



HAL
open science

Evaluation des futures technologies CMOS (

Manuel Sellier

► **To cite this version:**

| Manuel Sellier. Evaluation des futures technologies CMOS (

HAL Id: tel-00631246

<https://theses.hal.science/tel-00631246v1>

Submitted on 11 Oct 2011

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

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

Université de Provence

Année 2008

Numéro attribué par la bibliothèque :

□□□□□□□□

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE PROVENCE

Spécialité : Physique et Modélisation des Systèmes Complexes : Micro et
Nanoélectronique

Présentée et soutenue publiquement par :

MANUEL SELLIER

Ingénieur de l'Ecole Centrale Marseille

Evaluation des futures technologies CMOS (<50nm) au niveau circuit

Date de soutenance : 1^{er} Octobre 2008

Jury :

Rachid BOUCHAKOUR (Professeur, Université de Provence)

Président

Lionel TORRES (Professeur, Université de Montpellier II)

Rapporteur

Pascal MASSON (Professeur, Université de Nice)

Rapporteur

Jean-Michel PORTAL (Professeur, Université de Provence)

Directeur de Thèse

Bertrand BOROT (Responsable d'Equipe, STMicroelectronics)

Co-directeur de Thèse

Table des Matières

TABLE DES MATIERES.....	2
REMERCIEMENTS	9
TABLE DES ILLUSTRATIONS.....	11
TABLE DES TABLEAUX.....	19
TABLE DES SYMBOLES	21
TABLE DES ABREVIATIONS	25
INTRODUCTION	27
CHAPITRE I DE LA MICRO A LA NANOELECTRONIQUE : EVOLUTIONS, ENJEUX ET TECHNOLOGIES EMERGEANTES	29
I.1 Introduction	29
I.2 Evolution des technologies : critères de réduction des échelles et coûts associés	30
I.3 Enjeux actuels : vers la fin de la réduction des échelles ?	34
I.3.A Limitations à l'échelle du transistor	34
a) Le transistorMOS idéal.....	34
b) Les effets parasites	35
Dégradation de la pente sous le seuil	35
Effets Canaux Courts	36
Fuites de grille	38
Polydépétion	39
Confinement quantique des porteurs de charge	39
Dégradation de la mobilité des porteurs	41
Résistances séries.....	41
I.3.B Limitations à l'échelle du circuit.....	41
a) Variabilité.....	42

Origine	42
Conséquences	43
b) Augmentation des résistances d'interconnexion	45
Aspect néfaste de la réduction des échelles :	45
Augmentation de la résistivité du métal :	46
I.4 Technologies émergentes	48
I.4.A Technologies « Front End » (composants actifs)	48
a) « Boosters » technologiques	48
Grille métallique et diélectriques haute permittivité	48
Techniques d'augmentation de la mobilité des porteurs	49
Jonctions ultra fines	49
b) Nouvelles architectures de transistors	50
SOI (« Silicon On Insulator »)	50
SON (« Silicon On Nothing »)	51
DG	52
FinFET	53
I.4.B Technologies « Back End » (interconnexions)	53
Techniques de diminution de la permittivité des isolants des lignes d'interconnexion	53
I.4.C Au niveau des circuits	54
a) DFM	54
b) Circuits de topologie régulière	54
c) Conception à base de règle statistiques	55
I.5 Conclusion	56

CHAPITRE II DE LA TECHNOLOGIE AU CIRCUIT : KIT DE CONCEPTION ET CRITERES D'EVALUATION DES FUTURES TECHNOLOGIES..... 57

II.1 Introduction	57
II.2 Flot et kit de conception.....	58
II.2.A Le flot « full custom »	58
a) Dessin des masques selon le manuel des règles de dessin	58
b) Schématisation électrique	59
c) Simulations électriques	59
Simulation du schéma (pre-layout)	59
Simulation en tenant compte des effets parasites (post-layout)	59
d) Vérifications	60
II.2.B Le flot digital standardisé : synthèse et placement routage	60
a) Description du circuit au niveau RTL	60

b) Librairie de cellules standard	61
Définition	61
Les différentes vues	61
Caractérisation des délais.....	62
c) Synthèse logique.....	62
d) Placement Routage du circuit.....	62
II.2.C Rôle du kit de conception et éléments nécessaire à un flot prédictif.....	65
II.3 Les critères d'évaluation des technologies	66
II.3.A Les critères « Front End »	66
a) Oscillateur en anneaux	66
Délai.....	66
Puissance dynamique.....	67
Puissance statique	68
b) SRAM	69
Courants de fuite.....	69
Marge au bruit à la lecture	69
Marge au bruit à l'écriture	71
Courant de lecture.....	71
Critères de robustesse à la variabilité.....	72
II.3.B Les critères « Back End »	72
a) Délai de ligne.....	73
b) Nombre optimal de répéteurs	73
c) Chemin critique d'un bloc placé routé	74
II.3.C Flot digital.....	75
a) Nombre de violations de contraintes de temps	75
b) Nombre de violation de règles de dessin	75
c) Densité de cellules standard	75
d) Distribution des familles de cellules	76
e) Distribution des sortances des cellules	76
f) Distribution des longueurs d'interconnexions	76
II.4 Conclusion	77

CHAPITRE III METHODES DE GENERATION DE KITS DE CONCEPTION PREDICTIFS 79

III.1 Introduction79

III.2 Modélisation SPICE prédictive80

III.2.A Le modèle MASTAR80

a) Principe	80
b) Intérêt.....	80
c) Limites	82
III.2.B Modélisation SPICE	82
a) Principe du simulateur SPICE	82
b) Problématiques relatives à la modélisation SPICE	83
III.2.C Méthode de modélisation prédictive.....	83
a) Modifications empiriques de cartes technologiques existantes.....	83
b) Modèles réalisés à partir de simulations numériques (TCAD)	84
III.2.D Développement d'un nouveau modèle prédictif : MSIM	85
a) Principe	85
b) Génération de la carte modèle	86
c) Aspects non programmables sous BSIM	93
d) Aspects dynamiques	95
e) Implémentation	97
f) Sources d'inexactitude du modèle.....	98
g) Prise en compte de la variabilité stochastique (désappariement).....	99
III.2.E Validation du modèle MSIM	99
III.3 Modèles d'interconnexion	103
III.4 DRM (« Design Rule Manual »).....	104
III.4.A Règle de réduction d'échelle appliquée au DRM	104
III.4.B Paramétrage de l'outil d'implémentation physique	105
III.4.C Librairie de cellules standard.....	106
a) Dessin des masques / Règles de conceptions.....	106
b) Caractérisations électriques	106
III.5 Conclusion.....	107
CHAPITRE IV APPLICATIONS DES CRITERES D'EVALUATION AUX FUTURES TECHNOLOGIES.....	109
IV.1 Introduction.....	109
IV.2 SRAM.....	110
IV.2.A Analyse de la feuille de route technologique ITRS	110
a) Elaboration des modèles	110
b) Résultats des simulations	111
IV.2.B Analyse des nœuds 45nm et 32nm	112
a) Introduction.....	112
b) Résultats	112

c) Interprétation	114
IV.2.C Etude d'une nouvelle mémoire SRAM	115
a) Principe de la première innovation	115
b) Résultats	116
c) Deuxième innovation	118
IV.2.D Conclusion	119
IV.3 Critères « Back End ».....	120
IV.3.A Scénario ITRS	120
a) Données d'entrée	120
b) Oscillateur en anneau	120
c) Délai d'un bloc placé routé.....	121
d) Délais de ligne.....	122
e) Nombre optimal de répéteurs	124
f) Influence de la conductance des cellules (« drive »).....	126
IV.3.B Analyse du nœud 32nm.....	129
IV.4 Flot digital prédictif.....	131
IV.4.A Introduction	131
IV.4.B Impact de l'augmentation des résistances en technologie 65nm	132
IV.4.C Analyse du nœud 32nm.....	136
a) Validation du flot prédictif en technologie 45nm	136
b) Extrapolation en 32nm	136
IV.5 Conclusion	139
 CONCLUSION	 141
 RÉFÉRENCES.....	 145
 BIBLIOGRAPHIE DE L'AUTEUR.....	 149
 BREVET.....	 151
 ANNEXE.....	 153
 A.1 Exemple de fichier librairie paramétrant MSIM.....	153
A.2 Détails concernant le bloc arithmétique utilisé pour l'évaluation du nœud 32nm	162
A.3 Contraintes de délais dans un bloc placé routé	164

A.3.A	Temps de « hold » et « setup »	164
A.3.B	Définition des « slacks »	165

Remerciements

Au-delà du rôle de simples contributeurs ou collaborateurs, nombre de mes collègues et de mes supérieurs ont fait preuve d'une intelligence rare en ne se prêtant pas, pourtant dans un contexte difficile d'alliance technologique à trois, au culte de l'individualisme si souvent répandu dans nombre de services d'entreprises. Bien au contraire, ils ont cultivé un esprit d'émulation fertile mu par la seule volonté de faire progresser l'état de l'art de leur domaine. Ils ont su me faire comprendre que nulle invention notable ne peut s'établir sans la collaboration de chacun ni le partage systématique des informations.

Merci donc à toute l'équipe Full Custom Design et plus particulièrement à son chef et encadrant de cette thèse, Bertrand, ainsi qu'à l'ensemble de l'équipe Modules Avancés, et spécialement à Thomas et Fred pour m'avoir sans cesse aiguillé dans cette direction.

Merci à Jean-Michel, pour son esprit toujours novateur mais néanmoins rigoureux.

Merci à Stéphane, pour la confiance qu'il m'a accordée et son soutien.

Merci également à tous mes collègues et amis, bien trop nombreux pour être en mesure de dresser une liste exhaustive.

Merci à Steve, Manu, Pema, Robin, Marilyne, Brice, Yann, Maxime, Alex, Bart, Christophe, Yannick, Rodolphe, Richard, Anne-Laure, Gilles, Koos, Pierre, Matthieu, Jean-Pierre, Seb, Philou, Arnaud, Romain, Emilie, Alexis, Fabrice, ainsi que tous ceux que j'ai oubliés !

Merci à Angéline pour ses relectures et sa patience.

En espérant que ce manuscrit trouve lui aussi son utilité première dans le partage des connaissances !

Table des Illustrations

Figure	Intitulé	Page
Figure I-1	Evolution du nombre de transistors par puce en fonction de l'année de lancement de la production. La tendance énoncée par Moore en 1975 est parfaitement respectée (d'après [4]).	32
Figure I-2	Evolution du coût par transistor en fonction du temps (d'après [4]).	32
Figure I-3	Evolution des performances des microprocesseurs en fonction du temps (d'après [4]).	33
Figure I-4	Evolution du coût des usines en fonction du temps (d'après [2]).	34
Figure I-5	Principe de fonctionnement d'un transistor idéal.	35
Figure I-6	Caractéristique $I_{DS}=f(V_{GS})$ d'un transistor MOS en échelle logarithmique. La pente sous le seuil caractérise la décroissance exponentielle du courant en régime sous le seuil.	36
Figure I-7	Principe de l'abaissement de la barrière de potentiel (bande de conduction) par la réduction de la longueur de grille du transistor.	37
Figure I-8	Illustration de l'impact du dopage par poche afin de maintenir une barrière de potentiel la plus constante possible entre canal long et canal court et ainsi diminuer les effets canaux courts.	38
Figure I-9	Courant de fuite de grille pour différentes épaisseurs d'oxyde (issu de [7]).	38
Figure I-10	Impact des effets canaux courts et du courant de grille sur une caractéristique $I_{DS}-V_{GS}$ (d'après [6]).	39
Figure I-11	Courbure de bandes dans l'empilement de grille.	40
Figure I-12	Illustration des différentes capacités parasites s'ajoutant en série à la capacité d'oxyde et conduisant ainsi à une diminution de la capacité totale de grille.	40

Figure	Intitulé	Page
Figure I-13	Les différentes résistances séries parasites.	42
Figure I-14	Schéma d'un point mémoire SRAM.	43
Figure I-15	Description de la courbe en papillon.	44
Figure I-16	Impact de la variabilité des caractéristiques des transistors sur une courbe en papillon.	44
Figure I-17	Evolution des capacités et résistances de ligne calculée à partir du logiciel HF2D [13] et des données de l'ITRS [14].	45
Figure I-18	Illustration des phénomènes conduisant à l'augmentation de la résistivité des lignes d'interconnexion.	47
Figure I-19	Différentes contributions à l'augmentation de la résistivité des lignes d'interconnexion en cuivre (d'après [14]).	47
Figure I-20	Lignes de champ dans le cas de jonctions épaisses (A) et dans le cas de jonctions fines (B).	49
Figure I-21	Vue en coupe (A) d'une plaque SOI, (B) d'un transistor SOI partiellement déplété (PDSOI), (C) d'un transistor SOI complètement déplété (FDSOI).	51
Figure I-22	Vue en coupe d'un transistor SON.	52
Figure I-23	Description du procédé de fabrication SON (issu de [21]).	52
Figure I-24	Vue en coupe d'un transistor double grille.	53
Figure I-25	Vue 3D d'un transistor FinFET.	54
Figure I-26	Illustration d'un masque régulier	55
Figure II-1	Schéma d'un flot de conception de circuit de type « full custom ».	60

Figure	Intitulé	Page
Figure II-2	Description d'un circuit au niveau RTL.	61
Figure II-3	Exemple de différentes vues d'une cellule standard de type NAND2.	61
Figure II-4	Illustration d'un flot digital.	63
Figure II-5	Exemple de schéma d'un flot de conception de circuit digital.	64
Figure II-6	Vue de type masques (« layout ») d'un bloc digital placé routé (additionneur et multiplieur : 5000 transistors).	64
Figure II-7	Principe de l'oscillateur en anneau.	67
Figure II-8	Schéma équivalent d'un étage d'un oscillateur en anneau.	67
Figure II-9	Circuits permettant d'arrêter les oscillations d'un oscillateur en anneau pour mesurer sa puissance statique.	69
Figure II-10	SNM d'une mémoire SRAM.	70
Figure II-11	Illustration de la faiblesse du critère $\langle \text{SNM} \rangle$ seul. Ce qui compte, c'est la faible probabilité d'avoir des SNM négatives, qui est mieux traduite par le critère $\langle \text{SNM} \rangle / \sigma \text{SNM}$.	71
Figure II-12	Courant de lecture I_{read} d'une mémoire SRAM.	72
Figure II-13	Critère $\langle p \rangle / \sigma p$ minimal nécessaire pour avoir une probabilité de rejet de 100ppm en fonction de la taille de la mémoire.	73
Figure II-14	Mesure du délai entre deux portes séparées par une ligne d'interconnexion de longueur variable.	73
Figure II-15	Mesure du délai entre deux portes séparées par une ligne d'interconnexion de longueur fixe et présentant un nombre de répéteurs variable.	74

Figure	Intitulé	Page
Figure II-16	Principe du circuit « Flipper » développé par STMicroelectronics destiné au calcul du délai d'un chemin critique d'un bloc (additionneur ou multiplieur) placé routé.	75
Figure II-17	Distribution des longueurs d'interconnexion pour un circuit de type contrôleur de port USB.	77
Figure III-1	Vue d'écran du logiciel MASTAR.	81
Figure III-2	Exemple d'application de MASTAR.	82
Figure III-3	Génération d'une carte modèle d'une technologie N+1 à partir des données des précédentes générations de cartes modèles.	84
Figure III-4	Principe de modélisation MASTAR et MSIM.	86
Figure III-5	Modélisation des résistances d'accès et des courants de grille avec la méthodologie MSIM.	95
Figure III-6	Modélisation des jonctions source et drain.	97
Figure III-7	Influence du paramètre DELTA sur le délai d'un inverseur.	98
Figure III-8	Caractéristiques ID-VG et ID-VD NMOS et PMOS obtenues grâce au modèle MSIM à partir d'un profil MASTAR sans déplétion de polysilicium ni résistances d'accès.	100
Figure III-9	Caractéristiques ID-VG et ID-VD NMOS et PMOS obtenues grâce à la méthodologie MSIM à partir d'un profil MASTAR correspondant à une technologie 90nm.	101
Figure III-10	Caractéristiques ID-VG et ID-VD NMOS et PMOS obtenues grâce à la méthodologie MSIM à partir d'un profil MASTAR correspondant à une technologie 65nm.	101
Figure III-11	Caractéristique C-V obtenue à l'aide du modèle prédictif MSIM et du modèle standard 65nm.	101
Figure III-12	Oscillateur en anneau.	102

Figure	Intitulé	Page
Figure III-13	Comparaison des données ION/IOFF en 45nm mesurées sur silicium et simulées grâce à MSIM avec différentes longueurs de grille.	103
Figure III-14	Comparaison du compromis délai d'oscillateurs en anneaux / courant de fuite en 45nm mesuré sur silicium et simulé grâce à MSIM avec différentes longueurs de grille.	103
Figure III-15	Réseau RC distribué pour modéliser les interconnexions sous SPICE. Le réseau est composé de 100 éléments RC.	104
Figure III-16	Situation typique d'une ligne d'interconnexion dans un bloc placé routé utilisée pour extraire les valeurs de capacité de fil.	104
Figure III-17	Génération d'un fichier « techno.lef » prédictif.	105
Figure III-18	Génération d'un abstract de cellule standard d'une génération n+1 à partir de la génération n	106
Figure IV-1	Evolution du rapport $\langle \text{SNM} \rangle / \sigma \text{SNM}$ en fonction de l'année de production.	112
Figure IV-2	Simulation de courbes en papillon pour trois types de dispositifs différents (« Bulk », SOI et SON) pour une cellule de technologie 45nm présentant une surface de $0,25\mu\text{m}^2$.	113
Figure IV-3	Distributions de SNM associées aux simulations de la Figure IV-2.	113
Figure IV-4	Simulation de courbes en papillon en 32nm pour différentes tailles de cellules et différents types de dispositif.	114
Figure IV-5	Variation du nombre de dopants en fonction de la concentration moyenne en dopants (graphe du dessous) et variation de tension de seuil résultante (graphe du dessus) pour deux types de dispositifs (« Bulk » avec grille en polysilicium et SON avec grille métallique « midgap »).	115
Figure IV-6	Distribution de la SNM, de la WM, du courant Iread ainsi que courbe en papillon d'une cellule mémoire 22nm (32nm « half pitch », année 2012) avec dispositifs NMOS et PMOS non dopés élaborée d'après l'ITRS LOP.	117

Figure	Intitulé	Page
Figure IV-7	Distribution de la SNM, de la WM, du courant Iread ainsi que courbe en papillon d'une cellule mémoire 22nm (32nm « half pitch », année 2012) avec dispositifs NMOS et PMOS dopés élaborée d'après l'ITRS LOP.	118
Figure IV-8	Distribution de la SNM, de la WM, du courant Iread ainsi que courbe en papillon d'une cellule mémoire 22nm (32nm « half pitch », année 2012) avec dispositifs NMOS non dopé et PMOS dopé élaborée d'après l'ITRS LOP.	118
Figure IV-9	Evolution des critères $\langle \text{SNM} \rangle / \sigma \text{SNM}$, $\langle \text{Iread} \rangle / \sigma \text{Iread}$ et $\langle \text{WM} \rangle / \sigma \text{WM}$ d'une cellule 22nm avec transistors NMOS et PMOS dopés en fonction du facteur de dégradation de la mobilité des transistors PMOS.	119
Figure IV-10	Evolution des critères $\langle \text{SNM} \rangle / \sigma \text{SNM}$, $\langle \text{Iread} \rangle / \sigma \text{Iread}$ et $\langle \text{WM} \rangle / \sigma \text{WM}$ d'une cellule 22nm avec transistors NMOS non dopés et PMOS dopés en fonction du facteur de dégradation de la mobilité des transistors PMOS.	120
Figure IV-11	Evolution de la fréquence d'oscillateur en anneau simulé par MSIM d'après les données de l'ITRS HP et tendance CV/I prévue par MASTAR.	122
Figure IV-12	Délai du chemin critique d'un additionneur 10 bits en fonction de l'année de production.	123
Figure IV-13	Délai de ligne en fonction de la longueur pour différentes technologies de la feuille de route ITRS.	124
Figure IV-14	Evolution de la longueur critique en fonction de l'année de production d'après les données de l'ITRS.	125
Figure IV-15	Délai en fonction du nombre de répéteurs pour une ligne de longueur fixe et pour différentes technologies de la feuille de route ITRS.	126
Figure IV-16	Impact du scénario d'augmentation de la résistance d'interconnexion de l'ITRS sur le pourcentage de répéteurs requis dans le circuit dans le cas d'une distribution de longueur de fils équivalente à celle d'un contrôleur de port USB 65nm.	127
Figure IV-17	Equivalence électrique de la ligne de transmission.	128

Figure	Intitulé	Page
Figure IV-18	Délai de ligne en fonction de la longueur de ligne et de la sortance de la cellule chargeant la ligne pour différents scénarios de la feuille de route ITRS.	129
Figure IV-19	Délais d'oscillateur en anneau FO1 et FO3 en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.	131
Figure IV-20	Puissance par fréquence d'un oscillateur en anneau en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.	131
Figure IV-21	Délai et puissance dynamique d'un additionneur 10 bits en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.	131
Figure IV-22	Longueur critique d'interconnexion et délai optimal le long d'une ligne avec répéteurs en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.	132
Figure IV-23	Densité et pourcentage de « Buffers » dans un bloc contrôleur USB en fonction de différents scénarios d'augmentation des résistances d'interconnexion.	134
Figure IV-24	Répartition du type de cellules dans un bloc contrôleur de port USB.	135
Figure IV-25	Distribution des longueurs d'interconnexion pour un bloc contrôleur de port USB selon différents scénarios d'augmentation des résistances prévus par l'ITRS.	136
Figure IV-26	Distribution des longueurs d'interconnexion pour un bloc contrôleur de port USB selon différents scénarios d'augmentation des résistances prévus par l'ITRS.	136
Figure IV-27	Densité du circuit en fonction de la fréquence pour le flot 45nm prédictif et pour le flot 45nm standard.	137
Figure IV-28	Distribution de longueurs de fil dans le cas de l'implémentation 45nm et 65nm référence et dans le cas de la prédiction pour la technologie 32nm.	139

Figure	Intitulé	Page
Figure IV-29	Répartition du type de cellules dans un bloc arithmétique.	140
Figure A-1	Schéma du bloc arithmétique.	163
Figure A-2	Détail du sous bloc Mult/Add.	163
Figure A-3	Les différents délais d'un chemin dans un bloc logique.	165

Table des Tableaux

Tableau	Intitulé	Page
Tableau 1	Données d'entrée utilisées pour l'évaluation SRAM de la feuille de route ITRS.	111
Tableau 2	Récapitulatif des données utilisées pour simuler les critères d'évaluation technologiques « back end » de la feuille de route ITRS.	121
Tableau 3	Tableau récapitulatif de l'impact de l'augmentation de la résistivité des interconnexions en terme de répéteurs pour un circuit contrôleur de port USB 65nm.	127
Tableau 4	Facteur de pondération des résistances d'interconnexion prévu par l'ITRS par rapport au nœud 65nm	133
Tableau 5	Rapport d'implémentation du bloc contrôleur USB en 65nm en fonction du scénario d'augmentation des résistances utilisé.	134
Tableau 6	Récapitulatif des données relatives à l'implémentation du bloc arithmétique pour l'évaluation des performances 32nm.	139

Table des Symboles

Nom	Unité	Description
C_{dep}	F	Capacité de déplétion
$C_{overlap}$	F	Capacité de recouvrement
C_{ox_phys}	F	Capacité physique d'oxyde de grille
C_{ox_el}	F	Capacité électrique de l'oxyde de grille
d	-	Paramètre de pondération du courant par l'effet non uniforme du substrat le long du canal
E_c	V/m	Valeur du champ critique
I_{DS}	A	Courant passant entre le drain et la source d'un transistor MOS
$I_{D,sat0}$	A	Courant de saturation source drain sans effets de résistances parasites
I_{GS}, I_{GD}, I_{GB}	A	Courants de la grille respectivement vers la source, le drain et le substrat
I_{ON}	A	Courant I_{DS} traversant le canal d'un transistor dans l'état saturé
I_{OFF}	A	Courant I_{DS} traversant le canal d'un transistor dans l'état bloqué
I_{read}	A	Courant de lecture d'un point mémoire SRAM
k	$J K^{-1}$	Constante de Boltzmann ($k = 1,38 \times 10^{-23} J K^{-1}$)
K_B	$V^{1/2}$	Coefficient de substrat
L	m	Longueur d'une capacité ou d'un transistor
L_{eff}	m	Longueur effective du canal du transistor
N_B ou N_{ch}	m^{-3}	Concentration en dopant du canal du transistor
n	m^{-3}	Nombre de dopants dans le canal
P_{dyn}	W	Puissance dynamique dissipée par un circuit
P_{stat}	W	Puissance statique dissipée par un circuit
q	C	Charge élémentaire de l'électron

Nom	Unité	Description
Q_{dep}	$C\ m^{-2}$	Charge de la couche de déplétion
R_{con}	Ohm	Résistance de contact
R_{dp}	Ohm	Résistance de source/drain
R_{ext}	Ohm	Résistance d'extension
R_{ov}	Ohm	Résistance d'overlap
R_S	Ohm	Résistance parasite série totale
S	mV/dec	Pente sous le seuil
T	K	Température
T_{dep}	m	Profondeur de déplétion
T_{ox_el}	m	Epaisseur électrique d'oxyde de grille
T_{ox_phys}	m	Epaisseur physique d'oxyde de grille
$T_{Polydep}$	m	Profondeur de déplétion dans la grille
t_{poly}	m	Epaisseur du polysilicium de grille
τ_p	s	Temps typique de traversée d'un inverseur
V_{BS}	V	Tension de polarisation source substrat
V_{DS}	V	Tension appliquée entre le drain et la source
V_{dd}	V	Tension d'alimentation
$V_{D,sat}$	V	Tension de saturation du drain
V_{FB}	V	Tension de bandes plates
$V_{th,on}$	V	Tension de seuil en régime de forte inversion
$V_{th,off}$	V	Tension de seuil en régime sous le seuil
W	m	Largeur du transistor
X_j	m	Profondeur des jonctions du TMOS
ϵ_0	$F\ m^{-1}$	Permittivité du vide ($\epsilon_0 = 8,85 \times 10^{-12}\ F\ m^{-1}$)
ϵ_{ox}	$F\ m^{-1}$	Constante diélectrique de l'oxyde SiO_2 ($\epsilon_{ox} = 3,9 \cdot \epsilon_0$)
ϵ_S	$F\ m^{-1}$	Constante diélectrique du Silicium ($\epsilon_S = 11,9 \cdot \epsilon_0$)

Nom	Unité	Description
Ψ_s	V	Potentiel de surface dans le semi-conducteur
Ψ_e	-	Fonction d'onde d'un électron
μ_{eff}	$\text{m}^2\text{V}^{-1}\text{s}^{-1}$	Mobilité des électrons dans le canal
ρ	Ohm.m^{-1}	Résistivité du métal
Φ_F	eV	Niveau de Fermi

Table des Abréviations

BL	Bit Line
BSIM	Berkeley Short-channel Igfet Model
CAD	Computer Aided Design (voir CAO)
CAO	Conception Assistée par Ordinateur
CLK	CLocK
CMOS	Complementary Metal Oxide Semiconductor
DIBL	Drain Induced Barrier Lowering
DG	Double Gate
DLY	DeLaY cell
DRC	Design Rule Check
DRM	Design Rule Manual
FDSOI	Fully Depleted Silicon On Insulator
FinFET	Fin Field Effect Transistor
FO1/FO3	Fan Out 1 / Fan Out 3
GDS	Graphic Data System
HP	High Performance
IDG	Independent Double Gate
ITRS	International Technology Roadmap for Semiconductors
IV	InVerter
MASTAR	Model for AnalySis of Transistors And Roadmaps
MOS	Metal-Oxide-Semiconductor
MSIM	MASTAR SIMulator
LOP	Low Operating Power
LP	Low Power

LVS	Layout Versus Schematic
OPC	Optical Process Control
PD	Pull Down
PDSOI	Partially Depleted Silicon On Insulator
PG	Pass Gate
PIDS	Process, Integration, Devices, and Structures
PPF	Power Per Frequency
PU	Pull Up
RAM	Random Access Memory
RTL	Register Transfer Level
SPICE	Simulation Program with Integrated Circuits Emphasis
SRAM	Static Random Access Memory
SCE	Short Channel Effect
SNM	Static Noise Margin
SOI	Silicon On Insulator
SON	Silicon On Nothing
TCAD	Technology CAD
TMOS	Transistor MOS
TNS	Total Negative Slack
VDT	Voltage Doping Transformation
VHDL	Very High Speed Integrated Circuit Hardware Description Language
WL	Word Line
WM	Write Margin
WNS	Worst-case Negative Slack
ZCE	Zone de Charge d'Espace

Introduction

L'industrie de la microélectronique s'est lancée dans une course à la miniaturisation depuis les années 1960. A cette époque, les premiers circuits intégrés étaient constitués de seulement quelques transistors par puce et les plaquettes de silicium faisaient 2'' (environ 50mm). A l'heure actuelle, les derniers microprocesseurs possèdent plusieurs milliards de transistors et sont produits sur des plaquettes de silicium de 300mm de diamètre. Ce formidable essor, sans commune mesure dans le reste de l'industrie, n'a pas été possible sans surmonter un nombre important de difficultés technologiques dans les deux grandes branches de métier que sont la technologie et la conception de circuit.

Au niveau technologique, il a fallu intégrer des sources de longueur d'onde de plus en plus faible afin de graver des motifs toujours plus fins, améliorer sans cesse la qualité des produits chimiques impliqués dans les procédés de gravure et progresser dans la défektivité globale du processus de fabrication à l'aide d'infrastructures adaptées (salle blanche et outils de fabrication). Pour continuer la course à la réduction des échelles, il faudra savoir également intégrer dans un futur proche des nouveaux matériaux tels que les diélectriques haute permittivité pour les oxydes de grilles, les grilles métalliques ou bien les diélectriques poreux pour les interconnexions.

Au niveau de la conception de circuit, il a été nécessaire d'incorporer des outils informatiques (CAD) toujours plus performants afin que les coûts de conception des circuits ne croissent pas avec la complexité de ces derniers. Outre la gestion d'une complexité croissante, les grands enjeux actuels et futurs concernent la gestion de paramètres technologiques dégradés de génération en génération tout en maintenant la vitesse et la puissance dissipée à des niveaux acceptables.

Au-delà des problématiques citées ci-dessus propres à chaque branche de métier, il apparaît aujourd'hui qu'un certain nombre de sujets ne relèvent totalement d'aucune des deux branches. Ainsi, l'augmentation drastique des résistances d'interconnexions attendue pour les futures générations et la hausse continue de la variabilité des paramètres technologiques sont des réalités sur lesquelles les technologues n'ont que peu de marge de manœuvre et qui affectent tout particulièrement les concepteurs de circuits. La relation

entre la performance du transistor isolé et celle du circuit fini devient ainsi de moins en moins directe impliquant un manque de visibilité dans le choix des meilleures technologies pour les circuits de demain.

Face à ce problème, il est nécessaire de bâtir des passerelles entre le monde de la technologie et le monde de la conception de circuit afin de cerner au mieux l'impact de telle ou telle technologie au niveau des circuits.

Tout l'enjeu de cette étude est ainsi de définir des méthodes permettant de passer d'un univers à l'autre plus simplement. Elle est composée de quatre chapitres différents. Le premier chapitre présente la loi de réduction des échelles qui régit l'industrie des semi-conducteurs. Les principaux obstacles à cette réduction y sont abordés ainsi que les solutions technologiques innovantes envisagées pour le futur. Le Chapitre II est consacré à la conception de circuit. Il introduit aux différents flots de conception et présente les critères d'évaluation technologiques sur lesquels s'articule cette étude. Dans le chapitre III sont présentées les techniques développées permettant de réaliser des kits de conception dans le but d'évaluer les futures technologies. Ceux-ci regroupent des modèles électriques pour les dispositifs actifs et les interconnexions, ainsi que des règles de paramétrage des outils de CAO. Le dernier chapitre rassemble l'ensemble des résultats des évaluations menées. Deux types d'analyse sont réalisées : simulation de l'évolution au cours du temps d'un scénario fixé par l'ITRS et analyse poussée du prochain nœud 32nm. Les critères d'évaluation utilisés sont ceux définis chapitre II. Ils incluent les caractéristiques d'un bloc digital placé routé réalisé de manière entièrement prédictive.

Chapitre I De la micro à la nanoélectronique : évolutions, enjeux et technologies émergentes

La réduction des échelles bénéficie de deux attraits majeurs qui ont justifié sa poursuite depuis les années 1960 : augmentation à la fois de la fonctionnalité et des performances des circuits.

Cette réduction ne va pas sans poser des difficultés technologiques importantes à la fois au niveau des transistors (dégradation de la pente sous le seuil, effets canaux courts, fuites de grilles, polydéplétion, confinements quantiques des porteurs de charge, dégradation de la mobilité et augmentation des résistances parasites) mais également au niveau du circuit (variabilité, augmentation des résistances d'interconnexion entre autres).

Plusieurs technologies émergentes sont à l'étude afin de passer outre ces difficultés. Au niveau des transistors il s'agit de l'utilisation de grilles métalliques, de jonctions ultra fines, de techniques d'augmentation de la mobilité ou bien de nouvelles architectures de dispositifs (SOI, SON, Double grille). Concernant les interconnexions, des diélectriques à faible permittivité sont étudiés. Au niveau des circuits, des techniques telles que le DFM (« Design For Manufacturing »), les bibliothèques à topologie régulières ou bien la conception à base de règles statistiques semblent des solutions prometteuses.

I.1 Introduction

L'objet de ce chapitre est d'introduire le lecteur à la problématique spécifique à l'industrie de la microélectronique qui consiste à suivre la loi de réduction des échelles spécifiée par Gordon Moore en 1968 [1]. Les défis technologiques ayant permis de réduire les échelles, et donc de confirmer la loi, ont toujours été surmontés depuis cette date. Cependant, le nombre de difficultés à maîtriser ne cesse de croître à chaque nouvelle génération technologique.

Après avoir présenté dans une première partie en quoi la réduction des échelles est un enjeu important pour l'industrie des semi-conducteurs, les principaux obstacles à cette réduction sont abordés dans une deuxième partie. Finalement plusieurs solutions

technologiques permettant de surmonter les difficultés mentionnées sont décrites dans la troisième partie.

I.2 Evolution des technologies : critères de réduction des échelles et coûts associés

La microélectronique a été l'objet d'un formidable développement au cours de la deuxième moitié du siècle dernier. En effet, c'est la seule industrie qui a vu les critères de performance des biens qu'elle produit augmenter de manière exponentielle sans discontinuer pendant plus de cinquante ans. La plupart des graphiques montrant l'évolution d'un critère de performance en microélectronique sont ainsi souvent mis en échelle logarithmique. Même le profane en électronique peut se rendre compte d'une telle évolution : il a pu constater que, lors d'un achat d'un bien électronique (téléphone, ordinateur, lecteur MP3,...), il lui faut constamment réadapter ses échelles de valeur. Il ne compte ainsi plus les fréquences des appareils en MHz comme il y a quelques années mais en Ghz, ou bien les capacités mémoires non plus en Kbits ou Mbits mais en Gbits.

Gordon Moore, co-fondateur d'Intel, a su très tôt prévenir la venue de ce processus en édictant une « loi » de réduction des échelles [1] : le nombre de transistors par puce doit doubler tous les ans (cf. Figure I-1). Même si l'impact premier de cette loi est une réduction physiques des dimensions des transistors, il s'agit avant tout d'une loi économique. Elle tire son origine d'un principe physique qui, associé aux lois de la finance, assure une prospérité constante aux entreprises du secteur.

En effet, les puces électroniques sont composées de transistors dont le critère premier de performance est la capacité à charger d'autres transistors. Cette faculté est traduite par le ratio

$$\frac{C_{ox} V_{dd}}{I_{ON}} \quad (\text{I-1})$$

I_{ON} étant le courant du transistor à l'état passant, V_{dd} la tension d'alimentation et C_{ox} la capacité d'entrée du transistor. Or, I_{ON} peut se réécrire en première approximation sous la forme :

$$I_{ON} = \frac{1}{2} \mu \frac{\varepsilon_{ox}}{T_{ox}} \frac{W}{L} V_{dd}^2, \quad (I-2)$$

avec μ la mobilité des porteurs de charge, ε_{ox} la permittivité diélectrique de l'oxyde de grille, T_{ox} l'épaisseur d'oxyde et W, L respectivement la longueur et la largeur du transistor. C_{ox} peut également s'écrire sous la forme

$$C_{ox} = \frac{\varepsilon_{ox} W L}{T_{ox}} \quad (I-3)$$

D'une génération à l'autre, les dimensions des dispositifs sont souvent réduites d'un facteur $k=0,7$ de manière à réduire la surface totale du circuit d'un facteur $k^2=0,5$. En diminuant ainsi les échelles d'un facteur k (diminution de W, L, T_{ox} et V_{dd} d'un facteur k), le facteur de performance de la génération $n+1$ par rapport à la génération n devient :

$$\left(\frac{C_{ox} V_{dd}}{I_{ON}} \right)_{+1} = \frac{C_{ox_{n+1}} V_{dd_{n+1}}}{I_{ON_{n+1}}} \frac{\frac{\varepsilon_{ox} k W k L}{k T_{ox}} k V_{dd}}{\frac{1}{2} \mu \frac{\varepsilon_{ox}}{k T_{ox}} \frac{k W}{k L} k k V_{dd}^2} = \frac{\frac{\varepsilon_{ox} W L}{T_{ox}} V_{dd}}{\frac{1}{2} \mu \frac{\varepsilon_{ox}}{T_{ox}} \frac{W}{L} V_{dd}^2} = \left(\frac{C_{ox} V_{dd}}{I_{ON}} \right) \quad (I-4)$$

La réduction des échelles d'un facteur k d'une génération à l'autre implique donc deux effets bénéfiques : réduction de la taille des circuits (ou augmentation de la fonctionnalité par augmentation du nombre de transistors pour une même taille de puce d'un facteur $1/k^2$) et augmentation des performances d'un facteur $1/k$. En d'autres termes, tout en diminuant les coûts il est possible d'augmenter les performances. Ces deux phénomènes sont illustrés Figure I-2 et Figure I-3. Ils ont justifiés les efforts massifs que l'industrie des semi-conducteurs a réalisés dans la réduction des échelles. Cela lui a assuré en retour un taux de croissance moyen d'environ 17% par an grâce à la vente de produits toujours plus innovants, ce qui est remarquable.

Cependant, le nombre de défis techniques à surmonter augmente à mesure que les technologies deviennent fortement submicroniques. De ce fait, plus les technologies avancent, plus le coût de production par plaquette est élevé. Afin de conserver un retour sur investissement correct, il est impératif de contenir les coûts par puce à des niveaux acceptables. Un moyen simple consiste à augmenter la taille de la plaquette de manière à produire plus de puces sur une même plaquette. Le diamètre des plaquettes est ainsi passée de 12mm dans les années 1960 à 300mm aujourd'hui, et devrait passer à 450mm en 2012 [2].

Or, le coût de l'unité de production va de pair avec la taille de l'infrastructure, qui est liée à son tour à la taille de la plaquette. Il atteint actuellement plusieurs milliards d'euros pour une unité de production 45nm (cf. Figure I-4). Une seule entreprise, Intel, est capable d'investir seule dans le développement des technologies. Les entreprises du secteur ne peuvent plus supporter à elles seules un tel coût et elles ont du former des systèmes d'alliance technologique pour partager les investissements en recherche et développement à l'image de l'ancienne Alliance Crolles2 (STM, Philips, Freescale), de l'alliance TSMC (TSMC, NXP) ou de l'alliance ISDA (IBM, AMD, Chartered, Infineon, Samsung, Sony et Toshiba, rejoints récemment par Freescale et STM)

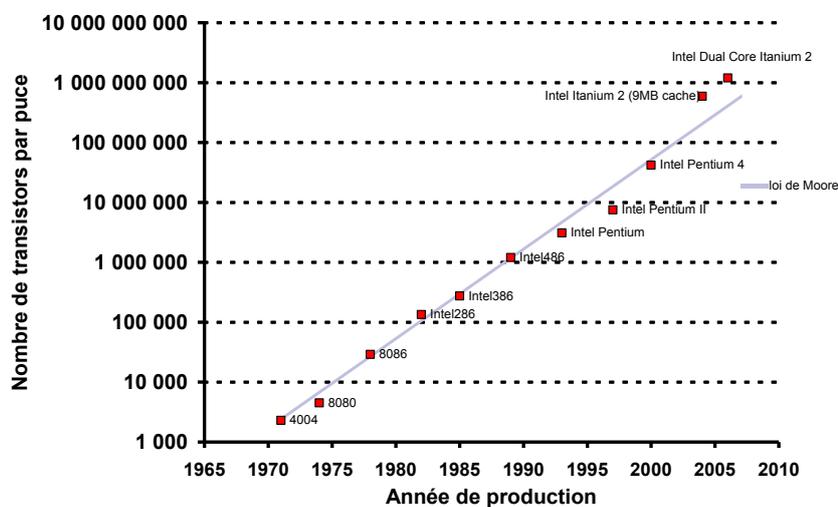


Figure I-1 : Evolution du nombre de transistors par puce en fonction de l'année de lancement de la production. La tendance énoncée par Moore en 1975 est parfaitement respectée (d'après [4]).

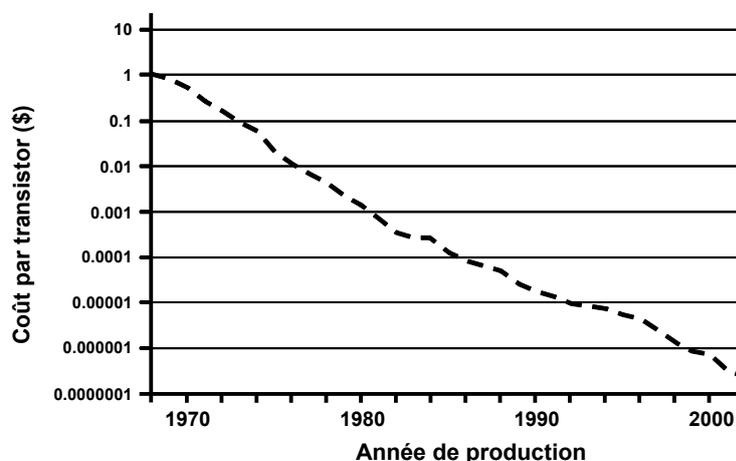


Figure I-2 : Evolution du coût par transistor en fonction du temps (d'après [4]).

Les risques financiers liés à des investissements de plusieurs milliards d'euros sont colossaux. En effet, compte tenu du nombre croissant de défis technologiques à surmonter, il n'est pas exclu que les investissements engagés ne soient pas couronnés de succès en terme de performance technologique. De ce fait, certains leaders comme Texas Instruments ont décidé d'adopter un nouveau modèle de fonctionnement basé sur la sous-traitance entière de leur production de façon à reporter les risques liés aux investissements technologiques sur d'autres entreprises [3]. Ainsi, la réduction du nombre d'entreprises investissant dans la fabrication des technologies CMOS avancées semble être le principal obstacle à la continuation de la loi de Moore, la course à la miniaturisation pouvant alors s'arrêter alors faute de compétiteurs .

Cependant, tous les fondeurs n'ont pas décidé de suivre le modèle dit « fabless » de Texas Instrument. Avant de s'engager dans des dépenses faramineuses pour développer une technologie, ces entreprises veulent s'assurer de la faisabilité de la technologie à développer et du retour sur investissement auquel elles peuvent s'attendre. Tout l'enjeu de cette étude est ainsi de fournir des critères d'évaluations technologiques nouveaux permettant d'enrichir l'état de l'art actuel à propos des risques liés au développement des nouvelles technologies.

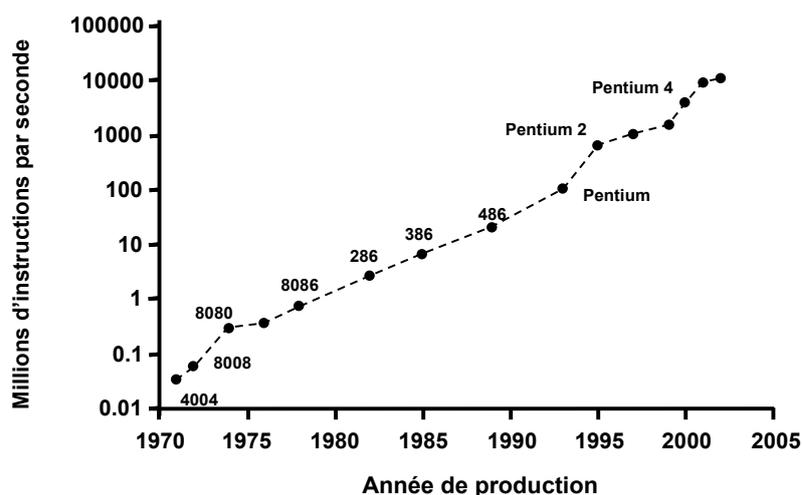


Figure I-3 : Evolution des performances des microprocesseurs en fonction du temps (d'après [4]).

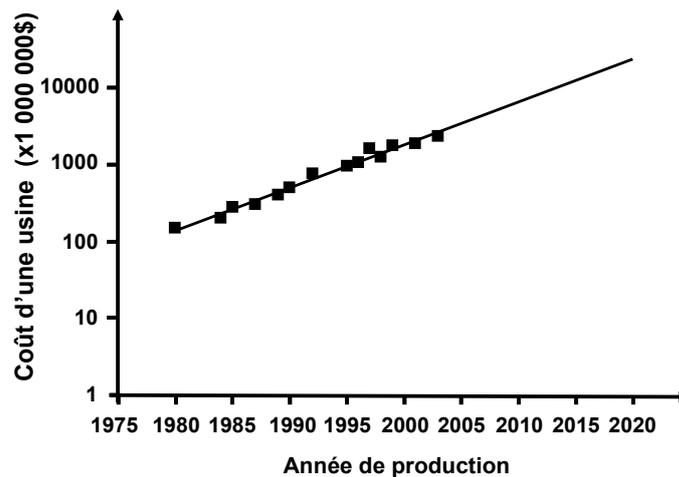


Figure I-4 : Evolution du coût des usines en fonction du temps (d'après [2]).

I.3 Enjeux actuels : vers la fin de la réduction des échelles ?

I.3.A Limitations à l'échelle du transistor

a) Le transistor MOS idéal

Le transistor MOS idéal est un interrupteur commandé en tension. Il a deux états : un état passant où il est capable de débiter un courant I_{ON} et un état bloqué où il ne circule aucun courant. Le principe de fonctionnement repose sur le principe de la capacité MOS (Métal Oxyde Semi-conducteur).

Lorsque la tension de grille est inférieure à la tension de seuil, aucune couche conductrice n'est présente dans le canal, c'est le régime de déplétion de la capacité MOS. Le transistor est alors équivalent à deux diodes têtes bêtes et aucun courant ne circule. Le transistor peut être assimilé en première approximation à un interrupteur ouvert (Figure I-5).

Lorsque la tension de grille est suffisante (ou dépasse un certain seuil appelé tension de seuil), une couche conductrice de faible résistivité se forme dans le canal, c'est le régime d'inversion forte. Un courant peut alors circuler entre la source (S) et le drain (D) sous l'application d'une tension drain source non nulle. Le transistor peut dans ce cas être assimilé en première approximation à un interrupteur fermé (cf. Figure I-5).

Le transistor MOS permet donc une modulation d'un flux d'électron à l'aide d'une commande présentant un coût énergétique très faible. Cela a favorisé son essor pour le traitement de gros volumes d'informations.

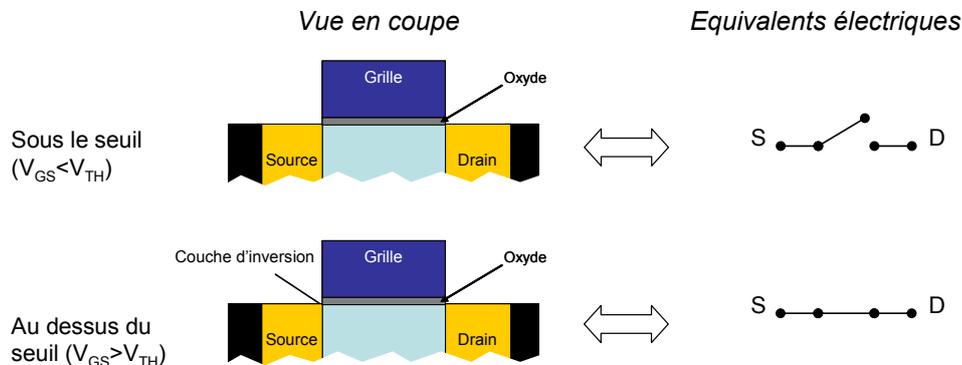


Figure I-5 : Principe de fonctionnement d'un transistor idéal.

b) Les effets parasites

Dégradation de la pente sous le seuil

En réalité, lorsque le transistor est à l'état bloqué, un courant de fuite circule entre la source et le drain. Ce courant contribue à l'augmentation de la puissance statique dissipée par les circuits. C'est donc un critère primordial pour les applications à faible consommation. Ce courant trouve son origine dans l'énergie d'activation thermique non nulle des porteurs dont un certain nombre arrive à franchir la barrière de potentiel dans le canal en régime d'inversion faible. Ce courant varie exponentiellement avec la tension de grille (Figure I-6). Un paramètre important à suivre est la pente sous le seuil qui traduit le facteur de proportionnalité entre la tension de grille et le courant de fuite en échelle logarithmique :

$$S = \frac{V_{th}}{\log(I_{OFF}) - \log(I_{th})} \quad (I-5)$$

Le paramètre S est appelé pente sous le seuil même s'il s'agit en fait de l'inverse de la pente. La valeur de S possède une limite inférieure de 60 mV/décade [5]. L'augmentation du dopage du canal contribue à fortement éloigner la pente sous le seuil de cette valeur idéale pour les dernières générations technologiques. A mêmes tensions de seuil, un

transistor présentant une plus forte valeur de pente sous le seuil aura ainsi un plus fort I_{OFF} impliquant une consommation statique plus importante.

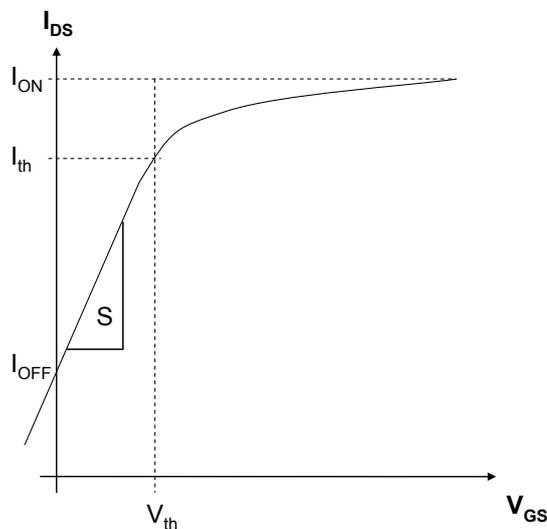


Figure I-6 : Caractéristique $I_{DS}=f(V_{GS})$ d'un transistor MOS en échelle logarithmique. La pente sous le seuil caractérise la décroissance exponentielle du courant en régime sous le seuil.

Effets Canaux Courts

L'effet canal court traduit la diminution de la tension de seuil par la diminution de la longueur de grille. Il est lié au fait que le potentiel électrostatique entre la source et le canal et entre le canal et le drain n'est pas abrupt mais graduel à cause des zones de charge d'espace le long des jonctions (cf. Figure I-7 A) [6]. Lorsque la distance entre la source et le drain atteint l'ordre de grandeur des zones de charge d'espace, la barrière de potentiel dans le canal chute et engendre une baisse de la tension de seuil appelé SCE (« Short Channel Effect ») comme décrit Figure I-7 B. La polarisation du drain influençant la profondeur de la zone de charge d'espace côté drain, l'effet canal court est accentué par la polarisation du drain (Figure I-7 C). Cet effet spécifique est appelé DIBL (« Drain Induced Barrier Lowering »). La dénomination « effets canaux courts » correspond aux deux effets combinés SCE et DIBL alors qu'« effet canal court » au singulier renvoie seulement à l'effet SCE.

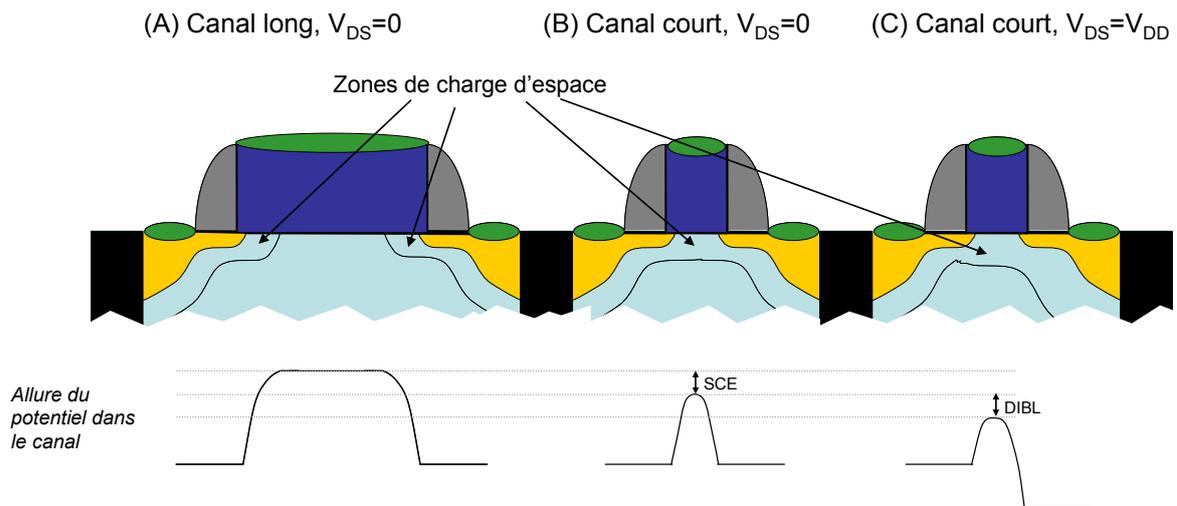


Figure I-7 : Principe de l'abaissement de la barrière de potentiel (bande de conduction) par la réduction de la longueur de grille du transistor. (A) La longueur du transistor est très grande devant la taille de la zone déplétée par les source et drain. (B) La longueur du transistor est inférieure à cette taille : c'est le régime d'effet de canal court (SCE). (C) Abaissement supplémentaire de la barrière par la tension de drain (DIBL).

Les conséquences des effets canaux courts sont une augmentation du courant de fuite liée à la diminution de tension de seuil, une baisse de fonctionnalité du transistor (différence de tension de seuil à V_{DS} faible et à V_{DS} fort) et également une disparité de tensions de seuil entre les transistors présentant différentes longueurs de grille, rendant plus complexe la tâche de conception des circuits.

Il existe un levier d'action immédiat permettant de diminuer les effets canaux courts qui consiste à augmenter le dopage afin de rendre plus abruptes les zones de charge d'espace et ainsi éviter leur recouvrement. Afin d'empêcher l'augmentation des tensions de seuil de tous les dispositifs (courts et longs), les implantations sont effectuées de manière locale près des jonctions grâce à des faisceaux d'ions inclinés permettant de former des poches.

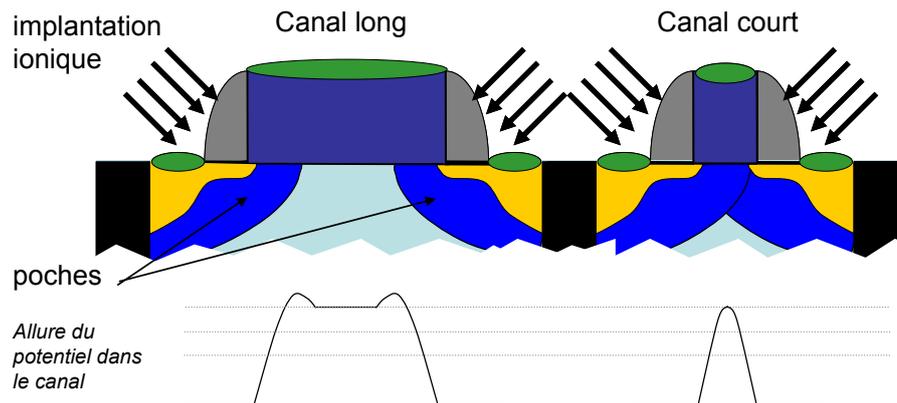


Figure I-8 : Illustration de l'impact du dopage par poche afin de maintenir une barrière de potentiel la plus constante possible entre canal long et canal court et ainsi diminuer les effets canaux courts.

Fuites de grille

Les épaisseurs d'oxyde de grille des dernières générations de circuit présentent des épaisseurs d'oxyde à peine supérieures à 10 Å, soit seulement quelques couches atomiques. A ces dimensions, le courant tunnel à travers le diélectrique de grille peut largement dépasser le courant OFF requis pour certaines applications (cf. Figure I-9). La Figure I-10 illustre l'impact du courant de grille ainsi que des effets canaux courts sur une caractéristique $I_{DS}-V_{GS}$.

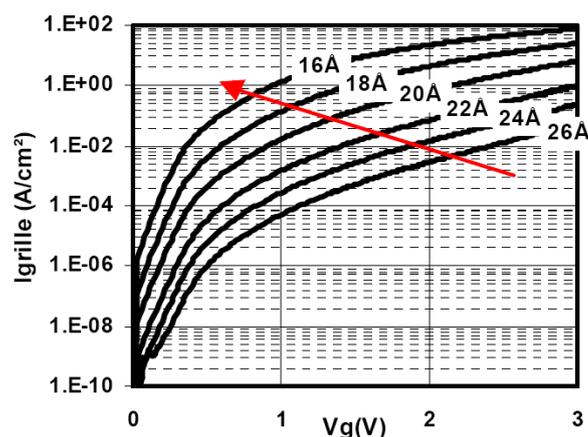


Figure I-9 : Courant de fuite de grille pour différentes épaisseurs d'oxyde (issu de [7]).

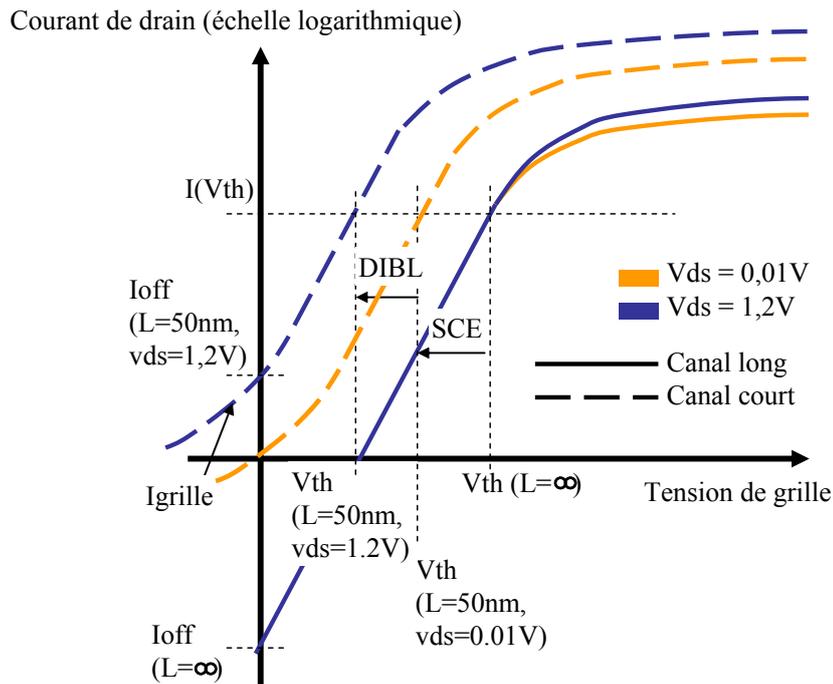


Figure I-10 : Impact des effets canaux courts et du courant de grille sur une caractéristique $I_{DS}-V_{GS}$ (d'après [6]).

Polydépétion

Le matériau de grille utilisé pour les dernières générations technologiques est du silicium poly cristallin (dit « polysilicium ») fortement dopé (N ou P en fonction du type de transistor MOS). Malgré le fait qu'il soit fortement dopé, le polysilicium présente une zone de déplétion vide de porteurs de charge [8]. Cette zone contribue à l'augmentation de l'épaisseur effective de l'oxyde de grille. Si cet effet est négligeable pour les anciennes générations technologiques, la profondeur de la zone de déplétion pour les dernières générations de circuit conduit à une augmentation de l'épaisseur de l'oxyde de grille de l'ordre de 30 à 40% (3 à 4Å environ) ce qui diminue d'autant la capacité de grille et donc l'effet de modulation de la charge dans le canal.

Confinement quantique des porteurs de charge

A cause de l'aspect non uniforme du puits de potentiel sous le canal en régime d'inversion forte (puits triangulaire), l'énergie des porteurs de charge est discrétisée (cf. Figure I-11). Une des conséquences de ce phénomène est que la probabilité maximale de présence des porteurs est décalée d'une certaine distance par rapport à l'interface Si-Oxyde, contribuant à l'existence d'une zone vide de porteurs appelée « darkspace » ou zone noire juste sous l'oxyde de grille [5]. Ce phénomène a été longtemps négligé à cause de son faible impact.

Cependant, pour les dernières générations de circuit, il contribue à augmenter fortement les épaisseurs effectives d'oxyde de grille et donc à diminuer les capacités de grille au même titre que la polydépétion. Des valeurs de 2 à 4 Å pour les électrons et de 6 à 8 Å pour les trous sont couramment reportées dans la littérature [5]

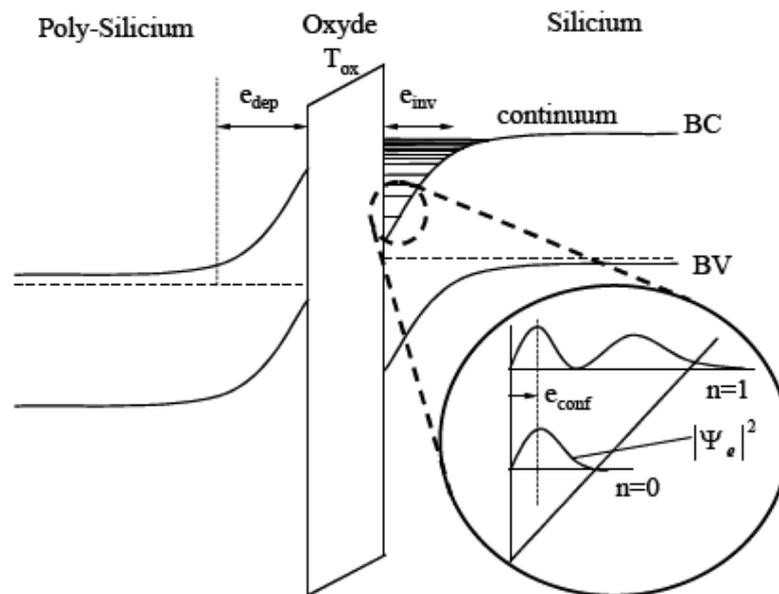


Figure I-11 : Courbure de bandes dans l'empilement de grille. Le régime de déplétion dans le poly-silicium augmente l'épaisseur effective de l'oxyde. Dans le canal, la courbure de bande provoque la quantification des niveaux d'énergies et donc la non uniformité de répartition de la charge près de l'interface oxyde-silicium : le pic de densité de probabilité de présence des électrons $|\Psi_e|^2$ est alors éloigné de e_{conf} de l'interface oxyde-silicium (d'après [6]).

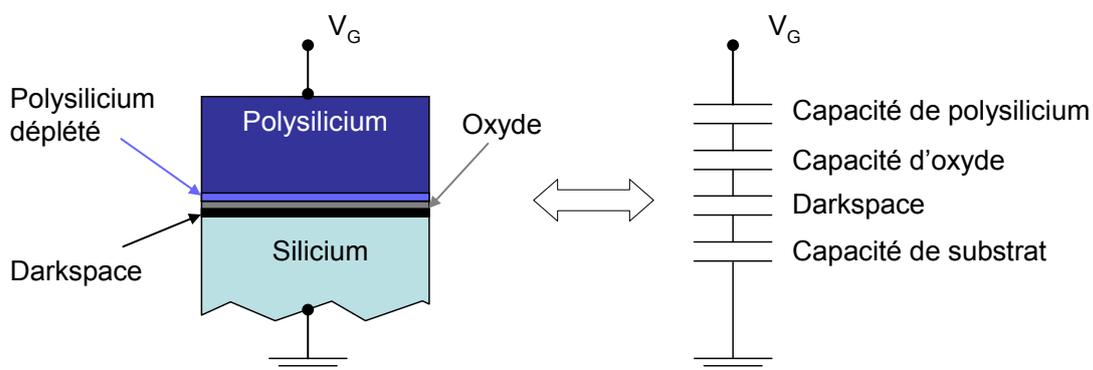


Figure I-12 : Illustration des différentes capacités parasites s'ajoutant en série à la capacité d'oxyde et conduisant ainsi à une diminution de la capacité totale de grille.

Dégradation de la mobilité des porteurs

La mobilité traduit la faculté des porteurs à se mouvoir dans le canal. Elle est inversement proportionnelle à la fréquence des interactions des porteurs avec le réseau cristallin [9]. Or, l'augmentation du dopage augmente de manière importante le nombre d'interactions coulombiennes (porteurs avec dopants ionisés) et par conséquent tend à diminuer la mobilité. La possibilité d'augmenter le dopage afin de diminuer les effets canaux courts est donc limitée.

Résistances séries

Les résistances séries du transistor MOS nanométrique ne sont plus du tout négligeables. Elles sont constituées des résistances de contact ainsi que de toutes les résistances d'accès le long de la source et du drain (cf. Figure I-13). Sans elles, le courant à l'état passant serait environ 15% supérieur pour les technologies 45nm (résistance série de l'ordre de $200\Omega\cdot\mu\text{m}$ pour des résistances de canal de $1600\Omega\cdot\mu\text{m}$). La diminution de la profondeur des extensions tend à augmenter fortement cette valeur, limitant là aussi un des leviers permettant de diminuer les effets canaux courts.

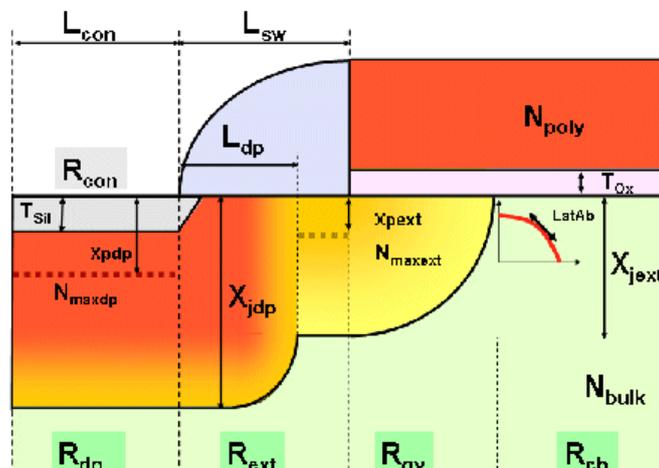


Figure I-13 : Les différentes résistances séries parasites : résistance de contact (R_{con}), résistance de source/drain (R_{dp}), résistance d'extension (R_{ext}) et résistance d'overlap (R_{ov}) (issu de [10]).

I.3.B Limitations à l'échelle du circuit

La fabrication d'un transistor isolé est complexe mais son intégration au sein des circuits l'est encore plus et pose aujourd'hui de plus en plus de problèmes.

Parmi ces problèmes figurent la gestion de la puissance dissipée par les circuits. Afin de diminuer celle-ci, des stratégies complexes de gestion des alimentations doivent être déployées, rendant la tâche de conception de circuits beaucoup plus compliquée. Garantir l'intégrité des signaux est également une thématique qui prend de plus en plus d'importance à mesure que les capacités de couplage entre les fils augmentent. La mise au point d'outils de conception sachant gérer des quantités de plus en plus importantes de portes (souvent plusieurs dizaines de millions) avec des temps de conception fixes est aussi une problématique récurrente du monde de la conception de circuit.

Il existe par ailleurs d'autres problématiques à l'échelle du circuit parmi lesquelles figurent la variabilité et l'augmentation des résistances d'interconnexion sur lesquelles la technologie a un impact non négligeable. Il a donc été choisi de traiter ces thèmes en priorité durant cette étude.

a) Variabilité

Origine

La variabilité des caractéristiques des transistors est une problématique qui prend de plus en plus d'importance avec chaque nouvelle technologie [11].

Sous certaines formes, la variabilité correspond à l'atteinte par la réduction des échelles de l'aspect discret de la matière. Par exemple, le dopage d'un dispositif 45nm est typiquement de l'ordre de $N_b=1 \times 10^{24}$ atomes/m³. Pour un dispositif présentant des dimensions classiques de règles de dessin ($L=40\text{nm}$, $W=120\text{nm}$) et une profondeur de déplétion de l'ordre de $T_{\text{dep}}=20\text{nm}$, le nombre de dopants définissant la tension de seuil est seulement de $N_b \cdot W \cdot L \cdot T_{\text{dep}} = 1 \cdot 10^{24} \times 40 \cdot 10^{-9} \times 120 \cdot 10^{-9} \times 20 \cdot 10^{-9} = 96$ dopants. A ces échelles la moindre variation du nombre de dopants va engendrer une variation de tension de seuil. Ces variations interviennent forcément car les techniques de dopage par implantation ionique utilisées pour doper les dispositifs induisent des variations stochastiques du nombre de dopants.

Une autre source de variabilité importante est la lithographie [12]. En effet, les techniques de photolithographie actuelles utilisent des sources de longueur d'onde bien supérieures aux plus petites dimensions du circuit (source de 193nm pour des résolutions de 45nm). La diffraction à ces échelles est donc très importante et le contrôle des dimensions particulièrement difficile.

Conséquences

Pour certains circuits tels que les mémoires SRAM où les transistors doivent être appareillés, la variabilité peut avoir des conséquences désastreuses. En effet, le point mémoire SRAM est constitué de deux inverseurs tête à tête avec chacun un transistor d'accès (Figure I-14). Les points de fonctionnement du circuit sont ceux correspondant au croisement des caractéristiques entrée-sortie des deux sous circuits inverseurs avec transistor d'accès. Ces points sont facilement visualisables sur une courbe en papillon (décrite Figure I-15). Lorsque les caractéristiques des transistors varient d'un point mémoire à l'autre, les caractéristiques entrée-sortie des inverseurs avec transistor d'accès varient également. A la limite, il est possible d'envisager des cas où les deux caractéristiques n'ont plus qu'un point de croisement stable (cf. Figure I-16). Dans ce cas, le point mémoire n'est plus capable de stocker qu'une seule valeur, autant dire le point mémoire ne fonctionne plus. Même si le pire cas décrit Figure I-16 a peu de chances de se produire, cela peut quand même avoir des impacts très importants compte tenu du nombre conséquent de cellules mémoires dans les circuits.

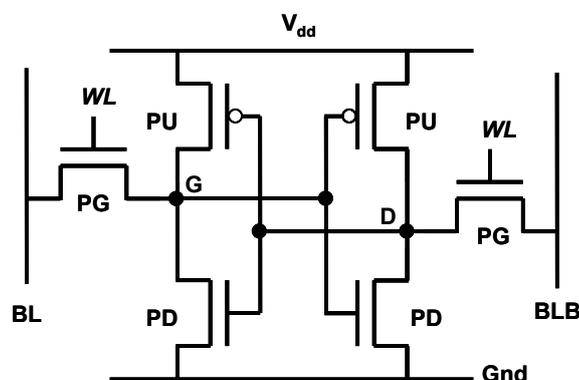


Figure I-14 : Schéma d'un point mémoire SRAM. PU est le « pull up », PD est le « pull down », PG est le « pass gate », WL est la « word line » et BL est la « bit line ».

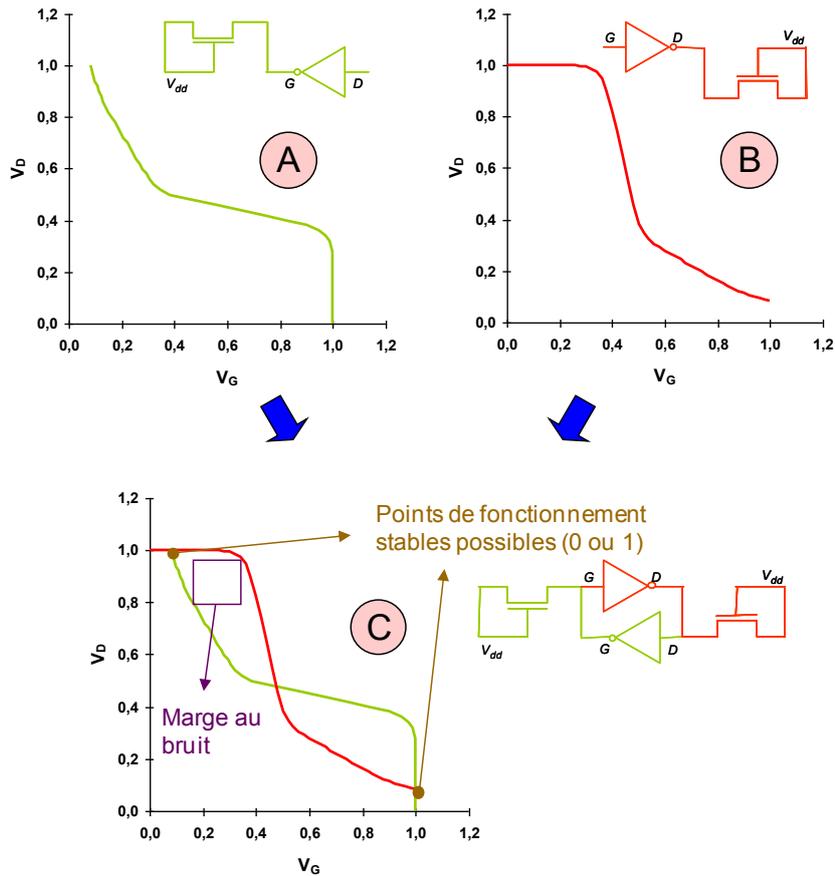


Figure I-15 : Description de la courbe en papillon (C). Celle-ci est composée des deux caractéristiques superposées des deux inverseurs et de leurs transistors d'accès (A et B) de la mémoire SRAM en condition de lecture (« Bit Line » et « Word Line » à 1). La marge au bruit est définie comme le côté du plus petit carré insérable dans les deux ouvertures de la courbe.

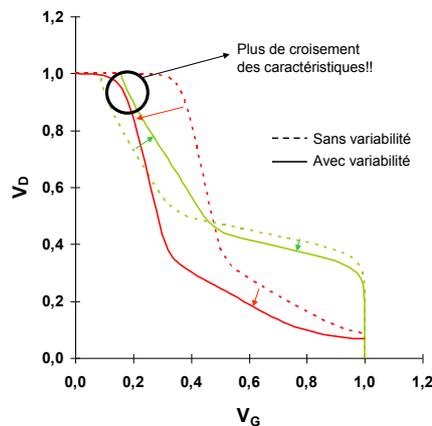


Figure I-16 : Impact de la variabilité des caractéristiques des transistors sur une courbe en papillon. A la limite, un des points de fonctionnement peut disparaître rendant le point mémoire non fonctionnel.

b) Augmentation des résistances d'interconnexion

La réduction de la largeur des lignes de métallisation devrait entraîner une augmentation des résistances d'interconnexion très importante pour les futures technologies (cf. figure ci-dessous).

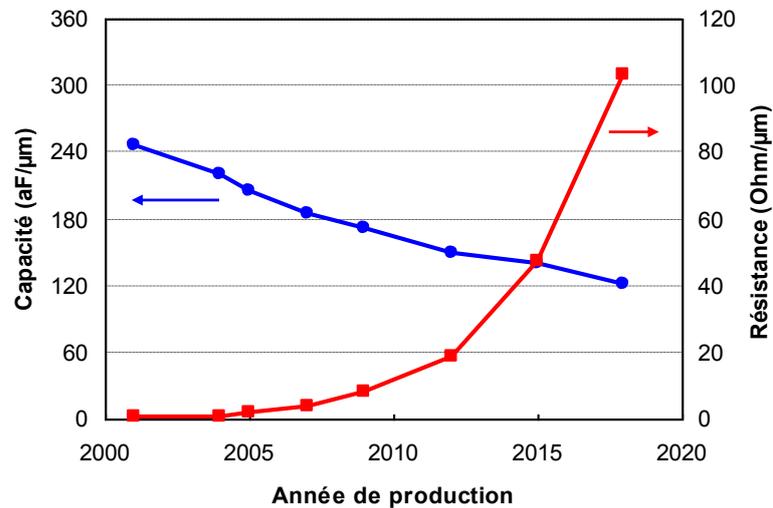


Figure I-17 : Evolution des capacités et résistances de ligne calculée à partir du logiciel HF2D [13] et des données de l'ITRS [14]. Une augmentation quasi exponentielle des résistances est attendue pour les années à venir.

Plusieurs phénomènes expliquent cette augmentation :

Aspect néfaste de la réduction des échelles :

La résistance d'un barreau de métal peut être exprimée sous la forme suivante :

$$R = \frac{\rho L}{S}, \quad (\text{I-6})$$

avec ρ la résistivité du métal, L la longueur du barreau et S sa section. S peut être mis sous la forme $S=hl$ avec h la hauteur du métal et l sa largeur. Réduire les échelles implique de devoir réduire toutes les dimensions d'un même facteur. En effet, à cause des problèmes de capacités de couplage entre métaux, le rapport h/l doit être maintenu à peu près constant et donc la dimension h doit être réduite de génération en génération. Avec cette hypothèse, en réduisant toutes les dimensions d'un facteur k , l'expression de la résistance devient :

$$R_{n+1} = \frac{\rho L}{k S} = \frac{R_n}{k}. \quad (\text{I-7})$$

Réduire les échelles d'un facteur k implique donc, à résistivité de métal constante, de diviser la résistance du métal par un facteur k . C'est un aspect non bénéfique de la réduction des échelles. Cet aspect empire même si l'on considère des lignes de longueur fixe. En effet, les circuits ont des dimensions plus ou moins constantes depuis plusieurs générations. Les plus longues lignes ne devraient donc pas voir leur longueur diminuée. En considérant ainsi que les longueurs des lignes ne diminuent pas d'une génération à l'autre, le facteur de dégradation de la résistance vaut alors $1/k^2$, c'est-à-dire la résistance double à chaque génération technologique ($k=0,7$ entre chaque génération).

Augmentation de la résistivité du métal :

Il y a deux raisons à l'augmentation de la résistivité ρ des métaux. La première concerne l'augmentation de la diffusion des électrons à cause de la réduction de l'épaisseur de la ligne et de la diminution de la taille des poly cristaux [15]. En effet, les dimensions atteintes par les épaisseurs de métaux sont si petites qu'elles deviennent de l'ordre de grandeur du libre parcours moyen des électrons. L'électron subit ainsi un nombre de collisions de plus en plus important pour pouvoir poursuivre sa route dans le barreau de métal. Par ailleurs, les poly cristaux de métal ont tendance à diminuer de taille à mesure que les épaisseurs diminuent créant ainsi une source de diffusion supplémentaire.

La deuxième raison de l'augmentation de la résistivité du métal est liée à la non réduction de taille de la barrière de diffusion du cuivre dans les interconnexions [15]. En effet, afin d'empêcher la diffusion du cuivre dans le substrat, une barrière de tantale (Ta) est insérée entre le diélectrique et le métal. Cette barrière possède une épaisseur qu'il est très difficile de réduire d'une génération à l'autre. La section effective du barreau de métal est donc plus faible que celle apparente et ce phénomène s'accroît de génération en génération, accroissant ainsi les effets de diffusion des électrons.

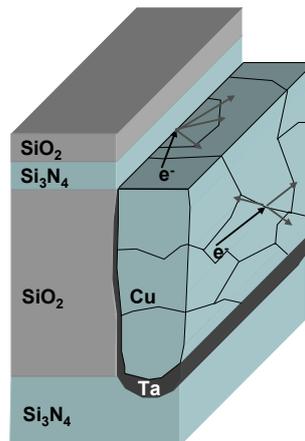


Figure I-18 : Illustration des phénomènes conduisant à l’augmentation de la résistivité des lignes d’interconnexion. La barrière de diffusion en tantale (Ta) empêchant la diffusion du cuivre dans le reste du circuit réduit la section effective du métal et accentue les phénomènes de diffusion d’électron à la fois au niveau des joints de grain mais également au niveau des frontières avec le diélectrique.

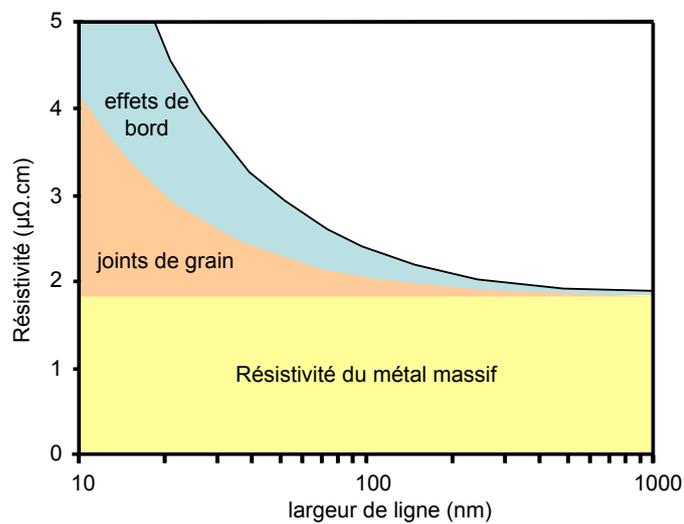


Figure I-19 : Différentes contributions à l’augmentation de la résistivité des lignes d’interconnexion en cuivre (d’après [14]).

I.4 Technologies émergentes

I.4.A Technologies « Front End » (composants actifs)

a) « Boosters » technologiques

Grille métallique et diélectriques haute permittivité

Pour contrer les effets de diminution de la capacité de grille par déplétion du polysilicium (cf. §I.3.Ab) p.35), l'utilisation de grilles métalliques pour les futures technologies semble être requise. En effet, les grilles métalliques ont l'avantage de ne pas présenter de zones de déplétion, ce qui constitue une vraie solution à cet effet parasite. Cependant, l'intégration de nouveaux matériaux de grille, dont il faut savoir contrôler parfaitement le travail de sortie, n'est pas sans poser d'importantes difficultés, et ce en considérant les problèmes de compatibilité avec des oxydes de grille à forte permittivité.

En effet, l'augmentation continue des courants de fuite de grille à mesure que les épaisseurs diminuent poussent aujourd'hui les technologues à se poser la question de la pertinence de l'utilisation de l'oxyde de silicium qui, s'il a permis un essor formidable des technologies silicium en partie grâce à la simplicité du procédé d'oxydation, présente l'inconvénient majeur de ne pas être le meilleur des isolants (faible constante diélectrique $k=3,9$). Ainsi d'autres matériaux, parmi lesquels figure l'oxyde d'Hafnium HfO_2 , présentent des facultés d'isolation électrique bien supérieures. La constante diélectrique du HfO_2 est ainsi de l'ordre de 20 à 25 [16], soit dix fois celle de l'oxyde traditionnel SiO_2 . Cette propriété implique qu'il est possible, en théorie, d'obtenir des capacités aussi importantes avec des épaisseurs de matériaux dix fois supérieures, rendant alors beaucoup moins probable le passage d'électrons par effet tunnel à travers l'isolant.

Dans l'optique de minimiser l'effort d'intégration, deux changements aussi importants de l'empilement de grille (grille métalliques et diélectriques haute permittivité) ne peuvent s'envisager que simultanément. C'est pourquoi, l'essentiel des recherches porte directement sur l'intégration du couple grille métallique et diélectrique haute permittivité.

L'intégration de tels matériaux est complexe d'autant plus qu'il existe des stratégies d'intégration très différentes (dépôt de 2 grilles métalliques, intégration d'une grille « mid gap », siliciuration totale du polysilicium...). Il faut de plus trouver des matériaux aux travaux de sortie symétriques de manière à avoir des tensions de seuil ajustées entre

NMOS et PMOS dans une gamme permettant d'avoir des tensions de seuil raisonnables, ce qui dépend fortement du type d'application voulue (basse consommation ou haute performance).

Malgré les difficultés d'intégration mentionnées, plusieurs sociétés ont annoncé récemment l'intégration fructueuse de grilles métalliques et d'oxyde haute permittivité pour les générations 45nm [17] [18].

Techniques d'augmentation de la mobilité des porteurs

Pour pallier la réduction de la mobilité des porteurs de charge dans les canaux à cause de l'augmentation des niveaux de dopage requis pour contrer les effets canaux courts, différentes techniques sont utilisées ou envisagées afin de créer des contraintes mécaniques dans le canal. Ces contraintes peuvent sous certaines conditions permettre une modification de la structure de bande du silicium et ainsi aboutir à de meilleures propriétés de transport des porteurs de charge [9].

Jonctions ultra fines

Des jonctions ultra fines sont utilisées afin de diminuer les effets canaux courts. En effet, le couplage électrostatique entre la source et le drain (qui détermine l'intensité des effets canaux courts) est directement proportionnel à la profondeur de jonction (cf. Figure I-20). Cela requiert l'utilisation de techniques d'activation des dopants innovantes telles que les recuits laser [19] afin d'empêcher une trop grande diffusion des dopants. Cependant, la diminution de la profondeur des jonctions ne doit pas être faite au détriment de l'augmentation des résistances série. Il faut donc trouver le meilleur compromis résistance d'accès/effets canaux courts.

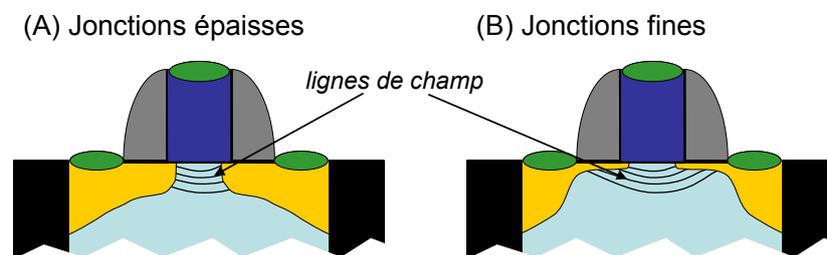


Figure I-20 : Lignes de champ dans le cas de jonctions épaisses (A) et dans le cas de jonctions fines (B). L'utilisation de jonctions fines permet d'augmenter la longueur moyenne des lignes de champ et donc de réduire le couplage entre source et drain et ainsi diminuer les effets canaux courts.

b) Nouvelles architectures de transistors

Les effets canaux courts constituent le principal frein à la réduction d'échelle des dispositifs. Jusqu'à maintenant des solutions temporaires (augmentation du dopage et diminution des profondeurs de jonction) ont permis de maintenir les effets canaux courts à des niveaux acceptables. Cependant, il est nécessaire de trouver des solutions viables à plus long terme. Ainsi il est envisagé de changer les architectures traditionnelles de dispositifs de manière à passer sur des architectures intrinsèquement moins sensibles aux effets canaux courts.

SOI (« Silicon On Insulator »)

Le SOI (« Silicon On Insulator » ou silicium sur isolant) est une technique qui consiste à utiliser des substrats présentant une fine couche d'isolant à proximité de la surface. Il existe deux types d'architectures SOI : PDSOI (« Partially Depleted SOI » ou SOI partiellement déplété) et FDSOI (« Fully Depleted SOI » ou SOI complètement déplété). Dans le premier cas, la profondeur de déplétion du dispositif est inférieure à l'épaisseur du film de silicium. Cette technologie est actuellement utilisée par plusieurs sociétés, principalement pour les applications haute performance. L'avantage du PDSOI est tout d'abord d'isoler parfaitement les transistors les uns des autres évitant ainsi tous les phénomènes de transistors bipolaires parasites connus sous le nom de « latch up ». En effet, la présence de ces transistors parasites pour les technologies conventionnelles impose d'espacer les transistors suffisamment les uns des autres réduisant ainsi la densité d'intégration. Le deuxième avantage est la présence d'une zone neutre dans le dispositif où vont venir se stocker les charges créées par ionisation par impact. Ces charges vont contribuer à diminuer la tension de seuil à l'état passant et donc à augmenter le courant. Le SOI permet également de réduire les capacités de jonctions lorsque celles-ci s'étendent en partie jusqu'à l'isolant et donc d'améliorer les vitesses de commutation.

Le vrai intérêt du SOI à long terme réside dans le FDSOI qui possède la faculté de réduire les effets canaux courts. En effet, grâce à la couche isolante qui limite l'extension verticale des jonctions, le couplage source drain est réduit et l'intégrité électrostatique du canal est améliorée. Il faut alors utiliser des substrats aux épaisseurs de silicium inférieures à la profondeur de déplétion des dispositifs. L'utilisation de technologies FDSOI implique de savoir surmonter plusieurs difficultés :

1. Augmentation des résistances séries : l'utilisation de films minces implique inexorablement l'augmentation des résistances séries qu'il faut savoir compenser par des techniques d'élévation source/drain par épitaxie notamment.
2. L'épaisseur du film a un impact direct sur les variations de tension de seuil des dispositifs. Or, les épaisseurs actuelles requises sont de l'ordre de 10nm soit une trentaine de couches atomiques seulement. Les technologies pour fabriquer ces transistors utilisent la plupart du temps des procédés de planarisation mécano-chimiques qui ne semblent pas encore en mesure d'aboutir à des épaisseurs de films contrôlées à l'atome près.
3. Pour les films les plus minces, la charge de déplétion est très inférieure à celle d'un transistor conventionnel impliquant des tensions de seuil naturellement très faibles qu'il faut contrôler par la modulation du travail de sortie des gilles.

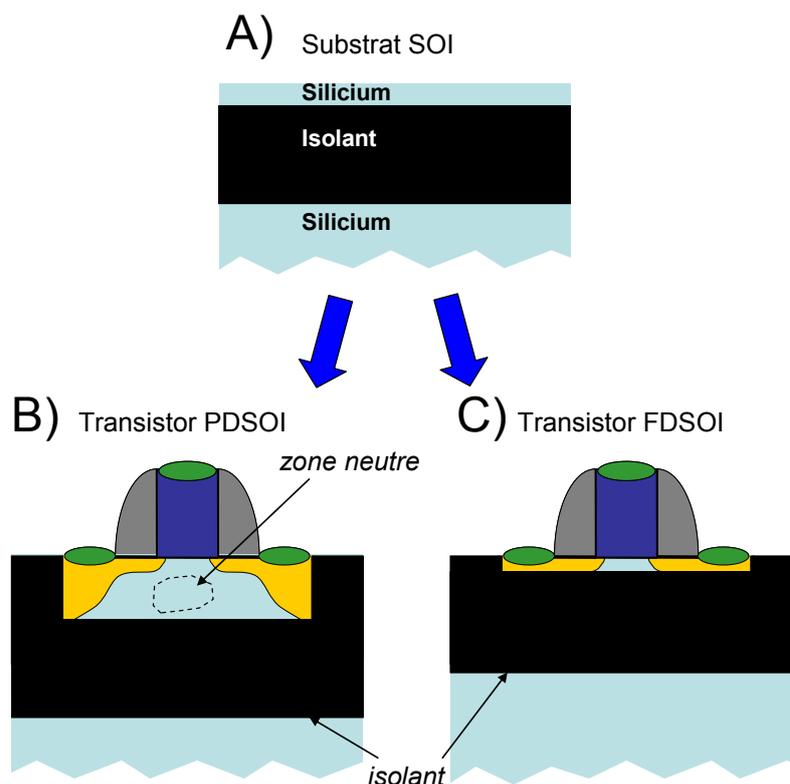


Figure I-21 : Vue en coupe (A) d'une plaque SOI, (B) d'un transistor SOI partiellement déplété (PDSOI), (C) d'un transistor SOI complètement déplété (FDSOI).

SON (« Silicon On Nothing »)

Le transistor SON (« Silicon On Nothing ») [20] est globalement similaire à l'architecture SOI. La différence principale se situe au niveau de l'isolant qui ne recouvre pas les jonctions (cf. Figure I-22). Cette architecture présente de multiples avantages. Tout d'abord, la technique de fabrication du SON aboutit à des épaisseurs de films de silicium extrêmement bien contrôlées. En effet, le procédé d'élaboration est basé sur l'élimination d'une couche sacrificielle de silicium-germanium (SiGe) définie par épitaxie (cf. Figure I-23) de manière très précise. En outre, le SON ne nécessite pas d'utiliser des substrats spécifiques comme dans le cas du SOI. Enfin, il peut être aisément co-intégré avec des dispositifs conventionnels sur silicium massif (« bulk ») au sein de la même puce.

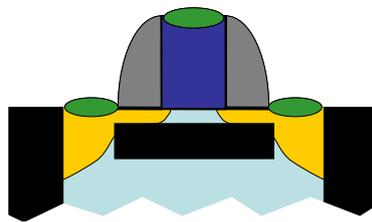


Figure I-22 : Vue en coupe d'un transistor SON.

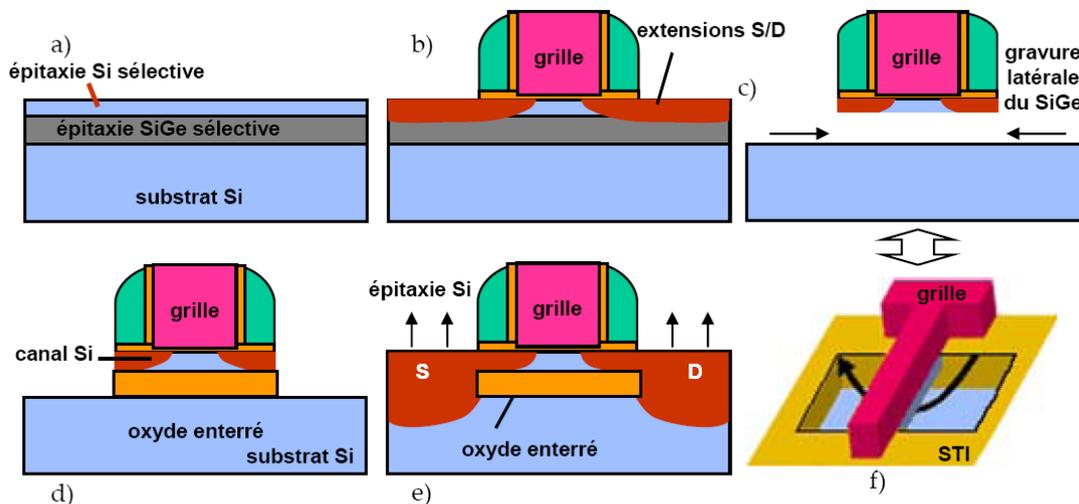


Figure I-23 : Description du procédé de fabrication SON (issu de [21]).

DG

La double grille (DG) [22] est l'architecture connue permettant de limiter au mieux les effets canaux courts. Elle est similaire à l'architecture SON à l'exception de l'oxyde enterré qui est remplacé par une autre grille isolée du canal par un oxyde mince. La grille

enterrée peut-être connectée à la grille supérieure ou bien indépendante. Dans ce dernier cas, il s'agira d'IDG (« Independent Double Gate »).

Les avantages de l'architecture double grille sont un contrôle optimal des effets canaux courts grâce à la grille enterrée et la présence de deux interfaces de conduction aboutissant au doublement du courant à l'état passant à même largeur dessinée de transistor.

L'inconvénient majeur réside dans la complication des procédés de fabrication pouvant potentiellement nuire aux rendements et détériorer la variabilité des caractéristiques des dispositifs.

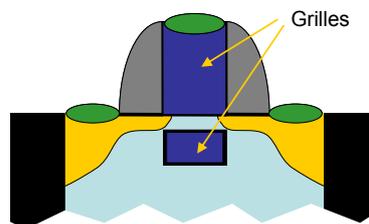


Figure I-24 : Vue en coupe d'un transistor double grille.

FinFET

Le FinFET (« Fin Field Effect Transistor » ou transistor à effet de champ à ailettes [23]) est l'équivalent du DG à la différence près de son orientation par rapport au substrat. En effet, la largeur « W » du transistor est définie dans sa hauteur, qui est fixée par l'épaisseur du film de silicium utilisée. Pour faire varier le « W » du transistor, il faut alors utiliser plusieurs transistors en parallèle ce qui implique de devoir complètement bouleverser les règles de conception traditionnelles utilisées dans le cas de transistors conventionnels planaires.

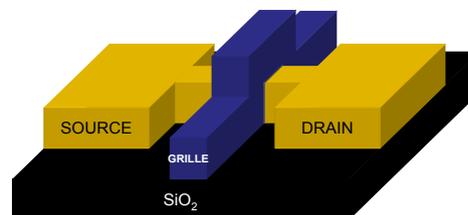


Figure I-25 : Vue 3D d'un transistor FinFET.

I.4.B Technologies « Back End » (interconnexions)

Techniques de diminution de la permittivité des isolants des lignes d'interconnexion

Le délai d'une ligne d'interconnexion est fixé en première approximation par le produit RC (résistance capacité) de la ligne. L'absence de solution connue pour limiter l'augmentation des résistances pour les années à venir (cf. I.3.Bb) implique que les capacités de ligne sont le seul levier pour jouer sur les délais. En agissant sur la nature des matériaux isolants, il est possible de diminuer ces capacités. Ainsi, en utilisant des matériaux poreux, il est possible de diminuer la constante diélectrique autour de 2,5. Cependant, il n'est pas possible d'atteindre des valeurs inférieures à 1 (constante du vide). Ce levier d'action est donc limité par nature.

I.4.C Au niveau des circuits

a) DFM

Le DFM (« Design For Manufacturing » ou conception orientée fabrication) est une méthodologie de conception des circuits permettant de maximiser les rendements. Certaines règles de cette méthodologie ont trait à la variabilité et ont pour but d'éviter les cas où celle-ci est maximale [24].

b) Circuits de topologie régulière

L'utilisation d'un jeu réduit de règles de dessin des circuits permet de limiter le nombre de topologies sur silicium et donc de faciliter leur contrôle. C'est le principe des mémoires dans lesquelles les éléments de stockage de l'information sont tous identiques. Ce principe s'applique également aux autres éléments des circuits, tels que les cellules standard élémentaires (NAND, NOR, MUX...).

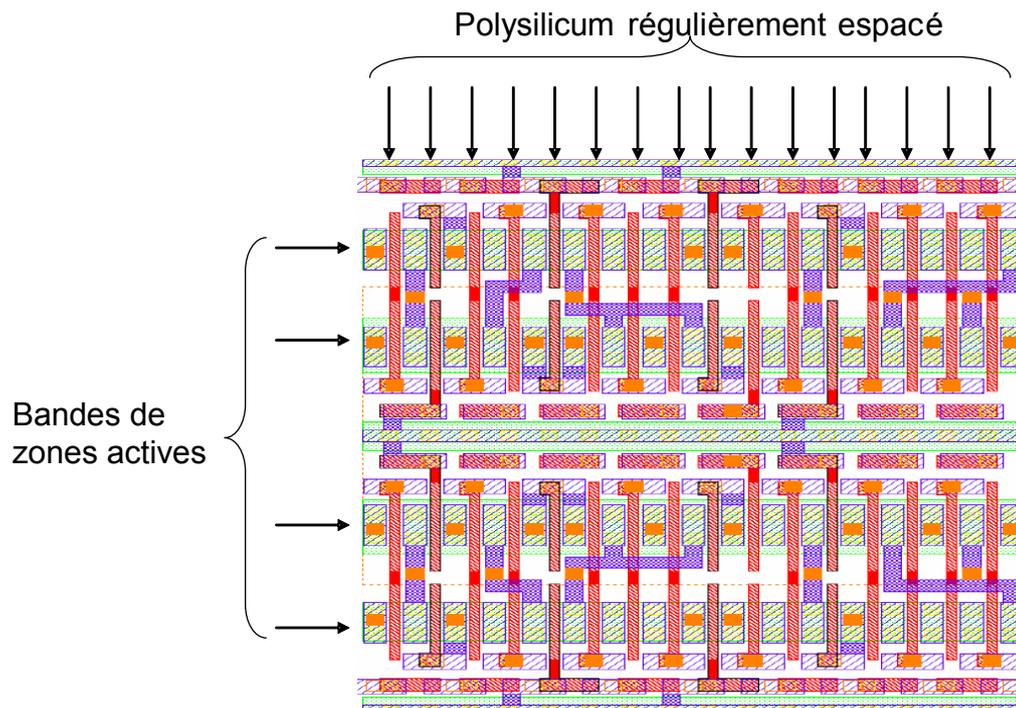


Figure I-26 : Illustration d'un masque régulier : les bandes de zones actives sont infinies et les ligne de polysilicium régulièrement espacées.

c) Conception à base de règle statistiques

Les règles de conception classiques des circuits s'appuient sur des modèles discrets de la technologie qui représentent le plus souvent les cas extrêmes : pire cas de vitesse des transistors, meilleur cas de vitesse des transistors. L'application de ces règles sous-entend que tous les transistors du circuit ont le même comportement : le concepteur vérifie par exemple que le circuit fonctionne dans le cas où tous les transistors sont lents (et respectivement dans le cas où tous les transistors sont rapides). Cette représentation est fautive pour les technologies récentes pour lesquelles on peut trouver dans le même circuit des transistors rapides et des transistors lents. De nouvelles méthodologies de conception permettent de prendre en compte ces variations en s'appuyant sur une représentation statistique de la technologie. Leur application change les critères de succès de la conception d'un circuit. Le critère de vitesse du circuit n'est plus binaire (soit en dessous soit au-dessus de la spécification), mais une probabilité (95% de probabilité d'être en dessous de la spécification par exemple).

I.5 Conclusion

Dans ce chapitre, les critères de réduction des échelles ont été présentés. Ceux-ci impliquent des investissements de plus en plus importants obligeant les entreprises à minimiser les risques liés à de tels investissements en effectuant des évaluations en amont du développement des technologies. Les principaux phénomènes parasites qui apparaissent avec la réduction des dimensions des transistors ont été introduits : effets canaux courts, réduction de la mobilité, fuites de grilles, variabilité, augmentation des résistances d'interconnexion. Les remèdes et technologies innovantes ont également été abordés : grille métallique, diélectrique « high-K », nouvelles architectures de transistors (SOI, SON, DG), DFM, circuits à topologies régulières, diélectriques « low-K » faible permittivité pour les interconnexions.

Choisir la meilleure technologie pour les années à venir implique de savoir définir des critères d'évaluation de ces technologies innovantes. L'objet du prochain chapitre est ainsi d'introduire plusieurs critères d'évaluation technologique de niveau circuit.

Chapitre II De la technologie au circuit : kit de conception et critères d'évaluation des futures technologies

Il existe principalement deux types de flots de conception : le flot « Full Custom » et le flot digital standardisé. Le premier est basé sur le dessin manuel des masques à l'aide d'un manuel (DRM) contenant toutes les règles de conception alors que l'autre travaille à partir de l'implémentation automatique de portes élémentaires caractérisées en terme de comportement électrique.

Exécuter un flot de manière prédictive est nécessaire pour obtenir des critères d'évaluations des futures technologies. Au niveau « full custom » sont associés les critères relatifs à l'oscillateur en anneau, à la robustesse à la variabilité d'un mémoire SRAM, aux délais le long de lignes d'interconnexion seules ou avec répéteurs. Au niveau du flot digital sont reliés la densité, le nombre de violation de contraintes de délai ainsi que la distribution des longueurs d'interconnexion d'un bloc placé routé.

II.1 Introduction

Le chapitre précédent a mis en évidence les nombreuses solutions technologiques envisagées pour pallier les différents phénomènes parasites qui contrarient les réductions des dimensions. Choisir la meilleure technologie pour les circuits de demain implique de savoir définir des critères d'évaluation de niveau circuit. Ceci est l'objet de ce chapitre.

La conception de circuits nécessite le déploiement d'un grand nombre d'outils de conception assistée par ordinateur (CAO). La succession des étapes nécessaire à l'élaboration des masques est appelée flot de conception. Il en existe principalement deux types : flot « Full Custom » et flot digital.

Après avoir présenté ces deux types de conception de circuit, différents critères d'évaluation technologiques sont abordés. Les critères proches du transistor élémentaire sont tout d'abord introduits puis des éléments de comparaison faisant intervenir les

interconnexions sont présentés. Finalement, les aspects relatifs à une évaluation de l'implémentation physique d'un bloc placé routé sont développés.

II.2 Flot et kit de conception

Le flot de conception d'un circuit est l'ensemble des étapes nécessaires à la conception de ce circuit, c'est-à-dire l'ensemble des étapes permettant d'aboutir à un fichier de données qui peut être envoyé à l'équipe d'assemblage des réticules avant les corrections des aberrations optiques (OPC) et l'envoi au fabricant de masque.

Il existe deux types de flot de conception : un flot « full custom » où chaque élément du circuit est dessiné manuellement et un flot digital standardisé où la plupart des étapes de conception sont automatisées. Le flot « full custom » est employé partout où le flot digital ne peut l'être (circuits analogiques, circuits de test, structures de test, circuits spécifiques haute performance, etc.)

Le kit de conception regroupe quant à lui l'ensemble des outils, avec leurs paramétrages associés, nécessaires au déroulement des flots de conception.

II.2.A Le flot « full custom »

a) Dessin des masques selon le manuel des règles de dessin

Dans le cas du flot « full custom », le concepteur dessine un à un tous les éléments du circuit. Il lui faut ainsi définir toutes les formes géométriques des masques pour la zone active, les caissons, les lignes de polysilicium, les lignes de métallisation, les contacts, les vias, etc. Il peut se servir de bibliothèques déjà existantes ou bien tout dessiner lui-même. La seule vraie contrainte qu'il a, est de respecter les règles de dessin de la technologie. Celles-ci sont consignées dans un manuel appelé DRM (« Design Rules Manual »). Les règles concernent aussi bien des données de dimensionnement des dispositifs (surface minimale des contacts, des zones actives,...), d'espacement géométrique (distance minimale entre 2 lignes de polysilicium ou 2 métaux, distance minimale poly-contact, etc.), que des données de densité (densité minimale et maximale de métal, etc.). Des logiciels permettent de contrôler automatiquement la conformité des circuits avec le DRM (logiciels DRC pour « Design Rules Check »).

b) Schématisation électrique

Lors d'un flot « full custom », le concepteur part d'un schéma électrique qui est transformé sous forme d'un ensemble de masques. Afin de l'assister dans sa tâche, des logiciels permettent de contrôler à tous niveaux l'homologie schémas/masques. Cette action est appelée LVS (« Layout Versus Schematic »).

c) Simulations électriques

Simulation du schéma (pre-layout)

Au niveau « full custom », les simulations des performances électriques sont effectuées à l'aide de simulateurs de type SPICE, dont le principe réside dans l'utilisation d'algorithmes de Newton-Raphson. Ces algorithmes permettent, par itérations successives, de trouver les valeurs de potentiel et de courant aux différents nœuds du circuit en résolvant les lois de Kirchhoff. En effectuant certaines approximations (approximation quasi-statique, linéarisation locale autour des points de polarisation), le simulateur peut prédire le comportement temporel du circuit. Les simulateurs de type SPICE se basent sur des modèles analytiques de dispositifs. De ce fait, pour ajuster les modèles à la technologie, les simulateurs proposent un ensemble d'équations paramétriques dont les paramètres sont ajustés de manière spécifique à la technologie considérée. Le jeu de paramètres décrivant le comportement d'un dispositif est appelé carte modèle.

Simulation en tenant compte des effets parasites (post-layout)

Une fois le circuit dessiné, le concepteur doit simuler les performances électriques en tenant compte des effets parasites additionnels par rapport au seul schéma électrique. En effet, les résistances et capacités d'interconnexions, de contacts, de vias et le dimensionnement effectif des transistors au niveau des zones actives (surface et périmètre des jonctions) peuvent avoir un impact considérable sur les performances du circuit. Ces résistances et capacités augmentent typiquement le délai d'un circuit aussi simple qu'un oscillateur en anneau de 30 à 40%. Il est ainsi nécessaire de procéder à une extraction de toutes ces capacités et résistances afin de les ajouter au schéma électrique. L'outil permettant d'effectuer cette action est un extracteur post-layout. Un fichier, appelé « nxtgrd » dans le cas de l'extracteur Star-RCXT [25], permet de paramétrer l'outil au niveau des interconnexions grâce à la description des propriétés physiques des empilements diélectriques/métaux.

d) Vérifications

Avant d’être livré pour fabrication, le circuit doit impérativement avoir passé avec succès les étapes DRC et LVS. En effet, compte tenu des coûts de fabrication des circuits, toute erreur de conception doit être exclue. Le schéma de la Figure II-1 illustre le principe du flot de conception « Full Custom »

II.2.B Le flot digital standardisé : synthèse et placement routage

a) Description du circuit au niveau RTL

Le point d’entrée d’un flot digital est la description du circuit au niveau RTL (« Register Transfer Level »). Dans cette description, l’ensemble du circuit est défini au niveau des registres (Figure II-2). Les registres sont des bascules reliées à l’horloge qui permettent de séquencer le traitement des informations. A chaque front d’horloge, la donnée en entrée de la bascule est transmise en sortie afin de poursuivre son traitement. Le langage RTL ne donne pas la description détaillée de la logique entre les registres. Il se contente d’une information synthétique sous forme d’équations combinatoires.

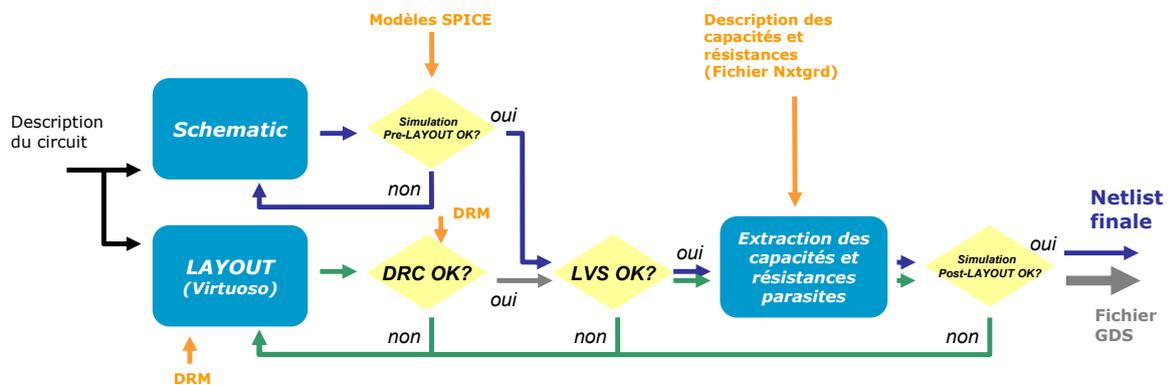


Figure II-1 : Schéma d’un flot de conception de circuit de type « full custom ».

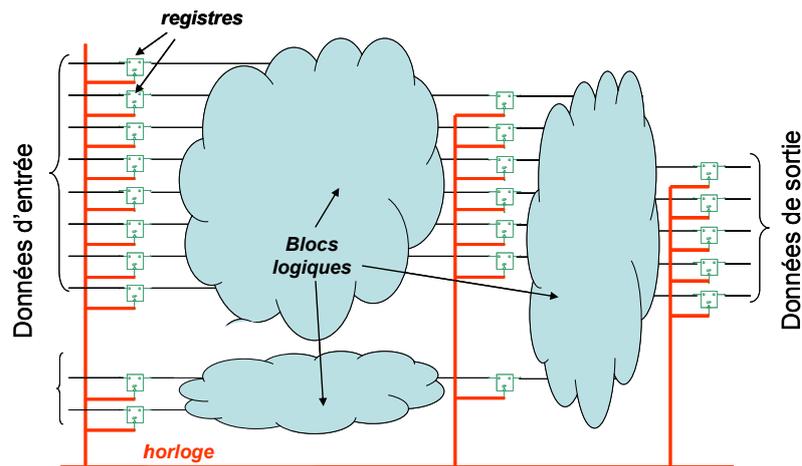


Figure II-2 : Description d'un circuit au niveau RTL.

b) Librairie de cellules standard

Définition

Les cellules standard (ou « standard cell ») sont des portes logiques possédant une fonctionnalité élémentaire (NAND, NOR, MUX, etc.). Les cellules standard sont conçues de manière à pouvoir être aboutées les unes aux autres. De cette façon la connexion des alimentations est immédiate. Les cellules standard sont ainsi des briques que le concepteur (ou l'outil) assemble pour construire le circuit.

Les différentes vues

Il existe différents types de «vues» des cellules standard en fonction du type d'informations auxquelles le concepteur veut avoir accès (cf. Figure II-3).

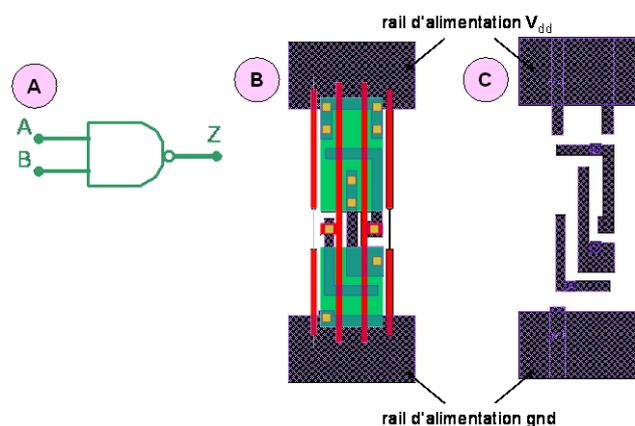


Figure II-3 : Exemple de différentes vues d'une cellule standard de type NAND2. A : vue schématique. B : vue masque. C : vue abstraite (seulement les masques correspondant aux différents niveaux de métaux et l'emplacement des connexions).

Caractérisation des délais

Des matrices de délais permettent de caractériser la vitesse des cellules. Celles-ci sont obtenues à l'aide de simulations électriques. Elles consistent en des tables de délais à deux dimensions (pente du signal d'entrée et capacité en sortie). En caractérisant ainsi la vitesse de chaque cellule et en connaissant leur capacité d'entrée ainsi que la capacité des fils, il est possible de calculer le délai de propagation d'un signal le long de n'importe quel chemin. Ce délai est dit statique et ne prend pas en compte les couplages entre signaux.

c) Synthèse logique

La synthèse a pour but de fournir une description de la fonctionnalité du circuit à partir d'une librairie de cellules standard. Le fichier obtenu en sortie est une netlist Verilog ou VHDL (« Very High Speed Integrated Circuit Hardware Description Language») de niveau porte. La synthèse travaille à partir des données RTL, des données schématiques des cellules, des données de délai des cellules ainsi que des contraintes de délai du circuit. Elle permet d'obtenir une première estimation des délais du circuit qui ne prend pas en compte les délais d'interconnexion. Elle peut cependant les estimer grâce à des modèles prédictifs (« wire load model »).

d) Placement Routage du circuit

L'étape de placement routage consiste comme son nom l'indique à placer les cellules dans un espace réservé (« floor plan ») puis à les router afin de réaliser les connexions décrites dans le fichier verilog ou VHDL. De nombreuses itérations de placement permettent d'améliorer la performance du circuit (délai ou puissance selon les cas). Des algorithmes perfectionnés permettent également d'échanger certaines cellules (changement de la sortance, décomposition en plusieurs autres cellules de la librairie) et de placer des cellules tampon (« buffer ») de manière à remplir les contraintes fixées en terme de fréquence d'horloge ou de consommation.

La Figure II-4 illustre les principales étapes d'un flot digital à partir d'un circuit élémentaire. La Figure II-5 décrit quant à elle les éléments nécessaires à la configuration d'un tel flot. La Figure II-6 est une vue de type masques d'un circuit complexe obtenu grâce à un flot digital (5000 transistors).

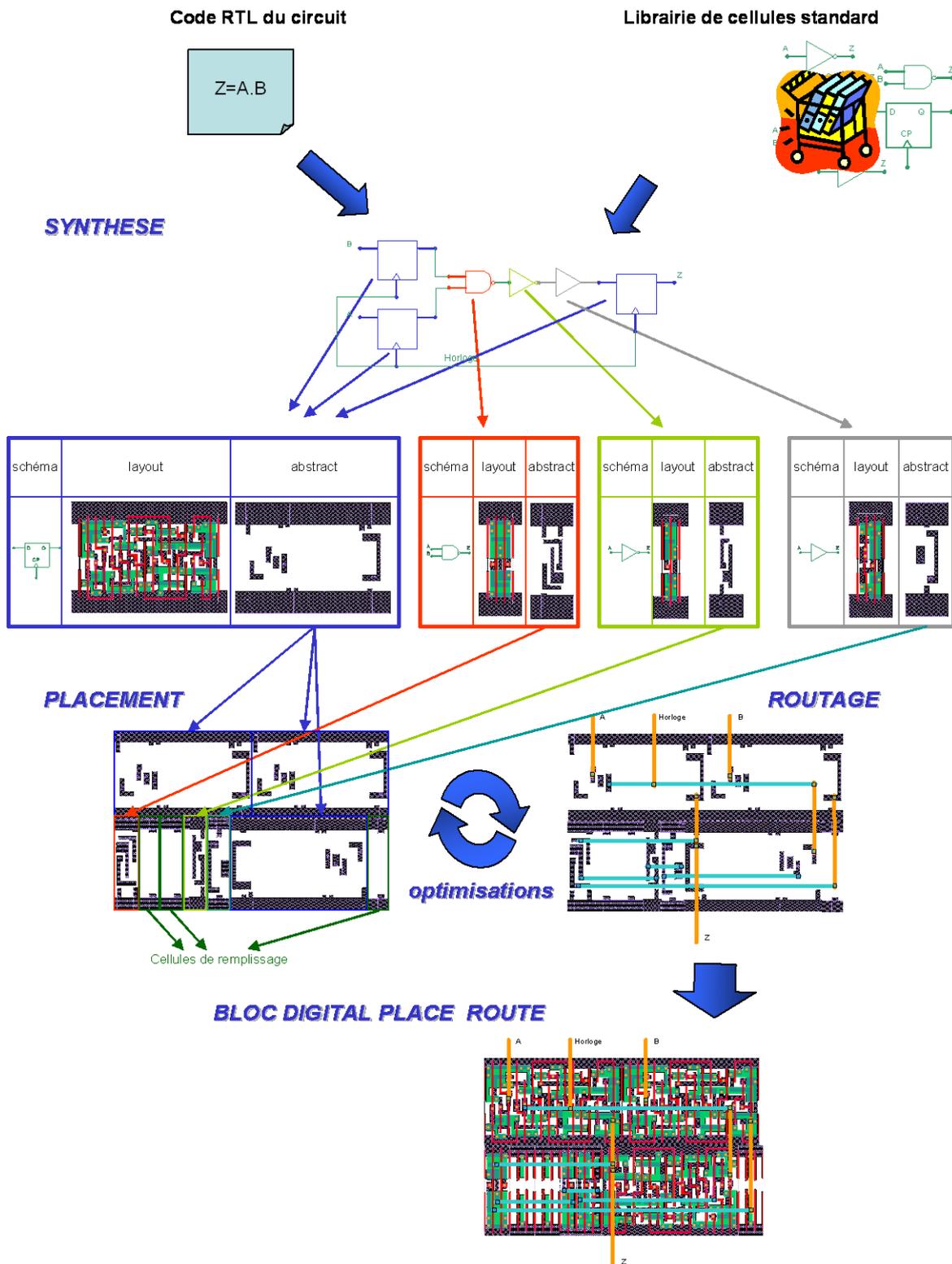


Figure II-4 : Illustration d'un flot digital.

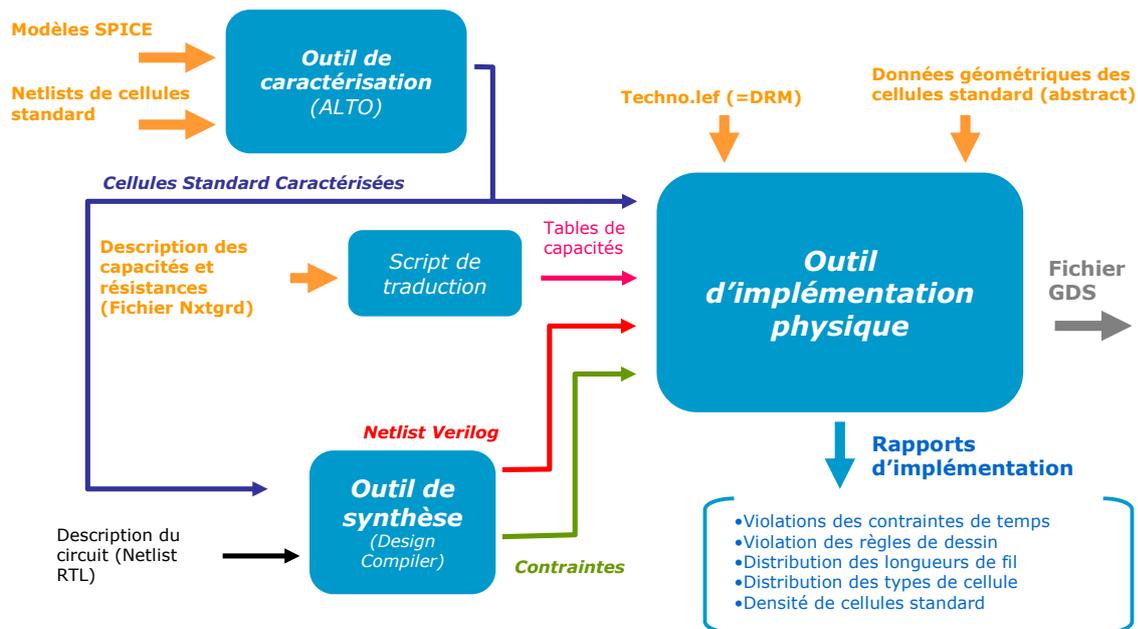


Figure II-5 : Exemple de schéma d'un flot de conception de circuit digital. Les données d'entrée sont les modèles SPICE, une bibliothèque de cellules standard (netlist et abstract), des modèles d'interconnexion (résistance et capacité), un fichier de paramétrage de l'outil d'implémentation physique contenant les règles de dessin de la technologie (fichier « techno.lef ») et la netlist RTL décrivant la fonctionnalité du circuit. La sortie consiste en un fichier GDSII (« Graphic Data System ») détaillant l'ensemble des masques afin de permettre leur conception chez le fabricant.

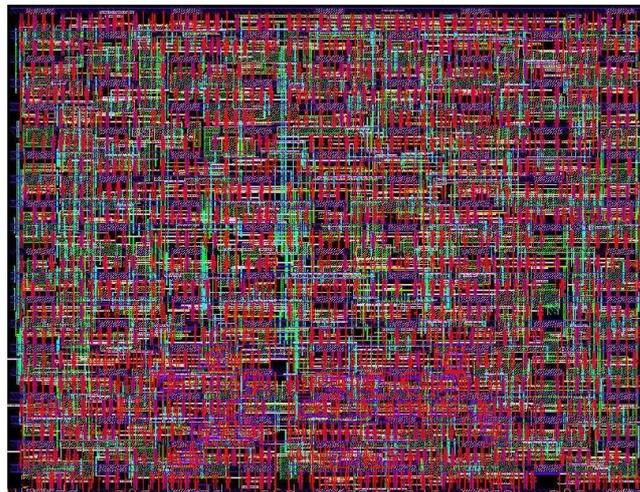


Figure II-6 : Vue de type masques (« layout ») d'un bloc digital placé routé (additionneur et multiplieur : 5000 transistors).

II.2.C Rôle du kit de conception et éléments nécessaire à un flot prédictif

Le kit de conception rassemble l'ensemble des paramètres nécessaires au fonctionnement des outils de conception. Il regroupe pour le flot full custom :

- Les modèles
 - des dispositifs (cartes modèle SPICE)
 - des interconnexions (fichier « nxtgrd »)
- Les paramétrages du DRM pour les outils de
 - vérification de règles de dessin (DRC)
 - vérification de l'homologie schémas/dessins (LVS)

Dans le cadre d'un flot digital, il faudra ajouter au minimum

- Les bibliothèques
 - description de niveau masques
 - description de niveau abstract
 - description schématique
 - fichier de connectique (netlist) avec capacités et résistances parasites
 - tables de délai (fichier .lib)
- Les paramétrages du DRM pour les outils de placement routage (fichier techno.lef)

Pour effectuer un flot prédictif, il est indispensable de construire un kit de conception prédictif regroupant l'ensemble des données listées ci-dessus correspondant au type de flot prédictif (full custom ou digital). Dans la mesure où le but d'un flot prédictif n'est pas d'obtenir des masques destinés à la fabrication mais des critères de performance sur les futures technologies, il est possible d'alléger le paramétrage des outils DRC et LVS.

II.3 Les critères d'évaluation des technologies

II.3.A Les critères « Front End »

a) Oscillateur en anneaux

Délai

L'oscillateur en anneau traduit la faculté d'une porte à charger une autre porte (cf. Figure II-7). C'est le cas typique qui se produit dans les circuits. C'est pourquoi les technologues utilisent souvent les oscillateurs en anneau pour caractériser la vitesse de la technologie lors des phases de développement de celle-ci. Plusieurs formules analytiques existent pour modéliser le délai moyen de transition du signal par porte. Même si cela se vérifie de moins en moins pour les technologies avancées, la charge capacitive mise en jeu dans les délais est dominée par la capacité d'oxyde C_{ox} des transistors d'entrée. Le délai d'un oscillateur en anneau dépend donc en grande partie de sa faculté à charger cette capacité. En regardant le transistor comme une résistance de valeur $R_{ON} = V_{dd}/I_{ON}$, le délai moyen pour une transition peut être vu comme le temps caractéristique de charge du circuit composé de la résistance R_{ON} et d'une capacité dépendant à un facteur de proportionnalité près de la capacité d'oxyde $K.C_{ox}$ (Figure II-8), soit un délai moyen par porte

$$\tau_p = \zeta_{ON} K C_{ox} = \zeta \frac{C_{ox} V_{dd}}{I_{ON}}, \quad (\text{II-1})$$

Afin de correspondre au temps moyen d'une transition (moyenne des temps de transition haut vers bas et bas vers haut), I_{ON} est égal au courant moyen des transistors NMOS et PMOS.

Pour mesurer le temps moyen d'une transition, il suffit de diviser la période du signal T par le nombre de transitions ayant eu lieu durant cette période, soit $2n$ avec n le nombre impair d'étages :

$$\tau_p = \frac{T}{2n} \quad (\text{II-2})$$

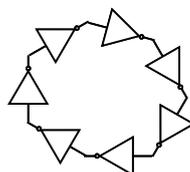


Figure II-7 : Principe de l'oscillateur en anneau. Ce dernier est composé d'un nombre n impair de portes inverseuses imbriquées les unes dans les autres. La fréquence d'oscillation caractérise la vitesse de la technologie.

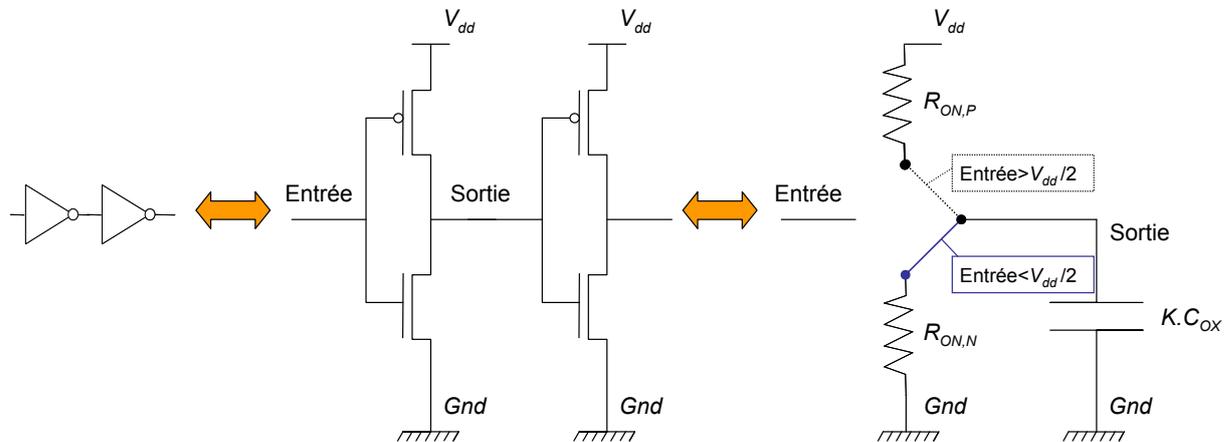


Figure II-8 : Schéma équivalent d'un étage d'un oscillateur en anneau.

Puissance dynamique

La puissance dissipée par l'oscillateur caractérise la puissance dynamique d'un circuit. Un paramètre intéressant à observer est le ratio puissance/fréquence. Plus ce ratio est bas et plus la technologie est à même de dissiper peu de puissance à une fréquence donnée.

La puissance dynamique lors d'une transition correspond au produit de la tension d'alimentation par le courant moyen dissipé par la cellule lors d'une transition, soit à une constante multiplicative près K :

$$P_{dyn} = \tau_{dd} K I_{ON} = \tau_{dd} K \frac{C_{ox} V_{dd}}{\tau_p} = \zeta C_{ox} V_{dd}^2 F, \quad (\text{II-3})$$

avec F la fréquence d'oscillation. Cette formule montre tout l'intérêt que les concepteurs de circuits ont à réduire les tensions d'alimentation et à diminuer les fréquences afin de réduire les puissances dissipées dans les circuits.

Il peut être intéressant de mesurer le ratio puissance par fréquence (PPF). D'après (II-3), il vaut :

$$PPF = P_{dyn} \tau_p = \zeta C_{ox} V_{dd}^2 \quad (\text{II-4})$$

Les formules analytiques (II-3) et (II-4) omettent de tenir compte de la puissance dite de court-circuit et ne considèrent que la puissance de charge des cellules. La puissance de

court-circuit provient de la faible résistance présente entre l'alimentation et la masse lorsque la tension de grille des portes est autour de $V_{dd}/2$. Cette puissance participe couramment à environ 10 à 15% de la puissance dynamique dissipée par le circuit.

Par simulation, la puissance P_{dyn} peut être obtenue en divisant le courant I_{alim} moyen dissipé par l'oscillateur en anneau sur une période de signal par le nombre de transitions durant une période ($2n$) :

$$P_{dyn} = \frac{\int_0^T \frac{I_{alim}(t)}{T} dt}{2n} \quad (\text{II-5})$$

Le ratio PPF vaut alors :

$$PPF = \frac{\int_0^T I_{alim}(t) dt}{4n^2} \quad (\text{II-6})$$

Puissance statique

La puissance statique correspond à la puissance dissipée par l'oscillateur lorsque celui-ci n'est pas en fonctionnement. Pour « arrêter » l'oscillateur, il suffit de placer une porte NAND à la place d'un des inverseurs. Ainsi, il existe un signal permettant d'arrêter le fonctionnement d'une des portes et donc d'arrêter la propagation du signal.

Etant donné que la porte NAND peut avoir des caractéristiques légèrement différentes du reste du circuit, il est préférable en simulation d'utiliser une tension V_{dd} placée entre deux étages afin d'arrêter la propagation du signal (Figure II-9).

Le critère important à mesurer est la puissance statique divisée par le nombre de portes :

$$P_{stat} = \frac{V_{dd} I_{alim,off}}{n} \quad (\text{II-7})$$

avec $I_{alim,off}$ le courant statique lorsque qu'il n'y a pas d'onde se propageant dans l'oscillateur en anneaux.

Une figure classique de mérite consiste à placer les points de mesure sur un graphe présentant le temps moyen par porte τ_p en fonction du courant statique I_{stat} . En effet, étant donné que τ_p est directement modulable par le courant I_{ON} (translation de la courbe $I_{DS}-V_{GS}$, cf. Figure I-10 p.39), qui lui-même est une fonction directe de la tension de seuil, il est

indispensable de vérifier que de bons temps de propagation ne sont pas obtenus au détriment d'un courant statique important.

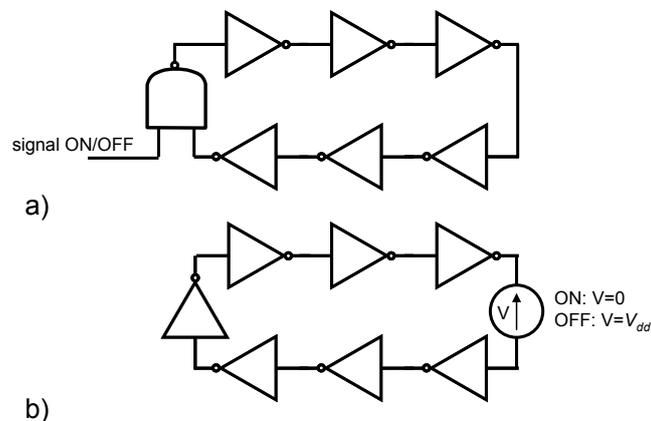


Figure II-9 : Circuits permettant d'arrêter les oscillations d'un oscillateur en anneau pour mesurer sa puissance statique. a) cas d'une implémentation silicium. b) cas de la simulation.

b) SRAM

L'évaluation des caractéristiques des mémoires SRAM des futures technologies revêt un intérêt particulier compte tenu de leur rôle extrêmement important dans les circuits (souvent plus de 50% de la surface des circuits est occupée par les mémoires SRAM). Plusieurs critères permettent de quantifier la qualité d'une mémoire SRAM.

Courants de fuite

L'utilisation de techniques basées sur la coupure des alimentations des blocs non utilisés permet souvent, au niveau de la conception des circuits, de réduire la puissance statique dissipée. Cependant ces techniques ne sont pas applicables aux SRAM car celles-ci doivent rester constamment alimentées sous peine de perdre l'information stockée. Le courant de fuite est donc un critère encore plus important pour les mémoires SRAM. Le courant de fuite est souvent exprimé en pA/cellule.

Marge au bruit à la lecture

Comme précisé au paragraphe I.3.Ba) p.42, la marge au bruit à la lecture (SNM ou « Static Noise Margin ») est le critère de qualité le plus fortement mis à mal avec les nouvelles générations de circuit à cause de la variabilité. Comme précisé ici [26], la SNM correspond à la tension parasite V_p maximale applicable aux bornes de la cellule sans que celle-ci change d'état en condition de lecture (WL et BL à 1) (Figure II-10). C'est cette définition

qui est retenue en simulation. Il suffit alors d'appliquer une tension V_p croissante et de noter la valeur minimale permettant de faire changer la valeur stockée.

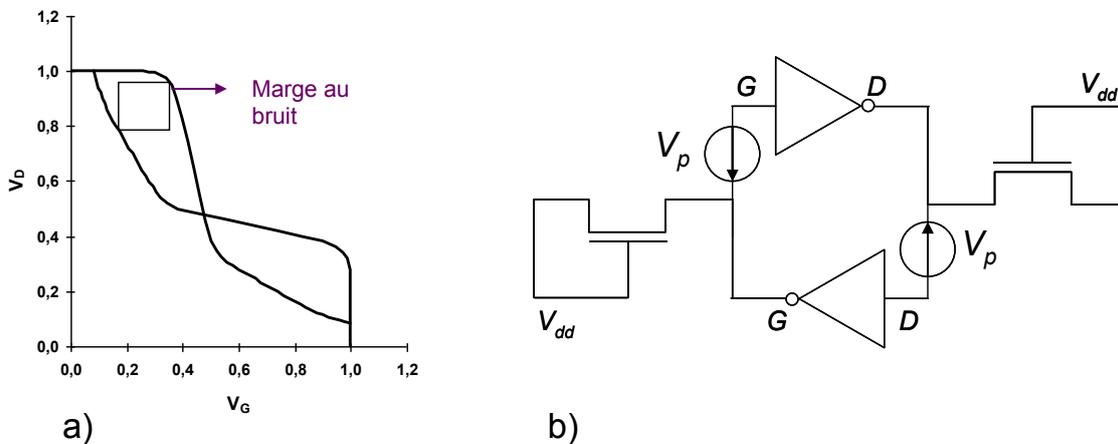


Figure II-10 : SNM d'une mémoire SRAM. Celle-ci peut être vue a) comme le côté du plus petit carré des deux plus grands carrés insérables dans les lobes de la courbe en papillon, b) comme la tension parasite V_p maximale applicable sans que la cellule ne change d'état.

Plus que la valeur absolue de la SNM, c'est la variabilité de la SNM qui pose problème. Ainsi, une technologie présentant une forte SNM en valeur moyenne peut présenter beaucoup plus de dispersion et donc avoir beaucoup plus de chances de donner lieu à des valeurs de SNM négatives qu'une technologie à valeur de SNM moyenne plus faible mais présentant moins de dispersion (Figure II-11). Le critère de mérite sera donc $\frac{\langle SNM \rangle}{\sigma_{SNM}}$,

c'est-à-dire la valeur moyenne de la SNM divisée par son écart type.

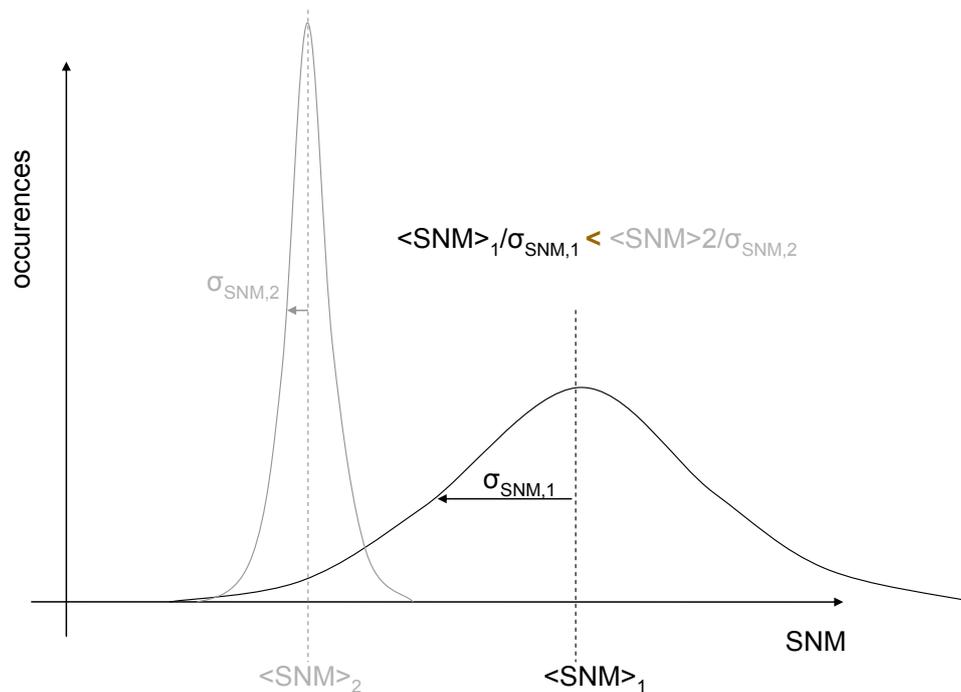


Figure II-11 : Illustration de la faiblesse du critère $\langle SNM \rangle$ seul. Ce qui compte, c'est la faible probabilité d'avoir des SNM négatives, qui est mieux traduite par le critère $\langle SNM \rangle / \sigma_{SNM}$.

Marge au bruit à l'écriture

Il arrive que les optimisations pour améliorer la SNM nuisent à la faculté d'écriture de la mémoire. Pour caractériser cela, les concepteurs de SRAM utilisent le critère WM (« Write Margin ») pour définir la marge à l'écriture. Celui-ci correspond à la tension de BL nécessaire à écrire un 0 dans la cellule mémoire (le 0 peut être considéré écrit pour une tension en dessous de 0.05V). Avoir une WM négative signifie que le point mémoire n'a pas de plage d'écriture dans la gamme de tension de « Bit Line » allant de 0 à V_{dd} . Ce qui compte est donc la probabilité que ce cas arrive, plus que la valeur absolue de WM . De la même manière que pour la SNM , le critère $\langle WM \rangle / \sigma_{WM}$ est utilisé pour définir la qualité d'écriture de la mémoire.

Courant de lecture

Lors de la lecture, les deux « Bits Line » sont préchargées à V_{dd} . L'ouverture d'une « Word Line » permet alors de décharger une des « Bit Line » dans la cellule sélectionnée (celle du côté où se trouve stocké le 0). Au bout d'un certain laps de temps, un amplificateur accroît la différence de tension de manière à lire la donnée (0 ou 1). Le laps de temps peut être d'autant plus court que la « Bit Line » se vide rapidement. Or, la décharge de la « Bit

Line » est proportionnelle au courant de lecture I_{read} (Figure II-12). Ce paramètre est donc important à suivre puisqu'il définit la vitesse de la mémoire. Par ailleurs, et de la même manière que pour la SNM et la WM, à cause de la variabilité, le cas définissant la vitesse de la mémoire sera le pire cas (I_{read} le plus faible). En effet, la spécification client définit les performances que ce dernier est en mesure d'attendre quelque soit les conditions, c'est-à-dire même dans les cas de variabilité extrême du courant de lecture. Il est donc nécessaire de travailler également à partir des critères $\langle I_{read} \rangle$ et σI_{read} .

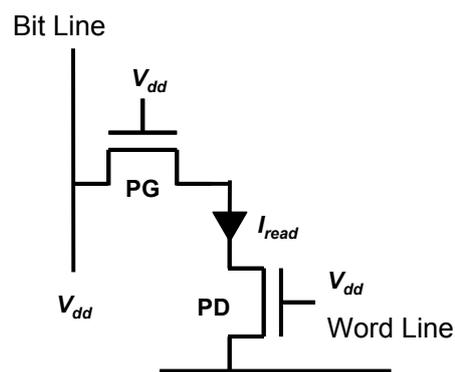


Figure II-12 : Courant de lecture I_{read} d'une mémoire SRAM.

Critères de robustesse à la variabilité

Comme discuté précédemment, le critère $\langle p \rangle / \sigma_p$ permet de quantifier l'immunité à la variabilité d'un paramètre p . Reste à définir une valeur acceptable pour ce critère. Si l'on suppose que la distribution de p est gaussienne (ce qui est le cas pour SNM, WM et I_{read} [27]), il est possible de connaître la probabilité d'avoir une chance que p soit négatif en fonction de $\langle p \rangle / \sigma_p$. De là il est alors possible d'en déduire la valeur $\langle p \rangle / \sigma_p$ à atteindre pour que la probabilité de défaillance d'une cellule mémoire parmi N soit de 100ppm (cf. Figure II-13). Ainsi, pour des mémoires de l'ordre de 10 Mbits (taille attendue pour le nœud 32nm), le critère suivant est requis :

$$\frac{\langle p \rangle}{\sigma_p} > 3,7 \quad (\text{II-8})$$

II.3.B Les critères « Back End »

Les critères définis précédemment ne font pas intervenir les interconnexions. Or, à cause de leurs capacités et résistances, celles-ci peuvent dégrader fortement les délais dans le cas

des lignes les plus longues. Il est donc nécessaire de définir des critères d'évaluation supplémentaires qui en tiennent compte.

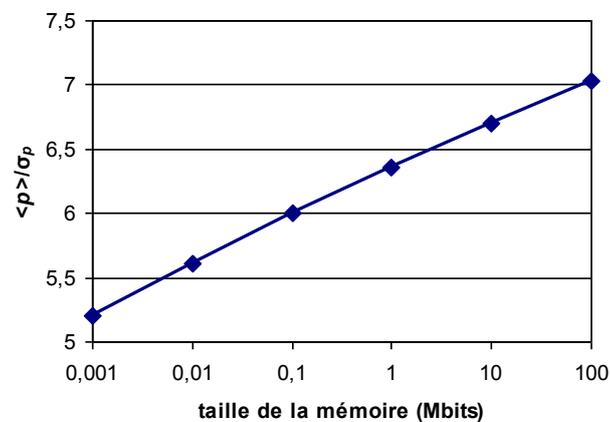


Figure II-13 : Critère $\langle p \rangle / \sigma_p$ minimal nécessaire pour avoir une probabilité de rejet de 100ppm en fonction de la taille de la mémoire.

a) Délai de ligne

Dans un circuit, les portes sont séparées par des lignes d'interconnexion plus ou moins longues. Il est donc particulièrement intéressant de regarder le délai de propagation d'un signal entre deux portes séparées par une ligne de longueur variable comme décrit Figure II-14.

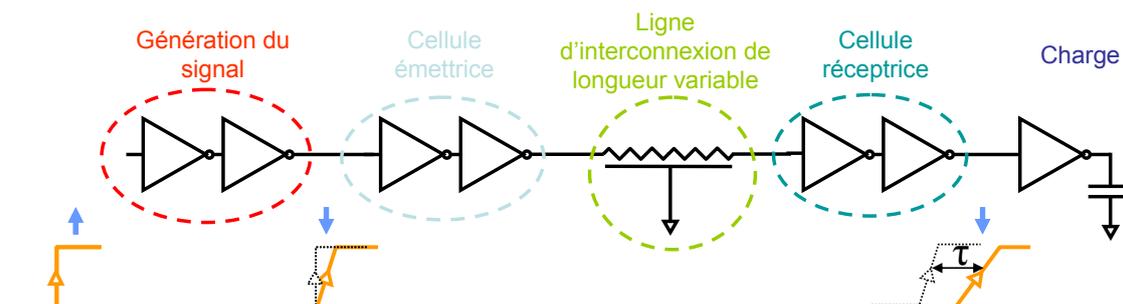


Figure II-14 : Mesure du délai entre deux portes séparées par une ligne d'interconnexion de longueur variable.

b) Nombre optimal de répéteurs

Pour réduire les délais de ligne, des portes répétrices peuvent être insérées uniformément le long des lignes. Un critère important à regarder est donc le nombre de répéteur optimal

permettant de minimiser les délais ainsi que le délai minimal atteignable à l'aide de répéteurs.

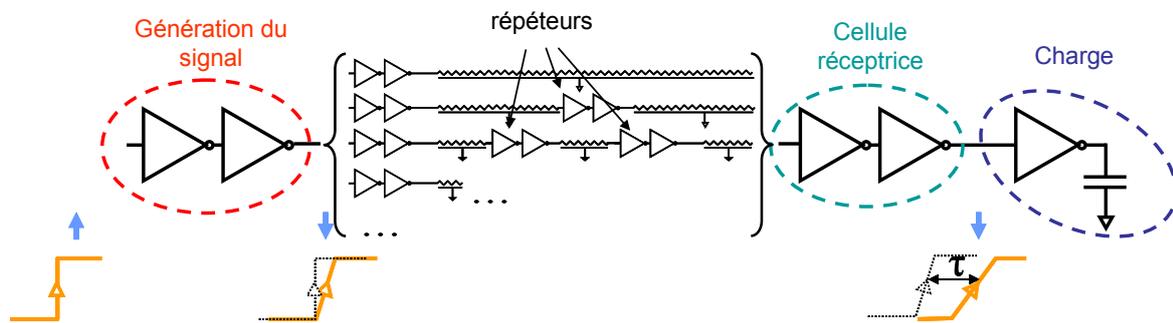


Figure II-15 : Mesure du délai entre deux portes séparées par une ligne d'interconnexion de longueur fixe et présentant un nombre de répéteurs variable.

c) Chemin critique d'un bloc placé routé

Le chemin critique d'un bloc placé routé correspond au chemin de propagation du signal présentant le délai le plus important du bloc. Il caractérise donc la vitesse du bloc. Une méthode astucieuse pour mesurer ce délai consiste à réaliser un bloc placé routé puis à faire réaliser l'exécution d'une opération par le bloc dont le résultat est connu à l'avance. En comparant le résultat fourni par le bloc à celui connu et enregistré, il est possible de mesurer le temps de calcul du bloc, c'est-à-dire le délai du chemin critique du bloc pour cette donnée d'entrée. En implémentant le bloc sous forme d'oscillateur en anneau, il est possible d'obtenir au final un signal oscillant dont la période sera proportionnelle au délai du chemin critique. Cela permet de minimiser les ressources en terme de test.

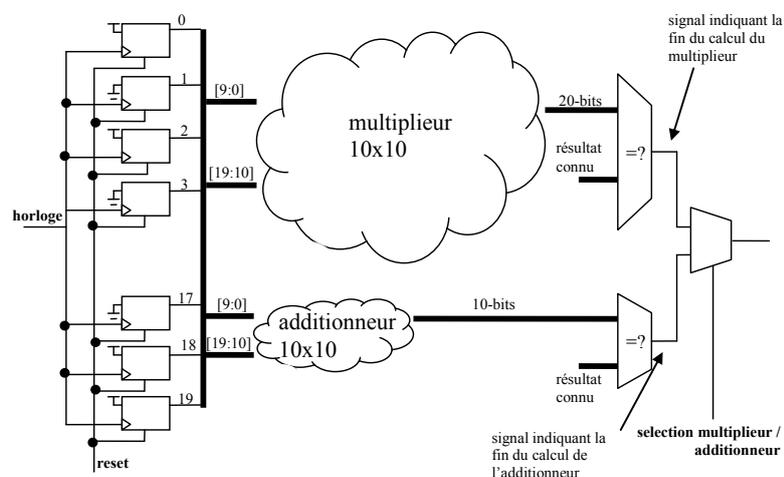


Figure II-16 : Principe du circuit « Flipper » développé par STMicroelectronics destiné

au calcul du délai d'un chemin critique d'un bloc (additionneur ou multiplieur) placé routé.

II.3.C Flot digital

Le critère le plus réaliste d'évaluation technologique consiste à réaliser entièrement un bloc placé routé et à analyser différents critères de performance.

a) Nombre de violations de contraintes de temps

Ce critère définit le nombre de contraintes temporelles que l'outil de placement routage n'a pas été en mesure d'atteindre. Ce chiffre doit être nul ou presque nul (dans ce dernier cas, des corrections manuelles peuvent suffire pour corriger les problèmes). Un trop grand nombre de violations temporelles signifie que le circuit est trop difficile à réaliser à partir des contraintes de surface et de délai fixées. Il faut alors donner de la souplesse à l'outil, soit en relâchant les contraintes temporelles (le plus souvent, il s'agit de la fréquence d'horloge), soit en élargissant la surface allouée (« Floor Plan »).

b) Nombre de violation de règles de dessin

Lors de la phase de placement routage, l'outil travaille par itérations. Il effectue dans les premières phases un premier routage grossier ne respectant pas forcément les règles de dessins (espacement de métaux, densité, etc.). C'est seulement au bout d'un certain nombre d'itérations que l'outil fixe les problèmes relatifs aux règles de dessin. Si les contraintes temporelles ou de surface sont trop importantes, l'outil ne peut corriger toutes les règles de dessin violées dans les premières itérations et un certain nombre de violations persistent. Ainsi, de la même façon que pour les violations de contraintes temporelles, si le nombre de violation de règles de dessin est faible, une action manuelle est envisageable. En revanche, un nombre élevé témoigne d'un circuit impossible à réaliser compte tenu des contraintes fixées.

c) Densité de cellules standard

Une fois le circuit réalisé, la densité de cellules standard correspond au ratio surface des cellules / surface allouée. Ce ratio traduit l'efficacité d'utilisation du silicium. Les circuits les plus denses peuvent dépasser une densité de 90%.

d) Distribution des familles de cellules

Il existe deux grands types de familles de cellules : les cellules séquentielles et les cellules combinatoires. Parmi les cellules combinatoires, il faut distinguer les cellules assurant la fonctionnalité logique du circuit et les autres cellules servant soit à accélérer le signal (« buffers », inverseurs), soit à le ralentir (cellules de temporisation).

e) Distribution des sortances des cellules

Un même type de cellule est souvent décliné en plusieurs versions disposant de sortances différentes. La sortance est fonction de la puissance du transistor de sortie de la cellule. Plus elle est élevée, plus le courant de sortie, et donc la faculté de la cellule à charger rapidement les capacités de sortie, est élevé. L'outil de placement routage a la possibilité d'échanger certaines cellules par des cellules de sortance plus élevée afin de minimiser les délais.

f) Distribution des longueurs d'interconnexions

La distribution des interconnexions dans une puce suit une loi phénoménologique dite de Rent [28]. Cette loi spécifie qu'il existe une relation simple entre le nombre d'entrées/sorties T d'un bloc et la complexité de ce bloc (nombre de porte g) :

$$T = g^p, \quad (\text{II-9})$$

t correspond au nombre moyen de connexions par porte et p à l'exposant de Rent qui reflète le degré d'optimisation du placement.

Une des conséquences de cette loi est que la distribution des longueurs d'interconnexion est linéaire en échelle logarithmique (cf. Figure II-17). Analyser les distributions des longueurs est particulièrement intéressant surtout lorsqu'il existe une longueur d'interconnexion critique pour la technologie considérée comme il sera étudié plus loin dans ce manuscrit.

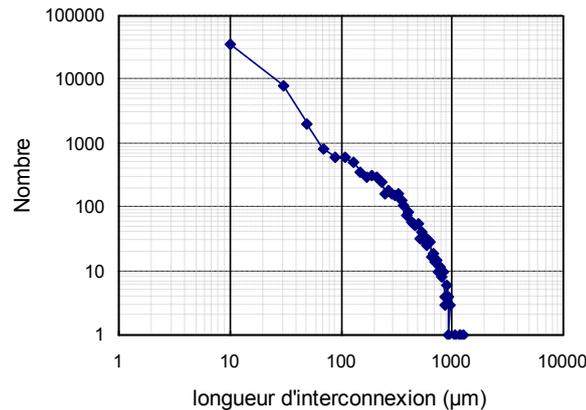


Figure II-17 : Distribution des longueurs d'interconnexion pour un circuit de type contrôleur de port USB. La loi de Rent implique qu'il y a beaucoup d'interconnexions de faible longueur et peu de longues.

II.4 Conclusion

Les deux types de flots de conception (« full custom » et digital) sont abordés dans ce chapitre. Ceux-ci nécessitent différentes étapes et outils. Dans le cas d'un flot « full custom », seuls des modèles SPICE et le paramétrage des outils DRC et LVS est requis. Pour le flot digital, il est nécessaire d'ajouter des bibliothèques de cellules standard caractérisées au niveau des délais ainsi qu'un paramétrage de l'ensemble des outils.

L'évaluation des technologies peut s'effectuer à différents niveaux. Le critère permettant de caractériser les aspects délai « front end » (niveau transistor) est l'oscillateur en anneau. La variabilité « front end » peut être caractérisé par les critères de performances de la SRAM (SNM, WM, I_{read} , I_{stat}). Les aspects « back end » (niveau interconnexions) sont couverts par les délais de ligne avec ou sans répéteurs. Finalement, les données de performance circuit mêlant des aspects à la fois « Front End » et « Back End » sont fournis par les mesures de délai de chemin critique à l'aide de circuits adaptés et par les rapports d'implémentation d'un bloc placé routé (nombre de violations de contraintes de temps et de règles de dessin, densité du circuit, distribution des cellules et des longueurs d'interconnexions).

Le chapitre suivant est consacré à l'élaboration des méthodes de générations de kits de conception prédictifs permettant d'appliquer les critères d'évaluation technologiques pour les futures technologies.

Chapitre III Méthodes de génération de kits de conception prédictifs

Le modèle MASTAR est un modèle de référence permettant de prédire l'évolution des futures technologies. A partir de MASTAR est bâti un modèle SPICE prédictif, MSIM. Ce modèle, basé sur une écriture des équations de MASTAR dans une carte modèle standard de type BSIM3 est capable de reproduire fidèlement le comportement électrique de transistors de plusieurs générations technologiques. En l'associant à un modèle prédictif d'interconnexion, et à un paramétrage correct des outils CAO, un kit complet de conception prédictif est élaboré.

III.1 Introduction

Comme décrit dans le chapitre précédent, pour établir un flot de conception prédictif, il est nécessaire de paramétrer correctement les outils d'aide à la conception de circuits. Dans le cas d'un flot « Full Custom », les données à adapter sont le manuel contenant les règles de dessins (DRM), les modèles SPICE et les modèles de capacités et de résistances des interconnexions (Figure II-1). Dans le cas de l'implémentation d'un bloc placé routé, il est nécessaire de fournir également aux outils de synthèse et de placement routage des bibliothèques de cellules standard caractérisées à partir des modèles SPICE ainsi qu'un fichier de paramétrage de l'outil d'implémentation physique décrivant les règles de dessin (Figure II-5).

La première partie de ce chapitre est ainsi consacrée à la modélisation SPICE prédictive des transistors. La modélisation des interconnexions est ensuite présentée dans une seconde partie. Finalement, dans une troisième partie, la façon d'obtenir des informations prédictives concernant le DRM et le paramétrage des outils CAO est abordée.

III.2 Modélisation SPICE prédictive

III.2.A Le modèle MASTAR

a) Principe

Le modèle MASTAR (Model for Analysis of CmoS Technology And Roadmaps) est un modèle analytique développé par STMicroelectronics. Il a pour but de modéliser les performances des dispositifs des futures technologies CMOS. Son principe repose sur quelques équations analytiques simples [29], dont la transformation tension dopage (VDT : Voltage Doping Transformation) [30] qui permet de modéliser simplement les effets canaux courts pour tout type d'architecture (Bulk, SOI/SON, DG). En outre, MASTAR inclut différents modèles ou facteurs correctifs permettant de prendre en compte les principaux phénomènes physiques nouveaux intervenant pour les futures technologies :

- déplétion du polysilicium
- confinement quantique des porteurs de charge (Darkspace)
- effets des contraintes sur la mobilité
- modèle de diélectrique à haute permittivité (high k)
- modèle de dopage par poches
- résistances d'accès
- facteur d'amélioration du courant par effet balistique
- dégradation de la pente sous le seuil

Ce modèle a été constamment validé en confrontant ses résultats calculés par rapport aux mesures sur silicium depuis les générations 90nm.

b) Intérêt

L'intérêt du modèle MASTAR est de fournir très rapidement, à partir seulement d'un jeu de paramètres principalement physique (dopage, travail de sortie de la grille, épaisseur d'oxyde, etc...), un ensemble de critères de performance pour les futurs dispositifs. Avec MASTAR, il est ainsi possible de mesurer facilement l'évolution de paramètres technologiques suivant plusieurs scénarios. Cette faculté a valu à MASTAR d'être utilisé

pour construire l'intégralité des feuilles de route technologiques (« roadmap ») ITRS 2005 et 2007. La Figure III-1 est un aperçu d'écran de l'interface MASTAR. La Figure III-2 donne un exemple d'évolution d'un paramètre (CV/I) au cours du temps simulé grâce au logiciel MASTAR.

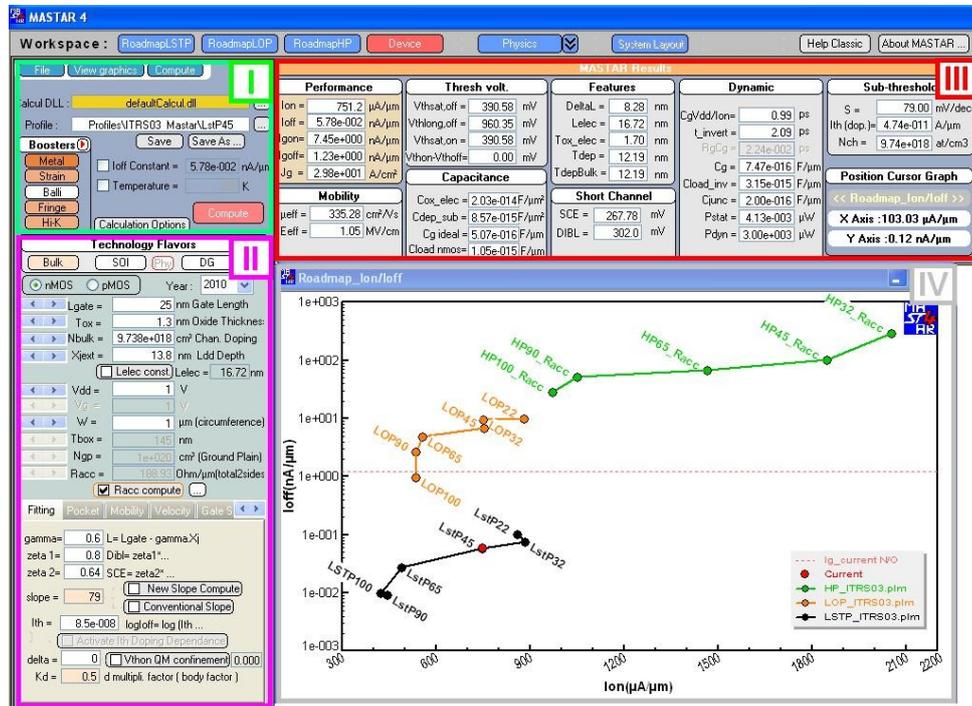


Figure III-1: Vue d'écran du logiciel MASTAR. La partie I gère les fichiers de paramètres, la partie II sert à rentrer les données de la technologie considérée (épaisseur d'oxyde, dopage, tension d'alimentation, architecture du transistor, etc.), la partie III récapitule les données de sortie du modèle sous forme tableau et la partie IV permet d'exploiter les données sous forme graphique.

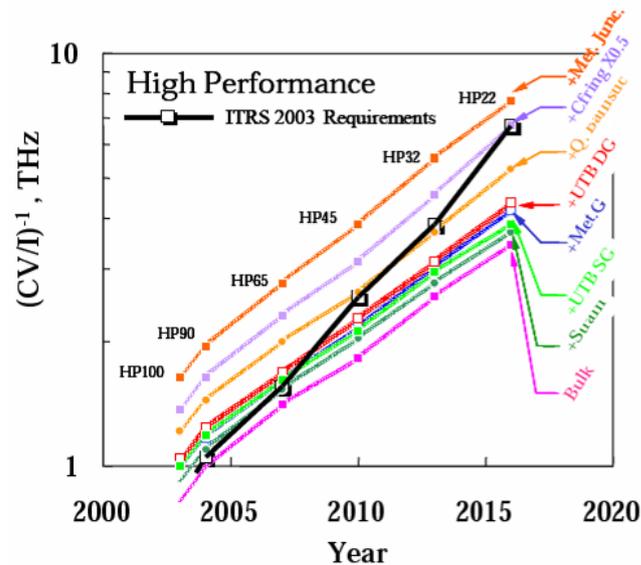


Figure III-2: Exemple d'application de MASTAR : évolution du paramètre CV/I au cours du temps en fonction de différents scénarios technologiques. (d'après [31])

c) Limites

MASTAR, dans la version disponible au début de cette étude – d'autres évolutions de MASTAR, inspirés en grande partie par cette étude, ont ensuite partiellement corrigé ce problème – ne fournit que quelques paramètres de sortie des dispositifs (principalement les courants aux points de polarisation I_{ON} et I_{OFF}). Or, dans le cadre de l'étude présente qui concerne la simulation de circuit, il est nécessaire de savoir reproduire l'ensemble des caractéristiques courants/tensions et charges/tensions. Le modèle MASTAR, bien que fort utile pour les technologues, n'est donc pas utilisable en tant que tel pour l'évaluation des performances au niveau circuit.

III.2.B Modélisation SPICE

a) Principe du simulateur SPICE

Pour effectuer des simulations électriques de circuit, l'industrie de la microélectronique utilise depuis plus de 30 ans le principe SPICE (Simulation Program with Integrated Circuit Emphasis). SPICE est une méthode de résolution d'un ensemble de flux/potentiels qui s'adapte très bien, comme son nom l'indique, aux simulations de circuits. Le flux est alors le courant et le potentiel le potentiel électrique. Le principe de base réside dans l'utilisation d'algorithmes de Newton-Raphson qui, par itérations successives, tente de

trouver les valeurs de potentiel et de courant aux différents nœuds du circuit en résolvant les lois de Kirchhoff pour chaque instant considéré. Une méthode d'évaluation de l'erreur commise permet de décider quand arrêter le processus d'itération. Le simulateur a ainsi besoin d'effectuer un nombre très important d'itérations avant de trouver une solution acceptable. Il est donc nécessaire que le calcul des courants et tensions aux bornes des dispositifs du circuit soit très rapide, ce qui implique que tous les dispositifs du circuit soient décrits de manière très simple en terme de relation entre les courants et les tensions. Il est donc nécessaire pour la description de ces éléments d'utiliser des équations **analytiques** de manière à éviter toutes les situations qui impliqueraient deux boucles d'itération (modèle et SPICE) et qui seraient rédhibitoires en temps de simulation du circuit

b) Problématiques relatives à la modélisation SPICE

Si la description analytique de la relation tension/courant d'une résistance idéale semble assez simple ($U=RI$), celle d'un MOS est déjà nettement plus difficile, surtout quand il s'agit de décrire fidèlement un comportement électrique qui dépend de centaines, voir de milliers de facteurs. Le domaine de la modélisation de circuit a ainsi très fortement évolué depuis les débuts de la microélectronique. En effet, les premiers modèle (SPICE level 1, 2, 3) n'avaient que quelques paramètres d'entrée physiques qu'il s'agissait d'ajuster par rapport au silicium afin de décrire scrupuleusement le comportement du transistor. La tâche de modélisation était donc relativement aisée. Or, à l'heure actuelle, les modèles de MOS utilisés (BSIM4 par exemple) possèdent plus de 400 paramètres d'entrée (carte modèle) afin de refléter tous les effets parasites (SCE, DIBL, fuites de grille, effets de proximité, etc.). Ces paramètres, bien que d'origine physique pour la plupart, ne sont souvent plus l'image directe de phénomènes physiques et sont utilisés dans leur grande majorité comme paramètres d'ajustement mathématiques. Les modèles ont ainsi perdu beaucoup de leur sens physique, abandonnant par là même tout caractère prédictif.

III.2.C Méthode de modélisation prédictive

a) Modifications empiriques de cartes technologiques existantes

A cause de ce caractère non physique des paramètres des cartes modèles, la modélisation SPICE prédictive (modélisation non issue de caractérisation sur silicium) s'avère être une tâche particulièrement ardue. La première méthodologie pour obtenir des cartes prédictives

consiste à établir des lois empiriques d'évolution des paramètres des précédentes générations technologiques et à les extrapoler à la génération voulue (Figure III-3). Cette façon de faire ne repose sur aucun principe physique valable et ne semble donc justifiée que pour extrapoler une technologie relativement similaire à celle de la carte modèle d'origine. En effet, savoir que l'on manipule des données qui ont perdu leur sens physique est une chose, extrapoler des « lois » sur ces paramètres dépourvus de sens physique en est une autre. Cette façon de procéder est proche de celle utilisée pour obtenir des modèles dits « tentative » où les cartes modèles sont élaborées à partir de données éparses et non cohérentes (TCAD, « bon sens », données morphologiques, extrapolation de précédents paramètres, etc.).

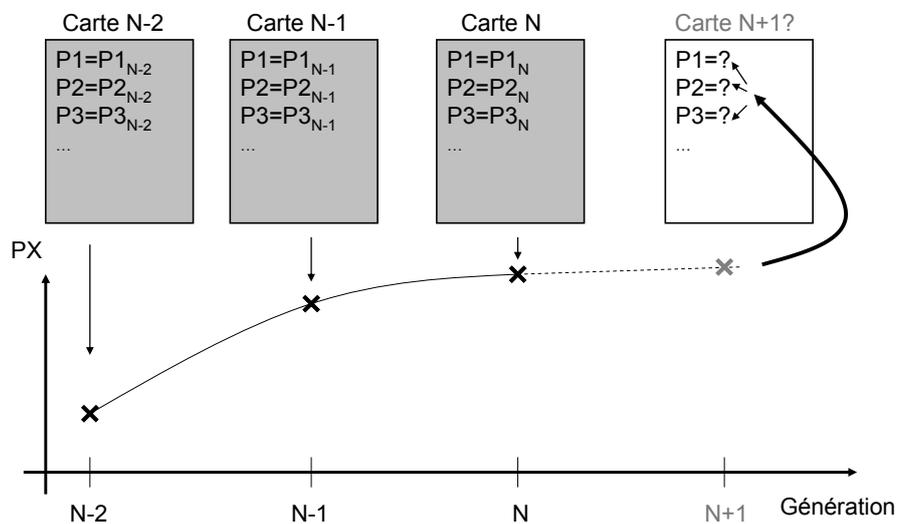


Figure III-3 : Génération d'une carte modèle d'une technologie N+1 à partir des données des précédentes générations de cartes modèles. Pour chaque paramètre, une « loi » d'évolution du paramètre en fonction des générations est établie et extrapolée à la génération N+1.

b) Modèles réalisés à partir de simulations numériques (TCAD)

L'autre méthode connue pour obtenir des cartes modèles consiste à ajuster les paramètres des modèles sur des données issues de simulations numériques (TCAD : « Technology Computer Aided Design»). Ces simulations permettent de reproduire l'ensemble des étapes du procédé de fabrication des transistors afin de simuler leur comportement électrique. Cette méthode est beaucoup plus fiable que la précédente grâce à la cohérence des données d'entrée. Les inconvénients sont, d'une part, son extrême complexité en représentation de la technologie, et de ressources informatiques et humaines, et, d'autre part, la lenteur pour

obtenir des résultats (plusieurs semaines). L'utilisation de TCAD implique donc de fixer les données du procédé de fabrication des dispositifs à modéliser une fois pour toute. Or, ces données sont en permanence susceptibles d'évoluer lors de l'étude de multiples scénarii technologiques qui est le but de cette thèse.

III.2.D Développement d'un nouveau modèle prédictif : MSIM

a) Principe

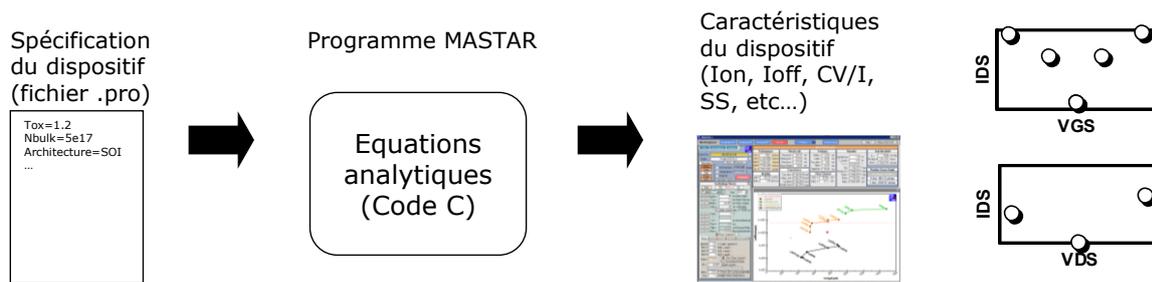
L'originalité de cette étude consiste à élaborer des cartes modèles directement à partir du logiciel MASTAR afin de s'affranchir des inconvénients des deux précédentes méthodologies de modélisation prédictives (Figure III-4). Le caractère prédictif du modèle est assuré par l'utilisation de MASTAR, outil reconnu par la communauté internationale, et la rapidité grâce à l'emploi intégral de formules analytiques (contrairement à la TCAD qui, elle, procède par résolution numériques itératives). Le principe de la méthodologie repose sur le constat que l'équation de courant de MASTAR est identique à celle de modèles SPICE standard à base de tension de seuil comme BSIM3 et 4 [32] ce qui n'est pas le cas de modèles à base de potentiels de surface de type PSP [33] :

$$I_{D,sat 0} = \frac{1}{2} \mu_{eff} C_{ox_el} \frac{W}{L} V_{GT} V_{D,sat} \quad (III-1)$$

avec $I_{D,sat 0}$ le courant de saturation source drain sans effets de résistances parasites, μ_{eff} la mobilité effective du canal, C_{ox_el} la capacité électrique équivalente du diélectrique de grille par unité de surface, W/L le ratio largeur sur longueur du transistor, V_{GT} la différence entre la tension grille et la tension de seuil et $V_{D,sat}$ la tension de saturation du drain.

Les différences résident d'une part dans la présence de fonctions de lissage pour faire le lien entre les différents régimes de fonctionnement dans le cas des modèles BSIM3 et 4, et d'autre part dans la façon de programmer les effets parasites (SCE, DIBL, etc...). La méthode MSIM développée dans le cadre de cette étude consiste à transcrire MASTAR dans une carte modèle de type BSIM3, c'est-à-dire à programmer la carte modèle (paramètres de BSIM3) de façon à bénéficier des calculs des effets parasites de MASTAR tout en profitant des fonctions de lissage de BSIM3. BSIM3 est choisi car il couvre un large spectre de phénomènes parasites tout en présentant une quantité limitée de paramètres par rapport à BSIM4. Par la suite le terme BSIM renvoie systématiquement à la carte modèle BSIM3.

Modélisation MASTAR



Modélisation MASTAR/SPICE

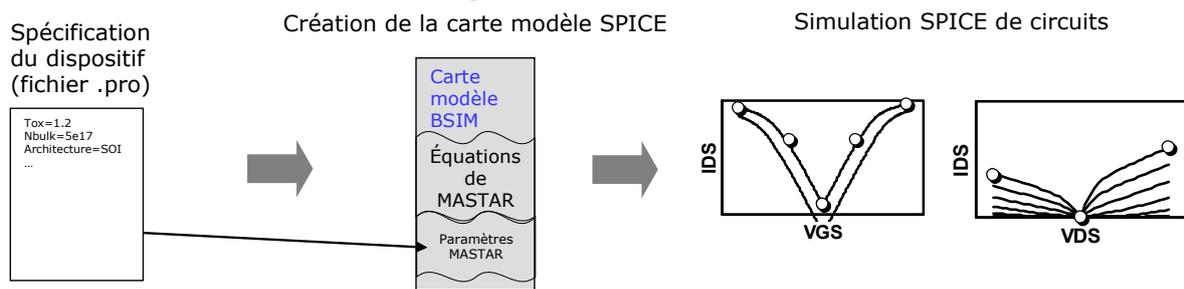


Figure III-4 : Principe de modélisation MASTAR et MSIM.

b) Génération de la carte modèle

La méthode peut être résumée de la façon suivante : lorsque MASTAR prévoit un effet de valeur P_{MASTAR} (dépendant principalement de paramètres physiques) et lorsque BSIM propose de programmer ce même effet P_{BSIM} à l'aide de plusieurs paramètres (V_1, V_2, \dots, V_n), il faut trouver un jeu de valeurs pour ces paramètres ($V_{01}, V_{02}, \dots, V_{0n}$), tels que $P_{BSIM}(V_{01}, V_{02}, \dots, V_{0n}) = P_{MASTAR}$. En appliquant cette méthode à l'ensemble des paramètres de MASTAR, il est possible de reprogrammer la majorité des éléments du modèle MASTAR dans la carte BSIM. Ci-dessous figure le détail des valeurs des paramètres BSIM programmés pour chaque élément de modélisation (les paramètres d'entrée du modèle BSIM sont marqués en gras, sans italique) :

Longueur de canal effective

Equations et paramètres :

$$\text{MASTAR : } L_{eff} = L_g - \Delta \tag{III-2}$$

$$\Delta = 1.8 X_j \tag{III-3}$$

$$\text{BSIM : } L_{eff} = L_g - \text{LINT} \quad (\text{III-4})$$

$$\text{Solution : } \text{LINT} = 1.4 X_j \quad (\text{III-5})$$

Définition des variables : L_{eff} est la longueur électrique effective du canal, L_g la longueur de grille dessinée et X_j est la profondeur des extensions source et drain.

Commentaire : L'intégration de la longueur effective de MASTAR dans la carte modèle BSIM est immédiate grâce au paramètre **LINT**.

Epaisseur d'oxyde (cas d'une grille métallique seulement)

Equations et paramètres :

$$\text{MASTAR : } T_{ox_el} = \tau_{ox_phys} + \tau_{darkspace} \quad (\text{III-6})$$

$$\text{BSIM : } \text{TOXE}$$

$$\text{Solution : } \text{TOXE} = \tau_{ox_phys} + \tau_{darkspace} \quad (\text{III-7})$$

Définition des variables : T_{ox_el} est l'épaisseur électrique de l'oxyde de grille, T_{ox_phys} est l'épaisseur physique (en équivalent SiO_2) et $T_{darkspace}$ est la profondeur de la « zone noire » dépourvue de porteurs de charges (en équivalent SiO_2)

Commentaire : La programmation de l'épaisseur électrique de l'oxyde de grille est directe dans le cas de la grille métallique (absence de déplétion du polysilicium). Dans ce cas, il suffit d'ajouter la valeur de correction des effets de confinement quantique des porteurs de charge $T_{Darkspace}$ prévue par MASTAR à la valeur de T_{ox_phys} . Le cas de la grille en polysilicium est discuté plus loin.

Effet de substrat

Equations et paramètres :

$$\text{MASTAR : } V_{D,sat} = \frac{L_{eff} E_c V_{GT}}{L_{eff} E_c + (1 + d) L_{eff} E_c} \quad (\text{III-8})$$

$$\text{avec } d = \frac{K_B}{2\sqrt{2}\phi - \tau_{BS}} \quad (\text{III-9})$$

$$\text{et } K_B = \frac{q N_{ch} T_{dep}}{C_{ox_el} \sqrt{2\phi_{BS}}} \quad (\text{III-10})$$

$$\text{BSIM : } V_{D,sat} = \frac{L_{eff} E_c V_{GT}}{L_{eff} E_c + A_{bulk} L_{eff} E_c} \quad (\text{III-11})$$

avec

$$A_{bulk} = 1 + \frac{\mathbf{K1}}{2\sqrt{2\phi_{BS}}} \left[\frac{\mathbf{A0} L_{eff}}{L_{eff} + 2\sqrt{X_J X_{dep}}} \left(1 - \mathbf{AGS} V_{gsteff} \left(\frac{L_{eff}}{L_{eff} + 2\sqrt{X_J X_{dep}}} \right)^2 \right) + \frac{\mathbf{B0}}{W + \mathbf{B1}} \right] \cdot \frac{1}{1 + \mathbf{KETA} V_{BS}} \quad (\text{III-12})$$

$$\text{Solution : } A_{bulk} = 1 + d \quad (\text{III-13})$$

$$\text{d'où } \mathbf{K1} = K_B \quad (\text{III-14})$$

$$\mathbf{A0} = \frac{L_{eff} + \sqrt{X_J X_{dep}}}{L_{eff}} \quad (\text{III-15})$$

$$\mathbf{AGS} = 0 \quad (\text{III-16})$$

$$\mathbf{B0} = 0 \quad (\text{III-17})$$

$$\mathbf{KETA} = -0,047 \text{ (valeur par défaut)} \quad (\text{III-18})$$

Définition des variables : N_{ch} est la valeur du dopage du canal, E_c est la valeur du champ critique, d est le paramètre de pondération du courant par l'effet non uniforme du substrat le long du canal, K_B est le coefficient de substrat, A_{bulk} est le coefficient de charge du substrat de BSIM et V_{BS} la tension de polarisation source substrat.

Commentaire : Afin de réaliser la correspondance $A_{bulk}=1+d$, les coefficients $\mathbf{K1}$, $\mathbf{A0}$, \mathbf{AGS} , $\mathbf{B0}$ et \mathbf{KETA} ont été programmés selon les équations (III-14) à (III-18). Il est important de noter que MASTAR ne prend pas en compte de variation du coefficient de substrat en fonction de la longueur de grille contrairement à BSIM. C'est pour cette raison que les paramètres \mathbf{AGS} et $\mathbf{B0}$ sont mis à 0 et que l'effet programmé par le paramètre $\mathbf{A0}$ est annulé. Le paramètre \mathbf{KETA} , qui traduit l'influence de l'effet de substrat par rapport à

la tension V_{BS} , est laissé à sa valeur par défaut étant donné que MASTAR ne sait pas modéliser l'effet du substrat sur ce paramètre.

Tension de seuil canal long :

Equations et paramètres :

$$\text{MASTAR : } V_{th,\infty \text{ MASTAR}} = \tau_{FB} + \psi_0 + \frac{1}{C_{ox_eot}} \sqrt{2\varepsilon_{Si} q N_B (2\psi_0 - \tau_{BS})} \quad (\text{III-19})$$

BSIM : **VTH0**

$$\text{Solution : } \mathbf{VTH0} = \tau_{th,\infty \text{ MASTAR}} \quad (\text{III-20})$$

Définition des variables : V_{FB} est la tension de bandes plates, φ_F le niveau de Fermi, ε_{Si} la permittivité diélectrique du silicium, q la charge élémentaire de l'électron, N_B le dopage du canal.

Commentaire : La formule de tension de seuil de MASTAR est directement intégrée au paramètre de tension de seuil **VTH0** de BSIM

Effet canal court

Equations et paramètres :

$$\text{MASTAR : } SCE_{MASTAR} = 1.64 \frac{\varepsilon_{ox}}{\varepsilon_{Si}} \left(1 + \frac{X_j^2}{L_{eff}^2} \right) \frac{\Gamma_{ox_el} T_{dep}}{L_{eff}} V_{bi} \quad (\text{III-21})$$

$$\text{BSIM : } SCE_{BSIM} = \mathbf{DVT0} \left[\exp\left(-\mathbf{DVT1} \frac{L_{eff}}{2L_t}\right) + \exp\left(-\mathbf{DVT1} \frac{L_{eff}}{2L_t}\right) \right] \psi_{bi} - \psi_0 \quad (\text{III-22})$$

$$\text{Solution : } \mathbf{DVT0} =) \quad (\text{III-23})$$

$$\mathbf{DVT1} =) \quad (\text{III-24})$$

$$\mathbf{VTH0} = \tau_{th,\infty \text{ MASTAR}} - SCE_{MASTAR} \quad (\text{III-25})$$

Définition des variables : ε_{ox} la permittivité diélectrique de l'oxyde de grille, T_{dep} la profondeur de déplétion, V_{bi} est la tension de seuil de la jonction source substrat.

Commentaire : L'effet canal court de MASTAR n'est pas programmé par les paramètres prévus par BSIM (**DVT0** et **DVT1**). Il est directement intégré au paramètre **VTH0** et le paramètre SCE_{BSIM} est annulé via (III-23) et (III-24).

Effet DIBL

Equations et paramètres :

$$\text{MASTAR : } DIBL_{\text{MASTAR}} = 1.8 \frac{\varepsilon_{\text{ox}}}{\varepsilon_{\text{si}}} \left(1 + \frac{X_j^2}{L_{\text{eff}}^2} \right) \frac{T_{\text{ox-el}}}{L_{\text{eff}}} \frac{T_{\text{dep}}}{L_{\text{eff}}} V_{\text{DS}} \quad (\text{III-26})$$

$$\text{BSIM : } DIBL_{\text{BSIM}} = \left[\exp\left(-\text{DSUB} \frac{L_{\text{eff}}}{2L_t}\right) + 2 \exp\left(-\text{DSUB} \frac{L_{\text{eff}}}{2L_t}\right) \right] \text{ETA0} V_{\text{DS}} \quad (\text{III-27})$$

$$\text{Solution : } \text{DSUB} = \text{DSUB0 (valeur arbitraire)} \quad (\text{III-28})$$

$$\text{ETA0} = \frac{0.8 \frac{\varepsilon_{\text{ox}}}{\varepsilon_{\text{si}}} \left(1 + \frac{X_j^2}{L_{\text{eff}}^2} \right) \frac{T_{\text{ox-el}}}{L_{\text{eff}}} \frac{T_{\text{dep}}}{L_{\text{eff}}}}{\left[\exp\left(-\text{DSUB0} \frac{L_{\text{eff}}}{2L_t}\right) + 2 \exp\left(-\text{DSUB0} \frac{L_{\text{eff}}}{2L_t}\right) \right]} \quad (\text{III-29})$$

Définition des variables : V_{DS} est la différence de potentiel drain source.

Commentaire : Pour programmer l'effet DIBL (et avoir au final $DIBL_{\text{BSIM}} = DIBL_{\text{MASTAR}}$), il a été décidé d'affecter une valeur arbitraire à **DSUB** de manière à réinjecter une partie du calcul du DIBL de BSIM dans le paramètre **ETA0**. La relation obtenue est donc :

$$DIBL_{\text{BSIM}} = \text{ETA0} \cdot K_{\text{BSIM}} = \frac{DIBL_{\text{MASTAR}}}{K_{\text{BSIM}}} \cdot K_{\text{BSIM}} = DIBL_{\text{MASTAR}}$$

Comportement sous le seuil

Equations et paramètres :

$$\text{MASTAR : } I_{DS_{MASTAR}} = I_{th} \frac{W}{L_{eff}} \left[1 - \exp\left(-\frac{V_{DS}}{v_{bi}}\right) \right] \exp\left\{ \frac{V_{GS} - V_{th} - \Delta}{\left(1 + \frac{C_{dep}}{C_{ox}}\right) v_{bi}} \right\} \quad (\text{III-30})$$

$$\text{BSIM : } I_{DS_{BSIM3}} = \mu_{\omega} C_{ox_el} v_{bi}^2 \frac{W}{L_{eff}} \left[1 - \exp\left(-\frac{V_{DS}}{v_{bi}}\right) \right] \exp\left\{ \frac{V_{GS} - V_{th} - \mathbf{VOFF}}{\left(1 + \frac{C_{dep}}{C_{ox}}\right) v_{bi}} \right\} \quad (\text{III-31})$$

$$\text{Solution : } \mathbf{VOFF} = \left(1 + \frac{C_{dep}}{C_{ox}}\right) v_{bi} \ln\left(\frac{\mu_{\omega} C_{ox_el} v_{bi}^2}{I_{th}}\right) - \Delta \quad (\text{III-32})$$

Définition des variables : Δ est la différence de tension de seuil entre les régimes d'inversion forte et sous le seuil, C_{dep} est la capacité de déplétion.

Commentaire : La façon de programmer le comportement sous le seuil avec BSIM est très similaire à celle de MASTAR. Seule la constante multiplicative du courant change. Il suffit alors de modifier le paramètre **VOFF** de manière appropriée comme indiqué dans l'équation (III-35).

Mobilité

Equations et paramètres :

$$\text{MASTAR NMOS : } E_{eff} = \frac{V_{GS} + V_{th,on}}{6T_{ox_el}} - \frac{V_{FB} + \phi}{6T_{ox_el}} \quad (\text{III-33})$$

$$\mu_{\omega} \left[\frac{\text{cm}^2}{\text{Vs}} \right] = 330 E_{eff}^{-0.3} \left[\frac{\text{MV}}{\text{cm}} \right] \quad (\text{III-34})$$

$$\mu_{\omega} \left[\frac{\text{cm}^2}{\text{Vs}} \right] = 1450 E_{eff}^{-2.9} \left[\frac{\text{MV}}{\text{cm}} \right] \quad (\text{III-35})$$

$$\text{MASTAR PMOS : } E_{eff} = \frac{V_{GS} + V_{th,on}}{9T_{ox,el}} - \frac{V_{FB} + \phi_s}{9T_{ox,el}} \quad (\text{III-36})$$

$$\mu_{ac} \left[\frac{\text{cm}^2}{\text{Vs}} \right] = 90 E_{eff}^{-0.3} \left[\frac{\text{MV}}{\text{cm}} \right] \quad (\text{III-37})$$

$$\mu_{sr} \left[\frac{\text{cm}^2}{\text{Vs}} \right] = 140 E_{eff}^{-1} \left[\frac{\text{MV}}{\text{cm}} \right] \quad (\text{III-38})$$

$$\text{NMOS \& PMOS : } \frac{1}{\mu_{ac}} = \frac{1}{\mu_{sr}} + \frac{1}{\mu_{sc}} \quad (\text{III-39})$$

$$\text{BSIM (mobmod=1) : } \mu_{eff} = \frac{U_0}{1 + \mathbf{UA} + \mathbf{UC} \cdot V_{BS} \left(\frac{V_{GS} + V_{th}}{T_{ox}} \right) + \mathbf{UB} \cdot \left(\frac{V_{GS} + V_{th}}{T_{ox}} \right)} \quad (\text{III-40})$$

$$\text{Solution : } \mathbf{U0} = \mu_{eff,MASTAR} \left(V_{GS} = V_{dd} \right) \quad (\text{III-41})$$

$$\mathbf{UA} = \mathbf{UB} = \mathbf{UC} = 0 \quad (\text{III-42})$$

Définition des variables : E_{eff} est le champ effectif dans le canal, $V_{th,on}$ est la tension de seuil en régime de forte inversion, μ_{ac} la mobilité correspondant au régime d'interaction avec les phonons acoustiques, μ_{sr} la mobilité en régime d'interaction des porteurs avec les rugosités à l'interface oxyde semi-conducteur, V_{dd} la tension d'alimentation.

Commentaire : Pour programmer la mobilité, aucun autre moyen que d'annuler les effets de pondération de la mobilité effective en fonction des polarisations n'a été trouvé. La mobilité effective est calculée pour un champ effectif correspondant au point de polarisation à l'état passant ($V_{GS} = V_{dd}$). L'annulation des dépendances de la mobilité en fonction des polarisations peut aboutir à une sous-estimation des courants à faible polarisation de grille. Cependant, ces courants à faibles polarisations ne jouent pas un rôle très important sur les délais (qui dépendent principalement des courants à forte polarisation). Par ailleurs, négliger la dépendance de la mobilité en fonction des polarisations est relativement usuel lors du développement de nouveaux modèles. Une telle approximation semble donc acceptable dans le cadre du développement d'un modèle prédictif.

c) Aspects non programmables sous BSIM

Différents aspects du modèle MASTAR ne sont pas directement programmable dans une carte modèle de type BSIM3. Ci-dessous sont décrits les moyens mis en œuvre afin de contourner ce problème.

Déplétion du polysilicium

Equations et paramètres :

$$\text{MASTAR : } T_{ox_el} = T_{ox_phys} + T_{Polydep} \quad (\text{III-43})$$

$$\text{BSIM : } V_{GS,eff} = V_{FB} + \varphi_{-} + \frac{q \text{ NGATE } \varepsilon_{ox} T_{ox}^2}{\varepsilon_{ox}^2} \left(\sqrt{1 + \frac{2\varepsilon_{ox}^2 (V_{GS} - V_{FB} - \varphi_{-})}{q \text{ NGATE } \varepsilon_{ox} T_{ox}^2}} - 1 \right) \quad (\text{III-44})$$

Solution : Ajustement itératif du paramètre **NGATE** de façon à atteindre le courant $I_{D,sat}$ de MASTAR.

Définition des variables : $T_{Polydep}$ est la profondeur de déplétion dans la grille, **NGATE** est le paramètre du dopage de grille de BSIM.

Commentaire : Pour les technologies à base de grille en polysilicium, le phénomène de déplétion du polysilicium doit être pris en compte. Cependant, la façon de prendre en compte ce phénomène est complètement différente entre MASTAR et BSIM. MASTAR suppose que la présence de la zone déplétée au niveau du polysilicium de grille induit une augmentation de l'épaisseur électrique de l'oxyde de grille de valeur fixe $T_{Polydep}$. Cette manière de procéder est vraie seulement pour $V_{GS}=V_{dd}$ (cas de MASTAR). Pour être valable pour toutes les polarisations, la profondeur de déplétion doit être recalculée pour chaque valeur de V_{GS} . Au lieu de recalculer cette valeur pour chaque polarisation, BSIM établit une chute de potentiel de grille variable $V_{GS,eff} - V_{GS}$ (équation (III-44)) basée sur la donnée du dopage du polysilicium de grille (**NGATE**). Afin de prendre en compte la déplétion du polysilicium de grille dans BSIM, il a été décidé d'ajuster manuellement le paramètre **NGATE** en se servant du courant I_{ON} (le paramètre **NGATE** est ajusté de manière à aligner le courant I_{ON} de BSIM sur celui de MASTAR à $V_{GS}=V_{dd}$).

Résistances d'accès

Equations et paramètres :

$$\text{MASTAR : } I_{D,sat} = \frac{I_{D,sat_0}}{1 + \frac{\gamma R_s I_{D,sat_0}}{V_{GT}} - \frac{R_s I_{D,sat_0}}{V_{GT} + \gamma_{et} E_c (1 + \gamma)}} \quad (\text{III-45})$$

$$\text{BSIM : } I_{D,sat} = \frac{I_{D,sat_0}}{1 + \frac{\gamma R_s I_{D,sat_0}}{V_{D,sat}}} \quad (\text{cas des résistances incluses dans le modèle}) \quad (\text{III-46})$$

Solution : Pas de programmation des résistances avec les paramètres du modèle.
Utilisation de résistances extérieures au modèle dans le fichier « netlist » (littéralement « liste des nœuds ») du modèle.

Définition des variables : $I_{D,sat}$ est le courant de saturation en tenant compte des résistances d'accès et R_s la résistance parasite série.

Commentaire : Plusieurs façons permettent de programmer les résistances d'accès sous BSIM. Soit l'effet des résistances est inclus dans l'équation de courant (cas de MASTAR) soit les résistances sont ajoutées *in fine* à l'extérieur du modèle et les effets de baisse de potentiel aux bornes de la source et du drain sont calculés par l'algorithme SPICE. Or, lorsque les résistances sont incluses dans l'équation de courant, il est nécessaire de faire certaines approximations sur les valeurs de chute de potentiel des côtés source et drain. Ces approximations étant différentes sous MASTAR et BSIM (équations (III-45) et (III-46)), il a été choisi de programmer les résistances à l'extérieur du modèle et d'effectuer une étape de calibration des valeurs par rapport à MASTAR (valeurs de résistance déterminées de manière itérative en faisant correspondre les courants de saturation).

Courant de grille

Equations et paramètres :

$$\text{MASTAR : } J_{gate} = 1.4 X_j W a e^{(bV_{GS}^2 + V_{GS})} e^{-T_{ox_phys}} \quad (\text{III-47})$$

BSIM : Pas de paramètres de courants de grille

Solution : Ajout de sources de courant dépendantes des polarisations dans la netlist du modèle selon l'équation (III-47)

Définition des variables : a , b , c et d sont des constantes d'ajustement du modèle et J_{gate} est le courant de grille entre la grille et la source.

Commentaire : BSIM3 n'offre pas la possibilité de programmer des courants de grille. Pour ajouter ce phénomène parasite, des sources de courant dépendant des polarisations (Figure III-5) ont été programmées à l'extérieur du modèle à la manière des résistances parasites.

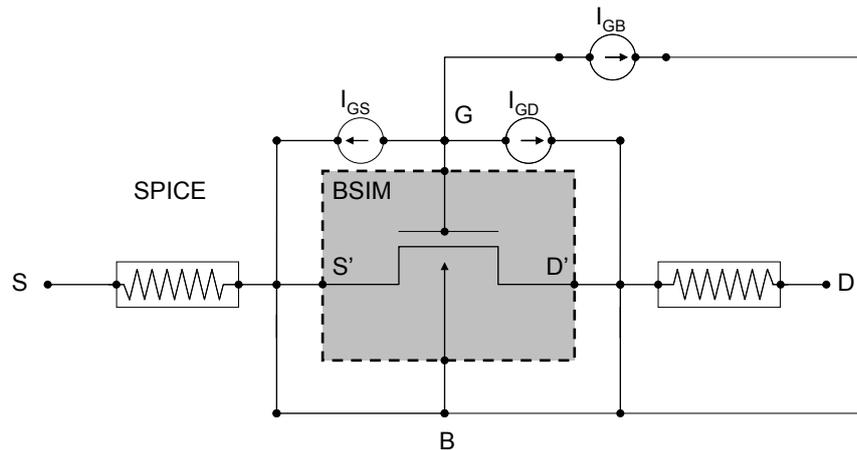


Figure III-5 : Modélisation des résistances d'accès et des courants de grille avec la méthodologie MSIM. Les résistances ne sont pas modélisées dans la carte modèle mais à l'extérieur de celle-ci. Les chutes de potentiel à la source et au drain sont alors calculées par l'algorithme SPICE. Les courants sont modélisés comme des sources de courant indépendantes.

d) Aspects dynamiques

La précision des simulations temporelles dépend en grande partie du modèle de charge utilisé par le modèle SPICE. Il est donc indispensable de ne pas négliger cet aspect qui est peu développé sous MASTAR.

Capacités intrinsèques

BSIM calcule la charge intrinsèque $Q_i(t)$ dans le canal de manière quasi-statique à partir des données géométriques et des paramètres précédemment définis (capacité d'oxyde, effet de substrat, ...). Un calcul de la répartition des charges de manière physique a été choisi (répartition « 40/60 » dans le cas saturé). C'est le cas le plus prédictif qui existe. Il n'est *a priori* pas obligatoire dans BSIM de programmer des paramètres spécifiquement pour le calcul des charges. Les paramètres programmés pour le calcul du courant doivent normalement suffire. Cependant, dans le cas où il n'est pas possible d'obtenir une

correspondance entre le modèle et la réalité à la fois pour les caractéristiques I-V et C-V, BSIM autorise de modifier certains paramètres pour le calcul des charges. Cette façon de procéder n'est pas physique puisqu'elle autorise la création de deux jeux de paramètres distincts (donc deux modèles différents) pour les régimes I-V et C-V. Etant donné que l'approche de modélisation de cette étude se veut proche de la physique, il a été décidé de ne pas utiliser de paramètres de correction entre les régimes I-V et C-V pour la création du modèle MSIM. BSIM calcule donc les caractéristiques C-V à partir seulement des paramètres I-V.

Capacités de recouvrement :

Equations et paramètres :

$$\text{MASTAR : } C_{\text{overlap}} = \Delta C_{\text{ox_el}} \quad (\text{III-48})$$

$$\text{BSIM : } \quad \mathbf{CGS0, CGD0}$$

$$\text{Solution : } \quad \mathbf{CGS0 = CGD0 = C_{\text{overlap}}} \quad (\text{III-49})$$

Définition des variables : C_{overlap} est la capacité de recouvrement, **CGS0** et **CGD0** sont les paramètres définissant les capacités de recouvrement respectivement côté source et côté drain.

Commentaire : La surface de recouvrement source/grille et grille/drain est immédiate sous MASTAR ($\Delta = 1.8 X_j$). Il est possible d'en déduire très rapidement la capacité de recouvrement résultant de la capacité formée par les deux plans parallèles extensions grille.

▪ Capacités de Miller

Equations et paramètres :

MASTAR : Pas de modèle.

$$\text{BSIM : } \quad \mathbf{CF}$$

$$\text{Solution : } \quad \mathbf{CF = \frac{2 \varepsilon_{\text{ox}}}{\pi} \ln \left(1 + \frac{t_{\text{poly}}}{t_{\text{ox_phys}}} \right)} \quad (\text{III-50})$$

Définition des variables : t_{poly} est la hauteur du polysilicium.

Commentaire : MASTAR n'ayant pas de modèle de capacité de Miller, un modèle de la littérature [34] a été retenu et implémenté. Ce modèle, unanimement reconnu, traduit la capacité existant entre deux plans métalliques perpendiculaires séparés d'une distance t_{ox_phys} dont l'un est infini et l'autre de hauteur t_{poly} .

Capacités de jonctions

Equations et paramètres :

MASTAR : Pas de modèle.

BSIM : **CJ, CJSW, PB, PBSW, CJSWG**

Solution : Annulation de tous les paramètres de jonction de BSIM et utilisation de modèles de jonctions existants.

Définition des variables : t_{poly} est la hauteur du polysilicium.

Commentaire : Comme les jonctions évoluent peu d'une génération à l'autre en terme de dopage contrairement à d'autres aspects du transistor comme le dopage de canal par exemple, il est légitime d'utiliser des modèles de générations précédentes à condition d'ajuster les données géométriques. Des jonctions extérieures au modèle ont ainsi été implémentées à la manière des résistances d'accès et des courants de grille.

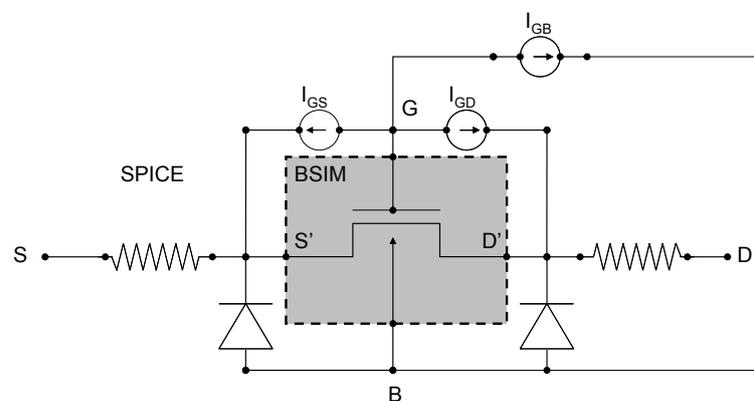


Figure III-6 : Modélisation des jonctions source et drain. Des modèles de jonctions de précédentes générations technologiques sont ajoutés à l'extérieur de la carte modèle.

e) Implémentation

Pour effectuer une simulation SPICE, le simulateur a besoin en entrée d'un fichier « liste de nœuds » (ou « netlist » en anglais) listant tous les éléments du circuit et les connexions associées. Dans ce fichier figurent également tous les paramètres de la simulation (type de

simulation, pas de résolution, paramètres d'extraction de données), ainsi que les fichiers « librairie » décrivant les modèles des éléments du circuit (principalement transistors et diodes). Le fichier librairie peut être écrit sous forme de « sous circuit » (« subckt »). Cette façon de procéder permet de rajouter facilement des éléments externes au modèle BSIM comme les résistances d'accès, les jonctions source et drain ou les courants de grille (Figure III-6). Par ailleurs, il est possible d'écrire des équations pour programmer les paramètres d'entrée du modèle. Un fichier modèle sous forme de « sous circuit » avec les équations décrites précédemment servant à programmer les paramètres du modèle a ainsi été élaboré (cf. Annexes p.149).

f) Sources d'inexactitude du modèle

Plusieurs sources d'incertitude peuvent contribuer à l'inexactitude du modèle prédictif.

Paramètre « DELTA »

Le paramètre **DELTA** est un paramètre de lissage de BSIM qui permet d'assurer la continuité des caractéristiques $I_{DS}-V_{DS}$ entre les parties ohmiques et saturées. Son origine est mathématique. C'est donc un pur paramètre d'ajustement qui n'est pas prévisible. Il n'y a donc d'autres choix que de le prendre égal à une valeur arbitraire. Il a été choisi de le prendre égal à 0,05 car c'est une valeur qui a permis d'ajuster correctement plusieurs générations technologiques (cf. ci-dessous). Par ailleurs, la Figure III-7 montre l'influence du paramètre **DELTA** sur le temps de transit d'un inverseur. Celui-ci ne dépend que très faiblement de **DELTA**. Le choix d'une valeur arbitraire ne semble donc pas problématique même dans le cadre d'un modèle prédictif.

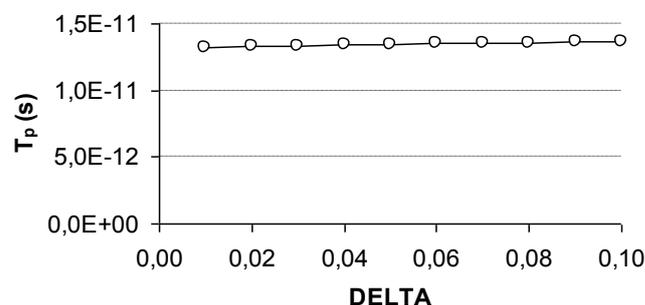


Figure III-7: Influence du paramètre **DELTA** sur le délai d'un inverseur.

Variation de la mobilité par polarisation de la grille

Comme spécifié au paragraphe b), la dépendance en V_{GS} de la mobilité a été omise. Cela amène à sous estimer quelque peu le courant pour les faibles polarisations de grille.

g) Prise en compte de la variabilité stochastique (désappariement)

Pour rendre compte de la variabilité aléatoire des caractéristiques des dispositifs, il est possible de programmer des variables aléatoires sur certains paramètres des modèles SPICE. Lors de simulations de type « Monte Carlo » ceux-ci sont alors tirés aléatoirement selon la loi de variation programmée. Pour le paramètre de concentration en dopant du canal N_B , une loi de variation gaussienne de valeur moyenne N_B et d'écart type σ_{NB} est programmée, avec :

$$\sigma_{NB} = \frac{2}{\sqrt{3}} \sqrt{\frac{N_B}{W L_{eff} T_{dep}}} \quad (\text{III-51})$$

Cette loi traduit le fait que le nombre de dopants suit une loi dite des grands nombres où l'écart type de la distribution en nombre de dopants n dans le canal est égale à la racine de ce même nombre [35][36] :

$$\sigma_n = \sqrt{n} \quad (\text{III-52})$$

III.2.E Validation du modèle MSIM

L'approche de modélisation est validée en trois temps :

- tout d'abord en comparant les résultats obtenus grâce à MSIM à ceux obtenus à l'aide du logiciel MASTAR,
- ensuite en vérifiant la cohérence du modèle par rapport aux modèles élaborés dans des technologies matures (90 & 65nm),
- enfin en évaluant l'écart du modèle par rapport aux données expérimentales pour des technologies avancées (45nm).

Cohérence entre MASTAR et MSIM

Etant donné d'une part la similitude des équations de courant entre MASTAR et BSIM, et d'autre part la programmation des principaux effets parasites (SCE, DIBL, mobilité, effet

de substrat, etc...), une égalité des courants entre MASTAR et BSIM est attendue lorsque les aspects non programmables de MASTAR dans BSIM (résistances parasites, déplétion du polysilicium de grille) ne rentrent pas en jeu. Cette égalité est vérifiée Figure III-8. Cette figure représente les caractéristiques I-V obtenues à partir d'un profil MASTAR d'un MOS obtenu sans ajustement itératif des résistance et de la polydéplétion (résistances séries nulles et utilisation d'une grille métallique). Les points I_{ON}/I_{OFF} prévus par MASTAR sont également reportés. Ces points coïncident parfaitement avec les caractéristiques simulées avec le modèle MSIM. Les courants calculés par MASTAR ont donc été reproduits fidèlement à l'aide d'une carte modèle de type BSIM.

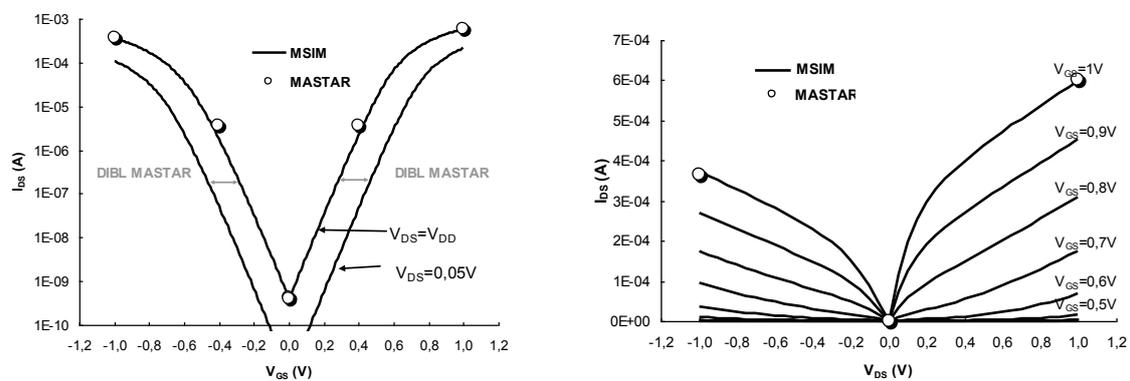


Figure III-8 : Caractéristiques ID-VG et ID-VD NMOS et PMOS obtenues grâce au modèle MSIM à partir d'un profil MASTAR sans déplétion de polysilicium ni résistances d'accès. Une parfaite concordance entre les courants I_{ON} et I_{OFF} de MASTAR et MSIM est observée. La valeur de DIBL calculée par MSIM est également correctement reproduite.

Cohérence entre MSIM et les modèles SPICE

Ci-dessous sont représentées les caractéristiques I-V obtenues avec la méthodologie MSIM et avec les modèles SPICE développés par les équipes de modélisation de STMicroelectronics pour deux générations technologiques matures (90 et 65nm). Les modèles MSIM ont été obtenus à partir de profils MASTAR représentatifs de la technologie considérée. Une très bonne correspondance est obtenue pour les I-V pour les deux technologies considérées. Le modèle de charge de MSIM est comparé à celui d'un modèle standard 65nm (cf. Figure III-11). Une bonne concordance est également observée.

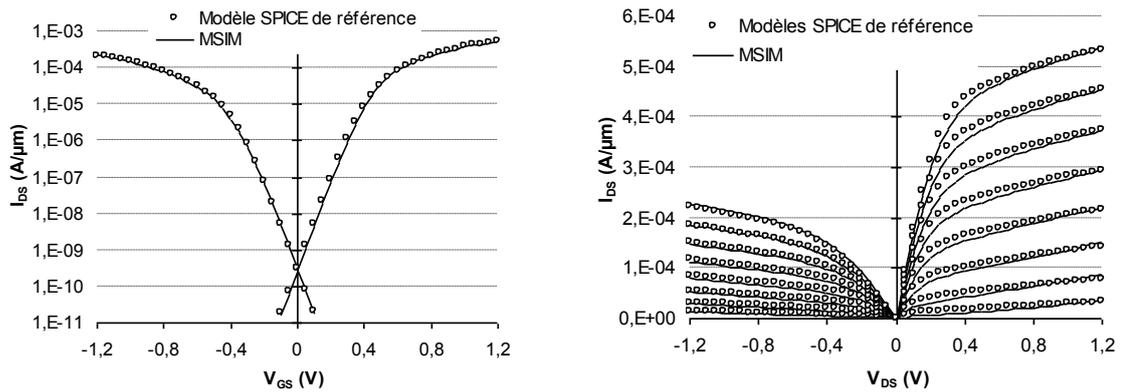


Figure III-9 : Caractéristiques ID-VG et ID-VD NMOS et PMOS obtenues grâce à la méthodologie MSIM à partir d'un profil MASTAR correspondant à une technologie 90nm. Les caractéristiques obtenues avec les deux modèles sont très proches.

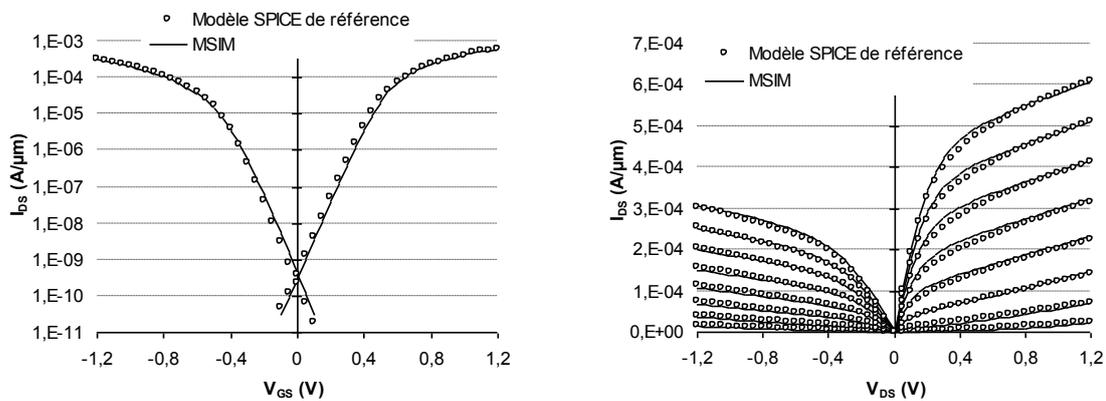


Figure III-10 : Caractéristiques ID-VG et ID-VD NMOS et PMOS obtenues grâce à la méthodologie MSIM à partir d'un profil MASTAR correspondant à une technologie 65nm. Les caractéristiques obtenues avec les deux modèles sont très proches.

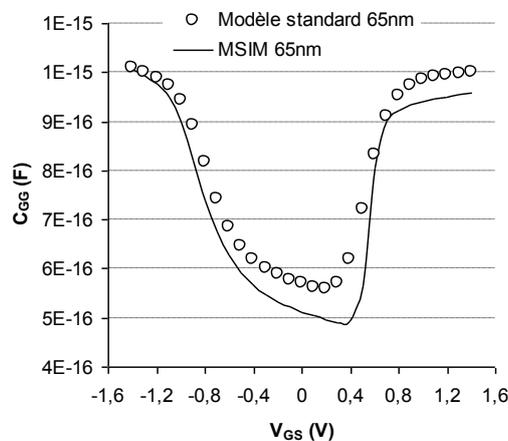


Figure III-11 : Caractéristique C-V obtenue à l'aide du modèle prédictif MSIM et du

modèle standard 65nm. Une bonne correspondance est observée entre les deux modèles.

Les résultats dynamiques sont présentés Figure III-12. Il s'agit de simulations de l'oscillateur en anneau représenté Figure III-12 a). Une bonne précision est obtenue à la fois pour les générations 90 et 65nm (+/- 4% d'erreur entre les deux modèles).

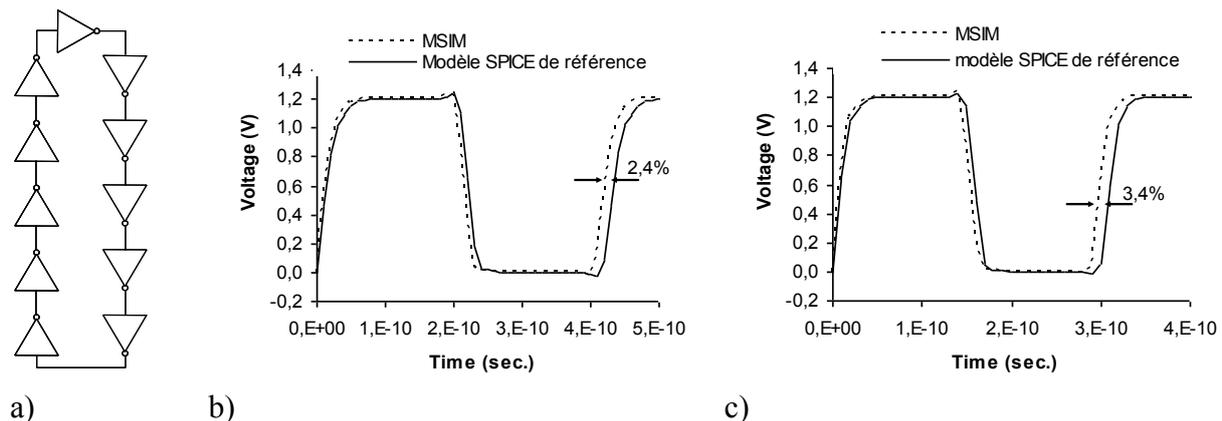


Figure III-12 : a) Oscillateur en anneau utilisé pour vérifier le comportement dynamique ; b) et c) : Simulation de l'oscillateur en anneau avec le modèle MSIM et avec les modèles de référence de la technologie considérée. b) technologie 90nm et c) technologie 65nm.

Cohérence entre MSIM et les données expérimentales

Afin de tester l'aptitude du modèle à prédire une technologie, un modèle MSIM est réalisé en 45nm à partir seulement de données 65nm. Ce modèle est comparé aux résultats silicium et à des modèles SPICE « tentative » (1^{er} modèle SPICE élaboré dans une technologie, cf. §III.2.C p.83) en 45nm. Les résultats sont donnés Figure III-13 et Figure III-14. Pour reproduire la dispersion du procédé de fabrication, des simulations avec plusieurs longueurs de grille ont été utilisées pour le modèle MSIM. Nous voyons que le modèle MSIM 45nm suit les tendances du silicium à la fois en statique et en dynamique, témoignant ainsi d'un bon degré de prédictibilité.

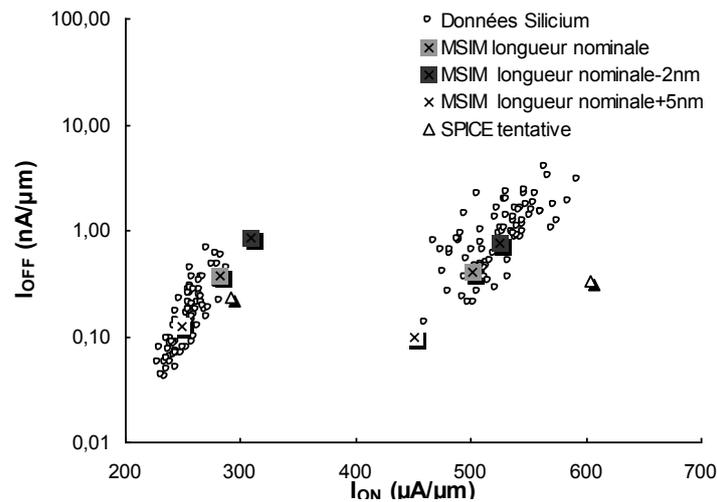


Figure III-13 : Comparaison des données I_{ON}/I_{OFF} en 45nm mesurées sur silicium et simulées grâce à MSIM avec différentes longueurs de grille.

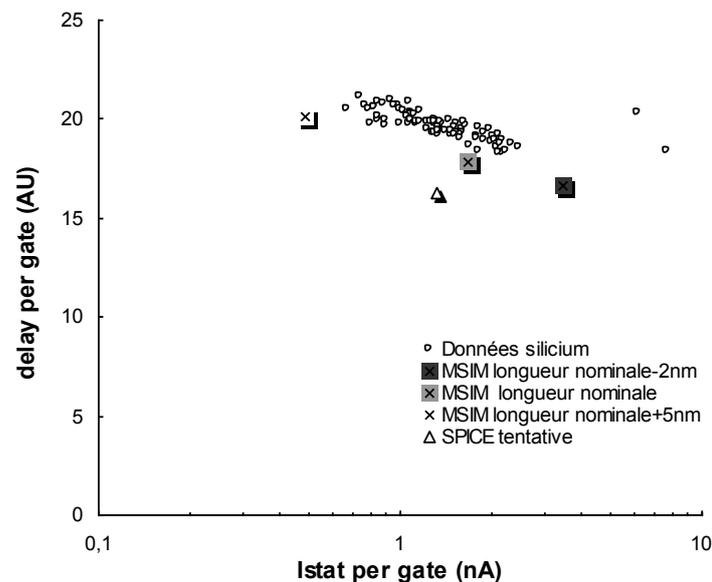


Figure III-14 : Comparaison du compromis délai d'oscillateurs en anneaux / courant de fuite en 45nm mesuré sur silicium et simulé grâce à MSIM avec différentes longueurs de grille.

III.3 Modèles d'interconnexion

Tous les fils ont été modélisés sous SPICE grâce à un réseau RC distribué comme illustré Figure III-15.

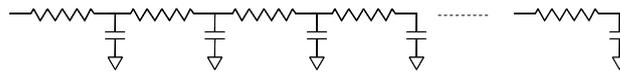


Figure III-15 : Réseau RC distribué pour modéliser les interconnexions sous SPICE. Le réseau est composé de 100 éléments RC.

Les capacités par unité de longueur des fils sont calculées grâce au logiciel HFO₂ [37] à partir des données matériau (permittivité effective k du diélectrique à ultra faible permittivité (ULK) et du diélectrique de recouvrement) et des données géométriques (pas de routage et rapports hauteur sur largeur) de l'empilement métal/diélectrique. La situation typique utilisée pour extraire les valeurs de capacité est illustrée Figure III-16. Il s'agit d'un fil avec deux voisins inséré entre deux plans de masse.

Les valeurs des résistances sont calculées à partir des données de résistivité effective (prenant en compte la barrière de diffusion) et des données géométriques figurant dans l'ITRS.

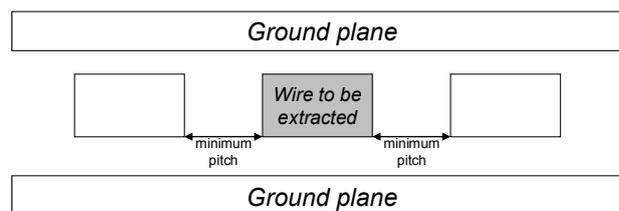


Figure III-16 : Situation typique d'une ligne d'interconnexion dans un bloc placé routé utilisée pour extraire les valeurs de capacité de fil.

III.4 DRM (« Design Rule Manual »)

III.4.A Règle de réduction d'échelle appliquée au DRM

La miniaturisation des technologies a conduit à une augmentation spectaculaire du nombre de règles du DRM, traduisant ainsi la complexité de la définition des règles de dessin de façon à garantir la fabrication des circuits. L'apparition de certaines règles peut avoir d'importants impacts sur la conception des circuits. A titre d'exemple, parmi les changements « brutaux » étudiés ces dernières années figure l'utilisation d'une seule orientation de grille en 45nm (règle qui a depuis disparu). N'étant pas à même de prédire l'apparition de ce genre de règles (qui dépendent des données de rendement extraites sur

silicium), il a été décidé de se baser seulement sur des règles existantes pour construire les DRM prédictifs.

D'une technologie à l'autre, les principales longueurs des circuits (longueur de grille, largeur du transistor, largeur des lignes d'interconnexions...) sont classiquement diminuées d'un facteur 0,7 (ou plus exactement $\sqrt{2} / 2$), et celles faisant intervenir des aires d'un facteur 2. En effet, en première approximation, la loi de Moore stipule que, lors du passage d'une technologie à l'autre, le nombre de transistors doit doubler pour une surface de circuit égale. Cela implique qu'à fonctionnalité identique (même nombre de transistors), la taille des circuits doit être diminuée par deux. Cela suppose donc que toutes les règles homogènes à des surfaces doivent être multipliées par un facteur 0.5, et, par voie de conséquence, toutes celles homogènes à des longueurs par un facteur $\sqrt{0.5} = \sqrt{2} / 2$.

III.4.B Paramétrage de l'outil d'implémentation physique

Les paramètres de la technologie pour l'outil de placement routage sont définis dans le fichier « techno.lef ». Toutes les règles de conception nécessaire à l'outil d'implémentation physique (dimensions des métaux, des vias, espacement des métaux, etc...) sont récapitulées dans ce fichier.

Pour établir un fichier « techno.lef » prédictif, un fichier « techno.lef » d'une technologie existante est utilisé. Les données homogènes à des surfaces et des longueurs sont repérées et ajustées comme décrit ci-dessus.

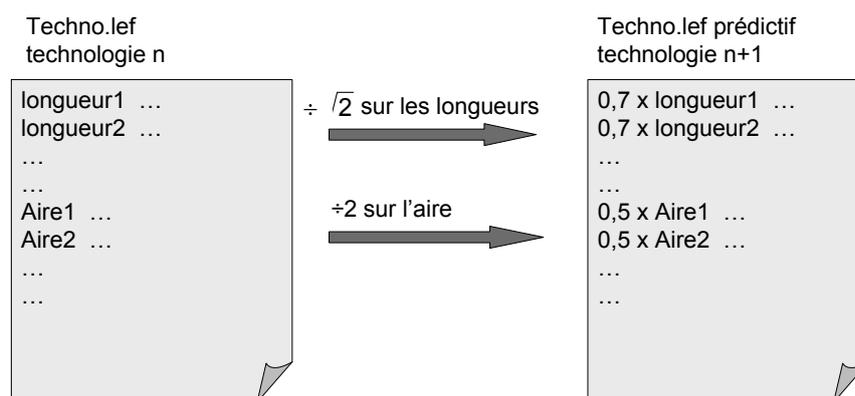


Figure III-17 : Génération d'un fichier « techno.lef » prédictif. Toutes les données relatives aux longueurs sont multipliées par 0,7 et toutes celles relatives aux aires par 0,5.

III.4.C Librairie de cellules standard

a) Dessin des masques / Règles de conceptions

Pour élaborer des bibliothèques de cellules standard prédictives, il faudrait en toute rigueur dessiner des cellules entièrement en flot « Full Custom » avec un DRM prédictif comme décrit au paragraphe III.4.A . Etant donné la façon d'obtenir le DRM prédictif, cela aboutirait directement aux cellules de la génération précédente, à un facteur d'échelle près. De manière à éviter un long travail fastidieux de conception de cellules standard, des facteurs géométriques sont directement appliqués sur les abstraits des cellules standard.

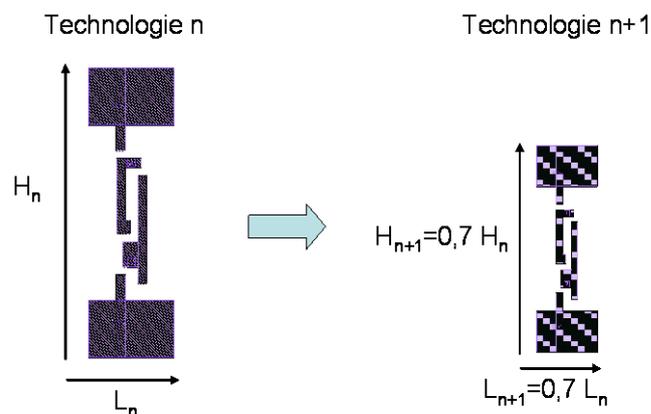


Figure III-18 : Génération d'un abstract de cellule standard d'une génération n+1 à partir de la génération n : toutes les dimensions sont diminuées d'un facteur 0,7.

b) Caractérisations électriques

De la même manière que les fichiers techno.lef prédictifs, les netlists de cellules standard sont obtenues à partir des netlists de la génération précédente en diminuant toutes les dimensions des transistors MOS d'un facteur 0,7 ou 0,5 selon qu'il s'agit de paramètres proportionnels à des longueurs ou bien des aires. Les facteurs à appliquer aux capacités et résistances de la netlist sont un peu plus compliqués. En effet, il faut distinguer deux sortes de capacités et résistances qui peuvent intervenir dans les circuits : les résistances et capacités parasites liées au transistor et les résistances et capacités d'interconnexion.

Concernant les capacités et résistances de la netlist, elles sont supposées proportionnelles aux longueurs de fil. Il est alors possible de les adapter de manière appropriée pour passer d'une génération à l'autre (cf. paragraphe III.3 page 103).

Une fois les netlists des cellules obtenues, l'étape de caractérisation est effectuée. Les gammes de valeurs de capacités sur lesquelles est effectuée la caractérisation sont pondérées d'un facteur 0,7 de manière à prendre en compte une réduction globale possible de la capacité minimale d'une génération à l'autre.

III.5 Conclusion

Une méthode de génération de kits de conception prédictifs a été présentée. Celle-ci regroupe différents aspects : modélisation SPICE prédictive de dispositifs et d'interconnexion, règles de réduction d'échelle du DRM, paramétrage des outils de conception et génération de bibliothèques de cellules standard. Les modèles MSIM reposent sur la traduction du logiciel MASTAR au sein d'une carte modèle BSIM. Les modèles d'interconnexion s'appuient sur un réseau SPICE RC-distribué basé sur les données d'évolution des résistances et des capacités des interconnexions. Des critères simples de réduction d'échelle applicables au DRM ont été définis et appliqués au paramétrage des outils de conception et à la génération de bibliothèques de cellules standard prédictives.

Chapitre IV Applications des critères d'évaluation aux futures technologies

Grâce aux critères d'évaluation technologiques introduits chapitre II et aux kits de conception prédictifs définis chapitre IV, il est possible de simuler les performances des futures technologies. Plusieurs évaluations de technologies hypothétiques sont ainsi réalisées : simulation de l'évolution au cours du temps d'un scénario fixé par l'ITRS et analyse poussée du prochain nœud 32nm. La première évaluation met en lumière des tendances de fond relatives à l'évolution des délais d'interconnexion qui n'avaient, jusqu'à présent, jamais été chiffrées de la sorte. Elle montre également une persistance des problèmes liés à la variabilité des dispositifs. La deuxième évaluation montre quant à elle une réduction des échelles convenable pour le nœud 32nm, où les problèmes mis en évidence au cours de la première évaluation n'ont pas encore un impact important. Elle démontre également la pertinence de l'utilisation d'une nouvelle technique pour réduire la variabilité des cellules mémoires SRAM par l'utilisation de dispositifs faiblement dopés.

IV.1 Introduction

Savoir évaluer l'ITRS au niveau circuit revêt un grand intérêt. En effet, l'ITRS est un consortium mondial regroupant de multiples acteurs du monde de la microélectronique (universitaires, fondeurs, équipementiers) qui a pour but d'assurer des avancées rentables au niveau des performances des circuits intégrés de façon à garantir la prospérité de cette industrie. L'ITRS détermine les principaux axes de recherche de ce domaine en fixant les objectifs à atteindre pour chaque métier (conception, test, intégration des dispositifs innovants, nouveaux dispositifs, procédés analogiques radiofréquences, nouveaux matériaux, lithographie, interconnexions, assemblage des usines, packaging, sécurité et environnement, amélioration des rendements, métrologie et modélisation). La partie intégration des dispositifs innovants (PIDS, « Process Integration Device and Structures ») de la feuille de route technologique n'inclut pas de critères d'évaluation au niveau des circuits. Un des objectifs de ce chapitre est donc d'utiliser l'ITRS, et en particulier la partie PIDS pour élaborer les modèles MSIM correspondant qui, associés aux données de la

partie « Interconnect » de l'ITRS, permettent d'évaluer l'évolution des performances au niveau des circuits.

Si l'évaluation à long terme est instructive, il est également nécessaire de savoir fournir des évaluations précises sur les technologies 32nm en cours de développement dans l'industrie, ce qui constitue le deuxième objectif de ce chapitre. Dans ce cas, les données d'entrée alimentant les kits prédictifs ne sont pas issues de l'ITRS mais des technologues travaillant sur l'intégration de ces technologies.

La première partie de ce chapitre est ainsi consacrée à l'évaluation de la variabilité des futurs points mémoires SRAM. Un scénario prévu par l'ITRS est tout d'abord étudié puis différentes options technologiques en 45 et 32nm sont évaluées. Un nouveau type de point mémoire innovant est finalement proposé et simulé. La deuxième partie du chapitre fournit les résultats des différents critères d'évaluation « Back End » définis précédemment pour un scénario tiré de l'ITRS dans un premier temps puis pour un scénario 32nm spécifique. Au cours de la troisième partie du chapitre, plusieurs implémentations de circuits sont effectuées d'une part pour évaluer l'impact de l'augmentation des résistances à technologie fixée, puis pour évaluer les performances du nœud 32nm.

IV.2 SRAM

IV.2.A Analyse de la feuille de route technologique ITRS

a) Elaboration des modèles

Pour une évaluation prédictive de SRAM, les données d'entrée nécessaires sont des modèles SPICE ainsi que des données relatives à la variabilité des dispositifs et à leur dimensionnement. L'ensemble de ces données est détaillé Tableau 1. Des modèles HP (« High Performance ») sont choisis. Le nœud technologique est relatif à la définition de l'ITRS (« half pitch metal 1 ») et ne correspond donc pas à la dénomination classique. Le dimensionnement des dispositifs a été choisi de manière à maintenir des ratios de courants constants pour les transistors de la cellule. La surface de la cellule résultante est indiquée et correspond à une évolution conforme à la loi de Moore (surface divisée par 2 à chaque génération). Les données relatives à la variabilité correspondent aux critères définis par l'ITRS concernant les variations de longueur de grille et d'épaisseur du film de silicium.

La loi de distribution du nombre de dopants est choisie conformément à celle définie par l'équation (III-51). Etant donné qu'il est difficile d'obtenir des informations précises sur l'amplitude des variations des travaux de sortie de grille, l'écart type de variation de travail de sortie de la grille est prise arbitrairement égale à 10mV pour les technologies polysilicium et 5mV pour les technologies à base de grille métallique (il est supposé que le passage en grille métallique permettra de diminuer l'ampleur des fluctuations).

Tableau 1 : Données d'entrée utilisées pour l'évaluation SRAM de la feuille de route ITRS.

Année	Nœud	Type de dispositif	W_{PU}/L_{PU} (nm)	W_{PD}/L_{PD} (nm)	W_{PG}/L_{PG} (nm)	σ_{L_g} (nm)	$\sigma_{T_{Si}}$ (nm)	σ_N	$\sigma\phi_{ms}$ (mV)	Surface de Bitcell (μm^2)
2005	90	Bulk	80/32	120/32	80/43	1,28	-	Distribution de Poisson [6] Pas de variation	10	0,761
2007	65	Bulk	65/25	98/25	34/65	1	-		5	0,501
2010	45	Bulk	45/18	68/18	45/24	0,72	-		5	0,241
2013	32	SOI	32/13	48/13	32/17	0,52	0,52		5	0,122
2016	22	DG	22/9	33/9	22/12	0,36	0,36		5	0,06

b) Résultats des simulations

Les résultats concernant l'évolution du rapport $\langle SNM \rangle / \sigma_{SNM}$ sont donnés Figure IV-1. Un rapport stable autour de 3 est observé quelque soit la technologie alors qu'un ratio de 7 est requis pour pouvoir fournir du rendement. Aucune amélioration majeure de la SNM ne doit donc être attendue d'après les données de l'ITRS pour les années à venir.

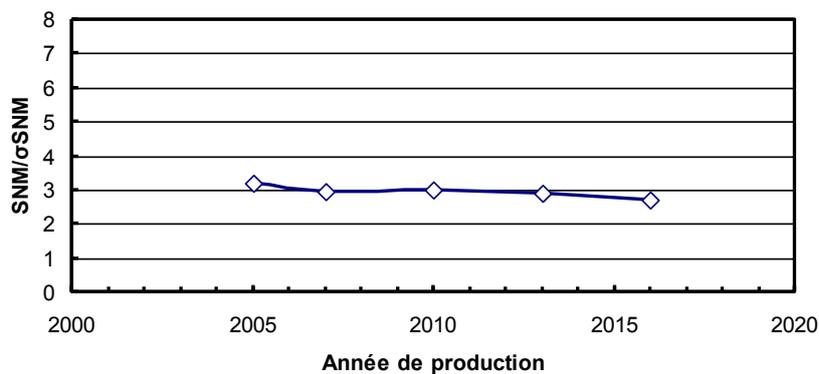


Figure IV-1 : Evolution du rapport $\langle SNM \rangle / \sigma_{SNM}$ en fonction de l'année de production.

IV.2.B Analyse des nœuds 45nm et 32nm

a) Introduction

D'après des études récentes [38], il a été montré que le contributeur majoritaire à la variation stochastique des caractéristiques des dispositifs est le dopage, ce dernier représentant environ 75% de ces variations. Afin de comprendre quels peuvent être les leviers pour réduire les conséquences des variations de dopage, plusieurs types de dispositifs sont simulés en 45nm et 32nm avec le dopage comme seule source de variabilité.

b) Résultats

Les Figure IV-2 et Figure IV-3 montrent les résultats de simulations en 45nm de différents types de dispositifs (Bulk, SOI, SON) pour une même taille de cellule ($0,25\mu\text{m}^2$). Le dispositif « Bulk » classique est le dispositif qui présente le plus de variabilité ($\langle SNM \rangle / \sigma_{SNM} = 4,7$ soit bien en dessous du critère $\langle SNM \rangle / \sigma_{SNM} > 7$ requis, et ceci en omettant 25% de la variabilité par des phénomènes autres que la variation du nombre de dopants). Les dispositifs sur film mince (SOI et SON) présentent de meilleurs résultats avec un net avantage pour le SON ($\langle SNM \rangle / \sigma_{SNM} = 11,6$ pour le SOI contre 6,4 pour le SON).

La Figure IV-4 rassemble des simulations de courbes en papillon pour différents types de dispositifs 32nm (« bulk » grille polysilicium, « bulk » grille métallique et SON grille métallique « midgap ») et différentes tailles de cellules. Le dispositif « bulk » avec grille polysilicium est celui qui présente le plus de dispersion suivi par le « bulk » grille métallique puis par le SON « midgap ». La dépendance en fonction de la taille de la cellule est également patente avec une augmentation des dispersions pour une réduction de la taille de la cellule.

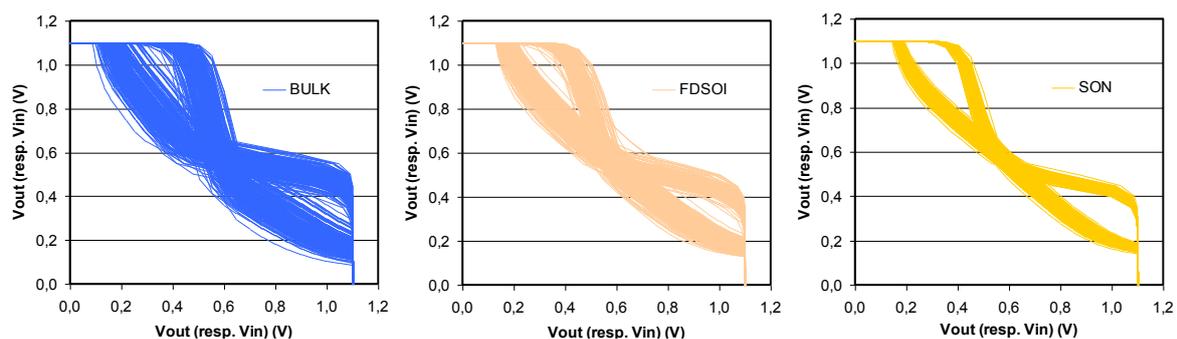


Figure IV-2 : Simulation de courbes en papillon pour trois types de dispositifs différents (« Bulk », SOI et SON) pour une cellule de technologie 45nm présentant une surface de 0,25µm².

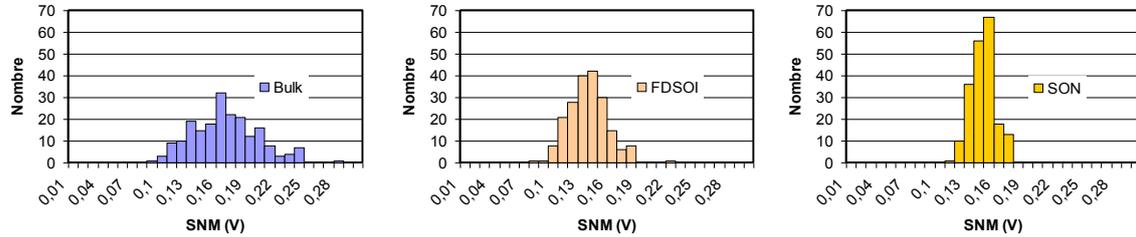


Figure IV-3 : Distributions de SNM associées aux simulations de la Figure IV-2.

Tableau 2 : Récapitulatif des résultats des simulations présentées Figure IV-2 et Figure IV-3.

	<SNM> (V)	σ_{SNM} (V)	<SNM>/ σ_{SNM}
Bulk	0,160	0,034	4,74
FDSOI	0,131	0,020	6,42
SON	0,139	0,012	11,62

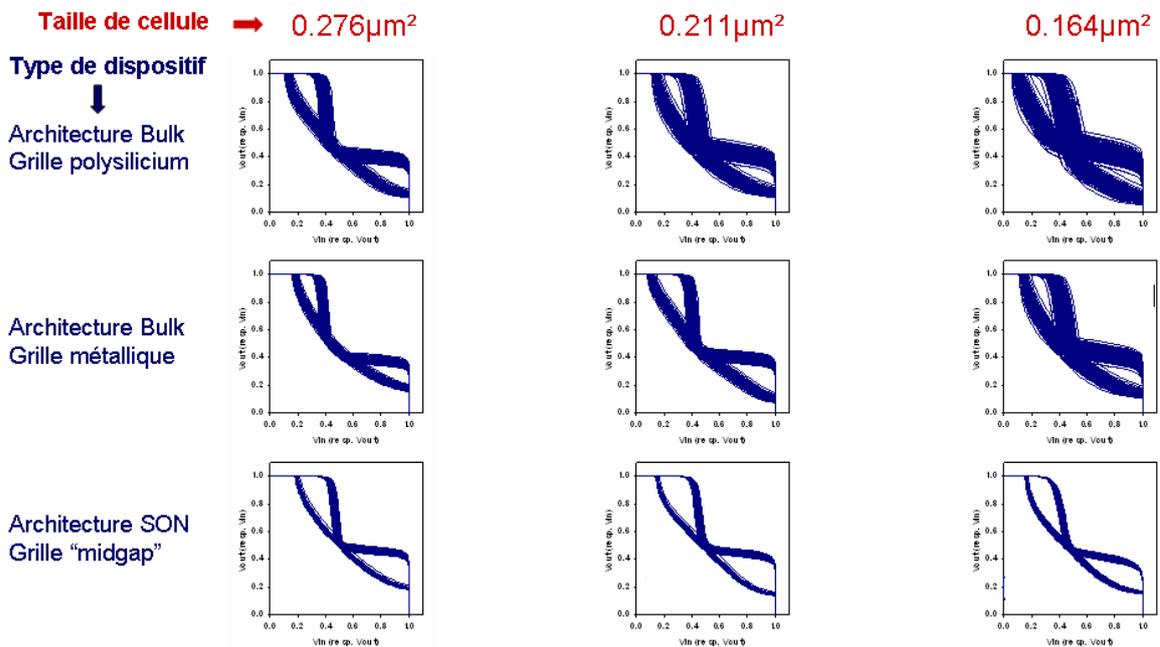


Figure IV-4 : Simulation de courbes en papillon en 32nm pour différentes tailles de cellules et différents types de dispositif.

c) *Interprétation*

Les résultats présentés au paragraphe précédent font apparaître deux propriétés :

1. plus les dimensions sont importantes et plus les dispersions sont faibles
2. plus le dispositif est dopé, plus les variations sont importantes.

La première propriété est explicable par le phénomène décrit par Pelgröm [39]. Lorsque deux dispositifs rectangulaires de taille identique (de largeur W et longueur L) sont séparés d'une distance D_x , la variance de la différence de n'importe quel paramètre P caractérisant ces dispositifs peut être mise sous la forme :

$$\sigma(\Delta) = \frac{A_p^2}{WL} + S_p^2 D_x^2, \quad (\text{IV-1})$$

avec A_p et S_p des constantes. Ainsi, plus les transistors sont grands, plus les variations de n'importe quel paramètre d'un transistor à l'autre sont faibles.

Pour illustrer le deuxième point, les variations de la quantité de dopants en fonction du dopage moyen ont été tracées grâce à la formule (III-51) pour deux des dispositifs simulés Figure IV-4 (« Bulk » avec grille en polysilicium et SON avec grille métallique « midgap »). Malgré le fait que le SON se distingue du « Bulk » par un niveau de dopage 2,5 fois moindre ($1,5 \cdot 10^{18}$ atomes/cm³ pour le SON « midgap » contre $3,7 \cdot 10^{18}$ atomes/cm³ pour le dispositif « Bulk »), la variation de quantité de dopants est à peu près identique pour les deux dispositifs ($\sigma_{\text{Nch}} = 0,3 \cdot 10^{18}$ atomes/cm³) à cause des différences de profondeur de déplétion. En traçant pour chacun des deux dispositifs la relation liant la tension de seuil au dopage grâce au logiciel MASTAR, il est possible de reporter les amplitudes de variation de tension de seuil résultant des variations de dopage. Etant donné que la tension de seuil du dispositif SON dépend moins de la concentration en dopants que le dispositif « Bulk », la variation résultante de tension de seuil est moindre.

Ainsi, il est possible d'observer une nette différence au niveau des variations de tensions de seuil entre les dispositifs « Bulk » et SON alors qu'il n'y a pas de différence au niveau de la variabilité de la concentration en dopants entre les deux dispositifs. Contrairement à ce qui est couramment affirmé, c'est donc avant tout la faible dépendance des dispositifs faiblement dopés vis-à-vis de la concentration en dopants plus que le faible nombre de dopant qui permet de minimiser les variations de tension de seuil.

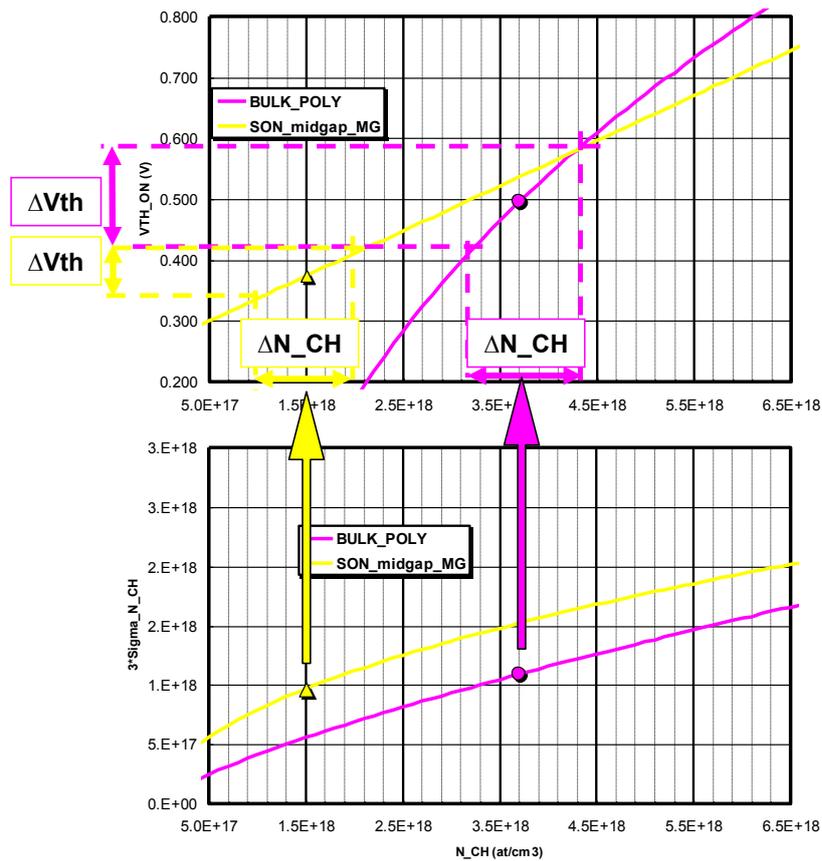


Figure IV-5 : Variation du nombre de dopants en fonction de la concentration moyenne en dopants (graphe du dessous) et variation de tension de seuil résultante (graphe du dessus) pour deux types de dispositifs (« Bulk » avec grille en polysilicium et SON avec grille métallique « midgap »). La variation du nombre de dopants est calculée d'après la formule (III-51) et la dépendance de la tension de seuil vis-à-vis de la concentration en dopants grâce au logiciel MASTAR.

IV.2.C Etude d'une nouvelle mémoire SRAM

a) Principe de la première innovation

Comme cela a été précisé précédemment, l'utilisation de dispositifs faiblement dopés permet de réduire les variations des caractéristiques des transistors. Par ailleurs, à cause de la réduction des dimensions et de la dépendance en $1/\sqrt{WL}$ de leur amplitude (cf. équation (IV-1)), il est de plus en plus difficile de contenir ces variations pour les technologies avancées. L'utilisation des dispositifs faiblement dopés est donc fortement requise. Or, l'intégration simultanée des transistors NMOS et PMOS faiblement dopés est complexe, surtout lorsque ceux-ci nécessitent l'intégration de deux grilles métalliques différentes. Compte tenu du fait que les caractéristiques des cellules SRAM dépendent principalement

des transistors NMOS [40], il peut sembler pertinent d'évaluer la robustesse à la variabilité d'une cellule faite uniquement de transistors NMOS faiblement dopés. Les transistors PMOS restant fortement dopés, ceci afin de minimiser les efforts d'intégration.

b) Résultats

Plusieurs simulations de points mémoires sont effectuées grâce au modèle MSIM en 22nm. Les données d'entrées ayant servi à la création des modèles sont les profils MASTAR ITRS (« Low Operating Power ») pour le nœud 22nm, contrairement à la précédente évaluation qui est basée sur la feuille de route HP. Seule la variabilité de la concentration en dopants est prise en compte.

La Figure IV-6 représente les caractéristiques d'une cellule mémoire 22nm simulée à partir des données de l'ITRS (cas 1, technologie « Low Operating Power » avec dispositifs NMOS et PMOS fortement dopés) et d'un scénario de réduction des échelles conforme à la loi de Moore (cellule de $0,0625 \mu\text{m}^2$). L'utilisation de technologies « bulk » classiques pour ce nœud très avancé ne laisse aucune chance d'obtenir une « fenêtre » SNM (les lobes de la courbe en papillon sont complètement fermés). En passant en technologie faiblement dopée (cas 2, Figure IV-7), la fenêtre s'ouvre fortement. Cependant, comme discuté précédemment, cela se fait au prix d'un effort d'intégration important à la fois pour le NMOS et le PMOS. En n'utilisant une technique de faible dopage seulement pour le NMOS (cas 3, Figure IV-8), les simulations montrent un gain important au niveau des critères $\langle SNM \rangle / \sigma_{SNM}$, $\langle I_{read} \rangle / \sigma_{I_{read}}$ et $\langle WM \rangle / \sigma_{WM}$ par rapport au cas où les deux dispositifs sont fortement dopés (cas 1), sans toutefois atteindre les performances en terme de SNM lorsque les deux types de dispositifs sont faiblement dopés (cas 2). Cependant, il n'y a pas de différence notable au niveau WM et I_{read} entre les cas 2 et 3 car ces deux critères dépendent quasiment exclusivement des transistors NMOS. Ainsi il peut être tout à fait intéressant d'utiliser des dispositifs faiblement dopés seulement pour les NMOS pour les futures SRAM, les dispositifs PMOS faiblement dopés n'étant pas indispensables.

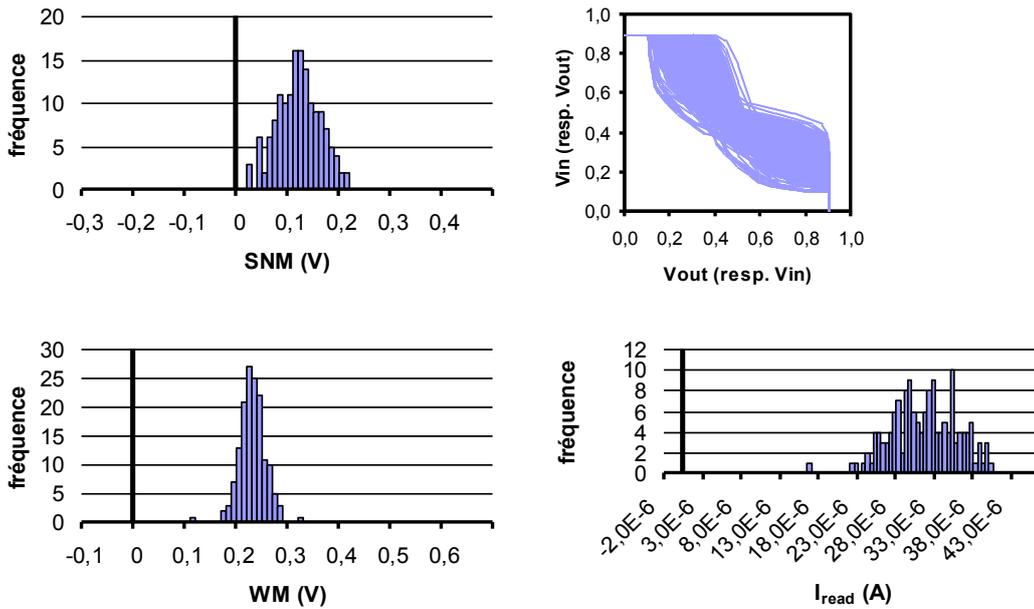


Figure IV-6 : Distribution de la SNM, de la WM, du courant I_{read} ainsi que courbe en papillon d'une cellule mémoire 22nm (32nm « half pitch », année 2012) avec dispositifs NMOS et PMOS non dopés élaborée d'après l'ITRS LOP. Seule la variation du nombre de dopants est prise en compte.

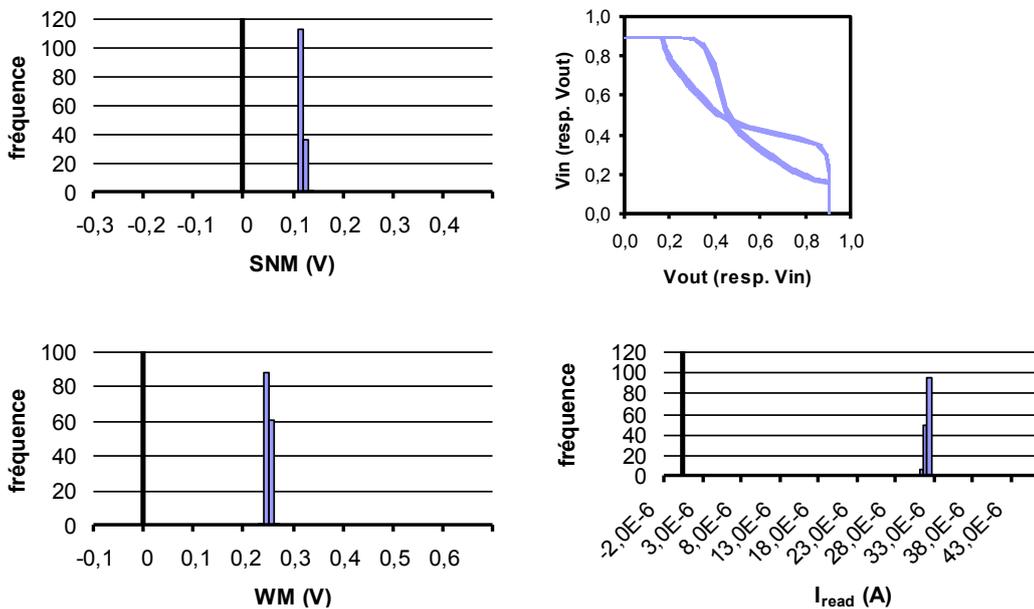


Figure IV-7 : Distribution de la SNM, de la WM, du courant I_{read} ainsi que courbe en papillon d'une cellule mémoire 22nm (32nm « half pitch », année 2012) avec dispositifs NMOS et PMOS dopés élaborée d'après l'ITRS LOP. Seule la variation du nombre de dopants est prise en compte.

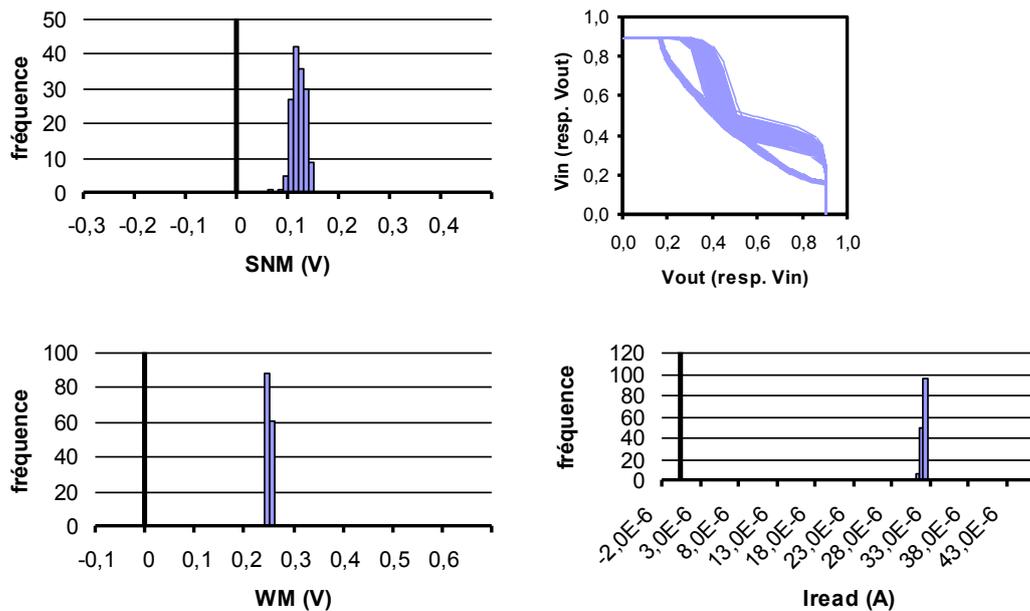


Figure IV-8 : Distribution de la SNM, de la WM, du courant I_{read} ainsi que courbe en papillon d'une cellule mémoire 22nm (32nm « half pitch », année 2012) avec dispositifs NMOS non dopé et PMOS dopé élaborée d'après l'ITRS LOP.

c) Deuxième innovation

Afin d'améliorer encore les performances de la cellule mémoire dans le cas d'un transistor NMOS faiblement dopé et d'un transistor PMOS fortement dopé, il est possible d'envisager de diminuer l'impact de la variabilité du transistor PMOS en diminuant sa conductance. Cette opération peut être réalisée concrètement de plusieurs manières : non siliciuration des sources/drain, dégradation de la mobilité par implantation de différents matériaux dans le canal... La Figure IV-9 donne les résultats des simulations concernant les critères $\langle SNM \rangle / \sigma_{SNM}$, $\langle I_{read} \rangle / \sigma_{I_{read}}$ et $\langle WM \rangle / \sigma_{WM}$ dans le cas de transistors NMOS et PMOS fortement dopés (cas 1) pour différents facteurs de dégradation de la mobilité. La Figure IV-10 fournit les mêmes résultats dans le cas d'un transistor NMOS faiblement dopé et d'un transistor PMOS fortement dopé (cas 3). Si réduire la conductance du transistor PMOS a peu d'impact dans le cas 1, il existe un optimum de dégradation de la conductance du PMOS qui permet de maximiser le ratio $\langle SNM \rangle / \sigma_{SNM}$ dans le cas où seul le transistor NMOS est faiblement dopé (cas 3). La valeur seuil de 7 permettant d'aboutir à des rendements significatifs est alors nettement dépassée. Cet optimum traduit le compromis existant entre la dégradation naturelle de la SNM par la réduction de la conductance du transistor PMOS (éloignement d'un point de fonctionnement optimal) et la réduction de l'influence de la variabilité de ce même transistor.

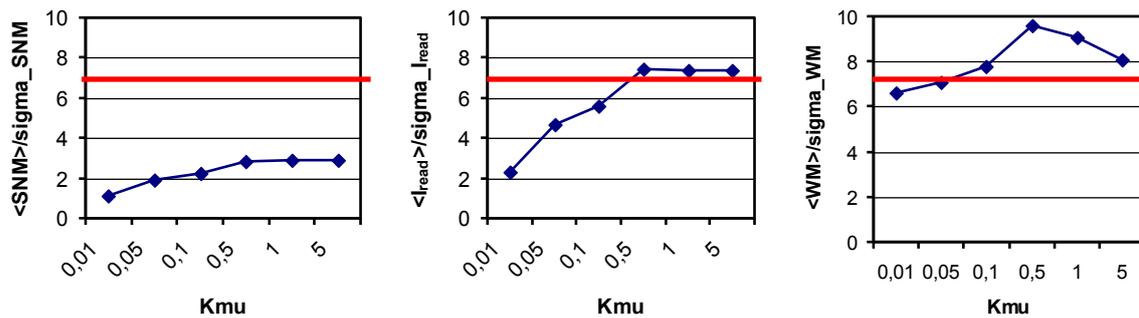


Figure IV-9 : Evolution des critères $\langle SNM \rangle / \sigma_{SNM}$, $\langle I_{read} \rangle / \sigma_{I_{read}}$ et $\langle WM \rangle / \sigma_{WM}$ d'une cellule 22nm avec transistors NMOS et PMOS dopés en fonction du facteur de dégradation de la mobilité des transistors PMOS. Seule la variation du nombre de dopants est prise en compte.

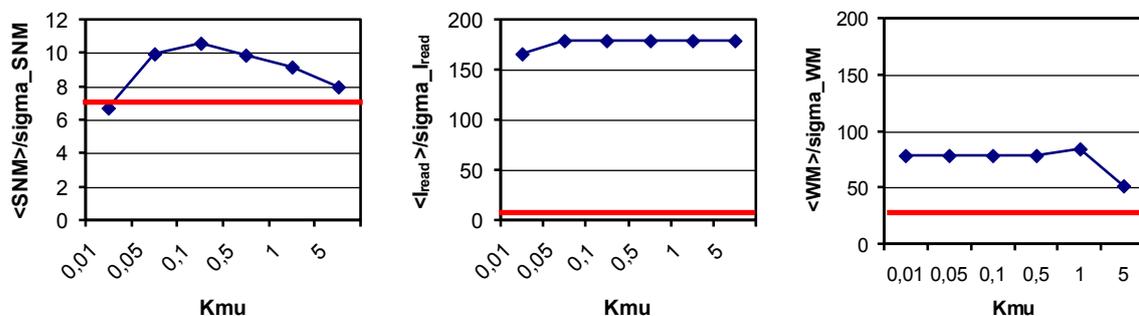


Figure IV-10 : Evolution des critères $\langle SNM \rangle / \sigma_{SNM}$, $\langle I_{read} \rangle / \sigma_{I_{read}}$ et $\langle WM \rangle / \sigma_{WM}$ d'une cellule 22nm avec transistors NMOS non dopés et PMOS dopés en fonction du facteur de dégradation de la mobilité des transistors PMOS.

IV.2.D Conclusion

Les cellules SRAM de dernières générations présentent une variabilité importante de leurs caractéristiques pouvant conduire à l'absence de rendement en production. L'évaluation de la feuille de route ITRS grâce aux modèles MSIM montre qu'aucun gain significatif n'est à attendre pour les années à venir. Pour les nœuds en cours de développement (45 et 32nm), deux leviers d'action permettant de réduire la variabilité se distinguent : augmentation des dimensions de la taille de cellule et utilisation de transistors faiblement dopés. L'augmentation de la taille des cellules est en contradiction complète avec la loi de Moore, elle ne peut donc constituer une solution viable. Le recours à des dispositifs faiblement dopés est prometteur, d'autant plus qu'il est montré que cette solution peut se limiter aux

seuls transistors NMOS. Dans ce dernier cas, un autre levier d'action permettant de réduire la variabilité apparaît. Il consiste à réduire l'impact de la variabilité du transistor PMOS en diminuant sa conductance.

IV.3 Critères « Back End »

IV.3.A Scénario ITRS

a) Données d'entrée

Les données relatives aux dispositifs simulés dans le cadre de l'évaluation « back end » de la feuille de route ITRS sont présentés Tableau 3. Les modèles de dispositifs sont générés à partir des données hautes performances de la section PIDS de l'ITRS et les modèles d'interconnexion à partir des données de résistivité et d'empilements de diélectriques décrits dans la partie « Interconnect ».

Tableau 3 : Récapitulatif des données utilisées pour simuler les critères d'évaluation technologiques « back end » de la feuille de route ITRS. Les modèles choisis sont tous issus de la section haute performance de l'ITRS.

Année	Noeud "half pitch"(nm)	Architecture du dispositif	Résistance d'interconnexion (Ohm/mm)	Capacité d'interconnexion (pF/mm)
2001	130	bulk	0,3	216
2005	90	bulk	1,8	190
2007	65	bulk	3,9	166
2010	45	bulk	10,8	158
2013	32	FDSOI	24,3	136
2016	22	DG	60,4	131
2020	14	DG	200,6	114

b) Oscillateur en anneau

La simulation avec MSIM de l'évolution du délai moyen par porte d'un oscillateur en anneau est donnée Figure IV-11. L'évolution des délais prévus par l'ITRS selon le critère CV/I est également reportée. Les deux courbes (ITRS CV/I et fréquence d'oscillateur en anneau prévue par MSIM) suivent la même tendance d'évolution (+17/an) avec un

décalage constant. Le décalage est du au facteur de proportionnalité existant entre le délai de l'oscillateur en anneaux et le critère CV/I (cf (II-1)).

La même tendance d'évolution observée montre la cohérence entre l'approche de simulation MSIM et l'approche par MASTAR utilisée par l'ITRS.

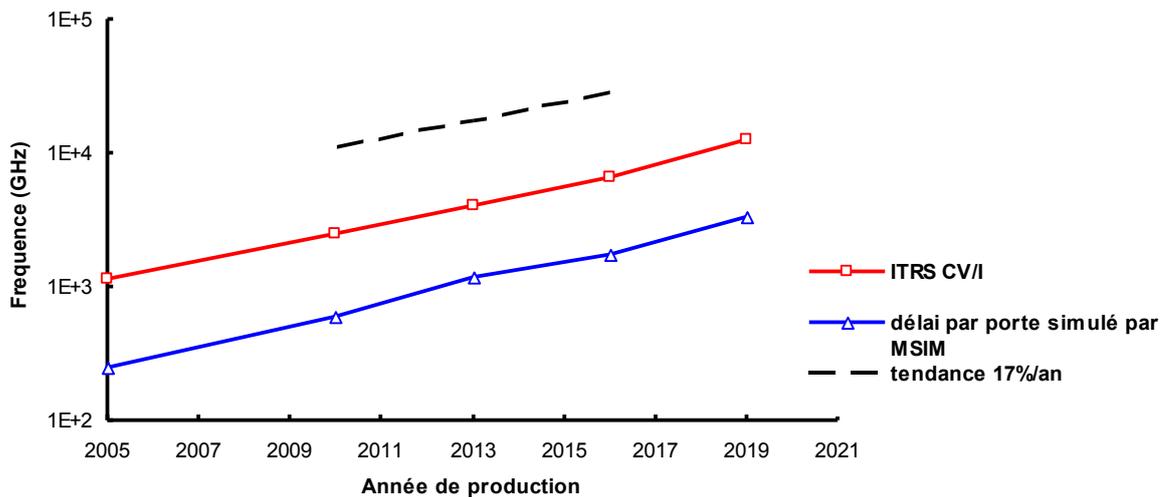


Figure IV-11 : Evolution de la fréquence d'oscillateur en anneau simulé par MSIM d'après les données de l'ITRS HP et tendance CV/I prévue par MASTAR. Une parfaite concordance entre les deux modèles est observée.

c) Délai d'un bloc placé routé

L'évolution du délai d'un additionneur 10 bits en fonction de l'année de production est donnée Figure IV-12. De la même manière que pour l'oscillateur en anneau, une tendance de réduction des délais proche de 17%/an est observée. Il n'y a donc pas d'effets particuliers liés à l'augmentation de la résistance des interconnexions pour ce type de bloc. Une des explications possibles est que le bloc n'est pas assez grand (environ $50\mu\text{m} \times 50\mu\text{m}$ en technologie 65nm), et donc les lignes d'interconnexion trop courtes pour mettre en évidence des problèmes liés à l'augmentation des délais d'interconnexion. Un circuit de taille plus importante semble donc être requis. Or la simulation d'un circuit plus complexe est trop coûteuse en termes de temps de simulation dans le cadre de simulations SPICE.

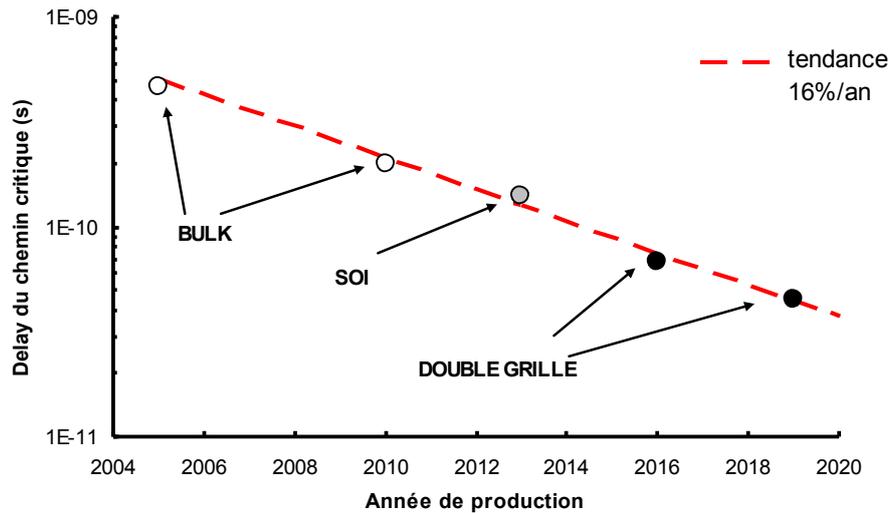


Figure IV-12 : Délai du chemin critique d'un additionneur 10 bits en fonction de l'année de production. Les simulations sont basées sur des facteurs géométriques appliqués à un bloc placé routé en technologie 65nm ainsi que sur les données de l'ITRS concernant les dispositifs HP et les données de capacité et résistance d'interconnexion extrapolés d'après la partie « interconnect ».

d) Délais de ligne

Afin de percevoir plus précisément l'impact de l'augmentation des résistances de ligne, des simulations de délais de ligne sont effectuées conformément au schéma électrique décrit Figure II-14. Les résultats de ces simulations sont donnés Figure IV-13. Plusieurs points marquants se dégagent :

- Pour chaque technologie, il existe une longueur d'interconnexion qu'il est possible de qualifier de « critique » à partir de laquelle les délais augmentent fortement.
- Cette longueur critique diminue avec la taille de la technologie.
- Alors que les délais diminuent de génération en génération pour les très faibles longueurs de ligne, l'inverse se produit pour les grandes longueurs de lignes.

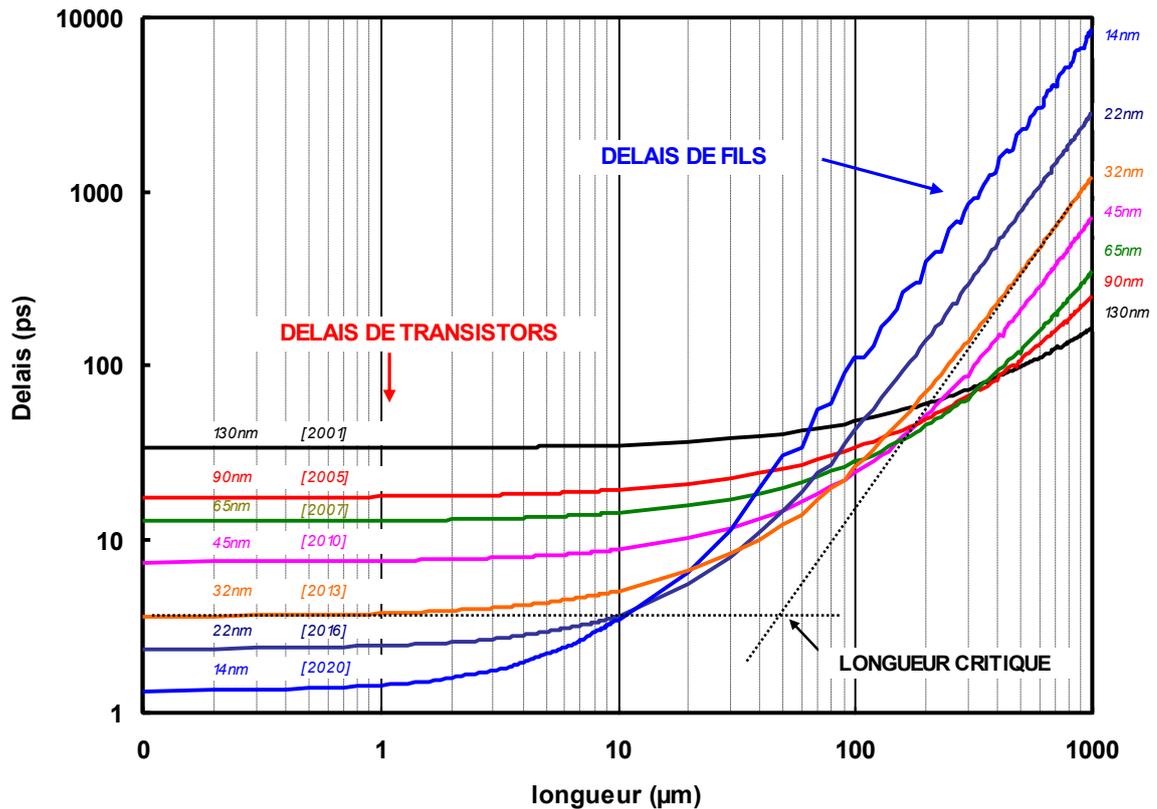


Figure IV-13 : Délai de ligne en fonction de la longueur pour différentes technologies de la feuille de route ITRS.

Il est possible de définir la longueur critique comme l'abscisse du croisement des deux asymptotes de la relation Délais=f(longueur). L'évolution de la longueur critique selon cette définition est donnée Figure IV-14. Une réduction continue de la longueur critique en fonction de l'année de production de la technologie est observée à la fois en valeur absolue mais également en valeur relative. La longueur critique passe ainsi de 776 μ m en technologie 2001 (soit 5000 carrés de métal pour cette technologie) à seulement 11 μ m en technologie 2020 (soit environ 700 carrés de métal pour cette technologie).

Cela suggère que les problèmes liés aux délais dans les interconnexions vont s'accroître pour les futures technologies non seulement dans le cas où les longueurs de fils ne diminuent pas d'échelle (augmentation de la fonctionnalité à bloc de taille fixe) mais également lorsque les longueurs diminuent d'un facteur $1/\sqrt{2}$ (même description comportementale d'une technologie à l'autre par exemple).

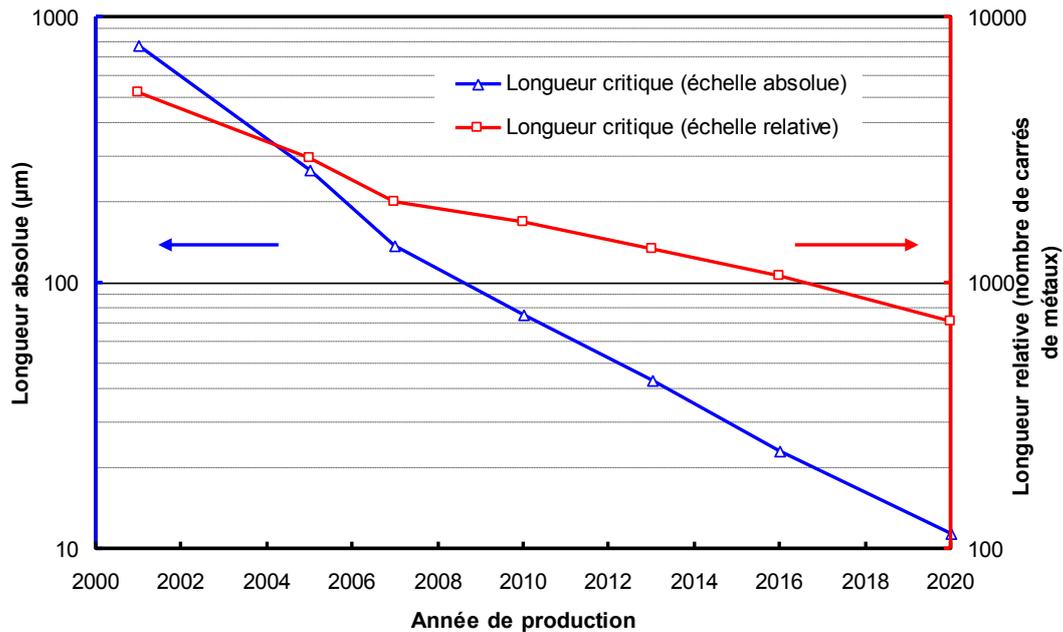


Figure IV-14 : Evolution de la longueur critique en fonction de l'année de production d'après les données de l'ITRS. La longueur critique est exprimée en valeur absolue (μm) et également en valeur relative (nombre de carrés de métaux).

e) Nombre optimal de répéteurs

Pour contrer les problèmes mentionnés ci-dessus, une solution consiste à s'interdire les interconnexions supérieures à la longueur critique. Pour cela, il faut segmenter les longues lignes en segments de longueur inférieure à la longueur critique grâce à l'ajout de répéteurs. Afin de savoir quel est le nombre optimal de répéteurs permettant de minimiser les délais de ligne, le circuit de la Figure II-15 est simulé. Les résultats sont donnés Figure IV-15. A nouveau, plusieurs points saillants se distinguent :

- Pour chaque technologie il existe un nombre optimal de répéteurs permettant de minimiser les délais.
- Le nombre optimal de répéteurs augmente fortement passant de 1 répéteur en technologie 90nm « half pitch » à plus de 100 en technologie 14nm « half pitch » pour une longueur de fil de 1mm.
- Le délai minimal absolu ne diminue pas mais augmente légèrement pour chaque technologie.

Ces trois points suggèrent que des problèmes importants d'ajout de répéteurs vont se poser pour les futures technologies, pénalisant non seulement la surface des circuits mais

également les délais comme le laissent penser les résultats de la Figure IV-15. Afin de quantifier au premier ordre quel peut être cet impact, le nombre de lignes supérieures à la longueur critique a été calculé sur un circuit de type contrôleur de port USB. Le nombre de répéteurs requis a été chiffré. Deux cas ont été étudiés : distribution de longueurs de fils identique à celle d'un contrôleur USB 65nm et diminution des longueurs de tous les fils selon la loi de Moore. Ces données sont résumées Tableau 4 et Figure IV-16. Dans le premier scénario, l'augmentation du nombre de répéteurs requis est très importante passant de 147 répéteurs (0,3% du nombre total de cellules) en technologie 65nm à plus de 4141 en technologie 2020 (8,3% du nombre total de cellules). Dans le cas du deuxième scénario, le nombre de répéteur est contenu (705 répéteurs soit 1,4% du nombre total de cellules en technologie 2020 contre 147 répéteurs soit 0,3% du nombre total de cellules en technologie 2005)

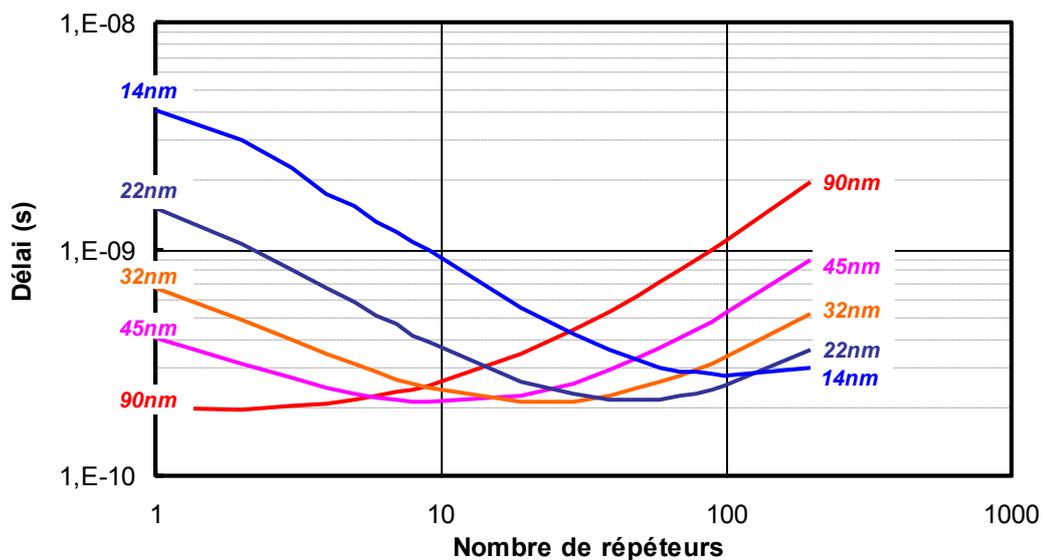


Figure IV-15 : Délai en fonction du nombre de répéteurs pour une ligne de longueur fixe et pour différentes technologies de la feuille de route ITRS.

Tableau 4 : Tableau récapitulant l'impact de l'augmentation de la résistivité des interconnexions en termes de répéteurs pour un circuit contrôleur de port USB 65nm.

Deux scénarios sont étudiés : distribution des longueurs de fils constante et distribution des longueurs de fils identique à un facteur de réduction d'échelle près.

Technologie		2005	2007	2010	2013	2016	2020
Longueur critique (µm)		265	137	76	43	23	11
Hypothèse de distribution de longueurs de fils équivalente au port USB 65nm	Nombre de répéteurs requis en valeur absolue	147	313	593	1081	1997	4141
	Nombre de répéteurs requis en valeur relative (% du nombre total de cellules)	0,3%	0,6%	1,2%	2,2%	4,0%	8,3%
Hypothèse de réduction des échelles par rapport à une distribution de longueur d'un port USB 65nm	Nombre de répéteurs requis en valeur absolue	147	212	279	363	474	705
	Nombre de répéteurs requis en valeur relative (% du nombre total de cellules)	0,3%	0,4%	0,6%	0,7%	0,9%	1,4%

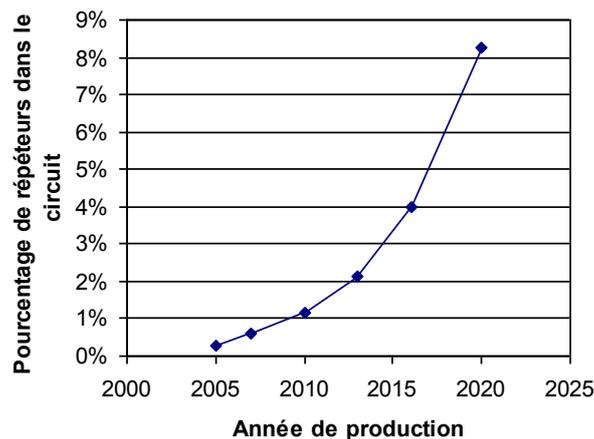


Figure IV-16 : Impact du scénario d'augmentation de la résistance d'interconnexion de l'ITRS sur le pourcentage de répéteurs requis dans le circuit dans le cas d'une distribution de longueur de fils équivalente à celle d'un contrôleur de port USB 65nm.

f) Influence de la conductance des cellules (« drive »)

Afin de mieux percevoir les modifications induites par l'augmentation des résistances d'interconnexion, une troisième dimension, la conductance de la cellule de charge de la ligne, est ajoutée aux simulations de la Figure IV-13. Les résultats sont donnés Figure IV-18. Ces résultats mettent en lumière un changement progressif de l'influence de la taille des cellules chargeant la ligne. En effet, pour les technologies allant de 2001 à 2007, les délais dépendent à la fois de la longueur d'interconnexion mais également de la taille des

transistors de charge. Il existe donc à ce niveau deux leviers permettant de diminuer les délais (découpage des lignes avec des répéteurs et augmentation de la conductance des cellules de charge). Or, pour les technologies au-delà de 2010, l'influence de la largeur des transistors de charge est fortement réduite jusqu'à devenir quasiment inexistante pour les technologies 2016 et 2020. L'utilisation de conductances fortes dans le cas de lignes de longueur supérieure à la longueur critique de la technologie considérée n'a donc aucun impact sur les délais.

Ce phénomène peut être expliqué en considérant que la propagation du signal à travers la ligne est modélisée par le schéma de la Figure IV-17. Dans ce cas, le temps caractéristique de charge du circuit vaut

$$\tau = (R_{ON} + R_{interco})(C_{interco} + C_{charge}), \quad (IV-2)$$

avec R_{ON} la résistance équivalente du dispositif, $R_{interco}$ la résistance d'interconnexion, et $C_{interco}$ la capacité de ligne. Lorsque la résistance d'interconnexion est faible ($R_{interco} \ll R_{ON}$), le délai est dominé par le produit de la résistance de ligne et de la somme de la capacité d'interconnexion et de la capacité de charge :

$$\tau = R_{ON}(C_{interco} + C_{charge}). \quad (IV-3)$$

Par contre lorsque la résistance de ligne devient très grande devant la résistance du dispositif ($R_{interco} \gg R_{ON}$), le délai est indépendant de R_{ON} :

$$\tau = R_{interco}(C_{interco} + C_{charge}). \quad (IV-4)$$

Dans ce cas, diminuer R_{ON} (en augmentant la largeur W du transistor par exemple) n'a aucun impact sur les délais.

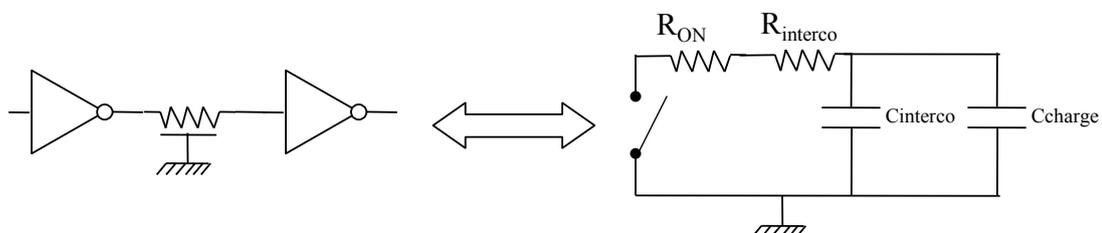


Figure IV-17 : Equivalence électrique de la ligne de transmission. R_{ON} est la résistance équivalente du transistor, $R_{interco}$ la résistance de ligne, $C_{interco}$ la capacité de ligne et C_{charge} la charge de la cellule réceptrice.

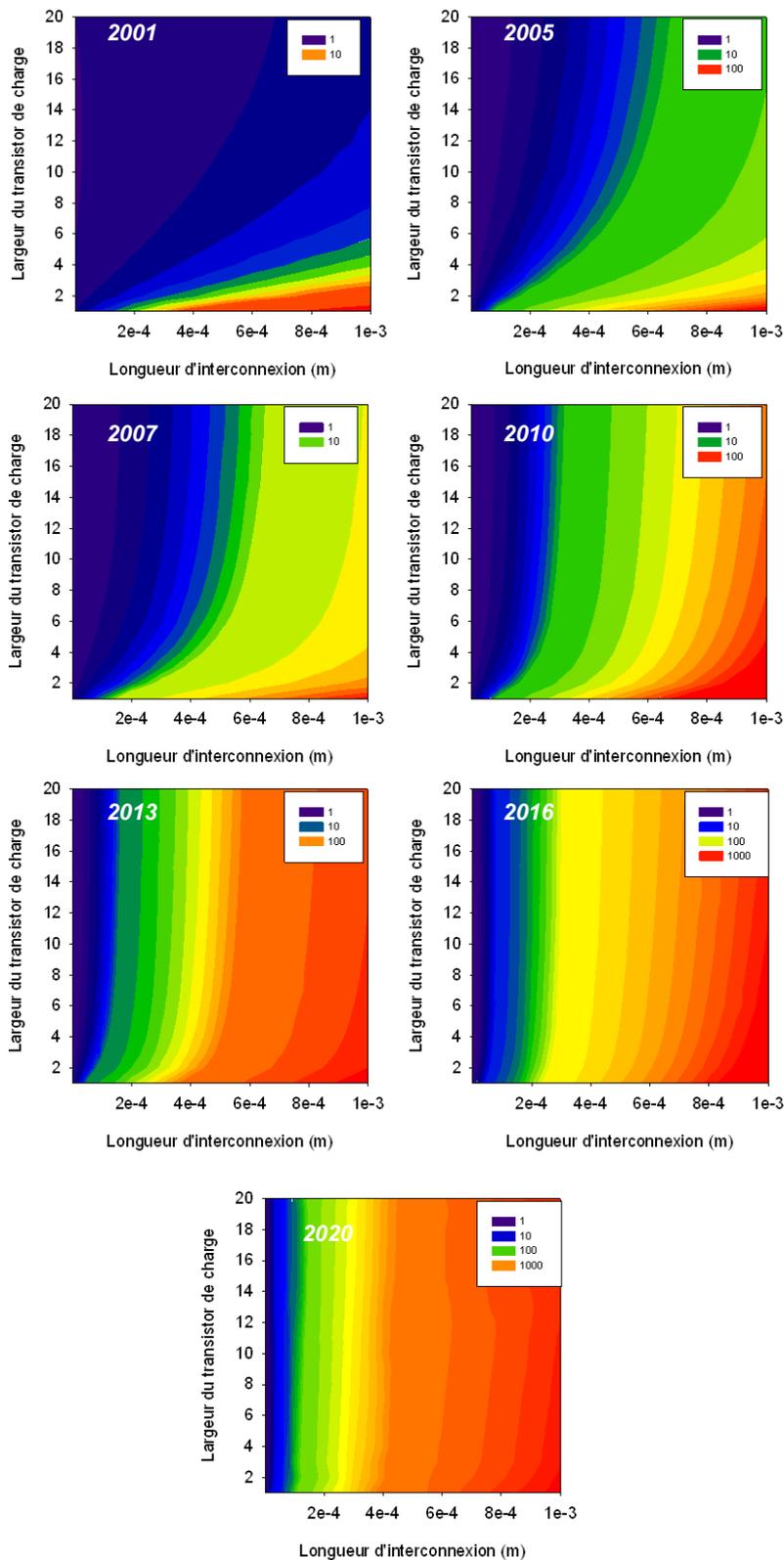


Figure IV-18 : Délai de ligne en fonction de la longueur de ligne et de la sortance de la cellule chargeant la ligne pour différents scénarios de la feuille de route ITRS. Les délais sont exprimés en nombre de fois le délai optimal (cas de la plus faible longueur de ligne et du plus large transistor).

IV.3.B Analyse du nœud 32nm

Les résultats précédents montrent les grandes tendances de fond d'évolution des technologies : diminution des délais de petits blocs conformément à l'oscillateur en anneau, diminution de la longueur critique d'interconnexion, augmentation des délais de ligne même en considérant l'ajout de répéteurs. Afin de comprendre dans quelle mesure ces évolutions vont influencer à court terme les technologies 32nm en cours de développement dans l'industrie, il convient d'appliquer les critères définis précédemment de manière plus spécifique. Pour cela, un modèle MASTAR est élaboré en 32nm LP (« Low Power ») d'après les spécifications fixées pour ce nœud. Les détails des paramètres utilisés sont livrés en annexe p. 149. Ce modèle est ensuite traduit en modèle SPICE MSIM. L'ensemble des critères d'évaluation « back end » est appliqué aux nœuds 65nm et 45nm à la fois en utilisant les modèles standard et des modèles prédictifs MSIM élaborés pour ces nœuds. Le nœud 32nm est évalué seulement à l'aide du modèle MSIM (les modèles SPICE pour ce nœud n'existant pas encore). Les résultats sont résumés de la Figure IV-19 à la Figure IV-22. Pour tous les critères d'évaluation, une bonne concordance est observée entre les modèles MSIM prédictifs et les modèles standard pour les nœuds 65nm et 45nm.

Concernant le 32nm, une diminution des délais d'oscillateur en anneau de 24% pour le FO1 et de 30% pour le FO3 est attendue (Figure IV-19). La puissance dissipée ramenée à la fréquence (PPF) suit également une forte tendance de diminution (-53% pour FO1 et FO3). Le délai du chemin critique de l'additionneur 10 bits suit également une tendance de diminution mais plus faible que l'oscillateur en anneau (-9%). Cependant, la tendance de diminution est conforme à celle observée pour le passage du 65nm au 45nm (-10,4%). Finalement, la longueur critique d'interconnexion diminue fortement (-39%) passant de 330 μ m à seulement 201 μ m. Une hausse de plus de 11% du délai optimal de propagation le long de lignes avec répéteurs est également observée (306ps/mm pour le 32nm contre 275ps/mm pour le 45nm).

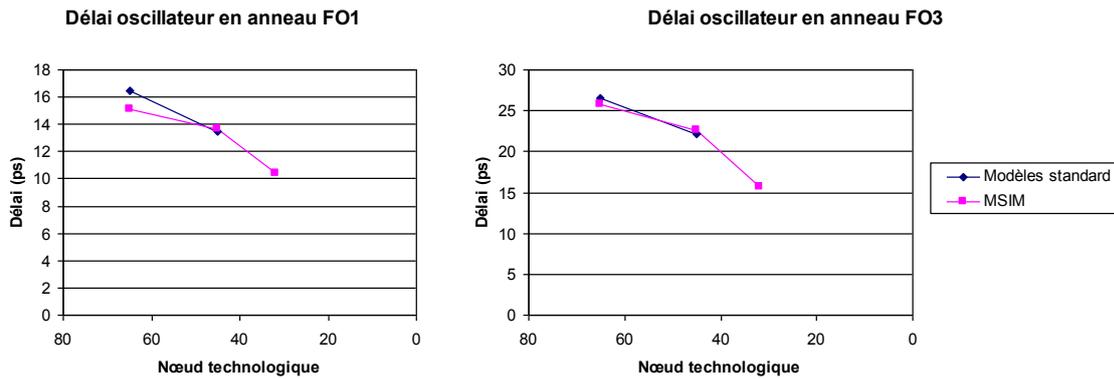


Figure IV-19 : Délais d'oscillateur en anneau FO1 et FO3 en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.

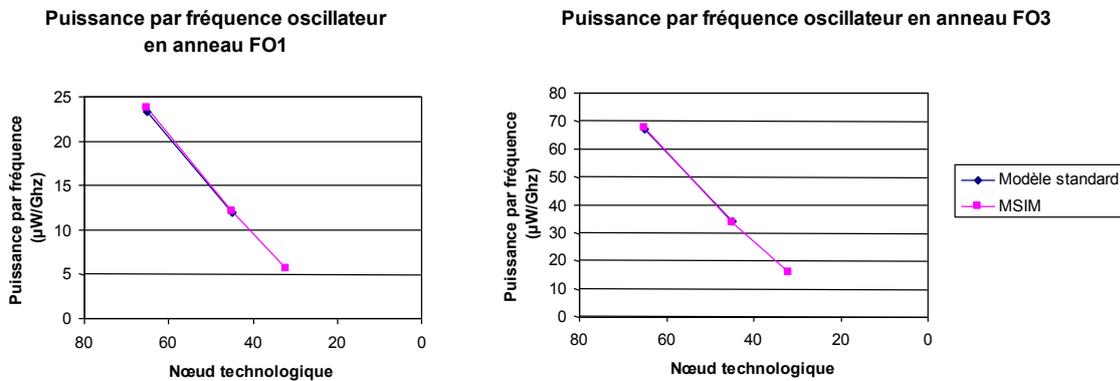


Figure IV-20 : Puissance par fréquence d'un oscillateur en anneau en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.

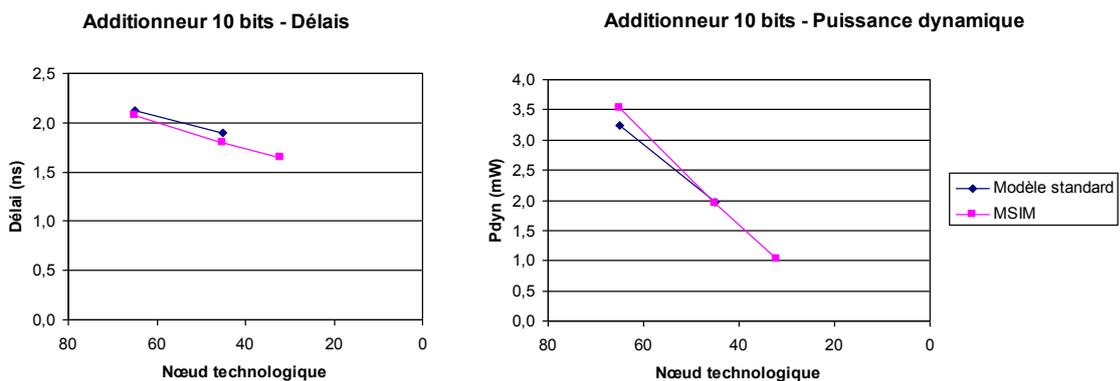


Figure IV-21 : Délai et puissance dynamique d'un additionneur 10 bits en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.

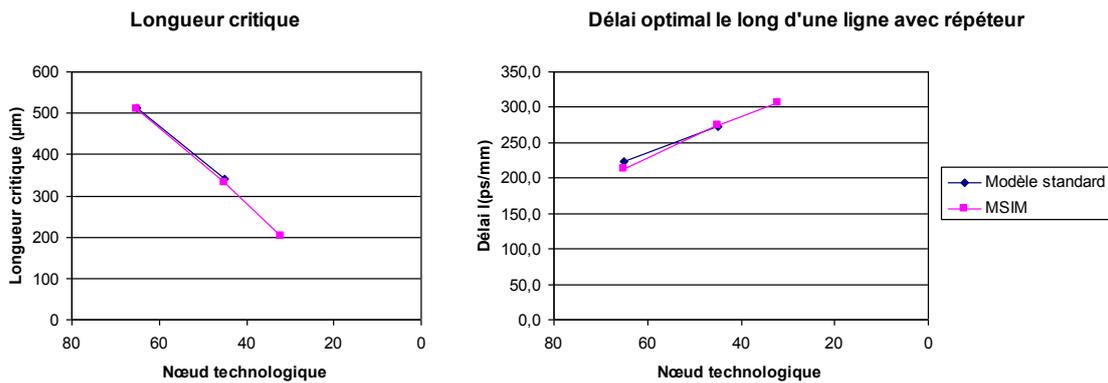


Figure IV-22 : Longueur critique d'interconnexion et délai optimal le long d'une ligne avec répéteurs en fonction du nœud technologique simulé grâce aux modèles standard en 65nm et 45nm et grâce aux modèles prédictifs MSIM en 65nm, 45nm, et 32nm.

IV.4 Flot digital prédictif

IV.4.A Introduction

Afin de mieux chiffrer l'impact des phénomènes parasites mis en évidence précédemment (augmentation des délais de ligne, augmentation du nombre de répéteurs, stagnation des délais avec répéteurs), il convient de ne pas se limiter aux simulations de circuits élémentaires (oscillateurs en anneau, lignes d'interconnexion) et d'élargir les simulations prédictives à l'implémentation de blocs entiers. Dans un premier temps, seules les résistances ont été modifiées par rapport au flot de conception classique selon les données prévues par l'ITRS, le reste des paramètres restant identique. Cette première évaluation a pour but de discriminer l'impact de l'augmentation des résistances d'interconnexion.

Dans un deuxième temps, un flot de conception prédictif est réalisé. Dans ce cas, toutes les modifications nécessaires à l'élaboration d'un flot prédictif décrites Chapitre III sont effectuées. Ce flot est tout d'abord accompli en 65nm et 45nm, à la fois de manière standard avec les outils CAO développés pour ces technologies, et également avec les outils de conception prédictifs élaborés durant cette étude, de manière à permettre la validation du flot prédictif. Une fois validé, ce dernier est alors utilisé pour l'évaluation du nœud 32nm.

IV.4.B Impact de l'augmentation des résistances en technologie 65nm

Il est possible d'affecter des facteurs de dégradation ou d'augmentation des résistances et capacité d'interconnexion sur les outils d'implémentation physique. Afin de rendre compte de l'augmentation des résistances pour les futures technologies, plusieurs implémentations d'un contrôleur USB sont réalisées en technologie 65nm en faisant varier le coefficient de pondération des résistances de l'outil d'implémentation selon un scénario prévu par l'ITRS (cf. Tableau 5).

Tableau 5 : Facteur de pondération des résistances d'interconnexion prévu par l'ITRS par rapport au nœud 65nm

Technologie (définition traditionnelle)	Année	Facteur de pondération
65	2005	1
25	2012	10
17	2015	26
15	2016	30

Les données relatives aux rapports d'implémentation sont fournies Tableau 6. Seul le cas d'une augmentation d'un facteur 30 des résistances montre une impossibilité totale de l'outil à router le bloc selon les contraintes fixées. La densité du circuit augmente de manière croissante avec le facteur de dégradation des résistances. Celle-ci semble être directement liée à l'augmentation du nombre de « Buffers » dans le circuit (cf. Figure IV-23). Ce fait est confirmé par l'analyse de la répartition du type de cellules dans le circuit (Figure IV-24). Seul le nombre de « Buffers » change d'une implémentation à l'autre. Celui-ci passe de 2% du nombre total de cellules pour le scénario standard 65nm à plus de 23% pour le scénario 2016. L'estimation théorique du nombre de « buffers » requis donnée Figure IV-16 semble donc sous-estimée dans le cas d'une réalisation pratique. De plus il est possible de constater que l'outil essaie bien d'éviter les longues lignes en insérant des « buffers ». La distribution des longueurs des fils s'éloigne ainsi de la loi de Rent classique à mesure que les résistances augmentent. Le nombre de fils de longueur intermédiaire se voit ainsi renforcé alors que le nombre de longs fils chute (Figure IV-25 et Figure IV-26). Le cas du scénario 2016 où le nombre de lignes de forte longueur augmente à nouveau laisse supposer que l'outil n'a pu réaliser complètement pour ce scénario la tâche de diminution des longues lignes par manque de place pour ajouter des « buffers » (densité de circuit de plus de 90%). Ceci expliquerait pourquoi le nombre de violations de

contraintes de délai est très important pour ce scénario (1959, cf. Tableau 6), une grande quantité d'informations devant circuler par des fils de longueur supérieure à la longueur critique.

Tableau 6 : Rapport d'implémentation du bloc contrôleur USB en 65nm en fonction du scénario d'augmentation des résistances utilisé. WNS est le « Worst Case Negative Slack » et TNS le « Total Negative Slack » (cf. annexe p. 164).

Scénario de résistance		x1 (2005)	x10 (2012)	x26 (2015)	x30 (2016)
SETUP	WNS (ns)	4,666	3,483	0,458	-1,089
	TNS (ns)	0	0	0	-1,089
	Nombre de violation	0	0	0	1
HOLD	WNS (ns)	-0,003	-0,009	-0,038	-215000
	TNS (ns)	-0,006	-0,029	-0,157	-2230000
	Nombre de violation	3	15	33	1959
Buffers (%)		3,0%	8,0%	16,3%	21,0%
Densité (%)		80,4%	83,1%	88,8%	94,8%

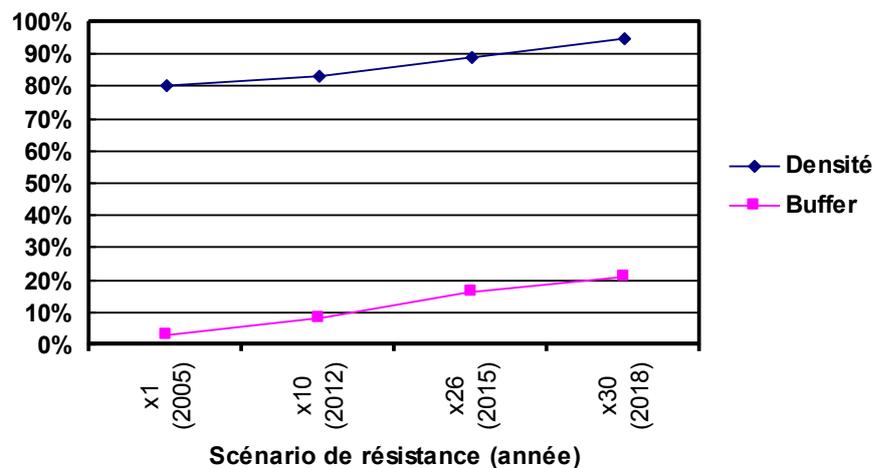


Figure IV-23 : Densité et pourcentage de « Buffers » dans un bloc contrôleur USB en fonction de différents scénarios d'augmentation des résistances d'interconnexion.

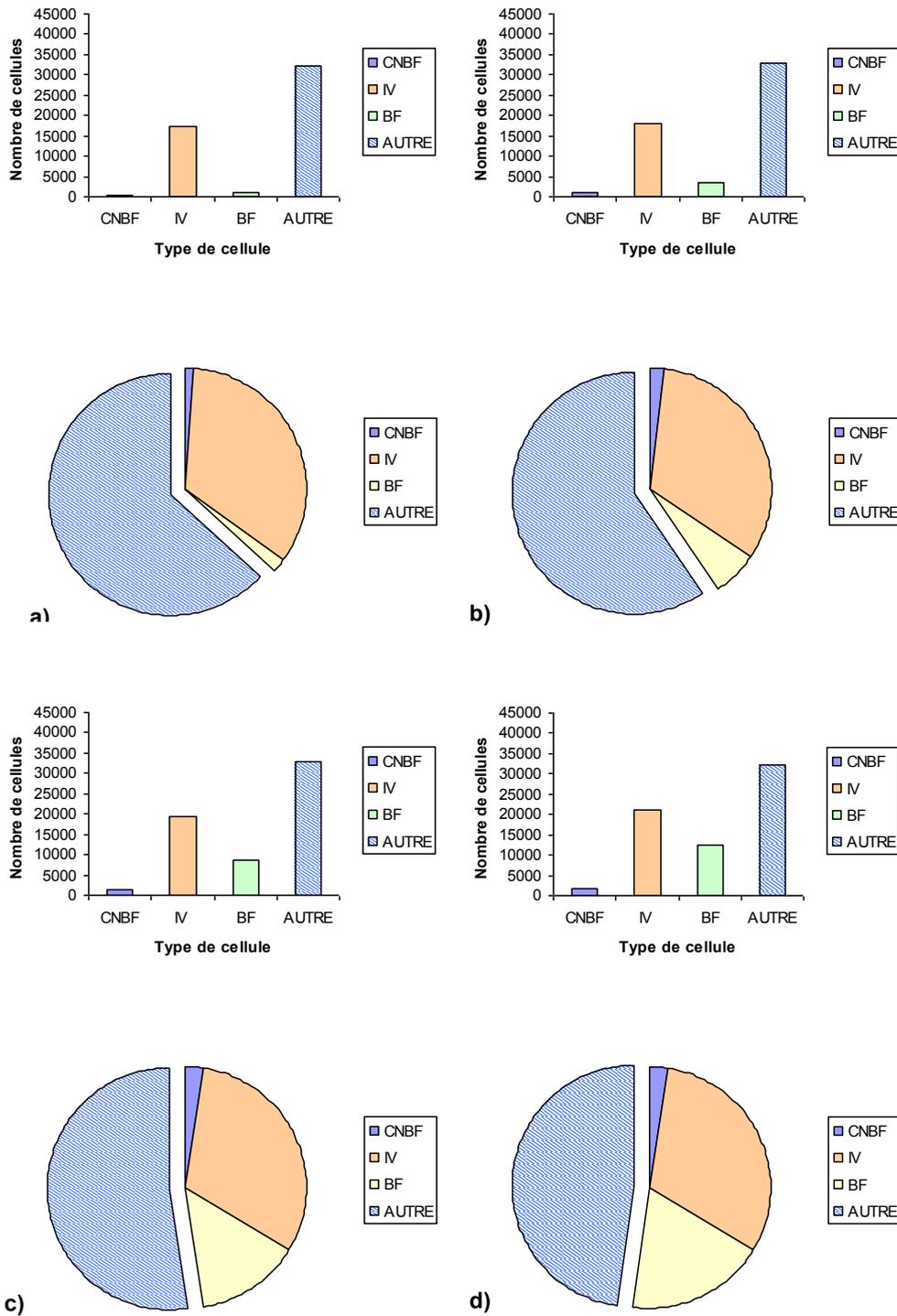


Figure IV-24 : Répartition du type de cellules dans un bloc contrôleur de port USB. a) scénario 65nm standard (2005), b) scénario 2012, c) scénario 2015, d) scénario 2016. CNBF (Clock Network Buffers) correspond aux répéteurs le long de l'arbre d'horloge, IV aux cellules inverseurs et BF aux « buffers ».

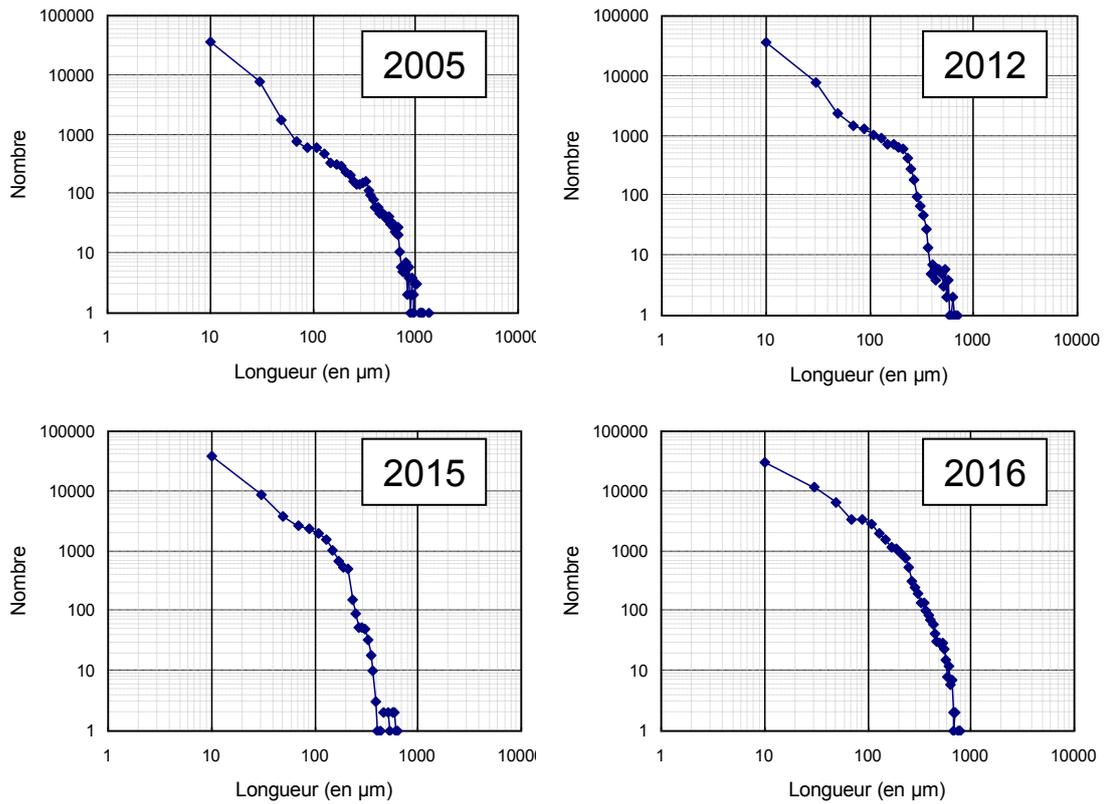


Figure IV-25 : Distribution des longueurs d’interconnexion pour un bloc contrôleur de port USB selon différents scénarios d’augmentation des résistances prévus par l’ITRS.

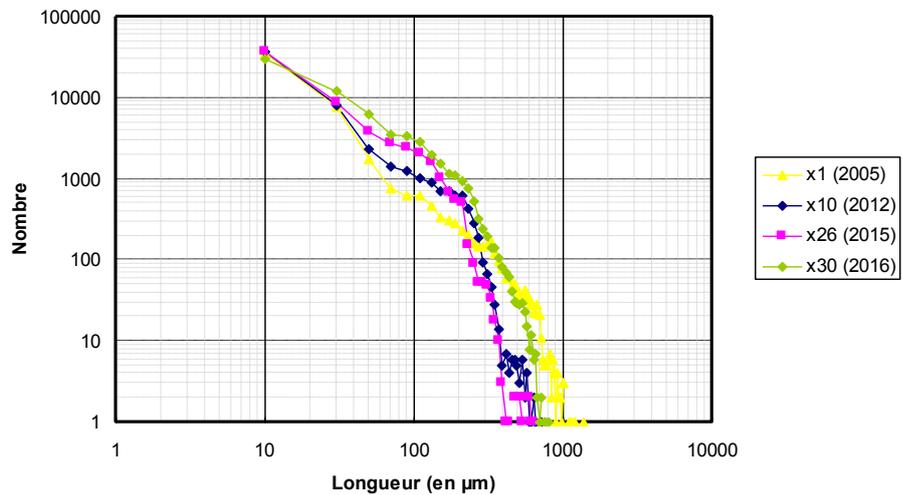


Figure IV-26 : Distribution des longueurs d’interconnexion pour un bloc contrôleur de port USB selon différents scénarios d’augmentation des résistances prévus par l’ITRS.

IV.4.C Analyse du nœud 32nm

a) Validation du flot prédictif en technologie 45nm

Afin de valider la précision du flot prédictif, l'ensemble des données décrites Chapitre III et nécessaires au bon déroulement du flot est prédit ou extrapolé en technologie 45nm d'après le nœud 65nm. Les résultats sont alors comparés par rapport à ceux obtenus avec les outils de CAO standard déjà développé pour la technologie 45nm. Le bloc simulé pour cette partie est un bloc arithmétique. Les détails de ce bloc sont donnés en annexe p. 162.

Différentes implémentations sont effectuées à différentes fréquences. Le comportement densité/fréquence des deux types de flot (standard et prédictif) est alors comparé. Les résultats donnés Figure IV-27 témoignent d'un comportement densité-fréquence très proche entre le flot standard et le flot prédictif. Cela valide complètement l'approche par flot prédictif élaboré durant cette étude.

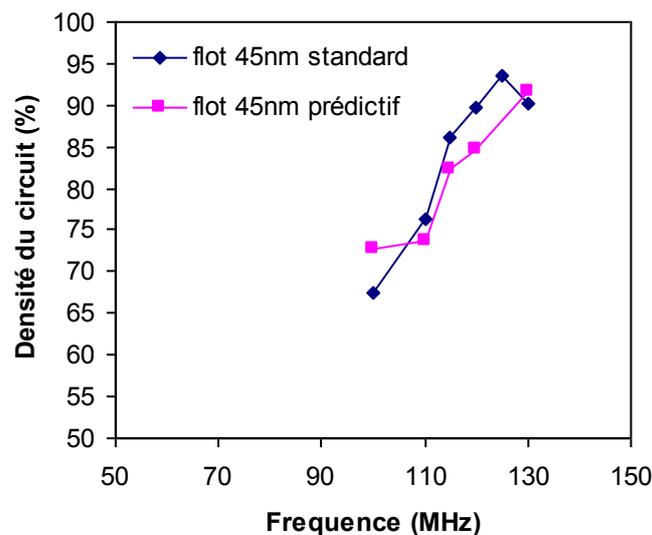


Figure IV-27 : Densité du circuit en fonction de la fréquence pour le flot 45nm prédictif et pour le flot 45nm standard.

b) Extrapolation en 32nm

De la même manière que pour la validation du flot prédictif en 45nm, l'ensemble des données nécessaires au bon déroulement du flot est prédit ou extrapolé en technologie 32nm d'après le nœud 45nm. Le Tableau 7 récapitule les données relatives à l'implémentation pour ce nœud et à celles de références utilisées pour les technologies

65nm et 45nm. Le scénario d'évolution des fréquences et des surfaces de circuit est proche de celui de Moore (surface divisée par deux à chaque génération et fréquence suivant une tendance d'accroissement de 20% à chaque nouvelle technologie). Les distributions de longueur d'interconnexion pour les 3 technologies (65nm et 45nm référence, et prédiction 32nm) sont données Figure IV-28. Le nœud 32nm présente une densité en hausse de 4% par rapport à l'implémentation 45nm de référence. Le nombre de « buffers » est en légère diminution. Les résultats concernant les distributions de familles de cellules sont fournis Figure IV-29. Aucune évolution notable n'est observable entre le 45nm référence et la prédiction 32nm.

Ces résultats montrent qu'aucun des phénomènes parasites mis en évidence pour l'évaluation de la feuille de route ITRS ne joue de rôle majeur dans le cadre du 32nm. Plusieurs raisons permettent d'expliquer cela. Tout d'abord, la technologie 32nm évaluée est une technologie « Low Power » dont les transistors présentent des délais de commutation plus faibles que dans le cas « High Performance » de l'évaluation ITRS. Ainsi la longueur à partir de laquelle les résistances d'interconnexion dépassent les résistances de dispositifs correspond à une longueur plus longue dans le cas « Low Power » que dans le cas « High Performance ». De plus, le bloc simulé présente des dimensions inférieures à la longueur critique (bloc de 155 μ m de côté dans le cas de la technologie 32nm pour une longueur critique de 201 μ m). Il paraît donc logique de ne pas enregistrer de phénomènes parasites importants à ces dimensions.

Ces simulations montrent donc que pour des blocs de taille raisonnable, le nœud 32nm ne sera pas influencé par les grandes tendances de fond mises en évidence au cours de l'évaluation de la feuille de route ITRS. Pour une étude plus complète de l'impact de l'augmentation des résistances d'interconnexion, il convient d'utiliser des blocs de plus grande taille. Cependant, cette étude n'a pas pu avoir lieu durant cette thèse en raison de contraintes de temps.

Tableau 7 : Récapitulatif des données relatives à l'implémentation du bloc arithmétique pour l'évaluation des performances 32nm.

Technologie (nm)		65	45	32
Fréquence (Mhz)		90	110	130
Dimension: hauteur x largeur (µm)		300x300	215x215	155x155
Type de flot		Standard	Standard	Prédicatif
SETUP	WNS (ns)	4,336	4,508	4,473
	TNS (ns)	0	0	0
	Nombre de violation	0	0	0
HOLD	WNS (ns)	0,056	0,081	0,075
	TNS (ns)	0	0	0
	Nombre de violation	0	0	0
Buffers (%)		3,6%	9,3%	6,4%
Densité (%)		76,2%	76,4%	80,9%

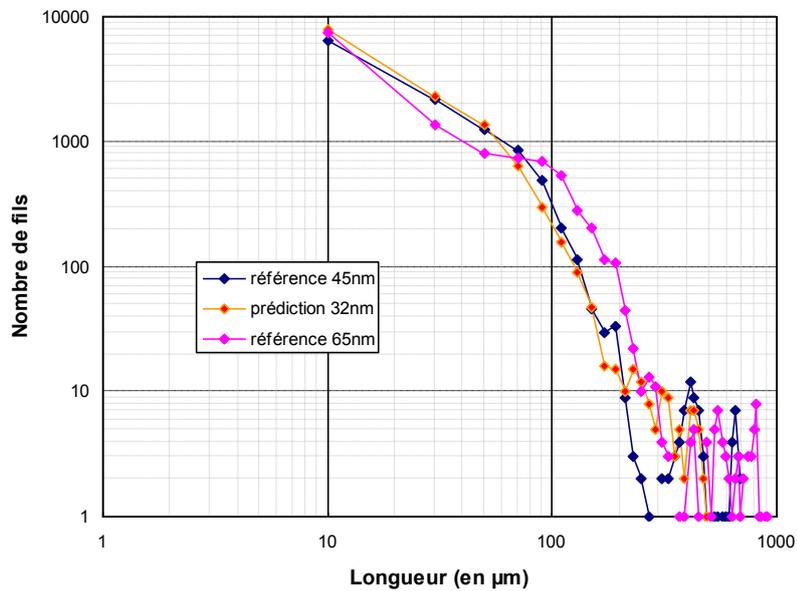


Figure IV-28 : Distribution de longueurs de fil dans le cas de l'implémentation 45nm et 65nm référence et dans le cas de la prédiction pour la technologie 32nm.

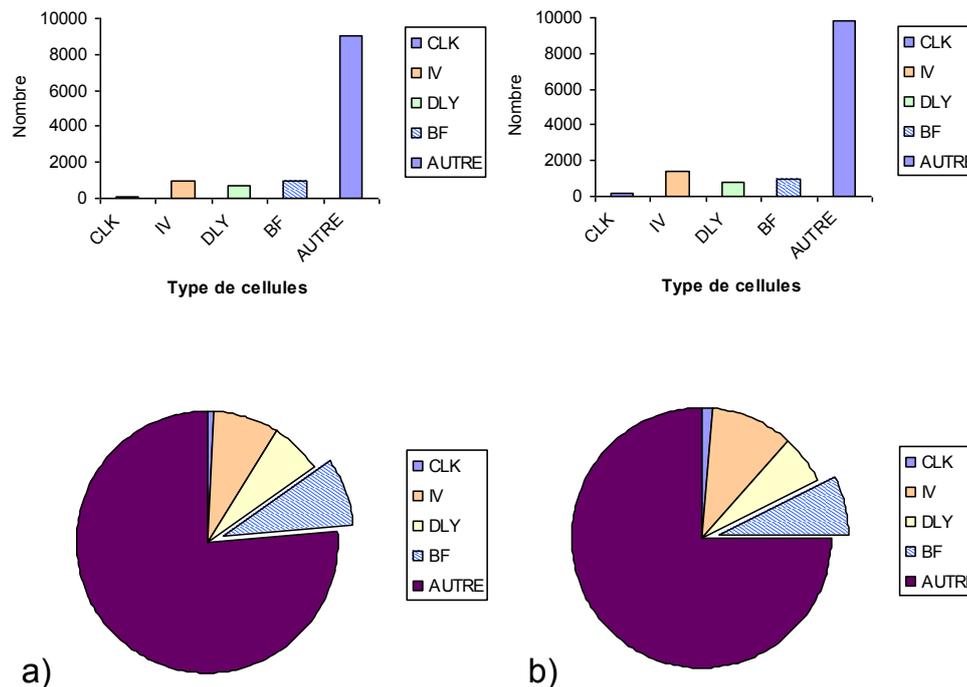


Figure IV-29 : Répartition du type de cellules dans un bloc arithmétique. a) scénario 45nm standard, b) scénario 32nm prédictif. CLK correspond aux cellules de l'arbre d'horloge, IV aux cellules inverseurs, DLY aux cellules de délai, BF aux « buffers ».

IV.5 Conclusion

Deux types d'évaluation des futures technologies à l'aide des kits de conception prédictifs élaborés durant cette étude sont effectués : évaluation de la robustesse à la variabilité à l'aide de simulations prédictives de SRAM et évaluation de l'impact de l'augmentation des résistances d'interconnexion pour les futures technologies (différents critères de délais de ligne et flot prédictif).

L'étude de la variabilité à travers l'évaluation de la feuille de route ITRS montre que cette dernière va rester une problématique récurrente à long terme. Cependant, l'utilisation à court et moyen terme de dispositifs faiblement dopés semble encourageant pour diminuer l'amplitude des variations. Cette option technologique semble d'autant plus prometteuse qu'il est montré que l'utilisation de dispositifs faiblement dopés peut se limiter aux seuls dispositifs NMOS, simplifiant l'intégration de ces dispositifs.

L'application des différents critères d'évaluation technologiques « back end » à la feuille de route ITRS met en lumière plusieurs phénomènes : existence d'une longueur d'interconnexion critique, diminution drastique de cette longueur critique au cours du temps rendant le recours à des répéteurs indispensables, et augmentation des délais de ligne malgré l'ajout de répéteurs. En outre, il est montré que le recours à des fortes conductances de transistor est caduc dans le cas de délais limités par les résistances d'interconnexion. De manière plus générale, cela traduit un changement majeur dans la conception des circuits où les délais ne sont plus dominés par des capacités mais des résistances. Alors que dans le cas des capacités les délais sont atteints grâce à un dimensionnement approprié des conductances des cellules, les délais des circuits présentant de fortes résistances d'interconnexion sont avant tout régis par la faculté à insérer une quantité importante de répéteurs à des endroits appropriés du circuit. Les outils de CAO doivent donc impérativement anticiper cette tendance.

A moins long terme (technologie 32nm), les phénomènes liés à l'augmentation de la résistivité des interconnexions ne semblent pas avoir atteint un niveau suffisamment conséquent pour avoir un impact très important au niveau des circuits. Il faut toutefois noter une augmentation des délais minimum de lignes (délais le long de lignes avec répéteurs) de 11% pour le nœud 32nm. Les blocs de faible dimension (inférieurs à 0,04 mm²) ne semblent pas encore touchés.

Conclusion

L'objectif initial de cette étude était de fournir des éléments d'évaluation des futures technologies au niveau des circuits. Dans ce but il convenait d'arriver à synthétiser des données prédictives ayant trait à la fois au monde de la technologie et au monde de la conception de circuits. Dans ce cadre, il était également nécessaire de fournir des éléments de comparaison de niveau circuit basés sur des kits de conception prédictifs de manière à mettre en évidence des problèmes méconnus jusqu'à présent ou bien d'évaluer l'opportunité d'utiliser certaines innovations technologiques.

Ces objectifs ont été tenus. Des critères d'évaluation ayant trait à différents aspects (variabilité, délais d'interconnexion, délai de blocs placé routé) ont été définis. Ceux-ci ont pu être évalués dans le cadre des futures technologies grâce à l'élaboration de kits de conception prédictifs. Ces kits reposent sur la modélisation prédictive des futurs dispositifs et des interconnexions, et sur le paramétrage correct des outils nécessaires au déroulement du flot digital pour les futures technologies (bibliothèques de cellules standard, paramétrage de l'outil d'implémentation physique).

Les principales informations fournies par l'évaluation de ces critères sont, pour ce qui concerne l'évolution des technologies à long terme, une augmentation drastique des délais d'interconnexion associée à une impossibilité de diminuer ces délais malgré l'ajout de répéteurs, laissant augurer une nouvelle façon de régir les délais dans les circuits par les outils de conception. Ceux-ci vont dépendre de manière croissante des possibilités d'insérer une quantité importante de répéteurs à des endroits appropriés du circuit. Concernant la variabilité des dispositifs, une stagnation à des niveaux non acceptables est mise en évidence.

Pour le court et moyen terme (32 et 22nm), des solutions consistant à utiliser des dispositifs faiblement dopés sont identifiés afin de diminuer le niveau de variabilité aléatoire affectant les points mémoires de type SRAM. L'intérêt d'une solution innovante consistant à n'utiliser que des dispositifs NMOS faiblement dopés a également été montré. A court terme (32nm), les problèmes posés par les délais d'interconnexion ne semblent pas encore jouer un rôle important pour les blocs de faibles dimensions. Il faut toutefois noter qu'une étude plus exhaustive portant sur l'évaluation de blocs de plus grande taille, plus proches des dimensions des produits finaux, et présentant un nombre important de lignes de longueur supérieure à la longueur critique, est nécessaire.

Dans le futur, il serait particulièrement intéressant d'élargir le champ des applications de l'outil d'évaluation technologique élaboré durant cette étude. Ainsi, un plus grand nombre de scénarios d'évolution des technologies pourrait être investigué. Par exemple, la sensibilité de l'évolution de la longueur critique d'interconnexion à différents scénarios d'empilement de diélectriques permettrait de mieux quantifier le gain à développer ces technologies. Egalement, il semblerait judicieux de poursuivre l'application du flot de conception prédictif à des technologies encore plus avancées et à partir de circuits mettant en exergue les problèmes de longueur critique de manière à bâtir une vraie feuille de route de niveau circuit. Cette étude pourrait être réalisée à partir de l'implémentation physique d'un bloc de grande taille couramment utilisé de type ARM par exemple. Il deviendrait alors possible de quantifier sur un cas concret la disjonction entre la performance du transistor et la performance du circuit de manière prospective.

Ce travail ouvre également un important champ d'investigation au niveau de l'évaluation prospective de la performance des technologies. En effet, l'outil développé se trouve être un excellent compromis entre précision et simplicité de mise en œuvre. En effet, avant cette étude, le concepteur de circuit désireux d'obtenir une évaluation rapide d'une future technologie n'avait en sa possession que deux choix : utilisation de modèles standards extrêmement complexes ou bien emploi d'équations analytique souvent trop simplistes. Afin de conforter la position de « troisième voie » de cet outil pour l'évaluation prédictive des technologies, un approfondissement du travail de quantification de l'erreur commise par rapport aux modèles de références devra être entrepris. En particulier, l'utilisation à cette fin des critères couramment utilisés pour les modèles standards (étude de la correspondance des dérivées par exemple) s'avèrera particulièrement judicieuse et permettra de mieux situer l'outil au niveau du compromis simplicité de mise en œuvre - précision. Ainsi pourrait être validé le fait d'utiliser cet outil de manière courante pour l'évaluation des technologies à moyen terme, par exemple lors de l'élaboration des premiers modèles d'une technologie (modèle dits « tentative »).

Mieux évaluer l'impact au niveau des circuits des évolutions de fond mises en évidence lors de cette étude est d'un intérêt majeur. Par exemple, à cause de la faible diminution - voire l'augmentation - des délais de ligne mise en évidence au cours de cette étude, un paramètre des circuits, le délai de propagation de l'arbre d'horloge (ou « insertion delay »), tend à stagner d'une génération à l'autre. Etant donné que les analyses de délais ne sont effectuées que pour un jeu de configurations de transistors réduit (transistors NMOS et PMOS rapides et transistors NMOS et PMOS lents par exemple) par rapport à un ensemble plus vastes de configurations possibles (incluant par exemple transistors NMOS rapides et transistors PMOS lents), il est nécessaire de prendre certaines marges lors des analyses de délai afin de couvrir l'ensemble

des configurations possibles. Certaines de ces marges sont directement proportionnelles au délai d'horloge, elles sont donc constantes en valeur absolue d'une technologie à l'autre. Or, pour continuer à augmenter les fréquences des circuits, il faudrait que ces marges diminuent ce qui n'est pas donc le cas. La non-diminution des délais de ligne peut ainsi avoir un impact très grand au niveau des circuits. Inclure dans le futur plus de critères de niveau circuits critiques directement en lien avec les précédents critères définis durant cette étude pourrait ainsi être d'un intérêt majeur pour le concepteur de circuit.

En conclusion, cette étude a ouvert la voie à un nouveau type de recherche transverse à de nombreux domaines. Elle a permis de chiffrer des phénomènes jusqu'à présent seulement pressentis. Elle a également montré tout l'intérêt que plusieurs disciplines ont à travailler de concert. Elle doit donc absolument être poursuivie afin d'étoffer l'étendue de ses applications (application analogiques radio fréquence, imageurs,...).

Références

- [1] MOORE, G.E., Cramming More Components Onto Integrated Circuits, *Proceedings of the IEEE*, Jan. 1998, Volume 86, Issue 1, p. 82 – 85
- [2] ITRS Starting Materials Sub-TWG, Position Paper for 450nm Development, June 2005. Disponible sur <http://www.itrs.net/papers.html>
- [3] LAPEDUS, M., TI to make fabless switch at 45 nm, *EE Times*, 29 Janvier 2007. Disponible sur www.eetimes.com .
- [4] Intel, Moore's Law. Disponible sur <http://www.intel.com/technology/mooreslaw/index.htm>
- [5] SKOTNICKI, T. et BOEUF, F., Chapitre 3 : Introduction à la physique du transistor MOS, *EGEM*, édition Hermès, 2003
- [6] SKOTNICKI, T., Cahier E 2 430 : Transistor MOS et sa Technologie de Fabrication, *Encyclopédie Techniques de l'Ingénieur*, Février 2000
- [7] MONFRAY, S., Conception et caractérisation de dispositifs CMOS en technologie SON, Th : Phys. : Université de Provence, 2003
- [8] JOSSE, E., Nouvelles architectures de grille pour les générations CMOS 0.1 μ m et en deçà, Th : Phys. : INPG, 2000
- [9] PAYET, F., Modélisation de transistors à canal de silicium contraint pour les nœuds technologiques CMOS 45nm et en deçà, Th : Phys. : Université de Provence, 2005.
- [10] Manuel du logiciel MASTAR. Disponible sur <http://www.itrs.net/models.html> .
- [11] BOWMAN, K.A., DUVALL, S.G., MEINDL, J.D., Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration, *IEEE Journal of Solid-State Circuits*, Volume 37, p.183, 2002
- [12] BORIVOJE, N., Measurement and Analysis of Process Variability in 90nm CMOS, *8th International Conference on Solid-State and Integrated Circuit Technology (ICSICT '06)*, Oct. 2006, p. 505 – 508.

- [13] CHARLET, F. *et al.*, Extraction of (R,L,C,G) interconnect parameters in 2D transmission lines using fast and efficient numerical tools, *Proc. of SISPAD*, 2000.
- [14] Edition ITRS Roadmap Interconnect 2005, Disponible sur <http://www.itrs.net/Links/2005ITRS/Interconnect2005.pdf>.
- [15] BESLING, W.F.A., ARNAL, V., GUILLAUMOND, J.F., GUEDJ, C., BROEKAART, M., CHAPELON, L.L., FARCY A., ARNAUD, L. et TORRES, J., *Proc. of the IEDM conference*, 2004, p. 325.
- [16] TAVEL, B., Intégration de diélectriques haute permittivité et de grilles métalliques dans un procédé CMOS sub-0,1 μ m, Th : Phys. : INPG, 2003.
- [17] MISTRY, K. *et al.*, A 45nm Logic Technology with High-K + Metal Gate Transistors, Strained Silicion, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging, *International Electron Device Meeting (IEDM)*, 2007, p.247-250
- [18] CHUDZIK, M. *et al.*, High Performance High-K/Metal Gates for 45nm CMOS and beyond With Gate-First Processing, *VLSI Technology Symposium*, 2007, p.194
- [19] POUYDEBASQUE, A. *et al.*, High Density and High Speed SRAM bit-cells and Ring Oscillators due to Laser Annealing for 45nm Bulk CMOS, *International Electron Device Meeting (IEDM)*, 2005, p.663-666
- [20] JURCZAK, M. *et al.*, Silicon-on-Nothing (SON)-an innovative process for advanced CMOS, *IEEE Trans. Electron Devices*, Nov. 2000, vol. 47, p. 2179-2187
- [21] CHANEMOUGAME, D., Conception et fabrication de nouvelles architectures CMOS et étude du transport dans les canaux de conduction ultra minces obtenus avec la technologie SON, Th : Phys. : Institut National des Sciences Appliquées de Lyon, 2005
- [22] HARRISON S., CORONEL P., LEVERD F., CERUTTI R., et al., Highly performant double gate MOSFET realized with SON process, *International Electron Device Meeting (IEDM)*, 2003, p. 449-452
- [23] TANG S.H., CHANG L., LINDERT N., YAN-KYU CHOI, WEN-CHIN LEE, et al.– a quasi-planar double-gate MOSFET, *International Solid-State Circuit Conference (ISSCC)*, 2001, p. 118-119

- [24] FARHANG ALI R., DEETER T., RIKHI S., Co-optimization of Product Design and Semiconductor Process Technology : The Core of a Winning DFM Strategy, FUTURE FAB International, Jan. 2007, vol. 22, p. 40-44
- [25] Solution Synopsis d'extraction des résistances et capacités parasites Star-RCXT. Documentation disponible sur http://www.synopsys.com/products/starxct/star_rcxt_ds.pdf
- [26] SEEVINCK, E., LIST, F.J., LOHSTROH, J., Static-noise margin analysis of MOS SRAM cells, *IEEE Journal of Solid-State Circuits*, 1987, Volume 22, p.748-754
- [27] GROSSAR, E., STUCCHI, M., MAEX, K., DEHAENE, W., Read Stability and Write-Ability Analysis of SRAM Cells for Nanometer Technologies, *IEEE Journal of Solid-State Circuits*, 2006, Volume 41, p.2577-2588
- [28] CHRISTIE, P., STROOBANDT, D., The interpretation and application of Rent's rule, *IEEE Trans. VLSI*, Dec. 2000, Volume 8, Issue 6, p.639-648.
- [30] SKOTNICKI, T., MERCKEL, G., PEDRON, T., The voltage-doping transformation: a new approach to the modeling of MOSFET short-channel effects, *IEEE Electron Device Lett.*, Mar. 1988, vol. 9, pp. 109-112
- [31] Rapport ITRS 2003 section PIDS. Disponible sur <http://www.itrs.net/Links/2003ITRS/Home2003.htm>
- [32] CHENMING, H., BSIM model for circuit design using advanced technologies, *VLSI Circuits Symposium*, 2001, p.5-10
- [33] GIDENBLAT, G., LI, X., WU, W., WANG, H., JHA, A., LANGEVELDE, R., et al., PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation, *IEEE Trans. Electron Devices*, Sept. 2006, vol. 53, p. 1979-1993
- [34] SUZUKI, K., Parasitic capacitance of submicrometer MOSFETs, *IEEE Trans. Electron Devices*, Sep. 1999; volume 46, p.1895-1900
- [35] MIZUNO, T., OKAMURA, J., TORIUMI, A., Experimental Study Of Threshold Voltage Fluctuations Using An 8k MOSFET's Array, *VLSI Technology Symposium*, 1993, p.41-42

- [36] KEYES, R. W., Effect of randomness in the distribution of impurity ions on FET thresholds in integrated electronics, *IEEE Journal of Solid-State Circuits*, 1975, Volume 10, p.245-247
- [37] CHARLET, F., BERMOND, C., PUTOT, S., LE CARVAL, G., FLECHET, B., Extraction of (R,L,C,G) interconnect parameters in 2D transmission lines using fast and efficient numerical tools, *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2000, p.87-89
- [38] CATHIGNOL, A., Caractérisation et modélisation des fluctuations locales des paramètres électriques du transistor des filières CMOS sub-65 nm, Th : Phys. : INPG, 2008
- [39] PELGROM, M. J. M., DUINMAIJER, A.C.J., WELBERS, A. P. G., Matching Properties of MOS Transistors, *IEEE Journal of Solid-State Circuits*, 1987, Volume 22, p.748-754
- [40] MAUFRONT, C., Crolles 2 Alliance, rapport interne.

Bibliographie de l'auteur

SELLIER, M., BOROT, B., PORTAL, J.M., BOEUF F., A Universal Predictive Model for Circuit Performance Assessment, *SiO2 Advanced Dielectrics and Related Devices*, 2006, p. 182-183

SELLIER, M., PORTAL J. M., BOROT B., COLQUHOUN S., FERRANT R.; BŒUF F., FARCY A., Predictive Delay Evaluation on Emerging CMOS Technologies: A Simulation Framework, *International Symposium on Quality Electronic Design (ISQED)*, 2008, p. 492-497

SELLIER, M., BOROT, B., PORTAL, J. M., Anticipation de l'introduction des technologies CMOS sub-50nm dans les flots classiques de conception de circuit, *Journée Nationale du Réseau Doctorale en Microélectronique, nanotechnologie et microsystèmes (JNRDM)*, 2006

BOEUF, F, SELLIER, M., DURIEZ, B., JOSSE, E., POUYDEBASQUE, A., MULLER, M., PAYET, F., BOROT, B., SKOTNICKI, T., Using MASTAR as a Pre-SPICE Model Generator for Early Technology Assessment and Circuit Simulation, *Solid-State Device Meeting (SSDM)*, 2006

BOEUF, F, SELLIER, M., PAYET, F., BOROT, B., SKOTNICKI, T., Using MASTAR as a Pre-SPICE Model Generator for Early Technology Assessment and Circuit Simulation, *Japanese Journal of Applied Physics*, in press

BOEUF, F., SELLIER, M., FARCY, A., SKOTNICKI, T., Impact of Layout, Interconnects and Variability on CMOS Technology Roadmap, *VLSI Technology Symposium*, 2007, p.24-25

BOEUF, F., SELLIER, M., FARCY, A., SKOTNICKI, T., An Evaluation of the CMOS Technology Roadmap from the Point of View of Variability, Interconnects, and Power Dissipation, *Trans. On Elec. Devices*, in press

SKOTNICKI T., FENOUILLET-BERANGER C., GALLON C., BOEUF F., MONFRAY S., PAYET F., POUYDEBASQUE A., FARCY A., ARNAUD F., CLERC S., SELLIER M., CATHIGNOL A., SCHOELLKOPF J. P., PEREA E., FERRANT R., MINGAM H.,

Innovative Materials, Devices and Circuits for Low Power Mobile Multimedia, *IEEE Trans. Electron Devices*, Jan. 2008, vol. 55, p. 96-130

FARCY, A., GALLITRE M., ARNAL, V., SELLIER M., GUIBE L., BLAMPEY B., BERMOND C., FLECHET B., TORRES J., Impact of process parameters on circuit performance for the 32 nm technology node, *Microelectronic Engineering*, Nov. 2007, Volume 84, Issue 11, p. 2738-2743

BIDAL, G., BOEUF, F., DENORME, S., LOUBET, N., LAVIRON, C., LEVERD, F., BARNOLA, S., SALVETAT, T., COSNIER, V., MARTIN, F., GROSJEAN, M., PERREAU, P., CHANEMOUGAME, D., HAENDLER, S., MARIN, M., RAFIK, M., FLEURY, D., LEYRIS, C., CLEMENT, L., SELLIER, M., MONFRAY, S., BOUGUEON, J., SAMSON, M.P., CHAPON, J.D., GOURAUD, P., GHIBAUDO, G., SKOTNICKI, T., Planar Bulk+ Technology using TiN/Hf-based gate stack for Low Power Applications, *VLSI Technology Symposium*, in press

Brevet

FERRANT R., SELIER, M., SKOTNICKI, T., Variability Immune SRAM bitcell, en cours de dépôt

Annexe

A.1 Exemple de fichier librairie paramétrant MSIM

```
.lib svtlp_TT
.subckt nsvtlp D G S B

*-----
*
*   PARAMETERS/CONSTANTS
*-----
*-----MISMATCH PARAMETERS-----

.param dev_ndep = 0 devx/gauss=sigmandep
.param sigmandep = 'valif(fluctdop==1,2/sqrt(3)*sqrt(ndeop0_s/(w1*leff1*xdep_v2_ch)),0) '

*-----PHYSICAL CONSTANTS-----
.param q = 1.6e-19
.param epsilon0 = 8.85e-12
.param epsilonSi = 11.7055
.param epsilonSiO2= epsrox1
.param k = '1.3806226e-23/q'
.param ni = '1.45e10*(T/300.15)*sqrt(T/300.15)*exp(21.5565981-eg/(2*k*(TC+273.15)))*1e6'
.param eg = '1.16-7.02e-4*T*(T+1108) '

*-----MOS PARAMETERS-----
*****
**N MOS**
*****

**MODEL SELECTOR**

    *mastar*
.param newSlope=1
.param S_manu='8.971e+001 *(1-newSlope) '
.param MastarCompute=1
.param DG=0
.param SOI='0 '
.param vdd = 1.100e+000

    *bsim*
.param capmod=3
.param correction_ION=1

**PROCESS PARAMETERS**

.param tox_mastar = 'Toxe_phys+darkspace+polydep+dev_toxe'
.param darkspace = '3.800e+000 *1e-10'
.param polydep = '3.600e+000 *1e-10'
.param Toxe_phys='1.700e+000 *1e-9'
.param ndep = '9.000e+017 '
.param r = 176.287583844015
.param XJ = '1.500e+001 *1e-9'
.param Tsi = '1.500e+001 *1e-9+dev_Tsi'
.param Tbox='1.450e+002 *1e-9'
.param Ngp=1.000e+020
.param ngate1 = 0
*valif(polydep!=0,1.300e+020 ,0)
.param phim1 = 4.050e+000
.param nsd1 = 9.000e+019
.param xl=0
.param tpoly = '90 *1e-9'
.param Qss='0.000e+000 *q*1e4*phi_c'

    *poches*

.param Poches = 1
.param Nsub = 2.500e+017
.param Energie_KeV = 4.900e+001
.param angle = 2.500e+001
.param diffusion = 1.000e+000
.param activation = 1.000e+002
.param Dose_cm2 = 5.380e+013
.param species = '1 +1'
*B = 1
*BF2 = 2
*IN = 3

    *High-K*

.param active_HK = 0
.param HiK_tox_A ='5.450e+000 *10'
.param epsilon_HK =2.500e+001
```

```

.param t_SiO2_pied_A =8.500e-001

**FITTING PARAMETERS**

    *mstar*
.param kmu=1.8
.param kw=1.000e-007
.param delta = 1.000e-001
.param gamma=4.000e-001
.param zeta1=8.000e-001
.param zeta2=6.400e-001
.param zeta3=2.000e+000
.param Kd = 1.000e+000
.param kvs = 1.000e+000
.param TIF=8.000e+001
.param kfield = 1.000e+000
.param vsat1 = ' 1e5*kvs'
.param Ith=1.000e-007
.param epsrox1 = 3.9018

    *bsim*
.param polydep_bsim=polydep
.param dtocv=0
*2.6e-10
.param noff=1
.param voffcv=0
.param pvag=0
.param pclm=1e-10
.param pdiblc1=1e-10
.param pdiblc2=1e-10
.param DELTAvd=0.1
.param drout=0.56
.param T=300
.param TC='300-273.15'

**MISMATCH PARAMETERS**

.param dev_l=0.0 devx/gauss=0e-9
.param dev_w=0.0 devx/gauss=0e-9
.param dev_r=0.0
.param fluctdop=1
.param dev_toxe=0.0 devx/gauss=0e-10
.param dev_Tsi=0.0 devx/gauss=0e-10

.param nrd=0
.param nrs=0
.param rd1='r/w/2'
.param rs1='rd1'
.param ll='l*tometer + dev_l'
.param wl='valif(DG1==0,w*Tometer,w*2*tometer) + dev_w'
.param as1='AS*tometer*tometer'
.param ad1='AD*tometer*tometer'
.param ps1='PS*tometer'
.param pd1='PD*tometer'
.param tometer=1e-6
.param MastarCompute1='valif(SOI1==1,1,MastarCompute)'
.param zeta21=zeta2
*'valif(zeta2==0,0.64*(1-SOI1)*(1-DG1)+SOI1*1+0.2*DG1,zeta2)'
.param zeta11=zeta1
*'valif(zeta1==0,0.8*(1-SOI1)*(1-DG1)+SOI1*0.77+0.35*DG1,zeta1)'
.param Kd1 = ' valif(Kd==0,1-0.5*SOI1-DG1,Kd)'
.param XJ1='Tsi/2*DG1 + Tsi*SOI1 + (1-DG1)*(1-SOI1)*XJ'
.param SOI1 = 'valif((DG==1),0,valif((SOI==2),1,0))'
.param DG1 = 'valif(((DG==1),1,0))'
.param correction_ION1 = 'valif(((correction_ION)==(0)),0,1)'
.param ro = '2.329e3'

*-----SCALED PARAMETERS-----
.param Ndep1 = 'valif(Poches==1,(nsub+Nmoy),ndep)'
.param Ndep0_s = 'Ndep1*1e6'
.param Ndep_s = 'abs(Ndep0_s + dev_ndep)'
.param Nsub_s = 'valif(Poches==1,nsub*1e6,Ndep_s)'
.param ngate1_s = 'ngate1 * 1e6'
.param nsdl_s = 'nsdl*1e6'
.param mueff1_s = 'mueff1/1e4'
.param Ngp_s = 'Ngp*1e6'

*---MASTAR CALCULATED PARAMETERS---

.param coxe_mastar = 'epsilon0*epsilonSiO2/((toxe_mastar))'
.param Kb = '-sqrt(2*q*Ndep_s*epsilon0*epsilonSi) /coxe_mastar'
.param d = 'Kd1*Kb/(2*sqrt(2*phif1))'
.param vfb1 = 'phim1-phisl - Qss/coxe_mastar'
.param phif1 = 'k*T*log(Ndep_s/ni)'
.param phisl = '4.05+0.56+phif1'
.param phid = 'k * T * log(nsdl_s*Ndep_s/(ni*ni))'
.param phid0 = 'k * T * log(nsdl_s*Ndep0_s/(ni*ni))'
.param lint = 'gamma/2*XJ1 + DG1*gamma/2*XJ1'
.param leff1 = 'll+xl-2*lint'
.param Tsub = '-(Tsi+3*Tbox)+sqrt(pow(((Tsi+3*Tbox),2)+2*epsilon0*epsilonSi*phid/(q*Ngp_s)-Nsub_s/Ngp_s*Tsi*Tsi))'
.param psibox='q*Ngp_s*Tsub/(coxe_mastar)'
.param psisup = 'phid-2*phif1-k*T*log(k*T/q/Nsub_s*coxe_mastar*exp((1/(k*T)+1)/2))'
.param vth0soi='vfb1 + phid + psibox + q*Nsub_s*Tsi/coxe_mastar'
.param vth0dg = '(vfb1 + phid - psisup + q*Nsub_s*xdep_v1/coxe_mastar)'
.param SCERSCE='RSCE-SCE_MASTAR-NCE'

*****POCHES*****
.param x = Energie_KeV
.param angle_rad = 'angle *pi/180.0'
.param pi= 3.1415928

```

```
.param Rp_A='valif(species==1,-2e-07*pow(x,5.0) + 4e-05*pow(x,4.0) - 0.003*pow(x,3.0) - 0.0231*pow(x,2.0) + 39.201*(x) +
15.333,valif(species==2, 2e-07*pow(x,5.0) - 6e-05*pow(x,4.0) + 0.0063*pow(x,3.0) - 0.3102*pow(x,2.0) + 15.862*(x) +
1.8744,valif(species==3,3e-07*pow(x,5.0) - 7e-05*pow(x,4.0) + 0.008*pow(x,3.0) - 0.4061*pow(x,2.0) + 14.488*(x) +
2.584,0))'
.param DeltaRp_A=' valif(species==1,2e-07*pow(x,5.0) - 6e-05*pow(x,4.0) + 0.0078*pow(x,3.0) - 0.561*pow(x,2.0) +
27.469*(x) + 1.4359,valif(species==2, 1e-07*pow(x,5.0) - 3e-05*pow(x,4.0) + 0.0036*pow(x,3.0) - 0.1815*pow(x,2.0) +
7.1725*(x) + 1.1073,valif(species==3, 1e-07*pow(x,5.0) - 3e-05*pow(x,4.0) + 0.0032*pow(x,3.0) - 0.1612*pow(x,2.0) +
4.9472*(x) + 0.9665, 0))'
.param DeltaRl_A=' valif(species==1,1e-07*pow(x,5.0) - 3e-05*pow(x,4.0) + 0.004*pow(x,3.0) - 0.3048*pow(x,2.0) +
19.505*(x) + 1.5642,valif(species==2, 8e-08*pow(x,5.0) - 2e-05*pow(x,4.0) + 0.0026*pow(x,3.0) - 0.1353*pow(x,2.0) +
5.355*(x) + 0.8419, valif(species==3, 7e-08*pow(x,5.0) - 2e-05*pow(x,4.0) + 0.0023*pow(x,3.0) - 0.1138*pow(x,2.0) +
3.673*(x) + 0.7998 , 0))'

.param Rp_um = ' Rp_A*cos(angle_rad)*1e-4'
.param DeltaRp_um = ' diffusion * DeltaRp_A*1e-4'
.param DeltaRl_um = 'diffusion * DeltaRl_A*1e-4'
.param Cpoches_cm2 = '(activation/100.0)*(Dose_cm2/4.0)'

.param Npoches = '0.5*Cpoches_cm2/(Rp_um*0.0001+2.0*DeltaRp_um*0.0001)'
.param Lpoches_um = '(Rp_um+2.0*DeltaRp_um)*sin(angle*pi/180.0)+2.0*DeltaRl_um*cos(angle*pi/180.0)-
gamma*XJ1*1e9/2000.0'
.param Lmin_um = 'Min((l1+x1)*1e6,Lpoches_um)'
.param Nmoy = '2.0*Npoches*Lmin_um/((l1+x1)*1e6)'

.param XJ2 = '2.0/3.0*1e9 * XJ1'
.param Lelec = 'leff1/1e-9'
.param alphaSlope = '0.5*(1.0+(14.0/Lelec))'
.param phi_d = phid
.param phi_c = 'k*T*log(Ndep_s/ni)+k*T*log(Ndep_s/ni)'
.param root1 = 'valif((phi_d- alphaSlope *phi_c)>0,(phi_d- alphaSlope *phi_c),0)'
.param root2 = 'valif((phi_d- alphaSlope *phi_c/2)>0,(phi_d- alphaSlope *phi_c/2),0)'
.param Nul = 'vdd+2*(phi_d - alphaSlope *phi_c)+2*sqrt(root1*(vdd+phi_d- alphaSlope *phi_c))'
.param Nu2 = 'vdd+2*(phi_d - alphaSlope *phi_c/2)+2*sqrt(root2*(vdd+phi_d- alphaSlope
*phi_c/2))'
.param Ndep_s_star1= 'Ndep_s - 2.0* epsilon0*epsilonSi * Nul / q /pow((Lelec)* 1e-9, 2.0)'
.param Ndep_s_star2= 'Ndep_s - 2.0* epsilon0*epsilonSi * Nu2 / q /pow((Lelec)* 1e-9, 2.0)'
.param Ttest1 =
'abs(Ndep_s_star1)/Ndep_s_star1*sqrt(abs(2*phi_c*epsilon0*epsilonSi/q/Ndep_s_star1))'
.param Ttest2 =
'abs(Ndep_s_star2)/Ndep_s_star2*sqrt(abs(phi_c*epsilon0*epsilonSi/q/Ndep_s_star2))'
.param Tdepl = 'valif((Ttest1*1e9 < XJ2) && (Ttest1 *1e9 > 0.0),Ttest1,sqrt((q*(Ndep_s-
Ndep_s_star1)*pow(XJ2*1e-9,2.0) + 2.0*phi_c * epsilon0*epsilonSi)/q/Ndep_s))'
.param Qdepl = 'valif((Ttest1*1e9 < XJ2) && (Ttest1 *1e9 > 0.0),q * Ndep_s_star1 * Tdepl,q*Tdepl
* Ndep_s - 2.0 * Nul *epsilon0*epsilonSi *(XJ2*1e-9)/pow(Lelec*1e-9,2.0))'
.param Tdep2 = 'valif((Ttest2*1e9 < XJ2) && (Ttest2 *1e9 > 0.0),Ttest2,sqrt((q*(Ndep_s-
Ndep_s_star2)*pow(XJ2*1e-9,2.0) + phi_c * epsilon0*epsilonSi)/q/Ndep_s))'
.param Qdep2 = 'valif((Ttest2*1e9 < XJ2) && (Ttest2 *1e9 > 0.0),q * Ndep_s_star2 * Tdep2,q*Tdep2
* Ndep_s - 2.0 * Nu2 *epsilon0*epsilonSi *(XJ2*1e-9)/pow(Lelec*1e-9,2.0))'
.param essai= log(10.0)
.param S_newslope = 'k*T*essai*(1.0+abs(Qss)/coxe_mastar+(Qdepl-Qdep2)/phi_c*2.0/coxe_mastar)'
.param S_MASTAR='valif(MastarCompute1==1,S_MASTAR_v1,S_MASTAR_v2)'
.param S_SOI='k*T*log(10)*(1+(epsilonSi/epsilonSiO2)*toxe_mastar/(xdep_v1+ (SOI1)*(epsilonSi/epsilonSiO2)*Tbox + Tsub))'
.param vthsat = 'vth0 - SCE_MASTAR - DIBL_MASTAR +RSCE-NCE+ delta'

.param vth0='(1-SOI1)*(1-DG1)*valif(MastarCompute1==1,vth0_v1,vth0_v2)+SOI1*vth0soi+DG1*vth0dg'
.param SCE_MASTAR = 'valif(MastarCompute1==1,SCE_v1,SCE_v2)'
.param DIBL_MASTAR = 'valif(MastarCompute1==1,DIBL_v1,DIBL_v2)'
.param RSCE = 'valif(Poches==1,sqrt(2*epsilonSi*epsilon0*q*Nsub_s*2*phif1)/coxe_mastar*(sqrt(Ndep_s/Nsub_s)-1),0)'
.param musr = 'valif(MastarCompute1==1,1450 * PWR(Eeff_v1,-2.9),1450 * PWR(Eeff_v2,-2.9))'
.param muac = 'valif(MastarCompute1==1,330 * PWR(Eeff_v1,-0.3),330 * PWR(Eeff_v2,-0.3))'

.param S='valif(S_manu==0,newSlope*(1-SOI1)*S_newslope+SOI1*S_SOI+(1-newSlope)*(1-SOI1)*S_MASTAR,S_manu/1000)'
.param NCE = '(epsilonSi/epsilonSiO2)*toxe_mastar*xdep_v1_ch/(w1*w1)*zeta3'

* phid (v1)
.param epsilon_moy = '(HiK_tox_A*epsilon_HK + t_SiO2_pied_A *epsilonSiO2) / (HiK_tox_A + t_SiO2_pied_A)'

.param xdep_v1 =
'min(sqrt(2*epsilon0*epsilonSi*phid/(q*Nsub_s)),Tsi)+min(sqrt(2*epsilon0*epsilonSi*phid/(q*Nsub_s)),Tsi/2)*DG1+sqrt(2
*epsilon0*epsilonSi*phid/(q*Nsub_s))*(1-SOI1) *(1-DG1) '
.param xdep_v1_ch =
'min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep_s)),Tsi)*SOI1+min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep_s)),Tsi/2)*DG1+sqrt(2
*epsilon0*epsilonSi*phid/(q*Ndep_s))*(1-SOI1) *(1-DG1) '
.param xdep_v2_ch =
'Tsi*SOI1+min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep0_s)),Tsi/2)*DG1+sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep0_s))*(1-
SOI1)*(1-DG1) '
* .param xdep_v2_ch =
='min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep0_s)),Tsi)*SOI1+min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep0_s)),Tsi/2)*DG1+s
qrt(2*epsilon0*epsilonSi*phid/(q*Ndep0_s))*(1-SOI1) *(1-DG1) '
.param vth0_v1 = 'vfb1 + phid + 1/coxe_mastar*sqrt(2*epsilon0*epsilonSi*q*phid*Nsub_s)'
.param SCE_v1 = 'zeta21 * (1+active_HK*(epsilon_moy-15.0)*0.0078)*(epsilonSi/epsilonSiO2) * (1 + XJ1*XJ1/PWR(leff1,2)) *
(toxe_mastar/leff1) * (xdep_v1_ch/leff1) * phid'
.param delta_DIBL_box = 'valif((pow(Tsi+3*Tbox,2) < (leff1*leff1)),0,(toxe_mastar * Tsi/(leff1*leff1)) *
(sqrt(pow(Tsi+3*Tbox,2)-(leff1*leff1)) / leff1) * Vdd)'
.param DIBL_v1 = 'zeta21 * (1+active_HK*((HiK_tox_A+t_SiO2_pied_A)-20)*0.015)*(epsilonSi/epsilonSiO2) * (1 +
XJ1*XJ1/PWR(leff1,2)) * (toxe_mastar/leff1) * ((xdep_v1_ch+SOI1*((leff1-Tsi)*3*Tbox)/(leff1-Tsi+3*Tbox))/leff1) * vdd +
0.16*delta_DIBL_box*SOI1'
.param S_MASTAR_v1 = 'k*T*log(10)*(1+epsilon0*epsilonSi*toxe_mastar/(epsilon0*epsilonSiO2*xdep_v1))'
*.param Eeff_v1 = 'kfield*((vdd + vthsat-2*(vfb1+phid-DG1*psisup+SOI1*psibox))/(6*toxe_mastar))*1e-8'

.param Qdep_Cox1 = 'vthsat-vfb1-phid -SOI1*psibox + DG1*psisup'
.param Qdep_Cox = 'valif(Qdep_Cox1<0,0, Qdep_Cox1)'

.param Eeff_v1 = 'kfield*((vdd-vthsat)/(6*toxe_mastar) + (2.0 * Qdep_Cox / (6*toxe_mastar)))*1e-8'
* + ((Vdd-Vthsat)/(9.0*toxe_mastar))*(choixMOS == PMOS)
* + (3.0 * Qdep_Cox / (9.0*toxe_mastar))*(choixMOS == PMOS);

* 2phif (v2)

.param vth0_v2 = 'vfb1 + 2*phif1 + 1/coxe_mastar*sqrt(2*epsilon0*epsilonSi*q*2*phif1*Nsub_s)'
.param xdep_v2 = 'sqrt(2*epsilon0*epsilonSi*2*phif1/(q*Nsub_s))'
```

```

.param SCE_v2 = 'zeta21 * (epsilonSi/epsilonSiO2) * (1 + XJ1*XJ1/PWR(leff1,2)) * (toxe_mastar/leff1) * (xdep_v2/leff1) *
2*phif1'
.param DIBL_v2 = 'zeta11 * (epsilonSi/epsilonSiO2) * (1 + XJ1*XJ1/PWR(leff1,2)) * (toxe_mastar/leff1) * (xdep_v2/leff1) *
vdd'
.param S_MASTAR_v2 = 'k*T*log(10)*(1+epsilon0*epsilonSi*toxe_mastar/(epsilon0*epsilonSiO2*xdep_v2))'
.param Eeff_v2 = 'kfield*(vdd + vthsat-2*(vfb1+2*phif1))/(6*Toxe_mastar)*1e-8'

*-----BSIM PARAMETERS-----
*
* Divers
.param dsub = drou
*
* Process parameters
.param toxe_bsim='toxe_mastar'
*-polydep_bsim'
*
* threshold voltage
.param vthsatlong = 'vth0 + delta'
*dev/gauss=(Avt/sqrt(2*w1*11/tometer/tometer))

.param lt0 = 'sqrt(epsilonSi*epsilon0*xdep/coxe_bsim)'
.param xdep='sqrt(2*epsilon0*epsilonSi*2*phif1/(q*Ndep_s))'
.param eta0 = 'DIBL_MASTAR/(vdd*(exp(-dsub*leff1/(2*lt0))+2*exp(-dsub*leff1/lt0)))'

*
* mobility
.param mueff1 = '(1+kw/w1)*kmu * musr * muac / (musr + muac)'

* SUBTHRESHOLD SLOPE
.param coxe_bsim = 'epsilon0*epsilonSiO2/(toxe_bsim)'
.param coxe='coxe_bsim'
.param cdep_bsim = 'epsilon0*epsilonSi/xdep'
.param vt='k*T'
.param nfactor = '(S/(vt*log(10))-1)*coxe_bsim/cdep_bsim'
.param I0='mueff1_s*sqrt(abs(q*epsilon0*epsilonSi*Ndep_s)/(2*2*phif1))*vt*vt'
.param Abulk='1+Kb/(2*sqrt(2*phif1))*a0*leff1/(leff1+2*sqrt(XJ1*xdep))'
.param n='1+nfactor*cdep_bsim/coxe_bsim'
.param voff1='n*vt*log(I0*Rearly/(Ith*Abulk))-delta'

* IGATE
.param a=1.44e9
.param b=-4.02
.param c=13.05
.param dl=1.17e10

* I fuites jonction
.param GIJ0=0
*1e-15

* OTHER
.param vgsteff0 = '2*n*vt*log(1+exp(-(vth0-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE)/(2*n*vt)))/(1+2*n*coxe_bsim/cdep_bsim*exp(
(vth0-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE+2*voff1)/(2*n*vt)))'
.param a0='Kdl*(leff1+2*sqrt(XJ1*xdep))/leff1'
.param esat='2*vsat1/mueff1_s'
.param litl='sqrt(epsilonSi/epsrox1*toxe_bsim*XJ1)'
.param vdsat='1/(1/(2*vsat1/mueff1_s*leff1)+Abulk/(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE)+2*vt))'
.param vdsat_mastar='1/(1/(esat*leff1)+(1+d)/(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE)))'
.param Rapport='ids0_mastar/ids0_bsim_eff'
.param vdsat_bsim='1/(1/(esat*leff1)+(1+d)/(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE)+2*vt))'
.param vdsat_eff='vdsat_bsim-0.5*(vdsat_bsim-vdd-DELTAvd+sqrt(POW((vdsat_bsim-vdd-DELTAvd),2)+4*DELTAvd*vdsat_bsim))'
.param ids0_bsim='0.5*w1/leff1*mueff1_s*coxe_mastar*(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE))*vdsat_bsim'
.param ids0_bsim_eff='w1/leff1*coxe_mastar*mueff1_s*(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-
NCE))/(1+vdsat_eff/(esat*leff1))*(1-Abulk*vdsat_eff/(2*(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-
NCE))+2*vt))*vdsat_eff'
.param ids0_mastar='0.5*w1/leff1*mueff1_s*coxe_mastar*(vdd-(vthsatlong-SCE_MASTAR-DIBL_MASTAR+RSCE-NCE))*vdsat_mastar'
.param C_correction_ION = 'correction_ION1*Rapport + (1-correction_ION1)*1'
.param Rearly='(1+(vdd-vdsat)/VA)'
.param VA='VAsat+VACLMIDL'
.param VAsat='(esat*leff1)'
.param VACLMIDL='(1+pvag*(esat*leff1))*1/(1/VACLM+1/VADIBL)'
.param VACLM='valif((vdd-vdsat)>1e-10,(Abulk*esat*leff1)*(vdd-vdsat)/(pclm*Abulk*esat*litl),exp(34))'
.param VADIBL='valif((tetarout)>=0,(2*vt)/(tetarout)*(1-Abulk*vdsat/(Abulk*vdsat+2*vt)),exp(34))'
.param tetarout='pdiblc1*(exp(-drou*leff1/(2*lt0))+2*exp(-drou*leff1/lt0))+pdiblc2'
.param cf='3.45e-11/1.57*log(1+tpoly/Toxe_phys)'
.param na='Ndep_s*w1*leff1*xdep_v1_ch' DEVM/gauss=10%
*{sqrt(Ndep_s*w1*leff1*xdep_v1_ch)}

*-----
*-----
*
* MODEL
*-----
*-----

gigd1 G D1 value='-1/TIF*lint*w1*a*exp(b*v(g,d1)*v(g,d1)+c*abs(v(g,d1)))*exp(-(d1*Toxe_phys))'
gigs1 G S1 value='1/TIF*(lint)*w1*a*exp(b*v(g,s1)*v(g,s1)+c*abs(v(g,s1)))*exp(-(d1*Toxe_phys))'
gigb1 G B value='1/TIF*(leff1)*w1*a*exp(b*v(g,b)*v(g,b)+c*abs(v(g,b)))*exp(-(d1*Toxe_phys))'

*gijsb S B value='GIJ0*v(s,b)'
*gijsb D B value='GIJ0*v(d,b)'

.lib "/soft/unicad/65nm_4_1_Feb1306/common/models/LPgp_all/bsim4/stress_def/eldo/LPdiode.lib" diodelp_typ

DSOURCE_NMOS B S1 dnsvtlp area=asl pj=ps1 pgate='w1'
DDRRAIN_NMOS B D1 dnsvtlp area=adl pj=pd1 pgate='w1'

rs S S1 rs1
rd D D1 rd1

```

```

M1 D1 G S1 B nsvt1p l='l1' w='w1' as=as1 pd=pd1 ad=ad1 ps=ps1 nrd=nrd nrs=nrs

.MODEL nsvt1p NMOS LEVEL = 53

* Threshold parameters
+ver=3,24
+delvt0=0
+vth0='vthsatlong+SCERSCE'
+vfb=vfb1
+k1='1*Kb'
+k2=0
+k3=0
+n1x=0
+dvt0=0
+dvt0w=0
+b0=0

* Subthreshold parameters

+voff=voff1
+nfactor = nfactor
+eta0=eta0
+drou=drout
+dsub=dsub

* Saturation parameters

+a0='a0'
+vsat='vsat1*C_correction_ION'
+mobmod=1
+u0='mueff1_s*C_correction_ION'
+ua=00
+ub=0
+uc=0
+cdsc=0
+cdscd=0
+cdscb=0
+pclm= 'pclm'
+pdiblc1= 'pdiblc1'
+pdiblc2='pdiblc2'
+pvag='pvag'
+pscbe2=1e-10

* Process parameters

+tox='toxe_bsim'
+ngate='ngate1'
+xj='XJ1'
+nch='Ndep_s*1e-6'
+lint = 'lint'
+x1=x1

* capacitance parameters
+capmod=capmod
+c1e=1
+c1c=0
+cgs0 = 'lint*coxe'
+cgs1 = 0
+cgd0 = 'lint*coxe'
+cgd1 = 0
+cf=cf
+noff=noff
+voffcv=voffcv
+dtocv=dtocv
+moin=20
+acde=1.4

* other
+delta=DELTAvd
+a1=0
+a2=1

+cjsw=0 cjswg=0 cj=0

.ends nsvt1p

.subckt psvt1p D G S B

*-----
*
*      PARAMETERS/CONSTANTS
*-----
*-----MISMATCH PARAMETERS-----

.param dev_ndep = 0 devx/gauss=sigmandep
.param sigmandep = 'valif(fluctdop==1,2/sqrt(3))*sqrt(ndep0_s/(w1*1eff1*xdep_v2_ch)),0)'

*-----PHYSICAL CONSTANTS-----
.param q = 1.6e-19
.param epsilon0 = 8.85e-12
.param epsilonSi = 11.7055
.param epsilonSiO2= epsrox1
.param k = '1.3806226e-23/q'
.param ni = '1.45e10*(T/300.15)*sqrt(T/300.15)*exp(21.5565981-eg/(2*k*(TC+273.15)))*1e6'
.param eg = '1.16-7.02e-4*T*(T+1108)'

*-----MOS PARAMETERS-----
*****

```

```

**PMOS**
*****

**MODEL SELECTOR**

    *mastar*
.param newSlope=1
.param S_manu='9.407e+001 *(1-newSlope) '
.param MastarCompute=1
.param DG=0
.param SOI='0 '
.param vdd = 1.100e+000

    *bsim*
.param capmod=3
.param correction_ION=1

**PROCESS PARAMETERS**

.param tox_mastar = 'Toxe_phys+darkspace+polydep+dev_toxe'
.param darkspace = '5.700e+000 *1e-10'
.param polydep = '3.700e+000 *1e-10'
.param Toxe_phys='1.700e+000 *1e-9'
.param ndep = '9.000e+017 '
.param r = 276.012970196425
.param XJ = '1.500e+001 *1e-9'
.param Tsi = '1.500e+001 *1e-9+dev_Tsi'
.param Tbox='1.450e+002 *1e-9'
.param Ngp=1.000e+020
.param ngatel = 0
*valif (polydep!=0,1.300e+020 ,0)
.param phim1 = 5.120e+000
.param nsd1 = 9.000e+019
.param xl=0
.param tpoly = '90 *1e-9'
.param Qss='-5.000e+011 *q*1e4*phi_c'

    *poches*

.param Poches = 1
.param Nsub = 5.000e+017
.param Energie_KeV =5.000e+001
.param angle = 2.500e+001
.param diffusion = 1.000e+000
.param activation = 1.000e+002
.param Dose_cm2 = 1.730e+013
.param species = '0 +1'
*AS = 1
*PH = 2
*SB = 3

    *High-K*

.param active_HK = 0
.param HiK_tox_A ='5.450e+000 *10'
.param epsilon_HK =2.500e+001
.param t_SiO2_pied_A =8.500e-001

**FITTING PARAMETERS**

    *mastar*
.param kmu=1.3
.param kw=6.000e-008
.param delta = 1.000e-001
.param gamma=4.000e-001
.param zeta1=8.000e-001
.param zeta2=1.000e-001
.param zeta3=1.000e+000
.param Kd = 1.000e+000
.param kvs = 1.000e+000
.param TIF=9.000e+001
.param kfield = 1.000e+000
.param vsat1 =' 1e5*kvs'
.param Ith=5.000e-008
.param epsrox1 = 3.9018
*AS = 1
*PH = 2
*SB = 3

    *bsim*
.param polydep_bsim=polydep
.param dtoxcv=0
.param noff=1
.param voffcv=0
.param pvag=0
.param pclm=1e-10
.param pdiblc1=1e-10
.param pdiblc2=1e-10
.param DELTAvd=0.1
.param drout=0.56
.param T=300
.param TC='300-273.15'

**MISMATCH PARAMETERS**

.param dev_l=0.0 devx/gauss=0e-9
.param dev_w=0.0 devx/gauss=0e-9
.param dev_r=0.0
.param fluctdop=1
.param dev_toxe=0.0 devx/gauss=0e-10
.param dev_Tsi=0.0 devx/gauss=0e-10

```

```

.param nrd=0
.param nrs=0
.param rd1='r/w/2'
.param rsl=rd1
.param ll='l*tometer + dev_l'
.param wl='valif(DG1==0,w*tometer,w*2*tometer) + dev_w'
.param asl='AS*tometer*tometer'
.param adl='AD*tometer*tometer'
.param psl='PS*tometer'
.param pdl='PD*tometer'
.param tometer=1e-6
.param MastarCompute1='valif(SOI1==1,1,MastarCompute)'
.param zeta21=zeta2
*'valif(zeta2==0,0.64*(1-SOI1)*(1-DG1)+SOI1*1+0.2*DG1,zeta2)'
.param zeta11=zeta1
*'valif(zeta1==0,0.8*(1-SOI1)*(1-DG1)+SOI1*0.77+0.35*DG1,zeta1)'
.param Kd1 = 'valif(Kd==0,1-0.5*SOI1-DG1,Kd)'
.param XJ1='Tsi/2*DG1 + Tsi*SOI1 + (1-DG1)*(1-SOI1)*XJ'
.param SOI1 = 'valif((DG==1),0,valif((SOI==2),1,0))'
.param DG1 = 'valif((DG==1),1,0)'
.param correction_ION1 = 'valif(((correction_ION)==(0)),0,1)'
.param ro = 2.329e3
.param midgap=4.57

*-----SCALED PARAMETERS-----

.param Ndep1 = 'valif(Poches==1,(nsub+Nmoy),ndep)'
.param Ndep0_s = 'Ndep1*1e6'
.param Ndep_s = 'abs(Ndep0_s + dev_ndep)'
.param Nsub_s = 'valif(Poches==1,nsub*1e6,Ndep_s)'
.param ngate1_s = 'ngate1 * 1e6'
.param nsd1_s = 'nsd1*1e6'
.param mueff1_s = 'mueff1/1e4'
.param Ngp_s = 'Ngp*1e6'

*---MASTAR CALCULATED PARAMETERS---

.param coxe_mastar = 'epsilon0*epsilonSiO2/((toxe_mastar))'
.param Kb = 'sqrt(2*q*Ndep_s*epsilon0*epsilonSi) /coxe_mastar'
.param d = 'Kd1*Kb/(2*sqrt(2*phif1))'
.param phim2='midgap-(phim1-midgap)'
.param vfb1 = 'phim2-phisl - Qss/coxe_mastar'
.param phif1 = 'k*T*log(Ndep_s/ni)'
.param phisl = '4.05+0.56*phif1'
.param phid = 'k * T * log(nsd1_s*Ndep_s/(ni*ni))'
.param phid0 = 'k * T * log(nsd1_s*Ndep0_s/(ni*ni))'
.param lint = 'gamma/2*XJ1 + DG1*gamma/2*XJ1'
.param leff1 = 'll+xl-2*lint'
.param Tsub = '-(Tsi+3*Tbox)+sqrt(pow((Tsi+3*Tbox),2)+2*epsilon0*epsilonSi*phid/(q*Ngp_s)-Nsub_s/Ngp_s*Tsi*Tsi)'
.param psibox='q*Ngp_s*Tsub/coxe_mastar'
.param psisup = 'phid-2*phif1-k*T*log(k*T/q/Nsub_s*coxe_mastar*exp( (1/(k*T)+1)/2 ) )'
.param vth0soi='- (vfb1 + phid + psibox + q*Nsub_s*Tsi/coxe_mastar)'
.param vth0dg='- (vfb1 + phid - psisup + q*Nsub_s*xdep_v1/coxe_mastar)'
.param SCERSCE='RSCE-SCE_MASTAR-NCE'

*****POCHES*****
.param x = 'Energie_KeV'
.param angle_rad = 'angle *pi/180.0'
.param pi= 3.1415928

.param Rp_A='valif(species==1,2e-07*pow(x,5.0) - 6e-05*pow(x,4.0) + 0.006*pow(x,3.0) - 0.2948*pow(x,2.0) + 13.097*(x) +
14.514,valif(species==2, 2e-07*pow(x,5.0) - 6e-05*pow(x,4.0) + 0.0062*pow(x,3.0) - 0.2834*pow(x,2.0) + 18.636*(x) +
1.7019,valif(species==3,3e-07*pow(x,5.0) - 8e-05*pow(x,4.0) + 0.0082*pow(x,3.0) - 0.4174*pow(x,2.0) + 14.607*(x) +
2.7257,0))'
.param DeltaRp_A= 'valif(species==1,1e-07*pow(x,5.0) - 3e-05*pow(x,4.0) + 0.0033*pow(x,3.0) - 0.165*pow(x,2.0) +
5.6693*(x) + 1.137,valif(species==2, 1e-07*pow(x,5.0) - 4e-05*pow(x,4.0) + 0.0041*pow(x,3.0) - 0.2071*pow(x,2.0) +
9.588*(x) + 1.2162,valif(species==3, 1e-07*pow(x,5.0) - 3e-05*pow(x,4.0) + 0.003*pow(x,3.0) - 0.1561*pow(x,2.0) +
4.8485*(x) + 1.0707, 0))'
.param DeltaRl_A= 'valif(species==1,8e-08*pow(x,5.0) - 2e-05*pow(x,4.0) + 0.0024*pow(x,3.0) - 0.122*pow(x,2.0) +
4.2986*(x) + 0.7027,valif(species==2, 9e-08*pow(x,5.0) - 3e-05*pow(x,4.0) + 0.0029*pow(x,3.0) - 0.1516*pow(x,2.0) +
6.9764*(x) + 0.9221, valif(species==3, 7e-08*pow(x,5.0) - 2e-05*pow(x,4.0) + 0.0022*pow(x,3.0) - 0.112*pow(x,2.0) +
3.6301*(x) + 0.7302, 0 ))'

.param Rp_um = 'Rp_A*cos(angle_rad)*1e-4'
.param DeltaRp_um = 'diffusion * DeltaRp_A*1e-4'
.param DeltaRl_um = 'diffusion * DeltaRl_A*1e-4'
.param Cpoches_cm2 = '(activation/100.0)*(Dose_cm2/4.0)'

.param Npoches = '0.5*Cpoches_cm2/(Rp_um*0.0001+2.0*DeltaRp_um*0.0001)'
.param Lpoches_um = '(Rp_um+2.0*DeltaRp_um)*sin(angle*pi/180.0)+2.0*DeltaRl_um*cos(angle*pi/180.0)-
gamma*XJ1*1e9/2000.0'
.param Lmin_um = 'Min((ll+xl)*1e6,Lpoches_um)'
.param Nmoy = '2.0*Npoches*Lmin_um/((ll+xl)*1e6)'

.param XJ2 = '2.0/3.0*1e9 * XJ1'
.param Lelec = 'leff1/1e-9'
.param alphaSlope = '0.5*(1.0+(14.0/Lelec))'
.param phi_d = phid
.param phi_c = 'k*T*log(Ndep_s/ni)+k*T*log(Ndep_s/ni)'
.param root1 = 'valif((phi_d- alphaSlope *phi_c)>0,(phi_d- alphaSlope *phi_c),0)'
.param root2 = 'valif((phi_d- alphaSlope *phi_c/2)>0,(phi_d- alphaSlope *phi_c/2),0)'
.param Nu1 = 'vdd+2*(phi_d - alphaSlope *phi_c)+2*sqrt(root1*(vdd+phi_d- alphaSlope *phi_c))'
.param Nu2 = 'vdd+2*(phi_d - alphaSlope *phi_c/2)+2*sqrt(root2*(vdd+phi_d- alphaSlope *phi_c/2))'
.param Ndep_s_star1= 'Ndep_s - 2.0* epsilon0*epsilonSi * Nu1 / q /pow((Lelec)* 1e-9, 2.0)'
.param Ndep_s_star2= 'Ndep_s - 2.0* epsilon0*epsilonSi * Nu2 / q /pow((Lelec)* 1e-9, 2.0)'
.param Ttest1 = 'abs(Ndep_s_star1)/Ndep_s_star1*sqrt(abs(2*phi_c*epsilon0*epsilonSi/q/Ndep_s_star1))'
.param Ttest2 = 'abs(Ndep_s_star2)/Ndep_s_star2*sqrt(abs(phi_c*epsilon0*epsilonSi/q/Ndep_s_star2))'

```

```

.param Tdep1 = 'valif((Ttest1*1e9 < XJ2) && (Ttest1 *1e9 > 0.0),Ttest1,sqrt((q*(Ndep_s-
Ndep_s_star1)*pow(XJ2*1e-9,2.0) + 2.0*phi_c * epsilon0*epsilonSi)/q/Ndep_s)')
.param Qdep1 = 'valif((Ttest1*1e9 < XJ2) && (Ttest1 *1e9 > 0.0),q * Ndep_s_star1 * Tdep1,q*Tdep1
* Ndep_s - 2.0 * Nu1 *epsilon0*epsilonSi *(XJ2*1e-9)/pow(Lelec*1e-9,2.0))'
.param Tdep2 = 'valif((Ttest2*1e9 < XJ2) && (Ttest2 *1e9 > 0.0),Ttest2,sqrt((q*(Ndep_s-
Ndep_s_star2)*pow(XJ2*1e-9,2.0) + phi_c * epsilon0*epsilonSi)/q/Ndep_s)')
.param Qdep2 = 'valif((Ttest2*1e9 < XJ2) && (Ttest2 *1e9 > 0.0),q * Ndep_s_star2 * Tdep2,q*Tdep2
* Ndep_s - 2.0 * Nu2 *epsilon0*epsilonSi *(XJ2*1e-9)/pow(Lelec*1e-9,2.0))'
.param essai= log(10.0)
.param S_newslope = 'k*T*essai*(1.0+abs(Qss)/coxe_mastar+(Qdep1-Qdep2)/phi_c*2.0/coxe_mastar)'
.param S_MASTAR='valif(MastarComputel==1,S_MASTAR_v1,S_MASTAR_v2)'
.param S_SOI='k*T*log(10)*(1+(epsilonSi/epsilonSiO2)*toxe_mastar/(xdep_v1+ (SOI1)*(epsilonSi/epsilonSiO2)*Tbox + Tsub))'
.param vthsat = 'abs(vth0 + SCE_MASTAR + DIBL_MASTAR -RSCE+NCE - delta)'

.param vth0='(1-SOI1)*(1-DG1)*valif(MastarComputel==1,vth0_v1,vth0_v2)+SOI1*vth0soi+DG1*vth0dg'
.param SCE_MASTAR = 'valif(MastarComputel==1,SCE_v1,SCE_v2)'
.param DIBL_MASTAR = 'valif(MastarComputel==1,DIBL_v1,DIBL_v2)'
.param RSCE = 'valif(Poches==1,sqrt(2*epsilonSi*epsilon0*q*Nsub_s*2*phif1)/coxe_mastar*(sqrt(Ndep_s/Nsub_s)-1),0)'
.param musr = 'valif(MastarComputel==1,140 * PWR(Eeff_v1,-1),140 * PWR(Eeff_v2,-1))'
.param muac = 'valif(MastarComputel==1,90 * PWR(Eeff_v1,-0.3),90 * PWR(Eeff_v2,-0.3))'

.param S='valif(S_manu==0,valif(newSlope==1,S_newslope,S_MASTAR),S_manu/1000)'
.param NCE = '(epsilonSi/epsilonSiO2)*((toxe_mastar*xdep_v1_ch)/(w1*w1))*zeta3'

* phid (v1)
.param epsilon_moy = '(HiK_tox_A*epsilon_HK + t_SiO2_pied_A *epsilonSiO2) / (HiK_tox_A + t_SiO2_pied_A)'

.param xdep_v1 =
'min(sqrt(2*epsilon0*epsilonSi*phid/(q*Nsub_s)),Tsi)*SOI1+min(sqrt(2*epsilon0*epsilonSi*phid/(q*Nsub_s)),Tsi/2)*DG1+sqrt(2
*epsilon0*epsilonSi*phid/(q*Nsub_s))*(1-SOI1)*(1-DG1)'
.param xdep_v1_ch =
'min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep_s)),Tsi)*SOI1+min(sqrt(2*epsilon0*epsilonSi*phid/(q*Ndep_s)),Tsi/2)*DG1+sqrt(2
*epsilon0*epsilonSi*phid/(q*Ndep_s))*(1-SOI1)*(1-DG1)'
.param xdep_v2_ch =
'min(sqrt(2*epsilon0*epsilonSi*phid0/(q*Ndep0_s)),Tsi)*SOI1+min(sqrt(2*epsilon0*epsilonSi*phid0/(q*Ndep0_s)),Tsi/2)*DG1+s
qrt(2*epsilon0*epsilonSi*phid0/(q*Ndep0_s))*(1-SOI1)*(1-DG1)'
.param vth0_v1 = '-(vfb1 + phid + 1/coxe_mastar*sqrt(2*epsilon0*epsilonSi*q*phid*Nsub_s))'
.param SCE_v1 = 'zeta21 * (1+active_HK*(epsilon_moy-15.0)*0.0078)*(epsilonSi/epsilonSiO2) * (1 + XJ1*XJ1/PWR(leff1,2)) *
(toxe_mastar/leff1) * (xdep_v1_ch/leff1) * phid'
.param delta_DIBL_box = 'valif((pow(Tsi+3*Tbox,2) < (leff1*leff1)),0,(toxe_mastar * Tsi/(leff1*leff1)) *
(sqrt(pow(Tsi+3*Tbox,2)-(leff1*leff1)) / leff1) * Vdd)'
.param DIBL_v1 = 'zeta11 * (1+active_HK*((HiK_tox_A+t_SiO2_pied_A)-20)*0.015)*(epsilonSi/epsilonSiO2) * (1 +
XJ1*XJ1/PWR(leff1,2)) * (toxe_mastar/leff1) * ((xdep_v1_ch+SOI1*((leff1-Tsi)*3*Tbox)/(leff1-Tsi+3*Tbox))/leff1) * vdd +
0.16*delta_DIBL_box*SOI1'
.param S_MASTAR_v1 = 'k*T*log(10)*(1+epsilon0*epsilonSi*toxe_mastar/(epsilon0*epsilonSiO2*xdep_v1))'
.param Eeff_v1 = 'kfield*((vdd + 2*vthsat-3*(vfb1+phid-DG1*psisup+SOI1*psibox))/(9*toxe_mastar))*1e-8'
.param Qdep_Cox1 = 'vthsat-vfb1-phid -SOI1*psibox + DG1*psisup'
.param Qdep_Cox='valif(Qdep_Cox1<0,0, Qdep_Cox1)'
.param Eeff_v1 = 'kfield*((vdd-vthsat)/(9*toxe_mastar) + (3.0 * Qdep_Cox / (9*toxe_mastar)))*1e-8'

* 2phif (v2)
.param vth0_v2 = '-(vfb1 + 2*phif1 + 1/coxe_mastar*sqrt(2*epsilon0*epsilonSi*q*2*phif1*Nsub_s))'
.param xdep_v2 = 'sqrt(2*epsilon0*epsilonSi*2*phif1/(q*Nsub_s))'
.param SCE_v2 = 'zeta21 * (epsilonSi/epsilonSiO2) * (1 + XJ1*XJ1/PWR(leff1,2)) * (toxe_mastar/leff1) * (xdep_v2/leff1) *
2*phif1'
.param DIBL_v2 = 'zeta11 * (epsilonSi/epsilonSiO2) * (1 + XJ1*XJ1/PWR(leff1,2)) * (toxe_mastar/leff1) * (xdep_v2/leff1) *
vdd'
.param S_MASTAR_v2 = 'k*T*log(10)*(1+epsilon0*epsilonSi*toxe_mastar/(epsilon0*epsilonSiO2*xdep_v2))'
.param Eeff_v2 = 'kfield*((vdd + 2*vthsat-3*(vfb1+2*phif1))/(9*toxe_mastar))*1e-8'

*-----BSIM PARAMETERS-----
* Divers
.param dsub = drout

* Process parameters
.param toxe_bsim='toxe_mastar'
*~polydep_bsim'

* threshold voltage

.param vthsatlong = 'vth0 - delta'
*dev/gauss={Avt/sqrt(2*w1*11/tometer/tometer)}
*dev/gauss={Avt/sqrt(2*w1*11/tometer/tometer)}

.param lt0 = 'sqrt(epsilonSi*epsilon0*xdep/coxe_bsim)'
.param xdep='sqrt(2*epsilon0*epsilonSi*2*phif1/(q*Ndep_s))'
.param eta0 = 'DIBL_MASTAR/(vdd*(exp(-dsub*leff1/(2*lt0))+2*exp(-dsub*leff1/lt0)))'

* mobility

.param mueff1 = '(1+kw/w1)*kmu * musr * muac / (musr + muac)'

* SUBTHRESHOLD SLOPE

.param coxe_bsim = 'epsilon0*epsilonSiO2/(toxe_bsim)'
.param coxe=coxe_bsim
.param cdep_bsim = 'epsilon0*epsilonSi/xdep'
.param vt='k*T'
.param nfactor = '(S/(vt*log(10))-1)*coxe_bsim/cdep_bsim'
.param I0='mueff1_s*sqrt(abs(q*epsilon0*epsilonSi*Ndep_s)/(2*2*phif1))*vt*vt'
.param Abulk='1+Kb/(2*sqrt(2*phif1))*a0*leff1/(leff1+2*sqrt(XJ1*xdep))'
.param n='1+nfactor*cdep_bsim/coxe_bsim'
.param voff1='n*vt*log(I0*Rearly/(Ith*Abulk))-delta'

* IGATE
.param a=1.44e9
.param b=-4.02
.param c=13.05
.param dl=1.17e10

```

```

* I fuites jonction
.param GIJ0=0
*1e-15

*
OTHER

.param vgsteff0 ='2*n*vt*log(1+exp(-(vth0+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE)/(2*n*vt)))/(1+2*n*coxe_bsim/cdep_bsim*exp(
(vth0+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE+2*voff1)/(2*n*vt)))'
.param a0='Kd1*(leff1+2*sqrt(XJ1*xdep))/leff1'
.param esat='2*vsat1/mueff1_s'
.param litl='sqrt(epsilonSi/epsrox1*toxe_bsim*XJ1)'
.param vdsat='1/(1/(2*vsat1/mueff1_s*leff1)+Abulk/(vdd+(vthsatlong+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE)+2*vt))'
.param vdsat_bsim='1/(1/(esat*leff1)+(1+d)/(vdd+(vthsatlong+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE)+2*vt))'
.param vdsat_mastar='1/(1/(esat*leff1)+(1+d)/(vdd+(vthsatlong+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE)))'
.param vdsat_eff='vdsat_bsim-0.5*(vdsat_bsim-vdd-DELTAvd+sqrt(POW((vdsat_bsim-vdd-DELTAvd),2)+4*DELTAvd*vdsat_bsim))'
.param Rapport='ids0_mastar/ids0_bsim_eff'
.param ids0_bsim='0.5*w1/leff1*mueff1_s*coxe_mastar*(vdd+(vthsatlong+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE))*vdsat_bsim'
.param ids0_mastar='w1/leff1*mueff1_s*coxe_mastar*(vdd+(vthsatlong+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE))/(1+vdsat_eff/(esat*leff1))*(1-Abulk*vdsat_eff/(2*(vdd+(vthsatlong+SCE_MASTAR+DIBL_MASTAR-RSCE+NCE))+2*vt))*vdsat_eff'
.param C_correction_ION = 'correction_ION1*Rapport+ (1-correction_ION1)*1'
.param Rearly='(1+(vdd-vdsat)/VA)'
.param VA='VAsat+VACLMDIBL'
.param VAsat='(esat*leff1)'
.param VACLMDIBL='(1+pvag*(esat*leff1))*1/(1/VACLMDIBL+1/VADIBL)'
.param VACLMDIBL='valif((vdd-vdsat)>1e-10,(Abulk*esat*leff1)*(vdd-vdsat)/(pclm*Abulk*esat*litl),exp(34))'
.param VADIBL='valif((tetarout)>0,(2*vt)/(tetarout)*(1-Abulk*vdsat/(Abulk*vdsat+2*vt)),exp(34))'
.param tetarout='pdiblc1*(exp(-drout*leff1/(2*lt0))+2*exp(-drout*leff1/lt0))+pdiblc2'
.param cf='3.45e-11/1.57*log(1+tpoly/Toxe_phys)'
* .param na='Ndep_s*w1*leff1*xdep_vl_ch' DEVG/gauss=10%
*{sqrt(Ndep_s*w1*leff1*xdep_vl_ch)}

*-----
*
* MODEL
*-----

gigd1 G D1 value='1/TIF*lint*w1*a*exp(b*v(g,d1)*v(g,d1)+c*abs(v(g,d1)))*exp(-(d1*Toxe_phys))'
gigs1 G S1 value='-1/TIF*(lint)*w1*a*exp(b*v(g,s1)*v(g,s1)+c*abs(v(g,s1)))*exp(-(d1*Toxe_phys))'
gigb1 G B value='-1/TIF*(leff1)*w1*a*exp(b*v(g,b)*v(g,b)+c*abs(v(g,b)))*exp(-(d1*Toxe_phys))'

*gijsb S B value='GIJ0*V(s,b)'
*gijdb D B value='GIJ0*V(d,b)'

.lib "/soft/unicad/65nm_4_1_Feb1306/common/models/LPGP_all/bsim4/stress_def/eldo/LPdiode.lib" diodelp_typ

DSOURCE_PMOS S1 B dpsvtlp area=as1 pj=ps1 pgate='w1'
DDRRAIN_PMOS D1 B dpsvtlp area=ad1 pj=pd1 pgate='w1'

rs S S1 rs1
rd D D1 rd1

M1 D1 G S1 B psvtlp l='11' w='w1' as=as1 pd=pd1 ad=ad1 ps=ps1 nrd=nrd nrs=nrs

.MODEL psvtlp PMOS LEVEL = 53

* Threshold parameters
+ver=3.24
+deltvt0='0'
+vth0='vthsatlong-SCERSCE'
+vfb='vfb1'
+k1='Kb'
+k2=0
+vbm='S'
+K3=0
+n1x=0
+dvt0=0
+dvt0w=0
+b0=0

* Subthreshold parameters
+voff='voff1'
+nfactor = 'nfactor'
+eta0='eta0'
+drout='drout'
+dsb=dsb

* Saturation parameters
+a0=a0
+vsat='vsat1*C_correction_ION'
+rds=0
+mobmod=1
+cdsc=0
+cdscd=0
+cdscb=0
+ua=0
+ub=0
+u0='mueff1_s*C_correction_ION'
+uc=0
+pclm='pclm'
+pdiblc1='pdiblc1'
+pdiblc2='pdiblc2'
+pvag=pvag
+psche2=1e-10

* Process parameters

```

```
+tox='toxe_bsim'  
+ngate=ngate1  
+xj=XJ1  
+nch='Ndep_s*1e-6'  
+lint = '1*lint'  
+xl=xl  
  
* Capacitance parameters  
  
+capmod=capmod  
+cle=1  
+clc=0  
+cgso = 'lint*coxe'  
+cgs1 = 0  
+cgdo = 'lint*coxe'  
+cgdl = 0  
+cf=cf  
+noff=noff  
+voffcv=voffcv  
+dtoxcv=dtoxcv  
+acde=1.4  
+moin=5  
  
* OTHER  
  
+delta=DELTAvd  
+a1=0  
+a2=1  
  
+cjsw=0 cjswg=0 cj=0  
  
.ends psvtlp  
.endl
```

A.2 Détails concernant le bloc arithmétique utilisé pour l'évaluation du nœud 32nm

Le schéma du bloc arithmétique ayant servi à évaluer les performances du nœud 32nm est présenté Figure A-1 et Figure A-2. L'ambition de ce bloc est de favoriser la présence de longues lignes d'interconnexion en réalisant des opérations dépendant de nombreux résultats d'autres opérations.

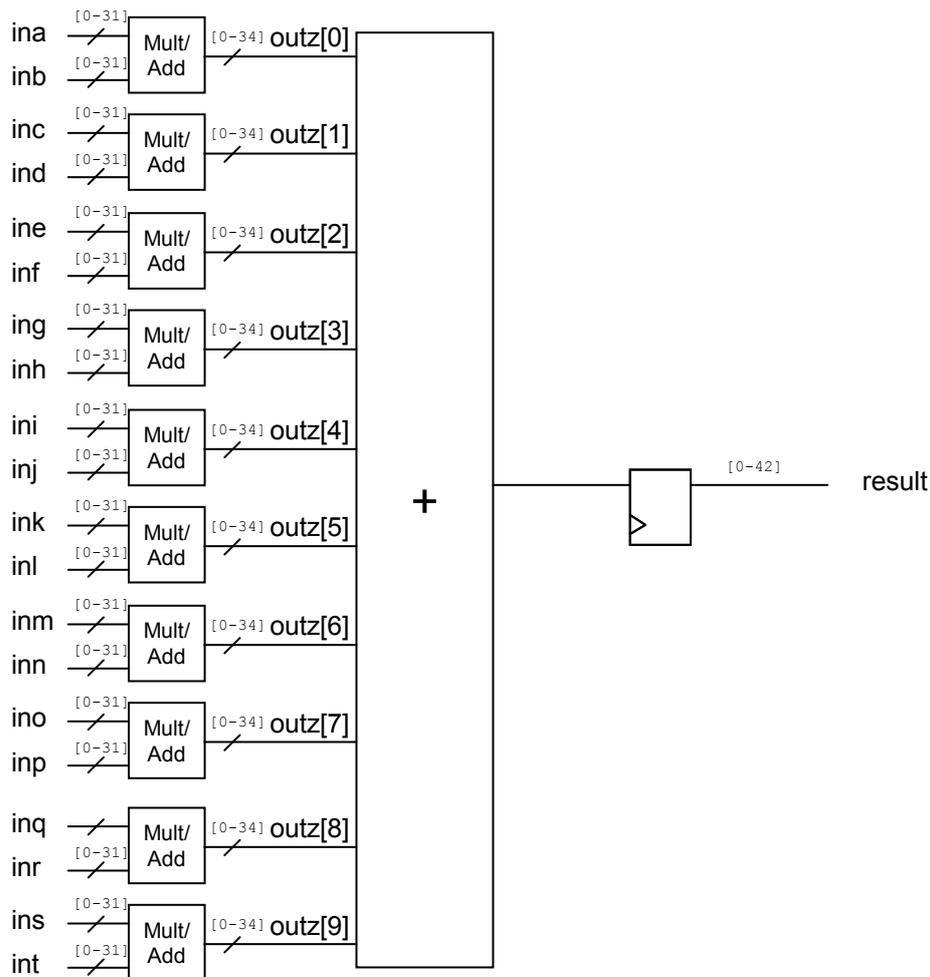


Figure A-1 : Schéma du bloc arithmétique.

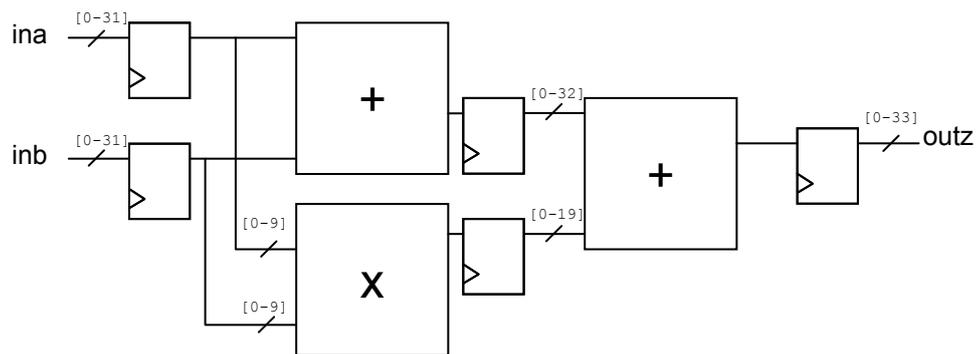


Figure A-2 : Détail du sous bloc Mult/Add.

A.3 Contraintes de délais dans un bloc placé routé

A.3.A Temps de « hold » et « setup »

Un bloc placé routé digital consiste en une succession de registres entre lesquels est insérée de la logique combinatoire. A chaque coup d'horloge, les données passent d'un côté à l'autre des registres pour poursuivre leur traitement. Les contraintes au niveau des délais de cellule sont duales. En effet, afin de garantir l'intégrité des informations transmises au niveau des bascules, il est nécessaire que les données soient présentes de manière stable avant et après le coup d'horloge. Le laps de temps à respecter avant le coup d'horloge est appelé temps de « setup » alors que celui après le coup d'horloge est appelé temps de « hold ».

Sur les chemins correspondant à un seul cycle d'horloge (« mono cycle path »), il est nécessaire de respecter les relations suivantes afin de garantir un traitement séquentiel correct des données (cf. Figure A-3) :

$$T_{CLK \rightarrow Q} + \tau_{comb} + \tau_{setup} < \tau_{CLK} + \tau_{skew}, \quad (A-1)$$

avec $T_{CLK \rightarrow Q}$ le temps de transmission de la donnée dans la bascule, T_{comb} le délai lié au traitement des données dans la partie combinatoire, T_{setup} le temps de « setup » de la bascule, T_{CLK} la période de l'horloge et T_{skew} la différence de temps entre les fronts d'horloge entre deux bascules successives.

Cette relation spécifie que la donnée doit être présente en avance d'un temps T_{setup} sur le front d'horloge. Il est possible de voir la relation (A-1) sous la forme d'une contrainte à respecter sur le temps de « setup » :

$$T_{setup} < \tau_{CLK} + \tau_{skew} - \tau_{CLK \rightarrow Q} - \tau_{comb}. \quad (A-2)$$

Si cette contrainte n'est pas respectée, il est dit qu'il y a une « violation de setup », c'est-à-dire que la donnée arrive trop en retard pour respecter les contraintes de « setup » de la bascule. Il y a alors un risque de transmission d'une valeur erronée.

Les délais doivent également respecter la contrainte suivante :

$$T_{hold} < \tau_{CLK \rightarrow Q} + \tau_{comb} - \tau_{skew}, \quad (A-3)$$

avec T_{hold} le temps de « hold ».

Cette relation spécifie qu'une donnée ne doit pas arriver en entrée de la bascule pendant le temps de « hold ». Si cela se produit, il y a une « violation de hold », c'est-à-dire que la

donnée change d'état trop tôt et qu'il y a un risque de transmettre une donnée erronée à travers la bascule.

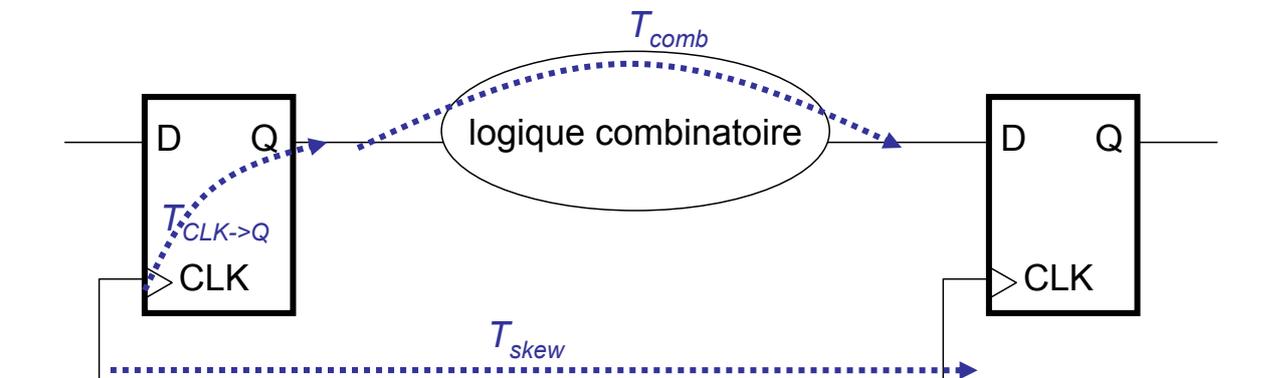


Figure A-3 : Les différents délais d'un chemin dans un bloc logique.

A.3.B Définition des « slacks »

Lors de la vérification des délais des circuits, les outils calculent des marges ou « slack » au niveau des temps de « hold » et de « setup » correspondant à la différence entre les temps simulés et les spécifications à respecter pour la bascule. Lorsque la marge nécessaire n'est pas respectée, la différence est négative et le terme « negative slack » est employé.

Afin de caractériser la marge au niveau des contraintes de délai d'un bloc entier, la plus faible marge (« Worst-case Negative Slack » ou WNS) est souvent repérée ainsi que la somme totale des marges négatives (« Total Negative Slack » ou TNS) à la fois pour les temps de « hold » et de « setup ».

L'objectif de cette étude est de fournir des éléments d'évaluation des futures technologies CMOS au niveau circuit. Dans ce but, des kits de conception prédictifs sont élaborés. Ces kits reposent sur la modélisation prédictive des futurs dispositifs et des interconnexions, ainsi que sur le paramétrage des outils nécessaires au déroulement d'un flot digital dans le cadre de futures technologies.

Les résultats des évaluations réalisées grâce à ces kits mettent en évidence une augmentation drastique des délais d'interconnexion laissant augurer d'importants problèmes d'ajout de répéteurs pour les futurs circuits. A court terme (32nm), l'évaluation réalisée dans le cadre d'un flot digital entièrement prédictif montre que les problèmes posés par les délais d'interconnexion ne semblent pas encore jouer un rôle important pour les blocs de faible dimension.

Concernant la variabilité des dispositifs, qui affecte tout particulièrement les circuits de type mémoires SRAM, une stagnation à des niveaux non acceptables est observée pour les technologies futures. Cependant, à court terme, des solutions consistant à utiliser des dispositifs faiblement dopés sont identifiées. L'intérêt d'une nouvelle mémoire SRAM, dont le principe réside dans l'utilisation de dispositifs faiblement dopés seulement pour les transistors NMOS, est également démontré.

The goal of this study is to perform circuit level assessment of future CMOS technologies. To this end, predictive design kits have been made. These kits rely on future devices and interconnect modeling, and on the digital flow tools parameter setting in the framework of future technologies.

The results of the evaluations carried out thanks to these kits show a drastic increase of interconnect delays suggesting that there will be an important issue of repeater adding for the future circuits. In the short run (32nm), the evaluation led with the predictive design flow show that the interconnect delay problem do not seem to play an important role for small blocs.

Concerning the variability of the devices which affect all particularly circuits such as SRAM memories, stagnation at non acceptable levels is highlighted although solutions consisting in using undoped devices are identified. The worth use of a new SRAM memory consisting of using undoped devices for NMOS transistors only is also shown.