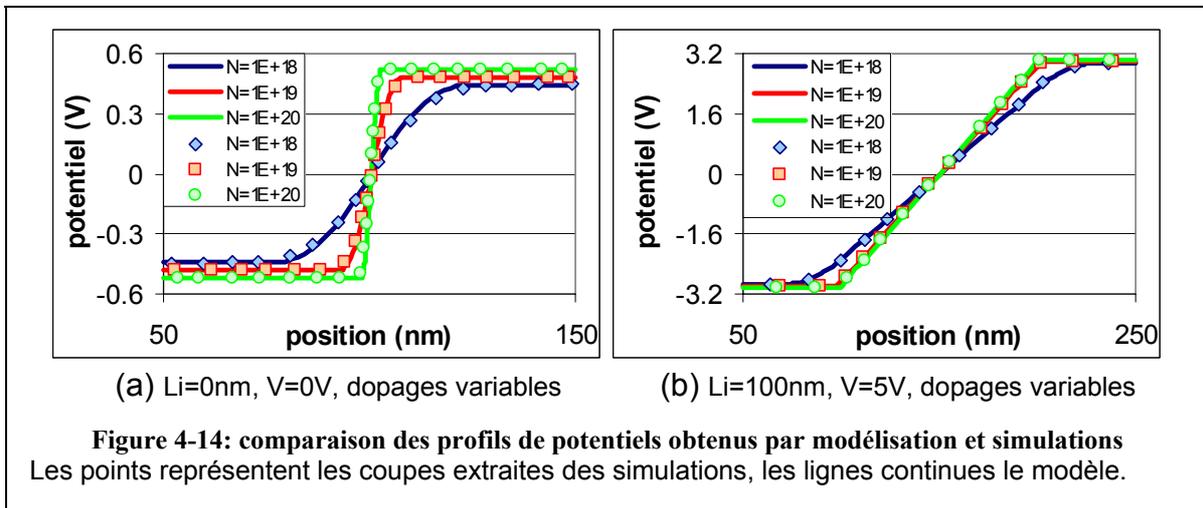
$$\begin{aligned} E(x=0) &= 0 \\ E(x=w) &= 0 \\ \phi(x=0) &= \phi_0 \\ \phi(x=w) &= \phi_w \end{aligned} \quad (4.24)$$

où ϕ_0 et ϕ_w sont les potentiels imposés par les électrodes.

La **figure 4-13** présente les profils de champ électrique et de potentiel obtenus après cette double intégration. La résolution complète de ce problème est donnée en annexe.



Les profils obtenus ont été comparés à ceux obtenus par simulations TCAD. Celles-ci ont été réalisées en utilisant le modèle de dérive diffusion, sans prise en compte de l'ionisation par impact. Les profils sont rigoureusement identiques, quels que soient les niveaux de dopage, les dimensions, la tension appliquée ou le matériau utilisé.



4.3.3. Tension d'avalanche d'une diode p-i-n

Comme pour une jonction p-n, l'avalanche s'observe lorsque l'on polarise la diode en inverse. C'est en effet sous l'action d'un champ électrique élevé que les porteurs peuvent acquérir suffisamment d'énergie pour générer une paire électron-trou par ionisation.

4.3.3.1. Résolution de l'intégrale d'ionisation

Le champ électrique étant déterminé dans toute la structure, nous pouvons à présent utiliser l'intégrale d'ionisation présentée au paragraphe 4.2.2.2 pour déterminer la tension de claquage:

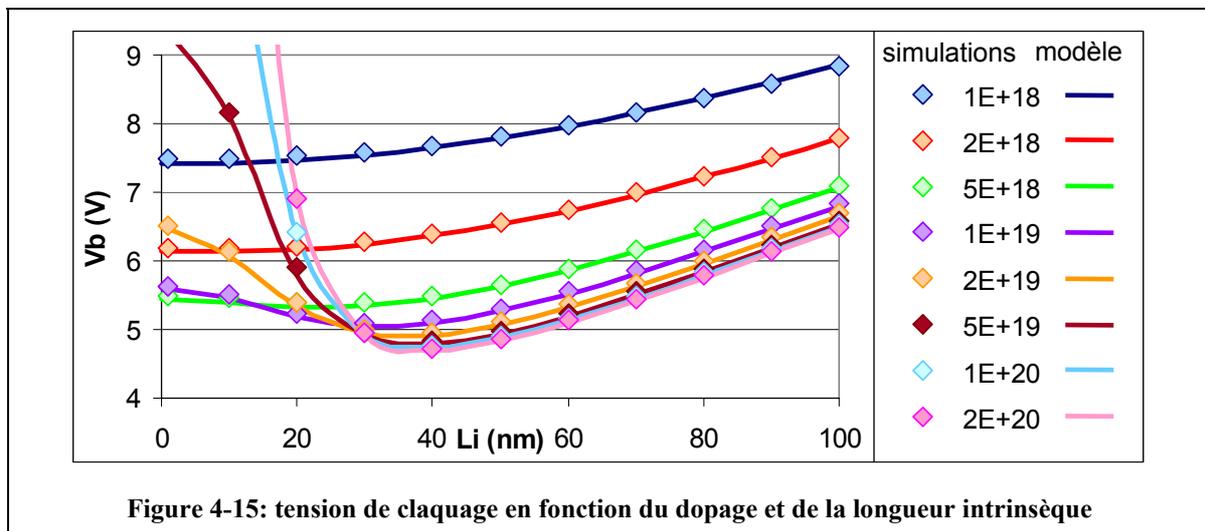
$$\int_0^w \alpha_p \exp\left(-\int_0^x (\alpha_p - \alpha_n) dx'\right) dx = 1$$

Cependant, il n'existe pas de solution analytique de l'intégrale d'ionisation [Sze'81]. De même, nous n'avons pas réussi à trouver une formule analytique approchée donnant des résultats

satisfaisants. Par contre, il est possible de résoudre numériquement cette équation à l'aide de la méthode de Müller [Müller'56].

L'utilisation du modèle analytique et de l'intégrale d'ionisation présente de nombreux avantages par rapport aux simulations:

- un temps de calcul réduit: une vingtaine de secondes contre une dizaine de minutes.
- moins de ressources.
- une plus grande plage d'utilisation: il a été impossible de faire converger les simulations dans les cas extrêmes (zones fortement dopées et longueur intrinsèque faible), alors que le couple modèle analytique/intégrale d'ionisation n'a pas posé de problème.
- une plus grande simplicité d'utilisation: le maillage lors des simulations doit s'adapter aux dimensions. Un ajustement automatique du maillage pour des structures de dimensions très



variées est un défi en soi. Or, ce problème est typiquement numérique et disparaît complètement lorsqu'on utilise le modèle analytique.

En outre, ces deux approches donnent des résultats parfaitement identiques (**figure 4-15**)

4.3.3.2. Remarques sur la tension d'avalanche

Deux choses apparaissent lorsque l'on examine les courbes $V_b(L_i)$ obtenues:

- en augmentant le dopage, les courbes tendent asymptotiquement vers une courbe limite
- sur cette courbe limite, il existe un minimum de tension de claquage.

Le premier point traduit le fait que la déplétion dans les zones dopées diminue lorsque le dopage augmente. Le cas limite est celui où ces déplétions sont nulles. Par conséquent, le champ électrique est simplement:

$$E = \frac{V}{L_i} \quad (4.25)$$

Dans ce cas, et en considérant que les coefficients d'ionisation des trous sont égaux à ceux des électrons, l'intégrale d'ionisation devient calculable analytiquement:

$$I = \alpha_n \cdot L_i = 1 \quad (4.26)$$

Dans le modèle de Van Overstraeten, le coefficient d'ionisation vaut:

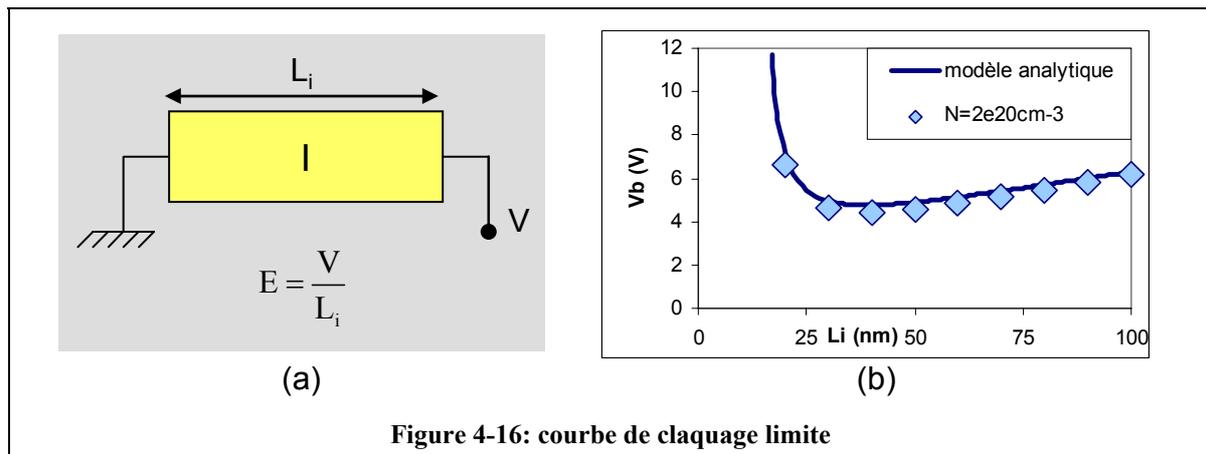
$$\alpha_n = a_n \exp\left(\frac{-b_n}{E}\right) \quad (4.27)$$

On peut donc déterminer analytiquement la courbe $V_b(L_i)$:

$$V_b = \frac{b_n \cdot L_i}{\ln(a_n \cdot L_i)} \quad (4.28)$$

La **figure 4-16** montre que c'est bien vers cette courbe que l'on tend en augmentant le dopage des zones n et p.

En outre, l'équation (4.26) permet d'expliquer le minimum de ces courbes. Pour avoir

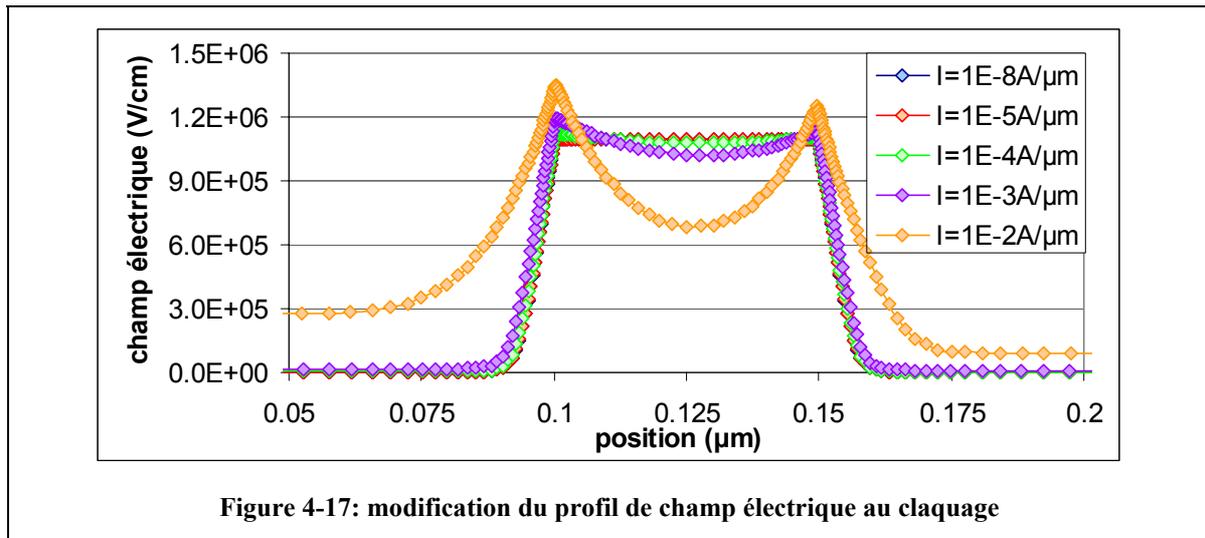


avalanche, il faut avoir un taux de génération suffisamment élevé sur une distance suffisamment longue. En diminuant la longueur L_i , le champ électrique et donc les coefficients d'ionisation augmentent. A grande longueur intrinsèque, faire varier le taux de génération est plus efficace que de faire varier la longueur. C'est le contraire aux faibles longueurs intrinsèques.

4.3.3.3. Limites de cette approche

Nous avons considéré que les charges créées par ionisation par impact ne viennent pas perturber le profil de charges considéré. Hors, ceci n'est vrai que pour de faibles valeurs de courant. Lorsque le courant qui traverse la structure augmente, le profil de charges est modifié, ce qui induit des modifications dans les profils de champ électrique et de potentiel.

La **figure 4-17** illustre ces modifications. Le profil de champ électrique est tracé à tension d'avalanche, pour différentes valeurs du courant. A faible courant, le profil est conforme à celui obtenu sans prise en compte de l'ionisation par impact. En revanche, lorsque le courant qui traverse la structure est élevé, le profil de champ électrique est fortement modifié, rendant alors l'approche de modélisation par perturbation inappropriée. La résolution auto-cohérente des équations de transport incluant la prise en compte de l'ionisation par impact est alors nécessaire.



En outre, d'autres phénomènes physiques peuvent intervenir lorsque la longueur intrinsèque tend vers 0, comme le tunnel bande à bande. La caractéristique sera alors perturbée, voire complètement imposée par ce phénomène.

4.3.4. Modélisation du courant d'une diode p-i-n

Modélisation du courant avant avalanche

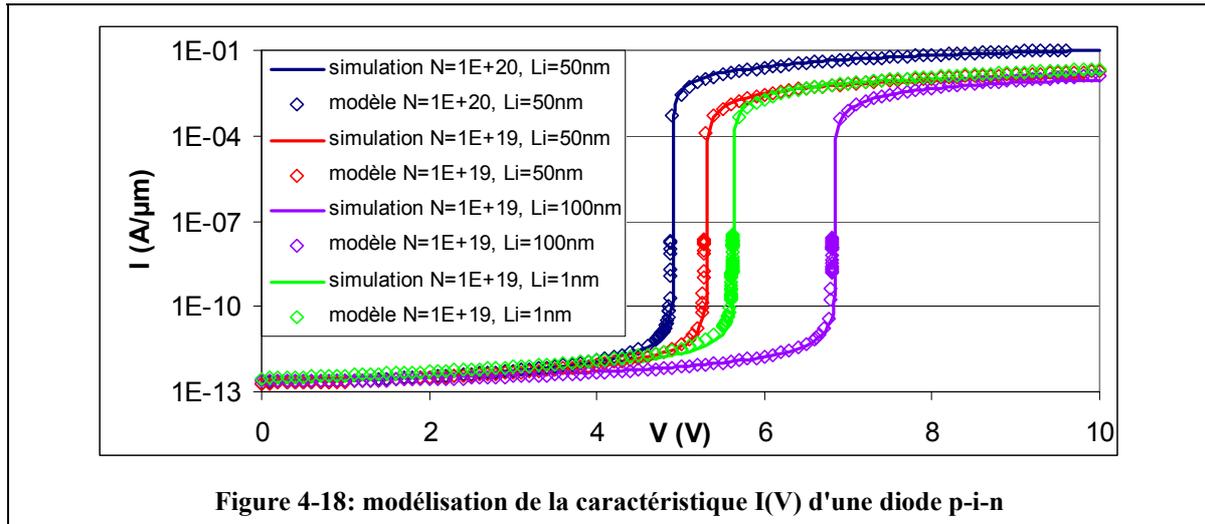
L'intégrale d'ionisation est un cas particulier du calcul général qui permet de déterminer le facteur de multiplication du à l'avalanche (annexe). La diode p-i-n étant modélisée quelle que soit la polarisation, et ce modèle étant valable jusqu'à l'avalanche, nous pouvons l'utiliser pour déterminer le coefficient de multiplication et donc le courant avant avalanche.

Modélisation du courant après avalanche

L'approche utilisée jusqu'ici n'est plus valable, nous l'avons souligné au paragraphe 4.3.3.3. Néanmoins, les essais menés avec cette approche ne présentent pas de limitation en terme de facteur de multiplication. En d'autres termes, le courant augmenterait indéfiniment s'il n'était

pas limité les résistances d'accès. Cela suggère de modéliser le dispositif par une simple résistance après la tension d'avalanche, qui représente les résistances d'accès.

La **figure 4-18** fournit la comparaison de caractéristiques $I(V)$ de diodes p-i-n entre la simulation TCAD et la modélisation décrite dans ce paragraphe. La reproduction est excellente, validant l'ensemble des hypothèses utilisées. Les résistances série variant avec les dimensions des zones dopées et leurs niveaux de dopage, chaque courbe se voit attribuer une valeur différente, reproduisant au mieux la simulation.



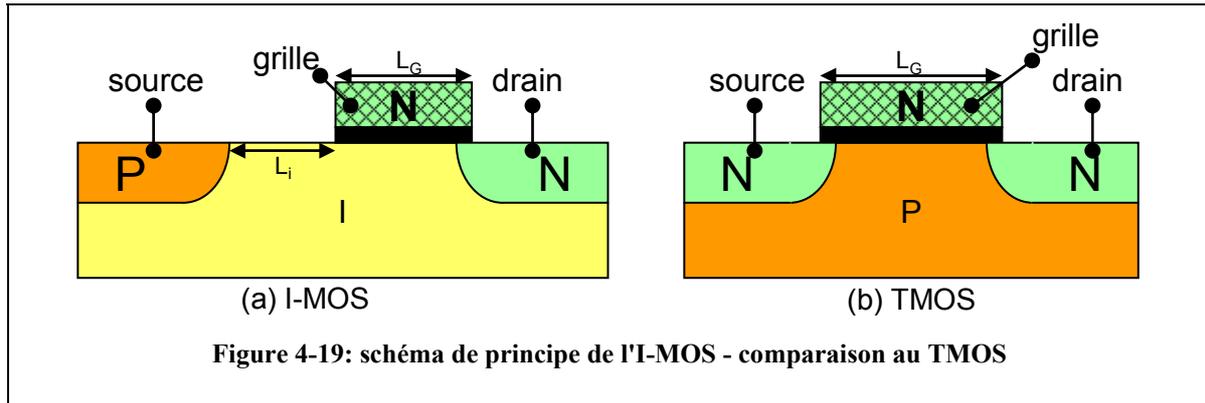
L'hypothèse selon laquelle le courant dépend au premier ordre des résistances d'accès est vérifiée expérimentalement [Charbuillet'06a,Charbuillet'06b], et sera abordée en détail au chapitre 6.

4.4. Le transistor I-MOS

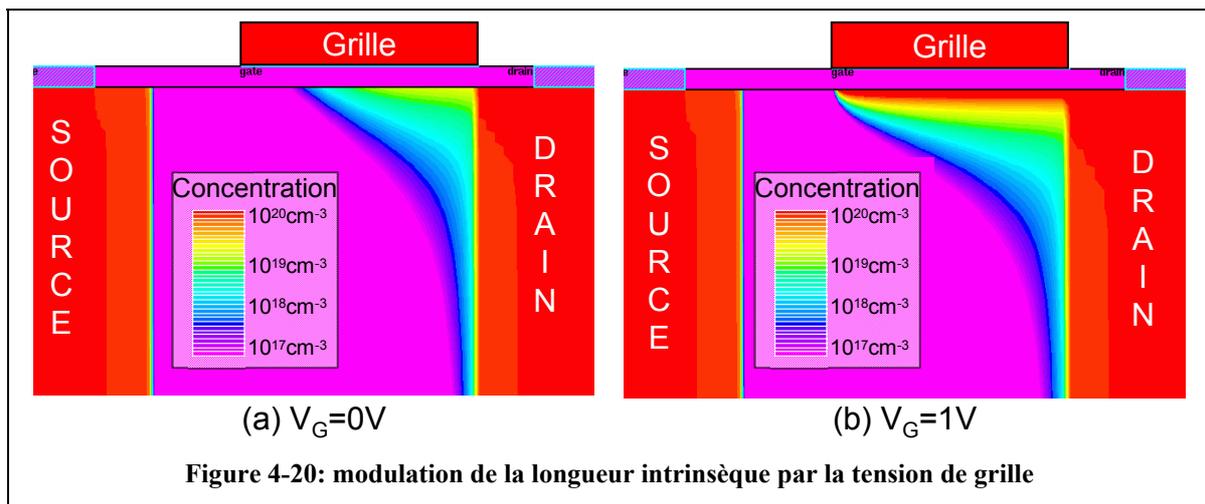
4.4.1. Présentation

Description

Le transistor I-MOS se présente comme d'une diode p-i-n partiellement surmontée d'une électrode isolée par une fine couche de diélectrique (**figure 4-19a**). Cette électrode recouvre une partie de la zone intrinsèque et s'étend jusqu'à une des deux zones dopées. Dans ce qui suit, nous considérons qu'il s'agit de la zone dopée n. Par analogie au TMOS, l'anode de la diode p-i-n est appelée source, la cathode est appelée drain et l'électrode supérieure grille. Notons que si la symétrie du TMOS ne permet de distinguer la source du drain, dans un I-MOS le drain est la zone légèrement couverte par la grille. Les dimensions caractéristiques d'un I-MOS sont la longueur intrinsèque recouverte par la grille L_G et longueur non recouverte L_i .



Par simulation, il est possible d'extraire une coupe unidimensionnelle pour déterminer le profil de charges près de l'interface semi-conducteur / diélectrique de grille. En fonction de la tension de grille, on peut se ramener à une situation comparable à une diode p-i-n de longueur intrinsèque $L_{ei}=L_i$ ou $L_{ei}=L_i+L_G$ (L_{ei} représente la distance électrique entre l'anode et la cathode).



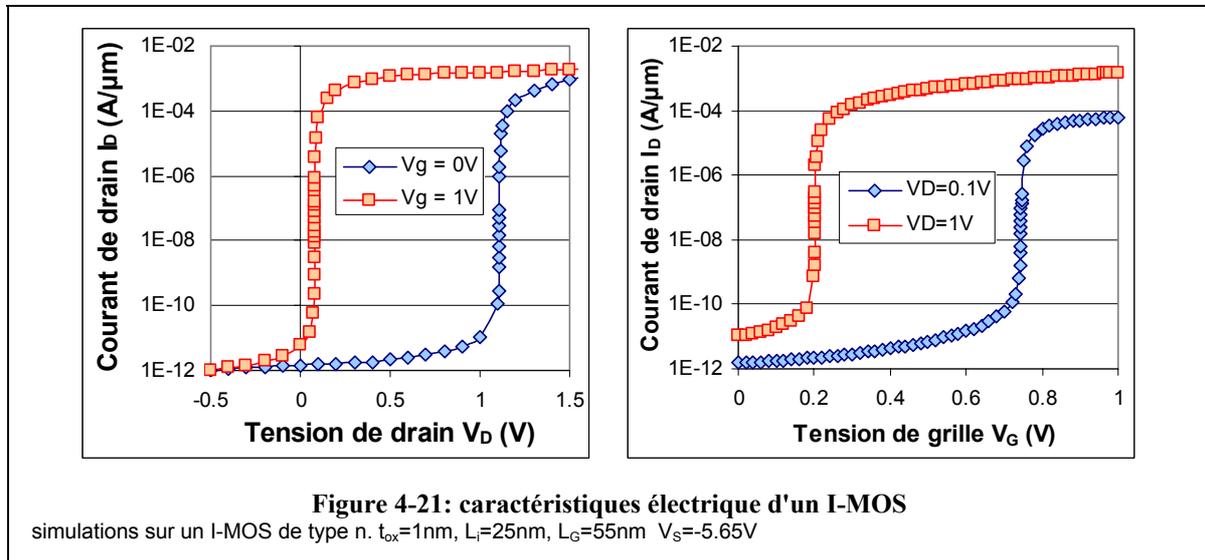
En polarisant cette diode en inverse grâce aux électrodes source et drain, il est possible de déclencher le phénomène d'avalanche. La tension de claquage étant dépendante de la distance anode/cathode, elle dépend donc de la polarisation de grille.

Fonctionnement

Les dimensions L_i et L_G sont choisies de manière à ce que les deux tensions de claquage $V_B(V_G=0V)$ et $V_B(V_G=V_{DD})$ soient de part et d'autre de la tension d'alimentation V_{DD} (figure 4-21a). Ainsi, en régime de blocage ($V_G=0$), le courant de drain est faible. En régime de conduction ($V_G=V_{DD}$), un courant d'avalanche apparaît, mettant le dispositif en conduction.

Un intérêt majeur est visible sur la caractéristique $I_D(V_G)$ (figure 4-21): la pente sous le seuil ou assimilée $S = \frac{\partial V_G}{\partial \log(I_D)}$ n'est plus limitée à 60mV/dec. Ceci sera discuté au **paragraphe**

4.4.3.3.



Un dispositif complémentaire

Dans la structure que nous venons de présenter, la grille est localisée du côté dopé de type n et la grille polarisée positivement. Par conséquent, les porteurs accumulés sous la grille sont des électrons. Par analogie au TMOS, nous appellerons donc ce dispositif I-MOS de type n. En déplaçant la grille du côté dopé p et en changeant le signe des polarisations, on aboutit à un dispositif symétrique dont les porteurs accumulés sous la grille sont des trous. On parlera donc d'un dispositif I-MOS de type p. Pour déclencher le phénomène d'avalanche, il convient toujours de polariser la diode p-i-n en inverse.

La publication d'origine de l'I-MOS [Gopalakrishnan'02] ne propose pas de symbole pour ce dispositif. Une notation proche de celle du TMOS a été proposée ailleurs, mais elle ne permet d'identifier rapidement le type du dispositif, et reste très proche de celle du TMOS. C'est pourquoi nous introduisons une autre notation (**figure 4-22**).

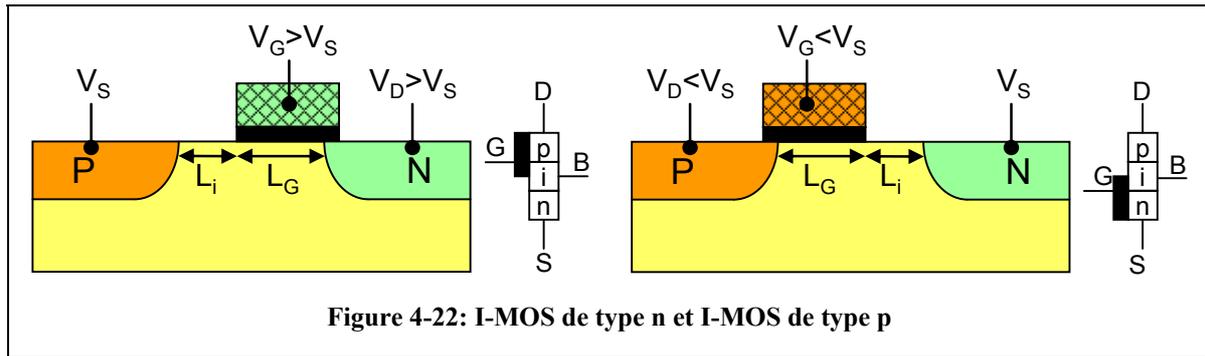


Figure 4-22: I-MOS de type n et I-MOS de type p

Contrairement au TMOS, l'I-MOS est un dispositif bipolaire, c'est-à-dire que le courant est à la fois un courant d'électrons et un courant de trous. En effet, les porteurs sont créés par paires par avalanche. Ce sont donc les deux types de porteurs qui participent au courant. Les électrons se déplacent vers la zone dopée de type n, de potentiel plus élevé, alors que les trous se déplacent vers la zone dopée de type p, de potentiel moins élevé (figure 4-23).

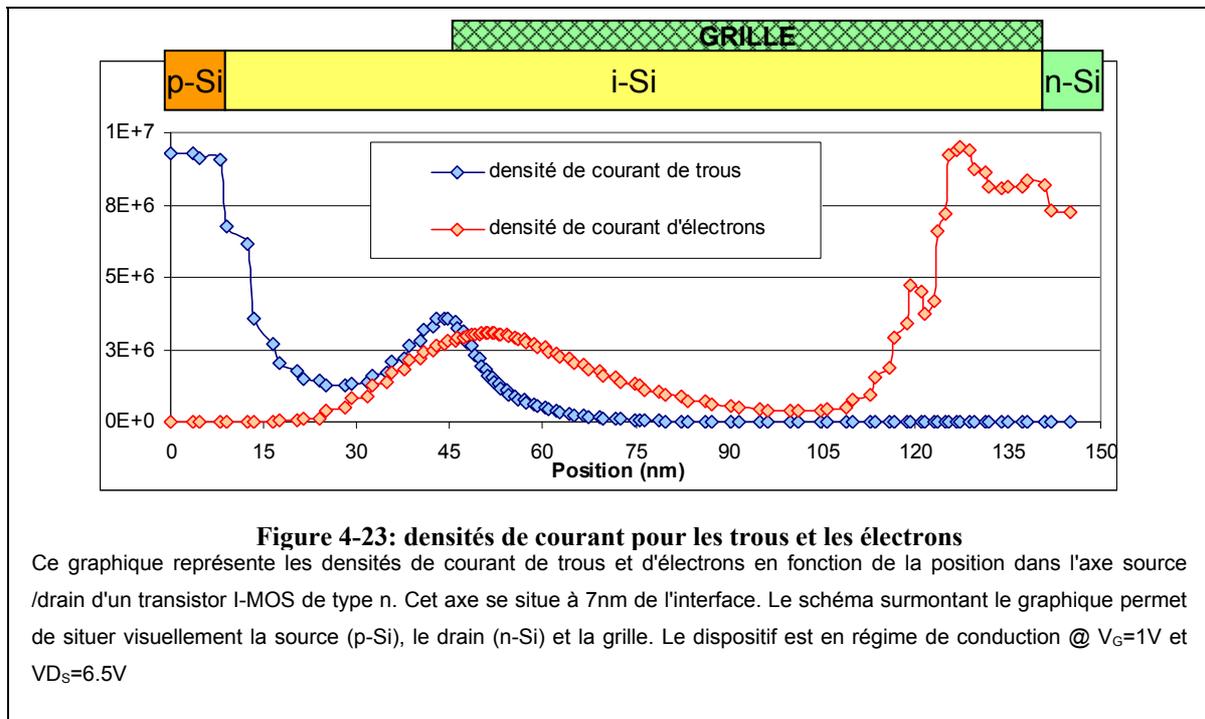


Figure 4-23: densités de courant pour les trous et les électrons

Ce graphique représente les densités de courant de trous et d'électrons en fonction de la position dans l'axe source /drain d'un transistor I-MOS de type n. Cet axe se situe à 7nm de l'interface. Le schéma surmontant le graphique permet de situer visuellement la source (p-Si), le drain (n-Si) et la grille. Le dispositif est en régime de conduction @ $V_G=1V$ et $V_{D_S}=6.5V$

4.4.2. Polarisation

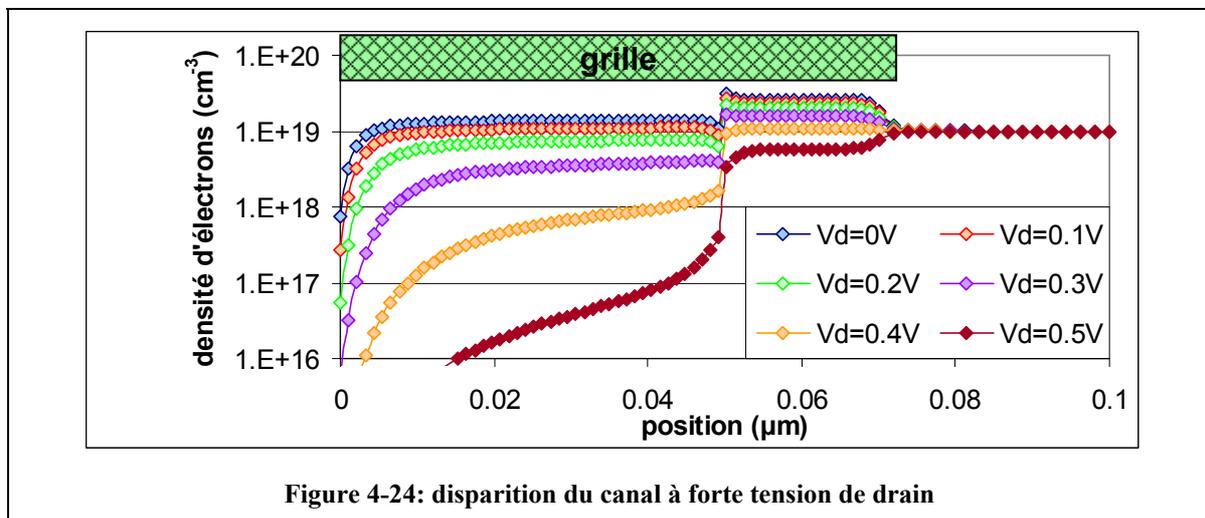
Le principe de fonctionnement de l'I-MOS incite à le polariser de manière identique au TMOS. Ceci permettrait en outre de conserver la même réalisation de circuits élémentaires. Cependant, l'architecture spécifique va impliquer de modifier le schéma de polarisation habituel.

4.4.2.1. Polarisation identique à un TMOS

- $V_s=0V$ à la source (masse)
- $0V$ à $V_{DD}>0V$ à la grille
- $0V$ à $V_{DD}>0V$ au drain

Dans le transistor MOS conventionnel, le canal est formé par des porteurs accumulés sous la grille par effet électrostatique. Ces porteurs peuvent *a priori* être fournis par la source, le drain ou le substrat. Cependant, le canal est formé par des porteurs minoritaires. Le substrat ne peut donc les fournir. De plus, le déplacement des porteurs se fait de la source vers le drain: dans un nmos, le drain est au potentiel le plus élevé alors que le canal est formé d'électrons. La source joue donc le rôle de réservoir de porteurs pour la formation du canal. Ceci est possible car la zone de source est légèrement recouverte par la grille.

Dans un I-MOS, la source est séparée du "canal" par la zone intrinsèque. Elle ne peut donc pas jouer son rôle de réservoir de porteurs. C'est alors le drain qui va fournir les porteurs nécessaires à la formation du canal. Mais à forte polarisation de drain, l'effet électrostatique de la grille est insuffisant: le canal disparaît. Ce phénomène est comparable au pincement dans un transistor MOS en régime de saturation. Il est amplifié par le fait que la source peut difficilement alimenter le canal en porteurs.



La **figure 4-24** représente les profils de densité d'électrons dans le canal en fonction de la tension de drain. La grille est alors polarisée à 1V. La densité de porteurs sous la grille chute dramatiquement, même pour des tensions de drain inférieures à la tension de grille.

Les charges formant le canal sont attirées par effet électrostatique. Si l'on se place à tension de grille donnée (1V) et que l'on fait croître la tension de drain par rampes temporelles

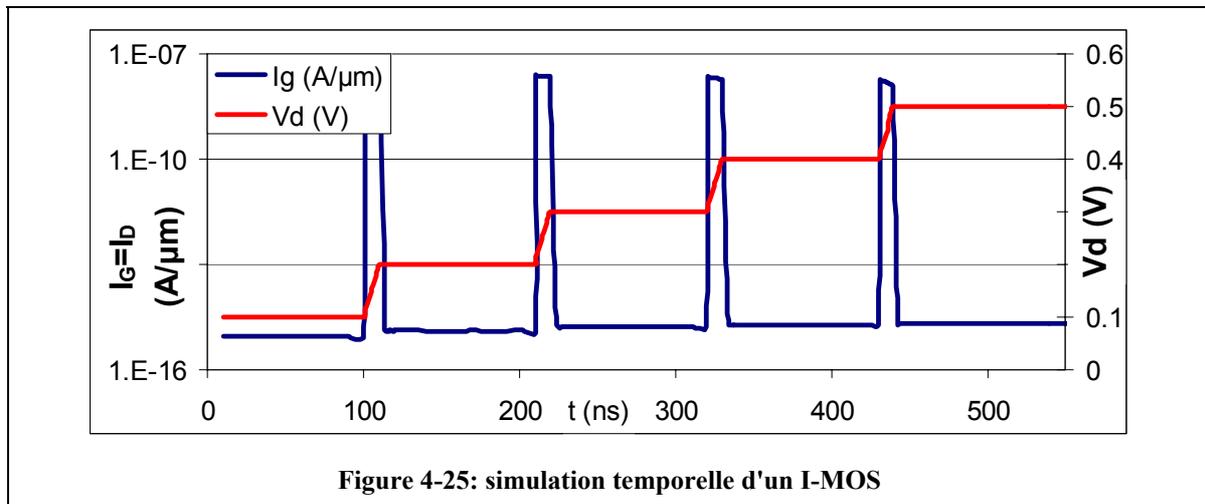


Figure 4-25: simulation temporelle d'un I-MOS

successives, on observe à la fois un courant d'électrons au niveau du drain et un courant de trous de même valeur au niveau de la grille (**figure 4-25**).

Le courant d'électrons est la manifestation dynamique de la perte de charge dans le canal observée en statique (**figure 4-24**). Le courant de trous s'explique par la neutralité de charges aux bornes de la capacité de l'oxyde de grille. D'autre part, nous vérifions la relation:

$$\Delta Q = C_{ox} \cdot \Delta V_D \quad (4.29)$$

où ΔQ la variation de charges constatée pour une variation de tension de drain ΔV_D .

4.4.2.2. Nouvelle polarisation

La polarisation classique ne permet pas de déclencher l'ionisation par impact. Une autre polarisation doit être considérée. Pour cela, deux points doivent être considérés:

- la grille doit être polarisée du même signe que le drain pour attirer le bon type de porteurs
- le champ électrique entre source et drain doit être important.

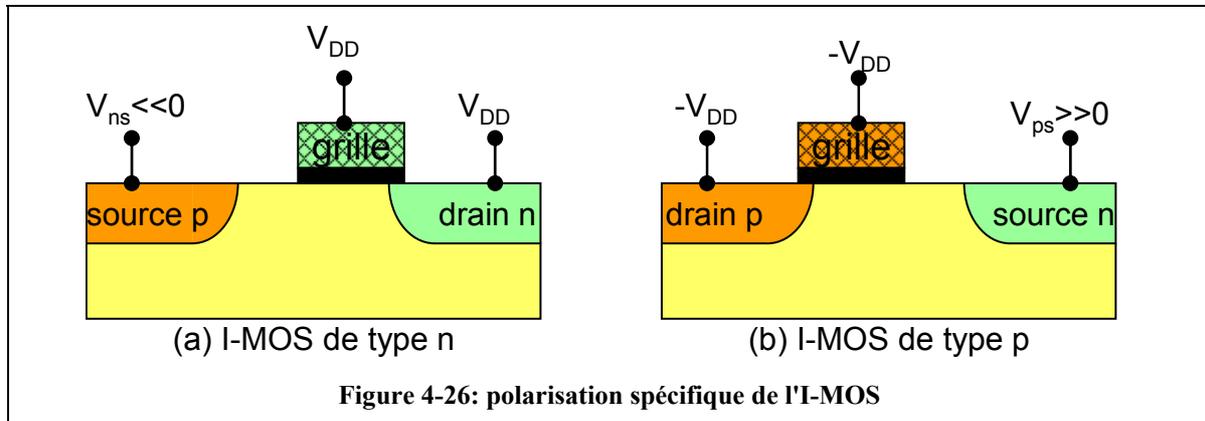
Une solution simple consiste à polariser la source: nous obtenons ainsi les deux effets escomptés. Ce principe a déjà été proposé dans la littérature ([Gopalakrishnan'05a], [Choi'05c]). Le principe de cette approche est de dissocier la tension d'alimentation de la tension du signal. Une tension élevée est nécessaire pour obtenir de l'avalanche, mais le signal qui parcourt un circuit n'a pas besoin de la même amplitude.

Ainsi, le transistor I-MOS de type n sera polarisé de la façon suivante (**figure 4-26a**):

- $V_{ns} \ll 0V$ constant à la source
- $0V$ à $V_{DD} > 0V$ à la grille
- $0V$ à $V_{DD} > 0V$ au drain

le pmos sera lui polarisé (**figure 4-26b**):

- $V_{ps} \gg 0V$ constant à la source
- $0V$ à $-V_{DD} < 0V$ à la grille
- $0V$ à $-V_{DD} < 0V$ au drain



Les implications de cette polarisation spécifique sur les circuits élémentaires seront discutées au chapitre 5.

4.4.2.3. Valeur des tensions de polarisation

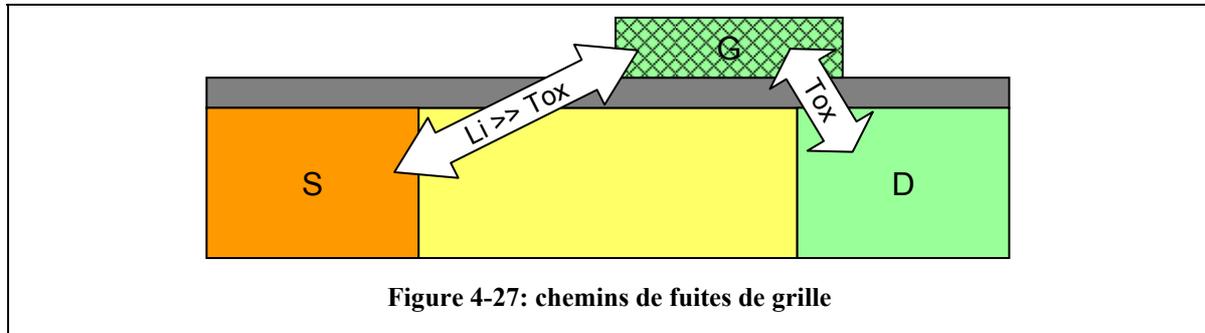
L'ionisation par impact dans les semi-conducteurs est un procédé qui apparaît à relativement haute tension. En effet, un porteur passe de la bande de valence à la bande de conduction. L'énergie nécessaire est donc de l'ordre de la hauteur de la bande interdite. Par conséquent, un porteur capable de générer des porteurs par avalanche doit avoir une énergie au moins égale à la bande interdite. Notons que dans le transistor MOS, l'avalanche est à l'origine de courant substrat, et ce même à des tensions inférieures au gap. Ceci peut être expliqué par l'activation thermique de l'ionisation par impact [Su'02]. Cependant, nous considérons ici l'avalanche non comme un mécanisme secondaire, mais bien comme fondamental pour le dispositif.

Pour le silicium, la tension minimale à appliquer est de l'ordre de 5V pour obtenir de l'avalanche [Charbuillet'06b]. Cette valeur est relativement élevée en regard des tensions d'alimentations actuelles, de l'ordre du volt. Ceci pose un certain nombre de questions: sur la tenue et la fiabilité de l'oxyde de grille, les courants de fuites, la consommation.

Tenue de l'oxyde de grille

Il a été rappelé au chapitre 1 que l'épaisseur de l'oxyde de grille est adaptée à la tension d'alimentation. Sous 5V, le champ de claquage du SiO₂ impose une épaisseur minimale de 5nm, ce qui est une valeur élevée en regard des épaisseurs actuelles (<15Å).

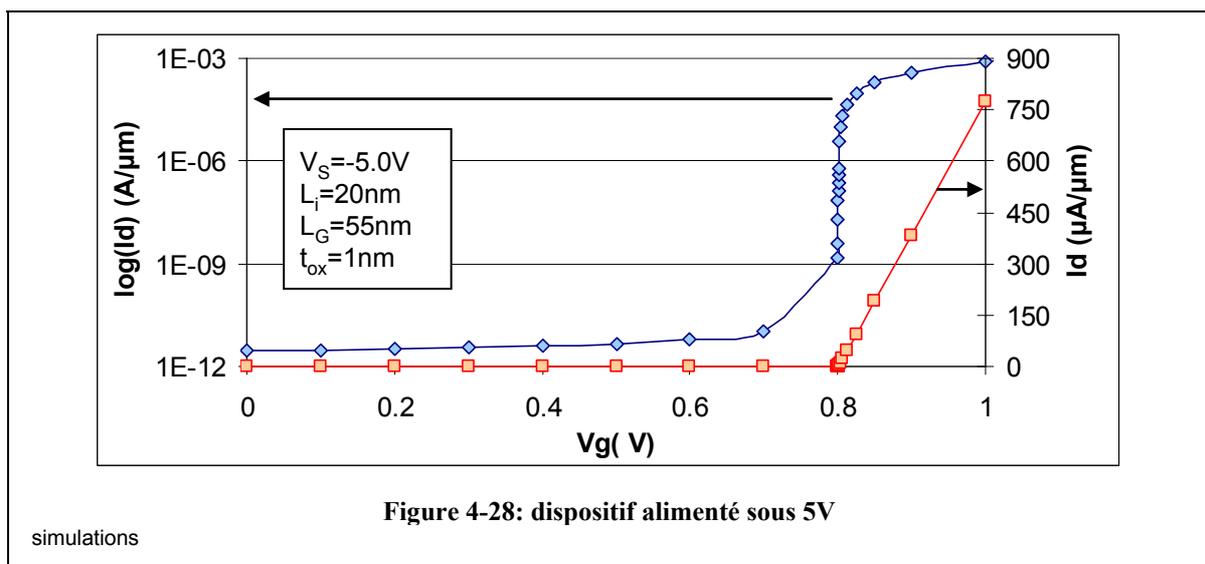
Cependant, la différence de potentiel entre grille et drain est conventionnelle, de l'ordre du volt. Elle est plus importante entre la source et la grille. Mais ces deux électrodes sont distantes de L_i (**figure 4-27**) soit environ 25nm. C'est une valeur bien supérieure aux 5nm correspondant au champ de claquage de la silice.



Nécessité d'une tension d'alimentation spécifique

Le nouveau schéma de polarisation impose l'introduction de deux sources de tension complémentaires V_{SN} et V_{SP} . Or, ajouter des sources de tensions sur un circuit prend énormément de place, et doit donc être évité autant que possible. C'est pourquoi nous devons adapter nos dispositifs aux tensions déjà existantes.

L'alimentation 0V/+5V est généralement disponible: c'est historiquement la tension de la logique TTL. Les simulations montrent qu'il est possible de réaliser des dispositifs dont la source est à polarisée à -5V, la grille et le drain à V_{DD} .



Consommation

La consommation électrique est un élément prépondérant. Pour les appareils nomades, cela impacte directement l'autonomie. Pour les appareils fixes, où le besoin d'une faible consommation est moins évident, logique est de consommer moins pour limiter l'échauffement, et ainsi éviter tout système de refroidissement supplémentaire, coûteux et source de bruit (ventilateur d'ordinateur par exemple).

La consommation de circuits élémentaires réalisés en I-MOS est discutée dans le chapitre 5.

4.4.3. Performances

L'I-MOS est basé sur un phénomène physique différent de celui du TMOS. Par conséquent, ses performances ne sont pas soumises aux mêmes limites. Dans cette partie, les performances statiques élémentaires (courant de conduction I_{on} , courant de fuite I_{off} , tension de seuil V_{th} et pente sous le seuil S) sont abordées d'un point de vue théorique enrichi par les données expérimentales disponibles dans la littérature et obtenues dans le cadre de nos travaux.

4.4.3.1. Courant de fuite

Dans ce régime, les polarisations sont:

$$-V_{ns} \ll 0V \sim -5V$$

$$-V_G = 0V$$

$$-V_D = V_{DD} \sim 1V$$

La tension de grille est trop faible pour obtenir de l'avalanche. Les origines possibles de courant de fuite sont donc le courant inverse, les fuites de grille à travers l'oxyde et les fuites de jonction, comme pour le TMOS

Courant inverse

Dans le transistor MOS conventionnel, le courant à l'état bloqué n'est pas nul. Les porteurs qui ont suffisamment d'énergie thermique peuvent passer de la source au drain par dessus la barrière de potentiel modulée par l'effet de champ de la grille. C'est ce courant thermoélectronique qui est l'origine de la non idéalité du blocage.

Dans un transistor I-MOS, la configuration bloquée est celle d'une jonction polarisée en inverse. Par conséquent, la non idéalité du blocage est due au courant en inverse de cette jonction, qui vaut:

$$J = J_s \cdot \left[\exp\left(-\frac{q \cdot V_{DS}}{k \cdot T}\right) - 1 \right] \quad (4.30)$$

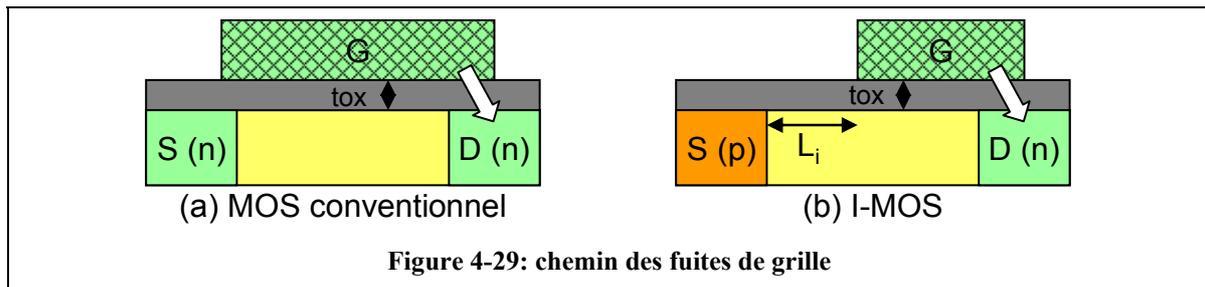
où J_S est le courant de saturation en inverse. Ce courant est généralement de l'ordre de 10^{-12} A/ μm à 10^{-15} A/ μm , ce qui est plus faible que le courant thermoélectronique du MOS.

Une autre façon de voir les choses est de considérer la jonction p-n de l'I-MOS comme une barrière de potentiel de hauteur $V_{DS}+V_{bi}\sim 7\text{V}$, ce qui est bien plus élevé que la barrière de potentiel du MOS conventionnel qui vaut $V_{TH}\sim 0.25\text{V}$.

Les fuites de grille

Ce sont les fuites de la grille vers le drain à travers l'oxyde de grille (**figure 4-29**). En effet, celui-ci n'est pas un isolant parfait, et ce d'autant plus qu'il est fin. Ce courant dépend essentiellement de l'épaisseur d'oxyde et de la différence de potentiel entre grille et drain. Or, les situations entre transistor MOS conventionnel et I-MOS sont semblables:

- les oxydes de grille sont de l'ordre de 2nm
- dans le pire des cas, la grille est polarisée à 0V et le drain à $V_{DD}\sim 1\text{V}$ pour les deux dispositifs.



Par conséquent, les fuites de grille sont semblables.

Dans le transistor MOS conventionnel, la recherche d'oxydes de grille extrêmement minces est motivée par l'augmentation de la capacité de grille $C_{ox}=\epsilon_{ox}/t_{ox}$, qui a pour effet d'améliorer le contrôle de la grille sur le canal d'inversion.

Les transistors I-MOS n'étant pas basés sur le transport d'une charge d'inversion, ils sont moins sensibles à la capacité d'oxyde. Il est donc possible de relâcher la contrainte sur l'épaisseur d'oxyde sans dégradation notable des performances [Gopalakrishnan'05a]. Ceci a pour effet de limiter les courants de fuite de grille.

4.4.3.2. Courant de conduction

Dans un transistor MOS conventionnel, le courant de conduction I_{on} dépend principalement:

- de la charge d'inversion
- de la mobilité des porteurs
- dans une moindre mesure des résistances d'accès dans une technologie bien maîtrisée

Ainsi, la saturation de la charge d'inversion à forte polarisation de drain (pincement du canal) est à l'origine de la saturation du courant, visible sur les caractéristiques $I_D(V_D)$ d'un transistor. L'amélioration de la mobilité, par exemple par contrainte du silicium, permet d'améliorer les performances des transistors [Payet'06].

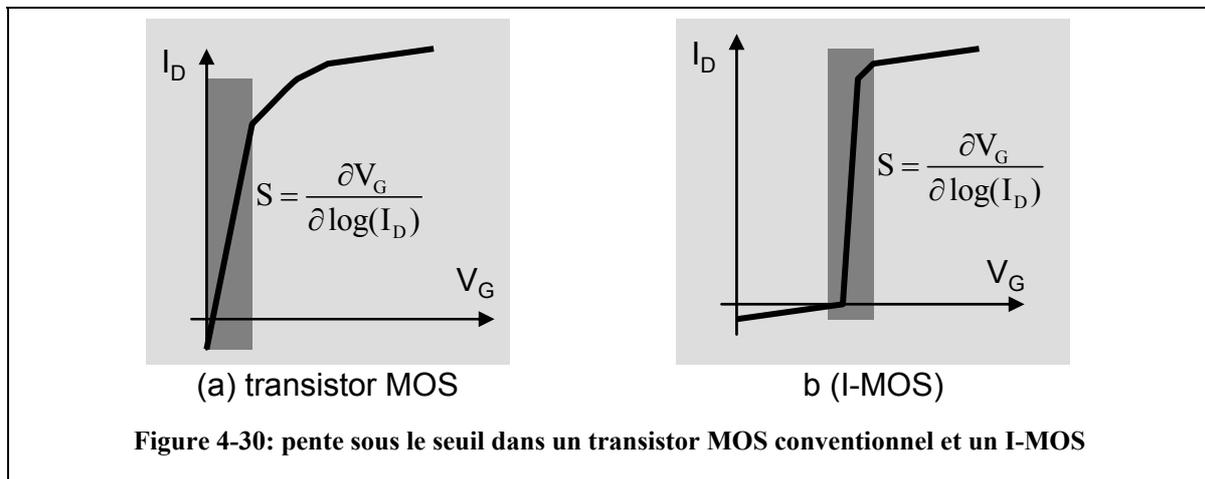
Dans un transistor I-MOS, le courant de conduction est le courant d'avalanche d'une diode p-i-n. Par conséquent, il est également principalement limité par les résistances d'accès [Charbuillet'06a]. C'est pourquoi des I-MOS ont permis de mesurer des valeurs de courant jamais atteintes pour les transistors MOS conventionnels ($4800\mu\text{A}/\mu\text{m}$, [Charbuillet'06b]).

4.4.3.3. La pente sous le seuil

La pente sous le seuil S est la quantité qui caractérise la brutalité avec laquelle un dispositif commute de l'état bloqué à l'état passant. Pour le transistor MOS conventionnel, le courant de drain variant exponentiellement avec la tension de grille en régime de blocage (**figure 4-30a**), la pente sous le seuil est donc définie par:

$$S = \frac{\partial V_G}{\partial \log(I_D)} \quad (4.31)$$

Pour le transistor I-MOS, bien que la caractéristique $I_D(V_G)$ soit différente, nous conserverons cette notion de pente sous le seuil définie de manière identique, dans la zone de transition entre régime de blocage et régime de conduction (**figure 4-30b**).



Dans le transistor MOS, celle-ci est intrinsèquement limitée à $60\text{mV}/\text{dec}$. Ceci est dû au fait que le courant est un déplacement de charges obéissant à une statistique proche de celle de Boltzmann (voir chapitre 2).

Dans un I-MOS, le courant est lié à la création de paires électron/trou par avalanche. Cette limite ne doit donc plus exister. Ceci est étayé de trois façons différentes: les mesures

expérimentales, les simulations, et une modélisation qui permet de donner une limite prospective.

Données expérimentales

Les pentes sous le seuil mesurées présentées dans la littérature varient entre 2 et 12mV/dec. Elles sont généralement plus faibles pour les dispositifs de type n que de type p.

publication	[Gopalakrishnan'02]	[Choi'04a]		[Choi'04b]		[Choi'05b]		[Toh'05]		[Mayer'06]
type	pmos	nmos	nmos	pmos	nmos	pmos	nmos	nmos	pmos	
S (mV/dec)	~10	7.2	5.5	12.2	7.5	11.8	4.5	24	2	

Tableau 4: pentes sous le seuil expérimentales

Simulations

Nous disposons de simulations à partir desquelles nous pouvons extraire des valeurs de pente sous le seuil. D'autres sont disponibles dans la littérature [Choi'05c]. Celles-ci varient de à .

Modélisation – approche au premier ordre

La pente sous le seuil peut être estimée à l'aide d'un modèle simple dérivé du modèle de la diode p-i-n. Une coupe unidimensionnelle extraite d'un I-MOS selon l'axe source/drain donne une structure de diode p-i-n (**figure 4-31**).

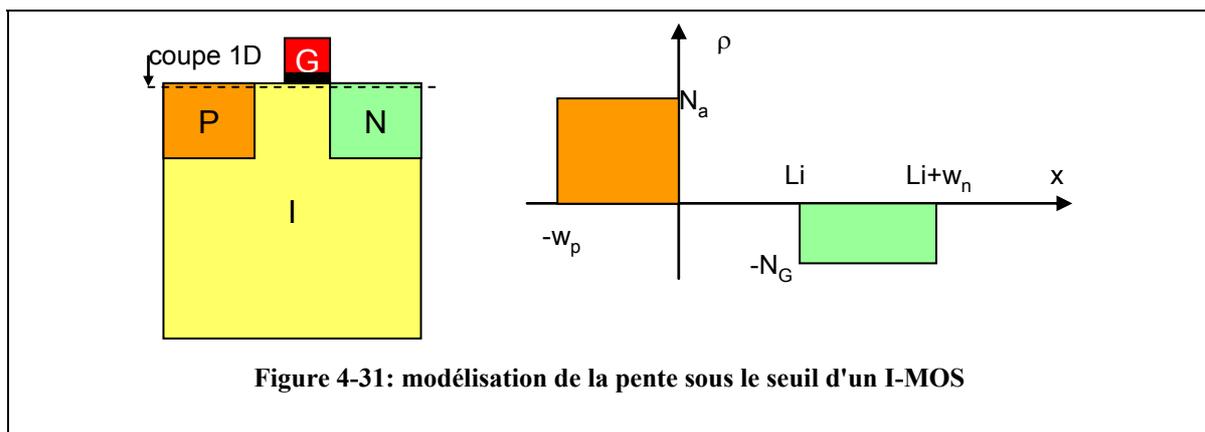


Figure 4-31: modélisation de la pente sous le seuil d'un I-MOS

En supposant que les porteurs sous la grille obéissent à une statistique de Boltzmann, et que le couplage entre tension de grille et potentiel de surface est parfait, la densité de porteurs sous la grille peut s'écrire en fonction de la tension appliquée:

$$N_G = N_{G0} \cdot \exp\left(\frac{V_G}{u_t}\right) \quad (4.32)$$

où N_{G0} est une constante homogène à un dopage.

D'autre part, le modèle de la diode p-i-n nous permet de calculer le gain de la structure associé à l'avalanche. En considérant le courant initial indépendant de la tension de grille, la pente peut s'exprimer en fonction de la tension de grille et du gain M_p :

$$S = \frac{\partial V_G}{\partial \log(M_p)} \quad (4.33)$$

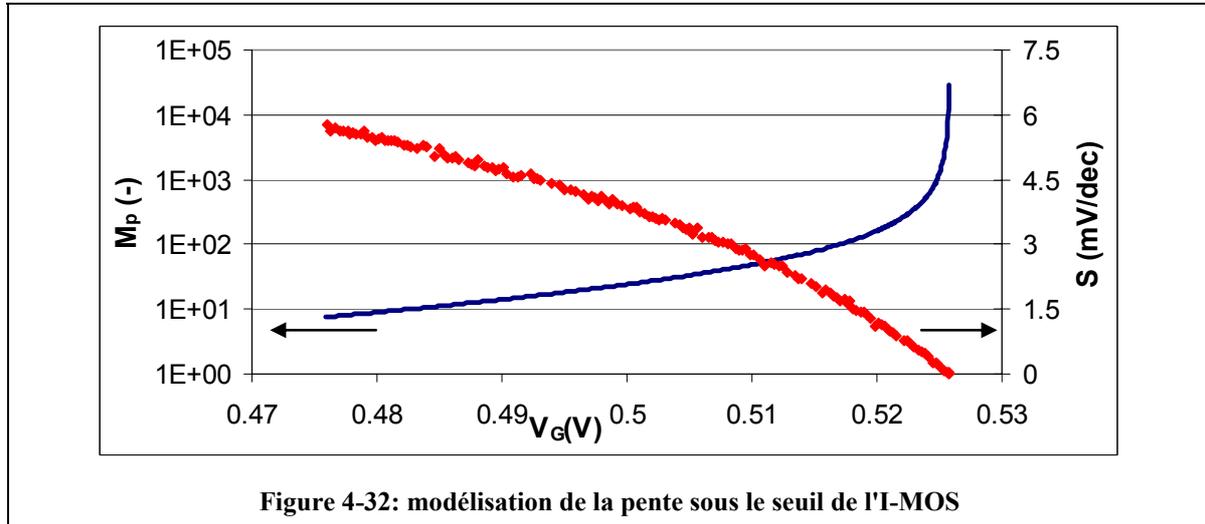


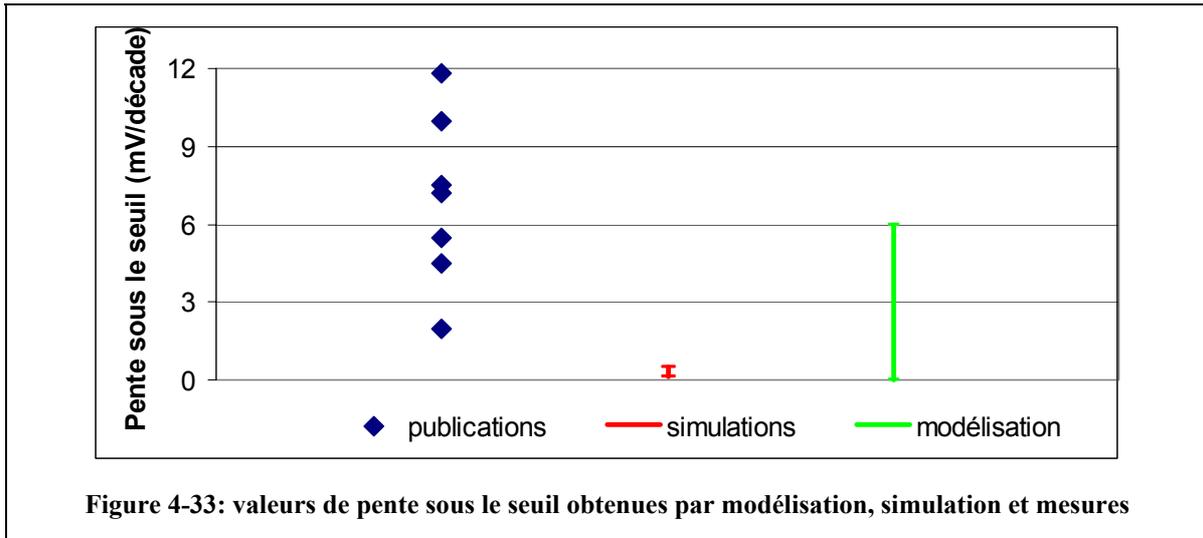
Figure 4-32: modélisation de la pente sous le seuil de l'I-MOS

Cette méthode permet de calculer la pente sous le seuil, qui contrairement au TMOS dépend de la longueur de grille.

Les valeurs calculées varient entre 0 et 6mV/dec près de la tension de claquage. A aucun moment nous n'avons relevé de limite inférieure.

Comparaison entre données expérimentales, simulations et modélisation

La **figure 4-33** compare les valeurs de pente sous le seuil expérimentales relevées dans le **tableau 4**, ainsi que les valeurs relevées pour nos simulations et celles résultant de notre modélisation.



La disparité des mesures expérimentales n'est absolument pas reproduite par nos simulations. Cela indique que ces dernières ne prennent pas en compte tous les phénomènes influant sur la pente.

En revanche, les résultats de modélisation couvrent les meilleures valeurs expérimentales. Ceci amène à penser que la pente sous le seuil des dispositifs I-MOS n'a pas de limite inférieure.

Conclusion du chapitre

L'I-MOS est un dispositif basé sur le phénomène d'avalanche. Il consiste à moduler la tension de claquage d'une diode p-i-n par effet de champ, de manière comparable à la façon dont la conductivité d'un semi-conducteur est modulée dans le TMOS.

Pour étudier ce dispositif, nous avons commencé par choisir un modèle d'avalanche approprié: simple et reflétant la réalité. Ceci nous a permis d'étudier et de modéliser la caractéristique $I(V)$ d'une diode p-i-n, structure servant de référence à l'I-MOS.

Puis, nous avons pu étudier le fonctionnement de l'I-MOS. En particulier, le schéma de polarisation habituel du TMOS ne peut être appliqué. Il doit être adapté. L'utilisation de tensions élevées (5V) pour obtenir l'avalanche ne semble pas poser de problème insurmontable.

Enfin, les performances statiques (courant de conduction, courant de fuite, tension de seuil et pente sous le seuil) ont été évaluées, montrant un gain par rapport au TMOS.

Cependant, les performances statiques d'un dispositif ne reflètent pas forcément les performances des circuits. Ceci est d'autant plus vrai que l'I-MOS diffère grandement du TMOS. Dans le chapitre suivant, nous étudierons les circuits élémentaires réalisés en I-MOS avec des dispositifs complémentaires (type n et type p). Nous verrons comment adapter le schéma de polarisation spécifique de l'I-MOS à ces circuits, ainsi que les performances statiques et dynamiques (fréquence de fonctionnement, puissance dissipée).

Chapitre 5. Les circuits en I-MOS

Ce chapitre propose d'étudier les circuits élémentaires réalisés avec des transistors de type I-MOS.

La première partie est consacrée à l'inverseur. Elle présente la méthodologie d'optimisation, ainsi que les différences fondamentales avec l'inverseur CMOS conventionnel. Les deux technologies sont ensuite comparées aux travers des figures de mérite de l'inverseur: la technologie I-MOS présente de meilleures performances en termes de vitesse de commutation et de consommation statique grâce aux excellentes performances en terme de I_{on}/I_{off} , mais est pénalisé par sa tension d'alimentation pour les consommations dynamiques et de court-circuit.

La seconde partie montre que les portes NAND et NOR réalisées en technologie I-MOS sont fonctionnelles, alors que la troisième partie montre que la marge de bruit statique d'une SRAM est suffisante (251mV @ $V_{DD}=1V$), avec un gain en terme de vitesse par rapport à une intégration CMOS conventionnelle.

Ce chapitre s'appuie essentiellement sur des simulations TCAD calibrées sur les mesures à notre disposition.

Introduction

Le chapitre précédent a présenté le transistor de type I-MOS d'un point de vue théorique. Cependant, les transistors ne sont jamais utilisés seuls, mais intégrés dans des circuits plus complexes. Ce sont les performances de ces circuits, plus que celles d'un transistor isolé, qui sont représentatives d'une technologie.

Ce chapitre se propose donc d'évaluer les performances de circuits élémentaires:

- l'inverseur est le composant d'évaluation de la vitesse de commutation par excellence. Un nombre impair d'inverseurs étant accolés les uns aux autres, le circuit va naturellement osciller à une fréquence directement dépendante du nombre d'inverseurs et de la fréquence propre à un inverseur. Cette structure est appelée oscillateur en anneau.
- l'utilisation de deux inverseurs en série rebouclés sur eux-mêmes permet de verrouiller l'information. C'est le principe de la mémoire SRAM (*Static Random Access Memory*)
- les portes logiques NAND et NOR sont à la base de tout circuit logique

Avec les circuits apparaît la notion de robustesse, qui n'existe pas pour un transistor isolé. Si les dispositifs qui les composent ne sont pas correctement calibrés, alors les performances s'en ressentent, pouvant aller jusqu'à rendre le circuit non fonctionnel.

Le manque de maturité des réalisations technologiques amènent donc à privilégier une évaluation basée sur de la modélisation simple et des simulations TCAD.

5.1. Conception d'un inverseur CI-MOS

5.1.1. L'inverseur CMOS

Principe de fonctionnement

L'inverseur est le circuit logique le plus élémentaire. Il transforme un signal à l'état logique haut signal à l'état logique bas et inversement. Sa table de vérité est donnée dans le **tableau 5**.

V_{IN}	V_{OUT}
0	1
1	0

Tableau 5: table de vérité d'un inverseur

Il existe plusieurs réalisations possibles pour les inverseurs. La plus utilisée est l'intégration en technologie CMOS (*Complementary MOS*), qui utilise un transistor de type n et un transistor de type p (**figure 5-1**). En effet, l'inverseur CMOS présente l'avantage de consommer très peu à l'état statique. Lorsque la l'entrée est à l'état bas, le transistor n-mos est bloqué. A l'état haut, c'est le transistor p-mos qui est bloqué. Par conséquent, pour des états haut et bas définis, il n'y a jamais de chemin de conduction entre la masse et l'alimentation.

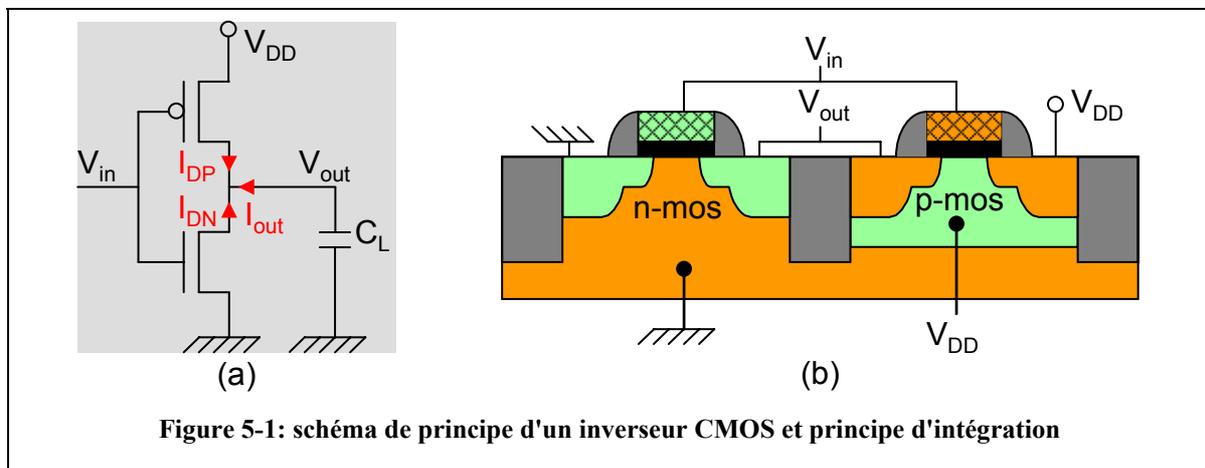


Figure 5-1: schéma de principe d'un inverseur CMOS et principe d'intégration

La caractéristique de transfert de l'inverseur, c'est-à-dire $V_{OUT}=f(V_{IN})$, peut se déterminer à partir des caractéristiques des transistors n-mos et p-mos qui le composent. La conservation du courant implique (**figure 5-1a**):

$$I_{DN} + I_{DP} + I_{OUT} = 0 \quad (5.1)$$

En régime statique, le courant de charge I_{out} est nul. L'équation (5.1) devient donc

$$I_{DN} + I_{DP} = 0 \quad (5.2)$$

Construction analytique de la caractéristique de transfert

Il est possible de résoudre analytiquement ce système à partir des équations introduites au chapitre 1. Cependant, il est nécessaire de déterminer au préalable dans quel état se trouve chacun des transistors selon l'état de polarisation, ce qui est résumé sur la **figure 5-2**.

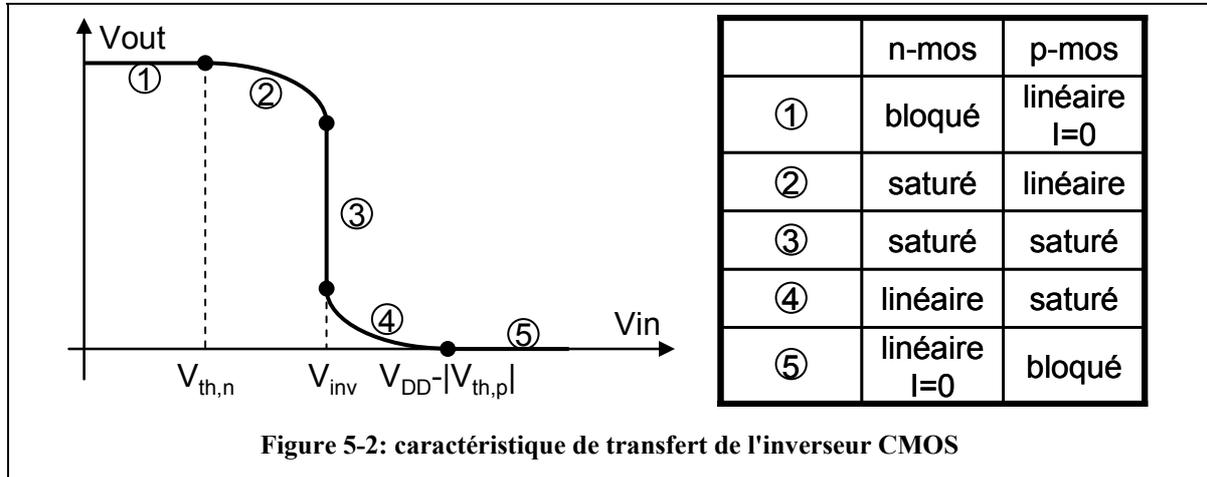


Figure 5-2: caractéristique de transfert de l'inverseur CMOS

En outre, il convient d'exprimer les tensions de drain et de grille du n-mos et du p-mos en fonction des tensions appliquées à l'inverseur:

$$\begin{aligned}
 V_{DN} &= V_{out} \\
 V_{DP} &= V_{out} - V_{DD} \\
 V_{GN} &= V_{in} \\
 V_{GP} &= V_{in} - V_{DD}
 \end{aligned}
 \tag{5.3}$$

L'ensemble des équations qui décrivent la caractéristique de sortie d'un inverseur CMOS est résumé dans le **tableau 6**.

V _{in}	V _{in} < V _{thn}	V _{thn} < V _{in} < V _{inv}	V _{inv}	V _{inv} < V _{in} < V _{thp}	V _{in} > V _{thp}
état nmos	bloqué	saturation	saturation	linéaire	linéaire
état pmos	linéaire	linéaire	saturation	saturation	bloqué
I _{dn}	0	$\frac{\beta_n}{2} (V_{in} - V_{thn})^2$	$\frac{\beta_n}{2} (V_{in} - V_{thn})^2$	$\beta_n \left(V_{in} - V_{thn} - \frac{V_{out}}{2} \right) V_{out}$	
I _{dp}	$\beta_p \left(V_{in} - V_{thp} - \frac{V_{out} - V_{dd}}{2} \right) (V_{out} - V_{dd})$		$\frac{\beta_p}{2} (V_{in} - V_{thp})^2$	$\frac{\beta_p}{2} (V_{in} - V_{thp})^2$	0
V _{out}	V _{DD}	$V_{in} - V_{thp} + V_{dd} + \sqrt{(V_{in} - V_{thp})^2 - \frac{\beta_n}{\beta_p} (V_{in} - V_{thn})^2}$			0
		$V_{in} - V_{thn} - \sqrt{(V_{in} - V_{thn})^2 - \frac{\beta_n}{\beta_p} (V_{in} - V_{thp})^2}$			

Tableau 6: expression analytique de la caractéristique statique d'un inverseur CMOS

L'expression de la tension d'inversion V_{inv} remarquable sur la **figure 5-2** est obtenue par résolution de l'équation:

$$\frac{\beta_n}{2}(V_{inv} - V_{thn})^2 = \frac{\beta_p}{2}(V_{inv} - V_{thp})^2 \quad (5.4)$$

soit:

$$V_{inv} = \frac{V_{thp} + V_{thn} \sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}} \quad (5.5)$$

Construction numérique de la caractéristique de transfert

La résolution analytique nécessite de connaître les relations courant-tension du transistor. Pour des dispositifs innovants, il est donc exclu d'utiliser cette approche, faute de modèle.

Dans ce cas, il est courant d'utiliser une approche numérique. La relation (5.2) se traduit graphiquement par l'intersection des courbes $I_D(V_{OUT})$ du p-mos et du n-mos (**figure 5-3b**). Or, ces courbes sont la transformation des courbes $I_D(V_D)$ accessibles par simulation.

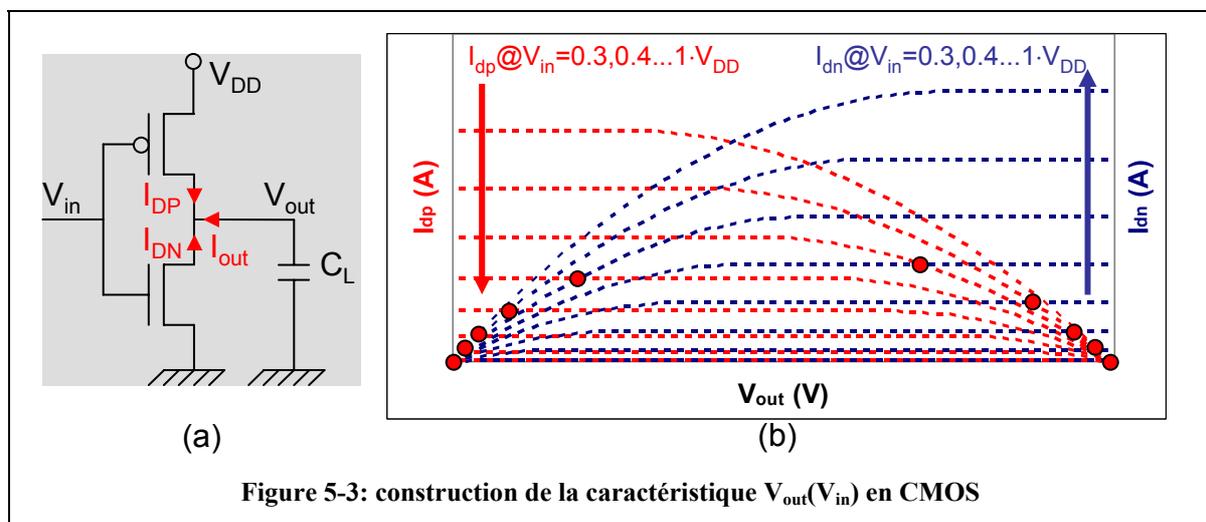


Figure 5-3: construction de la caractéristique $V_{out}(V_{in})$ en CMOS

Pour obtenir la caractéristique de transfert de l'inverseur, il suffit donc pour chaque tension d'entrée V_{IN} de simuler la caractéristique $I_D(V_{OUT})$ du p-mos et du n-mos, et de déterminer le point de fonctionnement (V_{OUT} , I_D) par construction. C'est cette méthodologie qui sera employée pour déterminer les caractéristiques des inverseurs I-MOS.

Remarques sur la caractéristique de transfert

La caractéristique de transfert de l'inverseur CMOS permet de relever deux points essentiels des circuits logiques:

- le signal doit être régénéré, sans quoi les circuits ne pourraient pas être cascadés
- cela passe par l'obligation d'un gain en tension $\frac{\partial V_{out}}{\partial V_{in}} > 1$.

Les zones hors de cette zone de gain sont appelées marge de bruit à l'état haut NM^H (*Noise Margin High*) et marge de bruit à l'état bas NM^L (*Noise Margin Low*). Elles sont intimement liées aux tension de seuil des transistors (**figure 5-4**).

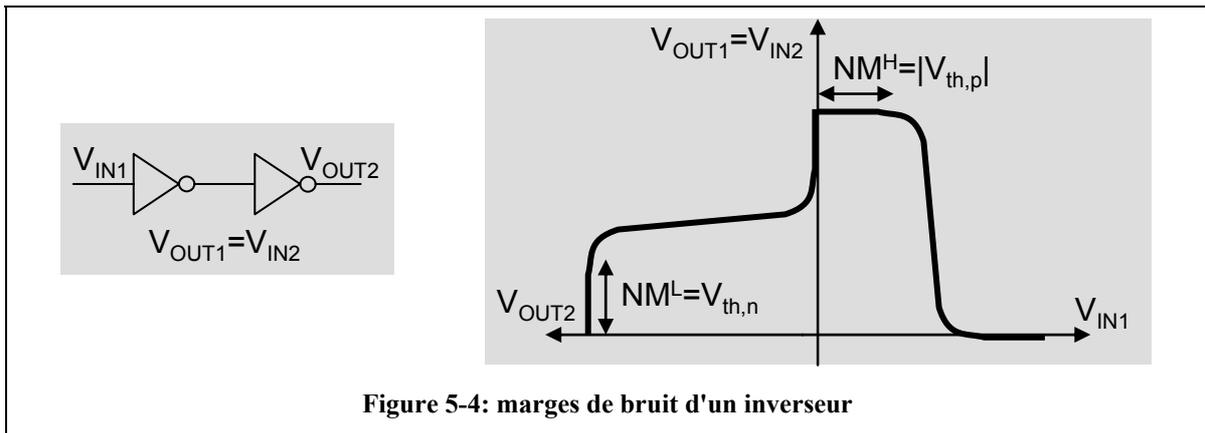


Figure 5-4: marges de bruit d'un inverseur

5.1.2. L'inverseur CI-MOS

Perte d'amplitude du signal

Soit un inverseur (indifféremment CMOS ou CI-MOS) à l'équilibre, dont l'entrée est à l'état bas (et par conséquent la sortie est à l'état haut). Lorsque l'entrée passe brusquement à l'état bas, le dispositif de type p ne conduit plus et l'inverseur peut être réduit au dispositif de type n. Dans ce cas, ce dispositif va décharger la capacité C_L associées aux éléments en aval (**figure 5-5a**), et la tension de sortie V_{OUT} va chuter au cours du temps.

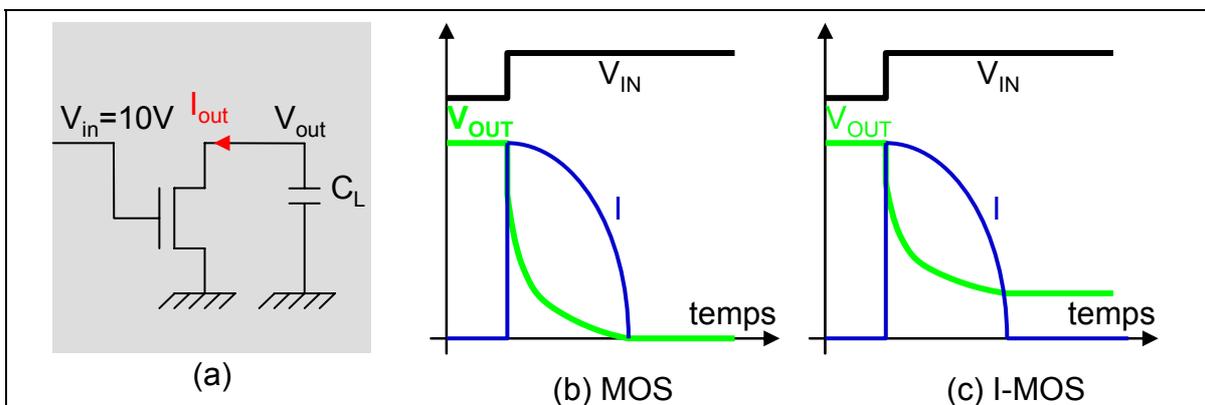
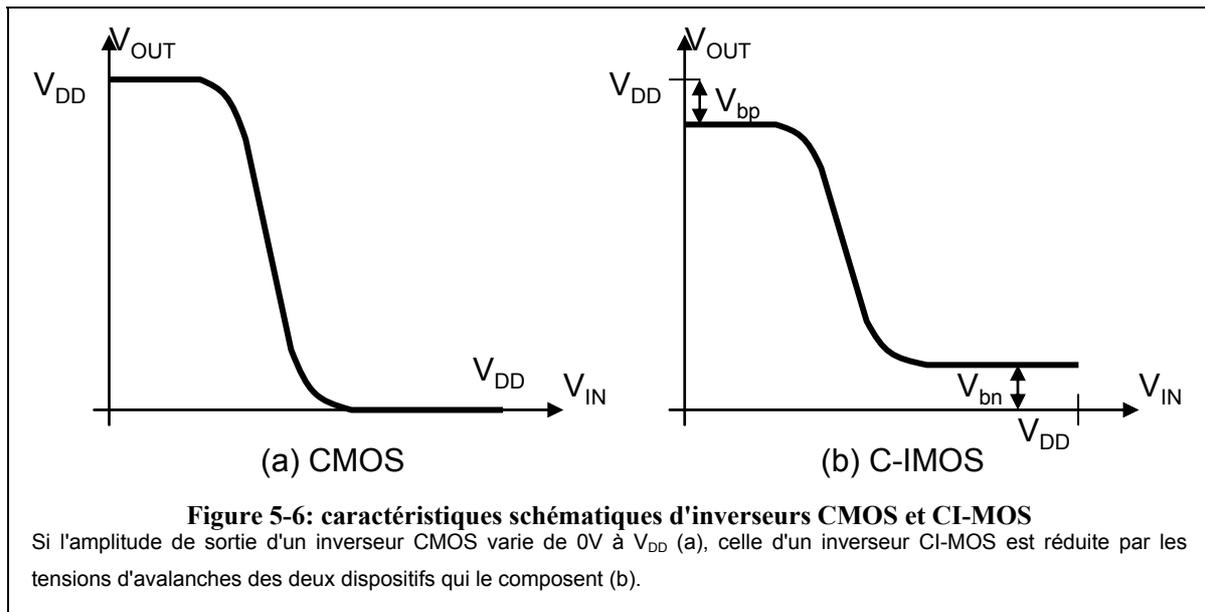


Figure 5-5: décharge d'une capacité à travers un MOS conventionnel et un I-MOS

Lors d'un basculement de l'état bas à l'état haut de la tension d'entrée d'un inverseur, la tension de sortie redescend à 0 en technologie CMOS conventionnel (a) alors qu'elle redescend à une valeur non nulle en technologie CI-MOS.

Des simulations TCAD transitoires permettent de relever une différence majeure entre un transistor conventionnel (**figure 5-5b**) et un I-MOS (**figure 5-5c**): pour ce dernier, la tension de sortie ne peut descendre en dessous de la tension d'avalanche V_{BN} . En effet, pour des valeurs inférieures, l'I-MOS ne conduit plus.

En considérant l'effet symétrique sur l'I-MOS de type p, il s'ensuit que l'amplitude de sortie d'un inverseur CI-MOS est plus petite que son amplitude d'entrée.



Par conséquent, outre la nécessité d'un gain supérieur à l'unité, une nouvelle condition est nécessaire pour un inverseur CI-MOS: l'amplitude du signal d'entrée doit être supérieure à la somme des tensions d'avalanche des dispositifs composant l'inverseur.

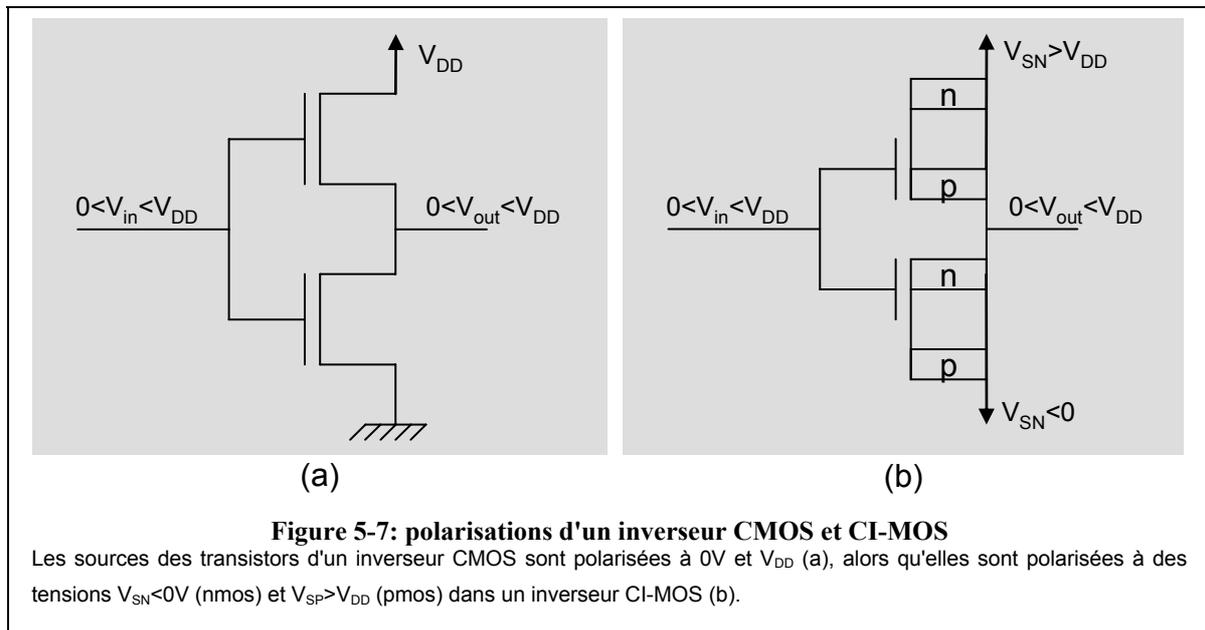
Cet diminution a été mentionnée dans [Gopalakrishnan'05a], et la simulation de dispositifs en germanium a montré qu'avec une tension d'avalanche d'environ 0.7V et une tension d'alimentation de 2V, le signal de sortie a une amplitude de 0.6V, suffisante pour commander un autre inverseur.

Cependant, le silicium présente des tensions d'avalanche plus élevées que celles du germanium, du fait de son gap plus élevé. L'étude des dispositifs I-MOS en silicium a montré que pour de tels dispositifs, la tension d'avalanche est supérieure à 5V [Charbuillet'06b]. Un inverseur complémentaire devrait donc être alimenté avec une tension supérieure à 10V (5V pour le n-mos, 5V pour le p-mos), ce qui conduirait à des consommations inacceptables.

Application d'un schéma de polarisation spécifique

Cependant, il a été souligné au chapitre précédent que la polarisation la plus adaptée pour l'I-MOS consiste à polariser la source négativement (n-mos). De fait, il s'agit de séparer la tension d'alimentation et les signaux de commande et de sortie.

Cette stratégie peut être appliquée à l'inverseur CI-MOS: la source du nI-MOS sera polarisée négativement ($V_{SN} < 0$), alors que la source du pI-MOS sera polarisée positivement ($V_{SP} > 0$) (**figure 5-7b**). De cette manière, les signaux d'entrée et de sortie varient entre 0 et V_{DD} , à l'instar du CMOS (**figure 5-7a**).



5.1.3. Conception d'un inverseur CI-MOS

La conception des inverseurs CMOS conventionnels est robuste: fort peu de choix mal faits entraîneront un inverseur non fonctionnel. En revanche, la conception d'un inverseur CI-MOS fonctionnel est plus délicate. Dans cette partie, tous les éléments associés à cette conception sont abordés, afin de proposer un inverseur fonctionnel et optimisé.

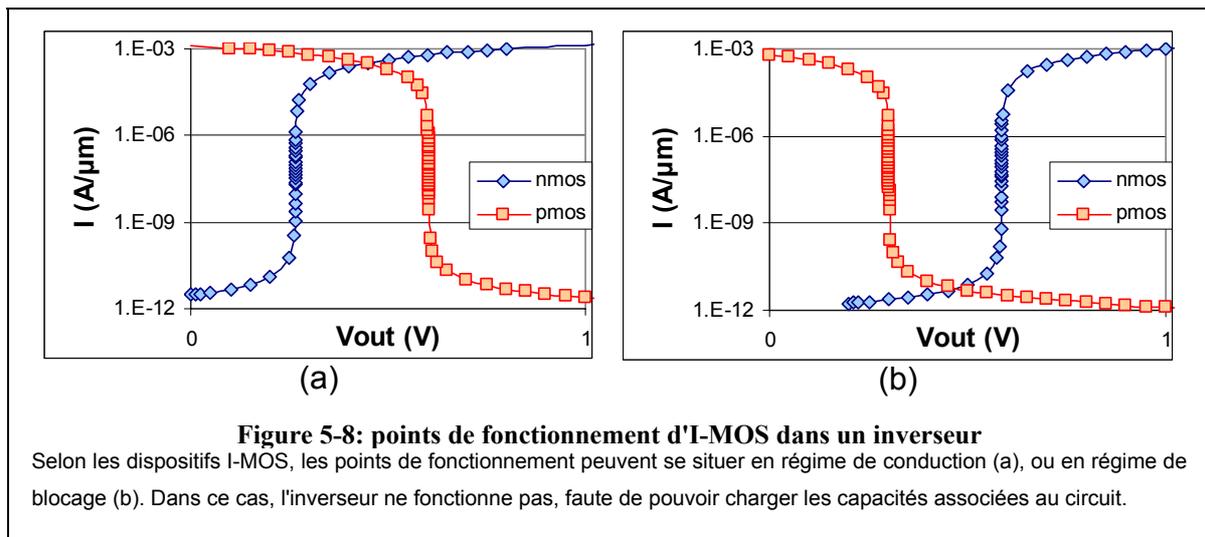
5.1.3.1. Tensions d'avalanche et points de fonctionnement

Dans un inverseur CMOS, un des deux transistors au minimum est passant (**figure 5-2**), quel que soit l'état de polarisation.

Dans un inverseur CI-MOS, cette configuration équivaut à ce que la tension d'avalanche du p-mos soit supérieure à celle du n-mos (**figure 5-8a**). Cependant, il existe une autre

configuration, inexistante en CMOS: lorsque la tension d'avalanche du p-mos est inférieure à celle du n-mos (**figure 5-8b**), les deux transistors sont bloqués.

Il est important de souligner que ce cas est à éviter: en effet, la valeur de la tension de sortie est déterminée par la condition $I_{D,n}=I_{D,p}$, qui seraient alors des courants de blocage, donc très faibles. L'inverseur ne pourrait donc pas commuter, ne pouvant charger ou décharger les capacités du circuit.



Pour éviter ce cas, et donc obtenir un inverseur fonctionnel, il faut par conséquent respecter les critères suivants:

- pour des niveaux logiques définis haut et bas ($V_{IN}=0$ et $V_{IN}=V_{DD}$), un transistor doit être passant et l'autre transistor bloqué.
- pour des tensions d'entrée intermédiaires, il doit exister une plage sur laquelle les deux dispositifs sont passants simultanément afin de permettre la charge/décharge de la capacité.

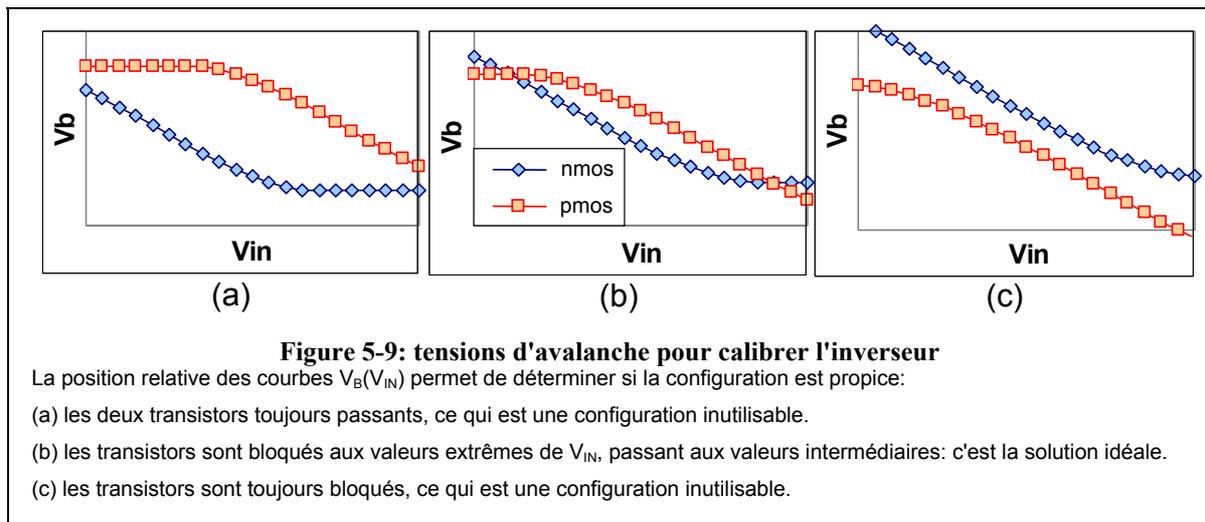
Ces conditions peuvent notamment se déterminer en examinant graphiquement les courbes de tension de claquage en fonction de la tension d'entrée $V_{B,N}(V_G)$ et $V_{B,P}(V_G)$.

5.1.3.2. Courbes $V_{B,N}(V_G)$ et $V_{B,P}(V_G)$

$V_{B,N}$ et $V_{B,P}$ représentent les tensions d'avalanche (*Breakdown Voltage*) d'un I-MOS de type n et d'un I-MOS de type p, respectivement. Dans les caractéristiques $I_D(V_D)$, ce sont les tensions de drain pour lesquelles le courant de drain augmente brusquement. Les caractéristiques $I_D(V_D)$ et donc les tensions d'avalanche dépendent de la tension de grille.

En traçant les courbes représentant les tensions d'avalanche V_B en fonction de la tension d'entrée V_{IN} , plusieurs configurations peuvent apparaître:

- la tension d'avalanche du n-mos est toujours inférieure à celle du p-mos (**figure 5-9a**): les deux transistors sont toujours passants. Les états logiques haut et bas ne sont pas clairement définis, ce qui rend cette configuration inutilisable.
- la tension d'avalanche du n-mos est toujours supérieure à celle du p-mos (**figure 5-9c**): les deux transistors sont toujours bloqués. A aucun moment il n'y a de courant capable de charger/décharger la capacité. Cette situation traduit un inverseur non fonctionnel.
- les courbes se croisent (**figure 5-9b**). On retrouve alors les conditions de fonctionnement évoquées au paragraphe précédent:



- les deux transistors sont bloqués aux états logiques haut et bas ($V_{IN}=0$ et $V_{IN}=V_{DD}$).

Il est intéressant de remarquer que cela entraîne l'absence de consommation à l'état statique, comme pour un inverseur CMOS.

- en zone centrale, les deux transistors sont passants, autorisant ainsi charge ou décharge.

En outre, il est intéressant de souligner que les marges de bruit en CI-MOS sont définies par le croisement entre $V_{B,n}$ et $V_{B,p}$, alors qu'elles sont définies par les tensions de seuil des transistors en CMOS conventionnel.

5.1.3.3. Paramètres d'ajustement

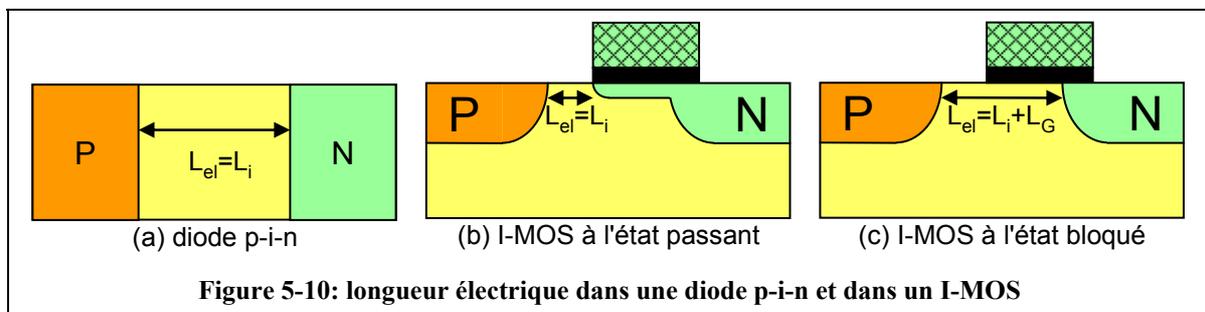
Le paragraphe précédent a permis de définir la configuration nécessaire à l'obtention d'un inverseur fonctionnel. Un certain nombre de leviers sont disponibles afin d'obtenir cette configuration. Certains sont d'ordre technologique (longueur intrinsèque, épaisseur d'oxyde), d'autres relèvent de la conception de circuit (polarisation).

Cette partie vise à établir la liste des leviers disponibles, ainsi que leur effet sur la caractéristique de l'inverseur CI-MOS, à des fins d'optimisation.

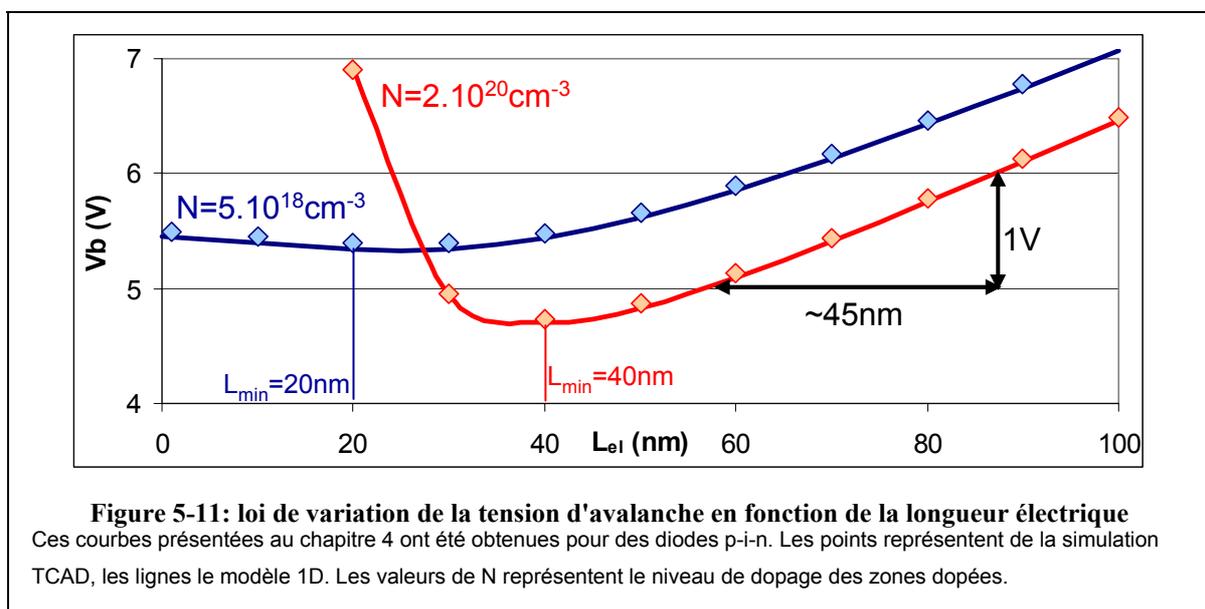
Dimensions latérales

Cependant, un transistor I-MOS est similaire à une diode p-i-n dont on modulerait la distance entre zone dopée de type n et zone dopée de type p. Dans une diode p-i-n, la longueur intrinsèque L_i représente la distance qui sépare la zone dopée de type n de la zone dopée de type p, alors que dans un I-MOS elle représente la dimension entre la source et la grille.

Pour homogénéiser les notations, cette dimension sera appelée longueur électrique L_{el} . Elle est égale à la longueur intrinsèque pour une diode p-i-n (**figure 5-10a**). Pour un I-MOS, elle varie en fonction de la tension de grille de la longueur intrinsèque (**figure 5-10b**) à la somme longueur intrinsèque + longueur de grille (**figure 5-10c**).



La loi de variation de la tension d'avalanche d'une diode p-i-n en fonction de la longueur intrinsèque a été déterminée au chapitre précédent (**figure 5-11**). Cette loi peut être utilisée pour déterminer les plages d'utilisation d'un I-MOS. La longueur de grille augmentant la longueur électrique de l'I-MOS, cette dernière est donc à rapprocher de la longueur intrinsèque.



L'une des hypothèses fondamentales du principe de l'I-MOS est que sa tension d'avalanche augmente avec la longueur électrique. Il s'ensuit qu'elle doit se situer dans une zone de courbe pour laquelle:

$$\frac{\partial V_b}{\partial L_{\text{el}}} > 0 \quad (5.6)$$

Si l'I-MOS se comportait comme une diode p-i-n fortement dopé, cela impliquerait une longueur intrinsèque au moins égale à 40nm. Cependant, pour des tensions de grille de l'ordre du volt, la densité de charges sous la grille est plus faible, environ quelques 10^{18}cm^{-3} . Or, l'étude de la diode p-i-n au chapitre 4 a montré que c'est la zone moins dopée qui fixe la caractéristique. Par conséquent, la longueur intrinsèque minimale est de l'ordre de 20nm.

La technologie CMOS actuelle dispose un tension d'alimentation de l'ordre du volt. Pour assurer la compatibilité, la différence de tension d'avalanche doit donc être légèrement supérieure au volt. Une longueur de grille de l'ordre de 50nm semble donc bien adaptée.

Un inverseur ayant pour longueur intrinsèque 25nm et pour longueur de grille 50nm devrait présenter une amplitude de sortie d'environ 1V. Pour optimiser cet inverseur, il faut garder à l'esprit que:

- augmenter la longueur de grille permet d'augmenter l'amplitude de sortie
- diminuer la longueur intrinsèque risque de diminuer l'amplitude de sortie, alors que l'augmenter ne devrait pas avoir d'impact.

Oxyde de grille

La tension d'avalanche est d'autant plus faible que la densité de porteurs dans le canal est élevée. Par conséquent, réduire l'épaisseur de l'oxyde de grille permet d'abaisser la tension d'avalanche, et est donc bénéfique pour les performances de l'inverseur.

Polarisation

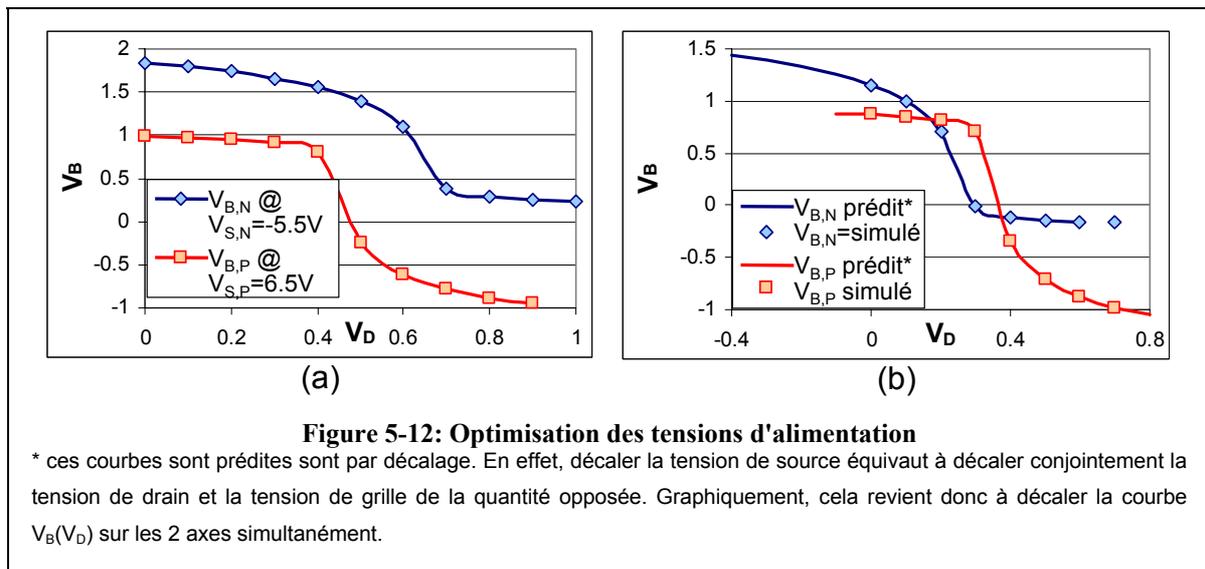
Les précédents leviers ne donnaient qu'une information qualitative. Par contre, l'impact de la polarisation peut être quantifié, et donc être utilisé de façon prédictive. En effet, décaler l'ensemble des polarisations d'une même valeur ne change pas la caractéristique:

$$I(V_s, V_G, V_D) \approx I(V_s + \Delta V, V_G + \Delta V, V_D + \Delta V) \quad (5.7)$$

Par conséquent, faire varier la tension d'alimentation du n-mos V_{sn} de ΔV aura le même effet que faire varier tension de drain et tension de grille de $-\Delta V$. La tension d'avalanche étant une valeur particulière de tension de drain, faire varier la tension d'alimentation revient donc à

déplacer la courbe $V_B(V_G)$ sur les deux axes simultanément. Il est donc possible de chercher graphiquement le meilleur point de polarisation pour l'inverseur.

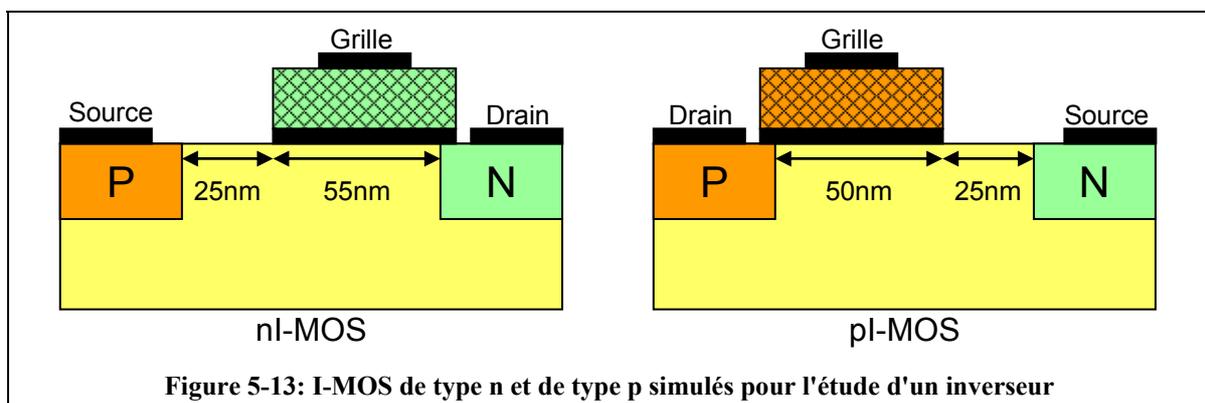
Cette méthodologie est illustrée sur la **figure 5-12**. Dans un premier temps, les transistors n-mos et p-mos sont simulés avec des tensions d'alimentation $V_{S,N}$ et $V_{S,P}$ estimées à partir de la littérature. La position relative des courbes $V_B(V_G)$ montrent que cet inverseur ne sera pas fonctionnel (**figure 5-12a**). Puis, un point de polarisation est optimisé par simple construction graphique. Les simulations effectivement réalisées avec ces polarisations permettent de valider cette approche (**figure 5-12b**).



5.1.3.4. Application

Dimensions et polarisation

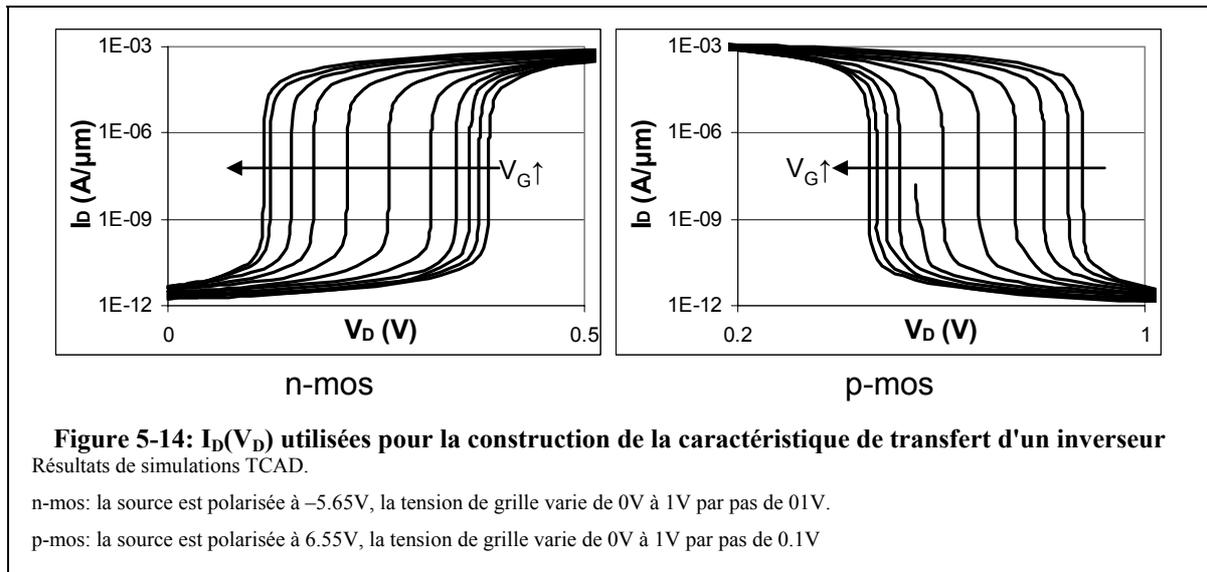
Ces considérations ont été appliquées pour calibrer des transistors I-MOS de type p et de type n. L'oxyde de grille mesure 1nm d'épaisseur, les transistors ont une longueur de grille de respectivement 50nm et 55nm, et une longueur intrinsèque de 25nm (**figure 5-13**). La source du n-mos est polarisée à $V_{s,n}=-5.65V$ et la source du p-mos à $V_{s,p}=6.55V$. Cette différence



d'1V est à rapprocher des tensions d'alimentation en CMOS conventionnel, où la source du n-mos est à la masse alors que la source du p-mos est à V_{DD} .

Construction de la caractéristique de transfert

Les caractéristiques $I_D(V_D)$ sont simulées pour le n-mos et pour le p-mos, pour des valeurs de tension de grille allant 0V à 1V (**figure 5-14**).

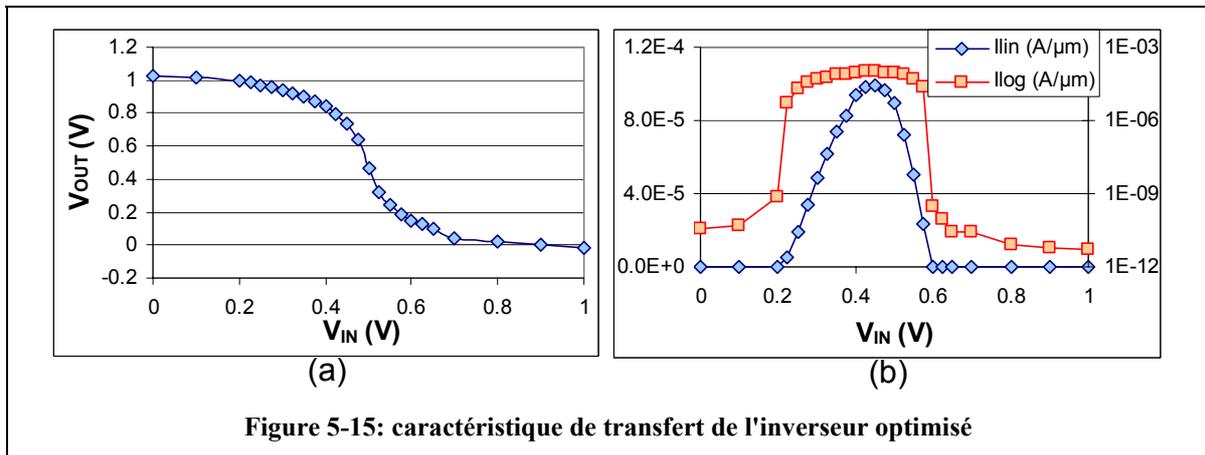


Pour chaque valeur de tension d'entrée $V_{IN}=V_G$, la tension de sortie V_{OUT} est la tension de drain V_D qui correspond au croisement des courbes. L'ensemble des points de fonctionnement ainsi obtenus compose la caractéristique de transfert.

Etude de la caractéristique de transfert

La caractéristique de sortie de la **figure 5-15a** permet de valider les des hypothèses faites:

- la caractéristique est indéniablement celle d'un inverseur.
- l'amplitude de sortie est égale à l'amplitude d'entrée répondent à l'objectif de 1V
- le gain supérieur à l'unité dans la zone de basculement permet de cascader des inverseurs.



- comme pour le CMOS, le courant de fuite est très faible à l'état statique (**figure 5-15b**). Le courant plus élevé dans la zone de basculement permet la charge/décharge de la capacité.

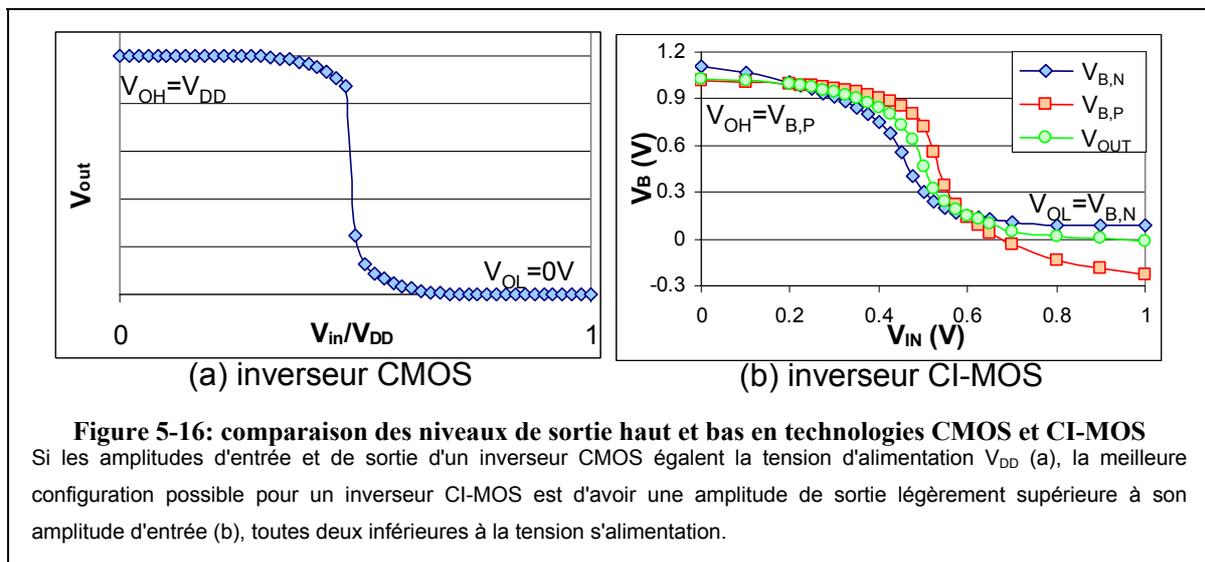
5.1.4. Spécificités de l'inverseur CI-MOS

L'étude réalisée jusqu'ici sur l'inverseur CI-MOS a permis d'en concevoir un fonctionnel. Cependant, celui-ci présente des différences fondamentales avec un inverseur CMOS,

Niveaux logiques haut et bas

Dans un inverseur CMOS, la tension d'entrée vient créer ou supprimer un chemin de conduction. La sortie est donc en court-circuit avec la masse ou l'alimentation. Par conséquent, le niveau de sortie haut V_{OH} (*Voltage Output High*) est égal à V_{DD} et le niveau de sortie bas V_{OL} (*Voltage Output Low*) est égal à $0V$ (**figure 5-16a**).

Dans un inverseur CI-MOS, en revanche, il n'existe pas de chemin de conduction. Le système fournit du courant lorsque la polarisation est favorable. Cela implique notamment que les niveaux de sortie haut et bas ne sont directement dépendants de l'alimentation, mais des tensions d'avalanche des dispositifs (**figure 5-16b**).



C'est cette particularité qui rend l'inverseur CI-MOS plus délicat à concevoir qu'un inverseur CMOS. Pour réaliser un inverseur CMOS fonctionnel, il suffit d'avoir une tension d'alimentation suffisante (supérieure à V_{th}). Quels que soient les transistors utilisés, l'amplitude du signal de sortie $V_{OH}-V_{OL}$ sera égale à l'amplitude d'entrée V_{DD} . Pour un inverseur CI-MOS, il convient d'abord de gérer correctement les tensions d'avalanche, et par conséquent les tensions d'alimentation $V_{S,N}$ et $V_{S,P}$.

5.2. Performances d'un inverseur CI-MOS

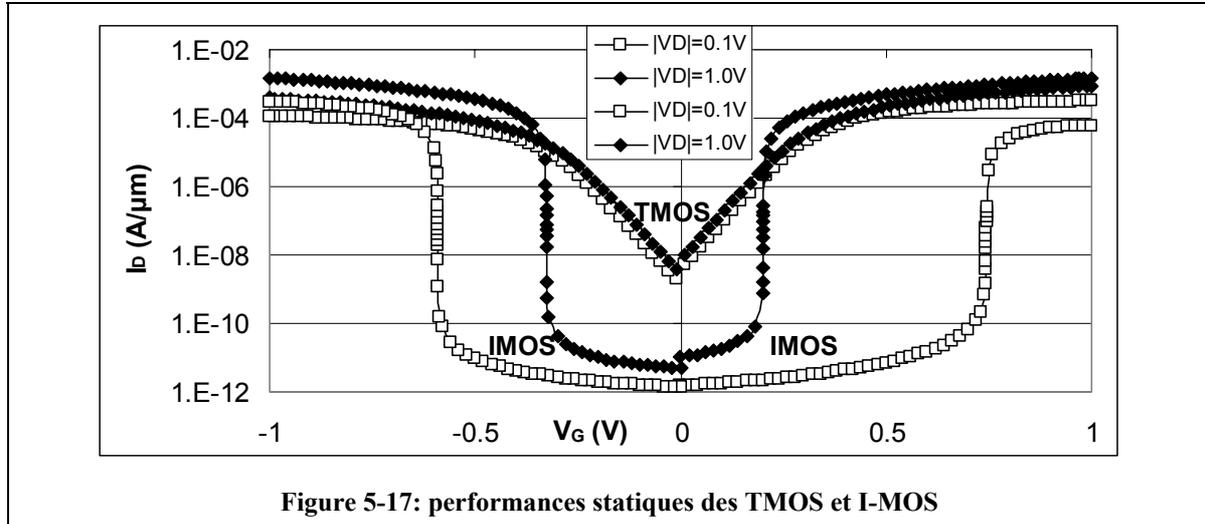
De même que les transistors isolés peuvent être évalués à travers des figures de mérite (I_{on}/I_{off} , $CV/I...$), les inverseurs peuvent être estimés à travers des indicateurs de performance: fréquence maximale de fonctionnement, consommation. Cette partie vise à établir les performances d'un inverseur réalisé en technologie I-MOS complémentaire, et de les comparer à celles d'un inverseur réalisé en technologie CMOS conventionnelle.

Deux approches seront abordées simultanément:

- les expressions analytiques permettant de souligner les avantages ou inconvénients intrinsèques à l'I-MOS seront utilisées tant que possible
- en illustration et à défaut, les performances d'un inverseur CI-MOS et d'un inverseur CMOS de dimensions identiques seront comparées, en se basant sur des simulations TCAD

5.2.1. Structures simulées – stratégie de comparaison

L'inverseur CI-MOS étudié est identique à celui présenté au paragraphe 5.1.3.4. Les transistors MOS simulés pour la comparaison sont de même dimensions que leur homologues I-MOS: longueurs de grille de 55nm (n-mos) et 50nm (p-mos), oxyde de grille 1nm.



Les caractéristiques statiques des quatre dispositifs simulés isolément sont montrées sur la **figure 5-17** et leurs performances sont résumées dans le **tableau 7**. Celles des I-MOS sont excellentes en ce qui concerne la figure de mérite I_{on}/I_{off} .

	TMOS		I-MOS	
	nmos	pmos	nmos	pmos
V_S	0V	0V	-5.65V	5.55V
V_{DD}	1V	1V	1V	1V
$V_{th\ lin}$	0.24V	-0.34V	0.75V	-0.6V
$V_{th\ sat}$	0.21V	-0.30V	0.2V	-0.32V
S	75mV/dec	75mV/dec	0.17mV/dec	0.11mV/dec
$I_{on} (\mu A/\mu m)$	856	388	1483	1606
$I_{off} (A/\mu m)$	$5 \cdot 10^{-9}$	$2 \cdot 10^{-9}$	10^{-11}	$5 \cdot 10^{-12}$

Tableau 7: performances statiques des TMOS et I-MOS

Les différences de ces caractéristiques sont telles qu'elles ne peuvent être directement comparées. C'est pourquoi les comparaisons seront faites à **courant de conduction égaux**. La largeur des dispositifs I-MOS est arbitrairement fixée à $1\mu m$, alors que la largeur des dispositifs MOS conventionnels est augmentée pour assurer l'égalité des courants de conduction ($W_n=1.7\mu m$ pour le n-mos et $W_p=4.1\mu m$ pour le p-mos).

5.2.2. Consommation statique

Lorsque l'inverseur est en régime statique, il y a toujours un des deux transistors qui est bloqué. Cependant, l'inverseur est quand même traversé par le courant de fuite I_{off} , à l'origine de la consommation statique:

$$P_{stat} = W \cdot I_{off} \cdot V_{alim} \quad (5.8)$$

Si l'I-MOS est désavantagé du fait d'une tension d'alimentation supérieure, son très faible courant de fuite lui permet de moins consommer en statique que son équivalent CMOS. Le **tableau 8** montre les consommations statiques des deux architectures. Dans les deux cas, c'est le n-mos, à l'origine de la consommation la plus importante, qui est considéré.

	TMOS	I-MOS
V_{alim}	1V	12.2V
P_{stat}	$8.7 \cdot 10^{-9} W$	$1.22 \cdot 10^{-10} W$

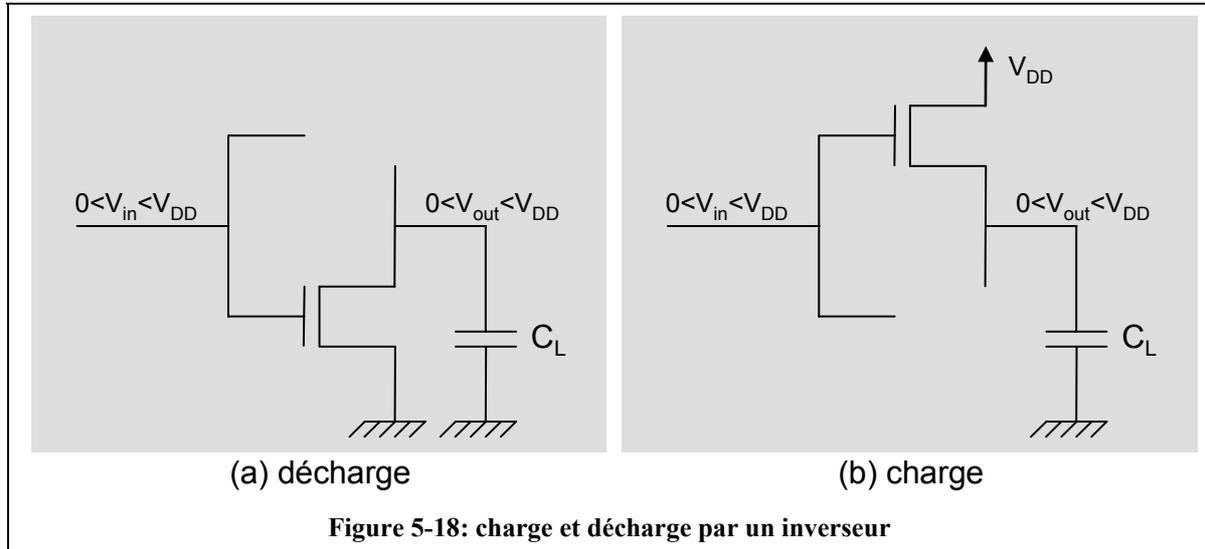
Tableau 8: consommations statiques

5.2.3. Consommation dynamique

Lorsque l'inverseur réalisée en technologie complémentaire commute, l'alimentation fournit de l'énergie pour charger la capacité de ligne. En outre, lorsque le transistor fournit du courant pour charger ou décharger la capacité de charge C_L (*Load Capacitance*) modélisant les éléments aval du circuit, il consomme de l'énergie. Ces deux phénomènes sont regroupés dans la consommation dynamique.

Les prochains paragraphes vont établir les expressions analytiques de la consommation statique, afin de pouvoir comparer les I-MOS avec les MOS conventionnels. Pour cela, un cycle complet est divisé en deux parties:

- la charge par le p-mos (**figure 5-18a**)
- le décharge par le n-mos (**figure 5-18b**)



5.2.3.1. Energie dissipée lors de la charge

L'entrée passe brusquement de V_{DD} à 0. La sortie passe de V_{OL} pour $t=0$ à V_{OH} pour $t=\infty$.

Energie dissipée par le p-mos

L'énergie dissipée par le p-mos est:

$$E_{\text{charge-pmos}} = \int_0^{\infty} i_p(t) \cdot [V_{SP} - V_{\text{out}}(t)] \cdot dt \quad (5.9)$$

Le courant peut être exprimé en fonction de la capacité de charge en sortie:

$$i_p(t) = C_L \cdot \frac{dV_{\text{out}}(t)}{dt} \quad (5.10)$$

Les équations (5.9) et (5.10) permettent d'exprimer la puissance dissipée par le p-mos:

$$E_{\text{charge-pmos}} = C_L \cdot V_{SP} \cdot (V_{OH} - V_{OL}) - \frac{1}{2} \cdot C_L \cdot (V_{OH}^2 - V_{OL}^2) \quad (5.11)$$

Energie stockée dans la capacité

L'énergie stockée dans la capacité est:

$$E_{\text{charge-}C_L} = \int_0^{\infty} i_p(t) \cdot [V_{\text{out}}(t) - V_{GND}] \cdot dt \quad (5.12)$$

Les relations (5.12), (5.10) et les conditions limites permettent de calculer:

$$E_{\text{charge-}C_L} = \frac{1}{2} \cdot C_L \cdot (V_{OH}^2 - V_{OL}^2) - C_L \cdot V_{GND} \cdot (V_{OH} - V_{OL}) \quad (5.13)$$

Energie fournie par l'alimentation

L'énergie fournie par l'alimentation est:

$$E_{\text{charge-alim}} = \int_0^{\infty} i_p(t) \cdot [V_{SP} - V_{GND}] \cdot dt \quad (5.14)$$

Les relations (5.14) et (5.10) et les conditions limites permettent de calculer:

$$E_{\text{charge-alim}} = C_L \cdot (V_{SP} - V_{GND}) \cdot (V_{OH} - V_{OL}) \quad (5.15)$$

5.2.3.2. Energie dissipée lors de la décharge

L'entrée passe brusquement de 0 à V_{DD} . La sortie passe de V_{OH} pour $t=0$ à V_{OL} pour $t=\infty$.

Energie dissipée par le n-mos

L'énergie dissipée par le n-mos est:

$$E_{\text{décharge-nmos}} = \int_0^{\infty} i_n(t) \cdot [V_{SN} - V_{out}(t)] \cdot dt \quad (5.16)$$

Le courant peut être exprimé en fonction de la capacité de ligne:

$$i_n(t) = -C_L \cdot \frac{dV_{out}(t)}{dt} \quad (5.17)$$

Les équations (5.16) et (5.17) permettent de calculer la puissance dissipée par le n-mos:

$$E_{\text{décharge-nmos}} = \frac{1}{2} \cdot C_L \cdot (V_{OH}^2 - V_{OL}^2) + C_L \cdot V_{SN} \cdot (V_{OH} - V_{OL}) \quad (5.18)$$

Energie stockée dans la capacité

L'énergie stockée dans la capacité est:

$$E_{\text{décharge-}C_L} = \int_0^{\infty} i_n(t) \cdot [V_{out}(t) - V_{GND}] \cdot dt \quad (5.19)$$

Les relations (5.19) et (5.17) et les conditions limites permettent de calculer:

$$E_{\text{décharge-}C_L} = -\frac{1}{2} \cdot C_L \cdot (V_{OH}^2 - V_{OL}^2) + C_L \cdot V_{GND} \cdot (V_{OH} - V_{OL}) \quad (5.20)$$

Energie fournie par l'alimentation

L'énergie fournie par l'alimentation est:

$$E_{\text{décharge-alim}} = \int_0^{\infty} i_n(t) \cdot [V_{\text{SN}} - V_{\text{GND}}] \cdot dt \quad (5.21)$$

Les relations (5.21) et (5.17) et les conditions limites permettent de calculer:

$$E_{\text{décharge-alim}} = C_L \cdot (-V_{\text{SN}} + V_{\text{GND}}) \cdot (V_{\text{OH}} - V_{\text{OL}}) \quad (5.22)$$

5.2.3.3. Synthèse

Les résultats sont synthétisés dans le **tableau 9**:

énergie dissipée par:	montée	descente
le transistor	$C_L \cdot (V_{\text{OH}} - V_{\text{OL}}) \cdot \left[V_{\text{SP}} - \frac{1}{2} \cdot (V_{\text{OH}} + V_{\text{OL}}) \right]$	$C_L \cdot (V_{\text{OH}} - V_{\text{OL}}) \cdot \left[V_{\text{SN}} + \frac{1}{2} \cdot (V_{\text{OH}} + V_{\text{OL}}) \right]$
la capacité	$C_L \cdot (V_{\text{OH}} - V_{\text{OL}}) \cdot \left[\frac{1}{2} \cdot (V_{\text{OH}} + V_{\text{OL}}) - V_{\text{GND}} \right]$	$-C_L \cdot (V_{\text{OH}} - V_{\text{OL}}) \cdot \left[\frac{1}{2} \cdot (V_{\text{OH}} + V_{\text{OL}}) - V_{\text{GND}} \right]$
l'alimentation	$C_L \cdot (V_{\text{SP}} - V_{\text{GND}}) \cdot (V_{\text{OH}} - V_{\text{OL}})$	$C_L \cdot (V_{\text{SN}} + V_{\text{GND}}) \cdot (V_{\text{OH}} - V_{\text{OL}})$
total par cycle	$C_L \cdot (V_{\text{SP}} + V_{\text{SN}}) \cdot (V_{\text{OH}} - V_{\text{OL}})$	

Tableau 9: consommations dynamiques

Il est intéressant de vérifier qu'à la charge comme à la décharge, l'énergie dissipée par l'alimentation est bien égale à la somme de l'énergie dissipée par le transistor et de l'énergie stockée dans la capacité de ligne.

Application au CMOS:

$$V_{\text{SP}}=V_{\text{DD}}, V_{\text{SN}}=0, V_{\text{OH}}=V_{\text{DD}}, V_{\text{OL}}=0$$

énergie dissipée par:	montée	descente
le transistor	$\frac{1}{2} \cdot C_L \cdot V_{\text{DD}}^2$	$\frac{1}{2} \cdot C_L \cdot V_{\text{DD}}^2$
la capacité	$\frac{1}{2} \cdot C_L \cdot V_{\text{DD}}^2$	$-\frac{1}{2} \cdot C_L \cdot V_{\text{DD}}^2$
l'alimentation	$C_L \cdot V_{\text{DD}}^2$	0
total par cycle	$C_L \cdot V_{\text{DD}}^2$	

Application au CI-MOS:

$$V_{\text{GND}}=0, V_{\text{SP}}, V_{\text{SN}}, V_{\text{OH}}=V_{\text{DD}}, V_{\text{OL}}=0$$

énergie dissipée par:	montée	descente
le transistor	$C_L \cdot (V_{OH} - V_{OL}) \cdot \left[V_{SP} - \frac{1}{2} \cdot (V_{OH} + V_{OL}) \right]$	$C_L \cdot (V_{OH} - V_{OL}) \cdot \left[V_{SN} + \frac{1}{2} \cdot (V_{OH} + V_{OL}) \right]$
la capacité	$\frac{1}{2} \cdot C_L \cdot (V_{OH}^2 - V_{OL}^2)$	$-\frac{1}{2} \cdot C_L \cdot (V_{OH}^2 - V_{OL}^2)$
l'alimentation	$C_L \cdot V_{SP} \cdot (V_{OH} - V_{OL})$	$C_L \cdot V_{SN} \cdot (V_{OH} - V_{OL})$
total par cycle	$C_L \cdot (V_{SP} + V_{SN}) \cdot (V_{OH} - V_{OL})$	

Comparaison

	CMOS	CI-MOS
V_{DD}	1V	1V
V_{OH}	0V	0V
V_{OL}	$=V_{DD}=1V$	$=V_{DD}=1V$
V_{SP}	0V	6.6V
V_{SN}	$V_{DD}=1V$	-5.6V
$P_{CC} = f \cdot C_L \cdot (V_{SP} + V_{SN}) \cdot (V_{OH} - V_{OL})$	$f \cdot 2.8 \cdot 10^{-14}$	$f \cdot 1.1 \cdot 10^{-13}$

La consommation dynamique est proportionnelle à la tension d'alimentation et à la capacité de charge. Un I-MOS nécessite une alimentation bien plus élevée (12.2V contre 1V), ce qui induit une consommation plus élevée, mais cet inconvénient est modéré par le fait qu'il présente une capacité plus faible qu'un transistor MOS conventionnel (à courants de conduction égaux). Cependant, la consommation dynamique d'un inverseur CI-MOS reste bien plus élevée que celle d'un inverseur CMOS conventionnel, environ 3.5 fois supérieure.

5.2.4. Puissance de court-circuit

Principe de calcul

Lorsque l'inverseur commute, il y a un bref instant où les deux transistors conduisent simultanément, court-circuitant alors la masse et l'alimentation. Ceci est à l'origine d'une puissance consommée supplémentaire, dépendante de la forme du signal d'entrée et de la fréquence de fonctionnement de l'inverseur.

Le calcul de la puissance de court-circuit peut être effectué analytiquement à partir d'hypothèses simplificatrices [Veendrick'84]:

- la capacité de charge est supposée nulle
- le signal d'entrée est supposé varier linéairement avec le temps (**figure 5-19**).

Dans ces conditions, la tension de sortie V_{out} dépend de la tension d'entrée V_{in} de la même manière que dans l'étude statique. En utilisant la relation $V_{in}(t)$ (**figure 5-19**), le courant de court-circuit I_{CC} peut alors être exprimé en fonction du temps t et être moyenné sur la période de fonctionnement:

$$I_{moyen} = \frac{1}{T} \cdot \int_0^T I_{cc}(t) \cdot dt \quad (5.23)$$

Le circuit dissipe une puissance de $I_{moyen} \cdot V_{alim}$ par cycle. La puissance de court-circuit est donc:

$$P_{CC} = I_{moyen} \cdot V_{alim} \quad (5.24)$$

Calcul analytique de la puissance de court-circuit en CMOS

On peut remarquer sur la **figure 5-19** que le courant est non nul uniquement entre t_1 et t_3 , ce qui correspond à la mise en conduction du n-mos et du p-mos respectivement. Pour simplifier, il est possible de considérer que n-mos et p-mos sont symétriques: même tension de seuil V_{th} , même courant de saturation I_{dsat} , même facteur de gain β . Il s'ensuit que les temps de montée et de descente sont égaux: $\tau_m = \tau_d = \tau$.

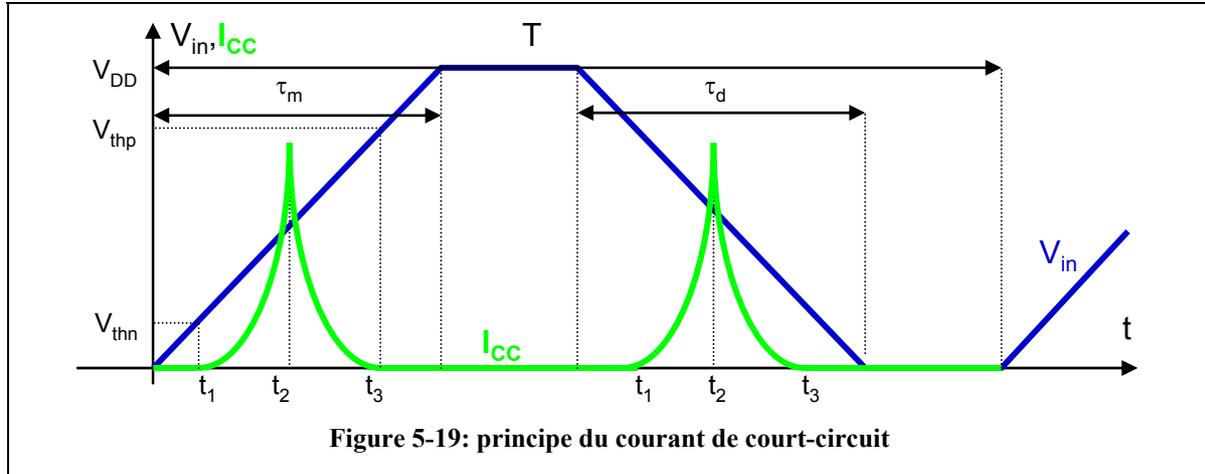


Figure 5-19: principe du courant de court-circuit

Avec ces hypothèses, le courant moyen devient:

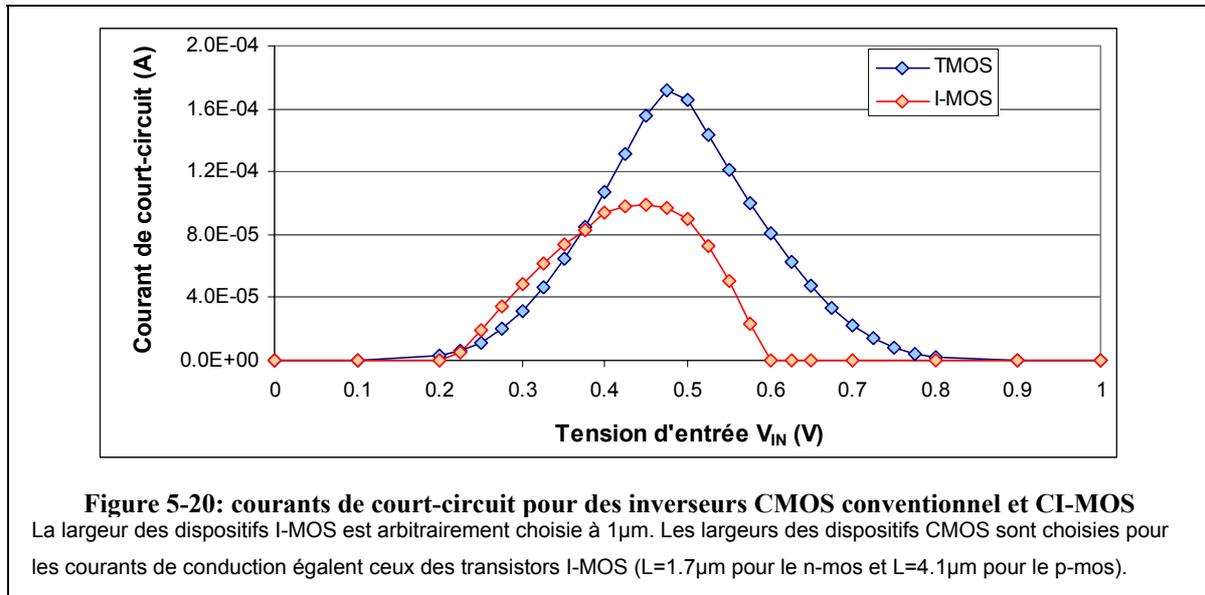
$$I_{moyen} = \frac{1}{12} \cdot \frac{\beta}{V_{DD}} \cdot (V_{DD} - 2V_t)^3 \cdot \frac{\tau}{T} \quad (5.25)$$

La puissance de court circuit $I_{moyen} \cdot V_{DD}$ s'écrit alors:

$$P_{cc} = \frac{\beta}{12} \cdot (V_{DD} - 2 \cdot V_t)^3 \frac{\tau}{T} \quad (5.26)$$

Calcul numérique de la puissance de court-circuit

Ce travail analytique ne peut pas être mené pour les I-MOS, faute de modélisation du courant. Il est donc impossible d'établir une relation équivalente à (5.26). Par contre, il est possible de calculer numériquement le courant moyen à l'aide de l'équation (5.23) à partir des caractéristiques obtenues par simulation. Par souci d'exhaustivité, cette approche a également été appliquée à l'inverseur CMOS conventionnel en complément de l'étude analytique.



Les courants de court-circuit en technologie CMOS conventionnelle et en technologie I-MOS complémentaire sont représentés sur la **figure 5-20**, et les résultats de ces calculs de puissance de court-circuit sont résumés dans le **tableau 10**. Le courant de court-circuit de l'I-MOS est plus faible que celui du CMOS. Ceci est dû à l'existence d'une tension d'avalanche sur la caractéristique $I_D(V_D)$, qui n'existe pas en CMOS conventionnel.

Néanmoins, l'I-MOS est fortement pénalisé par sa tension d'alimentation élevée. Il en résulte une consommation de court-circuit quatre fois plus élevée que celle de l'inverseur CMOS conventionnel.

	CMOS (analytique)	CMOS (numérique)	I-MOS (numérique)
$V_{\text{sn}}-V_{\text{SP}}$	1V	1V	12.2V
I_{moyen}	$5.6 \cdot 10^{-5} \text{ A}$	$8.2 \cdot 10^{-5} \text{ A}$	$2.4 \cdot 10^{-5} \text{ A}$
P_{CC}	$5.6 \cdot 10^{-5} \text{ W}$	$8.2 \cdot 10^{-5} \text{ W}$	$2.9 \cdot 10^{-4} \text{ W}$

Tableau 10: puissances de court-circuit

5.2.5. Fréquence maximum de fonctionnement

Après un brusque changement d'état du signal d'entrée, les capacités du circuit doivent être chargées ou déchargées. En l'absence de nouvelle commutation à l'entrée, la sortie atteindra son état stable après un délai dû à ces capacités. Ce délai sera dit temps de montée si la sortie

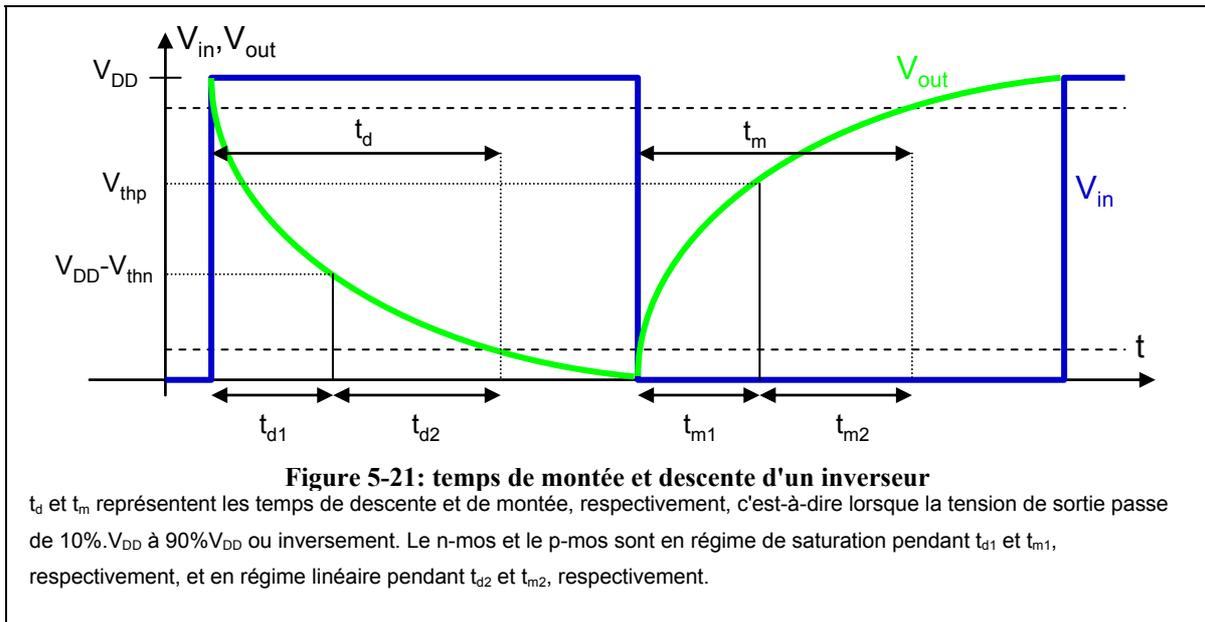
passé de l'état bas à l'état haut, temps de descente dans le cas contraire. On considère que le niveau haut est atteint lorsque:

$$V_{out} = 0.9 \cdot V_{DD} \tag{5.27}$$

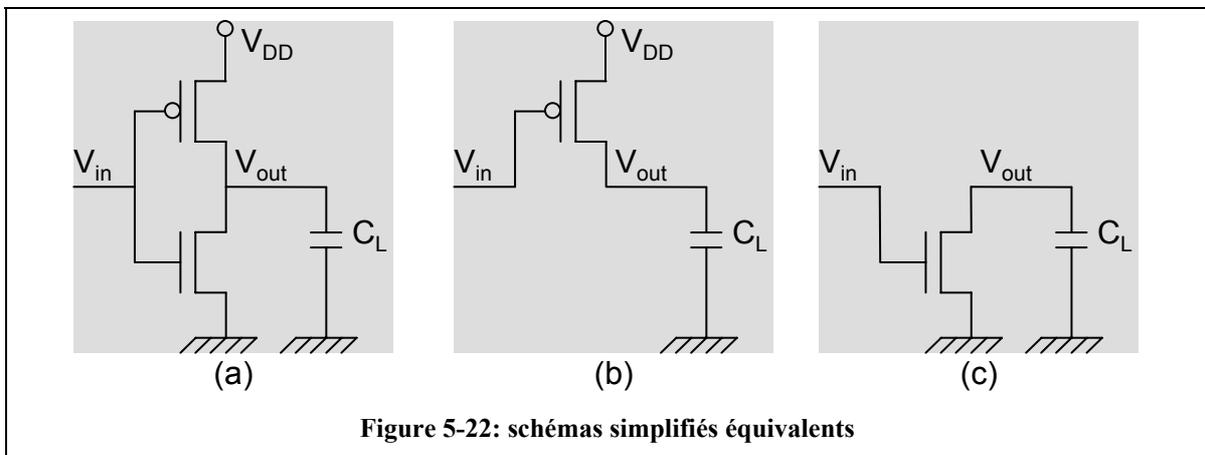
De manière équivalente, on considère que le niveau bas est atteint lorsque

$$V_{out} = 0.1 \cdot V_{DD} \tag{5.28}$$

Pour calculer les temps de montée et descente, il faut considérer les relations entre courant de charge et tension de sortie. D'une part, la relation entre I_c et V_{OUT} est connue: il s'agit de la courbe $I_D(V_D)$ du p-mos pour la montée, de la courbe $I_D(V_D)$ du n-mos pour la descente.



La tension d'entrée est supposée être un échelon idéal, ce qui implique que le transistor qui ne sert pas à la charge est bloqué. Par conséquent, le schéma de l'inverseur peut être réduit au p-mos pour la montée (**figure 5-21a**), au n-mos pour la descente (**figure 5-21a**).



La loi de charge/décharge d'une capacité permet de fournir une autre relation entre tension de sortie et courant de charge:

$$I = C_L \cdot \frac{dV_{out}}{dt} \quad (5.29)$$

C'est donc un système de deux équations qui est disponible pour trouver les solutions $V_{out}(t)$ et $I(t)$. Comme pour le calcul de puissance dynamique, la fréquence maximale de fonctionnement peut être calculée de deux manières:

- de façon analytique lorsque les équations qui régissent le courant sont connues
- à défaut de façon numérique à l'aide de simulations TCAD.

Calcul analytique des temps de montée et descente en CMOS

(a) Calcul du temps de descente

A l'état initial, l'entrée est à l'état bas ($V_{in}=0$) et la sortie à l'état haut ($V_{out}=V_{DD}$). L'entrée passe brusquement à l'état haut ($V_{in}=V_{DD}$). Dans ce cas, le n-mos devient passant et permet de décharger la capacité de charge.

Le transistor travaille d'abord en régime de saturation, de $t=0$ à $t=t_1$, puis en régime linéaire, de $t=t_1$ à $t=t_1+t_2$. La transition se fait lorsque la tension de drain $V_{DN}=V_{out}$ atteint la tension de pincement $V_p=V_G-V_{th}$, soit ici $V_{out}=V_{DD}-V_{thn}$.

L'utilisation des équations du courant du transistor introduites au chapitre 1 permet de déterminer:

$$t_{d1} = \frac{(V_{th} - 0.1 \cdot V_{DD}) C_L}{I_{dsatN}} \quad (5.30)$$

$$t_{d2} = \frac{C_L V_{GT}}{2I_{dsatN}} \ln \left(\frac{2V_{GT}}{0.1 \cdot V_{DD}} - 1 \right)$$

Le temps de descente est la somme de ces deux composantes. En technologie standard, il est courant de fixer:

$$V_{th} = \frac{V_{DD}}{4} \quad (5.31)$$

L'expression du temps de descente peut donc être simplifiée:

$$t_d = \frac{5}{4} \frac{V_{dd} \cdot C_L}{I_{dsatN}} \quad (5.32)$$

(b) Calcul du temps de montée

A l'état initial, l'entrée est à l'état haut ($V_{in}=V_{DD}$) et la sortie à l'état bas ($V_{out}=0$). L'entrée passe brusquement à l'état bas ($V_{in}=0$). Dans ce cas, le p-mos devient passant et permet de charger la capacité de charge.

Le transistor travaille d'abord en régime de saturation, de $t=0$ à $t=t_{m1}$, puis en régime linéaire, de $t=t_{m1}$ à $t=t_{m1}+t_{m2}$. La transition se fait lorsque la tension de drain $V_{DP}=V_{DD}V_{out}$ atteint la tension de pincement $V_P=V_G-V_{th}$, soit ici $V_{out}=V_{th}$.

De la même manière que pour le temps de descente, le temps de montée s'exprime par:

$$t_m = \frac{5 V_{DD} \cdot C_L}{4 I_{dsatP}} \tag{5.33}$$

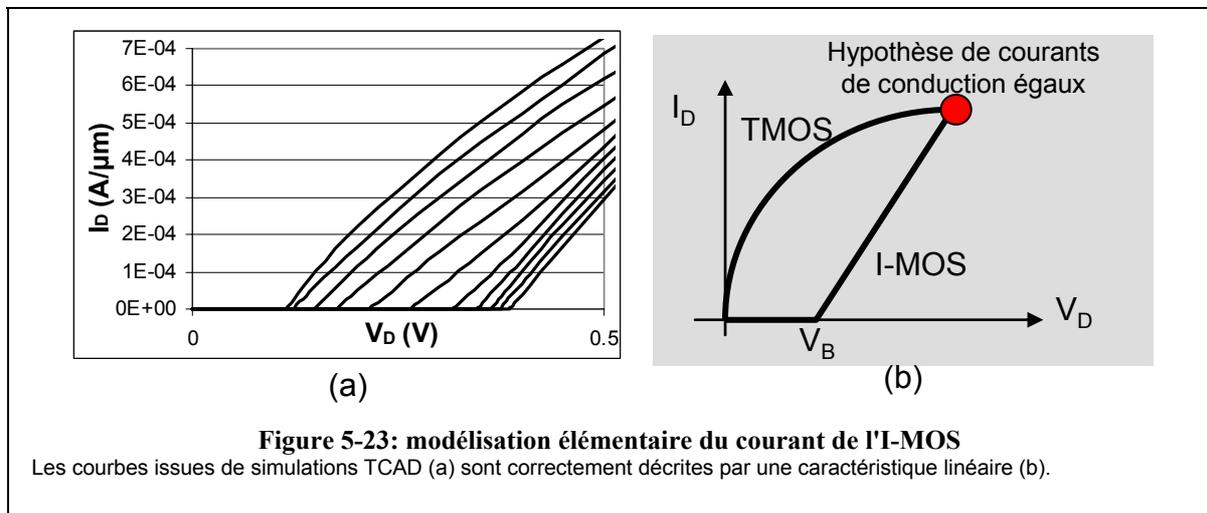
La fréquence maximale de fonctionnement est obtenue lorsque la tension de sortie oscille entre $0.1 \cdot V_{DD}$ et $0.9 \cdot V_{DD}$, soit:

$$f_{max} = \frac{1}{t_d + t_m} = \frac{4 \cdot (I_{dsatN} + I_{dsatP})}{5 \cdot V_{DD} \cdot C_L} \tag{5.34}$$

Calcul des temps de montée et descente en technologie I-MOS complémentaire

Bien que ne disposant pas d'expression analytique du courant pour l'I-MOS, il est possible de le modéliser de façon élémentaire à partir des simulations à notre disposition (**figure 5-23a**). A tension de grille donnée (hypothèse de tension d'entrée en échelons), le courant est négligeable avant la tension d'avalanche la caractéristique est linéaire pour des tensions supérieures. Cette courbe présentée sur la **figure 5-23b** a pour équation:

$$I_n(V_D) = \begin{cases} 0 & \text{si } V_D < V_{br} |_{V_G=V_{OH}} \\ I_{Dsat} \cdot \frac{V_D - V_{br} |_{V_G=V_{OH}}}{V_{DD} - V_{br} |_{V_G=V_{OH}}} & \text{sinon} \end{cases} \tag{5.35}$$



Dans le cas d'un p-mos, cette équation devient:

$$I_p(V_D) = \begin{cases} 0 & \text{si } V_D < V_{br} |_{V_G=V_{OL}} \\ I_{Dsat} \cdot \frac{V_D - V_{br} |_{V_G=V_{OL}}}{V_{DD} - V_{br} |_{V_G=V_{OL}}} & \text{sinon} \end{cases} \quad (5.36)$$

En utilisant les équations(5.29) et (5.36), des expressions similaires à celles obtenues pour le transistor MOS conventionnelles sont obtenues:

$$\begin{aligned} t_m &= 2 \cdot \frac{V_{DD} \cdot C_L}{I_{Dmax,n}} \\ t_d &= 2 \cdot \frac{V_{DD} \cdot C_L}{I_{Dmax,p}} \end{aligned} \quad (5.37)$$

Pour le transistor MOS conventionnel, le facteur de proportionnalité était de 1.25 au lieu de 2.

Calcul numérique des temps de montée et descente

L'approche analytique permet de comparer TMOS et I-MOS de façon globale. Elle permet de relever quels sont les avantages respectifs de ces deux technologies. Cependant, le résultat fourni est plutôt qualitatif. Un résultat plus quantitatif est accessible en déterminant numériquement les temps de montée et de descente à partir des simulations.

Pour cela, les triplets (temps, tension de sortie, courant de sortie) sont calculés par itérations successive.

- la condition initiale correspond à $t=0$, $V=0V$ ou $V=1V$, et le courant correspondant.
- chaque itération peut ensuite être calculée à partir de la précédente:
 - le courant est à $I_{i+1}=I_i+\Delta I$, ou ΔI est une valeur faible devant I_i fixée arbitrairement
 - la tension de sortie V_i est fournie la caractéristique $I(V_{out})$ simulée.
 - le temps entre ces deux itérations est calculé en utilisant la relation habituelle:

$$I = C_L \cdot \frac{dV_{out}}{dt} \quad (5.38)$$

et en considérant linéarisant la relation $I(t)$ entre ces deux itérations:

$$I(t) = \frac{I_{i+1} - I_i}{t_{i+1} - t_i} \cdot (t - t_i) + I_i \quad (5.39)$$

L'intégration des équations (5.38) et (5.39) de t_{i-1} à t_i , permet de calculer:

$$t_{i+1} = t_i + \frac{2 \cdot C_L \cdot (V_{i+1} - V_i)}{I_{i+1} + I_i} \quad (5.40)$$

Le **tableau 11** résume les résultats obtenus pour chacune des technologies et pour chaque approche.

	CMOS (analytique)	CMOS (numérique)	I-MOS (analytique)	I-MOS (numérique)
t_m	16.7 ps	23.78 ps	6.45 ps	8.16 ps
t_d	8.32 ps	9.79 ps	7.7 ps	10.78 ps
$t = t_m + t_d$	25 ps	33.56 ps	14.1 ps	18.94 ps
$F_{max} = 1/t$	40 GHz	29 GHz	70 GHz	53 GHz

Tableau 11: fréquence maximale de fonctionnement

Bien que le calcul analytique fournisse un résultat supérieur de 30% supérieur aux résultats numériques dans les 2 technologies, les deux résultats convergent vers l'idée que l'inverseur en technologie I-MOS complémentaire est 80% plus rapide que l'inverseur CMOS. Ceci est dû à la fois au courant supérieur (d'un facteur 4 pour le pmos), mais est limité par la forme de la caractéristique.

5.3. Circuits élémentaires

5.3.1. La porte NAND

La porte NAND réalise la fonction logique suivante: "la sortie est à l'état bas si et seulement si toutes les entrées sont à l'état haut.". Nous nous intéresserons au cas d'une NAND à deux entrées, dont la table de vérité est donnée dans le **tableau 12**.

a	b	$\overline{a.b}$
0	0	1
0	1	1
1	0	1
1	1	0

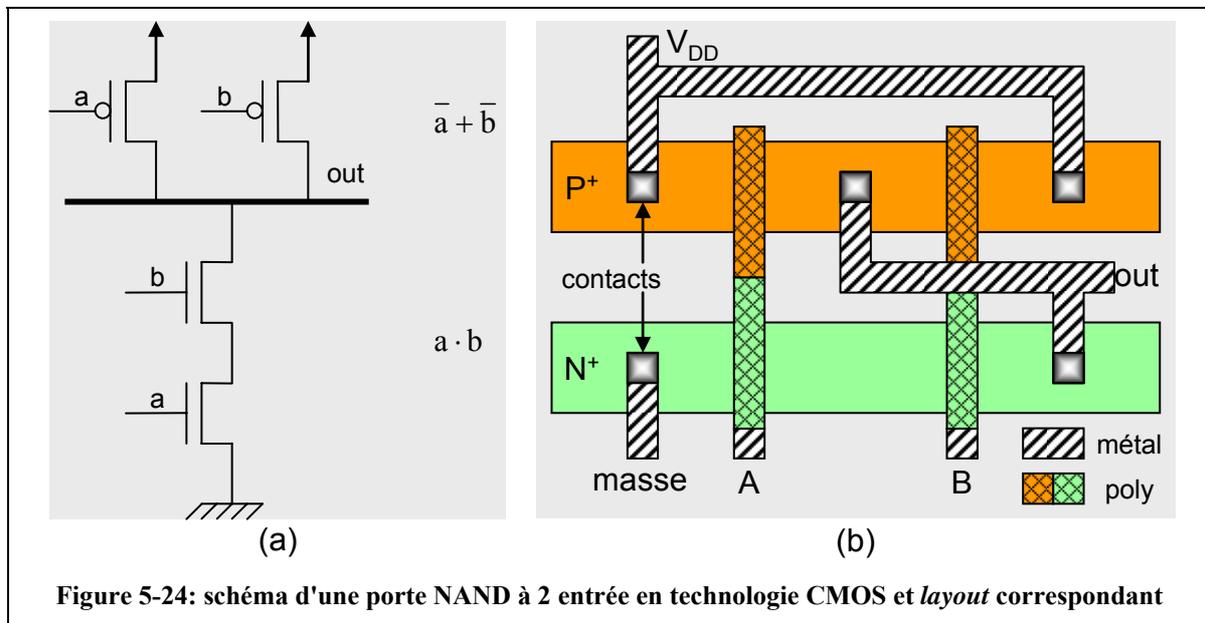
Tableau 12: table de vérité d'une porte NAND à deux entrées

5.3.1.1. Intégration CMOS

Comme pour l'inverseur, il est possible de réaliser cette fonction uniquement avec des transistors n-mos, ou bien en CMOS. La solution CMOS est préférable car le circuit consomme très peu en dehors des commutations.

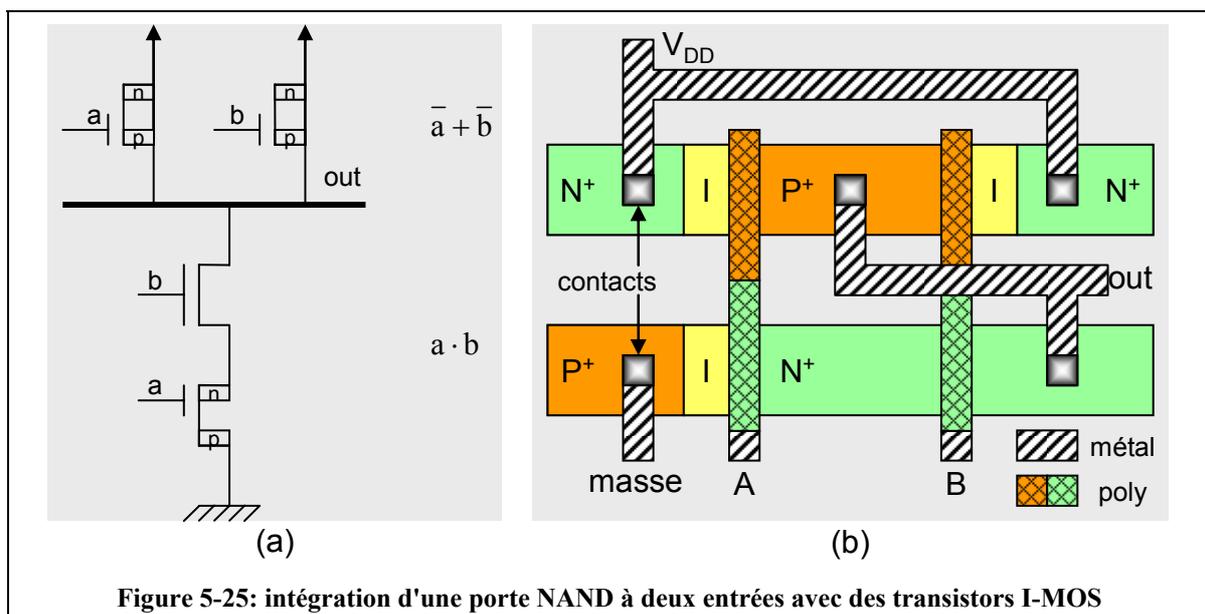
L'intégration CMOS (**figure 5-24**) peut être décomposée en deux parties:

- les deux transistors p-mos en haut réalisent la fonction logique $\bar{a} + \bar{b}$. Ils servent à mettre la sortie à l'état haut lorsque l'une au moins des entrées est à l'état bas.
- les deux transistors n-mos du bas réalisent la fonction logique $a \cdot b$. Ils servent à mettre la sortie à l'état bas lorsque les deux entrées sont à l'état haut.



5.3.1.2. Intégration I-MOS complémentaire

Les deux transistors n-mos sont en série. Par conséquent, s'ils étaient remplacés par des transistors I-MOS, il faudrait une différence de potentiel au minimum égale à deux fois la tension d'avalanche, pour qu'ils atteignent tous deux leur mode de conduction. Cette différence de potentiel s'ajoute à celle des p-mos, ce qui amènerait à une alimentation de



l'ordre de 15V.

Afin d'éviter cela, une solution consiste à conserver un n-mos de type MOS conventionnel, et de remplacer par des I-MOS les trois autres transistors. Le schéma d'intégration de cette intégration est présenté sur la **figure 5-25**.

Evaluation

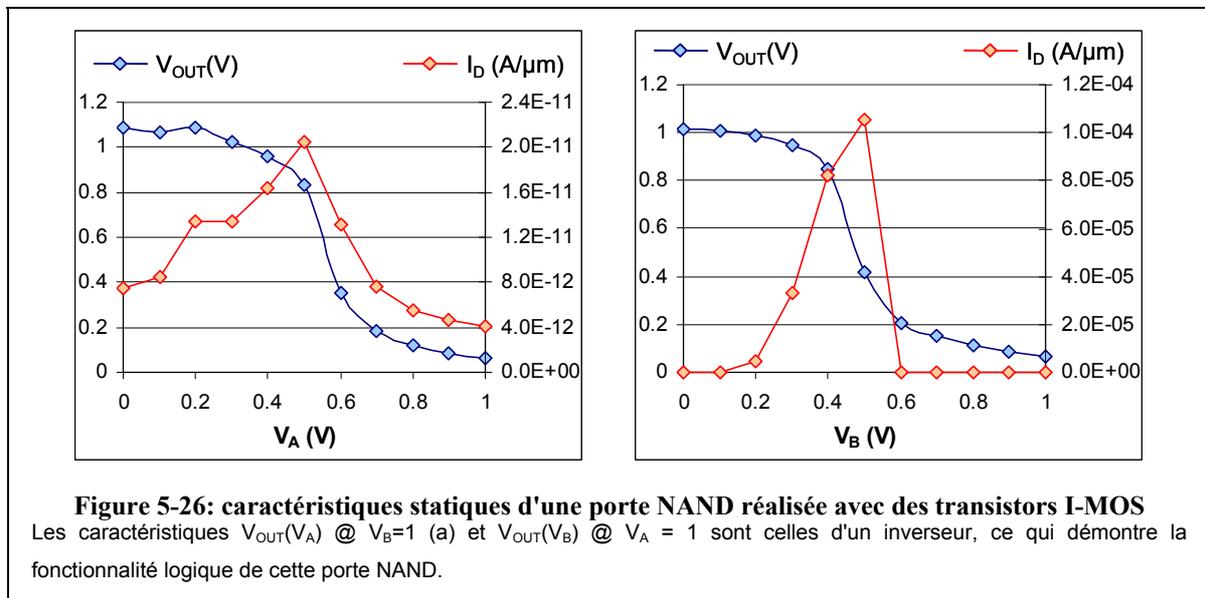
La table de vérité de la porte NAND permet de déterminer comment démontrer la fonctionnalité d'une porte NAND à deux entrées:

- à $V_A=0V$, la caractéristique $V_{out}(V_B)$ doit être celle d'un inverseur
- à $V_B=0V$, la caractéristique $V_{out}(V_A)$ doit être celle d'un inverseur

La **figure 5-26** présente ces caractéristiques obtenues à l'aide de simulations. Deux demi-circuits ont été simulés séparément (les 2 n-mos d'une part, les deux p-mos d'autre part), et la caractéristique totale a été obtenue par construction numérique, de façon similaire à un inverseur dont ont dispositif des caractéristiques de chacun des transistors le composant.

Ces caractéristiques, similaires à celle d'un inverseur, démontrent la fonctionnalité du circuit:

- la tension de sortie bascule de 1V à 0V lorsque la tension d'entrée varie de 0V à 1V
- le courant de court-circuit, très faible aux états logiques, est important lors du basculement, permettant la charge ou la décharge des capacités du circuits.



5.3.2. La porte NOR

La porte NOR réalise la fonction logique suivante: "la sortie est à l'état haut si et seulement si toutes les entrées sont à l'état bas". Nous nous intéresserons au cas d'une NOR à deux entrées, dont la table de vérité est donnée dans le tableau 13.

a	b	$\overline{a \cdot b}$
0	0	1
0	1	0
1	0	0
1	1	0

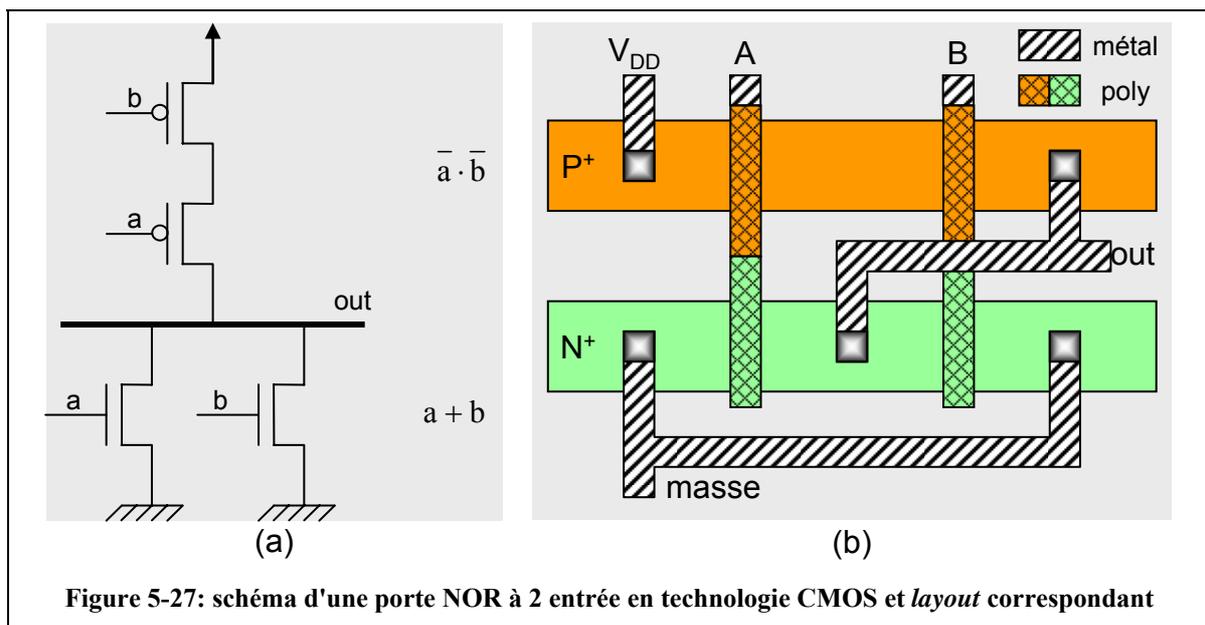
Tableau 13: table de vérité d'une porte NOR à deux entrées

5.3.2.1. Intégration CMOS

Comme pour l'inverseur, il est possible de réaliser cette fonction uniquement avec des transistors n-mos, ou bien en CMOS. La solution CMOS est préférable car le circuit consomme très peu en dehors des commutations.

L'intégration CMOS (**figure 5-27**) peut être décomposée en deux parties:

- les deux transistors p-mos en haut réalisent la fonction logique $\overline{a \cdot b}$. Ils servent à mettre la sortie à l'état haut lorsque les deux entrées sont à l'état bas.
- les deux transistors N-MOS du bas réalisent la fonction logique $a + b$. Ils servent à mettre la sortie à l'état bas lorsque l'une des entrées est à l'état haut.



5.3.2.2. Intégration I-MOS complémentaire

Les deux transistors n-mos sont en série. Par conséquent, s'ils étaient remplacés par des transistors I-MOS, il faudrait une différence de potentiel au minimum égale à deux fois la tension d'avalanche, pour qu'ils atteignent tous deux leur mode de conduction. Cette différence de potentiel s'ajoute à celle des p-mos, ce qui amènerait à une alimentation de l'ordre de 15V.

Afin d'éviter cela, une solution consiste à conserver un n-mos de type MOS conventionnel, et de remplacer par des I-MOS les trois autres transistors. Le schéma d'intégration de cette intégration est présenté sur la **figure 5-28**.

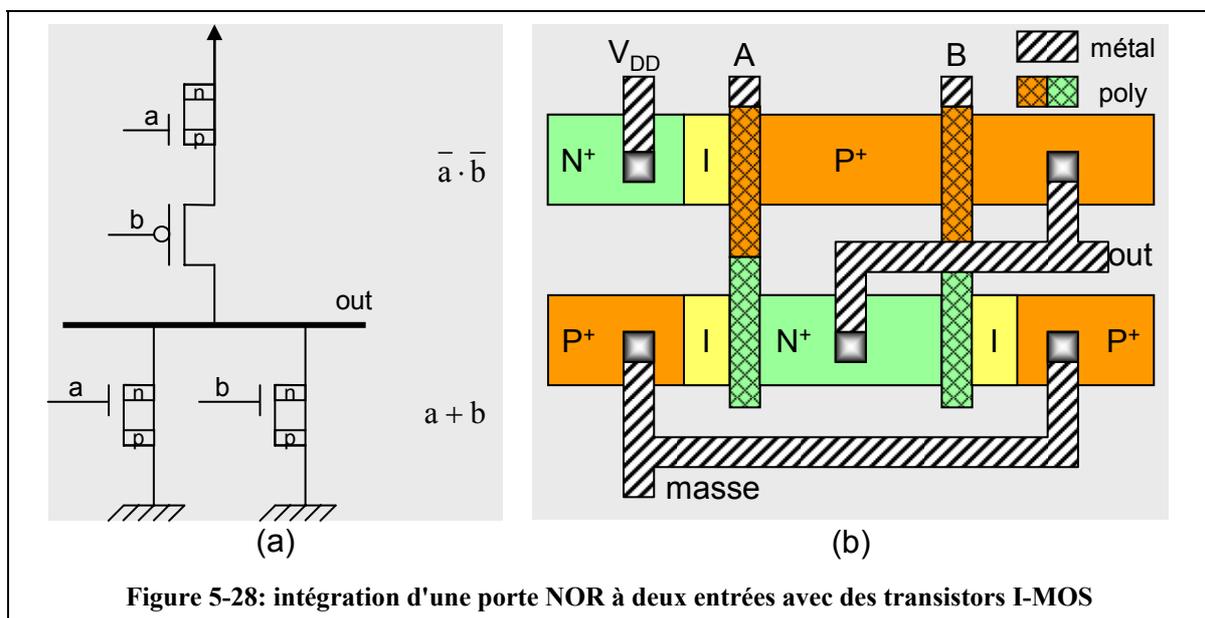


Figure 5-28: intégration d'une porte NOR à deux entrées avec des transistors I-MOS

Evaluation

La table de vérité de la porte NAND permet de déterminer comment démontrer la fonctionnalité d'une porte NAND à deux entrées:

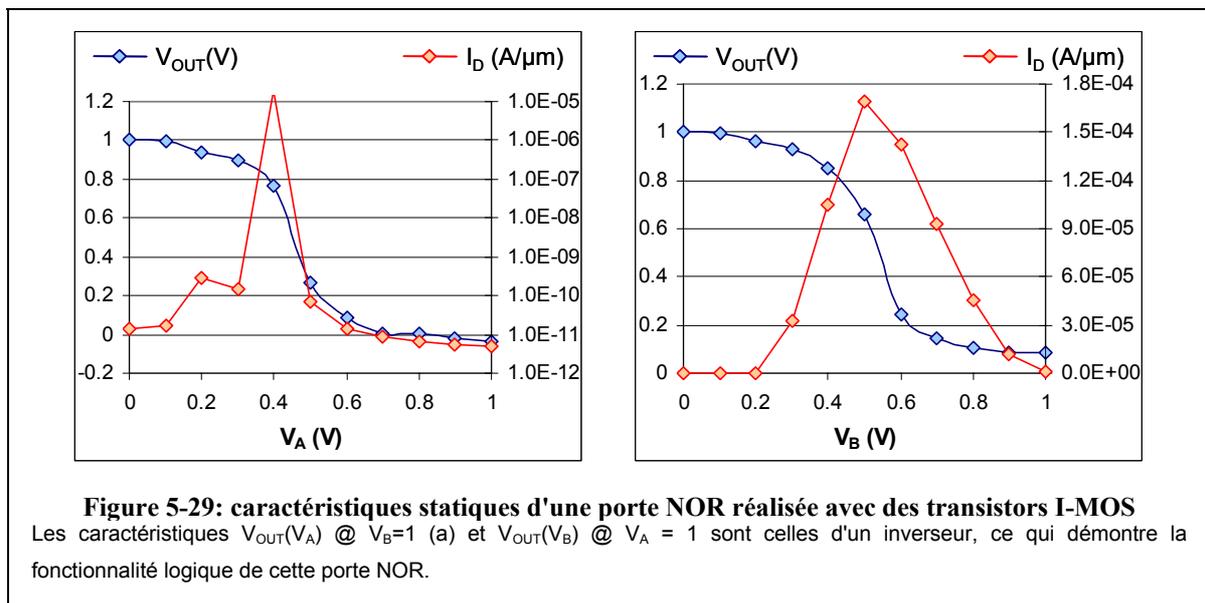
- à $V_A=0V$, la caractéristique $V_{out}(V_B)$ doit être celle d'un inverseur
- à $V_B=0V$, la caractéristique $V_{out}(V_A)$ doit être celle d'un inverseur

La **figure 5-29** présente ces caractéristiques obtenues à l'aide de simulations. Deux demi-circuits ont été simulés séparément (les 2 n-mos d'une part, les deux p-mos d'autre part), et la caractéristique totale a été obtenue par construction numérique, de façon similaire à un inverseur dont ont dispositif des caractéristiques de chacun des transistors le composant.

Ces caractéristiques, similaires à celle d'un inverseur, démontrent la fonctionnalité du circuit:

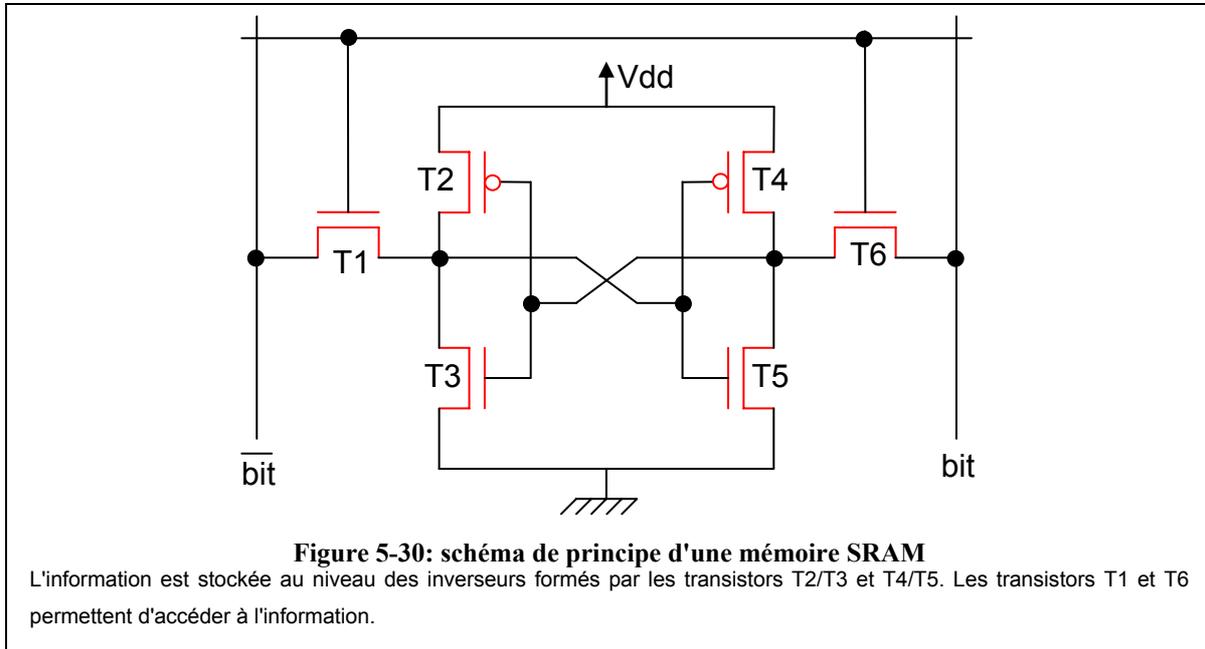
- la tension de sortie bascule de 1V à 0V lorsque la tension d'entrée varie de 0V à 1V

- le courant de court-circuit, très faible aux états logiques, est important lors du basculement, permettant la charge ou la décharge des capacités du circuits.



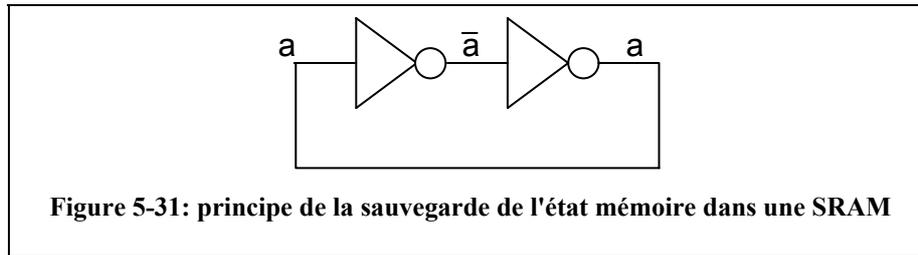
5.4. La SRAM

5.4.1. Présentation



Dans une SRAM (*Static Random Access Memory*), l'information n'a pas besoin d'être rafraîchie, contrairement à la DRAM (*Dynamic Random Access Memory*). C'est une mémoire volatile, c'est-à-dire que l'information n'est pas conservée en cas de coupure d'alimentation. Elle fonctionne sur le principe du verrouillage entre deux inverseurs formés sur la **figure 5-30**

par les transistors T2 et T3 d'une part, et T4 et T5 d'autre part. La sortie de l'un est reliée à l'entrée de l'autre. Ainsi, l'information circule de façon stable:



Les transistors d'accès T1 et T6 permettent de lire et d'écrire l'état de la cellule mémoire.

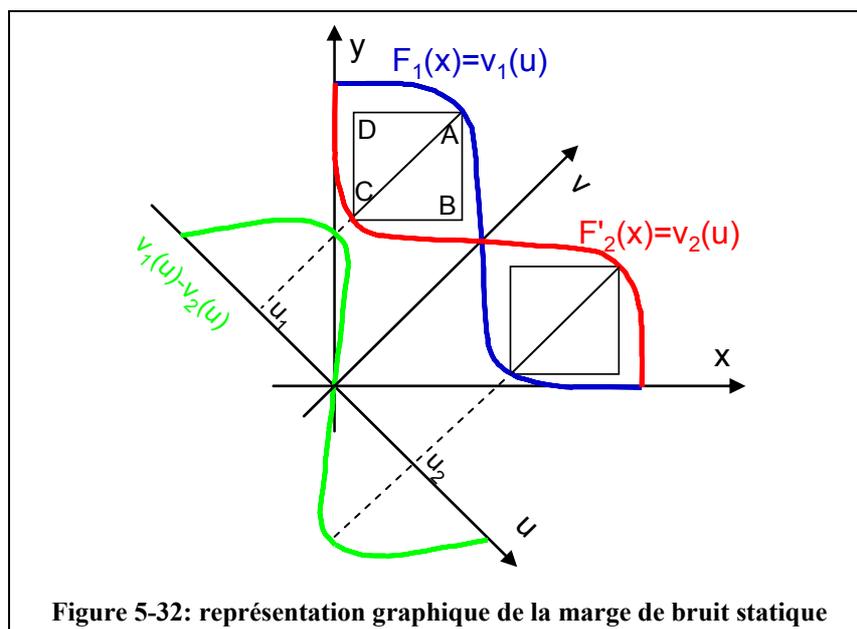
Les SRAM sont utilisées pour les applications logiques nécessitant de grandes fréquences de fonctionnement, typiquement la mémoire cache des ordinateurs.

Les deux paramètres importants d'une cellule SRAM sont son encombrement et sa stabilité. Cette dernière détermine le taux d'erreur ainsi que sa sensibilité aux fluctuations du procédé de fabrication. La stabilité peut être mesurée à l'aide de la marge de bruit statique SNM (*Static Noise Margin*), qui peut être calculé à partir des caractéristiques des inverseurs qui la composent. On considère qu'une cellule SRAM est stable si sa marge de bruit est supérieure à 10% de sa tension d'alimentation:

$$\text{SNM} > 10\% \cdot V_{\text{DD}} \quad (5.41)$$

5.4.2. Calcul de la marge de bruit statique

[Seevinck'87] propose une méthode pour calculer la marge de bruit statique d'une cellule SRAM à partir des caractéristiques statiques des inverseurs qui la composent. Pour cela, il faut d'abord tracer dans le repère usuel (x,y) la caractéristique de sortie $V_{\text{out}}(V_{\text{in}})$ du premier



inverseur. Sur la figure 5-32, il s'agit de la courbe $y=F_1(x)$. Puis, la symétrique par rapport à l'axe $y=x$ de la caractéristique $V_{out}(V_{in})$ du second inverseur est portée sur le même graphique. Sur la figure 5-32, il s'agit de la courbe $y=F'_2(x)$. Ces deux courbes forment la figure caractéristique en "ailes de papillons". Graphiquement, la marge de bruit statique représente le côté du plus grand carré qu'il est possible d'inscrire simultanément dans chacune des deux ailes.

Il est également possible de calculer la marge de bruit statique de façon analytique.

Soit (u,v) le repère (x,y) tourné de 45° dans le sens anti-trigonométrique. Dans ce repère, les courbes $y=F_1(x)$ et $y=F'_2(x)$ s'écrivent alors $v=v_1(u)$ et $v=v_2(u)$. Le changement de repère s'effectue à l'aide des relations:

$$\begin{aligned} u &= \frac{x-y}{\sqrt{2}} \\ v &= \frac{x+y}{\sqrt{2}} \end{aligned} \quad (5.42)$$

La diagonale d'un carré inclus entre ces deux courbes est simplement $v_1(u)-v_2(u)$. Maximiser le carré équivaut à maximiser sa diagonale, et donc à rechercher le maximum de la fonction $v_2(u)-v_1(u)$. Il existe deux solutions, une par "aile de papillon". Elles sont placées sur la **figure 5-32** aux abscisses u_1 et u_2 . La marge de bruit statique est la plus petite des deux différences v_2-v_1 à $\sqrt{2}$ près:

$$SNM = \frac{\min(v_2(u_1) - v_1(u_1), v_2(u_2) - v_1(u_2))}{\sqrt{2}} \quad (5.43)$$

Par ailleurs, Les quatre sommets du carré ont comme coordonnées:

	A	B	C	D
u	u_1	$u_1 + \frac{v_1(u_1) - v_2(u_1)}{2}$	u_1	$u_1 - \frac{v_1(u_1) - v_2(u_1)}{2}$
v	$v_1(u_1)$	$v_1(u_1) + \frac{v_1(u_1) - v_2(u_1)}{2}$	$v_2(u_1)$	$v_1(u_1) + \frac{v_1(u_1) - v_2(u_1)}{2}$

Tableau 14: coordonnées des sommet du carré déterminant la marge de bruit statique

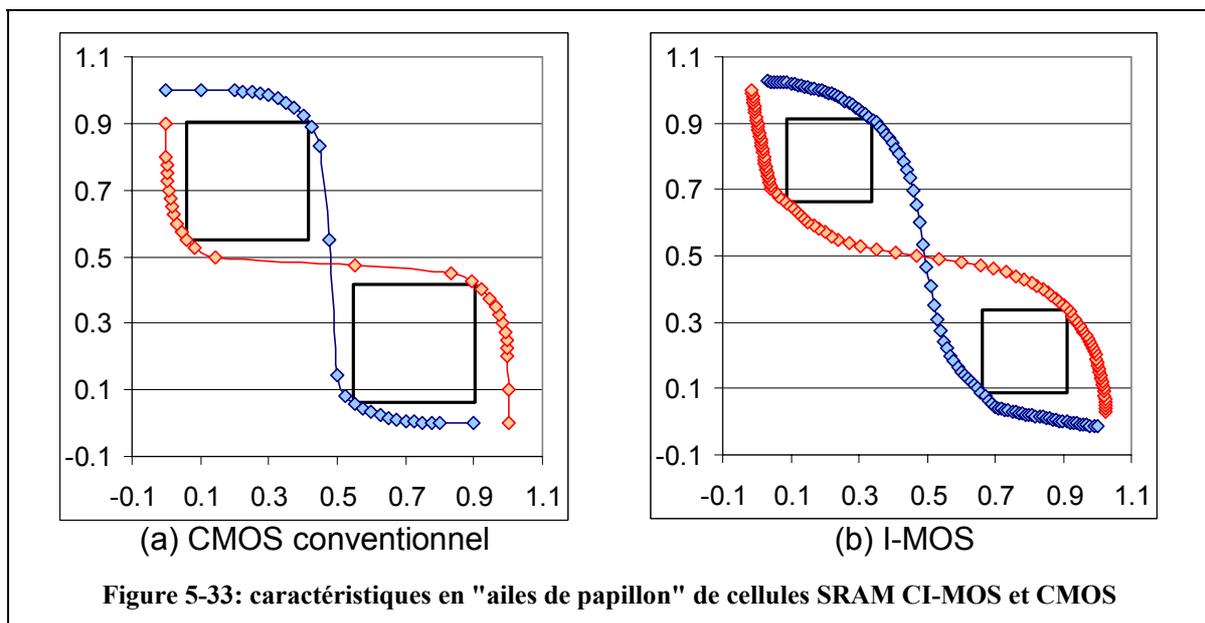
Pour avoir les coordonnées dans le repère (x,y) , il suffit d'utiliser les formules de changement de repère:

$$\begin{aligned} x &= \frac{1}{\sqrt{2}} \cdot u + \frac{1}{\sqrt{2}} \cdot v \\ y &= -\frac{1}{\sqrt{2}} \cdot u + \frac{1}{\sqrt{2}} \cdot v \end{aligned} \quad (5.44)$$

5.4.3. Marge de bruit d'une cellule SRAM réalisée en technologie I-MOS

Le calcul de la marge de bruit statique nécessite la caractéristique des inverseurs qui la composent. En supposant qu'ils sont identiques, à ceux présentés en première partie de ce chapitre, il est donc possible de déterminer la marge de bruit d'une SRAM réalisée avec des transistors I-MOS, et de la comparer avec la référence CMOS conventionnelle.

La **figure 5-33** montre les caractéristiques en "ailes de papillons" pour chacune des technologies. La marge de bruit en I-MOS est moindre qu'en CMOS conventionnel (251mV contre 365mV), mais largement au-dessus du seuil communément accepté ($0.1 \cdot V_{DD} = 100\text{mV}$).



Conclusion du chapitre

La première partie de ce chapitre a détaillé la démarche permettant d'établir les règles de conception d'un inverseur réalisé en technologie I-MOS complémentaire en terme de dimensions et de polarisations. A partir de là, il est devenu possible de comparer ses performances avec celles d'un transistor réalisé en technologie CMOS conventionnelle, en se fondant sur de la modélisation simple et des simulations TCAD. L'inverseur CI-MOS est plus rapide et présente une consommation statique inférieure, grâce à ses excellentes performances I_{on}/I_{off} . Il est cependant handicapé par la nécessité d'une tension d'alimentation élevée en ce qui concerne les consommations dynamique et statique.

Ce gain de vitesse est profitable pour d'autres circuits élémentaires, tels la porte NAND ou la porte NOR, ou bien la mémoire de type SRAM dont la marge de bruit statique est suffisante pour assurer un fonctionnement correct ($SNM=250mV@V_{DD}=1V$).

Les deux chapitres précédents ont montré le transistor I-MOS isolé ou en circuit en se fondant sur des simulations TCAD. Cependant, les simulations gommant un certain nombre de questions pratiques: comment réaliser des dispositifs complémentaires sur une même plaque? Peut-on réaliser des transistors aussi petits que désirés? Comment réaliser des dispositifs asymétriques? L'intégration est-elle compatible avec les procédés actuels?

C'est à ces questions auxquelles va essayer de répondre le prochain et dernier chapitre, en traitant de l'intégration de dispositifs I-MOS.

Chapitre 6. Intégration de l'I-MOS

L'intégration d'un dispositif I-MOS soulève le problème de ses asymétries:

- d'une part, la source et le drain sont dopés de nature différente.
- d'autre part, le drain doit être recouvert par la grille, mais pas la source.

En outre, l'étude théorique faite au chapitre 4 montre que la longueur intrinsèque (distance entre la source et le bord de la grille) est cruciale pour le comportement électrique du dispositif. Il convient donc d'obtenir cette dimension de façon précise et reproductible. Enfin, cette étude avait également montré que réduire les dimensions permet de réduire la tension d'avalanche du dispositif.

L'étude des différentes approches présentées dans la littérature présente la réduction des dimensions comme le principal moyen de réduire la tension d'avalanche.

En nous inspirant de ces différents procédés de fabrication, nous proposons une intégration horizontale "classique". L'asymétrie de dopage est effectuée par masquage successif de la source et du drain, alors que la longueur intrinsèque est déterminée par la taille d'espaces en diélectrique. Les dispositifs obtenus comptent parmi les plus petits jamais présentés, et les résultats électriques montrent des niveaux de courant et de conductivité exceptionnels en technologie silicium.

Cependant, la modélisation prédit que la taille de ces dispositifs ne pourra pas descendre sous les 20nm, sous peine de dégradation des performances. Nous proposons donc comme alternative une intégration verticale, permettant d'obtenir une dimension utile supérieure à 20nm, tout en présentant un encombrement réduit.

Introduction

Plusieurs solutions existent dans la littérature pour fabriquer des dispositifs de type I-MOS. La principale difficulté réside dans les asymétries du dispositif:

- d'une part, la source et le drain sont dopés de nature différente.
- d'autre part, la zone de drain doit s'étendre jusque légèrement sous la grille, comme pour un transistor MOS conventionnel, alors que la source doit être éloignée de la grille.

Parmi les critères de qualité d'une intégration, l'une des plus importantes est la **définition de la longueur intrinsèque**. En effet, l'étude théorique de l'I-MOS faite au chapitre 4 a montré que cette dimension est critique pour le comportement électrique du dispositif. Les autres qualités doivent être observées sont:

- la compatibilité avec le procédé de fabrication du transistor MOS conventionnel
- la possibilité de réaliser des dispositifs complémentaires
- la compacité.

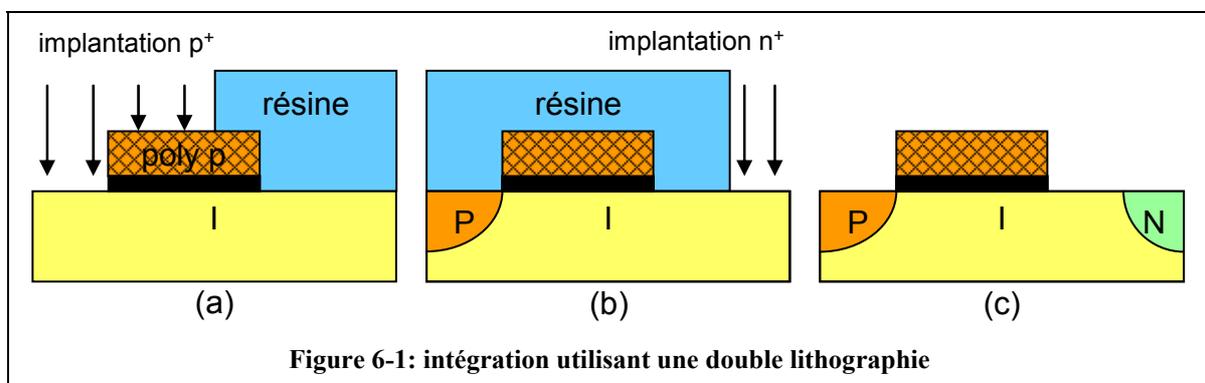
Les transistors à effet tunnel présentant la même asymétrie de dopage entre source et drain, leur réalisation est également considérée.

6.1. Etat de l'art

6.1.1. Double lithographie

C'est la première intégration proposée [Gopalakrishnan'02,Gopalakrishnan'05b]. Le dispositif de type p a été réalisé sur SOI.

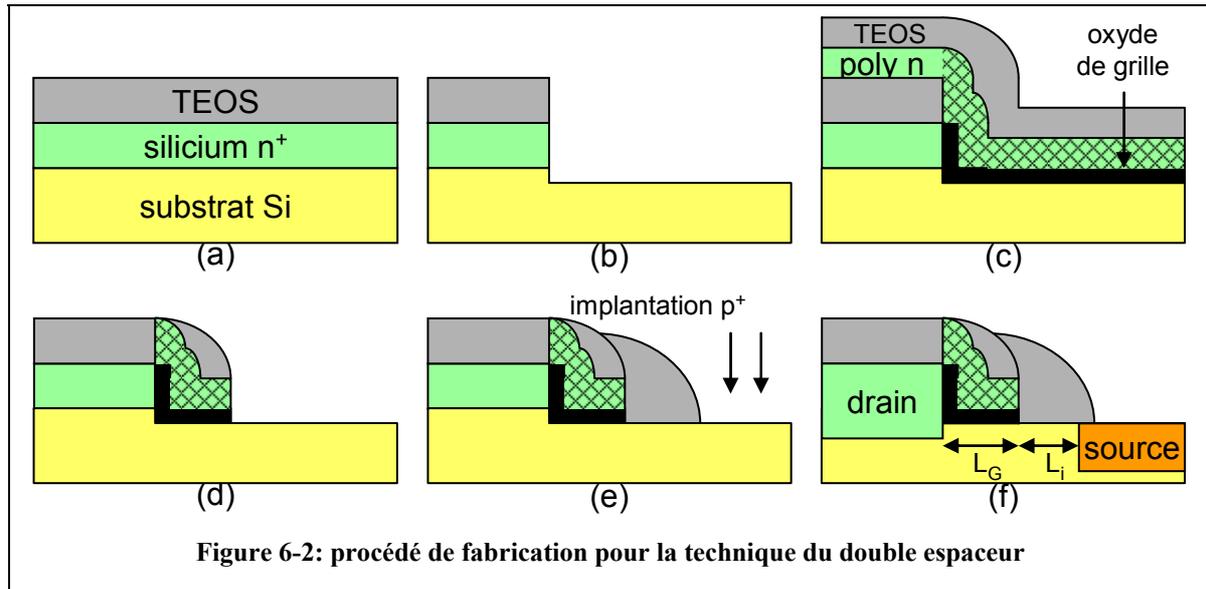
- la grille est réalisée de façon classique
- la source est masquée, ce qui permet d'implanter le drain (**Figure 6-1a**).
- puis le drain et la zone intrinsèque sont masqués, ce qui permet d'implanter la source (**Figure 6-1b**).



Ce procédé de fabrication est très simple à mettre en oeuvre. Il a permis de démontrer le principe de l'I-MOS. Cependant, il n'autorise pas la réalisation de petites dimensions: la définition de la zone intrinsèque, notamment, est définie par l'alignement relatif de deux photolithographies, ce qui est loin d'être le procédé le plus précis disponible.

Cette méthode a été reprise récemment dans [Mayer'06]. En utilisant un jeu de masques dédié, l'auteur a montré la co-intégration de transistors I-MOS et de transistors MOS conventionnels sur la même plaque, chacun étant réalisé en exemplaires complémentaires (n-mos et p-mos). Si la grande taille des dispositifs I-MOS impose des tensions de polarisation élevées (10 à 20V), ils présentent la pente sous le seuil la plus faible de la littérature (2mV/dec).

6.1.2. Double espaceur



Pour définir de façon précise la grille et la zone intrinsèque, deux espaceurs sont utilisés (**Figure 6-2**) [Choi'04a].

- une couche de silicium dopé est épitaxiée. Elle formera le drain. Puis elle est recouverte d'un dépôt de TEOS (**figure 6-2a**)
- ces deux couches sont gravées (**figure 6-2b**)
- l'oxyde de grille est réalisé par oxydation thermique. Puis, le polysilicium de grille est déposé et une seconde couche de TEOS est déposée (**figure 6-2c**).
- cette seconde couche sert à réaliser un premier espaceur qui détermine la longueur de grille (**figure 6-2d**)
- un second espaceur est formé pour repousser la source (**figure 6-2e**).

Cette intégration a permis de réduire les dimensions ($L_G=130\text{nm}$), d'où des tensions de claquages plus faibles et des performances globalement meilleures. C'est également la première publication montrant un dispositif de type n. Le procédé a été amélioré en travaillant les recuits d'activation [Choi'04b,Choi'05b].

Le principal inconvénient est la compacité: le dispositif réalisé est "circulaire", ce qui est spaciophage (**Figure 6-3a**). Ceci peut vraisemblablement être attribué à la réutilisation d'un jeu de masques déjà existant. Pour pallier ce problème [Choi'05a], l'auteur a utilisé un autre jeu de masques, similaire à ceux du transistor MOS (**Figure 6-3b**). Cela a permis de réduire encore les dimensions ($L_G=70\text{nm}$). Un substrat SOI a été utilisé pour réduire les courants de fuite. Notons que ce procédé est assez éloigné de celui du TMOS, rendant la co-intégration délicate.

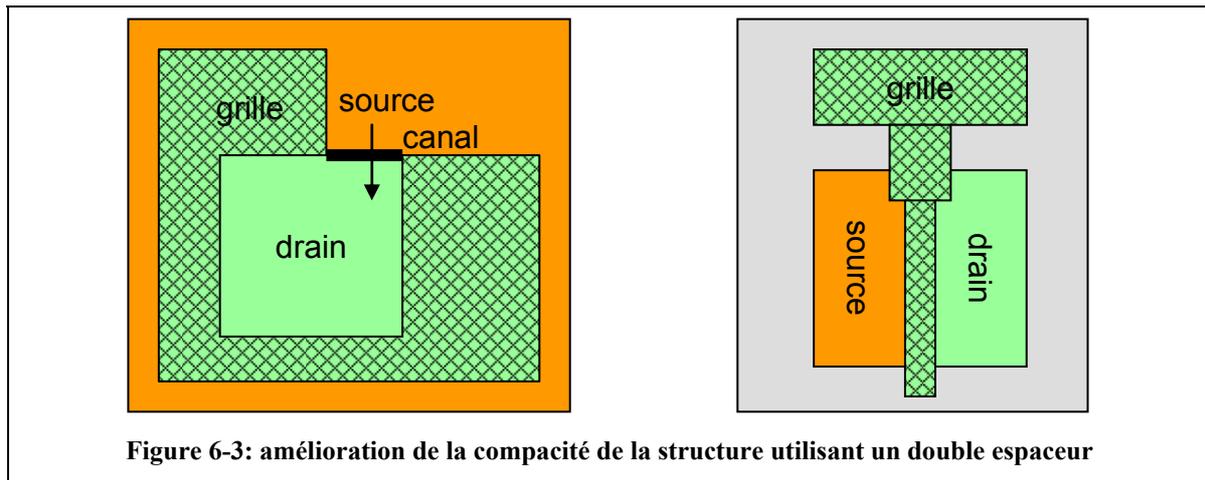


Figure 6-3: amélioration de la compacité de la structure utilisant un double espaceur

6.1.3. I-MOS en forme de L

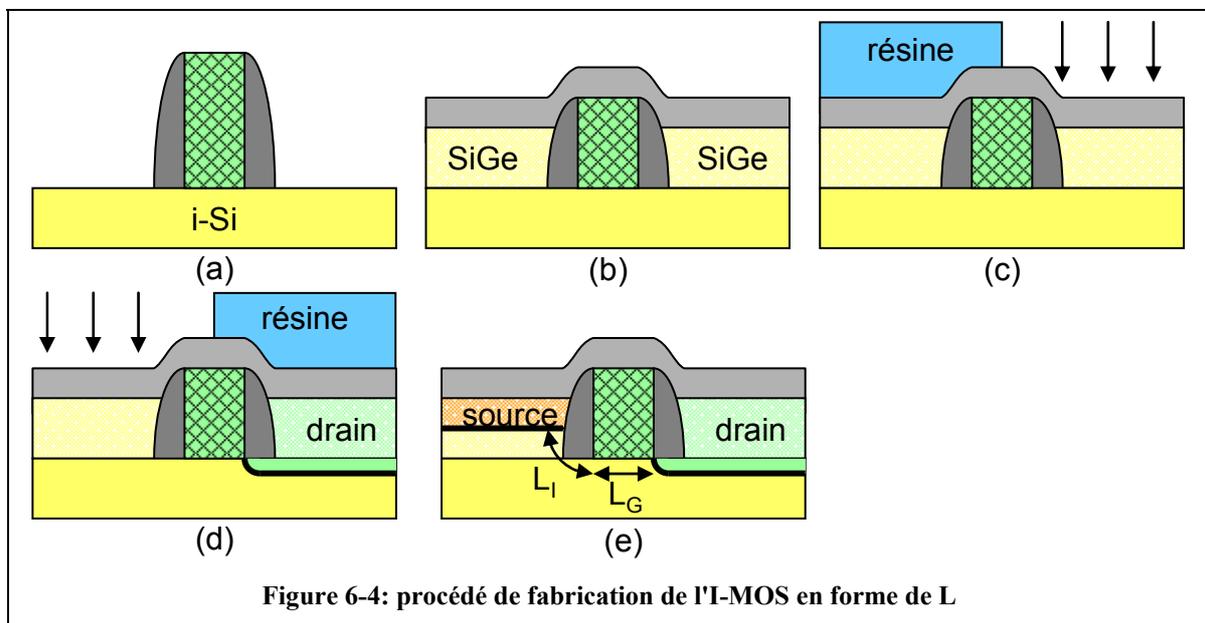


Figure 6-4: procédé de fabrication de l'I-MOS en forme de L

Cette proposition d'intégration est plus récente [Toh'05].

- la grille est réalisée de façon standard sur ce qui sera la zone intrinsèque (**figure 6-4a**)
- les zones source et drain sont surélevées par épitaxie (**figure 6-4b**)
- la source est masquée, ce qui permet d'implanter le drain (**figure 6-4c**)
- le drain est masqué, ce qui permet d'implanter la source (**figure 6-4d**)

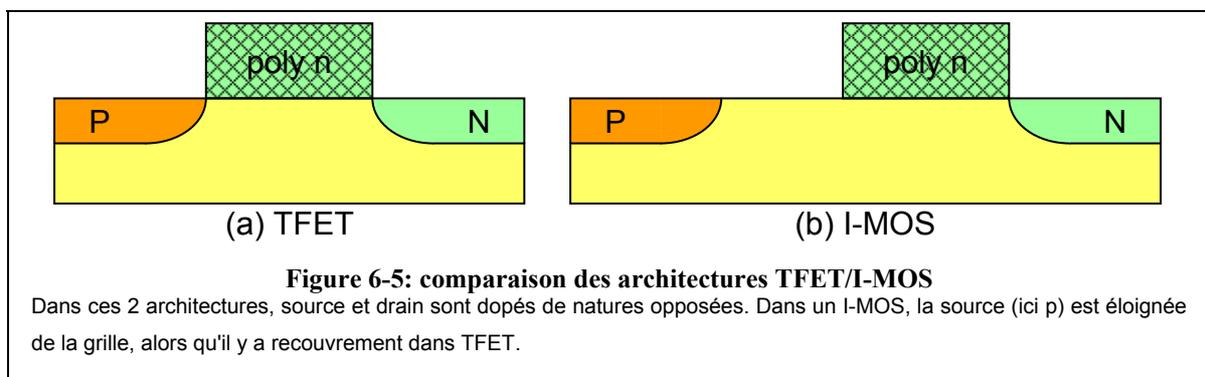
Dans cette intégration, les longueurs sont déterminées par la profondeur d'implantation et la longueur de diffusion. Ce procédé présente certains d'avantages:

- il est très proche du procédé conventionnel CMOS: la co-intégration semble possible
- le dispositif réalisé présente la même compacité qu'un transistor MOS
- il est possible d'intégrer un autre matériau dans l'épitaxie favorisant l'ionisation par impact (du SiGe par exemple), bien qu'une partie de la zone intrinsèque reste en silicium.

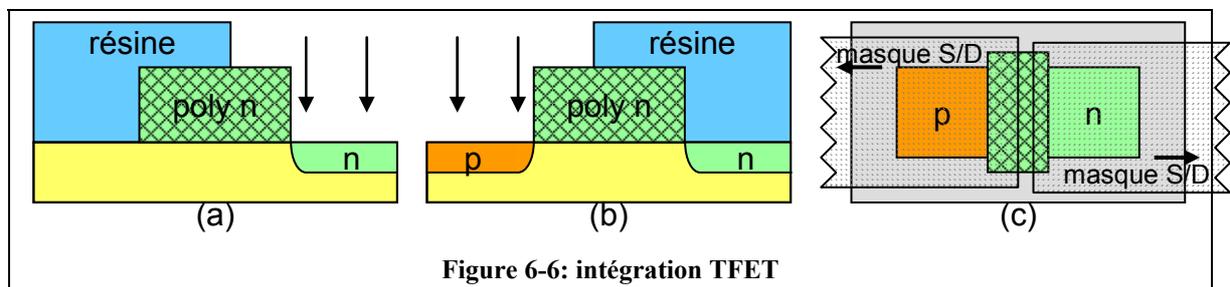
Cette architecture a permis de réaliser un dispositif encore plus petit ($L_G=60\text{nm}$). Mais malgré l'inclusion de SiGe dans les source/drain, les tensions de claquage sont encore élevées ($\sim 8\text{V}$).

6.1.4. Utilisation du masque S/D décalé

Le transistor à effet tunnel ou TFET est un dispositif qui se base sur un transfert de charges par effet tunnel. Une de ses intégrations se présente sous la forme d'une diode p-i-n surmontée d'une grille, de façon similaire à l'I-MOS. La différence est que la grille du TFET recouvre entièrement la zone intrinsèque (**Figure 6-5**). Cette similitude incite à étudier de quelle façon ces dispositifs sont intégrés.



Parmi ces intégrations, il a été proposé [Nirschl'05b] d'utiliser le masque S/D en le décalant pour masquer sélectivement le drain ou la source, de manière similaire à l'intégration par double lithographie (**Figure 6-6**). Cette solution n'offre que très peu de différence avec un procédé de fabrication CMOS conventionnel, permettant donc de se baser sur un procédé industriel. En outre, avec peu de développement à faire et pas de masque supplémentaire à créer, l'investissement reste très limité. La principale difficulté réside dans le fait d'aligner deux fois de suite une photolithographie sur une grille de dimensions réduites.



6.1.5. Comparaison des intégrations

Les différents facteurs à prendre en compte permettant de juger ces dispositifs ont été évoqués en introduction. Le **tableau 6-1** permet de les comparer, alors que **tableau 6-2** recense les performances électriques obtenues selon les différentes publications.

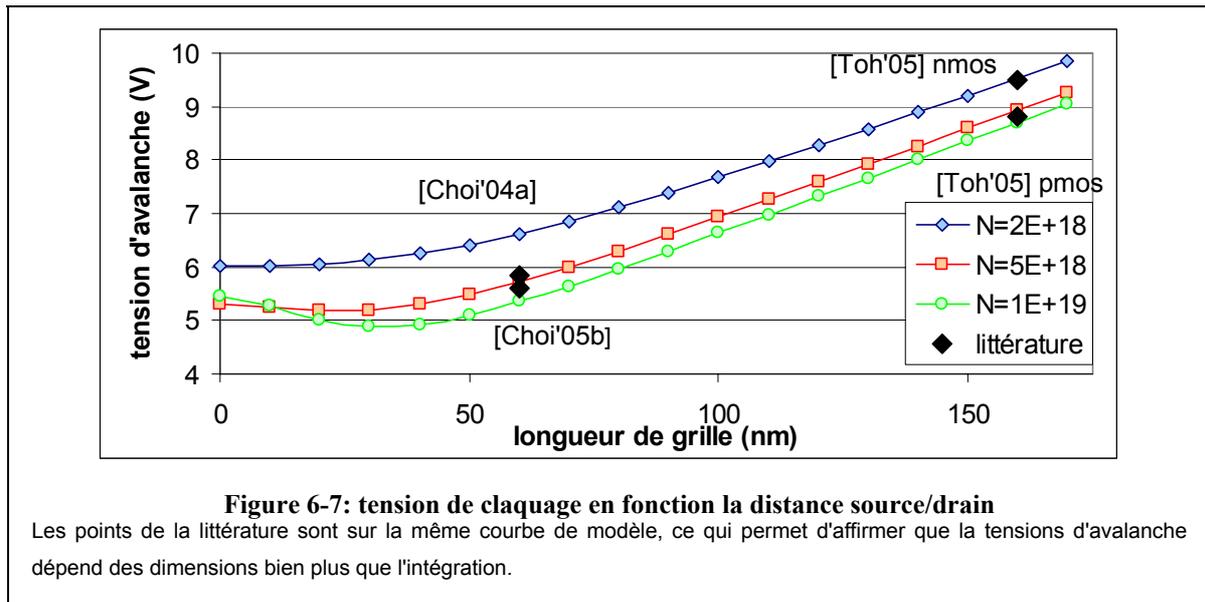
intégration	définition L_i	définition L_G	complémentarité	réduction des dimensions	co-intégration CMOS	compacité
double lithographie	- -	+	+	- -	+	-
double espaceur	+	+	-	+	-	\pm
en L	-	+	+	+	+	+
masque S/D décalé	+	+	+	+	+	+

Tableau 6-1: comparaison des qualités des différentes intégrations I-MOS

publication	[Gopalakrishnan'02]	[Choi'04a]		[Choi'04b]		[Choi'05b]		[Toh'05]		[Mayer'06]
	titre	I-MOS: A Novel Semiconductor	A new fabrication method	80nm self aligned...		100-nm n-/p-Channel		A Novel CMOS compatible		Co-integration of 2mV Subthreshold
méthode	double litho	double espaceur				en L		double lithographie		
substrat	SOI	bulk	bulk		bulk		bulk		SOI	
t_{ox}	20nm	4nm	3nm		3nm		3nm		9nm	
L_i (nm)	0.5 μ m	60	60		60		100	60	400	
L_g (nm)	2 μ m	130	80		100		60	60	1000	
type	p-mos	n-mos	n-mos	p-mos	n-mos	p-mos	n-mos	n-mos	p-mos	
V_s	-20V	-5.75			-5.5	5.5	-8.5	-7.8	18	
$V_{th,lin}$	-2V	0.45	0.6V	-0.45	0.6	0.55	1.7	4	-1 / -2.5	
$V_{th,sat}$		0.2	0.45V	-0.35	0.25	0.45				
I_{on} (μ A/ μ m)	~300	~50	394	355	81	78	400	400	200	
I_{off} (nA. μ m)	~1.5	~6000	300	890	2800	3400	600	10000	0.1	
S (mV/dec)	~10	7.2	5.5	12.2	7.5	11.8	4.5	24	2	

Tableau 6-2: comparaison des performances obtenues avec des dispositifs I-MOS

De ces deux tableaux, il apparaît que les réalisations présentant les meilleures performances sont celles qui utilisent l'intégration par double espaceur. En effet, cela permet de définir de façon précise la longueur intrinsèque et la longueur de grille, et d'obtenir de petites dimensions. Il est également intéressant de remarquer que la tension d'avalanche V_{DS} réduit avec les dimensions. De manière similaire à la figure $V_B(L_i)$ qui a été utilisée pour un dispositif unidimensionnel, la **figure 6-7** représente la tension de claquage en fonction de la longueur de grille.



Sur cette figure sont également portées les résultats du modèle 1D développé au chapitre 3. Il s'agit des tensions de claquage pour des diodes p-i-n 1D ayant des zones n et p identiquement dopées, pour trois valeurs de dopage: $2 \cdot 10^{18} \text{cm}^{-3}$, $5 \cdot 10^{18} \text{cm}^{-3}$, et $1 \cdot 10^{19} \text{cm}^{-3}$. Ce graphique permet de faire quelques observations:

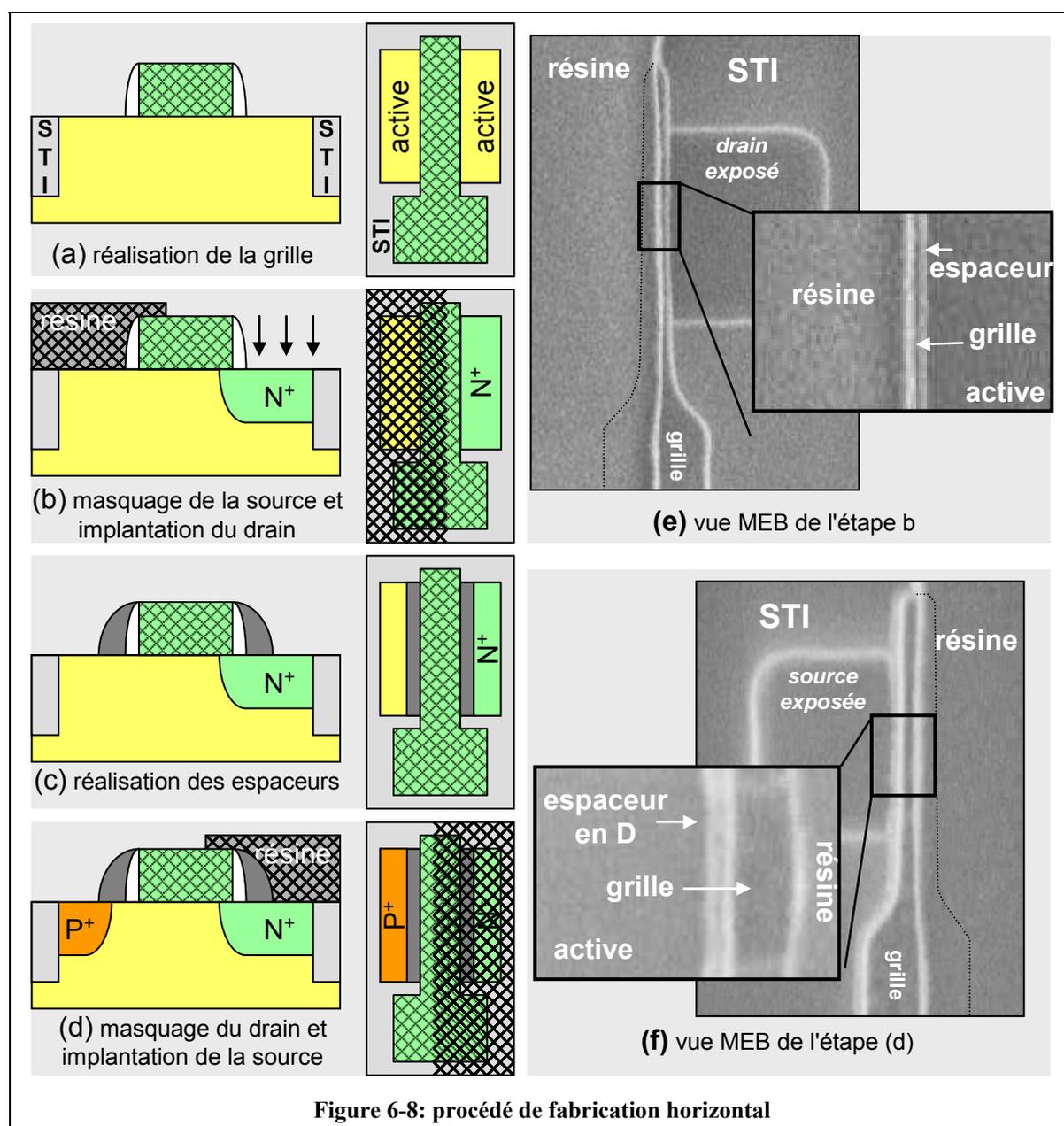
- les points de littérature semblent être sur la même courbe, bien que différentes intégrations aient été utilisées. Cela implique que seule la taille de grille L_G a de l'importance lorsque l'on cherche une tension d'avalanche faible.
- l'intégration de germanium dans les source/drain surélevés [Toh'05] n'a pas d'impact visible sur la tension d'avalanche.
- la valeur de dopage servant de paramètre dans le modèle 1D représente la plus faible densité porteurs entre zone n et zone p (chapitre 4). Ces zones sont ici le drain et le "canal" sous la grille. On attend donc une valeur de dopage théorique correspondant à la densité de porteurs sous la grille (en conduction). Or, la valeur qui reflète le mieux les tensions d'avalanche des I-MOS est de $5 \cdot 10^{18} \text{cm}^{-3}$, ce qui est tout à fait cohérent avec la densité de porteurs sous la grille.

Cela montre que l'approximation qui consiste à considérer l'I-MOS comme une diode 1D dont on module la longueur intrinsèque est raisonnable.

6.2. Intégration horizontale

6.2.1. Procédé de fabrication

L'analyse de l'état de l'art technologique montre qu'il est avantageux de réduire autant que possible la longueur de grille pour diminuer la tension de claquage. Or, ce sont les technologies auto-alignées qui ont le plus gros potentiel de réduction des dimensions (utilisation d'espaces). En outre, l'intégration de dispositifs de type TFET a montré qu'il est



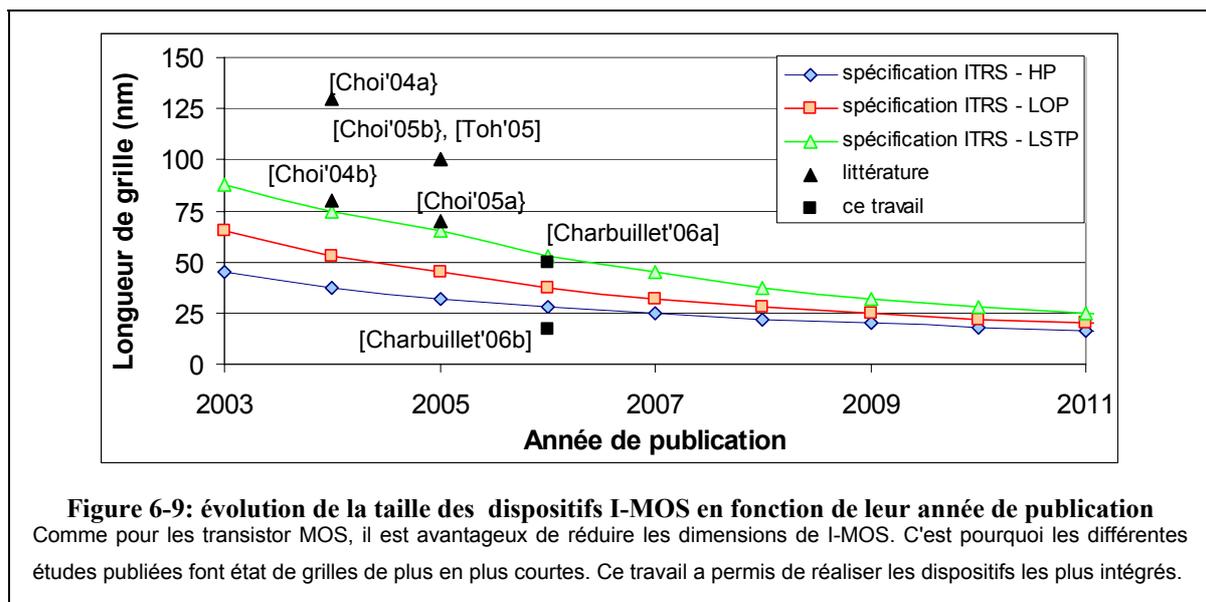
possible d'adapter simplement un procédé de fabrication standard pour créer l'asymétrie source/drain. Nous avons donc choisi d'utiliser une approche similaire, en ajoutant des espaceurs pour définir la taille de la zone intrinsèque (**figure 6-8**).

Cette approche présente plusieurs avantages:

- l'intégration est très proche du procédé de fabrication conventionnel. Il n'y a que très peu d'étapes spécifiques à développer, ni de masque de photolithographie supplémentaire à créer.
- on peut utiliser une technologie mature, stable. Dans notre cas, il s'agit de CMOS090.
- cette technologie donne accès à des dimensions réduites, autant pour la taille de grille que pour la taille de la zone intrinsèque.

La difficulté majeure est d'aligner successivement deux photolithographies sur une grille de petite dimension.

6.2.1.1. Dimensions

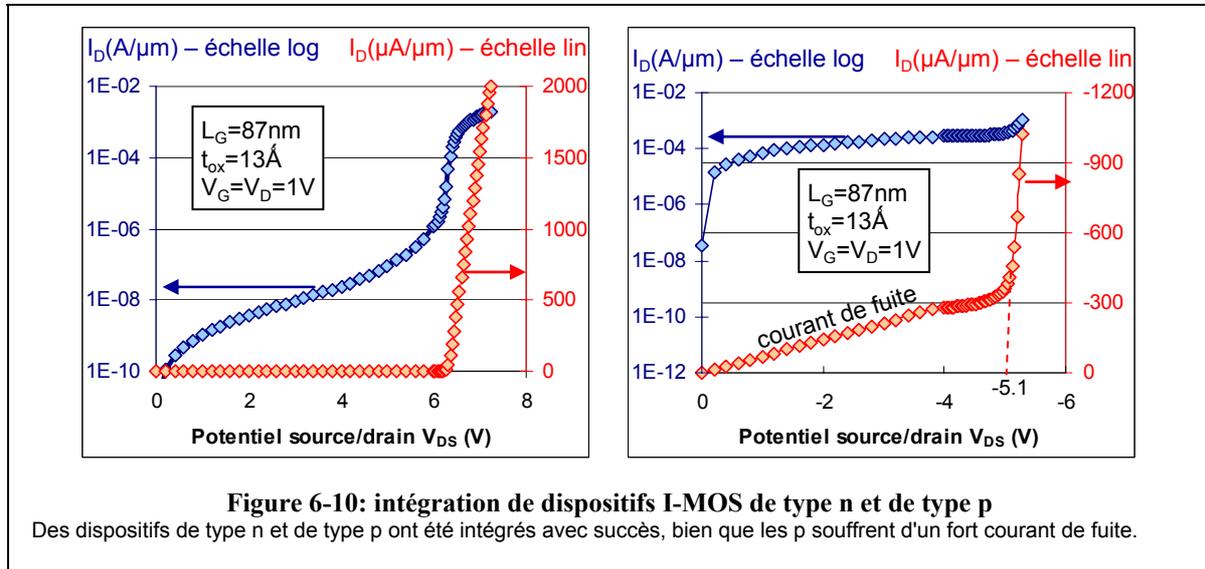


Comme pour le transistor MOS, il est avantageux de réduire les dimensions des dispositifs I-MOS. C'est pourquoi les études portent sur des dimensions de plus en plus réduites, comme en témoigne la **figure 6-9**. Les dispositifs les plus courts que nous avons réalisés avec ce procédé de fabrication ont une longueur de grille de **17nm** [Charbuillet'06a,Charbuillet'06b], ce qui en fait les plus compacts lors de leur publication.

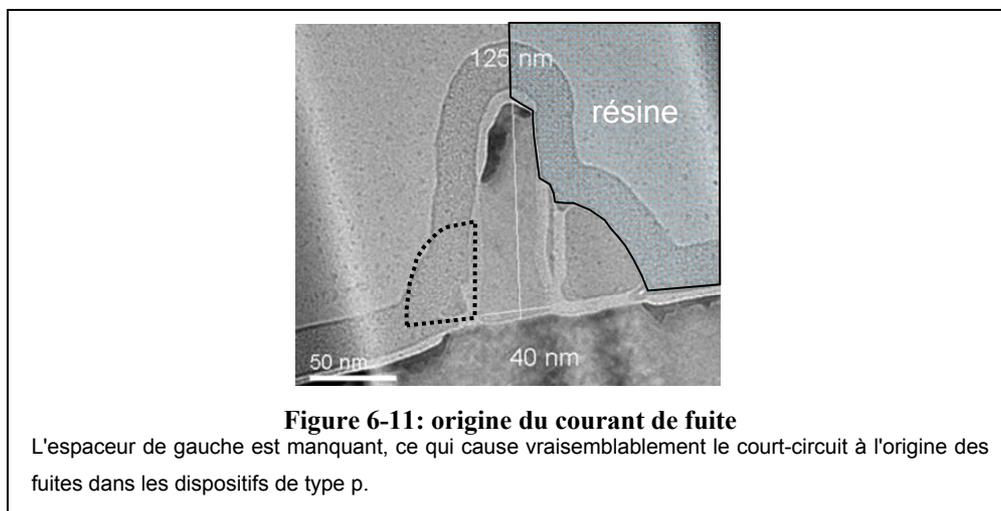
6.2.2. Résultats électriques – observations générales

6.2.2.1. Fuites

Ce procédé ne permet pas d'intégrer simplement des dispositifs de type n et de type p sur la même plaque. Par contre, il est possible d'intégrer l'un ou l'autre en adaptant les implantations.

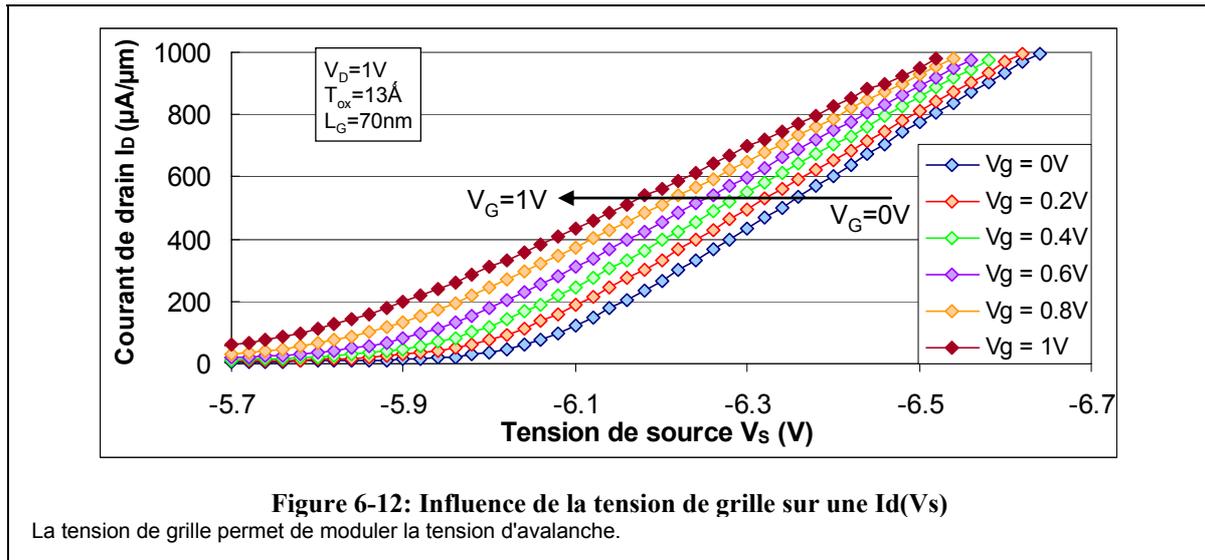


La **figure 6-10** montre que le courant de drain est entaché d'un courant de fuite, bien plus dommageable pour les dispositifs de type p que les dispositifs de type n. Il circule entre la source et la grille, et est dû à la perte de l'espaceur côté source. En effet, au cours du procédé de fabrication, le drain est protégé par de la résine pour pouvoir implanter la source. A cette étape, les espaceurs en D sont réalisés. Lors du nettoyage de la résine, l'espaceur côté source est attaqué alors que celui côté drain est protégé. Les photos TEM montrent qu'un espaceur a effectivement disparu (figure 6-11). Ceci est d'autant plus gênant que les dispositifs étudiés sont en batterie à source commune et grille commune. Le court-circuit d'un seul dispositif rend donc inutilisables tous des dispositifs la batterie.



6.2.2.2. Polarisation

La polarisation classique de l'I-MOS (de type n) consiste à polariser la source négativement, le drain et la grille entre 0 et V_{DD} . Or, le potentiel de grille de nos réalisations est perturbé par le potentiel de source à cause du courant de fuite entre ces électrodes. Cela limite donc les caractérisations possibles. Notamment, l'influence de la grille disparaît pour des dispositifs courts. Le dispositif le plus court sur lequel on mesure une influence de la tension de grille dispose d'une grille de 55nm (**figure 6-12**).



De plus, des courants de fuite entre grille et drain apparaissent. Compte tenu de ces difficultés, une polarisation particulière est appliquée: la grille et le drain sont à la masse pour éviter les fuites entre eux, et la source est polarisée négativement pour générer l'avalanche. Cela permet d'étudier la tension de claquage et les courants. D'autre part, l'étude a été focalisée sur l'étude des dispositifs de type n, ceux de type p étant d'avantage dégradés par les courants de fuite.

6.2.2.3. Caractéristiques $I_D(V_{DS})$

Le schéma de polarisation expliqué au paragraphe 6.2.2.2 donne accès à des caractéristiques $I_D(V_{DS})$ (**figure 6-13**). Quelle que soit la géométrie du dispositif, trois régions de fonctionnement peuvent être distinguées:

- ① à faible tension V_{DS} , le courant de drain est faible. C'est le courant de fuite de la diode.
- ② au-delà de la tension d'avalanche, la caractéristique est linéaire.
- ③ enfin, on trouve une ou plusieurs discontinuités. Celles-ci apparaissent au-delà d'un certain courant limite, et non au-delà d'une certaine tension V_{DS} .

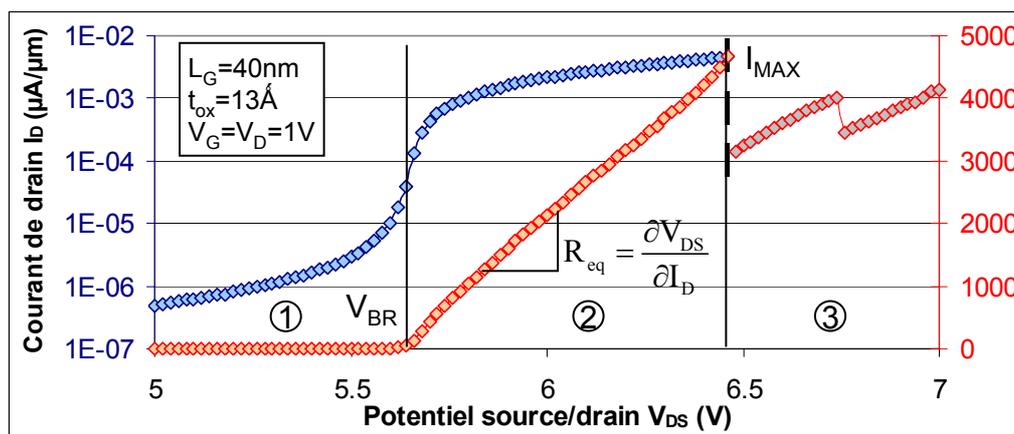


Figure 6-13: caractéristique typique $I_D(V_{DS})$

La caractéristique typique $I_D(V_{DS})$ typique est linéaire au-dessus de la tension d'avalanche V_{BR} . Cette linéarité est caractérisée par sa pente assimilable à une résistance.

Les grandeurs caractéristiques qui déterminent ces zones sont:

- la tension d'avalanche V_{br}
- la résistance équivalente de la zone linéaire $R_{eq} = \frac{\partial V_{DS}}{\partial I_D}$
- le courant maximal débité I_{max}

Ces trois grandeurs dépendent de la géométrie des dispositifs (**figure 6-14**).

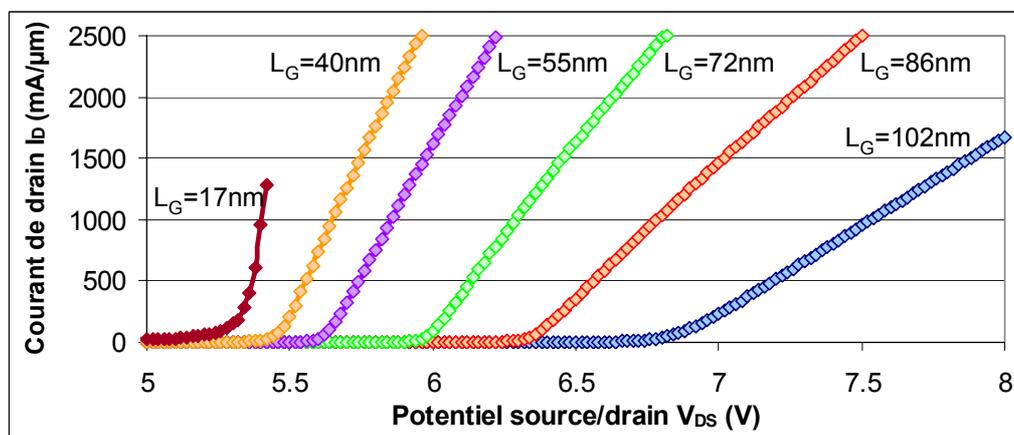


Figure 6-14: caractéristiques $I_D(V_{DS})$ en fonction de la longueur de grille

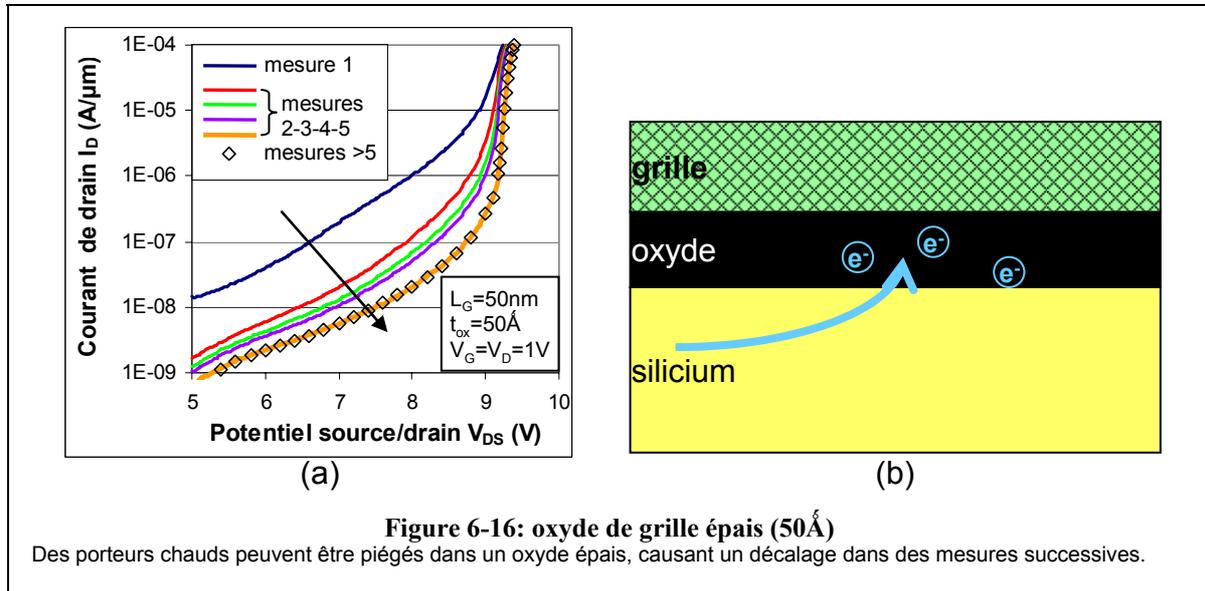
Lorsque la longueur de grille diminue, la tension d'avalanche et la résistance équivalente diminuent également.

6.2.3. Résultats électriques – impact de la géométrie des dispositifs

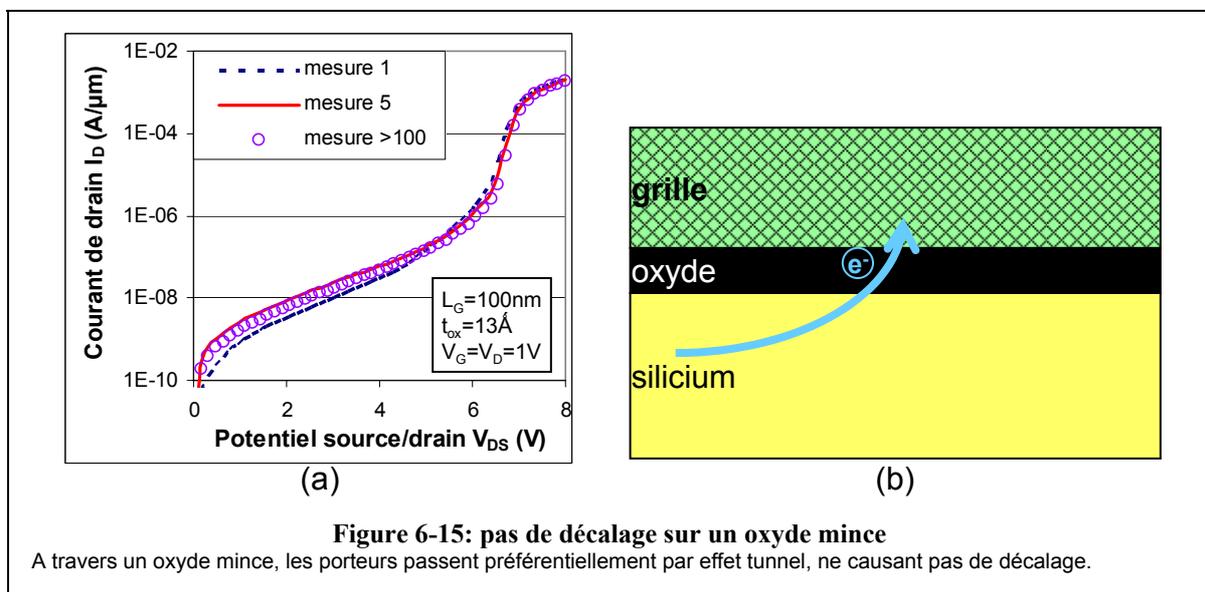
6.2.3.1. Effet de l'épaisseur de l'oxyde de grille

Un oxyde épais (50Å) et un oxyde mince (13Å) ont été intégrés. Ce dernier est le plus mince jamais intégré, l'épaisseur la plus faible de la littérature étant 30Å (**tableau 6-2**).

Pour un oxyde épais, les caractéristiques se décalent au cours des différents cycles de mesures (**figure 6-16**). Ceci a déjà été observé [Gopalakrishnan'05b] et est attribué à la présence de porteurs chauds dans l'oxyde. Cet effet semble se stabiliser après cinq cycles.



En revanche, ce décalage n'est pas observé sur un oxyde de 13 Å (**figure 6-15**). Ceci avait déjà été supposé [Gopalakrishnan'05b] et attribué au fait que sur un oxyde mince, un porteur ayant assez d'énergie pour se loger dans l'oxyde passera préférentiellement par effet tunnel à travers l'oxyde [Schuegraf'94].



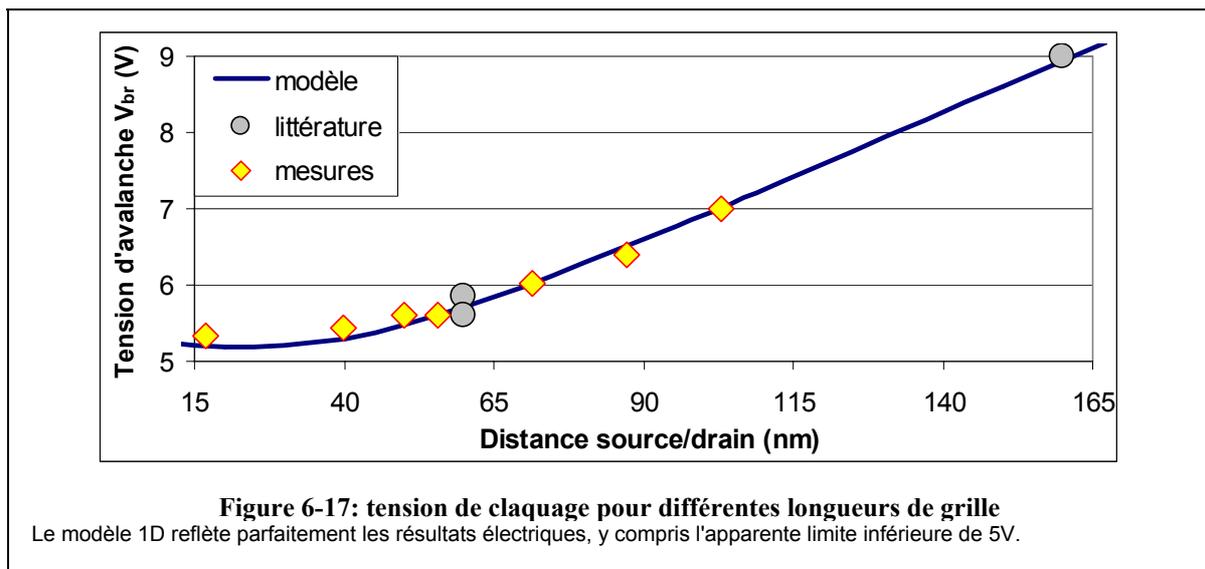
6.2.3.2. Dépendance de la tension de claquage avec la longueur de grille

Il a été souligné au paragraphe 6.1.5 que la tension d'avalanche dépend de la longueur de grille. C'est même le principal levier pour baisser la tension d'avalanche. En réalisant une coupe dans l'axe source/drain, l'I-MOS se présente comme une diode p-i-n 1D. Il est donc

intéressant de comparer le modèle de tension d'avalanche développé au chapitre 4 avec les résultats de dispositifs 2D. Cette comparaison est faite sur la **figure 6-17**. Elle représente la tension d'avalanche en fonction de la distance électrique. La longueur électrique (introduite au chapitre 4) est la distance entre les zones dopées. Cette notion permet de comparer des dispositifs 1D avec des dispositifs 2D.

En plus de nos propres dispositifs I-MOS, différents résultats de la littérature sont représentés. La modélisation 1D comporte un unique paramètre d'ajustement, la valeur de dopage des zones dopées. Cette valeur est prise identique à celle du paragraphe 6.1.5, qui compare les intégrations dans la littérature.

La modélisation 1D reflète parfaitement les résultats électriques, y compris le fait que la tension d'avalanche ne semble pas devoir descendre sous 5V avec des dispositifs en silicium. Cette limite a pu être soulignée parce que nos dispositifs sont plus courts que ceux jamais réalisés: 17nm [Charbuillet'06a,Charbuillet'06b] contre 60nm précédemment.



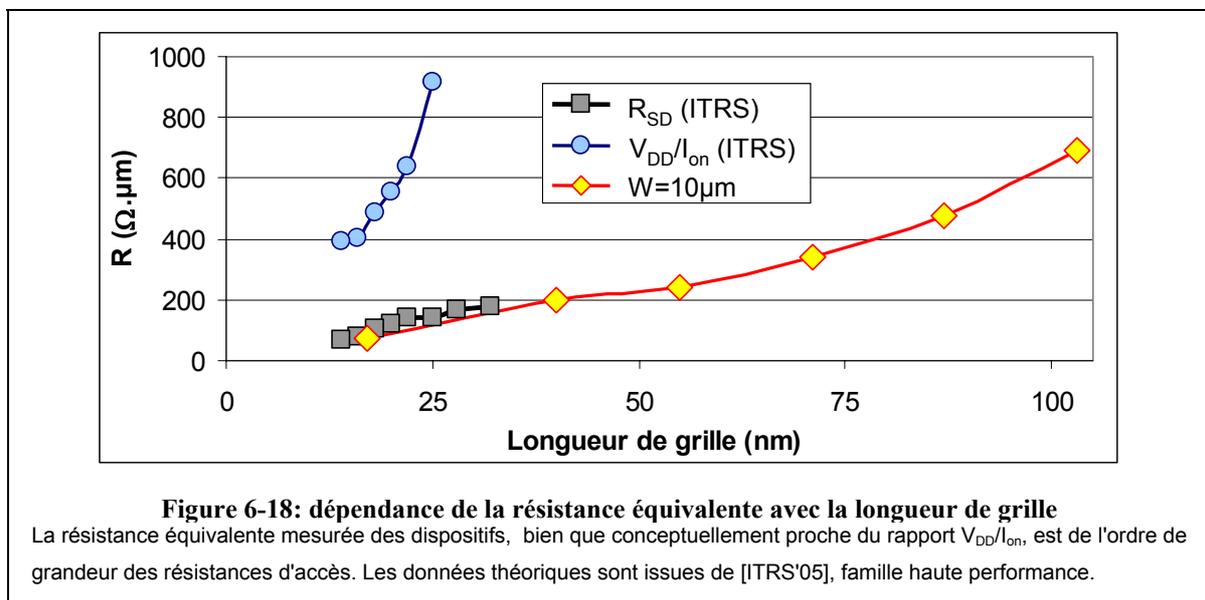
6.2.3.3. Dépendance de la résistance équivalente avec la longueur de grille

Il a été souligné que la caractéristique $I_D(V_{DS})$ est linéaire au-delà de l'avalanche. Par conséquent, le dispositif I-MOS peut être considéré comme une résistance équivalente au-delà du seuil d'ionisation par impact.

Cette résistance peut être séparée en deux composantes: la résistance du dispositif, qui varie avec la longueur, et une résistance parasite constante. Cette dernière a les mêmes origines que celles du transistor MOS conventionnel: contact, résistance du siliciure, résistance de couche des zones source/drain.

Ces valeurs mesurées sont comparées aux spécifications ITRS'05 pour les transistors hautes performances sur la **figure 6-18**. Pour ces spécifications, la résistance parasite est notée R_{SD} , et la résistance totale équivalente peut être associée au rapport V_{DD}/I_{on} .

La résistance équivalente mesurée est très proche des spécifications ITRS de résistance d'accès, ce qui signifie que la résistance du dispositif est faible devant la résistance d'accès. Elle n'est cependant pas négligeable, sans quoi la résistance équivalente ne varierait pas avec la longueur de grille. Dans un transistor MOS conventionnel, la résistance parasite est de l'ordre de 20% de la résistance du dispositif (du canal dans ce cas). Dans un I-MOS, elles sont du même ordre de grandeur. Cela signifie que les performances de nos dispositifs sont d'avantage limitées par les parasites que par le dispositif lui-même.



De fait, les dispositifs I-MOS ne montrent pas de saturation, contrairement au transistor MOS.

Cela est dû au fait que le phénomène physique à l'origine du transport n'est pas le même:

- dans un transistor MOS, le courant est le déplacement d'une charge d'inversion. La saturation apparaît aux valeurs élevées de V_D , lorsque le pincement apparaît.
- dans un I-MOS, c'est la génération de paires électron-trou qui est à l'origine du courant. La modélisation de la diode p-i-n a montré que celui-ci pourrait augmenter indéfiniment si la résistance du dispositif et les résistances d'accès étaient nulles. Le courant n'a aucune raison de saturer.

Le dernier point de la courbe expérimentale, correspondant à un dispositif de 17nm, est sensiblement désaligné du reste de la courbe. C'est parce que ce dispositif présente une largeur différente de celle des autres, $0.38\mu m$ au lieu de $10\mu m$.

6.2.3.4. Dépendance de la résistance équivalente avec la largeur du dispositif

La résistance série peut être séparée en deux composantes: la résistance des zones dopées, ou résistance de film, et la résistance du contact. La première dépend essentiellement de la géométrie du dispositif selon une loi:

$$R_{\text{film}} = \rho \cdot \frac{L}{W \cdot t} \quad (4.1)$$

où L représente la distance contact/dispositif, W la largeur du dispositif et t l'épaisseur utile du film.

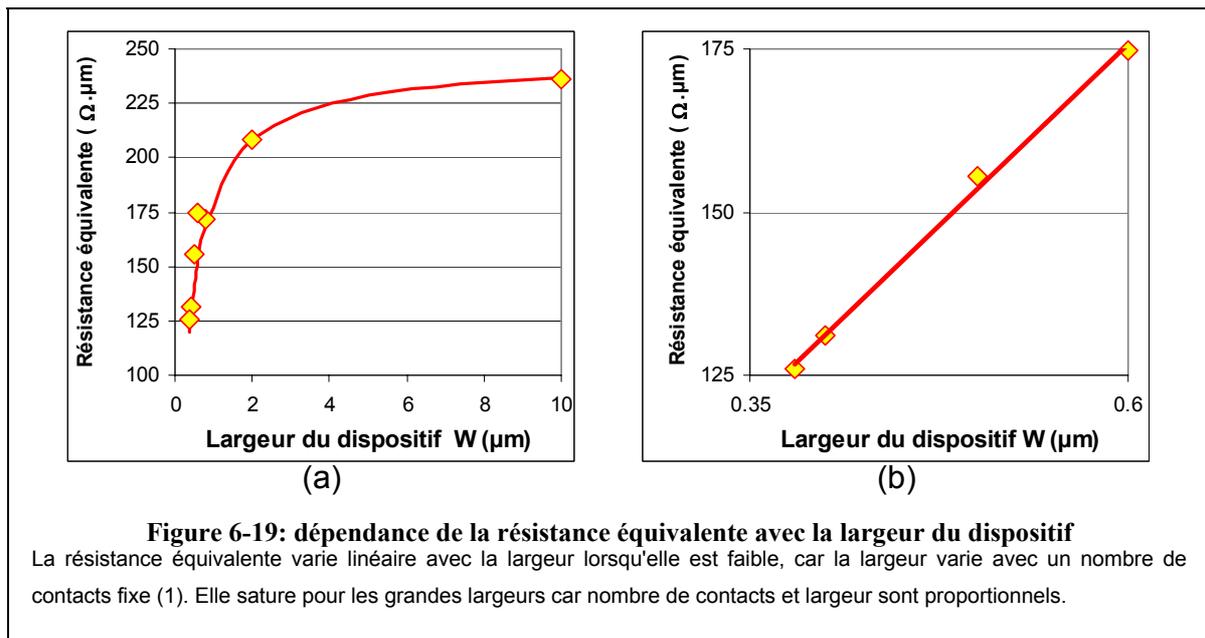
D'autre part, la résistance de contact est d'autant plus faible que les contacts sont nombreux:

$$R_{\text{contacts}} = \frac{R_c}{N_{\text{contacts}}} \quad (4.2)$$

où R_c est la résistance d'un contact, et N_{contacts} le nombre de contacts. Ces deux résistances sont en série et la résistance équivalente est normalisée par la largeur:

$$R_{\text{eq}} = W \cdot (R_{\text{film}} + R_{\text{contacts}}) = W \cdot \left(\rho \cdot \frac{L}{W \cdot t} + \frac{R_c}{N_{\text{contacts}}} \right) \quad (4.3)$$

La courbe qui porte la résistance équivalente en fonction de la largeur du dispositif (**figure 6-19**) présente deux asymptotes: la première est horizontale à W élevé, la seconde est linéaire à W faible. Ceci correspond à deux situations opposées: dans le premier cas, c'est la résistance de film qui est prépondérante, alors que dans le second c'est la résistance de contact qui gouverne le niveau de courant.



En effet, lorsque la largeur est élevée, le nombre de contacts est proportionnel à la largeur du dispositif:

$$N_{\text{contacts}} = \alpha \cdot W \quad (4.4)$$

Dans ce cas, l'expression de la résistance équivalente (4.3) devient:

$$R_{\text{eq}} = W \cdot \left(\rho \cdot \frac{L}{W \cdot t} + \frac{R_c}{\alpha \cdot W} \right) = \rho \cdot \frac{L}{t} + \frac{R_c}{\alpha} \quad (4.5)$$

Cette expression est bien indépendante de W .

En revanche, pour des dispositifs étroits, on ne peut intégrer qu'un seul contact, bien que la largeur du dispositif puisse varier:

$$N_{\text{contacts}} = 1 \quad (4.6)$$

Dans ce cas, l'expression de la résistance équivalente (4.3) devient:

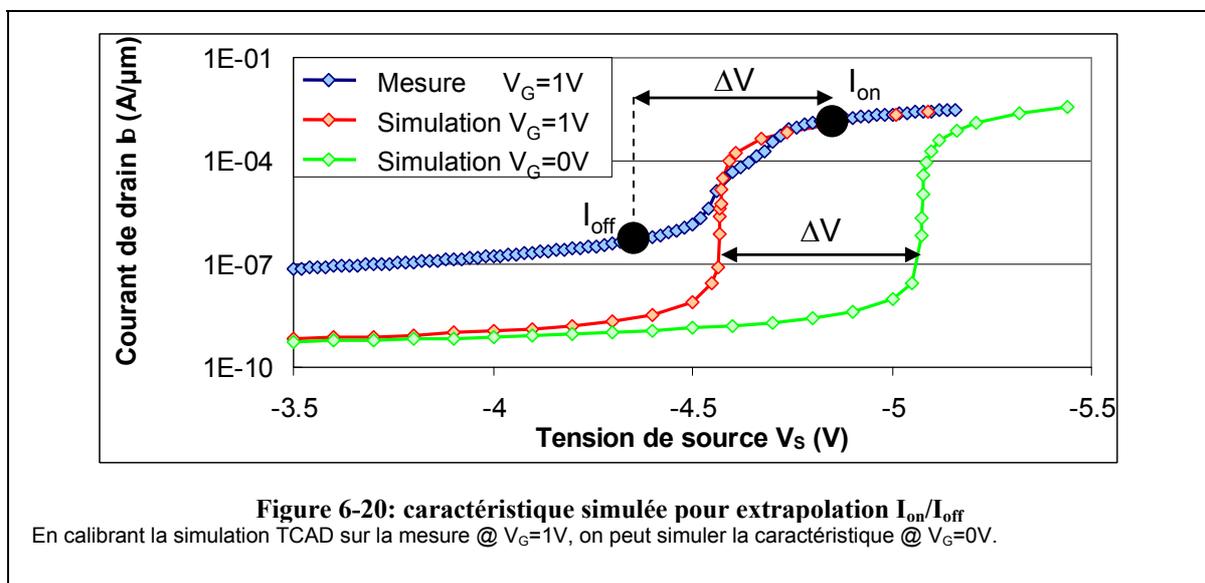
$$R_{\text{eq}} = W \cdot \left(\rho \cdot \frac{L}{W \cdot t} + \frac{R_c}{1} \right) = \rho \cdot \frac{L}{t} + W \cdot R_c \quad (4.7)$$

La relation est trouvée est bien linéaire en fonction de W . L'application numérique donne pour la résistance de contact $R_c=223\Omega$. Cette valeur englobe les deux contacts (un côté source, l'autre côté drain), soit une résistance par contact de 110Ω environ.

6.2.4. Extraction de la figure de mérite $I_{\text{on}}/I_{\text{off}}$

La figure de mérite $I_{\text{on}}/I_{\text{off}}$ est un moyen rapide d'évaluer les performances d'un dispositif. Si la mesure de I_{on} de nos dispositifs est accessible, il est plus délicat de mesurer I_{off} à cause des courants de fuite. I_{off} est donc extrait à l'aide de simulations TCAD.

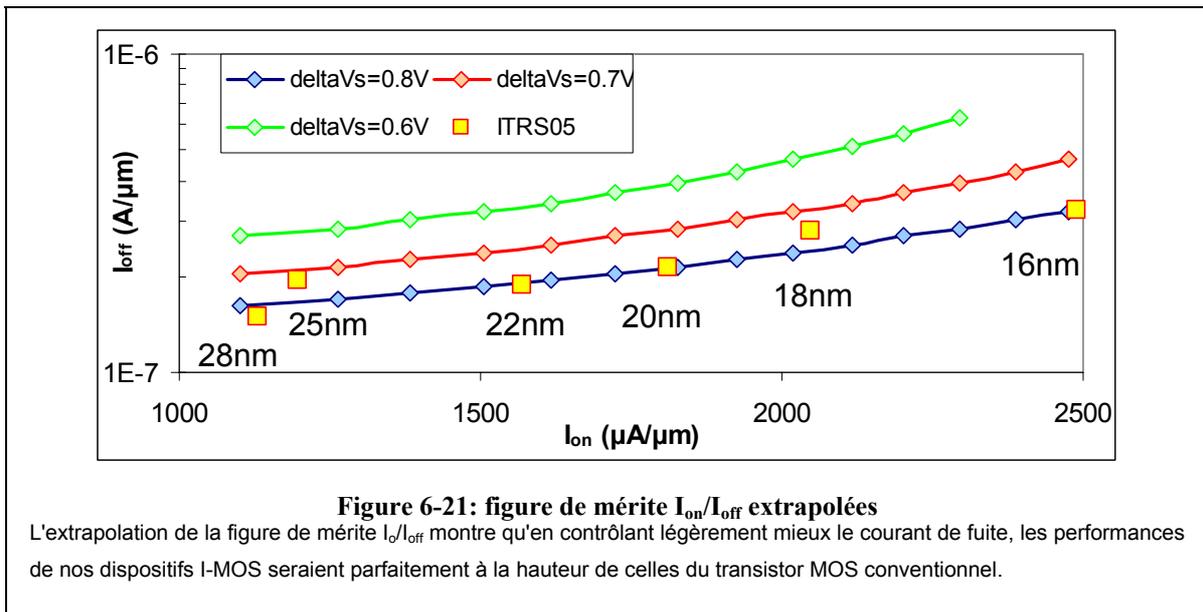
Dans un premier temps, la mesure $I_D(V_S)@V_G=V_D=V_{DD}$ est utilisée à des fins de calibration.



Puis, la courbe $I_D(V_S)@V_G=0, V_D=V_{DD}$ est simulée. La différence de tension de grille se traduit par un décalage de la courbe de ΔV (**figure 6-20**).

Par conséquent, tout couple de points de la courbe mesurée aux tensions $V_S, V_S+\Delta V$ peut être assimilé au couple I_{on}/I_{off} à la tension de source V_S .

La figure de mérite I_{on}/I_{off} ainsi extraite est représentée sur la **figure 6-21**. Avec le courant de fuite de nos dispositifs, une valeur de ΔV de 0.8V semble nécessaire pour atteindre les spécifications ITRS haute performance. Une autre solution consiste à améliorer le courant de fuite, pour le rendre comparable à celui des I-MOS sur oxyde épais.



6.3. Intégration verticale

La difficulté majeure de l'intégration de dispositifs I-MOS reste l'asymétrie n/p, rendant impossible l'auto-alignement. La solution proposée dans la partie précédente de discriminer les zones n et p par photolithographie devient de plus en plus difficile à mesure que les longueurs de grille diminuent.

De plus, la modélisation du dispositif prédit que les tensions d'avalanche cessent de diminuer avec la longueur de grille en dessous de 30 nm (**figure 6-22**). Cet effet s'avère extrêmement limitant pour la réduction des dimensions, pourtant incontournable.

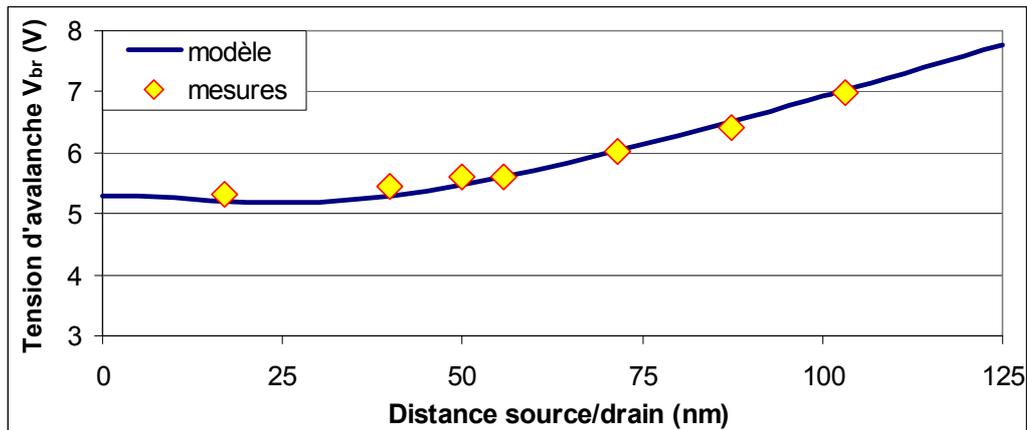
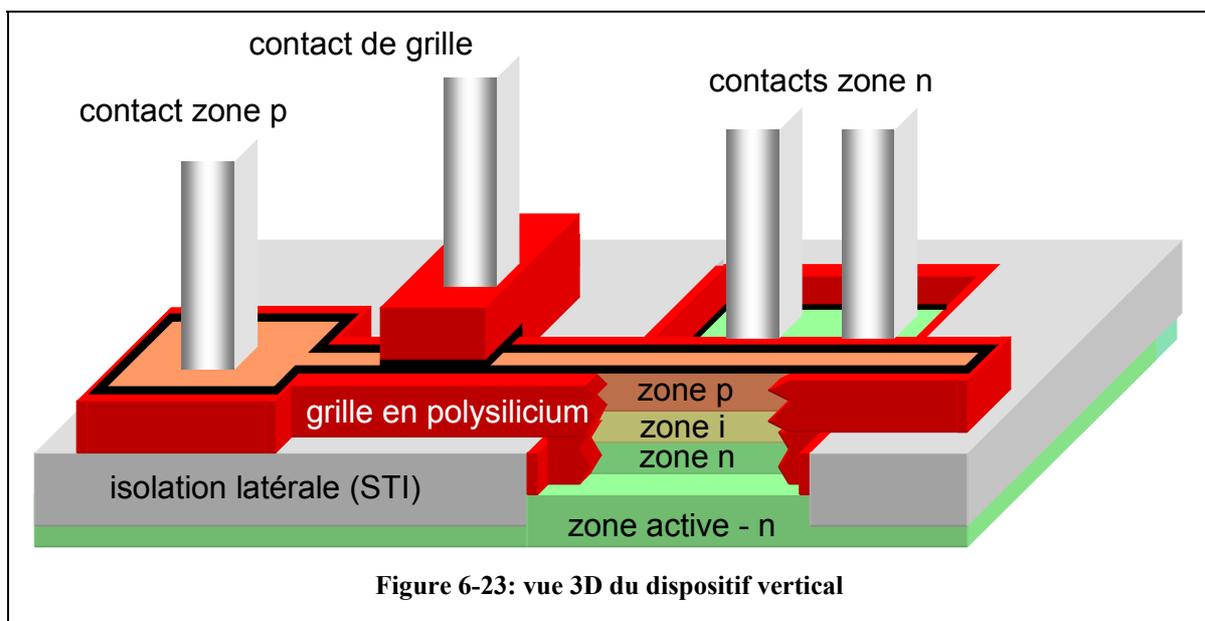


Figure 6-22: augmentation de la tension de claquage pour les dispositifs les plus courts

La tension d'avalanche sature aux environs de 5V pour les dispositifs les plus courts.

C'est pourquoi nous proposons une intégration verticale: les zones i et p sont épitaxiées, ce qui permet un contrôle excellent de ces dimensions. L'encombrement est réduit, car il ne dépend plus des dimensions électriques.

La zone n sera formée par la zone active, les zones n et p épitaxiées. La grille sera formée d'un espaceur en polysilicium, qui sera contactée sur un plot spécifique. Celui-ci nécessite un masque supplémentaire par rapport à une intégration CMOS conventionnelle.



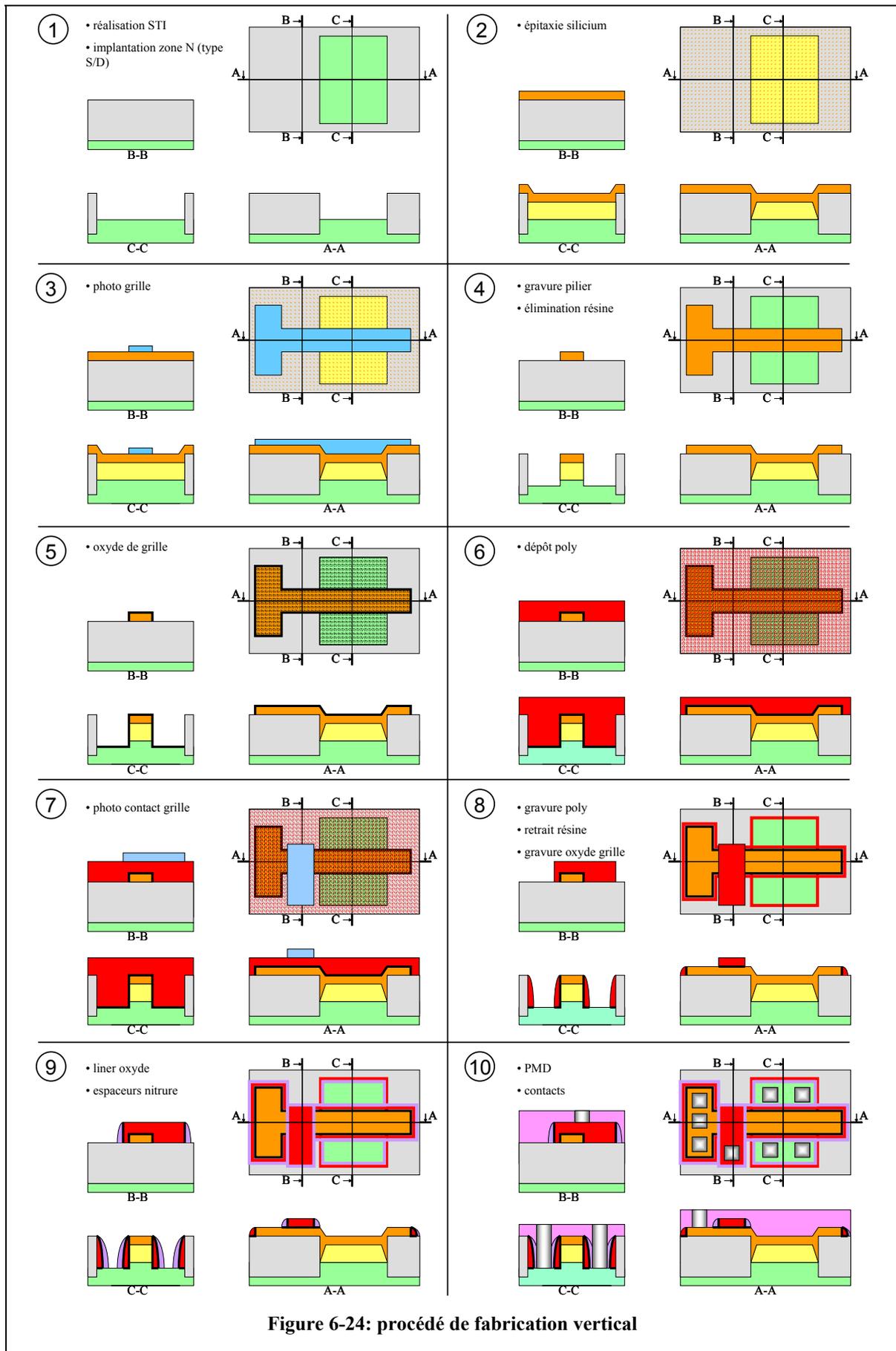


Figure 6-24: procédé de fabrication vertical

6.3.1. Procédé de fabrication

Chacune de ces étapes peut être considérée de façon relativement indépendante comme un module technologique élémentaire, dont l'assemblage mène à l'intégration du dispositif. Les plus importantes d'entre elles vont à présent être détaillées.

6.3.1.1. Epitaxie

Elle présente de nombreux avantages:

- excellent contrôle des dimensions
- possibilité d'intégrer des couches alliages $\text{Si}_x\text{Ge}_y\text{C}_{1-x-y}$.

Ce dernier point a deux possibilités d'utilisation:

- les alliages SiGe sont plus favorables à l'ionisation par impact que le silicium [Sze'66]
- la vitesse de diffusion des dopants variant selon l'alliage, ces couches peuvent servir de barrière de diffusion.

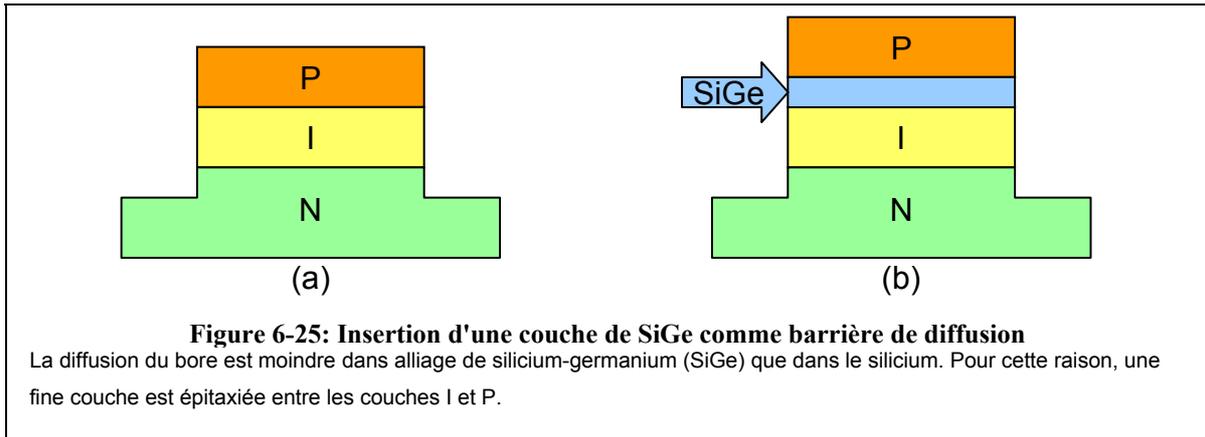
Incorporation de SiGe pour faciliter l'ionisation par impact

L'ionisation par impact est plus efficace dans le germanium que dans le silicium. C'est pourquoi il serait intéressant de l'intégrer pour diminuer les tensions d'avalanche. Le premier essai présenté dans la littérature [Toh'05] n'est cependant pas convaincant: la tension d'avalanche est alignées avec celles des dispositifs silicium (**figure 6-7**). En effet, la fraction de germanium que l'on peut incorporer sans désaccord de maille est relativement faible, ce qui limite l'intérêt de cette approche.

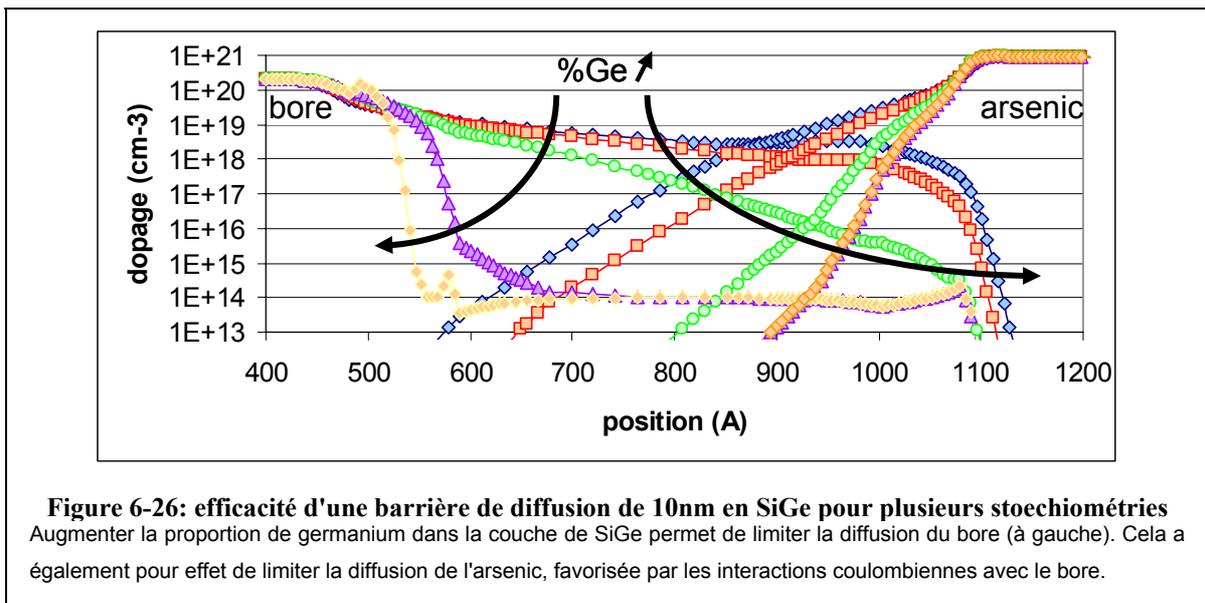
Incorporation de SiGeC en tant que barrière de diffusion

Si la diffusion de l'arsenic est moins importante dans le silicium que dans les alliages SiGeC, celui-ci est une barrière de diffusion efficace pour le bore [Pakfar'03]. Pour déterminer l'épaisseur et la quantité de dopants à incorporer, la diffusion des dopants a été simulée par simulation TCAD.

La structure initiale est un empilement idéal de couches de silicium dopé et de couches de SiGeC (**figure 6-25**). Cette idéalité est une approximation raisonnable du profil obtenu dans un bâti d'épitaxie. Puis, l'ensemble des opérations thermiques (= budget thermique) est simulé, afin d'estimer la diffusion des dopants.



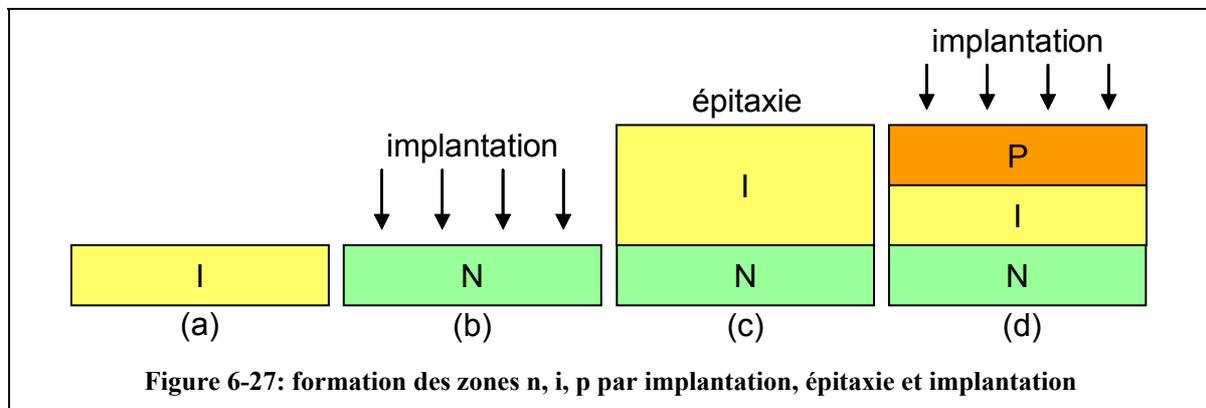
La **figure 6-26** montre les profils de bore (à gauche) et d'arsenic (à droite) pour différentes concentrations en germanium de la couche barrière. Pour une concentration nulle (pas de couche barrière), bore et arsenic vont diffuser dans la zone intrinsèque, qui sera alors dopée (courbe 1 en bleu). Dans ce cas, le dopage minimal sera de quelque 10^{18}cm^{-3} , ce qui est très éloigné de la valeur théorique d'un silicium intrinsèque qui ne devrait pas dépasser 10^{17}cm^{-3} .



Augmenter la fraction de germanium permet de limiter la diffusion du bore: les courbes de droite revêtent un caractère de plus en plus abrupt à mesure que la concentration augmente. Plus étonnant, la barrière semble avoir un effet bénéfique également sur la diffusion de l'arsenic, bien que celui-ci ne soit pas en contact avec la barrière. C'est le résultat de la limitation de la diffusion de l'arsenic activée par interactions coulombiennes. Les atomes de bore ionisés présentant une charge négative, ils favorisent la diffusion des atomes d'arsenic ionisés présentant une charge négative. La barrière de diffusion limitant la diffusion du bore, elle réduit par là même la diffusion de l'arsenic.

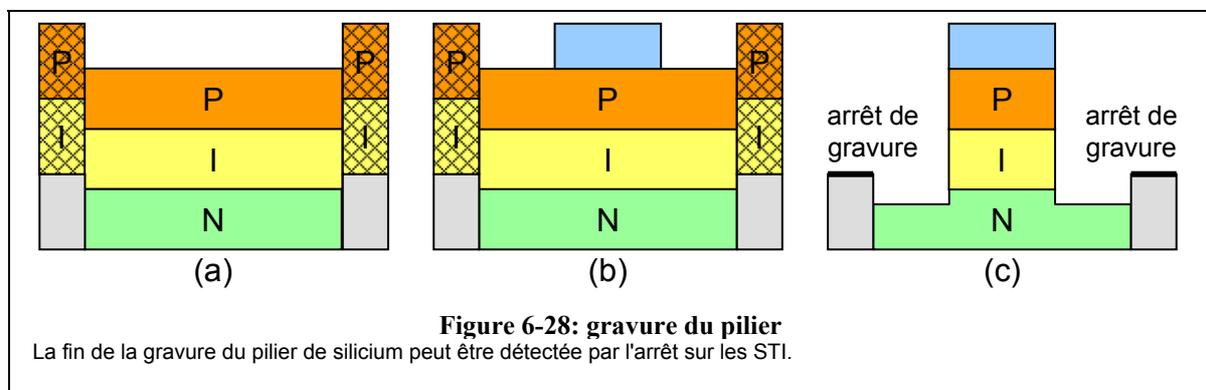
En se basant sur ces simulations TCAD, plusieurs barrières semblent efficaces: 10nm @ 30% de germanium ou 5nm @ 40% de germanium. Cependant, une couche de 10nm @ 30% présente un désaccord de maille sur une épaisseur trop importante: la couche épitaxiée va relaxer et générer des défauts étendus (dislocations par exemple). Par contre, la couche 5nm @ 40% de germanium devrait rester stable, bien que contrainte.

La grille étant en bas, le dopage de la partie basse détermine le type du transistor: dopage de type p (respectivement n) pour un dispositif de type p (respectivement n). Ce système d'épitaxie présente donc un inconvénient: il ne permet pas de réaliser de dispositifs complémentaires. Afin de contourner ce problème, nous avons décidé de ne pas doper les zones n et p pendant l'épitaxie, mais d'utiliser la séquence implantation/épitaxie/implantation.



6.3.1.2. Gravure du pilier

Cette opération doit permettre de graver la zone de silicium épitaxiée jusqu'à la zone dopée inférieure. Il s'agit donc de graver du silicium, et de s'arrêter dans du silicium dopé différemment. La sélectivité en fonction du dopage n'étant pas suffisante, la solution la plus simple est donc de réaliser une **gravure à temps fixe**.



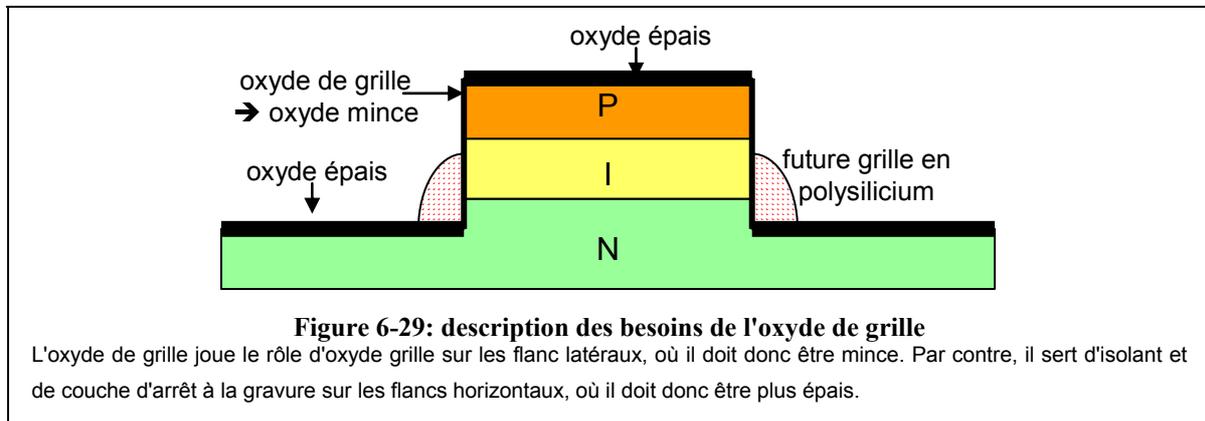
Il est cependant possible de rendre plus fiable l'arrêt de la gravure. La couche de silicium intrinsèque est épitaxiée de manière non sélective: du poly silicium monocristallin croît sur

les zones en silicium monocristallin (zone active), alors que du silicium poly-cristallin croît les zones d'oxyde (STI). Le polysilicium est environ 20% plus épais que le mono-silicium, à cause de la différence de compacité. Il est donc possible d'utiliser la zone de STI comme couche d'arrêt.

6.3.1.3. Oxyde de grille

Dans cette intégration, l'oxyde de grille a deux utilités:

- sur les flancs du pilier, il sert d'oxyde de grille à proprement parler. Il doit donc répondre aux exigences habituelles: faible épaisseur, bonne isolation électrique.
- sur le dessus du pilier et sur la zone active, il servira de couche d'arrêt à une gravure plasma. Par conséquent, l'oxyde doit être mince sur les flancs verticaux et épais sur les flancs horizontaux.



Cette problématique a déjà été rencontrée dans la résolution de transistors verticaux. Une des solutions employée consiste à réaliser une implantation avant l'oxydation [Josse'00]. En effet, l'implantation conduit à l'amorphisation des surfaces horizontales. Or, le silicium amorphisé présente une vitesse d'oxydation plus importante que le silicium cristallin, à cause de la plus grande surface développée. Cette méthode permet donc d'obtenir un oxyde plus épais sur les surfaces horizontales que sur les surfaces verticales.

Nous avons testé plusieurs implantations pour déterminer la plus efficace (**figure 6-30**).

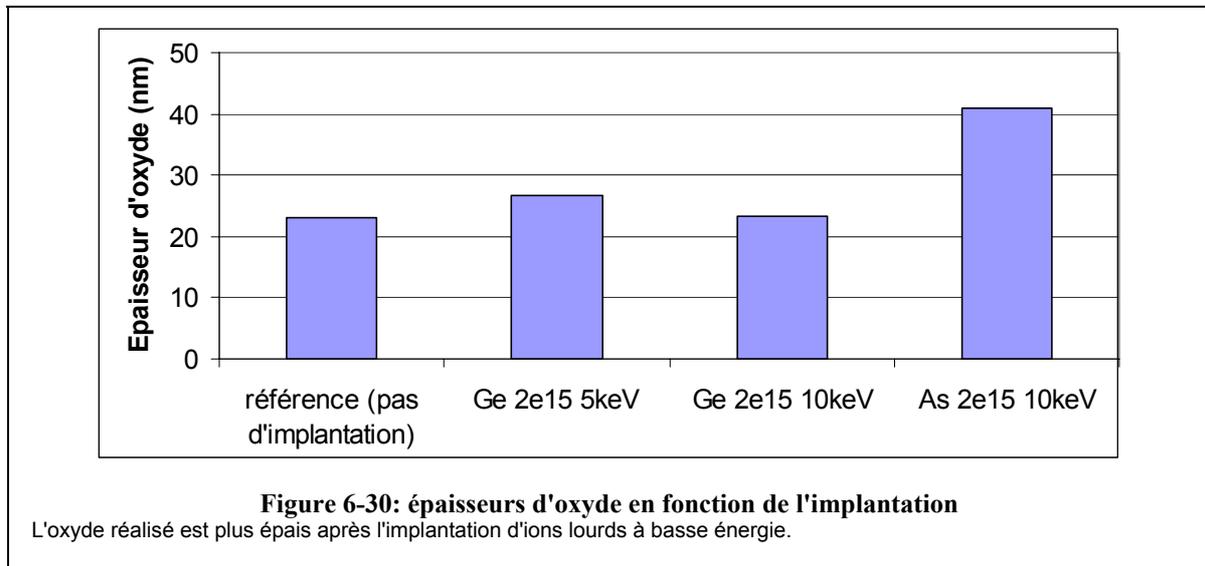
Il apparaît que le plus efficace correspond à:

- une forte dose
- une faible énergie
- des ions lourds

C'est ce qui correspond à une meilleure amorphisation.

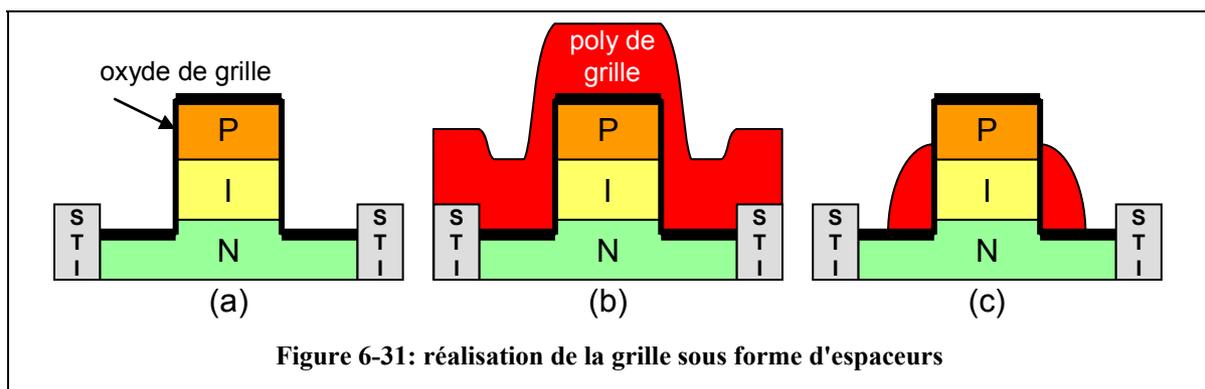
En outre, l'implantation de bore ou d'arsenic présente un inconvénient: cela va perturber le dopage des zones actives dans une intégration CMOS. Par conséquent, l'implantation de germanium, électriquement neutre, est préférable.

Malheureusement, les sources d'implantation de germanium n'étant pas disponibles, cette étape d'amorphisation n'a pas pu être intégrée.

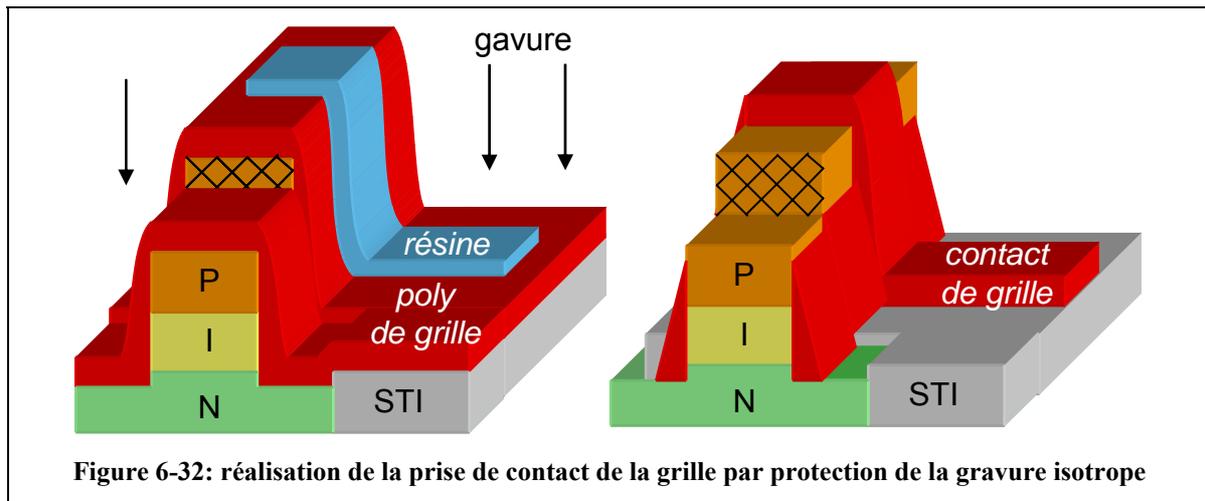


6.3.1.4. Polysilicium de grille - contact de grille

La grille et sa prise de contact sont réalisées en polysilicium dopé. La grille est formée par des espaceurs, résultant d'un dépôt de polysilicium et une gravure isotrope. La prise de contact est réalisée en protégeant le dépôt de la gravure.



Pour réaliser les espaceurs, le dépôt doit être mince (50nm à 100nm). Or, il est très difficile de doper une couche aussi mince par implantation. De plus, il serait nécessaire d'activer les dopants par recuit thermique, mais cela entraînerait une diffusion des zones épitaxiées, ce qui n'est pas souhaitable. Par conséquent, nous avons opté pour un dépôt dopé *in situ*. Il est réalisé sous forme d'épitaxie.

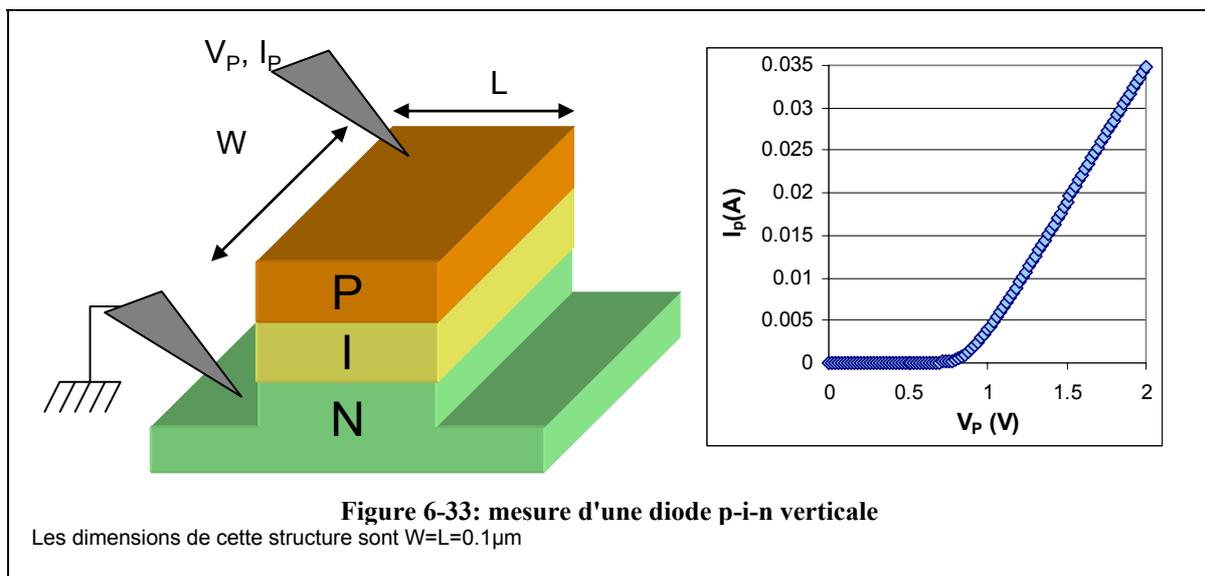


La photolithographie nécessaire à la réalisation de la prise de contact, inexistante en CMOS conventionnel, a été intégrée sur le niveau correspondant à l'implantation des LDD. Or, ce masque est normalement utilisé pour ouvrir des zones, alors que nous le détournons pour masquer une petite zone. Il est donc nécessaire d'adapter le temps et l'exposition de la photolithographie.

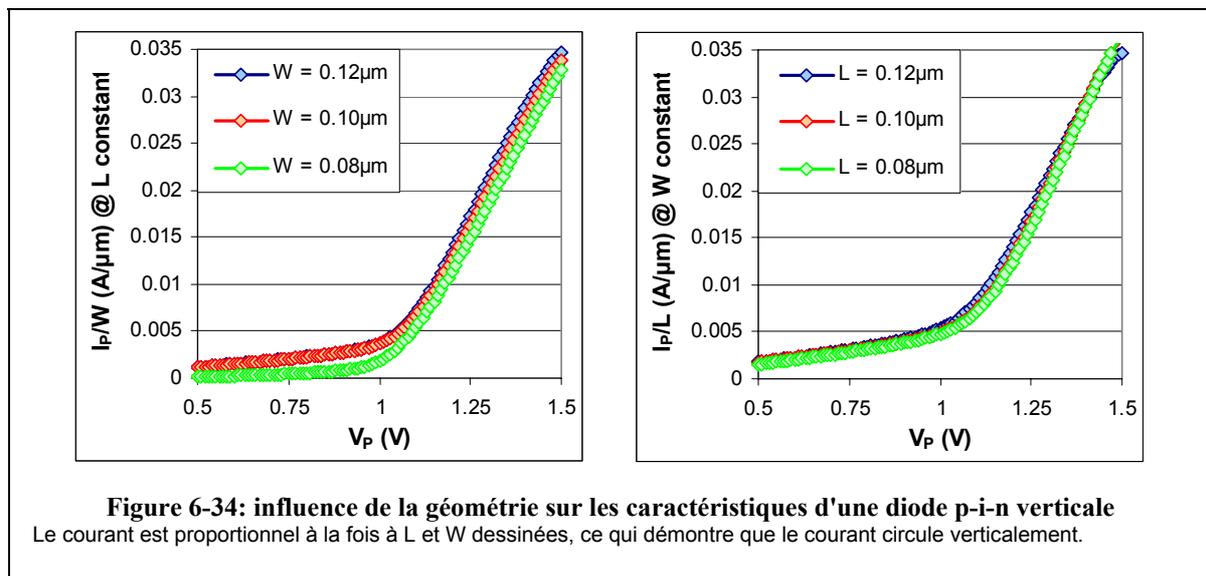
6.3.2. Mesures sur l'intégration verticale

6.3.2.1. Démonstration de la conduction verticale

La première étape est la réalisation de diodes p-i-n verticales. Celles-ci peuvent être testées simplement en étant polarisées en inverse. La **figure 6-33** présente le principe de mesure ainsi qu'une caractéristique typique. La diode présente un seuil à 0.8V, et un niveau de courant élevé du à la conduction volumique.

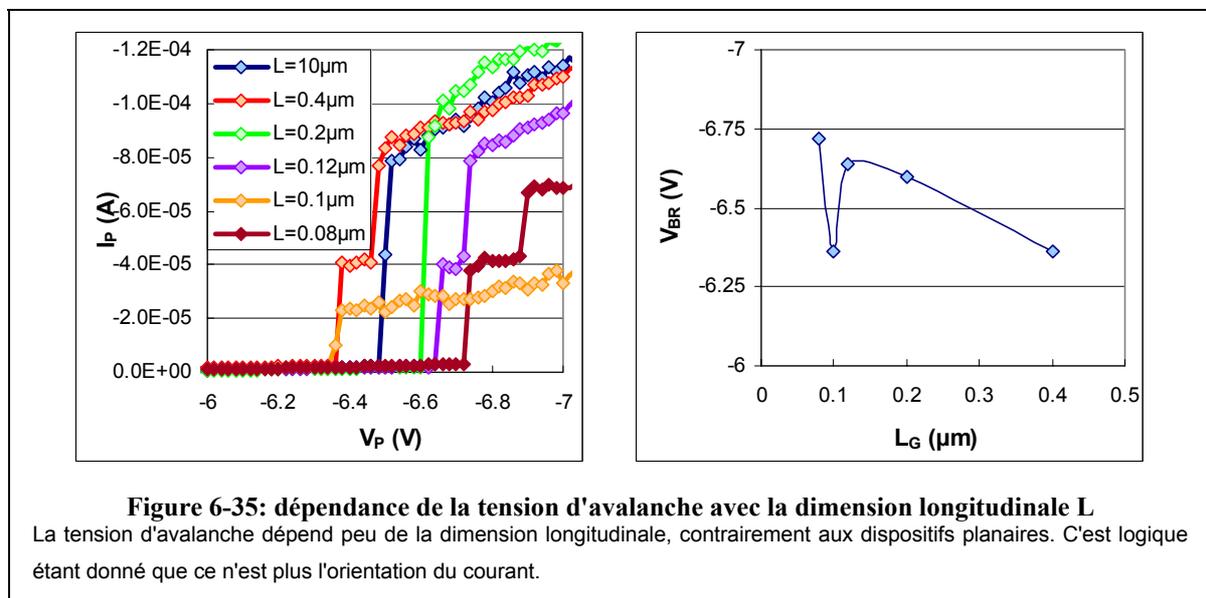


La **figure 6-34** présente les caractéristiques de différentes géométries. Le courant est proportionnel à la surface des dispositifs, ce qui est caractéristique d'une conduction verticale.



6.3.2.2. Tension d'avalanche

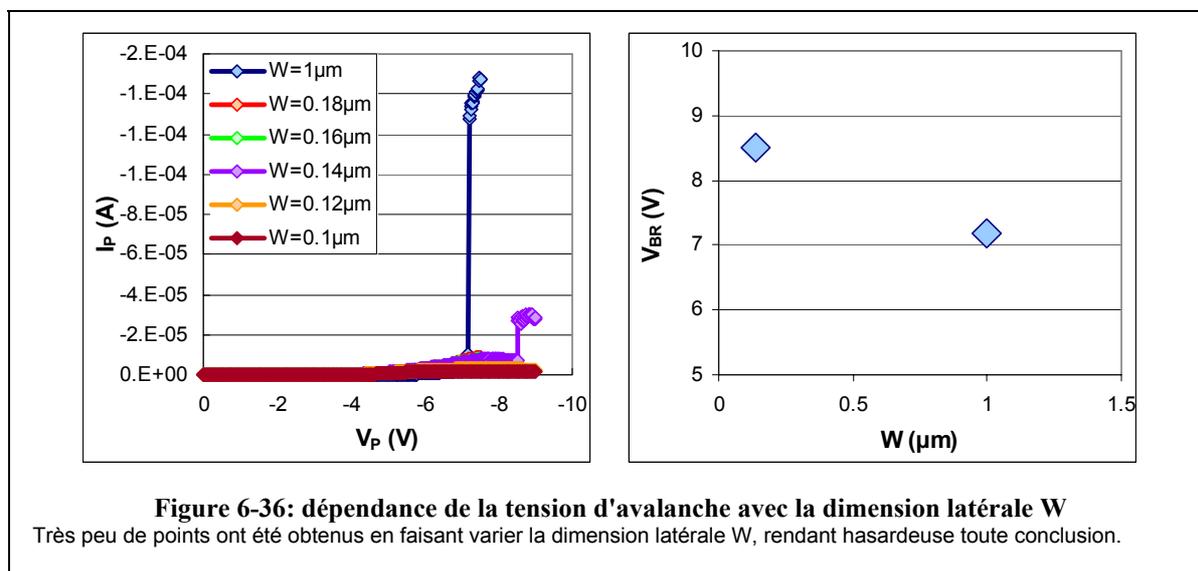
La tension d'avalanche des dispositifs mesurés dépend de leur géométrie. La dépendance avec la dimension longitudinale L est représentée sur la **figure 6-35**. La tension d'avalanche croît lorsque la dimension diminue.



De manière similaire, la tension d'avalanche diminue lorsque la dimension latérale W augmente, mais dans de plus grandes proportions (**figure 6-36**).

Cette dépendance de la tension d'avalanche avec la géométrie est étonnante. Si l'amplitude de variation dans la dimension longitudinale est raisonnable, celle dans la dimension latérale est

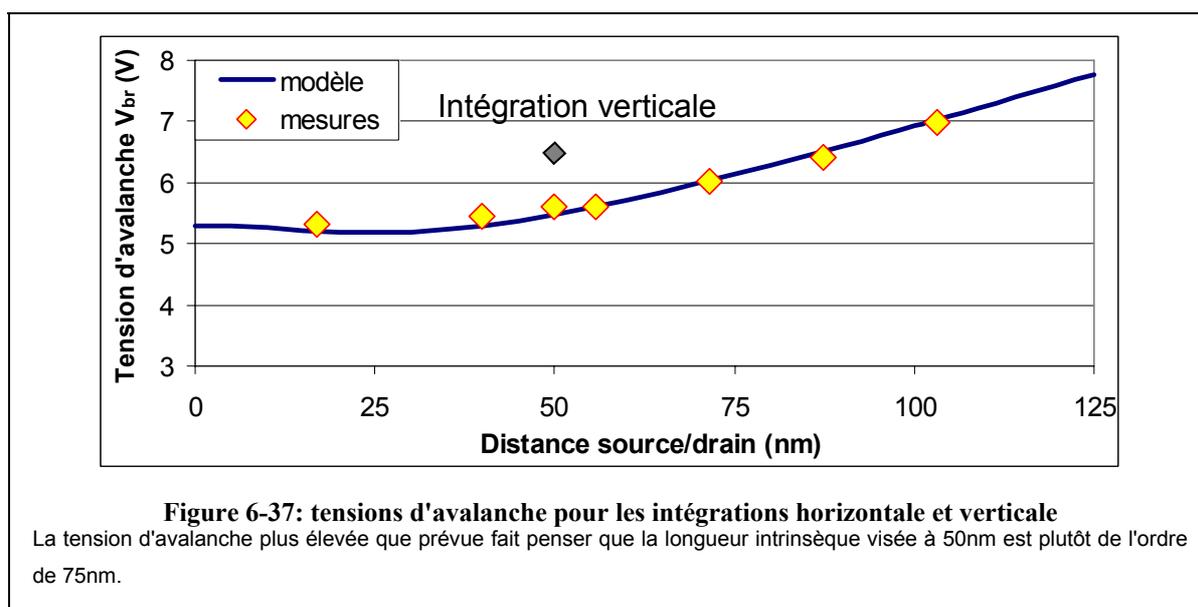
très élevée. Cependant, très peu de points ont été obtenus, et rendant difficile la détection d'un point aberrant.



Comparaison avec l'intégration horizontale

La **figure 6-37** porte sur la courbe tension d'avalanche/distance source/drain obtenue pour l'intégration horizontale le point de l'intégration verticale. La distance source/drain est supposée être égale à l'épaisseur de silicium intrinsèque épitaxié, soit 50nm.

Ce point n'est pas en adéquation avec les mesures horizontales. Cependant, la totalité des points expérimentaux se trouvent sur cette courbe. Par conséquent, ce décalage semble indiqué que la distance source/drain soit plutôt de l'ordre de 75nm.



Conclusion du chapitre

L'intégration de dispositifs I-MOS est plus délicate que l'intégration de dispositifs MOS conventionnels, à cause de ses asymétries: source et drain sont dopés de types différents, et la grille recouvre le drain mais pas la source. Ceci empêche notamment d'utiliser un procédé entièrement auto-aligné, rendant difficile la réalisation de petites dimensions.

L'étude des différentes solutions proposées dans la littérature a relevé deux idées judicieuses: la définition de la zone intrinsèque peut être réalisée à l'aide d'un espaceur, et la discrimination des zones sources et drain par photo-lithographie d'un masque existant, volontairement décalé.

Cette approche nous a permis de réaliser les dispositifs les plus courts jamais publiés (17nm de longueur de grille), dispositifs qui présentent des niveaux de courant de conduction élevés.

La comparaison de nos dispositifs avec ceux de la littérature en terme de tension d'avalanche a permis de souligner que celle-ci ne dépend pas de l'intégration, mais uniquement des dimensions. En particulier, il semble que la tension d'avalanche la plus faible que l'on puisse obtenir avec un dispositif silicium soit de l'ordre de 5V. Ceci est élevé, en regard des tensions d'alimentation actuelles: ~1V en logique, jusqu'à 2.5V pour les entrées/sorties. Ceci est un réel handicap pour l'industrialisation de dispositifs I-MOS. La solution la plus intuitive pour diminuer la tension d'avalanche est d'intégrer des dispositifs en germanium, matériau plus propice à l'ionisation par impact.

En revanche, le niveau de courant de conduction élevé reste attrayant, en particulier lorsque l'on vise une application à forte densité. Par exemple, il serait intéressant de pouvoir disposer d'un I-MOS pour contrôler une cellule P-RAM (mémoire à changement de phase).

Conclusion générale et perspectives

Les performances du transistor MOS sont caractérisées par le couple courant de conduction I_{on} / courant de fuite I_{off} . Ces deux grandeurs sont liées au premier ordre par la pente sous le seuil S , caractérisant le régime sous le seuil du transistor. Celui-ci a été étudié dans la première partie de ce manuscrit, aboutissant aux expressions analytiques de la tension de seuil, de la pente sous le seuil et du courant sous le seuil pour des substrats de type silicium massif et SOI. Ces expressions intègrent la modélisation des effets canaux courts au moyen de la transformation tension dopage (VDT) qui permet d'assimiler les effets de champs latéral à une variation du dopage apparent du canal sur un substrat standard. Dans le cas d'un substrat SOI, la démarche a été étendue en considérant une charge électrique apparente dans l'oxyde enterré. Cette approche analytique a été validée par des simulations TCAD jusqu'à des longueurs de grille de 20nm.

La pente sous le seuil du transistor MOS étant thermodynamiquement limitée à $k_B \cdot T/q = 60\text{mV/dec}$ @ 300K, d'autres dispositifs permettant de franchir cette limite sont passés en revue. Ceux-ci peuvent classés en trois catégories:

- les structures de rupture, qui utilisent de nouveaux matériaux (à transition de Mott par exemple), ou d'autres effets (mécanique par exemple). Elles sont encombrantes par rapport au transistor MOS conventionnel, semblent trop éloignées du comportement du transistor MOS conventionnel pour les assimiler à des structures à pente inférieure à 60mV/dec.
- les structures utilisant l'effet tunnel: en effet, la conduction par effet tunnel est différente de la conduction par diffusion, ce qui permet de s'affranchir de la limite usuelle de pente. Bien que des simulations suggèrent qu'un MOS à jonction Schottky puisse présenter une pente inférieure à 60mV/dec, nous pensons que ces simulations ont utilisé un modèle de tunnel bande à bande en dehors de son domaine d'utilisation. A ce jour, aucune démonstration de structure de type MOS utilisant l'effet tunnel pour diminuer la pente sous le seuil n'a été faite.
- les structures utilisant un phénomène d'amplification, comme le gain d'une structure bipolaire ou le phénomène d'ionisation par impact. Parmi eux, l'I-MOS, dispositif basé sur le phénomène d'amplification de courant par ionisation par impact, est intégrable en technologie silicium conventionnelle et a montré une pente sous le seuil aussi faible que 2mV/décade.

Ce dispositif complémentaire semblant prometteur, l'essentiel des travaux de recherche menés dans le cadre de cette thèse portent sur celui-ci. D'un point de vue théorique, un I-MOS peut

être considéré comme une jonction p-i-n dont la longueur intrinsèque varie avec la tension de grille. Cette analogie implique qu'un certain nombre de phénomènes peuvent être compris à partir de l'étude de la diode p-i-n, dispositif plus simple, pour expliquer le fonctionnement d'un transistor I-MOS. En particulier, la tension d'avalanche diminue avec la longueur de grille jusqu'à environ 5V: l'ionisation par impact est limitée par le champ électrique. Pour les dispositifs les plus courts, cette tension remonte légèrement. C'est alors la distance que parcourent les porteurs qui est trop faible pour générer d'autres porteurs par ionisation. Il est alors nécessaire d'augmenter leur énergie en augmentant la tension. Cette limite dépend du matériau: elle est de 5V environ pour le silicium, mais de l'ordre de 0.3V pour le germanium. En outre, ce dispositif ne présente pas de saturation: le courant qu'il débite n'est limité que par les résistances extrinsèques, ce qui implique de très forts niveaux de courant.

La pente sous le seuil très faible permet d'obtenir un courant de fuite faible, et donc de limiter la consommation statique. Le courant de conduction élevé permet à un inverseur en technologie I-MOS complémentaire d'être plus rapide que son homologue réalisé en technologie CMOS conventionnelle. Le handicap essentiel de l'I-MOS reste son schéma de polarisation: +5V sur la source du pmos, -5V à la source du nmos pour un signal variant entre 0V et 1V. Cette décade d'augmentation de la tension d'alimentation est en partie compensée par la forme spécifique du signal, mais les consommations dynamiques et statiques sont tout de même trois fois supérieures à celles d'un inverseur CMOS pour un gain de 25% en fréquence.

Ce gain de performances se retrouve sur des circuits plus complexes, bien que la polarisation spécifique nécessite d'adapter les circuits. Nous avons notamment montré comment réaliser des portes NAND et NOR en adaptant légèrement l'architecture usuelle. En outre, une SRAM réalisée en technologie I-MOS présente une marge de bruit statique (SNM) de 250mV/V, ce qui est tout à fait satisfaisant. Ces études permettent de considérer avec confiance l'I-MOS dans un environnement circuit.

Enfin, les études théoriques sur les dispositifs isolés ont été confirmées par des réalisations technologiques, dont le principal enjeu est l'asymétrie de dopage n/p.

Une intégration horizontale, proche de l'intégration CMOS classique, a permis confirmer le fonctionnement de l'I-MOS, à commencer par la relation entre dimension et tension d'avalanche. Le procédé de fabrication n'étant pas optimisé, les mesures sont perturbées par

des courants de fuites, rendant notamment impossible la mesure de pente sous le seuil. Ce procédé a cependant permis d'explorer ce dispositif de façon plus avancée que dans la littérature: un oxyde de grille plus mince, des longueurs de grilles jusqu'à 17nm (contre 60nm précédemment) sont à l'origine de tension d'avalanche et de niveaux de courant inégaux.

La caractéristique linéaire au-delà de l'avalanche traduit un comportement ohmique, dont la valeur est de l'ordre des résistances parasites. Toutes choses égales par ailleurs, ce comportement est fortement influencé par le nombre de contacts du dispositif. Cette faible résistance équivalente a permis de mesurer un courant jusqu'à $4700\mu\text{A}/\mu\text{m}$, avant que les contacts ne soient détruits.

Pour que cette architecture soit pérenne, il est nécessaire de pouvoir réduire ses dimensions. Hors, les dispositifs les plus courts présentent une saturation de la tension d'avalanche. Celle-ci s'explique par la trop courte distance parcourue par les porteurs pour générer des porteurs par ionisation par impact. Par conséquent, une seconde réalisation est proposée. Il s'agit d'intégrer verticalement les I-MOS, ce qui permet de conserver une distance électrique suffisante tout en améliorant la compacité. Les dispositifs réalisés montrent une conduction verticale, mais pas de contrôle du courant par la grille.

Le point de départ de ces travaux est la limite fondamentale d'une pente sous le seuil de kT/q pour le transistor MOS conventionnel. La modélisation du courant sous le seuil a permis de quantifier l'écart à cette valeur idéale pour les technologies planaire et SOI, mais souligne que cette limite n'est pas sensible aux leviers technologiques utilisés pour améliorer les performances, telle l'amélioration de mobilité.

La revue de différentes architectures suffisamment proches du transistor MOS pour le remplacer partiellement a mis en évidence les énormes atouts de celui-ci: compacité, performances, renforcées par des années de recherches et d'industrialisation, permises grâce à un enjeu économique à la fois énorme et croissant. Cependant, l'I-MOS a des aspects séduisants: il peut être intégré sur silicium sans rupture de technologie, existe en version complémentaire, et surtout dispose d'une pente sous le seuil non limitée. Les différentes réalisations ont démontré sa supériorité au transistor MOS en terme de courant de conduction, et l'étude sur les circuits que ce gain a de l'intérêt. En outre, sa capacité à fournir un courant élevé est particulièrement intéressante dans des applications nécessitant de la puissance, comme des mémoires à changement de phase.

Il est cependant terriblement handicapé par la tension d'alimentation nécessaire à produire de l'ionisation par impact, supérieure à 5V. Dans ces conditions, il est hors de question d'envisager une application industrielle classique pour l'I-MOS. La solution la plus intuitive qui pourrait assurer son avenir est l'utilisation du germanium, qui devrait faire tomber la tension d'avalanche à 0.3V environ. Cependant, on se heurte à la difficulté d'intégration du germanium, en particulier la réalisation d'un oxyde stable. Le transistor MOS ayant également à gagner à être intégré sur germanium dont la mobilité est plus élevée que le silicium, on peut espérer que les avancées de l'un profite à l'autre. D'autant plus qu'un des inconvénients de l'intégration d'un MOS à canal germanium, à savoir des fuites de jonctions plus élevées, ne se présente pas pour l'I-MOS.

Enfin, il y a encore bien des questions dont la réponse est importante dès lors qu'on envisage une application industrielle:

- bien que le transistor I-MOS semble pouvoir être utilisé dans des circuits, il n'y a pas encore de réalisation permettant d'évaluer le gain réel de performance
- jusqu'à quel point peut-on intégrer un dispositif asymétrique? Toutes les propositions de la littératures montrent une faiblesse de ce point de vue, alors qu'on a montré un effet transistor sur un MOS 6nm de longueur de grille
- peut-on décemment utiliser le phénomène d'avalanche, alors qu'il est à l'origine de dégradations dans le transistor MOS?
- quand on sait que les diodes zeners basée sur l'ionisation par impact étaient utilisées comme générateurs de bruit blanc, on peut s'interroger sur la pertinence d'utiliser un I-MOS dans un circuit analogique.

Cependant, ces questions restent tout de même d'un second ordre par rapport à ce qu'on sait déjà de l'I-MOS: il présente un effet transistor, est intégrable en technologie silicium conventionnelle, présente une pente sous le seuil inférieure à 60mV/dec @ 300K, son courant de conduction ne sature pas, il est utilisable dans des circuits. Bien peu d'autres dispositifs permettent de concurrencer autant le transistor MOS, malgré quelques inconvénients pour lesquels on imagine déjà des solutions.

Alors, que les recherches continuent...

Bibliographie

- [Abele'05a] **N. Abele.**
Suspended-Gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor, proceedings of IEDM, **2005a**
- [Abele'05b] **N. Abele, V. Pott, K. Boucart et al.**
Comparison of RSG-MOSFET and capacitive MEMS resonator detection. Electronics Letters, **2005b**, Vol.41, N°5, p. 242
- [Abele'06] **N. Abele, A. Villaret, A. Gangadharaiah et al.**
1T MEMS Memory Based on Suspended Gate MOSFET, proceedings of IEDM, **2006**
- [Barett'02] <http://www.vnunet.fr/fr/vnunet/news/2002/10/03/craig-barrett-futur-sera-numerique-sera>
- [Benvenuti'93a] **A. Benvenuti, F. Bonani, G. Ghione et al.**
Analysis of output ndr in power AlGaAs/GaAs hbt's by means of a thermal-fully hydrodynamic model, proceedings of ISDRS, **1993a**
- [Benvenuti'93b] **A. Benvenuti, G. Ghione et C.U. Naldi.**
Non-stationary transport hbt modeling under non-isothermal conditions, proceedings of SISDEP-5, **1993b**
- [Benvenuti'92] **A. Benvenuti, G. Ghione, M.R. Pinto et al.**
Coupled thermal-fully hydrodynamic simulation of InP-based HBTs, proceedings of IEDM, **1992**
- [Beteille'99] **F. Beteille, L. Mazerolles et J. Livage.**
Microstructure and metal-insulating transition of VO₂ thin films. Materials Research Bulletin, **1999**, Vol.34, N°14-15, p. 2177
- [Bhat'88] **A.K.S. Bhat.**
IGT-gated bipolar MOS hybrid cascode switch, proceedings of **1988**
- [Bhuwalka'05] **K.K. Bhuwalka, J. Schulze et I. Eisele.**
Scaling the vertical tunnel FET with tunnel bandgap modulation and gate workfunction engineering. Electron Devices, IEEE Transactions on, **2005**, Vol.52, N°5, p. 909
- [Bidaud'01] **M. Bidaud.** *Croissance et caractérisation d'oxydes de grille ultra-minces (1 - 2.5nm) pour les technologies CMOS 0.10 - 0.12µm*, Manuscrit de thèse
- [Blanks'98] **D.K. Blanks, G. Klimeck, R. Lake et al.**
NEMO: general release of a new comprehensive quantum device simulator, proceedings of **1998**
- [C2Alliance'03] <http://www.minatec.com/cgi-bin/charge.pl>

- [Callen'85] **H.B. Callen.** *Thermodynamics and an Introduction to Thermostatistics*, **1985**
- [Charbuillet'06a] **C. Charbuillet, E. Dubois, S. Monfray et al.**
Fabrication and Analysis of CMOS Fully-Compatible High Conductance Impact-Ionization MOS (I-MOS) Transistors, proceedings of ESSDERC, **2006a**
- [Charbuillet'06b] **C. Charbuillet, S. Monfray, E. Dubois et al.**
High Current Drive in Ultra-Short Impact Ionization MOS (I-MOS) Devices, proceedings of IEDM, **2006b**
- [Chen'93] **D. Chen, Z. Yu, K.-C. Wu et al.**
Dual energy transport model with coupled lattice and carrier temperatures, proceedings of SISDEP-5, **1993**
- [Choi'04a] **W.Y. Choi, B.Y. Choi, D.-S. Woo et al.**
A new fabrication method for self-aligned nanoscale I-MOS (impact-ionization MOS), proceedings of Device Research Conference, 2004. 62nd DRC. Conference Digest [Late News Papers volume included], **2004a**
- [Choi'04b] **W.Y. Choi, J.Y. Song, B.Y. Choi et al.**
80nm self-aligned complementary I-MOS using double sidewall spacer and elevated drain structure and its applicability to amplifiers with high linearity, proceedings of Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International, **2004b**
- [Choi'05a] **W.Y. Choi, J.Y. Song, J.D. Lee et al.**
70-nm Impact-Ionization Metal-Oxide-Semiconductor (I-MOS) Devices Integrated with Tunneling Field-Effect Transistors (TFETs), proceedings of IEDM, **2005a**
- [Choi'05b] **W.Y. Choi, J.Y. Song, J.D. Lee et al.**
100-nm n-/p-Channel I-MOS Using a Novel Self-Aligned Structure. *Electron Device Letters*, IEEE, **2005b**, Vol.26, N°4, p. 261-263
- [Choi'05c] **W.Y. Choi, J.Y. Song, J.D. Lee et al.**
A Novel Biasing Scheme for I-MOS (Impact-Ionization MOS) Devices. *Nanotechnology*, IEEE Transactions on, **2005c**, Vol.4, N°3, p. 322-325
- [Chynoweth'58] **Chynoweth.**
Ionization rates for electrons and holes in silicon. *Physical Review*, **1958**, Vol.109, N°5, p. 1537-1540
- [Davis'86] **J.R. Davis, A.E. Glaccum, K. Reeson et al.**
Improved subthreshold characteristics of n-channel SOI transistors. *Electron Device Letters*, IEEE, **1986**, Vol.7, N°10, p. 570
- [Duriez'04] **B. Duriez, B. Tavel, R. Boeuf et al.**
Gate stack optimization for 65 nm CMOS low power and high performance platform, proceedings of IEDM, **2004**
- [Eitan'82] **B. Eitan.**

IEEE transactions on electronic devices, **1982**, Vol.29, N°2, p. 254-265

[Fossum'87] **J.G. Fossum, R. Sundaesan et M. Matloubian.**
Anomalous subthreshold current-Voltage characteristics of n-channel SOI MOSFET's.
Electron Device Letters, IEEE, **1987**, Vol.8, N°11, p. 544

[Fu'84]**J.S. Fu.**
IEEE transactions on electronic devices, **1984**, Vol.31, N°4, p. 440-447

[Fu'97]**K.-Y. Fu.**
IEEE transactions on electronic devices, **1997**, Vol.44, N°5, p. 847-855

[Gartner'06] http://www.gartner.com/press_releases/asset_141290_11.html

[Ghani'03] **T. Ghani, M. Armstrong, C. Auth et al.**
A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors, proceedings of IEDM, **2003**

[Gopalakrishnan'02] **K. Gopalakrishnan, P.B. Griffin et J.D. Plummer.**
I-MOS: a novel semiconductor device with a subthreshold slope lower than kT/q , proceedings of Electron Devices Meeting, 2002. IEDM '02. Digest. International, **2002**

[Gopalakrishnan'05a] **K. Gopalakrishnan, P.B. Griffin et J.D. Plummer.**
Impact ionization MOS (I-MOS)-Part I: device and circuit simulations. Electron Devices, IEEE Transactions on, **2005a**, Vol.52, N°1, p. 69-76

[Gopalakrishnan'05b] **K. Gopalakrishnan, R. Woo, C. Jungemann et al.**
Impact ionization MOS (I-MOS)-Part II: experimental results. Electron Devices, IEEE Transactions on, **2005b**, Vol.52, N°1, p. 77-84

[Grant'73] **Grant.**
Electron and hole ionization rates in epitaxial silicon at high electric fields. Solid State Electronics, **1973**, Vol.16, p. 1189-1203

[Gwoziecki'02] **R. Gwoziecki et T. Skotnicki.**
Physics of the subthreshold slope - initial improvement and final degradation in short CMOS devices, proceedings of ESSDERC, **2002**

[Harrison'05] **S. Harrison.** *Dispositifs GAA en technologie SON: conception, caractérisation et modélisation en vue de l'intégration dans les noeuds CMOS avancés*, Manuscrit de thèse

[Harrison'04] **S. Harrison, P. Coronel, A. Cros et al.**
Poly-gate replacement through contact hole (PRETCH): a new method for high-k/metal gate and multi-oxide implementation on chip, proceedings of **2004**

[Ionescu'02] **A. Ionescu.**
proceedings of International Sympos. Quality Electronic Design, **2002**

[ISE'04] **ISE.** Dessis user's manual, **2004**

- [ITRS'05] **ITRS.**
[Jang'04] **M. Jang, Y. Kim, J. Shin et al.**
Novel properties of erbium-silicided n-type Schottky barrier MOSFETs, proceedings of SSDM, **2004**
- [Josse'00] **E. Josse.** *Nouvelles architectures de grille pour les générations CMOS 0.1 μ m et en-deçà*, Manuscrit de thèse
- [Jurczak'00] **M. Jurczak, T. Skotnicki, M. Paoli et al.**
Silicon-on-Nothing (SON)-an innovative process for advanced CMOS. Electron Devices, IEEE Transactions on, **2000**, Vol.47, N°11, p. 2179
- [Kim'04] **H.-T. Kim, B.-G. Chae, D.-H. Youn et al.**
Mechanism and observation of Mott transition in V02-based two- and three-terminal devices. New Journal of Physics, **2004**, Vol.6, N°52, p. 0
- [Kim'00] **S.D. Kim, C.-M. Park et W.J.C. S.**
Advanced model and analysis for series resistance in sub-100 nm CMOS including poly depletion and overlap doping gradient effect, proceedings of **2000**
- [Lackner'91] **T. Lackner.**
Avalanche multiplication in semiconductors: a modification of Chynoweth's law. Solid State Electronics, **1991**, Vol.34, N°1, p. 33-42
- [Lee'64] **Lee, Logan, Batdorf et al.**
Ionization rates of holes and electrons in silicon. Physical Review, **1964**, Vol.134, N°3A, p. 761-773
- [Livage'00] **J. Livage.**
Les procédés sol-gel. revue VERRE, **2000**, Vol.6, N°5, p. 1
- [Livage'01] **J. Livage.**
Sol-gel electrochromic coatings and devices: a review. Solar Energy Materials and Solar Cells, **2001**, Vol.68, p. 365-381
- [Lundstrom'00] **M.S. Lundstrom.** Fundamentals of Carrier Transport, **2000**
- [Mastar'04] **Mastar.** Mastar user's guide
- [Mathieu'98] **H. Mathieu.** Physique des semiconducteurs et des composants électroniques, **1998**
- [Mayer'06] **F. Mayer.**
Co-integration of 2 mV/dec Subthreshold Slope Impact Ionization MOS (I-MOS) with CMOS, proceedings of ESSDERC, **2006**
- [Mondot'05] **A. Mondot, M. Muller, D. Aim et al.**
Silicidation induced strain phenomena in totally silicided (TOSI) gate transistors, proceedings of ESSDERC, **2005**

- [Monfray'02] **S. Monfray, T. Skotnicki, B. Tavel et al.**
SON (Silicon-On-Nothing) P-MOSFETs with totally silicided (CoSi/sub 2/) polysilicon on 5 nm-thick Si-films: the simplest way to integration of metal gates on thin FD channels, proceedings of IEDM, **2002**
- [Moore'65] **G. Moore.**
Cramming more components in integrated circuits. Electronics, **1965**, Vol.38, N°8, p. 114
- [Mott'90] **S.N. Mott.** Metal-Insulator Transitions, **1990**
- [Müller'56] **D.E. Müller.**
A method for solving algebraic equations using an automatic computer. Mathematical Tables and Other Aids to Computation, **1956**, Vol.10, p. 208-215
- [Muller'05] **M. Muller, A. Mondot, D. Aime et al.**
CMP-less integration of 40nm-gate totally silicided (TOSI) bulk transistors using selective S/D Si epitaxy and ultra-low gates, proceedings of ESSDERC, **2005**
- [Nemati'99] **F. Nemati et J.D. Plummer.**
A novel thyristor-based SRAM cell (T-RAM) for high-speed, low-voltage, giga-scale memories, proceedings of IEDM, **1999**
- [Newell'66] **W. Newell.**
Novel circuit aspects of the resonant gate transistor, proceedings of **1966**
- [Nirschl'05a] **T. Nirschl.**
The 65nm Tunneling Field Effect Transistor (TFET) 0.68um² 6T Memory Cell and MultiVth Device, proceedings of ESSDERC, **2005a**
- [Nirschl'05b] **T. Nirschl, U. Schaper, J. Einfeld et al.**
Impact of mask alignment on the tunneling field effect transistor (TFET), proceedings of International Conference on Microelectronic Test Structures, **2005b**
- [Oica'03] <http://www.oica.net/htdocs/statistics/OICA%204%20VOLETS%20F.pdf>
- [Okuto'75] **Okuto et Crowell.**
Threshold energy effect on avalanche breakdown voltage in semiconductor junctions. Solid State Electronics, **1975**, Vol.18, p. 161-168
- [Ortolland'04] **C. Ortolland, S. Orain, J. Rosa et al.**
Electrical characterization and mechanical modeling of process induced strain in 65 nm CMOS technology, proceedings of ESSDERC, **2004**
- [Pakfar'03] **A. Pakfar.** *Modélisation de la diffusion des dopants dans les alliages SiGe et SiGeC*, Manuscrit de thèse
- [Payet'06] **F. Payet.** *Modélisation et intégration de transistors à canal de silicium contraint pour les noeuds technologiques CMOS 45nm et en-deçà*, Manuscrit de thèse

- [Plimmer'00] **S.A. Plimmer, J.P.R. David et Ong.**
The merits and limitations of local impact ionization theory. IEEE transactions on electronic devices, **2000**, Vol.47, N°5, p. 1080-1088
- [Schuegraf'94] **K.F. Schuegraf, D. Park et C. Hu.**
Reliability of thin SiO₂ at direct tunneling voltages, proceedings of IEDM, **1994**
- [Seevinck'87] **E. Seevinck, F.J. List et J. Lohstroh.**
Static-noise margin analysis of MOS SRAM cells. Solid-State Circuits, IEEE Journal of, **1987**, Vol.22, N°5, p. 748
- [Segueni'07] **K. Segueni.** *Etude et fabrication de micro systèmes électro-mécaniques réalisés en technologie SON,* Manuscrit de thèse
- [Shimizu'01] **A. Shimizu, K. Hachimine, N. Ohki et al.**
Local mechanical-stress control (LMC): a new technique for CMOS-performance enhancement, proceedings of IEDM, **2001**
- [Skotnicki] **T. Skotnicki.** cours de DEA
- [Skotnicki'00] **T. Skotnicki.**
Transistor MOS et sa technologie de fabrication. *techniques de l'ingénieur*, **2000**
- [Skotnicki'03] **T. Skotnicki et F. Boeuf.** EGEM, chapitre 3, Introduction à la physique du transistor MOS, **2003**
- [Skotnicki'88] **T. Skotnicki, G. Merckel et T. Pedron.**
A new punchthrough current model based on the voltage-doping transformation. Electron Devices, IEEE Transactions on, **1988**, Vol.35, N°7, p. 1076
- [Sodamos] **Sodamos.**
- [Su'02] **P. Su, K.-I. Goto, T. Sugii et al.**
A thermal activation view of low voltage impact ionization in MOSFETs. IEEE electron device letters, **2002**, Vol.23, N°9, p. 550-552
- [Sze'81] **S.M. Sze.** Physics of Semiconductor Devices, **1981**
- [Sze'66] **S.M. Sze et Gibbons.**
Avalanche breakdown voltages of abrupt and linearly graded p-n junctions in Ge, Si, GaAs and GaP. Applied Physics Letters, **1966**, Vol.8, N°5, p. 111-113
- [Tihanyi'75] **J. Tihanyi et H. Schlotterer.**
Properties of ESFI MOS transistors due to the floating substrate and the finite volume. Electron Devices, IEEE Transactions on, **1975**, Vol.22, N°11, p. 1017
- [Toh'05] **E.-H. Toh, G.H. Wang, G.-Q. Lo et al.**
A Novel CMOS Compatible L-Shaped Impact-Ionization MOS (LI-MOS) Transistor, proceedings of IEDM, **2005**
- [Valdinoci'99] **Valdinoci, Ventura, M.C. Vecchi et al.**

Impact-ionization in silicon at large operating temperature, proceedings of International Conference on Simulation of Semiconductor Processes and Devices, **1999**

[Van Overstraeten'70] **Van Overstraeten.**
Measurement of the ionization rates in diffused silicon P-N junctions. Solid State Electronics, **1970**, Vol.13, p. 583-608

[VanDamme'97] **E.P. VanDamme.**
IEEE electron device letters, **1997**, Vol.18, N°6, p.

[Veendrick'84] **H.J.M. Veendrick.**
Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits. Solid-State Circuits, IEEE Journal of, **1984**, Vol.19, N°4, p. 468

Bibliographie de l'auteur

S. Monfray, T. Skotnicki, B. Tavel, Y. Morand, S. Descombes, A. Talbot, D. Dutartre, C. Jenny, P. Mazoyer, R. Palla, F. Leverd, Y. Le Friec, R. Pantel, M. Haond, C. Charbuillet, C. Vizioz, D. Louis et N. Buffet.

SON (Silicon-On-Nothing) P-MOSFETs with totally silicided (CoSi/sub 2/) polysilicon on 5 nm-thick Si-films: the simplest way to integration of metal gates on thin FD channels, proceedings of IEDM, 2002

M. Muller, B. Froment, V. Carron, A. Beverina, R. Palla, R. Pantel, P. Morin, C. Charbuillet, A. Pouydebasque, F. Boeuf et T. Skotnicki.

New junction concepts for sub-50 nm CMOS transistors: slim spacers and Ni silicide, proceedings of ESSDERC, 2003

R. Ranica, A. Villaret, C. Fenouillet-Beranger, P. Malinge, P. Mazoyer, P. Masson, D. Delille, C. Charbuillet, P. Candelier et T. Skotnicki.

A capacitor-less DRAM cell on 75nm gate length, 16nm thin fully depleted SOI device for high density embedded memories, proceedings of IEDM, 2004

C. Charbuillet, E. Dubois, S. Monfray, P. Bouillon et T. Skotnicki.

Fabrication and Analysis of CMOS Fully-Compatible High Conductance Impact-Ionization MOS (I-MOS) Transistors, proceedings of ESSDERC, 2006a

C. Charbuillet, S. Monfray, E. Dubois, P. Bouillon, F. Judong et T. Skotnicki.

High Current Drive in Ultra-Short Impact Ionization MOS (I-MOS) Devices, proceedings of IEDM, 2006b

C. Gallon, C. Fenouillet-Beranger, A. Vandooren, F. Boeuf, S. Monfray, F. Payet, S. Orain, V. Fiori, F. Salvetti, N. Loubet, C. Charbuillet, A. Toffoli, F. Allain, K. Romanjek, I. Cayrefourcq, B. Ghyselen, C. Mazure, D. Delille, F. Judong, C. Perrot, M. Hopstaken, P. Scheblin, P. Rivallin, L. Brevard, O. Faynot, S. Cristoloveanu et T. Skotnicki.

Ultra-Thin Fully Depleted SOI Devices with Thin BOX, Ground Plane and Strained Liner Booster, proceedings of International SOI Conference, 2006

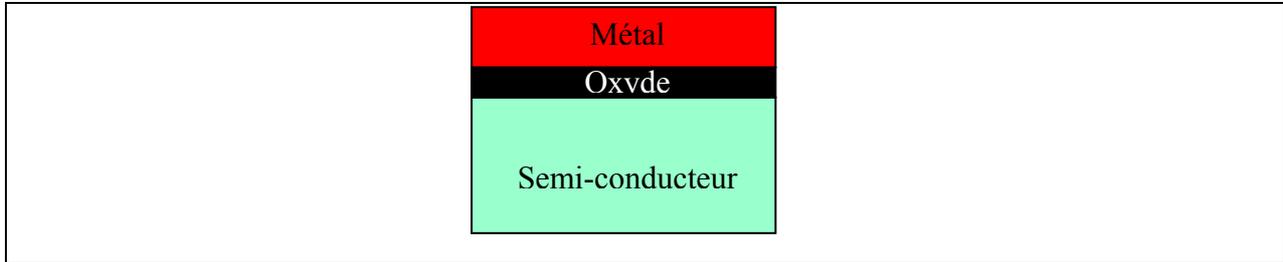
A. Pouydebasque, C. Charbuillet, R. Gwoziecki et T. Skotnicki.

Refinement of the sub-threshold slope modeling for advanced bulk CMOS devices, IEEE Transactions on Electron Devices, 2007, Vol.54, N°10, p. 2723

Annexes

Annexe 1 - Etude d'une capacité MOS

Structure



Remarques préalables

- les porteurs sont répartis selon une loi de Boltzmann:

$$p = p_0 \cdot \exp\left(\frac{-\psi \cdot q}{k \cdot T}\right) \quad (1)$$

$$n = n_0 \cdot \exp\left(\frac{\psi \cdot q}{k \cdot T}\right) \quad (2)$$

- en profondeur dans le semi-conducteur

- le potentiel est nul $\psi=0$. Il en découle $n=n_0$ et $p=p_0$.

- on peut exprimer les densités de porteurs n_0 et p_0 :

$$p_0 = n_i \exp\left(\frac{q \cdot \phi_F}{kT}\right) \quad (3)$$

$$n_0 = n_i \exp\left(\frac{-q \cdot \phi_F}{kT}\right) \quad (4)$$

- on a neutralité électrique:

$$Nd - Na = n_0 - p_0 \quad (5)$$

Principe de calcul

- on exprime la densité de charges $\rho(y)$.

- on applique l'équation de Poisson

$$\frac{d^2\psi}{dy^2} = -\frac{\rho}{\epsilon_{Si}} \quad (6)$$

- on intègre pour obtenir le champ électrique

- on utilise le théorème de Gauss pour obtenir la charge dans le semi-conducteur

$$\int \vec{E} \cdot d\vec{S} = \frac{Q_{SC}}{\epsilon_{Si}} \quad (7)$$

Calcul

- Densité de charges:

$$\rho(y) = q(p(y) - n(y) + N_a - N_d) \quad (8)$$

- Application de l'équation de Poisson (6)

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\epsilon_{Si}} (N_d - N_a + p(y) - n(y)) \quad (9)$$

- On remplace $N_d - N_a$ par $n_0 - p_0$ (5) et $p(y)$ et $n(y)$ par leurs valeurs définies en (1) et (2):

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\epsilon_{Si}} \left[n_0 - p_0 + p_0 \cdot \exp\left(\frac{-\psi \cdot q}{k \cdot T}\right) - n_0 \cdot \exp\left(\frac{\psi \cdot q}{k \cdot T}\right) \right] \quad (10)$$

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\epsilon_{Si}} \left[p_0 \left(\exp\left(\frac{-\psi \cdot q}{k \cdot T}\right) - 1 \right) - n_0 \left(\exp\left(\frac{\psi \cdot q}{k \cdot T}\right) - 1 \right) \right]$$

- Un petit peu de dérivation sauvage:

$$\frac{d}{dy} \left(\frac{d\psi}{dy} \right)^2 = 2 \frac{d\psi}{dy} \frac{d^2\psi}{dy^2}$$

$$\Leftrightarrow \int d \left(\frac{d\psi}{dy} \right)^2 = \int 2 \frac{d\psi}{dy} \frac{d^2\psi}{dy^2} dy \quad (11)$$

$$\Leftrightarrow \left(\frac{d\psi}{dy} \right)^2 = \int 2 \frac{d^2\psi}{dy^2} d\psi$$

- L'intégrale se fait du substrat à la surface, c'est-à-dire de $\psi=0$ à $\psi=\psi$

$$\begin{aligned}
\left(\frac{d\psi}{dy}\right)^2 &= \int_0^\psi 2 \frac{d^2\psi}{dy^2} d\psi \\
\Leftrightarrow \left(\frac{d\psi}{dy}\right)^2 &= \int_0^\psi -\frac{2q}{\epsilon_{Si}} \left[p_0 \left(\exp^{\frac{-\psi \cdot q}{k \cdot T}} - 1 \right) - n_0 \left(\exp^{\frac{\psi \cdot q}{k \cdot T}} - 1 \right) \right] d\psi \\
\Leftrightarrow \left(\frac{d\psi}{dy}\right)^2 &= -\frac{2q}{\epsilon_{Si}} \left[p_0 \left(-\frac{k \cdot T}{q} \exp^{\frac{-\psi \cdot q}{k \cdot T}} - \psi \right) - n_0 \left(\frac{k \cdot T}{q} \exp^{\frac{\psi \cdot q}{k \cdot T}} - \psi \right) \right]_0^\psi \\
\Leftrightarrow \left(\frac{d\psi}{dy}\right)^2 &= -\frac{2q}{\epsilon_{Si}} \left[p_0 \left(-\frac{k \cdot T}{q} \exp^{\frac{-\psi \cdot q}{k \cdot T}} - \psi + \frac{k \cdot T}{q} \right) - n_0 \left(\frac{k \cdot T}{q} \exp^{\frac{\psi \cdot q}{k \cdot T}} - \psi - \frac{k \cdot T}{q} \right) \right] \\
\Leftrightarrow \left(\frac{d\psi}{dy}\right)^2 &= \frac{2q}{\epsilon_{Si}} \frac{k \cdot T}{q} p_0 \left[\exp^{\frac{-\psi \cdot q}{k \cdot T}} + \frac{\psi \cdot q}{kT} - 1 + \frac{n_0}{p_0} \left(\exp^{\frac{\psi \cdot q}{k \cdot T}} - \frac{\psi \cdot q}{kT} - 1 \right) \right] \tag{12}
\end{aligned}$$

- on peut remplacer le rapport $\frac{n_0}{p_0}$ avec les expressions (3) et (4):

$$\begin{aligned}
\left(\frac{d\psi}{dy}\right)^2 &= \frac{2q}{\epsilon_{Si}} \frac{k \cdot T}{q} p_0 \left[\exp^{\frac{-\psi \cdot q}{k \cdot T}} + \frac{\psi \cdot q}{kT} - 1 + \exp^{-\frac{2 \cdot q \cdot \phi_F}{kT}} \left(\exp^{\frac{\psi \cdot q}{k \cdot T}} - \frac{\psi \cdot q}{kT} - 1 \right) \right] \\
\Leftrightarrow \left(\frac{d\psi}{dy}\right)^2 &= \frac{2q}{\epsilon_{Si}} \frac{k \cdot T}{q} p_0 \left[\exp^{\frac{-\psi \cdot q}{k \cdot T}} + \frac{\psi \cdot q}{kT} - 1 + \exp^{\frac{q}{kT} (\psi - 2 \cdot \phi_F)} - \exp^{-\frac{2 \cdot q \cdot \phi_F}{kT}} \left(\frac{\psi \cdot q}{kT} + 1 \right) \right] \tag{13}
\end{aligned}$$

- le champ électrique vaut

$$\begin{aligned}
E &= -\frac{d\psi}{dy} \\
\Rightarrow E &= \pm \sqrt{\frac{2q}{\epsilon_{Si}} \frac{k \cdot T}{q} p_0 \left[\exp^{\frac{-\psi \cdot q}{k \cdot T}} + \frac{\psi \cdot q}{kT} - 1 + \exp^{\frac{q}{kT} (\psi - 2 \cdot \phi_F)} - \exp^{-\frac{2 \cdot q \cdot \phi_F}{kT}} \left(\frac{\psi \cdot q}{kT} + 1 \right) \right]} \tag{14}
\end{aligned}$$

Le champ électrique est dirigé des potentiels décroissants. Par conséquent, le champ électrique est positif si et seulement si $\psi_S > 0$, soit

$$\text{Sign}(E) = \text{Sign}(\psi_S) \tag{15}$$

- Application du théorème de Gauss (7). On considère un cylindre d'axe Y. Le champ étant dirigé selon y, le flux latéral est nul. Dans le volume, le champ est nul. Par conséquent la charge est égale au flux à travers l'interface, soit

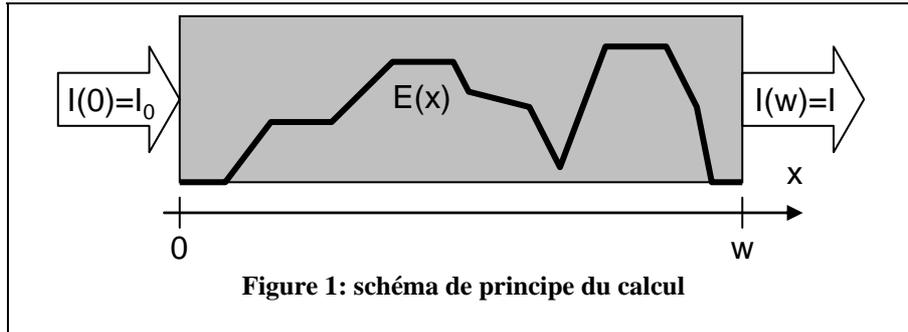
$$Q_{SC} = \epsilon_{si} \cdot E_s \quad (16)$$

$$\begin{aligned} Q_{SC} &= \text{Sign}(\psi_s) \epsilon_{si} \sqrt{\frac{2q}{\epsilon_{si}} \frac{k.T}{q} p_0} \left[\exp^{\frac{-\psi_s \cdot q}{k.T} + \frac{\psi_s \cdot q}{k.T} - 1} + \exp^{\frac{q}{k.T} (\psi_s - 2 \cdot \phi_F)} - \exp^{-\frac{2 \cdot q \cdot \phi_F}{k.T} \left(\frac{\psi_s \cdot q}{k.T} + 1 \right)} \right] \\ \Leftrightarrow Q_{SC} &= \text{Sign}(\psi_s) \sqrt{2kT \epsilon_{si} p_0} \left[\exp^{\frac{-\psi_s \cdot q}{k.T} + \frac{\psi_s \cdot q}{k.T} - 1} + \exp^{\frac{q}{k.T} (\psi_s - 2 \cdot \phi_F)} - \exp^{-\frac{2 \cdot q \cdot \phi_F}{k.T} \left(\frac{\psi_s \cdot q}{k.T} + 1 \right)} \right] \quad (17) \\ \Leftrightarrow Q_{SC} &= \text{Sign}(\psi_s) \psi_s \sqrt{2q \epsilon_{si} N_a} \left[\frac{u_t}{\psi_s} \exp^{u_t} + 1 - \frac{u_t}{\psi_s} + \frac{u_t}{\psi_s} \exp^{\frac{\psi_s - 2 \cdot \phi_F}{u_t}} - \exp^{-\frac{2 \cdot \phi_F}{u_t} \left(1 + \frac{u_t}{\psi_s} \right)} \right] \end{aligned}$$

Annexe 2 - Calcul du coefficient de multiplication

Lorsque de le champ électrique est suffisant dans une structure, des porteurs sont générés par ionisation par impact. Si la quantité de porteurs générés est suffisamment faible, on peut considérer l'ionisation comme un perturbation et calculer l'augmentation de courant qui lui est imputable.

Considérons une structure 1D de longueur w . Le courant $I(x)$ qui la traverse vaut $I(0)=I_0$ en entrant dans la structure, et $I(w)$ en sortant (**figure 1**). Si le champ électrique est suffisant pour créer des paires électron-trou par ionisation par impact, alors le courant va augmenter avec la distance le long de la structure.



L'incrément de courant dI_p du courant de trous I_p acquis sur la distance dx est:

$$dI_p = I_p \cdot \alpha_p \cdot dx + I_n \cdot \alpha_n \cdot dx$$

$$dI_p = I_p \cdot \alpha_p dx + (I - I_p) \cdot \alpha_n \cdot dx$$

$$dI_p = I_p \cdot (\alpha_p - \alpha_n) \cdot dx + I \cdot \alpha_n \cdot dx$$

$$\frac{dI_p}{dx} + I_p \cdot (\alpha_n - \alpha_p) = I \cdot \alpha_n$$

où $I_n(x)$ est le courant d'électrons, I le courant total ($I=I_n(x)+I_p(x)$), $\alpha_p(x)$ le coefficient de génération des trous et $\alpha_n(x)$ le coefficient génération des électrons. Cette expression est de la

forme $y'+P \cdot y=Q$ avec $\begin{cases} P=\alpha_n - \alpha_p \\ Q=I\alpha_n \end{cases}$.

$$\Rightarrow \begin{cases} I_p(x) = \frac{1}{M(x)} \cdot \left[\int_0^x I \cdot \alpha_n \cdot M(z) \cdot dz + C \right] \\ M(x) = \exp\left(\int_0^x (\alpha_n - \alpha_p) dx'\right) \end{cases}$$

Il faut à présent déterminer la constante C . Pour cela, on peut supposer que le courant de trous $I_p(w)$ qui sort de la structure est proportionnel au courant de trous $I_p(0)$ qui y rentre:

$$I_p(w) = M_p \cdot I_p(0)$$

$$I_p(0) = \frac{1}{M(0)} \cdot [0+C]$$

$$I_p(0) = \frac{1}{\exp\left(\int_0^0 (\alpha_n - \alpha_p) dx'\right)} \cdot [0+C]$$

$$I_p(0) = \frac{C}{\exp(0)} = C$$

$$\Rightarrow C = I_{p0} = \frac{I}{M_p}$$

En réinjectant la valeur de la constante:

$$\Rightarrow I_p(x) = \frac{1}{M(x)} \cdot \left[\int_0^x I \cdot \alpha_n \cdot M(z) \cdot dz + \frac{I}{M_p} \right]$$

$$\Leftrightarrow I_p(x) = \frac{I}{M(x)} \cdot \left[\frac{1}{M_p} + \int_0^x \alpha_n \cdot M(z) \cdot dz \right]$$

$$\Leftrightarrow \frac{I_p(x) \cdot M(x)}{I} = \frac{1}{M_p} + \int_0^x \alpha_n \cdot M(z) \cdot dz$$

$$\Leftrightarrow \frac{I_p(x) \cdot M(x)}{I} = \frac{1}{M_p} + \int_0^x (\alpha_n - \alpha_p + \alpha_p) \cdot M(z) \cdot dz$$

$$\Leftrightarrow \frac{I_p(x) \cdot M(x)}{I} = \frac{1}{M_p} + \int_0^x (\alpha_n - \alpha_p) \cdot M(z) \cdot dz + \int_0^x \alpha_p \cdot M(z) \cdot dz$$

$$D = \int_0^x (\alpha_n - \alpha_p) \cdot \exp\left(\int_0^{x'} (\alpha_n - \alpha_p) dz\right) dx'$$

$$u' \equiv \alpha_n - \alpha_p \Rightarrow u = \int \alpha_n - \alpha_p$$

$$D = \int_0^x u' \cdot \exp(u) dx'$$

$$\Leftrightarrow D = [\exp(u(x'))]_0^x$$

$$\Leftrightarrow D = \exp(u(x)) - \exp(u(0))$$

Or $u(0) = 0$

$$\Leftrightarrow D = \exp\left(\int_0^x (\alpha_n - \alpha_p) dx'\right) - 1$$

$$\Leftrightarrow D = \exp\left(-\int_0^x (\alpha_p - \alpha_n) dx'\right) - 1$$

$$\Leftrightarrow D = M(x) - 1$$

$$\Rightarrow \frac{I_p(x) \cdot M(x)}{I} = \frac{1}{M_p} + M(x) - 1 + \int_0^x \alpha_p \cdot M(z) \cdot dz$$

Par définition $I_p(w) = I$

$$\Leftrightarrow \frac{I \cdot M(w)}{I} = \frac{1}{M_p} + M(w) - 1 + \int_0^w \alpha_p \cdot M(z) \cdot dz$$

$$\Leftrightarrow 1 - \frac{1}{M_p} = \int_0^w \alpha_p \cdot M(z) \cdot dz$$

$$1 - \frac{1}{M_p} = \int_0^w \alpha_p \cdot \exp\left(-\int_0^x (\alpha_p - \alpha_n) \cdot dx'\right) dx$$

Le calcul peut être mené de manière symétrique sur le courant d'électrons. L'expression finale est alors

$$1 - \frac{1}{M_p} = \int_0^w \alpha_n \cdot \exp\left(-\int_0^x (\alpha_n - \alpha_p) \cdot dx'\right) dx$$

Ces deux expressions sont mathématiquement équivalentes.

Annexe 3 - Modélisation d'une diode p-i-n

La modélisation d'une diode p-i-n consiste à déterminer les profils de porteurs, de champ électrique et de potentiel dans la structure.

Profil de porteurs

Le profil de charges est supposé ne dépendre que des profils de dopage. En particulier l'influence de la tension appliquée est négligée et la génération de porteurs par ionisation par impact est supposée ne pas perturber ce profil de charges.

Nous considérons une structure idéale:

- la zone i est faiblement dopée de type p. En effet, les substrats les plus purs présentent 10^{15} défauts par cm^3 . Ce niveau de dopage est électriquement équivalent à du silicium intrinsèque. Cependant, lors des simulations numériques, l'utilisation d'un dopage à 10^{15}cm^{-3} améliore la convergence.

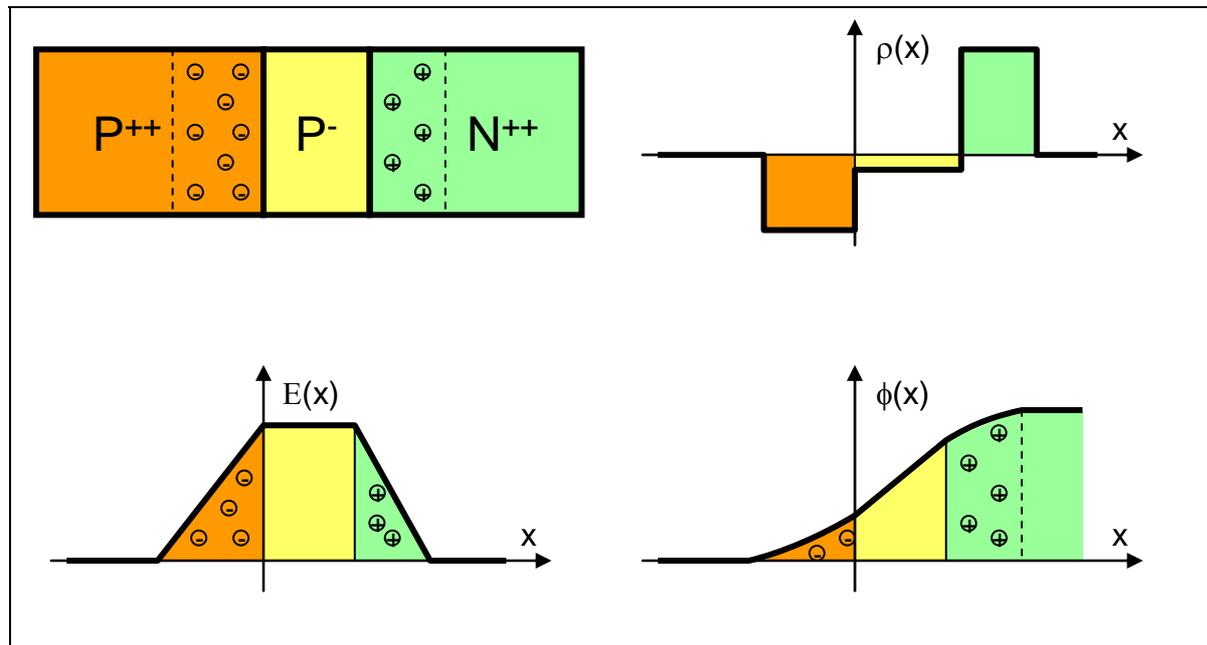
- les zones n et p sont uniformément dopées

- la transition entre zone dopée et zone non dopée est infiniment abrupte. Cette hypothèse de jonction abrupte est tout à fait valable dans le cas de jonctions implantées ou épitaxiées.

Des zones de charge d'espace (ZCE) se créent aux jonctions p-i et i-n. Nous utiliserons les hypothèses usuelles:

- la charge est nulle hors de la zone intrinsèque et des ZCE (neutralité électrique)

- la charge est égale au dopage dans les ZCE et la zone intrinsèque



Intégration de l'équation de Poisson

$$w_p \leq x \leq 0$$

Hypothèse:

$$\rho(x) = -qN_a$$

Poisson:

$$\frac{d^2V}{dx^2} = \frac{qN_a}{\epsilon}$$

Intégration: $\frac{dV}{dx} = \frac{qN_a}{\epsilon}(x + k_1)$

CL: continuité de E en $x=-w_p$: $\frac{dV}{dx}\Big|_{x=-w_p} = 0 = \frac{qN_a}{\epsilon}(-w_p + k_1) \Rightarrow k_1 = w_p$

Remplacement de k_1 : $\frac{dV}{dx} = \frac{qN_a}{\epsilon}(x + w_p)$

Intégration: $V(x) = \frac{qN_a}{2\epsilon}(x + w_p)^2 + k_2$

CL: continuité de V en $x=-w_p$: $V\Big|_{x=-w_p} = V_p \Leftrightarrow V(0) = k_2 = V_p$

Remplacement de k_2 : $V(x) = \frac{qN_a}{2\epsilon}(x + w_p)^2 + V_p$

$$0 \leq x \leq L_i$$

Hypothèse: $\rho(x) = -qN_i$

Poisson: $\frac{d^2V}{dx^2} = \frac{qN_i}{\epsilon}$

Intégration: $\frac{dV}{dx} = \frac{qN_i}{\epsilon}x + k_3$

CL: continuité de E en $x=0$: $\frac{dV}{dx}\Big|_{x=0} = \frac{qN_a w_p}{\epsilon} = k_3$

Remplacement de k_3 : $\frac{dV}{dx} = \frac{qN_i}{\epsilon}x + \frac{qN_a w_p}{\epsilon}$

Intégration: $V(x) = \frac{qN_i}{2\epsilon}x^2 + \frac{qN_a w_p}{\epsilon}x + k_4$

CL: continuité de V en $x=0$: $V\Big|_{x=0} = \frac{qN_a w_p^2}{2\epsilon} + V_p = k_4$

Remplacement de k_4 : $V(x) = \frac{qN_i}{2\epsilon}x^2 + \frac{qN_a w_p}{2\epsilon}(2x + w_p) + V_p$

$$L_i \leq x \leq L_i + w_n$$

Hypothèse: $\rho(x) = qN_d$

Poisson: $\frac{d^2V}{dx^2} = -\frac{qN_d}{\epsilon}$

Intégration: $\frac{dV}{dx} = -\frac{qN_d}{\epsilon}(x + k_5)$

CL: continuité de E en $x=L_i+w_n$: $\frac{dV}{dx}\Big|_{x=L_i+w_n} = 0 \Rightarrow k_5 = -L_i - w_n$

Remplacement de k_5 : $\frac{dV}{dx} = -\frac{qN_d}{\epsilon}(x - L_i - w_n)$

Intégration: $V(x) = -\frac{qN_d}{2\epsilon}(x - L_i - w_n)^2 + k_6$

Hypothèse: $V|_{x=L_i+w_n} = V_n = k_6$

Remplacement de k_6 : $V(x) = -\frac{qN_d}{2\epsilon}(x - L_i - w_n)^2 + V_n$

Calcul de w_n, w_p

CL: continuité de E en $x=L_i$: $\frac{dV}{dx}|_{x=L_i} = -\frac{qN_d}{\epsilon}(-w_n) = \frac{qN_a w_p}{\epsilon} + \frac{qN_i L_i}{\epsilon}$

Moralité: $N_d w_n = N_a w_p + N_i L_i$

CL: continuité de V en $x=L_i$: $\frac{qN_i L_i^2}{2\epsilon} + \frac{qN_a w_p}{2\epsilon}(2L_i + w_p) + V_p = -\frac{qN_d w_n^2}{2\epsilon} + V_n$

$$w_n^2 = \frac{N_a^2 w_p^2 + N_i^2 L_i^2 + 2N_a N_i w_p L_i}{N_d^2}$$

$$\frac{qN_i L_i^2}{2\epsilon} + \frac{2qN_a L_i}{2\epsilon} w_p + \frac{qN_a w_p^2}{2\epsilon} + V_p = -\frac{qN_d}{2\epsilon} \frac{N_a^2 w_p^2}{N_d^2} - \frac{qN_d}{2\epsilon} \frac{N_i^2 L_i^2}{N_d^2} - \frac{qN_d}{2\epsilon} \frac{2N_a N_i w_p L_i}{N_d^2} + V_n$$

Equation en w_p^2 :

$$\frac{qN_a}{2\epsilon} \left(1 + \frac{N_a}{N_d}\right) w_p^2 + \frac{qN_a L_i}{\epsilon} \left(1 + \frac{N_i}{N_d}\right) w_p + \frac{qN_i L_i^2}{2\epsilon} \left(1 + \frac{N_i}{N_d}\right) + \underbrace{(V_p - V_n)}_{=-V_d} = 0$$

$ax^2 + bx + c = 0$ avec:

$$\begin{cases} a = \frac{qN_a}{2\epsilon} \left(1 + \frac{N_a}{N_d}\right) \\ b = \frac{qN_a L_i}{\epsilon} \left(1 + \frac{N_i}{N_d}\right) \\ c = \frac{qN_i L_i^2}{2\epsilon} \left(1 + \frac{N_i}{N_d}\right) + (V_p - V_n) = -V_d + \frac{qN_i L_i^2}{2\epsilon} \left(1 + \frac{N_i}{N_d}\right) \end{cases}$$

déterminant:

$$\Delta = b^2 - 4ac = \left(\frac{qN_a L_i}{2\epsilon}\right)^2 4 \left(1 + \frac{N_i}{N_d}\right)^2 - 4 \frac{qN_a}{2\epsilon} \left(1 + \frac{N_a}{N_d}\right) \left[-V_d + \frac{qN_i L_i^2}{2\epsilon} \left(1 + \frac{N_i}{N_d}\right)\right]$$

expression: $w_p = \frac{-b + \sqrt{\Delta}}{2a}$

$$w_p = \frac{-\frac{qN_a L_i}{\epsilon} \left(1 + \frac{N_i}{N_d}\right) + \sqrt{4 \left(\frac{qN_a L_i}{2\epsilon}\right)^2 \left(1 + \frac{N_i}{N_d}\right)^2 + 4 \frac{qN_a}{2\epsilon} \left(1 + \frac{N_a}{N_d}\right) \left[V_d - \frac{qN_i L_i^2}{2\epsilon} \left(1 + \frac{N_i}{N_d}\right)\right]}{2 \frac{qN_a}{2\epsilon} \left(1 + \frac{N_a}{N_d}\right)}$$

suppression de $\frac{qN_a}{\epsilon}$:

$$w_p = \frac{-L_i \left(1 + \frac{N_i}{N_d}\right) + \sqrt{L_i^2 \left(1 + \frac{N_i}{N_d}\right)^2 + \frac{2\varepsilon}{qN_a} \left(1 + \frac{N_a}{N_d}\right) \left[V_d - \frac{qN_i L_i^2}{2\varepsilon} \left(1 + \frac{N_i}{N_d}\right)\right]}{1 + \frac{N_a}{N_d}}$$

simplification: $N_i \ll N_a, N_d$

$$w_p = \frac{-L_i \left(1 + \frac{N_i}{N_d}\right) + \sqrt{L_i^2 \left(1 + \frac{N_i}{N_d}\right)^2 + \frac{2\varepsilon}{qN_a} \left(1 + \frac{N_a}{N_d}\right) \left[V_d - \frac{qN_i L_i^2}{2\varepsilon} \left(1 + \frac{N_i}{N_d}\right)\right]}{1 + \frac{N_a}{N_d}}$$

$$w_p = \frac{-L_i + \sqrt{L_i^2 + \frac{2\varepsilon}{qN_a} \left(1 + \frac{N_a}{N_d}\right) \left(V_d - \frac{qN_i L_i^2}{2\varepsilon}\right)}}{1 + \frac{N_a}{N_d}}$$

simplification:

$$w_p = \frac{-L_i + \sqrt{L_i^2 + \frac{2\varepsilon}{q} \left(V_d - \frac{qN_i L_i^2}{2\varepsilon}\right) \left(\frac{1}{N_a} + \frac{1}{N_d}\right)}}{1 + \frac{N_a}{N_d}}$$

$$w_n = \frac{N_a}{N_d} w_p + \frac{N_i}{N_d} L_i \quad w_n = \frac{-L_i + \sqrt{L_i^2 + \frac{2\varepsilon}{q} \left(V_d - \frac{qN_i L_i^2}{2\varepsilon}\right) \left(\frac{1}{N_a} + \frac{1}{N_d}\right)}}{1 + \frac{N_a}{N_d}} + \frac{N_i}{N_d} L_i$$

Calcul de V_n, V_p

Ces potentiels dépendent de la façon dont est modélisée le contact. Le but est de pouvoir reproduire sous Excel ou Mathcad le profil de potentiel, et de pouvoir le comparer à des simulations. Par conséquent, la façon de calculer ces potentiels se calque sur celle utilisée dans les simulations. La documentation utilisée est celle de Dessis.

Contact ohmique

On suppose l'équilibre et la neutralité des charges:

$$n_o - p_o = N_d - N_a \quad n_o \cdot p_o = n_{i,eff}$$

Dans le cas de la statistique de Boltzman, le système d'équations peut être résolu:

$$V = V_{app} + \frac{kT}{q} \operatorname{asinh} \left(\frac{N_d - N_a}{2n_{i,eff}} \right)$$

Dans notre structure, la cathode est placée à la masse et l'anode est portée au potentiel V_d . On déduit donc les expressions de V_n et V_p :

$$V_n = V_d + \frac{kT}{q} \operatorname{asinh} \left(\frac{N_d}{2n_{i,eff}} \right) \quad V_p = \frac{kT}{q} \operatorname{asinh} \left(\frac{-N_a}{2n_{i,eff}} \right)$$

Calcul de la densité de porteurs intrinsèque $n_{i,eff}$

A fort dopage ($>10^{18} \text{cm}^{-3}$), la structure de bandes du semi-conducteur se modifie. Des paramètres fondamentaux tels que la valeur du gap E_g ou la densité de porteurs intrinsèques n_i change. On cherche ici à déterminer le potentiel induit par un contact ohmique sur une zone fortement dopée. Il est donc nécessaire de prendre en compte ces effets.

La densité de porteurs intrinsèques dépend de la valeur du gap selon l'expression:

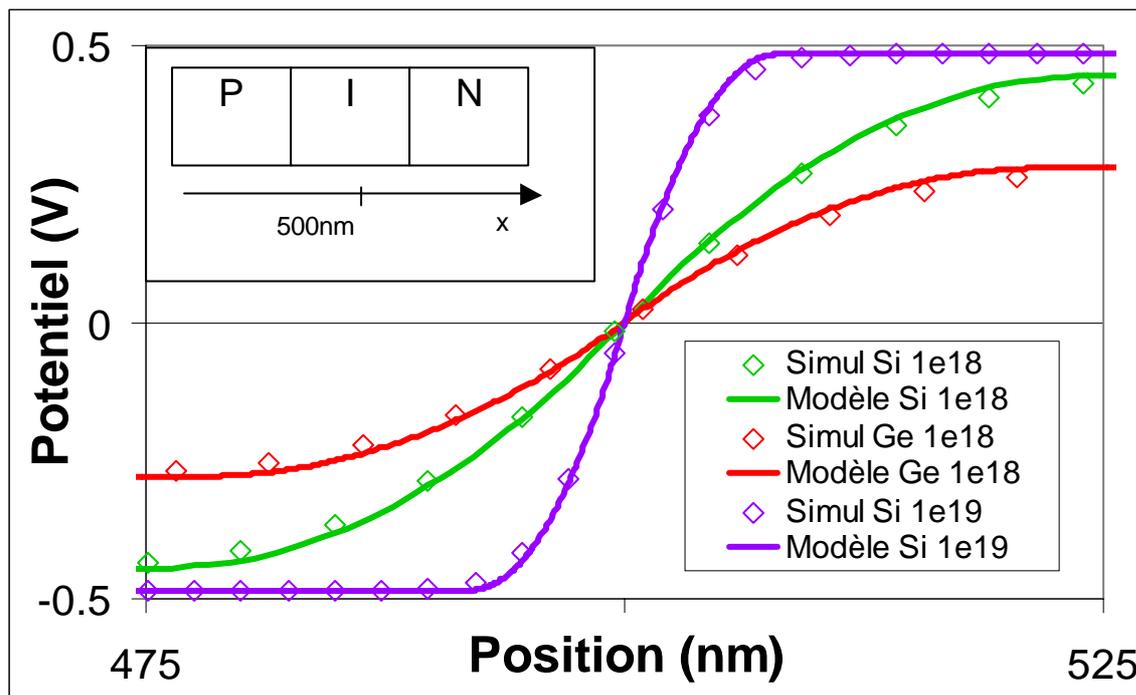
$$n_{i,eff} = n_i \exp\left(\frac{-\Delta E_g}{2kT}\right)$$

où ΔE_g est la variation de gap due au dopage. Le modèle de variation utilisé est celui proposé par Slotboom et implémenté sous Dessis par le mot-clef *OldSlotBoom*. Il suggère la formulation empirique:

$$\Delta E_g(N_A) = E_{bgn} \left[\ln\left(\frac{N_A}{N_{ref}}\right) + \sqrt{\left(\ln\left(\frac{N_A}{N_{ref}}\right)\right)^2 + 0.5} \right]$$

avec $E_{bgn} = 9.10^{-3} \text{eV}$ et $N_{ref} = 1.10^{17} \text{cm}^{-3}$.

Comparaison avec la simulation



Le modèle a été implémenté sous Excel, et les résultats ont été comparés avec la simulation pour tout un jeu de paramètres:

- matériau
- dopages
- dimension de la longueur intrinsèque
- tensions

Dans tous les cas, le modèle reproduit avec précision les résultats obtenus par simulation. Seuls trois exemples sont présentés ici à tension nulle et longueur intrinsèque nulle.

Bibliographie

ISE Dessis manual et références incluses:

J. W. Slotboom and H. C. de Graaff, "Measurements of Bandgap Narrowing in Si Bipolar Transistors," *Solid State Electron.*, vol. 19, pp. 857–862, 1976.

J. W. Slotboom and H. C. de Graaff, "Bandgap Narrowing in Silicon Bipolar Transistors," *IEEE Trans. on Electron Devices*, vol. ED-24, no. 8, pp. 1123–1125, 1977.

J. W. Slotboom, "The pn-Product in Silicon," *Solid State Electron.*, vol. 20, pp. 279–283, 1977.