



**HAL**  
open science

# METHODOLOGIES DE TEST DEIEES AUX MEMOIRES EMBARQUEES

Hassen Aziza

► **To cite this version:**

Hassen Aziza. METHODOLOGIES DE TEST DEIEES AUX MEMOIRES EMBARQUEES. Micro et nanotechnologies/Microélectronique. Aix Marseille Université (AMU), 2012. tel-03139646v2

**HAL Id: tel-03139646**

**<https://hal.science/tel-03139646v2>**

Submitted on 16 Feb 2021 (v2), last revised 22 Dec 2021 (v3)

**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.

# MEMOIRE

présenté devant

AIX-MARSEILLE UNIVERSITE

par

**Hassen AZIZA**

Ingénieur-Docteur  
Maître de Conférences

pour obtenir

## L'HABILITATION A DIRIGER DES RECHERCHES

**METHODOLOGIES DE TEST DEIEES AUX MEMOIRES  
EMBARQUEES**

Soutenue le 30 Novembre 2012 devant la commission d'examen :

Rapporteurs	<b>Amara AMARA</b>	ISEP, Paris
	<b>Gilles JACQUEMOD</b>	Université de NICE
	<b>Patrick GIRARD</b>	LIRMM, Montpellier
Examineurs	<b>Jean-Luc AUTRAN</b>	IM2NP, Aix-Marseille Université
	<b>Rachid BOUCHAKOUR</b>	IM2NP, Aix-Marseille Université
	<b>Pascal MASSON</b>	Université de NICE
	<b>Jean-Michel PORTAL</b>	IM2NP, Aix-Marseille Université
	<b>Paolo PRINETTO</b>	Politecnico di Torino, ITALIE



# Introduction générale

Les travaux de recherche présentés pour l'obtention de L'Habilitation à Diriger des Recherches ont été menés entre début 2001 et Juin 2012 à l'Institut de Matériaux Microélectronique Nanosciences de Provence (IM2NP, UMR CNRS 7334 - Aix-Marseille Université).

Mes travaux se sont focalisés sur les circuits hétérogènes intégrant sur un même substrat de silicium les technologies suivantes :

- Technologie CMOS avancée pour le numérique,
- Blocs analogiques, notamment pour la gestion de la puissance,
- Mémoires FLASH/EEPROM pour le stockage non volatile.

Ces technologies sont présentes dans la plupart des systèmes sur puce et plus particulièrement dans les circuits microcontrôleurs.

Un **système sur puce** ou système mono-puce (SoC<sup>1</sup>) est un circuit qui intègre de la mémoire, un ou plusieurs microprocesseurs, des circuits analogiques, des périphériques d'interface, des fonctions RF<sup>2</sup>, mais aussi des capteurs mécaniques, optoélectroniques, chimiques ou biologiques, le tout sur un unique substrat. Une application typique des systèmes sur puce se retrouve dans le domaine des systèmes embarqués. Un **système embarqué** est un système composé à la fois d'une partie logicielle et d'une partie matérielle, conçu de manière à réaliser une tâche spécifique dans un système plus vaste. Un système embarqué est généralement associé à des applications soumises à des contraintes de fonctionnement temps réel et basse consommation.

Les systèmes embarqués contiennent des cœurs de traitement qui sont généralement des **microcontrôleurs** ou des processeurs de traitement de signal numérique (DSP<sup>3</sup>). Contrairement à un système à microprocesseurs qui nécessite traditionnellement des éléments externes pour son fonctionnement (horloge, mémoires, périphériques, etc.), un microcontrôleur est un composant autonome, capable d'exécuter le programme contenu dans sa mémoire dès sa mise sous tension. Ainsi, les microcontrôleurs améliorent l'intégration et le coût (lié à la conception et à la réalisation) d'un système en rassemblant les éléments essentiels à leur fonctionnement sur un seul circuit intégré.

Les caractéristiques des **Mémoires Non Volatiles** (MNV) répondent à deux exigences majeures des systèmes embarqués les plus avancés : d'une part, la non volatilité qui permet de stocker des informations dans les applications portables durant plus de 10 ans sans alimentation électrique, d'autre part la programmation électrique dans l'application qui permet de modifier et donc de faire évoluer le système. Ainsi, les MNV constituent des composants stratégiques majeurs dans les systèmes complexes tels que les microcontrôleurs. Leur capacité ne cesse d'augmenter et les applications qui les utilisent sont de plus en plus nombreuses. De ce fait, cette technologie joue un rôle crucial dans le développement du marché des systèmes à microcontrôleurs qui représente la plus grosse partie des ventes dans le marché des microprocesseurs.

En effet, les microcontrôleurs sont utilisés dans de nombreux systèmes embarqués liés à la téléphonie mobile, à l'automobile, aux récepteurs GPS, à l'électroménager, au transport, à la robotique et aux cartes à puces. De plus, l'essor des systèmes embarqués ouvre les perspectives, à moyen terme, d'un internet des objets. Dès aujourd'hui, un nombre croissant d'objets de la vie quotidienne sont des objets connectés, entre eux ou à internet, qui tirent

---

<sup>1</sup> System on Chip

<sup>2</sup> Radio Fréquence

<sup>3</sup> Digital Signal Processor : processeur de traitement du signal

de cette connectivité des services innovants, avec une réelle valeur ajoutée pour les utilisateurs : interface plus ergonomique, accessibilité, stockage des données en ligne, etc.

**Le test des microcontrôleurs** passe par le test de chacun de ses éléments et conduit à la réalisation de flots de tests denses et complexes.

La principale cause de perte de rendement de ce type de produit est liée à la partie MNV. En effet, la structure et le principe de fonctionnement spécifique de la MNV (oxydes fins, tensions de programmation élevées, densité du plan mémoire élevée) en font un composant particulièrement sensible, contrairement aux mémoires de type ROM<sup>4</sup> et RAM<sup>5</sup> réputées pour être plus robustes.

Pour les technologies microcontrôleurs avancées intégrant de la MNV embarquée, un nouveau type de défaillance appelé « mismatch » et lié au non appariement des transistors des circuits analogiques est apparu. Ce phénomène, lié à des variations locales entre deux transistors MOS appariés peut générer une forte dispersion au niveau des réponses électriques des circuits analogiques.

Concernant la partie purement numérique des microcontrôleurs, l'augmentation de la densité d'intégration associée à une utilisation systématique des règles de dessin minimales introduit de nouveaux mécanismes de défaillance liés aux interconnexions métalliques. Ainsi, les microcontrôleurs sont devenus des circuits fortement sensibles à la qualité du procédé de fabrication, mais aussi aux méthodes de conception.

A tout cela s'ajoute un nombre réduit de broches externes qui est lié aux protocoles de communication série utilisés sur le microcontrôleur, qui réduisent l'accès aux éléments internes du circuit. Ainsi, afin d'analyser de manière détaillée la fiabilité des circuits microcontrôleurs plusieurs éléments, qui vont de structures de test très simples à des circuits de test complets, sont disponibles et vont être détaillés.

Des structures complexes appelées **véhicules de test** ou « **Test Chip** » sont utilisées pour l'étude de la fiabilité des MNV. Ces circuits de test présentent une architecture proche de celle des mémoires embarquées dans les produits. La différence réside dans leur capacité à embarquer des modes de test spécifiques afin de pouvoir déterminer les limites de fonctionnement du circuit dans tous ses modes opératoires. Il est ainsi possible d'obtenir une caractérisation précise du plan mémoire. De plus, la taille importante de ce type de circuit permet une étude statistique. Dans la pratique, ces circuits permettent un adressage indépendant de chaque cellule du plan mémoire de manière à extraire, par exemple, la distribution des **tensions de seuil**  $V_T$  du plan mémoire. Pour l'évaluation de la fiabilité des fonctions analogiques de base et notamment des fluctuations locales sur lesquelles se sont concentrés nos efforts, des mesures électriques sont réalisées sur des **TEG<sup>6</sup> paramétriques** constituées de paires de transistors MOS appariés [Laks86]<sup>7</sup>. Concernant la partie numérique, au vu des limites des TEG paramétriques, un autre type de TEG a été développé dans le but d'apporter un complément à ces dernières (et non de les remplacer). Ce sont les **TEG « logiques »** qui ont une architecture plus complexe que celle des TEG paramétriques. Ces TEG sont généralement constituées de portes logiques, d'oscillateurs en anneau ou encore de bascules. Elles sont optimisées pour le contrôle du procédé de fabrication et l'étude du BEOL<sup>8</sup>.

---

<sup>4</sup> Read Only Memory

<sup>5</sup> Read Access Memory

<sup>6</sup> Test Element Group

<sup>7</sup> K. R. Lakshmikummar, R. A. Hadaway, and M. A. Copeland, "Characterisation and modeling of mismatch in MOS transistors for precision analog design", IEEE Journal of Solid-State Circuits, vol. 21, pp. 1057-1066, 1986.

<sup>8</sup> Back-End Of Line

L'exploitation de toutes ces structures passe par la mise en œuvre de moyens de caractérisation électrique sur platines de test semi-automatiques ou automatiques mais aussi par l'utilisation de testeurs industriels dédiés. Cela est le cas pour les véhicules de test dédiés aux MNV avec l'extraction des distributions des tensions de seuil.

Le mémoire présenté pour l'obtention de L'Habilitation à Diriger des Recherches se compose de quatre chapitres dont une description succincte est donnée ci-dessous :

Le **premier chapitre** se compose de deux parties. Une première partie, consacrée aux circuits hétérogènes, replace l'étude de fiabilité présentée dans ce manuscrit par rapport aux produits existants, qui sont les SoC et plus particulièrement les microcontrôleurs. Ensuite, les notions de test de circuits et de rendement de fabrication sont développées. Une deuxième partie s'intéresse aux mécanismes de défaillance particuliers qui touchent la partie MNV embarquée, la partie analogique et la partie numérique des microcontrôleurs.

Le **deuxième chapitre** traite de la fiabilité des MNV. Le véhicule de test TCF9 est présenté avec une description du flot de test qui lui est associé en insistant de manière approfondie sur les tests de fiabilités (tests d'endurance et de rétention). Le TCF9 est ensuite utilisé pour mener à bien deux études. La première étude consiste à extraire la densité de pièges de l'oxyde tunnel. La deuxième étude a permis d'établir une corrélation entre les tests de stress électrique et les tests de rétention. Le véhicule de test, de par sa taille (nombre de cellules important) et ses fonctionnalités, s'est avéré un outil incontournable pour mener à bien ces deux expériences.

Dans le **troisième chapitre**, des solutions permettant de contrôler et d'accélérer les montées en rendement des technologies numériques CMOS sont proposées. Ces solutions sont basées sur des circuits numériques appelés TEG logiques, macro-cellules de test ou encore TEG hybrides. Un exemple pratique est ensuite présenté. Le développement et l'optimisation d'une macro-cellule de test, le RAPITAC, est proposé dans le but de fournir un outil d'analyse des variations du BEOL. L'architecture du RAPITAC ainsi que son programme de test sont décrits. Grâce à cette structure, une solution de caractérisation des variations du BEOL est mise en place.

Le **quatrième et dernier chapitre** est entièrement consacré à l'appariement des transistors MOS (« mismatch ») destinés aux applications analogiques faible consommation. Ce chapitre commence par un rappel sur les sources de fluctuations à l'origine des variations aléatoires des paramètres électriques des dispositifs. Ensuite l'effet « hump » est présenté. Cet effet a pour conséquence une dégradation de l'appariement des transistors MOS. Pour contrer cet effet parasite, plusieurs solutions sont proposées et validées expérimentalement à partir de TEG paramétriques. Dans une deuxième partie, une étude de fiabilité met en évidence une dégradation de l'appariement de la tension de seuil des transistors standards sous stress porteur chaud. Il est également démontré que les transistors « octogonaux », du fait de leur structure particulière, se dégradent moins que les transistors standards. Ainsi, une étude de fiabilité comparative est présentée entre les transistors octogonaux et les transistors standards.

Il est important de noter que l'ensemble de ce travail inclut les contributions de plusieurs thèses et stages de Master encadrés ou co-encadrés durant ces dernières années. De plus, tous ces travaux ont fait l'objet de nombreuses collaborations industrielles, notamment avec la société ST Microelectronics Rousset et Crolles.

# Table des symboles

Paramètre	Unité	Description
$\mu$		Mobilité des porteurs de charge dans la couche d'inversion
$\mu_0$		Mobilité des porteurs de charge sous faible champ électrique
$A_{cl}$	$mV.\mu m$	Paramètre d'appariement obtenu par régression linéaire (« méthode classique »)
$A_{rp}$	$mV.\mu m$	Paramètre d'appariement obtenu par régression pondérée
$A_{VT}$	$mV.\mu m$	Paramètre d'appariement de la tension de seuil $V_T$
$A_{VTc}$	$mV.\mu m$	Paramètre d'appariement de la tension de seuil $V_T$ lié au dopage du
$A_{VTc+g}$	$mV.\mu m$	Paramètre d'appariement de la tension de seuil $V_T$ lié au dopage du canal et de la grille
$A_{VTc+p}$	$mV.\mu m$	Paramètre d'appariement de la tension de seuil $V_T$ lié au dopage du canal et des poches
$C_{ox}$	F	Capacité de l'oxyde de grille
$C_{sc}$	F	Capacité de la couche de désertion dans le semi-conducteur
$E_C$	eV	Energie du niveau le plus bas de la bande de conduction du silicium
$E_F$	eV	Energie du niveau de Fermi dans le silicium
$E_i$	eV	Niveau d'énergie intrinsèque du silicium
$E_V$	eV	Energie du niveau le plus haut de la bande de conduction du silicium
$g_m$	A/V	Transconductance du transistor MOS
$I_B$	A	Courant substrat
$I_D$	A	Courant de drain du transistor MOS
$K$		Nombre de transistors bipolaires en parallèle
$L$	$\mu m$	Longueur du transistor MOS
$L_p$	$\mu m$	Longueur des poches
$m$		Nombre de transistors MOS identiques en parallèle
$M_x$ ou $N_x$		Dénomination d'un transistor MOS
$m_{\Delta P}$		Moyenne de la population des $\Delta P$
$N$		Rapport de courant dans un miroir de courant
$N_C$	atomes.cm <sup>-3</sup>	Concentration de dopants dans le semi-conducteur
$N_G$	atomes.cm <sup>-3</sup>	Concentration de dopants dans la grille
$N_{géo}$		Nombre de géométries différentes testées
$n_i$	m <sup>-3</sup>	Concentration intrinsèque d'électrons dans le silicium
$N_P$	atomes.cm <sup>-3</sup>	Concentration de dopants dans les poches
$N_{paires}$		Nombre de sites mesurés
$N_{S/D}$	atomes.cm <sup>-3</sup>	Concentration de dopants dans les zones de Source et Drain (S/D)
$Q_{it}$	C.m <sup>-2</sup>	Charge due aux états d'interface
$Q_{ox}$	C.m <sup>-2</sup>	Charge fixe équivalente à l'interface Si/SiO <sub>2</sub>
$Q_{sc}$	C.m <sup>-2</sup>	Charge de la zone de désertion dans le semi-conducteur
$Q_{scp}$	C.m <sup>-2</sup>	Charge de la zone de désertion dans les poches

$Q_X$		Dénomination d'un transistor bipolaire
$R_X$	$\Omega$	Dénomination et valeur d'une résistance
$S$	mV/dec	Pente sous le seuil
$T_{ox}$	$\text{\AA}$	Epaisseur de l'oxyde de grille
$U_T=kT/q$	V	Tension thermodynamique
$V_B$	V	Tension de polarisation du substrat (Effet « Bulk »)
$V_{BE}$	V	Tension Base-Emetteur d'un transistor bipolaire
$V_{BG}$	V	Tension de sortie d'une référence de tension « bandgap »
$V_D$	V	Tension de polarisation du drain
$V_{FB}$	V	Flat Band Voltage (Tension de Bandes Plates)
$V_G$	V	Tension de polarisation de la grille
$V_{GB}$	V	Différence de potentiels grille-substrat
$V_{offset}$	V	Tension de décalage en entrée d'une paire différentielle
$V_{ox}$	V	Chute de potentiel aux bornes de l'oxyde de grille du transistor MOS
$V_S$	V	Tension de polarisation de la source
$V_T$	V	Tension de seuil du transistor MOS
$W$	$\mu\text{m}$	Largeur du transistor MOS
$y_G$	nm	Profondeur de la couche de désertion dans la grille
$y_{SC}$	nm	Profondeur de la couche de désertion dans le semi-conducteur
$\beta$		Facteur de courant du transistor MOS
$\delta P$		Décalage dans le temps d'un paramètre P
$\Delta P$		Différence d'un paramètre P entre deux transistors MOS
$\epsilon_{ox}$	$\text{F.m}^{-1}$	Permittivité électrique de l'oxyde de grille
$\epsilon_{Si}$	$\text{F.m}^{-1}$	Permittivité électrique du silicium
$\eta$		Facteur de pente
$\theta_1$		Premier facteur de réduction de la mobilité
$\theta_2$		Deuxième facteur de réduction de la mobilité
$\sigma_{locales}$		Déviations standard liées aux fluctuations locales des transistors MOS uniquement
$\sigma_{mesurées}$		Déviations standard mesurées sur structure de test
$\sigma_{système}$		Déviations standard liées au système de mesure uniquement
$\sigma_{\Delta P}$		Déviations standard (écart-type) de la population des $\Delta P$
$\Phi_F$	V	Potentiel de volume dans le semi-conducteur
$\Phi_{Fg}$	V	Potentiel de volume dans la grille
$\Phi_{ms}$	V	Différence des travaux de sortie aux bornes de la structure MOS
$\Psi_G$	V	Potentiel de surface dans la grille
$\Psi_S$	V	Potentiel de surface dans le semi-conducteur



# Table des matières

## ***Chapitre I : Mémoire embarquée en technologie CMOS***

<b>I. SYSTEMES INTEGRES SUR PUCE ET MEMOIRE EMBARQUEE...</b>	<b>14</b>
A. SYSTEMES SUR PUCE .....	14
1 Présentation .....	14
2 Architecture.....	15
3 Interconnexions.....	17
4 Fiabilité des systèmes sur puce .....	18
5 Logiciel embarqué .....	19
B. CIRCUITS MICROCONTROLEURS .....	19
1 Présentation .....	19
2 Les unités de traitement de données analogiques .....	21
3 Les unités de traitement de données numériques.....	21
4 Mémoires Non Volatiles embarquées de type Flash et EEPROM .....	21
5 Circuiterie périphérique des Mémoires Non Volatiles.....	26
C. NIVEAUX DE TEST ET RENDEMENT DE FABRICATION.....	28
1 Placement des structures de test.....	28
2 Rendement de fabrication .....	29
3 Facteurs limitatifs du rendement.....	30
4 Outils de mesure et d'analyse des pertes de rendement.....	31
5 L'analyse de défaillance .....	33
<b>II. FIABILITE DES MNV EMBARQUEES .....</b>	<b>37</b>
A. PRINCIPAUX CRITERES DE FIABILITE.....	37
1 Caractéristiques en endurance .....	37
2 Caractéristiques en rétention .....	38
3 Extraction des distributions de $V_T$ .....	38
B. EVALUATION DE LA FIABILITE DES MNV .....	39
1 Les véhicules de test (« Test Chip ») .....	39
2 Le test sous pointe (« Probe Test »).....	40
<b>III. FIABILITE DE LA CIRCUITERIE LOGIQUE.....</b>	<b>41</b>
A. TECHNOLOGIE CMOS ET BEOL .....	41
B. EVALUATION DE LA FIABILITE DES CIRCUITS NUMERIQUES .....	42
1 Les TEG logiques .....	43
2 Les Macro-cellules de test.....	43
3 Les TEG hybrides.....	45
<b>IV. FIABILITE DES BLOCS ANALOGIQUE .....</b>	<b>47</b>
A. RAPPELS SUR LA CONCEPTION ANALOGIQUE .....	47
1 Les blocs de base .....	47
2 Les blocs fonctionnels.....	47
3 Variabilité inter dispositifs (« matching ») .....	48
B. FIABILITE DES BLOCS ANALOGIQUES.....	48
1 Test des blocs analogiques .....	48
2 Utilisation des TEG pour l'analyse du « matching » .....	49
<b>V. CONCLUSION .....</b>	<b>49</b>

## ***Chapitre II : Fiabilité des Mémoires Non Volatiles***

<b>I.</b>	<b>LE VEHICULE DE TEST TCF9</b>	<b>54</b>
<b>A.</b>	<b>PRESENTATION</b>	<b>54</b>
1	Description générale	54
2	Architecture	55
3	Modes de fonctionnement	55
<b>B.</b>	<b>LE FLOT DE TEST TCF9</b>	<b>56</b>
1	Présentation	56
2	Analyse des résultats de Test	59
3	Fiabilité des oxydes : résultats de tests d'endurance et de rétention	61
<b>II.</b>	<b>UTILISATION DU TCF9 POUR L'EXTRACTION DE LA DENSITE DE PIEGES DE L'OXYDE TUNNEL</b>	<b>63</b>
<b>A.</b>	<b>LE SILC : PRESENTATION</b>	<b>63</b>
1	Origine physique du SILC	63
2	Modélisation du SILC	64
3	Méthodes de caractérisation du SILC	65
<b>B.</b>	<b>EXTRACTION DE LA DENSITE DE PIEGES SUR VEHICULE DE TEST</b>	<b>67</b>
1	Dispositif expérimental	67
2	Résultats expérimentaux	68
3	Méthode d'extraction de la densité de piège des cellules EEPROM	69
<b>III.</b>	<b>UTILISATION DU TCF9 POUR L'ETUDE DE LA CORRELATION ENTRE TESTS DE STRESS ET TESTS DE RETENTION</b>	<b>71</b>
<b>A.</b>	<b>CORRELATION ENTRE TESTS DE STRESS ET TESTS DE RETENTION</b>	<b>71</b>
1	Enjeux du stress électrique sur le temps de test	71
2	Dispositif expérimental	72
3	Modèle de corrélation entre temps de stress et temps de rétention	73
<b>B.</b>	<b>COMPARAISON DES MECANISMES DE PERTES A CHAMP FORT ET FAIBLE</b>	<b>75</b>
<b>IV.</b>	<b>CONCLUSION</b>	<b>76</b>

## ***Chapitre III : Fiabilité des circuits logiques***

<b>I.</b>	<b>STRUCTURES DE TEST NUMERIQUES : TEG « LOGIQUES »</b>	<b>80</b>
<b>A.</b>	<b>LE TAC8M</b>	<b>80</b>
1	La structure du TAC8M	80
2	Limitations du TAC8M	83
<b>B.</b>	<b>LE TEGTACF9</b>	<b>84</b>
1	Conception du TEGTACF9	84
2	Limitation du TEGTACF9	86
<b>II.</b>	<b>APPLICATION A L'ANALYSE DES VARIATIONS DU BEOL</b>	<b>86</b>
<b>A.</b>	<b>INTRODUCTION</b>	<b>86</b>
<b>B.</b>	<b>DESCRIPTION DU RAPITAC</b>	<b>87</b>
<b>C.</b>	<b>TEST DU RAPITAC ET TRAITEMENT DES DONNEES</b>	<b>90</b>
1	Déroulement du flot de test	90
2	Traitement des données issues du test	91
<b>D.</b>	<b>CARACTERISATION DES VARIATIONS DU PROCEDE DE FABRICATION</b>	<b>95</b>
<b>III.</b>	<b>CONCLUSION</b>	<b>99</b>

## ***Chapitre IV : Fiabilité des circuits analogiques***

<b>I. FLUCTUATIONS LOCALES DANS LES TRANSISTORS MOS.....</b>	<b>102</b>
<b>A. FLUCTUATION DANS LE TRANSISTOR MOS .....</b>	<b>102</b>
1 Présentation .....	102
2 Sources de fluctuations locales liées au procédé de fabrication .....	103
<b>B. MESURE DES FLUCTUATIONS LOCALES DANS LE TRANSISTOR MOS .....</b>	<b>107</b>
1 Structures de mesure .....	107
2 Méthodes de mesure de l'appariement .....	108
3 Traitement statistique des résultats d'appariement.....	109
4 Extraction du paramètre d'appariement .....	110
<b>II. CONTRIBUTION A L'AMELIORATION DE L'APPARIEMENT DES CIRCUITS ANALOGIQUES POLARISES SOUS LE SEUIL .....</b>	<b>112</b>
<b>A. CONCEPTION DE CIRCUITS ANALOGIQUES FAIBLE CONSOMMATION.....</b>	<b>112</b>
<b>B. L'EFFET « HUMP ».....</b>	<b>113</b>
1 Présentation .....	113
2 Modélisation de l'effet hump.....	115
3 Impact de l'effet « hump » sur l'appariement en tension.....	115
4 Etude dimensionnelle.....	116
5 Etude en fonction de l'effet substrat.....	117
<b>C. SOLUTIONS POUR CONTRER L'EFFET « HUMP ».....</b>	<b>119</b>
1 Solution basée sur des transistors « multi-doigts » .....	119
2 Autres solutions .....	121
<b>III. CONTRIBUTION A L'AMELIORATION DE L'APPARIEMENT EN FONCTION DU STRESS PORTEURS CHAUDS.....</b>	<b>122</b>
<b>A. PRESENTATION DES DIFFERENTS STRESS ELECTRIQUES.....</b>	<b>122</b>
1 Principe du stress par injection de porteurs chauds (HCI).....	122
2 Protocole expérimental.....	122
<b>B. DEGRADATION DES TRANSISTORS NMOS ET PMOS SOUS STRESS HCI .</b>	<b>123</b>
1 Dégradation de la tension de seuil $V_T$ .....	123
2 Dégradation de l'appariement de la tension de seuil $\Delta V_T$ .....	123
<b>C. ETUDE DES TRANSISTORS OCTOGONAUX .....</b>	<b>124</b>
1 Présentation des transistors octogonaux.....	124
2 Comparaison entre transistors standard et octogonal .....	125
3 Inversion des sources et drains des transistors octogonaux .....	127
<b>IV. CONCLUSION.....</b>	<b>129</b>
CONCLUSION GENERALE.....	131
PERSPECTIVES .....	133
CURRICULUM VIAE.....	135
LISTE DES TRAVAUX.....	136
ACTIVITES D'ENCADREMENT .....	142
RAYONNEMENT INTERNATIONAL .....	145
RECAPITULATIF DES ACTIVITES D'ENSEIGNEMENTS .....	147
REFERENCES.....	149





# Chapitre I : Mémoire embarquée en technologie CMOS

<b>I. SYSTEMES INTEGRES SUR PUCE ET MEMOIRE EMBARQUEE ...</b>	<b>14</b>
A. SYSTEMES SUR PUCE .....	14
1 Présentation .....	14
2 Architecture .....	15
3 Interconnexions .....	17
4 Fiabilité des systèmes sur puce.....	18
5 Logiciel embarqué .....	19
B. CIRCUITS MICROCONTROLEURS .....	19
1 Présentation .....	19
2 Les unités de traitement de données analogiques.....	21
3 Les unités de traitement de données numériques.....	21
4 Mémoires Non Volatiles embarquées de type Flash et EEPROM .....	21
a. <i>Technologie à grille flottante</i> .....	22
b. <i>La mémoire EEPROM</i> .....	23
c. <i>La mémoire Flash EEPROM</i> .....	24
5 Circuiterie périphérique des Mémoires Non Volatiles .....	26
C. NIVEAUX DE TEST ET RENDEMENT DE FABRICATION.....	28
1 Placement des structures de test .....	28
2 Rendement de fabrication .....	29
3 Facteurs limitatifs du rendement .....	30
4 Outils de mesure et d'analyse des pertes de rendement .....	31
a. <i>Mesure de la défektivité en ligne</i> .....	31
b. <i>La caractérisation</i> .....	32
5 L'analyse de défaillance .....	33
a. <i>Les techniques de localisation</i> .....	33
b. <i>Les analyses physico-chimiques</i> .....	35
<b>II. FIABILITE DES MNV EMBARQUEES.....</b>	<b>37</b>
A. PRINCIPAUX CRITERES DE FIABILITE.....	37
1 Caractéristiques en endurance .....	37
2 Caractéristiques en rétention .....	38
3 Extraction des distributions de $V_T$ .....	38
B. EVALUATION DE LA FIABILITE DES MNV .....	39
1 Les véhicules de test (« Test Chip ») .....	39
2 Le test sous pointe (« Probe Test ») .....	40
<b>III. FIABILITE DE LA CIRCUITERIE LOGIQUE.....</b>	<b>41</b>
A. TECHNOLOGIE CMOS ET BEOL .....	41
B. EVALUATION DE LA FIABILITE DES CIRCUITS NUMERIQUES .....	42
1 Les TEG logiques.....	43
2 Les Macro-cellules de test .....	43
3 Les TEG hybrides.....	45
<b>IV. FIABILITE DES BLOCS ANALOGIQUE.....</b>	<b>47</b>
A. RAPPELS SUR LA CONCEPTION ANALOGIQUE .....	47
1 Les blocs de base.....	47
2 Les blocs fonctionnels .....	47
3 Variabilité inter dispositifs (« matching »).....	48
B. FIABILITE DES BLOCS ANALOGIQUES .....	48
1 Test des blocs analogiques .....	48
2 Utilisation des TEG pour l'analyse du « matching ».....	49
<b>V. CONCLUSION .....</b>	<b>49</b>

# I. Systèmes intégrés sur puce et mémoire embarquée

## A. Systèmes sur puce

### 1 Présentation

Les SoC<sup>9</sup> sont destinés à diminuer les coûts de production des systèmes tout en augmentant leur performance. Le principe est l'intégration sur un même substrat d'éléments hétérogènes. Afin de tirer parti du potentiel des systèmes sur puce, il est nécessaire de mettre en œuvre des méthodologies de test et d'analyse de fiabilité de chaque sous-système constitutif du SoC. Ces sous-systèmes, qui réalisent souvent des fonctions complexes, sont composés d'éléments fonctionnels (aussi appelés IP<sup>10</sup>), eux-mêmes composés de plusieurs millions de transistors. L'accroissement de la densité d'intégration que permettent les technologies de fabrication actuelles pose de nouveaux défis du point de vue de la fiabilité. Depuis la projection de Moore sur les tendances technologiques des semi-conducteurs en 1965 [Moo65] et des règles de réduction de taille des transistors MOS fixées par Dennard en 1974 [Denn74], la densité d'intégration a connu une croissance exponentielle. Cette tendance consiste à doubler le nombre de transistors par unité de surface tous les 2 ans, ce qui correspond à une réduction de la taille minimale de la technologie d'un facteur 0,7 pour les technologies CMOS [Boh09], comme le montre la Fig. I-1. Ainsi, tous les 2 à 3 ans, un nouveau nœud est annoncé pour les technologies CMOS.

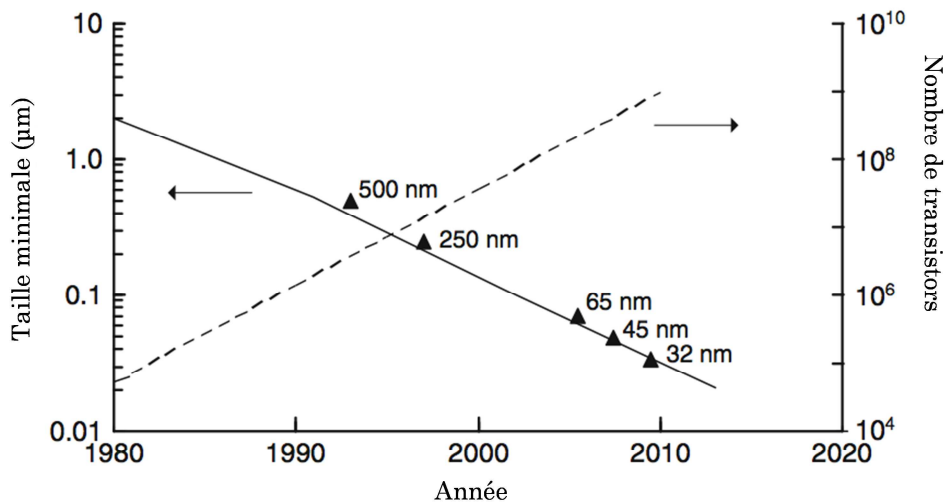


Fig. I-1 : Tendence des évolutions des nœuds technologiques et du nombre de transistors par unité de surface pour les technologies 500, 260, 65 et 32 nm.

A cela s'ajoute la nécessité de gérer la complexité non seulement en termes de nombre de composants présents dans les systèmes, mais aussi en termes de

<sup>9</sup> System On Chip

<sup>10</sup> Intellectual Property

diversité des composants dans la mesure où les systèmes actuels intègrent de plus en plus de blocs fonctionnels hétérogènes. De plus, l'augmentation de la densité d'intégration impacte les communications entre les blocs fonctionnels d'un SoC qui constituent aujourd'hui une part importante de la conception d'un système sur puce.

Ce changement d'échelle ne peut pas se faire au détriment de la qualité du produit fini. Un circuit, qu'il soit très simple ou très complexe, doit être réalisé conformément à ses spécifications pour être utilisable. L'étape de test est par conséquent une étape essentielle dans le processus de fabrication et de mise sur le marché de ces nouveaux produits.

La Fig. I-2 illustre l'évolution suivie par les systèmes électroniques : un système complexe, auparavant réalisé par l'association de blocs fonctionnels distincts, peut être intégré de façon monolithique.

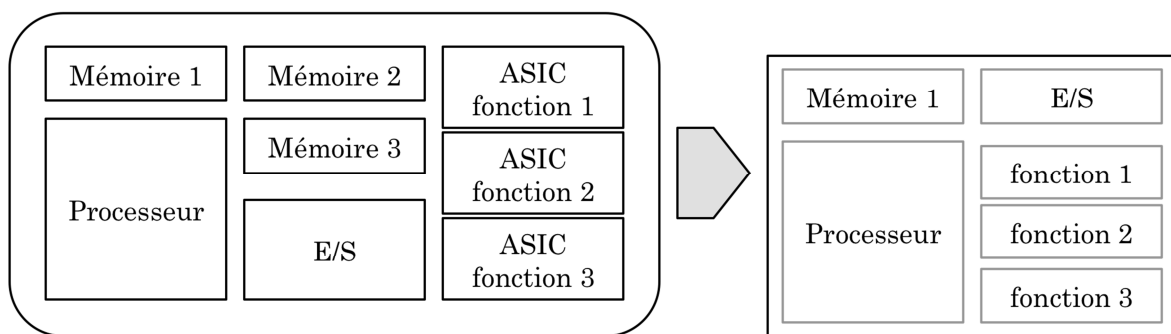


Fig. I-2 : Du circuit intégré vers le SoC

## 2 Architecture

L'architecture générale d'un système sur puce est présentée Fig. I-3. De manière schématique, il existe quatre principaux types de composants dans les SoC : les unités de traitements de données (microcontrôleurs et DSP<sup>11</sup>), les mémoires, les blocs analogiques et les interfaces de communication (filaires ou RF<sup>12</sup>).

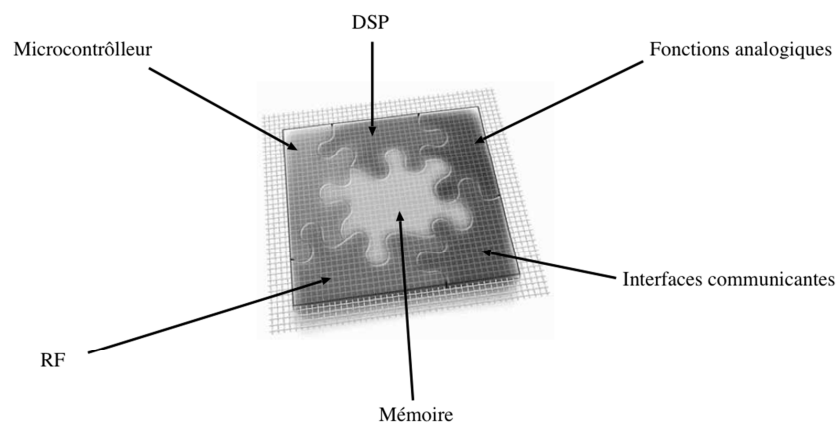


Fig. I-3 : Briques élémentaires d'un SOC

<sup>11</sup> Digital Signal Processor

<sup>12</sup> RadioFréquence



---

A partir des circuits logiques CMOS<sup>13</sup> de base, constitutifs de la partie numérique dédiée aux processeurs et DSP, les fabricants de semi-conducteurs ont progressivement ajoutés de la mémoire embarquée, des dispositifs mixte Analogique/Numérique, des dispositifs RF, voire des MEMS<sup>14</sup>.

Les éléments de traitement de données numériques sont des processeurs dédiés à l'exécution de logiciels embarqués ou des accélérateurs matériels spécifiques. Il existe deux catégories de cœur de traitement : les microcontrôleurs qui assurent l'exécution des instructions et le contrôle des périphériques, et les processeurs de signal numérique DSP optimisés pour la résolution d'algorithmes spécifiques au traitement du signal. Des efforts de développement aussi bien au niveau académique qu'industriel ont vu émerger un nouveau type de processeurs : les ASIP<sup>15</sup>, qui permettent une extension dynamique du jeu d'instructions par l'utilisateur en fonction des applications ciblées. Lorsqu'une approche logicielle ne peut pas permettre d'atteindre les performances requises pour une application donnée, les concepteurs de SoC se tournent vers des coprocesseurs dédiés. Ces coprocesseurs peuvent prendre en charge la gestion des codecs d'images JPEG<sup>16</sup>, des codecs vidéo MPEG-4<sup>17</sup>, des algorithmes de type Viterbi<sup>18</sup>, des moteurs de cryptage/décryptage, etc.

De nombreuses applications exigent des fonctions analogiques pour l'interfaçage avec le monde physique (biomédical, biométrie, automobile, authentification, applications sans fil et filaires, communications optiques, récupération d'énergie, multimédia). La prolifération de ces applications a ainsi conduit à une augmentation continue de la quantité et la complexité des circuits analogiques dans les SoC.

De manière à communiquer avec l'extérieur, le SoC propose généralement plusieurs types d'E/S standard. On peut trouver des IP prenant en charge les connexions Ethernet, USB<sup>19</sup>, des liaisons à haut débit série, des sorties audio/vidéo, des contrôleurs de mémoire, etc.

Un SoC classique utilise des centaines de blocs de mémoire interne. Il peut s'agir de blocs SRAM<sup>20</sup>, OPT<sup>21</sup>, ROM<sup>22</sup>, Flash ou EEPROM<sup>23</sup>. Leur configuration en termes de nombre de mots par page, de longueur de mots, de nombre de ports d'accès en lecture/écriture, et de vitesse d'accès est choisie de manière à répondre aux spécificités de l'application. Quand un bloc de mémoire a une taille trop importante pour être réalisé sur la puce, il est intégré en tant que mémoire externe à la puce à l'aide d'une solution basée sur un système en boîtier (SiP<sup>24</sup>). Ainsi, la majorité des SoC inclut un contrôleur de mémoire externe SDRAM<sup>25</sup>,

---

<sup>13</sup> Complementary Metal-Oxide Semiconductor

<sup>14</sup> Micro Electro-Mechanical Systems

<sup>15</sup> Application-Specific Instruction-set Processor

<sup>16</sup> Joint Photographic Experts Group

<sup>17</sup> Motion Picture Experts Group layer-4 video

<sup>18</sup> Permet de corriger les erreurs survenues lors d'une transmission à travers un canal bruité

<sup>19</sup> Universal Serial Bus

<sup>20</sup> Static Random-Access Memory

<sup>21</sup> One-Time Programmable

<sup>22</sup> Read Only Memory

<sup>23</sup> Electrically Erasable Programmable Read Only Memory

<sup>24</sup> System in Package

<sup>25</sup> Synchronous Dynamic RAM

Carte SD<sup>26</sup>, MMC<sup>27</sup> et Compact Flash, etc.

Un des problèmes principaux que rencontrent les concepteurs de SoC est la réalisation du réseau d'interconnexions qui permet aux blocs fonctionnels de communiquer entre eux. Le débit de données global à l'intérieur d'un SoC peut atteindre plusieurs dizaines de Gbit/s. Cette contrainte ne peut être atteinte que par la mise en œuvre de solutions architecturales performantes. En général, ce sont les interconnexions qui limitent les performances globales du système.

La communauté SoC a proposé plusieurs standards dont le plus populaire est la structure de communication avancée basée sur le bus AMBA<sup>28</sup> proposé par la société ARM<sup>29</sup>. Ce bus se compose d'un bus rapide (« Advanced High Bus » pour AHB) et d'un bus lent (« Advanced Peripheral Bus » pour APB). Le bus AHB permet de connecter des périphériques rapides comme les contrôleurs de mémoires SDRAM, DMA<sup>30</sup> et vidéo. Le bus APB permet quant à lui de s'interfacer avec des périphériques lents comme un clavier.

Au vu de l'évolution croissante de la quantité de données échangées au sein de la puce et de la complexité des composants, les performances d'un système sur puce sont de plus en plus dépendantes de celles du système d'interconnexion. Ainsi, les réseaux sur puce NoC<sup>31</sup> semblent être une solution appropriée. Ces réseaux reproduisent des concepts architecturaux et topologiques retenus dans le domaine des systèmes distribués et des réseaux d'ordinateurs pour établir des communications sur puce [Bert04][Ogra05].

### 3 Interconnexions

L'ITRS<sup>32</sup> indique que les technologies les plus avancées peuvent compter jusqu'à douze "niveaux" de métallisation comme le montre le Tableau I-1. La réduction du paramètre technologique conduit à une augmentation des éléments parasites (résistance des interconnexions) et donc à une augmentation du temps de transport de l'information. Comme par ailleurs le débit global à l'intérieur d'un SoC continue à augmenter, le système de communication constitue le plus souvent un goulot d'étranglement dans l'augmentation des performances des systèmes.

Pour l'ITRS les interconnexions font parties des "grand challenges" à résoudre à court terme pour poursuivre l'augmentation des performances des systèmes intégrés. D'autres défis liés à la consommation globale du système apparaissent comme un problème. La dispersion sur les paramètres technologiques n'est pas à négliger, notamment pour les blocs analogiques. En effet, cette variabilité rend difficile la prédictibilité des performances globales du système.

Les systèmes sur puces sont généralement évalués selon une grande variété de critères tels que la performance, le coût, la flexibilité, ainsi que la consommation d'énergie, leur taille et leur poids pour le cas des applications embarquées.

Ces spécifications sont très souvent contradictoires et il n'y a généralement pas

---

<sup>26</sup> Secure Digital

<sup>27</sup> MultiMediaCard

<sup>28</sup> Advanced Microcontroller Bus Architecture

<sup>29</sup> Advanced Risc Machine

<sup>30</sup> Direct Memory Access

<sup>31</sup> Network on Chip NoC

<sup>32</sup> International Technology Roadmap for Semiconductors

de conception unique optimale, mais une variété de choix. En conséquence, le concepteur du système n'est pas seulement intéressé par un choix de mise en œuvre unique, mais par un ensemble bien choisi de solutions optimales.

Tableau I-1 : Evolution de quelques paramètres importants des circuits intégrés

	2006	2007	2008	2009	2010	2012	2015	2018
Pitch local (nm)	170	152	134	120	120	84	60	42
Taille des puces à la production (mm <sup>2</sup> )	310	310	310	310	310	310	310	310
Longueur totale de l'interconnexion (m/cm <sup>2</sup> )	1002	1117	1401	1559	1784	2214	3544	5035
Horloge locale intégrée (GHz)	6,783	9,285	10,972	12,369	15,079	20,065	33,403	53,207
Nombre de niveaux métalliques	11	11	12	12	12	12	13	14
Puissance Maximale (W)	180	189	200	210	218	240	270	300

#### 4 Fiabilité des systèmes sur puce

Au niveau conception, les SoC sont généralement réalisés à partir de composants virtuels propriétaires ou IP, intégrés de manière à créer l'ensemble du système. Ces circuits de base sont complexes, préconçus et pré vérifiés. Ils sont généralement fournis par des sociétés spécialisées et réutilisés pour la conception des SoC. Cependant, l'importation et l'assemblage d'IP provenant de sources différentes entraînent souvent une perte de performance aussi bien au niveau de l'IP élémentaire qu'au niveau de l'ensemble du système. En effet, il reste plusieurs obstacles à une intégration rapide et efficace des IP au niveau système. Le plus important est le manque d'interfaces entre les circuits propriétaires, interfaces à partir desquelles le développement de tout le système pourrait être basé. Ainsi, pour remédier à ce problème, un certain nombre de groupe de travail ont été formés. Parmi eux, on peut citer le groupe VSIA<sup>33</sup> [Visa] et le groupe IEEE P1500 [P1500].

VSIA a été formé en Septembre 1996 dans le but d'unifier l'industrie du SoC en établissant des normes techniques facilitant l'intégration au niveau système. Des normes d'interface standard sont mises en places et permettent aux IP de s'adapter rapidement tant au niveau architectural qu'au niveau physique. Cela a permis aux fournisseurs d'IP de proposer des interfaces de communication standard compatibles entre elles.

Le groupe de travail IEEE P1500 a été créé pour élaborer un projet de norme de test pour la logique numérique et la mémoire embarquée. Les activités de ce groupe comprennent le développement d'un flot de test d'IP normalisé, une interface externe (« wrapper ») permettant l'amélioration de la testabilité et des lignes directrices concernant la fusion d'IP.

Le test des SoC est essentiellement un test composite constitué des tests individuels de chaque IP. Le développement d'un flot de test orienté SoC est particulièrement difficile, et cela pour plusieurs raisons.

En effet, les IP sont des composants propriétaires et les fournisseurs sont souvent réticents à divulguer des informations structurelles sur les produits qu'ils fournissent. Ainsi, les utilisateurs ne peuvent pas accéder directement aux

<sup>33</sup> Virtual Socket Interface Alliance

fichiers de descriptions comportementaux ou aux « netlists » de base de manière à y insérer des circuits additionnels permettant d'améliorer la testabilité. Au lieu de cela, un ensemble de modèles et de vecteurs de test est fourni par le vendeur d'IP qui garantit une couverture de fautes optimale. Ces vecteurs de test doivent être appliqués aux IP dans un ordre bien précis et à une fréquence d'horloge spécifique. D'autre part, les IP ne sont pas toujours directement accessibles à partir des E/S<sup>34</sup> de la puce. La propagation des stimuli de test nécessite donc des mécanismes de propagation dédiés.

## 5 Logiciel embarqué

Une implantation logicielle d'une partie d'une application a l'avantage de lui procurer une flexibilité liée à la possibilité de reprogrammation, contrairement à une implantation matérielle figée mais qui a l'avantage de satisfaire plus facilement les contraintes de performances.

Un logiciel embarqué se décline généralement en trois couches :

- La couche la plus basse est la couche d'abstraction matérielle HAL<sup>35</sup> qui permet de faire le lien avec le matériel en implémentant les pilotes d'E/S des périphériques, les contrôleurs de composants et les routines d'interruption ISR<sup>36</sup>,
- La couche système d'exploitation qui permet de porter l'application sur l'architecture en assurant la liaison entre les ressources matérielles, l'utilisateur et les applications,
- La couche application qui permet l'exécution des applications embarquées.

Un facteur important dans l'évolution de ces systèmes est l'apparition de nouvelles architectures ayant la flexibilité du logiciel et la performance du matériel. Ces architectures sont basées sur la programmation de circuits matériels tels que les FPGA<sup>37</sup>. Les propriétés attrayantes des FPGA ainsi que le rythme soutenu des progrès qu'a connus cette technologie (au niveau taille, intégration et performances) ont changé le rôle jusque-là joué par ces composants. Les FPGA sont passés du rôle d'outils de prototypage d'ASIC<sup>38</sup> ou de « glue logic » au rôle d'unités de calcul alternatives. On parle de plus en plus de systèmes (ou plateformes) reconfigurables qui intègrent sur un même substrat un ou plusieurs cœurs de processeurs et une matrice programmable.

## B. Circuits microcontrôleurs

### 1 Présentation

Un microcontrôleur, présenté Fig. I-4, est un composant réunissant sur un seul et même silicium une unité de calcul (cœur de processeur ou blocs de propriété

---

<sup>34</sup> Entrées/Sorties

<sup>35</sup> Hardware Abstraction Layer

<sup>36</sup> Interrupt Service Routine

<sup>37</sup> Field Programmable Gate Array

<sup>38</sup> Application Specific Integrated Circuit

intellectuelle numériques), divers dispositifs d'entrées/sorties et de contrôle d'interruptions ainsi que de la mémoire volatile et non volatile, notamment pour stocker le programme d'application. Le microcontrôleur embarque également un certain nombre de périphériques spécifiques (bus série, « Timers », convertisseur analogique numérique, capteurs, cœurs de processeurs cryptographiques, etc.). Il existe plusieurs familles de microcontrôleurs, se différenciant par la vitesse de leur processeur et par le nombre de périphériques qui les composent.

Les dernières générations de microcontrôleur se caractérisent par de nombreuses fonctionnalités conçues pour assurer la sécurité des données et le fonctionnement sûr du système, dont la réinitialisation flexible en cas de baisse anormale de la tension d'alimentation, de la mémoire Flash embarquée avec support de l'ECC<sup>39</sup>, une unité de protection de la mémoire MPU<sup>40</sup> et fusible JTAG<sup>41</sup> (norme portant sur les architectures pour le test et point d'accès de test). Dans de nombreux cas, une intégration du support au standard USB 2.0 « Full Speed » assure la compatibilité immédiate avec les combinés portables. A cela s'ajoute de nombreuses fonctionnalités qui permettent aux développeurs d'optimiser la consommation d'énergie de leurs applications. Ainsi, jusqu'à dix modes ultra-basse consommation permettent au dispositif d'exécuter les tâches qui lui sont confiées en consommant à tout moment le minimum d'énergie.

En termes de technologie de base, toutes ces fonctionnalités impliquent qu'on soit capable d'intégrer ensemble les options suivantes :

- Technologie **CMOS** avancée pour le numérique,
- Blocs **analogiques**, notamment pour la gestion de la puissance,
- Mémoires **FLASH/EEPROM** pour le stockage non volatile.

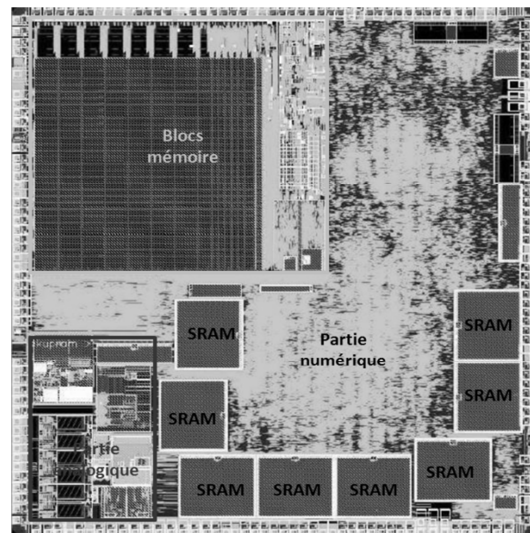


Fig. I-4 : Vue Layout<sup>42</sup> des principaux blocs composant un microcontrôleur

<sup>39</sup> Error Correction Code

<sup>40</sup> Memory Protection Unit

<sup>41</sup> Joint Test Action Group

<sup>42</sup> Dessin de masques

## 2 Les unités de traitement de données analogiques

La nécessité d'échanger des signaux avec le monde extérieur (traitement des signaux issus de capteurs, commande d'actionneurs) fait appel aux blocs analogiques du microcontrôleur. Ces signaux extérieurs doivent être amplifiés, filtrés, convertis d'une représentation analogique en une représentation numérique et vice-versa.

Il est donc nécessaire de savoir intégrer sur la même puce des blocs numériques et des blocs analogiques : on parle alors de technologie mixte. Aussi, la majorité des microcontrôleurs actuellement fabriqués sont-ils des circuits mixtes analogique/numérique, dans lesquels les fonctions analogiques sont de plus en plus fréquemment réalisées par des « transistors numériques ».

Au niveau des circuits de base, la plupart de ces blocs analogiques sont souvent définis autour de structures élémentaires telles que les miroirs de courant ou les paires différentielles. Un miroir de courant, composé de transistors appariés, a pour fonction de fournir un courant identique au courant de référence, dans une ou plusieurs branches de sortie. La paire différentielle sert souvent d'étage d'entrée dans les amplificateurs différentiels afin d'amplifier la différence des signaux d'entrée appliqués sur les grilles des transistors.

Dans la partie analogique, se trouve un circuit de gestion d'alimentation (« Power Management »). Cet élément analogique a une fonction primordiale puisqu'il fournit l'alimentation de nombreux circuits du microcontrôleur. Ainsi, si la tension fournie n'est pas suffisamment stable en température à cause de fluctuations trop importantes (notamment celles liées à l'appariement des dispositifs dans les miroirs de courant ou paires différentielles), cela peut entraîner la perte des données stockées en mémoire voire même le dysfonctionnement du microcontrôleur.

## 3 Les unités de traitement de données numériques

Un système numérique (ou processeur numérique) est un système logique qui opère sur des ensembles de bits appelés mots binaires. Ce système se décompose généralement en deux parties : une unité de traitement de données et une unité de commande. La réalisation de ces unités de traitement résulte de l'interconnexion de composants combinatoires et séquentiels. Le processeur (noté CPU, pour Central Processing Unit) est cadencé au rythme d'une horloge interne lors de l'exécution des instructions stockées en mémoire.

Les microcontrôleurs de dernière génération sont architecturés autour de cœurs de processeurs standards embarqués (notamment fournis par la société ARM) auxquels est associé de la MNV<sup>43</sup> de type Flash. Ces cœurs de processeurs ultra-basse consommation sont très présents dans les systèmes embarqués (téléphone mobile, console portable, tablette électronique).

## 4 Mémoires Non Volatiles embarquées de type Flash et EEPROM

Les mémoires peuvent se présenter sous une forme isolée (« stand alone ») ou embarquées (« embedded ») dans des produits spécifiques comme les

---

<sup>43</sup> Mémoires Non Volatiles

---

microcontrôleurs ou les cartes à puces. L'évolution du marché des dispositifs portatifs ou des supports de communication mobiles, influe directement sur l'utilisation des mémoires embarquées. En effet, la mémoire embarquée vient directement en support de la logique embarquée (programme d'application, variables temporaire des opérations de calcul, etc.). Ces mémoires sont une composante clé des SoC puisque la proximité de ces mémoires avec le CPU en fait des mémoires haute performance et haute capacité en raison de leur grande vitesse et de la largeur de leur bus d'adresses. Les mémoires embaquées sont souvent fournies sous la forme d'IP configurable pour être intégrées au niveau système.

Les principaux inconvénients des mémoires embarquées sont qu'elles sont généralement de taille plus importante et sont plus complexes à concevoir et fabriquer que les mémoires isolées. Par ailleurs, un compromis doit souvent être trouvé entre design et technologie puisque une technologie optimale pour une cellule mémoire l'est rarement pour des dispositifs logiques. De plus, l'intégration de ces mémoires devient encore plus complexe lorsque le concepteur intègre différents types de mémoire sur la même puce.

#### a. Technologie à grille flottante

Le fonctionnement du transistor de mémorisation à grille flottante consiste à stocker des charges électriques dans un matériau conducteur ou semi-conducteur en faisant passer ces charges à travers un diélectrique ( $\text{SiO}_2$ ). La différence entre les technologies à grille flottante est liée aux mécanismes d'injection de charges utilisés durant les phases d'écriture et d'effacement. Lorsque le mécanisme d'injection par porteurs chauds est utilisé en phase de programmation (écriture), la cellule est considérée comme étant de technologie SIMOS<sup>44</sup>. Cette technologie est utilisée principalement dans le cas des mémoires de type EPROM. Dans le cas où le mécanisme d'injection utilisé est de type Fowler-Nordheim à travers un oxyde mince (7 à 10 nm), la technologie est appelée FLOTOX<sup>45</sup>. Cette technologie est souvent utilisée dans les applications de type EEPROM et Flash.

L'idée de base de ces technologies est de garder l'information même après coupure de l'alimentation électrique. Ceci est possible, lorsqu'on stocke des charges au niveau de la grille flottante d'un transistor mémoire, par exemple. Ces charges stockées ont pour effet de modifier la valeur de la tension de seuil (notée  $V_T$ ) du transistor. Par convention, cette tension de seuil peut être associée à deux valeurs logiques : un état logique 0 défini comme un état effacé et un état logique 1 défini comme un état écrit. La Fig. I-5 représente les caractéristiques électriques  $I_d(V_{gc})$  du transistor mémoire dans l'état écrit et effacé. La détection de l'état logique du transistor mémoire est obtenue en appliquant une tension de lecture de grille comprise entre les tensions de seuil  $V_{Tefface}$  et  $V_{Tecrit}$ . Ainsi, pour ces conditions de lecture, le transistor est soit dans un état passant, soit dans un état bloqué.

---

<sup>44</sup> Stacked gate Injection MOS

<sup>45</sup> FLOating gate Thin OXide

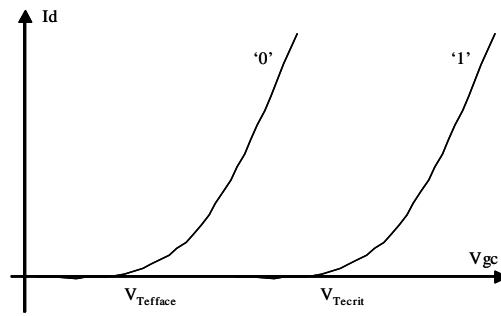


Fig. I-5 : Tensions de seuil et états logiques

### b. La mémoire EEPROM

La Fig. I-6 montre que la mémoire EEPROM se compose de deux transistors. Un transistor de sélection permet d'adresser le point mémoire et de sélectionner le mode d'utilisation (programmation ou lecture) [Yaro92]. Un transistor MOS à grille flottante appelé transistor d'état ou transistor mémoire permet de stocker l'information sous forme de charges électriques.

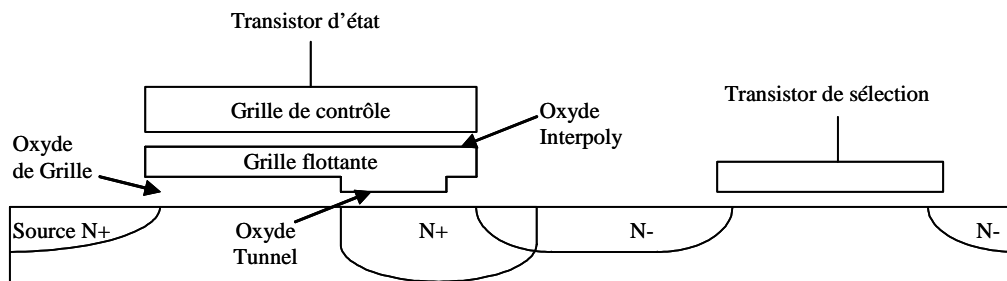


Fig. I-6 : Vue en coupe d'une mémoire EEPROM

La sélection du point mémoire se fait lorsque le transistor de sélection devient passant. Dans ce cas, la cellule peut être utilisée soit en mode lecture, soit en mode de programmation. L'écriture ou l'effacement de la cellule EEPROM se fait par injection de charges à travers l'oxyde tunnel. C'est une injection par courant tunnel Fowler-Nordheim (FN) qui a lieu à fort champ et à travers une épaisseur d'oxyde inférieure à 10 nm. On distingue trois états différents de la cellule : vierge, effacé et programmé. La cellule ne peut se trouver dans l'état vierge qu'après effacement sous UV<sup>46</sup>.

Le plan mémoire EEPROM est une matrice de lignes (« word line ») et colonnes (« bit line »). A chaque intersection correspond un point mémoire. Dans le cas des mémoires EEPROM, l'architecture NOR, présentée Fig. I-7, est principalement utilisée. Pendant les opérations de lecture, la cellule à lire est adressée en positionnant sa ligne à une tension positive alors que les autres lignes de la matrice sont connectées à la masse.

<sup>46</sup> Ultra-Violet



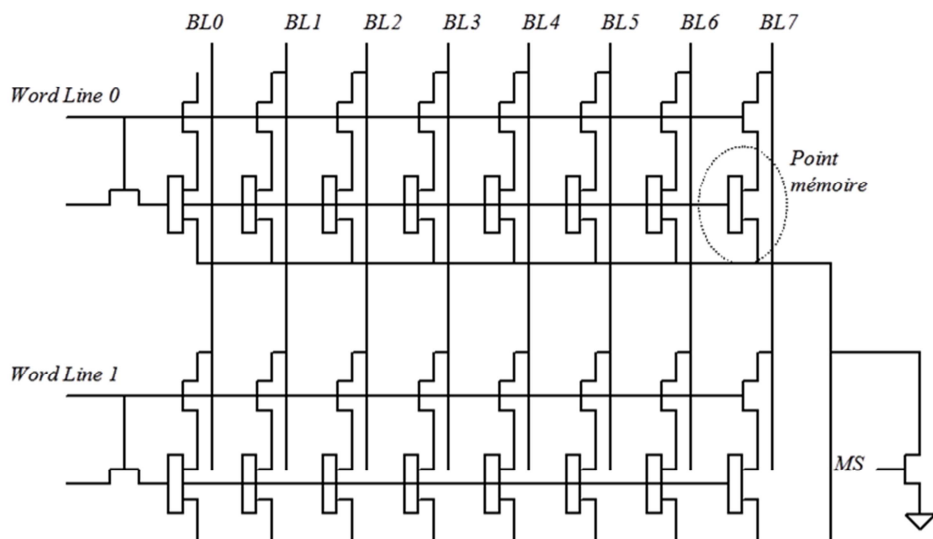


Fig. I-7 : Architecture de type NOR

Le principal inconvénient des architectures NOR est leur relative faible densité d'intégration. En effet, tous les points mémoire ont leur drain connecté à la « bit line » ainsi que leur électrode de source à une ligne commune. Ceci génère un nombre important de contacts. Dans les mémoires EEPROM de type FLOTOX, 8, 16 ou encore 32 bits adjacents forment un mot. La manière dont sont connectés ces mots (toutes les sources des transistors mémoire sont connectées entre elles) rend impossible une programmation sélective de quelques bits à l'intérieur d'un mot. Par conséquent, la programmation d'un bit au sein de cette architecture s'effectue en deux étapes.

Tout d'abord, les tensions de seuil de tous les transistors du mot sélectionné sont simultanément portées à des valeurs hautes. Ensuite certains bits, préalablement sélectionnés, sont écrits en diminuant la valeur de leur tension de seuil. L'opération qui permet une modification simultanée de l'état électrique de tous les bits d'un mot est appelée effacement. L'opération qui permet un accès sélectif à certains bits est une opération d'écriture. L'opération d'effacement entraîne donc une augmentation de la tension de seuil d'un mot mémoire complet vers une valeur positive, alors qu'une opération d'écriture entraîne une diminution de la tension de seuil de certains bits spécifiques d'un mot vers une valeur inférieure à zéro. Cette convention, adoptée pour une grande majorité de produits EEPROM, sera utilisée dans cet ouvrage.

### c. La mémoire Flash EEPROM

La cellule mémoire Flash EEPROM (appelée plus simplement mémoire Flash) se compose d'un unique transistor MOS à grille flottante comme le montre la Fig. I-8. L'idée étant d'avoir un seul transistor mémoire qui offre à la fois une rapidité de programmation, une haute densité d'intégration et un effacement électrique similaire à la cellule EEPROM.

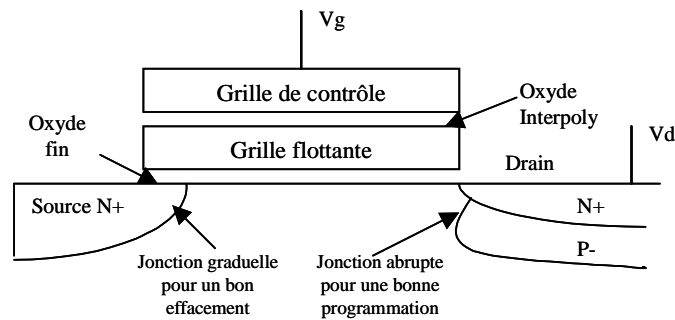


Fig. I-8 : Vue en coupe d'une mémoire Flash EEPROM

La programmation de la cellule s'effectue par porteurs chauds [Carv89]. Cette opération a pour but de stocker une quantité fixée de charges dans la grille flottante de la cellule. Les électrons sont accélérés le long du canal par une forte tension de drain ( $V_d = 5 \text{ V}$  typiquement) et acquièrent une énergie importante. Les chocs avec le réseau cristallin au voisinage du drain créent de nouvelles paires électrons-trous. C'est le phénomène d'ionisation par impact.

L'application d'une tension fortement positive sur la source (en maintenant la grille de contrôle à un potentiel nul) permet aux électrons stockés dans la grille flottante de passer à la source par conduction FN localisée [Huff80]. Durant cette opération, le drain est généralement laissé flottant de manière à limiter le courant de fuite entre drain et source. Ce type d'effacement nécessite d'une part un large recouvrement entre la jonction de source et la grille flottante (double diffusion) afin de limiter la génération de « porteurs chauds » dans la jonction source-substrat.

Pour la mémoire Flash, une meilleure densité d'intégration est obtenue grâce à une architecture de type NAND (Fig. I-9), uniquement valable pour les mémoires Flash [Brow98]. Dans ce cas, les « bit lines » sont composées de points mémoires connectés en série. Chaque colonne comprend deux transistors de sélection, le premier commandé par le signal SL (« Select Line »), permet de sélectionner la colonne adressée. Le second commandé par le signal GS (« Ground Select ») permet de relier les colonnes à la masse. Pour accéder à n'importe quelle cellule de la ligne, il faut non seulement activer la « word line » de la cellule sélectionnée, mais aussi toutes les « word lines » commandant les autres cellules de la ligne.

Lors d'une opération de lecture, on applique sur la grille du point mémoire sélectionné une tension de lecture assez faible (proche de zéro), alors qu'une tension de grille supérieure à la tension de seuil maximale de la technologie considérée est appliquée aux autres points mémoire. Ainsi la cellule sélectionnée impose le courant de « bit line » à lire.

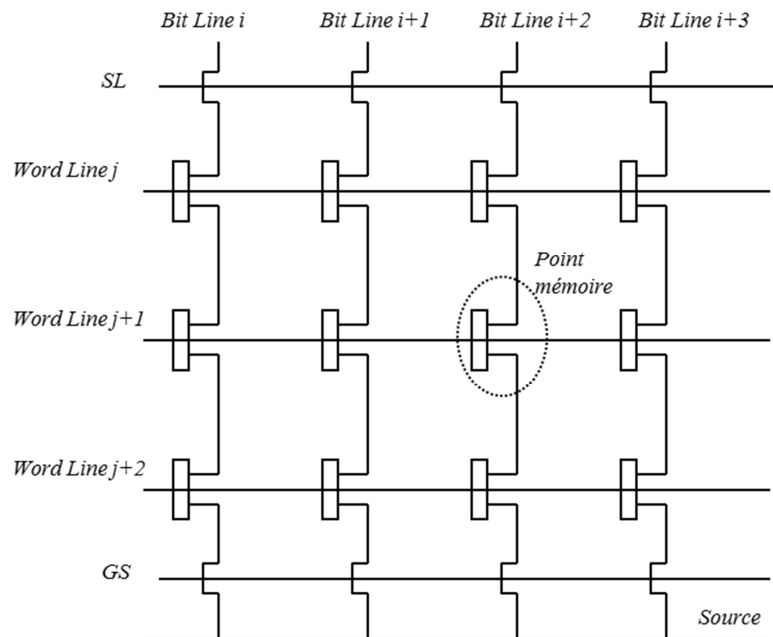


Fig. I-9 : Architecture NAND des mémoires Flash

## 5 Circuiterie périphérique des Mémoires Non Volatiles

La structure des mémoires non volatiles obéit à deux principes généraux :

- l'exploitation du plan mémoire qui nécessite l'élaboration d'une électronique analogique chargée de le mettre en œuvre (en tenant compte des contraintes imposées par la technologie) et de convertir l'état des cellules mémoires en grandeurs binaires.
- le traitement de l'information binaire, réalisé par une logique de contrôle chargée de la gestion de la communication entre le plan mémoire, les circuits périphériques et les entrées-sorties du composant

La Fig. I-10 fournit une représentation fonctionnelle d'une mémoire non volatile embarquée.

La fonction de mémorisation des mémoires non volatiles nécessite l'application d'un fort champ électrique à travers un oxyde tunnel de manière à obtenir un transfert de charges au niveau de la grille flottante. Et cela, pour des opérations d'écriture comme pour des opérations d'effacement. Ce fort champ est obtenu par l'application de tensions élevées aux bornes de la cellule mémoire. La production de la haute tension à partir d'une tension d'alimentation est l'un des points les plus critiques : cette haute tension doit, en effet, être stable pour des tensions d'alimentation comprises entre 1.5V et 3.5V, mais aussi dans la gamme de température dans laquelle sont testés les produits (- 40°C à +125°C). Ainsi, un signal de programmation haute tension (10V à 14V) est généré à partir d'un circuit, appelé pompe de charge, interne au produit. Le fonctionnement de la pompe de charge est basé sur un circuit multiplicateur de tension de type SCHENKEL [Dick76]. Plus précisément, le circuit de génération de la haute tension comprend un oscillateur qui attaque un multiplicateur de tension. La haute tension générée est ensuite prise en charge par le générateur de rampe de manière à produire un signal ayant une forme bien déterminée (signal double rampe).

Lors des opérations de lecture, l'application sur la grille de contrôle de la cellule mémoire adressée, d'une tension de lecture comprise entre la tension de seuil écrite et la tension de seuil effacée d'une cellule mémoire permet de différencier les deux états électriques de la cellule. Dans l'état électrique effacé, on sera en présence d'un transistor bloqué. Dans l'état électrique écrit, le transistor mémoire sera dans un état passant. Dans la pratique, la détection se fait par la lecture du courant traversant le point mémoire. On effectue en général une comparaison entre le courant fourni par une cellule vierge et celui fourni par la cellule à évaluer, toutes les deux polarisées avec une même tension de grille notée  $V_{ref}$  (noté Ref. I/V sur la Fig. I-10). Une conversion courant tension permet ensuite de traduire le résultat de la comparaison en niveau logique. Ce rôle est rempli par un circuit de lecture appelé amplificateur de lecture.

Tous les signaux utilisés au sein de la matrice mémoire durant les opérations d'adressage, de lecture et de programmation sont acheminés par l'intermédiaire des décodeurs lignes et colonnes. L'acheminement des signaux de programmation nécessite des bascules haute tension, encore appelées « latches ». Le rôle des bascules est d'acheminer les signaux de grille et de « bit line » durant les phases d'effacement et d'écriture.

Les bascules haute tension de « bit line » sont munies en sortie d'un commutateur. Lors d'une opération de lecture, ce commutateur déconnecte la bascule de la « bit line » pour que celle-ci soit reliée à l'amplificateur de lecture. On retrouve une structure similaire dans le bloc de décodage des lignes de manière à pouvoir appliquer sur la grille du transistor mémoire soit une tension constante lors de la phase de lecture, soit une rampe en tension lors de la phase d'effacement.

Une logique de contrôle représentée par la machine d'état appelée PLA<sup>47</sup>, ou un microcontrôleur sur certaines technologies Flash, est chargée de gérer le fonctionnement interne de la mémoire à partir d'informations contenues dans les registres de commande. Cette logique de contrôle est le cœur du système et permet de générer les différents signaux internes au circuit durant les phases de lecture et de programmation. En fonction des signaux qu'il reçoit, et selon un organigramme pré-établi, c'est lui qui génère les signaux de commande du circuit. La logique d'Entrée/Sortie gère le flux d'informations porté sur chaque broche externe du composant.

---

<sup>47</sup> Programmable logic array

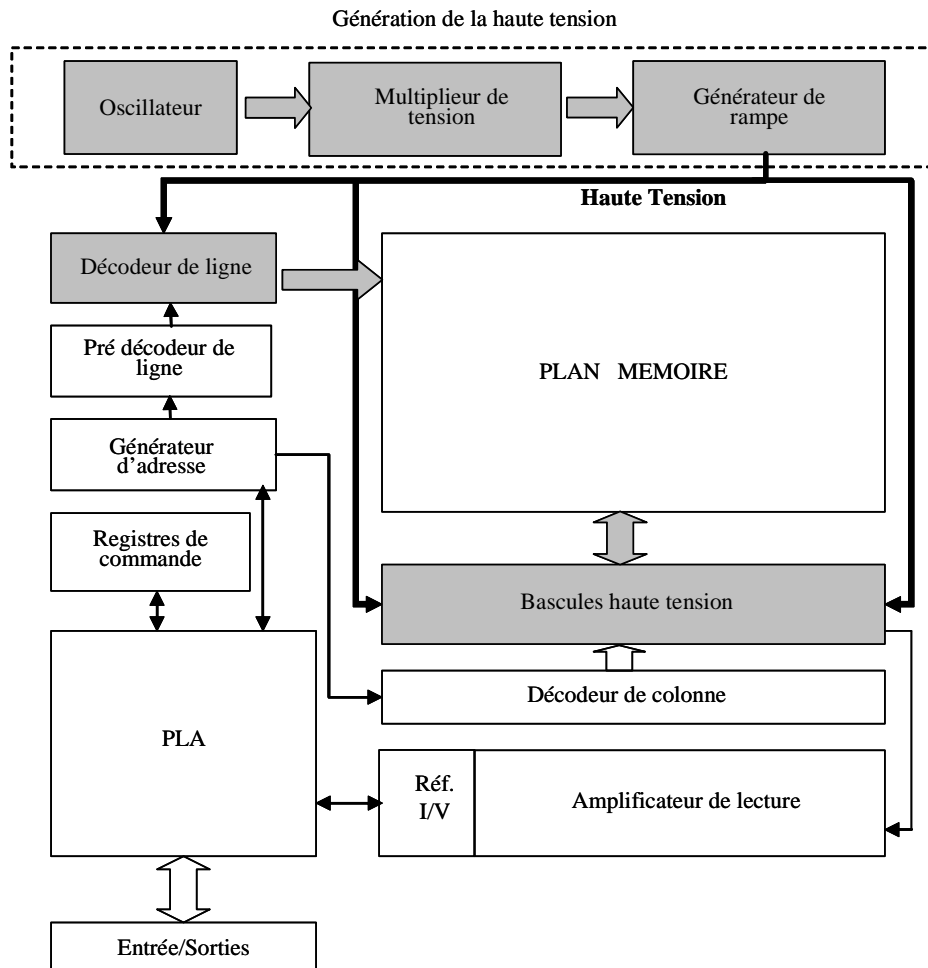


Fig. I-10 : Architecture classique d'une MNV

## C. Niveaux de Test et Rendement de Fabrication

### 1 Placement des structures de test

Le test des circuits intégrés est un sujet fondamental. En effet, le marché de la microélectronique est en perpétuelle évolution et concerne maintenant des applications grand public. Les volumes de production sont de plus en plus importants et il est impératif d'assurer une fiabilité satisfaisante pour un prix abordable. Cela nécessite de développer des techniques de test efficaces et le moins coûteuses possible pour les circuits en microélectronique.

Pendant le processus de fabrication, la qualité du processus est vérifiée en effectuant des mesures sur les structures de test élémentaires ou TEG<sup>48</sup>, situées dans les lignes de découpe de la plaquette de silicium, permettant, entre autre, de s'assurer de la bonne exécution de chaque étape du processus technologique.

Une technique complémentaire consiste à surveiller des paramètres électriques plutôt que le processus lui-même. Alors que certaines structures basiques permettent de vérifier des paramètres de bas niveau, il est tout à fait concevable d'effectuer des mesures de comportement sur des structures plus complexes à

<sup>48</sup> Test Element Group

condition naturellement d'être capable de relier les paramètres électriques mesurés, aux paramètres intrinsèques du processus technologique. Ces structures complexes appelées « véhicules de test » peuvent aussi bien cibler la fiabilité de la partie MNV ou de la partie digitale des microcontrôleurs.

La Fig. I-11a représente une plaquette de silicium ou « wafer » composée de zones rectangulaires indiquant les champs d'exposition lithographique. Le motif rectangulaire utilisé pour chaque couche du processus de fabrication est délimité par une plaque de verre ou masque connu sous le nom de réticule. Les lignes de découpe, de largeur égale à la largeur de la scie utilisée lors de la séparation des puces (de l'ordre de 30-100  $\mu\text{m}$ ), sont présentées Fig. I-11b. Ces lignes contiennent les TEG. Comme le montre les Fig. I-11c et Fig. I-11d, l'espace occupé par ces lignes de découpe (de couleur de sombre) est très limité en terme de surface s'il n'y a qu'une seule puce produit par champ de réticule, mais peut devenir beaucoup plus importante pour les réticules « multi-chip ». Ces derniers sont utilisés pour embarquer plusieurs véhicules de test ainsi que les produits correspondants durant, notamment, les phases d'introduction d'une nouvelle technologie.

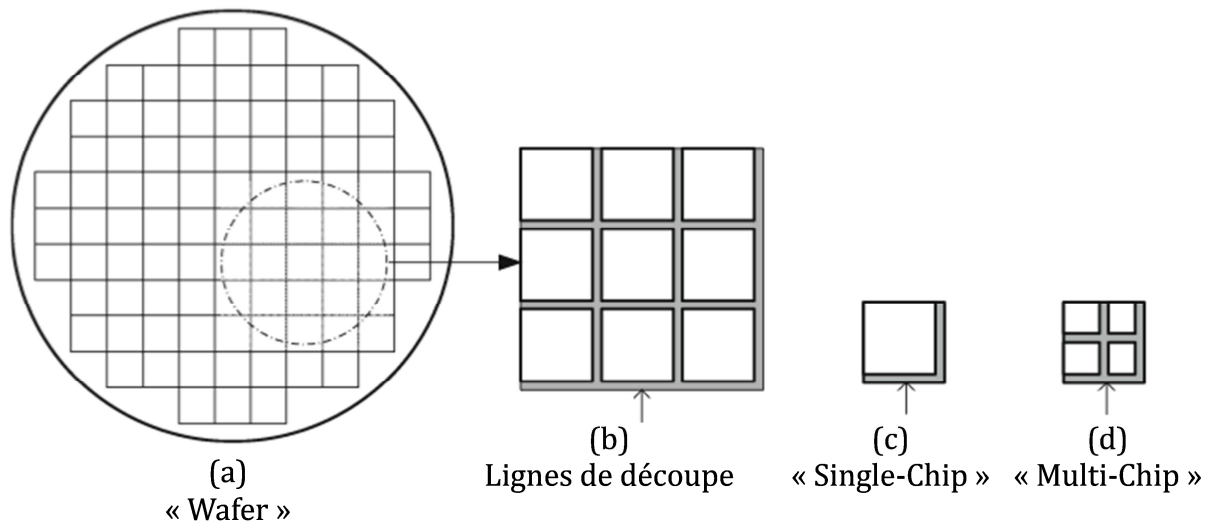


Fig. I-11 : Champs de réticule d'une plaquette de silicium (a) avec mise en évidence des lignes de découpe (b) & (c) et de l'augmentation de la surface de ces lignes pour les réticules multi puces (d).

## 2 Rendement de fabrication

Les unités de fabrication CMOS exigent un environnement ultra propre (salles blanches) et l'utilisation de systèmes de manipulation de plaquettes de silicium automatisés ou semi automatisés. Ces unités de fabrication produisent chacune des plaquettes de silicium de diamètres fixes, allant de 100 mm (4 pouces) à 300 mm (12 pouces). Le procédé de fabrication d'une puce électronique comporte de nos jours plus de 400 opérations et peut utiliser jusqu'à une cinquantaine de masques différents. Ainsi, lorsqu'une puce arrivée au test final est déclarée non fonctionnelle, il n'est pas aisé de déterminer l'origine de la défaillance de manière à corriger une ou plusieurs étapes de la chaîne de production.

Le rendement se définit comme le rapport entre le nombre de puces bonnes à l'issu du test sur le nombre de puces potentielles que contient la plaquette de silicium. Le nombre de puces potentielles correspondant au nombre de puces

---

entières, inscrites dans la surface de la plaquette et pouvant donc être, a priori, fonctionnelles.

### 3 Facteurs limitatifs du rendement

#### a. La défektivité

La défektivité notée  $D_0$ , représente en général le nombre de défauts par  $\text{cm}^2$  (ou  $\text{m}^2$ ) et par niveau de masquage. La défektivité peut être classée en deux catégories, la défektivité aléatoire et la défektivité systématique. La défektivité aléatoire peut être due à des « poussières », de la corrosion, ou de la contamination ionique provenant des équipements. Les défauts systématiques sont, quant à eux, souvent liés à une configuration du dessin de masque bien particulière. Soit cette dernière ne respecte pas une règle de dessin, soit les règles de dessin ne prennent pas en compte certaines configurations que l'on ne retrouve pas dans le produit. Les défauts systématiques peuvent également être causés par des problèmes de lithographie ou de planéité.

#### b. La variabilité

L'importance et la complexité des variations du procédé de fabrication augmentent avec la réduction des nœuds technologiques [Agar07]. En dépit de l'avancée des techniques optiques et notamment l'amélioration de la résolution [Lieb01], les variations de la lithographie continuent d'être un réel enjeu pour les technologies inférieures à 90nm. De plus, la réduction des tailles des dispositifs dans les technologies fortement nanométriques ont aussi pour conséquences de nombreuses autres sources de variations telles que :

- les variations de dopant [Keye75],
- les effets de proximité des caissons [Moro03],
- les variations de stress dépendantes du dessin de masque dans les technologies agressives [Moro06],
- les variations induites par la température lors du RTA<sup>49</sup> [Ahsa06].

Ces sources de variations doivent être caractérisées et modélisées pour améliorer la corrélation entre les modèles et le silicium. Il est possible de classer ces sources de variations en deux grandes catégories : les variations aléatoires et les variations systématiques.

Les variations aléatoires peuvent causer d'importantes différences de comportement entre deux dispositifs dessinés de la même façon et placés l'un à côté de l'autre. Ce phénomène est appelé « mismatch » et sera développé plus loin. Ces variations sont imprévisibles et sont causées par des incertitudes aléatoires au niveau du procédé de fabrication. Cela peut provenir de fluctuations microscopiques du nombre et de la localisation des atomes de dopant dans le canal [Keye75] ou de la rugosité des bords de lignes de poly silicium [Oldi00]. Les variations locales aléatoires sont critiques pour les structures sensibles au « matching » comme les paires de transistors appareillés ou les SRAM.

---

<sup>49</sup> Rapid Thermal anneal : recuit thermique rapide

## 4 Outils de mesure et d'analyse des pertes de rendement

Les outils disponibles pour mesurer et comprendre les pertes de rendement sont nombreux. Il existe cependant trois principaux outils pour mesurer les pertes de rendement. Ces outils sont la mesure de la défektivité en ligne (notée aussi D0), la caractérisation et le test électrique, évoqués précédemment.

Le D0 correspond à l'inspection optique des plaquettes de silicium. La caractérisation consiste à effectuer des mesures sur des éléments simples pour savoir si le procédé de fabrication est bien centré sur ses valeurs nominales. Le test électrique quant à lui, est utilisé pour discriminer les circuits fonctionnels des circuits défectueux.

Lorsqu'une perte de rendement est détectée, il est nécessaire de l'analyser pour comprendre son origine et corriger rapidement la chaîne de production. Si la perte de rendement provient d'une variation du procédé, l'analyse sera directement réalisée sur la TEG ayant détectée cette variation. Cependant, si la perte est liée à de la défektivité sur un produit, il faudra dans un premier temps localiser le ou les défauts pour pouvoir ensuite les analyser. Une localisation précise du défaut devra ensuite être réalisée de manière à le soumettre à certaines techniques d'analyses.

### a. Mesure de la défektivité en ligne

La mesure de la défektivité est une des premières étapes de vérification du processus de fabrication. Après certaines étapes de fabrication critiques, les plaquettes sont inspectées par des équipements optiques dédiés. Cette inspection permet de détecter des défauts de fabrication et des contaminations engendrées par des particules. Les équipements réalisent des comparaisons successives de puce à puce. Les puces sont scannées avec une résolution variable selon les besoins (typiquement de  $0.05\mu\text{m}$  à  $0.2\mu\text{m}$  [Team08]). Il est très important de bien choisir cette résolution, car elle impacte directement le temps de l'inspection (de  $11\text{cm}^2/\text{h}$  pour  $0.05\mu\text{m}$  à  $160\text{cm}^2/\text{h}$  pour  $0.2\mu\text{m}$ ).

Une deuxième étape, appelée la revue de défauts, peut être effectuée sur certaines plaquettes. Cette étape permet de se repositionner sur certains défauts de manière à les photographier pour essayer de déterminer leur origine. La Fig. I-12a est une photographie d'un défaut de type contamination englobée. Ce type de défaut est dû à la présence d'une particule parasite apparue avant l'étape de dépôt du cuivre, qui est ensuite englobée durant cette même étape. La Fig. I-12b correspond, quant à elle, à un défaut de type « sur-attaque » qui apparait lorsque le temps de gravure est trop long. Ce type de défaut agit sur la rugosité des lignes.

La mesure de défektivité est l'outil d'évaluation du rendement qui intervient le plus tôt dans le procédé de fabrication. De plus, grâce à la revue des défauts, qui intervient souvent avant et après chaque étape du procédé de fabrication, cette méthode d'inspection permet de connaître rapidement l'origine du problème. Le point négatif de cet outil est, qu'en raison de son temps d'exécution et du fait qu'il intervient plusieurs fois au cours de processus de fabrication, toutes les plaquettes ne sont pas inspectées et tous les défauts ne sont pas revus. Par conséquent certains problèmes aléatoires peuvent ne pas être décelés. De plus, la



détection de défaut se limite à une étude topographique de surface. De ce fait, les anomalies dans les vias par exemple ne peuvent pas être détectées.

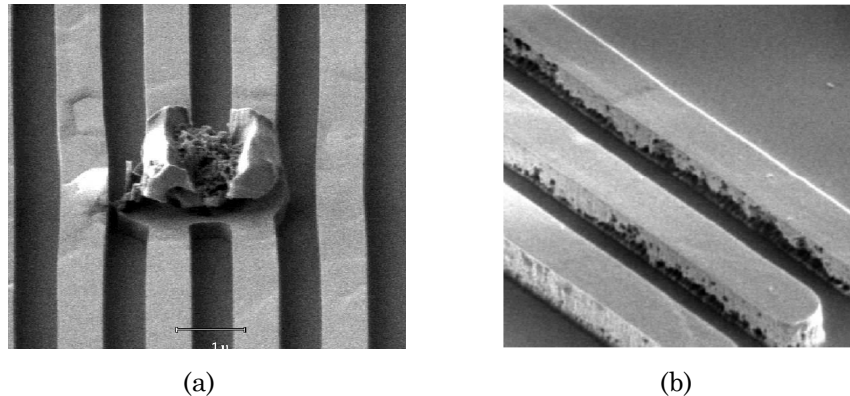


Fig. I-12 : Processus de détection de défaut utilisé durant le D0

#### b. La caractérisation

Plusieurs niveaux de contrôle du procédé de fabrication existent. Les TEG paramétriques permettent d'extraire tous les paramètres du Front End Of the Line (FEOL) et du Back End Of the Line (BEOL). Pour le FEOL, les paramètres peuvent être la tension de seuil ( $V_T$ ), la transconductance ( $g_m$ ), le courant du transistor en mode de saturation ( $I_{ON}$ ) etc. Pour le BEOL, les paramètres des résistances et capacités de ligne sont analysés. En général, une vingtaine de ces structures sont requises pour caractériser une technologie (avec plus de cent paramètres associés). De plus, en raison du grand nombre de structures à tester, et pour ne pas avoir un temps de test trop important, seulement quelques réticules de la plaquette sont mesurés dans le flot standard. En effet, 3h par plaquette sont nécessaires pour tester tous les sites (cf. Fig. I-11), alors qu'il ne faut que 6h à 8h pour un lot de 25 plaquettes en ne testant que cinq sites.

Lorsqu'une correction est apportée sur le procédé de fabrication, il faut attendre en moyenne six semaines pour avoir un retour sur le bien-fondé de cette correction. Une solution pour éviter cela est de concevoir des structures de test appelées « short loop » qui ne sont composées que de quelques niveaux de masques. Ainsi, ces structures sont fabriquées et testées plus rapidement, et le procédé de fabrication est ainsi corrigé en un temps beaucoup plus court. De plus, grâce au faible nombre de masques nécessaires, ces structures sont moins onéreuses que les véhicules de test. Elles sont en revanche limitées à de simples structures résistives [Hess01], mais fournissent tout de même de précieuses indications sur la défektivité des étapes de fabrication ciblées.

Après fabrication, les produits sont testés électriquement pour séparer les pièces défectueuses de celles qui vont être vendues. Ce test est composé de deux grandes étapes, la première étape est le test paramétrique DC, suivie d'une étape de test fonctionnel et (ou) structurel. Le test paramétrique DC évalue, dans un premier temps, la continuité électrique entre les PAD<sup>50</sup> et le testeur. Ensuite, des mesures statiques sont effectuées pour détecter les puces non fonctionnelles suivant les valeurs de leurs paramètres DC (mesures de courant de consommation, des

<sup>50</sup> PAD : surface métallique réalisée au dernier niveau de métal pour connecter les entrées/sorties du circuit avec l'extérieur

niveaux de tension  $V_{IH}$ ,  $V_{IL}$ ,  $V_{OH}$  et  $V_{OL}$ , etc.). La seconde partie du test électrique consiste à évaluer toutes les fonctionnalités du produit.

Le principal résultat de cette analyse est une cartographie des résultats issus du programme de test. Ces résultats sont traduits en classe de rejet, appelés « binning » ou « bin ». Ces classes de rejet sont des nombres qui traduisent la nature électrique de la défaillance d'un produit. Hormis le « bin 1 » qui correspond à une puce 100% fonctionnelle, chaque « bin » correspond à une fonctionnalité défaillante.

## 5 L'analyse de défaillance

Lorsqu'une perte de rendement est détectée, il est nécessaire de l'analyser pour comprendre son origine et corriger rapidement la chaîne de production. Si la perte de rendement provient d'une variation du procédé, l'analyse sera directement réalisée sur la TEG ayant détectée cette perte. Cependant, si la perte est liée à de la défektivité sur un produit, il faudra dans un premier temps localiser ce défaut pour pouvoir ensuite l'analyser.

### a. Les techniques de localisation

Les techniques de localisation sont dites optiques car elles nécessitent l'emploi de microscopes. Il existe deux grandes catégories de techniques de localisation optiques, les techniques statiques et les techniques dynamiques [Mach08].

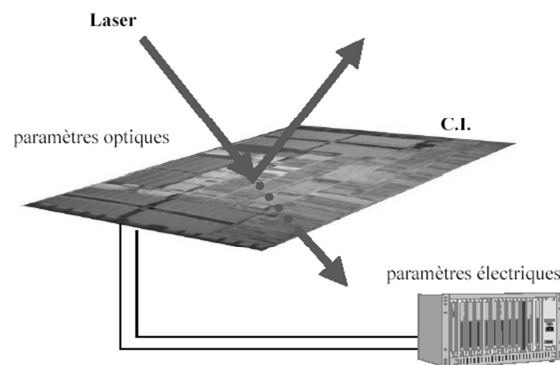


Fig. I-13 : Interaction du faisceau laser avec le circuit intégré

Les techniques statiques peuvent elles-mêmes être décomposées en deux catégories, les techniques à base de stimulation thermique laser (Fig. I-13) ou d'émission de lumière. Les techniques les plus répandues en termes de stimulation laser sont l'OBIRCh<sup>51</sup> et le TIVA<sup>52</sup>.

Pour le TIVA, le circuit est alimenté avec un courant constant et une variation de tension est mesurée (due à une variation de résistance traversée par un courant constant). L'effet recherché est donc une variation de la résistance du matériau, lorsque qu'il est soumis à un gradient de température, généré par la stimulation laser. En effet, le laser a pour effet d'augmenter localement la température du circuit sur le point stimulé.

Pour la méthode OBIRCh, le circuit est alimenté avec une tension constante et une variation de courant est mesurée. La conséquence est une variation du

<sup>51</sup> Optical Beam Induced Resistance Change

<sup>52</sup> Thermally-Induced Voltage Alteration

courant due à une variation de la résistance sur laquelle une tension constante est appliquée. Le laser balaye ainsi la surface de la puce, et les variations de courant (ou de tension) sont enregistrées sous la forme d'une cartographie. Les zones ayant subies des variations de courant apparaissent en rouge ou en vert, en fonction du sens de variation du courant. Cette cartographie est ensuite superposée sur la photo prise par le microscope pour localiser les défauts sur le circuit. La Fig. I-14 est un exemple de défaut localisé grâce à la technique OBIRCh.

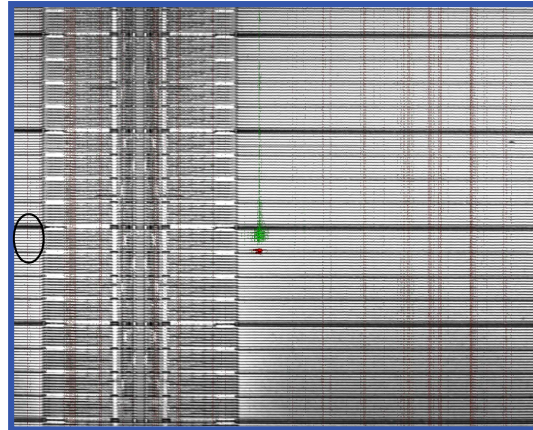


Fig. I-14 : Exemple de défaut localisé par la technique OBIRCh

Contrairement aux méthodes laser, la technique d'émission de lumière ne nécessite aucune stimulation optique externe. Cette technique est appelée l'EMMI<sup>53</sup>, elle consiste à observer l'émission de lumière lorsque le circuit est placé dans son mode de défaillance. L'observation peut se faire soit par la face avant, soit par la face arrière selon le cas. Les défauts pouvant potentiellement émettre des photons sont :

- Les claquages d'oxydes,
- La recombinaison radiative,
- Les porteurs chauds (transistors mos en régime de saturation),
- La radiation thermique proche de l'infrarouge.

Les techniques dynamiques sont utilisées lorsque les défauts ne sont pas visibles à partir des tests statiques. Ces défauts entraînent une baisse de performances du circuit qui n'est décelable qu'à l'aide de mesures fonctionnelles. Ces méthodes sont basées sur l'utilisation d'un outil très répandu dans le monde du test, le « shmoo plot ». Un exemple de ce type de graphique est donné Fig. I-15. Ce graphique représente la limite de fonctionnement du circuit pour différentes périodes d'horloge (en abscisse) et différentes tensions d'alimentations (en ordonnée).

La méthode dynamique consiste à placer le circuit sur un point du « shmoo plot », en bordure de la frontière de fonctionnement, et d'effectuer le test à chaque pas de déplacement du laser. Lorsque le laser est positionné sur le défaut, le test doit basculer de PASS à FAIL ou de FAIL à PASS.

<sup>53</sup> Emission Microscopy

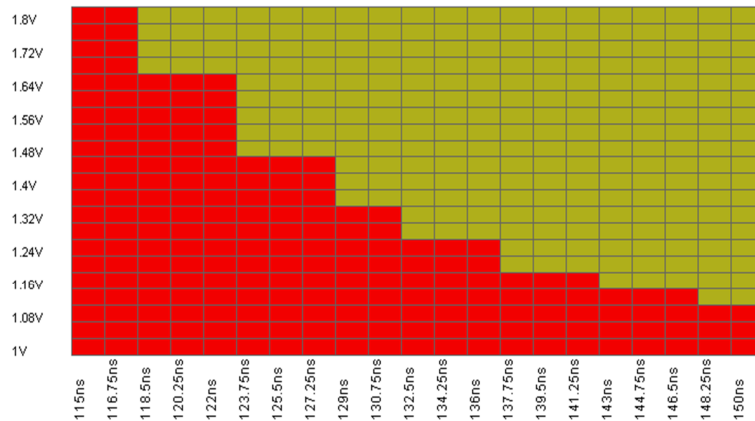


Fig. I-15 : Exemple de shmoo plot tensions-périodes

Les méthodes dynamiques utilisant les « shmoo plot » sont appelées DLS<sup>54</sup>, il en existe 2 types, les DLS synchrones et les DLS asynchrones.

Le type asynchrone, par définition, ne nécessite aucun couplage entre le testeur et le laser, il est donc plus facile à mettre en œuvre. Cependant, cette méthode doit être bien paramétrée pour donner des résultats satisfaisants. Il faut, par exemple, que la période du balayage laser soit plus grande que la durée d'exécution d'un flot de test.

Dans la seconde approche, le balayage et l'application du flot de test sont synchronisés, de manière à ce qu'une séquence de test complète soit réalisée à chaque pas de déplacement du laser. La résolution spatiale est alors optimale.

#### b. Les analyses physico-chimiques

Après avoir localisé les défauts, il faut ensuite les analyser, pour déterminer à quelle étape ils ont été créés, et quelle en est la cause. Le processus d'analyse physico-chimique comporte plusieurs étapes. Pour être analysé, l'échantillon doit tout d'abord être « préparé », dans le but d'être ensuite observé ou analysé physiquement.

L'observation de l'échantillon permet de visualiser et donc de trouver l'origine et la nature de certains défauts, ainsi que de mesurer des dimensions (lignes, jonctions etc.). Pour ce faire, deux classes de microscopes sont utilisées, les microscopes optiques et les microscopes à base de pointes [Fare08].

Dans la classe optique, il existe deux types de microscopes, ceux employant des récepteurs sensibles aux photons, et ceux employant des canons à électrons. Les premiers sont très employés pour observer la surface des plaquettes et peuvent fournir une résolution de 0.2 $\mu$ m. Les microscopes utilisant un canon à électrons ont l'avantage de fournir une meilleure profondeur d'image. Les deux principales techniques d'observation optique sont le SEM<sup>55</sup> et le TEM<sup>56</sup>.

<sup>54</sup> Dynamic Laser Stimulation

<sup>55</sup> Scanning Electron Microscopy

<sup>56</sup> Transmission Electron Microscopy

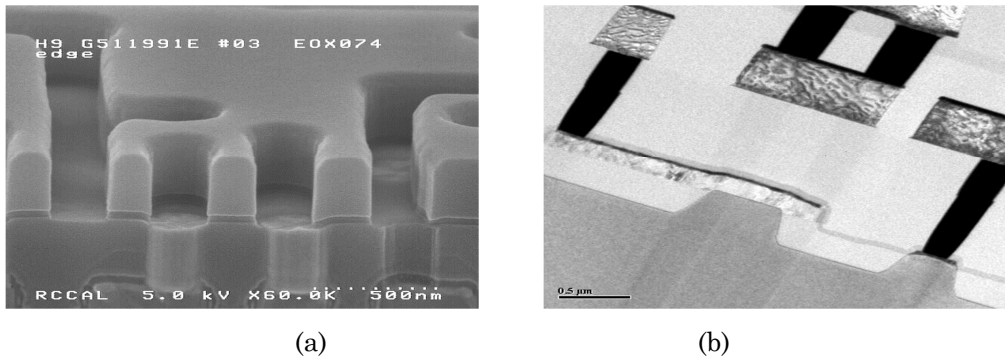


Fig. I-16 : Exemple de microscopie SEM (a) et TEM (b)

Le SEM possède une résolution moins importante, mais son champ d'observation est plus élevé (Fig. I-16a). Le TEM est, quant à lui, très employé pour obtenir une vue en coupe très précise permettant de distinguer les matériaux de densité différente (Fig. I-16b).

Dans le cas des microscopes à base de pointes, une pointe est utilisée pour balayer la surface de l'échantillon. Les mouvements verticaux de la pointe vont avoir une incidence sur l'orientation d'un faisceau laser pour l'AFM<sup>57</sup>, ou sur le courant induit dans un tube piézoélectrique pour le STM<sup>58</sup>. En plus de leurs performances pour la caractérisation de surface, ces microscopes permettent d'autres mesures. L'AFM pourra, par exemple, mesurer des épaisseurs d'isolants ou des dopages, tandis que le STM est capable de mesurer des densités de courant.

Les analyses chimiques constituent l'ultime étape de l'analyse de défaillance. Cette étape permet de déterminer l'origine de la défaillance [Fare08b]. Les 4 techniques d'analyses chimiques les plus répandues sont :

- l'Energy Dispersive Spectroscopy (EDS),
- l'Electron Energy Loss Spectroscopy (EELS),
- la Secondary Ion Mass Spectrometry (SIMS),
- l'Auger Electron Spectroscopy (AES).

L'EDS et EELS sont deux techniques non destructives, utilisant un canon à électrons. Dans la première, c'est l'émission de rayon-X qui va être étudiée, dans l'autre, c'est la perte d'énergie des électrons. L'EELS permet également de reconstituer une image avec une couleur par élément (Fig. I-17). Ensuite, la SIMS et l'AES sont deux techniques destructives. La SIMS consiste à analyser les ions secondaires générés par l'échantillon, lui-même attaqué par un canon à ions. Cette technique est très utilisée pour obtenir des profils de concentration de dopant dans le silicium. Finalement, l'AES est une technique qui consiste à analyser les électrons secondaires, émis par l'échantillon qui est attaqué par un faisceau d'ions et un faisceau d'électrons. L'AES permet de reconstituer une image de la concentration de chaque élément d'un échantillon.

<sup>57</sup> Atomic Force Microscope

<sup>58</sup> Scanning Tunneling Microscope

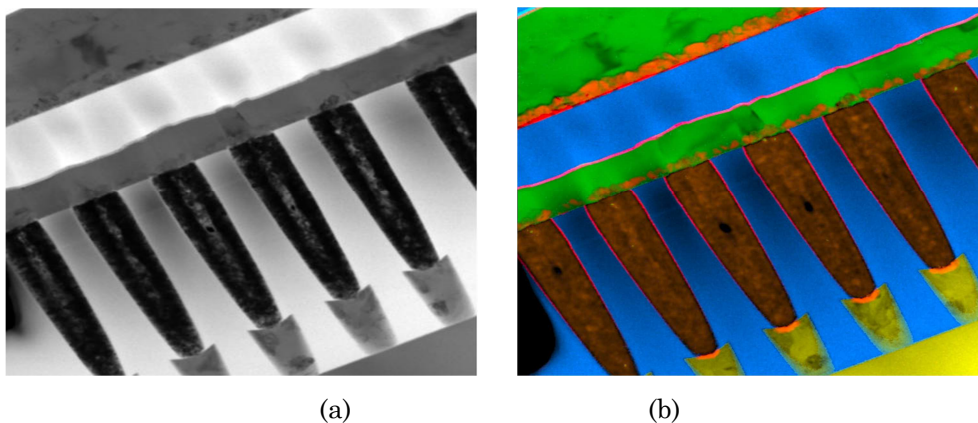


Fig. I-17 : Reconnaissance de matériaux par analyse EELS

## II. Fiabilité des MNV embarquées

### A. Principaux critères de fiabilité

Les technologies MNV embarquées nécessitent des tensions de programmation assez élevées pour être opérationnelles. Par conséquent, il existe au sein de ces structures des champs électriques relativement forts, notamment aux bornes des oxydes tunnels et oxydes de grille. De plus, dans le cas de mémoires embarquées (mémoires insérées dans un circuit logique conçu dans le procédé de fabrication CMOS classique), il a fallu introduire de nouvelles méthodes de conception et de fabrication. Dans cette technologie de fabrication mixte (logique et mémoire non volatile), il existe trois zones distinctes, fabriquées de manière différente. La première correspond au cœur mémoire. La seconde est une zone appelée « zone haute tension » : elle est composée de transistors ayant une épaisseur d'oxyde assez élevée pour résister à la haute tension de programmation. Enfin, la dernière zone, qualifiée de « zone basse tension », représente la partie CMOS.

La fiabilité des mémoires non volatiles repose principalement sur trois critères : la fonctionnalité bien sûr, mais surtout la rétention et l'endurance [Cape99]. Ces trois caractéristiques devront être validées durant les phases de test. A cela s'ajoutent des critères de performance incontournables comme le temps d'accès en lecture, la durée de programmation, la consommation en mode de lecture, d'écriture ou de veille, la surface occupée, etc.

#### 1 Caractéristiques en endurance

Les MNV peuvent être reprogrammées. Mais, contrairement aux mémoires volatiles, chaque programmation induit « une usure » permanente au niveau du point mémoire. Cela implique une limitation du nombre total de cycles de programmation. Le nombre de cycles supportés par la mémoire avant l'apparition de la première défaillance porte le nom d'endurance.

Le test d'endurance consiste à effectuer plusieurs cycles d'effacement-écriture sur une cellule élémentaire ou un plan mémoire, et à prélever la tension de seuil (en écriture et en effacement) ou les distributions de tension de seuil correspondant à

---

chaque nombre de cycles. La différence entre la tension de seuil en écriture et en effacement est appelée fenêtre de programmation.

L'endurance d'une MNV est principalement limitée par la tension de claquage de l'oxyde tunnel. Le claquage d'oxyde serait provoqué par des points faibles (défauts) situés dans l'oxyde. Ces défauts sont généralement activés par de forts champs électriques et engendrent des fuites de charges. Un autre phénomène est lié au piégeage de charge dans l'oxyde durant les phases de programmation. Il a pour effet de modifier le champ aux bornes de l'oxyde et dans le même temps la quantité de charges injectées. Cela peut induire un décalage insuffisant des tensions de seuil et perturber la détection de l'état de la cellule durant la phase de lecture.

## 2 Caractéristiques en rétention

Le but du test en rétention est d'estimer la capacité de la cellule mémoire ou du plan mémoire à sauvegarder l'information. Dans les conditions d'utilisation standard, c'est-à-dire à température ambiante, les cellules mémoires ne doivent pas perdre plus de 10% de la charge stockée pendant 10 ans. Afin d'activer et d'accélérer la perte de charge, le test en rétention se fait généralement à haute température (entre 150°C et 300°C). On effectue par la suite une mesure de la dérive de la tension de seuil à différentes températures. L'extrapolation de cette dérive à température ambiante, suivant la loi d'Arrhenus, permet d'évaluer la durée de vie de la cellule mémoire [Salv99].

## 3 Extraction des distributions de $V_T$

De manière à évaluer le comportement électrique global d'un nombre important de cellules d'un plan mémoire, l'extraction des distributions de tension de seuil ( $V_T$ ) est réalisée. La distribution des  $V_T$  des cellules d'un plan mémoire permet de mettre en évidence des populations de cellules marginales ( $V_T$  hors ou à la limite des spécifications). De plus, une dispersion importante de ces distributions permet d'évaluer la robustesse du procédé de fabrication.

L'extraction des distributions de  $V_T$  est primordiale lors des tests d'endurance et de rétention. Dans le cas des microcontrôleurs, l'accès et l'évaluation des  $V_T$  est très difficile. Une alternative à cette limitation a été d'introduire des circuits de test qui accompagnent les produits embarquant de la MNV. L'architecture de ces véhicules de test se rapproche des mémoires à accès parallèle et permet une caractérisation précise du plan mémoire.

La Fig. I-18 est une illustration des distributions de tension de seuil d'un plan MNV. La Fig. I-18.a est une représentation classique de la distribution de  $V_T$  alors que la Fig. I-18.b est une représentation en échelle logarithmique qui permet d'évaluer plus précisément les queues de distributions révélatrices des populations de cellules marginales.

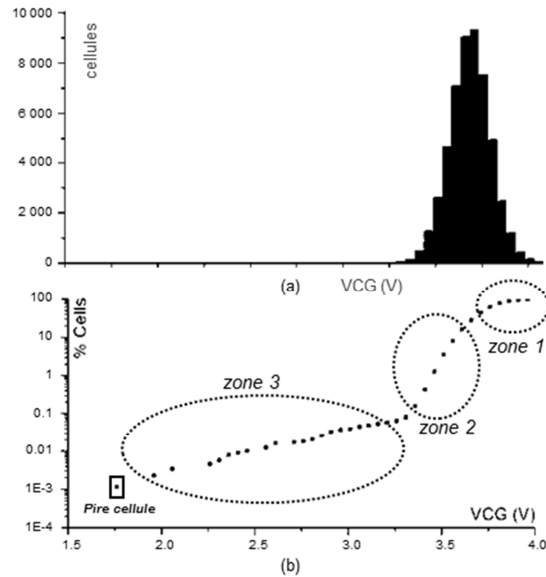


Fig. I-18 : Distribution des tensions de seuil

## B. Evaluation de la fiabilité des MNV

### 1 Les véhicules de test (« Test Chip »)

Les véhicules de test ont joué un rôle essentiel et indispensable dans la croissance et le succès de l'industrie des semi-conducteurs depuis l'invention du transistor bipolaire en 1947, jusqu'au développement généralisé des SoC actuels. La complexité croissante des circuits VLSI a fait de l'utilisation des véhicules de test des structures incontournables. Ces véhicules sont utilisés pour anticiper les caractéristiques et le rendement de certains éléments du circuit mais aussi de blocs de circuits fonctionnels complets dédiés à la logique et à la mémoire.

Parmi ces véhicules de test, ceux dédiés au MNV possèdent des propriétés particulières qui seront développées dans cette partie. En effet, leur taille importante, nécessaire à une étude statistique des caractéristiques du plan mémoire, ne permet pas une intégration dans les lignes de découpe. Ils sont généralement fabriqués sur des plaquettes de silicium comportant de plusieurs véhicules de test par réticule et nécessitent des testeurs industriels.

Les véhicules de test dédiés aux MNV peuvent être utilisés en mode fonctionnel (repos, lecture, écriture de données) ou en mode test suivant la configuration de certains registres de commande. Le mode test permet d'évaluer les performances de la mémoire et donne la possibilité d'effectuer certaines caractérisations. Ces véhicules de test permettent l'accès direct à des cellules isolées du plan mémoire à partir de modes de test particuliers, on distingue :

- le mode DMA<sup>59</sup>, il permet l'accès direct aux colonnes du plan mémoire en court-circuitant toute la logique d'adressage et autorise la lecture du courant traversant la cellule sélectionnée. Ce mode peut être utilisé pour extraire les caractéristiques  $I_d(V_{gc})$  et  $I_d(V_{ds})$  du point mémoire,

<sup>59</sup> Direct Memory Access



- 
- le mode CDMA<sup>60</sup> il s'agit d'un mode de lecture rapide de la mémoire par comparaison entre le courant traversant la cellule mémoire et un courant de référence.

Certaines mémoires permettent aussi de placer les lignes et colonnes du plan mémoire dans des conditions d'utilisation critiques (ou condition de « stress »). Ce stress peut être statique ou dynamique. Ces tests sont réalisés en appliquant une haute tension sur les lignes ou les colonnes pendant un temps donné, de manière à mettre en évidence les lignes ou colonnes dont les cellules mémoire présentent des oxydes défectueux.

## 2 Le test sous pointe (« Probe Test »)

L'extraction des distributions de  $V_T$  sur véhicules passe par le test sous pointe. Le test sous pointe est réalisé par un testeur pilotant une machine de test spécifique (appelé « wafer probe » ou encore « prober »). Le testeur est alors associé à un manipulateur qui va prendre en charge les tranches de silicium et le positionnement des pointes devant les contacts. Le test sous pointe est très souvent lié aux unités de fabrication de tranches pour des raisons de propreté d'environnement et de retour d'information rapide en cas de problème.

Cette étape de test est très délicate et nécessite la réalisation de cartes à pointes qui se posent sur les plots de contact du circuit à tester. Ces pointes sont le plus souvent en tungstène. Certains testeurs de taille importante utilisent des cartes à plus de 700 pointes, avec un pas d'espacement de 100 micromètres. Le but étant de tester un maximum de puces en parallèle.

La montée en puissance de la complexité des circuits augmente la difficulté du test, on pratique le plus souvent des tests simplifiés. Par exemple, pour un circuit de MNV, on exécutera une série de tests qui comprend un ensemble des vecteurs de test optimisés pour ne pas impacter le temps de test, plus la mesure des courants de consommation et de fuites sur les entrées. Des tests paramétriques plus poussés se feront au niveau de la puce en boîtier. Toutes les informations relatives au test sont sauvegardées sous un format spécifique à chaque testeur. Les résultats des tests (données logiques et analogiques, « binning », etc.) fournissent de précieuses informations aux ingénieurs en charge du test et des produits, qui peuvent les exploiter afin d'améliorer le rendement.

Le vieillissement est couramment pratiqué en microélectronique afin d'évaluer la fiabilité d'un composant à long terme. Il consiste en général à soumettre le circuit ou la puce à une forte température durant un temps donné, avant de le soumettre à de nouveaux tests paramétriques et logiques. Cela est dû à la célèbre courbe en forme de « baignoire », représentant le taux de défaillance en fonction du temps de fonctionnement du système. Le type de test exécuté ici dépend de la catégorie du composant.

Pour le cas des circuits MNV, une évaluation de la fiabilité du composant à long terme se fait en couplant les tests en température (rétention) et les tests de stress électrique (cyclage). Pour des produits dits à haute fiabilité (militaires ou spatiaux), on peut être amené à faire des corrélations de mesure entre les tests avant et après vieillissement (analyse de dérive). Le but recherché est alors

---

<sup>60</sup> Current Direct Memory Access

d'assurer que le circuit a atteint une stabilité suffisante. Pour apprécier ces dérives, il faut enregistrer des résultats de mesure et non plus seulement vérifier une fonctionnalité ou une caractéristique par rapport à une limite fixée. Les résultats sont enregistrés dans des fichiers qui serviront à des mesures de dérives ainsi qu'aux corrélations voulues.

### III. Fiabilité de la circuiterie logique

#### A. Technologie CMOS et BEOL<sup>61</sup>

Le test de circuits numériques avancés pose de nombreux problèmes :

- L'intégration devenant de plus en plus fine, la complexité et le temps de test augmentent de façon exponentielle et le coût du test également. Il pourrait dépasser le prix du silicium dans un proche avenir,
- L'augmentation des fréquences d'horloges permet à des puces de fonctionner à des vitesses que les testeurs conventionnels ne peuvent atteindre. On ne peut pas alors tester ces circuits à vitesse nominale. Cela est d'autant plus gênant qu'à ces vitesses, des parasites peuvent intervenir à cause des connexions entre la puce et les équipements de test,
- Enfin, les délais deviennent suffisamment importants par rapport aux cycles d'horloge pour engendrer des défaillances à haute vitesse. Le test à vitesse nominale est donc aujourd'hui indispensable.

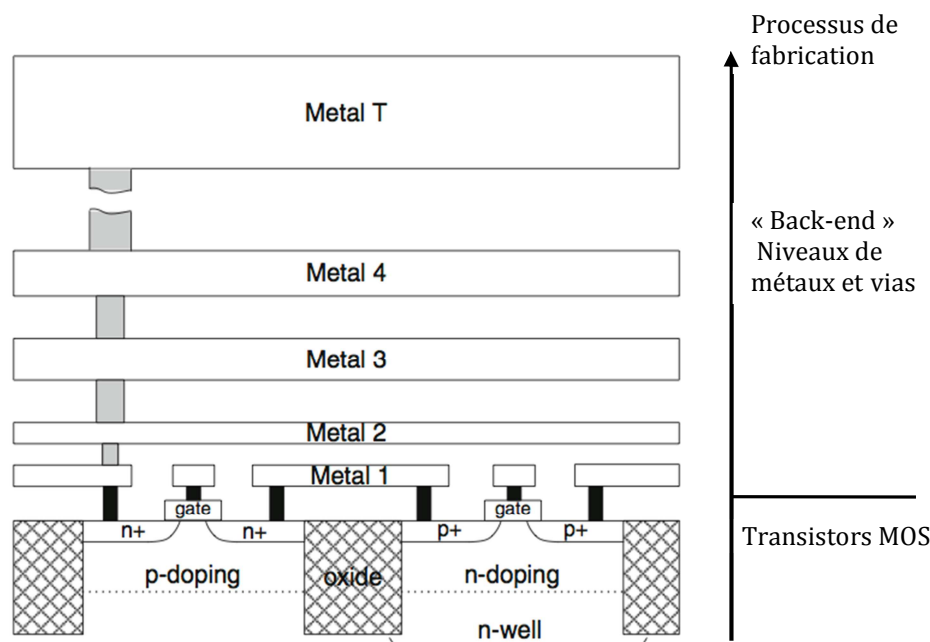


Fig. I-19 : Vue en coupe schématique d'un circuit CMOS comprenant « T » niveaux de métaux  
 Dans le cas des technologies CMOS actuellement utilisées, il s'avère qu'une très

<sup>61</sup> « Back-End Of Line » : étapes de fabrication liées aux interconnexions

---

large majorité des défauts induits par le processus de fabrication sont des défauts liés aux couches d'interconnexions. La Fig. I-19 montre que le BEOL comprend les niveaux de métaux et la gravure des contacts et des vias. Cette étape comprend aussi le dépôt des couches diélectriques de manière à isoler les composants et les couches de métaux entre elles et la passivation.

## **B. Evaluation de la fiabilité des circuits numériques**

Les défauts dans les couches d'interconnexion peuvent être détectés à partir de structures de test appropriées. De nombreuses structures ont été développées, leur but est de relier les paramètres mesurés aux paramètres intrinsèques du processus technologique. Une structure très populaire parmi les fondeurs consiste en un oscillateur en anneau. Un oscillateur en anneau est réalisé par un nombre impair d'inverseurs rebouclés sur eux-mêmes. Suivant sa conception, sa fréquence d'oscillation peut dépendre des paramètres caractéristiques de la technologie et rendre compte de la qualité du processus de fabrication.

La détection des défauts dans les couches d'interconnexion ne se limite pas à la détection des courts-circuits et des circuits ouverts (bien que les statistiques privilégient d'ailleurs l'occurrence des courts-circuits) mais aussi à la détection des variations du procédé de fabrication. Ces variations sont jugées critiques si leur impact entraîne une déviation de certains paramètres électriques hors des spécifications définies par la technologie.

La surface occupée par ces structures ainsi que leurs dimensions sont naturellement fonction de la qualité du procédé de fabrication (rendement ou nombre de défauts attendus par unité de surface) et des dimensions des défauts que l'on cherche à détecter.

La surveillance directe s'effectue par l'intermédiaire de modules de test spécialement conçus qui sont connus sous le terme TEG ou PCM<sup>62</sup> logiques. Durant le développement d'un nouveau nœud technologique CMOS, des structures de test sont utilisées tant en phase de développement de la technologie que pendant le processus d'optimisation de la technologie. Durant ce type de transition technologique, les structures de test jouent un rôle important de diagnostic de défauts durant la période critique de montée du rendement de fabrication. Des données précieuses, issues de ces structures de test et non disponibles sur le produit, permettent un contrôle précis du procédé de fabrication. En fonction de ces données, une rétroaction sur les paramètres du procédé de fabrication peut permettre un recentrage de certains paramètres électriques.

D'autres structures sont utilisées et cela, après une caractérisation poussée, de manière à fournir les données nécessaires à l'élaboration des modèles utilisés dans les technologies CMOS.

D'une manière générale, les structures de test accompagnant le produit final facilitent le suivi de la compréhension du rendement de fabrication, de la technologie, aident au « débogage » de certaines fonctionnalités du produit, et permettent de recueillir des données de fiabilité de manière précoce.

---

<sup>62</sup> Process Control Monitoring

## 1 Les TEG logiques

Aux vues des limites des TEG paramétriques, un autre type de TEG a été développé. Ce sont des TEG « logiques » qui ont une architecture complexe plus proche de celle d'un produit et sur lesquelles des mesures dynamiques sont effectuées. Ces architectures sont basées sur des portes logiques, des oscillateurs en anneau ou encore des bascules. De plus ces dernières sont pour la majorité fournies dans les bibliothèques de portes fournies dans le « Designe Kit » (DK). Ces structures sont généralement conçues de manière à être testées rapidement, par des testeurs logiques utilisés en production. De plus, leur dessin de masque est souvent réalisé de telle sorte que, si un défaut est détecté électriquement, il soit facilement localisable pour être analysé physiquement, et ainsi, déterminer l'origine de ce dernier. Ces architectures sont également conçues de façon à pouvoir être insérées dans les lignes de découpes des plaquettes de silicium. Grâce à ces structures, il est par exemple possible d'évaluer l'effet des variations de processus de fabrication sur le délai de transmission des portes logiques, en utilisant des oscillateurs en anneau [Khou005]. Pour cela, les oscillateurs doivent être configurables dans le but de changer le nombre d'inverseurs constituant l'oscillateur.

## 2 Les Macro-cellules de test

Les TEG logiques, de par leur petite taille, peuvent être insérées dans les lignes de découpes des produits. Cependant, pour la même raison, la probabilité de piéger des défauts est assez faible. Il est alors préférable de développer des structures qui, comme les TEG logiques, sont conçues avec le même genre d'architecture et de complexité que les produits, mais également le même ordre de dimension. Ces structures, appelées macro-cellules de test, sont conçues avec les cellules de base de la bibliothèque qui est associée au DK. Elles permettent de valider les cellules de base, ainsi que leur comportement face aux variations du procédé de fabrication, avant de les utiliser dans les produits. En règle générale, les macro-cellules sont souvent conçues sous forme de matrice ou de plan mémoire, pour pouvoir facilement localiser les défauts et les analyser. En effet, possédant une testabilité élevée, la localisation des défauts sur ces structures est précise. Par conséquent, la phase d'analyses de défaillance sera plus rapide. Les macro-cellules sont testées de la même façon que le produit, et un ATPG<sup>63</sup> peut également être utilisé pour générer les vecteurs de test. Les types de structures les plus utilisées sont les RAM pour analyser la densité de défauts et les oscillateurs en anneau qui sont très sensibles aux variations du procédé de fabrication.

Un exemple de circuit utilisant des RAM est décrit par [Fuji07]. Cette structure est composée de cellules FlipFlop-RAM (FF-RAM), assemblées en un plan mémoire de 1 Mbits, ainsi que de structures de défektivité pour le BEOL. Ce plan mémoire a été conçu en utilisant des cellules de la CORELIB ainsi qu'un outil de routage automatique pour se rapprocher le plus possible d'un produit. La Fig. I-20a donne le schéma structurel de la cellule FF-RAM comportant une bascule D avec une porte NAND de validation de l'horloge et un inverseur 3-états pour la

---

<sup>63</sup> Automatic Test Pattern Generator

lecture de la donnée. Les structures de défektivité pour le BEOL sont des peignes, des serpentins, des chaînes de vias empilés et des structures d'interconnexions métalliques, générées aléatoirement grâce à des logiciels de routage. Dans le but d'optimiser la surface de silicium occupée, ces structures sont positionnées au-dessus des cellules FF-RAM et couvrent les niveaux d'interconnexions qui ne sont pas utilisés par les points mémoire (Fig. I-20b).

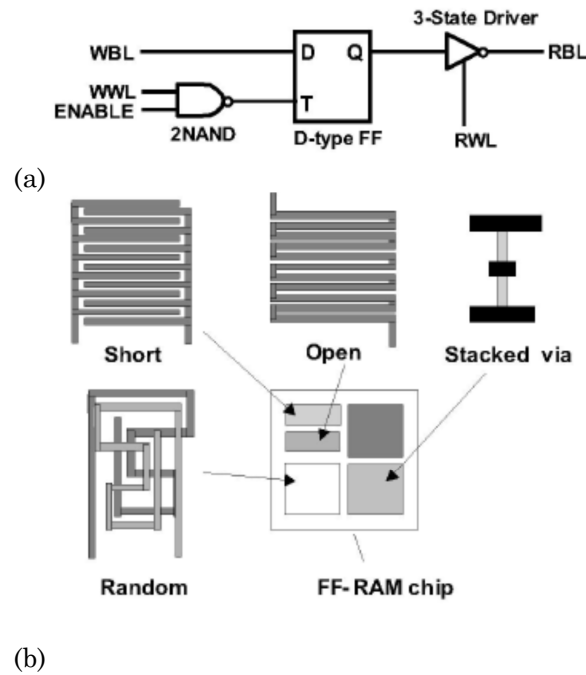


Fig. I-20 : Schéma d'une cellule FF-RAM (a) et structure de défektivité pour le BEOL (b)

Ensuite, lors du test, un jeu de multiplexeurs permet d'évaluer l'intégrité, soit du point mémoire, soit des différentes structures de défektivité situées au-dessus de ce dernier, comme le montre la Fig. I-21. Grâce à ce système d'adressage, l'intégralité du circuit peut être testée avec un testeur conventionnel et en utilisant les outils de génération de vecteur de test. De plus, cette configuration permet de faciliter les analyses de défaillance en fournissant une localisation de type bitmap<sup>64</sup> en X:Y mais également en Z (niveau de métallisation).

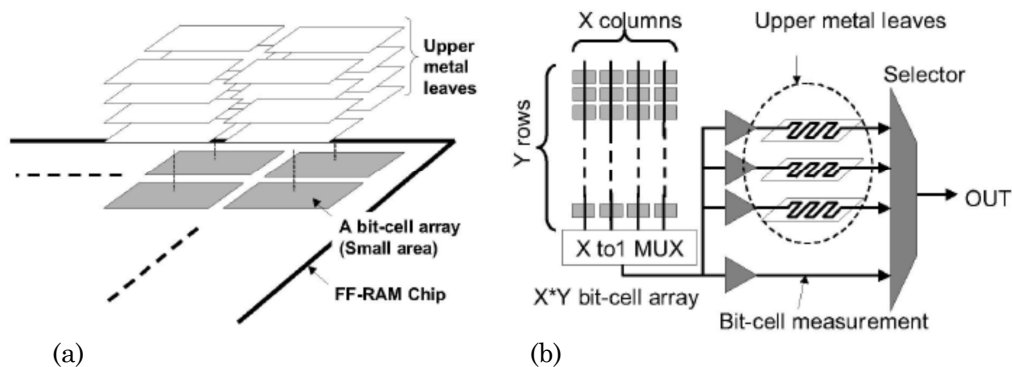


Fig. I-21 : Superposition des structures de défektivité (a) et adressage des niveaux (b)

<sup>64</sup> Bitmap : outil donnant une représentation topographique (2D ou 3D) des défauts dans un plan matriciel

### 3 Les TEG hybrides

Les TEG hybrides sont des structures qui fournissent des données paramétriques pour la caractérisation du procédé de fabrication, sur lesquelles, de la logique est ajoutée pour faciliter la communication avec l'extérieur et la corrélation avec le produit. Ces structures sont à la fois capables de faire des mesures analogiques, pour extraire des paramètres du processus de fabrication, mais embarquent également de la logique. Le but étant de rendre les informations extraites rapidement et facilement, lisibles par l'équipement de test. Ceci est rendu possible par les technologies submicroniques qui permettent d'intégrer de plus en plus de portes logiques dans la hauteur des lignes de découpes. De bons exemples de ces structures sont donnés par [Lafr06] et [Agar07].

[Lafr06] présente une structure qui permet de mesurer le délai induit par un DUT<sup>65</sup> au milieu d'un oscillateur en anneau. Ainsi, le délai d'une ligne, d'un transistor ou de tout autre circuit actif, peut être modélisé par un circuit RC classique (résistance série, et capacité en parallèle). Dans le but de déterminer l'effet des variations process, le buffer d'entrée possède une impédance variable. Ensuite, la fréquence d'oscillation est mesurée avec le DUT inséré dans la boucle de l'oscillateur, puis une seconde fois en le court-circuitant. Cette mesure de fréquence est effectuée à l'aide d'un compteur asynchrone de 20 bits, qui compte le nombre de périodes d'oscillation durant une durée  $T$  fixée. Le schéma de principe de cette structure est donné Fig. I-22. Cette structure est ensuite simulée avec différentes valeurs de  $R_d$  et  $C_d$ , dans le but de créer un abaque. Puis, lors du test sur silicium, l'abaque est interpolé pour caractériser les valeurs de  $R_d$  et  $C_d$  en fonction des fréquences mesurées. Cette étude permet également, de mettre en avant le fait que les capacités influent beaucoup plus que les résistances, sur la variation de délai de transmission.

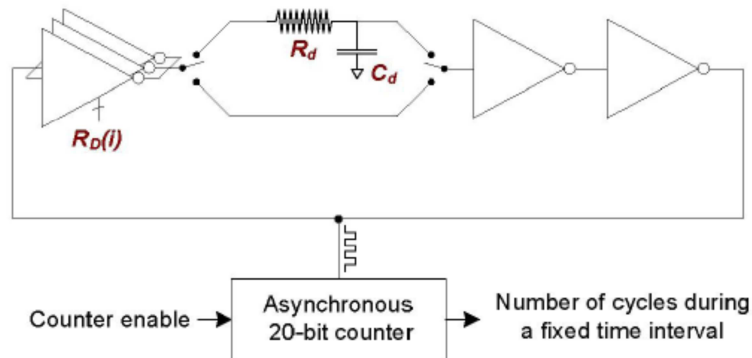
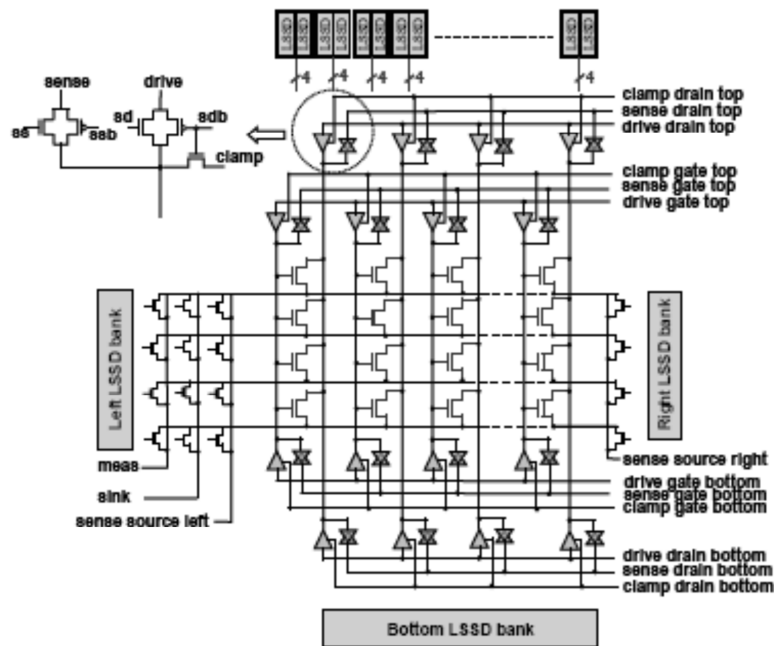


Fig. I-22 : Schéma de principe pour la modélisation de délai

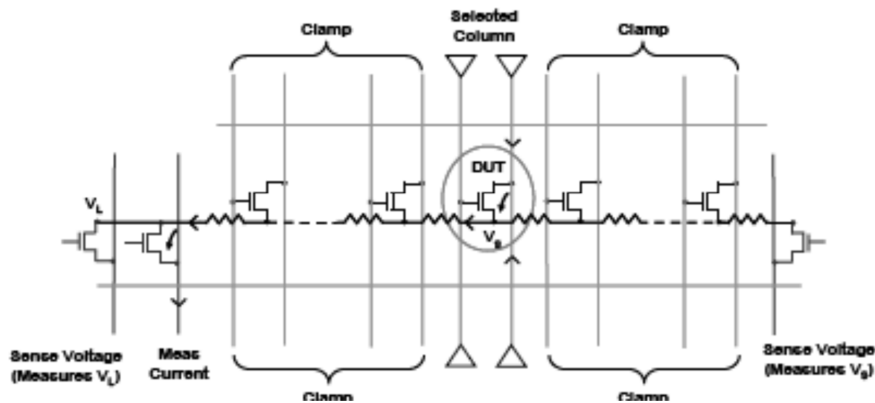
[Agar07] quant à lui, présente une structure permettant de mesurer précisément les tensions de seuil  $V_T$  d'un plan de 96000 transistors. Ces derniers sont organisés en matrice de 1000 lignes et 96 colonnes, l'adressage étant effectué par activation de la ligne et de la colonne correspondante. La Fig. I-23a met en avant la circuiterie utilisée pour appliquer et mesurer les différents signaux. Cette circuiterie est composée d'un transistor pour la mesure du courant, un pour l'écoulement du courant et un autre de chaque côté de la matrice pour la mesure

<sup>65</sup> Device Under Test : cellule à tester

des tensions de source. Se trouvent également en haut et en bas de la matrice, des circuits de mesure et de positionnement pour les tensions de grille et de drain.



(a)



(b)

Fig. I-23: Schéma de principe pour l'adressage d'un transistor (a) et évaluation de l'IR drop (b)

Une méthode d'évaluation de l'IR-drop<sup>66</sup> par mesure « quatre pointes » est aussi proposée. La Fig. I-23b montre que les drains sont forcés par le haut et mesurés par le bas. De même, des circuits de mesure sont ajoutés pour mesurer les tensions de la ligne à droite et à gauche de la matrice. Tout le courant est dirigé vers la gauche, ainsi la tension sur le circuit de mesure de droite est la tension exacte de la source. La différence entre la tension de gauche et de droite peut être utilisée pour évaluer l'IR-drop. De plus, pour réduire les fuites des cellules non sélectionnées, des clamps sont ajoutés à la matrice. Lors d'une mesure, la grille et le drain des cellules non sélectionnées sont mises à leur tension de clamp. Par exemple, une tension négative peut être appliquée sur la grille pour réduire les

<sup>66</sup> IR-drop : chute de tension due au passage du courant dans une ligne résistive

fuites de courant. Le courant du DUT est dirigé vers le circuit de mesure et le courant des autres lignes vers le circuit d'écoulement. Cette technique est utile pour mesurer les fuites lorsque le DUT est polarisé avec une faible tension de grille. Les cellules d'écoulement du courant sont réalisées avec un oxyde épais pour réduire le courant de fuite à travers la grille.

Cette technique présente donc une méthode très rapide pour l'évaluation du  $V_T$ . En effet, pour extraire ce paramètre, il faut en général tracer une courbe  $I_{DS}(V_{GS})$  et donc, faire un certain nombre de mesures pour un seul transistor. La méthode proposée est d'exploiter le fait que le courant  $I_{DS}$  est une fonction de  $(V_{GS} - V_T)$  et non de  $V_{GS}$  et  $V_T$  séparément. C'est pourquoi, si le même courant  $I_{DS}$  et la même tension  $V_{DS}$  sont forcés sur différents transistors, alors chacun devrait avoir le même  $V_{GS} - V_T$  quelle que soit la valeur absolue du  $V_T$ . Par conséquent, si une fluctuation de dopant cause une variation de  $V_T$  d'une certaine valeur, alors le  $V_{GS}$  aura la même variation pour garder une valeur de  $V_{GS} - V_T$  constante. De ce fait, la variation de  $V_T$  peut être évaluée, en mesurant simplement la variation de  $V_{GS}$ .

## IV. Fiabilité des blocs analogique

### A. Rappels sur la conception analogique

#### 1 Les blocs de base

Parmi les blocs de base de l'électronique intégrée analogique, on distingue les miroirs de courant, les paires différentielles et différents types de charges. Les miroirs de courant permettent de copier des courants mais également de réaliser des fonctions simples des courants, comme des additions, des soustractions et des valeurs absolues. Ils sont utilisés de manière extensive pour polariser les blocs fonctionnels analogiques. Ils utilisent le principe de similitude.

Les paires différentielles sont constituées de deux transistors identiques qui fonctionnent en aiguillage de courant. Ce circuit constitue l'étage de base d'un amplificateur différentiel, tel que beaucoup d'amplificateurs audio, et la majorité des amplificateurs opérationnels.

Quant aux charges, elles génèrent une tension à partir d'un courant. La résistance est une charge linéaire. Sa résistivité (en première approximation) ne dépend pas de sa tension. Les transistors MOS connectés en diode ou en source de courant fournissent un comportement non linéaire.

#### 2 Les blocs fonctionnels

Les blocs fonctionnels sont des assemblages d'éléments permettant de réaliser une fonction. Les principaux blocs fonctionnels analogiques, qu'on retrouve dans la plus grande partie des circuits intégrés, même digitaux, sont les amplificateurs, les comparateurs, les sources de courant/tension et les oscillateurs. De manière à répondre à la forte demande des systèmes embarqués, les concepteurs polarisent leurs structures analogiques avec des niveaux de



---

courant de plus en plus faibles. On parle alors d'applications basses consommation ou basse puissance (« low power » en anglais voire « ultra low power » pour les applications médicales). Ces structures à très faible consommation doivent souvent être polarisées dans des régimes de fonctionnement non optimaux pour les composants. Ainsi, les transistors MOS peuvent être polarisés sous le seuil ou en régime d'inversion faible [Vitt09].

De plus, pour réaliser des applications « ultra-faible consommation », un moyen simple est de réduire les tensions d'alimentation en fonction de chaque nœud technologique. Cela se traduit, pour les blocs analogiques, par des conditions de polarisations spécifiques dites sous le seuil, avec des courants de l'ordre de la centaine de nano ampère par branche.

### 3 Variabilité inter dispositifs (« matching »)

L'intégration sur un même circuit d'un nombre important de transistors MOS aux caractéristiques très différentes entraîne des difficultés accrues pour maîtriser la fonctionnalité des blocs de base tels les miroirs de courant et de paires différentielles. Ces structures élémentaires utilisées pour définir les étages d'adaptation d'impédance ainsi que les étages d'amplification sont sensibles aux fluctuations du procédé de fabrication. Les miroirs de courant et les paires différentielles sont composés de transistors appariés. Ainsi, la minimisation des variations entre les transistors est donc une des clés du bon fonctionnement de ce type de structure.

L'étude des variations inter dispositifs fait appel aux termes d'appariement (« matching ») ou désappariement (« mismatch ») qui permettent de quantifier respectivement la similitude ou la différence entre deux dispositifs. Les variations locales entre les deux transistors MOS appariés peuvent générer des erreurs de copie du courant d'entrée dans le cas du miroir de courant ou introduire une tension de décalage (« offset ») en entrée de la paire différentielle.

La mesure des fluctuations entre deux transistors requiert des équipements de mesure précis capables d'évaluer une différence de paramètres électriques, des structures de test particulières et une méthode d'extraction de ces paramètres adaptée à l'étude choisie. Un nombre conséquent de mesures est nécessaire suivi d'un traitement statistique permettant de conclure sur l'appariement des dispositifs.

## **B. Fiabilité des blocs analogiques**

### 1 Test des blocs analogiques

Bien que les parties analogiques ne représentent qu'une petite partie d'un SoC en terme de surface, son coût de test est très important. En effet, le test analogique nécessite des testeurs analogiques capables de générer des signaux analogiques avec une très grande précision et sont donc extrêmement coûteux.

De plus, l'analyse de la réponse est beaucoup plus complexe dans le domaine analogique que dans le domaine numérique. En effet, dans le cas numérique, la signature est bonne ou fautive, et il suffit de la comparer à la signature théorique pour voir s'il y a ou non une défaillance. Dans le cas d'un circuit analogique, la

sortie est toujours sujette à des fluctuations. Le circuit va être refusé si ses caractéristiques s'éloignent trop des caractéristiques voulues. A la notion de fautes (implicitement discrètes), il faut ajouter une notion de variation de paramètres, à celle de détection une notion de tolérance. Pourtant, un standard de test analogique existe déjà : il s'agit de la norme IEEE 1149.4 (inspirée de la norme IEEE 1149.1).

## 2 Utilisation des TEG pour l'analyse du « matching »

Comme évoqué précédemment, les motifs de test (TEG) sont des éléments qui sont introduits dans la plaquette de silicium par le fondeur. Ils sont placés sur les lignes de découpe de la plaquette de silicium. Les tests effectués sur ces motifs sont principalement de type analogique. Les motifs sont constitués d'éléments représentatifs du procédé de fabrication (transistors, résistances, capacités, etc.). Les mesures effectuées sur ces éléments vont permettre de mesurer des paramètres technologiques tout au long du processus de fabrication et vérifier que tous se situent bien dans les marges nominales.

La caractérisation du « mismatch » aléatoire requière des TEG particulières avec des éléments placés proches les uns des autres, et où chaque élément peut être mesuré au niveau transistor. En effet, des structures comme les oscillateurs en anneau ne peuvent pas être utilisées, car ils masquent les variations aléatoires en moyennant plusieurs éléments. De plus, un large nombre d'éléments devrait être mesuré pour obtenir une statistique correcte dans le cadre d'une prédiction de variance.

## V. Conclusion

Le test des microcontrôleurs passe par le test de chacun de ses éléments. De manière obtenir une caractérisation précise des sous-systèmes constitutifs du microcontrôleur, des structures de test spécifiques accompagnent le circuit, notamment durant la phase d'introduction de la technologie.

Ces structures de test peuvent être élémentaires et insérées dans les lignes de découpe de la plaquette de silicium ou disposer d'une architecture plus complexe. Ces dernières présentent une architecture proche de celle des produits. La différence réside dans leur capacité à embarquer des modes de test spécifiques afin de pouvoir déterminer les limites de fonctionnement du circuit dans tous ses modes opératoires.

Au niveau technologique, la fabrication d'un microcontrôleur implique que l'on soit capable d'intégrer sur un même substrat les technologies suivantes : CMOS avancée, blocs analogiques, mémoires FLASH/EEPROM.

La principale cause de perte de rendement des microcontrôleurs est liée à la partie MNV. En effet, la structure et le principe de fonctionnement spécifique des MNV en font un composant particulièrement sensible.

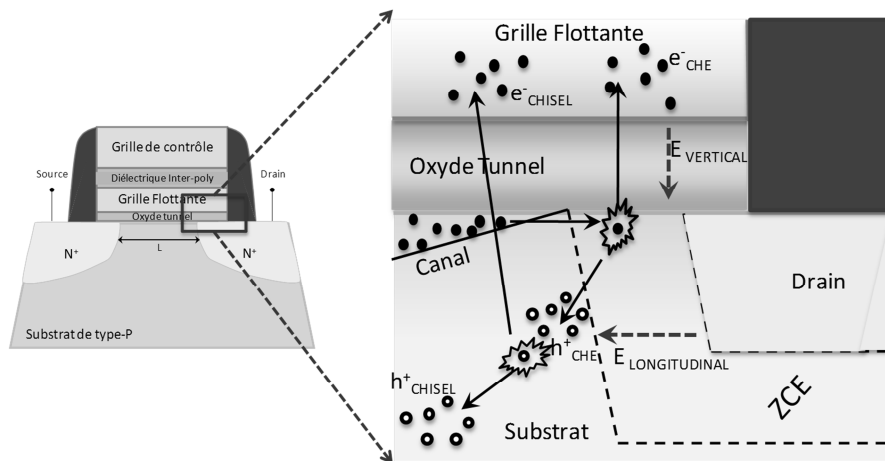
Concernant la partie numérique, les défauts dans les couches d'interconnexion sont généralement mis en cause. Quant à la partie analogique, les variations inter dispositifs (désignées sous le terme termes d'appariement ou « matching ») ne sont plus à négliger.





## Chapitre II : Fiabilité des Mémoires Non Volatiles

L'évaluation de la fiabilité des mémoires non volatiles est réalisée à partir de véhicules de test. L'architecture interne et les différents modes de fonctionnement de véhicule de test TCF9 sont présentés. Le flot de test du TCF9 est ensuite détaillé. En complément de ce flot standard, une description approfondie des tests de fiabilités (tests d'endurance et de rétention) est donnée en se basant sur des résultats expérimentaux. Enfin, le TCF9 est utilisé pour mener à bien deux études : la première consiste à extraire la densité de pièges de l'oxyde tunnel, la deuxième a permis d'établir une corrélation entre les tests de stress électrique et les tests de rétention. Le TCF9, de par sa taille (700kbits) et ses fonctionnalités, s'est avéré un outil incontournable pour mener à bien ces deux expériences.



<b>I.</b>	<b>LE VEHICULE DE TEST TCF9.....</b>	<b>54</b>
<b>A.</b>	<b>PRESENTATION .....</b>	<b>54</b>
1	Description générale.....	54
2	Architecture.....	55
3	Modes de fonctionnement.....	55
<b>B.</b>	<b>LE FLOT DE TEST TCF9.....</b>	<b>56</b>
1	Présentation .....	56
a.	<i>Les tests de continuité et de court-circuit (« Open/short »)</i> .....	57
b.	<i>Les tests de validation électrique</i> .....	58
c.	<i>Les Tests fonctionnels</i> .....	58
2	Analyse des résultats de Test.....	59
a.	<i>« Bitmap » logiques et analogiques</i> .....	59
b.	<i>Distributions de <math>V_T</math></i> .....	60
c.	<i>Distributions en <math>\Delta V_T</math></i> .....	61
3	Fiabilité des oxydes : résultats de tests d'endurance et de rétention .....	61
a.	<i>L'oxyde inter-polysilicium</i> .....	62
b.	<i>L'oxyde tunnel</i> .....	63
<b>II.</b>	<b>UTILISATION DU TCF9 POUR L'EXTRACTION DE LA DENSITE DE PIEGES DE L'OXYDE TUNNEL.....</b>	<b>63</b>
<b>A.</b>	<b>LE SILC : PRESENTATION .....</b>	<b>63</b>
1	Origine physique du SILC.....	63
2	Modélisation du SILC.....	64
3	Méthodes de caractérisation du SILC.....	65
<b>B.</b>	<b>EXTRACTION DE LA DENSITE DE PIEGES SUR VEHICULE DE TEST.....</b>	<b>67</b>
1	Dispositif expérimental .....	67
2	Résultats expérimentaux .....	68
a.	<i>Dégradation des distributions des tensions de seuil</i> .....	68
b.	<i>Technique de la cellule équivalente</i> .....	69
3	Méthode d'extraction de la densité de piège des cellules EEPROM .....	69
a.	<i>Modèle de cellule EEPROM</i> .....	69
b.	<i>Extraction de la densité de pièges <math>N_{it}</math></i> .....	70
c.	<i>Evolution du paramètre <math>N_{it}</math> en fonction du cyclage</i> .....	71
<b>III.</b>	<b>UTILISATION DU TCF9 POUR L'ETUDE DE LA CORRELATION ENTRE TESTS DE STRESS ET TESTS DE RETENTION .....</b>	<b>71</b>
<b>A.</b>	<b>CORRELATION ENTRE TESTS DE STRESS ET TESTS DE RETENTION.....</b>	<b>71</b>
1	Enjeux du stress électrique sur le temps de test .....	71
2	Dispositif expérimental .....	72
a.	<i>Test en rétention</i> .....	72
b.	<i>Test en stress</i> .....	72
3	Modèle de corrélation entre temps de stress et temps de rétention.....	73
<b>B.</b>	<b>COMPARAISON DES MECANISMES DE PERTES A CHAMP FORT ET CHAMP FAIBLE .....</b>	<b>75</b>
<b>IV.</b>	<b>CONCLUSION.....</b>	<b>76</b>

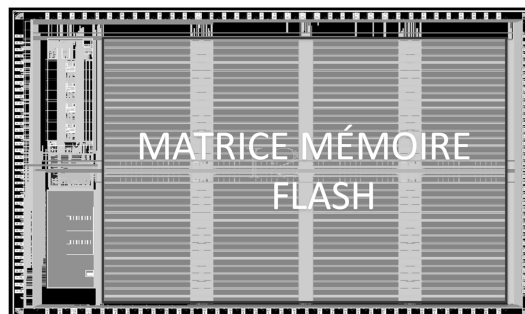
# I. Le véhicule de test TCF9

## A. Présentation

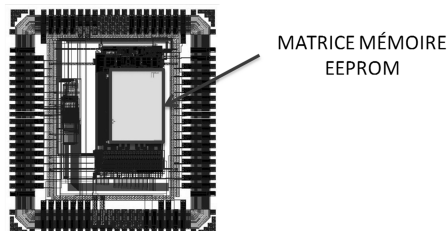
Comme nous l'avons vu au Chapitre I, un véhicule de test (« Test Chip » en anglais) est un composant conçu pour qualifier le procédé de fabrication d'une technologie. Généralement utilisé en recherche et développement, il comprend toutes les principales fonctionnalités que le produit final sera capable d'exécuter. La différence réside dans sa capacité à embarquer des modes de test spécifiques ayant pour but de mettre à l'épreuve la globalité du système, afin de pouvoir déterminer les limites de fonctionnement du circuit dans tous ses modes opératoires. Deux véhicules de test à base de MNV ont été étudiés :

- La véhicule de test « ANNA » qui embarque un microcontrôleur et 128Mbits de mémoire FLASH (Fig. II-1a),
- La véhicule de test « TCF9 », qui embarque 700kbits de mémoire EEPROM (Fig. II-1b).

Les principaux résultats expérimentaux présentés dans ce chapitre sont liés au véhicule de Test « TCF9 ».



a) ANNA : macro-cellule embarquant 128Mbits de FLASH NOR



b) TCF9 : macro-cellule embarquant 700Kbits d'EEPROM

Fig. II-1 : Véhicules de test dédiés aux mémoires EEPROM (a) et aux mémoires Flash (b)

### 1 Description générale

La véhicule de test TCF9 (« Test Chip » pour la technologie nommée « F9 ») est composé d'une matrice de cellules EEPROM et d'une interface logique permettant d'appliquer au plan mémoire des modes de test particuliers, afin d'évaluer ses performances ainsi que la fiabilité de la technologie.

Les caractéristiques principales du véhicule de test TCF9 sont les suivantes :

- Capacité de stockage de 18 Kbits avec des mots de 38 bits (32+6 ECC<sup>67</sup>),
- Programmation par mots ou par page,
- Effacement et programmation en un seul cycle,
- Temps d'effacement et de programmation inférieurs à 2 ms,
- Programmation de type Flash (plusieurs blocs en même temps),
- Consommation réduite au repos (mode « stand-by »).

La cellule élémentaire EEPROM est fabriquée en technologie « F9 » de chez STMicroelectronics. Le nœud technologique est 0.13µm et la surface de la cellule est de 1.3µm<sup>2</sup>. Le TCF9 se décline en plusieurs versions de manière à adresser trois différents domaines d'applications :

- Le TCF9 DSA (« Digital Secured Application ») pour des applications de type carte à puce,
- Le TCF9 MCD (« MicroControllers Devices ») pour les applications embarquées avec unité de calcul,
- Le TCF9 APG (« Automotive Product Group ») pour des applications dédiées au marché automobile.

Ces trois véhicules de test se distinguent principalement par les paramètres intrinsèques de la cellule. Par exemple pour le DSA, l'oxyde tunnel est plus fin que celui des cellules de la matrice APG.

## 2 Architecture

La matrice de cellules de type NOR EEPROM du TCF9 contient 18\*1024 mots. Un mot étant constitué de 32 bits de données et 6 bits d'ECC. Ce qui donne en tout 700 416 cellules ordonnées en 576 lignes et 1216 colonnes. Chaque ligne contenant 32 mots.

Toute la matrice est pilotée par des blocs indépendants tels que les bascules (« latches ») commandant les grilles de contrôle et des lignes de bit, les décodeurs de lignes et de colonnes. 32 amplificateurs de lecture permettent de lire les données. Le véhicule de test est accessible par le biais d'environ 80 broches externes, les bus d'adresses et de données ont une largeur de 16 bits. Cela implique que la lecture d'un mot de 32 bits s'effectue en deux phases. Il est aussi à noter que le TCF9 contient des registres de contrôle qui permettent de programmer différents modes de fonctionnement et modes de test.

## 3 Modes de fonctionnement

La mise en œuvre du plan mémoire commence par la configuration de registres permettant de choisir le mode de fonctionnement et l'opération à effectuer. Il est possible d'effacer et d'écrire les cellules mémoires de plusieurs manières par le biais des registres de contrôle :

- par mots : cette approche offre la granularité la plus fine,
- par pages : une page correspond aux cellules placées sur la même ligne,
- toute la matrice : on parle de mode « global »,
- par lignes « paires » ou par lignes « impaires »,

---

<sup>67</sup> « Error Correction Code » ou code de correction d'erreurs



- par programmation « flash » synchrone et asynchrone : c'est l'équivalent d'une programmation de type « globale » sans l'étape de sélection des colonnes, ce qui la rend plus rapide.

La configuration du registre de lecture est indispensable pour accéder à ces différents modes. Ce registre permet de fixer le courant de référence délivré par l'amplificateur de lecture. Le TCF9 offre 3 modes de lectures :

- le mode « normal » : qui permet de lire en appliquant une tension de lecture ( $V_{CG}$ ) paramétrable. On s'en sert aussi pour lire classiquement un mot de la matrice,
- le mode « verify erasing » : ce mode est utilisé pour vérifier la validité de l'opération d'effacement du plan mémoire, notamment après un effacement en mode global,
- le mode « strict » : c'est ce mode que qui est utilisé pour valider l'opération d'écriture du plan mémoire.

Les deux derniers modes sont utilisés pour détecter les cellules défailtantes après un effacement ou une écriture. Cette fonctionnalité permet aussi de vérifier si le taux de cellules marginales du plan mémoire est en accord avec les spécifications de la technologie.

Enfin, trois grands modes de test sont possibles sur le TCF9 :

- l'« Analog monitor » qui permet d'avoir accès à des signaux internes,
- le « Margin check » qui permet de faire varier la tension de lecture  $V_{CG}$  par le biais d'une source de tension externe, afin d'extraire les tensions de seuil des cellules mémoires. La Fig. II-2 représente la distribution d'une population de cellules EEPROM effacée en mettant en évidence les cellules marginales,
- le DMA qui permet de mesurer le courant des cellules pour une polarisation de grille donnée. Ce mode peut, par exemple, être utilisé pour extraire les caractéristiques  $I_D(V_{CG})$  des cellules mémoires.

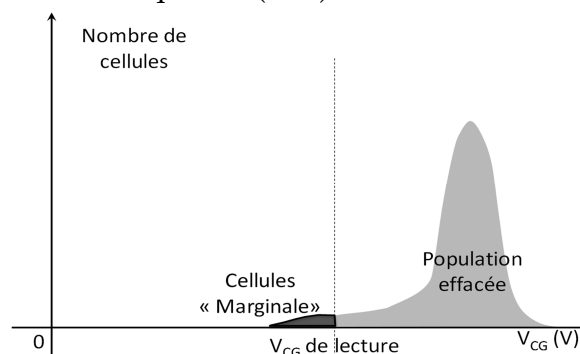


Fig. II-2 : Représentation d'une population de cellules EEPROM effacées mettant en évidence des cellules marginales

## B. Le Flot de Test TCF9

### 1 Présentation

Un flot de test, présenté Fig. II-3 pour le véhicule de test TCF9, est une suite de tests servant à valider le bon fonctionnement de la puce. Généralement la

qualification technologique d'un produit regroupe 4 flots de test EWS<sup>68</sup>. Ensuite, 2 flots seulement sont utilisés quand le produit est à maturité. Ces tests sont effectués sous pointes et sur la plaquette de silicium avant de mettre la puce en boîtier. A l'issue de chaque test EWS, des codes spécifiques sont inscrits dans certaines parties de la mémoire. Ces codes attestent que la puce a subi le flot de test considéré et renseignent sur le résultat du test.

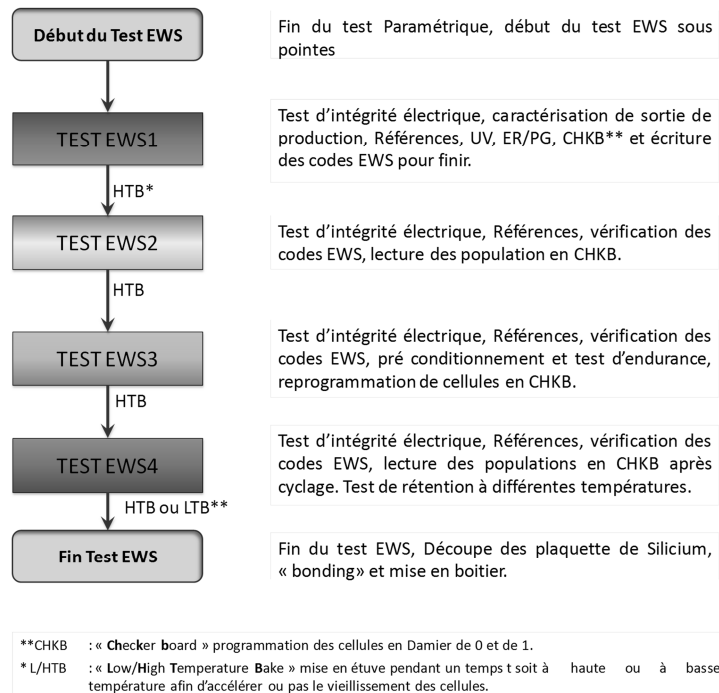


Fig. II-3 : Flots de Tests EWS utilisés pour une qualification technologique

La Fig. II-4 donne un aperçu du flot de test EWS1 servant à qualifier une technologie de type EEPROM. A l'issue de chaque test élémentaire, une vérification est effectuée pour qualifier le succès du test. Le flot de test est divisible en trois grandes familles de séquences de tests :

- Les tests « d'Open/Short » ou tests de continuité et de court-circuit,
- Les tests de validation électrique,
- Les tests fonctionnels.

Un rejet à l'issue d'un test élémentaire est traduit par un arrêt du test. Une classe de rejet ou « BIN » est un nombre qui représente l'étape du flot de test ayant subi le rejet. Les « BIN » sont classés suivant un ordre précis en fonction du circuit testé.

a. Les tests de continuité et de court-circuit (« Open/short »)

Ils servent à vérifier la continuité des pointes sur les plots de contact du circuit à tester, et à valider qu'il n'y ait pas de court-circuit au niveau des contacts. C'est la première étape du flot de test. C'est une étape importante. En effet, pour les tests sur plaquettes de silicium, l'opération d'alignement des pointes sur les plots du circuit est une étape délicate. Un mauvais alignement peut engendrer une perte de continuité ou un court-circuit entre les plots de contact. Il est souvent

<sup>68</sup> Electrical Wafer Sort : tests électriques sous pointes, effectués sur la plaquette de silicium

impératif de vérifier les traces des pointes sur les pads à l'aide d'une caméra haute précision.

b. Les tests de validation électrique

Ils sont effectués pour valider que la puce soit conforme aux spécifications attendues d'un point de vue électrique. Les tests effectués à ce niveau détectent d'éventuels dysfonctionnements des broches du composant ainsi qu'une consommation anormale. Ensuite, une vérification des registres de contrôle de la puce est réalisée afin de garantir son fonctionnement lors des prochaines phases de test. Enfin, une lecture des cellules UV<sup>69</sup> est réalisée après vérification de certains courants et tensions de référence utilisés lors de phases de programmation et de lecture de la mémoire.

c. Les Tests fonctionnels

Ces tests permettent entre autres de vérifier la fonctionnalité de la mémoire, c'est-à-dire la capacité de mémorisation des cellules. Ces tests comprennent deux étapes : une première étape de programmation va imposer à chaque cellule du plan mémoire un état logique donné (0 ou 1). Ensuite, une deuxième étape de lecture du plan mémoire va vérifier l'intégrité des informations écrites dans la mémoire.

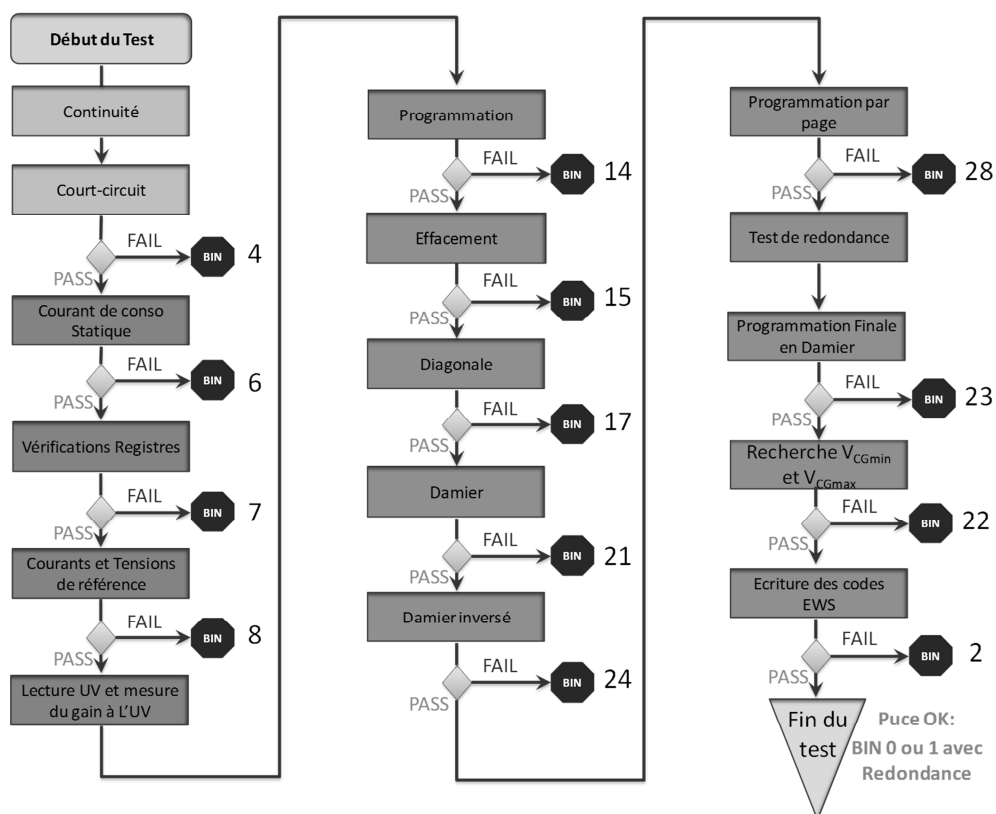


Fig. II-4 : Flot de Test EWS1 simplifié pour mémoires non-volatiles EEPROM

Tous les tests fonctionnels ne seront pas détaillés dans cette partie. Cependant, un exemple de test fonctionnel appelé « damier » (CHKB pour « CHeKerBoard »)

<sup>69</sup> « UltraViolet » : cellules dans leur état vierge, après effacement UV et avant toute opération de programmation

est présenté Fig. II-5 et analysé. Ce test consiste à programmer tout le plan mémoire de manière à réaliser un « damier » constitué de cellules écrites et de cellules effacées. Ce test est ensuite suivi d'un test appelé « damier inversé » (CHKI pour « CHeKer Board Inverse »). Ces tests consistent à écrire des données complémentaires dans des cellules adjacentes. Une défaillance typique révélée lors de ce test est un court-circuit au niveau du polysilicium constituant la grille flottante des cellules mémoires.

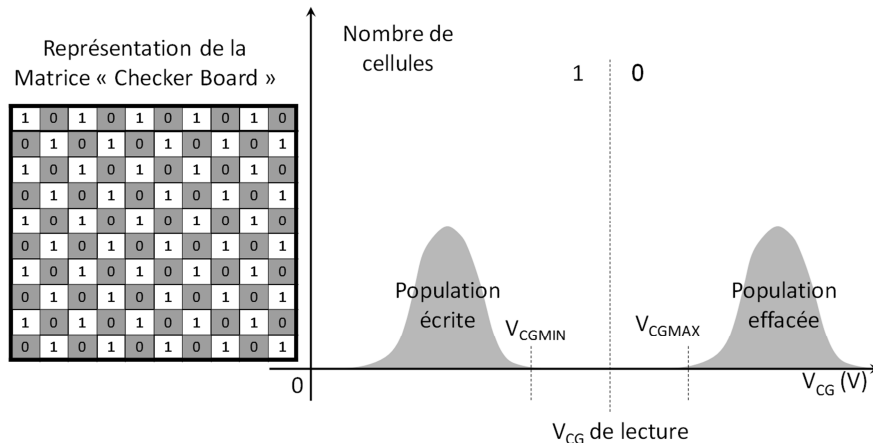


Fig. II-5: Représentation de la matrice EEPROM et de sa population lors d'un test de « Checker Board ».

Cet exemple permet de définir deux paramètres très importants en test de MNV :

- $V_{CGMIN}$ , qui représente la tension de lecture  $V_{CG}$  de la première cellule qui conduit et qui appartient à la population écrite ( $V_{CG}$  décroissant),
- $V_{CGMAX}$ , qui représente la tension de lecture  $V_{CG}$  de la première cellule qui conduit et qui appartient à la population effacée ( $V_{CG}$  croissant).

Ces deux paramètres sont extraits suite au test « programmation finale en damier » et sont très importants pour vérifier l'état des populations notamment de cellules mémoire après des tests d'endurance et de rétention, comme nous le verrons plus loin.

## 2 Analyse des résultats de Test

### a. « Bitmap » logiques et analogiques

En terme de diagnostic, une cartographie de bit logique (« bitmap logique ») est réalisée pour chacun des algorithmes de tests fonctionnels appliqués aux mémoires non volatiles de type EEPROM. Le « bitmap » logique est une représentation topologique des cellules défaillantes d'un plan mémoire. Il s'agit d'une cartographie du plan mémoire en deux dimensions mettant en évidence les cellules défaillantes. Dans le bitmap logique, comme son nom l'indique, seuls les états logiques « Pass » ou « Fail » sont représentés pour chaque cellule du plan mémoire.

Les bitmaps logiques donnent des informations binaires, alors que les mémoires EEPROM présentent un mécanisme de mémorisation de type analogique (décalage de  $V_T$ ). Ainsi, le bitmap analogique rajoute une nouvelle dimension correspondant aux valeurs des  $V_T$  des cellules mémoires tout en conservant la

dimension topologique. La cartographie résultante est une image tridimensionnelle du plan mémoire. La Fig. II-6 est une illustration du bitmap analogique obtenu sur le TCF9. La variation entre les  $V_T$  des différentes cellules du plan mémoire se manifeste par une différence de contraste. On distingue des signatures particulières au niveau des lignes et colonnes de la mémoire, dues dans cet exemple à la non homogénéité de la circuiterie périphérique.

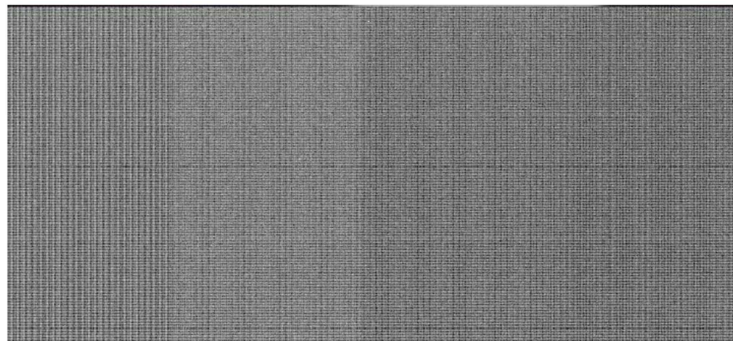


Fig. II-6 : Bitmap analogique permettant de visualiser l'effet de la circuiterie périphérique

#### b. Distributions de $V_T$

Grâce aux différents modes de test embarqués sur le TCF9, il est possible d'adresser chaque cellule de manière indépendante, et ensuite de faire varier la tension de lecture  $V_{CG}$  linéairement. Ainsi, les  $V_T$  de chaque cellule mémoire peuvent être extraites de manière précise.

Une représentation possible d'une distribution peut se faire en quantité cumulée (Fig. II-7). L'unité de la quantité est généralement le pourcentage (%) (Fig. II-7.a) ou les Parties Par Millions (PPM) (Fig. II-7.b). Cette représentation graphique permet d'évaluer les cellules marginales (faible  $V_T$ ) du plan mémoire en mettant en évidence les queues de distributions.

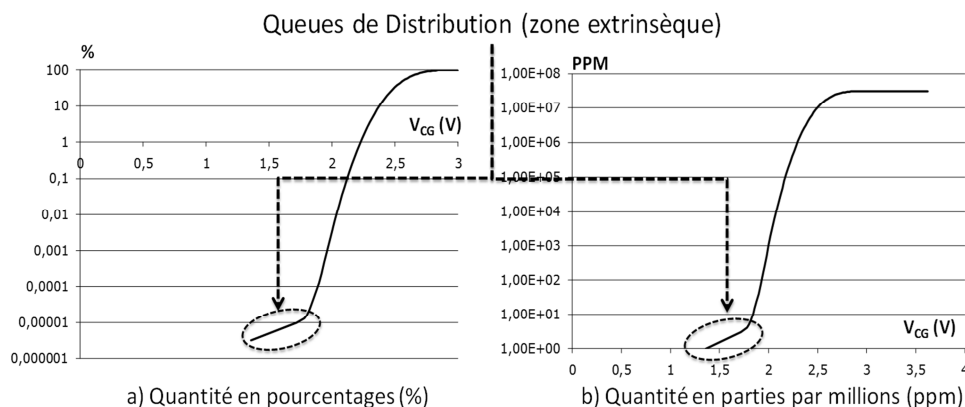


Fig. II-7 : Distributions de population effacée extraites sur le TCF9, en quantité cumulée, en pourcentage (a) et en parties par millions (b)

La queue de distribution est une zone d'intérêt pour l'analyse de fiabilité. Nous verrons par la suite qu'il est possible d'extraire de nombreux paramètres reflétant la fiabilité du dispositif à partir du comportement statistique des queues de distributions. Les cellules faisant partie de la queue de distribution sont les cellules marginales et constituent la zone extrinsèque de la distribution totale. En revanche, les autres cellules constituent la zone intrinsèque de la

distribution. Ces deux appellations sont souvent utilisées lors des tests de fiabilité des circuits mémoires.

### c. Distributions en $\Delta V_T$

$\Delta V_T$  est la différence entre  $V_T(t_0)$ , qui est la tension de seuil relevée à l'instant expérimental ( $t_0$ ), et la tension de seuil  $V_T(t)$ , relevée en général à la suite d'un ou plusieurs tests de fiabilité. Cette évaluation est effectuée pour chaque cellule du plan mémoire. Ce paramètre permet ainsi de discriminer les cellules dont les tensions de seuil ont été le plus dégradées, et cela indépendamment de leur position dans la distribution à  $t_0$ . La Fig. II-8 est une illustration de cette représentation. Cette technique de représentation permet d'aller plus loin dans l'étude de la fiabilité, en se basant sur un critère lié aux pertes de charges des cellules du plan mémoire.

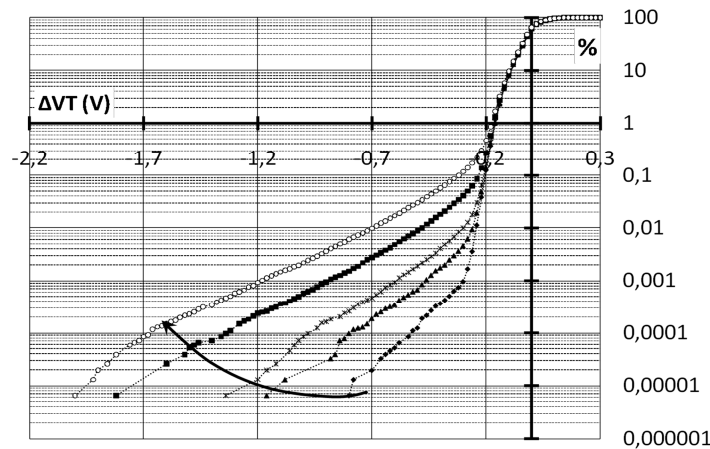


Fig. II-8 : Réseau de distributions en  $\Delta V_T$ , en pourcentage cumulé d'une population EEPROM (TCF9)

## 3 Fiabilité des oxydes : résultats de tests d'endurance et de rétention

Pour rappel, les tests d'endurance et de rétention garantissent respectivement l'aptitude du point mémoire à pouvoir subir un grand nombre de cycles d'écriture-effacement et la capacité de la cellule à maintenir l'information le plus longtemps possible. Ces tests ciblent la fiabilité des oxydes en altérant l'oxyde tunnel (à travers lequel transitent les charges pour être stockées dans la grille flottante). Ils permettent ainsi d'accélérer la fuite des charges de la grille flottante. Ces tests sont réalisés au niveau des flots de test EWS2, EWS3 et EWS4 (cf. Fig. II-3) et s'appliquent aux puces fonctionnelles obtenues à l'issue du flot de test EWS1.

La Fig. II-9 représente l'effet du cyclage (100k cycles) sur la fermeture de la fenêtre de programmation. La population totale prise en compte lors de ce test avoisine 34 millions de cellules EEPROM réparties sur 48 puces différentes. Ce nombre de cellules favorise une étude statistique. Les populations de cellules sont représentées sous la forme de distributions en pourcentage cumulé.

Lors des tests de rétention, les puces sont soumises à des températures comprises entre 25°C (température ambiante) et 250°C. Elles sont alors placées dans des étuves (« Bake ») de manière à simuler un vieillissement du dispositif et déceler des fuites de charges, principalement au travers de l'oxyde tunnel [Sob95] [Kuma01]. La température est un catalyseur des phénomènes de vieillissement des matériaux. Dans de nombreux cas, le test de rétention est précédé par le test

de cyclage de manière à affaiblir les oxydes. En traçant une fois de plus les distributions en pourcentage cumulé, on peut suivre l'évolution de la fenêtre de programmation. La Fig. II-10 représente la dégradation de la fenêtre de programmation après 24h, 168h, 500h, 1000h, 2000h, 3000h, et 6000h de rétention à 125°C.

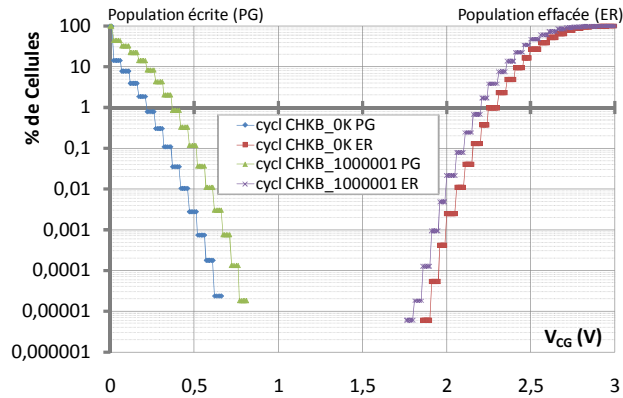


Fig. II-9 : Distributions en pourcentage cumulé, d'une population en « damier » d'un plan mémoire EEPROM, après 100K cycles (TCF9)

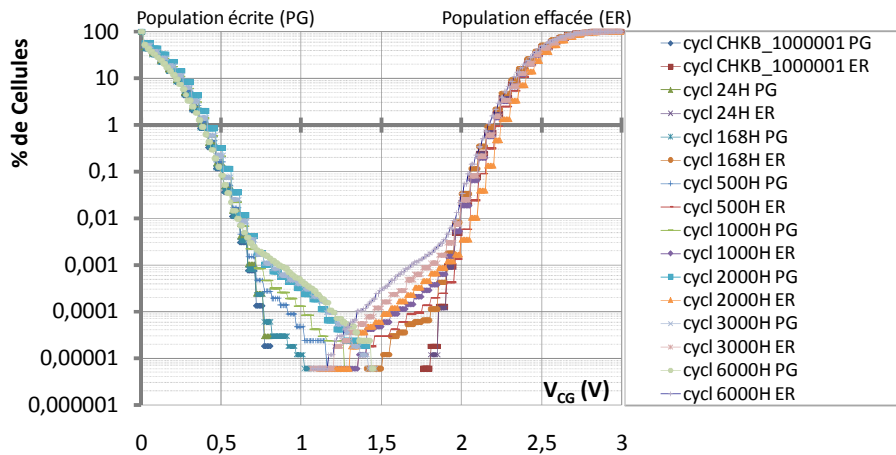


Fig. II-10 : Distributions en pourcentage cumulé, d'une population en « damier » d'un plan mémoire EEPROM après 100K cycles et après 24h, 168h, 500h, 1000h, 2000h, 3000h, et 6000h de rétention

Lors de la mise en œuvre d'une nouvelle technologie, l'élaboration des oxydes représente une des parties les plus sensibles du procédé de fabrication. Au niveau de la cellule mémoire EEPROM, deux oxydes garantissent la fiabilité de la mémoire : l'oxyde tunnel et, à moindre mesure, l'oxyde inter-polysilicium.

#### a. L'oxyde inter-polysilicium

Le diélectrique multicouche Oxyde-Nitruire-Oxyde ou « ONO » ( $\text{SiO}_2\text{-Si}_3\text{N}_4\text{-SiO}_2$ ) fut très tôt introduit dans les mémoires de type EPROM [Mori84] puis largement utilisé dans tous les types de mémoire à grille flottante [Yim03]. Cet oxyde inter-polysilicium qui sépare la grille flottante et la grille de contrôle doit assurer plusieurs fonctions. D'abord, il doit permettre un bon couplage capacitif entre la grille de contrôle et la grille flottante de manière à maintenir une bonne efficacité des phases de programmation. De plus, cet oxyde doit présenter un courant de fuite très faible pour ne pas altérer les performances en rétention. Pour garder

un coefficient de couplage constant d'une génération de cellule à l'autre, l'épaisseur de l'oxyde inter-poly a été réduite (autour de 0.5-0.6 nm). Durant ces dix dernières années, la réduction de cette épaisseur a pu se faire grâce à un amincissement de chacune des couches de l'ONO [Mori91], [Mori96].

#### b. L'oxyde tunnel

Les études réalisées dans ce chapitre se focalisent sur la fiabilité de l'oxyde tunnel. Les effets des courants de SILC<sup>70</sup> sur la fiabilité de la mémoire seront plus particulièrement développés.

Les principales conséquences des dégradations que peut subir l'oxyde tunnel sont : le claquage, le quasi-claquage et le SILC.

Le claquage correspond à une destruction locale de l'oxyde, celui-ci ne joue plus son rôle d'isolant. On observe alors de très forts courants pour des tensions faibles aux bornes de l'oxyde, c'est la dégradation ultime de l'oxyde.

La deuxième conséquence est le quasi-claquage, il se caractérise par une augmentation du courant mais cette fois à faibles champs électriques. A forts champs électriques, l'oxyde garde une caractéristique de type Fowler-Nordheim (FN). C'est une dégradation locale et plusieurs quasi-claquages peuvent se produire avant le claquage final [Okad94], [Depa95], [Brie97].

Le troisième type de dégradation est lié au courant de type SILC. Il est attribué à la création de charges positives dans l'oxyde qui permettent aux électrons de traverser l'oxyde tunnel par un mécanisme tunnel assisté par pièges [Mase82a]. Le SILC est aussi attribué à la génération de charges positives à l'interface Si/SiO<sub>2</sub>, qui produisent un abaissement local de la barrière tunnel [Oliv88].

En effet, dans les MNV à grille flottante, lors des phases de programmation, l'oxyde tunnel est soumis à de forts champs électriques (supérieur à 10MV.cm<sup>-1</sup>). La succession de ces contraintes électriques durant plusieurs cycles de programmation de la cellule provoque des dégradations de l'oxyde tunnel et l'apparition du SILC. Ce phénomène physique constitue la principale limitation à la diminution de l'épaisseur de l'oxyde tunnel.

## II. Utilisation du TCF9 pour l'extraction de la densité de pièges de l'oxyde tunnel

### A. Le SILC : Présentation

#### 1 Origine physique du SILC

La cause principale des pertes de charges durant le test de rétention est le courant de fuite induit par contrainte électrique ou « SILC ». Cette contrainte, subie par l'oxyde tunnel durant les phases d'écriture et d'effacement induit un courant de fuite qui contribue à la perte des charges stockées dans la grille flottante. Un critère de fiabilité pour les mémoires non volatiles est de garantir

---

<sup>70</sup> Stress Induced Leakage Current : mécanisme de courant de fuite induit sous stress



l'intégrité des données sur une durée de dix ans. Pour cela, il est nécessaire de s'assurer que le phénomène de SILC reste suffisamment faible pendant toute la durée de vie de la mémoire.

Il a été montré que l'injection de courant Fowler-Nordheim au travers de couches minces de SiO<sub>2</sub> (5-10 nm) induit des courants de fuite appelés SILC [Moaz92]. Dans le cas des mémoires EEPROM, il est maintenant généralement admis que le SILC est dû à un courant tunnel assisté par piège et que ces pièges sont créés suite aux champs électriques élevés présents au cours du processus d'écriture et d'effacement de la mémoire [Mode99], [Taka99]. A cause de ces pièges, les électrons vont pouvoir transiter à travers l'oxyde par effet tunnel. Lors de leur passage, les électrons vont perdre une partie de leur énergie (autour de 1.7eV). Le phénomène de SILC a été étudié aussi bien sur des capacités MOS [Rose97] que les mémoires EEPROM [Ielm02].

Le SILC est facilement observable lorsque l'on atteint une gamme d'épaisseur d'oxyde SiO<sub>2</sub> inférieure à 100Å. Ce phénomène se caractérise par une augmentation progressive du courant de fuite à travers l'oxyde à bas champ (4 à 7MV.cm<sup>-1</sup>), qui s'ajoute au courant tunnel Fowler-Nordheim obtenu pour des champs plus élevés (champs supérieurs à 9MV.cm<sup>-1</sup>). Le courant de SILC augmente lorsque l'on diminue l'épaisseur d'oxyde dans une plage 100Å-35Å. En dessous de 35Å, ce courant n'est plus visible car le courant tunnel direct devient dominant.

## 2 Modélisation du SILC

Le modèle de SILC que nous présentons ici [Regn07Th] a été un outil fondamental pour permettre l'extraction de la densité de piège dans l'oxyde tunnel. Le modèle de SILC considéré est un modèle de TAT<sup>71</sup> inélastique due à un piège. Le principe de base d'un courant tunnel assisté par pièges est que l'électron ne va pas traverser la barrière tunnel en une seule étape mais transiter par un défaut. L'approche utilisée pour la modélisation du SILC est basée sur le piégeage et dé-piégeage décrit par la statistique de « Schokley Read Hall ». Un piège est caractérisé par sa position spatiale  $X_t$  et un niveau énergétique  $E_t$ . Le courant TAT est constitué par une densité de courant de capture du piège,  $J_{n1}$ , et d'une densité de courant d'émission ou de dé-piégeage,  $J_{n2}$  comme le montre la Fig. II-11.  $J_{n1}$  est la densité de courant de remplissage du piège, dans le cas d'une grille de type N, elle va d'une part dépendre des mêmes paramètres que le courant tunnel classique, c'est-à-dire de la charge surfacique d'électrons à l'interface avec l'oxyde,  $Q_n$ , de la fréquence d'impact sur cette interface,  $f_n$ , et de la transparence tunnel,  $T_{r1}$ , pour atteindre le piège. Puis d'autre part, elle dépend des paramètres du piège c'est-à-dire de la section efficace de capture du piège,  $\sigma_n$ , de la densité de piège,  $N_{it}$ , et de la fonction d'occupation du piège,  $f_{occ}$ . Par conséquent  $J_{n1}$  s'écrit comme suit :

$$J_{n1} = Q_n \cdot f_n \cdot T_{r1}(\zeta_{ox}, E_t, X_t) \cdot \sigma_n \cdot N_{it} (1 - f_{occ}) \quad (II-1)$$

<sup>71</sup> « Trap Assisted Tunneling » ou Effet Tunnel Assisté par Piège (ETAP)

La section efficace de capture du piège,  $\sigma_n$ , correspond à une limitation spatiale de l'impact du défaut. On considère le piège comme une charge positive piégée qui a la forme d'un potentiel coulombien avec une surface de capture fixée à  $10^{-15}$  .cm<sup>-2</sup> comme montré dans [Saka96]. Pour décrire la sortie du piège, l'hypothèse que les électrons sont localisés à un niveau d'énergie  $E_t$  est faite. Ce courant va dépendre de la fonction d'occupation du piège,  $f_{occ}$ , de la transparence tunnel,  $T_{r2}$  et de la fréquence de sortie du piège,  $f_n$ . Ainsi, on obtient l'expression suivante pour  $J_{n2}$  :

$$J_{n_2} = f_n \cdot Tr_2(\zeta_{ox}, E_t, X_t) \cdot \sigma_n \cdot N_{it} \cdot f_{occ} \quad (\text{II-2})$$

Pour déterminer le courant assisté par piège, on considère que le régime permanent est atteint ( $J_{n1} = J_{n2}$ ), c'est-à-dire que l'occupation des pièges ne varie plus en fonction du temps. Ainsi, on calcule le courant tunnel assisté par pièges :

$$J_{TAT} = J_{n1} = Q_n \cdot f_n \cdot Tr_1(\zeta_{ox}, E_t, X_t) \cdot \sigma_n \cdot N_{it} (1 - f_p) \quad (\text{II-3})$$

Dans le cadre de cette étude, le paramètre qui nous intéresse est la densité de piège  $N_{it}$ . C'est en faisant l'extraction expérimentale du courant  $I_{TAT}$  que nous arriverons à extraire cette densité de piège.

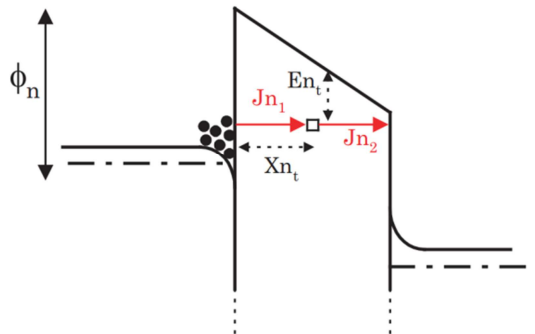


Fig. II-11 : Mécanismes tunnels assistés par un piège situé en  $X_t$  et ayant une énergie  $E_t$ . Le courant TAT est constitué d'un courant de capture  $J_{n1}$  et d'un courant d'émission du piège  $J_{n2}$

### 3 Méthodes de caractérisation du SILC

Dans notre étude, nous considérerons que les grilles flottantes sont dopées par des espèces de type N. La méthode la plus simple pour caractériser les propriétés électriques du SILC, notamment les paramètres des pièges intervenant dans le courant TAT, consiste à mesurer ce courant de fuite sur des capacités de grande taille [DeBl98]. Des caractéristiques  $I(V)$  sont extraites sur une capacité tunnel avec et sans création de pièges dans le volume. Le but est d'obtenir une information sur les propriétés des pièges qui interviennent dans les pertes de charges des cellules mémoires observées à température ambiante. Pour extraire ces paramètres et qu'ils soient cohérents avec les mesures de rétention, une capacité tunnel a été dégradée en la soumettant à 100K cycles d'écriture-effacement. La Fig. II-12 présente les caractéristiques  $I(V)$  d'une capacité tunnel standard avant et après 100K cycles. Lorsque la capacité tunnel est vierge, on

peut reproduire correctement ses caractéristiques en courant par une simulation de type Fowler-Nordheim [Regn06]. Avec la création de pièges par cyclage, on observe l'apparition du courant de SILC à faible champ électrique.

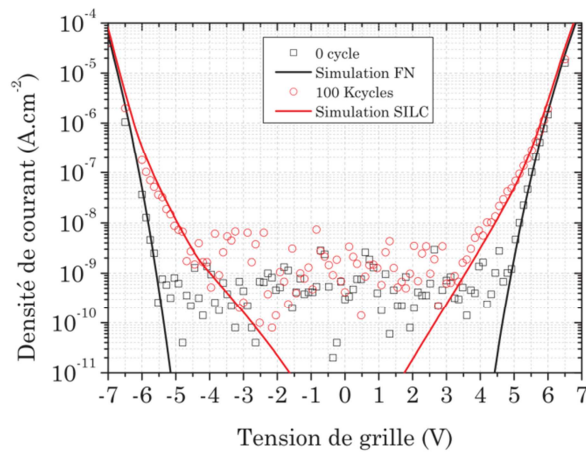


Fig. II-12 : Caractéristique I(V) d'une capacité tunnel de grande surface avant et après 100K cycles d'écriture-effacement [Regn07Th]

Le modèle de SILC utilisé dans cette partie est calibré pour obtenir une bonne corrélation entre mesures et simulations. On extrait les paramètres du piège comme son niveau d'énergie,  $E_t$ , sa position dans l'oxyde,  $X_t$ , puis on évalue la densité de piège  $N_{it}$ . Le Tableau II-1 récapitule les valeurs des paramètres SILC calibrés sur des capacités tunnels. On note que la surface de capture du piège a été fixée à  $1.10^{-15} \text{cm}^{-2}$  pour simplifier l'extraction des différents paramètres. Une information intéressante à mettre en évidence est la position du piège dans l'oxyde. En effet, pour une grille de type N, on note une quasi symétrie des caractéristiques du SILC qui peut être rattachée à une position du piège,  $X_t$ , au centre de l'épaisseur d'oxyde. Pour les deux polarités les porteurs qui constituent le courant de SILC sont les mêmes que ceux qui constituent les courants tunnels de programmation c'est-à-dire les électrons respectifs de la bande de conduction de la grille et du substrat.

Tableau II-1 : Résumé des paramètres du piège pour la simulation des caractéristiques I(V) du SILC d'une capacité à grille de type N

Paramètres	Valeurs
$\sigma_n$	$1.10^{-15} \text{cm}^{-2}$
$E_t$	2.6 eV
$N_{it}$	$4.10^{15} \text{m}^{-2}$
$X_t$	0.5 $t_{\text{tun}}$
$t_{\text{tun}}$	68 Å

La taille de la capacité tunnel étant importante, la valeur de la densité de piège résultante est une moyenne des pièges sur toute la surface de l'oxyde. La densité de piège extraite n'est donc pas comparable à celle que l'on peut trouver sur un oxyde de surface équivalente à celui présent dans une cellule EEPROM. Il est donc fondamental d'extraire la densité de piège sur cellule mémoire. Pour cela des mesures seront effectuées sur le véhicule de test TCF9.

## B. Extraction de la densité de pièges sur véhicule de test

### 1 Dispositif expérimental

Le plan expérimental se compose de trois étapes de test élémentaires répétées successivement : l'étape de cyclage, l'étape de rétention et l'étape de relecture. Le plan expérimental, présenté schématiquement sur la Fig. II-13, a été réalisé de manière à pouvoir suivre de manière précise l'évolution du courant de fuite  $I_{TAT}$  des cellules mémoires. Le but étant d'extraire la densité de piège  $N_{it}$ . Cette dernière est extraite après chaque étape de cyclage, durant les tests de rétention successifs. En effet, l'étape de cyclage va dégrader l'oxyde tunnel de manière significative alors que les étapes de rétention vont permettre la mesure des pertes de charges des cellules EEPROM lorsque celles-ci ne sont soumises à aucun champ électrique externe. Cependant, il est à noter que les charges stockées dans la grille flottante des cellules EEPROM induisent un champ électrique non nul (bas champ) aux bornes de l'oxyde tunnel.

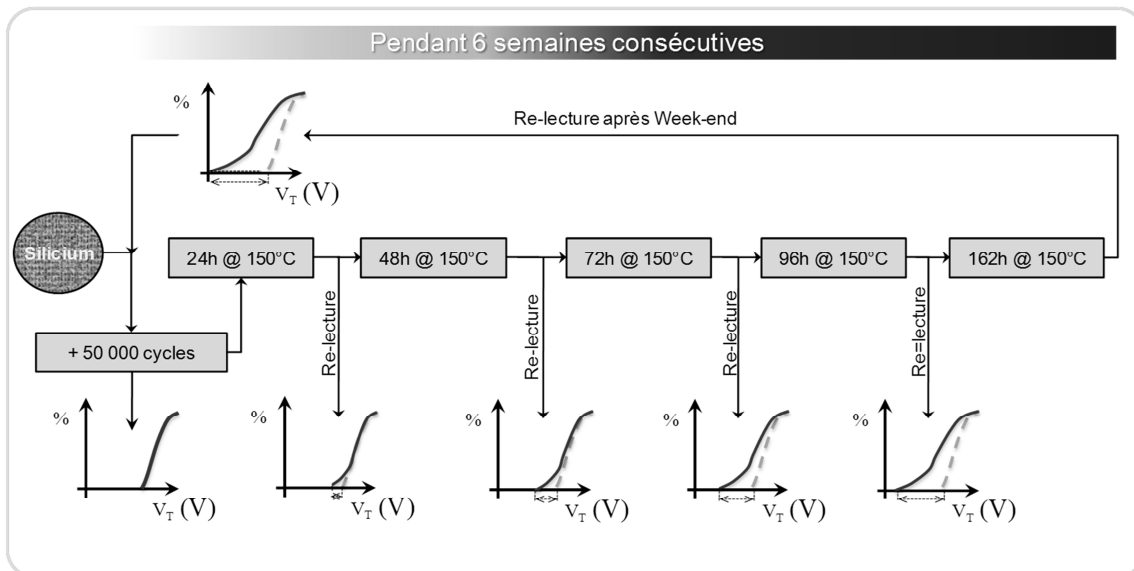


Fig. II-13 : Schéma représentatif du plan expérimental mis en place pour l'extraction de la densité de pièges générée tous les 50 000 cycles

Dans cette étude, les puces sont placées dans une étuve à 150°C pendant 24h. Cette température étant choisie de manière à ne pas induire un recuit des défauts.

Les étapes de relecture sont réalisées après le test de cyclage et ensuite toutes les 24h pendant toute la durée des tests de rétention, et cela, après avoir retiré les puces de l'étuve. Le but des étapes de rétention est d'extraire les distributions de  $V_T$  dégradées. Il est crucial de veiller à ne pas reprogrammer les cellules pendant les phases de relecture, ce qui fausserait la mesure en cours. Ainsi, la tension maximale de lecture est fixée à 4.5 Volts (tension appliquée sur les grilles de contrôle des cellules).

Tous les paramètres du plan expérimental sont résumés dans la Tableau II-2. Le plan expérimental s'étale sur 6 semaines et comprend 6 étapes de cyclage avec un nombre de cycles maximal subi par les cellules de 300 000 à la dernière étape.

Tableau II-2 : Récapitulatif des principaux paramètres du plan expérimental

Paramètres	Valeurs
Nombre total d'étapes	6
Nombre de cycles par étape	50 000
Heures de rétention pour chaque étape	162h
Nombre de Relecture	5 + 1 (à t=0)
Température de l'étuve pour la rétention	150°C
Tension de programmation	13.59V
Tension de lecture maximale	4.5V
Etat des cellules après cyclage et avant rétention	Effacées
Nombre total de cellules fonctionnelles à la fin du plan expérimental	18 911 232 (18,5Mbit)

## 2 Résultats expérimentaux

### a. Dégradation des distributions des tensions de seuil

Pour extraire le courant  $I_{TAT}$  pour une étape donnée, seules les cellules présentes dans la queue de distribution sont considérées. Plus précisément, on choisit un nombre de cellule fixe parmi la population de cellules marginales de la queue de distribution. Nous choisissons ainsi de suivre les 50 « pires » cellules de la queue de distribution qui sont identifiées par un critère simple : les cellules ayant la tension de seuil la plus faible. Les résultats à l'issue de chaque étape du plan expérimental (après 160h de rétention) sont présentés de la Fig. II-14 (50 00 cycles endurés) à la Fig. II-19 (300 000 cycles). Le nombre de cellules défectueuses augmente étape après étape. Cela se traduit par la translation vers le haut du point d'inflexion des distributions.

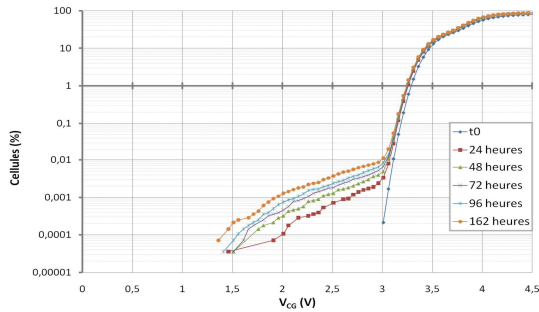


Fig. II-14 : Résultats des tests de cyclage/rétention après 50 000 cycles

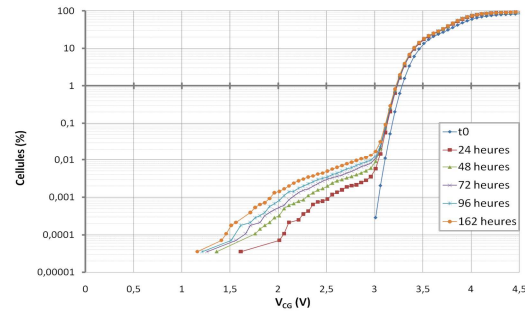


Fig. II-15 : Résultats des tests de cyclage/rétention après 100 000 cycles

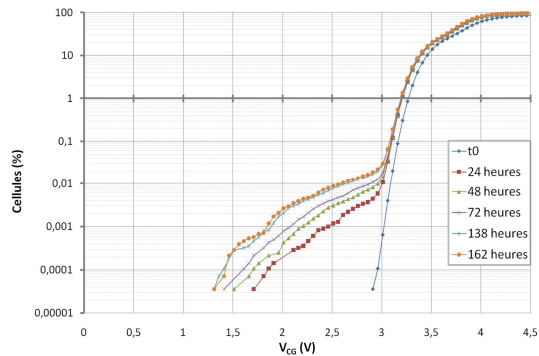


Fig. II-16 : Résultats des tests de cyclage/rétention à après 150 000 cycles

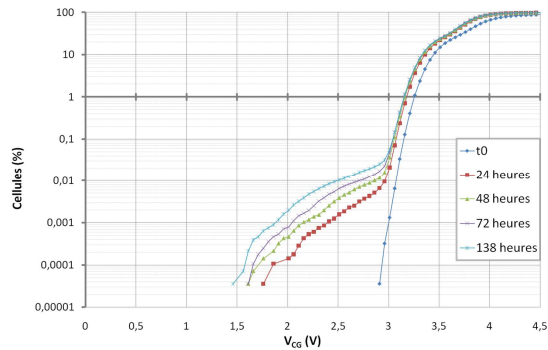


Fig. II-17 : Résultats des tests de cyclage/rétention après 200 000 cycles

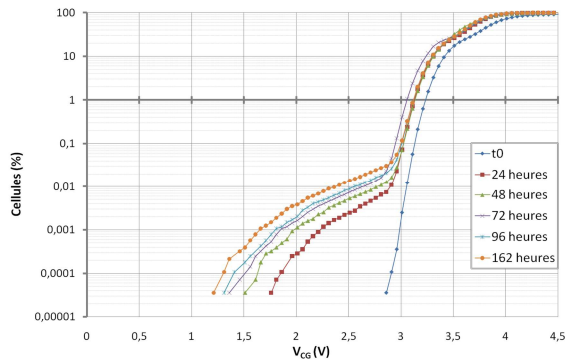


Fig. II-18 : Résultats des tests de cyclage/rétention après 250 000 cycles

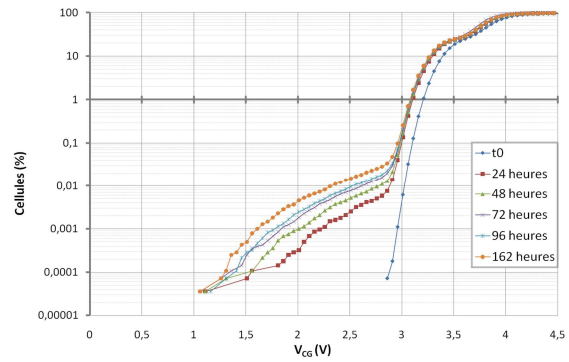


Fig. II-19 : Résultats des tests de cyclage/rétention après 300 000 cycles

### b. Technique de la cellule équivalente

L'approche utilisée de manière à suivre l'évolution des cellules marginales de la queue de distribution se base sur l'évolution de la moyenne des  $V_T$  des 50 pires cellules. Ainsi, l'analyse de la dégradation est ramenée à l'étude d'une cellule équivalente. Cette technique permet de faciliter l'extraction des paramètres du courant de fuite à partir des distributions de  $V_T$  dégradées.

La Fig. II-20 met en évidence la perte moyenne des 50 pires cellules à l'issue de chaque étape du plan expérimental. Si nous nous focalisons sur les extremums, nous pouvons constater que la perte moyenne pour les 50 000 premiers cycles est de 1.165 V alors qu'à l'issue des 300 000 cycles, la perte moyenne sur les 50 pires cellules est de 1.389V, soit une différence de 224mV. Il est à noter que le comportement constaté sur les cellules est erratique, ce qui veut dire que ce ne sont pas à chaque fois les mêmes cellules qui constituent la population des 50 pires cellules.

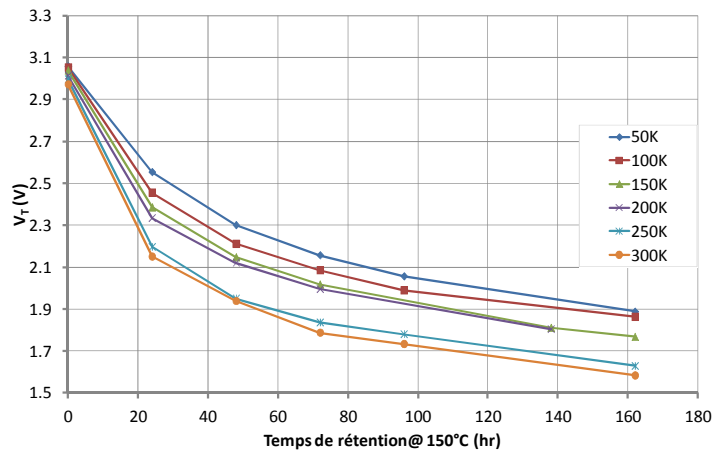


Fig. II-20 : Moyenne des  $V_T$  calculée sur les 50 pires cellules après chaque étape du plan expérimental (de 50 000 cycles à 300 000 cycles)

## 3 Méthode d'extraction de la densité de piège des cellules EEPROM

### a. Modèle de cellule EEPROM

L'extraction de la densité de pièges  $N_{it}$  utilise un modèle compact de cellule EEPROM [Regn06]. Ce modèle est basé sur l'utilisation d'un transistor

intrinsèque sur lequel les éléments spécifiques à l'EEPROM sont ajoutés. Ces éléments spécifiques sont les suivants :

- $C_{PP}$  : capacité modélisant la zone de diélectrique située entre la grille flottante et la grille de contrôle,
- $C_{TUN}$  : capacité modélisant la zone tunnel située entre la grille flottante et le drain,
- $I_{FN}$  : une source de courant pour modéliser l'injection Fowler-Nordheim,
- $I_{TAT}$  : une source de courant pour modéliser le courant TAT résultant du SILC.

Ces modifications permettent au modèle de calculer de manière indépendante la composante SILC et la composante Fowler-Nordheim du courant global traversant l'oxyde tunnel. L'extraction des valeurs de  $C_{PP}$  et de  $C_{TUN}$  est issue directement des mesures effectuées pendant les phases de fabrications des cellules EEPROM. Ce modèle, qui prend en compte les variations statiques et dynamiques des paramètres électriques du modèle, va nous permettre de relier les résultats expérimentaux liés à l'extraction des paramètres du SILC (et en particulier la densité de pièges générés en fonction du cyclage) aux résultats de simulation obtenus à partir du modèle.

b. Extraction de la densité de pièges  $N_{it}$

L'extraction de la densité de pièges à l'aide du modèle est réalisée par une méthode de superposition graphique.

Le modèle calcule dans un premier temps une courbe de dégradation de  $V_T$  pour chaque nombre de cycles endurés par les cellules mémoires, en tenant compte de la valeur par défaut des paramètres du modèle (valeurs présentées Tableau II-2 et obtenues à partir de capacités tunnel). Ensuite, un réajustement du paramètre de la densité de pièges  $N_{it}$  inclus dans le modèle de SILC est effectué de manière à obtenir une concordance entre les courbes de dégradation expérimentales ( $V_T$  de cellule équivalente) et les courbes de dégradation de  $V_T$  obtenues à partir du modèle.

En l'occurrence pour cette étude, seule la densité de pièges  $N_{it}$  a été ajustée par superposition graphique entre la courbe de dégradation de  $V_T$  calculée par le modèle et la courbe de dégradation expérimentale. Le Tableau II-3 présente le nouveau jeu de paramètres du SILC avec une valeur de  $N_{it}$  adaptée aux cellules mémoires EEPROM, après 50 000 cycles. Nous avons donc réalisé l'extraction du premier point de l'évolution de la densité de pièges générée après la première étape du plan expérimental.

Tableau II-3 : Paramètre  $N_{it}$  réajusté à la perte de charge de l'étape 1

Paramètres	Valeurs
$\phi_n$	$1.10 \cdot 10^{-15} \text{ cm}^{-2}$
$E_t$	2.6 eV
$N_{it}$	$1.247 \cdot 10^{16} \text{ m}^{-2}$
$X_t$	0.5 $t_{tun}$
$t_{tun}$	68 Å

### c. Evolution du paramètre $N_{it}$ en fonction du cyclage

La méthode d'extraction du paramètre  $N_{it}$  précédemment développée est réalisée pour l'ensemble des étapes du plan expérimental (i.e. après chaque étape de cyclage). Nous obtenons ainsi l'évolution de la densité de pièges en fonction du nombre de cycle (Fig. II-21). En utilisant une fonction d'interpolation reliant les points représentant l'évolution de la densité de pièges  $N_{it}$ , une loi d'évolution de type exponentiel est obtenue avec un coefficient de corrélation de 98,6%. Cette équation établit une relation simple entre le nombre de cycles et le taux de génération de pièges dans l'oxyde tunnel pour une technologie donnée.

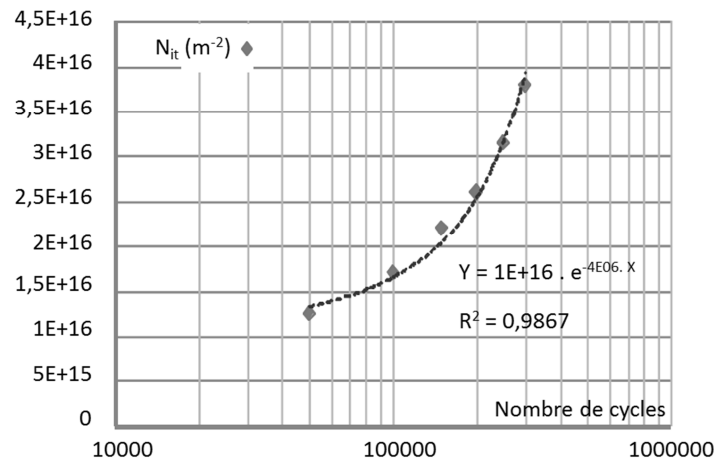


Fig. II-21 : Dynamique de génération de pièges dans l'oxyde tunnel en fonction du nombre de cycles

## III. Utilisation du TCF9 pour l'étude de la corrélation entre tests de stress et tests de rétention

### A. Corrélation entre tests de stress et tests de rétention

#### 1 Enjeux du stress électrique sur le temps de test

Les tests de stress électriques statiques consistent à appliquer un champ constant pendant une certaine période de temps de manière à accélérer la dégradation en  $V_T$  des cellules les plus faibles de la matrice mémoire. Les tests de stress électriques ont l'avantage principal d'offrir des temps de test beaucoup plus courts, comparativement aux tests en rétention. L'étude abordée dans ce paragraphe a pour but d'établir une corrélation entre les pertes de charges à la suite d'un test de rétention et celles obtenues à l'issue d'un test de stress électrique.

Les tests de rétention et de stress présentés dans cette étude sont effectués sur des populations de cellules effacées (grille flottante chargée en électrons).



## 2 Dispositif expérimental

Le plan expérimental est présenté Fig. II-22. Pour une même population de puces, après un test en endurance de 100K cycles, des tests en rétention et des tests de stress électriques sont effectués sur le véhicule de test TCF9. L'étape d'extraction des distributions de  $V_T$  (phase de lecture notée « R ») est réalisée pour différents temps de stress et de rétention.

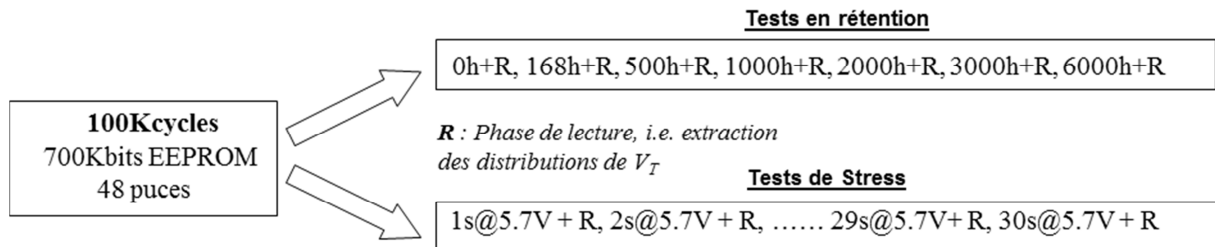


Fig. II-22 : Présentation du plan expérimental servant à faire la corrélation entre les tests de rétention et les tests de stress électrique

### a. Test en rétention

Les tests en rétention sont réalisés en plaçant les circuits sous étuve à une température de 150°C. Comme l'indique la Fig. II-22, des relectures de l'état des cellules sont effectuées après 0, 24, 168, 500, 1000, 2000, 3000 et 6000 heures de rétention. La Fig. II-23 illustre les distributions cumulées en  $\Delta V_T$  extraites suite aux étapes de lecture.

La première phase de la dégradation en rétention des données montre un taux de dégradation rapide de la distribution en  $\Delta V_T$ , tandis que la seconde phase présente une vitesse de dégradation plus lente.  $\Delta V_T$  peut être interprété comme une image de la perte de charge des cellules mémoire.

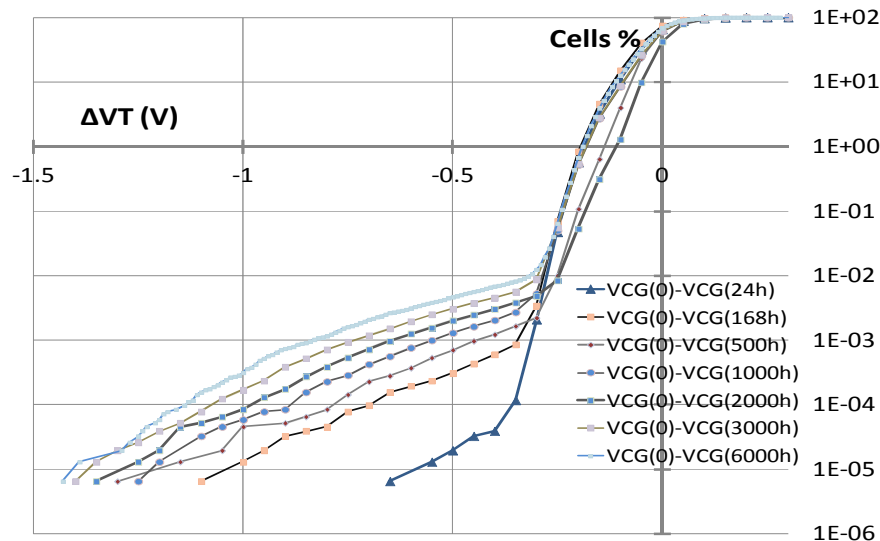


Fig. II-23 : Distributions en pourcentage cumulé en fonction de la perte de charge ( $\Delta V_T$ ) des cellules, pour toutes les étapes de rétention

### b. Test en stress

Les tests de stress électriques statiques génèrent un courant de fuite de la grille flottante vers le drain, sous l'effet d'un champ électrique aux bornes de l'oxyde

tunnel. Un point important des tests de stress est la maîtrise de la valeur du champ électrique de manière à ne pas se retrouver dans les conditions d'injections Fowler-Nordheim (champs électriques supérieurs à  $7\text{MV.cm}^{-1}$ ). Ces valeurs de champ électrique engendreraient une écriture des cellules.

La Fig. II-24 présente les distributions en  $\Delta V_T$  extraites toutes les secondes sur une période de temps de 30 secondes à température ambiante. Des temps de stress très courts sont nécessaires à la dégradation des distributions. Après chacun des tests de stress, le  $\Delta V_T$  de chaque cellule du plan mémoire est relevé.

On peut d'ores et déjà noter que les courbes de dégradation présentent la même allure que celles obtenues après les tests en rétention pour les mêmes puces. Cependant, la dégradation en stress est obtenue pour des temps de test beaucoup plus courts.

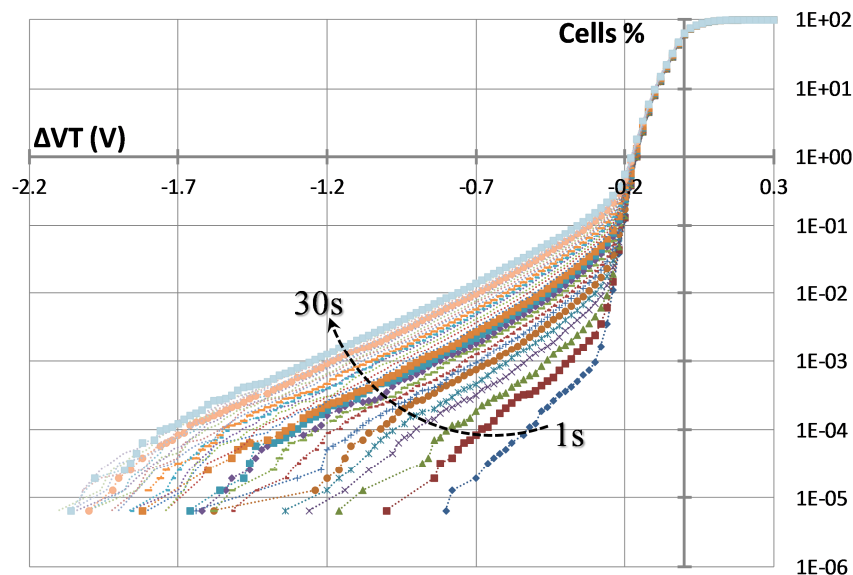


Fig. II-24 : Distributions en pourcentage cumulé en fonction de la perte de charge ( $\Delta V_T$ ) des cellules, pour toutes les étapes de stress.

### 3 Modèle de corrélation entre temps de stress et temps de rétention

Pour établir une corrélation entre les résultats des tests après stress électrique et ceux après rétention, les résultats de stress électriques obtenus entre 2s et 10s et les résultats de rétention obtenus entre 168h à 6000h sont comparés et présentés Fig. II-25a. Etant dans le cadre d'une étude de la défektivité en fonction du temps, les queues de distribution sont analysées, avec un intervalle d'étude  $\Delta V_T$  situé entre -0.4V et -0.8V comme l'illustre la Fig. II-25b. On note à partir cette dernière figure l'existence d'une corrélation entre les courbes de dégradation à l'issue des tests de stress et celles obtenues à l'issue des tests de rétention.

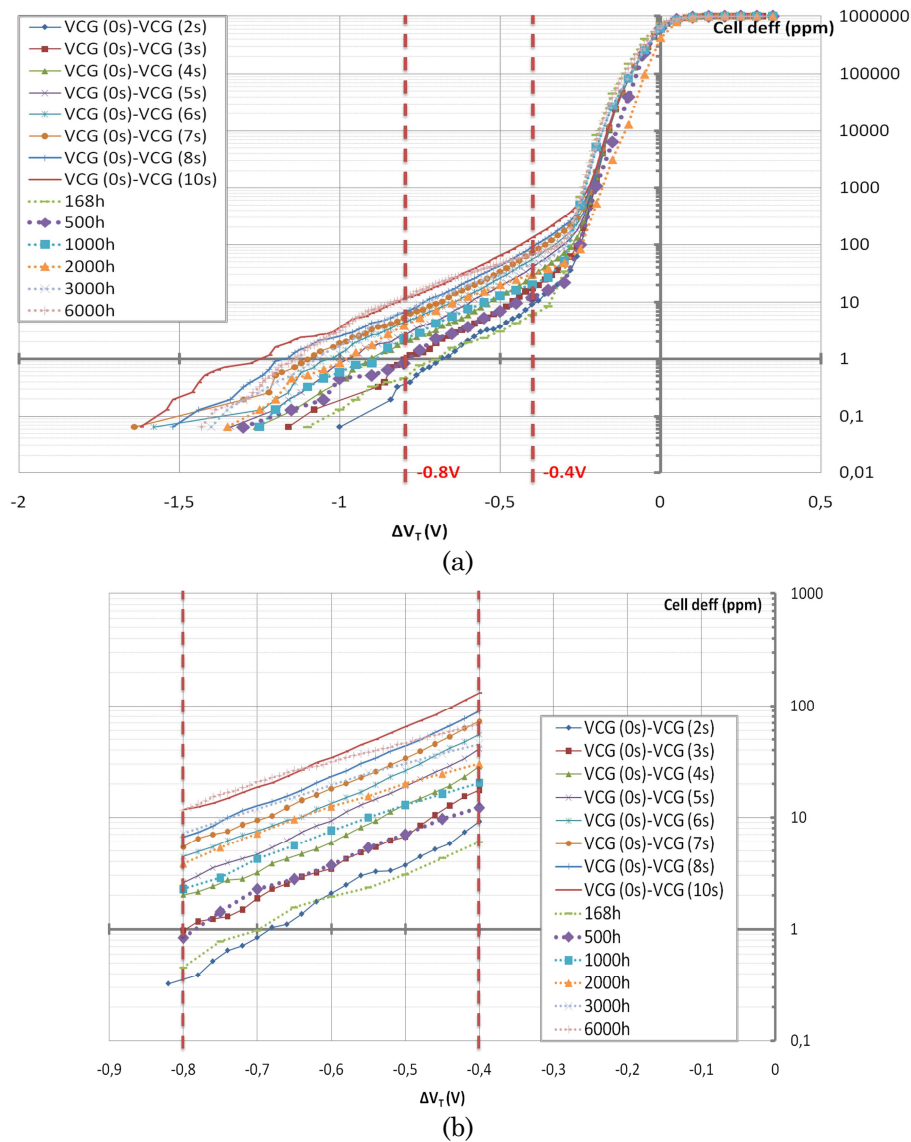


Fig. II-25 : Superposition des résultats de stress et des résultats de rétention en présentant les distributions en  $\Delta V_T$  (a) et les distributions en  $\Delta V_T$  limitées à l'intervalle d'étude (b).

La Fig. II-26 est tracée afin de trouver une fonction de corrélation entre les tests de stress électriques et les tests en rétention. Sur cette figure, chaque point est obtenu à partir d'un couple de valeurs communes (ppm et  $\Delta V_T$ ) des courbes de stress et de rétention extraites de la Fig. II-25. Ces points correspondent à une superposition entre les courbes de stress et les courbes de rétention pour des valeurs de  $\Delta V_T$  situées dans notre intervalle d'étude. On précise que nous autorisons une marge de corrélation, en ppm, entre les valeurs de défektivité issues du stress et de la rétention ( $\pm \log(3\text{ppm})$ ).

La fonction de corrélation entre stress et rétention est extrapolée avec un facteur de corrélation de bonne qualité, égal à 0.9732.

La fonction extrapolée (II-4) suit une loi de puissance de type :

$$y = \alpha x^\beta \quad (\text{II-4})$$

avec  $\alpha = 147848$  et  $\beta = 2,1706$

Dans cette équation,  $\alpha$  et  $\beta$  représentent les paramètres utilisés pour exprimer le temps de stress en fonction du temps de rétention. Selon (II.4), un test de stress électrique (court) de 10s peut être associé à un test de rétention de  $21.8 \times 10^6$  secondes (ce qui représente 6082 heures).

Ainsi, les tests en rétention (garant de la fiabilité des circuits mémoires), consommateurs en temps de test, peuvent être anticipés de manière rapide sur la base des résultats des tests de stress. Cette propriété peut être utilisée pour considérablement réduire le temps des tests de fiabilité en rétention. De plus, cette propriété peut s'avérer critique lors de l'introduction de nouvelles technologies, où une estimation de la fiabilité dans un temps très court est nécessaire de manière à ajuster le processus de fabrication.

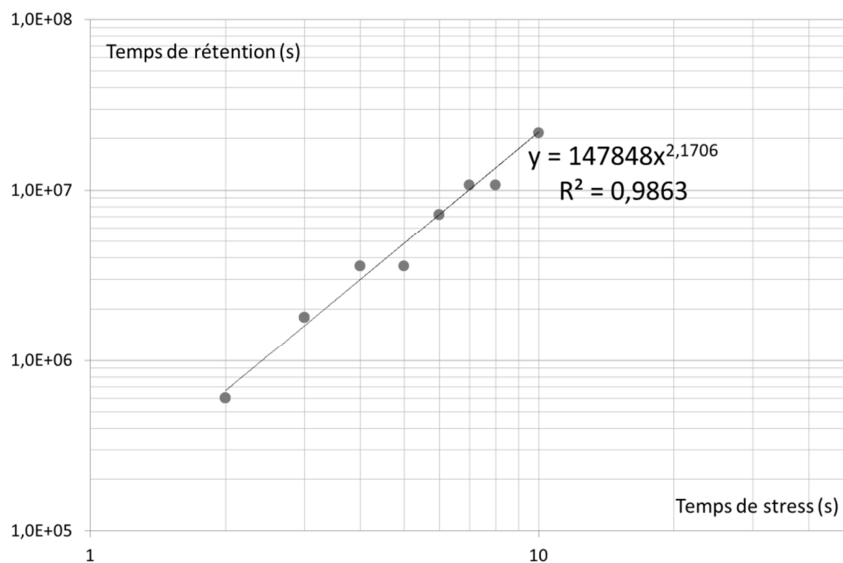


Fig. II-26 : Fonction de corrélation entre stress et rétention

## B. Comparaison des mécanismes de pertes à champ fort et champ faible

Dans cette partie, une corrélation entre les pièges dans l'oxyde activés lors des tests de rétention et les pièges activés lors des tests de stress est proposée. L'étude se base sur l'extraction et l'analyse des courants de fuite issus des distributions de  $V_T$ .

Ces courants de fuites sont mesurés pour les deux types de tests et comparés à des courants de fuite simulés à partir du modèle de SILC introduit au §II.A. La comparaison des courants de fuite extraits à partir des mesures et des courants de fuite simulés permet de fournir une explication physique de la corrélation entre les tests de rétention et de stress.

Les courants de fuite en rétention (« Measure Retention », après 100KCycles) et les courants de fuite en stress (« Measure Stress », après 100KCycles) sont reportés Fig. II-27. Ces courants sont extraits des distributions de  $V_T$  obtenues expérimentalement après les tests de rétention et de stress. On peut remarquer que les courants de fuite générés après les tests de rétention sont associés à des valeurs de tension de grille flottante (FG) comprises entre 0 V et -1V. En effet,

lors des tests de rétention, la perte de charge est observée sous de faibles champs électriques aux bornes de l'oxyde tunnel. Concernant les tests de stress, les mesures de courants de fuite sont obtenues pour des champs aux bornes de l'oxyde tunnel plus élevés, avec des valeurs de tension de grille flottante comprises entre -4V et -5.5V.

Sur la Fig. II-27, l'évolution de la densité de courant simulée  $J_{FG}^{72}$  ( $J_{FN}^{73} + J_{TAT}^{74}$ ) en fonction de la tension de grille flottante  $V_{FG}$  est représentée. La courbe de courant  $J_{FG}$  peut être obtenue pour toute valeur du paramètre  $N_{it}$  grâce à un modèle EEPROM calibré sur la technologie considérée pour cette étude.

Une bonne corrélation entre le courant de  $J_{FG}$  simulé et les courants de fuite extraits à partir des mesures est obtenue pour un paramètre densité de pièges égal à  $1.7 \times 10^{16} \cdot m^{-2}$  (i.e. une même courbe de courant  $J_{FG}$  correspond aux courbes expérimentales de rétention et de stress).

Cette observation démontre que la migration des électrons à travers l'oxyde tunnel est réalisée pour la même configuration de pièges (i.e. même densité de pièges  $N_{it}$ ) sous des champs électriques à travers l'oxyde tunnel aussi bien bas que haut. Par conséquent, les tests de stress électriques et les tests de rétention activent les mêmes pièges d'oxyde tunnel.

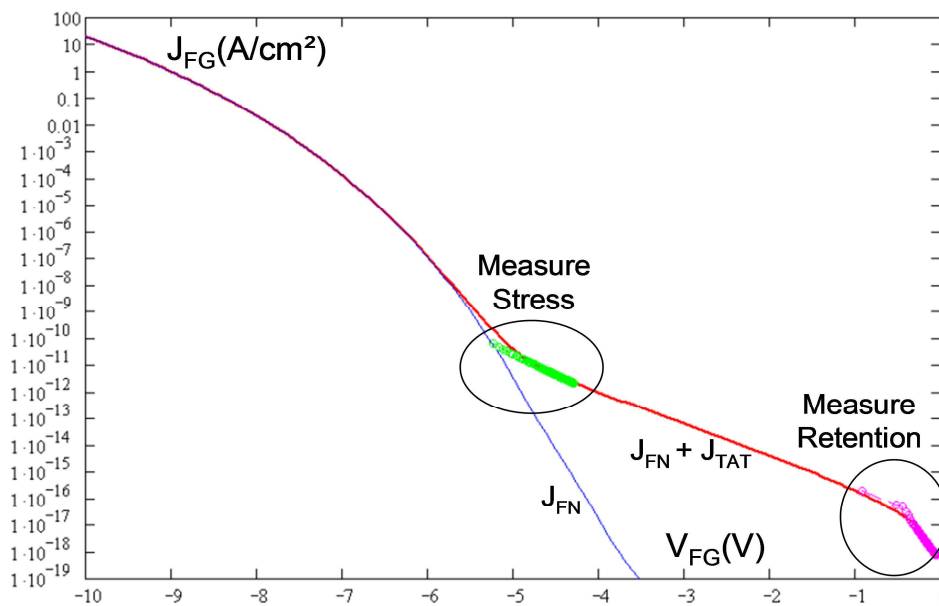


Fig. II-27 : Analyse des phénomènes d'activation des défauts dans l'oxyde tunnel durant les tests de stress et de rétention à partir de l'étude des courants de fuite de grille flottante

## IV. Conclusion

Dans ce chapitre, un véhicule de test a été utilisé de manière à réaliser deux études expérimentales : la première a consisté à extraire la densité de pièges de l'oxyde tunnel, la deuxième a permis d'établir une corrélation entre les tests de

<sup>72</sup> Densité de courants issus de la grille flottante chargée en électrons

<sup>73</sup> Densité de courant Fowler-Nordheim (FN)

<sup>74</sup> Densité de courant TAT

stress électrique et les tests de rétention.

L'extraction de la densité de pièges de l'oxyde tunnel a été motivée par le fait que la fiabilité des mémoires non-volatiles à grille flottante est directement liée aux défauts de l'oxyde tunnel. Ces défauts, modélisés sous forme de pièges, facilitent le transfert de charges depuis la grille flottante vers le substrat via un courant assisté par pièges à l'origine du SILC. Dans l'étude présentée, un modèle de cellule EEPROM prenant en compte le courant de SILC est utilisé de manière à extraire la densité de pièges  $N_{it}$  générée après différents tests de cyclage. La partie expérimentale alterne des tests en cyclage (à l'origine de la génération des défauts dans l'oxyde) et des tests en rétention (nécessaires à la mise en évidence de ces défauts). Durant ces tests, le paramètre expérimental mesuré a été la moyenne des  $V_T$  des cellules marginales du plan mémoire (i.e. technique de la cellule équivalente).

Ainsi, les tests en rétention ont permis de quantifier la fuite de charge à travers l'oxyde tunnel des cellules marginales en termes de variation de tension de seuil. L'analyse de la dégradation de la tension de seuil après chaque étape de cyclage et après des temps de rétention variables a permis d'extraire le paramètre  $N_{it}$  par corrélation entre les résultats expérimentaux et les résultats de simulation du modèle. Ainsi, une loi d'évolution de la densité de pièges en fonction du nombre de cycles a été proposée.

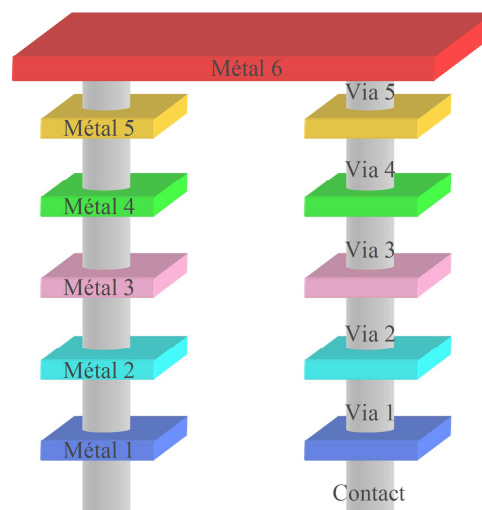
Concernant l'étude de la corrélation entre les tests de stress électrique et les tests de rétention, une technique inédite de prévision de la dégradation de la durée de vie des mémoires EEPROM a été présentée. Cette étude a montré que la dégradation des cellules EEPROM en rétention pouvait être anticipée de manière rapide et précise à partir de tests de stress électriques. La corrélation entre le temps de stress et le temps de rétention a été donnée par une fonction mathématique simple pour une technologie avancée de mémoire EEPROM. D'un point de vue physique, il est démontré à partir d'un modèle compact de cellule EEPROM que des mêmes pièges de l'oxyde tunnel, sont responsables des faibles courants de fuite lors des tests en rétention et de forts courants de fuite lors des tests de stress électrique. Ainsi, cette corrélation établit que deux tests de fiabilité différents activent le même type de défauts dans l'oxyde.

## Chapitre III : FIABLITE DES CIRCUITS LOGIQUES

Dans le premier chapitre, les différentes phases du développement d'une technologie ont été décrites et les deux principales sources de pertes de rendement qui sont la défektivité et les variations du procédé de fabrication ont été présentées. Ensuite, les outils employés pour mesurer les pertes de rendement ont été introduits. Les principaux outils sont le D0, la caractérisation et le test électrique. Le flot d'analyses de défaillance a également été présenté.

Dans ce chapitre, des solutions permettant de contrôler et d'accélérer les montées en rendement des technologies CMOS logiques sont proposées.

Un exemple pratique est ensuite développé : la réalisation et l'optimisation d'une macro-cellule de test, le RAPITAC. Son but est de fournir un outil d'analyse des variations du BEOL. Une description précise de l'architecture du RAPITAC est tout d'abord présentée. Ensuite, le programme de test du RAPITAC ainsi que le logiciel développé pour analyser les résultats de test sont décrits. Grâce à cette structure une solution de caractérisation des variations du BEOL est mise en place.



<b>I.</b>	<b>STRUCTURES DE TEST NUMERIQUES : TEG « LOGIQUES »</b>	<b>80</b>
<b>A.</b>	<b>LE TAC8M</b>	<b>80</b>
1	La structure du TAC8M	80
2	Limitations du TAC8M	83
<b>B.</b>	<b>LE TEGTACF9</b>	<b>84</b>
1	Conception du TEGTACF9	84
2	Limitation du TEGTACF9	86
<b>II.</b>	<b>APPLICATION A L'ANALYSE DES VARIATIONS DU BEOL</b>	<b>86</b>
<b>A.</b>	<b>INTRODUCTION</b>	<b>86</b>
<b>B.</b>	<b>DESCRIPTION DU RAPITAC</b>	<b>87</b>
<b>C.</b>	<b>TEST DU RAPITAC ET TRAITEMENT DES DONNEES</b>	<b>90</b>
1	Déroulement du flot de test	90
2	Traitement des données issues du test	91
a.	<i>Traitement et représentation des données fonctionnelles</i>	<i>92</i>
b.	<i>Traitement des données paramétriques</i>	<i>92</i>
c.	<i>Représentation des données paramétriques</i>	<i>92</i>
<b>D.</b>	<b>CARACTERISATION DES VARIATIONS DU PROCEDE DE FABRICATION</b>	<b>95</b>
<b>III.</b>	<b>CONCLUSION</b>	<b>99</b>



## I. Structures de test numériques : TEG « logiques »

Pour répondre à la problématique liée au contrôle de procédés de fabrication de plus en plus complexe, des structures de test sont développées. Elles sont plus proches des configurations et des fonctionnalités présentes dans les produits. Ces structures peuvent être classées en trois catégories : les TEG logiques, les macro-cellules de test et les TEG hybrides.

De manière à se familiariser avec les TEG « logiques ». Deux structures de test vont être présentées. Ces structures sont le TAC8M et le TEGTACF9, elles sont toutes deux dérivées de macro-cellules de test.

### A. Le TAC8M

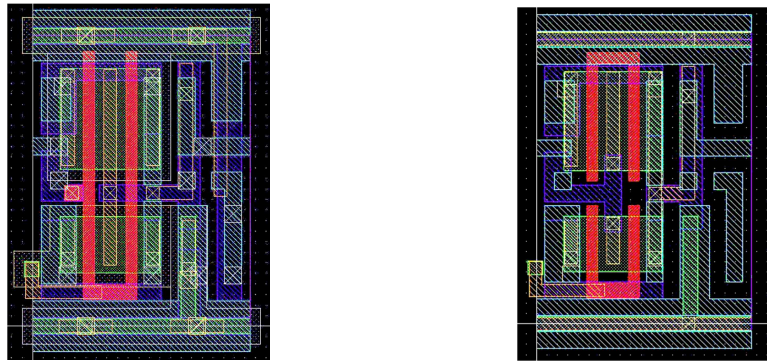
Le TAC8M [Forl05] est la version TEG d'une macro-cellule de test appelée TAC8 (Technology Assesment Chip), le "8" indiquant le niveau technologique et le "M" pour « modifié ». Cette structure a pour but d'évaluer la densité de défauts du BEOL. Pour ce faire, cette structure a été conçue avec le cahier des charges suivant :

- Pouvoir être facilement redimensionnée, de manière à être utilisé aussi bien en tant que TEG ou macro-cellules, ou encore être positionnée dans un produit,
- Obtenir rapidement la densité et la localisation des défauts,
- Placer les objets technologiques de manière à limiter le nombre de coupes nécessaires pour l'analyse de l'élément défaillant.

Il est important de souligner que le passage du TAC8 en TAC8M devait être effectué en modifiant le moins de masques possible. En effet, c'est à partir des inverseurs contenus dans le TAC8 que les différentes portes logiques du TAC8M ont été réalisées. La structure de ce dernier va maintenant être présentée, en utilisant une approche ascendante, en partant de l'inverseur unitaire jusqu'à la structure complète.

#### 1 La structure du TAC8M

L'élément de base du TAC8M est un inverseur « IV4 », c'est-à-dire un inverseur composé de deux inverseurs en parallèle. De ce fait, en cas d'un défaut intervenant au niveau du FEOL, selon sa taille, il n'impactera qu'un seul des deux inverseurs. En effet, la structure ayant pour cible le BEOL, il est important que le FEOL soit plus robuste. La Fig. III-1 montre le dessin des masques de l'« IV4 » ainsi que la façon dont il est modifié pour obtenir une porte de transmission « PT4 ». Un multiplexeur « MUX21 » ainsi qu'une bascule « D » sont également réalisés à partir de portes « IV4 ».



(a)

(b)

Fig. III-1 : Inverseur de base du TAC8M IV4 (a) et modification en porte de transmission PT4 (b)

Puis, ces différentes portes logiques sont assemblées pour former des chaînes de deux ou trois « IV4 », appelés respectivement bloc « 2IV4 » et bloc « 3IV4 ». La Fig. III-2 est une illustration du bloc « 2IV4 ». Ces chaînes sont précédées par un « MUX21 » qui sert à sélectionner en entrée, le signal ligne ou colonne, selon le niveau du signal « SEL ». Les cellules « CIE » et « CINE » sont des cellules d'interconnexions avec, respectivement, des vias empilés et non empilés décrits Fig. III-3.

Pour créer le niveau de hiérarchie supérieur, appelé « GROUPE », les blocs sont assemblés sous forme de matrice. Cette matrice, présentée Fig. III-4, est composée de 4 lignes et de 4 colonnes, chaque point de la matrice étant un bloc d'inverseurs « 3IV4 » et un bloc « 2IV4 » chaînés. Cette matrice permet d'évaluer la propagation d'un signal à travers les inverseurs, en mode de propagation ligne ou colonne. Le signal se propage à travers 4 blocs « 3IV4 » et 4 blocs « 2IV4 » soit 20 inverseurs. Un groupe possède 4 entrées et 4 sorties lignes ( $E_{Lx}$ ,  $S_{Lx}$ ) ainsi que 4 entrées et 4 sorties colonnes ( $E_{Cx}$ ,  $S_{Cx}$ ).

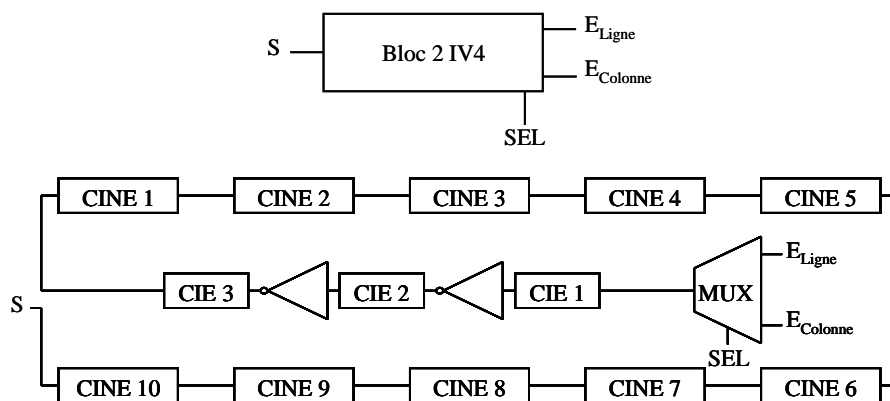


Fig. III-2 : Schéma fonctionnel du bloc 2IV4

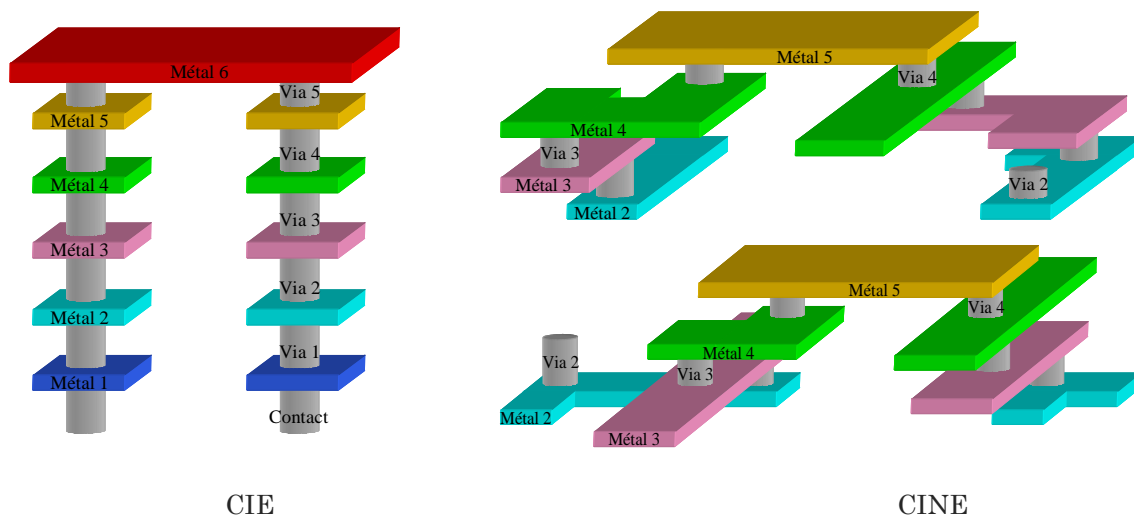


Fig. III-3 : Configuration du BEOL avec vias empilés (CIE) et vias non empilés (CINE)

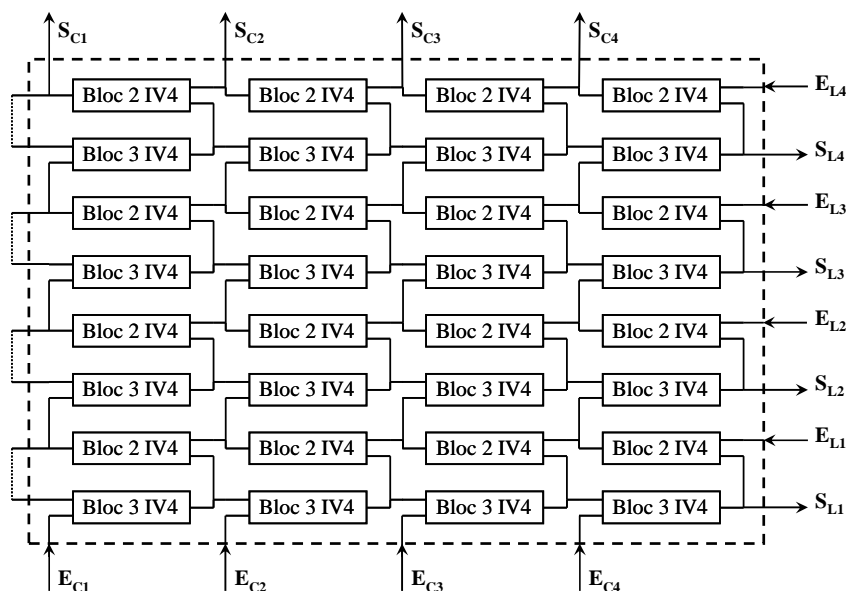


Fig. III-4 : Représentation schématique d'un GROUPE (4 lignes et 4 colonnes)

Finalement, le plus haut niveau de hiérarchie du TAC8M est composé de 2 groupes côte à côte et de registres à décalage, plus communément appelés « chaînes de scan ». Le système complet est donné Fig. III-5, où chaque rectangle représente un point de « GROUPE ». Finalement, les sorties des chaînes de scan ligne et colonne, sont connectées aux entrées d'un multiplexeur 2 vers 1, qui est lui-même la sortie du système (OUT). Le fonctionnement est alors le suivant :

- Initialisation des entrées IN1 et IN2 (respectivement '1' et '0'), mode de propagation colonne (SEL= '1') et positionnement en mode d'acquisition (AD='1'),
- Inversion des signaux IN1 et IN2, attente de la propagation durant une demi-période d'horloge et positionnement en mode de décalage (AD='0'),
- Lecture des données de la chaîne de scan colonne sur la sortie OUT,
- Positionnement en mode de propagation ligne (SEL= '0') et en mode d'acquisition (AD='1'),

- Inversion des signaux IN1 et IN2, attente de la propagation et positionnement en mode de décalage,
- Lecture des données de la chaîne de scan ligne sur la sortie OUT.

Puis, en analysant les signaux sur la sortie, d'éventuels défauts peuvent être détectés et localisés. En effet, lors du déchargement des registres sur la sortie, au moment d'un front d'horloge descendant, chaque donnée présente correspond à la propagation du signal le long d'une ligne ou d'une colonne (selon la valeur du niveau SEL).

Ainsi, lors d'un défaut « hard »<sup>75</sup>, les données de la ligne et de la colonne correspondantes sont erronées. Finalement, cette structure permet également de détecter des défauts « soft »<sup>76</sup>, qui, par définition, sont des défauts légers qui n'empêchent pas le fonctionnement du circuit, mais qui limitent ses performances. Sachant que la durée, entre le moment où les entrées sont inversées et le moment où les registres passent en mode décalage, correspond à une demi-période d'horloge, la détection des défauts « soft » est faite en appliquant le pattern de test à différentes fréquences. En effet, si un défaut « soft » est présent sur un point de la matrice, il modifiera les temps de propagation de la ligne et de la colonne correspondante.

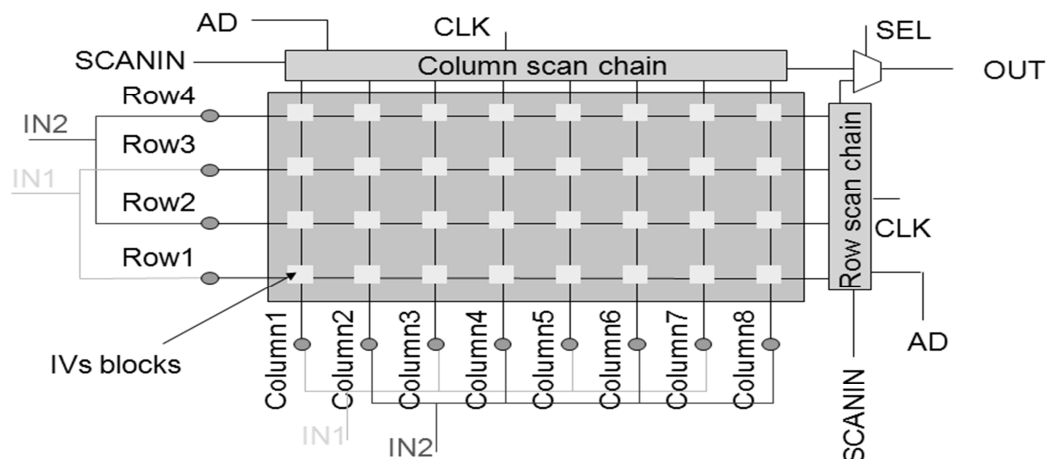


Fig. III-5 : Système complet du TAC8M

## 2 Limitations du TAC8M

Lors des premiers tests effectués sur silicium, il est apparu des différences entre la mesure et la simulation. En effet, alors qu'en simulation la fréquence maximale de fonctionnement était de 100MHz, en mesure, elle était limitée à environ 12MHz. Par conséquent, à cette fréquence les défauts « hard » peuvent être détectés mais pas les « soft ». Le second problème rencontré est qu'en raison de sa faible surface, de très rares défauts ont été détectés. Cependant, grâce à des défauts créés artificiellement au niveau du dessin des masques, sur d'autres versions de cette structure, la localisation de défauts ainsi que la procédure d'analyse physique ont pu être validées.

<sup>75</sup> Défaut « hard » : court-circuit ou circuit ouvert franc, avec respectivement une résistance quasi-nulle ou infinie

<sup>76</sup> Défaut « soft » : entraîne une limitation des performances et non une défaillance

## B. Le TEGTACF9

Tout comme le TAC8M, cette structure est également une contraction d'une macro-cellule de test appelée TATITAC. Son but est d'évaluer certains paramètres critiques du procédé de fabrication et de fournir, rapidement et simplement, un aperçu de la densité de défauts. Les paramètres visés sont la défektivité des vias empilés et non empilés, ainsi que les espacements minimums entre les lignes de métaux. La structure a été conçue en suivant les règles suivantes :

- Utilisation des dimensions minimales du DRM,
- Evaluer le plus grand nombre de fonctionnalités « produit » possible,
- Facilité d'extraction de la densité de défauts,
- Organiser de la structure de manière à faciliter les analyses de défaillances,
- Soucis de simplicité du programme de test.

### 1 Conception du TEGTACF9

La structure de ce circuit de test va maintenant être présentée, en utilisant une approche ascendante. Comme la plupart des structures de test, l'élément de base du TEGTACF9 est un inverseur. Chaque inverseur a été conçu avec des montées et descentes de vias empilés et non empilés. La Fig. III-6a montre que les vias et les contacts<sup>77</sup> ont été placés dans un nombre de plans horizontaux et verticaux limités dans le but de réduire le nombre de plans de coupes pour les analyses physiques. De plus, des marques d'alignements ont été placées à intervalles réguliers pour faciliter la localisation des inverseurs lors des analyses physiques (Fig. III-6b). Ces inverseurs sont ensuite utilisés avec la hiérarchie suivante :

- Un « bloc » est composé de 49 inverseurs chaînés,
- Ce « bloc » est répété 6 fois pour faire une « ligne »,
- Une ligne est répétée 2 fois pour faire une « boucle »,
- Une boucle est répétée 7 fois pour faire une « chaîne »,
- La structure comporte une seule « chaîne ».

---

<sup>77</sup> Contact : interconnexions métalliques entre le polysilicium ou l'active et le premier niveau de métal

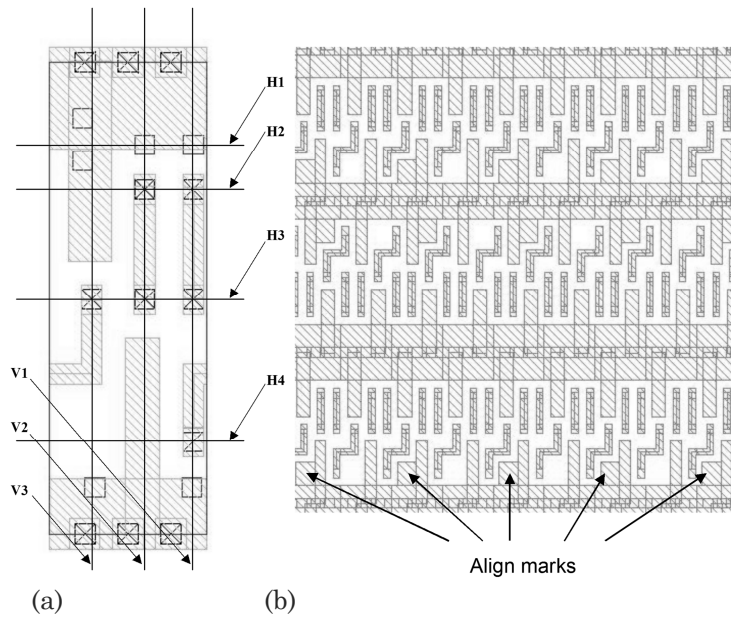


Fig. III-6 : Plan de positionnement des vias (a) et marques d'alignement (b) pour les analyses physiques

Ensuite, un registre à décalage ligne et une « chaîne de scan » colonne ont été ajoutés pour faciliter la localisation des inverseurs défaillants. Le registre à décalage ligne a ses entrées d'horloge connectées à la fin de chaque boucle et l'entrée de donnée de la première bascule est connectée à  $V_{DD}$  (Fig. III-7). De ce fait, lorsqu'un créneau<sup>78</sup> est appliqué sur l'entrée « Chainin », il est propagé à travers la chaîne d'inverseurs. Lorsque le créneau atteint une fin de « boucle » la bascule correspondante passe à '1'. Ainsi, quand le créneau a traversé toute la chaîne, la dernière bascule (la sortie « ScanOutColumn ») passe à '1', attestant d'une puce sans défaut.

Si la puce comporte un défaut, la sortie « ScanOutColumn » ne passera jamais à '1'. Dans ce cas, les entrées d'horloge du registre à décalage ligne, peuvent être connectées à l'entrée « CkColumn » pour décharger le registre à décalage ligne. Cela pour savoir à partir de quelle « boucle » le créneau ne s'est plus propagé. De plus, grâce à des multiplexeurs, les sorties de chaque « bloc » de la boucle défaillante sont connectées à la « chaîne de scan » colonne. Par conséquent en déchargeant cette dernière, il est possible de savoir quel « bloc » est défaillant.

Finalement l'analyse du temps de propagation de la « chaîne », entre le moment où le créneau est appliqué sur « Scanin » et le moment où « ScanOutColumn » passe à '1', permet d'évaluer le procédé de fabrication. En effet, toute dérive du procédé de fabrication va entraîner une variation du temps de propagation.

<sup>78</sup> Créneau : passage de l'état logique '1' à '0' ou inversement

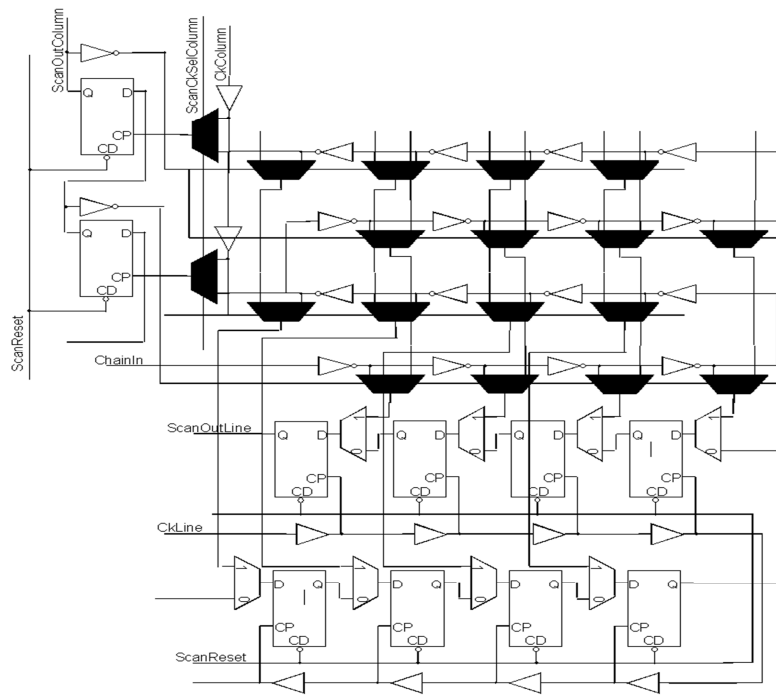


Fig. III-7 : Schéma électrique partiel du TEGTACF9

## 2 Limitation du TEGTACF9

Sachant que le niveau de granularité correspond à la détection des défauts est un « bloc », il est nécessaire et chercher les défauts parmi 49 inverseurs, ce qui est plus complexe que sur les deux ou trois inverseurs du TAC8M. Puis, ne pouvant obtenir que le temps de propagation de la « chaîne complète », si un défaut « soft » est présent sur un des inverseurs, le retard introduit par ce dernier, ne sera pas détectable sur le temps de propagation des 4116 inverseurs. Finalement, comme pour le TAC8M, la petite taille de la structure fait que la probabilité de capturer un défaut est assez faible.

## II. Application à l'analyse des variations du BEOL

### A. Introduction

Dans le but d'obtenir des systèmes sur puce (SOC) de plus en plus performants, des procédés technologiques avancés sont employés. Ces derniers sont plus sensibles à la défektivité qui impacte directement la montée en rendement des produits. De plus, les concepteurs et les ingénieurs « process » travaillent en étroite collaboration pour réduire l'impact des variations du procédé de fabrication sur les performances des produits. De manière à contrôler ces effets indésirables sur les produits, l'industrie du semi-conducteur a adopté une solution basée sur l'utilisation de macro-cellules de test. Les plans mémoires SRAM sont connus pour être de bon véhicules pour détecter les défauts « hard ».

D'autre part, les oscillateurs en anneau sont massivement utilisés pour caractériser les variations du procédé de fabrication.

Le principal objectif de cette partie est de présenter une macro-cellule de test conçue en combinant les avantages de ces deux dernières structures. Son but est de caractériser les pertes de rendement dues à la densité de défauts et aux variations du procédé de fabrication du BEOL. Les défauts considérés sont les circuits ouverts ou les courts-circuits dus à des particules ou à des défaillances critiques du procédé de fabrication. De la même façon, les dérives du BEOL concernées sont les variations des dimensions critiques (CD) des couples via/métaux. Cette structure, appelée RAPITAC, a été développée à l'IM2NP en collaboration avec ST-Microelectronics ROUSSET. Elle a été ensuite implémentée à ST-Microelectronics CROLLES.

La première section de cette partie est consacrée à la description de cette structure. Cette description montre comment une cellule SRAM oscillante basique est conçue et assemblée en un plan mémoire de 4096 cellules. Dans une deuxième section, le programme de test associé à la structure ainsi que les traitements effectués sur les fichiers contenant les résultats de test sont détaillés. Les différents types de représentations graphiques utilisés pour détecter les défauts et les variations du procédé de fabrication sont présentés. Enfin, la dernière section se focalise sur la caractérisation des variations du procédé de fabrication.

## **B. Description du RAPITAC**

La structure du RAPITAC est présentée en utilisant une approche ascendante, en partant de l'élément de base, jusqu'à la description de la structure complète. Cette structure a été conçue en utilisant la technologie H9, qui est une technologie 130nm de ST-Microelectronics ROUSSET. Le circuit est basé sur une cellule SRAM classique avec deux portes de transmission (PT1 et PT2) et deux inverseurs (IV1 et IV2). Un troisième inverseur, optionnel (IV3), est ensuite ajouté dans le but de créer un oscillateur en anneau. Cette cellule appelée OSRAM est représentée schématiquement Fig. III-8. Le signal OS permet d'insérer ou de court-circuiter le troisième inverseur. La cellule est en mode oscillation quand le signal OS est à '1', ou en mode SRAM quand OS est à '0'. Chaque cellule OSRAM est sélectionnée individuellement en plaçant les signaux ROW et COL correspondants, à '1'. En effet, le signal LM, contrôlant les portes de transmission permettant d'accéder à la cellule, est contrôlé par une porte ET.



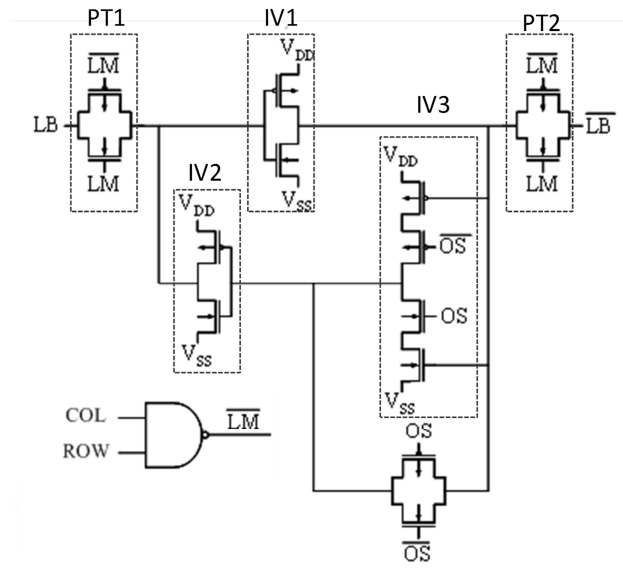


Fig. III-8 : Vue schématique de la cellule de base comportant 2 portes de transmission, 2 inverseurs plus un troisième inverseur optionnel

La Fig. III-9 représente le dessin de masques de la cellule OSRAM dont la surface est de  $94\mu\text{m}^2$ . La logique de contrôle est visible sur la gauche du dessin de masques. La cellule SRAM oscillante se trouvant sur la droite est entourée par les deux portes de transmission. Alors que la partie logique est conçue à l'aide de portes standards de la librairie CORELIB, les portes de transmission et les cellules mémoires sont dimensionnées manuellement. Ceci pour deux raisons :

- Obtenir des fréquences d'oscillation mesurables par un testeur de type Verigy HP93K,
- Réaliser les connections entre les inverseurs avec des niveaux différents.

En effet, en augmentant la longueur  $L$  des transistors, le rapport  $W/L$  diminue, entraînant une diminution du courant  $I_{DS}$ . Les capacités des transistors sont alors chargées plus lentement. Cela induit un délai plus important dans chaque inverseur et donc une fréquence d'oscillation plus faible. La fréquence d'oscillation obtenue est d'environ 65MHz.

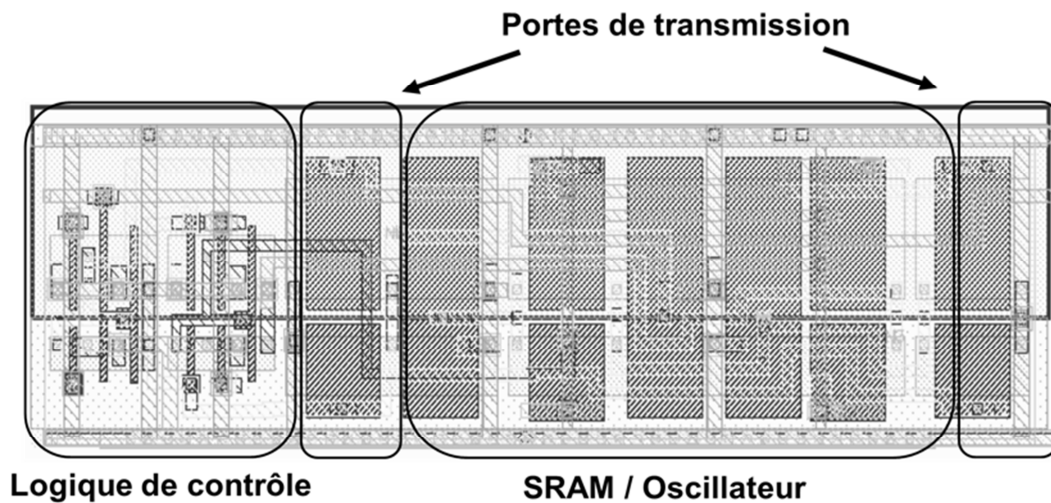


Fig. III-9 : Layout customisé de la cellule OSRAM ( $94\mu\text{m}^2$ )

Les cellules OSRAM sont routées avec des niveaux de métaux différents. La Fig. III-10a montre que les inverseurs d'une cellule M1 sont interconnectés uniquement avec du le niveau de métal 1 (OSRAM<sub>M1</sub>). De même, pour router les inverseurs d'une cellule M4, du métal 4 ainsi que les vias 1 à 3 sont employés. En effet, la fréquence d'oscillation d'un oscillateur en anneau dépend du FEOL des inverseurs, mais également du routage entre ces inverseurs. Par conséquent, en employant des cellules avec différents routages, il est possible de cibler le niveau de métal défaillant. Ces cellules sont ensuite assemblées en sous-blocs de 4 colonnes et 32 lignes d'un même type (Fig. III-10b). Les différents sous-blocs sont finalement assemblés en plan mémoire de 4096 bits avec une configuration en damier, dans le but d'avoir des sous-blocs de chaque type (M1, M2, M3 et M4), répartis sur toute la structure (Fig. III-10c).

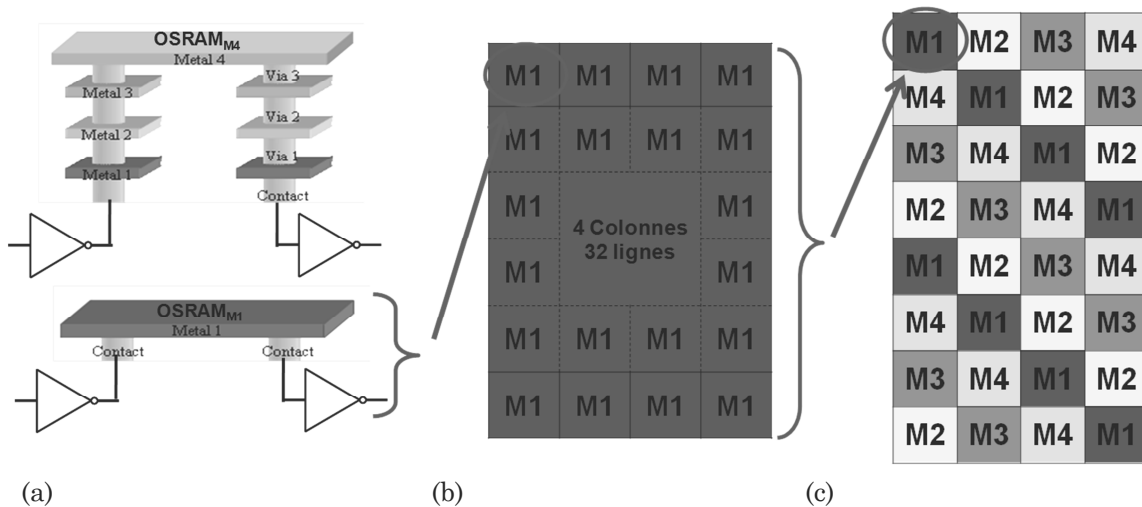


Fig. III-10 : Configuration du plan mémoire du RAPITAC

L'adressage du plan mémoire est réalisé à l'aide de registres à décalages. La Fig. III-11 présente le bloc d'adressage constitué d'un démultiplexeur « 1 vers 4 » en entrée, quatre registres à décalage et un multiplexeur « 4 vers 1 » en sortie. Le démultiplexeur est utilisé pour diriger l'entrée d'horloge CLK vers un des 4 registres à décalage à l'aide des signaux SEL0 et SEL1. L'entrée E<sub>SCAN</sub> étant connectée à tous les registres, les données présentes sur ce signal sont chargées en série dans le registre, vers lequel l'horloge est dirigée. De la même façon, le multiplexeur sélectionne la sortie de ce même registre pour la diriger sur la sortie S<sub>SCAN</sub>.

Le premier registre à décalage concerne une fonction non exploitée dans le cadre de cette étude. Le deuxième registre est le décodeur lignes (RDL), il est composé de 256 bits. Un '1' logique sur le bit n sélectionne la ligne n du plan mémoire. Le troisième registre est le décodeur colonnes RDC, qui lui est composé de 16 bits. Comme pour RDL, un '1' logique sur le bit n active la colonne n. Enfin, le dernier registre est le registre d'écriture qui est également composé de 16 bits. La donnée présente sur le n<sup>ième</sup> bit est écrite sur le n<sup>ième</sup> bit de la ligne activée. Une deuxième sortie S<sub>DATA</sub> (non représentée sur la Fig. III-11) permet de récupérer le signal de sortie de la cellule OSRAM sélectionnée.

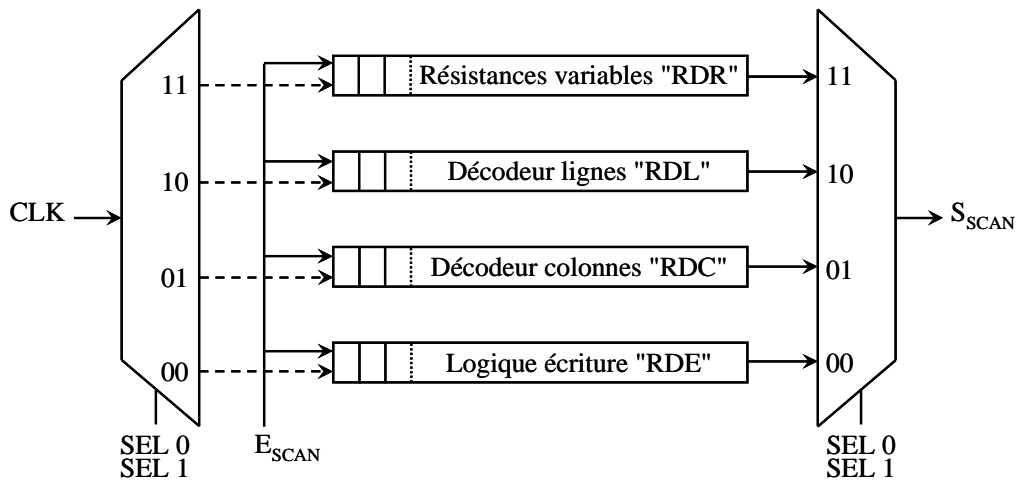


Fig. III-11 : Registres à décalage utilisés pour l'adressage du plan mémoire

## C. Test du RAPITAC et traitement des données

### 1 Déroulement du flot de test

Le programme de test est développé sur un testeur Verigy HP93K. Ce programme de test comporte trois parties : les tests paramétriques standards, les tests fonctionnels et les procédures de capture analogique du signal. Ces trois parties sont décrites ci-dessous :

- Test paramétriques standards : test de continuité, mesures de courants statiques et dynamiques,
- Test fonctionnels : écriture et lecture des registres, écriture et lecture du plan mémoire. Les algorithmes de test utilisés sont ceux employés pour le test des mémoires (tout à '0', tout à '1', damier et damier inverse),
- Capture analogique du signal : le signal d'oscillation de chaque point mémoire est capturé et enregistré dans des fichiers.

La capture analogique est effectuée à l'aide d'une carte électronique optionnelle, qui peut être rajoutée sur les testeurs, appelée digitiseur<sup>79</sup>. Ce dernier est utilisé pour convertir des signaux analogiques en mots binaires. Comme tout convertisseur analogique/numérique, il est composé d'un échantillonneur-bloqueur et d'une partie chargée de la quantification du signal. La première partie sert à maintenir le signal d'entrée à une valeur stable durant la quantification du signal. Le digitiseur étant un périphérique du testeur, il est commandé par un « trigger ». Ce dernier est une sortie virtuelle, qui est déclarée dans le programme de test et qui est connectée par des relais au digitiseur, elle doit être positionnée à '1' pour activer la conversion. Une « testMethod » doit alors être écrite pour récupérer les mots binaires provenant du digitiseur.

Le diagramme de la Fig. III-12 montre comment le test et le traitement des données analogiques sont effectués. Dans un premier temps, un testeur de production standard, équipé d'un digitiseur, envoie les signaux de test sur les

<sup>79</sup> Digitiseur : Carte électronique, composée d'un échantillonneur/bloqueur et d'un convertisseur analogique/numérique, utilisée pour retranscrire un signal analogique en mots binaires

plaquettes de silicium. Ensuite, le testeur enregistre les résultats des tests paramétriques et fonctionnels dans un fichier « .std ». Cette extension correspond au format standard des fichiers issus des ATE. Les signaux d'oscillations des points mémoires sont enregistrés dans d'autres fichiers indépendants. Finalement, un logiciel développé spécialement pour le RAPITAC calcule la TFD<sup>80</sup> de chaque point mémoire et donne des résultats de sortie sous forme de graphiques et de cartographies. Ce logiciel fait également le lien entre l'adresse des données (Lignes et Colonnes) et les différents types de sous-blocs dans le but d'obtenir des graphiques par niveau de métal. Le calcul de la TFD sert à obtenir la fréquence et l'amplitude des harmoniques les plus élevés. En effet, si un point mémoire présente une dérive de sa fréquence d'oscillation notée  $x$ , cette dérive devrait être de  $3x$  sur l'harmonique 3, de  $5x$  sur l'harmonique 5 et ainsi de suite. Par conséquent, pour une mesure de fréquence avec une précision donnée, plus l'harmonique mesuré est de rang élevé, plus la mesure de la dérive est précise. Quelques chiffres importants sont donnés ci-dessous. Pour une plaquette ayant un rendement de 100%, il faut :

- 4h pour tester les 150 puces,
- 1.4Go pour stocker les signaux de sortie des 150x4096 cellules OSRAM,
- 1h15 pour calculer les TFD des 150\*4096 cellules OSRAM (Intel® Pentium® M 1.6GHz et 512Mo de RAM).

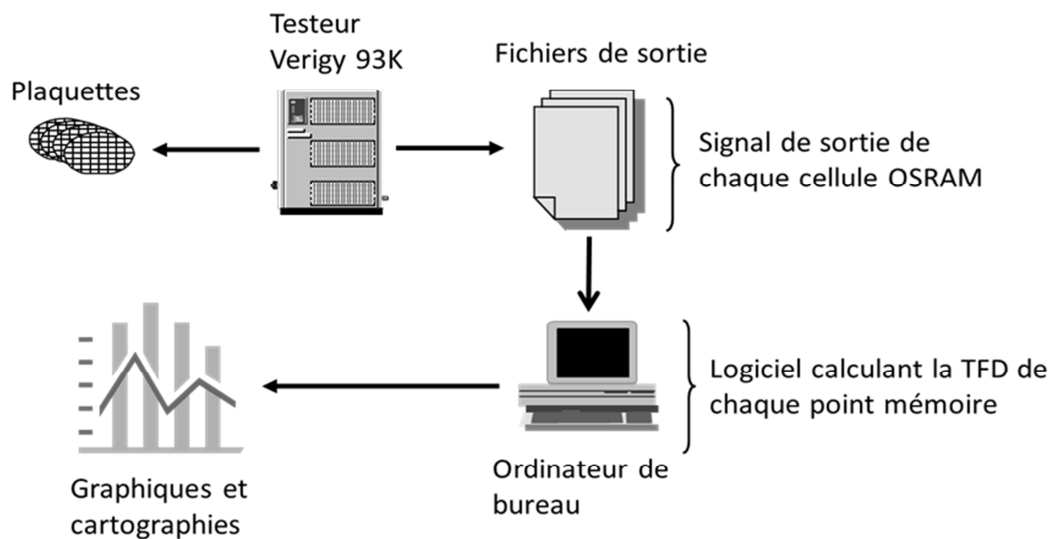


Fig. III-12 : Déroulement du test et analyses de résultats de test

## 2 Traitement des données issues du test

Le programme de test génère une énorme quantité de données dont le traitement est long et complexe. Par conséquent, il est choisi de développer un outil de traitement et d'analyse spécifique au RAPITAC. Ce dernier permet d'automatiser le processus d'analyse des données, pour éviter de reproduire des centaines de fois les mêmes manipulations.

<sup>80</sup> Transformée de Fourier Discrète : algorithme utilisé pour extraire d'un signal l'amplitude et la fréquence des harmoniques

La phase de mesure paramétrique étant la même que celles des TEG logiques, seuls les résultats de lecture/écriture du plan mémoire et des signaux d'oscillation sont traités dans cette partie.

L'outil de traitement des données est développé en langage Java. Ce dernier permet de générer facilement des interfaces graphiques et dispose d'une bibliothèque de code en libre accès très riche. L'algorithme de traitement des données est composé de deux parties, le traitement des données fonctionnelles et le traitement des données paramétriques.

#### a. Traitement et représentation des données fonctionnelles

Le programme de test génère, pour chaque puce testée, quatre résultats de lecture correspondant aux quatre modes de programmation du plan mémoire : tout à '0' (All0), tout à '1' (All1), damier (CHK), et damier inversé (CHKI). Pour chacun des 4096 points mémoire, le logiciel enregistre les informations fonctionnelles (« PASS » ou « FAIL », ligne, colonne et puce testée) sur une plaquette complète.

#### b. Traitement des données paramétriques

Les données paramétriques sont les signaux d'oscillation des points mémoire. Pour rappel, 4096 signaux sont extraits pour une puce, puis échantillonnés à la fréquence de 320MHz. Chaque signal est décrit en ASCII<sup>81</sup> et comporte 1024 échantillons.

Puis, un algorithme de traitement du signal est utilisé de manière à représenter la TFD des signaux d'une puce. Les signatures suivantes sont ensuite extraites pour chaque puce :

- L'amplitude de la composante continue,
- La fréquence du fondamental,
- L'amplitude du fondamental,
- La fréquence de l'harmonique 3,
- L'amplitude de l'harmonique 3,
- ...

#### c. Représentation des données paramétriques

La Fig. III-13 est une impression d'écran de l'onglet « cartographie analogique » du logiciel de traitement de données. Cet onglet représente les composantes des TFD de tous les signaux d'oscillation d'une puce (4096 cellules). Chaque image de la première ligne de cartographie correspond aux amplitudes des harmoniques, et celle de la ligne du bas à leurs fréquences. Chaque point adressable de la cartographie correspond à une cellule, et plus l'amplitude ou la fréquence est élevée, plus la cellule est représentée en clair. Comme expliqué ultérieurement, le plan mémoire est organisé en blocs avec des oscillateurs en anneau routés avec différents niveaux de métaux. Par conséquent, des différences de fréquences d'oscillation peuvent être observées sur les cartographies qui montrent un effet damier. Comme il avait été anticipé, le damier se distingue plus facilement sur les harmoniques de rang élevés.

---

<sup>81</sup> American Standard Code for Information Interchange: Norme de codage de caractère informatique la plus répandue

Il apparait également qu'un gradient de fréquence, suivant l'axe des Y, est présent sur les puces. Ceci est attribué à une mauvaise distribution des canaux d'alimentation. En effet, les rails d'alimentation sont connectés au haut de la structure et se propagent de haut en bas. Du fait de la chute de tension le long de ces rails d'alimentation, les cellules du bas sont alimentées avec une tension plus faible et donc oscillent plus lentement. Pour corriger ce problème, une correction peut être apportée sur la table, sélectionnée dans l'onglet table de l'outil, en appuyant sur le bouton « correction ». Ce bouton sert à recalculer la fréquence d'oscillation de chaque cellule de la table sélectionnée, en fonction de la ligne sur laquelle elle se trouve. Une correction linéaire est alors effectuée, par régression par rapport au delta entre le haut et le bas de la puce. La cartographie de la puce est alors redessinée avec ces nouvelles valeurs. Un exemple est donné Fig. III-13 où, l'avant dernière cartographie de la deuxième ligne (noté « harm4Hz ») correspond à l'harmonique sélectionné, et la dernière de cette ligne représente la même cartographie après correction. Sur cette cartographie, le damier peut être observé jusqu'en bas de la cellule. Ceci montre que la correction a fortement réduit l'effet du gradient de tension d'alimentation.

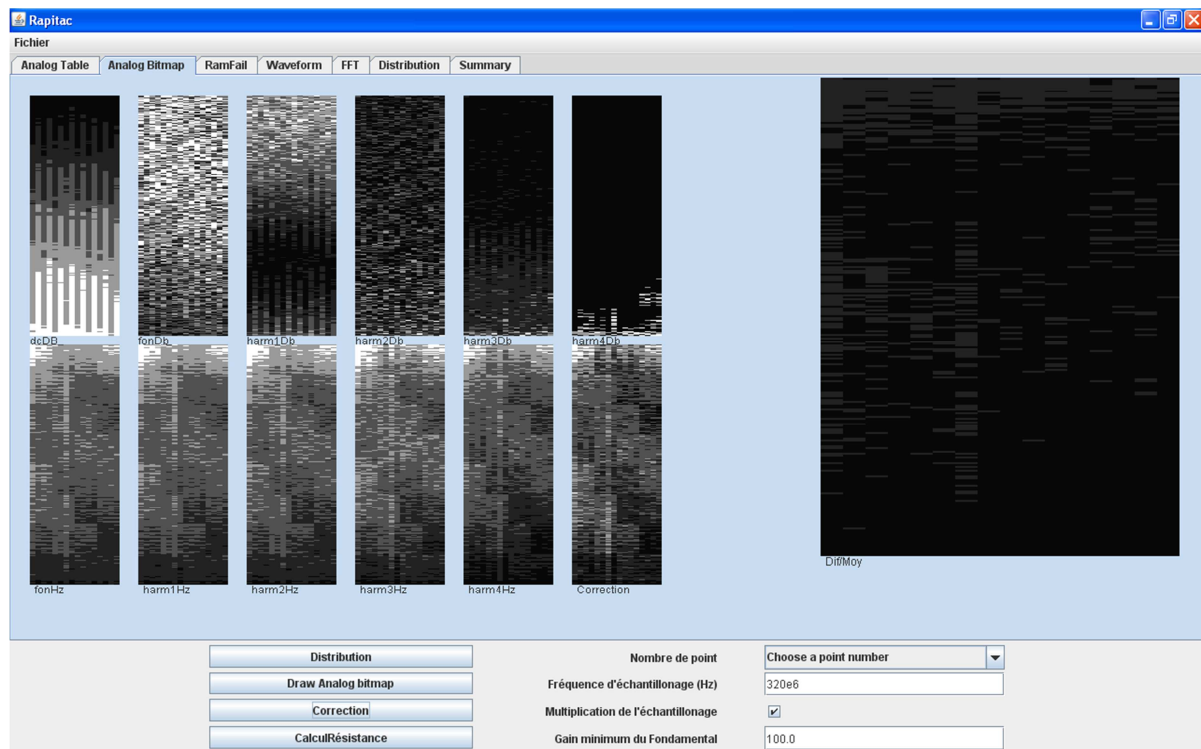


Fig. III-13 : Représentation graphique des différentes composantes de la TFD d'une puce RAPITAC

Cependant, en raison des différences de fréquences d'oscillation de chaque type de blocs, une petite variation intra-puce du procédé de fabrication ne peut pas être détectée. Dans le but de pouvoir détecter ces variations, il est possible de tracer des cartographies relatives à « la moyenne du type ». Ainsi pour chaque point mémoire  $OSRAM_{Mx}$ , la fréquence relative à la moyenne des fréquences d'oscillation de tous les blocs  $Mx$  ( $F_{avr}$ ) est calculée à l'aide de l'équation (III-1). Dans cette équation,  $F$  représente la fréquence d'oscillation mesurée. Ce lissage est représenté sur la cartographie (noté « DifMoy ») à droite de la Fig. III-13.

Grâce à cela, l'effet damier a complètement disparu et des effets de zone peuvent être visualisés sur la puce.

$$F_{avr} = F - \frac{\sum F_{OSRAM_{Mx}}}{\# OSRAM_{Mx}} \quad (III-1)$$

La Fig. III-14a et la Fig. III-14b sont respectivement, une distribution cumulée et un « box plot » des fréquences d'oscillation des 4 types de cellules OSRAM. En comparaison avec une représentation de type cartographie, ces distributions permettent d'évaluer la dispersion fréquentielle d'un type de cellules et de détecter quel niveau de métal a subi une variation. En effet, plus la distribution cumulée est étalée (ou le box plot est haut), plus il y a de variations intra-puce.

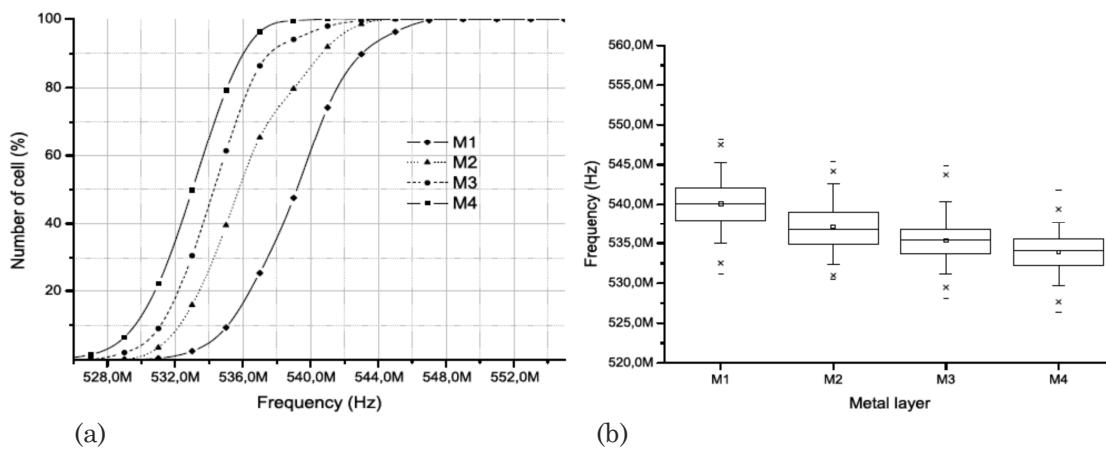


Fig. III-14 : Distribution cumulée (a) et box plot (b) par type de cellule OSRAM

Tous les résultats collectés au niveau puce, sont ensuite reportés au niveau plaquette dans le but de pouvoir comparer les puces entre elles ou d'observer des variations puces à puces. La cartographie de la Fig. III-15 représente les moyennes des fréquences d'oscillation des puces de la plaquette. Plus une puce de la cartographie est claire, plus les fréquences d'oscillation de la puce sont élevées. Les cases vides sont dues aux puces non fonctionnelles pour lesquelles le mode d'oscillation n'est pas testé.

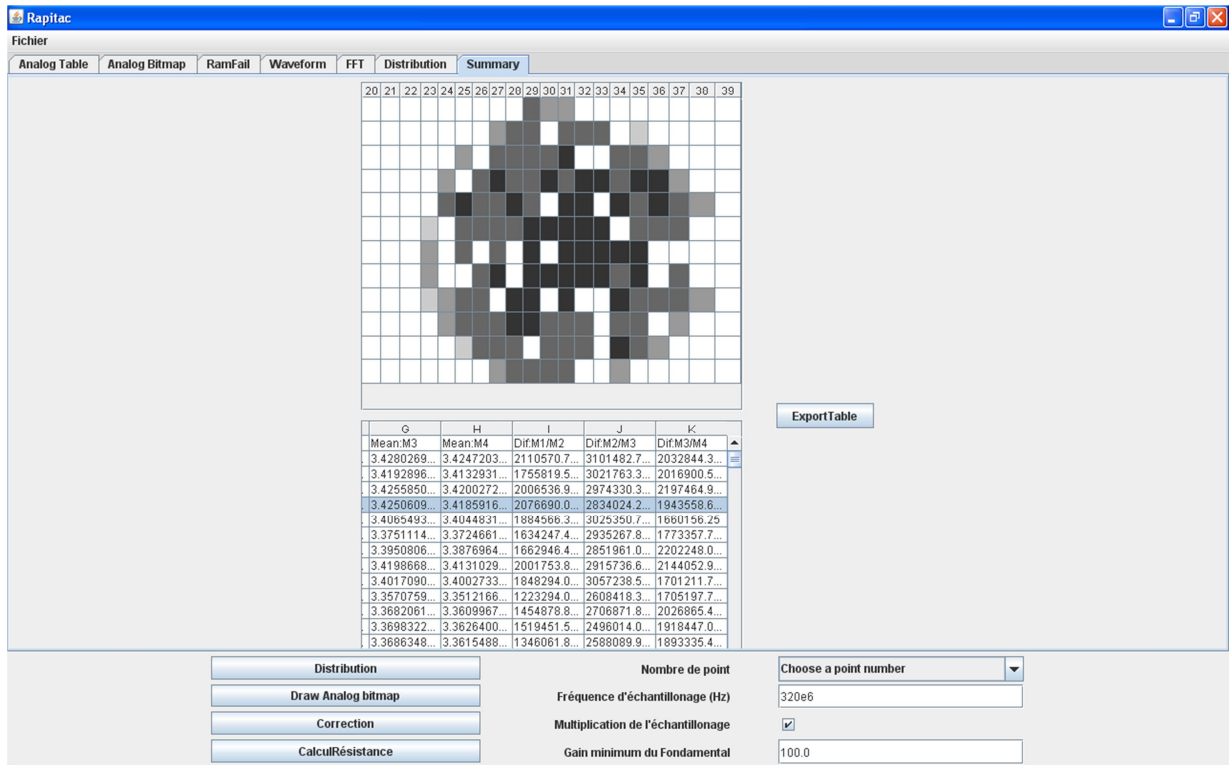


Fig. III-15 : Représentation graphique des fréquences d'oscillation le long de la plaquette (a) et table résumant les différents comportements de chaque puce (b)

#### D. Caractérisation des variations du procédé de fabrication

Le flot de traitement de données présenté précédemment montre comment détecter les variations grâce aux données paramétriques. Cependant, en analysant séparément les fréquences d'oscillation des différents types de blocs, il est possible, comme sur la Fig. III-16, de détecter des variations du procédé de fabrication. Cependant la variation de la fréquence d'oscillation des différents blocs peut provenir du BEOL comme du FEOL. Une analyse croisée doit alors être effectuée pour caractériser cette variation. En effet, la première hypothèse est que, si la variation du procédé de fabrication est causée par le FEOL, les fréquences d'oscillation des différents types de bloc vont afficher la même variation car leur FEOL est le même. Par exemple, de larges distributions peuvent être observées sur tous les types de blocs de la Fig. III-16. Au contraire si la variation provient, par exemple du métal 2, seuls les blocs M2 vont voir leur fréquence d'oscillation bouger.

Ce problème peut être contourné en analysant les différences de fréquences d'oscillation entre les types de point mémoire. Grâce à cela, il est possible de tracer la distribution des différences de fréquences d'oscillation entre chaque type de cellule OSRAM. Ces différences sont donc identifiées comme Dif(M1M2), Dif(M2M3) et Dif(M3M4).

En considérant que les différents types de cellules sont conçus avec le même FEOL et avec des longueurs de métaux identiques, les seules différences en terme de résistance du BEOL proviennent des via ajoutés et de la résistance du métal utilisé pour l'interconnexion. Par exemple, la différence de fréquence d'oscillation entre une cellule OSRAM<sub>M1</sub> et une cellule OSRAM<sub>M2</sub> est due à la



résistance du via 1 ( $R_{V1}$ ) et à la différence de résistivité entre le métal 1 et le métal 2 (cf. Fig. III-10). La Fig. III-17 représente la distribution schématisée des fréquences d'oscillation de chaque type de cellule OSRAM. Cette distribution permet d'illustrer la construction des différences de fréquence d'oscillation.

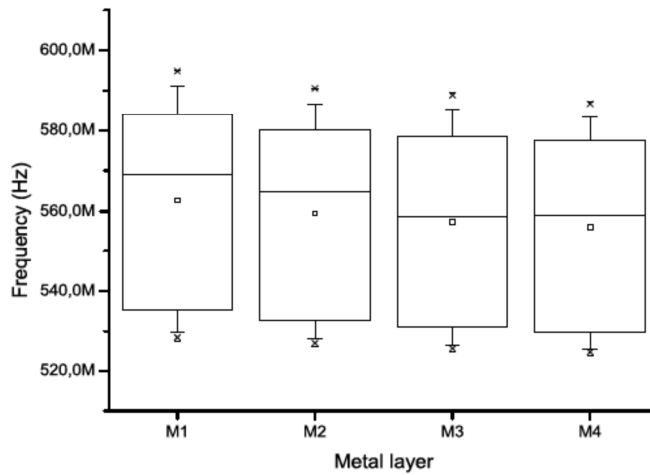


Fig. III-16 : « Box plot » des fréquences d'oscillation d'une plaquette de silicium par type d'OSRAM

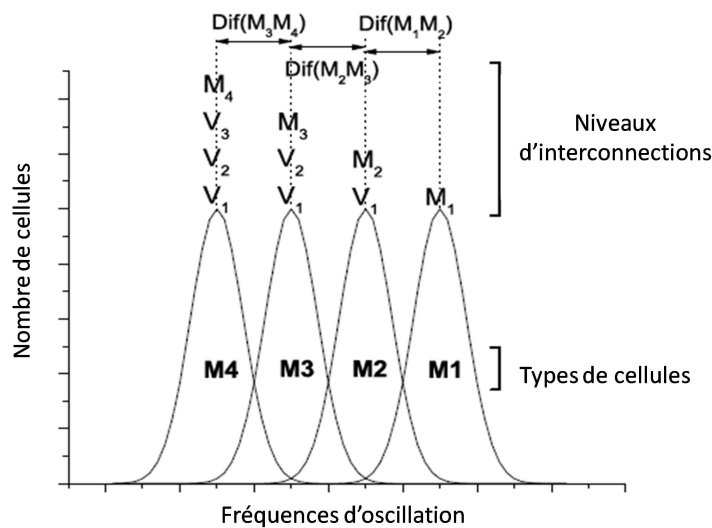


Fig. III-17 : Distribution schématisée des fréquences d'oscillation par type de cellule

A l'aide de cette approche, l'impact des variations du BEOL peut être dissocié du FEOL et peut être calculé. Dans ce but, une variable  $R_{VxMx+1}$  est introduite pour représenter la variation de résistance du couple via  $x$  / métal  $x+1$ . Cette approche est adoptée puisque la technologie utilisée pour réaliser cette structure est une technologie avec un BEOL cuivre où la ligne et les via correspondants sont conçus lors d'une seule et même opération. Le Tableau III-1 résume l'impact des variations de chaque couple  $R_{VxMx+1}$  sur les différences des fréquences d'oscillation. Par exemple, une augmentation de la résistance du via 1 ou du métal 2 doit se traduire par une diminution de  $Dif(M1M2)$  et une augmentation de  $Dif(M2M3)$ .

Tableau III-1: Variations des  $Dif(M_x M_{x+1})$  en fonction de la variation des couples via / métaux

Variation des interconnexions	Variation de $Dif(M1M2)$	Variation de $Dif(M1M2)$	Variation de $Dif(M1M2)$
$R_{M1} \nearrow$	$\nearrow$	$\rightarrow$	$\rightarrow$
$R_{M1} \searrow$	$\searrow$	$\rightarrow$	$\rightarrow$
$R_{V1M2} \nearrow$	$\searrow$	$\nearrow$	$\rightarrow$
$R_{V1M2} \searrow$	$\nearrow$	$\searrow$	$\rightarrow$
$R_{V2M3} \nearrow$	$\rightarrow$	$\searrow$	$\nearrow$
$R_{V2M3} \searrow$	$\rightarrow$	$\nearrow$	$\searrow$
$R_{V3M4} \nearrow$	$\rightarrow$	$\rightarrow$	$\searrow$
$R_{V3M4} \searrow$	$\rightarrow$	$\rightarrow$	$\nearrow$

De plus, les différences peuvent être exprimées en fonction des résistances d'interconnexions. Les équations linéaires (III-2), (III-3) et (III-4) peuvent être obtenues à l'aide d'outils de régressions linéaires à variables multiples. Dans le but d'obtenir des valeurs de coefficients  $a_x$ ,  $b_x$  et  $c_x$  précises et de vérifier ces expressions, des FEM ont été réalisées sur deux plaquettes de silicium. La première plaquette a été produite avec une FEM produisant une variation linéaire des CD du couple via1/métal2 alors que la deuxième plaquette comporte la même variation au niveau du couple via3/métal4.

$$Dif(M_1 M_2) = a_1 \times R_{M1} + b_1 \times R_{V1M2} + c_1 \quad (III-2)$$

$$Dif(M_2 M_3) = a_2 \times R_{M2} + b_2 \times R_{V2M3} + c_2 \quad (III-3)$$

$$Dif(M_3 M_4) = a_3 \times R_{M3} + b_3 \times R_{V3M4} + c_3 \quad (III-4)$$

Dans un premier temps, l'effet des variations de CD sur les résistances des couples via/métaux est évalué. Pour cela, ces plaquettes sont mesurées au test paramétrique en mode « full mapping », c'est-à-dire que tous les sites de la plaquette sont mesurés. La Fig. III-18 confirme la bonne corrélation entre les mesures effectuées sur la ligne de production après la photolithographie de la ligne 2 (mesures optiques de largeur de ligne) et la résistance de cette même ligne. Un effet de marche peut être observé sur le graphique, ceci est dû à l'erreur de mesure du CD de la ligne (précision de 10nm), alors que la mesure de la résistance est plus précise (précision de l'ordre de l'ohm). La mesure du PT reflétant mieux la dimension réelle de la ligne, c'est cette mesure qui sera utilisée pour évaluer la capacité du RAPITAC à caractériser les variations du BEOL. Les mesures du test paramétrique et les mesures du RAPITAC sont donc mises en commun pour évaluer la corrélation entre les deux. Ceci est une opération souvent complexe car les systèmes de coordonnées et souvent, l'orientation de la plaquette ne sont pas les mêmes. Concernant le RAPITAC, les mesures prises en compte sont les valeurs moyennes des  $Dif(M_x M_{x+1})$  calculées pour chaque puce.

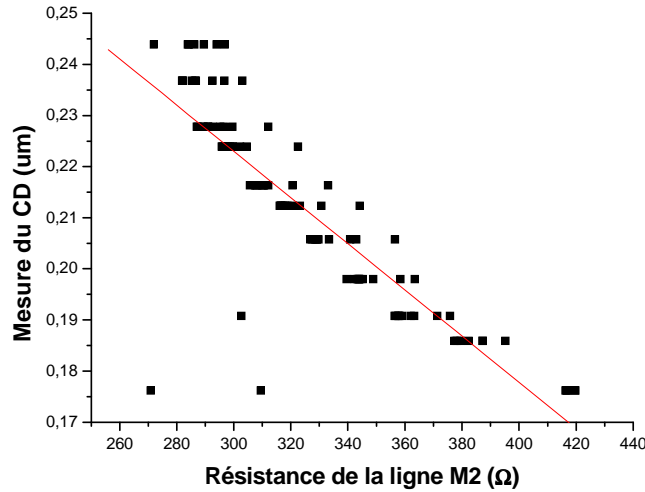


Fig. III-18 : Corrélation entre les mesures effectuées après la photolithographie de la ligne 2 et la résistance de la ligne 2 mesurée au test paramétrique

La Fig. III-19a montre la variation des différences  $Dif(M_1M_2)$  et  $Dif(M_2M_3)$ , en fonction de la résistance du couple  $R_{V1/M2}$  induite par la FEM. Il apparaît qu'une diminution de couple produit une augmentation de  $Dif(M_1M_2)$  et une diminution de  $Dif(M_2M_3)$ . Ceci confirme les estimations reportées dans la Tableau III-1. De la même façon, la Fig. III-19b met en évidence l'effet de la résistance du couple  $R_{V3/métal4}$  sur la  $Dif(M_3M_4)$ . Toujours en concordance avec Fig. III-19b, une diminution de  $R_{V3M4}$  entraîne une augmentation de  $Dif(M_3M_4)$ .

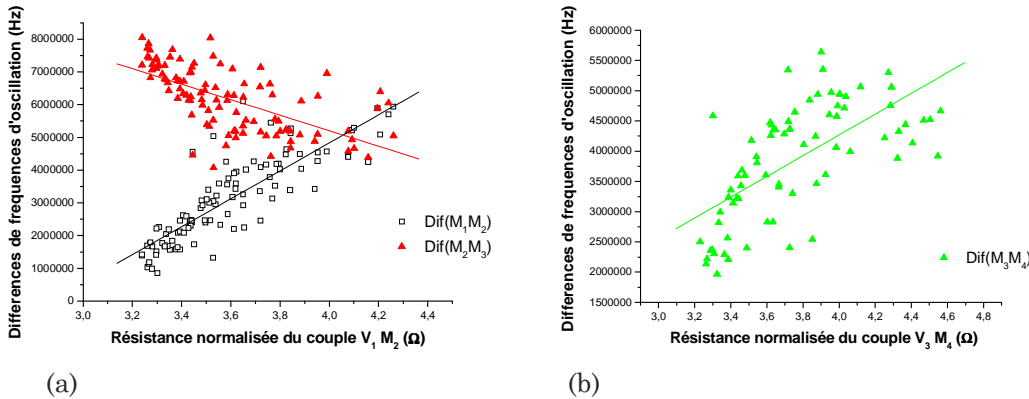


Fig. III-19 :  $Dif(M_1M_2)$  et  $Dif(M_2M_3)$  en fonction de la résistance du couple  $R_{V1/métal2}$  (a) et  $Dif(M_3M_4)$  en fonction de la résistance du couple  $R_{V3/métal4}$

Grâce aux bonnes corrélations obtenues sur la Fig. III-19a, il est maintenant possible d'inverser le système des équations (III-2), (III-3) et (III-4) pour exprimer la résistance du couple  $R_{VxM_{x+1}}$  en fonction des différences (III-5).

$$R_{V1M2} = a \times Dif(M_1M_2) + b \times Dif(M_2M_3) + c \quad (III-5)$$

Une régression multiple est calculée dans le but d'obtenir les coefficients  $a$ ,  $b$  et  $c$ . Finalement, la Fig. III-20 montre la bonne corrélation entre les résistances du couple via1/métal2 estimées grâce aux mesures fréquentielles et aux mesures du test paramétrique. Ceci valide le fait que le calcul des différences des fréquences d'oscillation permet une estimation précise de la variation de résistance. Cependant, cette approche nécessite l'emploi de FEM pour obtenir les coefficients et ne peut pas être appliquée au premier et au dernier niveau de métal.

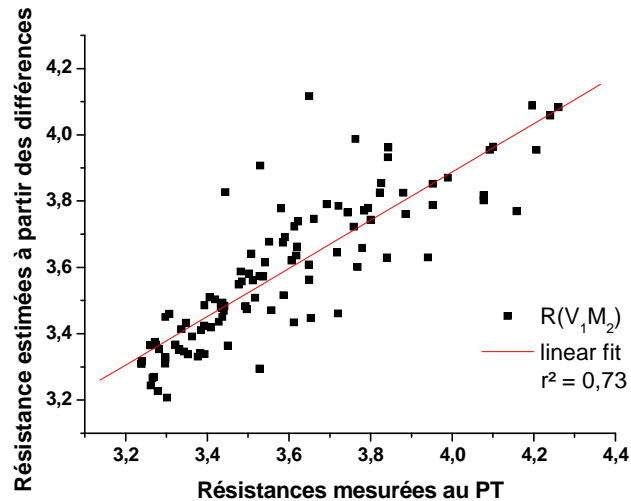


Fig. III-20 : Corrélation entre les résistances du couple via1/métal2 estimées grâce aux mesures fréquentielles et celles mesurées au test paramétrique

### III. Conclusion

Le RAPITAC est une macro-cellule de test, conçue dans le cadre d'une collaboration entre l'IMN2P, ST ROUSSET et ST CROLLES, avec pour objectif de combiner les fonctionnalités des mémoires SRAM et des oscillateurs en anneau. Tout d'abord, une brève description de l'architecture de la structure est donnée en partant de la cellule SRAM oscillante (OSRAM) jusqu'au plan mémoire composé de différents types de cellules OSRAM. Le programme de test est ensuite présenté et décrit. Durant ce test, le signal d'oscillation de chaque point mémoire est extrait.

Il a été choisi de développer un outil d'analyse spécifique au RAPITAC. Cet outil, développé en langage java permet le traitement des données fonctionnelles et le traitement des signaux d'oscillation. Une fonctionnalité intéressante de l'outil est l'exploitation des signaux d'oscillation provenant des cellules OSRAM. Cet outil réalise le calcul de la TFD de chaque cellule et fournit des cartographies analogiques (composantes de la TFD) de puces et de plaquettes.

La dernière partie de ce chapitre a été dédiée au développement d'une méthode de caractérisation des variations du procédé de fabrication. En effet, il apparaît que l'étude des différences de fréquences d'oscillation, entre les types de cellules OSRAM, permet de savoir quel niveau de métal subit une variation. De plus, il a été possible d'estimer la valeur de la résistance d'un couple via/métal.

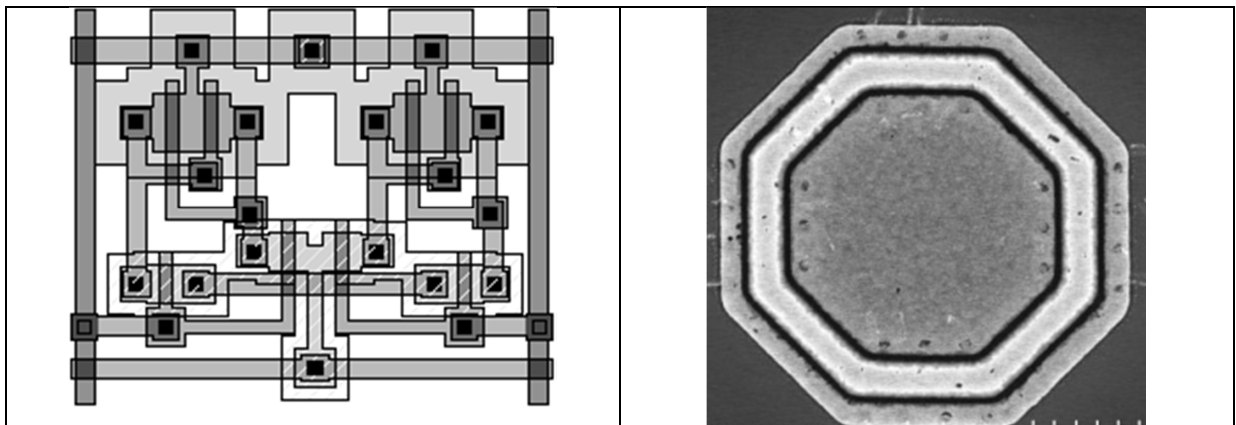
---

## Chapitre IV : FIABILITE DES CIRCUITS ANALOGIQUES

Le but de ce chapitre est d'analyser et d'optimiser l'appariement des transistors MOS destinés aux applications analogiques faibles consommation, pour une technologie mémoire embarquée non volatile 90nm (« 90nm embedded Non Volatile Memory »).

Ce chapitre commence par un bref rappel sur les sources de fluctuations à l'origine des variations aléatoires des paramètres électriques des dispositifs avec une présentation des différentes techniques de mesure et d'analyse des fluctuations. Dans le cas des technologies incluant des mémoires non volatiles embarquées, des étapes de fabrication supplémentaires peuvent créer des transistors en bord d'active responsable de l'effet « hump ». Ainsi, dans une deuxième partie, l'effet « hump » est présenté en mettant en évidence ses conséquences sur la dégradation de l'appariement des transistors MOS. Pour contrer cet effet parasite, plusieurs solutions seront proposées.

Enfin, une étude de fiabilité est réalisée sous stress porteurs chauds. Cette étude met en évidence une dégradation de l'appariement de la tension de seuil des transistors. Une étude de fiabilité comparative est présentée entre les transistors octogonaux et les transistors standards. Il est ainsi démontré que les transistors octogonaux se dégradent moins que les transistors standards.



<b>I.</b>	<b>FLUCTUATIONS LOCALES DANS LES TRANSISTORS MOS.....</b>	<b>102</b>
<b>A.</b>	<b>FLUCTUATION DANS LE TRANSISTOR MOS .....</b>	<b>102</b>
1	Présentation .....	102
2	Sources de fluctuations locales liées au procédé de fabrication .....	103
a.	<i>Fluctuations liées au dopage</i> .....	103
b.	<i>Fluctuations liées à la définition de la grille</i> .....	105
c.	<i>Fluctuations liées aux paramètres géométriques des transistors</i> .....	105
d.	<i>Bilan des fluctuations locales dans un transistor MOS</i> .....	106
<b>B.</b>	<b>MESURE DES FLUCTUATIONS LOCALES DANS LE TRANSISTOR MOS .....</b>	<b>107</b>
1	Structures de mesure .....	107
2	Méthodes de mesure de l'appariement .....	108
3	Traitement statistique des résultats d'appariement .....	109
4	Extraction du paramètre d'appariement .....	110
<b>II.</b>	<b>CONTRIBUTION A L'AMELIORATION DE L'APPARIEMENT DES</b>	
	<b>CIRCUITS ANALOGIQUES POLARISEES SOUS LE SEUIL.....</b>	<b>112</b>
<b>A.</b>	<b>CONCEPTION DE CIRCUITS ANALOGIQUES FAIBLE CONSOMMATION.....</b>	<b>112</b>
<b>B.</b>	<b>L'EFFET « HUMP ».....</b>	<b>113</b>
1	Présentation .....	113
2	Modélisation de l'effet hump .....	115
3	Impact de l'effet « hump » sur l'appariement en tension.....	115
4	Etude dimensionnelle.....	116
5	Etude en fonction de l'effet substrat .....	117
<b>C.</b>	<b>SOLUTIONS POUR CONTRER L'EFFET « HUMP ».....</b>	<b>119</b>
1	Solution basée sur des transistors « multi-doigts » .....	119
2	Autres solutions.....	121
<b>III.</b>	<b>CONTRIBUTION A L'AMELIORATION DE L'APPARIEMENT EN</b>	
	<b>FONCTION DU STRESS PORTEURS CHAUDS .....</b>	<b>122</b>
<b>A.</b>	<b>PRESENTATION DES DIFFERENTS STRESS ELECTRIQUES.....</b>	<b>122</b>
1	Principe du stress par injection de porteurs chauds (HCI) .....	122
2	Protocole expérimental.....	122
<b>B.</b>	<b>DEGRADATION DES TRANSISTORS NMOS ET PMOS SOUS STRESS HCI .</b>	<b>123</b>
1	Dégradation de la tension de seuil $V_T$ .....	123
2	Dégradation de l'appariement de la tension de seuil $\Delta V_T$ .....	123
<b>C.</b>	<b>ETUDE DES TRANSISTORS OCTOGONAUX .....</b>	<b>124</b>
1	Présentation des transistors octogonaux .....	124
2	Comparaison entre transistors standard et octogonal .....	125
a.	<i>Dégradation de la tension de seuil</i> .....	126
b.	<i>Dégradation de l'appariement de la tension de seuil</i> .....	127
3	Inversion des sources et drains des transistors octogonaux.....	127
<b>IV.</b>	<b>CONCLUSION.....</b>	<b>129</b>

# I. Fluctuations locales dans les transistors MOS

## A. Fluctuation dans le transistor MOS

### 1 Présentation

L'appariement des transistors MOS est très important en conception analogique, notamment dans le cas des circuits analogiques comportant des miroirs de courant et des paires différentielles. Le désappariement des dispositifs peut également se faire ressentir sur d'autres circuits comme les convertisseurs Analogiques-Numériques (« ADC ») [Pelg98] [Doll01], les convertisseurs Numériques-Analogiques (« DAC ») [Bast98] en impactant le rendement ou la précision par exemple, mais aussi la vitesse et la puissance consommée des références de tension « Bandgap » [Gupt02] [Gupt05] [Seng05]. Les sources de variations sont présentes à différents niveaux de la chaîne de fabrication comme le montre la Fig. IV-1 L'amplitude de ces variations ou fluctuations dépend de la distance séparant les dispositifs.

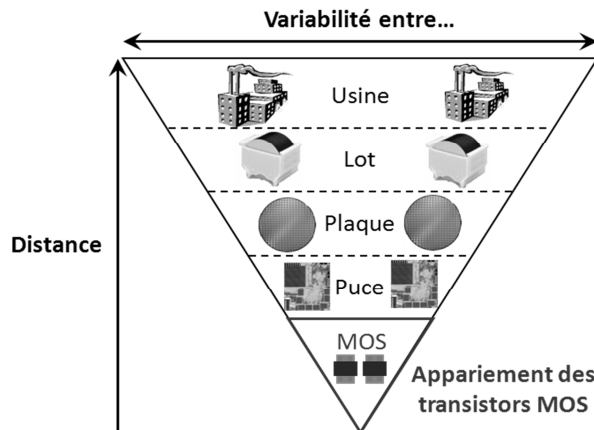


Fig. IV-1 : Variabilité des dispositifs suivant la distance

Les variations entre deux usines ou entre deux lots sont plutôt importantes et peuvent être dues à des différences ou dérives d'un ou plusieurs équipements lors du procédé de fabrication. Différentes positions dans un four par exemple sont sources de variabilité de plaque à plaque. Des effets de gradients sur les plaques peuvent augmenter les fluctuations des paramètres électriques. Ces phénomènes de gradient se retrouvent également au sein d'une même puce de manière plus ou moins importante suivant la distance séparant les deux dispositifs (fluctuations intra-puce). Toutes les variations précédentes sont des fluctuations dites « systématiques ». Enfin, les fluctuations entre deux dispositifs dessinés l'un à côté de l'autre sont dites « locales » et sont aléatoires.

Etant donné l'importance que peuvent avoir les fluctuations locales dans les performances des circuits analogiques, le nombre d'articles sur le sujet n'a cessé d'augmenter depuis maintenant près de 40 ans. Un des premiers travaux relatant des problèmes de fluctuations du nombre de dopants dans le canal induisant des variations de tension de seuil a été publié en 1972 [Hoen72]. Dans [Keye75], un modèle de prédiction des fluctuations de tension de seuil est proposé en considérant une distribution de Poisson des dopants dans le canal. La première loi de dépendance dimensionnelle a été étudiée dans [McCr81] en

s'intéressant aux fluctuations locales des capacités MOS. Ces dépendances sont ensuite expliquées dans [Shyu82]. C'est à partir de cette publication que les fluctuations systématiques sont associées à la moyenne de la distribution d'un paramètre et les fluctuations stochastiques à sa déviation standard. Peu après dans [Shyu84], un modèle d'appariement est proposé pour les capacités MOS et transistors MOS. La loi donnant l'amplitude des fluctuations locales en fonction des dimensions d'un dispositif est observée expérimentalement dans [Laks86]. Finalement, c'est dans [Pelg89], un des articles de référence dans le domaine de l'appariement de dispositifs, que cette loi de dépendance dimensionnelle des fluctuations locales est démontrée en séparant les variations locales et globales. Ainsi, les fluctuations locales d'un paramètre P de deux dispositifs appariés (notées  $\sigma_{\Delta P}$ ) sont inversement proportionnelles à la racine de la surface du dispositif étudié et de la technologie par l'intermédiaire du paramètre  $A_P$  (IV-1). Par la suite, la plupart des travaux relatant des fluctuations locales des dispositifs se sont inspirés de cette loi dimensionnelle<sup>1</sup>.

$$\sigma_{\Delta P} = \frac{A_P}{\sqrt{WL}} \quad (\text{IV-1})$$

## 2 Sources de fluctuations locales liées au procédé de fabrication

La plupart des étapes d'un procédé de fabrication peuvent être sources de fluctuation. Cela est particulièrement vrai dans le cas des étapes liées au dopage qui sont souvent considérées comme étant les principaux contributeurs des fluctuations des paramètres des transistors MOS et par conséquent de leur appariement. Le deuxième principal contributeur est lié à la définition de la grille. Enfin, les fluctuations liées aux paramètres géométriques du transistor MOS sont également à prendre en compte. Tous ces phénomènes étant aléatoires pour chaque transistor MOS, ils jouent un rôle non négligeable sur leur appariement.

### a. Fluctuations liées au dopage

Beaucoup de travaux ont été consacrés à l'étude des fluctuations de tension de seuil liées au dopage du canal et cela, principalement à cause de la variation aléatoire du nombre de dopants qui suit une distribution Gaussienne [Mizu96]. Dans [Take97], une dépendance en fonction de la profondeur des dopants dans la charge de désertion est mise en évidence. La diminution de la contribution des dopants du canal sur les fluctuations de tension de seuil en fonction de la profondeur est également démontrée dans [Stol98].

Ainsi, en considérant un nombre aléatoire de dopants du canal contrôlant la charge de désertion  $Q_{sc}$ , les fluctuations de la tension de seuil peuvent se mettre sous la forme (IV-2).

---

<sup>1</sup> Également appelée « loi de Pelgrom »



$$\sigma_{VT} = \frac{1}{C_{ox}} \sqrt{\frac{q \cdot Q_{SC}}{3WL}} \quad (IV-2)$$

Cette expression permet de voir l'impact du dopage canal sur le paramètre d'appariement (noté ici  $A_{VTc}$ ) de la tension de seuil d'après l'équation (IV-3). L'appariement se dégrade avec l'augmentation du dopage canal.

$$A_{VTc} = \frac{T_{ox}}{\epsilon_{ox}} \cdot \frac{\sqrt{2}}{\sqrt{3}} \cdot \sqrt[4]{2q^3 \epsilon_{Si} N_C (2\phi_F - V_B)} \quad (IV-3)$$

Cette dépendance en  $N_C^{0.25}$  du paramètre d'appariement de la tension de seuil n'est cependant pas valide pour toutes les géométries de transistors MOS [Difr03a]. Cette déviation par rapport à la loi de surface est principalement due à l'effet canal court SCE<sup>12</sup> avec une dégradation accentuée pour les transistors courts [McGi04].

Pour limiter l'effet canal court, des régions très dopées près du canal autour des diffusions de drain et de source appelées poches ou « halos » sont utilisées. Cependant, les étapes d'implantation nécessaires à la réalisation de ces poches sont elles-mêmes sources de fluctuations pouvant dégrader l'appariement des transistors MOS.

Une modélisation de l'impact de ces poches basée sur les contributions du canal et des poches en fonction de leur dopage et de leur longueur a été proposée dans [Difr00] [Rios02]. Le paramètre d'appariement  $A_{VTc+p}$  (canal + poches) pour un transistor MOS avec poches peut s'écrire par la relation (IV-4).

$$A_{VTc+p} = \frac{T_{ox}}{\epsilon_{ox}} \cdot \frac{\sqrt{2}}{\sqrt{3}} \cdot \sqrt{qQ_{SCp} \cdot \frac{2L_p}{L} + qQ_{SC} \cdot \frac{L - 2L_p}{L}} \quad (IV-4)$$

Avec  $Q_{SCp}$  et  $L_p$  représentant respectivement la charge de désertion et la longueur des poches et  $Q_{SC}$  et  $L$  représentant les mêmes grandeurs appliquées au canal.

Cependant, avec l'augmentation de la longueur  $L$  des transistors MOS, la différence de dopage entre le canal (faiblement dopé) et les poches (fortement dopées) est importante et donc non-uniforme. Cette différence devenant vraiment significative, les barrières de potentiels ainsi formées côté source et côté drain contrôlent les variations de  $V_T$  et ceci indépendamment de la longueur  $L$  [Cath08a]. L'impact des poches est aussi dépendant de la polarisation de grille qui vient abaisser les hauteurs des barrières de potentiel et ainsi diminuer les fluctuations de tension de seuil (l'ensemble canal + poches étant quasi uniforme) [Cath08a]. Ces effets sont confirmés en technologie 32nm avec des grilles métalliques sur isolant à haute permittivité (« high-k ») [Hook10].

Un modèle d'appariement prenant en compte ces fluctuations de tension de seuil en fonction de la polarisation de grille et pour les transistors longs est proposé dans [Cath09]. Ce dernier modèle ayant également des limitations (pas de diminution de l'appariement pour les transistors très longs et représentation

---

<sup>2</sup> Short Channel Effect

seulement qualitative), un modèle physique amélioré a été développé. Ce modèle prend en compte tous les effets cités précédemment [Mezz10b].

#### b. Fluctuations liées à la définition de la grille

L'aspect granulaire de la grille est un contributeur important des fluctuations locales des transistors MOS. Tout d'abord, les variations aléatoires du nombre de dopants dans la grille sont une des raisons de la dégradation de l'appariement des transistors MOS [Asen98] [Asen99]. Un modèle prenant en compte les fluctuations du nombre de dopants dans la grille en plus des dopants dans le canal ( $A_{V_{Tc+g}}$ ) est présenté en (IV-5) [Difr01].

$$A_{V_{Tc+g}} = \sqrt{\frac{2}{3} \cdot \left(\frac{T_{ox}}{\epsilon_{ox}}\right)^2 q Q_{sc} + (2\phi_F - V_B) \cdot \frac{N_C}{N_G} \cdot \frac{q}{Q_{sc}}} \quad (IV-5)$$

Du fait de la pénétration de dopants dans le canal à travers l'oxyde [Tuin97a] [Yu98] [Schm99] qui se fait essentiellement le long des joints de grains situés dans la grille, le dopage de la grille est par conséquent moins uniforme et certaines zones proche de l'oxyde de grille peuvent se retrouver partiellement désertées de dopants. De plus, la tension de seuil peut fluctuer localement suivant la position le long du canal (sous un grain ou sous un joint de grain) [Asen00]. Cet aspect est également démontré dans d'autres travaux [Cao98] [Kris96].

Partant de l'hypothèse de diffusion des dopants le long des joints de grains, il est concevable de s'attendre à une diminution des fluctuations locales en diminuant la taille des grains par une homogénéisation de la distribution des dopants dans la grille. Dans [Difr01], une amélioration notable de l'appariement des transistors est obtenue en passant d'une grille déposée en silicium amorphe à une grille déposée en silicium poly-cristallin. Après recuit, ces deux procédés conduisent à une grille de type poly-cristalline mais avec une diminution de la taille des grains dans le deuxième cas. A partir de cette observation, un modèle d'appariement de la tension de seuil basé sur les fluctuations aléatoires du nombre de grains a été proposé [Difr03b].

Une autre étude basée sur une grille en poly-silicium de type « nano grains » a démontré une amélioration de l'appariement des transistors MOS [Chen03]. En plus de présenter des grains encore plus petits que dans une grille poly-silicium de type colonnaire, aucune orientation cristalline n'est privilégiée pour les « nano grains » rendant la diffusion des dopants le long des joints de grains encore plus difficile.

#### c. Fluctuations liées aux paramètres géométriques des transistors

Un récapitulatif de l'appariement de la tension de seuil en fonction de l'épaisseur d'oxyde  $T_{ox}$  pour les technologies de  $0.5\mu m$  à  $32nm$  est proposé dans [Mezz11]. Toutes ces données permettent d'obtenir une tendance (IV-6) du paramètre d'appariement confirmée par des simulations atomistiques [Asen00].

$$A_{VT} = 3.2 \times 10^{-3} \cdot N_C^{0.4} \left( T_{ox} + \frac{\epsilon_{ox}}{\epsilon_{Si}} \cdot y_G \right) \quad (IV-6)$$

Un autre facteur responsable de fluctuations des paramètres des transistors MOS est lié à la rugosité de la grille notée LER<sup>3</sup> [Bara07] [Bara08] ayant pour conséquences des variations de longueur de canal. Ainsi, la variance d'un paramètre électrique P modifiée par les fluctuations de longueur à cause de la rugosité de la grille peut se mettre sous la forme (IV-7) [Croo02b].

$$\sigma_{\Delta P}^2 \cong \left( \frac{\partial \Delta P}{\partial L} \right)^2 \sigma_{<L>}^2 \quad (IV-7)$$

Où  $\sigma_{<L>}^2$  représente la déviation standard de la longueur moyenne du transistor MOS. Cet effet en L ( $\Delta L$ ) est considéré comme un facteur limitant majeur pour les futures technologies étant donné qu'il prend une part de plus en plus importante par rapport à la diminution des longueurs de grille [Croo02b] [Croo03] [Asen03].

Enfin, l'importance de prendre en compte les fluctuations aléatoires des états d'interface notamment pour les technologies récentes avec grilles métalliques « high-k » est démontrée dans [Andr09b] et des charges fixes dans l'oxyde [Putr08].

#### d. Bilan des fluctuations locales dans un transistor MOS

Toutes les sources de fluctuations décrites précédemment peuvent se représenter sur la Fig. IV-2.

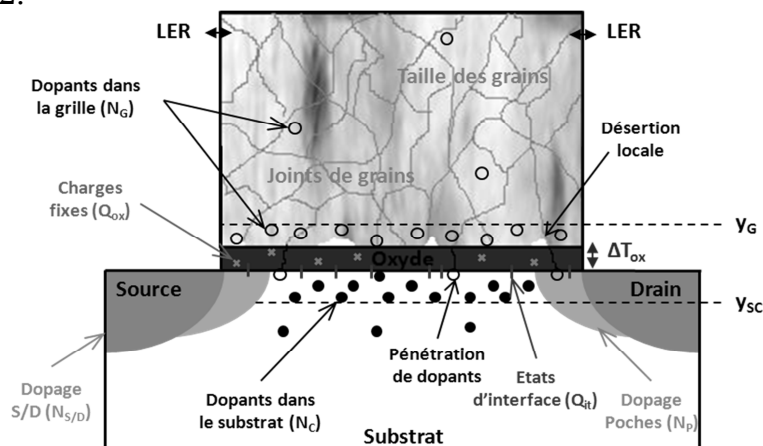


Fig. IV-2 : Représentation des principales sources de variabilité du transistor MOS

Ces différentes contributions ayant un impact sur l'appariement des transistors MOS sont difficiles à classer par ordre d'importance. En effet, le poids de chaque contribution peut varier notamment en fonction du nœud technologique, du type de grille, de l'épaisseur d'oxyde... Cependant, il est commun de citer les fluctuations liées au dopage du canal ( $N_C$ ), des sources et drains ( $N_{S/D}$ ) et des poches ( $N_P$ ) comme principales sources de variabilité [Cath08b] [Mezz11]. Ces fluctuations sont souvent regroupées sous le nom de RDD<sup>4</sup> dans la littérature. Ensuite, les fluctuations liées à la définition de la grille (taille des grains, concentration de dopants dans la grille  $N_G$ , joints de grains, zones localement désertées, pénétration de dopants, etc...) sont la deuxième cause de désappariement des transistors MOS. La rugosité de la grille a, quant à elle, un

<sup>3</sup> Line Edge Roughness

<sup>4</sup> Random Discrete Dopants

impact mineur sur les fluctuations des transistors MOS. Cet effet LER devient cependant important pour les nœuds technologiques avancés (32nm et 22nm) avec la réduction des dimensions [Roy06]. Enfin, les fluctuations des états d'interface à l'interface Si/SiO<sub>2</sub> et des charges fixes dans l'oxyde de grille peuvent impacter l'appariement des transistors MOS [Putr08] [Andr09b].

Pour les futures technologies (32nm, 28nm...), les grilles métalliques « high-k » sont souvent utilisées et, du fait d'une concentration de dopage dans le canal moins importante, peuvent améliorer l'appariement comparativement aux grilles en poly-silicium [Yuan11]. Afin de continuer la miniaturisation des transistors, de nouvelles architectures doivent être mises en place. Quelques travaux s'intéressent aux fluctuations de ces nouvelles architectures comme les transistors FinFET<sup>5</sup> [Gust06], à double grille (« DG »), à grille enrobante ou GAA<sup>6</sup> [Cath06a], les transistors silicium sur isolant complètement déserté FD-SOI<sup>7</sup> ou SON<sup>8</sup>

## B. Mesure des fluctuations locales dans le transistor MOS

La mesure des fluctuations locales requiert des équipements de mesure précis pour mesurer une différence de paramètres électriques, des structures de test particulières et une méthode d'extraction de ces paramètres adaptée à l'étude choisie. Un nombre conséquent de mesures est nécessaire suivi d'un traitement statistique permettant de conclure sur l'appariement des dispositifs. Le plan de cette partie suit la méthodologie de mesure couramment utilisée pour mesurer les fluctuations locales des transistors MOS illustrée Fig. IV-3.

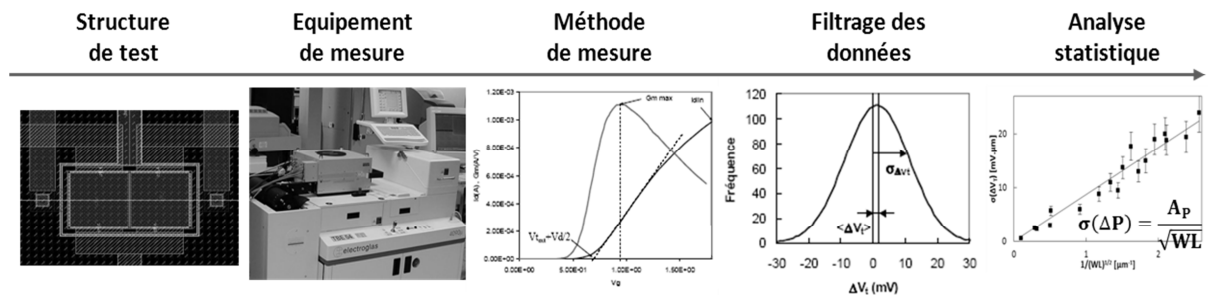


Fig. IV-3 : Méthodologie de mesure des fluctuations locales

### 1 Structures de mesure

Généralement, les mesures électriques d'évaluation des fluctuations locales peuvent être réalisées sur trois types de structures : des matrices adressables de transistors MOS ou des paires de transistors MOS appariés.

Ces dernières seront utilisées dans la plupart de nos études. Elles sont basées sur deux transistors MOS appariés et ont été proposées en 1986 [Laks86]. Malgré un nombre limité de mesures par plaque (de l'ordre de plusieurs dizaines contre plusieurs milliers pour les matrices de transistors MOS) donnant des incertitudes statistiques relativement élevées mais rarement problématiques, ce type de

<sup>5</sup> Fin Field Effect Transistor

<sup>6</sup> Gate All Around

<sup>7</sup> Fully Depleted Silicon On Insulator

<sup>8</sup> Silicon On Nothing

structure est facile à concevoir. Ainsi, pour des raisons de simplicité et de rapidité de conception, de flexibilité lors de la mesure et de précision dans les résultats, l'appariement des transistors est étudié dans ce manuscrit en utilisant principalement des structures de test composées de transistors MOS appariés. Une structure de test contient 22 plots de contacts et plusieurs paires de transistors MOS appariés (cf. Fig. IV-4). Un plot permet de réaliser les connexions de source (S) de toutes les paires, et un autre plot relie les prises substrat (B) (permettant une polarisation avec effet substrat). La connexion de grille (G) est également commune à tous les transistors (tous les tests d'appariement présentés dans ces travaux de thèse sont réalisés sur des transistors ayant un oxyde de grille de 65Å). Enfin, toutes les connexions de drain sont propres à chaque transistor (D1 et D2). Un anneau de garde (« Guard ring ») est dessiné autour de chaque paire pour limiter les effets systématiques

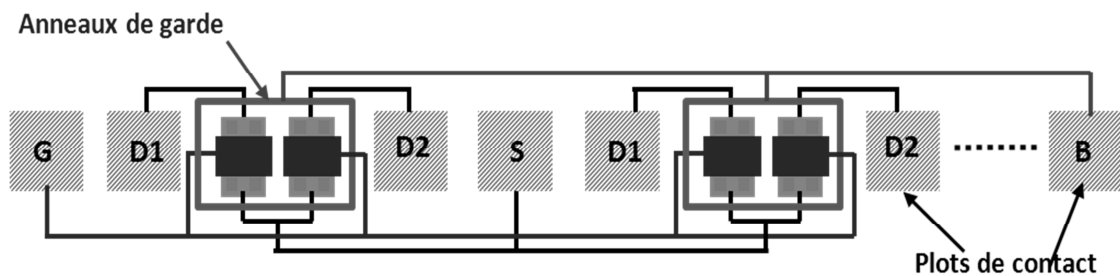


Fig. IV-4 : Structure de test de paires de transistors MOS

## 2 Méthodes de mesure de l'appariement

Dans ce manuscrit, les appariements en courant et en tension des transistors MOS sont étudiés grâce aux configurations de mesure présentées Fig. IV-5. La connexion de source est à la masse ainsi que la polarisation du substrat (sauf indication contraire).

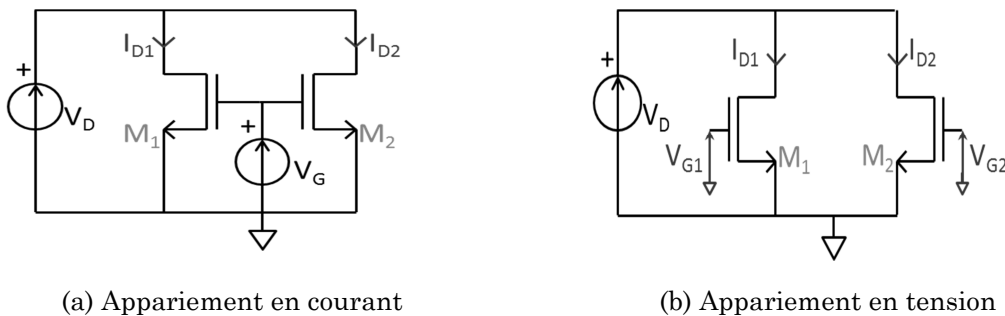


Fig. IV-5 : Méthodes de mesure pour étudier l'appariement en courant et en tension des transistors MOS

La Fig. IV-5a présente la méthode de mesure de l'appariement en courant. Cette méthode consiste simplement à imposer un potentiel de drain  $V_D$ , à faire varier la tension de grille  $V_G$  et à relever les courants de drain  $I_{D1}$  et  $I_{D2}$  des deux transistors (correspond à une extraction de caractéristique  $I_D-V_G$ ). La différence relative des courants  $\Delta I_D/I_D$  est ensuite déterminée par la relation (IV-9).

La Fig. IV-5b présente la méthode de mesure de l'appariement en tension de grille. Ici, la méthode de polarisation d'une paire différentielle consiste à imposer

le même courant de drain  $I_D$  dans chaque branche et à venir mesurer les tensions de grille  $V_{G1}$  et  $V_{G2}$  des deux transistors afin d'en déterminer leur appariement. Les connexions de source et de substrat sont à la masse. Cependant, cette méthode nécessite d'imposer le courant  $I_D$  mais également le potentiel de drain des transistors  $M_1$  et  $M_2$  afin de polariser la paire différentielle dans le régime souhaité. La solution utilisée afin de pouvoir mesurer cet appariement consiste à imposer le potentiel de drain et relever les tensions de grille des deux transistors pour chaque valeur de courant de drain  $I_D$  définie (correspond à une caractéristique  $V_G-I_D$ ). La différence des tensions de grille  $\Delta V_G$  est finalement calculée en utilisant la relation (IV-8).

De nombreuses définitions de la tension de seuil existant dans la littérature, il est souvent difficile de s'entendre sur la méthode d'extraction du  $V_T$  à utiliser. Cependant, l'étude de l'appariement de la tension de seuil des transistors MOS simplifie le travail en ne s'intéressant uniquement à la dispersion du  $V_T$  et non à sa valeur absolue. Une comparaison de plusieurs méthodes de mesure est proposée dans [Croo02c] avec des résultats obtenus ne présentant pas de différences réellement significatives. Les résultats d'appariement présentés dans ce manuscrit sont obtenus par la méthode du maximum de transconductance ou à courant constant. Pour les deux méthodes de mesure de la tension de seuil, l'extraction du paramètre  $V_T$  se fait à partir des caractéristiques  $I_D-V_G$  mesurées en régime linéaire ( $|V_D|=0.1V$ ) en faisant varier la tension de grille  $V_G$  de  $0V$  à  $V_{DD}=3.3V$ .

### 3 Traitement statistique des résultats d'appariement

La mesure de la différence des paramètres électriques entre deux dispositifs ( $P_1$  et  $P_2$ ) se représente généralement soit en valeur absolue (IV-8), soit en relatif ou pourcentage (IV-9).

$$\Delta_a(P) = P_1 - P_2 \quad (IV-8)$$

$$\Delta_{\%}(P) = 200 \cdot \frac{P_1 - P_2}{P_1 + P_2} \quad (IV-9)$$

Selon le théorème central-limite (la somme d'un très grand nombre de variables aléatoires de loi quelconque suit une loi normale), la population  $P$  mesurée suit une loi normale. A partir de cette distribution, la moyenne  $\langle m_{\Delta P} \rangle$  et l'écart-type (déviation standard)  $\sigma_{\Delta P}$  sont calculées à l'aide des équations (IV-10) et (IV-11) respectivement.

$$m_{\Delta P} = \frac{1}{N} \sum_{i=1}^N \Delta P_i \quad (IV-10)$$

$$\sigma_{\Delta P} = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (\Delta P_i - m_{\Delta P})^2} \quad (IV-11)$$

La moyenne  $\langle m_{\Delta P} \rangle$  donne une indication sur l'appariement systématique. Cette moyenne est théoriquement égale à zéro (distribution centrée sur zéro). La

déviatiion standard  $\sigma_{\Delta P}$  représente quant à elle, l'appariement stochastique (ou aléatoire) provenant des variations ou fluctuations aléatoires des paramètres électriques d'un dispositif.

L'extraction de l'écart-type est obtenue à partir de mesures sur une population de  $N_{\text{paires}}$  de transistors dont les mesures se trouvent dans un intervalle de confiance. La dispersion (variance) de ce paramètre  $\sigma$  est donnée par la relation (IV-12) [Cath06b].

$$\text{Var}(\sigma_{\Delta P}) = \frac{\sigma_{\Delta P}^2}{2(N_{\text{paires}} - 1)} \quad (\text{IV-12})$$

A partir de cette relation, la dispersion peut s'exprimer en pourcentage par rapport à l'estimation de  $\sigma$  permettant de définir les barres d'erreur sur les différentes représentations d'appariement.

#### 4 Extraction du paramètre d'appariement

Le paramètre  $A_P$  qui caractérise l'appariement (de la tension de seuil, du facteur de courant, etc...) des transistors MOS pour une technologie donnée possède une incertitude (provenant de l'incertitude des écart-types mesurés). La méthodologie classique pour extraire ce paramètre qui est noté  $A_{cl}$  dans le cas de l'utilisation d'une régression linéaire. Les écart-types ( $\sigma_{\Delta P}$ ) sont représentés en fonction de l'inverse de la racine de la surface des transistors MOS étudiés ( $1/\sqrt{WL}$ ). La valeur du paramètre  $A_{cl}$  est donnée par l'équation (IV-13) [Cath06b].

$$A_{cl} = \frac{\sum_{i=0}^{N_{\text{geo}}-1} (\sigma_{\Delta P_i} / \sqrt{W_i L_i})}{\sum_{i=0}^{N_{\text{geo}}-1} (1/\sqrt{W_i L_i})^2} \quad (\text{IV-13})$$

Avec  $N_{\text{geo}}$  le nombre de géométries différentes testées,  $i$  représente la  $i^{\text{ème}}$  paire testée (parmi  $N_{\text{paires}}$  au total). La variance (incertitude de ce paramètre  $A_{cl}$ ) peut s'écrire suivant (IV-14).

$$\text{Var}(A_{cl}) = \frac{\sum_{i=0}^{N_{\text{geo}}-1} (\sigma_{\Delta P_i} / \sqrt{W_i L_i})^2}{2(N_{\text{paires}} - 1) \left[ \sum_{i=0}^{N_{\text{geo}}-1} (1/\sqrt{W_i L_i})^2 \right]^2} \quad (\text{IV-14})$$

Dans [Cath06b], une nouvelle méthode d'estimation du paramètre  $A_P$  par régression pondérée (noté  $A_{rp}$ ) est proposée puis comparée à la méthode classique. Cette technique propose de pondérer les contributions de chacune des variances et ainsi minimiser l'incertitude lors de l'estimation du paramètre  $A_P$ . Cela se fait en moyennant les termes  $\sigma_{\Delta P_i} / \sqrt{W_i L_i}$  (déviatiion standard normalisée) d'après la relation (IV-15) et dont la variance s'exprime selon (IV-16).

$$A_{rp} = \frac{1}{N_{\text{geo}}} \sum_{i=0}^{N_{\text{geo}}-1} \sigma_{\Delta P_i} / \sqrt{W_i L_i} \quad (\text{IV-15})$$

$$\text{Var}(A_{rp}) = \frac{\sum_{i=0}^{N_{\text{geo}}-1} \left( \frac{\sigma_{\Delta P_i}}{1/\sqrt{W_i L_i}} \right)^2}{2(N_{\text{paires}} - 1) N_{\text{geo}}^2} \quad (\text{IV-16})$$

Ces deux méthodes sont représentées Fig. IV-6a pour la régression linéaire et Fig. IV-6b pour la régression pondérée. Les mesures sont réalisées sur 18 géométries de transistors NMOS appariés sur 32 sites ( $N_{\text{paires}}=32$ ) en utilisant la méthode du maximum de transconductance.

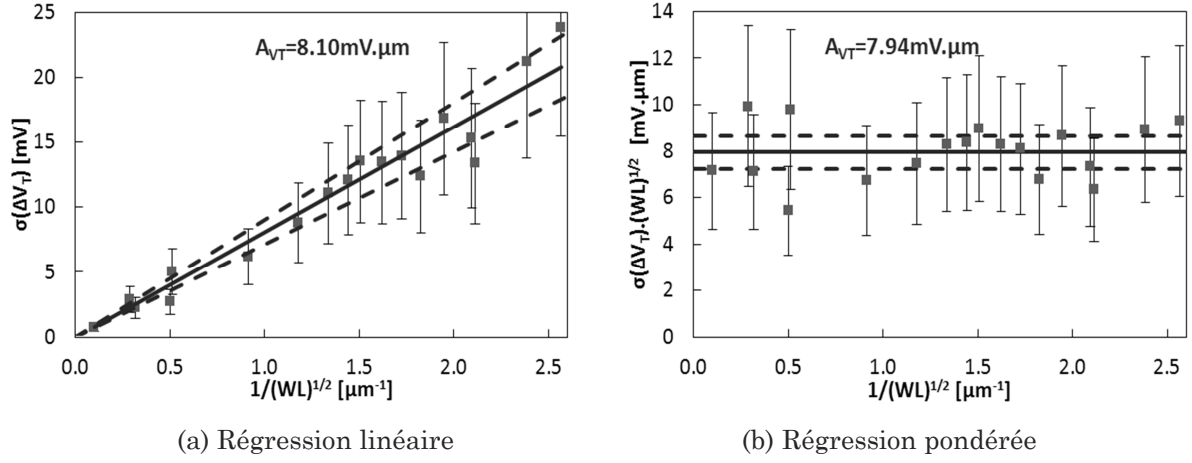


Fig. IV-6 : Différentes représentations de l'appariement de la tension de seuil des transistors MOS. Les incertitudes (pointillé rouge) et barres d'erreur sont données dans un intervalle de confiance de 99.7%

La méthode par régression pondérée est indépendante des dimensions testées alors que le niveau de fluctuation en utilisant l'estimation par régression classique dépend de la géométrie des transistors étudiés [Cath06b]. De plus, Les écart-types étant normalisés par rapport à la racine de la surface du transistor MOS, ils peuvent se représenter en fonction d'autres grandeurs dimensionnelles. Souvent, une représentation en fonction de  $W$  ou  $L$  en abscisse permet de révéler des effets non visibles sur une représentation linéaire (l'impact de la dose des poches par exemple [Cath08a]). Enfin, la minimisation de l'incertitude lors de l'estimation du paramètre  $A_p$  peut parfois permettre de comparer deux procédés différents sans ambiguïté en utilisant une régression pondérée alors que les incertitudes se recouvrent avec une estimation par régression linéaire. Cette minimisation est démontrée par la relation (IV-17) à l'aide de l'inégalité de Cauchy-Schwarz [Cath06b].

$$\frac{\text{Var}(A_{rp})}{\text{Var}(A_{cl})} = \frac{\left[ \sum_{i=0}^{N_{\text{geo}}-1} (1/\sqrt{W_i L_i})^2 \right]^2}{N_{\text{geo}} \sum_{i=0}^{N_{\text{geo}}-1} (1/\sqrt{W_i L_i})^4} \leq 1 \quad (\text{IV-17})$$

Ainsi pour les résultats présentés Fig. IV-6,  $A_{cl}=8.10 \text{ mV.}\mu\text{m}$  et  $A_{rp}=7.94 \text{ mV.}\mu\text{m}$  avec une variance divisée par deux en utilisant la régression pondérée.

Une légère surestimation des fluctuations locales (et donc de l'estimation du paramètre  $A_p$ ) est observée lors de la mesure à courant constant de l'appariement des transistors MOS (cf. Fig. IV-7). Cette surestimation peut s'expliquer par le fait que l'extraction de la tension de seuil à courant fixe se fait à la limite entre le régime d'inversion faible et d'inversion forte [Croo02c]. Cependant, les valeurs obtenues sur l'une ou l'autre méthode étant relativement proches, les deux méthodes sont utilisées dans ce manuscrit.



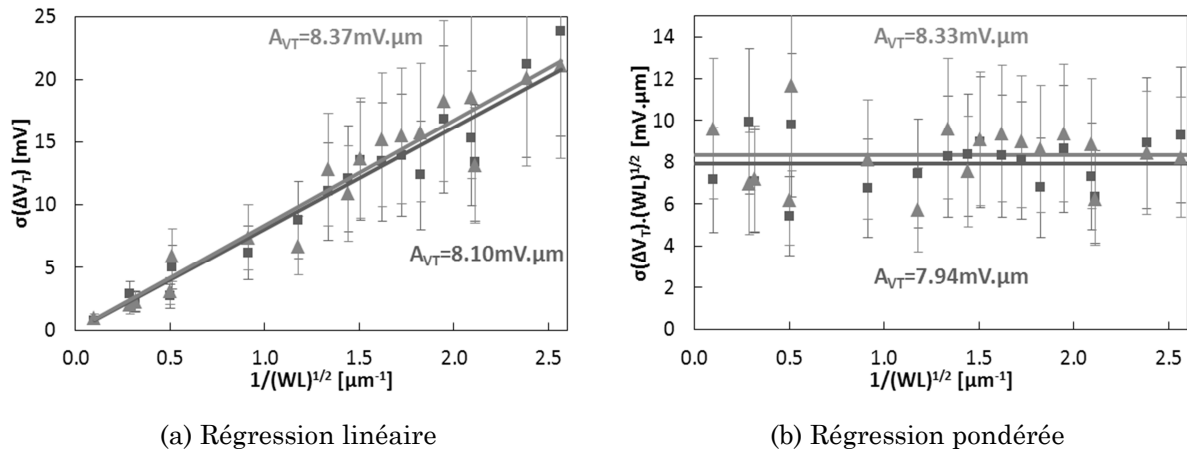


Fig. IV-7 : Comparaison de l'extraction du paramètre  $A_{VT}$  par la méthode du maximum de transconductance (carré bleu) ou à courant constant (triangle vert). Barres d'erreur données dans un intervalle de confiance de 99.7%

## II. Contribution à l'amélioration de l'appariement des circuits analogiques polarisés sous le seuil

### A. Conception de circuits analogiques faible consommation

Pour réaliser des applications « ultra-faible consommation », un moyen simple est de réduire les tensions d'alimentation pour chaque nœud technologique. Cela se traduit, pour les blocs analogiques, par des courants de l'ordre de la centaine de nano-ampère par branche.

Le rapport transconductance/courant de drain ( $g_m/I_D$ ) est représenté Fig. IV-8 en fonction de la différence  $V_G - V_T$  (appelé « overdrive »). Ce rapport est faible pour une forte polarisation de grille, augmente autour de la tension de seuil et devient particulièrement important dans la zone sous le seuil. Cette évolution du rapport  $g_m/I_D$  explique le fait que la plupart des circuits analogiques faibles consommation sont polarisés sous le seuil.

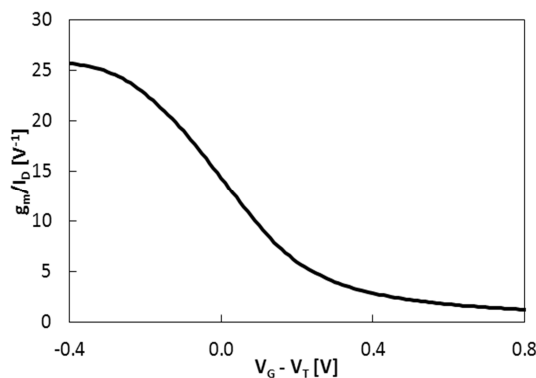


Fig. IV-8 : Comparaison de l'extraction Rapport Transconductance  $g_m$  / Courant de drain  $I_D$

Les concepteurs de circuits se basent sur les résultats de simulation pour prédire le fonctionnement de leur circuit et vérifier qu'il répond correctement aux cahiers des charges. Ils peuvent également simuler l'impact que peut avoir l'appariement

des transistors MOS sur les performances de leur circuit. De nombreux travaux étudient cet appariement à l'aide de simulations de type Monte Carlo (MC) sur des circuits simples (miroirs de courant, paires différentielles...) [King96] [King05] [King07] ou des circuits plus complets [Gupt02] [Gupt05]. Afin d'évaluer les fluctuations des paramètres des transistors MOS, des simulations MC sont utilisées. Les simulations MC permettent de générer aléatoirement des valeurs de paramètres parmi la distribution gaussienne de chaque paramètre en utilisant l'expression (IV-18).

$$P = P_0 + \text{Gaussienne}\left(\frac{A_P}{\sqrt{WL}}\right) \quad (\text{IV-18})$$

$P$  étant un paramètre électrique ( $V_T$ ,  $\beta$ , etc...) et  $P_0$  ce même paramètre électrique  $P$  ne tenant pas compte des fluctuations. Le terme « Gaussienne  $\left(\frac{A_P}{\sqrt{WL}}\right)$  » permet de rajouter des fluctuations supplémentaires (dépendant de la surface et de l'appariement du dispositif) au terme initial  $P_0$  en choisissant des valeurs parmi une distribution Gaussienne.

## B. L'effet « hump »

### 1 Présentation

L'effet « hump » est un phénomène pouvant apparaître dans les procédés de fabrication à base de STI<sup>9</sup>. Afin de réaliser les blocs de mémoire NVM embarqués, les procédés de fabrication avec mémoire embarquée présentent des étapes supplémentaires par rapport à un procédé CMOS classique. Ces étapes supplémentaires permettent de réaliser les transistors constituant les points mémoires. Le procédé de fabrication utilise la méthode de Tranchée Isolante peu Profonde (STI<sup>10</sup>) pour augmenter les performances et la densité d'intégration des transistors grâce à une séparation abrupte entre la tranchée d'isolation et la zone d'active [Schw05]. Cette séparation abrupte entre STI et la zone active recouverte de poly-silicium peut être responsable de la création de transistors parasites étroits de largeur  $W_{\text{HUMP}}$  (en rouge vif sur la Fig. IV-9) situés de part et d'autres du transistor principal ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ ).

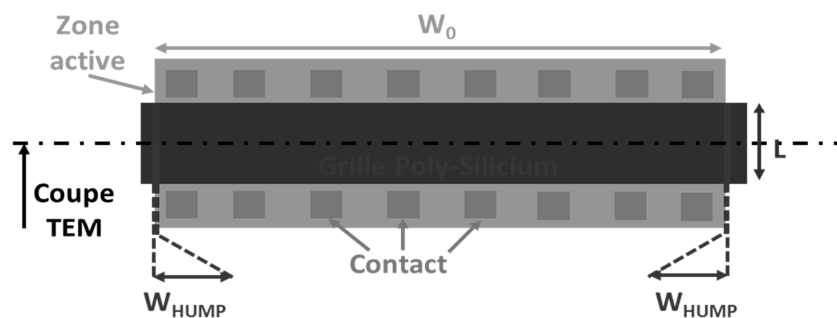


Fig. IV-9 : Vue « layout » simplifiée du transistor étudié ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ )

<sup>10</sup> Shallow Trench Isolation

Afin de voir l'aspect morphologique de ces transistors parasites, une coupe TEM<sup>11</sup> est réalisée et présentée Fig. IV-10 dans la direction de la largeur du transistor comme indiquée sur la vue « layout » de la Fig. IV-9.

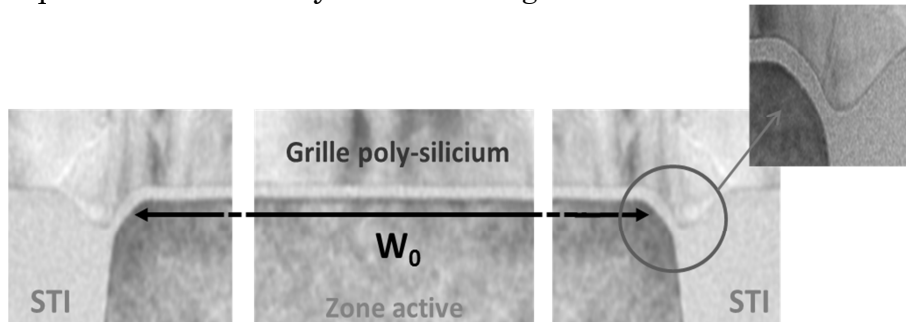


Fig. IV-10 : Coupe TEM dans le sens de la largeur  $W$  de la structure de test étudiée ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ )

Comme le montre la coupe TEM, la morphologie des transistors parasites est visible de part et d'autre du transistor principal. Ces derniers se localisent dans une « cuvette » présente dans le STI et remplie de poly-silicium générée lors de différentes étapes du procédé de fabrication.

Il est important de noter que ces transistors parasites possèdent une largeur  $W_{\text{HUMP}}$  très petite comparée à la largeur  $W_0$  du transistor principal. Les transistors de bord possèdent une largeur très faible et souffrent également d'une réduction du dopage sur les bords du canal [Sall96] par rapport au transistor principal. Ces transistors parasites, ayant un dopage canal beaucoup plus faible, présentent par conséquent une tension de seuil inférieure à celle du transistor principal [Schw05]. De plus, ces transistors présentent une sensibilité réduite à la polarisation de substrat du fait de leur tension de seuil plus faible [Sall96]. Ainsi, l'effet substrat est souvent utilisé pour mettre en évidence cet effet « hump » et distinguer clairement les caractéristiques électriques du transistor central de celles des transistors parasites. Les caractéristiques  $I_D$ - $V_G$  avec effet substrat sont présentées Fig. IV-11.

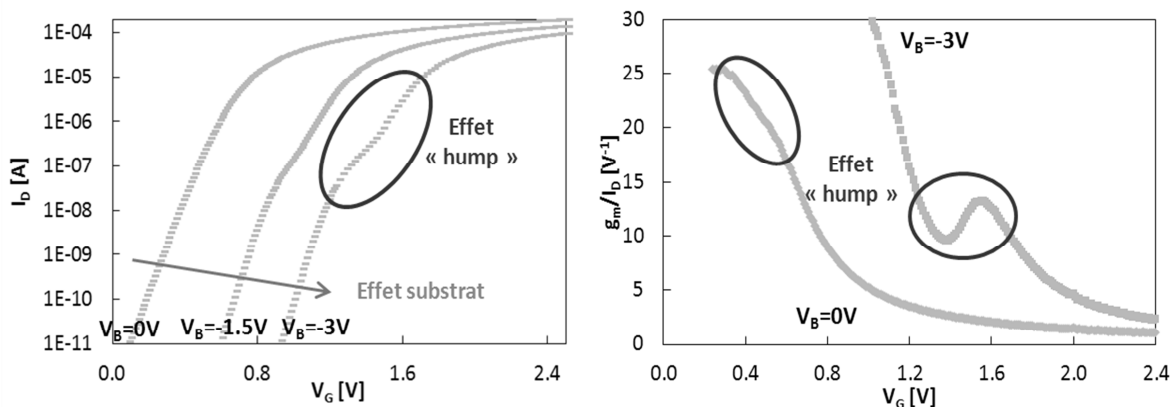


Fig. IV-11 : (a) Caractéristiques  $I_D$ - $V_G$  de la structure de test étudiée ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ ) en fonction de la polarisation de substrat ( $V_B=0\text{V}$ ,  $-1.5\text{V}$  et  $-3\text{V}$ ) (b) Rapport  $g_m/I_D$  de la structure de test étudiée ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ ) avec ( $V_B=-3\text{V}$ ) et sans effet substrat ( $V_B=0\text{V}$ )

<sup>11</sup> Transmission Electronic Microscopy

La déformation de la caractéristique  $I_D$ - $V_G$  est nettement visible avec effet substrat et il est plus compliqué de la déceler lorsque  $V_B=0V$ . Un autre moyen de voir parfois plus facilement cet effet « hump » est de tracer le rapport  $g_m/I_D$  car le fait de calculer la dérivée du courant de drain  $I_D$  par l'intermédiaire de la transconductance  $g_m$  puis de diviser par ce même courant  $I_D$  accentue encore plus cet effet « hump ». Cette méthode peut parfois permettre de déceler un impact de l'effet « hump » sans effet substrat alors qu'il n'est pas visible sur les caractéristiques  $I_D$ - $V_G$ .

## 2 Modélisation de l'effet hump

La modélisation de l'effet hump utilise une approche du premier ordre. Dans cette modélisation, le transistor est composé de trois transistors en parallèle : le transistor principal  $M_0$  et deux transistors modélisant les transistors de bord  $M_{HUMP}$  [Park08]. Ces deux transistors sont considérés comme identiques en utilisant les propriétés de symétrie d'un transistor. Ce modèle, présenté Fig. IV-12 se justifie d'après la coupe TEM présentée Fig. IV-10.

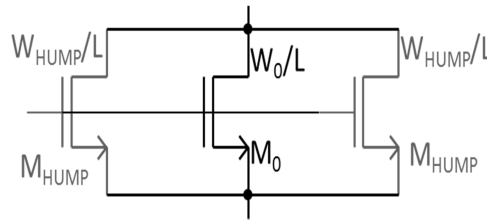


Fig. IV-12 : Macro-modèle pour modéliser l'effet « hump »

La tension de seuil  $V_T$  d'un transistor augmente en fonction de la largeur  $W$ . Ainsi, en ajustant le paramètre  $W_{HUMP}$ , il est possible de déterminer la tension de seuil  $V_{T_{HUMP}}$  de ces transistors de bord. Ainsi, ils conduisent plus ou moins de courant et plus ou moins rapidement par rapport au transistor principal en fonction de la tension de grille appliquée (cf. Fig. IV-13).

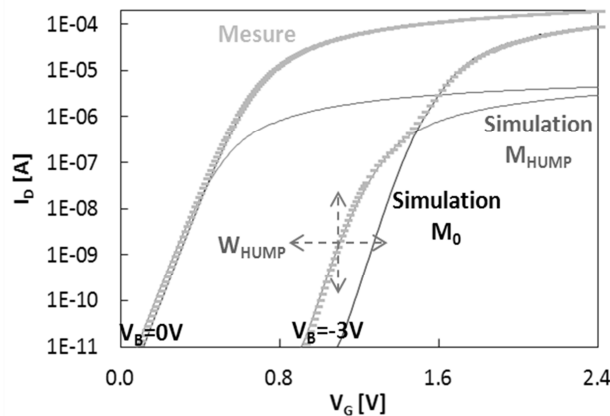


Fig. IV-13 : Comparaison entre mesure (en vert) et la simulation du courant fourni par le transistor principal (en noir) et par les transistors parasites (en bleu) pour  $V_B=0V$  et  $V_B=-3V$

## 3 Impact de l'effet « hump » sur l'appariement en tension

Les mécanismes de conduction du courant en régime de faible inversion décrits ci-dessus permettent d'expliquer le comportement de l'appariement des tensions

de grille d'une paire différentielle représenté Fig. IV-14. Autour de la tension de seuil, l'appariement est égal à sa valeur théorique minimale définie par l'équation (IV-1). Cependant, au lieu de rester constant dans la zone sous le seuil et égal à ce minimum ( $A_{VT}/\sqrt{W_0L}$ ), l'appariement en  $V_G$  se dégrade à cause de la conduction plus rapide des transistors de bord et tend vers l'appariement en  $V_T$  de ces transistors parasites ( $A_{VT}/\sqrt{2.W_{HUMP}L}$ ). Le facteur 2 vient de la présence de deux transistors parasites sur chaque bord du transistor principal. En utilisant le macro-modèle présenté précédemment et calibré sur les courbes  $I_D-V_G$ , la simulation Monte Carlo permet de correctement modéliser l'appariement en tension dans la zone sous le seuil comme le prouve la Fig. IV-15.

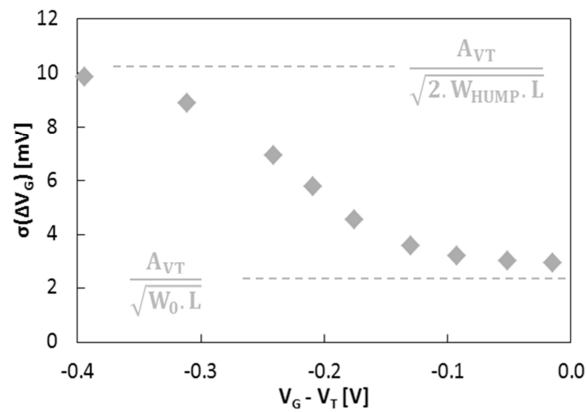


Fig. IV-14 : Dégradation de l'appariement en tension d'une paire différentielle polarisée sous le seuil

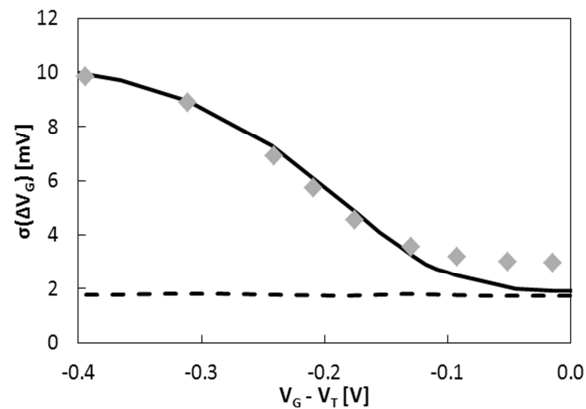


Fig. IV-15 : Appariement en tension dans la zone sous le seuil : Mesure (♦), simulation sans (pointillé) et avec (trait plein) le macro-modèle effet « hump »

#### 4 Etude dimensionnelle

Une étude des caractéristiques  $I_D-V_G$  avec effet substrat est nécessaire pour connaître l'impact des transistors de bord en fonction des dimensions du transistor principal. La Fig. IV-16 illustre cette étude dimensionnelle d'appariement en tension en normalisant cette valeur par la racine carrée de la surface. Ce changement d'échelle permet de mettre en avant d'éventuels effets liés aux tailles des transistors étudiés.

L'étude en fonction de la longueur  $L$  (Fig. IV-16) permet d'affirmer que la longueur du transistor sur l'effet « hump » n'est pas significatif.

L'étude en fonction de la largeur  $W$  (Fig. IV-16b) montre que quelle que soit la longueur du transistor, l'étude en fonction de  $W$  permet de confirmer que l'impact des transistors de bord est d'autant plus important que le transistor principal est large. L'appariement se dégrade clairement pour les transistors larges par une mise en évidence plus prononcée de l'effet « hump ». Ce phénomène de remontée sous le seuil est donc d'autant plus important que la différence entre  $W_0$  et  $W_{\text{HUMP}}$  est importante.

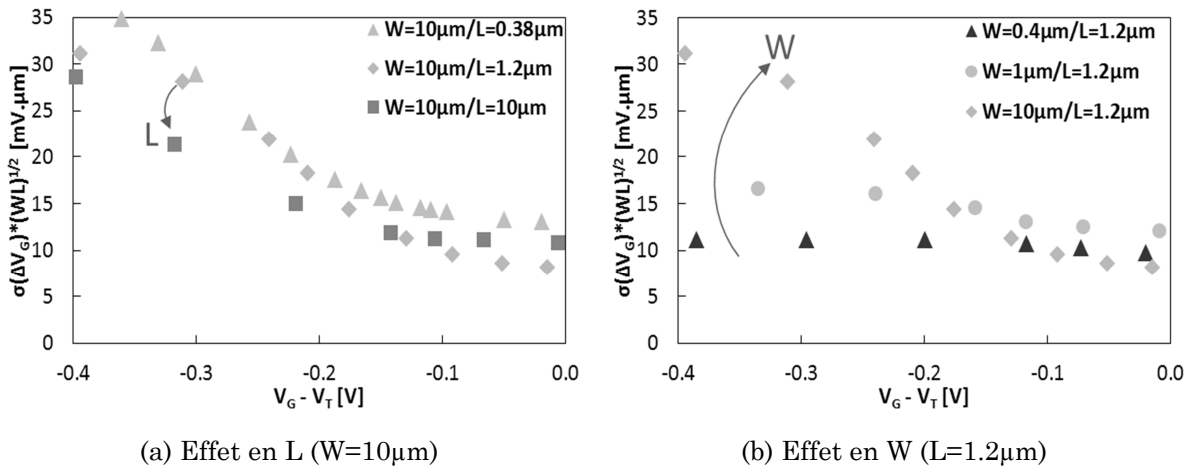


Fig. IV-16 : Mesure de l'appariement en tension pour plusieurs dimensions

(a) La mesure de l'appariement est effectuée sur trois structures de test ayant une largeur  $W=10\mu\text{m}$ . La longueur  $L$  du transistor varie en suivant les valeurs suivantes :  $L=0.38\mu\text{m}$ ,  $L=1.2\mu\text{m}$ ,  $L=10\mu\text{m}$ . Comme observé Fig. IV-16c et d, l'appariement en tension ne dépend que très peu de la longueur  $L$ .

(b) La longueur reste constante et égale à  $1.2\mu\text{m}$  et c'est la largeur  $W$  du transistor que l'on fait varier :  $W=0.4\mu\text{m}$ ,  $W=1\mu\text{m}$ ,  $W=10\mu\text{m}$  (Augmentation représentée par une flèche bleue).

## 5 Etude en fonction de l'effet substrat

Comme cela est expliqué dans [Sall96], les transistors parasites présentent une sensibilité réduite à la polarisation de substrat notamment à cause de leur dopage plus faible, c'est pourquoi l'effet substrat est utilisé pour mettre en évidence ce phénomène pas forcément visible à  $V_B=0\text{V}$  (○ Fig. IV-17). Du fait de cette sensibilité réduite, l'écart des tensions de seuil entre le transistor principal et les transistors de bord augmente avec la polarisation négative du substrat. Cet écart rend de plus en plus visible l'effet « hump » (- Fig. IV-17). En suivant le raisonnement inverse, appliquer une tension de substrat légèrement positive (◆ Fig. IV-17) peut permettre de diminuer l'écart entre  $V_{T_0}$  et  $V_{T_{\text{HUMP}}}$  afin de limiter la conduction des transistors parasites.

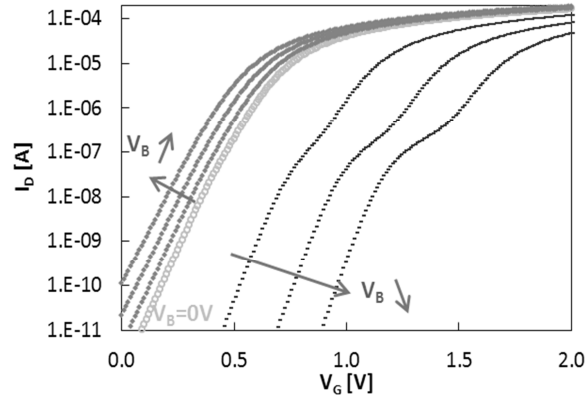


Fig. IV-17 : Caractéristiques  $I_D$ - $V_G$  de la structure de test étudiée ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ ) en fonction de la polarisation de substrat ((-)  $V_B=-3\text{V}, -2\text{V}, -1\text{V}$ ), (( $\circ$ )  $V_B=0\text{V}$ ), (( $\blacklozenge$ )  $V_B=0.1\text{V}, 0.2\text{V}, 0.3\text{V}$ )

La déformation des caractéristiques  $I_D$ - $V_G$  due aux transistors de bord n'étant pas facilement décelable sans effet substrat, il est difficile de voir une éventuelle amélioration avec une polarisation positive du substrat. L'impact de la polarisation substrat directe peut être mis en évidence par l'étude de l'appariement en tension dans la zone sous le seuil comme le montre la Fig. IV-18.

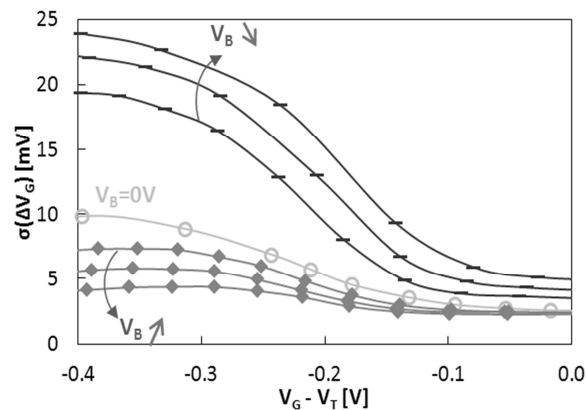


Fig. IV-18 : Appariement en tension de la structure de test étudiée ( $W=10\mu\text{m}/L=1.2\mu\text{m}$ ) en fonction de la polarisation de substrat ((-)  $V_B=-3\text{V}, -2\text{V}, -1\text{V}$ ), (( $\circ$ )  $V_B=0\text{V}$ ), (( $\blacklozenge$ )  $V_B=0.1\text{V}, 0.2\text{V}, 0.3\text{V}$ )

L'étude de l'appariement en tension confirme le comportement de l'effet « hump » observé sur les caractéristiques  $I_D$ - $V_G$  en fonction de la polarisation du substrat.

- Polarisation de substrat négative : Une telle polarisation augmente la tension de seuil d'un transistor NMOS. A cause d'une sensibilité à la polarisation de substrat réduite, la tension de seuil des transistors parasites augmente mais moins rapidement que le  $V_T$  du transistor principal ce qui accentue l'écart entre les deux et permet de visualiser facilement l'effet « hump ». La conduction des transistors de bord est encore plus précoce par rapport au transistor principal et dégrade l'appariement en tension dans la zone sous le seuil (-). La polarisation de substrat négative (appelée RBB pour « Reverse Body Bias ») augmente la tension de seuil d'un transistor MOS ce qui permet notamment de limiter les fuites.

- Polarisation de substrat positive : Devant une sensibilité réduite à la polarisation de substrat, lorsque  $V_B$  augmente, la tension de seuil du transistor principal diminue plus rapidement que  $V_{T_{\text{HUMP}}}$  et rend de moins en moins

important l'impact des transistors de bord avec une amélioration de l'appariement en tension dans la zone sous le seuil (♦). Une polarisation de substrat positive (appelée FBB pour « Forward Body Bias ») est utilisée quant à elle afin de diminuer la tension de seuil du transistor MOS, permettant ainsi d'accélérer la vitesse des transistors MOS et réduire les délais dans les circuits [Nare03]. L'inconvénient majeur d'une telle polarisation de substrat est l'augmentation des courants de fuite pouvant aller jusqu'à la polarisation directe des jonctions P/N.

L'étude de l'effet substrat sur l'appariement de la tension de grille aide à la compréhension de l'effet « hump » et de l'appariement en tension dans la zone sous le seuil. En effet, la remontée de cet appariement visible en régime de faible inversion dépend essentiellement de la conduction du courant plus ou moins rapide des transistors parasites par rapport au transistor principal en fonction de la tension de grille.

### C. Solutions pour contrer l'effet « hump »

Les solutions proposer pour contrer l'effet « hump » s'appliquent aussi bien au niveau circuit, au niveau du procédé de fabrication ou encore, en concevant de nouveaux transistors. Toutes les propositions sont présentées dans le cas de transistors NMOS mais peuvent très bien être appliquées aux transistors PMOS.

#### 1 Solution basée sur des transistors « multi-doigts »

L'appariement en tension d'une paire différentielle polarisée sous le seuil est lié à l'appariement en  $V_T$  des transistors parasites d'après l'équation (IV-19).

$$\sigma(\Delta V_G) = \frac{A_{VT}}{\sqrt{2 \cdot W_{HUMP} \cdot L}} \quad (IV-19)$$

Les transistors parasites pouvant être modélisés avec des transistors ayant une largeur  $W_{HUMP}$  très faible, leur surface est donc petite et l'appariement se dégrade. L'idée de ces transistors est donc d'augmenter cette largeur  $W_{HUMP}$  en « découpant » le transistor principal en « m » transistors plus étroits en parallèle. m étant le nombre de transistors identiques placés en parallèle. La Fig. IV-19 illustre le concept de configuration « multi-doigts ».

Le but est de limiter au maximum l'effet « hump » en augmentant la largeur équivalente des transistors parasites tout en gardant les mêmes caractéristiques pour le transistor principal. A l'aide du paramètre m introduit précédemment, la largeur du transistor principal est donnée par la relation (IV-20).

$$W_0 = m * W_{multi} \quad (IV-20)$$

L'équation (IV-19) qui décrit la valeur théorique vers laquelle tend l'appariement en tension sous le seuil devient, dans le cas d'un transistor « multi-doigts », la relation (IV-21).

$$\sigma(\Delta V_G) = \frac{A_{VT}}{\sqrt{m \cdot 2 \cdot W_{HUMP} \cdot L}} \quad (IV-21)$$



D'après cette équation (IV-21), plus le nombre de transistors placés en parallèle est important (plus  $m$  est grand), meilleur sera l'appariement en tension dans la zone sous le seuil (diminué d'un facteur  $\sqrt{m}$ ).

Pour valider cette hypothèse de manière expérimentale, plusieurs structures de test d'appariement ont été mesurées mettant en jeu des transistors MOS GO2 dont les dimensions totales restent constantes ( $W=12\mu\text{m}/L=1\mu\text{m}$ ) mais dont la zone d'active est plus ou moins « découpée » (mise en parallèle des transistors) afin de réaliser des transistors « multi-doigts » (cf. Tableau IV-1).

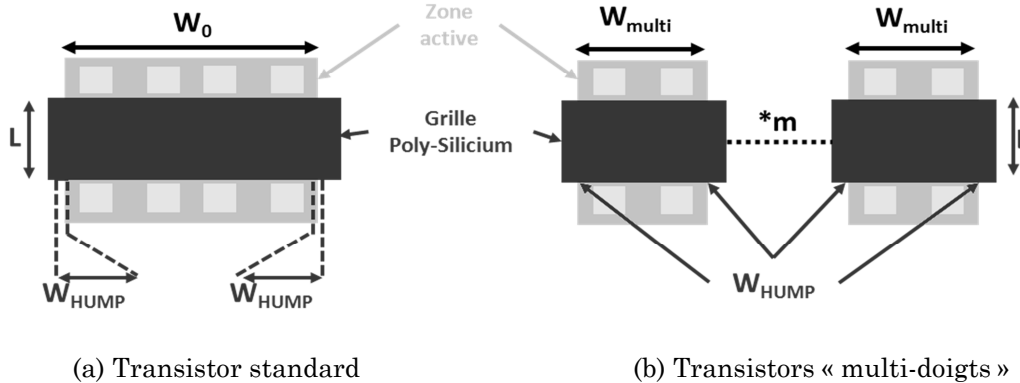


Fig. IV-19 : Vue « layout » simplifiée d'un transistor en « multi-doigts »

Tableau IV-1 : Description des structures de test « multi-doigts »

Largeur (W)	12	4	1.2	0.4
Nombre de transistors en parallèle (m)	1	3	10	30

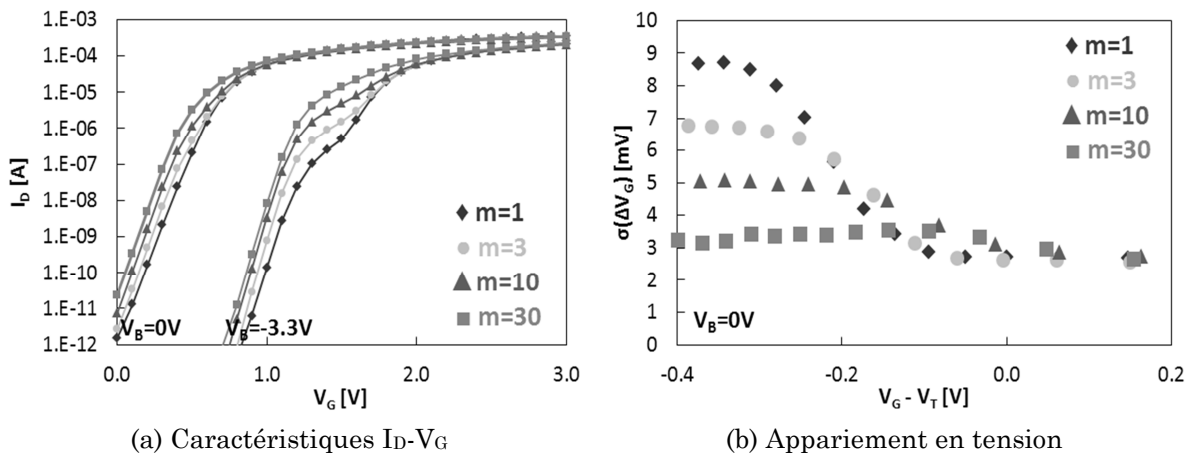


Fig. IV-20 : Comparaisons des caractéristiques  $I_D-V_G$  et de l'appariement en tension en fonction du nombre de doigt des transistors ( $W=12\mu\text{m}/L=1\mu\text{m}$ )

Les caractéristiques  $I_D-V_G$  sans et avec effet substrat sont représentées Fig. IV-20a et l'appariement en tension dans la zone sous le seuil Fig. IV-20b. Plus le nombre de transistors en parallèle est important ( $m$  grand), plus la largeur totale équivalente du dispositif est grande ( $m$  transistors parasites en parallèle), plus le niveau de courant est élevé comme le montre le décalage des courbes  $I_D-V_G$  vers le haut. Ce niveau de courant plus élevé dû aux transistors de bord « efface »

l'épaulement normalement visible avec effet substrat. Même si les transistors parasites conduisent le courant de plus en plus tôt avec l'augmentation de  $m$ , l'appariement en tension est nettement amélioré grâce à la surface équivalente totale du dispositif beaucoup plus grande.

Le « découpage » du transistor principal et la mise en parallèle de transistors plus étroits est, au même titre que la polarisation du substrat inversée, une méthode simple qui permet de limiter l'impact de l'effet « hump ».

## 2 Autres solutions

La solution la plus simple à mettre en œuvre est basée sur le « Forward Body Bias » ou polarisation du substrat directe. Etant donné que l'effet « hump » est mis en évidence avec effet substrat, l'idée est tout simplement d'utiliser le comportement inverse (polarisation du substrat positive dans le cas des transistors NMOS) afin de rendre inactif ces transistors parasites. Cette méthode donne d'excellents résultats sans avoir besoin d'appliquer une tension de substrat élevée. Cependant, le circuit doit être conçu de manière à avoir une tension  $V_B$  positive et le concepteur doit prendre en compte les fuites supplémentaires générées en utilisant ce mode de polarisation ce qui peut aller à l'encontre d'une conception ultra-basse consommation.

L'effet « hump » est aussi mis en évidence à cause d'un  $V_T$  des transistors en bord d'active plus faible. Ainsi, plusieurs solutions liées au procédé de fabrication permettent d'augmenter la tension de seuil des transistors parasites en augmentant (localement) l'épaisseur d'oxyde ONO [Schw05] ou le dopage canal en utilisant des oxydes nitrurés qui limitent la diffusion des dopants [Ito82]. Dans [Kim00], ou encore en diminuant le dopage de la grille dans [Park08].

D'autres solutions liées à la conception de nouveaux transistors (transistors carrés, octogonaux et circulaires) peuvent être utilisées. En effet, ces architectures ne présentent pas par construction de jonction active/STI recouverte par du poly-silicium et n'ont, par conséquent, pas de transistors de bord à l'origine de l'effet « hump ».

---

### III. Contribution à l'amélioration de l'appariement en fonction du stress porteurs chauds

#### A. Présentation des différents stress électriques

##### 1 Principe du stress par injection de porteurs chauds (HCI<sup>12</sup>)

Lorsque le transistor MOS est polarisé en régime saturé (forte valeur de la tension de drain  $V_D$ ), un champ électrique très important est présent dans la zone de pincement entre l'extrémité du canal et le drain. Ce champ électrique élevé accélère fortement les porteurs libres du canal qui acquièrent alors une énergie suffisante (« porteurs chauds ») pour induire le phénomène d'ionisation par impact en générant des paires électron-trou additionnelles. La présence de porteurs chauds peut être à l'origine de la dégradation de l'oxyde et de l'apparition d'un courant substrat  $I_B$  constitué de trous issus de l'ionisation par impact dans le cas des transistors NMOS ou d'électrons pour les PMOS. Dans le cas d'un stress porteurs chauds, la dégradation des paramètres d'un transistor MOS peut avoir pour effets [Bena08Th] :

- Une augmentation de la tension de seuil  $V_T$ ,
- Une dégradation de la mobilité des porteurs,
- Une diminution du courant de drain  $I_D$  et de la transconductance  $g_m$ ,
- Une diminution de la pente sous le seuil due à l'influence des états d'interface,
- Une augmentation du GIDL [Lope04Th].

La dégradation par stress porteurs chauds est asymétrique, les défauts se trouvent principalement du côté du drain.

Les premiers travaux sous contraintes électriques datent des années 60 [Deal67] et fin des années 80 pour l'étude de l'appariement des transistors MOS [Laks86] [Pelg89]. Un des premiers travaux étudiant l'évolution de l'appariement des transistors MOS sous stress électrique montre clairement une dégradation accentuée sur les transistors courts et les transistors PMOS [Mich92b]. A la fin des années 90, il faut souligner les nombreux travaux liés aux étages différentiels [Thew94], aux miroirs et sources de courant [Thew96], et plus généralement aux applications analogiques [Thew99] [Thew01]. Dans [Chen01], une étude porteurs chauds portant sur les transistors NMOS et une étude NBTI (« Negative Bias Temperature Instabilities ») pour les PMOS montre qu'une légère variation de la tension de drain ou de grille lors du stress accentue le désappariement des transistors. L'appariement des transistors NMOS et PMOS sous stress porteurs chauds est également étudié dans [Lin05] et sous stress NBTI [Wu08].

##### 2 Protocole expérimental

D'après l'état de l'art existant, l'étude de la dégradation des paramètres des transistors MOS se focalise particulièrement sur l'évolution de l'appariement sous stress porteurs chauds. Sachant que ce stress porteurs chauds est accéléré avec la diminution de la longueur  $L$  [Bena08Th], l'analyse est faite sur trois

---

<sup>12</sup> « Hot Carrier Injection »

transistors GO2 (oxyde de grille de 65Å) de différentes largeurs ( $W= 0.6, 1$  et  $10\mu\text{m}$ ) et ayant une longueur fixe  $L=0.38\mu\text{m}$ . Les conditions les plus dégradantes (déterminées pour la technologie 90nm étudiée) sont utilisées pour stresser les transistors. A savoir, la méthode du courant de substrat maximum pour les transistors NMOS (avec  $V_G=2.5\text{V}$  et  $V_D = 5.4\text{V}$ ) et la condition de polarisation  $V_G=V_D = -5.4\text{V}$  pour les transistors PMOS.

Même si la probabilité de voir fonctionner les transistors dans de telles conditions de polarisation dans un circuit analogique est extrêmement faible, ces tensions de grille et de drain sont utilisées afin d'obtenir une dégradation significative avec un temps de stress limité permettant une analyse statistique correcte.

Le nombre de puces testées pour chaque étude est de 64. Le temps total de stress est de 3000s avec trois séquences de mesures intermédiaires à 100, 300 et 1000s.

## B. Dégradation des transistors NMOS et PMOS sous stress HCI

### 1 Dégradation de la tension de seuil $V_T$

En prenant les valeurs médianes (valeurs à 50%) de distributions cumulées pour les transistors NMOS et PMOS pour chaque dimension, l'évolution du  $V_T$  en fonction de la largeur des transistors est présentée Fig. IV-21.

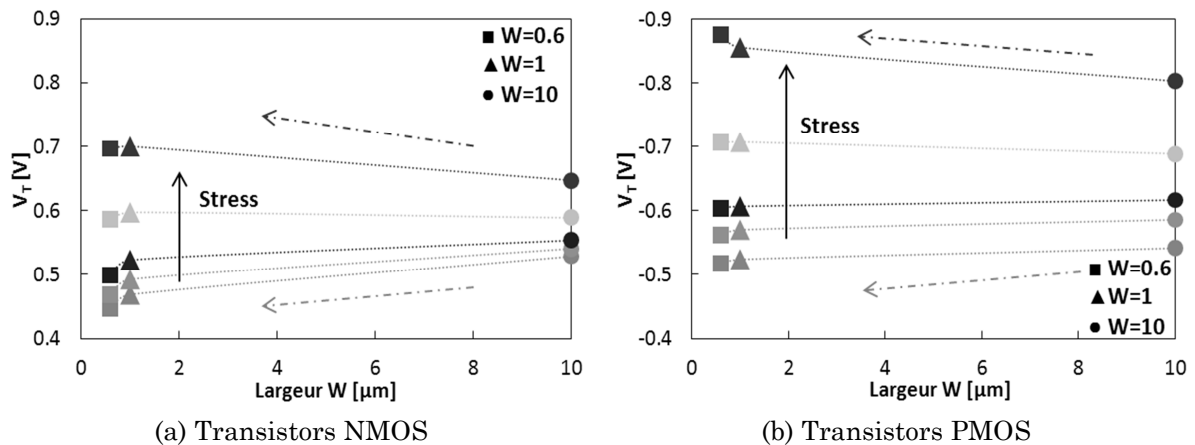


Fig. IV-21 : Evolution du  $V_T$  en fonction du stress

On peut noter qu'au cours du stress (0, 100, 300, 1000 et 3000s), la dégradation est d'autant plus importante que  $W$  est petit. Cette constatation a déjà été observée dans [Nish96] sous stress porteurs chauds puis plus tard dans [Math08] sous stress NBTI ou sur transistors TFT (« Thin Film Transistor ») [Huan08]. Cet effet est attribué aux transistors de bord se dégradant plus vite que le transistor principal. Ainsi, à cause d'une dégradation plus importante sur les transistors étroits, la tension de seuil des dispositifs  $W=0.6\mu\text{m}$  et  $W=1\mu\text{m}$  augmente plus rapidement et devient supérieure au  $V_T$  du transistor  $W=10\mu\text{m}$  après un stress porteurs chauds.

### 2 Dégradation de l'appariement de la tension de seuil $\Delta V_T$

Comme pour le  $V_T$ , la dégradation du paramètre  $\Delta V_T$  augmente de manière aléatoire avec le nombre de défauts générés. Cela se traduit par une augmentation des fluctuations locales avec le stress. Cette augmentation est

représentée Fig. IV-22 sur les transistors NMOS et PMOS pour les trois largeurs étudiées.

La loi d'appariement définie par [Pelg89] est également tracée (traits pleins) afin de représenter le comportement de l'appariement de la tension de seuil au cours du stress. Cela permet de constater que, pour un stress donné, les données de mesure sont toujours en accord avec la règle d'appariement.

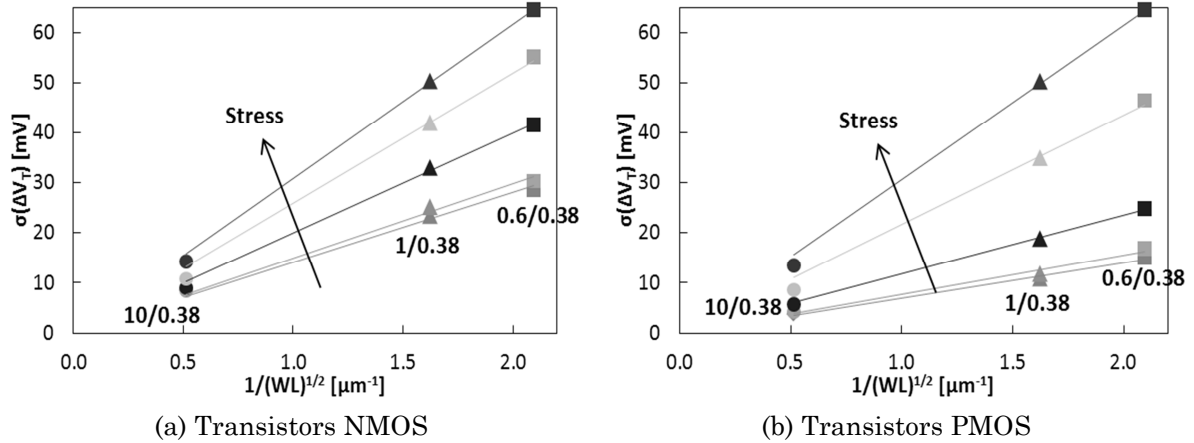


Fig. IV-22 : Comparaison NMOS/PMOS de l'appariement de la tension de seuil en fonction du stress

Les valeurs du paramètre d'appariement  $A_{VT}$  obtenues à partir des données de la Fig. IV-22 sont reprises Tableau IV-2. Comme pour les valeurs de  $V_T$ , l'appariement de la tension de seuil est également fortement dégradé par le stress HC. Cette dégradation est également plus importante sur les transistors PMOS.

Tableau IV-2 : Comparaison NMOS/PMOS du paramètre d'appariement de la tension de seuil en fonction du stress

Temps de stress [s]	0	100	300	1000	3000
$A_{VT}$ [mV.μm] - NMOS	14.1	15.0	134	20.0	30.8
$A_{VT}$ [mV.μm] - PMOS	7.0	7.8	11.8	21.8	30.7

## C. Etude des transistors octogonaux

### 1 Présentation des transistors octogonaux

Les transistors octogonaux ne présentent pas, par conception, de poly-silicium recouvrant une jonction Active/STI et donc de transistors de bord. Ils sont composés d'une zone d'active sur laquelle est dessiné un anneau octogonal de poly-silicium pour définir la grille du transistor. Le drain correspond à la surface d'active au centre du dispositif et la source à la zone d'active à l'extérieur de l'anneau octogonal de poly-silicium. Les contacts de grille sont placés directement sur la grille en poly-silicium sur active. Ce « layout » de contact n'est généralement pas autorisé dans les circuits standards. Cependant, aucun problème de fiabilité n'a été détecté lors de nos mesures.

La largeur  $W$  est donnée par la valeur du périmètre moyen de l'anneau octogonal de poly-silicium mesuré en son milieu. Pour des raisons simples de proportions

géométriques, la conception de tels transistors impose une largeur  $W$  minimale en fonction de la longueur choisie. Ainsi, pour  $L=0.38\mu\text{m}$ , les dimensions données Tableau IV-3 sont étudiées :

Tableau IV-3 : Largeurs  $W$  mesurées suivant le type de transistor ( $L=0.38\mu\text{m}$ )

Transistors	Largeur $W$ [ $\mu\text{m}$ ]
Standard	0.6, 1, 10
Octogonal	4.5, 7.2, 10, 20

Ces valeurs permettront d'obtenir le paramètre d'appariement de  $V_T$ . Cependant, l'étude se focalise sur la seule largeur commune pour comparer les transistors standards et les transistors octogonaux :  $W=10\mu\text{m}/L=0.38\mu\text{m}$ . Fig. IV-23 présente le « layout » des structures d'appariement étudiées.

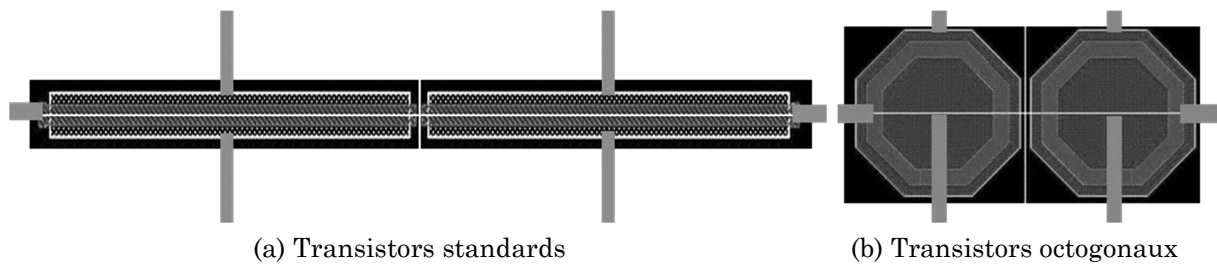


Fig. IV-23 : Vue « layout » simplifiée de la structure de test étudiée ( $W=10\mu\text{m}/L=0.38\mu\text{m}$ )

## 2 Comparaison entre transistors standard et octogonal

L'impact de ces transistors octogonaux sur les appariements en tension et en courant autour de la tension de seuil et dans la zone sous le seuil est présenté Fig. IV-24. La remontée sous le seuil, conséquence de la conduction du courant assurée par les transistors parasites, n'est plus présente sur le transistor octogonal car il ne présente pas d'effet « hump ». De plus, le meilleur appariement en tension est obtenu en régime de faible inversion et reste constant suivant la tension de grille.

L'augmentation de l'appariement en courant dans la zone sous le seuil à cause des transistors de bord est également démontrée en étudiant les fluctuations locales du courant de drain. En effet, les variations changent clairement autour de  $0V$  et augmentent par rapport aux fluctuations des transistors parasites. Ce phénomène n'est bien sûr pas présent sur le transistor octogonal où l'appariement en courant est nettement amélioré dans la zone sous le seuil par rapport au transistor standard.

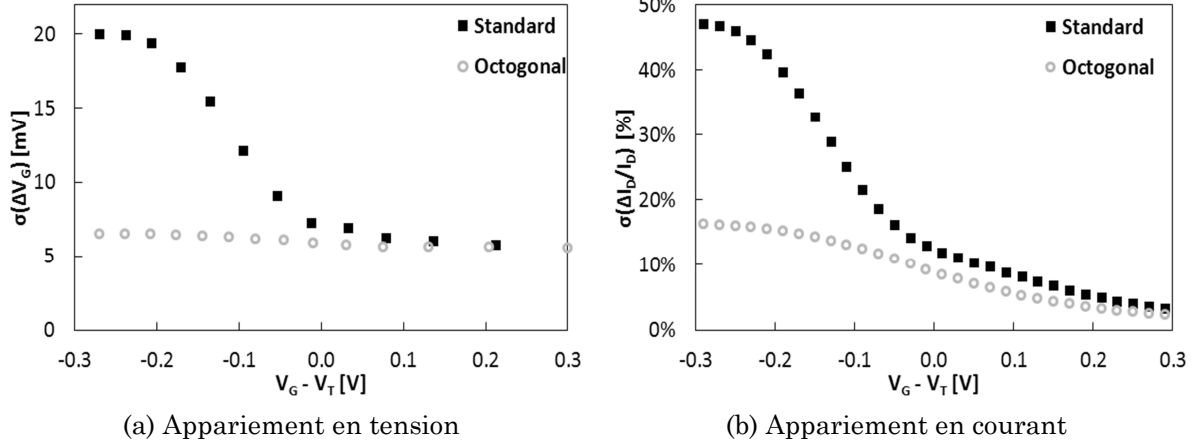


Fig. IV-24 : Comparaison de l'appariement en tension et en courant entre un transistor standard et un transistor octogonal ( $W=10\mu\text{m}/L=0.38\mu\text{m}$ )

### a. Dégradation de la tension de seuil

L'évolution de la tension de seuil en fonction de la largeur des transistors est tracée au cours du stress Fig. IV-25.

- Pour les transistors standards : sans stress, le  $V_T$  diminue avec la diminution de  $W$  (flèche verte). Au cours du stress, la dégradation étant plus importante pour les  $W$  faibles, la tendance s'inverse (flèche rouge) et la tension de seuil des transistors étroits devient supérieure aux transistors larges.

- Pour les transistors octogonaux : une très légère augmentation de la tension de seuil avec la diminution de la largeur est observable sans stress. Avec les défauts générés pendant le stress, cette tendance est amplifiée et peut devenir significative.

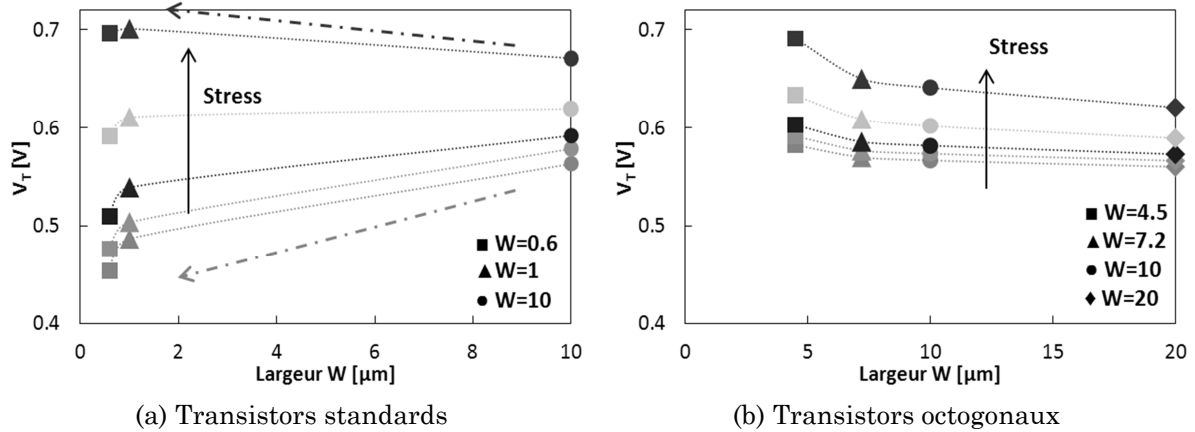


Fig. IV-25 : Evolution du  $V_T$  en fonction du stress

Tableau IV-4 : Evolution de  $V_T$  et  $\delta V_T$  en fonction du stress ( $W=10\mu\text{m}/L=0.38\mu\text{m}$ )

Temps de stress [s]	0	100	300	1000	3000
$V_T$ [V] - Standard	0.563	0.578	0.592	0.620	0.671
$V_T$ [V] - Octogonal	0.567	0.573	0.582	0.602	0.641
$\delta V_T$ [V] - Standard		15.3	28.8	56.6	108.0
$\delta V_T$ [V] - Octogonal		6.9	15.1	35.1	74.5

Le Tableau IV-4 reprend les valeurs de  $V_T$  des transistors standards et octogonaux ( $W=10\mu\text{m}/L=0.38\mu\text{m}$ ) et donne le  $\delta V_T$  au cours du stress. Comme vu précédemment, l'évolution de  $V_T$  est moins importante sur les transistors octogonaux étant donné l'absence de transistors de bord. Ainsi, le décalage de  $V_T$  est donc plus faible et confirme une nouvelle fois qu'une amélioration de la fiabilité des circuits peut être obtenue avec l'utilisation de transistors sans effet « hump ».

### b. Dégradation de l'appariement de la tension de seuil

L'étude de l'appariement de la tension de seuil durant la contrainte électrique HCI pour les transistors standards et octogonaux est donnée Fig. IV-26.

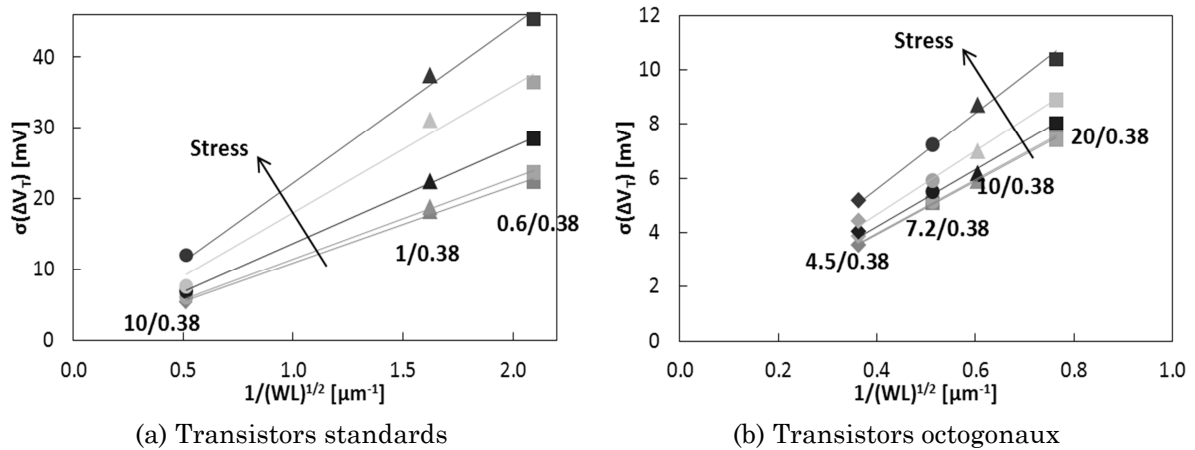


Fig. IV-26 : Comparaison entre transistors standards et octogonaux de l'appariement de la tension de seuil en fonction du stress

Les valeurs du paramètre d'appariement de la tension de seuil  $AV_T$  obtenues à partir de la figure précédente sont résumées Tableau IV-5. On note une nette amélioration de d'appariement avec l'utilisation de transistors octogonaux. Cela vient du fait que la détérioration des transistors standards est principalement due aux transistors parasites. Etant donné que les transistors octogonaux n'ont pas de transistors de bord, ils présentent intrinsèquement un meilleur appariement en  $V_T$  et bénéficient d'une dégradation sous stress HCI plus faible ce qui leur confère un avantage pour la conception de circuits analogiques précis et fiables.

Tableau IV-5 : Comparaison entre transistors standards et octogonaux du paramètre d'appariement en  $V_T$  en fonction du stress

Temps de stress [s]	0	100	300	1000	3000
$AV_T$ [mV.μm] - Standard	10.9	11.4	13.7	18	22.2
$AV_T$ [mV.μm] - Octogonal	9.8	9.9	10.5	11.7	14.0

### 3 Inversion des sources et drains des transistors octogonaux

L'impact d'une zone active de drain à l'intérieur ou à l'extérieur de la grille polysilicium est étudié sur les mêmes dimensions. Les connexions de drain et de source sont simplement inversées.



Précédemment réalisée entre les transistors standards et octogonaux, l'évolution de la tension de seuil en fonction de la largeur des transistors octogonaux est comparée cette fois-ci entre les structures ayant la zone de drain à l'intérieur et à l'extérieur. Cette comparaison est illustrée Fig. IV-27.

Lors d'un stress porteurs chauds, la zone la plus stressée se situe proche du drain. Ainsi, en utilisant une connexion de drain à l'extérieur, les effets de coin lorsque le drain est à l'intérieur sont significativement atténués. En effet, lorsque le drain se trouve à l'extérieur, la section des huit coins proche du drain est plus importante ce qui dans ce cas, donne lieu à une dégradation moins prononcée.

De plus, la très légère augmentation de la tension de seuil avec la diminution de  $W$  (et qui reste constante au cours du stress) observée sur le transistor avec drain extérieur peut s'expliquer par le fait que localement, toujours avec ces effets de coin, la longueur du transistor est légèrement plus grande. Ainsi, pour les très faibles longueurs (comme celle de la structure de test étudiée), la tension de seuil se trouve localement augmentée.

Les valeurs du paramètre d'appariement de la tension de seuil  $AV_T$  pour les transistors octogonaux avec les drains intérieurs et avec les drains extérieurs sont données Tableau IV-6.

Comme démontré sur les transistors NMOS, l'appariement en  $V_T$  des transistors PMOS est clairement moins dégradé par un stress porteurs chauds en utilisant des transistors octogonaux avec la connexion de drain à l'extérieur.

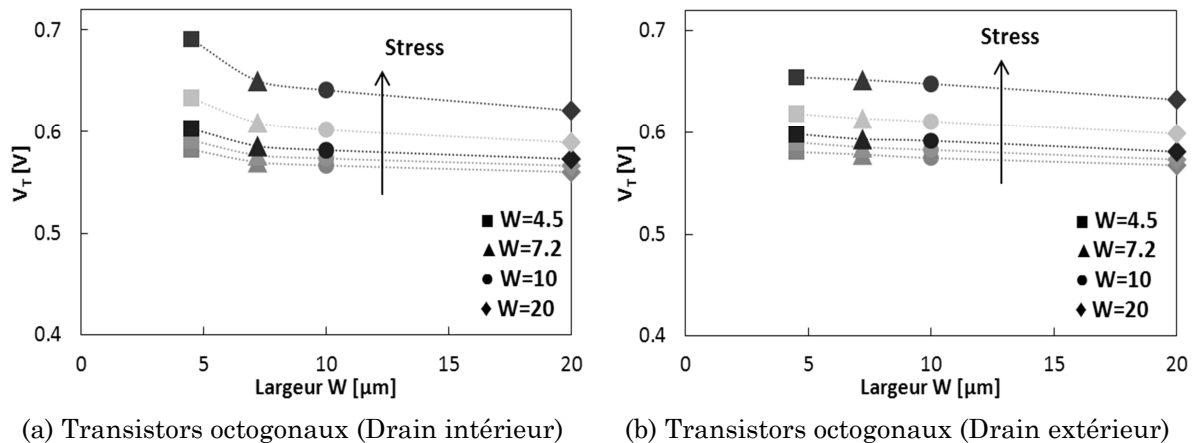


Fig. IV-27 : Evolution du  $V_T$  en fonction du stress

Tableau IV-6 : Comparaison du paramètre d'appariement en  $V_T$  entre transistors NMOS octogonaux et octogonaux (drain extérieur) en fonction du stress

Temps de stress [s]	0	100	300	1000	3000
$AV_T$ [mV.μm] - Octogonal, Drain intérieur	9.8	9.9	10.5	11.7	14.0
$AV_T$ [mV.μm] - Octogonal, Drain extérieur	9.3	9.5	9.9	10.5	11.4

## IV. Conclusion

Il a été montré dans ce chapitre que les fluctuations des dopants dans le canal, les poches et la grille sont les principales sources de variabilité, responsables de la dégradation de l'appariement des transistors MOS. La méthodologie de mesure appliquée pour mesurer les fluctuations locales des transistors MOS a été étudiée en décrivant les équipements de mesure et les méthodes de mesures utilisés pour la caractérisation de l'appariement des transistors MOS.

Afin d'avoir un bon appariement en tension et un gain important, les paires différentielles sont souvent polarisées sous le seuil en régime de faible inversion. Il a été prouvé que l'effet « hump » est responsable de la dégradation de l'appariement en tension d'une paire différentielle polarisée dans la zone sous le seuil. Ce phénomène est dû à l'apparition de transistors parasites étroits sur les bords du transistor principal à la jonction zone active/STI recouverte par du polysilicium. Un macro-modèle a été utilisé pour modéliser ces transistors de bord et simuler correctement la dégradation de l'appariement en tension observé en régime de faible inversion. Ces résultats ont confirmé le comportement de l'effet « hump » et montré qu'il est nécessaire pour les concepteurs d'utiliser le macro-modèle pour prédire le mieux possible l'appariement en tension d'une paire différentielle.

Une deuxième étude a consisté à évaluer l'impact du stress porteurs chauds sur la fiabilité des transistors MOS. Il a été montré que la courbure morphologique des transistors de bord des transistors MOS donne lieu localement à un champ électrique plus important notamment par rapport à celui présent sur le transistor principal. Cela ayant pour conséquence une augmentation de la dégradation de ces transistors de bord, notamment durant le stress porteurs chauds (plus le transistor est étroit, plus il est sensible aux transistors de bords ce qui augmente sa dégradation). A partir de ces résultats, une corrélation entre la dégradation de l'appariement de la tension de seuil et de la dérive de la tension de seuil au cours du temps a été démontrée.

Des transistors octogonaux ont été introduits et comparés avec des transistors standards. Ces transistors octogonaux permettent de supprimer l'effet « hump » du fait de l'absence de jonction STI/zone active recouverte par du poly-silicium. Cette suppression a été confirmée sur des caractéristiques  $I_D-V_G$  avec effet substrat. Par conséquent, les transistors octogonaux présentent un appariement en tension stable et minimal dans la zone sous le seuil. L'évolution de l'appariement des transistors octogonaux en fonction du stress porteurs chauds a montré une amélioration significative. Enfin, l'analyse des transistors octogonaux ayant le drain à l'extérieur de la grille poly-silicium a permis d'améliorer encore la fiabilité de ce type de transistors en supprimant des effets de coin de la grille octogonale. Tous ces résultats font des transistors octogonaux, en particulier ceux ayant le drain situé à l'extérieur, une solution prometteuse pour les applications analogiques faible consommation.



# Conclusion générale

Le test des **microcontrôleurs** passe par le test de chacun de ses éléments et conduit à la réalisation de flots de tests denses et complexes. La principale cause de perte de rendement de ce type de produit est liée à la partie Mémoire Non Volatile (MNV). Pour les technologies microcontrôleurs avancées intégrant de la MNV embarquée, un nouveau type de défaillance appelé « mismatch » et lié au non appariement des transistors des circuits analogiques est apparu. Ce phénomène, lié à des variations locales entre deux transistors MOS appariés peut générer une forte dispersion au niveau des réponses électriques des circuits analogiques. Concernant la partie purement numérique des microcontrôleurs, l'augmentation de la densité d'intégration associée à une utilisation systématique des règles de dessin minimales introduit de nouveaux mécanismes de défaillance liés aux interconnexions métalliques. Ainsi, les microcontrôleurs sont devenus des circuits fortement sensibles à la qualité du procédé de fabrication, mais aussi aux méthodes de conception. Les travaux de recherche présentés dans ce mémoire proposent des méthodes d'analyse de la fiabilité des circuits microcontrôleurs (partie mémoire, logique et analogique). Ces analyses utilisent plusieurs structures, qui vont de structures de test très simples embarquées dans les lignes de découpe, à des circuits de test plus complets appelés véhicules de test.

Les **MNV** sont de plus en plus confrontées à des problèmes de fiabilité dus essentiellement à la réduction des dimensions de la cellule mémoire élémentaire, réduction imposée par un accroissement de la densité d'intégration au niveau de la matrice de cellules mémoires. Dans ce contexte, le bon fonctionnement du dispositif est strictement lié à une meilleure compréhension des mécanismes de défaillance affectant les MNV. Par conséquent, la mise en place de solutions de test permettant d'évaluer la fiabilité de ce type de mémoire est devenue primordiale. Dans ce contexte, un véhicule de test, nommé TCF9, a été utilisé pour mener à bien deux études :

La première étude a consisté à extraire la densité de pièges de l'oxyde tunnel. Ce paramètre étant un élément majeur de la fiabilité des MNV. Ainsi, une loi d'évolution de la densité de pièges en fonction du nombre de cycles a été proposée.

La deuxième étude a permis d'établir une corrélation entre les tests de stress électrique et les tests de rétention. L'objectif étant d'anticiper les résultats des tests de rétention (lents) à partir de tests de stress (rapides).

Pour ces deux études, le véhicule de test TCF9, de par sa taille et ses fonctionnalités, s'est avéré un outil incontournable.

Dans un contexte de réduction de taille de transistors **numériques** et d'augmentations du nombre de niveaux de métaux, il est nécessaire de disposer d'outils de contrôle des paramètres du BEOL. Ainsi, l'optimisation d'une macro-cellule a été présentée dans le Chapitre III. Cette structure est un plan mémoire SRAM de 4096 bits dont chaque cellule peut osciller pour évaluer l'état des interconnexions entre les inverseurs qui la composent. Le mode mémoire a permis de détecter et de localiser des défauts « hard », et le mode d'oscillation a permis de détecter des défauts « soft » ainsi que les variations du procédé de fabrication. De plus, le plan mémoire possède des cellules interconnectées avec des niveaux de métaux différents (Métal1 à Métal4) dans le but de déterminer quel niveau a subi une variation. L'architecture de cette structure, nommée RAPITAC, est tout d'abord décrite. Le programme de test est ensuite présenté. Ce programme a permis d'évaluer l'état fonctionnel du plan mémoire, et de mesurer la fréquence d'oscillation de chaque point mémoire. Un logiciel a été développé pour calculer, entre autre, la transformée de Fourier discrète de chaque cellule. Le logiciel permet de représenter, à divers niveaux (puce, plaquette, lot), les

données fonctionnelles et fréquentielles. Grâce à ces données, il a été possible de savoir quel niveau du BEOL subit une variation et de caractériser cette variation.

En conception **analogique**, les performances de structures élémentaires comme les miroirs de courant ou les paires différentielles dépendent de l'appariement des dispositifs. Dans ce contexte, les travaux de ce manuscrit ont permis d'analyser et d'optimiser l'appariement des transistors MOS destinés aux applications analogiques. Cette étude a été réalisée dans un contexte de circuits faibles consommations fabriqués en technologie CMOS 90nm incluant une option de mémoire non-volatile embarquée. Les contraintes de faible consommation obligent parfois les concepteurs à polariser leurs structures sous le seuil. Cela est particulièrement vrai dans le cas des paires différentielles qui, en plus de proposer un gain important dans ce régime, présentent un appariement en tension théorique minimal. Cependant, la comparaison entre la mesure et la simulation des dispositifs étudiés a mis en évidence une nette dégradation attribuée à l'effet « hump » de cet appariement en tension en diminuant  $V_G$ . Cet effet parasite se traduit par la présence de transistors en bord d'active de chaque côté du transistor principal. Avec la diminution de la polarisation de grille, le courant passe de plus en plus par ces transistors parasites à cause de leur tension de seuil plus faible. L'appariement en tension ne dépend donc plus de la surface du transistor principal mais de celle des transistors parasites qui est beaucoup plus faible ayant pour conséquence une dégradation de l'appariement sous le seuil. Un macro-modèle basé sur trois transistors en parallèle permet de modéliser cet effet. Les origines physiques de l'effet « hump » ont été investiguées. L'étude de la concentration de dopants dans le canal a montré que cet effet parasite est lié à une ségrégation des atomes du canal en bord d'active. Différentes études ont été menées sur plusieurs types de transistor en introduisant une mesure paramétrique simple pour mieux appréhender ce phénomène. Finalement, à partir de toutes ces observations plusieurs solutions pour contrer cet effet parasite (dont l'utilisation de transistors « multi-fingers ») ont été proposées. Afin d'étudier le vieillissement de l'appariement des transistors MOS, une étude sous stress porteurs chauds a été réalisée. Cette étude a montré une corrélation entre la dégradation de l'appariement de la tension de seuil et de la dérive de la tension de seuil au cours du temps. Afin de limiter les problèmes liés à la présence de transistors de bord, des transistors octogonaux ont été introduits. Ces derniers présentent un appariement en tension stable et minimal dans la zone sous le seuil. De plus, les transistors octogonaux se dégradent nettement moins sous stress porteurs chauds que les transistors standards.

# Perspectives

L'évolution des mémoires permet de distinguer à l'heure actuelle plusieurs stades de maturité, à savoir :

- les mémoires de base (« Base line Memory »), actuellement utilisées dans la majorité des applications ;
- les mémoires prototypes (« Prototype Memory »), qui sont présentées dans l'ITRS<sup>1</sup> comme des remplaçants potentiels des mémoires actuelles en suivant la loi de Moore (« More Moore ») avec comme principal critère leur capacité d'intégration sub-nanométrique ;
- les mémoires émergentes (« Emerging Memory ») qui pourraient apparaître au-delà de la loi de Moore (« More than Moore »). Dans ce dernier cas, la recherche d'une mémoire universelle capable à terme de remplacer les mémoires non volatiles aussi bien que volatile est le premier critère de choix.

L'équipe « Mémoires » du laboratoire IM2NP, dans laquelle j'effectue mes travaux de recherche positionne ses activités dans ce contexte, en ciblant un large panel de solutions mémoires à tous les stades de maturité, à savoir mémoires de base, prototypes et émergentes.

L'ITRS identifie comme critique la fiabilité des mémoires prototypes et émergentes et, des circuits où elles sont embarquées. Ces circuits (aussi bien analogiques que numériques) sont nécessaires à la mise en œuvre de la mémoire durant les étapes d'écriture, d'effacement et de lecture. Les problèmes de fiabilité sont liés à de la réduction des dimensions (effet lithographique, variabilité, densité d'intégration, etc.), mais aussi à l'introduction de nouvelles options technologiques (High-k, grille métallique etc.). L'ensemble de ces éléments peut modifier les modes de défaillances dans le temps des circuits selon les sollicitations. Ainsi, mes travaux se poursuivent dans cette direction, en se basant sur l'expertise acquise à partir des études de fiabilité réalisées sur les MNV plus conventionnelles de type Flash et EEPROM.

Mes perspectives de travaux dans cette thématique à court et moyen terme ciblent plus particulièrement les mémoires prototypes de type ReRAM<sup>2</sup>. Il s'agit d'une nouvelle filière de mémoires non volatiles destinée à remplacer la filière Flash au-delà du nœud technologique 20 nm. Les caractéristiques des ReRAM en terme de performances pourraient séduire les fabricants de mémoires dans un avenir proche, grâce principalement à sa rapidité (écriture/effacement  $\approx 5$  ns), sa taille plus réduite ( $\approx 10$  nm), mais également son coût de fabrication moins élevé grâce à son intégration possible dans le BEOL. Dans la famille des ReRAM, les dispositifs OxRRAM<sup>3</sup> sont formés à partir de couches de matériaux incluant des oxydes métalliques dont la résistance peut être commutée entre un état haut et un état bas par application d'une tension aux bornes de la cellule mémoire [Waser07]<sup>4</sup>.

---

<sup>1</sup> International Technology Roadmap for Semiconductor 2009 (PIDS, ERD, Design)

<sup>2</sup> Resistive Random Access Memory

<sup>3</sup> Oxide-based Resistive RAM (OxRRAM)

<sup>4</sup> R. Waser, M. Aono, "Nanoionics- based resistive switching memories," Nature Materials, Vol.6, p. 833 (2007).

L'étude de fiabilité des OxRRAM se déroulera suivant deux niveaux :

- La fiabilité de la cellule mémoire élémentaire, qui est un réel verrou technologique avec la réduction des tailles des points mémoires et l'apparition de nouveaux modes de défaillance liés à la variabilité de la cellule [Chen11]<sup>5</sup>.
- La fiabilité et l'optimisation du plan mémoire, avec notamment des solutions basées sur l'architecture de la matrice mémoire ou l'optimisation des circuits périphériques [Yi11]<sup>6</sup>.

Un autre volet de mes perspectives de recherche est lié au développement de solutions mémoires innovantes. Cette thématique ambitieuse vise la compréhension des mécanismes de mémorisation avec un couplage entre caractérisation physique et électrique dans un premier temps, puis la modélisation de ces mécanismes au niveau physique et au niveau électrique (modèle compact) dans un second temps. Enfin, la proposition d'architectures au niveau circuit est envisagée. L'objectif final de ces études étant d'apporter des éléments pertinents quant à la viabilité de la mémoire considérée au niveau industriel.

Un autre axe de travail cible l'impact des radiations naturelles sur le fonctionnement des circuits intégrés et plus particulièrement des circuits mémoire. L'approche utilisée se base sur l'injection de fautes au niveau d'un simulateur électrique pour l'évaluation de la sensibilité du circuit aux radiations [Caste11]<sup>7</sup>. Cette étude sera étendue aux mémoires émergentes. D'autre part, des architectures matricielles de détection de particules sont envisagées [Mico12]<sup>8</sup>.

Pour mener à bien ces missions, des moyens matériels et un travail d'équipe sont nécessaires. Les moyens matériels (et logiciels) relativement complets (banc de caractérisation, testeur, outils de conception et de simulation électrique) permettent de suivre l'évolution des technologies. Quant aux moyens humains de l'équipe « Mémoires », ils m'ont permis et me permettront de continuer à m'épanouir dans ces nouvelles thématiques puisque la majorité des travaux réalisés et à venir sont le fruit d'un travail d'équipe.

---

<sup>5</sup> An Chen et al., "Variability of resistive switching memories and its impact on crossbar array performance", Proceedings of the IEEE International Reliability Physics Symposium (IRPS), pp. MY.7.1 - MY.7.4, 2011.

<sup>6</sup> Jaeyun Yi et al., "Requirements of bipolar switching ReRAM for 1T1R type high density memory array", IEEE Proc. of International Symposium on VLSI Technology, Systems and Applications (VLSI - TSA), p. 1- 2, 2011.

<sup>7</sup> Castellani-Coulié et al., "Optimization of SEU Simulations for SRAM Cells Reliability under Radiation", Journal of Electronic Testing: Theory and Application (JETTA), DOI: 10.1007/s10836-012-5281-8, 2011.

<sup>8</sup> Micolau G. et al. "SITARe: a SIMulation Tool for Analysis and diagnosis of Radiation Effects", IEEE Proceedings of International Latin American Test Workshop, LATW'12, 2012, accepted for publication.

# Curriculum vitae

## NOTICE INDIVIDUELLE ET FORMATION

Rubrique	Informations
Nom	<b>Hassen AZIZA</b>
Date et lieu de naissance	16 Mai 1976, à Nador (Maroc)
Nationalité	Française
Situation de Famille	Célibataire
Adresse personnelle	Le clos du puisatier, Bat F Le Logis Neuf, 13190, Allauch
Fonction	Maître de conférence – Ecole Polytechnique Universitaire de Marseille, département Microélectronique et télécommunications, Aix-Marseille Université
Adresse professionnelle	IMT – Technopôle de Château Gombert, 38 rue Frédéric Joliot-Curie, 13451 Marseille Cedex 20
Téléphone	+33-04-91-05-47-84
Fax	+33-04-91-05-47-82
Email	Hassen.Aziza@polytech.univ-mrs.fr

Année	Diplôme	Lieu
2001-2004	Doctorat Ecole doctorale : Physique, Modélisation et Sciences pour l'Ingénieur	Aix-Marseille Université
2000-2001	D.E.A. de micro et nano électronique	Aix-Marseille Université
1999-2001	Diplôme d'ingénieur en systèmes électroniques, spécialisation : microélectronique	Aix-Marseille Université



# Liste des Travaux

Type d'article	Nombre
Revues	13
Conférences Internationales avec actes	42
Conférences Internationales	8
Chapitre d'ouvrage	1
Concours internationaux	2
Brevets	2

## Revues

- [R1] **Aziza H.**, Portal J-M, Plantier J., Reliaud C., Regnier A, Ogier J.L., “Non Volatile Memory Reliability Prediction based on Oxide Defects Generation Rate during Stress and Retention Tests”, *Solid State Electronics*, 10.1016/j.sse.2012.05.069, pp. 151-155, **2012**.
- [R2] Y. Joly, L. Lopez, J. M. Portal, **Aziza H.**, J. L. Ogier, Y. Bert, F. Julien, and P. Fornara “Threshold Voltage Asymmetric Degradation on Octagonal MOSFET during HCI Stress”, Vol. 48, Issue 14, pp. 879 – 881, **2012**.
- [R3] Micolau G., Castellani K., **Aziza H.**, Portal J-M., “Contribution to SER Prediction: A New Metric Based on RC Transient Simulations”, *IEEE Transactions on Nuclear Science*, pp. 797 - 802, Vol. 59, Issue 5, **2012**.
- [R4] Castellani K, **Aziza H.**, Micolau G., Portal J-M., “Optimization of SEU Simulations for SRAM Cells Reliability under Radiation”, *Journal of Electronic Testing: Theory and Application (JETTA)*, 10.1007/s10836-012-5281-8, **2011**.
- [R5] Y. Joly, L. Lopez, J. M. Portal, **H. Aziza**, J. L. Ogier, Y. Bert, F. Julien, and P. Fornara, “Matching degradation of threshold voltage and gate voltage of NMOSFET after Hot Carrier Injection stress”, *Microelectronics Reliability*, vol. 51, N° 9-11, pp. 1561-1563, **2011**.
- [R6] Castellani K., Toure G., Portal J.-M., Ginez O., **Aziza H.**, Lesea A., “Circuit effect on Collection Mechanisms involved in Single Event Phenomena: Application to the Response of a NMOS Transistor in a 90nm SRAM Cell”, *IEEE Transactions on Nuclear Science*, Vol. 58, N° 3, pp. 870 - 876, **2011**.
- [R7] Rigaud F., Portal J-M., **Aziza H.**, Nee D., Vast J., Argoud F., Borot B., “Back-end soft and hard defect monitoring using a single test chip”, *Journal of Microelectronic Reliability*, , Vol. 51, N° 6, pp. 1136-1141, **2011**.
- [R8] Plantier J., **Aziza H.**, Portal J.M., Reliaud C, Regnier A., Ogier J.L., “EEPROM tunnel oxide lifetime reliability prediction based on fast electrical stress tests”, *Electronics Letters*, Vol. 46, N° 23, pp. 1568 - 1569, **2010**.
- [R9] Plantier J., **Aziza H.**, Portal J.M., Reliaud C., Regnier A., “An New Experimental Method To Extract EEPROM Tunnel Oxide Trap Density From Threshold Voltage Distributions”, *Journal of Non-Crystalline Solids*, Vol. 357, N° 8-9 (178 p.), pp.1971-1976, **2010**.

- [R10] **Aziza H.**, Perez A., Bergeret E., “Analog Blocks Design Automation”, *Journal of WSEAS transactions on circuits and systems (WSEAS)*, Vol. 8, N° 8, pp.686-695, **2009**.
- [R11] **Aziza H.**, Bergeret E., Portal J.M., Ginez O., “A Low-Power Oriented Design Methodology for Analog Blocks”, *Journal of Low Power Electronics (JOLPE)*, Vol. 4 N° 1, pp.60-67, **2008**.
- [R12] **Aziza H.**, Delsuc B., “Device and Memory array Models for Flash EEPROM Technology”, *Journal of WSEAS transactions on circuits and systems (WSEAS)*, Vol. 7, N° 4, pp.249-258, **2008**.
- [R13] Portal J.M., **Aziza H.**, Née D, “EEPROM Diagnosis Based on Threshold Voltage Embedded Measurement”, *Journal of Electronic Testing: Theory and Application (JETTA)*, Vol. 21, pp.33-42, **2005**.

---

### Conférences internationales avec actes et comité de lecture à diffusion publique

---

- [CP1] **Aziza H.**, Portal J-M, Plantier J., Reliaud C., Regnier A, Ogier J.L., “Development of an Empirical Model for Fast Prediction of NVM reliability”, *IEEE Proceedings of International Symposium Advanced Dielectrics and Related Devices, SiO2’12*, accepted for publication.
- [CP2] Zhao W.S., Zhang Y., Klein J.O., Querlioz D., Chabi D., Ravelosona D., Chappert C., Portal J.M., Bocquet M., **Aziza H.**, Deleruyelle D., Muller C., “Crossbar Architecture Based on 2R Complementary Resistive Switching Memory Cell”, *Proceedings of International Symposium on Nanoscale Architectures, NANOARCH’12*, accepted for publication.
- [CP3] G., Castellani K., **Aziza H.**, Portal J-M, Micolau G., “Investigation of a CMOS Oscillator Concept for Particle Detection and Diagnosis”, *IEEE Proceedings of International Latin American Test Workshop, LATW’12*, accepted for publication.
- [CP4] Haddad F., Rahajandraibe W., **Aziza H.**, Castellani K., Portal J-M, “Built-in Tuning of RFIC Passive Polyphase Filter by Process and Thermal Monitoring”, *IEEE Proceedings of International Latin American Test Workshop, LATW’12*, accepted for publication.
- [CP5] Micolau G., Castellani K., **Aziza H.**, Portal J-M, “SITARE: a SIMulation Tool for Analysis and diagnosis of Radiation Effects”, *IEEE Proceedings of International Latin American Test Workshop, LATW’12*, accepted for publication.
- [CP6] Joly Y., Lopez L., Portal J.M., **Aziza H.**, Bert Y., Julien F., Fornara P., “Active “multi-fingers”: Test structure to improve MOSFET matching in sub-threshold area”, *IEEE Proceedings of International Conference on Microelectronic Test Structures, ICMTS’12, USA, 2012*, accepted for publication.
- [CP7] **Aziza H.**, Portal J.M., Plantier J., Reliaud C., Regnier A., Ogier J.L., “Non Volatile Memory Reliability Prediction based on Oxide Defect Generation Rate during Stress and Retention Test”, *IEEE Proceedings of International Semiconductor Device Research Symposium, ISDRS’11, USA, 2011*, pp. 1-6.
- [CP8] **Aziza H.**, Bocquet M., Portal J-M., Muller C., “Bipolar ReRAM Reliability Evaluation based on Fault Injection and Simulation”, *IEEE Proceedings of International Workshop For Design & Test, IDT’11, Lebanon, 2011*, pp. 78-81.
- [CP9] **Aziza H.**, Bocquet M., Portal J.M., Muller C., “OxRRAM Cell Variability Impact Evaluation on Memory Performances using Electrical Simulations”, *IEEE*

- Proceedings of Non Volatile Memory Technology Symposium, NVMTS'11, China, **2011**, pp. 1-5.
- [CP10] **Aziza H.**, Portal J.-M., Castellani K., “Using design of experiment to diagnose analog blocks geometrical defects: Application to current reference circuits”, IEEE Proceeding of International conference on Design & Technology of Integrated Systems in nanoscale era, DTIS'11, Greece, **2011**, pp. 1-5.
- [CP11] Joly Y., Lopez L., Portal J.M., **Aziza H.**, Masson P., Ogier J., Bert Y., Julien F., Fornara P., “Octagonal MOSFET: Reliable device for low power analog applications”, in of Solid-State Device Research Conference, ESSDERC'11, Finland, **2011**, pp. 295-298.
- [CP12] Joly Y., Delalleau J., Lopez L., Portal J. M., **Aziza H.**, Bert Y., Julien F., Fornara P., “Poly-Silicon gate pre-doping implantation impact on MOSFET matching performances”, Proceedings of International Conference on Design & Technology of Integrated Systems in Nanoscale Era, DTIS'11, Greece, **2011**, pp. 1-4.
- [CP13] Joly Y., Truphemus L., Lopez L., Portal J.M., **Aziza H.**, Julien F., Fornara P., “Temperature and hump effect impact on output voltage spread of low power bandgap designed in the sub-threshold area”, in IEEE Proceedings of International Symposium on Circuits and Systems, ISCAS'11, Brazil, **2011**, pp. 2549-2552.
- [CP14] Castellani-Coulie K., Portal J.-M., Micolau G., **Aziza H.**, “Analysis of SEU parameters for the study of SRAM cells reliability under radiation”, IEEE Proceeding of Latin American Test Workshop, LATW'11, Brazil, **2011**, pp. 1-5.
- [CP15] Micolau G., **Aziza H.**, Castellani K., Portal J.-M. ,“Impact of SEU configurations on a SRAM cell response at circuit level”, IEEE Proceeding of Latin American Test Workshop, LATW'11, Brazil, **2011**, pp. 1-5.
- [CP16] Pulicani R., **Aziza H.**, Perez A., Bergeret E., Goducheau O., Degoirat H., “Simulation of Intrinsic Bipolar Transistor Mechanisms for future capacitor-less eDRAM on bulk substrate”, IEEE Proceedings of International conference on Electronics, Circuits, and Systems, ICECS'10, Greece, **2010**, pp. 966-969.
- [CP17] Castellani K., Toure G., Portal J.-M., Ginez O., **Aziza H.**, Lesea A., “Circuit effect on Collection Mechanisms involved in Single Event Phenomena: Application to the Response of a NMOS Transistor in a 90nm SRAM Cell”, IEEE Proceedings of International conference on Radiation and its Effects on Components and Systems, RADECS'10, Austria, **2010**, pp. 1-5.
- [CP18] Joly Y., Lopez L., Portal J.-M., **Aziza H.**, Bert Y., Julien F., Fornara P., Impact of hump effect on MOSFET mismatch in the sub-threshold area for low power analog applications, IEEE Proceedings of International conference on Solid-State and Integrated Circuit Technology, ICSICT'10, China, **2010**, pp. 1817-1819.
- [CP19] **Aziza H.**, Plantier J, Portal J.-M., Non Volatile Memory Signatures Extraction for Defects Diagnosis Purpose, IEEE Proceedings of International conference on Design & Technology of Integrated Systems in nanoscale era, DTIS'10, Hammamet, Tunisia, **2010**, pp. 1-5.
- [CP20] **Aziza H.**, NFC technology in Mobile Phone next-Generation Services, IEEE Proceedings of the 2nd International Workshop on Near Field Communication, NFC'10, Monaco, France, **2010**, pp. 21-26.
- [CP21] Plantier J., **Aziza H.**, Portal J.-M., “Retention test and electrical stress correlation to anticipate EEPROM tunnel oxide reliability issues”, IEEE Proceedings of

- International Semiconductor Device Research Symposium, ISDRS'09, USA, **2009**, pp. 1-5.
- [CP22] **Aziza H.**, Plantier J., Portal J-M., “A Realistic Fault Simulation Model for EEPROM Memories”, IEEE Proceedings of International Workshop For Design & Test, IDT'09, Saudi Arabia, **2009**, pp. 1-6.
- [CP23] Plantier J., **Aziza H.**, Portal J.M., Reliaud C., Investigation of EEPROM memories reliability during endurance and retention tests, IEEE Proceedings of International Conference on Design & Technology of Integrated Systems in Nanoscal Era, **DTIS'09**, Egypt, 2009, pp. 241-246.
- [CP24] Ginez O., Portal J-M., **Aziza H.**, An on-line testing scheme for repairing purposes in Flash memories, IEEE Proceedings of International Symposium on Design and Diagnostics of Electronic Circuits & Systems, DDECS '09, **2009**, pp. 120-123.
- [CP25] Rigaud F., Portal J.M., Dreux P., Vast J., **Aziza H.**, Bas G., Fast Embedded Characterization of FEOL Variations in MOS Devices, Microelectronic Test Structures, IEEE Proceedings of International Conference on Microelectronic Test Structures, ICMTS'09, USA, **2009**, pp. 205-208.
- [CP26] **Aziza H.**, Plantier J., Portal J.M., Reliaud C., Ginez O., Née D., A novel Diagnosis Methodology Based On I-V Signatures Extraction, IEEE Proceedings of Non Volatile Memory Technology Symposium, **NVMTS'08**, USA, 2008, pp.1-6.
- [CP27] **Aziza H.**, Bergeret E., Pérez A., A novel Design Methodology for Current Reference Circuits, IEEE Proceedings of International Conference on Electronics, Circuits and Systems , ICECS'08, Malta, **2008**, pp. 238-241.
- [CP28] Ginez O., Portal J-M., **Aziza H.**, A High-Speed Structural Method for Testing Address Decoder Faults in Flash Memories, Proceedings of the IEEE International Test Conference, ITC'08, USA, **2008**, pp.1-10.
- [CP29] Rigaud F., Portal J.M., **Aziza H.**, Nee D., Vast J., Argoud F., Borot B., Mixed test structure for soft and hard defect detection, IEEE Proceedings of International Conference on Microelectronic Test Structures, USA, ICMTS'08, **2008**, pp. 52-55.
- [CP30] Ginez O., Portal J-M., **Aziza H.**, “Reliability Issues in Flash Memories: An On-Line Diagnosis and Repair Scheme for Word Line Drivers”, Proceedings of Memory Technology and Design conference, ICMTD'08, France, **2008**, pp. 1-6.
- [CP31] **Aziza H.**, Portal J-M., Ginez O., Bergeret E., “An Efficient Diagnosis Methodology for Charge Pump Circuits: Application to Flash EEPROM devices”, IEEE Proceedings of International Conference on Design & Technology of Integrated Systems in Nanoscal Era, DTIS'08, Tunisia, **2008**, pp. 1-6.
- [CP32] **Aziza H.**, Portal J.M., Née D., Reliaud C., Argoud F., “Peripheral Circuitry Impact on EEPROM Threshold Voltage”, IEEE Non Volatile Memory Technology Symposium, NVMTS'07, USA, **2007**, pp. 20-24.
- [CP33] **Aziza H.**, Bergeret E., Pérez A., “An Automated Design Methodology for Charge Pump Circuits”, IEEE Proceedings of International Conference on Electronics, Circuits and Systems, ICECS'07, Morocco, **2007**, pp. 108-113.
- [CP34] **Aziza H.**, Bergeret E., Pérez A., Portal J.M., “An Efficient Charge Pump Model to Evaluate Impact of Design parameters on Charge Pump Circuits Performances: Application on RFID circuits”, IEEE Proceedings of Design & Test of Integrated Systems, DTIS'07, Morocco, **2007**, pp. 112-117.

- [CP35] Rigaud F., Portal J.M., **Aziza H.**, Née D., Vast J., Auricchio C., “Test Structure for Process and Product Evaluation”, Proceedings of the IEEE International Conference on Microelectronic Test Structures, ICMTS’07, Japan, **2007**, pp. 140-144.
- [CP36] **Aziza H.**, Delsuc B., Portal J.M., Née D., “Speeding up simulation time in EEPROM memory designs”, Proceedings of the Design & Test of Integrated Systems, DTIS’06, Tunisie, **2006**, pp. 285-288.
- [CP37] Regnier A., Portal J.M., **Aziza H.**, Masson P., Bouchakour R., Reliaud C., Née D., Mirabel J.M., “EEPROM Compact Model with SILC Simulation Capability”, IEEE Non Volatile Memory Technology Symposium, NVMTS’06, USA, **2006**, pp. 26-30.
- [CP38] **Aziza H.**, Portal J.M., Née D., “EEPROM Threshold Current Extraction: Silicon Validation”, Proceedings of the European Test Symposium, ETS’04, France, **2004**, pp. 81-87.
- [CP39] Portal J.M., **Aziza H.**, Née D., “EEPROM Memory: Threshold Voltage Built In Self Diagnosis”, Proceedings of the IEEE International Test Conference, ITC’03, USA, **2003**, pp. 23-28.
- [CP40] Portal J.M., **Aziza H.**, Née D., “EEPROM Memory Diagnosis Based on Threshold Current Extraction”, Proceedings of the IEEE Conference on Design of Circuits and Integrated Systems, DCIS’03, Spain, **2003**, pp. 133-139.
- [CP41] Portal J.M., Forli L., **Aziza H.**, Née D., “An Automated Methodology to Diagnose Geometric Defect in the EEPROM Cell”, Proceedings of the IEEE International Test Conference, ITC’02, USA, **2002**, pp. 31-36.
- [CP42] Portal J.M., Forli L., **Aziza H.**, Née D., “An Automated Design Methodology for EEPROM Cell (ADE)”, Proceedings of the IEEE International Workshop on Memory Technology Design and Testing, MTDT’02, France, **2002**, pp. 137-142.

### Conférences internationales

---

- [CR1] Pulicani, R., Goducheau, O., Degoirat, **H.**, **Aziza**, H., Pérez, A. & Bergeret, E. (2010). Embedded DRAM Future Trends: from 1T1C eDRAM to a 1T-Bulk gen2 Capacitor-less DRAM, In 40th European Solid-State Device Research Conference ESSDERC Fringe Poster Session, Seville (Spain), 13-17 Sept. **2010**
- [CR2] **Aziza H.**, Pannier P., “An access control solution based on NFC compliant cell phones”, Web proceeding of SmartMobility conference, SmartEvent’09, Sophia Antipolis, France, **2009**, pp.1-6.
- [CR3] **Aziza H.**, Portal J-M., Ginez O., Bergeret E., “An efficient Diagnosis Methodology for analog blocks: Application to Current Reference Circuits”, Proceedings of the European Test Symposium, ETS’08, Italie, **2008**.
- [CR4] Rigaud F., Portal J.M., **Aziza H.**, Née D., Vast J., Argoud F., Borot B., “Mixed Test Structure for Soft and Hard Defect Detection”, Proceedings of the 2<sup>nd</sup> IEEE International Workshop on Design for Manufacturability and Yield, DFMY’07, USA, **2007**.
- [CR5] Rigaud F., Portal J.M., **Aziza H.**, Née D., Vast J., Auricchio C., “Test Structure for Process and Product Performances Evaluation”, Proceedings of the 1<sup>st</sup> IEEE International Workshop on Design for Manufacturability and Yield, DFMY’06, USA, **2006**.

- [CR6] Portal J.M., Forli L., **Aziza H.**, Née D., Borot B., “BEOL Diagnosis based on Ring Oscillator Characterization”, 2nd IEEE International Workshop on Infrastructure IP, I-IP’04, USA, **2004**.
- [CR7] Portal J.M., **Aziza H.**, Née D, “EEPROM Memory: Threshold Voltage Built In Self Diagnosis”, Proceedings of the IEEE European Test Workshop, ETW’03, Hollande, **2003**, pp.81-87.
- [CR8] Portal J.M., Forli L., **Aziza H.**, Née D, “An Automated Geometric Defect Diagnosis Methodology for EEPROM Cell (AGDE) ”, Proceedings of the IEEE European Test Workshop, ETW’02, Grèce, **2002**, pp.343-344.

## Chapitre d’ouvrages

---

Pérez, A., Huynh Van Thieng, C., Charbouillot, S. & **Aziza, H.**, “An En/Decryption Machine based on Statistical Physics, In Cryptography”, InTech Publisher, ISBN 979-953-307-863-1, Nov. **2011**.

## Concours internationaux

---

[D1] **Lauréat (médaille d’argent) au concours international SIMagine** qui récompense les solutions innovantes dans le domaine des applications embarquées sur cartes à puce de type SIMCard avec l’application « TaggyNet », **2009**, Mobile World Congress, Barcelone. <http://www.simagine.info/nominees>.

Référencé sur le site de la société Gemalto, premier fabricant de cartes à puces : <http://www.gemalto.com/simagine/>.

[D2] **Finaliste au concours international SIMagine** qui récompense les solutions innovantes dans le domaine des applications embarquées sur cartes à puce de type SIMCard avec l’application « ParKeur », **2010**, Mobile World Congress, Barcelone. <http://www.simagine.info/nominees>.

## Brevets

---

- [B1] Portal J.M., Forli L., **Aziza H.**, *L2MP - ST Microelectronics*, Née D, Procédé de modélisation mathématique de composants électroniques et son utilisation pour la simulation, la détermination de géométrie et le diagnostic de défauts dans le composant, *N° de dépôt 0208652, N° de dossier 02-RO-029*.
- [B2] Portal J.M., **Aziza H.**, Née D, *L2MP - ST Microelectronics*, Circuit de caractérisation des tensions de seuils dans les mémoires non volatiles sous format numérique et traitement en vue du diagnostic des données obtenues, *N° de dépôt 0306184, N° de dossier 03-RO-059*.

# Activités d'encadrement

## Encadrement et co-encadrement de thèses

Type d'encadrement	Nombre	% d'encadrement global
Thèses	4 + 1*	220% + 30%*
Stages de Master	6	175%

### Thèses (*Ecole Doctorale : Physique, Modélisation et Sciences pour l'Ingénieur*)

- **Thèse : Loïc WELTER – 50%**

Sujet de thèse : *Contribution à l'amélioration de l'observabilité et de la reproductibilité des défauts, dans les circuits numériques pouvant embarquer des mémoires.*

Depuis le **1er novembre 2011**.

Directeurs de thèse : H. AZIZA (50%) et J.-M. PORTAL (50%)

Financement : *Contrat CIFRE, STMicroelectronics-Rousset*

- **Thèse : Yoan JOLY – 70%**

Sujet de thèse : *Etude de la robustesse du procédé de fabrication et de la conception des blocs analogiques pour les applications microcontrôleur avec mémoire non volatile embarquée.*

Depuis le **1er octobre 2008**. Soutenue le vendredi 16 décembre 2011.

Directeurs de thèse : H. AZIZA (70%) et J.-M. PORTAL (30%)

Financement : *Contrat CIFRE, STMicroelectronics-Rousset*

- **Thèse: Fabrice RIGAUD – 50%**

Sujet de thèse : *Mise en place d'une infrastructure permettant le diagnostic des défauts et des variations du procédé de fabrication en technologies 120 nm, 110nm et 90nm ciblant les composants circuits digitaux.*

Depuis le **15 Septembre 2006**. Soutenue le 24 Novembre 2010.

Directeurs de thèse : H. AZIZA (50%) et J.-M. PORTAL (50%)

Financement : *Contrat CIFRE, STMicroelectronics-Rousset*

- **Thèse: Jeremy PLANTIER – 50%**

Sujet de thèse : *Méthodes de test et de diagnostic dans les mémoires non volatiles : application à la mémoire Flash.*

Depuis le **1er octobre 2007**. Date de soutenance prévue en **Décembre 2012**.

Directeurs de thèse : H. AZIZA (50%) et J.-M. PORTAL (50%)

Financement : *Contrat CIFRE, STMicroelectronics-Rousset*

- \* **Thèse : Rémi PULLICANI – 30%**

Sujet de thèse : *Nouvelles architectures et solutions innovantes pour mémoires dynamiques embarquées en technologie CMOS 28nm et au-delà.*

Depuis le **1er novembre 2008**. Abandon à l'issue des 3 années pour raisons personnelles.

Directeurs de thèse : H. AZIZA (30%), E. BERGERET (30%) et A. PEREZ (40%)

Financement : *Contrat CIFRE, STMicroelectronics-Crolles*

# Références

- [Agar07] S. N. K. Agarwal, "Characterizing Process Variation in Nanometer CMOS," in DAC, San Diego, 2007, pp. 396-399.
- [Agga05] D. Aggarwal, A. Jain, P. Rohilla, and A. Vegetti, "KNOSSOSM10," ST Microelectronics datasheet, 2005.
- [Ahsa06] I. Ahsan, et al., "RTA-Driven Intra-Die Variations in Stage Delay, and Parametric Sensitivities for 65nm Technology," in Symp. on VLSI Technology, 2006, pp. 170-171.
- [Andr09b] P. Andricciola, H. P. Tuinhout, B. De Vries, N. A. H. Wils, A. J. Scholten, and D. B. M. Klaassen, "Impact of interface states on MOS transistor mismatch", IEEE International Electron Devices Meeting, 7-9 Dec. 2009, pp. 1-4.
- [Asen00] A. Asenov and S. Saini, "Polysilicon gate enhancement of the random dopant induced threshold voltage fluctuations in sub-100 nm MOSFETs with ultrathin gate oxide", IEEE Transactions on Electron Devices, vol. 47, 2000, pp. 805-812.
- [Asen03] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs", IEEE Transactions on Electron Devices, vol. 50, 2003, pp. 1837-1852.
- [Asen98] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 $\mu$ m MOSFET's: A 3-D atomistic simulation study", IEEE Transactions on Electron Devices, vol. 45, 1998, pp. 2505-2513.
- [Aseno99] A. Asenov and S. Saini, "Suppression of random dopant-induced threshold voltage fluctuations in sub-0.1  $\mu$ m MOSFET's with epitaxial and delta-doped channels", IEEE Transactions on Electron Devices, vol. 46, 1999, pp. 1718-1724.
- [Bast98] J. Bastos, A. M. Marques, M. S. J. Steyaert, and W. Sansen, "A 12-bit intrinsic accuracy high-speed CMOS DAC", IEEE Journal of Solid-State Circuits, vol. 33, 1998, pp. 1959-1969.
- [Bena08Th] C. Benard, PhD Thesis, "Etude de phenomenes de degradation des transistors MOS de type porteurs chauds et negative bias temperature instability (NBTI)", University of Provence, FRANCE, 2008.
- [Bert04] D. Bertozzi, L. Benini, Xpipes: a network-on-chip architecture for gigascale systems-on-chip, IEEE Circuits and Systems Magazine, Issue 2, 2004, pp. 18-31.
- [Brie97] O. Brière, A. Halimaoui et G. Ghibaudo, "Breakdown characteristics of ultra thin gate oxides following field and temperature stresses". Solid State Electronics, Vol. 41, 1997, pp. 981-985.
- [Brow98] W.D. Brown, J.E. Brewer, Nonvolatile semiconductor memory technology, IEEE PRESS, New York, 1998.
- [Cao98] M. Cao, P. Vande Voorde, M. Cox, and W. Greene, "Boron diffusion and penetration in ultrathin oxide with poly-Si gate", IEEE Electron Device Letters, vol. 19, 1998, pp. 291-293.
- [Capp99] P. Cappelletti, C. Golla, P. Olivo, E. Zanoni, FLASH Memories, Kluwer Academic Publishers, 1999.
- [Carv89] G. le Carval, Contribution de l'étude physique de l'injection par porteurs chauds à travers l'oxyde de grille des TMOS dans le cadre du fonctionnement des mémoires non-volatiles, Thèse, Institut National Polytechnique de Grenoble, 1989.
- [Caste11] Castellani-Coulié K, Aziza H., Micolau G., Portal J-M. "Optimization of SEU Simulations for SRAM Cells Reliability under Radiation", Journal of Electronic Testing: Theory and Application (JETTA), 2011, pp. 85-124.
- [Cath08a] A. Cathignol, S. Bordez, A. Cros, K. Rochereau, and G. Ghibaudo, "Abnormally High Current Local Fluctuations in Heavily Pocket-implanted Bulk Long MOSFET", International Symposium on VLSI Technology, Systems and Applications, 21-23 April 2008, pp. 167-168.



- [Cath08b] A. Cathignol, B. Cheng, D. Chanemougame, A. R. Brown, K. Rochereau, G. Ghibaudo, and A. Asenov, "Quantitative Evaluation of Statistical Variability Sources in a 45-nm Technological Node LP N-MOSFET", IEEE Electron Device Letters, vol. 29, 2008, pp. 609-611.
- [Cath09] A. Cathignol, S. Bordez, A. Cros, K. Rochereau, and G. Ghibaudo, "Abnormally high local electrical fluctuations in heavily pocket-implanted bulk long MOSFET", Solid-State Electronics, vol. 53, pp. 127-133.
- [Catha] A. Cathignol, C. Antoine, H. Samuel, C. Robin, C. Philippe, P. Arnaud, R. Krysten, S. Thomas, and G. Gerard, "High Threshold Voltage Matching Performance on Gate-All-Around MOSFET", European Solid-State Device Research Conference, Sept. 2006, pp. 379-382.
- [Cathb] A. Cathignol, K. Rochereau, S. Bordez, and G. Ghibaudo, "Improved methodology for better accuracy on transistors matching characterization", IEEE International Conference on Microelectronic Test Structures, 6-9 March 2006, pp. 173-178.
- [Chen01] Y. Chen, Z. Jonathon, S. Tedja, H. Frank, and A. S. Oates, "Stress-induced MOSFET mismatch for analog circuits", IEEE International Integrated Reliability Workshop, 2001, pp. 41-43.
- [Chen03] K.-C. Chen, S. Hsueh-Hao, H. Yaw-Lin, H. Cheng-Chen, H. Chung, S. Pan, and L. Chih-Yuan, "Applications of single-wafer rapid-thermal processing to the manufacture of advanced flash memory", IEEE Transactions on Semiconductor Manufacturing, vol. 16, 2003, pp. 128-137.
- [Chen11] An Chen et al., "Variability of resistive switching memories and its impact on crossbar array performance", Proceedings of the IEEE International Reliability Physics Symposium (IRPS), 2011, pp. MY.7.1-MY.7.4.
- [Croo02b] J. A. Croon, G. Storms, S. Winkelmeier, I. Pollentier, M. Ercken, S. Decoutere, W. Sansen, and H. E. Maes, "Line edge roughness: characterization, modeling and impact on device behavior", International Electron Devices Meeting, 2002, pp. 307-310.
- [Croo02c] J. A. Croon, H. P. Tuinhout, R. Difrenza, J. Knol, A. J. Moonen, S. Decoutere, H. E. Maes, and W. Sansen, "A comparison of extraction techniques for threshold voltage mismatch", IEEE International Conference on Microelectronic Test Structures, 8-11 April 2002, pp. 235-240.
- [Croo03] J. A. Croon, L. H. A. Leunissen, M. Jurczak, M. Benndorf, R. Rooyackers, K. Ronse, S. Decoutere, W. Sansen, and H. E. Maes, "Experimental investigation of the impact of line-edge roughness on MOSFET performance and yield", Conference on European Solid-State Device Research, 16-18 Sept. 2003, pp. 227-230.
- [Deal67] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of the Surface-State Charge ( $Q_{ss}$ ) of Thermally Oxidized Silicon", Journal of The Electrochemical Society, vol. 114, 1967, pp. 266-274.
- [DeBl98] J. De Blauwe, J. Van Houdt, D. Wellekens, G. Groeseneken et H.E. Maes, "SILC-related effects in flash EEPROM's part 1 : Prediction of steady- state SILC-related disturb characteristics". IEEE Trans. Electron Devices, Vol. 45, pp. 1151-1760, 1998.
- [Denn74] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, Design of Ion-Implanted MOSFET's with Very Small Dimensions, IEEE J. Solid-State Circuits SC-9, 1974, 256-268.
- [Depa95] M. Depas, M.M. Heyns et P.W. Mertens, "Soft breakdown of ultra-thin gate oxide layers". ESSDERC, 1995.
- [Dick76] J. Dickson, On chip high voltage generation in MNOS integrated circuits using an improved voltage multiplier technique", IEEE J. Sol. St. Cir., Vol.11, 1976, p.374.
- [Difr00] R. Difrenza, P. Llinares, G. Ghibaudo, E. Robillart, and E. Granger, "Dependence of Channel Width and Length on MOSFET Matching for 0.18  $\mu$ m CMOS Technology", European Solid-State Device Research Conference, 11-13 September 2000, pp. 584-587.
- [Difr01] R. Difrenza, P. Llinares, G. Morin, E. Granger, and G. Ghibaudo, "A New Model for Threshold Voltage Mismatch Based on the Random Fluctuations of Dopant Number in the MOS Transistor Gate", European Solid-State Device Research Conference, 11-13 September 2001, pp. 299-302.

- [Difr03a] R. Difrenza, P. Llinares, and G. Ghibaudo, "The impact of short channel and quantum effects on the MOS transistor mismatch", *Solid-State Electronics*, vol. 47, 2003, pp. 1161-1165.
- [Difr03b] R. Difrenza, J. C. Vildeuil, P. Llinares, and G. Ghibaudo, "Impact of grain number fluctuations in the MOS transistor gate on matching performance", *IEEE International Conference on Microelectronic Test Structures*, 17-20 March 2003, pp. 244-249.
- [Doll01] A. Dollberg, J. Oehm, R. Wunderlich, and K. Schumacher, "Influence of local matching effects on the accuracy of a sequential A/D-converter", *IEEE International Conference on Electronics, Circuits and Systems*, 2001, pp. 389-392.
- [Fare08] L. Fares, "Observation from die to atom," Cours de master "Microelectronics Technology & Manufacturing Management", ST university, rousset, 2008.
- [Forl05] L. Forli, et al., "Test et Diagnostic de Défauts dans les Interconnexions Métalliques des Circuits Numériques par Infrastructures "IP", Université de Aix-Marseille 1 Thèse, 2005.
- [Fuji07] M. Fujii, et al., "A Large Scale, Flip-Flop RAM imitating a logic LSI for fast development of process technology," in *ICMTS*, 2007, pp. 131-134.
- [Gupt02] V. Gupta and A. Rincon-Mora, "Predicting the effects of error sources in bandgap reference circuits and evaluating their design implications", *Midwest Symposium on Circuits and Systems*, 4-7 Aug. 2002, pp. 575-578.
- [Gupt05] V. Gupta and G. A. Rincon-Mora, "Predicting and designing for the impact of process variations and mismatch on the trim range and yield of bandgap references", *International Symposium on Quality of Electronic Design*, 21-23 March 2005, pp. 503-508.
- [Gust06] C. Gustin, A. Mercha, J. Loo, V. Subramanian, B. Parvais, M. Dehan, and S. Decoutere, "Stochastic Matching Properties of FinFETs", *IEEE Electron Device Letters*, vol. 27, 2006, pp. 846-848.
- [Hess01] C. Hess, et al., "Fast Extraction of Defect Size Distribution Using a Single Layer Short Flow NEST Structure," *IEEE transactions on semiconductor manufacturing*, vol. 14, no. 4, 2001, pp. 330-337.
- [Hoen72] B. Hoeneisen and C. A. Mead, "Current-voltage characteristics of small size MOS transistors", *IEEE Transactions on Electron Devices*, vol. 19, 1972, pp. 382-383.
- [Hook03] T. B. Hook, et al., "Lateral Ion Implant Straggle and Mask Proximity Effect," *IEEE Trans on Elec. Devices*, vol. 50, 2003, pp. 1946-1951.
- [Hook03] T. B. Hook, J. Brown, P. Cottrell, E. Adler, D. Hoyniak, J. Johnson, and R. Mann, "Lateral ion implant straggle and mask proximity effect", *IEEE Transactions on Electron Devices*, vol. 50, 2003, pp. 1946-1951.
- [Hook05] T. B. Hook, J. Brown, and T. Xiaowei, "Proximity effects and VLSI design", *International Conference on Integrated Circuit Design and Technology*, 9-11 May 2005, pp. 167-170.
- [Hook10] T. B. Hook, J. B. Johnson, H. Jin-Ping, A. Pond, T. Shimizu, and G. Tsutsui, "Channel Length and Threshold Voltage Dependence of Transistor Mismatch in a 32-nm HKMG Technology", *IEEE Transactions on Electron Devices*, vol. 57, 2010, pp. 2440-2447.
- [Huan08] C.-F. Huang, P. Cheng-Yi, Y. Ying-Jhe, S. Hung-Chang, C. Hung-Chih, K. Ping-Sheng, C. Huan-Lin, L. Chee-Zxaing, and L. Chee Wee, "Stress-Induced Hump Effects of p-Channel Polycrystalline Silicon Thin-Film Transistors", *IEEE Electron Device Letters*, vol. 29, 2008, pp. 1332-1335.
- [Huff80] H. R. Huff, R. D. Halvorson, T. L. Chiu, D. Guterman, Experimental observations on conduction through polysilicon oxide, *J. Electrochem. Soc.*, Vol 127, 1980, p.2482.
- [Ielm02] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and A. Modelli, Modeling of anomalous SILC in flash memories based on tunneling at multiple defects, *Solid-State Electronics*, Vol. 46, 2002, pp. 1749-1756.
- [Ito82] T. Ito, T. Nakamura, and H. Ishikawa, "Advantages of Thermal Nitride and Nitroxide Gate Films in VLSI Process", *IEEE Journal of Solid-State Circuits*, vol. 17, 1982, pp. 128-132.

- [Keye75] R. W. Keyes, "The Impact of Randomness in the Distribution of Impurity Atoms on FET Threshold," J. of App. Phys, 1975, pp. 251-259.
- [Keye75] R. W. Keyes, "Effect of randomness in the distribution of impurity ions on FET thresholds in integrated electronics", IEEE Journal of Solid-State Circuits, vol. 10, 1975, pp. 245-247.
- [Khou05] B. Z. Khouas and Abdelhakim, "Measurement of Delay Mismatch Due to Process Variation by Means of Modified Ring Oscillator," in ISCAS, 2005, pp. 5246-5249.
- [Kim00] J. Kim, K. Taewoo, P. Jaebeom, K. Woojin, H. Byungseop, and Y. Gyuhan, "A shallow trench isolation using nitric oxide (NO)-annealed wall oxide to suppress inverse narrow width effect", IEEE Electron Device Letters, vol. 21, 2000, pp. 575-577.
- [King05] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits", IEEE Journal of Solid-State Circuits, vol. 40, 2005, pp. 1212-1224.
- [King07] P. R. Kinget, "Device Mismatch: An Analog Design Perspective", IEEE International Symposium on Circuits and Systems, 27-30 May 2007, pp. 1245-1248.
- [King96] P. Kinget and M. Steyaert, "Impact of transistor mismatch on the speed-accuracy-power trade-off of analog CMOS circuits", IEEE Custom Integrated Circuits Conference, 5-8 May 1996, pp. 333-336.
- [Kma01] S. KUMAR, E.L. RUSSEL, "Model to predict reliability of ONO non-volatile memory", Integrated Reliability Workshop Final Report, IEEE International, 2001, pp.34-40.
- [Kris96] K. S. Krisch, M. L. Green, F. H. Baumann, D. Brasen, L. C. Feldman, and L. Manchanda, "Thickness dependence of boron penetration through O<sub>2</sub> and N<sub>2</sub>O-grown gate oxides and its impact on threshold voltage variation", IEEE Transactions on Electron Devices, vol. 43, 1996, pp. 982-990.
- [Lafr06] L.-P. Lafrance and Y. Savaria, "A Delay Characterization Method for Integrated Devices," in MWSCAS, 2006, pp. 540-544.
- [Laks86] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterisation and modeling of mismatch in MOS transistors for precision analog design", IEEE Journal of Solid-State Circuits, vol. 21, 1986, pp. 1057-1066.
- [Lieb01] L. W. Liebmann, et al., "TCAD Development for Lithography Resolution Enhancement," IBM J. of Res. and Dev, vol. 45, 2001, pp. 651-665.
- [Lin05] J. C. Lin, S. Y. Chen, H. W. Chen, H. C. Lin, Z. W. Jhou, S. Chou, J. Ko, T. F. Lei, and H. S. Haung, "Matching variation after HCI stress in advanced CMOS technology for analog applications", IEEE International Integrated Reliability Workshop, 17-20 Oct. 2005, pp. 107-110.
- [Lope04Th] L. Lopez, PhD Thesis, "Modelisation et caracterisation de cellules memoires eDRAM", University of Provence, FRANCE, 2004.
- [Mach08] A. Machouat, "Développement et application d'une méthode d'analyse de défaillances fonctionnelles et contribution à l'amélioration de l'utilisation des techniques optiques statiques et dynamiques," Université de Bordeaux 1, Thèse, 3709, 2008
- [Mase82] J. Maserjian et N. Zamani, "Observation of positively charged state ge-neration near the Si/SiO<sub>2</sub> interface during Fowler-Nordheimtunne-ling". Journal Vac. Sci. Tech., Vol. 20, 1982, p. 743.
- [Math08] G. Math, C. Benard, J. L. Ogier, and D. Goguenheim, "Geometry effects on the NBTI degradation of PMOS transistors", IEEE International Integrated Reliability Workshop, 12-16 Oct. 2008, pp. 60-63.
- [McCr81] J. L. McCreary, "Matching properties, and voltage and temperature dependence of MOS capacitors", IEEE Journal of Solid-State Circuits, vol. 16, 1981, pp. 608-616.
- [McGi04] J. Mc Ginley, O. Noblanc, C. Julien, S. Parihar, K. Rochereau, R. Difrenza, and P. Llinares, "Impact of pocket implant on MOSFET mismatch for advanced CMOS technology", IEEE International Conference on Microelectronic Test

- Structures, 22-25 March 2004, pp. 123-126.
- [Mezz09] C. M. Mezzomo, M. Marin, C. Leyris, and G. Ghibaudo, "Mismatch Measure Improvement Using Kelvin Test Structures in Transistor Pair Configuration in Sub-Hundred Nanometer MOSFET Technology", IEEE International Conference on Microelectronic Test Structures, March 30 2009-April 2 2009, pp. 62-67.
- [Mezz10b] C. M. Mezzomo, A. Bajolet, A. Cathignol, E. Josse, and G. Ghibaudo, "Modeling local electrical fluctuations in 45 nm heavily pocket-implanted bulk MOSFET", Solid-State Electronics, vol. 54, pp. 1359-1366.
- [Mezz11] C. M. Mezzomo, A. Bajolet, A. Cathignol, R. Di Frenza, and G. Ghibaudo, "Characterization and Modeling of Transistor Variability in Advanced CMOS Technologies", IEEE Transactions on Electron Devices, vol. 58, pp. 2235-2248.
- [Mich92a] C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits", IEEE Journal of Solid-State Circuits, vol. 27, 1992, pp. 154-166.
- [Mich93] C. Michael, C. Abel, and C. S. Teng, "A flexible statistical model for CAD of submicrometer analog CMOS integrated circuits", International Conference on Computer-Aided Design, 7-11 Nov 1993, pp. 330-333.
- [Mich96] C. Michael, S. Hua, M. Ismail, A. Kankunnen, and M. Valtonen, "Statistical techniques for the computer-aided optimization of analog integrated circuit", IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 43, 1996, pp. 410-413.
- [Mico12] Micolau G., Castellani-Coulié K., Aziza H., Portal J-M, "SITARE: a Simulation Tool for Analysis and diagnosis of Radiation Effects", IEEE Proceedings of International Latin American Test Workshop, LATW'12, accepted for publication.
- [Mizu93] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental Study Of Threshold Voltage Fluctuations Using An 8k MOSFET's Array", Symposium on VLSI Technology, 1993, pp. 41-42.
- [Moaz92] R. Moazzami, Chenming Hu, Stress-induced current in thin silicon dioxide films, Proc. Of Electron Devices Meeting, Technical Digest., 1992, pp. 139-142.
- [Mode99] A. Modelli, Reliability of thin dielectric for nonvolatile applications, Microelectron. Eng., vol.48, 1999, pp. 403-410.
- [Mons02Th] F. Monsieur, PhD Thesis, "Etude des mecanismes de degradation lors du claquage des oxydes de grille ultra minces application a la fiabilité des technologies CMOS sub-0.12µm", INP Grenoble, FRANCE, 2002.
- [Moor65] Moore GE, Cramming more components onto integrated circuits, Electronics Magazine, 1965.
- [Mori84] S. Mori et al, "Poly-oxyde/nitride/oxide structure for highly reliable EPROM cells". VLSI, 1984.
- [Mori91] S. Mori et al, "ONO inter-poly dielectric scaling for nonvolatile memory application". IEEE Trans. on Electron Devices, Vol. 38, 1991, pp. 386–390.
- [Mori96] S. Mori et al, "Thickness scaling limitation factors of ONO interpoly die-lectric for nonvolatile memory devices". IEEE Trans. on Electron Devices, Vol. 43, 1996, pp. 47–53.
- [Moro06] V. Moroz, L. Smith, X. W. Lin, D. Pramanik, and G. Rollins, "Stress-Aware Design Methodology," in Intl. Symp. Quality Elec. Design, 2006
- [Nish96] M. Nishigohri, K. Ishimaru, M. Takahashi, Y. Unno, Y. Okayama, F. Matsuoka, and M. Kinugawa, "Anomalous hot-carrier induced degradation in very narrow channel nMOSFETs with STI structure", International Electron Devices Meeting, 8-11 Dec. 1996, pp. 881-884.
- [Ogra05] U. Ogras and R. Marculescu, Energy - and performance - driven NoC communication architecture synthesis using a decomposition approach, In Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE), 352–357, 2005.
- [Okad94] K. Okada, S. Kawasaki et Y. Hirofuji, "New experimental findings on stress induced leakage currents of ultrathin silicon dioxides". Int.Conf. on Solid State Devices and Materials, 1994.

- [Oldi00] P. Oldiges, et al., "Modeling Line Edge Roughness Effects in sub 100 Nanometer Gate Length Devices," in SISPAD, 2000, pp. 131-134.
- [Oliv88] P. Olivo et al, "High-Field-induced degradation in ultra-thin SiO<sub>2</sub> films", IEEE Trans. Electron Dev., Vol. 35, 1988, pp. 2258–2267.
- [P1500] IEEE P1500 Standard for Embedded Core Test: <http://grouper.ieee.org/groups/1500>
- [Park08] B.-C. Park, L. Sung-Young, C. Dong-Ryul, B. Kee-In, K. Sung-Jun, Y. Sang-Bae, and J. Eun-Seung, "A Fermi Level Controlled High Voltage Transistor preventing subthreshold hump", IEEE International Conference on Solid-State and Integrated Circuit Technology, 20-23 Oct. 2008, pp. 172-175.
- [Pelg89] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors", IEEE Journal of Solid-State Circuits, vol. 24, 1989, pp. 1433-1439.
- [Pelg98] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications", International Electron Devices Meeting, 6-9 Dec 1998, pp. 915-918.
- [Putr08] A. T. Putra, T. Tsunomura, A. Nishida, S. Kamohara, K. Takeuchi, and T. Hiramoto, "Impact of fixed charge at MOSFETs SiO<sub>2</sub>/Si interface on V<sub>th</sub> variation", International Conference on Simulation of Semiconductor Processes and Devices, 9-11 Sept. 2008, pp. 25-28.
- [Regn06] Regnier, J.M. Portal, H.Aziza, EEPROM Compact Model with SILC Simulation Capability, Proc. Of the 7th annual Non-Volatile Memory Technological Symposium, 2006, pp. 26-30
- [Regn07Th] A. Regnier, Thèse de doctorat, Concept, modélisation et intégration d'architectures alternatives de mémoires non-volatiles à grille flottante, 2007.
- [Rios02] R. Rios, S. Wei-Kai, A. Shah, S. Mudanai, P. Packan, T. Sandford, and K. Mistry, "A three-transistor threshold voltage model for halo processes", International Electron Devices Meeting, 2002, pp. 113-116.
- [Rosen97] E. Rosenbaum, L.F.Register, Mechanism of stress-induced leakage current in MOS capacitors, IEEE Transactions on Electron Devices, Volume 44, Issue 2, 1997, pp. 317-323.
- [Roy06] G. Roy, R. B. Andrew, A.-L. Fikru, R. Scott, and A. Asen, "Simulation Study of Individual and Combined Sources of Intrinsic Parameter Fluctuations in Conventional Nano-MOSFETs", IEEE Transactions on Electron Devices, vol. 53, 2006, pp. 3063-3070.
- [Saka96] K. Sakakibara, N. Ajika, M. Hatanaka et H. Miyoshi, "A Quantitative analysis of stress induced excess current (SIEC) in SiO<sub>2</sub> films". IEEE IRPS, 1996.
- [Sall96] P. Sallagoity, M. Ada-Hanifi, M. Paoli, and M. Haond, "Analysis of width edge effects in advanced isolation schemes for deep submicron CMOS technologies", IEEE Transactions on Electron Devices, vol. 43, 1996, pp. 1900-1906.
- [Salv99] B. De. Salvo, Etude du transport électrique et de la fiabilité des isolants mémoires non volatiles à grille flottante, Thèse, Institut National Polytechnique de Grenoble, 1999.
- [Schm99] J. Schmitz, H. P. Tuinhout, A. H. Montree, Y. V. Ponomarev, P. A. Stolk, and P. H. Woerlee, "Gate polysilicon optimization for deep-submicron MOSFETs", European Solid-State Device Research Conference, 13-15 Sept. 1999, pp. 156-159,
- [Schr98] D. K. Schroder, "Semiconductor material and device characterization", 2nd ed., Wiley-interscience, 1998.
- [Schw05] S. Schwantes, J. Fuerthaler, T. Stephan, M. Graf, V. Dudek, T. Barry, G. Miller, and J. Shen, "Characterisation of a new hump-free device structure for smart power and embedded memory technologies", Microelectronic Engineering, vol. 81, 2005, pp. 132-139.
- [Seng05] S. Sengupta, L. Carastro, and P. E. Allen, "Design considerations in bandgap references over process variations", IEEE International Symposium on Circuits and Systems, 23-26 May 2005, pp. 3869-3872.

- [Seo04] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, Design of Ion-Implanted MOSFET's with Very Small Dimensions, IEEE J. Solid-State Circuits SC-9, 1974, pp.256-268.
- [Sob95] SIK-HAN SOB, C. MESSICK, CHIH-CHANG CHEN, CHUAN-PU LIU, B. BAKEL, R. PRTERSON, L. SADWICK, "Fast test at wafer-level for endurance of tunnel oxide", Integrated Reliability Workshop, Final Report., International, 1995.
- [Taka99] S. Takagi, N. Yasuda, A. Toriumi, Experimental evidence of inelastic tunneling in stress-induced leakage current, IEEE Trans. Electron Devices, vol.46, 1999, pp. 335-341.
- [Team08] D. Team, "Formation D0," ST Rousset support de formation, 200810/10/2012
- [Thew01] R. Thewes, R. Brederlow, C. Schlunder, P. Wieczorek, B. Ankele, A. Hesener, J. Holz, S. Kessel, and W. Weber, "Evaluation of MOSFET Reliability in Analog Applications", European Solid-State Device Research Conference, 11-13 September 2001, pp. 73-80.
- [Thew94] R. Thewes, K. Goser, and W. Weber, "Characterization and model of the hot-carrier-induced offset voltage of analog CMOS differential stages", International Electron Devices Meeting, 11-14 Dec 1994, pp. 303-306.
- [Thew96] R. Thewes, K. F. Goser, and W. Weber, "Hot carrier induced degradation of CMOS current mirrors and current sources", International Electron Devices Meeting, 8-11 Dec. 1996, pp. 885-888.
- [Thew99] R. Thewes, R. Brederlow, C. Schlunder, P. Wieczorek, A. Hesener, B. Ankele, P. Klein, S. Kessel, and W. Weber, "Device reliability in analog CMOS applications", International Electron Devices Meeting, 1999, pp. 81-84.
- [Tuin97a] H. P. Tuinhout, A. H. Montree, J. Schmitz, and P. A. Stolk, "Effects of gate depletion and boron penetration on matching of deep submicron CMOS transistors", International Electron Devices Meeting, 7-10 Dec 1997, pp. 631-634.
- [Visa] [Virtual ket Interface Alliance: http://www.vsi.org](http://www.vsi.org)
- [Vitt09] E. A. Vittoz, Weak inversion for ultra low-power and very low-voltage circuits, IEEE Asian Solid-State Circuits Conference, 2009, pp. 129-132.
- [Waser07] R. Waser, M. Aono, "Nanoionics- based resistive switching memories," Nature Materials, Vol.6, 2007, p. 833.
- [Yaro92] G. Yaron, S. Prasard, M. Ebel, B. Leong, A 16K EEPROM employing new array architecture and new designed-in reliability features, IEEE J. Sol. St. Circuit, Vol SC-17, no.5, 1992, p.933.
- [Yi11] Jaeyun Yi et al, "Requirements of bipolar switching ReRAM for 1T1R type high density memory array," IEEE Proc. of International Symposium on VLSI Technology, Systems and Applications (VLSI - TSA), 2011, pp. 1-2.
- [Yim03] Y.S. Yim et al., "70nm NAND Flash Technology with 0.025µm<sup>2</sup> Cell Size for 4Gb Flash Memory", IEDM, 2003.
- [Yu98] B. Yu, J. Dong-Hyuk, L. Wen-Chin, N. Kepler, K. Tsu-Jae, and H. Chenming, "Gate engineering for deep-submicron CMOS transistors", IEEE Transactions on Electron Devices, vol. 45, 1998, pp. 1253-1262.
- [Yuan11] X. Yuan, T. Shimizu, U. Mahalingam, J. S. Brown, K. Z. Habib, D. G. Tekleab, S. Tai-Chi, S. Satadru, C. M. Olsen, L. Hyunwoo, P. Li-Hong, T. B. Hook, H. Jin-Ping, P. Jae-Eun, N. Myung-Hee, and K. Rim, "Transistor Mismatch Properties in Deep-Submicrometer CMOS Technologies", IEEE Transactions on Electron Devices, vol. 58, 2011, pp. 335-342.

