

THÈSE

pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE PROVENCE

présentée et soutenue publiquement par
Pascal LEMOIGNE

Équipes d'accueil :
IM2NP et STMicroelectronics

Titre :

Simulation de la variabilité du transistor MOS

Soutenue le 1^{er} décembre 2011 devant le jury composé de :

Rachid Bouchakour, HDR	Président	CNRS/IM2NP, Marseille
Gérard Ghibaudo, HDR	Rapporteur	CNRS/IMEP, Grenoble
Pascal Masson, HDR	Rapporteur	CNRS/LEAT, Nice
Lakhdar Zaïd, HDR	Directeur	CNRS/IM2NP, Marseille
André Juge, Docteur	Directeur	STMicroelectronics, Crolles
Christian Dufaza, HDR	Invité	CNRS/IM2NP, Marseille

Merci à vous !

Je tiens à remercier Jean-Daniel Arnould, mon directeur de stage de M2R pour m'avoir incité à poursuivre en thèse et à « ne pas se poser de limites. »

Merci à Gérard Ghibaudo et Pascal Masson de m'avoir fait l'honneur d'accepter d'être rapporteurs de cette thèse malgré leurs agenda déjà plein.

Je veux témoigner ma gratitude à mes premiers directeurs de thèse Christian Dufaza et Denis Rideau, ainsi qu'à Hervé Jaouen et Clément Tavernier qui m'ont accordé leur confiance en m'accueillant dans leurs équipes.

Ces travaux n'auraient pas aboutis sans mes actuels directeurs de thèse André Juge et Lakhdar Zaïd qui ont accepté de reprendre l'encadrement de cette thèse à mi-parcours à la suite de réorganisations à Crolles comme à Marseille, et qui m'ont soutenu et encouragé jusqu'au bout (et au delà!). Encore merci.

Merci à Albane Sangiovanni pour les corrections de ce manuscrit.

Je tiens à remercier l'équipe de Crolles avec qui j'ai énormément appris et avec qui j'ai passé trois très bonnes années : Lorenzo, Adi, Clément, Denis, Blandine, Xavier, Thierry, Pierre, François, Olivier, Vincent, Florian, Daniel, David, Grégory. . . Merci aux thésards pour les bons moments et la bonne ambiance (et quelques soirées mémorables ;)) Maxime, Vincent, Marc, Mehdi, Sébastien, Alban, Andres, Romuald, Carine, Jean et Giancarlo.

Merci à l'équipe de *Process Integration* : Maud Bideau, Laurence Boissonnet, Emmanuel Josse et Dominique Golanski auprès de qui j'ai eu un aperçu de ce qui se passait en dehors du monde de la simulation et qui m'ont fourni l'ensemble des mesures utilisées dans cette thèse. Merci à Alexander Burenkov pour notre collaboration dans le cadre du projet ATOMICS.

Un merci à mes amis pour tout ce que nous avons vécu ensemble et pour m'avoir supporté pendant cette époque : Audrey et Philippe, Romain et Émilie, Caro et Simon, Pierre et Valéria, Jojo et Séverine, Be, Jack, Pierrot et Olivia, Benoit et Nathalie, David, Cobo, Olivier et Pauline, Émilie, Rémy et Céline, Bao, Julien, Anaïs, Katia, Alex et Géo, Gaspard, Rom2.

Enfin je remercierai ma famille, sans qui je ne serais pas là, avec qui nous traversé des moments tristes et des moments heureux pendant cette intervalle

de temps. Merci pour vos encouragements.

Table des matières

Table des matières	vi
Introduction	vii
1 Contexte – Problématique	1
1.1 La variabilité et ses conséquences	4
1.2 Sources de déviation	7
1.2.1 Observations des effets	7
1.2.2 Sources intrinsèques	13
1.2.3 Sources originaires du dessin	21
1.2.4 Sources originaires du procédé	27
1.3 Modélisation des déviations	32
1.3.1 Variabilité du process	32
1.3.2 Variabilité du transistor	34
1.3.3 Variabilité des circuits	36
1.4 Conclusion	37
2 Caractérisation du dopage des dispositifs MOSFET	39
2.1 Caractérisation de la variabilité	39
2.2 Caractérisation de la température de recuit	41
2.2.1 Estimation de la variation spatiale de la température recue	42
2.3 Caractérisation électrique du profil latéral du dopage	45
2.3.1 Principe de la méthode	45
2.3.2 Application de la méthode sur des simulations TCAD d'un procédé industriel	48
2.3.3 Étude de l'impact de la variabilité du procédé à l'aide d'un plan d'expériences	50
2.3.4 Mesures sur des nMOS 65 nm	52
2.3.5 Conclusion	54
2.4 Caractérisation de la polarisation du substrat	54
2.4.1 L'effet substrat	54

2.4.2	Extraction des paramètres gamma et phif	55
2.4.3	Caractérisation de l'effet avec un dopage réel	56
2.4.4	Effet de la longueur de grille sur cette caractérisation	58
2.5	Détermination du dopage canal	59
2.5.1	Méthode	59
2.5.2	Simulations TCAD	60
2.5.3	Détermination du profil vertical à partir de l'effet sub- strat	60
2.6	Variabilité du profil de dopage vertical	61
2.6.1	Présentation des données expérimentales	61
2.6.2	Application de la méthode	62
2.6.3	Étude des corrélations	64
2.7	Conclusion	66
3	Étude expérimentale du nœud technologique 45nm	68
3.1	Les variables étudiées	69
3.1.1	Échantillonnage spatial	69
3.2	Observation des données	72
3.2.1	Les facteurs	72
3.2.2	Variabilité intra-plaque	74
3.2.3	Corrélations entre variables technologiques	79
3.2.4	Indépendance des expériences	80
3.3	Étude du plan d'expériences	83
3.3.1	Le plan expérimental	83
3.3.2	Régression multilinéaire d'un modèle polynomial	86
3.3.3	Estimation de la contribution de chaque facteur	93
3.4	Conclusions	103
4	Simulation des variations dans les modèles compacts	104
4.1	Simulations Monte-Carlo	105
4.1.1	Construction de la simulation Monte-Carlo	105
4.2	Génération de Corners	107
4.2.1	Méthode standard	108
4.2.2	Statut par rapport aux résultats Monte-Carlo	110
4.2.3	Génération spécifique à une performance	113
4.2.4	Détermination non paramétrique multidimensionnelle	114
4.3	Modification du modèle de variation	114
4.3.1	Résultats	117
4.4	Conclusion	121
	Conclusion générale	122

A Annexes	125
A.1 Figures du chapitre 4	125
A.2 Boite à moustaches	136
A.3 Extraits de code relatifs aux simulations Monte-Carlo	138
 Bibliographie	 152
 Liste des figures	 158
 Liste des tableaux	 159
 Publications	 160

Introduction

La constance des performances des produits est une priorité pour les fabricants de circuits intégrés. Ils s'engagent auprès de leurs clients à fournir des puces dont la dispersion des performances est très faible, ce qui implique que chaque circuit fabriqué qui n'entre pas dans les marges de tolérance est mis au rebut et donc fait monter le coût de production moyen par plaque de silicium. Il faut donc savoir contrôler la fabrication afin d'uniformiser les performances. Les produits de l'industrie de la micro-électronique, des circuits électroniques, sont fabriqués par des techniques organisées autour de la photolithographie sur des tranches de silicium cristallin. Une tranche peut contenir de quelques dizaines à des centaines de puces électroniques. Dans une usine, les tranches de silicium sont groupées par lots de vingt-cinq plaques pour subir certaines étapes du procédé de fabrication. Enfin, un fondeur¹ peut choisir de fabriquer un même produit dans plusieurs usines. Il y a donc différentes échelles auxquelles vont intervenir les dispersions de performances. Les composants de base des circuits intégrés sont soumis à différentes sources de fluctuations. Certaines sont dues aux matériaux utilisés. En effet, l'implantation volontaire d'impuretés dans le silicium afin de régler les propriétés des composants – le dopage – n'agit pas de manière continue dans tout matériau. C'est une action locale à l'échelle atomique, et plus les composants sont petits, moins les effets sont lissés. De la même manière, l'utilisation de silicium polycristallin introduit des effets locaux aux interfaces entre les cristaux le composant. L'emploi de résines en matériaux polymères entraîne une certaine imperfection du bord des gravures, dont l'importance croît avec la miniaturisation des dispositifs.

D'autres imperfections sont induites par les outils du procédé². Les étapes

1. c'est le nom qu'on donne aux fabricants de puces, car c'est historiquement l'industrie de la métallurgie qui possédait les outils nécessaires pour fondre le silicium afin d'en faire des cristaux.

2. ou *process* en anglais qui remplace souvent le mot français dans le jargon

du process comportent des procédés chimiques (gravure, retrait de résine), physiques (gravure, implantation d'ions, dépôt de matériau), thermiques (recuit en four, recuit laser, croissance d'oxyde) ou encore optiques (insolation des résines pour appliquer le dessin des circuits). Les machines réalisant ces étapes introduisent chacune des non uniformités qui leurs sont propres.

Le placement-routage du circuit ou la proximité d'un circuit avec différentes zones du circuit peut aussi influencer la manière dont le circuit subit le procédé. La variabilité est un sujet qui concerne toutes les spécialités.

En amont de la fabrication, les concepteurs de circuits utilisent des modèles pour prédire le comportement électrique des circuits qu'ils dessinent, et ces modèles doivent permettre de prédire la dispersion des performances. Avec ces modèles, ils s'attachent à concevoir des circuits dont les performances simulées prévoient que la production des circuits atteindra un certain seuil de rendement.

Il existe différentes méthodes de simulation de la variabilité, avec différents niveaux de complexité. Certaines s'appliquent aux simulations de circuits complexes, d'autres aux simulations de petits circuits standards, et celles auxquelles nous nous intéressons en particuliers sont les simulations du transistor métal-oxyde-semiconducteur *MOS*, composant de base des circuits intégrés.

Reproduire et prédire les variations font l'objet de recherche sur diverses méthodes de simulations. Certaines simulations se focalisent sur les sources dues aux matériaux afin de mieux comprendre les phénomènes mis en jeu. Ces résultats peuvent servir de points d'entrées pour des simulations de procédé de type éléments finis appelées *TCAD* et qui permettent de reproduire virtuellement les recettes de fabrication et de prédire les effets au niveau des performances électriques. Du côté des simulations de circuits, l'utilisation de méthodes Monte-Carlo et de pire-cas reste la référence. La méthode Monte-Carlo permet de propager sans simplification la variabilité des paramètres technologiques et la méthode pire-cas fournit aux concepteurs un critère de sanction des performances raisonnable en termes de temps de calcul, contrairement à la méthode Monte-Carlo lorsqu'elle est appliquée à un circuit de plusieurs millions de transistors.

Ce manuscrit commence par une introduction et un état de l'art des études faites sur la variabilité du transistor MOS et les moyens pour la simuler dans le chapitre 1 [Contexte – Problématique](#).

Dans un premier temps, il est important de connaître expérimentalement

les fluctuations des facteurs technologiques. Le dopage du canal peut être mesuré par spectroscopie d'ions secondaires émis après un bombardement d'ions, mais pas sans détruire le dispositif. Cette caractérisation lourde en préparation ne permet pas un échantillonnage statistique. Dans le chapitre 2 intitulé [Caractérisation du dopage des dispositifs MOSFET](#), nous nous intéressons au facteur de variabilité qu'est le dopage du semiconducteur. Nous proposons une méthode d'estimation du dopage canal en se basant sur des mesures électriques. Nous présentons également une méthode de calibration du dopage latéral des simulations TCAD Ainsi qu'un moyen d'obtenir une image de la dispersion des recuits thermiques sur la surface d'une plaque.

Le chapitre 3 traite de l'[Étude expérimentale du nœud technologique 45nm](#). Cette étude expérimentale du nœud technologique 45 nanomètres nous permet de déterminer et confirmer la liste des facteurs qui participent principalement à la variabilité des performances du transistor et nous renseigne sur les valeurs de dispersion auxquelles nous avons à faire.

Pour finir, dans le chapitre 4 intitulé [Simulation des variations dans les modèles compacts](#), nous proposons d'améliorer la prise en compte des déviations des facteurs process dans les simulations Monte-Carlo et pire-cas appliquées aux modèles compacts au regard d'observations expérimentales.

Chapitre 1

Contexte – Problématique



FIGURE 1.1 – Réplique du premier transistor réalisé aux Bell Labs.

La fabrication du transistor MOS à grande échelle et à coût réduit a été rendue possible par une série d'avancées majeures dans le milieu du xx^e siècle. Après l'invention du transistor¹ en 1947 par John Bardeen, Walter Houser Brattain et William Shockley aux Bell Laboratories [1], Gordon Teal fabrique un premier transistor en silicium en 1954 chez Texas Instruments. Le premier MOS-FET a été réalisé par Kahng [2] et Atalla [3] en 1960 au Bell Labs [4] et enfin Jean Hoerni développa en 1962 chez Fairchild Semiconductor des procédés planaires que nous utilisons encore aujourd'hui pour la fabrication du dispositif MOS.

Le succès du transistor MOS tient à quelques éléments clefs : Le procédé de lithographie qui permet de fabriquer les composants comme les interconnexions, la possibilité d'utiliser l'oxyde de silicium comme isolant, l'auto-alignement du transistor autour de sa grille et l'association de dispositifs complémentaires dans les circuits logiques.

Depuis, la course à la miniaturisation, pour augmenter le nombre de puces

1. *Transistor* est un mot-valise pour *transfer resistor* :

« The way I provided the name, was to think of what the device did. And at that time, it was supposed to be the dual of the vacuum tube. The vacuum tube had transconductance, so the transistor would have *transresistance*. And the name should fit in with the names of other devices, such as *varistor* and *thermistor*. And... I suggested the name *transistor*. » — John R. Pierce

par surface de tranche de silicium et donc augmenter la rentabilité des produits, a suivi la fameuse « loi », énoncé par Gordon Moore, co-fondateur de Intel, en 1965 dans *Electronics magazine* qui prévoit que le nombre de transistors par puce va doubler tous les deux ans. Cette augmentation passe par la diminution des dimensions de ces composants. De dimensions de l'ordre du micromètre dans les années 70 [5], les dimensions nominales des récents microprocesseurs atteignent 32 nm, le noeud technologique 22 nm étant prévu pour 2013 selon l'ITRS [6].

La fabrication de circuits électroniques intégrés est une longue succession de procédés appliqués à une tranche de silicium monocristallin, ou *wafer*, sciée dans un lingot créé par la méthode Czochralski. La méthode Czochralski (CZ) qui est utilisée pour créer un cristal consiste à plonger un petit morceau de silicium monocristallin, la graine, à la surface de silicium en fusion² dans un creuset en rotation³. La graine, montée avec une orientation cristalline précise sur une tige en rotation dans le sens opposé, est mise en contact avec la surface du liquide. En refroidissant lentement, le silicium en fusion se solidifie en suivant la structure cristalline de la graine. Le cristal se forme au fur et à mesure de la remontée de la tige sur laquelle il est monté. Le diamètre du lingot est déterminé par le contrôle des vitesses de rotation, de tirage vers le haut et de la température. Les diamètres utilisés aujourd'hui sont 200 mm et 300 mm. L'utilisation de wafers de 450 mm est prévu pour 2012 [8]. Une fois que le lingot a fini de croître et qu'il a refroidi, il est scié en tranches de 200 μm à 800 μm . Les tranches sont ensuite arrondies sur leur bord et leur surface est polie par CMP (*Chemical Mechanical Polishing*) jusqu'à atteindre le niveau de planéité requis. Un certain nombre de méplats ou d'encoches sont faits sur le bord des plaques pour indiquer le type de dopage



FIGURE 1.2 – Un lingot de silicium. À l'extrémité supérieure se trouvait la graine, ou cristal initial à partir duquel a crû le cylindre entier.

2. La température de fusion du silicium est de 1414 °C.

3. Voir <http://youtu.be/aWVyhzuHnQ>, une vidéo montrant le procédé de fabrication des wafers (de 0 :59 à 4 :20) mais également la page d'explication du fabricant de wafers MEMC (en anglais) : [7]

et l'orientation du cristal. Le silicium à partir duquel est fait le wafer peut être initialement dopé, N ou P, à des niveaux de l'ordre de 10^{13} cm^{-3} à 10^{16} cm^{-3} .

Dans cette succession de procédés ou *process* qui est le terme anglais couramment employé, on trouve la fabrication des dispositifs de base tels que les transistors, les éléments passifs (capacités, inductances, résistances) et les connexions métalliques. On répartit tous ces dispositifs sur différentes couches empilées les unes après les autres sur le substrat de silicium, en fonction des procédés requis par chaque dispositif. Le nombre de niveaux utilisés dépend de la complexité du circuit, en particulier des interconnexions.

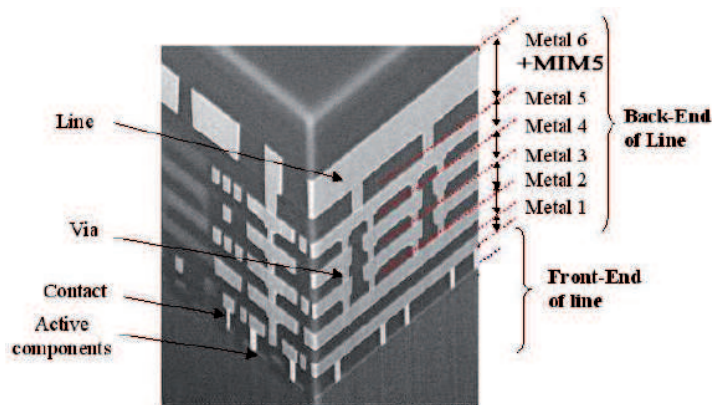


FIGURE 1.3 – Empilement des niveaux dans une technologie CMOS $0,25 \mu\text{m}$ avec 6 niveaux d'interconnexions en cuivre [9].

Les transistors et les capacités de type MOS sont fabriqués dans le premier niveau, et donc en premier à partir de la plaque de silicium initiale. On appelle ce niveau et l'ensemble de ce qui s'y trouve le *front-end of line*. Dans les niveaux suivants on trouvera les interconnexions, et parmi les niveaux d'interconnexions supérieurs on trouvera des composants passifs comme des inductances ou des capacités MIM par exemple. On appelle ces niveaux et l'ensemble de ce qui s'y trouve le *back-end of line*⁴. Les différents niveaux sont connectés entre eux par des *vias*. Ce sont des colonnes de métal. La figure 1.3 donne un aperçu de ces empilements dans une technologie CMOS $0,25 \mu\text{m}$

4. Back-end of line (BEOL) et front-end of line (FEOL) ne doivent pas être confondus avec back-end of chip et front-end of chip qui sont respectivement la mise en boîtier du circuit intégré et sa fabrication.

avec 6 niveaux d'interconnexions. Après la réalisation du circuit, la plaque est sciée pour séparer les puces (chip) qui vont être mises en boîtier. Nous nous intéresserons ici au *front-end of line*, là où sont réalisés les transistors.

1.1 La variabilité et ses conséquences

La miniaturisation du transistor se traduit par une diminution de la longueur de grille et de la largeur du composant. Ces réductions de dimensions impliquent la réduction de l'épaisseur d'oxyde de grille, l'augmentation des dopages et la diminution des tensions de fonctionnement.

Depuis le transistor simple tel qu'il a été inventé, il a fallu ajouter lors des passages à de nouveaux noeuds technologiques des caractéristiques techniques pour contrebalancer les effets qui apparaissent avec la miniaturisation. Ces étapes du procédé de fabrication supplémentaires sont, par exemple, un dopage anti-perçage, un dopage d'ajustement de la tension de seuil V_{th} , des régions *ldd* légèrement, accompagnées d'espaceurs pour repousser les implants source/drain. Ce sont autant de nouvelles sources de fluctuations des performances.

Dans une première approche, on peut voir que si une dimension D d'un transistor diminue, alors les variations ΔD de cette dimension vont représenter une part $\Delta D/D$ croissante. Le comportement électrique du transistor étant fonction de ces paramètres technologiques qui ont servi à sa fabrication, ses fluctuations augmentent aussi. On pourrait penser que les outils progressent sur la répétabilité autant que sur la miniaturisation, mais d'une part, rien ne garantit que ces progrès aillent de pair, et d'autre part l'augmentation de $\Delta D/D$ n'est pas le seul mécanisme de la dégradation de la variabilité.

La dispersion d'une performance électrique est le produit de la dispersion d'un paramètre technologique auquel elle est sensible par la sensibilité les liant. Pour une performance y et un seul paramètre technologique θ , on peut écrire :

$$\Delta_y = \frac{dy}{d\theta} \cdot \Delta_\theta \quad (1.1)$$

en faisant l'hypothèse d'une relation linéaire entre y et θ . Dans le cas de N

paramètres technologiques indépendants, donc orthogonaux, on a :

$$\Delta_y^2 = \sum_{i=1}^N \left(\frac{\partial y}{\partial \theta} \cdot \Delta_\theta \right)^2. \quad (1.2)$$

L'augmentation de la variabilité peut donc venir de l'augmentation de Δ_θ comme de l'augmentation de la sensibilité $\partial y / \partial \theta$. Sitte *et al.* comparent les sensibilités de dispositifs de 1,5 μm et 0,1 μm [11]. Ils observent avec le passage d'un grand transistor à un petit, l'augmentation des déviations des performances et l'augmentation des sensibilités des performances avec même un changement de signe dans certains cas. Cela montre la nécessité de mettre à jour les informations sur les fluctuations à chaque nouveau noeud technologique. Ils vérifient aussi l'indépendance des paramètres technologiques qui s'explique principalement par le fait que les étapes de fabrication sont séquentielles.

Les technologues ont la possibilité d'agir sur les deux éléments, déviations technologiques et sensibilité du dispositif à ces variations. Ils doivent réussir à réduire les variations des procédés afin de réduire les déviations électriques au niveau des composants et, en ce qui nous concerne, au niveau du transistor MOS. Ceci passe par une meilleure maîtrise des outils de production par les ateliers.

Ils peuvent également améliorer la robustesse du dispositif en trouvant la combinaison des réglages qui réduisent les sensibilités ou remplacer quelques point-clefs du process. Intel l'a fait au noeud 45 nm avec un isolant de grille *high- κ* et un matériau de grille métallique qui permettent de réduire les fuites d'isolant de grille. Ce procédé permet de s'affranchir de la désertion du polysilicium de grille tout en ayant une épaisseur physique supérieure pour une même capacité [12, 13]. La variation du travail de sortie⁵ a été reportée comme étant une source de variation dans les dispositifs à grille métallique [14].

Des architectures structurellement différentes sont également à l'étude pour remplacer le dispositif planaire tel qu'il existe depuis cinquante ans. Asenov *et al.*, étudient les couches de dopage *delta* [15]. Ils montrent que ces structures à canaux épitaxiés, formant une couche de dopage *delta* peu dopée, de l'ordre de 10^{15} cm^{-3} , sont plus robustes. Une solution radicale

5. différence d'énergie entre le niveau de Fermi du matériau et le niveau d'énergie du vide, utilisé comme référence.

pour supprimer la source de variation qui résulte du placement aléatoire des dopants serait d'avoir un canal non dopé, mais pour garder le contrôle électrostatique, il faut imaginer de nouvelles architectures. L'isolation de la couche d'oxyde des substrats SOI (*Silicon On Insulator*) permet de réduire le dopage du canal partiellement ou complètement déserté, réduisant ainsi le désappariement [16, 17]. L'utilisation de transistor à double grille, triple grille, quadruple grille ou *gate all around* permet de réduire fortement le dopage canal mais introduit d'autres sources de variation comme celles liées à la géométrie du dispositif [18].

Dans les dispositifs SOI, la réduction du dopage permet un gain sur la variation des performances. Les dispersions de longueur de grille et d'épaisseur de silicium sont attendues comme une limitation dans ces dispositifs [19].

Les concepteurs de circuits, les *designers*, subissent cette variabilité des procédés. Ils ne contrôlent pas les marges sur les performances des composants (transistors, résistances, etc.) du circuit qu'ils dessinent. Ils doivent pour autant s'assurer que les circuits qu'ils conçoivent présenteront des performances dont les dispersions seront dans les marges requises par le cahier des charges avec un rendement maximum. Pour cela, ils peuvent rendre plus robuste leur circuit, c'est à dire réduire sa sensibilité aux fluctuations de la technologie, grâce à des méthodes de conception [20] ou des méthodes d'optimisation du circuit [21, 22]. Du côté du placement des circuits – le *layout* – on peut citer la technique appelée *common centroid* permettant de compenser les effets d'un gradient spatial de paramètre technologique en dessinant les composants de manière symétrique (figure 1.4). Cette compensation des effets

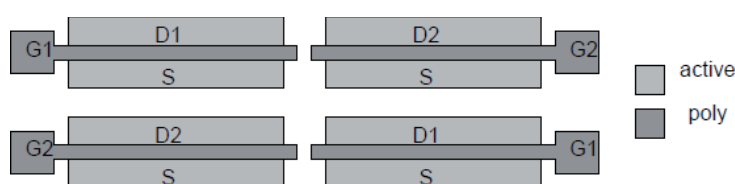


FIGURE 1.4 – Placement de 2 transistors en *common centroid*. Les transistors 1 et 2 ont été séparés chacun en une somme de deux transistors qui sont placés de manière symétrique par rapport au centre du dessin du circuit. Les gradients spatiaux de paramètres sont ainsi autocompensés autour de la valeur centrale du paramètre. [23]

spatiaux améliore en particulier l'appariement des dispositifs dont la surface est grande et où l'amplitude des variations de paramètres est significative.

Elzinga utilise des structures de tests suivant ce placement des transistors avec un accès indépendant aux quatre transistors pour différencier les effets totalement aléatoires des effets paramétriques spatiaux [24]. Les transistors par paires cablés côte-à-côte – horizontalement ou verticalement – présentent des effets aléatoires et paramétriques, tandis que les paires cablées de façon croisées masquent les effets paramétriques et ne présentent que des effets aléatoires.

En général, ces techniques de compensation par la conception ou le dessin impliquent des circuits plus complexes qui utilisent plus de composants et de surface de silicium. Donc, non seulement la variabilité réduit les rendements de production mais elle peut augmenter le coût des circuits en réduisant leur densité d'intégration. Pour s'assurer que le circuit sera conforme au rendement attendu, les concepteurs utilisent les fonctions de prédiction de la variabilité que leur offrent les modèles de simulation.

Le travail des modélistes consiste à développer des modèles de transistors prenant en compte la variabilité des dispositifs afin que les concepteurs puissent simuler le comportement statistique des performances critiques de leur circuit. Nous verrons en section 1.3 les différentes méthodes possibles. Les méthodes de simulation les plus couramment employées restent le « tir Monte-Carlo » et les « pire-cas » – appelés couramment *corner*. Dans tous les cas, créer des modèles prenant en compte la variabilité nécessite de connaître les sources et de comprendre leurs comportements.

1.2 Sources de déviation

1.2.1 Observations des effets

Le manque de répétabilité des performances d'un transistor à un autre, conçus identiquement, peuvent se manifester de différentes manières. Les différentes sources de variations agissent à des échelles différentes selon leur nature et s'observent différemment.

Croon *et al.* étudient l'influence de la méthode de caractérisation de la tension de seuil V_{th} [25]. En effet, la précision de la mesure peut apparaître dans les résultats comme un phénomène de désappariement mais sa nature est différente et il ne faut pas confondre ces effets. Ils comparent en termes de répétabilité et d'erreur systématique quatre appareils de mesures et cinq méthodes d'extraction du V_{th} : a) pente maximum, b) méthode trois points,

c) méthode quatre points, d) méthode à courant constant et e) méthode d'ajustement du désappariement du courant. Ils concluent que chaque méthode peut donner de bon résultats tant qu'elle est définie avec soin, mais la méthode trois points semble plus sûre.

Au niveau local

À des distances de l'ordre de la taille d'un transistor, les différences d'accord, ou d'appariement⁶ sont de nature aléatoire et sont attribuées aux variations locales des matériaux constituant les composants : silicium dopé, polysilicium, oxyde de grille principalement. L'appariement est mesuré sur des paires de transistors identiques et dessinés côte à côte afin de s'affranchir des sources de variations intervenant à plus grande distance. La figure 1.5 montre le placement d'une paire de transistors.

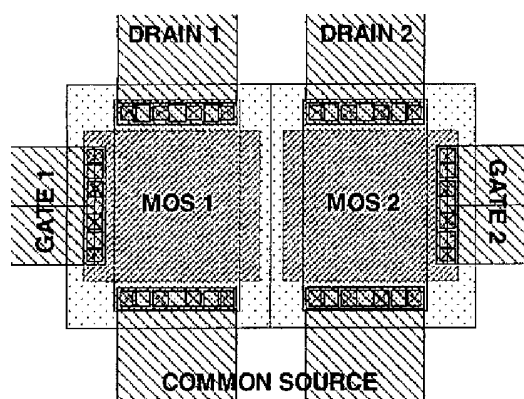


FIGURE 1.5 – Une paire de transistors identiques, placés au plus proche l'un de l'autre. La différence de caractéristiques électriques entre les deux est reconnue comme étant uniquement aléatoire car leurs paramètres de fabrication sont considérés identiques à cette distance minimale. [26]

Le désappariement provenant de ce type d'observation a pour cause une certaine quantité d'un même phénomène dont les effets sont tellement petits qu'ils peuvent être sommés. Leur distance de corrélation étant inférieure aux dimensions du transistor, le théorème centrale limite explique que les déviations engendrées par ces causes suivent une distribution normale, *i.e.* de loi gaussienne. Leur représentation fréquentielle est un bruit blanc (un

6. Pour le désappariement entre deux transistors identiques on utilise le terme anglais de *mismatch* (du verbe *to match* : correspondre à).

spectre continu). Les variations de performances en fonction du placement des transistors qui apparaissent à l'échelle d'une puce ou à l'échelle d'une tranche de silicium ont un spectre spatial composé de basses fréquences. Pelgrom *et al.* expriment, *via* l'espace de Fourier, l'écart type d'une performance du transistor causé par ces deux types de sources indépendantes [27] :

$$\sigma(\Delta P)^2 = \frac{A_P^2}{WL} + S_P^2 D_x^2 \quad (1.3)$$

où l'écart-type des déviations du paramètre P dépend d'une part de l'inverse de la surface du transistor pour les variations aléatoires (*via* le coefficient A_P dépendant de la technologie) et dépend de la distance entre les deux transistors considérés (*via* le paramètre S_P). La figure 1.5 montre une paire de transistors utilisée pour caractériser de cette variabilité. La distance entre les deux transistors est minimale. Par conséquent seul le premier terme de (1.3) reste :

$$\sigma(\Delta P) = \frac{A_P}{\sqrt{WL}}. \quad (1.4)$$

On appelle aujourd'hui couramment (1.4) la loi de Pelgrom et le tracé de $\sigma(\Delta P)$ en fonction de $1/\sqrt{WL}$ un graphe de Pelgrom. La figure 1.6 est un graphe de Pelgrom du V_{th} de différentes technologies. La croissance de l'écart-type de V_{th} est linéaire en fonction de l'inverse de la racine de la surface. Le facteur $A_{V_{th}}$ sert d'indicateur de la robustesse de telle ou telle technologie.

À l'échelle d'une puce

À cette échelle interviennent des sources modélisées par le second terme de l'équation (1.3), ce sont les gradients des paramètres technologiques sur la superficie de la tranche de silicium. Une autre source intervenant à cette échelle est l'interaction de la conception du circuit sur le procédé. Les paramètres de placement des éléments du circuit vont influencer le déroulement des étapes de fabrication. On peut citer :

- le nombre de grilles d'un transistor,
- la distance entre un transistor et ses voisins,
- l'orientation des dispositifs,
- la forme de la zone active,
- la largeur des tranchées d'isolation,...

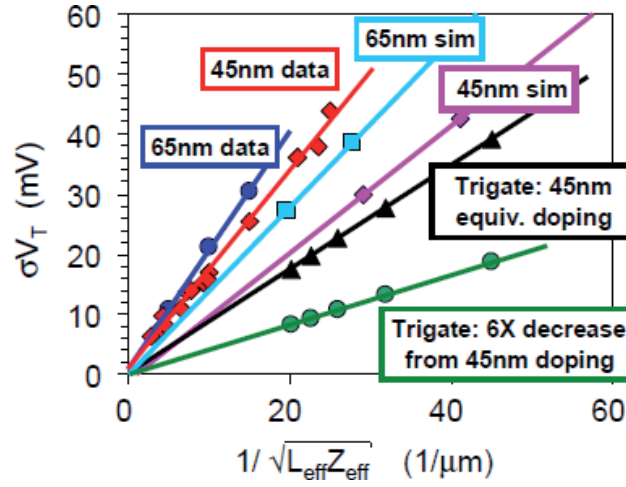


FIGURE 1.6 – Graphe de Pelgrom du V_{th} de différentes technologies mesurées ou simulées [13].

Ces paramètres relèvent du choix des concepteurs lors du placement routage de leur circuit – qu'on appellera couramment *layout*. Il a été montré que l'utilisation de dessins les plus réguliers possibles permet de minimiser ce genre de variations [28, 29, 30, 31].

Il est nécessaire pour mesurer ces effets de prévoir des structures de test à l'intérieur des puces et pas uniquement entre les puces sur les chemins de découpe comme il est courant de le faire pour la variabilité de puce à puce. L'étude expérimentale des différents effets de *layout* augmente considérablement le nombre de transistors à mesurer. Aussi, pour gagner en surface, on peut multiplexer ces structures et n'utiliser qu'un jeu de plots de mesure par groupe de structure de test pour poser les pointes de mesure [32]. Une étape de calibrage afin de retirer les contributions parasites des lignes d'accès est alors nécessaire.

À l'échelle d'une plaque

De puce à puce, les variations de paramètres existent lorsqu'un paramètre technologique varie d'un endroit à l'autre de la plaque, comme par exemple l'épaisseur d'oxyde déposé avant l'implantation des *ldd*⁷ représentée sur la figure 1.7. L'étendue et les motifs typiques de ces variations peuvent tout

⁷. *lightly doped drain* ou drain dopé légèrement. Ce sont les extensions des implants source/drain dans le canal.

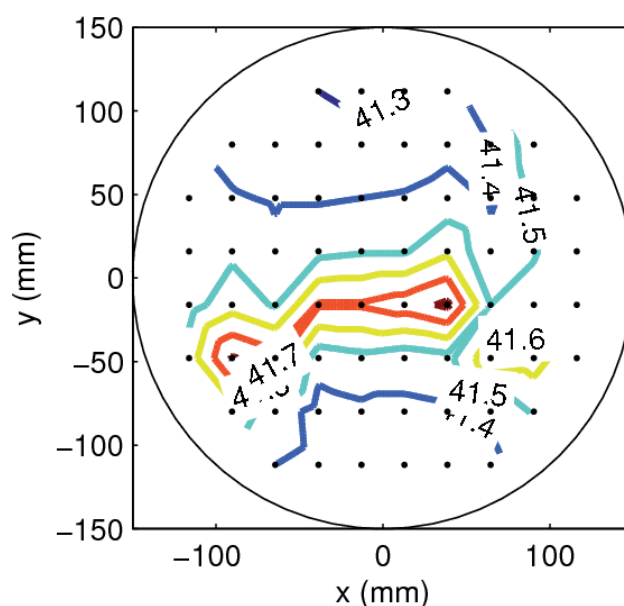


FIGURE 1.7 – Répartition spatiale de l'épaisseur d'oxyde *liner* mesurée sur la plaque n° 5 d'un lot de 12 plaques. Les variations suivent un motif radial et la plage de variation est de 0,3 %.

à fait être connues pour les paramètres mesurables. Cependant, du fait que ces variations interviennent d'une puce à l'autre, on ne peut pas les intégrer de manière déterministe dans la chaîne de simulation. En effet, les variations locales et intra-puce (cf. précédemment) font partie de la conception d'un même circuit. On peut donc dans une certaine mesure modéliser et prévoir le comportement de ces transistors, situé aux différents endroits de la puce et dont l'environnement est connu. Par contre les puces ne gardent pas l'information de leur position sur la tranche de silicium une fois qu'elles sont découpées. Ces variations ne peuvent donc être simulées que de manière statistique.

À l'échelle d'un lot

On peut noter que sur les 25 plaques – ou *wafers* – constituant un lot, les paramètres technologiques changent aussi. Certaines étapes technologiques se font plaque par plaque alors que d'autres se font simultanément – par *batch* –, c'est à dire 25 *wafers* dans la même machine. Dans le premier cas la répétabilité de l'opération seule importe, dans le second s'ajoute l'homogénéité spatiale de l'opération à l'intérieur de la machine. La figure 1.8 montre

les variations entre plaques de l'épaisseur de nitrure déposée pour la formation des espaceurs dans une technologie 45 nm en cours de développement. Chaque boîte à moustache⁸ représente la mesure de la population intra-plaque. On peut observer d'une plaque à l'autre l'évolution de la médiane qui rend compte de la variabilité *W2W* – *wafer to wafer*. On observe aussi que l'uniformité du dépôt évolue d'un bout à l'autre d'un lot.

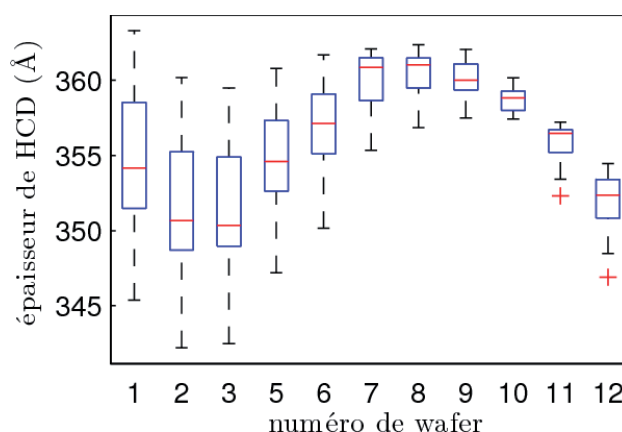


FIGURE 1.8 – Boîtes à moustaches représentant la population du paramètre HCD_{th} par plaque, pour les 12 plaques d'un lot. Cette visualisation permet de comparer les variations à l'intérieur d'une plaque et dans un lot. La tendance de la médiane sur les 25 plaques est une signature des équipements servant à cette étape de dépôt de nitrure.

Mesures

Les grandeurs technologiques que l'on peut mesurer directement et de manière non destructive sont les épaisseurs des différentes couches (dépôts ou oxyde thermique) et les largeurs (dimensions critiques ou CD, *Critical Dimensions*). Les épaisseurs sont mesurées par ellipsométrie. Le principe de l'ellipsométrie est de mesurer la polarisation d'une onde lumineuse de polarisation initiale connue après sa réflexion sur la surface à analyser. Les mesures de largeur se font par microscopie SEM (Scanning Electron Microscopy) ou par ellipsométrie spectroscopique (SCD, Spectroscopic Critical Dimension). Les grandeurs technologiques sont mesurées *en ligne*, c'est à dire entre deux étapes du procédé de fabrication, tant que la mesure est possible.

8. Voir [A.12](#), page 137

Pour mesurer les profils de dopages, on utilise la technique des SIMS qui consiste en l'analyse par spectrométrie de masse d'ions secondaires éjectés de la surface par le bombardement d'un faisceau d'ions primaires. Ce faisceau présente un rayon d'un micron environ et les ions utilisés peuvent être le césium Cs^+ ou l' O_2^+ selon les espèces mesurées. La mesure permet de compter le nombre d'ions secondaires par seconde. Un échantillon étalon est nécessaire pour calibrer quantitativement les résultats. Après la mesure de la profondeur creusée par le faisceau, on peut remonter au profil de la concentration de dopants en fonction de la profondeur par proportionnalité. Ce type de mesures nécessite la préparation de l'échantillon qui sera par conséquent détruit [33, 34].

Les mesures électriques se font après la fin du procédé de fabrication. Les structures de tests sont conçues avec des plots d'accès, c'est à dire des petites surfaces de métal où l'on vient poser les pointes des appareils de mesures. Des méthodes de multiplexages sont mises en œuvre pour accéder à plusieurs structures de test par les mêmes plots de mesures afin d'économiser de la surface [35].

Vickers rapporte le développement d'une technique de mesure sans contact [36]. Il présente un oscillateur en anneau comme structure de test, alimentée *via* un laser par une photodiode et dont la fréquence d'oscillation est déterminée par une mesure radiofréquence.

1.2.2 Sources intrinsèques

Une partie des sources de variabilité n'est pas attribuable au procédé de fabrication mais est causée par les matériaux constituant le transistor. Si, pour des dispositifs de grande dimension, les propriétés paraissent continues, dans les dispositifs décanométriques, les imperfections des matériaux et leurs propriétés à l'échelle locale voire atomique sont à prendre en compte. Les conséquences de ces imperfections ont été étudiées par simulation par plusieurs équipes de recherche.

Rugosité de ligne de grille

Le trait de polysilicium formant la grille du transistor n'est pas un parallépipède parfait. Ses bords verticaux en particulier ne sont pas parfaitement plans et souffrent de ce qu'on appelle la rugosité du bord de ligne – *Line Edge Roughness* ou LER – qui entraîne une variation de la longueur effective

du canal dans la direction correspondant à la largeur du transistor. En effet, la grille en polysilicium sert de masque lors de l'implantation des *ldd* et positionne donc les jonctions canal-source/drain. C'est le principe d'autoalignement.

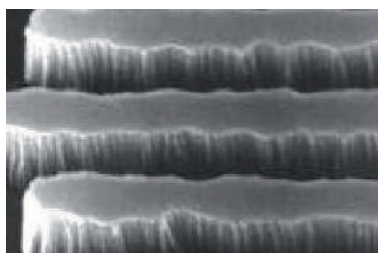


FIGURE 1.9 – Image SEM de traits de polysilicium. On voit nettement l'imperfection du bord de grille [37].



FIGURE 1.10 – Exemple de faible (a) et forte (b) rugosité de ligne [38].

On définit la LER par l'écart type de la largeur du trait de polysilicium [39]. Selon le positionnement et l'étendue de ces mesures, on observe sur une courte étendue des variations plutôt hautes fréquences et sur une étendue supérieure à une longueur caractéristique, dépendante de la technologie, la somme quadratique de variations basses fréquences et hautes fréquences [39, figure 2]. Cependant, l'influence de l'écart-type de la longueur de grille se moyenne quand le dispositif est plus large, et décroît naturellement pour le transistor long de par sa dépendance en $\Delta L/L$.

Difrenza *et al.* [40] montrent que la fluctuation des dopants n'est pas la seule source de désappariement et mettent en évidence l'impact de la rugosité du trait de polysilicium. Ils différencient aussi l'effet statistique de la réduction des dimensions et du changement de sensibilité à la longueur de grille qui en résulte. En effet, dans les canaux courts où la tension de seuil chute, la sensibilité $\partial V_t / \partial L$ augmente nettement.

L'origine de cette rugosité est attribuée au développement de la résine servant de masque pour la grille. Les propriétés du matériau polymère de la résine du masque, les propriétés optiques de la lumière utilisée pour insoler le masque et l'ensemble optique sont incriminés. Des traitements de cette résine ont été proposés pour améliorer la LER, comme un dépôt par dessus la résine gravée ou encore un traitement thermique [41, 42]. Une surgravure de la grille est aussi proposée pour réduire la composante haute-fréquence de

la rugosité [43]. Pargon *et al.* montrent que l'utilisation de masque en SiO_2 ou en carbone amorphe aide à réduire la rugosité du bord de grille [44].

Asenov *et al.* étudient cette source par simulation 3D [45]. Le profil de la grille est caractérisé par une amplitude moyenne et une longueur de corrélation. À partir de ces deux paramètres, ils génèrent aléatoirement des profils de grille qu'ils utilisent pour des simulations Monte-Carlo de courants de transistors de différentes dimensions. Ils prévoient une variation σ/μ de I_{on} de 8% pour des dispositifs de 50 nm de large et 30 nm de long avec une forte dépendance à longueur de grille ($\propto 1/L$).

Linton *et al.* prévoient à partir d'un modèle de transistor en parallèle que σ_{LER} devra être réduit à 1 nm au noeud 32 nm pour que son influence sur la variation du courant n'augmente pas par rapport au noeud 80 nm [38]. Ils proposent aussi, à partir de ce modèle, une méthode pour extraire des mesures la part de variation attribuable à la LER.

Pincement du potentiel de surface par les joints de grain du polysilicium

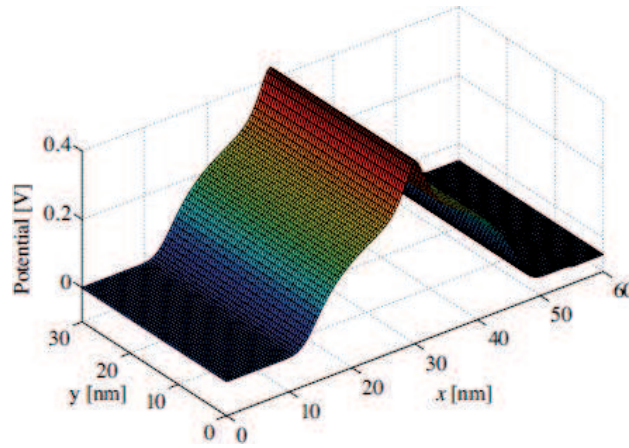
La grille est faite de silicium polycristallin – ou *polysilicium* – qui est composé de petits cristaux d'orientations différentes. Pendant le dépôt du matériau, la croissance du cristal démarre à plusieurs endroits en même temps avec des orientations aléatoires. À partir de ces points, des cristaux croissent jusqu'à être en contact. On parle de joints de grains pour ces surfaces de contact. La taille des grains est également aléatoire.

Les propriétés électriques locales du matériau aux joints sont différentes de celles du cristal lui-même. En particulier, à l'interface entre le polysilicium de grille et l'isolant, le travail de sortie du matériau se trouve modifié le long des joints de grains. Le potentiel de surface et la tension de seuil sont impactés en conséquence.

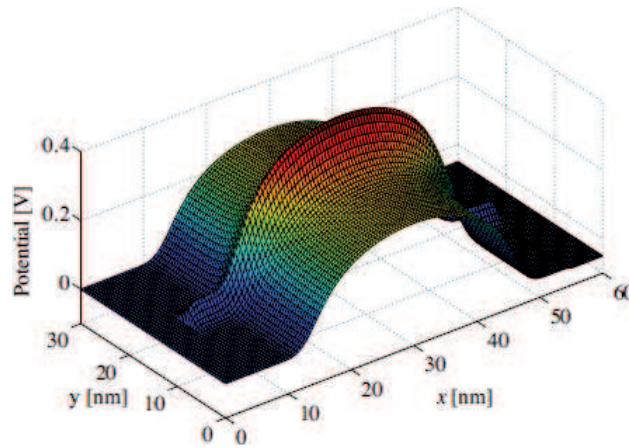
Difrenza *et al.* observent expérimentalement l'importance du procédé de formation de la grille dans les déviations de V_{th} en comparant le matériau déposé (silicium amorphe ou polycristallin) et le type de recuit (en four ou recuit rapide). Ils construisent un modèle compact afin d'évaluer la contribution de ce phénomène [46]. Ils prennent en compte un dopage différent sur les joints de grain et sur les parties cristallines du polysilicium qui implique un écart de V_{th} via une zone de déplétion dans les parties moins dopées. Ils différencient l'effet du nombre de grains par rapport au nombre de dopants

grâce à la dépendance en V_B des modèles qu'ils ont écrit pour le nombre de dopants et le nombre de grains.

Asenov montre comment le placement aléatoire des joints de grains et en particulier leur orientation influencent le courant du dispositif. Il montre les cas extrêmes de deux transistors ayant un joint de grain dans la grille [47, 48]. Sur l'un, le joint est longitudinal aux sens du canal, dans l'autre il est transversal (figure 1.11). La barrière de potentiel du premier est augmentée sur toute la largeur du canal alors que celle du second n'est surelevée que sur la largeur d'un joint de grain.



(a) Grain transverse au canal



(b) Grain longitudinal au canal

FIGURE 1.11 – Potentiel de surface sous l'influence d'un seul joint de grain [48].

Dans les grands transistors une grande quantité de joints de grains est

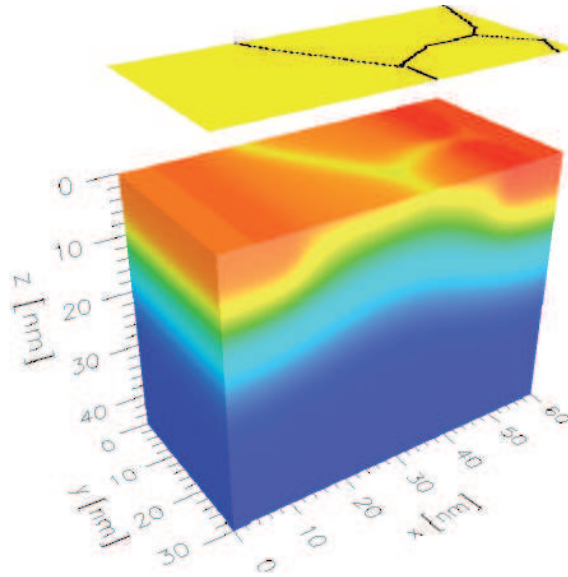


FIGURE 1.12 – Échantillon d’une simulation Monte-Carlo 3D atomistique prenant en compte l’influence des joints de grains sur le potentiel de surface [48].

présente et leurs influences sur une grande surface de canal sont globalement identiques d’un transistor à l’autre. Quand les dimensions diminuent et que la surface du canal ne comporte alors que quelques joints, leurs différences d’un dispositif à l’autre deviennent très nettes et donc la variabilité du seuil de conduction augmente dans les petits dispositifs.

Placement aléatoire des dopants

Dans des transistors de petite surface, on ne peut pas considérer que le dopage est uniforme comme dans un grand dispositif. Lors de l’implantation ionique, les atomes de dopants se répartissent globalement selon un profil bien connu à l’échelle macroscopique, mais à l’échelle microscopique leurs emplacements individuels restent aléatoires. Une concentration par exemple de 10^{17} atomes/cm³ ne représente que peu d’atomes au milieu du cristal de silicium qui en contient 5×10^{22} par centimètre cube : 2 atomes de dopants pour un million d’atomes de silicium. Pour donner un ordre de grandeur, un volume de $W \times L = 120 \text{ nm} \times 40 \text{ nm}$ sur disons 10 nm de profondeur, qui peut être le volume du canal d’un transistor de dimensions nominales,

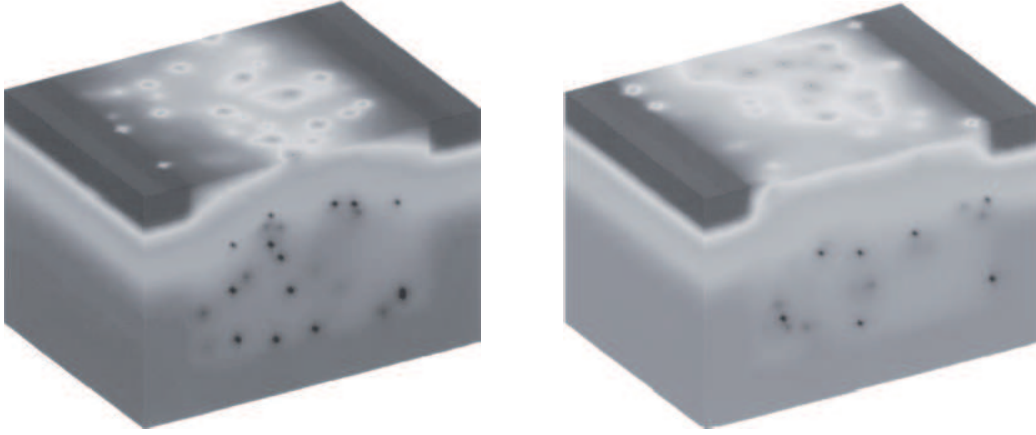


FIGURE 1.13 – Simulation de 2 transistors qui diffèrent uniquement par le positionnement local des dopants dans le canal.

contiendrait en moyenne :

$$120 \times 40 \times 10 \times 10^{-21} \text{ cm}^3 \times 10^{17} \text{ cm}^{-3} = 48000 \times 10^{-4} = 4,8 \text{ atomes.}$$

On comprend que ces cinq atomes auront un impact très différent s'il sont espacés régulièrement dans le volume ou s'ils sont plutôt regroupés au milieu du canal, ou encore sur les bords.

Plusieurs laboratoires ont développé des outils de simulation dits *atomistiques*, car ils calculent le potentiel électrique en prenant en compte la contribution individuelle de chaque atome dopant. Le placement des atomes dans le cristal est soit le résultat d'une simulation atomistique du process reproduisant l'implantation puis le recuit, soit un placement aléatoire suivant un profil de concentration prédéfini. Le courant électrique est ensuite calculé par un simulateur *drift-diffusion*⁹ utilisant le potentiel électrique calculé de manière atomistique.

Gross *et al.* montrent à partir de simulations atomistiques Monte-Carlo prenant en compte les interactions électron-électron et électron-ion que le positionnement des impuretés dans la zone de déplétion (qui occupe les 15 nm supérieur du canal) est fortement corrélé au V_{th} . De même, les fluctuations du courant sont très corrélées au placement des dopants dans les huit premiers

9. La densité de courant J est exprimée par la somme du courant résultant du gradient de dopage $n(x)$ et du courant résultant du gradient de potentiel électrique E . On a $J_n = qn(x)\mu_n E(x) + qD_n \partial n / \partial x$ pour les électrons et $J_p = qn(x)\mu_p E(x) - qD_p \partial n / \partial x$ pour les trous, avec μ la mobilité des porteurs et D leur coefficient de diffusion.

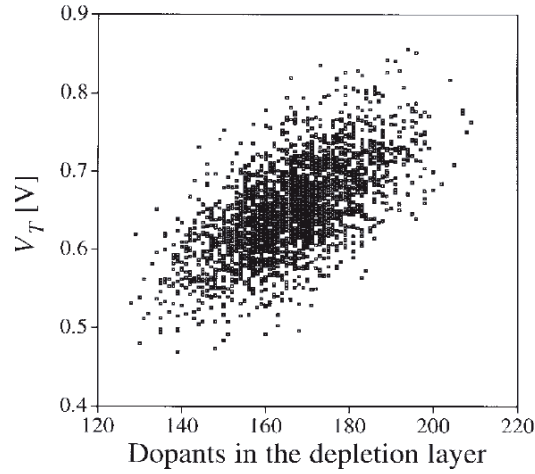


FIGURE 1.14 – Tensions de seuils de 2500 transistors de la figure 1.13 en fonction du nombre de dopants dans la couche de désertion [49].

nanomètres (couche d’inversion) [50].

Sano *et al.*, puis Ezaki *et al.*, proposent une décomposition du potentiel coulombien associé à chaque charge pour éviter une singularité de calcul au niveau de l’impureté en introduisant une longueur d’écrantage [51]. L’expression du potentiel est séparée en un terme d’interaction longue distance et un terme d’interaction proche. Ezaki *et al.* utilisent un simulateur d’implantation et de recuit pour déterminer le placement des impuretés [52].

Dollfus *et al.* développent aussi leur simulateur d’implantation/diffusion et leur simulateur atomistique [53]. Ils étudient l’influence du positionnement et du nombre de dopants dans la couche d’inversion sur le champ électrique, la concentration et la vitesse des électrons dans un transistor de dimensions $50\text{ nm} \times 50\text{ nm}$. Ils soulignent l’importance de s’intéresser aux architectures de transistors non dopés pour les technologies à venir.

De la même manière que pour le pincement du potentiel de surface, Ase-nov *et al.* simulent plusieurs dispositifs dont la densité de dopants est identique mais dont le placement de chaque atome d’impureté reste aléatoire [49, 54, 15, 55, 56, 57]. On peut observer, sur la figure 1.13, deux dispositifs globalement identiques mais dont le placement des dopants est aléatoire. On observe les différences de potentiel de surface qui apparaissent à cause de la nature discrète du dopage. Les valeurs de V_{th} obtenues à partir d’un tir Monte-Carlo de 2500 simulations de ce type présentent un écart-type de 50 mV à ces dimensions, pour une épaisseur d’oxyde de grille de 3 nm.

Imperfection du diélectrique de grille

Le diélectrique de grille présente également des imperfections. En effet, les derniers nœuds technologiques emploient des épaisseurs de 2 à 3 nanomètres et les variations d'épaisseurs d'un atome sont alors significatives. L'image SEM de la figure 1.16 illustre cette variation. En plus d'apporter des fuites par effets tunnel direct à travers l'oxyde [58], cette rugosité de l'interface apporte sa part de variabilité.

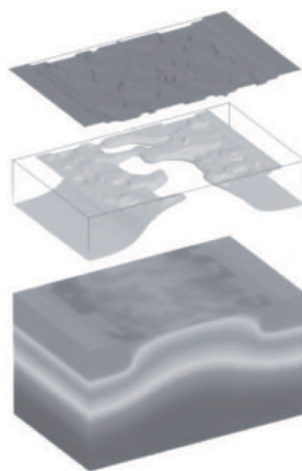


FIGURE 1.15 – Simulation prenant en compte les irrégularités de la surface SiO₂/Si [59]. On peut observer séparément de haut en bas l'irrégularité de la surface oxyde/silicium, une isosurface de niveau de dopage de 10^{17} cm^{-3} et la distribution de potentiel à la surface du canal.

L'interface SiO₂/Si est affectée par la surface du cristal de silicium dont les sauts d'un plan atomique à l'autre sont de l'ordre de 0,2–0,3 nm pour le plan (001) [59]. La figure 1.15 montre une simulation prenant en compte cette interface imparfaite [59]. Dans ce travail, la surface est modélisée grâce à une fonction d'autocorrélation gaussienne ou exponentielle, paramétrée avec une longueur d'autocorrélation Λ et une hauteur moyenne Δ . On note que Δ et Λ sont déterminés à partir d'image TEM et AFM. Le potentiel de surface du dispositif subit des variations locales aléatoires qui causent une déviation de la tension de seuil du dispositif. Sur ce dispositif de $30 \text{ nm} \times 30 \text{ nm}$ avec un oxyde d'épaisseur moyenne de 1,05 nm, un tir Monte-Carlo sur 200 échantillons prévoit un écart-type de la tension de seuil pouvant aller jusqu'à 25 mV sur ce dispositif selon les paramètres de rugosité utilisés. L'influence de cette rugosité augmente aussi avec le niveau de dopage mais les variations de ces

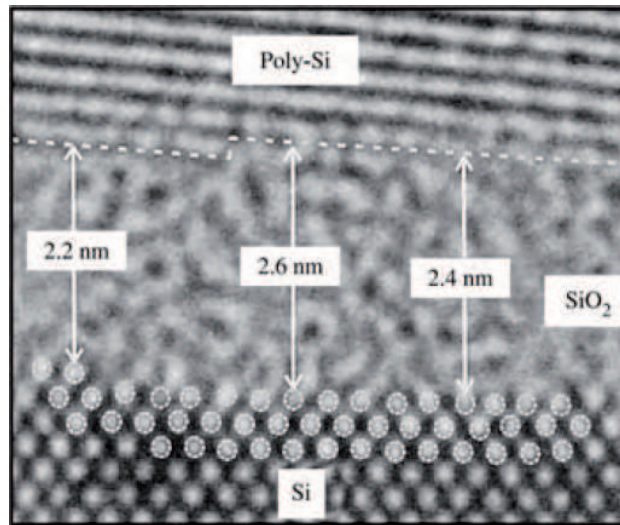


FIGURE 1.16 – Image SEM de l'oxyde de grille illustrant les imperfections d'épaisseur [60].

deux effets agissent indépendamment.

1.2.3 Sources originaires du dessin

La topologie dessinée impacte les chemins critiques des circuits numériques et les circuits analogiques (à signal mixte ou radiofréquences) à travers les effets de proximités dépendants du dessin (STI, WPE, lithographie...).

Orientation

L'orientation des dispositifs place les transistors dans des situations différentes vis-à-vis du cristal dont les propriétés électriques ne sont pas isotropes. La mobilité est différente selon le plan choisi pour la surface du silicium et pour l'orientation du canal [62]. Le tableau 1.1 rappelle les mobilités maximales pour différents plans de surface de la tranche de silicium et différentes orientations du canal.

Pelgrom *et al.* rapportent une augmentation d'un facteur 5 du coefficient de désappariement lorsque les paires de transistors mesurés sont perpendiculaires et non pas parallèles [27].

On a vu que les effets systématiques dus au placement peuvent être en partie modélisés et donc pris en compte dans la simulation du circuit, au coût

NMOS				
Substrat	(100)	(110)	(111)	
Canal	[110]	[110]	[100]	[112]
Mobilité max. ($\text{cm}^2/\text{V} \cdot \text{s}$)	380	280	140	280
PMOS				
Substrat	(100)	(110)	(111)	
Canal	[110]	[110]	[100]	[112]
Mobilité max. ($\text{cm}^2/\text{V} \cdot \text{s}$)	80	190	120	120

TABLE 1.1 – Mobilité maximale selon les orientations [62]. La mobilité est fonction du champ vertical et de la température. Seules les valeurs maximum sont rapportées ici.

d'un effort de caractérisation de structures de test dédiées. D'autres effets du *layout* sont ré-évalués par simulation numérique après le placement.

Perturbation de la lithographie

En plus du caractère aléatoire du bord de trait de polysilicium dont nous avons parlé au paragraphe 1.2.2, page 13, l'étape de révélation de la résine de lithographie est influencée par le routage des circuits. Les échelles atteintes par les motifs des transistors modernes sont de l'ordre de la longueur d'onde de la lumière utilisée (193 nm). La forme ou le degré de proximité des lignes de polysilicium vont affecter l'énergie reçue par une zone de la couche de résine et donc la forme finale de la résine après révélation. Par exemple les bouts de lignes et les angles vont être arrondis, la largeur du trait peut varier en fonction de la distance avec le trait voisin. Un motif trop fin ne s'imprimera pas du fait que sa résine aura été irradiée par les zones voisines etc. Pour contrebalancer cela, on met en place des techniques d'amélioration de la résolution – RET pour Resolution Enhancement Technics. On trouve en premier lieu les corrections optiques de proximité – OPC pour *Optical Proximity Correction* – dont le principe est de modifier le masque de manière à anticiper les effets non voulus de l'illumination de la résine. Les masques à décalage de phase – PSM pour Phase Shift Mask [63] –, quant à eux, permettent de graver au delà de la résolution optique en plaçant en opposition de phase deux zones encadrant un motif fin.

La figure 1.17 montre un exemple de masque rectangle avec et sans cor-

rection optique. Le motif du masque initialement rectangle produira sur la résine une forme dégradée par rapport à la forme souhaitée par les concepteurs. Les corrections apportées au masque entre la conception du circuit et la lithographie modifient en particulier les angles du rectangle et ajoutent des motifs pour contrôler l'environnement optique du rectangle.

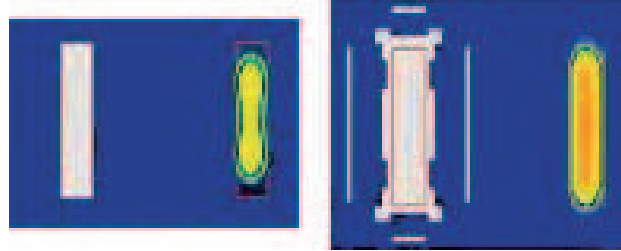


FIGURE 1.17 – Correction de proximité optique. À gauche, sans correction, le motif initial rectangle est arrondi et raccourci après exposition et révélation de la résine. L'échelle de couleur indique le taux de révélation de la résine. À droite, avec correction, la forme du motif sur le masque anticipe les dégradations et permet un motif proche du rectangle initial voulu [64].

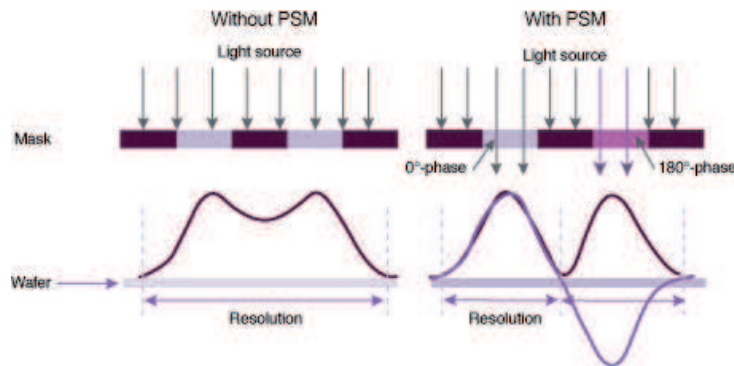


FIGURE 1.18 – Principe du masque à décalage de phase [65]. En mettant en opposition de phase les ondes passant à droite et à gauche du motif à imprimer, on conserve une zone d'ombre sur le motif en question. On augmente ainsi la résolution maximale.

La figure 1.18 montre le principe de la technique PSM. Sans PSM, à gauche, un motif trop étroit reçoit de l'énergie des zones voisines par diffraction. En effet la longueur d'onde est relativement trop grande et le seuil d'énergie nécessaire pour permettre la dissolution de la résine est dépassé. Par conséquent, le motif sera effacé à la révélation de la résine. La méthode

PSM à droite permet d'obtenir deux zones en opposition de phase en cadrant le motif afin que ce dernier ne reçoive pas d'énergie non voulue. Cette technique est possible grâce à une couche déposée sur la résine et dont l'épaisseur est choisie en fonction du décalage de phase souhaité.

Interaction entre le dessin et les variations du procédé

Le recuit arrivant en fin des procédés *front-end* est un recuit rapide à haute température. Le but de ce recuit est d'activer électriquement les dopants en les faisant diffuser le moins possible puisqu'on veut garder des jonctions peu profondes. Le recuit est réalisé par des lampes émettant dans le domaine du proche infrarouge. Localement sur la puce, l'absorption d'énergie dépend du matériau présent. L'absorption du silicium, de l'oxyde des régions *STI* et du polysilicium est différente. Il en résulte que l'énergie reçue lors du recuit dépend du dessin du circuit. L'activation et la diffusion des dopants seront donc influencées par la densité de polysilicium et d'oxyde à la surface du silicium.

La variabilité intrapuce de la température reçue lors du recuit spike peut être évaluée par simulation optique et thermique. Fiori, Cacho *et al.* étudient les propriétés optiques de plusieurs motifs réguliers en fonction de la distance entre deux zones de polysilicium, de la largeur des traits de polysilicium et de la présence ou pas du *STI*. Ils combinent ces résultats avec les informations contenues dans le fichier GDS¹⁰ afin d'évaluer les valeurs d'absorption sur un vrai circuit. La carte d'absorption est ensuite convertie en température grâce à des mesures préalablement effectuées sur des structures de test – des résistances de surface – dédiées à la corrélation entre le motif et la température, *via* la sensibilité de la résistance à la température. Ils étudient ainsi le recuit flash et le recuit laser [66, 67]. La figure 1.19 montre le résultat dans le cas de la simulation d'un recuit laser. Les différences de température atteignent 50 °C dans le domaine de simulation.

Dans le cas du recuit spike, il a été montré que le dépôt d'une couche de carbone amorphe avant le recuit aide grandement à l'amélioration de l'uniformité du recuit [68].

10. *graphic database system*, le fichier décrivant les masques utilisés pour la photolithographie.

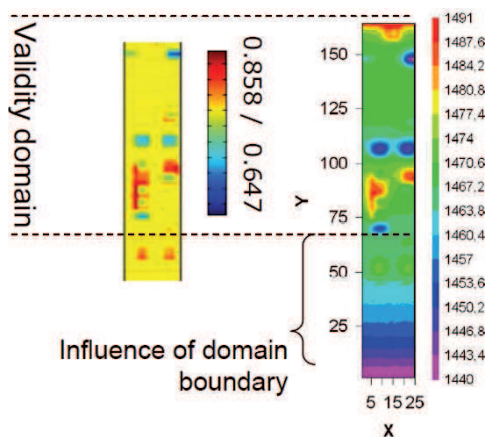


FIGURE 1.19 – Simulation de l’absorption énergétique et de la température associée lors d’un recuit laser [67]. Sur ce domaine d’étude, la température reçue varie d’une cinquantaine de degrés.

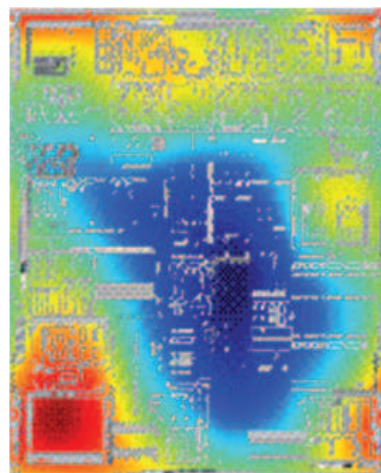


FIGURE 1.20 – Simulation de la température reçue lors d’un recuit spike [66]. Sur ce domaine d’étude, la température reçue varie de deux degrés.

Effet de proximité du *well*

En bord de *well*¹¹, le dopage est modifié à cause de diffractions des ions implantés. Les ions qui sont implantés sur le bord du masque sont déviés et viennent s’accumuler avec les ions normalement implantés dans le bord de la zone active, comme le montre la figure 1.21a. Ce surplus de dopants crée un gradient de dopant dans le canal [69, 70, 71]. On appelle *well proximity effect* (WPE) cet effet agissant en bord de *well*. C’est un effet systématique qui est pris en compte dans les modèles. Wils *et al.* montrent que cet effet n’apporte pas de variabilité supplémentaire au transistor [72].

Influence du dessin sur le stress

Tsuno *et al.* étudie l’influence de l’espacement des grilles, de la largeur des *STI*¹² et de la distance entre la grille et les *STI* [73]. Ils modélisent les variations causées par ces facteurs systématiques par deux effets : le stress

11. Zone délimitant le dopage substrat des dispositifs

12. Shallow Trench Isolation, tranchée d’isolation peu profonde.

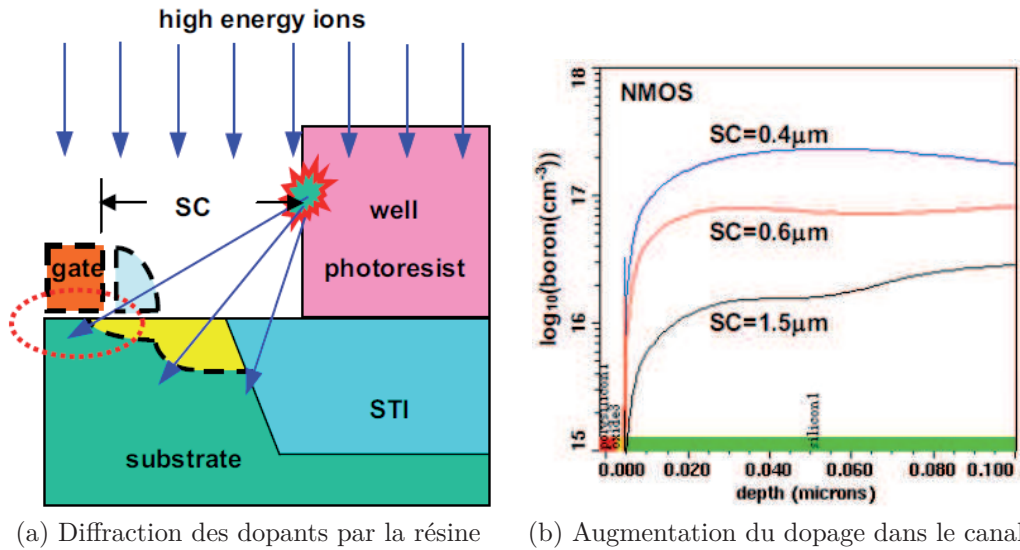


FIGURE 1.21 – Effet de proximité du well [70]

mécanique et la TED¹³. La TED varie à cause de sa proximité ou non avec l'interface silicium/oxyde des STI, où la ségrégation des interstitiels modifie localement la diffusion des dopants¹⁴. Le stress est supposé provenir du recuit de la siliciuration des source/drain et de la couche d'arrêt de gravure des

13. Transient Enhanced Diffusion, Diffusion accélérée transitoire. Un mécanisme de diffusion des dopants accélérés par leurs interactions avec la diffusion des défauts de silicium interstitiels. Une fois que tous les défauts sont réparés, la diffusion n'est plus accélérée, d'où son nom de transitoire [74, 75](voir figure 1.22).

14. Cette influence sur la diffusion est dépendante de l'espèce qui diffuse, ce qui implique que les effets sont différents sur les transistors N et P.

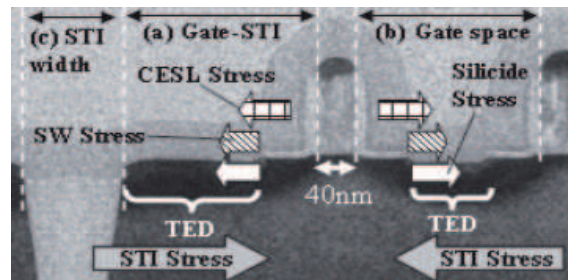


FIGURE 1.22 – Les performances du transistors sont influencées par des paramètres d'instance du transistor comme la distance entre grilles, la distance grille-STI, la largeur des STI, *via* leurs influences sur le stress et la diffusion. [73]

contacts, et des espaceurs. Tsuno *et al.* introduisent dans le modèle compact

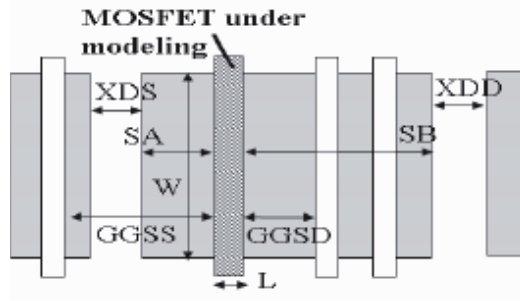


FIGURE 1.23 – Paramètres d’instance proposés par Tsuno pour simuler les interactions du layout sur le process. [73]

BSIM les effets de la distance au STI, de la taille du *STI* et la distance entre la grille et le premier trait de polysilicium, en ajoutant des termes correctifs à la tension de seuil V_{th} et à la mobilité μ_{eff} :

$$Vt = Vt_0 + f(SA, SB, XDS, XDD, L) + g(SA, SB, GGSS, GGSD, L), \quad (1.5)$$

$$\mu_{eff} = \mu_{eff0} \times h(SA, SB, XDS, XDD, GGSS, GGSD), \quad (1.6)$$

où f est une fonction du stress, g une fonction de la TED, et les autres paramètres sont décrits dans la figure 1.23.

1.2.4 Sources originales du procédé

Dépôts

Dans le *front-end* du procédé, des dépôts d’oxyde, de polysilicium (la grille) et de nitrure sont effectués. La technique de dépôt est le dépôt chimique en phase vapeur – CVD, Chemical vapor deposition – qui consiste à exposer la surface des plaquettes à une molécule contenant le matériau à déposer, appelée précurseur. Les dépôts existent sous un bon nombre de variantes selon la pression dans la chambre de réaction (utilisation de plasma ou de phase vapeur). Sous l’action de l’atmosphère présente dans la chambre à réaction, le précurseur va réagir selon les équations du tableau 1.2 et ainsi déposer le matériau voulu.

Le dépôt de dioxyde porte couramment le nom de son précurseur, le TEOS

TABLE 1.2 – Réactions utilisées pour les dépôts.

Polysilicium	SiH_4	$\longrightarrow$	$\text{Si} + 2\text{H}_2$
Dioxyde de silicium	$\text{SiH}_4 + \text{O}_2$	$\longrightarrow$	$\text{SiO}_2 + 2\text{H}_2$
	$\text{SiCl}_2\text{H}_2 + 2\text{N}_2\text{O}$	$\longrightarrow$	$\text{SiO}_2 + 2\text{N}_2 + 2\text{HCl}$
	$\text{Si}(\text{OC}_2\text{H}_5)_4$	$\longrightarrow$	$\text{SiO}_2 + \text{résidus}$
Nitrure	$3\text{SiH}_4 + 4\text{NH}_3$	$\longrightarrow$	$\text{Si}_3\text{N}_4 + 12\text{H}_2$
	$3\text{SiCl}_2\text{H}_2 + 4\text{NH}_3$	$\longrightarrow$	$\text{Si}_3\text{N}_4 + 6\text{HCl} + 6\text{H}_2$
	$n_1(\text{SiCl}_3)_2 + n_2\text{NH}_3$	$\longrightarrow$	$n_3\text{Si}_3\text{N}_4 + n_4\text{HCl}$ [76]

tétra-ethyl-ortho-silicate $\text{Si}(\text{OC}_2\text{H}_5)_4$. De même on appelle HCD, *hexa-chloro-disilane* $(\text{SiCl}_3)_2$, le dépôt de nitrure.

Les variations d'épaisseur de ces dépôts s'observent par plaques avec un motif typiquement radial. Pour les dépôts issus d'un procédé par *batch*, c'est à dire qu'on place le lot entier dans la chambre de réaction, on observe à la fois une dérive de la valeur moyenne et une dispersion de l'épaisseur déposée par plaquette d'un bout à l'autre du lot. On observera des mesures de ce type dans le chapitre 3.

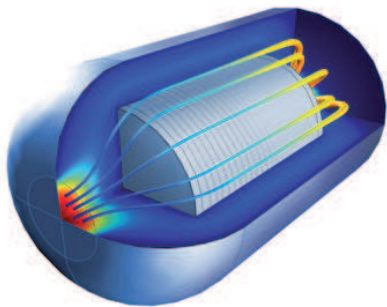


FIGURE 1.24 – Simulation des taux de croissance locaux dans le réacteur en fonction du flux de gaz [77].

Gravures

Parmi les gravures entrant en jeu dans la réalisation du transistor MOS, il y a la gravure de la tranchée d'isolation STI, la gravure du polysilicium de grille, la gravure du HCD des espaceurs et la gravure du diélectrique pré-métal pour le contact de grille, source, drain et bulk au niveau *Métal 1*.

Ces quatre gravures sont des gravures sèches et isotropiques, pour graver uniquement dans le plan vertical. La littérature ne rapporte pas de source de variabilité particulière associée aux gravures anisotropes.

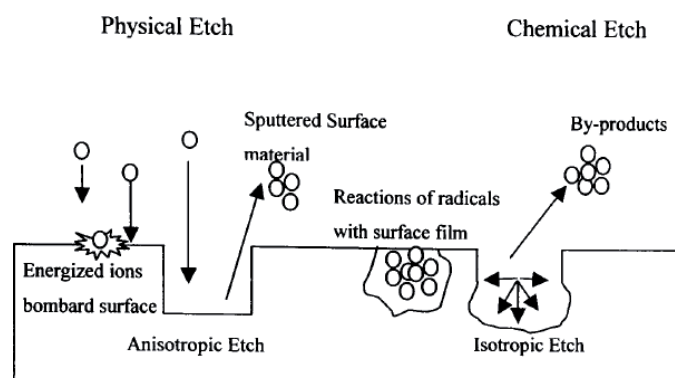


FIGURE 1.25 – Gravure sèche et humide [78].

Implantations

L'implantation des ions est en soi un procédé réputé comme ayant une bonne répétabilité. La dose est contrôlée en courant dans l'implanteur. C'est l'intégrale du flux d'ions pendant le temps d'implantation.

La miniaturisation des dispositifs a poussé à augmenter les niveaux de dopage. Pour régler le V_{th} , il a été nécessaire d'ajouter des implantations de dopants pour contrer les effets de canal court. Les implants appelés *poches*, *pockets* en anglais, ou encore *halo* sont des zones surdopées, situées sous les *ldd*. Leur fonction est de ralentir l'avancée des zones de charge d'espace des source/drain vers le canal et donc la perte de contrôle du canal par la grille, quand L_g diminue. Ce dopage est réalisé par des implantations ioniques inclinées par rapport à la verticale pour envoyer les dopants sous la grille. On effectue généralement quatre implants en pivotant la plaque d'un quart de tour pour doper les zones d'ombre de la grille.

Croon *et al.* séparent les fluctuations aléatoires de V_{th} en fluctuation du dopage, de la charge d'interface, de l'épaisseur d'oxyde et les différencie grâce à l'observation de A_{Vt} en fonction de V_B . Le modèle correspondant avec une dépendance en V_B prends en compte l'épaisseur de la couche de désertion W_D [79]. L'implant halo vient contredoper le dopage de la grille à travers les grains aléatoires de polysilicium et augmente la part de variation due à

la charge de grille. Ils montrent une réduction de A_{V_t} d'un facteur 2 sur le pMOS en réduisant l'inclinaison des *pockets* de 45° à 35° .

Le problème avec les implantations ioniques vient aussi et surtout du fait qu'il est nécessaire de réparer le cristal par traitement thermique, ce qui implique une diffusion des dopants.

Recuits

Le recuit est une opération intervenant plusieurs fois, après les implantations ioniques pendant les opérations de *front-end*. Sa fonction est de reformer le cristal de silicium qui a été abîmé, voire complètement amorphisé durant le bombardement ionique, tout en activant les dopants, c'est à dire les intégrer au maillage cristallin. Les recuits sont faits à des températures de l'ordre de 1000°C . On distingue les recuits *soak* avec profil en plateau pendant un temps de l'ordre de quelques dizaines de secondes, les recuits *spike*, en pic pendant quelques secondes et les recuits *flash* et laser dont le temps d'action est inférieur à la seconde.

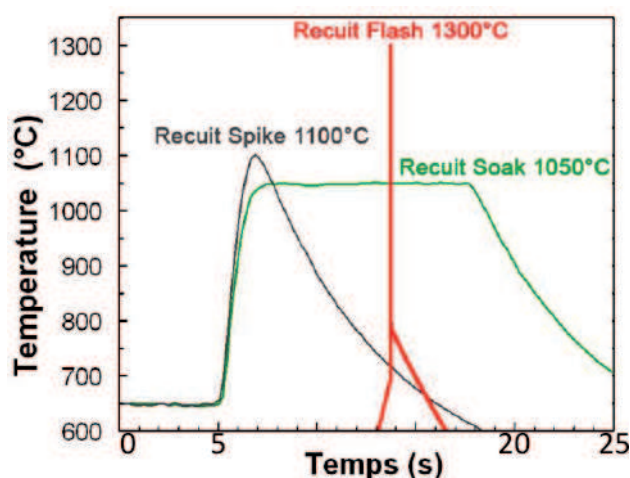


FIGURE 1.26 – Profils en température des recuits *spike*, *soak* et *flash* [80].

Les recuits spike (et laser) montent en pic à des températures de l'ordre de 1300° et ne chauffent qu'une couche superficielle du substrat. La chaleur diffuse ensuite dans le substrat. Les wafers sont chauffés plaque par plaque par des lampes halogènes ou au tungstène. Les wafers sont en rotation sur leur support pour améliorer l'uniformité du recuit. Le recuit flash utilise des lampes à arc à l'argon qui émettent autour de 800 nm pour le pic principal.

Pendant le recuit, l'apport d'énergie sous forme de chaleur permet la diffusion des dopants. La diffusion des dopants dans le temps suit le gradient spatial de concentration (lois de Fick) :

$$J = -D_A \frac{\partial C(x,t)}{\partial x} \quad (1.7)$$

$$\text{avec } D_A = -D_0 \exp\left(\frac{-E_a}{kT}\right) \quad (1.8)$$

$$\text{et } \frac{\partial C(x,t)}{\partial x} = \frac{\partial J(x,t)}{\partial x}. \quad (1.9)$$

J est le courant de diffusion, D_A le coefficient de diffusion associé à l'espèce A (bore, arsenic ou phosphore) et C la concentration de A . D_0 est appelé coefficient pré-exponentiel de la loi d'Arrhenius (1.8), E_a est l'énergie d'activation, k la constante de Boltzman et T la température.

La diffusion ne peut se faire que par des mécanismes d'interactions des impuretés avec les défauts. On définit différents types de placement pour décrire ces défauts :

un atome substitutionnel A_S qui se place dans le réseau cristallin,

un atome interstitiel A_I se trouve entre les atomes du réseau,

un auto-interstitiel I est un atome de silicium interstitiel.

Une lacune V^{15} est l'absence d'un atome à un endroit du cristal.

Ces différents défauts interagissent selon plusieurs mécanismes (lacunaire, interstitiel, *Franck-Turnbull*, *kick-out*, recombinaison interstitiel-lacune) qui résultent en la diffusion des impuretés. Ces défauts se regroupent en défauts étendus (agglomérats, défauts 311, boucles) et évoluent au cours du recuit, influençant la diffusion. On pourra se référer à [80] pour une description complète des mécanismes de diffusion.

Les recuits des implants du substrat ne sont pas à proprement parler critiques. Les recuits des source/drain le sont car on veut des jonctions étroites et donc peu diffusées. La diffusion des dopants dans les recuits rapides est accélérée par les défauts en particulier au début du recuit par la *diffusion accélérée et transitoire* (TED—transient enhanced diffusion) qui peut conduire à des coefficients de diffusion beaucoup plus élevés qu'à l'équilibre. Pour ces raisons, le profil final de dopage est très sensible à la température de recuit.

15. *Vacancy* en anglais.

La non uniformité du budget thermique sur une plaque est critique. On observera ce problème au chapitre 3.

Les conditions de *process* peuvent tout de même être optimisées afin de réduire la sensibilité à cette source de variabilité. Difrenza *et al.* [26] étudient la sensibilité du coefficient de désappariement aux paramètres du procédé. Ils montrent en particulier la sensibilité du facteur de *mismatch* A_{V_t} à la température du recuit RTA – pour *Rapid Thermal Annealing* – et aux doses des implants *ldd*. Ils modifient leur procédé en conséquence pour garder les conditions de fabrication de grille minimisant le désappariement de $4,1 \text{ mV} \cdot \mu\text{m}$ à $2 \text{ mV} \cdot \mu\text{m}$. L'amélioration de l'appariement induit, au niveau du circuit, un accroissement de la linéarité des convertisseurs numérique-analogique de l'ordre de 40 % et une amélioration de 1 bit sur la résolution des convertisseurs analogique-numérique. La marge statique au bruit des circuits SRAM est aussi augmentée de 25 %.

1.3 Modélisation des déviations

Modéliser les mécanismes de variabilité est essentiel à plusieurs niveaux. Les concepteurs doivent pouvoir prédire le rendement des circuits qu'ils dessinent, les modélistes doivent leur fournir des modèles permettant de dissocier les variations globales et locales et prévoyant au mieux les déviations systématiques. Les technologues ont besoin de modèles pour étudier les phénomènes mis en jeu afin d'annuler ou de minimiser leurs influences.

Les problématiques matériau, transistor et circuit utilisent des outils de simulation différents.

1.3.1 Variabilité du process

Simulation avancée de procédés

L'équipe de Martin Jaraiz de l'université de Valladolid a développé plusieurs outils de simulation atomistique des procédés de fabrication [81] pour étudier l'implantation, le recuit et la croissance de matériaux. Dans un simulateur atomistique, ce sont les relations entre atomes qui sont modélisées, le résultat de la simulation étant la conséquence à grande échelle de ces règles

intervenant à l'échelle atomique¹⁶.

Molecular Dynamics permet de reproduire la formation des dommages subis par le silicium et sa recristallisation. Cet outil simule tous les atomes du réseau, avec un pas de simulation constant, ayant pour inconvénient un temps de calcul important, mais les informations obtenues peuvent être pertinentes pour le développement d'autres modèles.

DADOS contrairement à *Molecular Dynamics* se concentre sur les événements de diffusion selon leur probabilité d'apparition et ne simule pas le réseau entier. L'implantation utilise l'approximation de collisions entre deux atomes¹⁷. La prise en compte, au niveau atomique, des interactions des défauts et des dopants pendant le recuit permet d'étudier la dynamique de la diffusion avec une grande prédictivité. Cela donne une référence pour le développement des modèles continus de diffusion. *DADOS* a été acquis par Synopsys et intégré au simulateur Sprocess [82]. Les résultats de ce type d'implantations prennent intrinsèquement en compte la nature discrète des dopants et permettent l'étude des effets électriques de cette source de variabilité [83].

Hane *et al.* ont développé un outil de simulation analogue avec lequel ils génèrent des dispositifs statistiquement différents pour les simuler électriquement dans un simulateur *drift-diffusion* dont le potentiel électrique est calculé à partir des potentiels coulombiens de chaque atome [84].

Modélisation de type TCAD

Les simulations TCAD¹⁸ sont des outils de simulation numérique permettant de simuler en premier lieu la fabrication des dispositifs, puis, à partir du résultat du procédé, de simuler son comportement électrique. Au sens usuel du terme, ce sont des simulations par éléments finis résolvant des équations continues. Par conséquent, elles ne sont pas destinées à prendre en compte la nature atomistique des matériaux. Cependant, les outils dont nous avons parlé au paragraphe précédent sont intégrés dans les suites logicielles de TCAD en tant qu'alternative à la résolution d'équations continues, dans le cas des implantations et des recuits. La nature aléatoire des dopants est ainsi utilisable dans les simulations électriques.

16. à la différence de simulations par éléments finis où les équations résolues portent sur la concentration des impuretés et au flux de diffusion.

17. BCA : Binary Collision Approximation

18. Technology Computer Aided Design

Les variations de largeur de trait de polysilicium dues aux effets de proximité peuvent être prédites à partir du masque à l'aide d'outils de simulation optique, utilisés en particulier pour mettre au point les corrections d'amélioration de la résolution [85]. La rugosité de ligne de polysilicium peut-être introduite dans les simulations de procédé, en trois dimensions, si l'utilisateur définit de son côté les profils de masque de grille à utiliser. Les profils peuvent être générés aléatoirement à partir de l'analyse spectrale de vrais profils observés sur polysilicium et dont on extrait l'amplitude moyenne et la longueur d'autocorrélation. Il est également possible de modéliser cette source par une somme de transistors dont les longueurs suivent statistiquement la rugosité de ligne [86]. Les simulateurs commerciaux sont capables de simuler des transistors présentant des irrégularités aléatoires de grille mais c'est à l'utilisateur d'introduire les géométries correspondantes.

1.3.2 Variabilité du transistor

Simulations TCAD

Les outils de simulation TCAD sont par nature paramétriques. Leur fonction est de pouvoir reproduire les mêmes étapes technologiques qu'en salle blanche [87]. Les données d'entrée d'un simulateur process sont ceux de la recette technologique correspondante : profils de température de recuit, dessin de masque de lithographie, temps et atmosphère d'oxydation, énergie d'implantation, température de recuit, vitesse de gravure, etc. À partir d'un fichier de simulation calibré sur le nœud technologique que l'on étudie, on pourra étudier l'influence des variations d'un paramètre *process* de manière indépendante et en l'absence de perturbation.

Sans moyen de calculs énormes, faire un tir Monte-Carlo sur les paramètres d'entrée d'une simulation du procédé puis de la réponse électrique du transistor, n'est pas vraiment envisageable simplement à cause du temps de simulation. Selon la complexité des modèles choisis dans les options d'une simulation de procédé, le temps de simulation d'un seul transistor peut aller d'une demi-heure à quatre heures¹⁹ pour un fichier de simulation 2D éléments finis de typiquement deux ou trois mille nœuds.

Pour représenter la variabilité, il est alors plus raisonnable d'extraire un modèle en réponse de surface à partir d'un nombre réduit de combinaisons de

19. auquel cas on s'inquiètera de la convergence de la simulation.

paramètres procédé [88]. Le choix de ces points de l'espace des paramètres procédé se fait généralement selon un plan d'expérience. Un plan d'expérience, ou *DoE* pour *Design of Experiment* en anglais, définit un jeu minimal de n points disposés de manière optimale dans l'espace des paramètres du procédé. L'objectif est d'extraire un modèle analytique d'une performance choisie et ce, sans confondre les facteurs, c'est à dire attribuer un effet observé à la mauvaise variable technologique. L'idée est simplement d'avoir un nombre suffisant mais minimal d'équations pour calculer les inconnues que sont les sensibilités de la performance aux facteurs du plan. Une fois le modèle en réponse de surface calculé, il peut être utilisé pour optimiser le procédé technologique [89]. Le même type de méthode peut être utilisé pour la réduction de la variabilité d'un circuit [21].

L'exploitation de plans d'expériences dans le cas de données expérimentales sera abordée au chapitre 3.

Carte modèle paramétrique

Pour faire le lien entre cette méthode et la simulation des circuits, les variables des procédés ont été introduites dans les modèles compacts. On appelle ce genre de modèles des *PCM* pour *Process Compact Model*. Le principe est d'ajouter, à un certain nombre de variables M d'un modèle compact calibré sur un procédé nominal, un terme dépendant des n variables technologiques θ :

$$M_j = M_0 + \sum_{i=1}^n f(\theta_i) \quad (1.10)$$

La figure 1.27 montre l'utilisation du logiciel *Paramos* de Synopsys pour l'extraction de cartes modèles paramétriques. Il est nécessaire de commencer par une extraction classique d'un modèle compact à partir des courbes de performances du transistor nominal ($I_d(V_G)$, $C(V_G)$, etc). Ensuite, on peut passer ce modèle nominal ainsi que les données simulées en présence de variations technologiques selon un plan d'expérience, afin de déterminer la fonction $f(\theta_i)$ lors d'une seconde série d'extractions. Au final, on obtient une carte modèle, *i.e.* un jeu de paramètres que l'on pourra utiliser dans un simulateur pour étudier l'influence des variations du process sur un circuit [90, 91, 92, 93].

Au niveau du dispositif MOSFET, le modèle de Pelgrom (équ. 1.4) est une

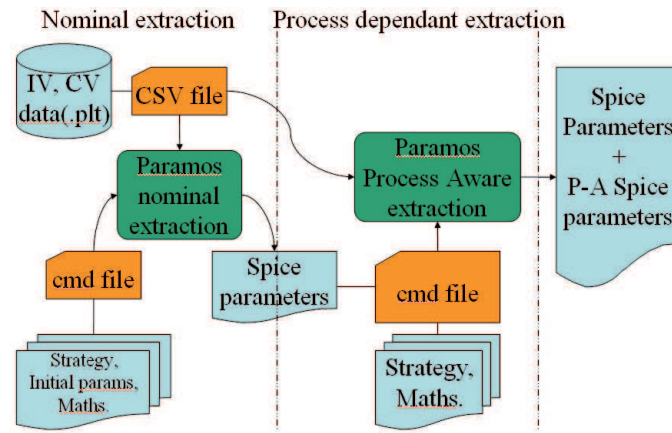


FIGURE 1.27 – Extraction d’un modèle compact paramétrique (PCM) avec le logiciel *PARAMOS* de Synopsys

approche descriptive de la déviation des performances [27]. C’est un facteur de mérite du niveau de variabilité.

1.3.3 Variabilité des circuits

Monte-Carlo et Pire-cas

La méthode de référence pour la simulation des performances d’un circuit reste la méthode Monte-Carlo consistant à pratiquer un nombre statistiquement significatif de simulations en utilisant un groupe de paramètres d’entrée tirés aléatoirement dans leur gamme de variations. L’intérêt est de propager, sans faire d’hypothèse, les fluctuations des paramètres technologiques. Ces simulations étant longues dans le cas de circuit comportant un grand nombre de transistors, les concepteurs utilisent des modèles pires-cas, où *corners*, qui, comme leur nom l’indique, ne simulent que les cas extrêmes d’une performance donnée. Leur défaut principal est d’être spécifiques à un seul groupe de performances (vitesse, consommation, ...) et de supposer des corrélations maximales entre les facteurs technologiques.

Ces deux types de simulation font l’objet du chapitre 4.

Simulation à base de fonction de densité de probabilité

Une alternative consiste à travailler sur les densités de probabilité plutôt qu’itérativement sur chaque évènement statistique. D’Agostino *et al.* ap-

pliquent cette méthode aux courants de fuite [94, 95]. Considérons une relation h entre le courant de fuite et la longueur de grille L telle que $I = h(L)$ (et $g = h^{-1}$), et $f_x(L)$ la densité de probabilité de L . Si on considère pour L une distribution gaussienne de moyenne μ_L et un écart-type σ_L , alors la densité de probabilité du courant causée par les variations de la longueur de grille $f_y(I)$ s'exprime :

$$f_y(I) = \frac{1}{h'(L)} \cdot \frac{1}{\sigma_L \sqrt{2\pi}} \cdot \exp\left(\frac{-(L - \mu_L)^2}{2\sigma_L^2}\right) \quad (1.11)$$

Pour prendre en compte deux paramètres indépendants du procédé (ou plus), ici la longueur de grille L et le dopage canal N_{ch} , on peut multiplier ces fonctions de densité de probabilité ainsi :

$$f_{tot}(I) = \sum_i f(I)_L[i] \cdot f(N_{ch})[i] \quad (1.12)$$

L'équation 1.11 peut être implémentée dans le modèle compact BSIM ou PSP pour chaque paramètre technologique. Les résultats montrés sont très proches des simulations Monte-Carlo prises pour références avec un temps de simulation moindre.

1.4 Conclusion

Nous avons présenté les sources de variabilité causant la dispersion de performances observées dans les dispositifs MOS sur substrat massif d'aujourd'hui. Ces sources peuvent être classées selon trois types : celles causées par la nature des matériaux qui sont aléatoires et dont on ne peut qu'estimer l'amplitude, celles causées par l'influence du placement-routage sur certaines étapes technologiques et qui peuvent en partie être prises en compte dans les modèles de conception et enfin, celle causées par la non uniformité et le manque de répétabilité des procédés de fabrication eux-même.

Ces fluctuations sont étudiées par l'ensemble des spécialités de la micro-électronique et différents types de modélisation existent. Les simulations des matériaux à leur niveau atomique ont accéléré la compréhension des mécanismes clefs de l'implantation et de la diffusion des dopants. L'utilisation de méthodes Monte-Carlo atomistiques pour la simulation électrique de l'influence des sources aléatoires donne une référence pour différencier ces causes

qui ne sont expérimentalement observables qu'en tant que bruit. Les outils de TCAD apportent des informations précieuses aux technologues pour l'optimisation de leur recettes de fabrication et peuvent être associés au développement de modèles compacts prenant en compte les variables technologiques.

La modélisation des effets de la variabilité nécessite la caractérisation des paramètres technologiques. Si la mesure des dimensions géométriques des transistors ne pose pas de problème technique, obtenir un échantillonnage statistique du dopage et des températures de recuit est plus délicat. Le chapitre 2 traite de ce problème.

Ensuite, au chapitre 3, nous étudierons expérimentalement un lot du nœud 45 nm afin de faire un point sur l'état de la variabilité de cette technologie qui servira de base au chapitre 4 où nous aborderons le cas des simulations Monte-Carlo et des simulations « pire-cas » des modèles compacts.

Chapitre 2

Caractérisation du dopage des dispositifs MOSFET

Pour étudier la variabilité des transistors, nous avons besoin de données quantitatives qui représentent, d’une part, la fabrication – les facteurs du procédé – et d’autre part, les performances électriques – les indicateurs. Nous traitons des indicateurs de variabilité et des facteurs du procédé dans la section 2.1. Dans la section 2.2, nous montrons une méthode permettant d’avoir un point d’entrée pour la température de recuit. Dans la section 2.3, nous nous intéressons à la caractérisation de la diffusion latérale des dopants des source/drain en particulier dans le cas de la calibration des simulations TCAD. Dans la section 2.4, nous étudions l’effet de polarisation du substrat dans l’optique de s’en servir dans la section 2.5 où nous tentons d’apporter une méthode permettant de mesurer la variabilité du dopage des dispositifs MOSFET.

2.1 Caractérisation de la variabilité

Les paramètres que nous étudions, de manière générale, sont des performances électriques. La figure 2.1 donne un aperçu d’ensemble de l’utilisation des indicateurs. On y représente l’utilisation des facteurs et indicateurs selon différents niveaux de simulation, verticalement et selon différents niveaux de complexité, horizontalement. Le niveau le plus bas n’est pas une simulation, ce sont les mesures directes. De gauche à droite, on a les facteurs technologiques, puis les indicateurs au niveau transistor, puis les indicateurs au

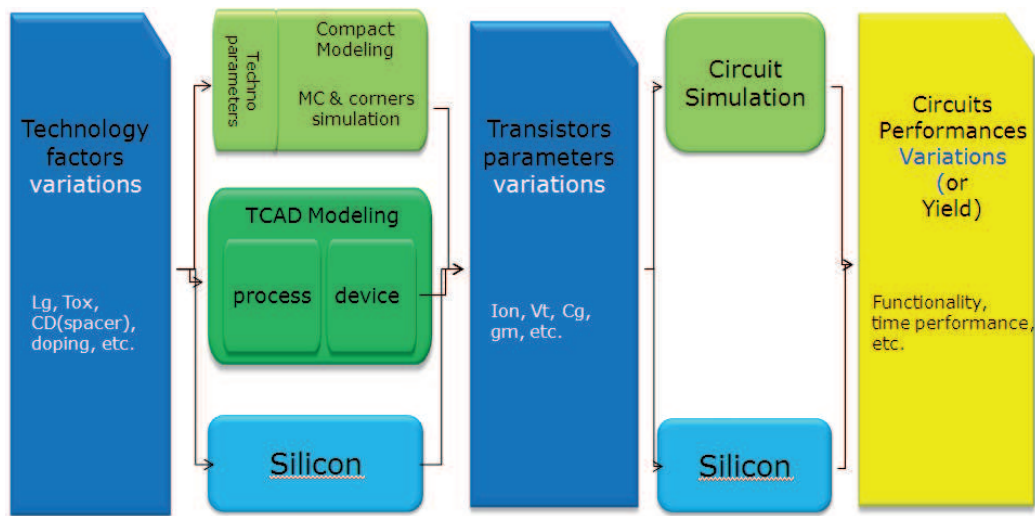


FIGURE 2.1 – Vue d’ensemble de l’utilisation des facteurs et indicateurs de variabilité en microélectronique

niveau cellules standards¹ et enfin les indicateurs de performance au niveau circuit.

Les simulations du procédé reproduisent les performances du transistor en utilisant les facteurs de la recette. On peut utiliser les mêmes variations en TCAD que celles mesurées en fabrication. Les paramètres physiques qui définissent le procédé de fabrication sont utilisés directement par les simulations de type TCAD qui, à l’idéal, reproduisent virtuellement toutes les étapes de fabrication des semiconducteurs afin de simuler toutes les caractéristiques électriques. Il n’est pas coutume de faire des simulations statistiques de ce type, vu la lourdeur de calcul, mais nous verrons dans le chapitre suivant une manière de le faire.

Au niveau de la simulation SPICE, un jeu simplifié de ces facteurs technologiques est utilisé dans les modèles pour représenter la technologie simulée et seuls les facteurs de première importance sont utilisés en simulation statistique. Ces simulations statistiques peuvent être des simulations Monte-Carlo destinées à estimer les distributions de performances ou des simulations *pire-cas* que nous appelons couramment *corners*², destinées à simuler la variation d’un critère de performance donnée (vitesse, consommation). Les per-

1. Les cellules standards sont des circuits comme des portes NAND, NOR, des bascules etc. Les mesures de ces circuits servent à vérifier les simulations sur des performances de circuit mais à un niveau de complexité réduit.

2. Nous détaillerons les simulations Monte-Carlo et corners dans le dernier chapitre.

formances qui sont vérifiées au niveau transistor en simulation SPICE sont les courants, les tensions V_{th} et les capacités. Ensuite, au niveau *standard-cell*, les indicateurs sont des temps de propagation, des courants de cellule, etc. Les méthodes Monte-Carlo ne sont plus utilisées à ce niveau compte tenu du nombre élevé de cellules et des variations de type environnemental (tension d'alimentation, température, ...) qui se superposent aux variations du procédé. On vérifie que les pires-cas répondent aux performances du cahier des charges (tolérances des délais, fonctionnalité des cellules). Au niveau circuit, les simulations ne sont plus de type modèle compact mais plutôt des simulations avec des modèles de haut niveau de type VHDL. On vérifie les délais des chemins critiques via des blocs de circuits mais pas le circuit en entier.

Les indicateurs sont tous mesurables directement à l'étape du test paramétrique. Ce sont des performances électriques comme des courants, des tensions ou des capacités, et on dispose de l'équipement nécessaire. La majorité des facteurs sont des dimensions spatiales. Les épaisseurs, comme celles de l'oxyde de grille par exemple, sont mesurables entre deux étapes de process avec des techniques telles que la scattérométrie. Les largeurs, comme celle des espaceurs ou de la ligne de polysilicium de grille, sont mesurables par microscopie SEM. Il y a principalement deux types de grandeurs que l'on ne peut pas mesurer directement de manière statistique : ce sont les températures et les dopages. Dans ce chapitre, nous allons voir comment quantifier les variabilités de ces grandeurs, dans le cas de la température du recuit rapide et du dopage du canal.

2.2 Caractérisation de la température de recuit

La plupart des caractéristiques des transistors sont fortement influencées par le profil du dopage selon la profondeur dans le substrat depuis l'interface. En particulier, le dopage du canal qui contribue à la définition de la tension de seuil du transistor. Le dopage du canal est déterminé par l'implantation d'ajustement de V_{th} et par le bilan thermique subi par le dispositif. Le résultat de la diffusion des dopants de cet implant soumis à ce bilan thermique donne le profil final des dopants.

La dose implantée est considérée comme étant très fiable car elle est

contrôlée en courant dans les implanteurs. Mais l'information de la température reçue par chaque puce n'est pas accessible car pendant le recuit on n'a pas de sonde donnant la température reçue par chaque site de chaque tranche.

Un des recuits est très critique dans le procédé : c'est le recuit rapide arrivant à la fin des étapes de *front-end*. Durant ce recuit, des lampes soumettent la plaque de silicium à une température de l'ordre de 1000 °C pendant un temps de l'ordre de la seconde. La fonction de cette étape est d'activer les dopants des source/drain sans pour autant les faire diffuser trop. Le problème est que, si petites que soient les déviations de la température de ce recuit, la sensibilité des performances électriques du transistor est suffisamment grande. Ce recuit n'est pas le seul bilan thermique important dans la diffusion de l'implant V_{th} mais il intervient après l'implantation des source/drain, implants auxquels l'électrostatique est très sensible.

2.2.1 Estimation de la variation spatiale de la température reçue

Il est possible de corréler la résistance de surface R_S du silicium aux températures de recuit rapide *spike* de la fin du process.

Pour connaître l'effet de la température de ce recuit, nous devons nous trouver dans une situation où nous pouvons nous affranchir des variabilités des autres paramètres du transistor. On propose d'utiliser des plaques n'ayant subi qu'un implant et qu'un recuit.

Les plaques ont un dopage de substrat N. On y crée un oxyde de 24 Å d'épaisseur puis on implante du bore à 2 keV et à $3 \cdot 10^{15} \text{ at} \cdot \text{cm}^{-2}$. On applique à un lot de plaques trois températures de recuit différentes. Un tiers du lot subi 997 °C, un autre 1000 °C et le dernier 1003 °C. Ensuite, les plaques subissent une désoxydation au *HF* puis les résistances sont mesurées.

Si on fait l'hypothèse que les profils de dopant *as-implanted* sont suffisamment peu variables, alors les seules variations que nous verrons sont dues au recuit. On peut alors mesurer la résistance de surface, qui est fonction du dopage du semiconducteur, sur toutes les plaques, sur toute leur surface. La figure 2.2 montre la cartographie qui a été utilisée pour cette étude.

Ensuite, on peut calculer la moyenne des résistances par groupe de plaques. La figure 2.3b montre ces moyennes. La corrélation avec la température appliquée devient évidente et on peut établir la sensibilité de l'indicateur résistance

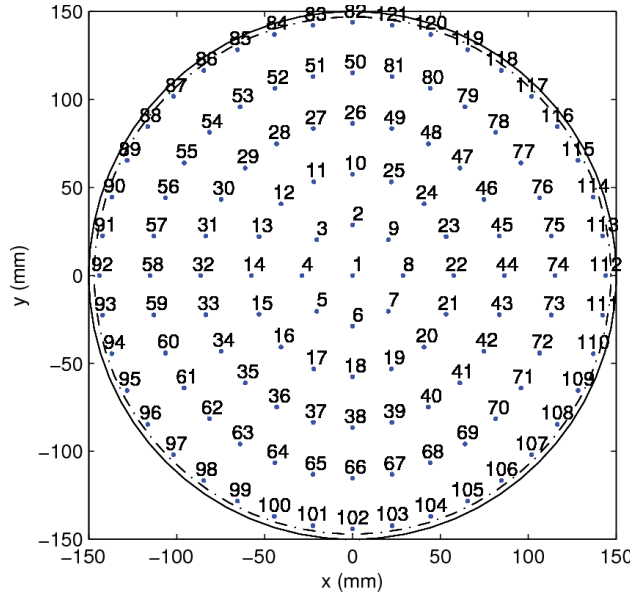


FIGURE 2.2 – Mapping des mesures de R_s sur une plaque ayant subi uniquement un implant et un recuit flash pour évaluer la fluctuation de température reçue.

R_s au facteur température reçue $\text{Spike}_{\text{temp}}$. On établit que $R_s = (aT_{\text{spike}} + b) \Omega$, avec $a = -2,7674$ et $b = 3111,3$, donc $T_{\text{spike}} = -0.36R_s + 1123^\circ\text{C}$.

Une fois cette sensibilité établie, la transformation des valeurs de résistance mesurées à chaque site en valeurs de température reçue est directe. Les figures 2.4a et 2.4b montrent pour la plaquette numéro 15 les résistances de surface et les températures associées.

Cette grandeur, T_{spike} , va nous servir de facteur – ou grandeur d’entrée – pour les études à venir. Pour l’instant, nous pouvons noter l’homogénéité de la distribution spatiale de la température sur les différentes plaques. Cette distribution est plus ou moins radiale, comme d’autres indicateurs que nous verrons plus tard.

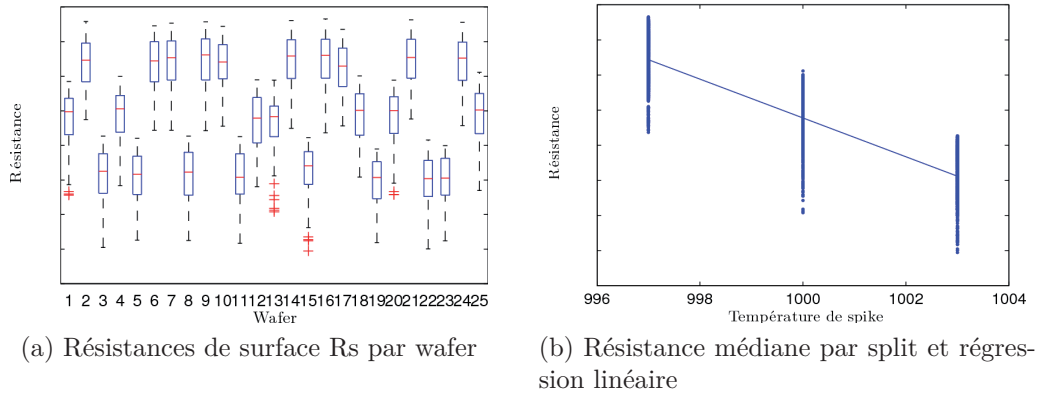
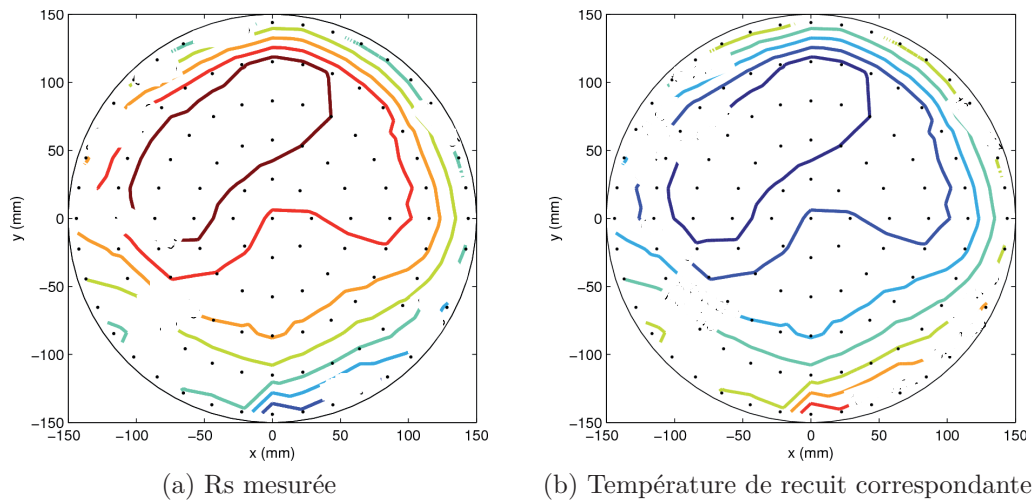
FIGURE 2.3 – Correlation de R_s à la température appliquée.

FIGURE 2.4 – Estimation de la distribution de la température de recuit spike reçue par le wafer.

2.3 Caractérisation électrique du courant de perçage comme indicateur du profil latéral du dopage des source et drain

Un contrôle parfait des distributions de dopage est impératif dans les technologies CMOS avancées et cela, à cause de la sensibilité des performances nominales des transistors aux dopages des source/drain et du canal. La profondeur de la jonction source/drain impacte directement les effets de canal court. La position latérale des jonctions détermine la longueur effective du canal ainsi que les capacités de recouvrement entre la grille et les source/drain. Le contrôle de ces positions dans les technologies d'aujourd'hui doit avoir une précision de l'ordre du nanomètre. Bien qu'un certain nombre de méthodes permettent la mesure du dopage vertical avec cette précision, la mesure des profils latéraux est compliquée et requiert d'appliquer des mesures spécifiques à des échantillons préparés spécialement [96, 97]. Dans cette section, nous discutons de la sensibilité des mesures électriques par rapport à la distribution latérale du dopage et de la possibilité de caractériser ledit profil.

2.3.1 Principe de la méthode

La méthode proposée repose sur l'utilisation d'un jeu d'échantillons où l'on fait varier uniquement la longueur de grille [98]. La variation de la longueur de grille entraîne un recouvrement progressif des profils de source et de drain et donc des zones de déplétions associées (figure 2.5). Le recouvrement sous la grille par ces profils latéraux est défini par la longueur de grille qui doit être connue et par la diffusion des dopants des source/drain sous la grille. Ce recouvrement change significativement le comportement de la structure de test qui peut, à son tour, être utilisée pour caractériser le profil de dopage latéral.

En premier lieu, l'influence du recouvrement des profils sur la capacité de grille et sur les courants de drain a été étudiée *via* des simulations technologiques et électriques à l'aide de la suite logicielle *Sentaurus* de *Synopsys* [99]. Il en résulte que la sensibilité des courants de drain est bien plus prononcée que celle des capacités. Un schéma de la mesure proposée est en figure 2.6.

Hormis la longueur de grille, les structures de test dont on étudie le profil

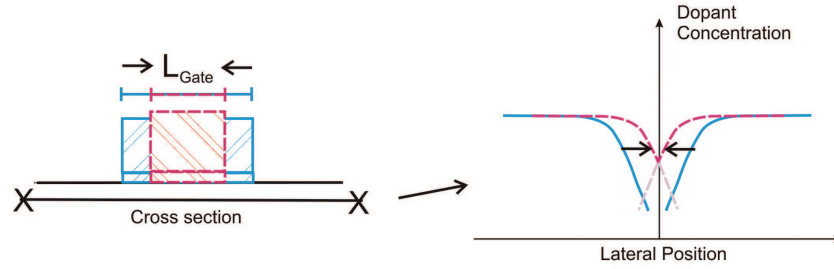


FIGURE 2.5 – Représentation schématique de l'influence de la réduction de la longueur de grille sur le profil latéral de la concentration de dopants.

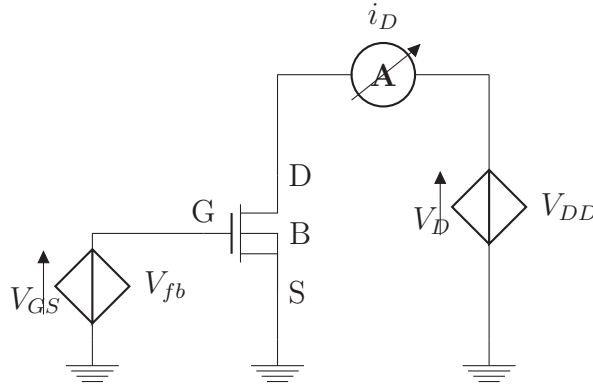


FIGURE 2.6 – Schéma de mesure électrique pour caractériser le profil latéral de dopants.

latéral sont identiques. Afin de concentrer l'étude sur les effets du dopage latéral source/drain, nous utilisons un dopage canal uniforme. Le dopage des source/drain est formé par implantation d'ions puis par recuit d'activation.

Pour caractériser la distribution latérale du dopage, nous suggérons de mesurer le courant de fuite I_D à faible tension de drain V_{DS} et à une tension de grille constante et proche de la tension V_{fb} de bandes plates. La forme du profil latéral est alors estimée par ingénierie à rebours en utilisant les simulations technologiques et électriques. La méthode présentée ici se distingue des autres techniques de modélisation inverse [100, 101] par l'utilisation de longueurs de grille causant un recouvrement très significatif des profils de source et de drain.

Une première analyse de sensibilité à été conduite sur une architecture pMOS simplifiée avec un dopage constant de $10^{17} \text{ at} \cdot \text{cm}^{-3}$ en appliquant des modèles de recuit rapide finement calibrés [102, 103]. La figure 2.7, où le

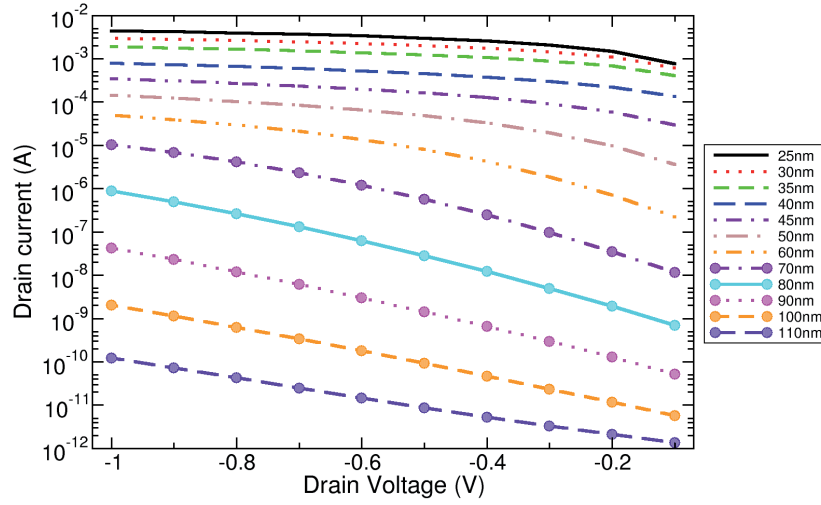
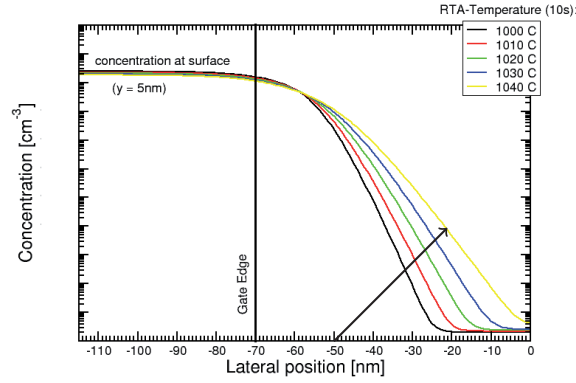


FIGURE 2.7 – Courant de drain I_d en fonction de la tension de drain V_{DS} dans un pMOSFET et pour différentes longueurs de grille. ($W = 1 \mu\text{m}$, $V_g = +0.6 \text{ V}$)

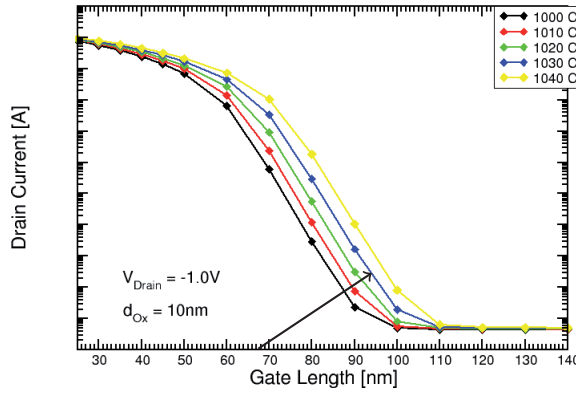
courant de drain I_D est tracé en fonction de la tension de drain V_{DS} , montre le résultat de ces simulations. Le courant I_D dépend fortement de la longueur de grille et varie d'environ 9 ordres de grandeurs de 25 à 110 nm, suggérant une forte sensibilité à la position et la forme du dopage source/drain.

Une sensibilité similaire est attendue si la variation de position des jonctions source-canal et drain-canal est causée par une variation de température lors du recuit. La figure 2.8 montre des simulations où la position de la jonction canal-drain a été modifiée en balayant volontairement la température maximale du recuit post-implantation par pas de 10° K (figure 2.8a). On observe que les décalages de la jonction sont également bien représentés sur les caractéristiques électriques avec une résolution de 2 nm (figure 2.8b).

La méthode présentée dans cette section, basée sur la mesure du courant I_D pour différentes longueurs de grille, permet donc d'estimer l'extension latérale du profil de dopage. Dans les sections suivantes, la précision de la méthode sur un procédé MOS complet ainsi que sa sensibilité à la variabilité de ce process sont évaluées.



(a) Profils de dopage latéraux pour différentes température de recuit rapide.



(b) Effets du décalage de la jonction sur la caractéristique électrique.

FIGURE 2.8 – Simulation d’une variation de température de recuit.

2.3.2 Application de la méthode sur des simulations TCAD d’un procédé industriel

Pour appliquer cette méthode, nous simulons un procédé générique du nœud 65 nm en utilisant des longueurs de grille allant de 35 à 90 nm. Les simulations de procédé ont été réalisées avec l’outil Sentaurus de Synopsys en utilisant les modèles de diffusion récents calibrés pour les jonctions ultra-fines (USJ) [104], tels que décrits dans [105].

Les principales caractéristiques de ce procédé MOS sont : un oxyde de grille nitruré de 1,45 nm d’épaisseur, un dopage rétrograde du substrat comprenant l’implant d’ajustement de la tension de seuil V_{th} et l’implant anti-perçage, suivi d’un recuit rapide RTP. Les LDD et source/drain sont implantés puis activés *via* un recuit rapide autour de 1050 °C à la fin du procédé.

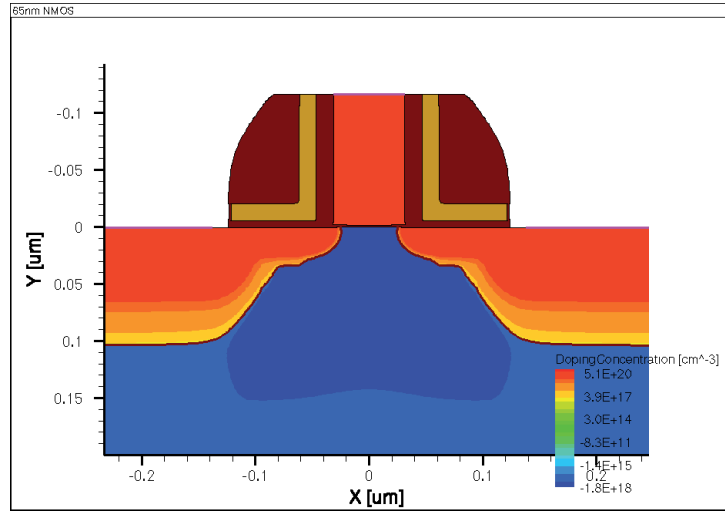


FIGURE 2.9 – Carte de dopage de la simulation 2D du NMOS.

La distribution 2D du dopage obtenue à la fin de la simulation est montrée dans la figure 2.9. Dans ce profil, la pente de la jonction entre *ldd* et canal est d'environ 2-3 nm par décade dans la direction du canal et de 10 nm par décade verticalement entre les source/drain et le substrat.

Comme expliqué dans la section précédente, la caractérisation du profil latéral est basée sur la mesure du courant I_D résultant du perçage entre le drain et la source. La caractéristique électrique est extraite en balayant V_{DS} avec V_G proche de V_{fb} , pour éviter une redistribution des porteurs due à la tension de grille. Le résultat est montré en figure 2.10 pour un n-MOSFET avec une longueur de grille allant de 35 à 90 nm et une largeur de 1 μm .

Les simulations électriques du nMOSFET sont basées sur le modèle de dérive-diffusion incluant les modèles les plus avancés, à l'exception de l'effet tunnel bande-à-bande. Les effets quantiques sont aussi négligés dans ce régime de fonctionnement.

La figure 2.10 démontre la forte dépendance de I_d sur la longueur de grille dans ce procédé NMOS comprenant toutes les étapes principales et valide le principe de la méthode. Cette validation est importante car ce flot de fabrication inclut des étapes conduisant à un profil à fort caractère 2D, comme des implants rétrogrades, avec anti-perçage, dans le *p-well* et des implants *poches* qui élèvent la concentration de bore dans le canal, et donc la barrière de potentiel, afin de limiter le courant de fuite des jonctions.

Afin d'éviter l'influence du mécanisme d'effet tunnel bande-à-bande qui

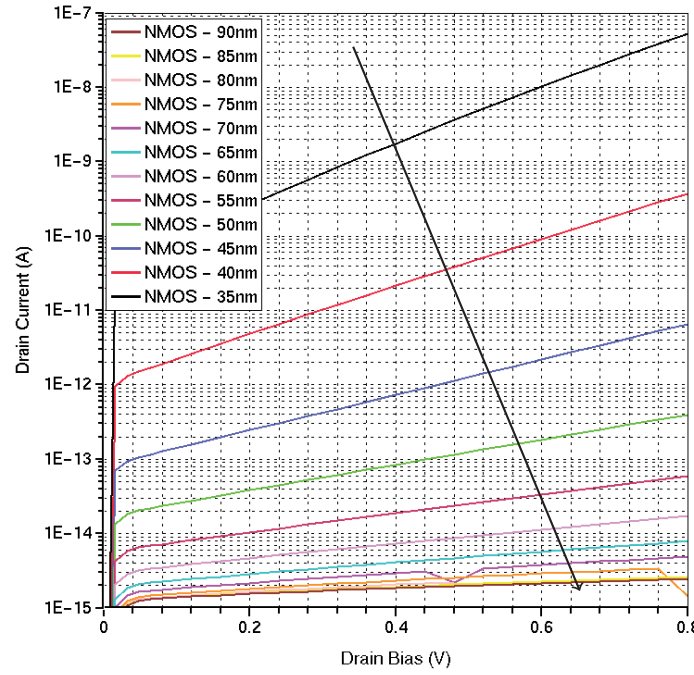


FIGURE 2.10 – I_d vs. V_{DS} ($V_g = -0,5$ V) pour différents transistors. La flèche indique les L_g croissant.

peut arriver dans un transistor réel, le courant de fuite n'est pas mesuré à $V_D = V_{DD}$, même si la sensibilité de la caractérisation serait maximisée dans ces conditions. Le courant de fuite I_D est donc mesuré à $V_D = 0,4$ V. La figure 2.11 montre I_D pour différentes longueurs nominales de grille. Le comportement est similaire à celui de la figure 2.8b.

2.3.3 Étude de l'impact de la variabilité du procédé à l'aide d'un plan d'expériences

La variabilité a de nombreuses sources dans les recettes de fabrication des semiconducteurs modernes [10]. Les effets parasitant la sensibilité de la méthode de caractérisation qui nous intéressent sont les décalages non intentionnels du procédé que nous appellerons *variabilité paramétrique*.

Dans cette section, nous évaluons l'incertitude sur la mesure de I_D en présence de variations de paramètres du procédé. Pour cela, le comportement de I_D doit être connu et modélisé en une fonction des facteurs technologiques. Un modèle en surface de réponse (RSM) peut être établi à partir de l'analyse

Facteur (nm)	Data Exp. ^{les}		Points du plan				
	Moyenne	Écart-type	-2	-1	0	+1	+2
T_{ox}	1.945	0.0056	1.89	1.92	1.945	1.98	2.0
L_{gate}	65	0.6	59	62	65	68	71
TH_{teos}	70	1.1	59	64.5	70	75.5	81
HCD_{th}	41	0.95	31.5	36.2	41	45.7	50.5

TABLE 2.1 – Construction du plan d’expériences

d’un plan d’expériences.

Les paramètres les plus influents ont été identifiés dans une étude précédente [106] et sont : 1) la longueur de grille (CD-gate, L_g), 2) l’épaisseur de l’oxyde de grille (T_{ox}), 3) l’épaisseur du *liner* (TEOS) et 4) la largeur de l’espaceur, paramétrée *via* l’épaisseur du dépôt de nitrure associé (HCD-TH.)

Nous avons opté pour un plan centré composite (cube+étoile) [107]. Ce type de plan a cinq niveaux par facteur et nécessite 25 simulations pour les quatre facteurs. La matrice des facteurs normalisés est parfaitement orthogonale et permet de construire un modèle RSM du second ordre. Le domaine d’étude est fixé à $\pm 5\sigma$ de la variabilité des facteurs technologiques observée. Les valeurs des facteurs du plan d’expériences sont détaillées dans le tableau 2.1 pour une longueur nominale de grille de 65 nm.

Ce plan de simulations a été appliqué autour des longueurs de grille nominales 35, 60 et 85 nm. Ces trois plans résultent en trois modèles RSM obtenus par régression linéaire multiple, en gardant seulement les termes significatifs du polynôme. Les minima et maxima des trois polynômes sont ensuite déterminés en bornant les quatre facteurs d’entrée à $\pm 2\sigma$ des variabilités mesurées, représentant ainsi 95% de la population (admise normale).

La figure 2.11 montre les courants I_D extraits de la figure 2.10 superposés aux barres de variabilité estimées à partir des RSM. La méthode a une très bonne sensibilité au profil latéral pour les plus courtes longueurs de grilles et au fur et à mesure que le phénomène de perçage s’installe. La résolution est de l’ordre de 2 décades pour 10 nm de longueur de grille autour de 45 nm.

Une nette augmentation de l’incertitude sur I_D apparaît pour les faibles valeurs de L_g . On peut observer que l’erreur causée par la variabilité paramétrique est forte dans les petits transistors lorsque la sensibilité de I_D est grande. Comme le montre la figure 2.11, la variabilité du procédé mène à une incertitude sur la localisation du profil de dopage de ± 2 nm, comme détaillé

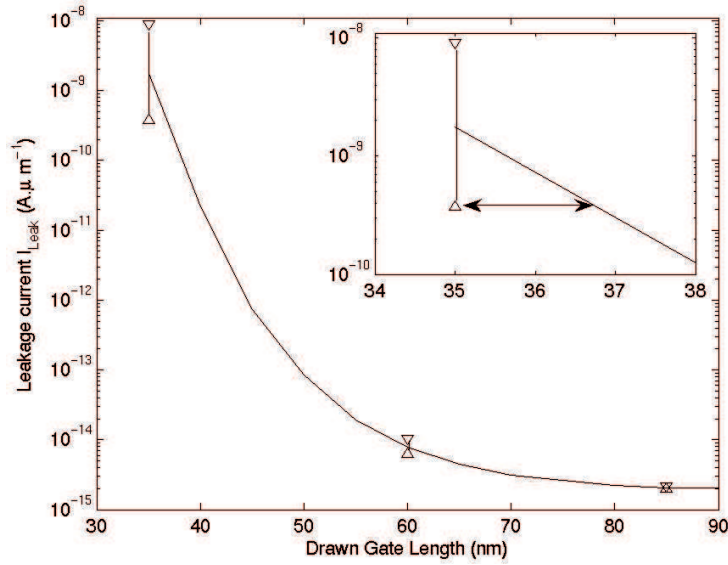


FIGURE 2.11 – Simulations TCAD de I_D vs L_g (ligne) et intervalles d'incertitude calculés à partir du modèle pour $\pm 2\sigma$ (barres verticales)

dans l'insert.

Cette variabilité est à comparer aux gradients de dopage des jonctions des technologies actuelles qui est de l'ordre de 2-3 nm par décade. On peut donc dire que la sensibilité à la variabilité du process est un facteur limitant à l'utilisation de cette méthode de caractérisation électrique du profil latéral. Toutefois, la précision reste meilleure que celle des méthodes conventionnelles.

2.3.4 Mesures sur des nMOS 65 nm

Nous avons ensuite testé la faisabilité de notre méthode de caractérisation électrique sur une technologie mature 65 nm. Ce nœud est basé sur un oxyde de grille d'environ 20 Å. Ensuite, viennent les implants substrats standards suivis par les recuits RTP standards. Les spécifications des implants *ldd* et source/drain sont dans la gamme des énergies et des doses définies dans le cadre du projet ATOMICS. Un recuit *spike* termine le flot d'étapes du procédé.

Les mesures de I_d vs V_{DS} à $V_g = -0.4$ V sont tracées sur la figure 2.12 pour des longueurs de grille allant de 55 nm à 10 μm, correspondant aux courbes A, B, C... dans l'ordre croissant. Ces caractéristiques électriques

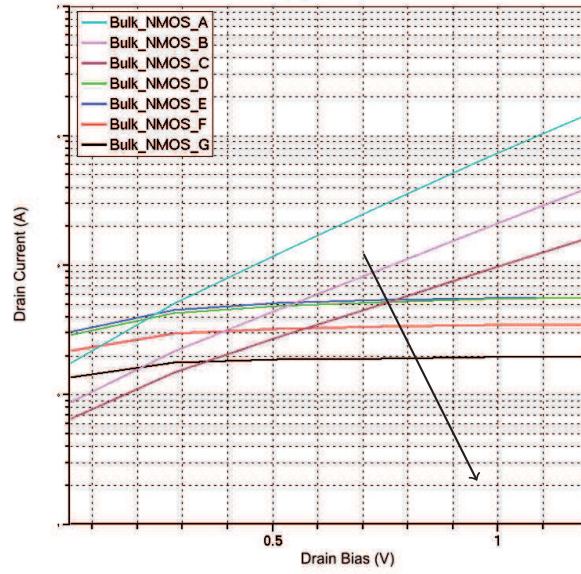


FIGURE 2.12 – Mesures de I_d vs V_{DS} ($V_g = -0,4$ V). L_g croît dans le sens de la flèche.

montrent la croissance exponentielle du courant de fuite I_D comme décrite dans la section précédente. La figure 2.12 montre un effet inattendu pour les faibles valeurs de V_{DS} résultant en un courant plus faible pour les petits transistors que pour les grands. Cela est dû à la redistribution des poches et du dopage substrat dans le cas des canaux courts, appelée *RSCE* pour *reverse short channel effect*. La croissance non monotone de I_D qui peut être expliquée qualitativement, doit être étudiée plus en détail. Comme le procédé réel présente un RSCE important, la méthode devrait être appliquée sur des structures de test dépourvues d'implants de poche.

Au final, la méthode montre sa capacité à distinguer des différences dans le profil de dopants qui apparaissent si la température de recuit change de seulement quelques degrés. La précision absolue de la méthode tient à la précision des modèles de simulations et peut toujours être améliorée par un calibrage précis des modèles pour une longueur de grille donnée. D'autre part, des informations qualitatives sur la distribution de dopants peuvent être extraites avec une résolution spatiale importante à partir de caractéristiques électriques. C'est donc un outil valorisable dans le cadre du calibrage de simulations TCAD des technologies submicroniques.

2.3.5 Conclusion

La possibilité de caractériser le profil de dopage latéral source/drain à partir de mesures électriques a été explorée. Les mesures électriques concernant le perçage entre drain et source se trouvent être suffisamment sensibles à la distribution latérale des dopants pour permettre la caractérisation du profil des jonctions. Cette méthode a été vérifiée *via* des simulations numériques TCAD et sa sensibilité aux fluctuations du procédé a été évaluée. Les mesures électriques sur des dispositifs réels de technologie 65 nm démontrent la bonne sensibilité des caractéristiques électriques et ont indiqué les difficultés potentielles d'interprétation sur des transistors présentant un dopage non uniforme du canal. Toutefois, cette technique apparaît plus précise que les techniques de caractérisation physique et se présente comme un bon outil de vérification du calibrage des profils de dopants en TCAD.

2.4 Caractérisation électrique de l'effet de la polarisation du substrat comme indicateur du profil vertical dans le canal

Dans cette section, nous détaillons une méthode visant à déterminer le dopage du canal de manière non destructive, à partir de mesures électriques, ce qui permet un large échantillonnage de cette grandeur. Les mesures que nous utilisons sont la tension de seuil V_{th} pour différentes tensions de substrat V_B ainsi que la capacité de grille C_{gg} . Ces mesures sont aujourd'hui des mesures standards et ne provoquent donc pas de surcout de caractérisation.

Comme nous déterminons une section du profil de dopage, nous obtenons deux informations qui sont :

- le niveau de dopage dans la région observée,
- et la pente du dopage.

2.4.1 L'effet substrat

L'effet substrat est l'influence de la tension de substrat V_B sur la tension de seuil V_{th} du dispositif. Cette influence de V_B sur V_{th} passe par la zone de déplétion du substrat. En effet, la valeur de V_{th} dépend de la charge associée

à la déplétion. On a :

$$V_{th} = \Phi_{ms} - 2\Phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \quad (2.1)$$

où Φ_{ms} est la différence de travaux de sortie du matériau de grille et du semiconducteur du canal, Q_{ox} représente les charges d'interfaces et Q_B est la charge de la zone de déplétion. Le comportement de V_{th} va dépendre du dopage puisque la charge de déplétion Q_B est liée à la somme du dopage N_a sur la zone de déplétion de profondeur w_D : $Q_B = q \int_0^{w_D} N_a(x) dx$.

Cette dépendance de V_{th} à V_B s'écrit classiquement dans les modèles compacts comme :

$$\Delta V_{th} = \gamma \sqrt{2\Phi_f + V_b} - \gamma \sqrt{2\Phi_f}, \quad (2.2)$$

$$\text{avec } \gamma = \frac{\sqrt{2qN_a\epsilon_{si}}}{C_{ox}} \quad (2.3)$$

$$\Phi_f = \frac{k_B T}{q} \ln(N_a/n_i) \quad (2.4)$$

où γ est appelé *body factor*, Φ_f est le niveau de Fermi dans le canal, n_i est la quantité de porteurs intrinsèques et ϵ_{si} est la constante diélectrique du silicium. Les deux termes γ et Φ_f dépendent du dopage N_a . Le facteur γ dépend aussi de l'épaisseur d'oxyde *via* $C_{ox} = \epsilon_{ox}/T_{ox}$.

Dans cette équation, la dépendance à V_B est linéaire selon $\sqrt{2\phi_f + v_b}$. Si on trace ΔV_{th} en fonction de $\sqrt{2\phi_f + v_b}$, les valeurs de V_{th} de chaque polarisation s'alignent sur une droite de pente γ , comme le montre la figure 2.13b.

2.4.2 Extraction des paramètres γ et ϕ_f

En pratique, nous n'avons pas accès au potentiel ϕ_f ni au paramètre γ . Pour tracer cette caractéristique, il nous faut connaître leur valeurs. Nous procédons alors à un ajustement des mesures à (2.2) en jouant sur la valeur du dopage dont dépendent γ et ϕ_f . Nous avons codé cet ajustement à l'aide du logiciel *Matlab*® et, en particulier de la fonction *fminsearch* qui pratique itérativement une minimisation non linéaire non contrainte. La fonction que nous lui donnons à minimiser est la somme des carrés de la différence entre les valeurs de $\Delta V_{th}(V_b) = V_{th}(V_b) - V_{th}(V_b = 0)$ pour les deux premières polarisations en V_B et l'équation (2.2). Le résultat de cette minimization est un couple de valeurs correspondant à un même dopage N_a , cSe qui consiste

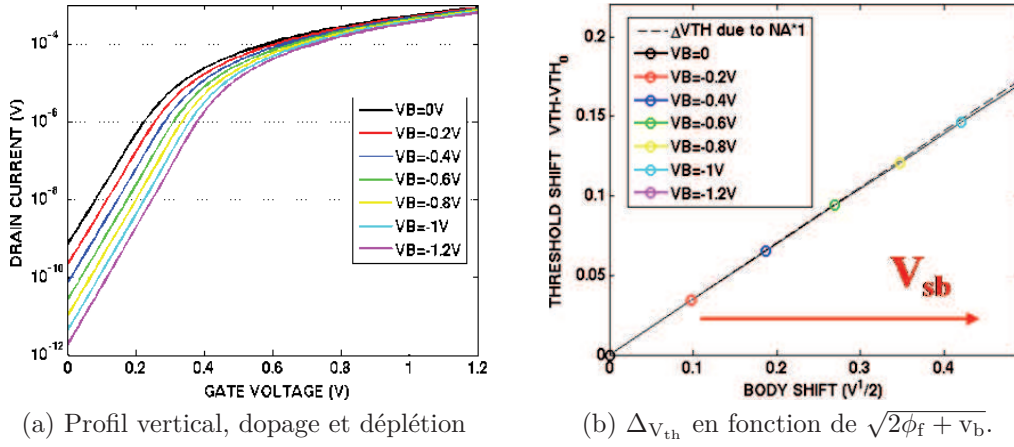


FIGURE 2.13 – a) Courant de drain en fonction de la tension de grille pour des tensions de substrat de 0 à $-1,2$ V. b) L'extraction de γ et ϕ_f permet de modéliser la déviation de la tension de seuil en fonction de V_B .

à ajuster une droite avec le paramètre N_a . La figure 2.13b montre cette caractéristique en racine de V_B pour un dispositif ayant un dopage constant. L'équation (2.2) reproduit parfaitement l'effet du substrat.

Dans ce cas, l'hypothèse est que le dopage est constant en fonction la profondeur dans le canal, hypothèse non applicable dans les canaux des dispositifs actuels ayant un implant d'ajustement du V_{th} et surtout si on considère les applications circuit où l'on ajuste V_B , soit pour améliorer les performances, soit pour baisser la consommation à l'état de veille [108]. La figure 2.14 montre le cas d'une simulation d'un dispositif ayant un dopage représentatif des technologies d'aujourd'hui et comporte un implant *well* en profondeur, un implant *anti punch-through* à la profondeur des jonctions source/drain et surtout un implant *adjust-vt* en surface.

2.4.3 Caractérisation de l'effet substrat dans le cas d'un dopage réel

La figure 2.15b montre le résultat de la caractérisation en $\sqrt{2\phi_f + v_b}$ sur les tensions V_{th} issus d'un transistor ayant un dopage réel. Comme on peut s'y attendre, la droite d'ajustement passe par les deux premières polarisations, mais le modèle (2.2) ne reproduit plus les autres déviations de V_{th} . L'écart entre les points mesurés et la droite d'ajustement, que nous pouvons

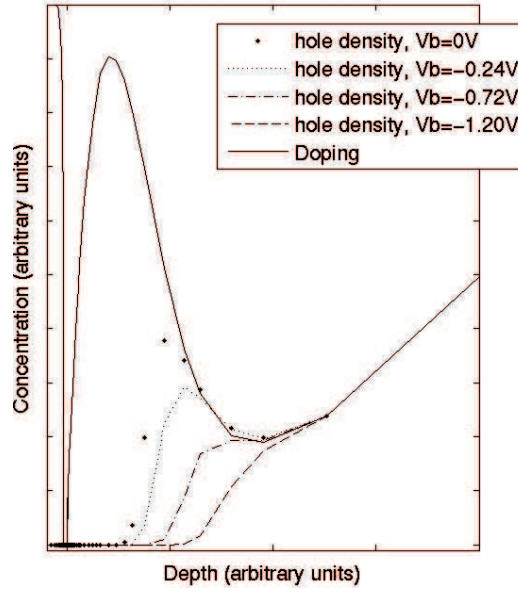


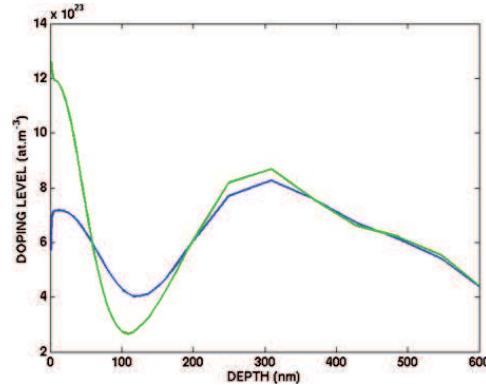
FIGURE 2.14 – Profil vertical de dopants et concentration de trous à $V_g = V_{th}$ pour plusieurs V_B . Le front de déplétion balaie le profil du dopage.

appeler la droite du dopage réel équivalent, reflète les caractéristiques de non uniformité du profil de dopage.

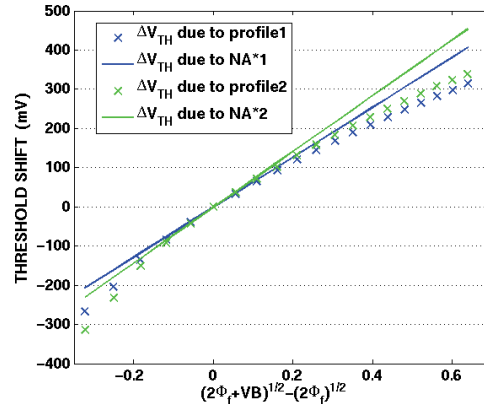
Grâce aux outils de TCAD, nous avons observé le comportement de cette caractéristique face à différents dopages dans des dispositifs de grande longueur de grille. Nous avons étudié le cas de dopages dont le niveau moyen est décalé, puis le cas de différentes pentes de dopage dans la gamme de profondeur où la couche de déplétion varie selon V_B .

Le niveau de dopage, comme dans le cas du dopage uniforme, change la pente de la caractéristique. Cet effet était attendu car γ est fonction d'un dopage supposé constant dans le canal : on s'attend donc à un effet analogue en fonction de la valeur moyenne du dopage. Ensuite, nous avons observé l'effet d'un changement de pente sur la caractéristique électrique pour des valeurs de V_B fortement positives ($V_B > 0,2$ V) ou négatives ($V_B < -0,2$ V). Pour cela, nous avons simulé des profils arbitrairement gaussiens ayant un écart-type différent. Les caractéristiques obtenues présentent une déviation des ΔV_{th} par rapport à la droite de dopage constant équivalent qui s'accroît quand l'évolution du dopage est plus rapide.

Ces informations peuvent nous permettre de comparer qualitativement deux transistors et même servir de guides supplémentaires pour calibrer un



(a) 2 profils simulés représentatif du nœud 45 nm.



(b) ΔV_{th} en fonction de $\sqrt{2\phi_F + v_b}$ dans le cas du dopage de la figure 2.15a.

FIGURE 2.15 – Caractérisation de l’effet de la polarisation substrat sur des dispositifs simulés en TCAD.

jeu de simulations TCAD, mais ne nous donnent pas d’information quantitative sur le profil lui même. Pour cela, nous proposons une méthode dans la section 2.5.

2.4.4 Effet de la longueur de grille sur cette caractérisation

La caractérisation que nous venons de détailler est clairement faite pour les dispositifs de grande longueur où l’influence des source et drain est négligeable. Dans le cas de dispositifs courts, le contrôle électrostatique du canal

par les source/drain vient dominer celui de la grille et masque l'influence du dopage canal sur l'effet substrat. Dans ce cas, nous observons aussi une forte déviation des ΔV_{th} mais qui ne reflète pas uniquement la non-uniformité du dopage vertical.

2.5 Détermination du dopage canal

2.5.1 Méthode

La figure 2.16 présente la validation de la méthode de détermination du profil vertical de dopage qui sera détaillée dans la section 2.5.3.

1. En premier lieu, nous simulons plusieurs transistors à canal long ($10\text{ }\mu\text{m}$). Nous connaissons donc le profil de départ.
2. Nous simulons les courbes $I_d(V_G)$ à différents V_B .
3. La valeur de la capacité de grille en inversion est également déterminée par simulation.
4. Nous appliquons notre méthode détermination du dopage.
5. Enfin, nous comparons le résultat avec le profil de départ.

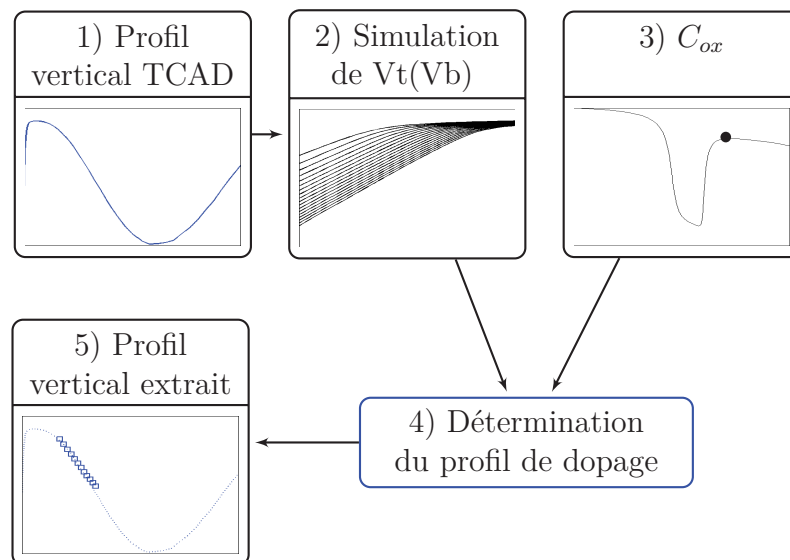


FIGURE 2.16 – Validation de la méthode de détermination du profil de dopage du canal grâce aux simulations TCAD.

2.5.2 Simulations TCAD

Pour cette validation, nous avons utilisé des simulations représentant un MOSFET de type N du noeud 45 nm. Les deux différents profils viennent de l'utilisation de paramètres de diffusion et de ségrégation à l'interface oxide silicium différents.

Nous simulons, ensuite, les tensions de seuils pour une polarisation du substrat allant de 0,45 V à -2,1 V. Le choix de la gamme de polarisation détermine la profondeur de déplétion et donc la zone d'exploration du profil de dopage vertical car le terme de la charge de déplétion Q_B , entrant en jeu dans la tension de seuil, est dépendant de la tension de substrat (voir (2.1)). La gamme de profondeur explorée sera fonction également du niveau de dopage. Plus le dopage est fort, moins la zone de déplétion est profonde. Et moins son augmentation en fonction de la tension de substrat V_B est grande.

2.5.3 Détermination du profil vertical à partir de l'effet substrat

Dans son article de 1987 [109], Wright modélise l'effet substrat. Dans les détails de ses calculs, en annexe, il réduit à une dimension l'équation de Poisson et un de ses calculs intermédiaires exprime la profondeur de déplétion ainsi que le dopage associé en fonction de la tension de seuil V_{th} et de la tension de substrat V_B .

$$x = -\frac{\varepsilon_{Si}}{C_{OX}} \frac{dV_{BS}}{dV_{th}}, \quad (2.5)$$

$$N(x) = -\frac{C_{OX}^2}{q\varepsilon_{Si}} \left(\frac{d^2V_{BS}}{dV_{th}^2} \right)^{-1}, \quad (2.6)$$

C'est exactement ce qu'il nous faut. Notons que ces calculs viennent d'une approximation en une dimension de l'équation de Poisson et que, par conséquent, ils ne sont valables que dans le cas du transistor long où l'électrostatique du canal est largement majoritaire sur celle des jonctions.

Nous appliquons (2.5) et (2.6) aux résultats électriques correspondants aux profils en trait plein de la figure 2.17. La figure 2.17 montre l'adéquation des 2 profils testés avec le résultat de la méthode. On observe la gamme de profondeur vue électriquement via cette technique. Le point le plus profond correspond à la tension de substrat la plus négative (NMOS) que nous

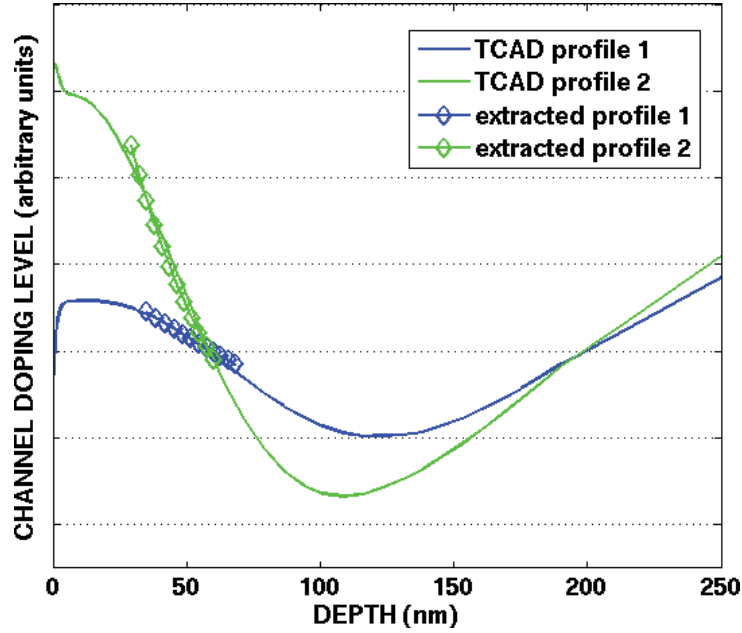


FIGURE 2.17 – Dopage en fonction de la profondeur. Profils déterminés avec les équations de Wright et superposés aux profils TCAD ayant servi à générer les simulations.

balayons et la moins profonde correspond à la tension de substrat la plus positive que nous pouvons utiliser sans placer la diode de jonction source/body en direct.

2.6 Variabilité du profil de dopage vertical

Dans cette section, nous appliquons à des mesures la méthode détaillée et validée précédemment.

2.6.1 Présentation des données expérimentales

Le lot mesuré est un lot 45 nm dont toutes les plaques ont subi un process unique – pas de split. Ce lot contient 25 plaques, 66 sites par plaque, ce qui nous donne 1650 puces mesurées. Le jeu de données contient la tension de seuil V_{th} pour des valeurs de V_B balayées de $-0,6$ V à $+0,6$ V, sur 2 longueurs de grille. Le V_B le plus positif que nous balayons est $+0,6$ V pour ne pas déclencher la jonction de source/drain vers le substrat. Les valeurs de V_{th} sont

extraites à courant de drain constant. Le jeu de données contient également la capacité de grille en inversion pour chaque échantillon. On appellera V_{th0} la tension $V_{th}(V_B = 0)$ et ΔV_{th} sera la différence entre $V_{th}(V_B = -0,6)$ et $V_{th}(V_B = +0,6)$. $V_{th}(V_B) = V_{th0} + \Delta V_{th}$.

Les figures 2.18a et 2.18c montrent les répartitions spatiales de V_{th0} et ΔV_{th} moyennées sur 25 tranches. Les distributions des 1650 mesures sont aussi montrées sur les figures 2.18b et 2.18d. On peut remarquer que V_{th0} montre une double population qu'on explique directement par le caractère très radial de la distribution spatiale de V_{th} .

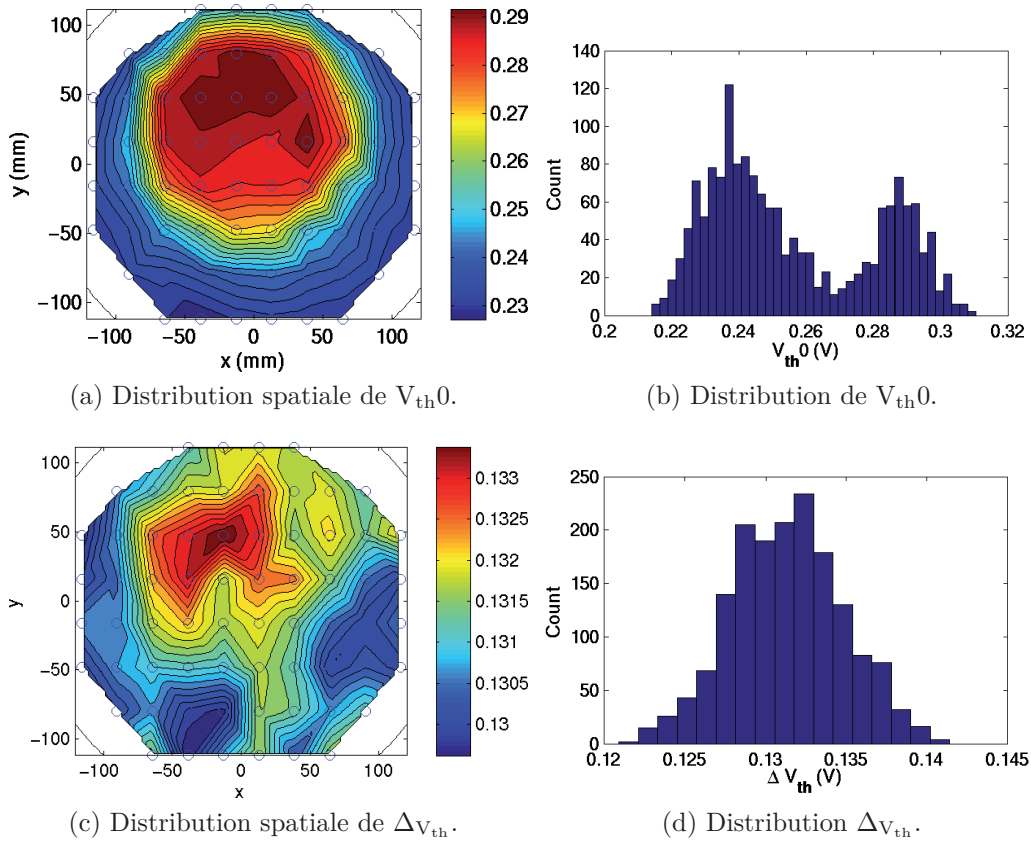


FIGURE 2.18 – Cartographie moyennée sur 25 wafers et histogrammes de V_{th0} et ΔV_{th} . $W = 0,6\mu m$; $L_{gate} = 1\mu m$.

2.6.2 Application de la méthode

Nous appliquons maintenant la méthode présentée dans 2.5.1. Le résultat est un segment de profil de dopage entre 30 nm et 60 nm de profondeur

pour chaque transistor mesuré, correspondant à la gamme de V_B balayé. La figure 2.19b montre 3 profils déterminés à partir des V_{th0} expérimentaux de la figure 2.19a.

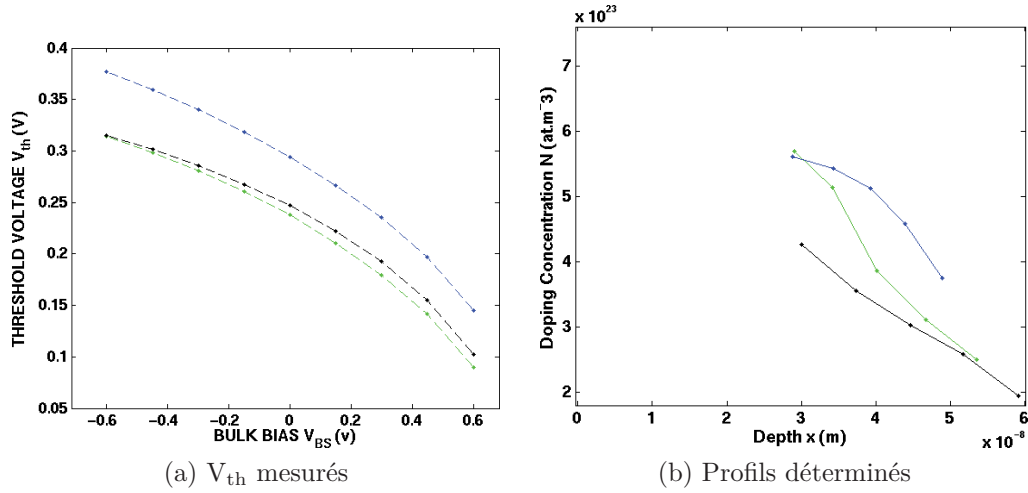


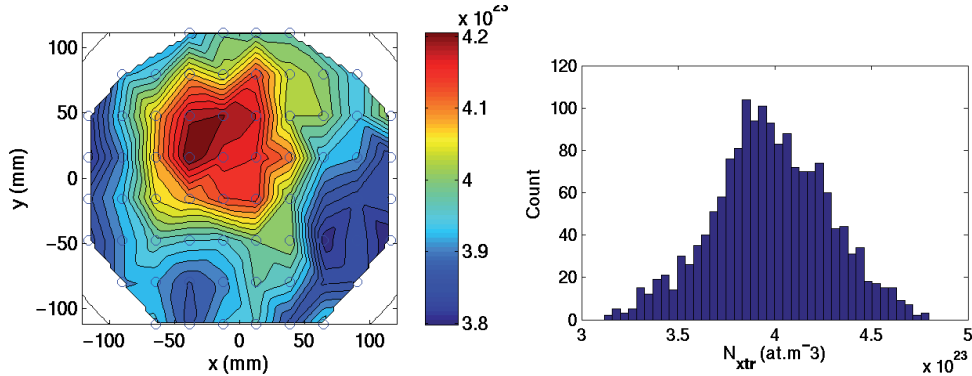
FIGURE 2.19 – Détermination des profils à partir de $V_{th}(V_B)$ pour trois transistors de 1 μm de long.

Pour étudier la variabilité des résultats, nous simplifions l'information en définissant 2 indicateurs qui sont :

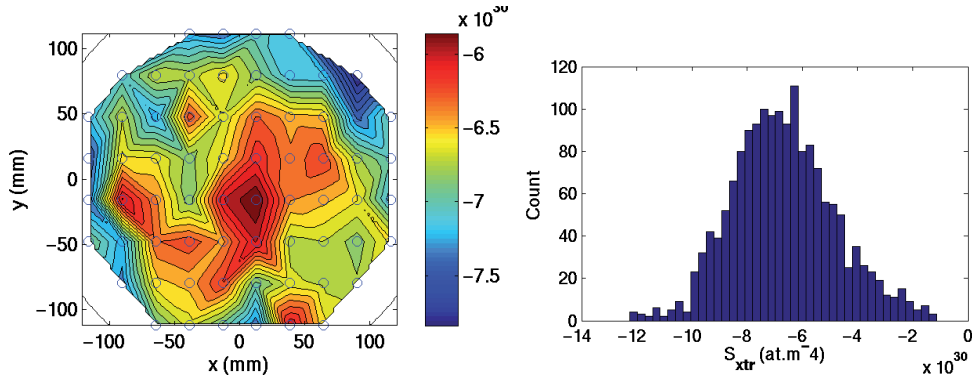
- N_{xtr} , la valeur moyenne du segment de profil extrait,
- et S_{xtr} , la pente de ce même profil (ou plus précisément la moyenne des gradients de ce segment de profil).

Nous étudions ces indicateurs par rapport aux variations de V_{th0} et ΔV_{th} .

La Fig. 2.20 présente, dans le cas d'un canal long, les histogrammes des 2 indicateurs et leurs répartitions spatiales moyennées sur 25 plaques. N_{xtr} a une répartition radiale qui suit une répartition analogue à ΔV_{th} dans la figure 2.18c. La fluctuation de plaque à plaque est montrée par les boîtes à moustaches sur la figure 2.21. La fluctuation intra-wafer est cohérente sur les 25 wafers. Les médianes de N_{xtr} et S_{xtr} montrent une légère tendance en variation intra-lot, indiquant que les opérations faites par lots, comme certains recuits, ont une uniformité imparfaite.



(a) Distribution spatiale du niveau de dopage N_{xtr} . (b) Population du niveau de dopage N_{xtr} sur un lot.



(c) Distribution spatiale de la pente S_{xtr} du dopage. (d) Population de la pente S_{xtr} du dopage.

FIGURE 2.20 – Cartographie moyennée sur 25 plaques et population du dopage N_{xtr} et de sa pente S_{xtr} . $W = 1\mu\text{m}$; $L_{gate} = 1\mu\text{m}$.

2.6.3 Étude des corrélations

V_{th0} et ΔV_{th} ont des signatures complètement différentes. Dans la figure 2.22a, nous avons tracé le nuage de points de V_{th0} en fonction de ΔV_{th} et calculé le coefficient de corrélation R . La figure 2.22a montre bien la faible corrélation entre V_{th0} et ΔV_{th} ($R = 0,26$). De même, on ne peut pas mettre en évidence de corrélation entre N_{xtr} et S_{xtr} (figure 2.22b $R = -0,13$).

L'étude a porté jusque là sur les transistors longs. Nous avons également observé le comportement des dispositifs courts. (V_{th0} en fonction de ΔV_{th} et N_{xtr} en fonction de S_{xtr}). Les résultats sont très différents (voir figure 2.23). En effet, dans les dispositifs courts, une part très significative des charges de déplétion est partagée entre les source/drain et la grille [110]. Nous attribuons

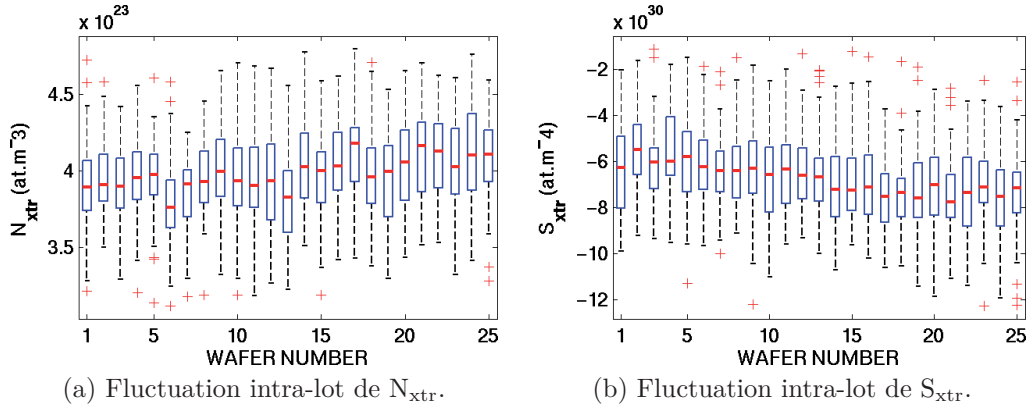
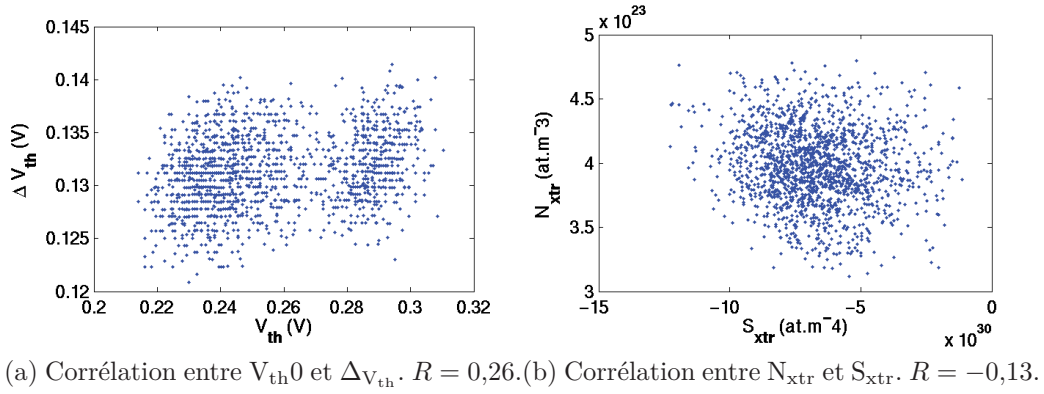
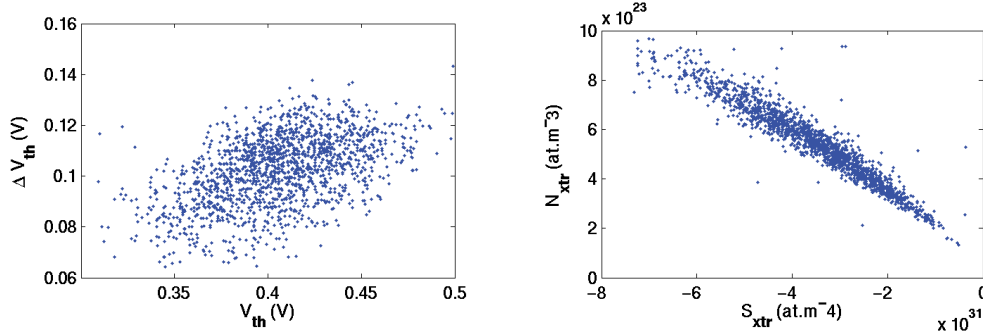


FIGURE 2.21 – Dispersion par plaque du dopage extrait et de sa pente.

FIGURE 2.22 – Transistors longs de $L_g = 1\mu m$. a) Corrélations pour 1650 puces entre V_{th0} et ΔV_{th} et b) entre N_{xtr} et S_{xtr} .

la corrélation systématique montré par la figure 2.23 aux effets de canal court (SCE). La méthode présentée dans la section 2.5.1 est clairement faite pour être utilisée sur des transistors longs où le dopage du canal a une variation quasiment unidimensionnelle car il n'est pas influencé par la redistribution des dopants des *ldd* et des *poches*.



(a) Corrélation entre V_{th0} et ΔV_{th} . $R = 0.50$. (b) Corrélations entre N_{xtr} et S_{xtr} . $R = 0.95$.

FIGURE 2.23 – Transistors courts ($L_g = 45nm$). a) Corrélations pour 1650 puces entre a) V_{th0} et ΔV_{th} et b) N_{xtr} et S_{xtr} .

2.7 Conclusion

Dans ce chapitre, nous nous sommes intéressés à des grandeurs technologiques qui ne sont pas mesurées de manière directe ni de manière statistiques, à savoir le dopage de canal, le profil de dopage des jonctions canal-source/drain et la température de recuit. Ces grandeurs sont pourtant de première importance dans la réalisation des transistors.

Le profil vertical de dopage définit le comportement des charges dans le canal et donc les performances de premier ordre du dispositif MOS. On peut accéder au profil vertical du dopage avec des SIMS mais il n'est pas envisageable d'appliquer cette méthode destructive dans l'optique d'un suivi statistique. Nous avons donc proposé une méthode nécessitant la mesure électrique des tensions de seuil en fonction de la tension de substrat et la mesure de l'épaisseur d'oxyde de grille permettant, ainsi, une détermination des fluctuations du niveau de dopage dans la zone observable.

Le profil transverse du dopage définit la longueur de grille effective. Il est le résultat de la redistribution des dopants de source/drain sur le dopage du substrat après le recuit rapide d'activation. Nous proposons une caractérisation électrique de ce profil qui consiste en la mesure du courant de drain pour toute une gamme de longueur de grille, dans des conditions de bandes plates pour éviter la conduction par le canal et se placer dans un régime de perçage du canal. La caractéristique se révèle être très sensible à la longueur de grille et peut constituer, ainsi, un critère de calibrage des jeux de simulations

TCAD.

La température de recuit, qui est une des causes de la variabilité du dopage, varie localement sur la puce, comme nous l'avons vu au premier chapitre, mais elle est également non uniforme sur une plaque à cause des équipements. Nous détaillons une méthode, nécessitant la fabrication et la mesure d'un lot dédié, mais qui permet de caractériser la signature de l'équipement de recuit et ainsi de connaître les fluctuations qu'il cause.

Dans la suite de ce travail (chapitre 3) , nous utiliserons en particulier la mesure de la température de recuit $\text{Spike}_{\text{temp.}}$ dans le cadre d'un plan d'expérience sur les facteurs principaux de la variabilité du transistor MOS au nœud 45 nm.

Chapitre 3

Étude expérimentale du nœud technologique 45nm

Dans ce chapitre, nous étudions les fluctuations des variables technologiques et des performances électriques. Nous exploitons, pour cela, les données expérimentales d'un plan d'expérience exécuté sur un lot de R&D. L'observation de ces variations nous permettra de faire un état des lieux et de le comparer aux hypothèses faites dans les jeux de simulations. Nous nous intéresserons aux déviations brutes, aux corrélations entre variables et aux motifs récurrents. L'étude du plan d'expériences, en particulier, nous apportera l'information des sensibilités des performances aux facteurs technologiques et confortera le choix des variables à utiliser en simulation.

La recette employée pour ce lot comporte l'utilisation d'un *pré-espaceur*, appelé dans le jargon *offset spacer* ou *spacer-0* et qui consiste, en gardant la même longueur effective de canal, à réaliser la grille plus courte tout en repoussant les implants *ldd* vers l'extérieur afin de réduire le recouvrement de la grille sur les diffusions.

Maitra et Bhal [111] étudient les performances circuit vis-à-vis des variations du recouvrement de grille. Ils rappellent la sensibilité des performances du dispositif à ce paramètre procédé. Le courant I_{on} y est très sensible *via* la résistance des diffusions source/drain. Le courant de fuite I_{off} et donc la consommation du circuit y est également sensible. Et enfin, la capacité totale de grille qui est très sensible influence très nettement la performance de délai des inverseurs. Cette sensibilité augmente en fonction du recouvrement de grille, soulignant la nécessité de garder un recouvrement le plus petit possible.

3.1 Les variables étudiées

Le choix des variables du plan d'expérience est effectué en considérant, dans un premier temps, celles dont l'influence sur le comportement électrique du dispositif est bien établie. On a aussi mesuré des variables ne faisant pas partie du plan (fig 3.1). Les variables du plan sont l'épaisseur d'oxyde de grille T_{ox} , la longueur de grille L_g , l'épaisseur de nitrure déposé pour former l'espaceur HCD_{th} , l'épaisseur de la couche de *SMT* (Stress Memory Technique) et la température de recuit Spike $Spike_{temp.}$. Les autres variables monitorées sont l'épaisseur $TEOS_{th}$ de la couche d'oxyde *liner* situé entre la grille et l'espaceur, la longueur du *spacer-zéro*, et enfin la longueur L_{spacer} de l'espaceur, conséquence directe de l'épaisseur déposée de nitrure qui est une variable du plan. Le *spacer-zéro* est un décalage du L_{spacer} classique qui a été introduit afin d'éloigner les implants *ldd* vers l'extérieur afin de réduire les capacités de recouvrement entre la grille et les source/drain.

Au premier ordre, on sait que :

- T_{ox} influence la capacité de grille et les tensions de seuil, et donc les courants mesurés à tension fixe,
- L_g va influencer le courant canal et la capacité de grille,
- $TEOS_{th}$ et *spacer-zéro* positionnent les implants *ldd* et définissent donc la longueur de canal effective. Ils influencent aussi les capacités grille-drain C_{gd} et d'overlap C_{ov} .
- HCD_{th} définit la longueur du L_{spacer} et donc le positionnement des source/drain. Ce paramètre n'est pas censé influencer la longueur effective du canal mais il modifie le contrôle électrostatique du canal en profondeur, donc le DIBL¹ et agit ainsi sur le courant de drain I_{on} .
- Le recuit spike est là pour activer les dopants des source/drain avec une diffusion limitée. Les variations de budget thermique reçu induisent une variation de la diffusion latérale des jonctions source/drain en direction du canal et donc $Spike_{temp.}$ doit influencer L_{eff} et C_{ov} .

3.1.1 Échantillonnage spatial

Les mesures sont de deux types : mesures électriques et mesures physiques (tableau 3.1). Toutes les mesures n'utilisent pas la même cartographie

1. *Drain Induced Barrier Lowering* ou réduction du seuil de conduction causée par la tension de drain et mesuré en termes de différence entre V_{Tlin} et V_{Tsat} .

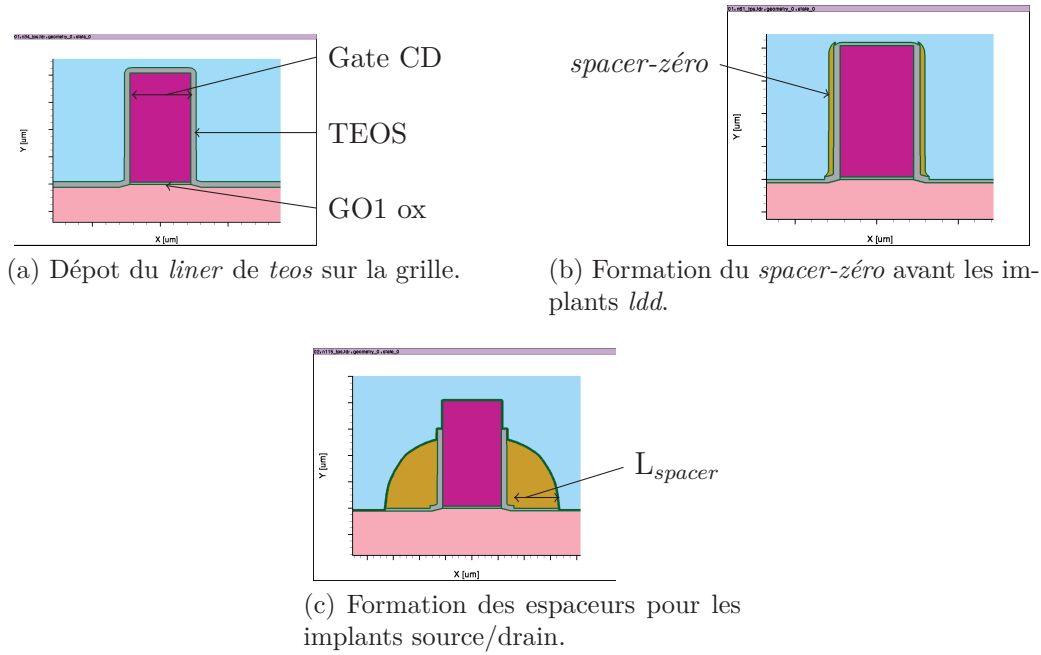


FIGURE 3.1 – Facteurs d’entrée du DOE. L’épaisseur d’oxyde de grille (a), la longueur de grille et l’épaisseur de la couche de nitrure servant à la création des espaceurs prennent plusieurs valeurs. Non représentée ici, l’épaisseur de la couche SMT située au-dessus de la grille, destinée à stresser le canal et la température de recuit font aussi partie du plan d’expérience. L’épaisseur du *liner* (a) et l’épaisseur du *spacer-zéro* (b) sont mesurées mais ne subissent pas de variations volontaires.

ou *mapping*² – ou répartition spatiale sur la plaquette des points de mesure. La figure 3.2 montre ces différents *mappings*. Chacun de ces *mappings* échantillonne différemment dans l’espace. L’emplacement des points sur la surface du *wafer* est choisi de manière à couvrir au mieux l’espace pour être capable de mesurer la dispersion intra-wafer. L’emplacement idéal des mesures est toutefois biaisé par la contrainte suivante : on ne mesure pas les transistors sur la puce. On mesure les caractéristiques de dispositifs dédiés aux tests électriques de suivi industriel qui sont situés dans les chemins de découpe,

2. Pour l’anecdote, les mesures proviennent chacune de leur ateliers respectifs et chacun de ces ateliers utilise sa manière de numérotter les puces mesurées et son format de fichiers de données. Pour pouvoir étudier les mesures les unes par rapport aux autres de manière cohérente, il est nécessaire de convertir toutes les données dans un *mapping* commun. Il a fallu développer des scripts dédiés à l’importation et à la conversion des fichiers. J’utilise le *mapping* à 66 points de mesures en ligne comme *mapping* commun.

c'est-à-dire entre les puces. Ces structures de test sont mesurées puis naturellement détruites lors de la découpe dans le cas des lots non R&D. Le cas des échantillonnages à 9 et à 17 points est un sous-ensemble de la population présente sur chaque tranche, ce qui implique que les intervalles de confiance sur la moyenne et sur la variance de ces mesures sont plus larges que dans le cas où toutes les puces sont mesurées.

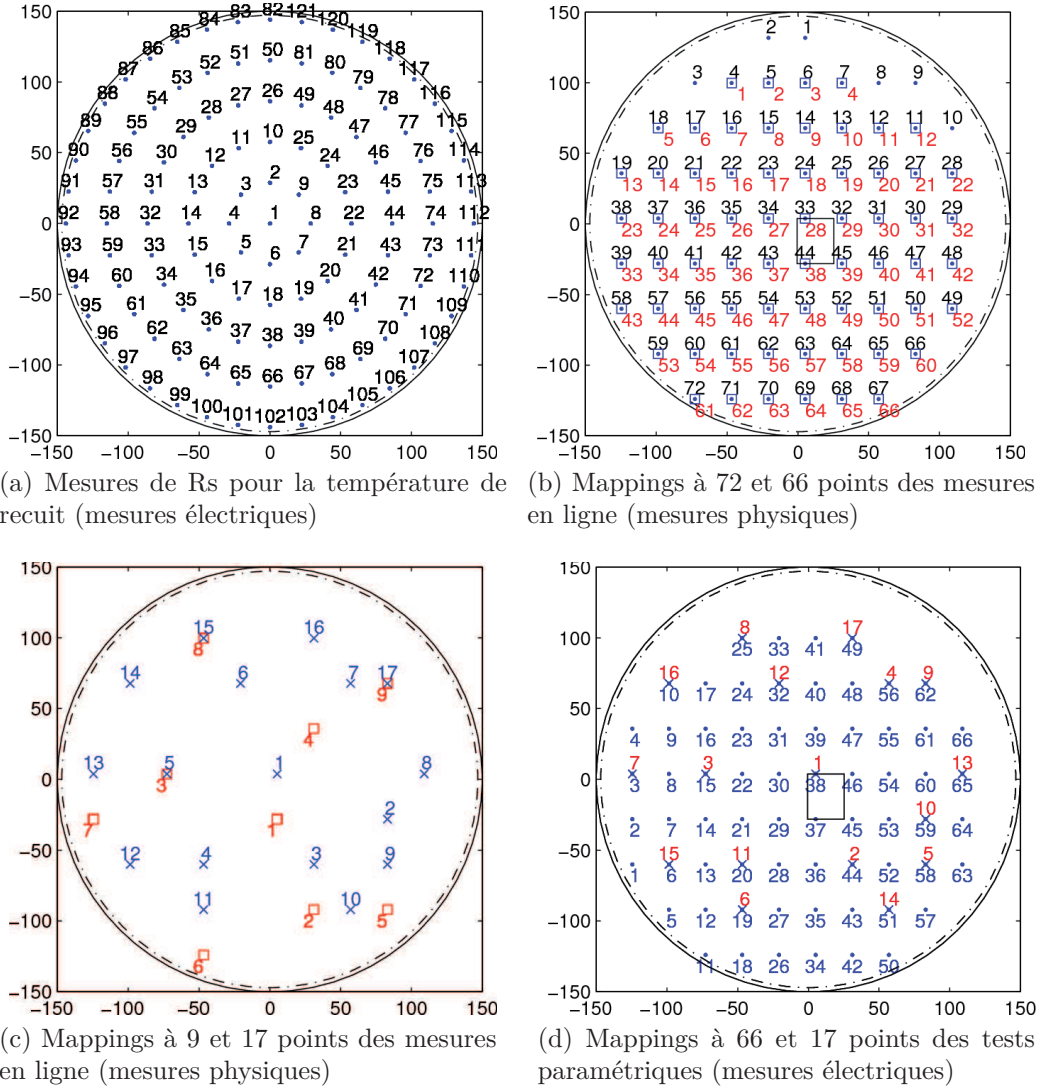


FIGURE 3.2 – Différentes cartographies utilisées sur tranche de silicium dans les différents jeux de données provenant des différents ateliers.

TABLE 3.1 – Mesures des facteurs sur un lot.

Définition		Facteurs	unité	type	p ^{ts} /pl.
Épaisseur d'oxyde de grille	X1	Tox	Å	ellipsométrie	9
Longueur de grille	X2	Lgate	nm	SCD ^a	66
Séparateur d'espaceurs		liner th	Å	ellipsométrie	17
Pré-espaceur		<i>spacer-zéro</i>	nm	ellipsométrie	17
Dépot de nitrure de l'espaceur principal	X3	HCD spacer	nm	ellipsométrie	17
Couche génératrice de stress	X4	SMT th	nm	← pas de mesures	
Température de recuit rapide	X5	Spike T	K	Rs équivalent	66 ^b

^a. Spectroscopic Critical Dimension

^b. réduit à 66 pour correspondre aux emplacements des puces des autres mesures.

3.2 Observation des données

Avant d'exploiter les résultats de mesures du plan d'expérience pour modéliser les réponses et définir les sensibilités, on observe les mesures brutes à titre de vérification du bon déroulement du plan et d'observation des variabilités à l'intérieur du lot et à l'intérieur des plaques.

3.2.1 Les facteurs

Les variables technologiques ayant servi de facteurs ont été mesurées pendant la réalisation des plaques. Les mesures d'épaisseur sont des mesures ellipsométriques et les mesures de dimensions horizontales sont des mesures sur photo SEM³. Les mesures de températures sont, en fait, des mesures de résistance de surface, comme expliqué au paragraphe 2.2, page 41.

La figure 3.3 présente les facteurs du plan d'expérience avec les mesures correspondantes⁴. L'abscisse de ces graphes représente le numéro d'expé-

3. *Scanning Electron Microscopy* ou MEB, microscopie électronique à balayage

4. D'un point de vue pratique, ces graphes me permettent aussi de vérifier que l'acquisition des données provenant de différents formats s'est bien passée.

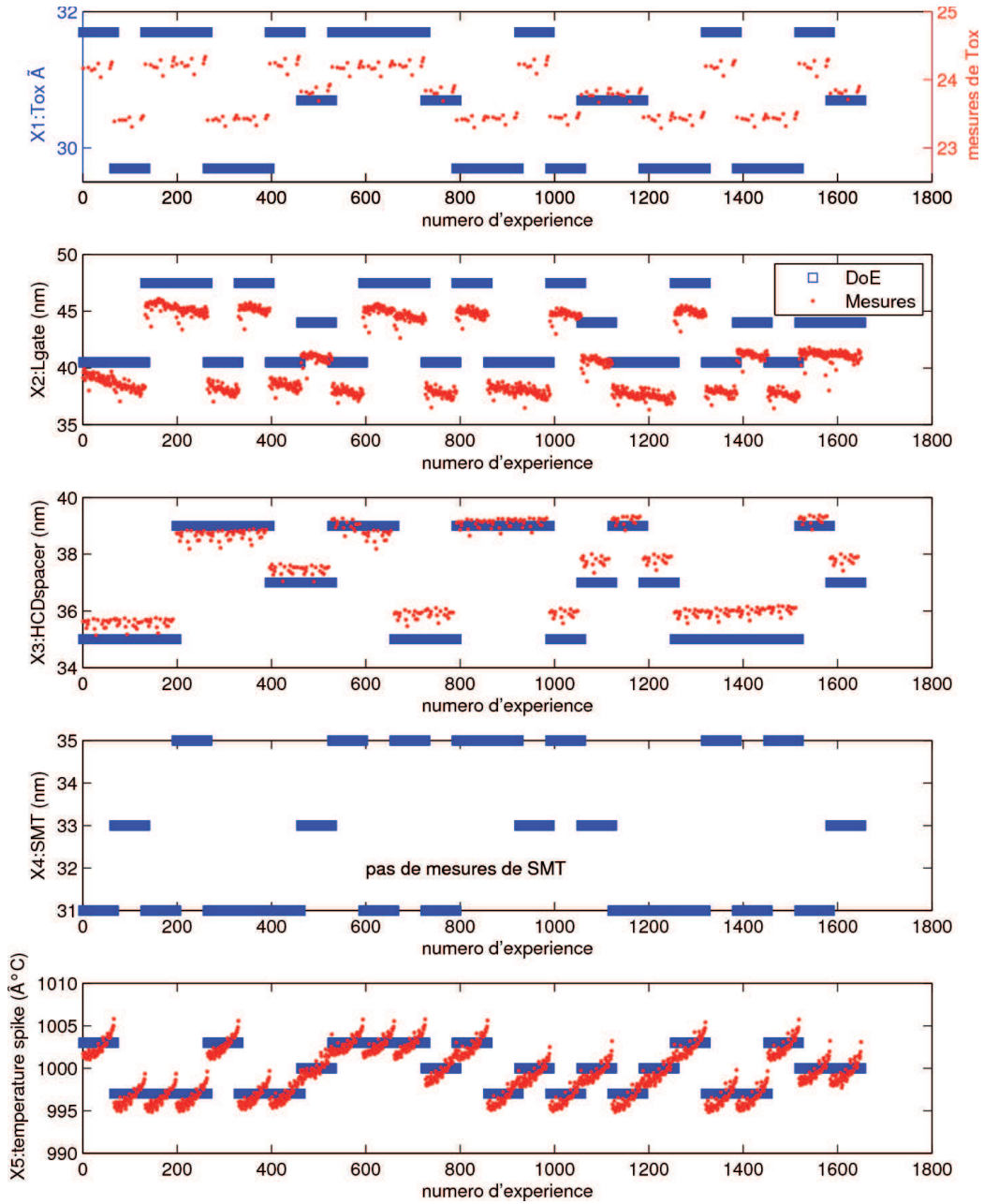


FIGURE 3.3 – Comparaison des mesures des paramètres physiques utilisés dans le DoE avec leurs valeurs nominales. Les expériences sont triées par ordre de plaque et de numéro de site. On observe un décalage et une dérive de la réalisation des paramètres entre la plaque 1 et la plaque 25 pour les paramètres X2 et X3. Dans le cas de X5, on apprécie l'importance de la variabilité intraplaque par rapport à l'excursion du DoE.

rience, trié par numéro de wafer puis par numéro de site sur le wafer. C'est à dire que les 66 premiers points proviennent du premier wafer, les points 67 à 132 du second. . . Sur chaque facteur, on peut observer un écart systématique entre les valeurs spécifiées au DoE et les valeurs mesurées. On peut aussi observer une dérive entre les lots et un motif associé aux tranches. La longueur de grille X2 présente un motif se répétant à chaque plaque. On observera cette dispersion spatiale dans la figure 3.4b. La longueur mesurée est systématiquement inférieure à la longueur définie. On observe aussi une décroissance de la valeur moyenne de X2 ainsi qu'une légère croissance de X3. Nous n'avons pas de mesures de X4, l'épaisseur de la couche SMT⁵. Les mesures de X5 ne montrent pas de décalage de leur valeur moyenne par plaque. Cette étape est manifestement très reproductible. On remarque aussi le motif de X5 se répétant sur toutes les plaques, un motif propre au recuit flash que l'on observera dans la figure 3.8b, page 78.

3.2.2 Variabilité intra-plaque

Observons les variations des données technologiques : les facteurs Xi et les variables TEOS et Spacer-zéro. La figure 3.4a montre les variations de l'oxyde de grille en fonction de la distance au centre de la tranche et ce, pour les trois valeurs du paramètre X1. L'évolution radiale est très nette. Elle suit une sorte de vague et une telle oscillation systématique – on peut l'observer sur toutes les plaques – est la marque du procédé sur cette variabilité. On note aussi que la dispersion autour de cette vague est relativement faible.

La figure 3.4b montre les valeurs de la longueur de grille pour le dispositif nominal – X2=44 nm – sous la forme de lignes de niveau sur le wafer 17 qui est un point au centre du plan d'expérience. Les variations ne montrent pas de motif reconnaissable. Il semble bien y avoir un effet systématique qu'on observait sur la figure 3.3 mais rien de très marqué au regard de la variation aléatoire. Le seul point commun que l'on peut observer sur toutes les plaques de ce lot sont les 3 puces situées en haut à gauche de la plaque et dont les longueurs de grille sont toujours plus faibles que le reste de la plaque.

La figure 3.5 présente les variations de la variable TEOS⁶ qui est une

5. Stress Memory Technique

6. tetra-ethyl-ortho-silicate, $\text{Si}(\text{OC}_2\text{H}_5)_4$. Le TEOS est une molécule utilisée en CVD – chemical vapor deposition – pour déposer du dioxyde de silicium. La réaction est $\text{Si}(\text{OC}_2\text{H}_5)_4 + 2\text{H}_2\text{O} \longrightarrow \text{SiO}_2 + 4\text{C}_2\text{H}_5\text{OH}$. À titre de rappel, le dioxyde silicium de la grille est créé par croissance de l'oxyde natif présent à la surface du silicium.

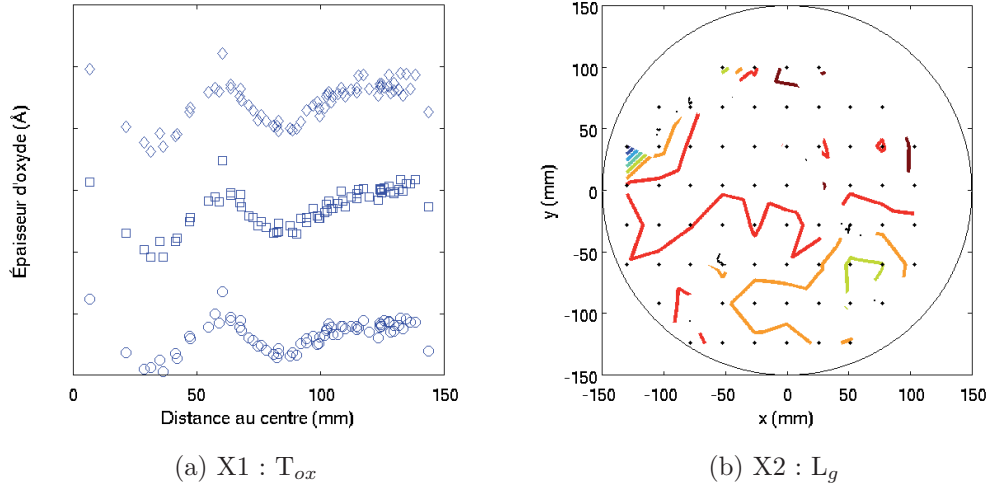


FIGURE 3.4 – a) Distribution radiale de l'oxyde de grille pour 3 valeurs de X1. La variation bien définie, en vague, est typique d'une cause process alors que la variation aléatoire autour de cette vague est relativement faible. b) Distributions spatiales de X2 sur le wafer 17 qui est un point au centre du plan d'expériences. X2 montre une variation aléatoire mais avec une zone de plus faibles valeurs en haut à gauche. Ce comportement se retrouve qualitativement sur toutes les plaques.

couche d'oxyde déposée entre le polysilicium de grille et le *spacer-zéro*. La figure 3.5a présente la variation de plaque à plaque de cette variable. L'échantillonnage est de 17 points par plaques. On peut observer que la médiane, qui est le trait central des boîtes à moustaches, prend des valeurs discrètes, avec un pas de 0,1 Å. Soit on a atteint les limites de l'appareil de mesures, soit ces chiffres ont subi une troncature. Cette variable présente à la fois des variations systématiques intra-lot et intra-plaque. La variation intra-lot de la médiane par wafer est de l'ordre de 1 Å. La distribution spatiale sur tranche est systématiquement très radiale et inférieure à 1 Å.

Les distributions de *spacer-zéro* dans la tranche et dans le lot sont représentées sur la figure 3.6. La longueur de *spacer-zéro* ne fait pas partie du plan d'expérience et sa valeur nominale est de 6 nm. La distribution de la population n'est pas normale (3.6a). Sa valeur moyenne est bien de 6 nm mais elle présente une forte asymétrie à droite. On peut en constater la provenance sur la figure 3.6b. En effet la médiane de chaque plaque dérive nettement d'un

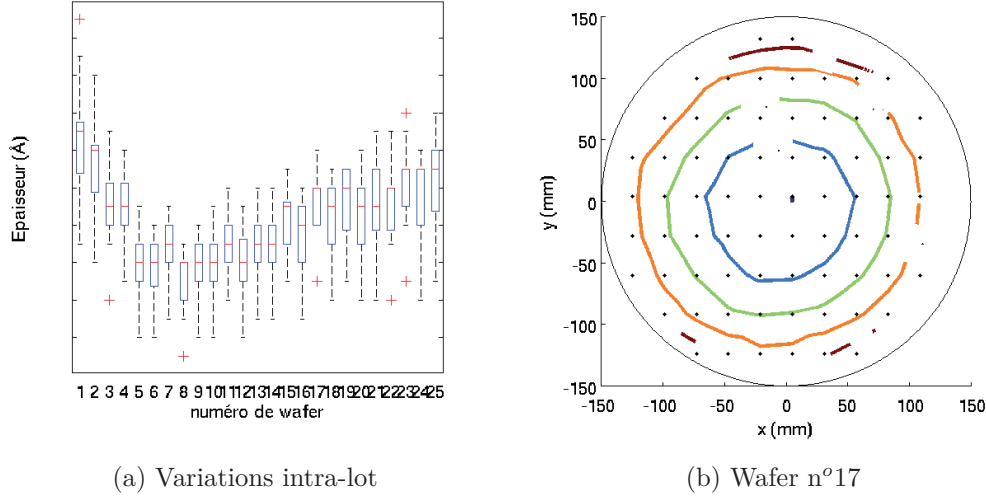


FIGURE 3.5 – Distributions spatiales intra-lot (a) et intra-plaque (b) de l'oxyde *liner* TEOS.

bout à l'autre du lot – le dépôt de nitrure se fait par *batch*, le lot entier dans la machine.

Par dessus cette dérive, vient s'ajouter une dispersion intra plaque plutôt⁷ du type bord centre (3.6c, 3.6d) dont la dispersion se réduit de moitié entre les premiers et les derniers wafers (3.6b).

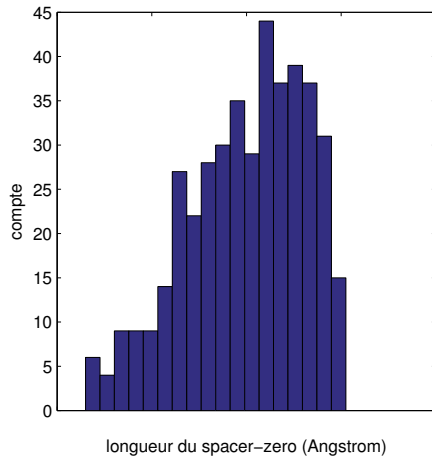
Passons au facteur du DOE X3, à savoir l'épaisseur de nitrure HCD⁸. L'histogramme en figure 3.7a montre la population des plaques qui ont été fabriquées avec une valeur nominale. La distribution asymétrique marque encore la dérive intra-lot de ce dépôt de nitrure. La médiane bouge d'environ 5 Å dans le lot. Les valeurs sur plaque (figure 3.7c) couvrent environ la même gamme.

La figure 3.8 montre les variations de la température *estimée*⁹ subies par les tranches de silicium lors du recuit flash. La variation intra-plaque est de 6°C.

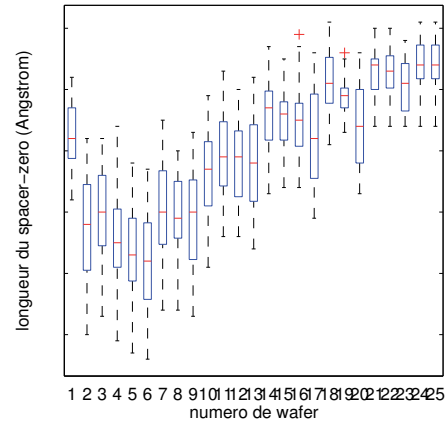
7. je dis *plutôt* car le centre de ce motif a tendance à se déplacer d'une plaque à l'autre

8. HexaChloroDisilane, Si_2Cl_6 , est utilisé pour le dépôt CVD du nitrure Si_3N_4 .

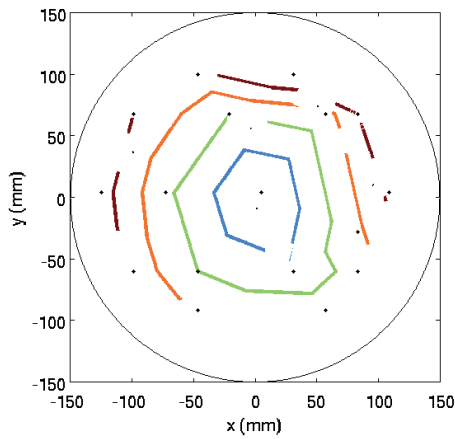
9. c'est en fait une mesure de résistance de surface convertie en température comme expliqué au paragraphe 2.2, page 41



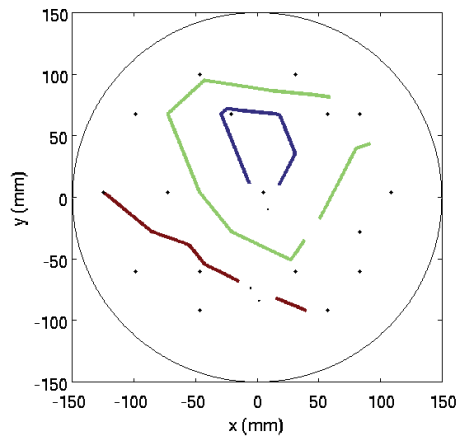
(a) Intra-lot



(b) Intra-lot / plaque à plaque

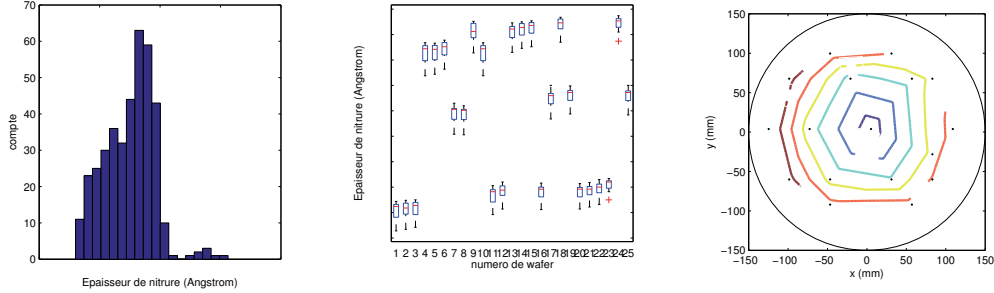


(c) Intra-plaque wafer 8



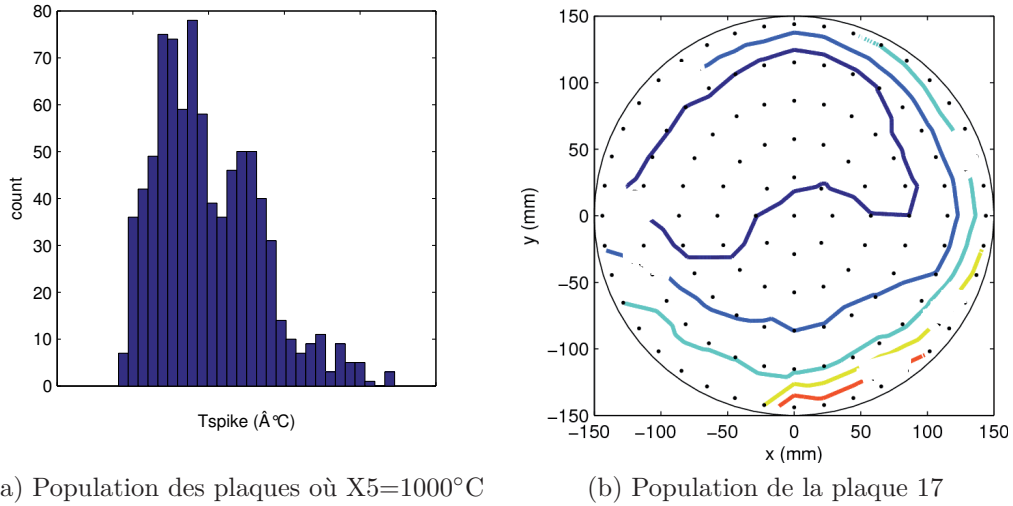
(d) Intra-plaque wafer 25

FIGURE 3.6 – Distributions spatiales intra-lot (a,b) et intra plaque (c,d) de *spacer-zéro*



(a) Population de HCD des wafer dont X3=0 (b) Wafer to wafer de HCD. 17 points par plaque. (c) Variation intra plaque de HCD sur le wafer n°17.

FIGURE 3.7 – Distributions intra-lot (a), plaque à plaque (b) et intra-plaque (c) de l'épaisseur de niture déposé pour la formation des espaceurs.



(a) Population des plaques où X5=1000°C

(b) Population de la plaque 17

FIGURE 3.8 – Distributions de X5 sur les plaques où X5=1000°C sur le wafer 7 qui est un point au centre du plan d'expériences. La population globale (a) est déséquilibrée à gauche par la répartition spatiale (b) X5 qui montre un motif plus ou moins radial et monotone avec une dynamique de 4°C. Ce comportement se retrouve qualitativement sur toutes les plaques.

3.2.3 Corrélations entre variables technologiques

Les facteurs, qui sont les variables explicatives du modèle que nous allons calculer dans l'analyse du plan d'expérience, sont supposés indépendants. La matrice (3.1) est la matrice de corrélation des variables technologiques, celles du plan plus $TEOS_{th}$ et *spacer-zéro*. Elle nous indique que toutes les variables ne sont pas indépendantes car tous les termes non-diagonaux ne sont pas nuls. On trouve notamment quelques termes supérieurs à 0,5.

$$R = \begin{bmatrix} 1 & -0.06 & 0.05 & -0.4 & -0.11 & -0.2 & 0.61 \\ -0.06 & 1 & -0.24 & -0.08 & -0.2 & -0.41 & -0.27 \\ 0.05 & -0.24 & 1 & 0.81 & 0.93 & 0.6 & -0.17 \\ -0.41 & -0.08 & 0.81 & 1 & 0.87 & 0.36 & -0.26 \\ -0.11 & -0.2 & 0.93 & 0.87 & 1 & 0.43 & -0.03 \\ -0.2 & -0.41 & 0.6 & 0.36 & 0.43 & 1 & -0.34 \\ 0.61 & -0.27 & -0.17 & -0.26 & -0.03 & -0.34 & 1 \end{bmatrix} \quad (3.1)$$

Regardons plutôt la figure 3.9 qui nous montre ces mêmes corrélations mais de manière graphique. Cette figure montre des graphes de corrélations par paire de variables. Les cases diagonales montrent les histogrammes de chaque variable seule. La matrice de graphes est symétrique. Lorsque une case montre X_i en fonction de X_j , sa symétrique par rapport à la diagonale montre X_j en fonction de X_i . La population utilisée ici est celle des trois plaques correspondant aux points au centre du DOE. Nous n'observons donc ici que les variations naturelles du process nominal sans aucune variation volontaire du plan d'expérience. On remarque que, d'une case à l'autre, le nombre de points change et on peut avoir près de 200 points ou seulement 9 points expérimentaux. Cela vient des différentes cartographies utilisées. Par exemple, pour les cases T_{ox} vs $TEOS_{th}$, T_{ox} vs *spacer-zéro* et T_{ox} vs HCD_{th} , il n'y a que 9 points. Ils correspondent à 3 fois 3 points par plaque et ces 3 points sont les 3 seuls points communs des *mappings* à 9 et 17 points utilisés pour des mesures en ligne (voir figure 3.2c). Pour ces trois combinaisons, il sera difficile de conclure quoi que ce soit avec si peu de points. Pour les autres combinaisons, celles où une des 2 variables est mesurée 66 fois par plaques, on aura 3×9 , 3×17 ou 3×66 points.

On a une corrélation de 0,6 entre T_{ox} et $Spike_{temp.}$ que l'on peut observer sur 27 points. Avec un taux de confiance de 95 %, ce taux de corrélation est compris entre 0,3 et 0,8. Ce qui signifie que c'est soit seulement le hasard de

l'échantillonnage qui fait que ce nuage de points apparaît corrélé, soit qu'il y a un phénomène physique à l'origine de cette situation. Il n'y a pas de raison *a priori* pour qu'une plus grande épaisseur d'oxyde de grille influence la résistance de surface.

Le groupe de variables $TEOS_{th}$, *spacer-zéro*, HCD_{th} et L_{spacer} présente des corrélations évidentes. D'abord HCD_{th} et L_{spacer} font partie de la même brique technologique. La longueur L_{spacer} est le résultat de la gravure anisotropique de la couche de nitrure HCD. Leur taux de corrélation est à 95 %, sur cet échantillon et est compris entre 0,17 et 0,63. Les variables $TEOS_{th}$, *spacer-zéro* et HCD_{th} varient très nettement ensemble. Leur point commun est clairement le dépôt CVD. On a un dépôt de SiO_2 pour $TEOS_{th}$, un dépôt de Si_3N_4 pour HCD_{th} et un dépôt de chaque pour *spacer-zéro*¹⁰. Il n'est donc pas surprenant que ces variables subissent les mêmes déviations systématiques puisque le process est très proche. On a déjà vu des similarités entre les comportements bord-centre des répartitions sur plaques sur les figures 3.5b, 3.6c et 3.7c, et dans une moindre mesure entre les tendances dans le lot.

Les autres cases ne présentent pas de corrélations particulières. À l'exception du facteur SMT pour lequel nous n'avons pas de mesure, nous avons vérifié que les facteurs du plan d'expérience T_{ox} , L_g , HCD_{th} , $Spike_{temp.}$ sont indépendants, et nous avons établi que les variables $TEOS_{th}$ et *spacer-zéro* qui ne font pas partie du plan ont une variation corrélée à celle de HCD_{th} .

3.2.4 Indépendance des expériences

Chaque expérience du plan est supposée indépendante des expériences la précédant. C'est à dire que le facteur *temps* n'influe pas sur les résultats et que les résultats ne sont pas fonction des résultats précédents. Pour vérifier cela l'outil classique est le graphe d'autocorrélation. On trace la variable en question, ici le courant mesuré, en fonction de la même variable décalée d'une expérience. Classiquement on trace Y_n vs. Y_{n-1} . Ici nous avons 66 points par plaque, chaque plaque étant une expérience. Nous traçons alors Y_n vs Y_{n-66} . La figure 3.10 montre l'autocorrélation de NIDS1L004LS. On observe l'indépendance des expériences.

10. Le spacer zéro est fait des mêmes étapes technologiques que le spacer, il est juste plus petit, 6 nm au lieu de 32 nm

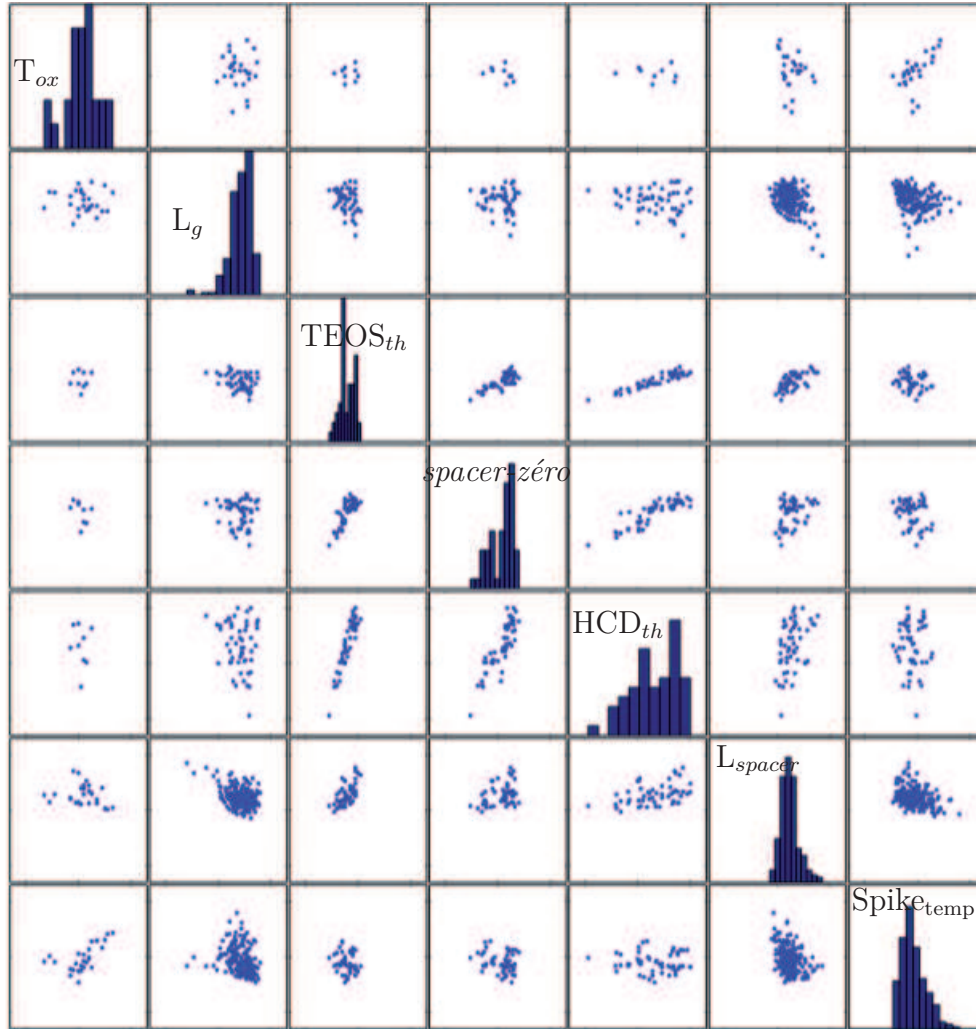
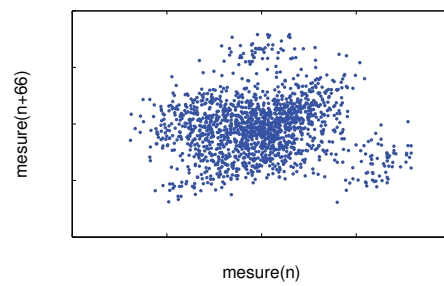


FIGURE 3.9 – Corrélation par paire des variables technologiques. Seules les 3 plaques ayant suivi un process nominal sont affichées de sorte qu'on observe que des variations naturelles. La matrice est symétrique. La diagonale montre les histogrammes de chaque variable seule. Les axes x et y sont communs à chaque colonne et ligne respectivement.



(a) autocorrelation de Y

FIGURE 3.10 – Autocorrélation de NIDS1L004LS

3.3 Étude du plan d'expériences

3.3.1 Le plan expérimental

Le lot étudié a été conçu avec un plan d'expérience sur 25 tranches de silicium qui permet de faire un état des lieux des variations technologiques du nœud 45 nm. Le plan d'expérience, tel qu'il est défini, a 5 facteurs notés X1, X2, ... X5, qui prennent 3 niveaux. Leurs domaines d'exploration sont listés dans le tableau 3.2. Le choix des niveaux est fait selon ce que l'on connaît de chaque facteur, selon ce que l'on veut explorer et selon ce que l'atelier peut raisonnablement faire. Le domaine $[+1 : -1]$ de chaque facteur doit être plus grand que la variabilité naturelle de ce facteur. La population d'un facteur à son niveau nominal doit être nettement différenciée de la population du même facteur à son niveau $+1$ ou -1 . Le nombre d'expériences est limité à

TABLE 3.2 – Plan d'expériences du lot étudié.

Facteur	Description	unité	-1	0	+1
X1	épaisseur d'oxyde de grille	Å	29,7	30,7	31,7
X2	longueur de grille	nm	40,5	44	47,5
	épaisseur du <i>liner</i>	Å		41,6	
	longueur du <i>spacer zero</i>	nm		6	
X3	épaisseur de nitrure d'espaceur	nm	35	37	39
	longueur de l'espaceur	nm		32,4	
X4	épaisseur de la couche SMT	nm	31	33	35
X5	température de recuit	K	-3	+0	+3

25 car les lots de notre fabrique contiennent 25 plaques et les paramètres du DoE sont des opérations qui se font par plaque entière. Les 25 expériences sont listées dans le tableau 3.3. On note que ce plan définit trois points au centre, répartis sur les plaques 8, 17 et 25 du lot. L'affectation de l'ordre des expériences à un numéro de plaque est aléatoire, de manière à éviter qu'une dérive des expériences vienne biaiser les résultats. Dans notre cas cette dérive se retrouverait dans une variation wafer-to-wafer. Sur chacune des 25 plaques, 66 puces sont mesurées. Cela permet d'avoir une bonne image de la variabilité de chaque indicateur électrique mesuré et donc, une bonne (faible) incertitude quant à l'estimation des effets du modèle polynomial que l'on va calculer pour représenter le comportement de chaque indicateur en

fonction des facteurs du plan.

$$R(i,j) = \frac{Cov(i,j)}{\sqrt{Cov(i,i)Cov(j,j)}} \quad (3.2)$$

$$R = \begin{bmatrix} 1 & 0.0006 & 0.1008 & 0.0006 & 0.0521 \\ 0.0006 & 1 & -0.0006 & 0.0697 & 0.0374 \\ 0.1008 & -0.0006 & 1 & -0.0006 & 0.0539 \\ 0.0006 & 0.0697 & -0.0006 & 1 & 0.0374 \\ 0.0521 & 0.0374 & 0.0539 & 0.0374 & 1 \end{bmatrix} \quad (3.3)$$

La matrice R (3.3) des corrélations entre les facteurs du modèle, définie par (3.2), ne contient pas d'élément non diagonaux supérieur à 0,1 , ce qui reste une corrélation très faible quant à une confusion des coefficients. La même matrice calculée pour un modèle du second ordre avec interactions¹¹ donne un taux de 0,579 entre les termes carrés. Ce plan expérimental reste très bon pour calculer un modèle de premier ordre avec interactions.

11. que je n'imprime pas pour vous épargner une matrice 21x21

TABLE 3.3 – Plan d'expériences du lot étudié. Les 25 expériences tiennent dans les 25 plaques que contient un lot. L'ordre des expériences est aléatoire pour éviter une confusion entre l'ordre d'exécution des expériences et un facteur du plan. Ce plan contient 3 points au centre : les expériences 1, 13 et 25.

n° exp.	n° de wafer	X1	X2	X3	X4	X5
1	17	0	0	0	0	0
2	2	-1	-1	-1	0	-1
3	23	-1	-1	-1	1	1
4	19	-1	-1	0	-1	0
5	5	-1	-1	1	-1	1
6	14	-1	-1	1	1	-1
7	22	-1	0	-1	-1	-1
8	20	-1	1	-1	-1	1
9	16	-1	1	-1	1	-1
10	6	-1	1	1	-1	-1
11	13	-1	1	1	1	1
12	12	0	-1	-1	-1	0
13	8	0	0	0	0	0
14	18	0	-1	1	-1	-1
15	1	1	-1	-1	-1	1
16	21	1	-1	-1	1	-1
17	7	1	-1	0	-1	-1
18	15	1	-1	1	0	0
19	9	1	-1	1	1	1
20	24	1	0	1	-1	0
21	3	1	1	-1	-1	-1
22	11	1	1	-1	1	1
23	10	1	1	1	-1	1
24	4	1	1	1	1	-1
25	25	0	0	0	0	0

3.3.2 Régression multilinéaire d'un modèle polynomial

Nous voulons modéliser Y par un modèle polynomial fonction des facteurs technologiques X_i . Une fois le lot réalisé, nous avons le vecteur colonne Y qui réunit les résultats de mesures des indicateurs électriques :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

avec N le nombre d'expériences – ici 25×66 . Nous postulons un modèle de la forme :

$$Y = k_0 + \sum_{i=1}^m k_i X_i + \sum_{i=1, j>i}^m k_{ij} X_i X_j + \sum_{i=1}^m k_{ii} X_i^2 + e \quad (3.4)$$

m étant le nombre de facteurs du plan, 5 dans notre cas. Y comprend un terme k_0 pour la moyenne μ_Y , m termes linéaires, C_m^2 termes de combinaison de facteurs linéaires et m termes quadratiques. Dans notre cas, cela fait un total de $1 + 5 + 10 + 5 = 21$ coefficients à calculer. Nous écrivons la matrice X dont chacune des 21 colonnes correspond à un terme du modèle de Y . Elle est composée d'une colonne de 1 pour le terme d'ordre zéro, des 5 colonnes X_i décrites dans le tableau 3.3 et des colonnes des termes d'interactions et de carrés :

$$X = \begin{bmatrix} X^0 & X_1 & \dots & X_5 & X_1 X_2 & \dots & X_4 X_5 & X_1^2 & \dots & X_5^2 \end{bmatrix},$$

avec chaque ligne répétée 66 fois car il y a 66 mesures par expérience. On peut réécrire (3.4) ainsi :

$$Y = XK + e \quad (3.5)$$

et on veut trouver K tel que $e = Y - \hat{Y} = Y - XK$ soit minimal au sens des moindres carrés. Dans ces conditions ($e = 0$), K s'écrit [107, 112] :

$$Y = XK \quad (3.6)$$

$$X'Y = X'XK \quad (3.7)$$

$$K = (X'X)^{-1}(X'Y). \quad (3.8)$$

Appliquons ce calcul en prenant pour Y la variable NIDS1L004LS qui est le courant de drain I_{on} à $V_{DS}=V_{dd}$ et $V_G=V_{dd}$ mesuré sur un transistor N de $1\text{ }\mu\text{m}$ de large pour 40 nm de longueur de grille. Nous calculons $K_{NIDS1L004LS}$ et estimons $\hat{Y}_{NIDS1L004LS}$. La figure 3.11 montre l'ajustement des données par le modèle quadratique. On voit clairement les 25 expériences regroupées en lignes horizontales de 66 répétitions. Le modèle représente bien leur valeurs moyennes par plaque, mais pas leurs variations naturelles intra-plaque (et c'est normal).

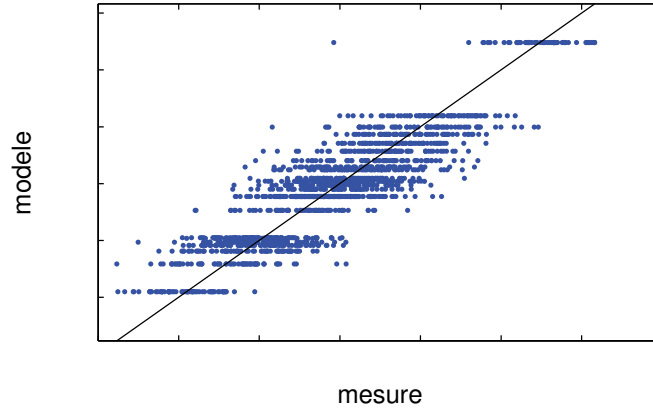


FIGURE 3.11 – Ajustement du modèle $\hat{Y}$ en fonction de Y pour le courant de drain en saturation d'un transistor N de $1\text{ }\mu\text{m}$ de large pour 40 nm de long.

On utilise souvent le rapport R^2 pour représenter le taux d'ajustement d'un modèle sur un nuage de points. Ce taux s'exprime par :

$$R^2 = 1 - \frac{\sum e^2}{\sum (Y - \bar{Y})^2}. \quad (3.9)$$

R^2 exprime la part de variabilité de Y expliquée par le modèle $\hat{Y}$. Il décroît mécaniquement lorsqu'on simplifie le modèle en excluant des effets non significatifs. On définit aussi le $R^2_{\text{ajusté}}$ qui prend en compte le nombre de degrés de liberté et qui peut donc servir à comparer des modèles ayant un nombre de degrés de liberté différents :

$$R^2_{\text{ajusté}} = 1 - (1 - R^2) \frac{N - 1}{N - k - 1} \quad (3.10)$$

avec N le nombre total d'essais et k le nombre de coefficients (degrés de

libertés) du modèle.

R^2 n'est toutefois qu'un rapport de deux variances, celles des observations et celles des erreurs e . Pour juger de la qualité d'un modèle il ne suffit pas de constater un R^2 élevé, ou inversement un R^2 faible qui n'est pas nécessairement signe d'un mauvais modèle. Si le phénomène étudié a une variabilité aléatoire importante, aucun terme du modèle ne prend cette part de variance en compte. Il faut passer par l'analyse du résidu.

Analyse du résidu

L'analyse du résidu consiste à observer $e = Y - \hat{Y}$ depuis différents angles de vues. On vérifie plusieurs hypothèses de la régression linéaire¹². La fi-

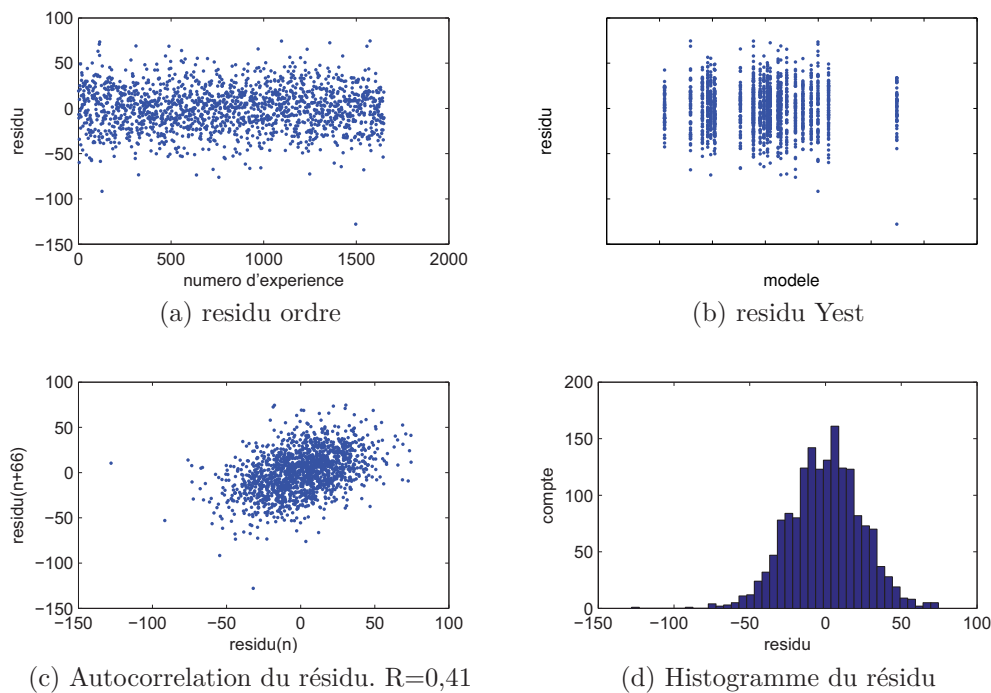


FIGURE 3.12 – Analyse du résidu du modèle de NIDS1L004LS (μA)

gure 3.12a vérifie que le résidu est bien centré autour de zéro et que, par conséquent, le modèle est bien spécifié en moyenne. La figure 3.12b vérifie l'homoscédasticité du résidu. C'est à dire que la variance est constante en fonction de Y . En pratique, on regarde aussi si une variation particulière

12. http://www.statelem.com/regression_lineaire_multiple.php

suggère d'appliquer une transformation à Y et de refaire la régression. La figure 3.12c vérifie la non autocorrélation du résidu. Cette hypothèse doit être vraie pour le calcul des intervalles de confiance sur les coefficients du modèle. La figure 3.12d vérifie la normalité du résidu et son centrage autour de zéro.

La figure 3.13 nous montre le résidu en fonction de chaque facteur du plan. Elle nous permet de vérifier que le modèle rend bien compte de chacun des effets. Ici, le résidu pour chaque facteur est bien centré en zéro et ne présente pas de tendance linéaire ou quadratique qui indiquerait qu'un effet supplémentaire est à introduire dans le modèle. Une forte hétéroscédasticité d'un de ces résidus indiquerait qu'une transformation du facteur, par une fonction logarithme ou exponentielle par exemple, serait à envisager pour appliquer la régression linéaire.

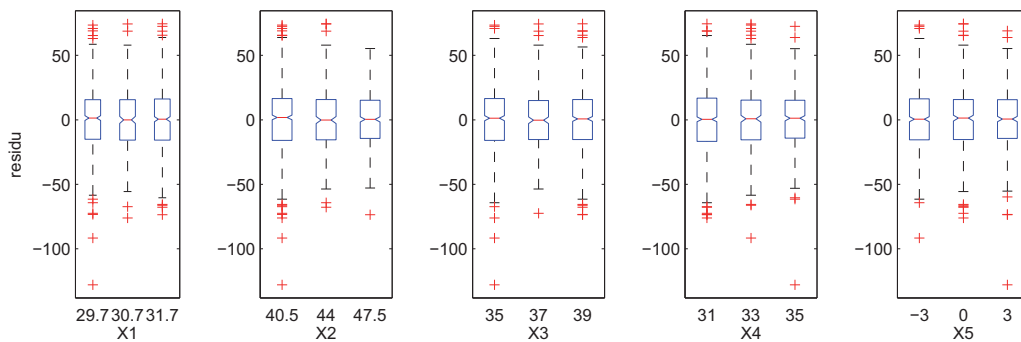


FIGURE 3.13 – Résidu du modèle de NIDS1L004LS en fonction des 5 facteurs du plan. Pour chaque facteur, les trois boîtes à moustaches représentent la distribution de l'erreur de modélisation en fonction de la valeur du facteur. On observe que l'erreur est centrée en zéro et que sa dispersion est homogène.

Effets moyens

L'effet d'un facteur est la déviation qu'il induit à $\hat{Y}$ lorsqu'il varie de sa valeur nominale vers sa valeur +1 ou -1, les autres facteurs du plan étant à leurs valeurs nominales :

$$\text{Effet} = k_i \times (F_{+1} - F_0)$$

La gamme de variation des facteurs étant normalisée à $[-1; +1]$ pour le calcul, l'effet est en fait la sensibilité de la sortie à ce facteur. La figure 3.14 montre chacun des effets des facteurs du modèle de NIDS1L004LS, les autres facteurs étant à leurs valeur nominales. Les mesures de NIDS1L004LS sont superposées au modèle. On peut apprécier les sensibilités relatives du courant de drain à ces variables et vérifier que le modèle se comporte suivant la physique du transistor MOS :

- X1 Le courant diminue avec l'accroissement de l'épaisseur T_{ox} qui augmente V_{th} et réduit la conductance ($\mu \cdot C_{ox}$).
- X2 Le courant diminue avec l'agrandissement de la longueur L_g et donc l'augmentation de la résistance du canal.
- X3 L'agrandissement des espaceurs ne devrait pas changer la longueur effective du canal puisqu'elle est déterminée par les *ldd* qui ne sont normalement pas recouverts par les diffusions latérales des source/drain. Le contrôle électrostatique du canal en particulier sur ce transistor, est affecté par les jonctions source/drain-canal. Un éloignement de ces jonctions augmente le V_{th} par effet électrostatique et diminue donc le courant I_{on} .
- X4 L'augmentation de l'épaisseur de la couche SMT accroît le courant par une augmentation de la mobilité due au stress mécanique.
- X5 Un recuit plus chaud va faire diffuser les source/drain un peu plus et, par conséquent, raccourcir le canal et donc donner plus de courant.

TABLE 3.4 – Variations des paramètres du procédé, utilisées dans la propagation de la variance.

Facteur		unité	μ	σ
X1	épaisseur d'oxyde de grille	Å	23,95	0,06
X2	longueur de grille	nm	40,89	0,5
	épaisseur du <i>liner</i>	Å	41,62	0,3
	longueur du <i>spacer zero</i>	nm	6,02	0,1
X3	épaisseur de nitrure d'espaceur	nm	37,27	0,22
X4	épaisseur de la couche SMT	nm	33	0,22 ^a
X5	température de recuit	K	999,47	2,23

^a. Il n'y pas de mesures. Je fais l'hypothèse que cette couche SMT a la même variation que la couche HCD des espaceurs.

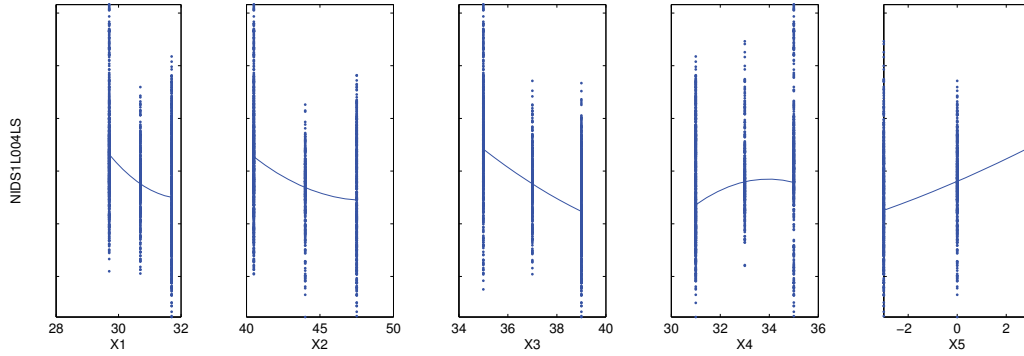


FIGURE 3.14 – Effet de chaque variable technologique du modèle indépendamment sur le modèle de NIDS1L004LS, dans son domaine de définition. Les mesures du courant sont superposées au modèle.

Significativité des effets

Les effets que nous avons calculés sont une estimation à partir d'un échantillon et donc leur valeur est connue à un intervalle de confiance près. Le test de Student permet de tester la significativité de ces coefficients en déterminant si leur intervalle de confiance contient la valeur 0. Auquel cas, il est alors probable qu'un coefficient soit nul, c'est à dire qu'il n'est pas significatif, ce qui nous permet de retirer le terme correspondant du modèle. Si la probabilité que 0 ne soit pas contenu dans l'intervalle de confiance du coefficient est supérieure à $\alpha = 5\%$, alors c'est que ce coefficient est significatif et on le conserve dans le modèle.

Avec l'hypothèse de normalité du résidu vérifiée, les coefficients du modèle calculé, k_i , suivent une loi de Student à $n - p - 1$ degrés de liberté, où n est le nombre d'observations et p le nombre de coefficients du modèle. Leurs intervalles de confiance avec un taux de confiance α s'écrivent ¹³ :

$$\left[\bar{k}_i - t_{(1-\alpha/2)} \frac{S}{\sqrt{n}}, \bar{k}_i + t_{(1-\alpha/2)} \frac{S}{\sqrt{n}} \right] \quad (3.11)$$

En pratique, on calcule le t de Student en divisant les effets par les erreurs standards qui sont les éléments diagonaux de $(X'X)^{-1}$ multipliés par l'écart-

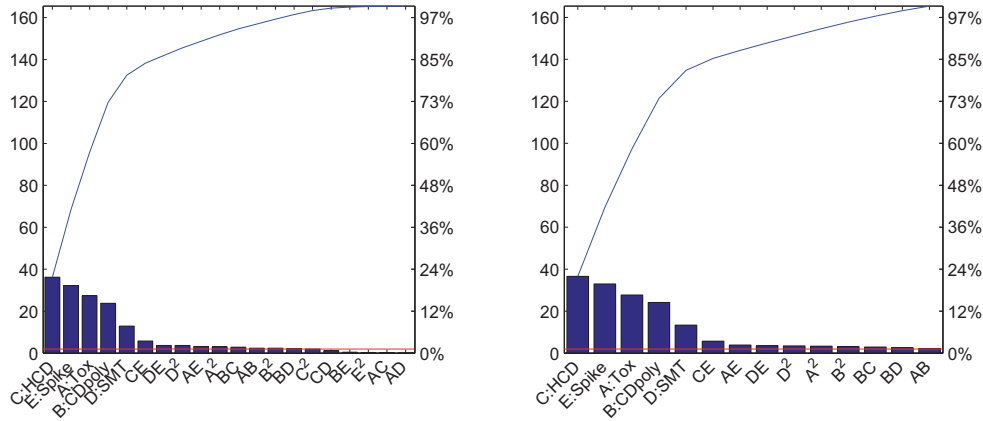
13. http://www.statelem.com/loi_de_student.php

type résiduel :

$$t \text{ de Student} = \text{Effets Standardisés} = \frac{K}{\sigma_e \times \text{diag}(X'X)^{-1}}. \quad (3.12)$$

On associe à ces t une probabilité P -value qui est l'image de la distribution de Student à $n - p - 1$ degrés de liberté. Les effets dont la valeur de risque est inférieure à 5 % sont jugés significatifs. Dans la figure 3.15 la barre rouge horizontale marque cette limite. On peut aussi comparer les effets standardisés à la densité de probabilité cumulée de la distribution de Student pour laquelle on a une probabilité de 97,5 % ($1 - \alpha/2$).

La figure 3.15a est un diagramme de Pareto. Les effets standardisés en valeurs absolues sont triés par ordre décroissant. On y apprécie à quels facteurs l'indicateur I_{on} est le plus sensible. La significativité des effets étant établie,



(a) Diagramme de Pareto du modèle complet

(b) Diagramme de Pareto du modèle simplifié

FIGURE 3.15 – Diagramme de Pareto de l'indicateur NIDS1L004 pour un modèle du premier ordre avec interactions. Pour alléger la lecture de l'axe des abscisses, les facteurs du plan sont désignés par une lettre et les termes d'interactions et les termes carrés par les combinaisons de ces lettres correspondantes.

on peut itérativement retirer du modèle les effets les moins significatifs et recalculer les modèles. On ne les retire pas tous simultanément. Comme les effets que l'on supprime se redistribuent sur les autres effets, on peut avoir

un effet non significatif qui le devient. La figure 3.15b présente le Pareto résultant de cette simplification.

3.3.3 Estimation de la contribution de chaque facteur

Le diagramme de Pareto 3.15, résultant du plan d'expériences et tel qu'il est défini dans les logiciels de statistiques, classe les *sensibilités* normalisées à l'écart type résiduel par valeurs absolues décroissantes. Cette représentation est très utile dans le cas où le plan d'expérience est utilisé pour une optimisation de process par exemple. Il indique sur quels facteurs agir en priorité pour régler le process. Dans notre cas, nous cherchons à déterminer quelle part de variance des indicateurs électriques est imputable à chaque facteur technologique θ .

Nous pouvons appliquer une propagation de la variance de chaque facteur pour estimer leurs contributions. En approximant le modèle au premier ordre autour du point de fonctionnement, la variance de l'indicateur Y est la somme des produits de sa sensibilité au carré par la variance des i variables technologiques X_i . On a :

$$\mu_y \approx y(\bar{\theta}) \quad (3.13)$$

$$\sigma_y^2 = \sum_{i=1}^n \frac{\partial y}{\partial \theta_i}^2 \sigma_{\theta_i}^2. \quad (3.14)$$

Dans le cas où le modèle de $Y(x_i)$ est peu non linéaire, cette approximation est correcte. La figure 3.16a illustre la propagation de variance dans le cas d'un modèle linéaire. La distribution de Y est au coefficient de sensibilité près la même que celle de X, sans déformation. Dans le cas d'un modèle quadratique (3.16b) la distribution est censée se déformer à cause de la courbure du modèle. La moyenne de la distribution de Y est également décalée par rapport à l'image de la valeur nominale de X par le modèle. L'approximation de linéarité peut sous-estimer la variance propagée et introduire une erreur sur la moyenne de sortie si la courbure est forte. On peut aussi estimer la propagation de la variance en faisant l'hypothèse d'une courbure du modèle du second degré autour du point nominal. Lei *et al.* [113] calculent la propagation de variance en faisant l'approximation d'une relation quadratique entre y et x. Ils repartent du développement du modèle en série de Taylor et gardent les termes de second ordre. Ils expriment alors la valeur moyenne

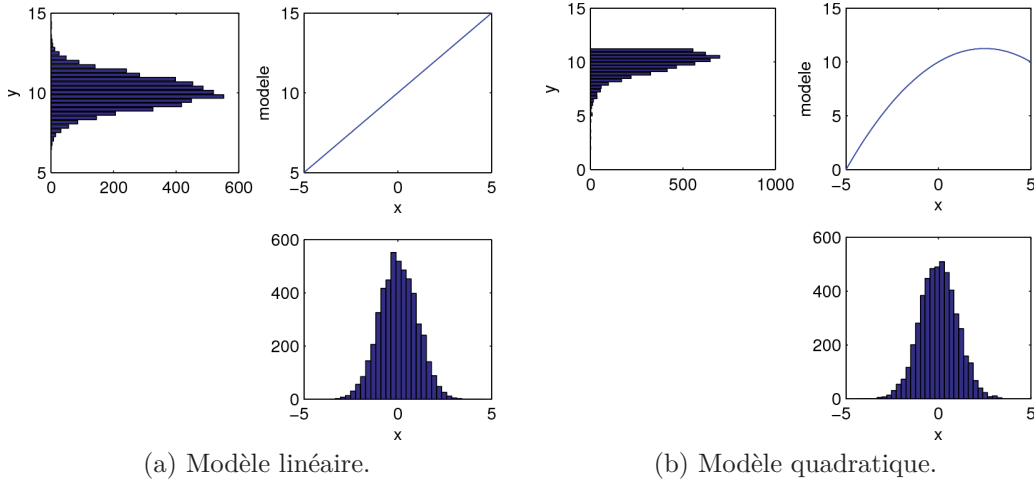


FIGURE 3.16 – Illustration de la propagation de la variance et de la déformation de distribution causée par un système non linéaire.

μ_y :

$$\mu_y \approx y(\bar{\theta}) + \frac{1}{2} \sum_{i=1}^t \frac{\partial^2 y}{\partial \theta_i^2} \sigma_{\theta_i}^2 \quad (3.15)$$

et la variance σ_y^2 :

$$\sigma_y^2 \approx \sum_{i=1}^t \left[\frac{\partial y}{\partial \theta_i} \right]^2 \sigma_{\theta_i}^2 + \frac{1}{2} \sum_{i=1}^t \left[\frac{\partial^2 y}{\partial \theta_i^2} \right]^2 \sigma_{\theta_i}^4 + \sum_{i < j}^t \left[\frac{\partial^2 y}{\partial \theta_i \partial \theta_j} \right]^2 \sigma_{\theta_i}^2 \sigma_{\theta_j}^2 \quad (3.16)$$

avec θ_i les t paramètres technologiques. Cette formulation introduit un terme supplémentaire dans μ_y qui corrige la valeur moyenne de la distribution de y et ajoute des termes à σ_y^2 en prenant en compte la courbure de la fonction $y(x)$. Si on compare les propagations de variance (3.15) et (3.16) à un tir Monte-Carlo de variables aléatoires normales sur un modèle quadratique, on constate à l'essai que (3.15) corrige bien l'erreur sur la moyenne de Y causée par la non-linéarité. Par contre, la variance estimée est assez différente de la variance du résultat du tir. Graupner *et al.* remarquent et étudient cet écart, mais concluent que, comme en pratique les termes quadratiques du modèle sont faibles devant les sensibilités linéaires, il n'est en général pas nécessaire d'utiliser les termes quadratiques dans la propagation de variance [114].

Application

Dans la suite de cette section, nous avons reporté pour chaque indicateur étudié les racines des variances propagées calculées par (3.16). Nous avons utilisé comme variation initiale les écarts-types des paramètres process mesurés sur les trois plaques qui sont les points au centre du plan, c'est à dire le process nominal. Pour l'épaisseur de la couche SMT pour laquelle nous n'avons pas de mesure, nous avons supposé que ces variations sont de l'ordre de celles du dépôt de nitrure des espaceurs.

La hauteur des barres (figures 3.17 à 3.22) est donc l'écart type estimé que causerait la variation d'une variable en l'absence de variation des autres variables¹⁴. La variance serait une grandeur plus adaptée à ce type de graphes car les variances se somment linéairement mais j'ai préféré montrer l'écart-type qui présente une grandeur homogène à la moyenne de la variable et qui est donc plus intuitif. La somme de ces variances est de l'ordre du rapport du R^2 et de la variance mesurée¹⁵. La moyenne μ_Y , l'écart-type σ_Y et le taux de déviation σ/μ de l'indicateur sont indiqués sur chaque diagramme à titre de repère pour les ordres de grandeur.

Courants

La figure 3.17 montre les contributions des facteurs technologiques pour les courants de drain en saturation, sur le dispositif N et le dispositif P de largeur nominale W de $0,6\ \mu\text{m}$ pour 2 longueurs de grille $L=40\ \text{nm}$ et $10\ \mu\text{m}$. Les valeurs moyennes des courants des transistors N sont le double de celles des transistors P de mêmes dimensions. Ceci est un point tout à fait normal car la mobilité des trous est deux fois moindre que celle des électrons et les concepteurs équilibrent les courants en adoptant des transistors P plus larges que les devices N.

Les variations relatives sur le dispositif long sont de l'ordre de 1 % et sur le court de 5 %. On note la contribution prépondérante de $\text{Spike}_{\text{temp}}$ aux variations des courants de saturation. Comme on peut s'y attendre son influence sur le transistor court est de loin la principale alors qu'elle est partagée avec T_{ox} sur le transistor long, où le contrôle électrostatique du canal par la grille est bon. Dans le dispositif court, où interviennent les effets

14. Pour rappel, les diagrammes de Pareto de la figure 3.15 représentaient les effets standards, c'est à dire rapportés à l'erreur standard.

15. sans déviation volontaire. Sur les plaques aux centres.

canaux courts, les facteurs arrivant juste après $\text{Spike}_{\text{temp.}}$ sont HCD_{th} et L_g .

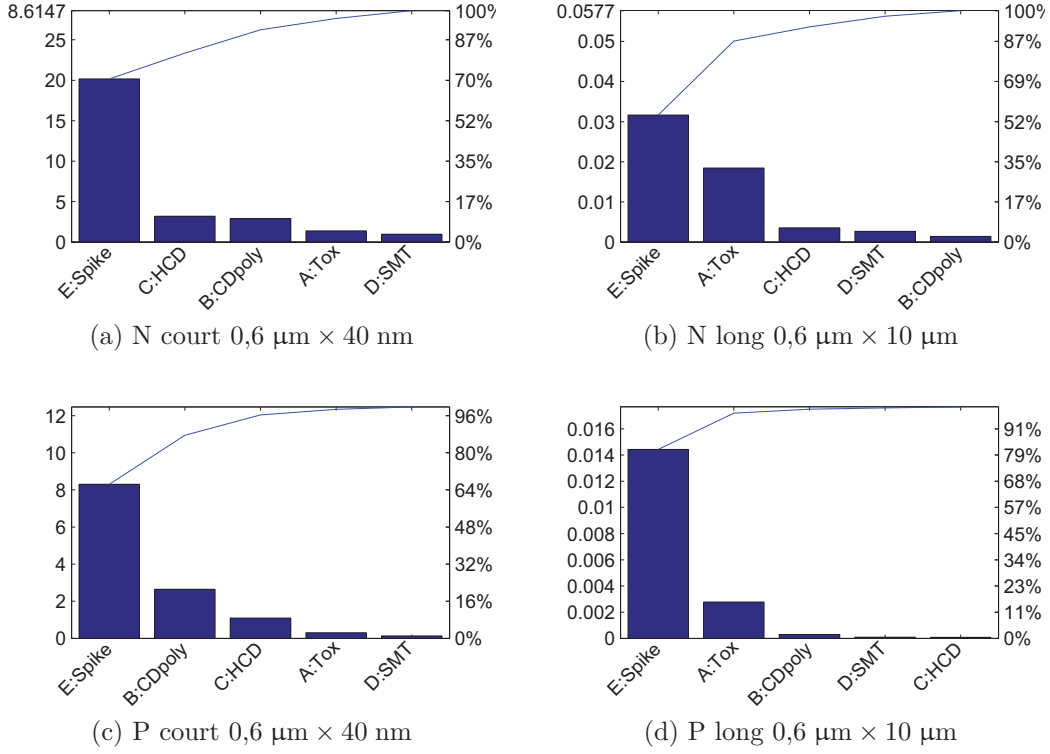


FIGURE 3.17 – Contribution des facteurs technologiques à la variation de I_{on} .

La figure 3.18 montre les contributions des facteurs pour les courants de drain en régime linéaire ($V_G=0,05 \text{ V}$). Le classement des deux facteurs principaux est identique à celui des régimes de saturation.

La différence de grandeur entre ces Pareto où la variable $\text{Spike}_{\text{temp.}}$ est prépondérante vient du fait que nous regardons l'influence résultante de la sensibilité *et* de la dispersion des facteurs. Dans le graphe standardisé, $\text{Spike}_{\text{temp.}}$ apparaît minimisé par la standardisation car sa variation naturelle est plutôt grande devant les variations volontaires du plan d'expériences. Les autres variables ne sont pas aussi dispersées par rapport aux variations volontaires du DoE – on peut revoir la figure 3.3, page 73 pour s'en convaincre – et donc leur erreur standard est de ce fait moindre.

Tensions de seuils

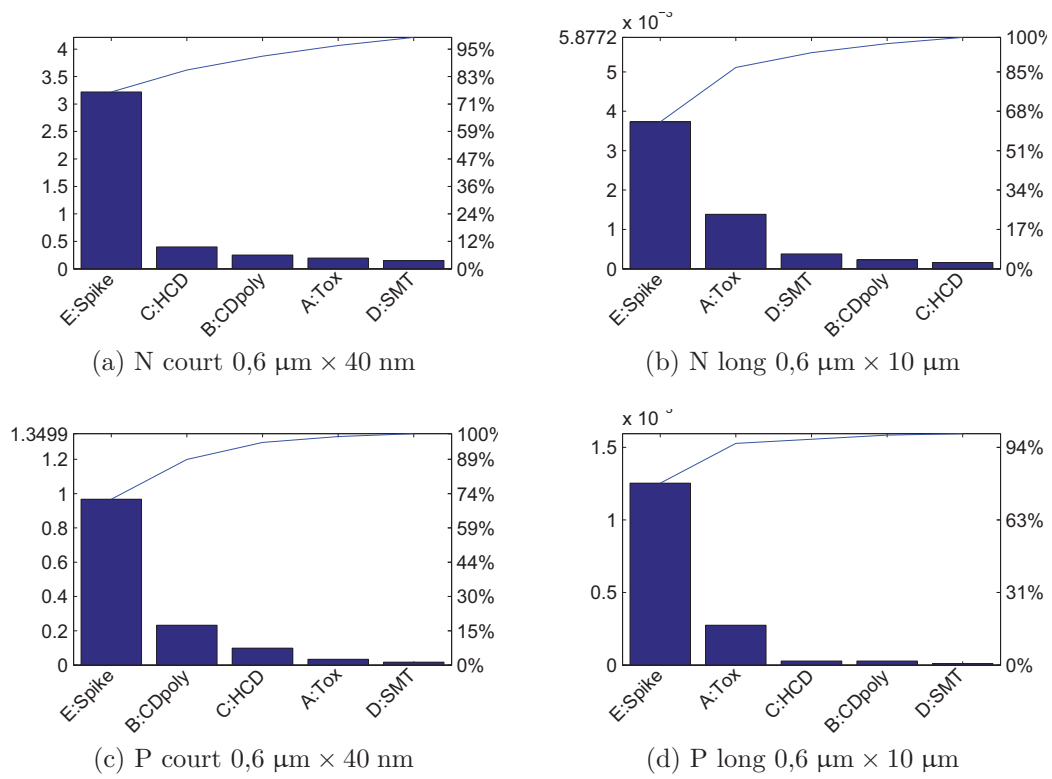
La figure 3.19 montre les contributions à la variation du V_{th} , en N et P, pour des devices de 40 nm et 10 μm de long, la largeur étant toujours de 0,6 μm . La tension de seuil est mesurée à densité de courant $I_d \cdot W/L$ constante.

Les contributions à la fluctuation du DIBL, la différence entre les tensions de seuils de saturation et de régime linéaire, sont montrées sur la figure 3.20.

Capacités

La figure 3.21 montre les contributions à la déviation de la capacité de grille C_{gg} , normalisées par W/L . Les longueurs de grille des dispositifs longs et courts sont 1 μm et 40 nm. Le transistor court montre une capacité surfacique (μ_y) supérieure due aux capacités parasites qui augmentent le C_{gg} apparent. Les taux de variations des dispositifs N et P sont similaires. Sur les dispositifs N longs, le premier facteur de variation est l'épaisseur d'oxyde de grille. La longueur de grille devient prépondérante sur le dispositif court. Dans le dispositif P, la température de recuit apporte plus de variabilité que T_{ox} et L_g . On peut raisonnablement penser que la diffusion des dopants de type P implantés dans les source/drain est plus sensible au recuit que celle des transistors N.

La figure 3.22 montre les contributions à la déviation de la capacité d'overlap C_{ov} pour les longueurs de grille de 1 μm et 40 nm. La température de recuit apparait comme facteur de variabilité dominant. En effet, la diffusion des *ldd* sous la grille donne directement la surface de cette capacité. On peut noter que la moyenne de cette capacité est inférieure dans le transistor court à la moyenne du transistor long. On peut supposer que la redistribution des implants de poche et du canal ralentissent la diffusion des implants *ldd*.

FIGURE 3.18 – Contribution des facteurs technologiques à la variation de I_{lin} .

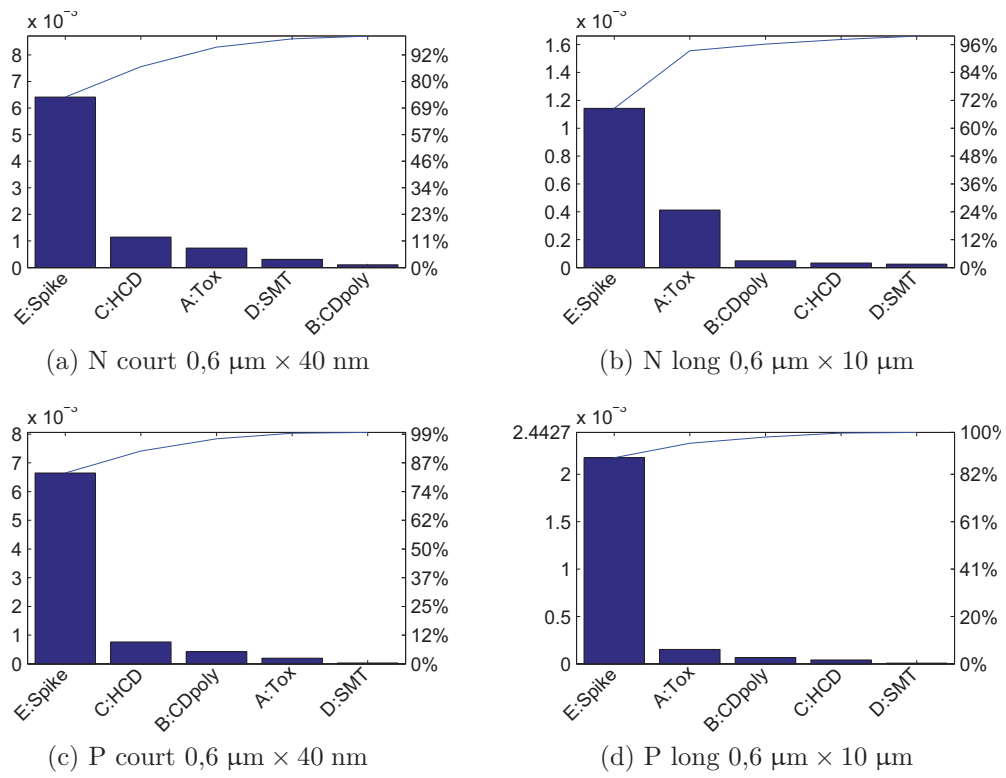


FIGURE 3.19 – Contribution des facteurs technologiques à la variation de V_{Tlin} .

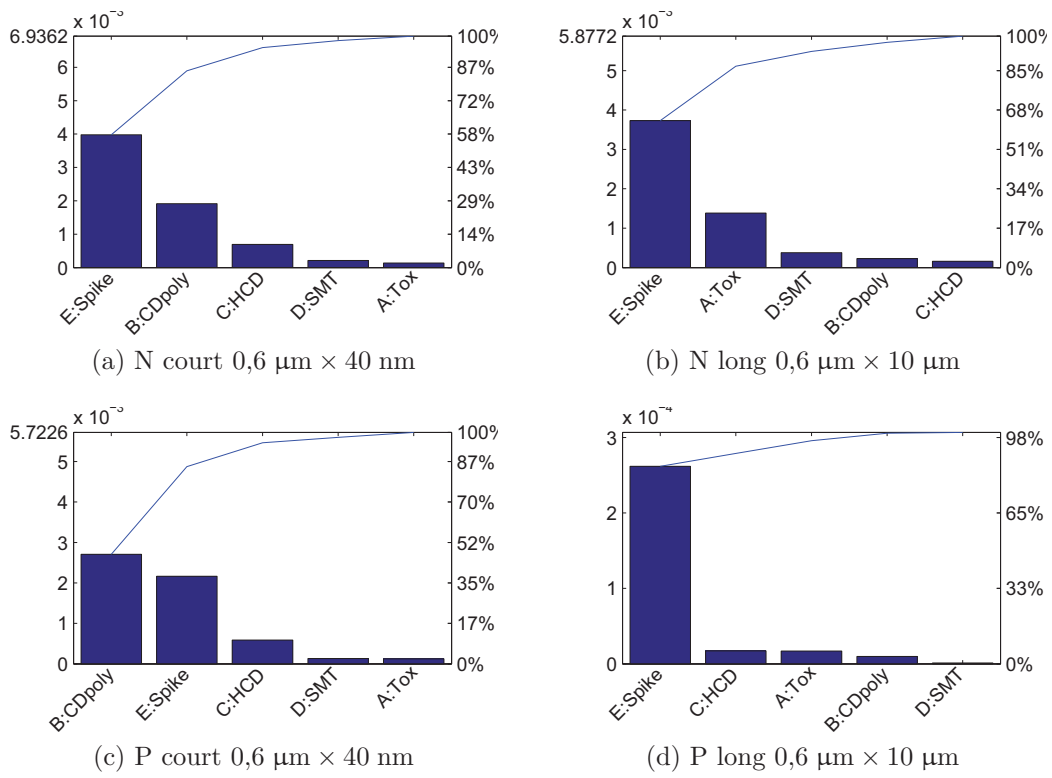


FIGURE 3.20 – Contribution des facteurs technologiques à la variation du DIBL.

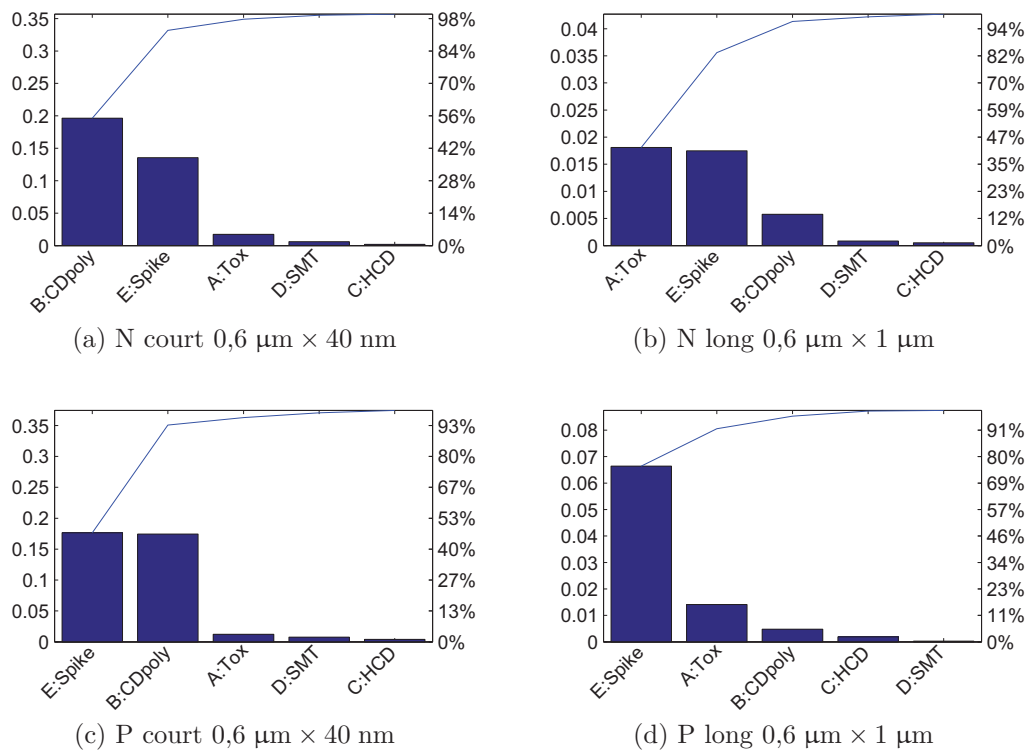


FIGURE 3.21 – Contribution des facteurs technologiques à la variation de C_{gg} .

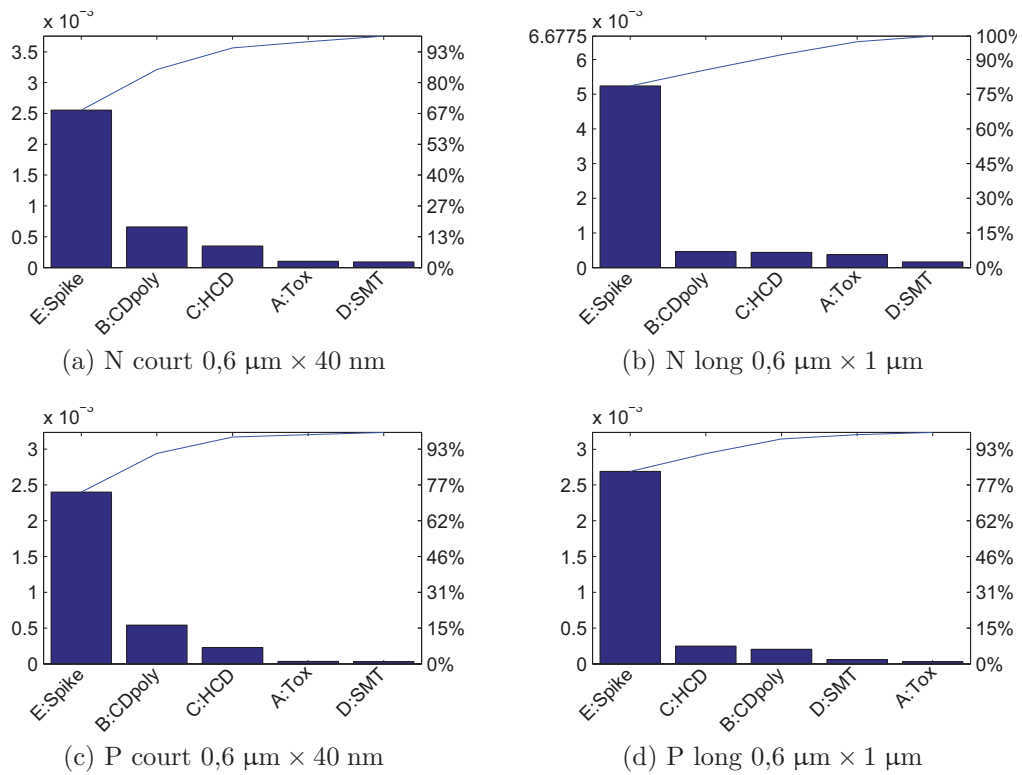


FIGURE 3.22 – Contribution des facteurs technologiques à la variation de C_{ov} .

3.4 Conclusions

Dans ce chapitre, nous avons examiné un jeu de données comprenant des mesures de paramètres technologiques et électriques importants des dispositifs MOSFET. Nous avons pu observer et analyser les variabilités de chacune de ces données en mettant en œuvre des méthodes statistiques adaptées.

L'exploitation du plan d'expériences et l'utilisation de la propagation de la variance nous ont permis d'extraire un modèle pour chaque indicateur électrique en fonction des variables technologiques, puis d'estimer l'importance relative des facteurs du plan dans la variabilité des indicateurs électriques.

La variabilité de la température de recuit apparaît comme le facteur prépondérant sur la majorité des observations. La variabilité de la variable HCD_{th} et donc des espaceurs apparaît comme le second facteur principal.

Chapitre 4

Simulation des variations dans les modèles compacts

Les modèles compacts, ou modèles de type SPICE, sont utilisés pour simuler le comportement électrique des circuits. Ces modèles sont un jeu d'équations analytiques associé à une liste de paramètres représentant les caractéristiques des transistors. Une partie des paramètres, définis pour chacun des transistors du circuit (W , L ...), sont appelés paramètres d'instance. Ces paramètres sont choisis par le *designer* lors de la conception du circuit. Une autre partie des paramètres est propre à la technologie employée par le transistor simulé et est donc commune à tous les dispositifs de cette technologie. Certains de ces paramètres d'entrée du modèle sont représentatifs des paramètres physiques comme les dimensions, les dopages, les caractéristiques des matériaux... Ce jeu de paramètres est fixé par les équipes d'ingénieurs de modélisation. Le choix du jeu de paramètres modèle, ou *calibration*, se fait en ajustant les paramètres par comparaison des caractéristiques importantes du transistor aux mesures correspondantes. Les modèles compacts doivent aussi permettre la simulation de la variabilité. Pour simuler la variabilité électrique, le modèle compact dispose de variables représentant les déviations des paramètres d'entrée du modèle. C'est sur ces déviations que joue le simulateur SPICE, Mentor Eldo dans notre cas.

Dans le cas d'une simulation Monte-Carlo, Eldo simule un grand nombre de fois le transistor en appliquant à chaque déviation une distribution gaussienne (ou uniforme, selon le paramètre) centrée en zéro. En choisissant quels paramètres d'entrée participent au tir Monte-Carlo, on peut simuler séparément les variabilités paramétriques et aléatoires. Pour chaque simulation

ayant des déviations différentes, on a des caractéristiques électriques simulées différentes. Un grand nombre de tirs Monte-Carlo est nécessaire pour obtenir une distribution des résultats qui soit statistiquement représentative.

Les modèles compacts sont également livrés avec des réglages de simulation *pire-cas* ou *corners*. Dans le cas des circuits numériques où l'on s'intéresse en premier lieu à la tolérance sur les délais des portes logiques, ce genre de simulation permet d'éviter un temps de calcul énorme en ne sortant que 5 valeurs censées encadrer la variabilité globale des performances principales du transistor. Dans un cas plus général (circuits analogiques), la construction des modèles pire-cas doit être adaptée spécifiquement à chaque critère de performance (gain, produit gain-bande...)

4.1 Simulations Monte-Carlo

La simulation Monte-Carlo est la référence sur laquelle s'appuieront les modèles corners. Elle doit permettre de propager les déviations des paramètres technologiques à travers le modèle SPICE, sans faire d'approximation sur les corrélations entre les facteurs ou sur le modèle lui-même.

4.1.1 Construction de la simulation Monte-Carlo

La figure 4.1 donne une vue d'ensemble de la propagation des déviations technologiques dans le modèle PSP. Les données provenant des mesures ne

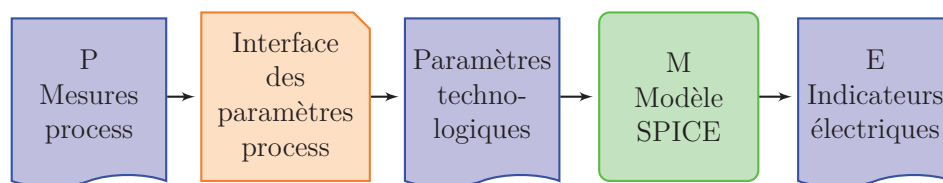


FIGURE 4.1 – Propagation des paramètres process

sont pas introduites directement dans le cœur du modèle compact car les différents modèles compacts (PSP, BSIM, hSpice...) n'utilisent pas nécessairement les mêmes grandeurs physiques comme paramètres d'entrée et n'utilisent certainement pas toutes les grandeurs physiques. L'interface évoquée dans le schéma de la figure 4.1 permet la transformation des données de l'espace des paramètres process P vers les paramètres modèle M . Les paramètres

process pris en compte comme points d'entrée sont listés dans le tableau 4.1. Ils sont associés à des paramètres indépendants du modèle qui propagent les variations aux variables du modèle.

TABLE 4.1 – Interfaçage des données process pour introduire les déviations dans le modèle PSP.

Paramètre process	Paramètre indépendant d'entrée du modèle	Paramètre du modèle
<i>Variabilité globale</i>		
épaisseur d'oxyde	TOXO	CTO, CTL, TOXOVO, CFL, ALPL, IGINVLW, IGOVW
longueur du polysilicium	LVAR	CSLW, VFBLW, DPHIBLW
dimension de la zone active	WVAR	
mobilité en champ faible	dmu	UO
tension de seuil	dvth0	NSUB0
Effets de canal court	dphib	DPHIBL
<i>Variabilité locale</i>		
mismatch de V_t	dvton_dev	VFB0
mismatch du facteur de	dbeta_dev	U0
gain	correlation	

Simulation des variations globales et locales

Nos fichiers de simulations, tels qu'ils sont prévus, donnent le choix de simuler la variabilité paramétrique ou globale indépendamment de la variabilité aléatoire ou locale. Les variations globales sont induites dans l'espace M par un jeu de paramètres P différents de celui qui induit la variabilité locale, tel qu'indiqué dans les deux parties du tableau 4.1. Les paramètres T_{ox} , L , W et la mobilité sont utilisés pour prendre en compte les déviations systématiques. Les fluctuations de dopage sont traduites par des variations de V_{th} et de Φ_B . Les variations locales sont prises en compte *via* V_{fb} et μ_0 .

La variabilité globale est ce que simule naturellement un tir Monte-Carlo. À chaque itération du tir Monte-Carlo, le simulateur tire aléatoirement les valeurs des variables technologiques selon la définition qui leur est attribuée dans les fichiers de paramétrage de la simulation. En présence de variabilité globale uniquement, tous les dispositifs d'un même circuit sont simulés avec les mêmes valeurs de paramètres process. En pratique, on indique

quel type de simulation on veut lancer en sélectionnant une bibliothèque différente : corners, Monte-Carlo, *user-defined*. . . On y définit quels paramètres *process* à utiliser ainsi que la nature de la distribution et l'amplitude de leurs déviations à appliquer à la simulation.

La variabilité locale est simulée en mettant en place un schéma reproduisant le principe des paires de transistors utilisées pour mesurer le désappariement. À chaque transistor dont on veut évaluer la variabilité intrinsèque on associe un transistor identique en termes de paramètres d'instance. À chaque itération Monte-Carlo, les paramètres choisis pour représenter le désappariement sont différents pour chaque dispositif du circuit et suivent une loi normale – c'est-à-dire gaussienne – définie dans les fichiers de configuration de la simulation. Des extraits de *netlist* qui mettent en œuvre ces simulations sont donnés en annexe A.3, page 138.

Nous considérons les variations locales statistiquement indépendantes et de même écart-type :

$$\text{Définissons } \Delta V_t = V_{t1} - V_{t2} \quad (4.1a)$$

$$\text{Si les variations de } V_{t1} \text{ et } V_{t2} \text{ sont faibles } \partial V_t = \partial V_{t1} - \partial V_{t2} \quad (4.1b)$$

$$\text{et on peut donc écrire } \sigma_{\Delta V_t}^2 = \sigma_{\Delta V_{t1}}^2 + \sigma_{\Delta V_{t2}}^2. \quad (4.1c)$$

$$\text{Si on part de l'hypothèse que } \sigma_{V_{t1}} = \sigma_{V_{t2}} = \sigma_{V_t} \quad (4.1d)$$

$$\text{alors } \sigma_{\Delta V_t}^2 = 2\sigma_{V_t}^2, \quad (4.1e)$$

$$\text{et donc } \sigma_{V_t} = \frac{\sigma_{\Delta V_t}}{\sqrt{2}}, \quad (4.1f)$$

d'où les lignes 26 et 35 du code A.2 en annexe, page 139. Cette formulation permet de garder l'information du désappariement découplée de la variation globale dans le cas où les deux types de variations sont activés.

4.2 Génération de Corners

Afin de simuler le comportement des circuits sous l'effet des variations locales et globales – ou respectivement intrinsèques et paramétriques telles que nous les avons nommées jusque là – la méthode de référence reste le tir Monte-Carlo. Cependant, faire un tir Monte-Carlo en prenant en compte des variations des paramètres géométriques et de dopage nécessite un échan-

tillonnage suffisant pour que les distributions des indicateurs simulés soient valables. Si le nombre de simulations du circuit entier est de 1000, le temps total de simulation pour un circuit comportant des millions de transistors devient rédhibitoire.

Pour pallier à cela, on définit des conditions de simulations représentant les pires cas de fonctionnement. On utilise couramment le terme anglais *corners* pour désigner ces points de fonctionnement. Ces *corners* sont un jeu de valeurs de déviations des paramètres process.

4.2.1 Méthode standard

On construit les corners pour un type de performance choisi. On donne une certaine variation à tout un jeu de paramètres d'entrée du modèle afin de reproduire une certaine gamme de variations dans les caractéristiques électriques du transistor auxquelles les performances circuit sont les plus sensibles. Le schéma en figure 4.2 illustre l'encadrement des variations par les corners destinés aux applications numériques.

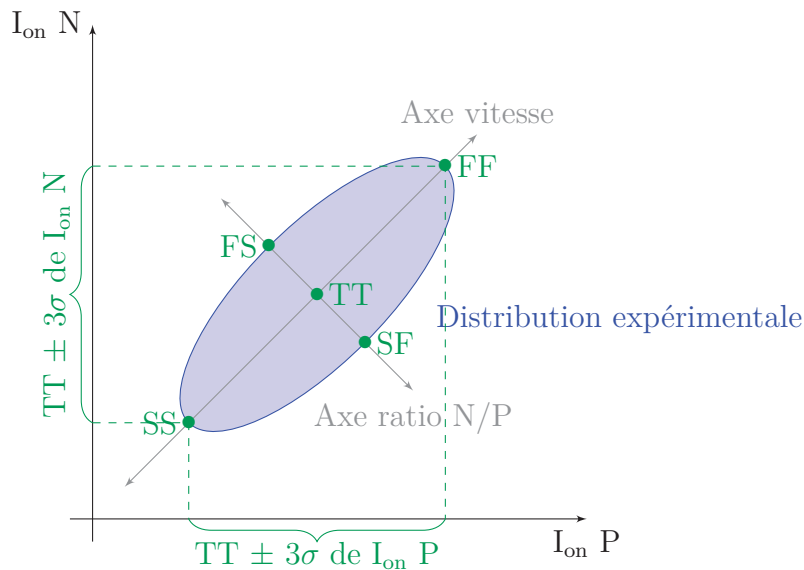


FIGURE 4.2 – Placement des *corners* dans le plan de corrélation $I_{on}N/I_{on}P$

Le tableau 4.2 détaille les déviations appliquées aux variables d'entrée du modèle. Les différents corners sont nommés par deux lettres, la première pour le dispositif nMOS et la deuxième pour le dispositif pMOS. Les différents corners sont :

- TT : les paramètres du nMOS et du pMOS sont à leur valeur typique,
- FF : les paramètres nMOS et pMOS ont des valeurs qui accélèrent le circuit numérique (F comme *fast*),
- SS : idem pour *slow*
- SF : on ralentit le nMOS et on accélère le pMOS,
- FS : on accélère le nMOS et on ralentit le pMOS.

La première ligne du tableau 4.2 donne la déviation appliquée au paramètre XL , longueur du transistor, pour chaque corner. La performance visée est la vitesse des circuits logiques. Donc, dans le cas du corner FF où l'on veut accélérer les transistors N et P, on choisit de diminuer le paramètre XL , ainsi, le canal est plus court et on a plus de courant. On peut faire un raisonnement similaire paramètre par paramètre pour remplir le tableau.

TABLE 4.2 – Entrée des Corners pré-définis.

Paramètre	Unités	SS	SF	TT	FS	FF
XL	m	$+2\sigma$	0	0	0	-2σ
XW	m	-2σ	0	0	0	$+2\sigma$
TOX	m	$+2\sigma$	0	0	0	-2σ
VTH0 N	V	$+2\sigma$	$+3\sigma$	0	-3σ	-2σ
VTH0 P	V	$+2\sigma$	-3σ	0	$+3\sigma$	-2σ
RDSW N	$\Omega \cdot m$	$+2\sigma$	0	0	0	-2σ
RDSW P	$\Omega \cdot m$	$+2\sigma$	0	0	0	-2σ
Mobilité N	m^2/Vs	-2σ	0	0	0	$+2\sigma$
Mobilité P	m^2/Vs	-2σ	0	0	0	$+2\sigma$
CJ surf. N	F/m^2	$+\Delta$	$+\Delta$	0	$-\Delta$	$-\Delta$
CJ surf. P	F/m^2	$+\Delta$	$-\Delta$	0	$+\Delta$	$-\Delta$
CJ périm. N	F/m	$+\Delta$	$+\Delta$	0	$-\Delta$	$-\Delta$
CJ périm. P	F/m	$+\Delta$	$-\Delta$	0	$+\Delta$	$-\Delta$
I diode N	A	$\times 0.1$	0	0	0	$\times 10$
I diode P	A	$\times 0.1$	0	0	0	$\times 10$

Par construction, les corners reflèteront les variations d'un groupe de performances, mais une famille de corners ne peut pas être correcte pour toutes les performances. Le tableau 4.3 indique dans quel sens varie tel paramètre quand on augmente telle variable d'entrée. Par exemple, pour la première case, on note que I_{on} décroît si on augmente L_g . Donc, on sait qu'on doit augmenter L_g pour définir le corner *Slow*. En disposant de ces informations, on connaît un jeu de déviations pour chaque ligne, soit : un corner. On peut observer que la définition des corners est la même pour I_{on} et I_{off} . Pour V_{th} , les déviations sont opposées sur toute la ligne. La définition est donc la même

à un facteur -1 près. Ce corner encadrera simplement la performance dans l'autre sens (SS sera supérieur au nuage de points et FF inférieur). Dans le cas de C_{gg} , le corner est une autre combinaison de déviations. C'est pour cela que le corner défini pour capter les variations des performances de vitesse et donc de I_{on} ne captera pas nécessairement les variations de C_{gg} .

TABLE 4.3 – Les corners digitaux ne peuvent pas, par construction, couvrir les performances analog.

Perf.	Paramètre	LW	XW	Tox	VTH0	RDSW	Mob.
Digital	I_{on}	↘	↗	↘	↘	↘	↗
	I_{off}	↘	↗	↘	↘	↘	↗
	V_{Tlin}	↗	↘	↗	↗	↗	
	V_{Tsat}	↗	↘	↗	↗	↗	
Analog	C_{gg}	↗	↗	↘			

4.2.2 Statut par rapport aux résultats Monte-Carlo

Nous avons simulé électriquement les transistors N et P de longueur 40 nm et de quatre largeurs de canal 1 μm , 600 nm, 300 nm et 150 nm. Dans les graphes suivants, les simulations de corners sont superposées aux résultats Monte-Carlo.

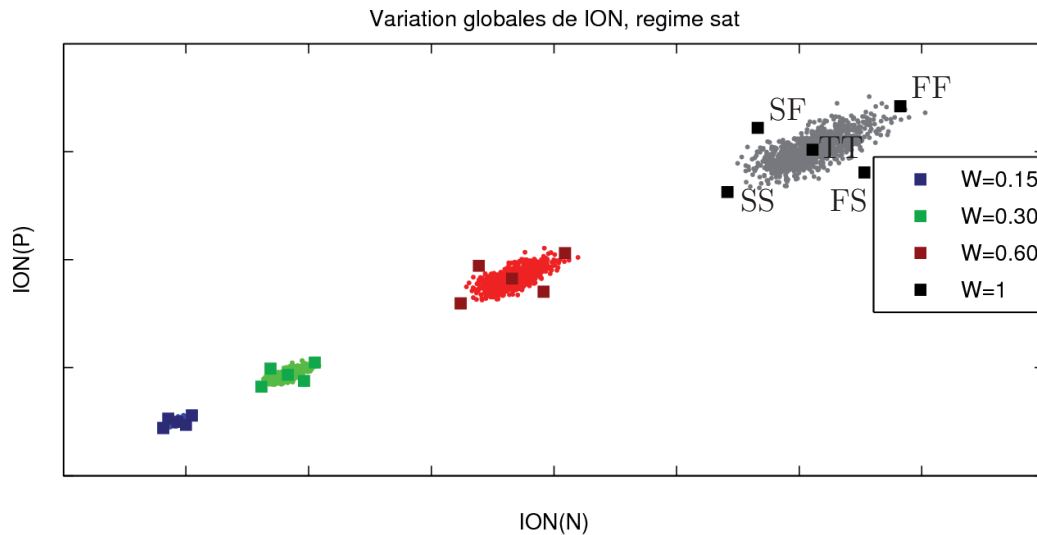


FIGURE 4.3 – Corrélation de I_{on} N et I_{on} P et placement des corners pour des transistors de longueur 40 nm.

La figure 4.3 présente la caractéristique principalement visée par les corners digitaux, la corrélation du courant I_{on} entre les transistors N et P. Pour chaque largeur, le corner TT est bien placé au milieu de la population Monte-Carlo générée en prenant en compte les variations globales. Les quatre autres corners encadrent bien, sans prendre trop de marge, la distribution. Le corner FF se trouve à l'extrémité de la distribution où les courants du transistor N et du transistor P sont majorés, entraînant un comportement plus rapide que le cas nominal. Inversement pour le point SS, N et P débitent moins de courant. Les points FS et SF encadrent les variations dans le cas où les dispositifs N et P varient dans des directions opposées.

Les (nombreuses) figures de cette section ont été déplacées en annexe A.1, page 125.

Sur la figure A.1, nous observons les distributions du courant I_{on} normalisé à 1 μm de largeur issues de simulations Monte-Carlo des variations locales – *intradie* –, des variations globales – *interdie* – et des variations globales et locales en même temps. De bas en haut sont affichés les résultats des 4 transistors, du plus large au plus étroit. La légende de l'axe x de chaque graphe indique les dimensions. Par exemple, `Classic_IONU_WOP15U_L40N` indique le modèle de variation `Classic` qui est le modèle de référence et l'indicateur `IONU` (I_{on} normalisé à 1 μm) pour un canal de 0,15 μm de large (`WOP15`) et 40 nm de long (`L40N`).

On note la dérive de la valeur moyenne du courant qui augmente sur les transistors étroits. Les déviations globales du courant unitaire sont constantes selon la largeur W . Les déviations locales augmentent et prennent le dessus dans les petits transistors, comme le décrit l'équation de Pelgrom (1.4).

En trait rouge sont précisées les valeurs de la moyenne m et des déviations à trois écarts-type 3σ . Le corner TT correspond bien à la moyenne et les corners FF et SS contiennent bien l'étendue à 3σ des variations globales. C'est le comportement attendu des modèles corner.

La figure A.2 représente les simulations de I_{off} selon le même formalisme. `LIOFFU` est $\log(I_{off})/W$. La valeur moyenne de ce courant unitaire décroît dans les transistors étroits, c'est le *inverse narrow width effect*. Comme pour I_{on} , les déviations locales et globales sont de même amplitude pour le transistor large. Les déviations globales suivent la diminution des dimensions mais les variations locales dominent dans le dispositif étroit. Les corners SS et FF encadrent bien les variations globales à 3σ .

Le comportement de V_{th} est totalement analogue au comportement de I_{on} (figure A.3). Les deux types de variations sont du même ordre de grandeur sur le transistor large et les variations locales sont prépondérantes dans le dispositif étroit. Le corner *fast* limite les valeurs hautes de V_{th} , le corner *slow* limite les valeurs basses.

Sur la figure A.4 où sont tracés les résultats de C_{gg} , on peut remarquer que les variations locales sont très faibles par rapport aux variations globales et la tendance s'accroît dans les dispositifs plus étroits. Comme l'expliquait le tableau 4.3, les corners numériques ne sont pas prévus pour couvrir également les variations de C_{gg} par construction. En effet, ici, les points FF et SS sont extrêmement optimistes et ne rendent pas compte des déviations globales.

La figure A.5 représente les corners de C_{gg} superposées aux simulations Monte-Carlo avec le modèle actuel. On y voit, selon ce point de vue différent des histogrammes, le manque de déviations des corners FF et SS dans l'axe vitesse. Les déviations non corrélées (l'axe FS-SF), par contre, semblent bien représentées.

La figure A.6 représente les histogrammes des variabilités de C_{gd} .

La figure A.7 représente les histogrammes des variabilités de C_{bd} .

La figure A.8 représente les déviations de la transconductance g_m . Le modèle est très largement optimiste pour cet indicateur.

La figure A.9 représente les variations de la conductance de sortie g_d . Bien que moins extrêmes, les corners sont moins optimistes pour ce gain.

Un oscillateur en anneau – RO, *Ring Oscillator* – est un circuit composé d'une série d'inverseurs en boucle fermée, comme décrit par la figure 4.4. Un inverseur est composé d'un PMOS et d'un NMOS câblé comme indiqué sur la figure 4.5. Un *ring oscillator* oscille car son nombre de portes N est impair.

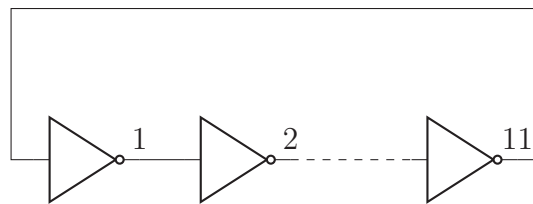


FIGURE 4.4 – Schéma d'un oscillateur en anneau.

En effet, la sortie de la $N^{\text{ième}}$ porte passe au niveau logique opposé à celui de l'entrée après un temps correspondant à N fois le temps de basculement d'une porte. Comme la sortie de la dernière porte est rebouclée sur l'entrée

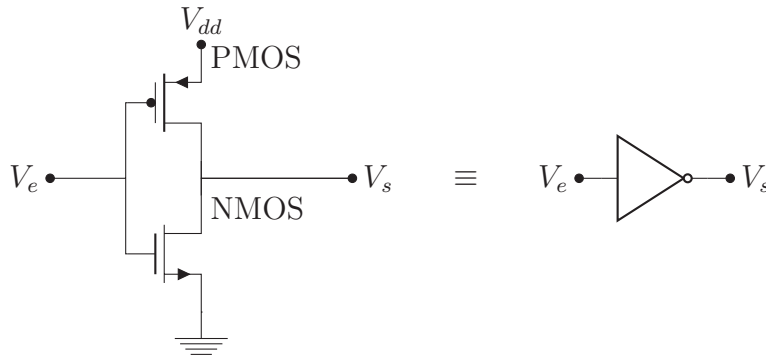


FIGURE 4.5 – Schéma d’un inverseur logique en technologie CMOS et sa représentation en porte logique.

de la première porte, les portes recommencent à basculer. La fréquence de fonctionnement d’un oscillateur en anneau dépend du nombre de portes que le signal doit parcourir et du temps de basculement de chacune d’entre elles.

La figure A.10 représente les variations de la fréquence d’oscillation d’un oscillateur composé de 11 inverseurs. Les corners y sont correctement placés, encadrant la gamme $\pm 3\sigma$ des déviations globales, mais sont assez pessimistes, c’est à dire qu’ils donnent un peu trop de marge.

Les variations d’origine locale de cette fréquence sont notamment faibles en comparaison aux déviations globales alors que ces déviations sont prépondérantes sur les caractéristiques transistors seuls. Les variations locales sont moyennées à l’échelle de l’oscillateur.

La figure A.11 montre la population du temps de montée T_{rise} d’un des transistors du circuit lors d’un basculement. Les variations locales sont très présentes à l’échelle d’un seul inverseur.

4.2.3 Génération spécifique à une performance

Lin *et al.* [115] proposent une méthode de choix des corners basée sur les résultats Monte-Carlo d’une performance en particulier. Ils choisissent parmi les jeux de paramètres process générés aléatoirement pour le tir Monte-Carlo ceux qui donnent les résultats médians, à 1σ , 2σ , ... De cette manière, les jeux de corners sont très spécifiques à une performance et une amplitude de variation donnée mais on ne fait pas de supposition quant aux corrélations entre facteurs de variations, puisqu’on choisit simplement les corners dans l’espace modèle selon les résultats de simulation.

4.2.4 Détermination non paramétrique multidimensionnelle

Rappitsch *et al.* proposent une manière fondamentalement différente de construire les modèles corners [116]. La méthode est basée sur des outils d'analyse de données permettant de définir ce qu'on appellera la *profondeur* d'un point dans un nuage de données. Cet outil statistique (*Location Depth*) a été développé par Rousseeuw et Struyf, comme une extension multidimensionnelle de la profondeur de Tukey. Le point de profondeur maximale est la médiane pour une variable aléatoire unidimensionnelle de n points, les extrema ont une profondeur minimale de 1 et la médiane a une profondeur maximale $n/2$. Dans un cas multidimensionnel ($d > 1$), on a une surface (ou une hypersurface si $d > 3$) de points de profondeur minimale et un point de profondeur maximale. En appliquant cette méthode à une sélection de paramètres process mesurés, on détermine les dispositifs les plus éloignés du dispositif médian. Ensuite, on applique une transformation des jeux de paramètres process des dispositifs de profondeur minimale pour déterminer le jeu de paramètres modèle correspondant. Ces points sont finalement des corners déterminés de manière non paramétrique, c'est-à-dire sans utiliser d'hypothèse de distribution gaussienne et ses paramètres associés m et σ . On utilise seulement l'observation des mesures.

On peut également prendre de la marge par rapport à cette surface extérieure du nuage de points en étendant la surface définissant l'enveloppe du jeu de données selon les directions radiales des points extérieurs (une transformation linéaire). Les points de la nouvelle enveloppe ainsi déterminée ne représentent pas directement des points mesurés mais anticipent les possibles fluctuations supérieures existantes sur les dispositifs non mesurés dans le jeu de données.

Cette méthode a été peu reportée dans la littérature depuis mais elle est assez prometteuse quant à la possibilité d'avoir des corners qui ne soient pas pessimistes et donc, entraînant un effort de design supérieur à ce qui est vraiment nécessaire et qui soit plus générique par rapport aux performances.

4.3 Modification du modèle de variation

Nous avons montré, dans le chapitre précédent, l'existence de la variation des espaceurs parmi les principales sources de variabilité dans le transistor

TABLE 4.4 – Paramètres process et les paramètres modèle les représentant *ou pouvant les représenter* dans PSP.

Paramètre process	PSP	
X1 : T_{ox}	TOXO	
X2 : L_g	LVAR	
TEOS _{th} , <i>spacer-zéro</i>	LAP, LOV	LAP pour les courants, LOV pour les capacités
X3 : HCD _{th}	CFL	Présence de diffusion des source/drain influant sur l'électrostatique en profondeur
X4 : <i>SMT</i>		
X5 : Spike _{temp.}	LAP, LOV	

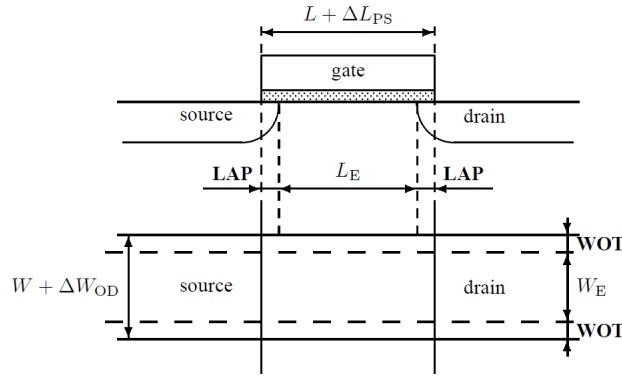


FIGURE 4.6 – Spécification des dimensions d'un transistor MOS dans PSP.

MOS. La construction du modèle de variations telle qu'elle est aujourd'hui ne prend pas en compte de déviation pour le paramètre LAP, *i.e.* pas de variabilité du positionnement de l'implantation du LDD par rapport au bord de grille. Dans le cas d'une architecture classique de MOSFET, la déviation de l'oxide *liner* présent lors de l'implantation des *ldd* est faible et la part de variabilité de la capacité de recouvrement des grille/drain est négligeable en comparaison de la variabilité de la capacité grille/canal. Dans le cas du lot étudié, l'architecture à double espaceur, introduite pour repousser les implants *ldd* apporte sa part de variabilité. La figure 4.6 présente les paramètres relatifs à la géométrie du MOSFET dans PSP.

La modification que nous apportons au modèle afin d'introduire la variabilité du spacer porte sur les paramètres LAP et LOV. En effet, dans PSP ces deux paramètres expriment séparément le recouvrement de grille/-drain. Le paramètre LAP impacte le calcul des courants dans le canal et la capacité intrinsèque de grille, et le paramètre LOV impacte les capacités

grille-source/drain. Au paramètre LAP nous définissons une variation DEV-LAP. C'est une variation de longueur qui viendra s'ajouter à LAP lors des simulations avec prise en compte des variations. DEVLAP agrandit LAP et diminue L_{eff} d'autant. La fluctuation que nous définissons pour le tir Monte-Carlo est une variation gaussienne, d'écart type $\sigma_{LAP} = 0,6$ nm, d'après les observations que nous avons faites en mesure.

Les résultats que nous attendons sont :

- peu de changement au niveau de la capacité de grille C_{gg} ,
- une augmentation de l'ordre de 10% pour les courants de canal I_{lin} , I_{on} et I_{off} ,
- une augmentation significative des fluctuations de la capacité C_{gd} et des caractéristiques des ring oscillators.

$$L_{poly} = L_D + LVAR, L_D \text{ pour } L \text{ dessiné.} \quad (4.2)$$

$$Tspacer = Tspacer_{typ} + \Delta spacer, \quad (4.3)$$

avec $\sigma(\Delta spacer) = 0,6$ nm et $m(\Delta spacer) = 0$

$$LVAR = LVAR_{typ} + \Delta LVAR \quad (4.4)$$

$$LAP = LAP_{typ} + \Delta LAP \quad (4.5)$$

avec $LVAR_{typ} = 2$ nm et $\Delta LVAR$ centré en 0 avec $\sigma(\Delta LVAR) = 1,2$ nm. $LAP_{typ} \simeq 5$ nm pour le device N comme pour le P, et $\Delta LAP = -\Delta Spacer$.

$$\begin{aligned} Le_{channel} &= L_{poly} - 2 \cdot Lap \\ &= L_D + LVAR - 2 \cdot Lap \\ &= L_D + LVAR_{typ} + \Delta LVAR - 2(LAP_{typ} + \Delta LAP) \\ &= L_D + LVAR_{typ} - 2 \cdot LAP_{typ} + \Delta LVAR + 2 \cdot \Delta Spacer \end{aligned} \quad (4.6)$$

avec $\sigma(\Delta LVAR) = 1,2$ nm et $\sigma(\Delta Spacer) = 0,6$ nm.

$$Lechannel_{typ} = 30 + 2 - 2 \times 5 = 22 \text{ nm typiques} \quad (4.7)$$

$$\sigma(Lechannel_{typ}) = \sqrt{1,2^2 + 4 \times 0,6^2} = 1,7 \text{ nm} \quad (4.8)$$

$$\frac{\sigma(Lechannel_{typ})}{Lechannel_{typ}} = \frac{1,7}{22} = 7,7\% \text{ avec } \Delta Spacer \quad (4.9)$$

$$= \frac{1,2}{22} = 5,5\% \text{ sans } \Delta Spacer \quad (4.10)$$

Dans PSP, $Le_{channel}$ contrôle I_{on} , $dCg_{channel}$. LOV définit l'overlap pour la capacité C_{gd} .

$$LOV = LOV_{typ} + \Delta LOV \quad (4.11)$$

$$\simeq LAP_{typ} + \Delta lap \quad (4.12)$$

$$= LAP_{typ} - \Delta Spacer \quad (4.13)$$

$$\frac{\sigma(LOV)}{LOV} = \frac{0,6}{5} = 12\% \quad (4.14)$$

4.3.1 Résultats

Nous régénérons les graphes de la section 4.2.2 avec l'introduction de la variabilité du paramètre DEVLAP.

La figure 4.7 montre le placement des corners dans le plan $I_{on} N / I_{on} P$. On observe une dispersion légèrement supérieure au modèle d'origine (figure 4.3), comme attendu. De par l'introduction d'une source de variation supplémentaire (Δ_{spacer}) pour le contrôle de la longueur effective du canal. Les corners encadrent correctement la distribution Monte-Carlo.

La figure 4.8 représente les histogrammes des distributions Monte-Carlo de C_{gg} pour 4 longueurs de grille différentes et superpose les corners. On ne note pas de différence majeure avec le modèle d'origine (figure A.4), comme prévu du fait de la compensation entre la variation supplémentaire de la longueur électrique du canal d'une part, et de la somme des variations de longueur de recouvrement grille-source/drain d'autre part.

La figure 4.9 montre la comparaison des corners et des distributions Monte-Carlo pour C_{gd} . On trouve une augmentation très significative de la dispersion Monte-Carlo de C_{gd} par rapport au modèle d'origine (figure A.6) qui ignore la variation Δ_{spacer} . Les corners, par contre, n'encadrent pas mieux

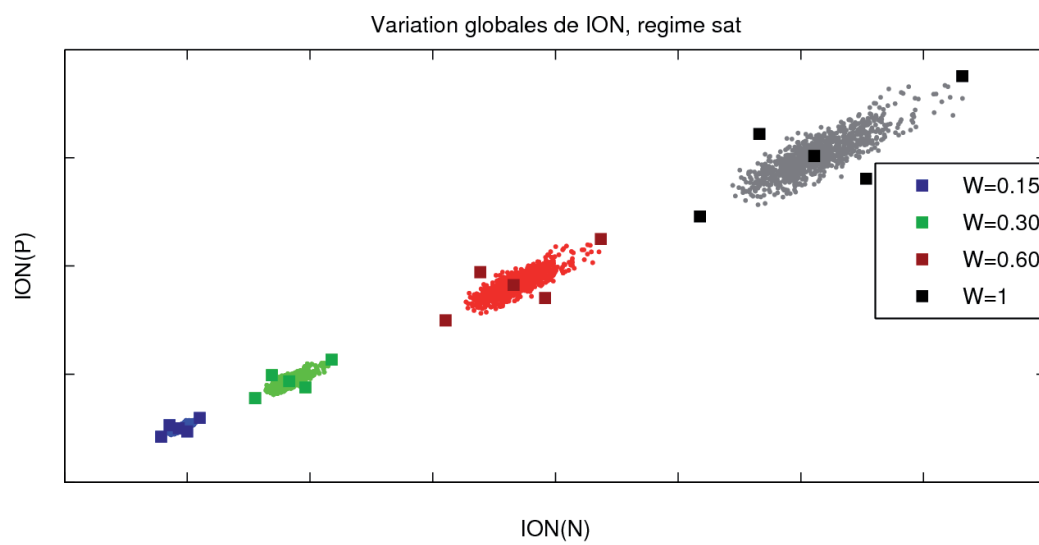


FIGURE 4.7 – Corrélation de $I_{on} N$ et $I_{on} P$ et placement des corners.

la distribution statistique.

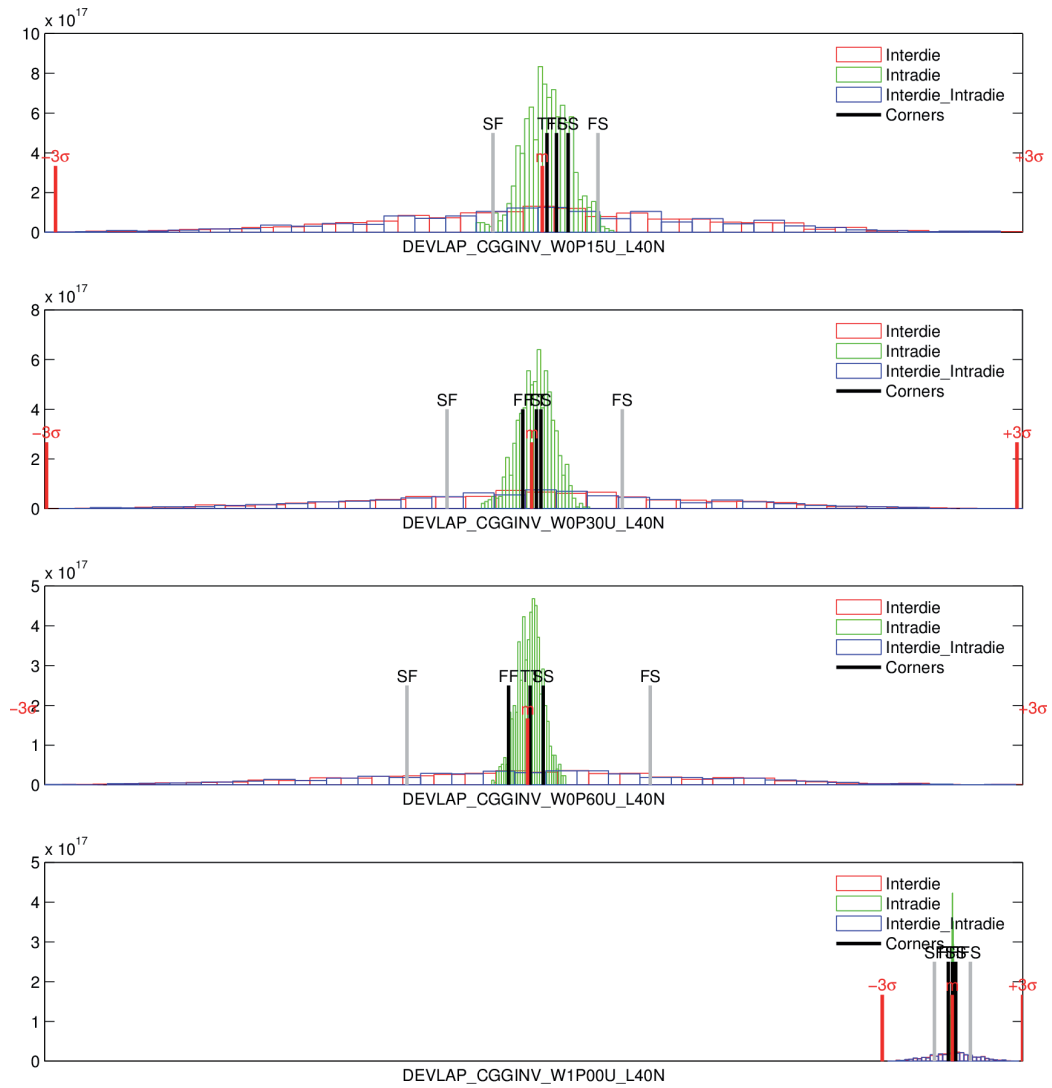


FIGURE 4.8 – Histogramme de C_{gg} résultant de la simulation Monte-Carlo avec le modèle modifié.

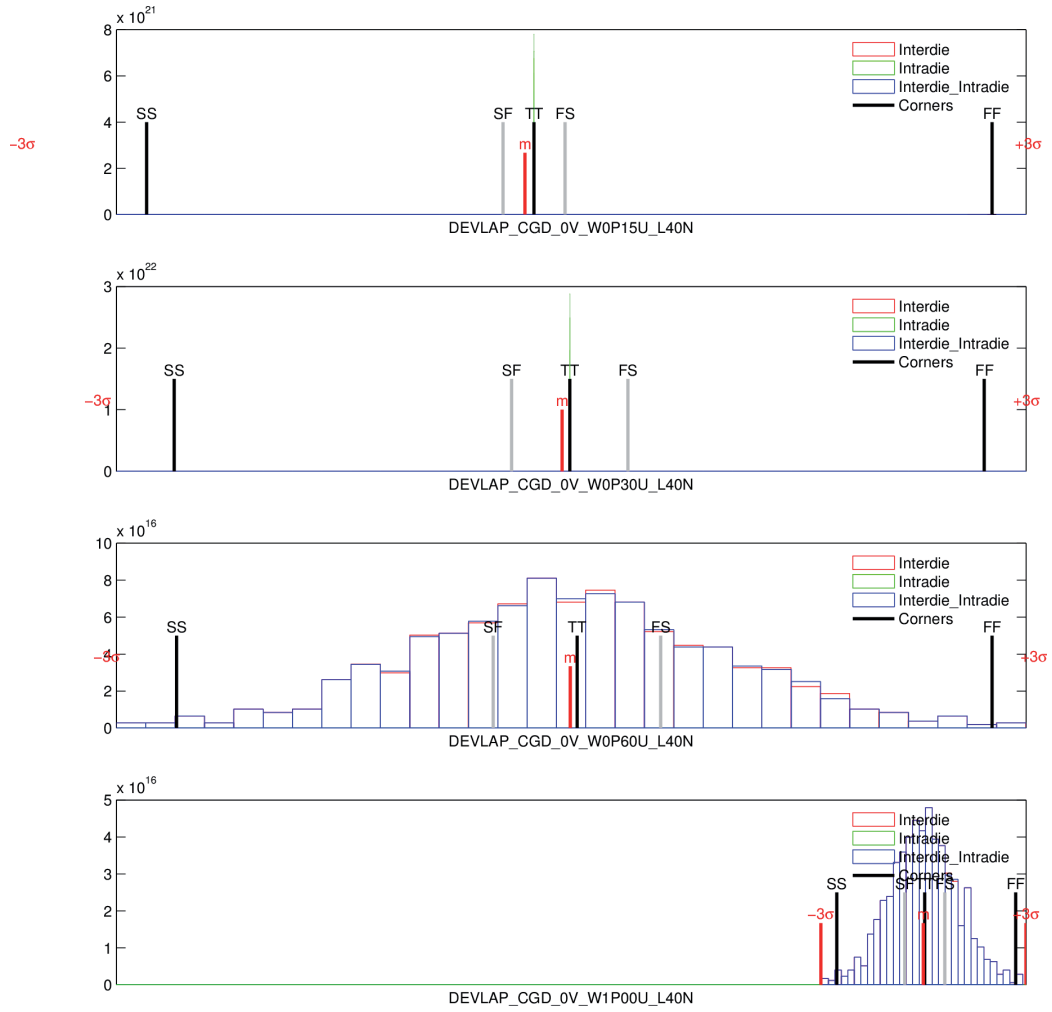


FIGURE 4.9 – Histogramme de C_{gd} résultant de la simulation Monte-Carlo avec le modèle modifié.

4.4 Conclusion

Dans ce chapitre, nous avons revu la méthode Monte-Carlo appliquée aux modèles compacts et décrit les possibilités d'effectuer des simulations distinctes des variabilités locales et globales. Nous avons ensuite revu la méthode de génération classique de construction des modèles pire-cas ainsi que deux méthodes de la littérature qui méritent d'être explorées.

Nous établissons ensuite un état des lieux du placement des modèles pire-cas par rapport aux modèles Monte-Carlo dans la technologie 45 nm faible puissance de STMicroelectronics. Au regard des résultats d'exploitation du plan d'expériences du chapitre précédent, nous proposons d'introduire dans ces modèles une variation de la longueur de recouvrement des source/drain sous la grille afin de représenter la variabilité des diffusions latérales des implants source/drain induite par le recuit rapide. La prise en compte de cette source de variabilité supplémentaire impacte de manière significative les variations de courant, ainsi que celles des capacités de recouvrement grille-diffusion, sans augmenter les variations du bilan capacitif vu de la grille. Cette première approche de l'impact de cet effet au niveau du transistor met en évidence la limitation de la méthode standard de construction des modèles pire-cas. Son impact au niveau circuit sera considéré dans la suite de cette étude, avec un intérêt tout particulier pour les applications circuits sensibles aux variations de l'effet Miller (lié au couplage entre grille et diffusion du transistor).

Conclusion générale

L'augmentation de la densité d'intégration des circuits intégrés nous a amenés à étudier, dans le cadre du développement de la technologie CMOS 45 nm, les sources de variabilité inhérentes aux procédés de fabrication utilisés pour ce nœud technologique, et à en déterminer les composantes principales, dans le but ultime de permettre la simulation précise de l'impact de la variabilité technologique à la fois au niveau transistor et circuit.

Afin de pouvoir reproduire les effets de variabilité en simulation, il a fallu dans un premier temps disposer d'observations expérimentales statistiquement représentatives, accessibles par des méthodes de caractérisation non destructives, et à même de fournir un ensemble cohérent de paramètres technologiques et électriques. Un effort important a donc été consacré à la compréhension des sources de variabilité, à la mise en œuvre de méthodes de caractérisation électrique et d'analyse statistique, préalablement à la mise en œuvre de la simulation de la variabilité jusqu'au niveau transistor et circuit.

Parmi les variabilités d'origine matériau, on peut mesurer la rugosité de ligne du polysilicium à partir de photographies de lignes de polysilicium. Similairement, une photographie d'une surface de polysilicium peut servir à caractériser la taille des grains qui participent à la variabilité du potentiel de surface dans le canal du transistor. Ces données ont été utilisées dans la littérature comme points d'entrée de simulations TCAD. Le placement aléatoire des dopants n'est pas mesurable. Il est reproduit par simulations atomistique de type Monte-Carlo pour les étapes d'implantation de dopants.

Nous avons étudié des indicateurs technologiques non accessibles à la mesure statistique que sont la température de recuit ainsi que les profils de dopage horizontaux et verticaux dans le canal et aux jonctions source/substrat et drain/substrat. Ces grandeurs sont de première importance dans le fonctionnement du transistor. Le dopage des source/drain est implanté dans le but d'obtenir des jonctions peu profondes. Pour activer les dopants en

gardant un profil de dopage élevé avec une pente forte, il faut appliquer un recuit à haute température mais pendant un temps très court. Ce sont des conditions où la répartition des impuretés est très sensible aux variations de paramètres process. Afin de se doter de moyens d'étudier la sensibilité des paramètres électriques, nous avons proposé des méthodes pour mesurer indirectement la température de recuit et le dopage du canal. La signature du four – sa contribution paramétrique à la température de recuit rapide reçue par chaque puce – a été évaluée *via* la résistance de surface du silicium. Une analyse de sensibilité a permis de déterminer une carte de température sur les plaques de silicium. Ces données ont permis de connaître la variabilité de cette grandeur.

Les diffusions des dopants des implants source/drain sont un problème à deux dimensions. En particulier dans les canaux courts, la redistribution de ces dopants va avoir un impact majeur sur la variabilité des performances du transistor. La caractérisation que nous proposons comme critère de calibrage est une mesure du courant de drain en régime bloqué en fonction de la longueur de grille. La tension de grille est autour de la tension de bandes plates pour s'affranchir de la conduction dans le canal. Une précision de l'ordre de 2 nm a été obtenue avec cet outil.

Pour estimer la variabilité du dopage du canal, nous avons utilisé la réponse de la tension de seuil en fonction de la tension de substrat. En effet, la zone de désertion qui varie en fonction de la polarisation du substrat suit le dopage vertical. Nous avons donc proposé de remonter au dopage à partir des tensions de seuil et de mesures de l'épaisseur d'oxyde de grille. La méthode proposée a permis de reproduire un profil connu dans une simulation TCAD et a ensuite été appliquée aux mesures d'un lot entier.

Ensuite nous exploitons les données mesurées sur un lot afin de faire un point sur les variations et les sensibilités au nœud 45 nm. Ce lot comprend un plan d'expériences sur cinq paramètres technologiques connus pour être parmi les plus influents sur les caractéristiques du transistor MOS. L'exploitation du plan d'expériences nous confirme les influences prépondérantes de la température de recuit rapide des source/drain et de la taille des espaceurs. L'ensemble des mesures de paramètres technologiques et électriques nous donne aussi les valeurs des déviations qui servent de point d'entrée pour les simulations.

En conséquence, nous avons introduit la prise en compte des variations des espaceurs dans les modèles compacts afin de refléter cette variabilité

des modèles Monte-Carlo. Les résultats attendus et estimés par simulation concernent en particulier une augmentation de la variabilité des capacités entre grille et source/drain.

Les perspectives de ce travail sont multiples ; en premier lieu, il est nécessaire de généraliser les approches de calibration TCAD pour prendre en compte la variabilité. La calibration des jeux de simulation TCAD doit se faire non seulement sur les caractéristiques électriques d'un point nominal, mais aussi sur leurs sensibilités par rapport aux paramètres technologiques dont l'influence est majeure. C'est déjà le cas pour la longueur de grille. Il est usuel d'utiliser la caractéristique de la tension de seuil en fonction de la longueur de grille pour calibrer. Il devrait en être de même pour l'épaisseur d'oxyde, la longueur des espaceurs, l'épaisseur de la grille par exemple. Des essais (non reportés ici) basés sur la comparaison des sensibilités mesurées avec des résultats de plan d'expériences simulés ont pu montrer qu'une simulation peut paraître calibrée en termes de valeur nominale alors que sa sensibilité à un facteur technologique est fautive. Calibrer cette sensibilité demande un effort de caractérisation supplémentaire mais permettra de supprimer des degrés de liberté (et donc d'aider) à la calibration.

En second lieu, il faut reconsidérer les approches de simulation électrique de la variabilité à chaque nœud technologique. Nous avons vu que précédemment que le choix de variables technologiques d'entrée des simulations Monte-Carlo SPICE n'était pas en phase avec les résultats de mesures. Dans la continuité du travail du dernier on pourra réviser la stratégie de simulation Monte-Carlo afin d'affiner la reproduction des variabilités observées en mesures.

Ensuite, la génération de simulation *pire-cas* pourra être développée avec comme objectif de permettre la création d'un jeu de *corner* spécifique à une application, car comme on l'a vu, par construction, un *corner* ne peut pas être représentatif de toutes les familles de performances.

De manière plus générale, la réduction des dimensions continuera d'augmenter l'impact de la variabilité statistique (locale). L'introduction de nouveaux matériaux tel que les high- κ / metal-gate (28nm) pourra ralentir la tendance en réduisant la sensibilité des dispositifs. L'évolution de l'architecture vers des dispositifs comme le FDSOI planaire ou le *triple gate* introduira de nouvelles sources de variations globales. Les efforts porteront sur la maîtrise des procédés et sur des architectures de dispositifs robustes.

Annexe A

Annexes

A.1 Figures du chapitre [4](#)

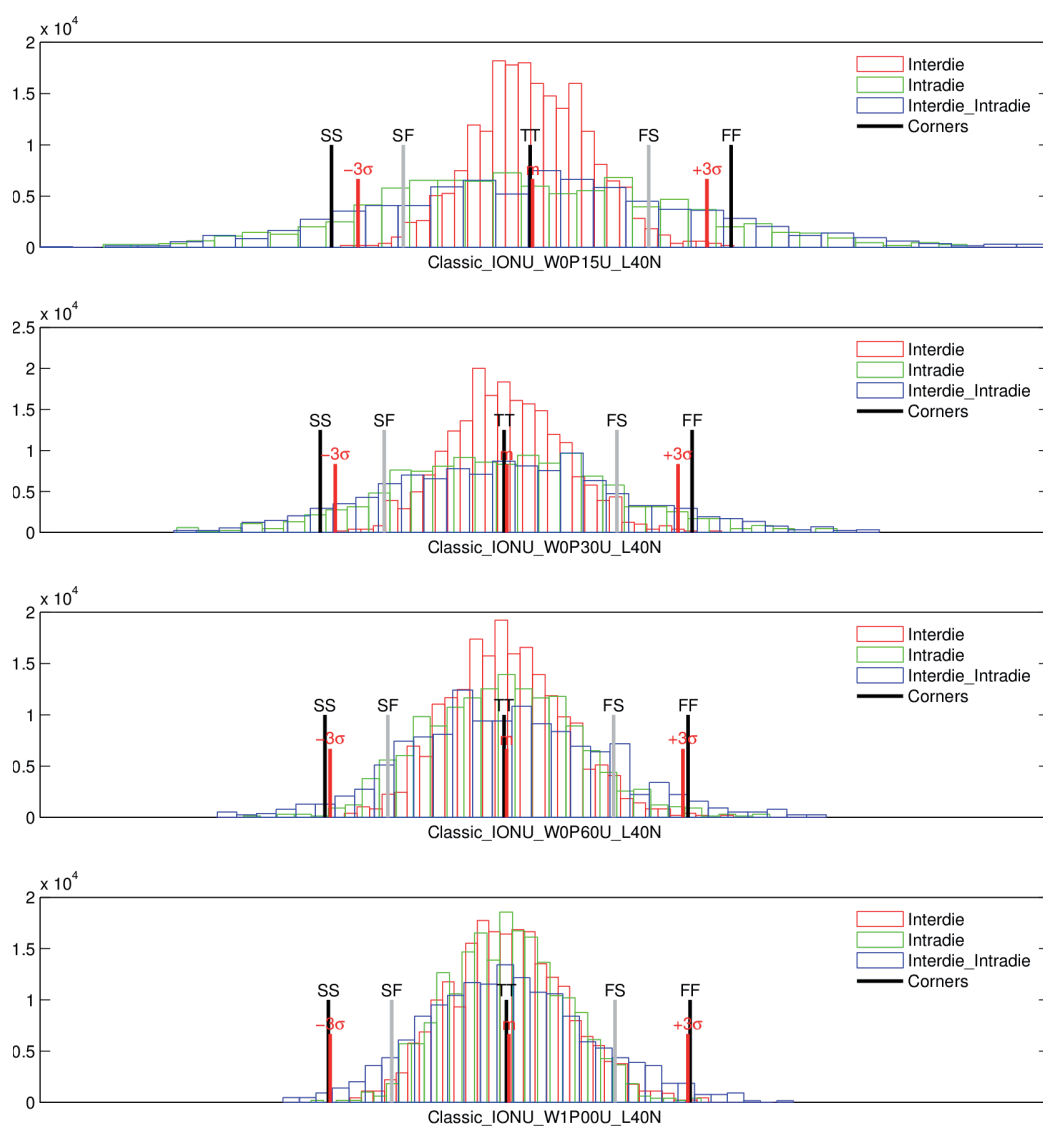


FIGURE A.1 – Histogramme de I_{on} résultant de la simulation Monte-Carlo avec le modèle actuel.

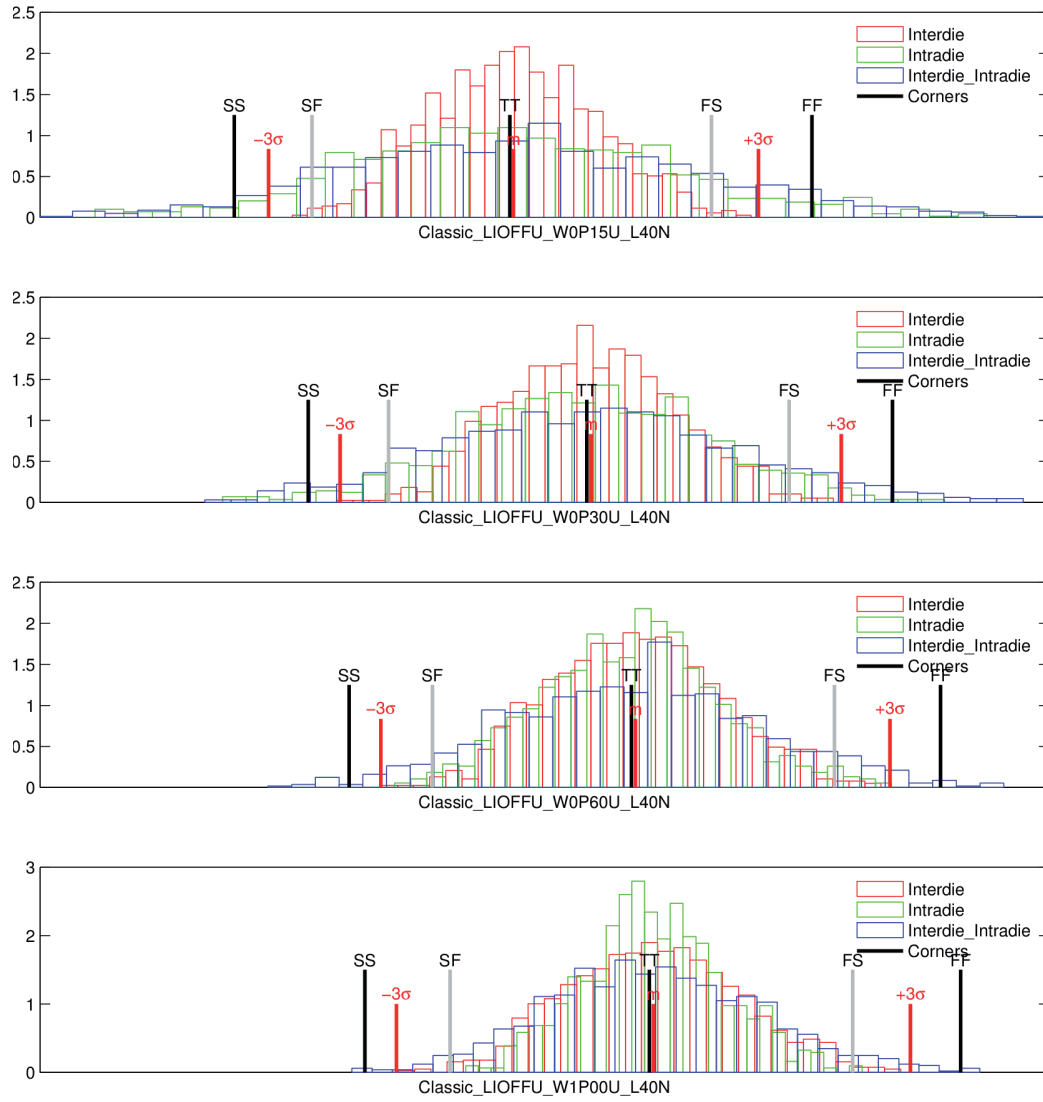


FIGURE A.2 – Histogramme de I_{off} résultant de la simulation Monte-Carlo avec le modèle actuel.

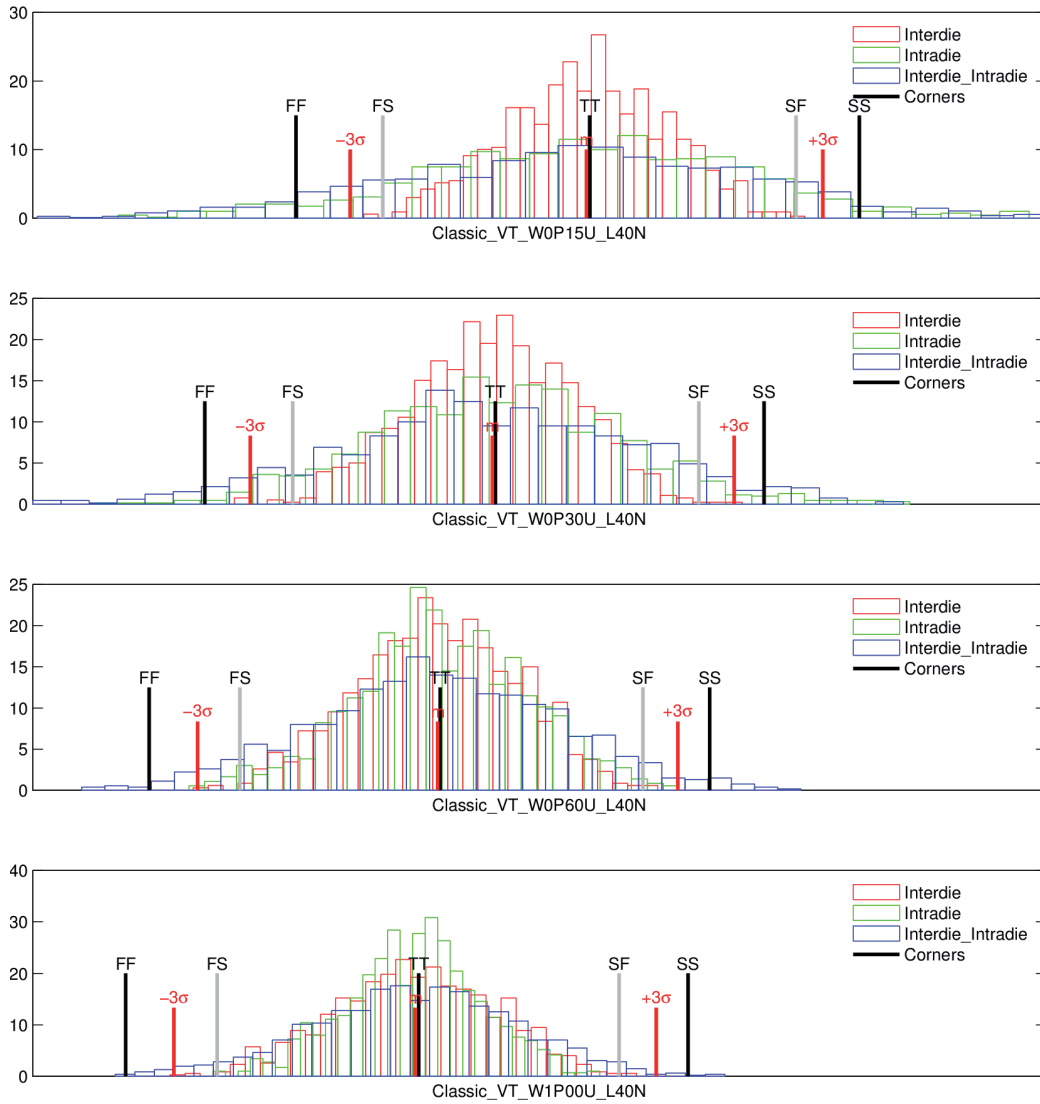


FIGURE A.3 – Histogramme de V_{Tsat} résultant de la simulation Monte-Carlo avec le modèle actuel.

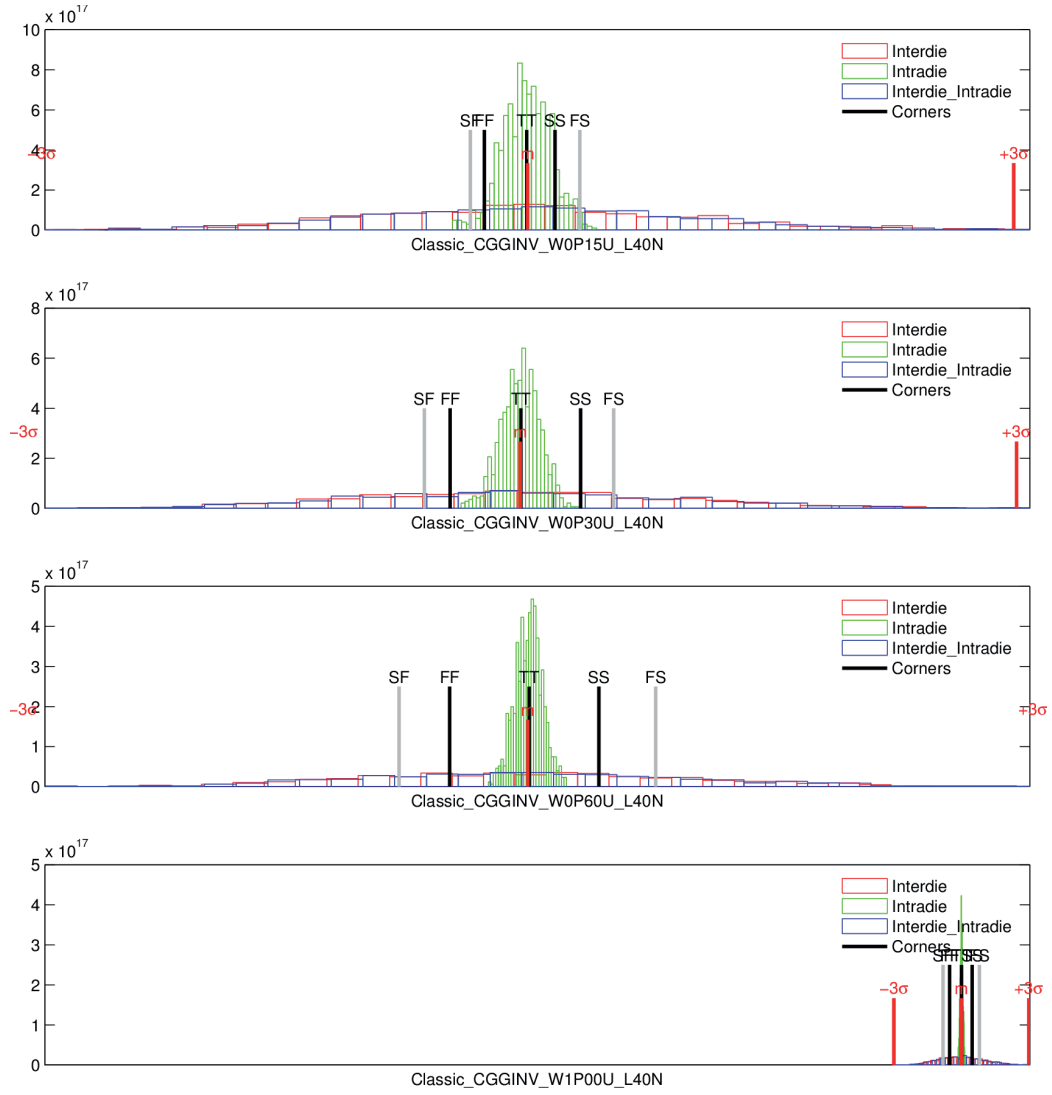


FIGURE A.4 – Histogramme de C_{gg} résultant de la simulation Monte-Carlo avec le modèle actuel.

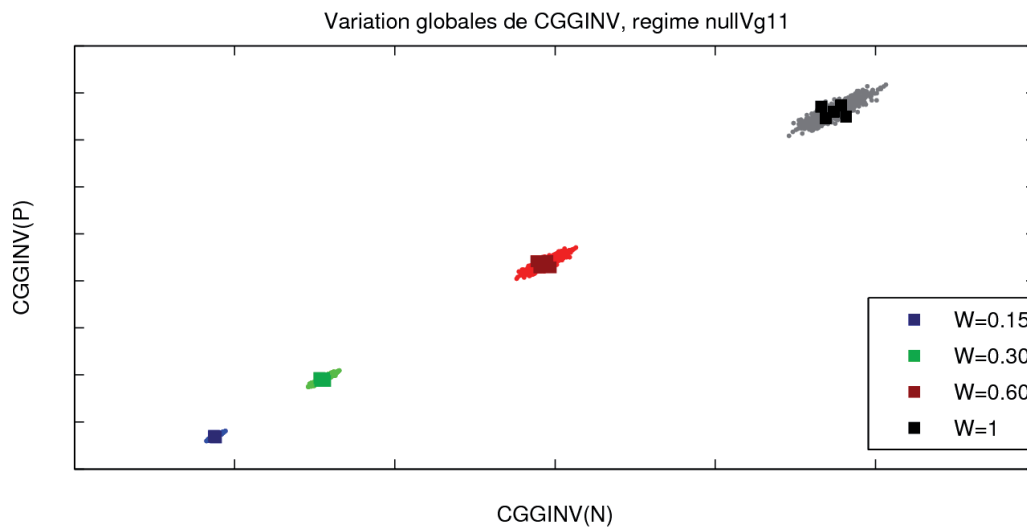


FIGURE A.5 – Corrélation des capacités C_{gg} du NMOS et du PMOS résultant de la simulation Monte-Carlo avec le modèle actuel.

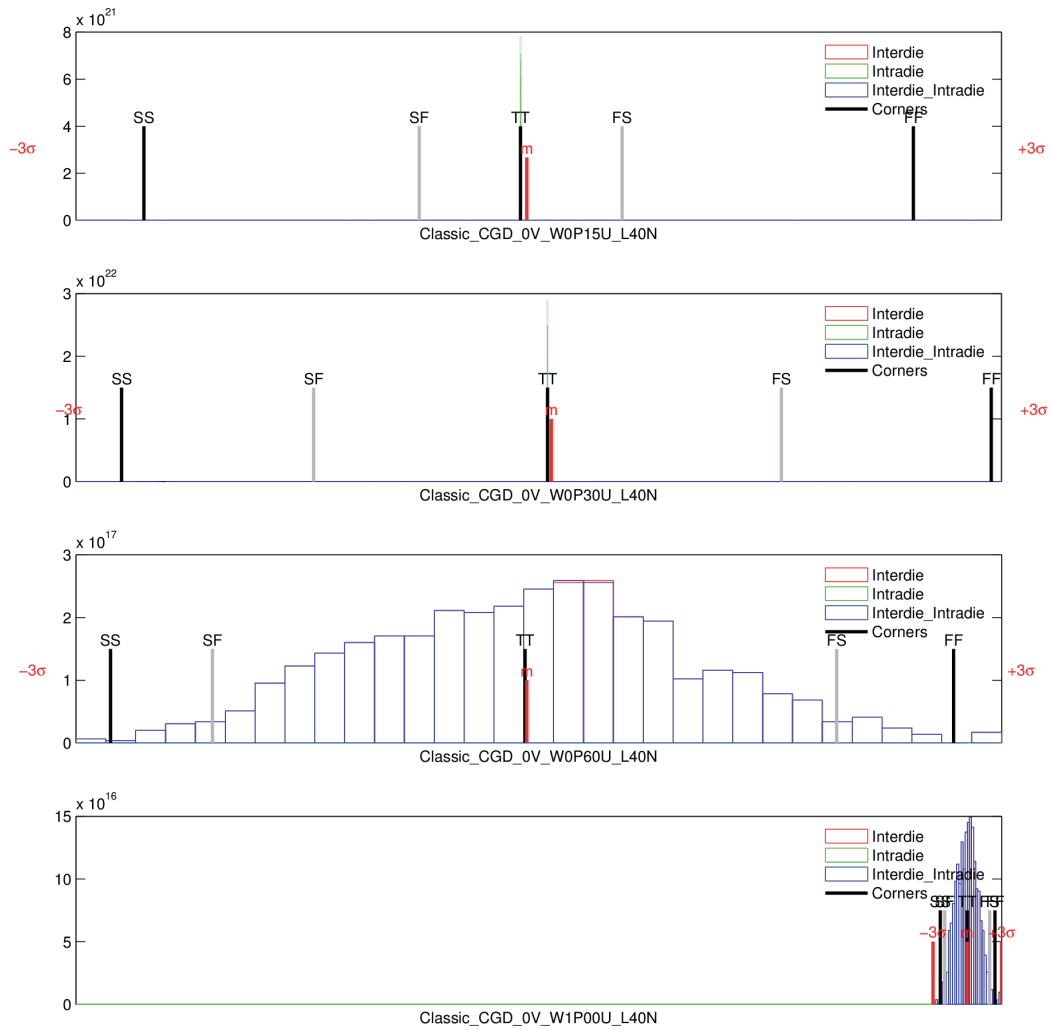


FIGURE A.6 – Histogramme de C_{gd} résultant de la simulation Monte-Carlo avec le modèle actuel.

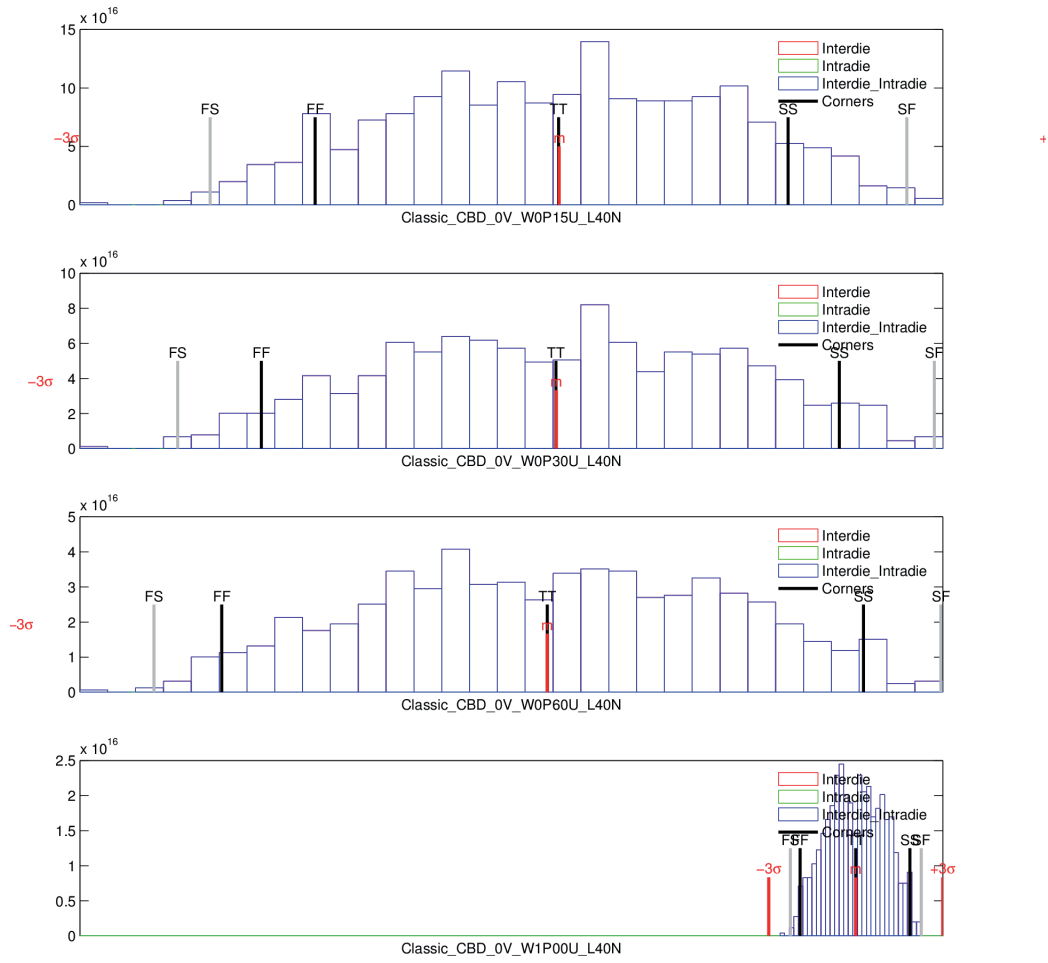


FIGURE A.7 – Histogramme de C_{bd} résultant de la simulation Monte-Carlo avec le modèle actuel.

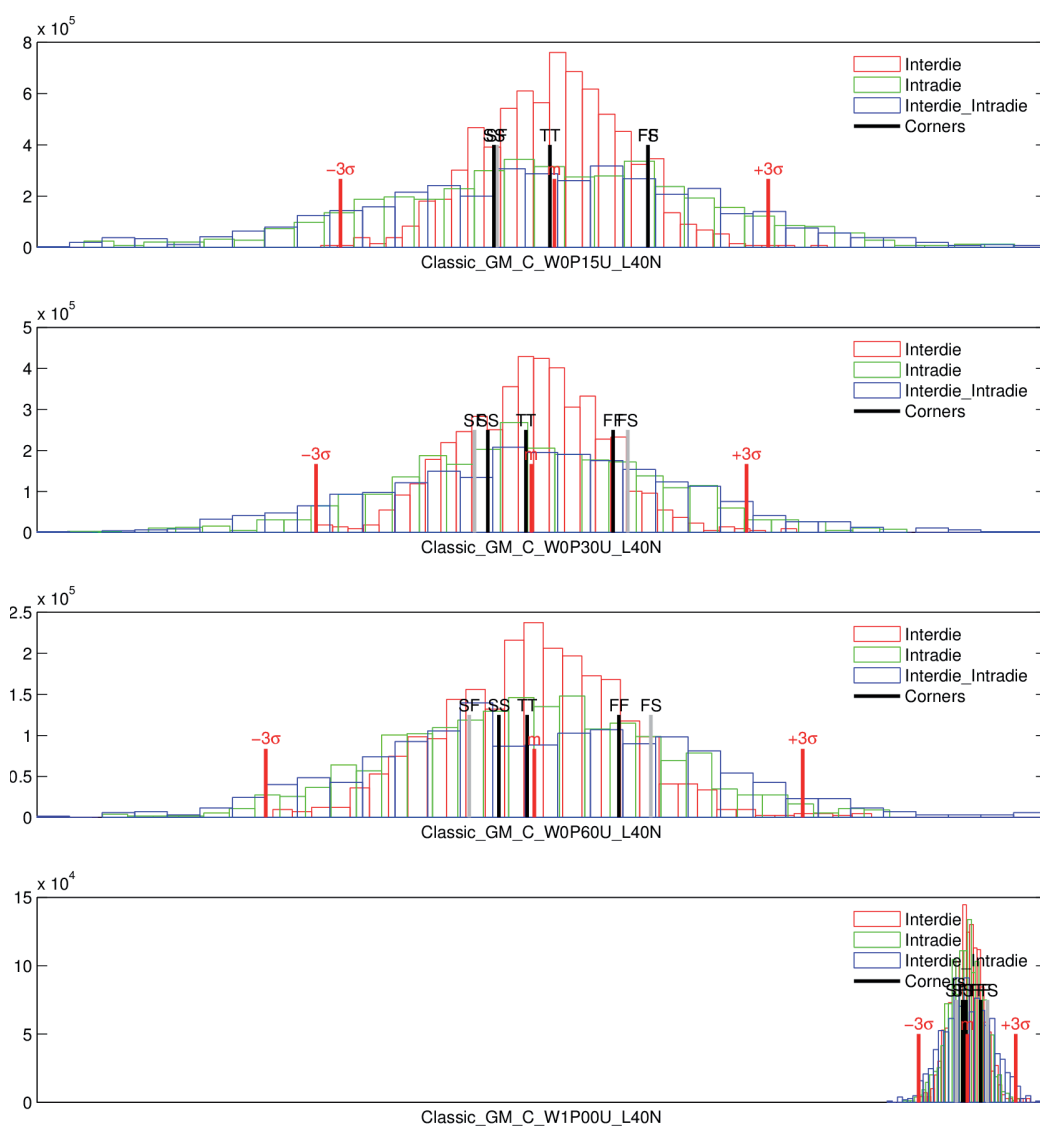


FIGURE A.8 – Histogramme de g_m résultant de la simulation Monte-Carlo avec le modèle actuel.

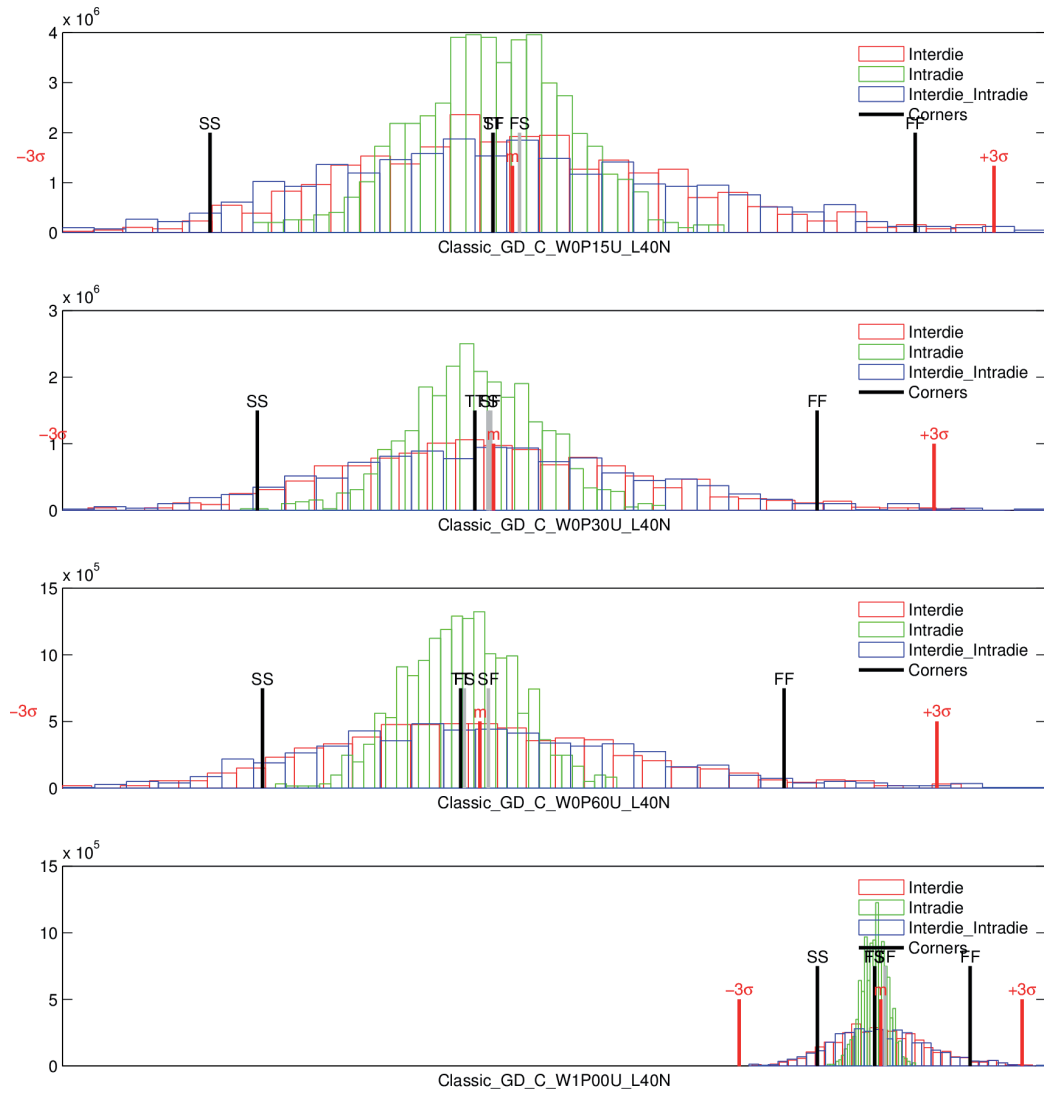


FIGURE A.9 – Histogramme de g_d résultant de la simulation Monte-Carlo avec le modèle actuel.

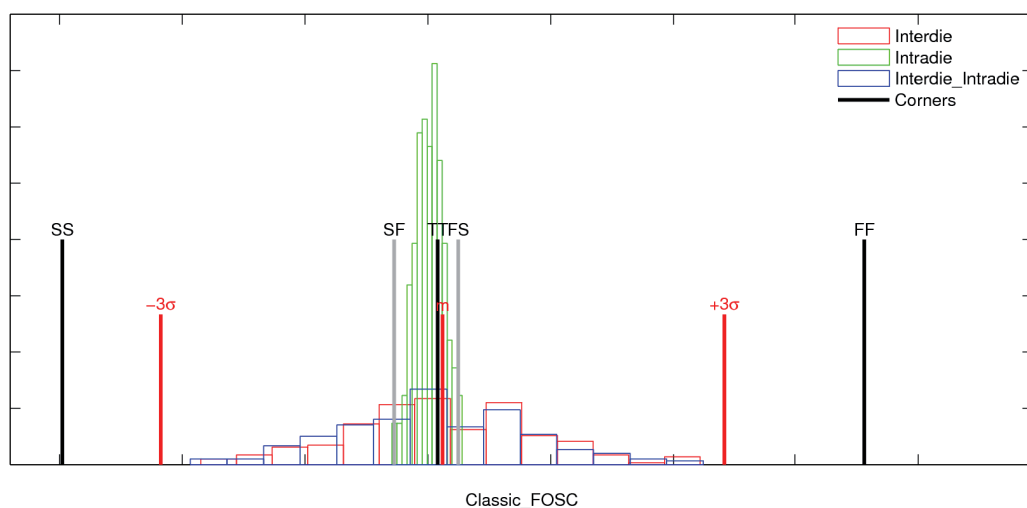


FIGURE A.10 – Histogramme de la fréquence d’oscillation des osculateurs en anneaux résultant de la simulation Monte-Carlo avec le modèle actuel.

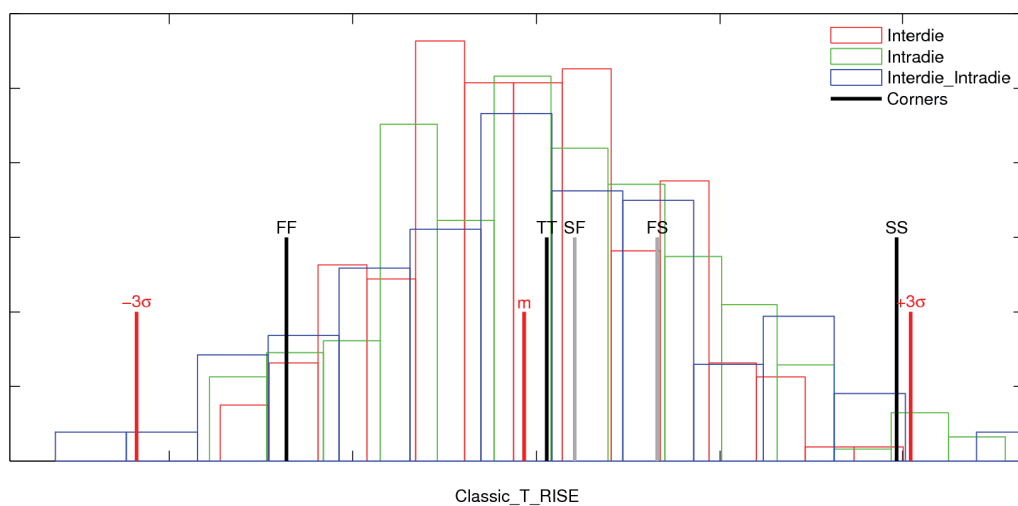


FIGURE A.11 – Histogramme de la fréquence de fonctionnement des oscilateurs en anneaux résultant de la simulation Monte-Carlo avec le modèle actuel.

A.2 Boite à moustaches

La boite à moustache, ou *box and whiskers plot* en anglais ou encore boxplot, est une représentation simplifiée d'une distribution statistique univariée. Elle a été inventée par John W. Tukey en 1977. Au lieu de représenter la densité de probabilité de la variable, la boite à moustaches représente la médiane et les quartiles. Les quartiles Q_1 , Q_2 et Q_3 sont les 3 valeurs pour lesquelles il y existe respectivement 25, 50 et 75 % d'observations dont les valeurs sont inférieures. La médiane est donc équivalente à Q_2 .

La figure [A.12](#) montre une boite à moustaches représentant une distribution gaussienne. La barre rouge est la médiane. Le rectangle, ou espace *inter-quartile* ($Q_3 - Q_1$) englobe 50 % de la population et les moustaches s'étendent au plus jusqu'à 1,5 fois l'espace inter-quartile (IQR), s'approchant de $\pm 2,7\sigma$ si l'on considère une répartition normale (*i.e.* gaussienne). L'IQR contient alors $99,6 - 0,4 = 99,2\%$ de la population.

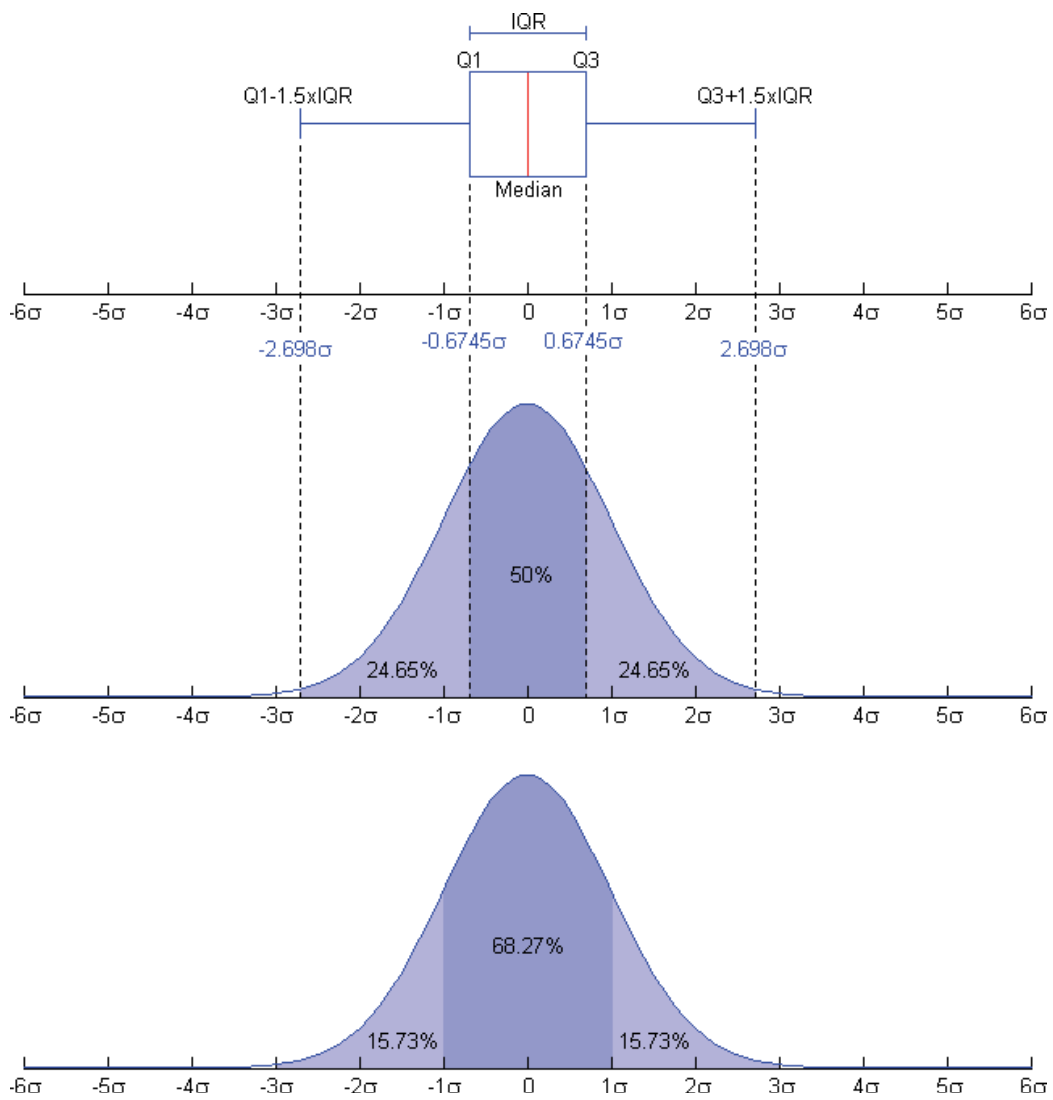


FIGURE A.12 – Description d’une boite à moustaches et comparaison avec une distribution gaussienne.

A.3 Extraits de code relatifs aux simulations Monte-Carlo

Dans les extraits de code A.1, nous suivons l'utilisation de la variable `cdvar`, que nous appelons ici `polylp_cdvar_cmos045` et qui définit la déviation de la longueur de grille par rapport à sa valeur nominale. Premièrement, cette variable peut être définie différemment selon que l'on veuille lancer une simulation Monte-Carlo avec ou sans variations globales (Code A.1, ligne 3). Ici on appelle la library `PRO_statcrolles` qui est prévue pour la simulation Monte-Carlo. Dans cette library (ligne 8), on définit que la déviation de la longueur de polysilicium suit une loi normale d'écart-type 1,2 nm et centrée en 0. Le mot-clef du simulateur LOT/ indique qu'il s'agit d'une simulation globale. Ensuite la valeur de ce paramètre est interfacé – *via* `nsvt1p_devdell_cmos045` – aux paramètres modèle `devdell`, `lvar` et `le` (lignes 12 à 16). Le paramètre interne à PSP `lvar0` est indirectement impacté.

Code source A.1 – Définition des variations des variables technologiques aux travers des différents fichiers de paramétrage de la simulation.

```

1 * eldo.lib
2 * on definit l'utilisation de la library PROstatcrolles
   pour le poly
3 .lib key=poly      ../models/common_poly.lib
   PRO_statcrolles
4
5 * common_poly_cd.lib
6 * .lib PRO_statcrolles
7 * cdvar suit une gaussienne centree en 0, d'ecart-type
   1.2nm
8 .param polylp_cdvar_cmos045 = 0.0 LOT/gauss='1.2e-09'
9
10 * LPmos_psp_svt.lib
11 *on propage cdvar aux parametres modele
12 .param nsvt1p_devdell_cmos045 = 'polylp_cdvar_cmos045 '
13 * dans le sous-circuit MOS nsvt1p
14 .param devdell = 'valif(((lm < 6e-8) &&
   (stylelayout>0.9)),
   (0.733*(nsvt1p_devdell_cmos045+2.0*devlap)),
```

```

    (nsvt1p_devdell_cmos045+2.0*devlap))'
15 .param lvar =
    '2.0E-09*(1.0+0*len/lm)*(1.0+0*wen/(wm/nfing)) +
    devdell'
16 .param le = 'lm+lvar-2*lap'* dans la carte modele
17 * dans la carte modele
18 +
    lvaro='2.0E-09+devdell/((1.0+0*len/lm)*(1.0+0*wen/wm))',

```

La figure A.13 et le code source A.2 décrivent cette implémentation. La différence de caractéristique entre ces deux devices est le mismatch. Dans le code A.2 les lignes 40 et 42 extraient les mesures de mismatch pour cette paire de transistors.

Code source A.2 – Implémentation Eldo d'une paire de transistors pour la simulation des courants et V_{th} en régime linéaire.

```

1 .param wn1=1.0    // dimensions du transistor
2 .param ln1=0.04
3 .param XVDD = 1.1 // parametre de l'analyse DC
4 * densite de courant pour l'extraction de VT
5 .param jdvt=10e-9*1/0.04
6
7 Vdd vd 0 dc 0.05 // tension de drain
8 R1  vd 0 1e6
9 vgn vg 0 dc XVDD // tension de grille parametree
10 R2  vg 0 1e6
11
12 edn1 vdn1 0 vd 0 1.0 // polarisation du premier device
13 egn1 vgn1 0 vg 0 1.0
14 vidn1 vdn1 vdn1i 0 // mesure du courant de drain
15
16 edn1b vdn1b 0 vd 0 1.0 // polarisation du second device
17 egn1b vgn1b 0 vg 0 1.0
18 vidn1b vdn1b vdn1bi 0
19
20 * transistors
21 * XMname D G S B modele W L SA SB
22 XMN1  vdn1i  vgn1  0 0 nsvt1p W=wn1 L=Ln1 SA=0.29
    SB=0.29

```

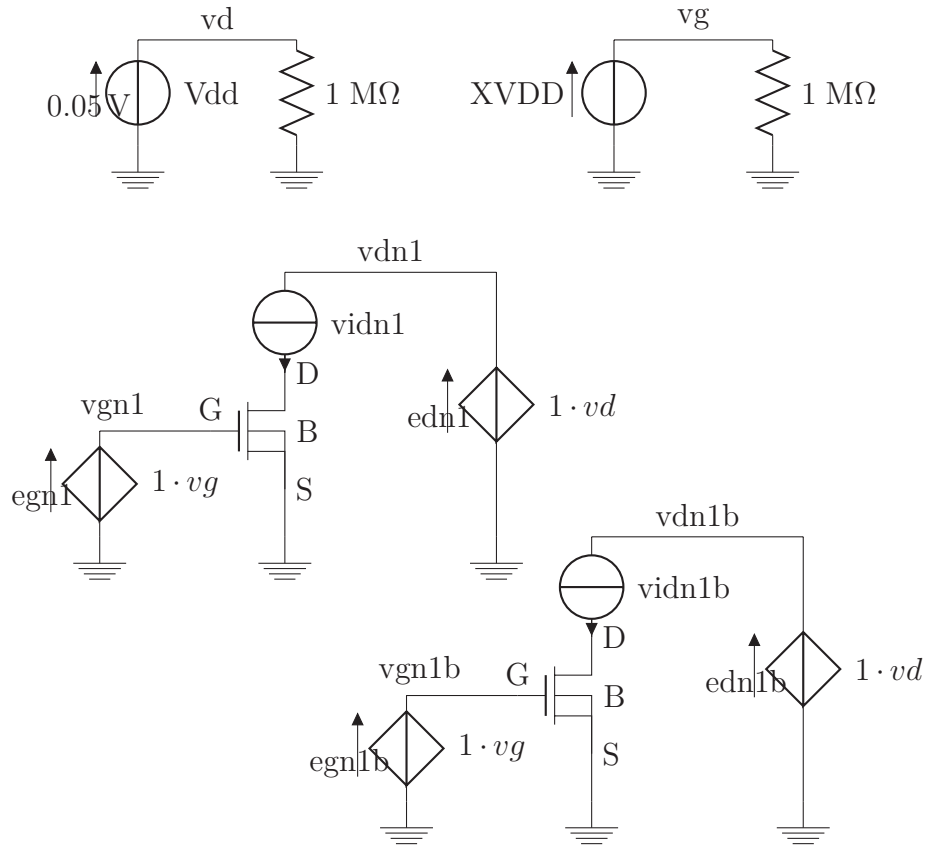


FIGURE A.13 – Paire de transistors pour la simulation des fluctuations locales. Les sources indépendantes Vdd et $XVDD$ contrôlent respectivement les sources $egn1(b)$ et $edn1(b)$. Les sources de tension nulle $vidn1(b)$ mesurent les courants de drains.

```

23 XMN1b vdn1bi vgn1b 0 0 nsvt1p W=wn1 L=Ln1 SA=0.29
    SB=0.29
24
25 * Fonction d'extraction des fluctuation locale de Vt
26 .defmac dVtcc(a,b)=((a-b)/1.414)
27 * Courant de drain (valeur absolue)
28 .defwave id_n1 'abs(i(vidn1))'
29 * Log du courant de drain normalise par W
30 .defwave lidu_n1 'LOG10(abs(i(vidn1))/wn1)'
31 * Courant de drain (valeur absolue, normalisee par
    micron de largeur)
32 .defwave idu_n1 '(abs(i(vidn1))/wn1)'

```

```
33 .defwave idu_n1b '(abs(i(vidn1b))/wn1)'  
34 * Courant de drain (Mismatch en %)  
35 .defwave did_n1  
    '(i(vidn1)-i(vidn1b))/(i(vidn1)+i(vidn1b))*2*100/1.414',  
36  
37 * Extraction des resultats dans le fichier de sortie  
38 .extract dc label=ionu_N_w1p00u_l40n Yval(W(idu_n1),1.1)  
39 .extract dc label=lioffu_P_w1p00u_l40n  
    Yval(W(lidu_p1),0.0)  
40 .extract dc label=dion_N_w1p00u_l40n Yval(W(did_n1),1.1)  
41 .extract dc label=Vt_N_w1p00u_l40n  
    xthres(w(idu_n1),(jdvt))  
42 .extract dc label=dVt_N_w1p00u_l40n  
    $dVtcc((xthres(w(idu_n1),(jdvt))),  
    (xthres(w(idu_n1b),(jdvt))))  
43  
44 * Parametres de l'analyse  
45 .dc vgn 0 XVDD 0.01 // rampe en tension sur Vg de 0 a  
    Xvdd par pas de 0.001  
46 .temp 25 // temperature  
47 .mc 1000 all // tir Monte-Carlo avec  
    enregistrement des donnees brutes  
48 .end
```

Bibliographie

- [1] A. G. Levine, “John bardeen, william shockley, walter brattain : Invention of the transistor – bell laboratories.” [Online]. Available : <http://www.aps.org/programs/outreach/history/historicsites/transistor> 1
- [2] D. Kahng, “Mos transistor.” [Online]. Available : http://www.invent.org/2009induction/1_3_09_induction_kahng.asp 1
- [3] M. M. John Atalla, “Mos transistor.” [Online]. Available : http://www.invent.org/2009induction/1_3_09_induction_atalla.asp 1
- [4] Wikipedia, “Mosfet.” [Online]. Available : <http://en.wikipedia.org/wiki/Transistor> 1
- [5] M. Bohr, “A 30 years retrospective on Dennard’s mosfet scaling paper,” Intel Corporation. [Online]. Available : <http://bit.ly/fL0qWx> 1
- [6] “International technology roadmap for semiconductor,” 2009. [Online]. Available : <http://www.itrs.net/Links/2009ITRS/Home2009.htm> 1
- [7] MEMC, “Process animations.” [Online]. Available : <http://www.memc.com/index.php?view=Process-Animations-> 3
- [8] “Intel, samsung, tsmc reach agreement for 450 mm wafer manufacturing transition,” May 6 2008. [Online]. Available : <http://www.physorg.com/news129301282.html> 1
- [9] P. Lemoigne, J.-D. Arnould, P.-E. Bailly, N. Corrao, P. Benech, M. Thomas, A. Farcy, and J. Torres, “Extraction of equivalent electrical models for damascene MIM capacitors in a standard 120 nm CMOS technology for ultra wide band applications,” in *IECON*, 2006. 1.3
- [10] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, and D. J. P. N. J. Rohrer, “High-performance CMOS variability in the 65-nm regime and beyond,” *IBM Journal of Research and Development*, vol. 50, no. 4/5, pp. 433–448, Jul./Sep. 2006. [Online]. Available : <http://www.research.ibm.com/journal/rd/504/bernstein.html> 2.3.3

- [11] R. Sitte, S. Dimitrijević, and H. B. Harrison, "Device parameter changes caused by manufacturing fluctuations of deep submicron mosfet's," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 2210–2215, Nov. 1994. [1.1](#)
- [12] C. Kenyon, A. Kornfeld, K. Kuhn, M. Liu, A. Maheshwari, W. Shih, S. Sivakumar, G. Taylor, P. VanDerVoorn, and K. Zawadzki, "managing process variation in intel's 45nm cmos technology.," *Intel Technology Journal*, Jun. 2008. [Online]. Available : <http://www.intel.com/technology/itj/2008/v12i2/3-managing/1-abstract.htm> [1.1](#)
- [13] K. Kuhn, "Reducing variation in advanced logic technologies : Approaches to process and design for manufacturability of nanoscale cmos," in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 2007, pp. 471–474. [1.1](#), [1.6](#), [A.3](#)
- [14] A. Brown, N. Idris, J. Watling, and A. Asenov, "Impact of metal gate granularity on threshold voltage variability : A full-scale three-dimensional statistical simulation study," *Electron Device Letters, IEEE*, vol. 31, no. 11, pp. 1199 –1201, nov. 2010. [1.1](#)
- [15] A. Asenov and S. Saini, "Suppression of random dopant-induced threshold voltage fluctuations in sub-0.1- μ m MOSFETs with epitaxial and δ -doped channels," *IEEE Trans. Electron Devices*, vol. 46, no. 8, p. 1718, Aug. 1999. [1.1](#), [1.2.2](#)
- [16] K. Takeuchi, R. Koh, and T. Mogami, "A study of the threshold voltage variation for ultra-small bulk and SOI CMOS," *IEEE Trans. Electron Devices*, vol. 48, no. 9, pp. 1995–2001, Sep. 2003. [1.1](#)
- [17] C. Fenouillet-Beranger, S. Denorme, B. Icard, F. Boeuf, J. Coignus, O. Faynot, L. Brevard, C. Buj, C. Soonekindt, J. Todeschini, J. Le-Denmat, N. Loubet, C. Gallon, P. Perreau, S. Manakli, B. Mmghetti, L. Pain, V. Arnal, A. Vandooren, D. Aime, L. Tosti, C. Savardi, F. Martin, T. Salvétat, S. Lhostis, C. Laviron, N. Auriac, T. Kormann, G. Chabanne, S. Gaillard, O. Belmont, E. Laffosse, D. Barge, A. Zaurner, A. Tarnowka, K. Romanjec, H. Brut, A. Lagha, S. Bonnetier, F. Joly, N. Mayet, A. Cathignol, D. Galpin, D. Pop, R. Delsol, R. Pantel, F. Pionnier, G. Thomas, D. Bensahel, S. Deleombus, T. Skotnicki, and H. Mmgam, "Fully-depleted soi technology using high-k and single-metal gate for 32 nmnode l1tp applications featuring 0.179 microm²6t-sram bitcell," in *Electron Devices Meeting*, 2007. [1.1](#)
- [18] S. Xiong and J. Bokor, "Sensitivity of double-gate and finfet devices to process variations," *IEEE Trans. Electron Devices*, vol. 50, no. 11, pp. 225–2261, Nov. 2003. [1.1](#)

- [19] F. Boeuf, M. Sellier, A. Farcy, and T. Skotnicki, "Impact of layout, interconnects and variability on CMOS technology roadmap," *Symposium on VLSI Technology*, pp. 24 – 25, Jun. 2007. 1.1
- [20] A. E. Carlson, "Device and circuit techniques for reducing variation in nanoscale sram," Ph.D. dissertation, EECS Department, University of California, Berkeley, May 2008. [Online]. Available : <http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-51.html> 1.1
- [21] A. Mutlu, C. Kwong, A. Mukherjee, and M. Rahman, "Statistical circuit performance variability minimization under manufacturing variations," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, 2006, pp. 4 pp.–. 1.1, 1.3.2
- [22] J. Singh, V. Nookala, Z.-Q. Luo, and S. Sapatnekar, "Robust gate sizing by geometric programming," in *DAC*, 2005. 1.1
- [23] R. R. Harrison, "Ece/cs 5720/6720 : Analog integrated circuit design," 2010. [Online]. Available : http://www.ece.utah.edu/~harrison/ece5720/Common_Centroid.pdf 1.4
- [24] H. Elzinga, "On the impact of spatial parametric variations on mos transistor mismatch," in *IEEE ICMTS*, 1996, pp. 173 – 177. 1.1
- [25] J. A. Croon, H. P. Tuinhout, R. Difrenza, J. Knol, A. J. Moonen, S. Decoutere, H. E. Maes, and W. Sansen, "A comparison of extraction techniques for threshold voltage mismatch," *IEEE ICMTS*, vol. 15, Apr. 2002. 1.2.1
- [26] R. Difrenza, K. Rochereau, T. Devoivre, B. Tavel, B. Duriez, D. Roy, S. Jullian, A. Denazzi, R. Boulestin, P. Stolk, and F. Arnaud, "Mosfet matching properties improvement in 65nm technology providing gain on both analog and sram performances," *IEEE International Conference on Microelectronic Test Structures*, vol. 18, Apr. 2005. 1.5, 1.2.4
- [27] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of mos transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, p. 1433, Oct. 1989. 1.2.1, 1.2.3, 1.3.2
- [28] K. Y. Tong, V. Rovner, L. Pileggi, and V. Kheterpal, "Design methodology of regular logic bricks for robust integrated circuits," in *Computer Design, 2006. ICCD 2006. International Conference on*, oct. 2006, pp. 162 –167. 1.2.1
- [29] J. Kuei and Y. Weng, "Yield impact from physical design at advanced technology nodes," in *ASIC, 2007. ASICON '07. 7th International Conference on*, oct. 2007, pp. 1104 –1109. 1.2.1

- [30] M. Smayling and V. Axelrad, "Simulation-based lithography optimization for logic circuits at 22nm and below," in *Simulation of Semiconductor Processes and Devices, 2009. SISPAD '09. International Conference on*, sept. 2009, pp. 1–4. 1.2.1
- [31] A. Strojwas, "Cost-effective variability reduction approaches to enable future technology nodes," in *Simulation of Semiconductor Processes and Devices (SISPAD), 2010 International Conference on*, sept. 2010, pp. 117–118. 1.2.1
- [32] K. Gettings and D. Boning, "Test circuit for study of cmos process variation by measurement of analog characteristics," in *Microelectronic Test Structures, 2007. ICMTS '07. IEEE International Conference on*, 2007, pp. 37–41. 1.2.1
- [33] Evans Analytical Group, "Depth profiling." [Online]. Available : http://www.eaglabs.com/training/tutorials/sims_theory_tutorial/profil.php 1.2.1
- [34] B. Boyer, "Technique d'analyse sims." [Online]. Available : <http://www.sciencefrance.com/techniques/SIMS/SIMS.html> 1.2.1
- [35] K. Agarwal, F. Liu, C. McDowell, S. Nassif, K. Nowka, M. Palmer, D. Acharyya, and J. Plusquellic, "A test structure for characterizing local device mismatches," in *VLSI Circuits, 2006. Digest of Technical Papers. 2006 Symposium on*, 0-0 2006, pp. 67–68. 1.2.1
- [36] J. S. Vickers, "In-line non-contact measurement of process induced indie parametric variability," in *SEMI Technology Symposium*, 2008. 1.2.1
- [37] L. Sponton, "From manufacturing variability to process-aware circuit simulation," Ph.D. dissertation, ETH Zurich, 2007. 1.9, A.3
- [38] T. Linton, M. Chandhok, B. J. Rice, and G. Schrom, "Determination of the line edge roughness specification for 34 nm devices," *International Electron Devices Meeting*, 2002. 1.10, 1.2.2, A.3
- [39] J. A. Croon, G. Storms, S. Winkelmeier, I. Pollenter, M. Ercken, S. Decoutere, W. Sansen, and H. Maes, "Line edge roughness : characterization, modeling and impact on device behavior," in *IEDM*, IEEE, Ed., 2002, pp. 307–310. 1.2.2
- [40] R. Difrenza, P. Llinares, S. Taupin, R. Palla, C. Garnier, and G. Ghi-
baudo, "Comparison between matching parameters and fluctuations at the wafer level," *IEEE International Conference on Microelectronic Test Structures*, vol. 15, Apr. 2002. 1.2.2
- [41] Z. Lu and C.-L. Yin, "Pattern transfer in device fabrication," 2004. 1.2.2

- [42] H.-L. Huang, B. S.-M. Lin, C.-C. Chen, and G. Liu, "Profile improvement method for patterning," 2007. 1.2.2
- [43] L. H. A. Leunissen, R. Jonckheere, K. Ronse, and G. B. Derksen, "Influence of gate patterning on line edge roughness," *Journal of Vacuum Science & Technology B*, vol. 21, p. 4, 2003. 1.2.2
- [44] E. Pargon, M. Martin, J. Thiault, O. Joubert, J. Foucher, and T. Lill, "Linewidth roughness transfer measured by critical dimension atomic force microscopy during plasma patterning of polysilicon gate transistors," *Journal of Vacuum Science & Technology B*, vol. 26, p. 10, 2008. 1.2.2
- [45] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer mosfets introduced by gate line edge roughness," *IEEE Trans. Electron Devices*, vol. 50, no. 5, p. 1254, May 2003. 1.2.2
- [46] R. Difrenza, J. C. Vildeuil, P. Llinares, and G. Ghibaudo, "Impact of grain number fluctuations in the mos transistor gate on matching performance," *IEEE International Conference on Microelectronic Test Structures*, 2003. 1.2.2
- [47] A. Asenov and S. Saini, "Polysilicon gate enhancement of the random dopant induced threshold voltage fluctuations in sub-100 nm mosfet's with ultrathin gate oxide," *IEEE Trans. Electron Devices*, vol. 47, no. 4, p. 805, Apr. 2000. 1.2.2
- [48] A. R. Brown, G. Roy, and A. Asenov, "Impact of fermi level pinning at polysilicon gate grain boundaries on nano-mosfet variability : A 3-d simulation study," *ESSDERC*, 2006. [Online]. Available : http://userweb.elec.gla.ac.uk/a/abrown/papers/ESSDERC_2006.pdf 1.2.2, 1.11, 1.12, A.3
- [49] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 m mosfet's : A 3-d "atomistic" simulation study," *IEEE Trans. Electron Devices*, vol. 45, no. 12, p. 2505, Dec. 1998. 1.14, 1.2.2
- [50] W. J. Gross, D. Vasileska, and D. K. Ferry, "Ultrasmall mosfets : The importance of the full coulomb interaction on device characteristics," *IEEE TRANSACTIONS ON ELECTRON DEVICES*, vol. 47, no. 10, p. 1831, Oct. 2000. 1.2.2
- [51] N. Sano, K. Matsuzawa, M. Mukai, and N. Nakayama, "Role of long-range and short-range coulomb potentials in threshold characteristics under discrete dopants in sub-0.1 pm si-mosfets," *IEDM*, p. 275, 2000. 1.2.2

- [52] T. Ezaki, T. Ikezawa, and M. Hane, "Investigation of realistic dopant fluctuation induced device characteristics variation for sub-100nm cmos by using atomistic 3d process/device simulator," *International Electron Devices Meeting*, 2002. 1.2.2
- [53] P. Dollfus, A. Bournel, S. Galdin, S. Barraud, and P. Hesto, "Effect of discrete impurities on electron transport in ultrashort mosfet using 3-d mc simulation," *IEEE TRANSACTIONS ON ELECTRON DEVICES*, vol. 51, no. 5, p. 749, May 2004. 1.2.2
- [54] A. Asenov, R. Balasubramaniam, A. R. Brown, J. H. Davies, and S. Saini, "Random telegraph signal amplitudes in sub 100 nm (decanano) mosfets : A 3d atomistic simulation study," *International Electron Devices Meeting*, p. 279, 2000. 1.2.2
- [55] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, "Increase in the random dopant induced threshold fluctuations and lowering in sub-100 nm mosfets due to quantum effects : A 3-d density-gradient simulation study," *IEEE Trans. Electron Devices*, vol. 48, no. 4, p. 722, Apr. 2001. 1.2.2
- [56] —, "Quantum mechanical enhancement of the random dopant induced threshold voltage fluctuations and lowering in sub 0.1 micron mosfets," *International Electron Devices Meeting*, 1999. 1.2.2
- [57] B. Cheng, S. Roy, G. Roy, and A. Asenov, "Integrating 'atomistic', intrinsic parameter fluctuations into compact model circuit analysis," in *33rd Conference on European Solid-State Device Research (ESSDERC'03)*, Sep. 2003, pp. 437–440. 1.2.2
- [58] D. Z.-Y. Ting, E. S. Daniel, and T. McGill, "Interface roughness effects in ultra-thin tunneling oxides," in *VLSI Design*, 1998, pp. 47–51. 1.2.2
- [59] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic threshold voltage fluctuations in decanano mosfets due to local oxide thickness variations," *IEEE Trans. Electron Devices*, vol. 49, no. 1, p. 112, Jan. 2002. 1.15, 1.2.2, A.3
- [60] A. Asenov, "Statistical device variability and its impact on design," 2008. 1.16, A.3
- [61] H. Tuinhout, "Impact of parametric fluctuations on performance and yield of deep-submicron technologies," in *IEEE ESSDERC*, 2002.
- [62] S.-L. Chen, "Mobility dependence on crystal orientation and strain engineering," june, 11 2007. [Online]. Available : <http://large.stanford.edu/courses/2007/ap273/chensl1/> 1.2.3, 1.1

- [63] C. A. Mack, “Field guide to optical lithography.” [Online]. Available : http://spie.org/x648.html?product_id=665802&origin_id=x32403 1.2.3
- [64] R. C. Hodgkin, “Ucla finds layout design important in transistor performance.” [Online]. Available : <http://www.tomshardware.com/news/ucla-finds-layout-design-important-transistor-performance,4585.html> 1.17
- [65] Synopsys, “Alt-psm technology for tighter control of chip performance and increased yield.” [Online]. Available : <http://www.synopsys.com/Tools/Manufacturing/MaskSynthesis/Pages/PSMCreate.aspx> 1.18
- [66] F. Cacho, “Mtm09_017 : Pattern effet during spike anneal process,” STMicroelectronics, Tech. Rep., 2008. 1.2.3, 1.20, A.3
- [67] —, “Mtm09_018 : Pattern effects during dsa process,” STMicroelectronics, Tech. Rep., 2008. 1.2.3, 1.19, A.3
- [68] M. Bidaud, H. Bono, C. Chaton, B. Dumont, V. Huard, P. Morin, L. Proencamota, R. Ranica, and G. Ribes, “High-activation laser anneal process for the 45nm cmos technology platform,” in *Advanced Thermal Processing of Semiconductors*, 2007. 1.2.3
- [69] P. Drennan, M. Kniffin, and D. Locascio, “Implications of proximity effects for analog design,” in *Custom Integrated Circuits Conference*, 2006. 1.2.3
- [70] Y.-M. Sheu, K.-W. Su, S.-J. Yang, H.-T. Chen, C.-C. Wang, M.-J. Chen, and S. Liu, “Modeling well edge proximity effect on highly-scaled mosfets,” in *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005*, sept. 2005, pp. 831 –834. 1.2.3, 1.21, A.3
- [71] J. Watts, K.-W. Su, and M. Basel, “Netlisting and modeling well-proximity effects,” *Electron Devices, IEEE Transactions on*, vol. 53, no. 9, pp. 2179 –2186, sept. 2006. 1.2.3
- [72] N. Wils, H. Tuinhout, and M. Meijer, “Characterization of sti edge effects on cmos variability,” *Semiconductor Manufacturing, IEEE Transactions on*, vol. 22, no. 1, pp. 59 –65, feb. 2009. 1.2.3
- [73] H. Tsuno, K. Anzai, M. Matsumura, S. Minami, A. Honjo, H. Koike, Y. Hiura, A. Takeo, W. Fu, Y. Fukuzaki, M. Kanno, H. Ansai, and N. Nagashima, “Advanced analysis and modeling of mosfet characteristic fluctuation caused by layout variation,” in *VLSI Technology, 2007 IEEE Symposium on*, 2007, pp. 204–205. 1.2.3, 1.22, 1.23, A.3
- [74] P. A. Stolk, H.-J. Gossmann, D. J. Eaglesham, D. C. Jacobson, C. S. Rafferty, G. H. Gilmer, M. Jaraíz, J. M. Poate, H. S.

- Luftman, and T. E. Haynes, "Physical mechanisms of transient enhanced dopant diffusion in ion-implanted silicon," *Journal of Applied Physics*, vol. 81, no. 9, pp. 6031–6050, 1997. [Online]. Available : <http://link.aip.org/link/?JAP/81/6031/1> 1.3
- [75] A. T. Fiory, S. G. Chawda, S. Madishetty, V. R. Mehta, N. M. Ravindra, S. P. McCoy, M. E. L. ois, K. K. Bourdelle, J. McKinley, H.-J. L. Gossmann, and A. Agarwal, "Transient enhanced diffusion in shallow junction formation." [Online]. Available : web.njit.edu/~fiory/Papers/TransientDiffusionJunctions.pdf 1.3
- [76] L. Lee. Methods for forming silicon comprising films using hexachlorodisilane in a single-wafer deposition chamber. [Online]. Available : <http://osdir.com/patents/Coating-processes/Methods-forming-silicon-comprising-films-hexachlorodisilane-single-wafer-deposition-chamber.html> 1.2
- [77] Comsol, "Boat reactor for low pressure chemical vapor deposition." [Online]. Available : <http://www.comsol.com/showroom/gallery/249/1.24,A.3>
- [78] S. Tugrul, "Improving wafer temperature uniformity for etch applications," Ph.D. dissertation, Texas tech university, 2003. 1.25, A.3
- [79] J. A. Croon, E. Augendre, S. Decoutere, W. Sansen, and H. Maes, "Influence of doping profile and halo implantation on the threshold voltage mismatch of a 0.13 μm cmos technology," in *ESSDERC*, 2002, pp. 579–582. 1.2.4
- [80] E.-M. Bazizi, "Modélisation physique et simulation de défauts étendus et diffusion des dopants dans le si, soi et sige pour les mos avancés," Ph.D. dissertation, Université de Toulouse, 2010. [Online]. Available : <http://thesesups.ups-tlse.fr/905/> 1.26, 1.2.4, A.3
- [81] P. Castrillo, "Atomistic process simulation for microelectronics , universidad de valladolid, departamento de electricidad y electrónica." [Online]. Available : <http://www.ele.uva.es/~simulacion/> 1.3.1
- [82] "Sentaurus process simulation of a 30-nm gate-length nmos transistor with hybrid continuum and atomistic diffusions," Synopsys Inc., Tech. Rep., 2007. [Online]. Available : http://www.synopsys.com/tools/tcad/capsulemodule/sentapro_30nm_hybrid.pdf 1.3.1
- [83] M. Jaraiz, P. Castrillo, R. Pinacho, and J. E. Rubio, "Current capabilities and future prospects of atomistic process simulation," in *IEEE International Electron Device Meeting*, 2007, pp. 951–954. 1.3.1

- [84] M. Hane, T. Ikezawa, and T. Ezaki, "Atomistic 3d process/device simulation considering gate line-edge roughness and poly-si random crystal orientation effects," *International Electron Devices Meeting*, 2003. 1.3.1
- [85] L. Sponton, L. Bomholt, D. Pramanik, and W. Fichtner, "A full 3d tcad simulation study of line-width roughness effects in 65 nm technology," in *International Conference on Simulation of Semiconductor Processes and Devices*, Oct. 2006, pp. 1306 – 1308. 1.3.1
- [86] S.-D. Kim, H. Wada, and J. Woo, "Tcad-based statistical analysis and modeling of gate line-edge roughness effect on nanoscale mos transistor performance and scaling," *Semiconductor Manufacturing, IEEE Transactions on*, vol. 17, no. 2, pp. 192–200, 2004. 1.3.1
- [87] D. Pramanik, L. Bomholt, and W. Fichtner, "What designers need to know about tcad," Apr.4 2005. [Online]. Available : <http://www.eetimes.com/showArticle.jhtml?articleID=160401660> 1.3.2
- [88] H. Sato, H. Kunitomo, K. Tsuneno, K. Mori, and H. Masuda, "Accurate statistical process variations analysis for 025um cmos with advanced tcad methodology," *IEEE Trans. Semicond. Manuf.*, 1998. 1.3.2
- [89] P. Alotto, M. Gaggero, G. Molinari, and M. Nervi, "A “design of experiment” and statistical approach to enhance the “generalised response surface” method in the optimisation of multim minima problems," *Magnetics, IEEE Transactions on*, vol. 33, no. 2, pp. 1896–1899, 1997. 1.3.2
- [90] L. Sponton, F. Carbognani, D. Pramanik, and W. Fichtner, "Tcad analysis for vlsi-application-oriented process optimization," in *8th International Conference on Solid-State and Integrated Circuit Technology*, Oct. 2006, pp. 1306 – 1308. 1.3.2
- [91] T. Ma, D. Pramanik, and R. Borges, "Bringing manufacturing into design through TCAD," Dec. 2005. 1.3.2
- [92] L. Peters, "DFM takes on process variability," Jan.11 2006. [Online]. Available : <http://www.semiconductor.net/article/CA6385511.html> 1.3.2
- [93] S. Tiramala, Y. Mahotin, X. Lin, V. Moroz, L. Smith, S. Krishnamurthy, L. Bomholt, and D. Pramanik, "Bringing manufacturing into design via process-dependent spice models," in *7th International Symposium on Quality Electronic Design*, vol. 1, Mar. 2006. 1.3.2
- [94] C. D'Agostino, P. Flatresse, E. Beigne, and M. Belleville, "Statistical leakage modeling in cmos logic gates considering process variations," in *Integrated Circuit Design and Technology and Tutorial, 2008. ICICDT 2008. IEEE International Conference on*, june 2008, pp. 301 –304. 1.3.3

- [95] C. D'Agostino, J. Le Coz, P. Flatresse, E. Beigne, and M. Belleville, "An accurate approach for statistical estimation of leakage current considering multi-parameter process variations in nanometer cmos technologies," in *Solid State Device Research Conference, 2009. ESSDERC '09. Proceedings of the European*, sept. 2009, pp. 427–430. [1.3.3](#)
- [96] N. Duhayon, *J. Vac. Sci. Technol.*, vol. B 22, p. 385, 2004. [2.3](#)
- [97] P. Eyben, *Materials Science and Engeneering B*, pp. 124–125, 2005. [2.3](#)
- [98] Fraunhofer IISB, "D501 - report on lateral diffusion characterisation methodology," Atomics IST project no.027152, Tech. Rep., feb 2007. [2.3.1](#)
- [99] *TCAD Sentaurus, Release A-2007.12, Synopsys Inc., 2007.* [2.3.1](#)
- [100] N. Khalil, *IEEE Electron Dev. Lett.*, vol. 16, p. 17, 1995. [2.3.1](#)
- [101] I. J. Dijomehri and D. Antoniadis, *IEEE Trans. Electron Dev.*, vol. 46, p. 568, 2002. [2.3.1](#)
- [102] J. Shermer, in *ESSDERC*, 2007. [2.3.1](#)
- [103] A. Martinez-Limia, in *GADEST*, 2007. [2.3.1](#)
- [104] *Advanced Calibration User Guide, version Z-2007.03, Mountain View, California : Synopsys Inc. 2003.* [2.3.2](#)
- [105] STMicroelectronics, "D603 - report on test applications," ATOMICS, IST project no. 027152, Tech. Rep., 2006. [2.3.2](#)
- [106] Crolles Variability group, "internal report," STMicroelectronics, Tech. Rep., 2007. [2.3.3](#)
- [107] G. Sado and M.-C. Sado, *Les plans d'expériences. De l'expérimentation à l'assurance qualité.* AFNOR, 2000. [2.3.3](#), [3.3.2](#)
- [108] Gammie, Wang, Chau, Gururajarao, Pitts, Jumel, Engl, Royannez, Flagerquist, Mair, Vaccani, Baldwin, Heragu, Mandal, Clinton, Arden, and Co, "A 45 nm 3.5g baseband-and-multimedia application processor using adaptative body bias and ultra-low-power techniques," in *International Conference on Solid State Circuits session 13.2, ISCC 2008. IEEE International Conference*, 2008, pp. 258–259. [2.4.2](#)
- [109] G. T. Wright, "Physical and CAD models for the implanted-channel VLSI MOSFET," *Transaction on Electron Devices*, 1987. [2.5.3](#)
- [110] V. Quenette, D. Rideau, R. Clerc, S. Retailliau, C. Tavernier, and H. Jaouen, "Dynamic charge sharing modeling for surface potential based models," in *Workshop on Compact Modeling*, 2009. [2.6.3](#)

- [111] K. Maitra and N. Bhat, "Impact of gate-to-source/drain overlap length on 80-nm cmos circuit performance," *IEEE Trans. Electron Devices*, vol. 51, no. 3, p. 409, Mar. 2004. 3
- [112] J. Goupy and L. Creighton, *Introduction aux plans d'expériences*, DUNOD, Ed. L'usine nouvelle, 2006. [Online]. Available : <http://www.plansdexpériences.com/plansdexpériences/> 3.3.2
- [113] J. Lei, P. Lima-Filho, and M. Styblinski, "Quadratic propagation of variance (qpov) in statistical circuits design," in *Statistical Metrology, 1998. 3rd International Workshop on*, 1998, pp. 88–91. 3.3.3
- [114] A. Graupner, W. Schwarz, and R. Schüffny, "Statistical analysis of analog structures through variance calculation," *IEEE Transactions on Circuits and Systems I : Fundamental theory and applications*, vol. 49, no. 8, Aug. 2002. 3.3.3
- [115] C.-H. Lin, A. M. N. Mohan V. Dunga, Darsen Lu, and C. Hu, "Statistical compact modeling of variations in nano mosfets," 2008. 4.2.3
- [116] G. Rappitsch, E. Seebacher, M. Kocher, , and E. Stadlober, "Spice modeling of process variation using location depth corner models," *TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING*, vol. 17, no. 2, pp. 201–213, May 2004. 4.2.4

Table des figures

1.1	Réplique du premier transistor réalisé aux Bell Labs.	1
1.2	Un lingot de silicium. À l'extrémité supérieure se trouvait la graine, ou cristal initial à partir duquel a crû le cylindre entier.	2
1.3	Empilement des niveaux dans une technologie CMOS 0,25 μm	3
1.4	Placement de 2 transistors en <i>common centroid</i>	6
1.5	Une paire de transistors identiques et placés au plus proches.	8
1.6	Graphe de Pelgrom du V_{th} de différentes technologies mesurées ou simulées [13].	10
1.7	Répartition spatiale de l'épaisseur d'oxyde <i>liner</i> mesurée sur la plaque n° 5 d'un lot de 12 plaques. Les variations suivent un motif radial et la plage de variation est de 0,3 %.	11
1.8	Population du paramètre HCD_{th} par plaque	12
1.9	Image SEM de traits de polysilicium. On voit nettement l'imperfection du bord de grille [37].	14
1.10	Exemple de faible (a) et forte (b) rugosité de ligne [38].	14
1.11	Potentiel de surface sous l'influence d'un seul joint de grain [48].	16
1.12	Échantillon d'une simulation Monte-Carlo 3D atomistique prenant en compte l'influence des joints de grains sur le potentiel de surface [48].	17
1.13	Simulation de 2 transistors qui diffèrent uniquement par le positionnement local des dopants dans le canal.	18
1.14	Tensions de seuils de 2500 transistors	19
1.15	Simulation prenant en compte les irrégularités de la surface SiO_2/Si [59]. On peut observer séparément de haut en bas l'irrégularité de la surface oxyde/silicium, une isosurface de niveau de dopage de 10^{17} cm^{-3} et la distribution de potentiel à la surface du canal.	20
1.16	Image SEM de l'oxyde de grille illustrant les imperfections d'épaisseur [60].	21
1.17	Correction de proximité optique	23
1.18	Principe du masque à décalage de phase	23

1.19	Simulation de l'absorption énergétique et de la température associée lors d'un recuit laser [67]. Sur ce domaine d'étude, la température reçue varie d'une cinquantaine de degrés.	25
1.20	Simulation de la température reçue lors d'un recuit spike [66]. Sur ce domaine d'étude, la température reçue varie de deux degrés.	25
1.21	Effet de proximité du well [70]	26
1.22	Influences du dessin sur le stress subi par le dispositif.	26
1.23	Paramètres d'instance proposés par Tsuno pour simuler les interactions du layout sur le process. [73]	27
1.24	Simulation des taux de croissance locaux dans le réacteur en fonction du flux de gaz [77].	28
1.25	Gravure sèche et humide [78].	29
1.26	Profils en température des recuits <i>spike</i> , <i>soak</i> et <i>flash</i> [80].	30
1.27	Extraction d'un modèle compact paramétrique (PCM) avec le logiciel <i>PARAMOS</i> de Synopsys	36
2.1	Vue d'ensemble de l'utilisation des facteurs et indicateurs de variabilité en microélectronique	40
2.2	Mapping des mesures de R_s sur une plaque ayant subi uniquement un implant et un recuit flash pour évaluer la fluctuation de température reçue.	43
2.3	Correlation de R_s à la température appliquée.	44
2.4	Estimation de la distribution de la température de recuit spike reçue par le wafer.	44
2.5	Représentation schématique de l'influence de la réduction de la longueur de grille sur le profil latéral de la concentration de dopants.	46
2.6	Schéma de mesure électrique pour caractériser le profil latéral de dopants.	46
2.7	Courant de drain I_d en fonction de la tension de drain V_{DS} dans un pMOSFET et pour différentes longueurs de grille.	47
2.8	Simulation d'une variation de température de recuit.	48
2.9	Carte de dopage de la simulation 2D du NMOS.	49
2.10	I_d vs. V_{DS} ($V_g = -0,5$ V) pour différents transistors. La flèche indique les L_g croissant.	50
2.11	Simulations TCAD de I_D vs L_g (ligne) et intervalles d'incertitude calculés à partir du modèle pour $\pm 2\sigma$ (barres verticales)	52
2.12	Mesures de I_d vs V_{DS} ($V_g = -0,4$ V). L_g croît dans le sens de la flèche.	53

2.13	a) Courant de drain en fonction de la tension de grille pour des tensions de substrat de 0 à $-1,2\text{ V}$. b) L'extraction de γ et ϕ_f permet de modéliser la déviation de la tension de seuil en fonction de V_B	56
2.14	Profil vertical de dopants et concentration de trous à $V_g = V_{th}$ pour plusieurs V_B . Le front de déplétion balaie le profil du dopage.	57
2.15	Caractérisation de l'effet de la polarisation substrat sur des dispositifs simulés en TCAD.	58
2.16	Validation de la méthode de détermination du profil de dopage du canal grâce aux simulations TCAD.	59
2.17	Dopage en fonction de la profondeur. Profils déterminés avec les équations de Wright et superposés aux profils TCAD ayant servi à générer les simulations.	61
2.18	Cartographie moyennée sur 25 wafers et histogrammes de V_{th0} et $\Delta_{V_{th}}$. $W = 0,6\mu m$; $L_{gate} = 1\mu m$	62
2.19	Détermination des profils à partir de $V_{th}(V_B)$ pour trois transistors de $1\mu m$ de long.	63
2.20	Cartographie moyennée sur 25 plaques et population du dopage N_{xtr} et de sa pente S_{xtr} . $W = 1\mu m$; $L_{gate} = 1\mu m$	64
2.21	Dispersion par plaque du dopage extrait et de sa pente.	65
2.22	Transistors longs de $L_g = 1\mu m$. a) Corrélations pour 1650 puces entre V_{th0} et $\Delta_{V_{th}}$ et b) entre N_{xtr} et S_{xtr}	65
2.23	Transistors courts ($L_g = 45nm$). a) Corrélations pour 1650 puces entre a) V_{th0} et $\Delta_{V_{th}}$ et b) N_{xtr} et S_{xtr}	66
3.1	Facteurs d'entrée du DOE	70
3.2	Différentes cartographies utilisées sur tranche de silicium dans les différents jeux de données provenant des différents ateliers.	71
3.3	Comparaison des mesures des paramètres physiques utilisés dans le DoE avec leurs valeurs nominales. Les expériences sont triées par ordre de plaque et de numéro de site. On observe un décalage et une dérive de la réalisation des paramètres entre la plaque 1 et la plaque 25 pour les paramètres X2 et X3. Dans le cas de X5, on apprécie l'importance de la variabilité intraplaque par rapport à l'excursion du DoE.	73

3.4	a) Distribution radiale de l'oxyde de grille pour 3 valeurs de X1. La variation bien définie, en vague, est typique d'une cause process alors que la variation aléatoire autour de cette vague est relativement faible. b) Distributions spatiales de X2 sur le wafer 17 qui est un point au centre du plan d'expériences. X2 montre une variation aléatoire mais avec une zone de plus faibles valeurs en haut à gauche. Ce comportement se retrouve qualitativement sur toutes les plaques.	75
3.5	Distributions spatiales intra-lot (a) et intra-plaque (b) de l'oxyde <i>liner</i> TEOS.	76
3.6	Distributions spatiales intra-lot (a,b) et intra plaque (c,d) de <i>spacer-zéro</i>	77
3.7	Distributions intra-lot (a), plaque à plaque (b) et intra-plaque (c) de l'épaisseur de nitrure déposé pour la formation des espaceurs.	78
3.8	Distributions de X5 sur les plaques où X5=1000°C sur le wafer 7 qui est un point au centre du plan d'expériences. La population globale (a) est déséquilibrée à gauche par la répartition spatiale (b) X5 qui montre un motif plus ou moins radial et monotone avec une dynamique de 4°C. Ce comportement se retrouve qualitativement sur toutes les plaques.	78
3.9	Corrélation par paire des variables technologiques. Seules les 3 plaques ayant suivi un process nominal sont affichées de sorte qu'on observe que des variations naturelles. La matrice est symétrique. La diagonale montre les histogrammes de chaque variable seule. Les axes x et y sont communs à chaque colonne et ligne respectivement.	81
3.10	Autocorrélation de NIDS1L004LS	82
3.11	Ajustement du modèle $\hat{Y}$ en fonction de Y pour le courant de drain en saturation d'un transistor N de 1 μm de large pour 40 nm de long.	87
3.12	Analyse du résidu du modèle de NIDS1L004LS (μA)	88
3.13	Résidu du modèle de NIDS1L004LS	89
3.14	Effet de chaque variable technologique du modèle indépendamment sur le modèle de NIDS1L004LS, dans son domaine de définition. Les mesures du courant sont superposées au modèle.	91

3.15	Diagramme de Pareto de l'indicateur NIDS1L004 pour un modèle du premier ordre avec interactions. Pour alléger la lecture de l'axe des abscisses, les facteurs du plan sont désignés par une lettre et les termes d'interactions et les termes carrés par les combinaisons de ces lettres correspondantes.	92
3.16	Propagation de la variance.	94
3.17	Contribution des facteurs technologiques à la variation de I_{on}	96
3.18	Contribution des facteurs technologiques à la variation de I_{lin}	98
3.19	Contribution des facteurs technologiques à la variation de V_{Tlin}	99
3.20	Contribution des facteurs technologiques à la variation du DIBL.	100
3.21	Contribution des facteurs technologiques à la variation de C_{gg}	101
3.22	Contribution des facteurs technologiques à la variation de C_{ov}	102
4.1	Propagation des paramètres process	105
4.2	Placement des <i>corners</i> dans le plan de corrélation I_{onN}/I_{onP}	108
4.3	Corrélation de I_{onN} et I_{onP} et placement des <i>corners</i>	110
4.4	Schéma d'un oscillateur en anneau.	112
4.5	Schéma d'un inverseur logique en technologie CMOS et sa représentation en porte logique.	113
4.6	Spécification des dimensions d'un transistor MOS dans PSP.	115
4.7	Corrélation de I_{onN} et I_{onP} et placement des <i>corners</i>	118
4.8	Histogramme de C_{gg} résultant de la simulation Monte-Carlo avec le modèle modifié.	119
4.9	Histogramme de C_{gd} résultant de la simulation Monte-Carlo avec le modèle modifié.	120
A.1	Histogramme de I_{on} de la simulation Monte-Carlo avec le modèle actuel.	126
A.2	Histogramme de I_{off} résultant de la simulation Monte-Carlo avec le modèle actuel.	127
A.3	Histogramme de V_{Tsat} résultant de la simulation Monte-Carlo avec le modèle actuel.	128
A.4	Histogramme de C_{gg} résultant de la simulation Monte-Carlo avec le modèle actuel.	129
A.5	Corrélation des capacités C_{gg} du NMOS et du PMOS résultant de la simulation Monte-Carlo avec le modèle actuel.	130
A.6	Histogramme de C_{gd} résultant de la simulation Monte-Carlo avec le modèle actuel.	131
A.7	Histogramme de C_{bd} résultant de la simulation Monte-Carlo avec le modèle actuel.	132

A.8	Histogramme de g_m résultant de la simulation Monte-Carlo avec le modèle actuel.	133
A.9	Histogramme de g_d résultant de la simulation Monte-Carlo avec le modèle actuel.	134
A.10	Histogramme de la fréquence d'oscillation des osculateurs en anneaux résultant de la simulation Monte-Carlo avec le modèle actuel.	135
A.11	Histogramme de la fréquence de fonctionnement des oscillateurs en anneaux résultant de la simulation Monte-Carlo avec le modèle actuel.	135
A.12	Description d'une boîte à moustaches et comparaison avec une distribution gaussienne.	137
A.13	Paire de transistors pour la simulation des fluctuations locales. Les sources indépendantes V_{dd} et $XVDD$ contrôlent respectivement les sources $egn1(b)$ et $edn1(b)$. Les sources de tension nulle $vidn1(b)$ mesurent les courants de drains.	140

Liste des tableaux

1.1	Mobilité maximale selon les orientations	22
1.2	Réactions utilisées pour les dépôts.	28
2.1	Construction du plan d'expériences	51
3.1	Mesures des facteurs.	72
3.2	Plan d'expériences du lot étudié	83
3.3	Plan d'expériences du lot étudié	85
3.4	Variations des paramètres du procédé, utilisées dans la propagation de la variance.	90
4.1	Interfaçage des données process pour introduire les déviations dans le modèle PSP.	106
4.2	Entrée des Corners pré-définis.	109
4.3	Les corners digitaux ne peuvent pas, par construction, couvrir les performances analog.	110
4.4	Paramètres process et les paramètres modèle les représentant <i>ou pouvant les représenter</i> dans PSP.	115

Liste des publications

- [117] Vincent QUENETTE, Pascal LEMOIGNE, Denis RIDEAU, Raphael CLERC, Lorenzo CIAMPOLINI, Michel MINONDO, Clément TAVERNIER et Hervé JAOUEN : Electrical characterization and compact modeling of MOSFET body effect. *In 9th International Conference on Ultimate Integration of Silicon*, pages 163–166. IEEE, 2008.
- [118] Pascal LEMOIGNE, Christian DUFAZA, Lakhdar ZAÏD et André JUGE : Variability in mos transistors with tcad tools. Journée des doctorants de l'école doctorale 353, décembre 2008.
- [119] Pascal LEMOIGNE, Clément TAVERNIER, Lorenzo CIAMPOLINI, Ardéchir PAKFAR, Christian DUFAZA, Lakhdar ZAÏD, Alexander BURENKOV, Johann SCHERMER et Peter PICHLER : Caractérisation électrique de la distribution latérale du dopage source/drain dans les mosfets nanométriques. *In 12ème édition des Journées Nationales du Réseau Doctoral de Micro-Nanoélectronique*, mai 2009.
- [120] Pascal LEMOIGNE, Vincent QUENETTE, André JUGE, Denis RIDEAU, Sylvie RETAILLEAU, Lakhdar ZAÏD, Christian DUFAZA, Clément TAVERNIER et Hervé JAOUEN : Monitoring variability of channel doping profile in the 45nm node mosfet through reverse engineering of electrical back-bias effect. *In Solid State Device Research Conference*, pages 383–386, septembre 2009.

