



Modélisation et caractérisation de transistors MOS appliquées à l'étude de la programmation et du vieillissement de l'oxyde tunnel des mémoires EEPROM

Juliano Razafindramora

► To cite this version:

Juliano Razafindramora. Modélisation et caractérisation de transistors MOS appliquées à l'étude de la programmation et du vieillissement de l'oxyde tunnel des mémoires EEPROM. Micro et nanotechnologies/Microélectronique. Université de Provence - Aix-Marseille I, 2004. Français. NNT : . tel-00009378

HAL Id: tel-00009378

<https://theses.hal.science/tel-00009378>

Submitted on 3 Jun 2005

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.

Année : 2004

THESE

présentée devant

L'UNIVERSITÉ DE PROVENCE (AIX-MARSEILLE I)

par

Juliano Brahim RAZAFINDRAMORA

Ingénieur ENSPM

pour obtenir

LE GRADE DE DOCTEUR

Ecole doctorale : Physique, Modélisation et Sciences pour l'Ingénieur

Spécialité : Physique et Modélisation des Systèmes Complexes

**Modélisation et Caractérisation de Transistors MOS
appliquées à l'étude de la programmation et du
vieillissement de l'oxyde tunnel des mémoires
EEPROM**

Directeur de thèse : Pierre CANET

Soutenue le 17/12/2004 devant la commission d'examen :

Rapporteurs :

Carole PLOSSU
Mireille COMMANDRÉ

Professeur
Professeur

INSA Lyon
EGIM Marseille

Examineurs :

Yves DANTO
Rachid BOUCHAKOUR
Pierre CANET
Frédéric LALANDE

Professeur
Professeur
Maître de Conférences
Professeur

Université de Bordeaux I
Université de Provence
Université Paul Cézanne
Université Paul Cézanne

Invité :

Philippe BOIVIN

Ingénieur

STMicroelectronics

À mon frère et à mes sœurs

À ma mère

À la famille Vaysse

À Fanny

"Ny hazo no vanon-ko lakana, ny tany naniriany tsara"

*Ohabolana malagasy
(Proverbe malgache)*

Remerciements

Cette thèse a été effectuée au Laboratoire Matériaux et Microélectronique de Provence (L2MP), dirigé par Monsieur Michel Lannoo, Directeur de Recherches au CNRS, dans l'équipe Microélectronique, dirigée par Monsieur Rachid Bouchakour, Professeur à l'Université de Provence. Je les remercie de m'avoir accueilli.

Toute ma gratitude va à mon directeur de thèse, Pierre Canet, Maître de Conférences à l'Université Paul Cézanne, dont l'encadrement très professionnel, la présence et les encouragements constants m'ont permis de mener cette thèse à son terme. Je voudrais qu'il trouve ici toute l'expression de ma reconnaissance.

Je remercie Monsieur Yves Danto, Professeur à l'Université de Bordeaux I, d'avoir bien voulu présider le jury chargé de juger ma thèse. Je remercie également Mesdames Carole Plossu, Professeur à l'Institut National des Sciences Appliquées de Lyon (INSA-Lyon), et Mireille Commandré, Professeur à l'Ecole Généraliste d'Ingénieurs de Marseille (EGIM), de s'être intéressées à ce travail en ayant accepté d'en être les rapporteurs. Qu'elles trouvent ici toute l'expression de ma reconnaissance.

Je remercie la Société ST-Microelectronics d'avoir fourni les plaquettes de silicium sans lesquelles cette thèse n'aurait pas été possible. Je voudrais remercier en particulier Monsieur Philippe Boivin, Ingénieur sur le site de Rousset, de m'avoir partagé sa grande expérience dans le domaine des mémoires non-volatiles à travers les nombreuses réunions et discussions enrichissantes que nous avons eues tout au long de ces trois années.

Je voudrais également remercier Monsieur Frédéric Lalande, Professeur à l'Université Paul Cézanne, qui m'a beaucoup aidé tout au long de cette thèse, et ce tant sur le plan expérimental que théorique. Son approche analytique et son esprit de synthèse ont été très précieux pendant la rédaction de ce manuscrit. Je lui adresse ici mes remerciements les plus vifs.

J'ai effectué mes premiers pas dans le monde de l'enseignement au département Génie Electrique et Informatique Industrielle (GEII) de l'IUT de Marseille Saint-Jérôme et à l'EGIM. Je voudrais remercier mes collègues Olivier Bosch, de l'IUT GEII, ainsi que Caroline Fossati et Damien Barakel, de l'EGIM, de m'avoir permis de travailler dans une ambiance sereine.

Cette thèse s'inscrit dans la suite du travail qu'a effectué Nadia Harabech sur la modélisation d'une cellule mémoire EEPROM. Je la remercie vivement pour le temps qu'elle a consacré pour m'expliquer le fonctionnement du modèle et sa grande sympathie.

Je voudrais également remercier les personnes suivantes pour toute leur aide pendant ces années de DEA et de thèse : Philippe Pannier, Cyril Vidal, Wenceslas Rahajandraibe, Isabelle Jimenez, Evelynne Santacroce, Cathy Paitel-Haldenwang, Claude Arnold et Sandrine Fischer. Leur disponibilité m'a permis de surmonter les mille et un soucis que rencontre un étudiant en thèse sur le plan administratif, informatique, etc.

Je voudrais remercier tous les thésards, anciens thésards et étudiants en DEA que j'ai connus pendant ces trois années pour leur jovialité et leur bonne camaraderie : Valéry, Hassen, Henia, Sandrine B., Laurent, Bertrand, Vincent, Matthieu F., Nicolas, Arnaud, Lionel, Romain, Anna, Marc, Sophie, Eric, Karim, Samir S., Aurélie, Daniele, Matthieu E., Yannick, Amir, Jérémy, Manon, Samir B., Samir M., Mahat, Bruno et tous ceux que j'ai maladroitement oubliés. Je souhaite bon courage à ceux qui soutiendront leur thèse bientôt et un parcours professionnel brillant à tous ceux qui, comme moi, viennent de soutenir leur thèse.

Mes années d'études supérieures en France ont été possibles et se sont déroulées sereinement grâce en grande partie à la famille Vaysse : Fanny, David, Claude, Dany, Romain et Hélène, voyez en ce manuscrit l'un des fruits de l'amitié qui nous unit depuis plus de dix ans. Merci du fond du cœur !

Merci également à Isabelle Ramos, Liliane Ralefy, Andronic Razafindrabe et Nadia Ratsiambakaina pour leur soutien, leur amitié et tous les innombrables services qu'ils m'ont rendus.

Merci enfin à ma famille et à Fanny pour leurs encouragements quotidiens et leur grande compréhension dans les moments les plus difficiles.

TABLE DES MATIERES-----	1
LISTE DES SYMBOLES ET DES ABREVIATIONS-----	5
INTRODUCTION GENERALE-----	11

CHAPITRE I : ETAT DE L'ART DE LA TECHNOLOGIE DES MEMOIRES NON-VOLATILES.

I.1 LES MEMOIRES A SEMI-CONDUCTEURS.....	21
I.1.1 Mémoires volatiles.....	21
I.1.1.1 SRAM.....	21
I.1.1.2 DRAM.....	22
I.1.2 Mémoires non volatiles.....	22
I.1.2.1 NOVRAM	23
I.1.2.2 ROM.....	23
I.1.2.3 EPROM.....	24
I.1.2.4 EEPROM.....	24
I.1.2.5 Flash EEPROM.....	25
I.1.3 Mémoires émergentes.....	25
I.1.3.1 MRAM	25
I.1.3.2 Les mémoires à changement de phase.....	26
I.1.4 Marché des mémoires à semi-conducteurs.....	27
I.2 TECHNOLOGIE DES MEMOIRES NON VOLATILES.....	28
I.2.1 Technologies à grille flottante.....	29
I.2.2 Technologies à piégeage de charges.....	31
I.2.2.1 Technologies MNOS et SNOS	31
I.2.2.2 Technologie SONOS.....	32
I.2.3 Technologie ferroélectrique	33
I.3 MEMOIRE EEPROM.....	33
I.3.1 Structure de la cellule mémoire EEPROM	33
I.3.2 Principe de fonctionnement	34
I.3.3 Architecture des matrices mémoires EEPROM.....	35
I.4 FIABILITE DES MNV A GRILLE FLOTTANTE.....	38
I.4.1 Endurance	39
I.4.2 Rétention	40
I.4.2.1 Mécanismes intrinsèques de perte de charge	41
I.4.2.2 Mécanismes extrinsèques de perte de charge.....	42
I.4.2.3 Test en rétention	42
I.4.3 Modèles prédictifs de rétention de données	43
I.4.3.1 Modèle en 1/T	43

I.4.3.2 Modèle en T	44
CONCLUSION.....	44
Références bibliographiques du chapitre I	45

CHAPITRE II : MODELISATION DE LA CELLULE MEMOIRE EEPROM.

II.1 EFFET TUNNEL DANS LES OXYDES MINCES.....	53
II.1.1 Transparence tunnel.....	53
II.1.2 Effet tunnel dans les structures MOS et SOS.....	55
II.1.3 Barrière triangulaire : mécanisme d'injection FN	56
II.2 MODELISATION DE LA CELLULE EEPROM	58
II.2.1 Modélisation du transistor d'état	58
II.2.1.1 Modèle statique du courant de drain.....	58
II.2.1.2 Modèle dynamique.....	60
II.2.2 Modélisation de la capacité de la zone tunnel	62
II.2.2.1 Capacité constante	62
II.2.2.2 Prise en compte de la capacité de substrat.....	62
II.2.2.3 Prise en compte de la désertion de grille flottante.....	64
II.2.2.4 Comparatif.....	67
II.2.3 Modélisation du transistor de sélection	68
II.2.4 Bilan des paramètres du modèle	69
II.3 EXTRACTION DES PARAMETRES.....	71
II.3.1 Méthodologie.....	71
II.3.2 Extraction de α et β	73
II.3.2.1 Méthode générale	73
II.3.2.2 Approche de Croci et al.	74
II.3.2.3 Analyse de Chiou et al.....	75
II.3.3 Mesures $C_{tun}(V_{gb})$ et $I_{FN}(V_{gb})$	77
II.4 SIMULATIONS	79
II.4.1 Implantation du modèle dans ELDO	79
II.4.2 Simulations statiques	79
II.4.3 Simulations dynamiques	82
II.5 BANC DE MESURE.....	86
II.5.1 Caractérisation d'une structure Si-Poly/SiO ₂ /Si-N+	86
II.5.2 Ecriture, effacement et lecture de cellules EEPROM	87
II.5.3 Mesure de V_{th} et validation du modèle	88
Références bibliographiques du chapitre II.....	90

CHAPITRE III : OPTIMISATION DES SIGNAUX ET PROGRAMMATION RAPIDE.

III.1 INFLUENCE DU TRANSISTOR DE SELECTION.....	96
III.2 OPTIMISATION DES SIGNAUX DE PROGRAMMATION.....	97
III.2.1 Démarche phénoménologique	98
III.2.2 Résultats de simulations avec un signal optimisé	105
III.2.3 Méthodologie de conception du signal optimisé	107
III.3 PROGRAMMATION RAPIDE.....	108
III.3.1 Principe de la programmation rapide.....	109
III.3.2 Mesure de la fenêtre de programmation	111
III.3.3 Résultats en endurance	112
III.3.3.1 Fermeture de la fenêtre de programmation.....	112
III.3.3.2 Courbe caractéristique.....	112
III.4 PROGRAMMATIONS NON-STANDARDS.....	113
III.4.1 Programmation répartie : impact sur la fiabilité.....	113
III.4.1.1 Programmmations classique et répartie	113
III.4.1.2 Simulations	115
III.4.1.3 Mesures	117
III.4.1.4 Test en endurance et interprétation	118
III.4.2 Programmation par trains d'impulsions	119
Références bibliographiques du chapitre III	125

CHAPITRE IV : MODELISATION DU VIEILLISSEMENT DE L'OXYDE TUNNEL.

IV.1 ETUDE BIBLIOGRAPHIQUE.....	131
IV.1.1 Problématique	131
IV.1.1.1 De la difficulté de prévoir l'endurance d'une MNV.....	131
IV.1.1.2 Définitions	131
IV.1.2 Fiabilité des diélectriques : les différents types de contraintes électriques.....	131
IV.1.2.1 Stress à tension constante	132
IV.1.2.2 Stress à courant constant.....	132
IV.1.2.3 Rampe de tension (RVS ou Ramped Voltage Stress)	134
IV.1.2.4 Rampe de courant exponentielle.....	135
IV.1.3 Etude du piégeage de charges dans les oxydes	136
IV.1.3.1 Technique de Di Maria.....	136
IV.1.3.2 Méthode de R. Kies et al.	137
IV.1.3.3 Loi en puissance de Papadas et al.	139
IV.1.3.4 Limite de la loi en puissance par Scarpa et al.	139

IV.2 METHODOLOGIE DE L'ETUDE	140
IV.2.1 Introduction.....	140
IV.2.2 Détermination des tensions appliquées.....	141
IV.2.3 Extraction des paramètres	142
IV.2.4 Simulation de la fenêtre de programmation	144
IV.2.5 Comparaison de la dégradation en effacement et en écriture.....	145
IV.3 ETUDE DE LA MESURE $I_{FN}(V_{GB})$	146
IV.3.1 Mise en évidence de la dégradation due à la mesure $I_{FN}(V_{gb})$	146
IV.3.2 Méthode de mesure « 2 points en transitoire ».....	147
IV.3.3 Nouvelle estimation de l'évolution de α et β	150
Références bibliographiques du chapitre IV.....	152
CONCLUSION GENERALE-----	157
VALORISATION SCIENTIFIQUE-----	160
ANNEXES-----	162

Liste des symboles et des abréviations

Symbole	Signification	Unité
Bath-tub	Baignoire (courbe en)	-
BL	Bit Line	-
CCS	Constant Current Stress	-
CD	Compact Disc	-
CMOS	Complementary Metal Oxide Semiconductor	-
CPU	Central Processing Unit	-
CVS	Constant Voltage Stress	-
DRAM	Dynamic Random Access Memory	-
DVD	Digital Versatile Disc	-
DVD-RW	ReWritable Digital Versatile Disc	-
EAROM	Electrically Alterable Read Only Memory	-
EEPROM	Electrically Erasable and Programmable Read Only Memory	-
EEPROM2	Modèle de cellule EEPROM avec une capacité tunnel de type MIM	-
EEPROM3	Modèle de cellule EEPROM avec une capacité tunnel de type MIS	-
EEPROM4	Modèle de cellule EEPROM avec une capacité tunnel de type SIS	-
EPROM	Electrically Programmable Read Only Memory	-
ERCS	Exponentially Ramped Current Stress	-
FeRAM	Ferroelectric Random Access Memory	-
FIT	Failure unIT	-
FLOTOX	FLOating gate Thin OXide	-
FN	Fowler-Nordheim	-
GL	Gate Line (Ligne de grille)	-
GMR	Giant Magneto-Resistance	-
LPCVD	Low Pressure Chemical Vapor Deposition	-
MIM	Métal Isolant Métal	-
MIS	Métal Isolant Semiconducteur	-
MNOS	Métal Nitrure Oxyde Semiconducteur	-
MOS	Métal Oxyde Semiconducteur	-
MTJ	Magnetic Tunnel Junction	-
NOVRAM	NOn Volatile Random Access Memory	-
ONO	Oxyde Nitrure Oxyde	-

Symbole	Signification	Unité
OUM	Ovonic Universal Memory	-
PROM	Programmable Read Only Memory	-
RAM	Random Access Memory	-
ROM	Read Only Memory	-
RVS	Ramped Voltage Stress	-
Sampling	Echantillonnage	-
SIA	Semiconductor Industry Association	-
SILC	Stress Induced Leakage Current	-
SIMOS	Stacked Injection Metal Oxide Semiconductor	-
Si-N+	Silicium hyper-dopé	-
SiO ₂	Dioxyde de silicium	-
Si-Poly	Polysilicium	-
SIS	Semiconducteur Isolant Semiconducteur	-
SNOS	Silicium Nitrure Oxyde Silicium	-
SONOS	Silicium Oxyde Nitrure Oxyde Silicium	-
SRAM	Static Random Access Memory	-
Stress	Contrainte électrique	-
TPFG	Textured Poly Floating Gate	-
USB	Universal Serial Bus	-
UV	Ultra-Violets (Rayonnements)	-
WKB	Wentzel Kramers Brillouin (Approximation)	-
WL	Word Line (Ligne de mot)	-
WS	Word Selection (Ligne de sélection de mot)	-
μ_0	Mobilité des électrons à champ faible	$\text{m}^2.\text{V}^{-1}.\text{s}^{-1}$
μ_n	Mobilité des électrons (notation générale)	$\text{m}^2.\text{V}^{-1}.\text{s}^{-1}$
C_{bf}	Capacité basse fréquence d'une structure SIS	F
C_d	Capacité associée au substrat Si-N+ (drain du transistor d'état)	F
C_{j0}	Capacité surfacique de la jonction PN (drain, source)/substrat	F
C_{ox}	Capacité de l'oxyde de grille du transistor d'état	F
C_{oxd}	Capacité de l'oxyde tunnel	F
C'_{ox}	Capacité surfacique de l'oxyde de grille du transistor d'état	F.m^{-2}
C'_{oxd}	Capacité surfacique de l'oxyde tunnel	F.m^{-2}
C_p	Capacité de désertion associée à la grille en polysilicium	F
C'_p	Capacité surfacique de désertion associée à la grille en polysilicium	F.m^{-2}

Symbole	Signification	Unité
C_{pp}	Capacité associée à l'isolant ONO	F
C_{scmin}	Capacité minimale associée au substrat d'une structure MOS	F
C_{tun}	Capacité de la zone tunnel	F
C_x	Somme de C_{pp} , C_{ox} et C_{tun}	F
$C_{\Psi=2\Phi_{bd}}$	Capacité de la zone tunnel en limite d'inversion forte	F
D	Dérivée de $\ln(J_{FN}/E^2)$ par rapport à $1/E$	$V.m^{-1}$
D_m	Maximum de D	$V.m^{-1}$
E_{om}	Champ électrique correspondant à D_m	$V.m^{-1}$
$E_{c,drain}$	Niveau le plus bas de la bande de conduction dans le drain	eV
$E_{c,grille}$	Niveau le plus bas de la bande de conduction dans la grille	eV
$E_{début}$	Champ électrique initial lors d'un RVS	$V.m^{-1}$
E_{eff}	Champ électrique effectif à travers un oxyde dégradé	$V.m^{-1}$
$E_{f,drain}$	Niveau de Fermi dans le drain	eV
$E_{f,grille}$	Niveau de Fermi dans la grille	eV
E_{ox}	Champ électrique à travers l'oxyde de grille du transistor d'état	$V.m^{-1}$
E_{pas}	Pas de variation du champ électrique lors d'un RVS	$V.m^{-1}$
E_{pp}	Champ électrique à travers l'isolant ONO	$V.m^{-1}$
E_{tun}	Champ électrique à travers l'oxyde tunnel	$V.m^{-1}$
$E_{v,drain}$	Niveau le plus haut de la bande de valence côté drain	eV
$E_{v,grille}$	Niveau le plus haut de la bande de valence côté grille	eV
g	Conductance associée au modèle linéaire de courant FN	$A.V^{-1}$
h	Constante de Planck réduite	J.s
I^*	Rapport de deux mesures successives de courant lors d'un CVS	-
$I_{claquage}$	Courant à travers l'isolant au claquage	A
$I_{dérive}$	Courant de dérive dans le modèle de dérive-diffusion	A
$I_{diffusion}$	Courant de diffusion dans le modèle de dérive-diffusion	A
I_{ds}	Courant drain-source d'un transistor MOS	A
I_{FN}	Courant d'injection Fowler-Nordheim	A
I_{FN0}	Courant d'injection Fowler-Nordheim associé à un signal optimisé	A
I_n	$n^{ème}$ mesure de courant lors d'un CVS	A
I_s	Courant de saturation de la diode PN (drain, source)/substrat	A
$J_{début}$	Densité de courant initiale lors d'un stress de type ERCS	$A.m^{-2}$
J_{FN}	Densité de courant Fowler-Nordheim	$A.m^{-2}$
K_e	Coefficient de couplage entre la grille flottante et la grille de contrôle	-

Symbole	Signification	Unité
K_w	Coefficient de couplage entre la grille flottante et le drain	-
L	Longueur de grille d'un transistor MOS	m
L_c	Longueur effective de canal du transistor d'état	m
M_j	Facteur de gradualité de la jonction PN (drain, source)/substrat	-
$m_{ox,BC}^*$	Masse effective des électrons dans la bande de conduction de l'oxyde	kg
$m_{Si,BC}$	Masse effective des électrons dans la bande de conduction du silicium	kg
N_g	Niveau de dopage de la grille en polysilicium	m^{-3}
n_i	Concentration intrinsèque de porteurs dans le silicium à $T=300K$	m^{-3}
N_{sub}	Niveau de dopage du substrat du transistor d'état	m^{-3}
N_{subd}	Niveau de dopage du substrat du drain en Si-N+	m^{-3}
q	Charge élémentaire	C
Q_b	Quantité de charge de désertion dans le substrat	C
Q'_b	Densité surfacique de charge de désertion dans le substrat	$C.m^{-2}$
Q_{bd}	Quantité de charges piégées dans l'oxyde au claquage	C
Q_d	Quantité de charges dans le drain en Si-N+	C
Q'_d	Densité surfacique de charge dans le drain en Si-N+	$C.m^{-2}$
Q_{fg}	Quantité de charge emmagasinée dans la grille flottante	C
Q'_{fg}	Densité surfacique de charge de grille flottante	$C.m^{-2}$
Q_{fgini}	Quantité de charge emmagasinée dans la grille flottante à $t=0$	C
Q_g	Quantité de charge dans la grille d'une structure MOS	C
Q'_g	Densité surfacique de charge dans la grille d'une structure MOS	$C.m^{-2}$
Q_{inj}	Densité surfacique de charges injectées dans l'oxyde lors d'un stress	$C.m^{-2}$
Q_{ox}	Densité surfacique de charges piégées dans l'oxyde lors d'un stress	$C.m^{-2}$
Q_n	Quantité de charge d'inversion dans le canal	C
Q'_n	Densité surfacique de charge d'inversion dans le canal	$C.m^{-2}$
Q_{sc}	Quantité totale de charge dans le substrat semi-conducteur	C
Q'_{sc}	Densité surfacique de charge dans le substrat semi-conducteur	$C.m^{-2}$
r	Dérivée temporelle de la rampe du signal de programmation	$V.s^{-1}$
R_d	Résistance d'accès côté drain	Ω
R_s	Résistance d'accès côté source	Ω
S_{ox}	Surface de la capacité associée à l'oxyde de grille	m^2
S_{pp}	Surface de la capacité associée à l'isolant ONO	m^2
S_{tun}	Surface de la capacité associée à la zone tunnel	m^2
t_F	Temps de descente du signal de programmation	s

Symbole	Signification	Unité
t_{ox}	Epaisseur de l'oxyde de grille	m
t_{pas}	Pas temporel lors d'un stress de type RVS ou ERCS	s
$t_{plateau}$	Durée du plateau du signal de programmation	s
t_{pp}	Epaisseur de l'isolant ONO	m
t_R	Temps de montée du signal de programmation	s
t_{Rampe}	Durée de la rampe du signal de programmation	s
t_{tun}	Epaisseur de l'oxyde tunnel	m
U_t	Potentiel électrothermique	V
V^*	Rapport de deux mesures successives de tension lors d'un CCS	-
V_b	Potentiel de substrat	V
V_d	Potentiel de drain	V
V_{fb}	Tension de bandes plates du transistor d'état	V
V_{fbd}	Tension de bandes plates d'une structure Si-Poly/SiO ₂ /Si-N+	V
V_{fg}	Potentiel de grille flottante	V
V_{fg0}	Potentiel de grille flottante correspondant à un signal optimisé	V
V_{gb}	Tension grille/substrat	V
V_{gc}	Potentiel de grille de contrôle	V
V_{gcmax}	Potentiel de grille de contrôle maximal pour un signal optimisé	V
V_{gcmin}	Potentiel de grille de contrôle minimal pour un signal optimisé	V
V_{gs}	Tension grille/source	V
V_j	Barrière de potentiel de la jonction PN (drain, source)/substrat	V
V_n	$n^{\text{ème}}$ mesure de tension lors d'un stress de type CCS	V
V_{prog}	Tension de programmation	V
$V_{prog,eff}$	Potentiel régnant effectivement sur le drain du transistor d'état	V
$V_{ref,basse}$	Tension de référence basse lors de la lecture d'une cellule EEPROM	V
$V_{ref,haute}$	Tension de référence haute lors de la lecture d'une cellule EEPROM	V
V_s	Tension de seuil dans le modèle linéaire de courant FN	V
V_{sat}	Vitesse de saturation des électrons	m.s ⁻¹
V_{sb}	Tension source/substrat du transistor d'état	V
V_{sel}	Tension de la grille de sélection	V
V_{th}	Tension de seuil (notation générale)	V
V_{th0}	Tension de seuil pour une charge de grille flottante nulle	V
$V_{thécrit}$	Tension de seuil à l'état écrit	V
$V_{theffacé}$	Tension de seuil à l'état effacé	V

Symbole	Signification	Unité
V_y	Potentiel le long du canal d'un transistor MOS à effet de champ	V
W	Largeur de la grille d'un transistor MOS	m
$W_{\max}$	Largeur maximale de la zone de désertion dans le substrat	m
x_c	Barycentre des charges piégées dans l'oxyde	m
X_d	Largeur de la zone de désertion dans la grille en polysilicium	m
z	Exposant de Q_{inj} dans la loi de Papadas <i>et al.</i>	-
$\alpha_{écriture}$	Coefficient FN pré-exponentiel en écriture	$A.V^{-2}$
$\alpha_{effacement}$	Coefficient FN pré-exponentiel en effacement	$A.V^{-2}$
$\beta_{écriture}$	Coefficient FN exponentiel en écriture	$V.m^{-1}$
$\beta_{effacement}$	Coefficient FN exponentiel en effacement	$V.m^{-1}$
γ	Facteur de substrat du transistor d'état	$V^{1/2}$
γ_{oxd}	Facteur de substrat associé à la zone tunnel	$V^{1/2}$
ΔV_{g^-}	Décalage en tensions négatives dans la méthode de Di Maria	V
ΔV_{g^+}	Décalage en tensions positives dans la méthode de Di Maria	V
ΔV_{th}	Largeur de la fenêtre de programmation	V
$\Delta V_{th,initial}$	Largeur initiale de la fenêtre de programmation	V
ϵ_0	Permittivité diélectrique du vide	$F.m^{-1}$
ϵ_{Si}	Permittivité diélectrique relative du silicium	-
ϵ_{SiO_2}	Permittivité diélectrique relative du dioxyde de silicium	-
θ	Facteur de réduction de la mobilité	V^{-1}
λ	Barycentre réduit des charges piégées dans l'oxyde	-
Ξ	Fonction d'onde associée à l'électron	-
Ψ	Potentiel de surface à l'interface $SiO_2/Si-N^+$	V
Ψ_p	Potentiel de surface à l'interface $Si-Poly/SiO_2$	V
Ψ_s	Potentiel de surface dans le canal du transistor d'état	V
Ψ_{ss}	Potentiel de surface au niveau de la source du transistor d'état	V
Ψ_{sd}	Potentiel de surface au niveau du drain du transistor d'état	V
Φ_0	Hauteur de barrière à l'interface cathode/oxyde	eV
Φ_b	Potentiel de substrat associé au transistor d'état	V
Φ_{bd}	Potentiel de substrat associé à la zone tunnel	V

Introduction générale

Un peu d'histoire...

La miniaturisation des composants électroniques a atteint un tel niveau de complexité, de nos jours, qu'il devient de plus en plus difficile de repousser les limites de leurs performances, qui semblent buter désormais sur des obstacles fondamentaux imposés par les lois de la physique.

Notre vie quotidienne est envahie d'objets divers produits par l'industrie des semi-conducteurs : appareils photos numériques, mémoires amovibles (clés USB, cartes Compact Flash, SmartMedia...), disques durs de très grande capacité, frôlant le millier de giga-octets, disques compacts optiques réinscriptibles...

Tous ces objets qui facilitent et agrémentent notre vie quotidienne n'auraient pas vu le jour sans la recherche fondamentale entreprise en physique d'une manière générale, mais plus particulièrement en physique des solides et des semi-conducteurs, après la fin de la deuxième guerre mondiale. Cette recherche fondamentale était nourrie en grande partie par la compétition militaire féroce que se livraient les Etats-Unis et l'ex-Union soviétique pendant la guerre froide. L'un des objectifs majeurs, à cette époque, était la mise au point de calculateurs performants pour calculer la trajectoire des missiles intercontinentaux, composants stratégiques de l'équilibre de la terreur, mais aussi des vaisseaux spatiaux, telles les capsules Apollo qui ont amené les premiers hommes sur la lune. De tels calculateurs nécessitaient des composants fiables, susceptibles de fonctionner pendant de longues périodes sans intervention humaine.

On peut raisonnablement affirmer que trois dates marquent un tournant important dans l'histoire, récente, de l'industrie des semi-conducteurs :

i. 1947 : invention du transistor par W. Shockley, Bardeen et Brattain

Pendant la première moitié du XX^{ème} siècle, l'industrie électronique était dominée par les tubes à vide. Ces composants étaient peu pratiques : ils étaient fragiles, encombrants, peu fiables, consommaient et chauffaient beaucoup. Par ailleurs, les réseaux téléphoniques étaient constitués de relais mécaniques qui tombaient souvent en panne. Ces problèmes ont été résolus par l'invention du transistor en 1947 dans les laboratoires Bell, invention qui valut à ses trois co-auteurs le prix Nobel de physique en 1956 : les transistors étaient minuscules par rapport aux tubes à vide, étaient plus fiables, avaient une longévité plus grande, consommaient et chauffaient beaucoup moins. L'invention du transistor stimula les ingénieurs pour inventer des circuits toujours plus complexes contenant des milliers de composants tels que les transistors, les diodes, les résistances et les condensateurs.

Malgré ce grand progrès, un obstacle de taille demeurait : ces composants devaient être reliés, individuellement, par des kilomètres de fils électriques qui rendaient l'entreprise coûteuse. Le défi consistait à trouver un procédé efficace et peu onéreux pour produire ces composants et les relier entre eux. Ce défi fut relevé, indépendamment, par Jack Kilby, ingénieur chez Texas Instruments, et Robert Noyce, co-fondateur de Fairchild Semiconductor et de Intel.

ii. 1958 : invention du circuit intégré par J. Kilby et R. Noyce

J. Kilby eut l'idée de fabriquer sur un même substrat de silicium tous les composants d'un circuit électronique donné (diodes, résistances, transistors, condensateurs...) alors qu'il travaillait sur le projet « Micro-Module », chez Texas Instruments. Le projet « Micro-Module » était censé résoudre les problèmes posés par l'interconnexion des composants, mais Kilby pensait qu'il était voué à l'échec. Parallèlement, et indépendamment, R. Noyce mettait au point le procédé « planar ». Ce procédé est basé sur une oxydation en surface afin de séparer, par une couche isolante, plusieurs niveaux d'interconnexion sur un même substrat de silicium. Le procédé « planar » a connu un grand succès industriel qui a failli coûter à J. Kilby la paternité du circuit intégré. Par la suite, le Bureau américain des brevets reconnut à Kilby la paternité de l'invention du circuit intégré monolithique tandis que l'invention du procédé « planar » a été attribuée à Noyce. Le circuit intégré, plus communément appelé puce, a été unanimement reconnu comme l'une des inventions majeures de l'humanité en raison de l'impact considérable qu'il a eu dans notre vie quotidienne. Il valut le prix de Nobel de physique à Jack Saint-Clair Kilby en 2000.

iii. 1968 : mise au point du premier microprocesseur par T. Hoff

L'invention du circuit intégré a rendu possible, entre autres, la mise au point des supercalculateurs. Pendant les années cinquante et soixante, seuls les grands organismes gouvernementaux et les grandes firmes avaient les fonds nécessaires pour s'offrir des machines avec de telles puissances de calcul. Cette situation allait changer en 1968, quand Ted Hoff, un ingénieur de la firme Intel, invente, sous la houlette de Robert Noyce et de Gordon Moore, les co-fondateurs d'Intel, le premier microprocesseur : le 4004. Il fut commercialisé par Intel en 1971. Cette invention a ouvert la voie à la production en masse des ordinateurs personnels.

Les mémoires

Par définition, une mémoire est un dispositif capable d'emmagasiner et de restituer, au bout d'une durée plus ou moins longue, avec plus ou moins d'altération, une certaine quantité d'information.

La mémoire qui nous est la plus familière est le cerveau humain. D'autres types de mémoires existent, comme les livres, les bibliothèques ou les mémoires électroniques, objet de cette thèse.

L'invention du circuit intégré et du microprocesseur a rendu possible le traitement d'une quantité sans cesse croissante de données. Ces données doivent être stockées quelque part avant, pendant, et après leur traitement.

Historiquement, la première mémoire moderne a été la carte perforée, inventée par l'Américain Hermann Hollerith en 1880 pour la saisie de données statistiques. Le succès de son invention stimula Hollerith qui, en 1896, fonda la compagnie IBM. A titre d'exemple, l'ENIAC (Electronic Numerical Integrator And Computer, ordinateur utilisé par l'armée américaine dans le cadre du projet Manhattan) utilisait des cartes perforées en tant qu'interface d'entrée/sortie.

Selon les applications visées, divers types de mémoires sont apparus sur le marché. On peut cependant les classer en trois grandes catégories : les mémoires magnétiques, les mémoires optiques et les mémoires à semi-conducteurs.

i. Les mémoires magnétiques

Deux exemples courants de mémoires magnétiques sont la disquette et le disque dur des ordinateurs. Dans ces mémoires, l'information est stockée sur une fine couche d'oxyde de fer aimantée qui recouvre un film en plastique (disquettes) ou un disque en aluminium (disques durs). L'écriture et la lecture de l'information se fait à l'aide d'un dispositif mécanique, appelé tête, constitué principalement d'un électroaimant qui peut modifier l'aimantation de la couche d'oxyde de fer (écriture) ou transformer cette aimantation, sans la modifier, en signal électrique (lecture).

ii. Les mémoires optiques

La mémoire optique la plus familière est le disque compact : l'information binaire s'y traduit par la présence, ou non, à la surface du disque, de minuscules cavités qui réfléchissent plus ou moins un faisceau laser produit par le système de lecture. Un disque compact optique de type DVD peut, typiquement, emmagasiner 4.9 giga-octets de données sur une surface de 12cm de diamètre.

iii. Les mémoires à semi-conducteurs.

Les mémoires à semi-conducteurs (cf. §I.1) sont constituées de composants électroniques tels que les diodes ou les transistors MOS. L'une des principales caractéristiques des mémoires à semi-conducteurs est l'absence de parties mécaniques. Les mémoires à semi-conducteurs comprennent essentiellement les mémoires volatiles, où l'information est perdue quand l'alimentation est coupée, et les mémoires non-volatiles, qui gardent l'information même en l'absence d'alimentation électrique.

Les mémoires EEPROM

Une classe particulière de mémoires à semi-conducteurs non-volatiles comprend les mémoires Flash EEPROM et les mémoires EEPROM (Electrically Erasable and Programmable Read Only Memory – mémoire inscriptible et effaçable électriquement). A titre d'exemple, on trouve respectivement les Flash EEPROM et les EEPROM dans la carte mémoire d'un appareil photo numérique et dans les cartes bancaires.

Dans la société de l'information où nous vivons actuellement, la tendance générale consiste à développer des produits « nomades » (téléphones mobiles, ordinateurs de poche...), capables de restituer et d'enregistrer rapidement, mais aussi d'emmagasiner une grande quantité d'information sur une surface réduite le plus longtemps possible. Par ailleurs, le caractère « nomade » de ces produits impose l'utilisation d'une source d'alimentation de faible puissance.

L'évolution de la technologie des mémoires EEPROM est dictée par ces besoins : les mémoires EEPROM doivent être plus rapides, plus denses et fonctionner sous faible tension. Cette course aux performances doit surmonter des obstacles : l'une

des principales difficultés est posée par la fiabilité de ces nouvelles mémoires. Notre travail s'inscrit dans ce contexte.

Cette thèse, divisée en quatre chapitres, apporte quelques éléments de réponse sur la façon dont on peut accroître les performances d'une mémoire EEPROM en termes d'endurance, de vitesse d'écriture et de baisse des tensions de polarisation. Elle explore également la possibilité d'établir une expression analytique de la fermeture de la fenêtre de programmation en fonction du nombre de cycles écriture/effacement.

Le premier chapitre est une étude bibliographique qui a pour but de situer les mémoires EEPROM dans la famille des mémoires à semi-conducteurs. Les mémoires volatiles (Static Random Access Memory et Dynamic Random Access Memory) sont abordées succinctement tandis que les mémoires non-volatiles sont abordées en fonction de leur technologie de fabrication. Par ailleurs, l'évolution récente du marché des mémoires à semi-conducteurs ainsi que la description des mémoires émergentes permettent d'estimer la part occupée par chaque type de mémoire et son évolution future. Enfin, les principales difficultés posées par la fiabilité des mémoires non-volatiles sont exposées dans la dernière partie de ce chapitre afin de situer nos travaux.

Le deuxième chapitre décrit le modèle de cellule EEPROM que nous avons utilisé. Il s'agit d'un modèle en feuillets basé sur le calcul du potentiel de grille flottante et du potentiel de surface dans le canal du transistor d'état. Ce modèle physique permet de s'affranchir des discontinuités introduites par les modèles de type SPICE dans les dérivées successives du courant drain/source. Ce modèle prend en compte la désertion dans la grille flottante en polysilicium. La description de la méthodologie que nous avons mise en œuvre pour extraire les paramètres du modèle y est complétée par une étude bibliographique afin de déterminer les paramètres les plus influents lors de l'extraction des paramètres Fowler-Nordheim α et β .

Généralement, on augmente la fiabilité d'une cellule mémoire en mettant en œuvre des matériaux plus performants. Une autre solution consiste à minimiser la contrainte électrique subie par la cellule avec des signaux optimisés. Le troisième chapitre explique la mise en œuvre des signaux optimisés et s'articule autour de deux objectifs : réduire la durée de programmation d'une cellule EEPROM et réduire ses tensions de programmation. Des travaux antérieurs [Canet'01] ont démontré la possibilité de programmer une cellule EEPROM en 550 μ s, ce qui représente un gain de facteur 8 par rapport à la durée de programmation habituelle de 4ms. Nous avons réduit davantage cette durée et avons quantifié les conséquences des champs élevés mis en jeu sur l'endurance de la cellule. La réduction des tensions de programmation est investiguée en répartissant la polarisation entre la grille de contrôle et le drain et à l'aide de la programmation multi-impulsionnelle.

Le dernier chapitre est consacré à l'émulation du vieillissement de l'oxyde tunnel d'une cellule EEPROM sur une capacité de grande surface équivalente à la zone tunnel. L'objectif, à terme, est d'établir une expression analytique de la fermeture de la fenêtre de programmation d'une cellule mémoire EEPROM en fonction du nombre de cycles écriture/effacement. En effet, la fermeture de la fenêtre de programmation résulte d'une dégradation de l'oxyde tunnel où des charges sont piégées. Ce piégeage de charges provoque une variation des paramètres α et β et une diminution du courant tunnel. Dans une telle étude, l'aspect critique réside dans l'extraction des paramètres Fowler-Nordheim α et β : les valeurs extraites dépendent de la durée pendant laquelle le courant Fowler-Nordheim est mesuré. Nous avons mis en œuvre différentes techniques de mesures de ce courant afin d'obtenir les valeurs pertinentes de α et β . Ce travail s'appuie sur une étude bibliographique des techniques établies comme la méthode de Di Maria [Di Maria'76], ou la loi en puissance de Papadas *et al.* [Papadas'93], pour étudier les oxydes soumis à une contrainte électrique.

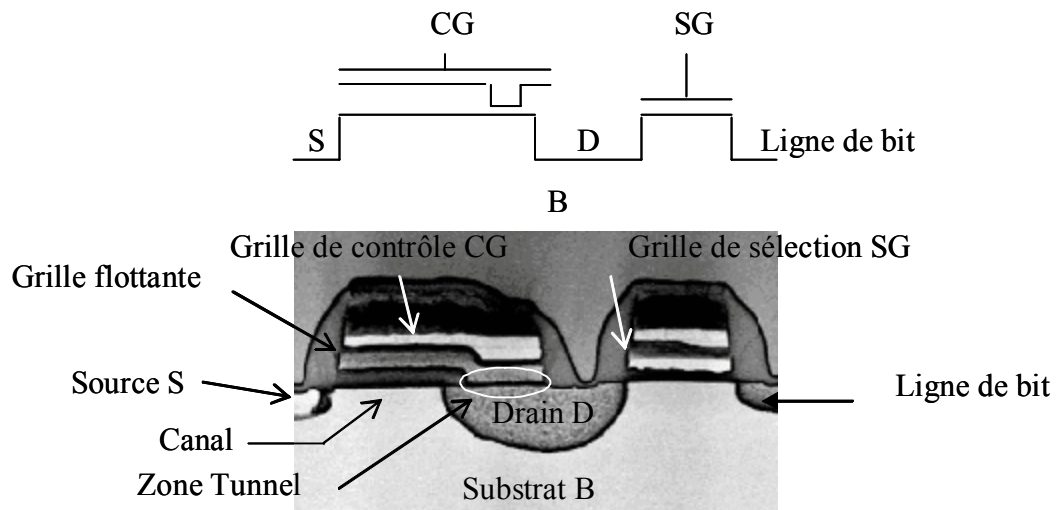
Références bibliographiques de l'introduction générale

[DiMaria'76] D.J. Di Maria, "Determination of insulator bulk trapped charge densities and centroids from photocurrent-voltage characteristics of MOS structures", Journal of Applied Physics, Vol. 47, No. 9, September 1976.

[Papadas'93] C. Papadas, G. Ghibaudo, G. Pananakakis, C. Riva, Solid-State Electronics, Vol. 36, p. 1303, 1993.

[Canet'01a] P. Canet, R. Bouchakour, N. Harabech, "EEPROM Programming Study: Time and Degradation Aspects", Proceedings of the IEEE International Symposium on Circuits And Systems (ISCAS), Sydney, Australia, p. IV.846-849, 2001.

Chapitre I : Etat de l'art de la technologie des mémoires non volatiles



Vue en coupe d'une cellule mémoire EEPROM.

I.1 LES MEMOIRES A SEMI-CONDUCTEURS.....	21
I.1.1 Mémoires volatiles.....	21
I.1.1.1 SRAM	21
I.1.1.2 DRAM.....	22
I.1.2 Mémoires non volatiles.....	22
I.1.2.1 NOVRAM	23
I.1.2.2 ROM.....	23
I.1.2.3 EPROM.....	24
I.1.2.4 EEPROM.....	24
I.1.2.5 Flash EEPROM.....	25
I.1.3 Mémoires émergentes.....	25
I.1.3.1 MRAM	25
I.1.3.2 Les mémoires à changement de phase.....	26
I.1.4 Marché des mémoires à semi-conducteurs.....	27
I.2 TECHNOLOGIE DES MEMOIRES NON VOLATILES.....	28
I.2.1 Technologies à grille flottante.....	29
I.2.2 Technologies à piégeage de charges.....	31
I.2.2.1 Technologies MNOS et SNOS	31
I.2.2.2 Technologie SONOS.....	32
I.2.3 Technologie ferroélectrique	33
I.3 MEMOIRE EEPROM.....	33
I.3.1 Structure de la cellule mémoire EEPROM	33
I.3.2 Principe de fonctionnement	34
I.3.3 Architecture des matrices mémoires EEPROM.....	35
I.4 FIABILITE DES MNV A GRILLE FLOTTANTE.....	38
I.4.1 Endurance	39
I.4.2 Rétention	40
I.4.2.1 Mécanismes intrinsèques de perte de charge	41
I.4.2.2 Mécanismes extrinsèques de perte de charge.....	42
I.4.2.3 Test en rétention.....	42
I.4.3 Modèles prédictifs de rétention de données	43
I.4.3.1 Modèle en $1/T$	43
I.4.3.2 Modèle en T	44
CONCLUSION.....	44
Références bibliographiques du chapitre I.....	45

Ce premier chapitre est dédié à une description des différentes mémoires non-volatiles.

Dans un premier temps, nous présentons une classification des mémoires à semi-conducteurs disponibles sur le marché. Les mémoires volatiles (SRAM, DRAM...) sont traitées rapidement tandis que les mémoires non-volatiles (FLASH, EEPROM...) sont décrites en détail. Nous nous attachons notamment à comparer les différentes mémoires non-volatiles sur plusieurs critères de fonctionnalités. Cette démarche s'inspire du fait que d'importants efforts sont actuellement déployés pour mettre au point une mémoire qui serait non-volatile, bas coût, effaçable et programmable rapidement, avoir une endurance et une rétention infinies, de faibles dimensions et enfin compatible avec les technologies (CMOS) existantes. Actuellement, les mémoires disponibles sur le marché remplissent tout ou partie de ces critères de manière plus ou moins satisfaisante.

Après cette comparaison, nous décrivons la technologie des différentes mémoires non-volatiles puis le fonctionnement d'une cellule EEPROM.

Enfin, nous terminons par une étude bibliographique sur la fiabilité des mémoires non-volatiles comprenant les phénomènes physiques associés, ainsi que les solutions proposées afin d'améliorer la fiabilité de ces dispositifs.

I.1 Les mémoires à Semi-conducteurs

Divers types de mémoires sont apparus depuis l'invention du premier circuit intégré monolithique par Jack Kilby chez Texas Instruments en 1958. Cette diversité (fig. I. 1) s'explique par une grande variété de besoins. En effet, chaque système requiert un type de mémoire aux performances précises (durée de rétention de l'information, tension d'alimentation, consommation, vitesse d'écriture, vitesse de lecture, dimensions). La recherche de ces performances induit un cahier de charges qui exigera un procédé de fabrication plus ou moins coûteux. Le compromis entre les performances recherchées, les limites technologiques et les contraintes économiques définira les caractéristiques principales de chaque type de mémoire.

On divise généralement les mémoires à semi-conducteurs en deux grandes classes : les mémoires volatiles appelées également RAM (Random Access Memories) et les mémoires non-volatiles.

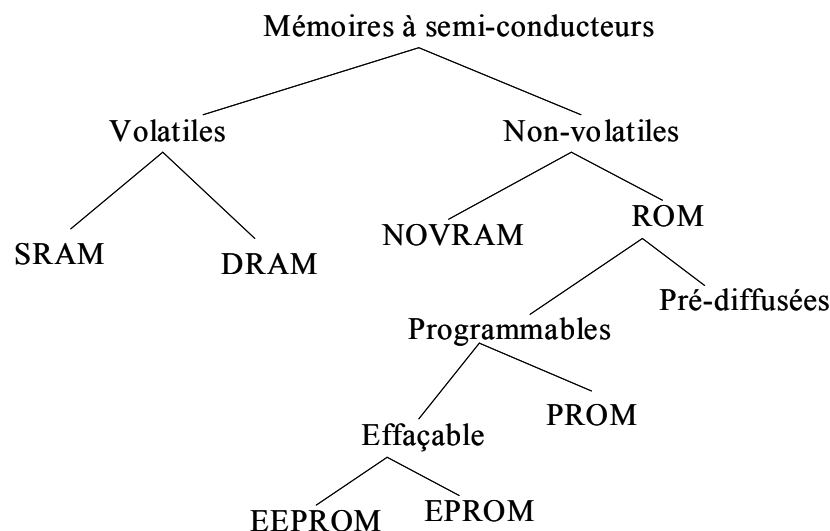


Figure I. 1 : *Classification des différentes mémoires à semi-conducteurs.*

I.1.1 Mémoires volatiles

Les mémoires volatiles, comme leur nom l'indique, perdent l'information en l'absence d'alimentation électrique. Les deux types les plus connus de mémoires volatiles sont la SRAM (Static RAM) et la DRAM (Dynamic RAM).

I.1.1.1 SRAM

La SRAM est la plus rapide de toutes les mémoires à semi-conducteurs, mais elle est aussi d'un coût plus élevé. Elle retient les données aussi longtemps que

l'alimentation électrique est maintenue et n'exige pas de rafraîchissements à intervalles réguliers. La SRAM est surtout utilisée dans les mémoires cache des ordinateurs.

1.1.1.2 DRAM

La DRAM peut être réalisée avec un seul transistor (fig. I.2 : 1-T DRAM). Son fonctionnement schématisé est le suivant : pour écrire la cellule, la ligne de bit est forcée à l'état logique « 1 » (ou « 0 ») tout en activant la ligne de mot. Ainsi, le transistor d'accès est actif et on peut charger la capacité de stockage à l'état logique désiré. Pour lire la cellule, il suffit d'activer le transistor d'accès, permettant ainsi de propager sur la ligne de bit la différence de potentiel aux bornes de la capacité de stockage. La variation de potentiel provoquée par les charges stockées dans la capacité sur la ligne de bit permet ensuite à des circuits annexes de déterminer l'état logique de la cellule.

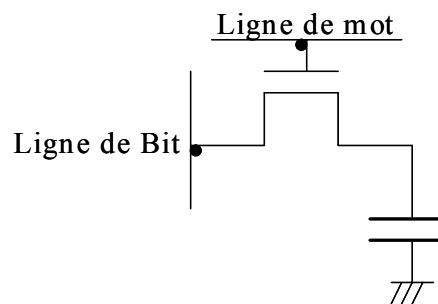


Figure I. 2 : *Cellule mémoire DRAM à un transistor.*

Cette structure lui confère un faible coût et une surface réduite, qui permet une grande densité d'intégration. L'opération de lecture d'une DRAM provoque la décharge de la capacité de stockage dans la ligne de bit. Ainsi, des circuits annexes doivent copier la valeur qui a été lue pour la réécrire ensuite dans la cellule. Le deuxième problème posé par la DRAM provient des fuites de la capacité de stockage : la cellule doit être rafraîchie régulièrement.

I.1.2 Mémoires non volatiles

A l'opposé des mémoires volatiles, les mémoires non-volatiles gardent l'information même en l'absence d'alimentation électrique : c'est leur principal intérêt.

Généralement, les mémoires non-volatiles sont considérées selon trois critères qui sont : les limitations imposées par la technologie, les besoins des utilisateurs et enfin les contraintes économiques. Dans ce paragraphe, nous allons détailler les trois grandes générations de mémoires non-volatiles qui sont apparues successivement sur le marché.

On peut classer les mémoires non-volatiles en deux catégories principales : les mémoires non-volatiles de type ROM (Read Only Memory) et les mémoires non-volatiles de type NOVRAM.

I.1.2.1 NOVRAM

La NOVRAM (NOVolatile RAM) allie dans la même structure la non volatilité des EEPROM (Electrically Erasable and Programmable Read Only Memory) et la très grande vitesse d'écriture et de lecture des SRAM [Harari'78][Drori'81]. Le fonctionnement de cette structure peut être décrit simplement à l'aide d'un exemple. Dans un ordinateur, le CPU (Central Processing Unit) lit et écrit les données dans la SRAM en fonctionnement normal. Lorsque la SRAM détecte un début de coupure d'alimentation, des circuits spéciaux copient rapidement (en quelques millisecondes) les données de la SRAM vers l'EEPROM, empêchant ainsi la perte des données. Dès que l'alimentation est restituée, les données sont recopiées dans la SRAM et le dispositif fonctionne comme s'il n'y avait eu aucune coupure d'alimentation. Toutefois, la NOVRAM offre une faible densité d'intégration en raison de la surface importante occupée par la cellule SRAM et la cellule EEPROM. La figure I.3 représente un exemple de NOVRAM.

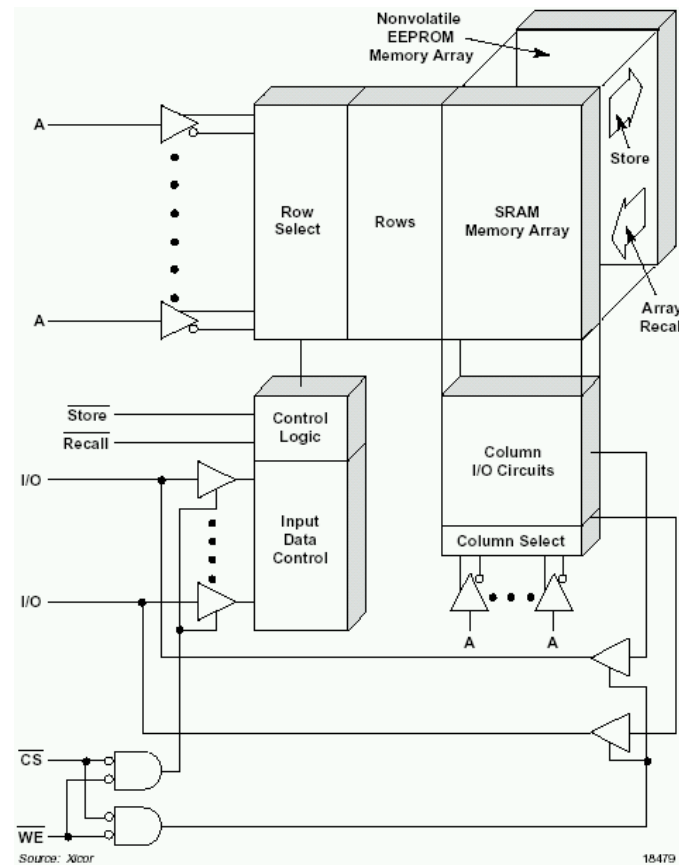


Figure I. 3 : Exemple de structure NOVRAM (Source : Xicor)

I.1.2.2 ROM

La mémoire ROM (Read Only Memory) est destinée uniquement à être lue. Les ROM elles-mêmes se subdivisent en deux catégories : les ROM pré-diffusées et les ROM programmables, ces dernières pouvant être effaçables par la suite ou non. Elle est programmée soit lors de la fabrication (activation ou non de transistors par masquage), soit par l'utilisateur avec des structures à base de

fusibles [De Graaf'96]. Elle sert par exemple de support au jeu d'instruction des microprocesseurs.

1.1.2.3 EPROM

Une mémoire EPROM (Electrically Programmable Read Only Memory) est écrite électriquement mais peut être effacée avec des rayons ultraviolets. La phase d'écriture peut durer de 100µs à 1ms. Quant à la phase d'effacement, elle nécessite plus de 20 minutes d'exposition aux rayons ultraviolets [Brown'98]. L'écriture se fait par injection d'électrons chauds à travers l'oxyde qui sépare la grille flottante du canal du transistor MOS. La figure I.4 montre une cellule EPROM constituée d'un seul transistor, ce qui permet une haute densité d'intégration, comparable à celle des DRAM [Maes'89].

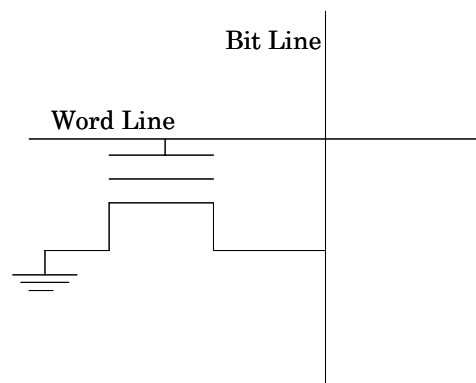


Figure I. 4 : *Représentation schématique d'une mémoire EPROM.*

1.1.2.4 EEPROM

L'inconvénient majeur de la mémoire EPROM est sa durée d'effacement (20 minutes). La solution consiste à utiliser des mémoires EEPROM (Electrically Erasable and Programmable Read Only Memory) qui sont effaçables et programmables électriquement.

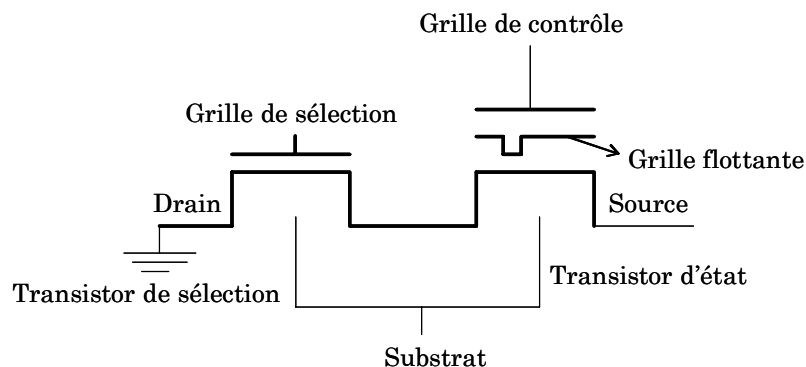


Figure I. 5 : *Cellule mémoire EEPROM montrant le transistor d'état en série avec le transistor de sélection.*

La figure I.5 présente une cellule EEPROM constituée de deux transistors MOS mis en série dont un transistor dit d'état qui comporte une grille flottante et un

transistor de sélection [Yaron'82]. Contrairement à la mémoire EPROM, constituée d'un seul transistor, la mémoire EEPROM nécessite une grande surface à cause de la présence de ces deux transistors. Les EEPROM sont donc peu compétitives en terme de densité d'intégration.

1.1.2.5 Flash EEPROM

La mémoire Flash EEPROM est née de la recherche d'une cellule EEPROM qui, en plus d'être effaçable électriquement, puisse également offrir une grande densité d'intégration. Une telle structure est réalisable en utilisant une cellule mémoire composée d'un seul transistor à grille flottante écrite à l'aide d'une injection d'électrons chauds et effacée par injection Fowler-Nordheim.

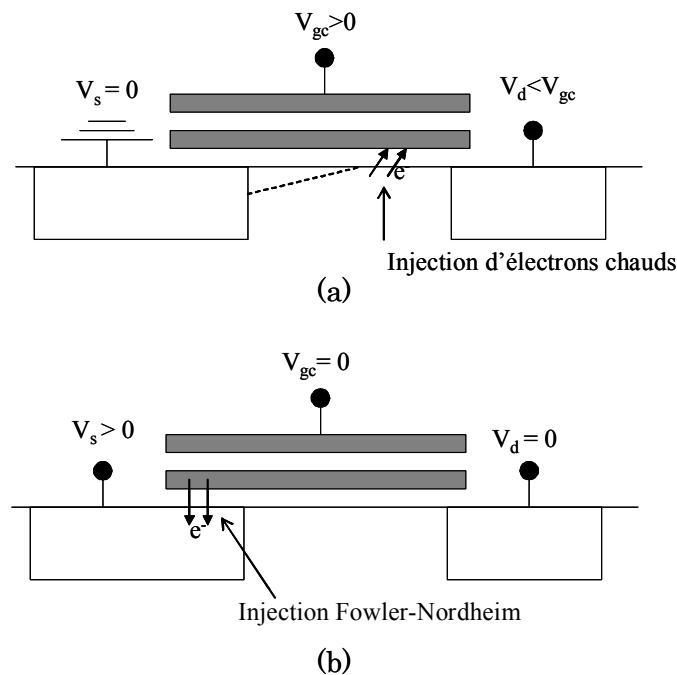


Figure I. 6 : *Cellule Flash EEPROM écrite par injection d'électrons chauds (a) et effacée par injection Fowler-Nordheim (b).*

1.1.3 Mémoires émergentes

La recherche actuelle essaie de rassembler tous les critères précédemment cités dans les mémoires du futur. Deux types de mémoires sont prometteuses dans cette direction : la MRAM (Magnetoresistive Random Access Memory) et la mémoire à changement de phase ou OUM (Ovonic Universal Memory).

1.1.3.1 MRAM

Le fonctionnement de la MRAM repose sur le changement de résistivité d'une multicouche ferromagnétique en fonction de l'aimantation relative de ses constituants. Le phénomène physique à l'origine de l'invention de la MRAM est la magnétorésistance ou TMR (Tunnel MagnetoResistance), mise en évidence

expérimentalement par Jullière en 1975 [Jullière'75]. Cependant, il a fallu attendre la découverte en 1988 de la magnétorésistance géante (Giant MagnetoResistance) par l'équipe d'Albert Fert [Baibich'88] pour que des applications à l'échelle industrielle voient le jour. En effet, la variation de résistivité peut atteindre 60% dans le cas de la GMR alors qu'elle vaut à peu près 15% pour la TMR. Le premier prototype de MRAM opérationnel a été présenté par IBM en 1999 [Parkin'99] et a démontré de hautes potentialités en exhibant un temps d'écriture/lecture inférieure à 3ns. Les multicouches ferromagnétiques étaient des jonctions tunnel magnétiques (MTJ) intégrées dans une technologie IBM CMOS 0.25 μ m.

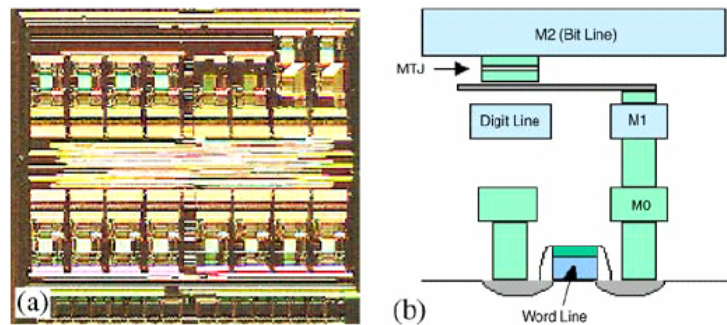


Figure I. 7 : Microphotographie d'une matrice MRAM 64kb (a) et structure verticale d'une cellule MRAM (b) [Koh'04].

La MRAM est non-volatile grâce à la présence d'une jonction tunnel magnétique (MTJ) dans chaque cellule mémoire. La MTJ la plus simple est un sandwich dans lequel deux films ferromagnétiques sont séparés par une fine couche d'isolant tunnel. La résistance de la MTJ dépend de l'orientation relative des moments magnétiques des deux couches ferromagnétiques. Si les deux moments magnétiques sont parallèles, la résistance de la structure est faible et lorsqu'ils sont antiparallèles, cette résistance est grande. Les principaux avantages de la MRAM sont la non-volatilité et le nombre illimité de cycles écriture-effacement. Malheureusement, la MRAM pose encore de nombreux problèmes, parmi lesquels un courant d'écriture élevé (entre 4 et 5mA par cellule), une résistance qui varie exponentiellement avec l'épaisseur de l'isolant tunnel et enfin une faible compatibilité avec les technologies « Logic CMOS ».

1.1.3.2 Les mémoires à changement de phase

Les mémoires à changement de phase, constituées de chalcogénures, sont des candidates prometteuses pour les prochaines générations de mémoires non-volatiles. Ces mémoires ont aussi appelées O.U.M. (Ovonic Universal Memory). Le matériau utilisé [Yamada'91] appartient à la même famille que ceux utilisés pour les disques compacts optiques réinscriptibles de type CD-RW ou DVD-RW. Dans les disques optiques, un faisceau laser de faible puissance est focalisé en un point sur la surface du disque afin de chauffer localement celui-ci et fait basculer le matériau d'un état cristallin à un état amorphe. L'état de la mémoire optique est déterminé par la réflectivité du matériau en ce point. Dans les mémoires à changement de phase, le passage d'un courant électrique à travers le matériau considéré (cf. « Programmable volume », figure I.8) provoque, par effet joule, le

changement de phase de ce dernier. Avant le passage du courant, le matériau se trouve dans une phase dont la résistivité est faible. Après le passage du courant, le matériau se trouve dans une phase amorphe, de résistivité plus élevée. La durée requise pour basculer à l'état amorphe est typiquement inférieure à 30 nanosecondes.

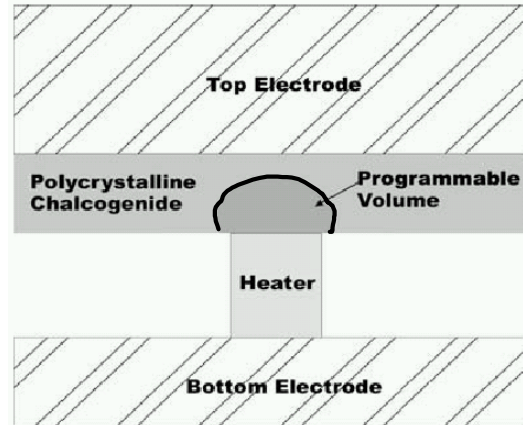


Figure I. 8 : *Vue en coupe de l'élément fondamental d'une mémoire à changement de phase.*

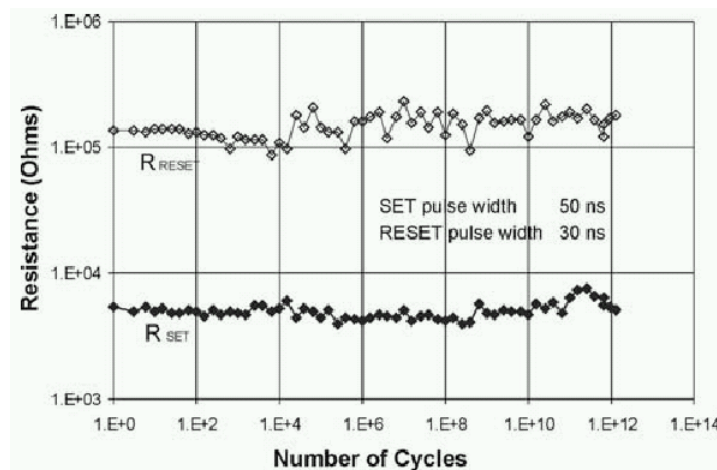


Figure I. 9 : *Test en endurance d'une mémoire à changement de phase.*

Plusieurs équipes de recherche ont fabriqué des prototypes de mémoires O.U.M. [Lai'01][Gill'02][Hwang'03]. Actuellement, la recherche se focalise sur trois domaines : l'étude de la physique de la cellule, la réduction du courant de programmation et le développement de matrices O.U.M. de haute densité.

La figure I.9 présente les résultats d'un test en endurance effectué sur une cellule mémoire O.U.M. On constate que la variation relative de résistance demeure inchangée après 10^{12} cycles.

I.1.4 Marché des mémoires à semi-conducteurs

Parallèlement à l'industrie des semi-conducteurs, le volume de marché occupé par les mémoires n'a cessé de croître. Cette croissance a été notamment

due à la très forte demande en ordinateurs personnels dans les années 2000. Les prévisions de la SIA (Semiconductor Industry Association) annonçaient respectivement un volume global de 204 et 319 milliards de dollars en 2000 et en 2003 pour toute l'industrie des semi-conducteurs. Les mémoires occuperaient 20% et 27% de ce marché en 2000 et en 2003 avec respectivement 49 et 88 milliards de dollars.

Prévisions de volumes du marché des mémoires à semi-conducteurs pour les années 2000 et 2003

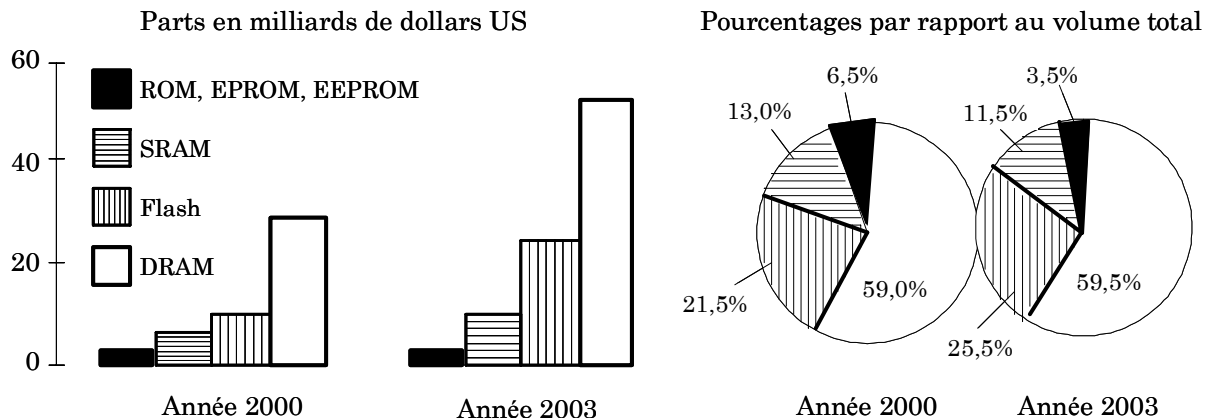


Figure I. 10 : *Prévisions de volumes du marché des mémoires à semi-conducteurs pour les années 2000 et 2003. (Source : Semiconductor Industry Association).*

Ce marché est inégalement réparti : en 2000, la DRAM arrive largement en tête avec 59% de l'ensemble, suivie de la Flash (21,5%), de la SRAM (13%) et enfin des ROM, EPROM et EEPROM (6,50%). Lorsque l'on regarde l'évolution de la proportion respective de chaque type de mémoire sur la période s'étendant de 2000 à 2003, on constate que la Flash connaît la plus forte augmentation avec 4% de part gagnée, suivie de la DRAM, à peu près stable avec 0,5% d'augmentation. A l'opposé, la SRAM et les ROM, EPROM et EEPROM sont en net recul avec respectivement 1,5% et 3% de baisse. On peut donc supposer que les mémoires Flash et DRAM, moteurs de la croissance de l'industrie des mémoires, resteront sur une pente ascendante dans le futur proche.

I.2 Technologie des mémoires non volatiles

Le principe consiste à garder l'information en l'absence d'alimentation électrique, en stockant des charges qui vont modifier la tension de seuil d'un transistor MOS. Par convention, cette tension de seuil peut prendre deux valeurs : un état logique « 0 » est défini comme un effacement et un état logique « 1 » comme une écriture. On peut accéder à cette information binaire en appliquant une tension comprise entre les deux tensions de seuil sur la grille du transistor : dans un cas, le transistor est conducteur (état logique « 1 ») et dans l'autre cas il ne l'est pas (état logique « 0 »). En matière de stockage de charge, les technologies des mémoires non-volatiles peuvent être divisées en deux catégories [Brown'98] :

la technologie à grille flottante et la technologie à piégeage de charges. On peut citer une troisième technologie non-volatile basée sur l'effet ferroélectrique : dans ce cas, l'état effacé ou écrit est fixé par l'état de polarisation du matériau [Brown'98]. On fixe cet état de polarisation ferroélectrique à l'aide d'un champ électrique.

I.2.1 Technologies à grille flottante

Le principe de la grille flottante consiste à isoler des charges dans une grille isolée en matériau conducteur ou semi-conducteur afin de faire varier la tension de seuil d'un transistor MOS. Dans le cas le plus courant les charges sont injectées à travers le diélectrique, souvent constitué par un oxyde (SiO_2), situé entre la grille flottante et le canal, (fig. I.11). Une deuxième grille, appelée grille de contrôle, est placée au dessus de la grille flottante. Ces deux grilles sont généralement séparées par une couche isolante de type ONO (Oxyde-Nitruure-Oxyde) [Brown'98].

Les technologies à grille flottante se distinguent par le mécanisme d'injection de charges dans la grille flottante en programmation (écriture) et en effacement. Lorsque l'on utilise une injection par électrons chauds en phase de programmation, on a une technologie SIMOS (Stacked Injection MOS). Cette technologie est utilisée dans les mémoires de type EPROM. Lorsqu'on utilise le mécanisme d'injection Fowler-Nordheim à travers une fine couche d'oxyde de 8 à 10nm d'épaisseur en effacement et en programmation, on a une technologie FLOTOX (FLOating gate Thin OXide). Cette technologie est utilisée pour les mémoires EEPROM et NOVRAM. Enfin, il existe une troisième technologie, appelée TPFPG (Textured Poly Floating Gate) dans laquelle on utilise une injection Fowler-Nordheim mais à travers une couche en poly-oxyde [Klein'79][Landers'80].

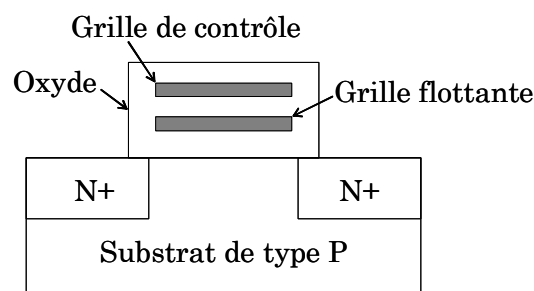


Figure I. 11 : *Structure à grille flottante.*

La figure I.12 représente une cellule SIMOS dont la grille flottante et la grille de contrôle sont en polysilicium. C'est la configuration typique d'une cellule de type EPROM [Rössler'77]. On programme la cellule par injection d'électrons chauds à travers l'oxyde de grille. L'opération inverse pour décharger la grille flottante par le même mécanisme est impossible et on procède par insolation UV en effacement. La programmation par injection d'électrons chauds consomme beaucoup d'énergie et exclut toute technologie basse tension d'alimentation. Par ailleurs, l'efficacité de l'injection par électrons chauds dépend fortement du

dopage de substrat, de la longueur effective du canal et de la longueur de la zone de recouvrement entre la grille flottante et le drain [Brown'98].

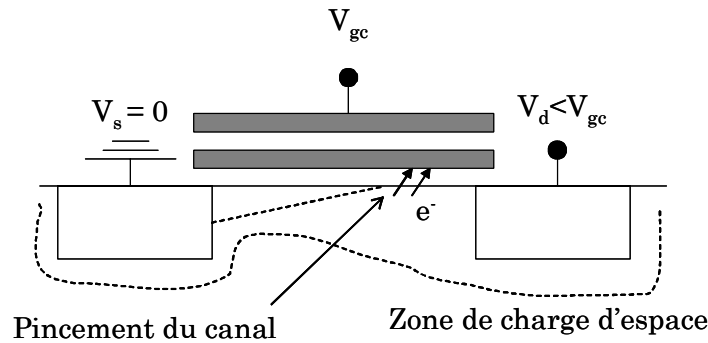


Figure I. 12 : Phase de programmation d'une cellule en technologie SIMOS montrant le mécanisme d'injection d'électrons chauds dans la grille flottante.

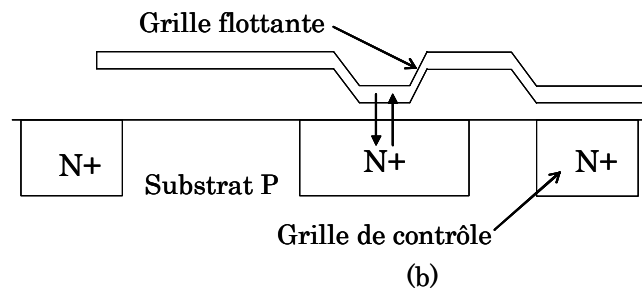
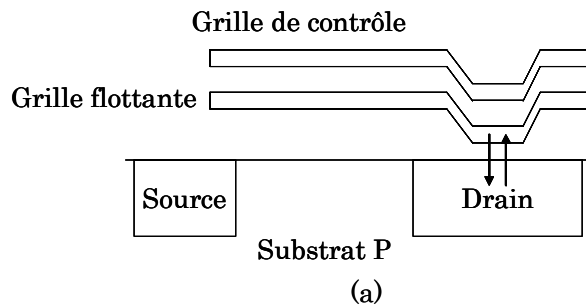


Figure I. 13 : Structures en technologie FLOTOX « Double Poly » (a) et « Simple Poly » (b).

La figure I.13 présente une structure FLOTOX. Dans cette technologie, on utilise le mécanisme d'injection Fowler-Nordheim, mis en oeuvre pour la première fois, en programmation et en effacement, dans une cellule mémoire non-volatile de type RAM [Harari'78]. Cette technologie a été à l'origine de l'apparition de la mémoire EEPROM au début des années quatre vingt [Johnson'80]. La technologie FLOTOX peut être « Simple Poly » ou « Double Poly ». En « FLOTOX Simple Poly », il y a un seul niveau de silicium poly-cristallin qui est la grille flottante. La grille de contrôle est enterrée dans le substrat et est séparée de la grille flottante par une couche d'oxyde de 20 à 30 nm d'épaisseur. En « FLOTOX Double Poly », un second niveau de silicium poly-cristallin est séparé de la grille flottante par un isolant appelé oxyde inter-poly qui est souvent une couche ONO (Oxyde-Nitruire-Oxyde). Que ce soit en FLOTOX « Simple » ou « Double Poly », une mince couche d'oxyde (8 à 10nm) sépare le drain de la grille flottante. Des

électrons traversent cette couche d'oxyde par effet tunnel en présence d'un champ électrique intense. Suivant le sens de ce champ électrique, les électrons sont injectés du drain vers la grille flottante ou inversement, ce qui a pour effet d'augmenter ou de diminuer la tension de seuil du transistor.

Enfin, dans la technologie TPF_G, où l'injection de porteurs est aussi de type Fowler-Nordheim, la couche en poly-oxyde est obtenue en faisant croître thermiquement de l'oxyde sur une couche de poly-silicium. Les mécanismes de conduction à travers ce type de structure requièrent des champs électriques plus faibles (2MV/cm) que dans le cas d'un oxyde formé à partir d'un silicium monocristallin. Ceci est dû à la présence d'aspérités à l'interface du poly-silicium et du poly-oxyde. En comparaison avec la technologie FLOTOX, la technologie TPF_G permet d'utiliser une épaisseur d'oxyde plus grande puisque le champ requis est plus faible, ce qui provoque une moindre dégradation et assure une meilleure fiabilité.

I.2.2 Technologies à piégeage de charges

Cette technologie est basée sur le piégeage de charges dans une couche diélectrique isolante.

I.2.2.1 Technologies MNOS et SNOS

Dans la technologie MNOS (Métal Nitrure Oxyde Semiconducteur), au lieu de stocker les charges dans une grille flottante isolée électriquement, on utilise les pièges situés dans une couche de nitrure. Les électrons sont injectés dans le nitrure à travers une fine couche d'oxyde de 1.5 à 3nm d'épaisseur par effet tunnel. Ces électrons, une fois piégés, modifient la tension de seuil du transistor. La figure I.14 représente une cellule EAROM (Electrically Alterable semiconductor Read Only Memory) en technologie MNOS. Le transistor de cette cellule EAROM possède une grille conventionnelle en métal. Les charges sont injectées à travers la couche d'oxyde centrale située entre le canal et la couche de nitrure Si_3N_4 . L'inconvénient majeur de cette mémoire réside dans sa lenteur, sa densité limitée ainsi que la nécessité d'appliquer 2 à 3 tensions différentes pour son fonctionnement.

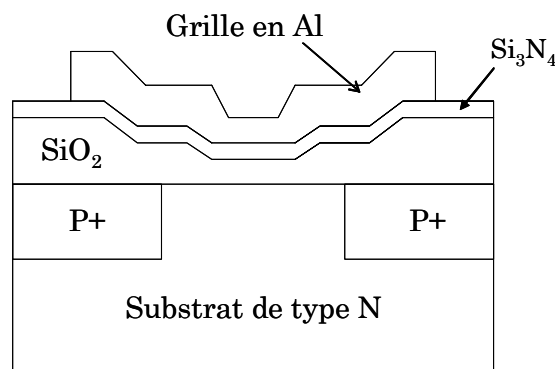


Figure I. 14 : Structure en technologie MNOS.

En 1980, la technologie MNOS a été nettement améliorée en réalisant la technologie SNOS (Silicium-Nitruire-Oxyde-Silicium). La fiabilité de cette dernière est basée sur l'utilisation de la méthode LPCVD (Low-Pressure Chemical Vapor Deposition) pour le dépôt du poly-silicium sur la couche de nitruire et la « pré-métallisation » par un recuit d'hydrogène haute température pour améliorer la qualité des interfaces nitruire-oxyde mince-silicium [Brown'98].

1.2.2.2 Technologie SONOS

Cette technologie (Silicium-Oxyde-Nitruire-Oxyde-Silicium) a été mise en œuvre afin de réduire l'injection de charges de la grille vers le nitruire comme dans le cas de la technologie MNOS. Une couche d'oxyde de 2 à 3nm d'épaisseur est introduite entre le nitruire et la grille en poly-silicium [Chen'77].

Cette couche d'oxyde est obtenue par oxydation du nitruire, ce qui diminue l'épaisseur de cette dernière, ou par dépôt. Toutefois, cette réduction de l'épaisseur du nitruire provoque la fuite des trous vers l'électrode de grille. En effet, comme la longueur de piégeage dans le nitruire est plus importante pour les trous (15 à 20nm) que pour les électrons (5 à 10nm), en réduisant l'épaisseur du nitruire (<20 nm), les trous sont piégés à l'interface entre le nitruire et l'électrode de grille alors que les électrons sont piégés dans le nitruire.

La majorité des trous sera perdue à travers l'électrode de grille même en présence de la fine couche d'oxyde entre le nitruire et l'électrode de grille. Ceci provoque une diminution significative de la tension de seuil.

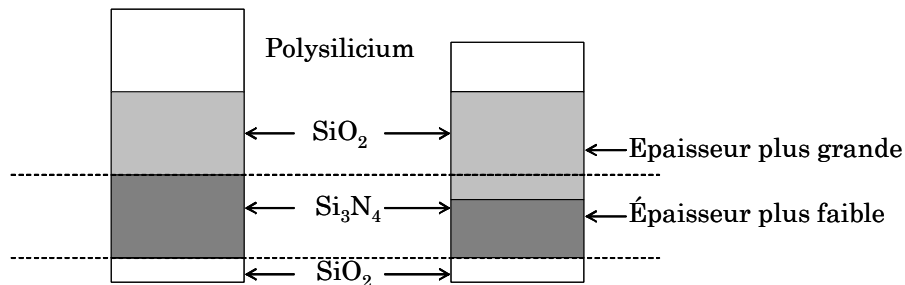


Figure I. 15 : *Évolution de la technologie SONOS où la couche d'oxyde entre le nitruire et le poly-silicium devient plus épaisse, la couche de nitruire moins épaisse et l'ensemble de la structure moins épaisse.*

Afin de pallier à cette fuite des trous piégés vers l'électrode de grille, un nouveau concept de la technologie SONOS a été mis en œuvre tout en conservant une haute densité d'intégration ainsi que des basses tensions de programmation [Suzki'83][Chao'87][Dellin'87]. La nouvelle structure (fig. I.15) est donc composée d'une mince couche d'oxyde, d'une épaisseur fine de nitruire (<10nm) et d'un oxyde plus épais (>3nm) [Suzki'83][Dellin'87]. Ceci permet de réduire l'épaisseur totale de la structure et les tensions de programmation (<5V); ce qui implique une dégradation moindre et une meilleure rétention [Libsch'86].

I.2.3 Technologie ferroélectrique

Dans cette technologie, les opérations d'écriture et d'effacement sont basées sur l'effet ferroélectrique. En effet, certains matériaux se polarisent spontanément sous l'effet d'un champ électrique et conservent cette polarisation, dite rémanente, même en l'absence de champ électrique. Les mémoires non-volatiles ferroélectriques sont basées sur l'état de polarisation d'une capacité en film mince ferroélectrique. L'information stockée est lue en quantifiant l'interaction entre un champ de « lecture » et l'état de polarisation de la capacité.

Il existe deux candidats potentiels pour le type de matériau ferroélectrique utilisé : la céramique de type Perovskite PZT (Perovskite-type ferroelectric : $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$) et la céramique SBT (Aurivilius layered ferroelectric : $\text{SrBi}_2\text{Ta}_2\text{O}_9$). Cette technologie permet une haute densité d'intégration, une rapidité d'écriture, une endurance élevée ainsi qu'une faible consommation d'énergie [Brown'98].

Les différentes technologies énumérées précédemment ont chacune des avantages et des inconvénients. Pour certaines technologies, les efforts entrepris dans la compréhension des mécanismes de dégradation ont permis de mettre au point des structures plus fiables (SONOS). Cependant, des critères de coût et de compatibilité avec les technologies standard limitent leur domaine d'utilisation. C'est pourquoi il est important de poursuivre l'étude et l'amélioration des technologies classiques déjà industrialisées.

I.3 Mémoire EEPROM

I.3.1 Structure de la cellule mémoire EEPROM

La mémoire EEPROM est constituée de deux transistors MOS mis en série : un transistor à grille flottante appelé « transistor d'état » et un transistor haute tension appelé « transistor de sélection ».

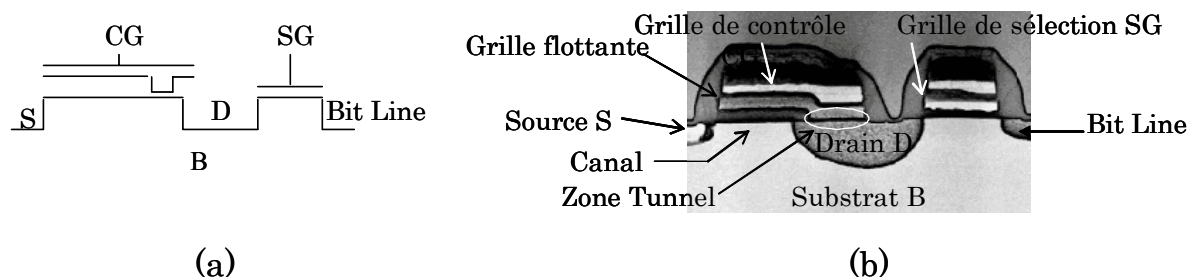


Figure I. 16 : Vue schématique (a) et en coupe (b) d'une cellule EEPROM avec les cinq électrodes (drain, source, substrat, grille de contrôle et grille de sélection).

Le transistor d'état contient l'information sous forme de charges stockées dans la grille flottante. Ces charges sont injectées dans la grille flottante par effet tunnel Fowler-Nordheim à travers la mince couche d'oxyde située entre la grille flottante et le drain. Les cellules EEPROM que nous avons étudiées (fig. I.16) sont des structures « FLOTOX Double Poly ».

I.3.2 Principe de fonctionnement

Le fonctionnement d'une cellule mémoire EEPROM repose sur le décalage de la tension de seuil du transistor d'état, MOS à effet de champ, lorsque des charges sont piégées dans son oxyde de grille. Par convention, une cellule dont la grille flottante est neutre est dite « vierge ». Une cellule dont la grille flottante est chargée positivement est dite « écrite » et à l'état logique « 1 », le transistor est passant. Une cellule dont la grille flottante est chargée négativement est dite « effacée » et à l'état logique « 0 », le transistor est bloqué.

Pour écrire une cellule EEPROM, il faut injecter des charges positives dans la grille flottante. Ceci est réalisé en polarisant le drain à un potentiel positif (de 12 à 14V), la grille de contrôle et le substrat étant à la masse et la source flottante (haute impédance). La grille du transistor de sélection est polarisée à un potentiel fortement positif (de 14 à 16V) afin de rendre ce transistor passant. La présence de charges positives dans la grille flottante va décaler la caractéristique $I_{ds}(V_{gs})$ du transistor d'état du côté des tensions négatives, par rapport à la caractéristique d'une cellule vierge : la tension de seuil $V_{th\text{écrit}}$ du transistor d'état d'une cellule écrite est donc inférieure à la tension de seuil $V_{th\text{vierge}}$ du transistor d'état d'une cellule vierge.

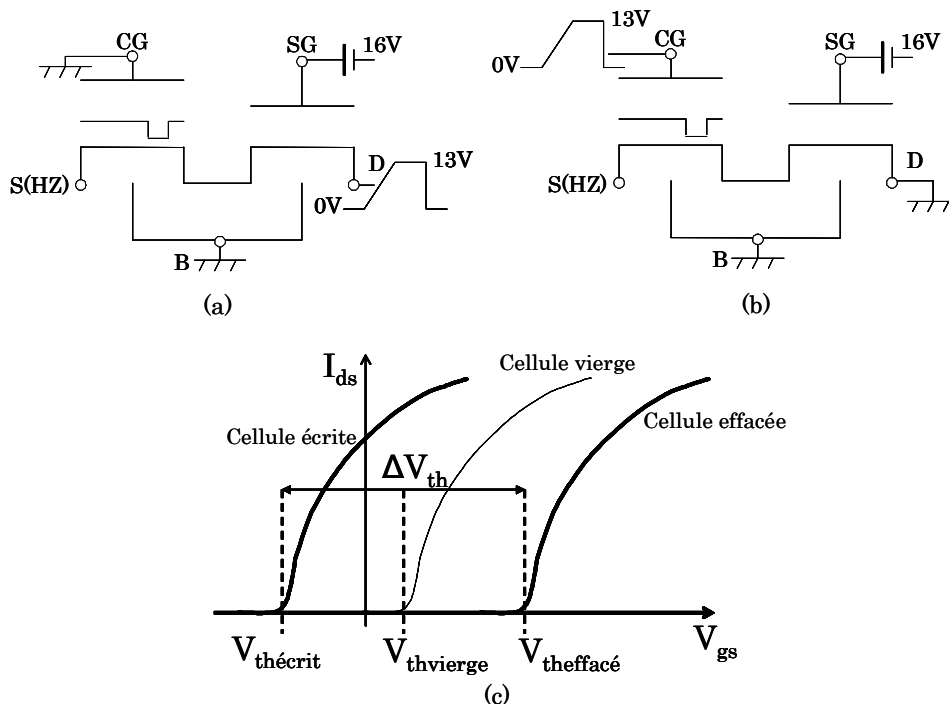


Figure I. 17 : Polarisation d'une cellule EEPROM en écriture (a), en effacement (b) et caractéristiques $I_{ds}(V_{gs})$ dans les états vierge, écrit et effacé (c).

Pour effacer une cellule EEPROM, il faut injecter des électrons dans la grille flottante. Il faut inverser le sens du champ électrique par rapport à la phase d'écriture tout en gardant la même intensité. On réalise cela en polarisant cette fois la grille de contrôle à un potentiel positif (de 12 à 14V), le drain et le substrat étant à la masse et la source restant flottante. La grille du transistor de sélection est polarisée à un potentiel fortement positif pour le rendre passant (de 14 à 16V). Cette fois-ci, la caractéristique $I_{ds}(V_{gs})$ est décalée vers les tensions positives par rapport à celle d'une cellule vierge et la tension de seuil $V_{th\text{effacé}}$ du transistor d'état est plus grande que celle d'une cellule vierge. La figure I.17 représente les polarisations des électrodes d'une cellule EEPROM en phase d'écriture et en phase d'effacement ainsi que la caractéristique $I_{ds}(V_{gs})$ correspondant aux trois états (écrit, vierge et effacé).

L'état logique d'une cellule EEPROM est lue en forçant un courant de référence dans le canal du transistor d'état. On lit ensuite la tension de grille de contrôle correspondante. On compare cette tension à deux tensions de référence haute ($V_{ref,haute}$) et basse ($V_{ref,basse}$): si la tension obtenue est supérieure à la tension de référence haute, la cellule est reconnue comme étant effacée et correspond à l'état logique « 0 ». A l'opposé, si la tension obtenue est inférieure à la tension de référence basse, la cellule est reconnue comme étant écrite et correspond à l'état logique « 1 ». La figure I.18 montre les deux cas.

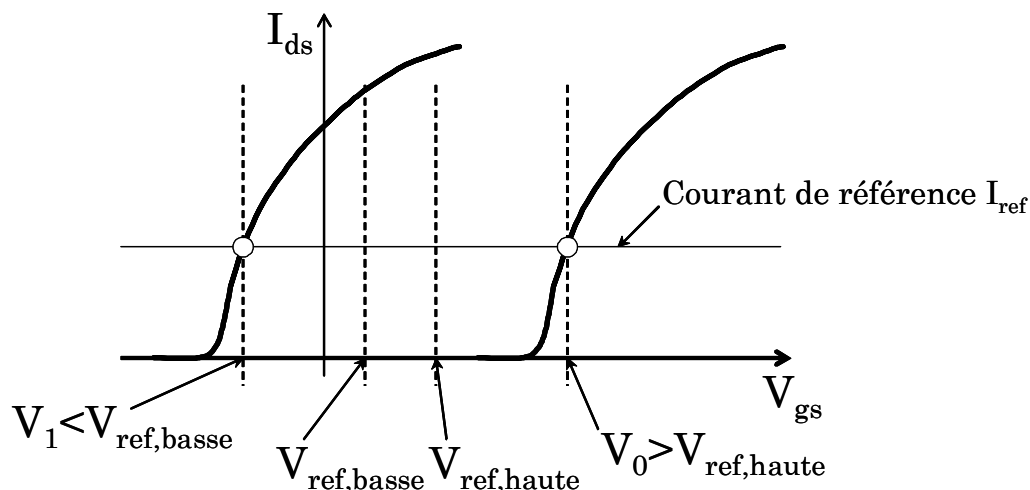


Figure I. 18 : *Lecture de l'information contenue dans une cellule EEPROM : état logique « 0 » pour $V_0 > V_{ref,haute}$ et état logique « 1 » pour $V_1 < V_{ref,basse}$*

I.3.3 Architecture des matrices mémoires EEPROM

Le point mémoire EEPROM est constitué de deux transistors MOS (transistors d'état et de sélection) reliés au reste du plan-mémoire par quatre électrodes : la ligne de mot (WL), la ligne de sélection de mot (WS), la ligne de bit (BL) et la ligne de masse (fig. I.19). Le transistor de sélection agit comme un interrupteur placé entre le drain du transistor d'état et la ligne de bit BL. Ainsi, le drain du transistor de sélection est connecté à la ligne de bit et sa source est connectée au drain du transistor d'état. Sa grille est commandée par la ligne de

sélection de mot WS alors que celle du transistor d'état est commandée par la ligne de mot WL. Enfin, le substrat des deux transistors et la source du transistor d'état sont reliés à la masse (dans certaines technologies, la source du transistor d'état peut également rester flottante via un autre transistor).

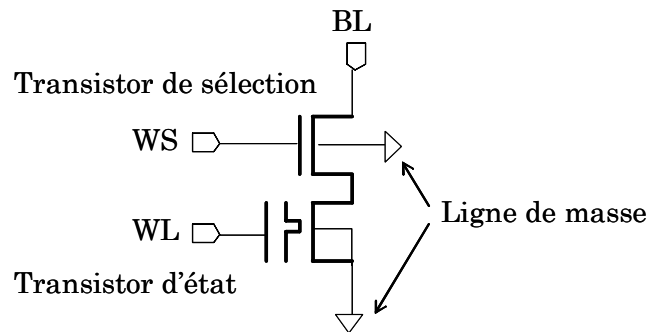


Figure I. 19 : *Point mémoire EEPROM.*

Les matrices mémoires EEPROM sont organisées selon m lignes et n colonnes de points mémoires. Il y a un point mémoire à chaque nœud (intersection d'une ligne et d'une colonne). Une matrice mémoire EEPROM comporte donc m x n points mémoires. Les points mémoires qui se trouvent sur une même rangée forment un mot de n bits, auquel s'ajoute un transistor de sélection de mot.

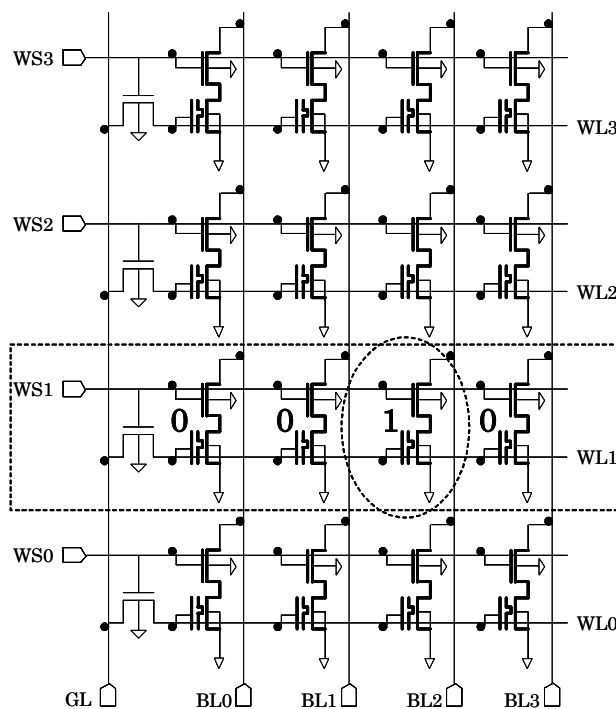


Figure I. 20 : *Exemple de matrice EEPROM 4x4.*

Sur la figure I.20, nous avons l'exemple d'une matrice mémoire EEPROM 4 x 4 : c'est une matrice mémoire de quatre mots de 4 bits. Ce plan mémoire comporte quatre lignes de mot (WL3, WL2, WL1 et WL0), quatre lignes de sélection de mot (WS3, WS2, WS1 et WS0), quatre transistors de sélection de mot, quatre lignes de bits (BL3, BL2, BL1 et BL0) et une ligne de grille (GL).

La brique de base d'un plan mémoire est le mot : il est constitué de n points mémoires, un transistor de sélection de mot, une ligne de mot WLi , une ligne de sélection de mot WSi , une ligne de grille GL et n lignes de bit BLj . Le transistor de sélection de mot joue le même rôle que celui d'un point mémoire : sa grille est reliée à la ligne de sélection de mot tandis que son drain est relié à une ligne de grille et sa source à la ligne de mot WLi .

Nous allons illustrer le processus par l'exemple de l'écriture du mot 1 ($WS1$ et $WL1$) sur la matrice de la figure I.20. Écrivons le mot 1101.

L'écriture du mot comporte quatre étapes :

i. Sélection du mot 1

On sélectionne ce mot en appliquant une tension V_{sel} sur $WS1$. Les autres mots, non sélectionnés, ont leur ligne de sélection WSi à la masse et ne peuvent donc être ni écrits ni programmés. Ainsi, seuls les transistors de sélection des points mémoires du mot 1 sont passants.

ii. Effacement de tous les points mémoires du mot 1

Il faut charger négativement la grille flottante de tous les points mémoires du mot sélectionné afin de les effacer. Pour cela, on élève la ligne de grille GL au potentiel V_{EE} et on relie toutes les lignes de bit BLj à la masse. Ainsi, toutes les grilles flottantes du mot 1 sont portées à un potentiel positif par couplage capacitif avec les grilles de contrôle. Il y a injection d'électrons du drain vers la grille flottante de chaque point mémoire qui se retrouve alors dans l'état effacé. A la fin de l'opération, on ramène la ligne de grille au potentiel de masse pour arrêter l'injection.

iii. Programmation des points mémoires qui ne doivent pas être effacés

A la fin de la deuxième étape, nous avons écrit le mot 0000. Or, nous voulons écrire le mot 0010. Nous devons donc changer le deuxième bit à partir de la droite en programmant le point mémoire de la ligne de bit BL_2 : il faut charger la grille flottante de ce point mémoire positivement. Il suffit de maintenir la ligne de grille GL et toutes les lignes de bit BLj à la masse, à l'exception de BL_2 . Cette dernière est élevée au potentiel V_{EE} et amène la grille flottante à un potentiel inférieur à V_{EE} par couplage capacitif. Ainsi, des électrons sont injectés de la grille flottante vers le drain. Le point mémoire est programmé et nous avons écrit le mot 0010.

iv. Fin de sélection

A la fin de l'opération d'écriture, la ligne de sélection $WS1$ est ramenée à la masse.

Voyons à présent comment on lit un mot dans une matrice mémoire. L'opération comporte cinq étapes.

i. Sélection du mot 1

On sélectionne le mot 1 en polarisant la ligne de sélection WS_1 à V_{PP} comme en écriture.

ii. Polarisation de toutes les lignes de bit à BL_j au potentiel V_d

iii. *Polarisation de la ligne de grille GL à V_g*

iv. *Mesure des courants de drain et comparaison à une référence unique*

Le courant de drain de chaque point mémoire est mesuré et comparé à une référence unique. Par exemple, le point mémoire de la ligne de bit BL2 est programmé : sa tension de seuil est basse. Par conséquent, son courant de drain est élevé et supérieur au courant de référence. A l'opposé, tous les autres points mémoires de ce mot sont effacés : leur tension de seuil est élevée. Par conséquent, leur courant de drain est faible et inférieur au courant de référence.

v. *Fin de sélection du mot 1*

A la fin de la lecture, la ligne de sélection WS1 est ramenée à la masse.

Dans cette troisième partie, nous avons examiné en détail le fonctionnement des mémoires EEPROM en exposant les phénomènes physiques qui ont rendu possible leur conception, les éléments constitutifs de la plus petite unité (point mémoire) et enfin en expliquant l'architecture d'une matrice mémoire et les opérations que l'on peut y effectuer.

Néanmoins, cet exposé demeure incomplet car nous avons examiné le fonctionnement des EEPROM de façon ponctuelle : nous n'avons pas pris en compte les effets de l'environnement, du temps ou d'un grand nombre d'utilisations sur le fonctionnement du dispositif. Cet aspect essentiel est l'objet de la dernière partie de cet état de l'art sur les mémoires non-volatiles à semi-conducteurs.

I.4 Fiabilité des MNV à grille flottante

La fiabilité est une préoccupation d'importance grandissante dans l'industrie microélectronique. Ceci est particulièrement vrai pour les fabricants de mémoires non-volatiles. En effet, la course à la miniaturisation entraîne une réduction de la taille de tous les composants dans un circuit intégré. D'autre part, la baisse des tensions d'alimentation impose la réduction des épaisseurs d'oxyde de grille des transistors MOS : l'épaisseur de l'oxyde de grille d'un transistor MOS est désormais inférieure à 2 nm pour les technologies de pointe. A titre d'exemple, une polarisation de 2V seulement provoque un champ électrique de 10 MV/cm à travers de tels oxydes. Des champs aussi importants sont particulièrement néfastes à la fiabilité des mémoires car ils peuvent provoquer des transports de charges indésirables à travers le diélectrique qui entoure la grille flottante, modifiant ainsi l'état logique de la cellule et, à terme, dégrader fortement l'oxyde tunnel et rendre la cellule mémoire inutilisable.

La fiabilité est la capacité d'un composant électronique à fonctionner correctement dans le temps. On peut déterminer quantitativement la fiabilité en mesurant le taux de défaillance d'un dispositif en fonction du temps : l'unité communément adopté du taux de défaillance est le FIT (Failure unIT).

Par définition,

$$1 \text{ FIT} = \frac{1 \text{ défaillance}}{10^9 \text{ dispositifs} \times \text{heure}} = 10^{-9} \text{ h}^{-1} \quad (\text{I. 01})$$

Typiquement, la courbe de défaillance d'un dispositif est une courbe dite « en baignoire » (« bath-tub ») [NVSM'98-Chapitre 6]. Cette courbe comporte typiquement trois parties : une première partie (A) qu'on appelle mortalité infantile, une deuxième partie appelée zone utile et une dernière partie appelée période d'usure ou fin de vie. La durée de chacune de ces trois parties dépend à la fois des caractéristiques du produit, des conditions d'utilisation et des contraintes auxquelles il est soumis.

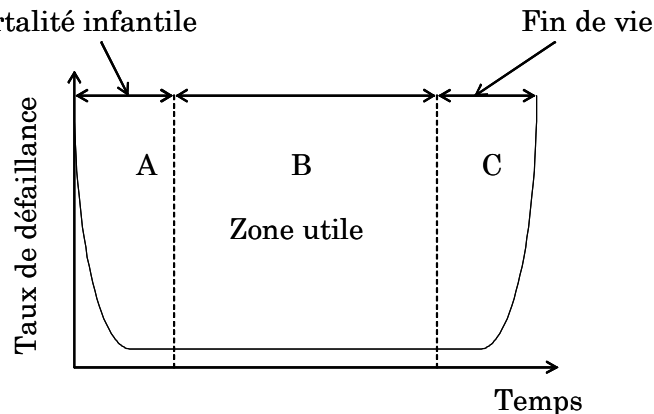


Figure I. 21 : *Courbe classique en baignoire (« bath-tub ») de la fiabilité.*

Pour les mémoires non-volatiles, par fiabilité, l'on entend généralement deux concepts :

i. *La capacité de la mémoire à garder des caractéristiques physiques qui la rendent encore utilisable après un certain nombre de cycles écriture/effacement.*

Ce concept est souvent appelé « endurance » de la mémoire. Typiquement, une cellule EEPROM doit garder une fenêtre de programmation exploitable au bout d'un million de cycles écriture/effacement.

ii. *La capacité de la mémoire à garder sans défaillance pendant une durée donnée l'information stockée par l'utilisateur.*

Ce concept est appelé « rétention » de la mémoire. Typiquement, une cellule EEPROM standard doit avoir une rétention de dix ans.

I.4.1 Endurance

L'endurance est un problème crucial pour toute mémoire non-volatile [Crisenza'91]. En effet, des cycles programmation/effacement répétés finissent par réduire fortement les performances des cellules mémoires [Cappelletti'94][Haddad'89][Aritome'90][Yamada'93] en endommageant l'oxyde à travers lequel les charges sont injectées du drain vers la grille flottante ou inversement (dans le cas des Flash EEPROM, ce sont les électrons chauds provenant du canal qui sont néfastes à l'endurance de la cellule).

Le courant qui passe à travers cet oxyde crée des états d'interfaces, des pièges et une accumulation de charges dans l'oxyde. Les principales conséquences de ce stress sont au nombre de trois :

- i. Dégradation des caractéristiques électriques du transistor (surtout pour les cellules Flash).*

Cette dégradation se caractérise notamment par une diminution de la mobilité, de la transconductance et du courant de drain. La réduction de la transconductance et du courant de drain peut conduire à des lectures erronées.

- ii. Ecrantage de la charge emmagasinée dans la grille flottante par les charges piégées dans l'oxyde.*

Dans le cadre de notre travail, nous avons surtout analysé ce deuxième effet. Les charges piégées dans l'oxyde, en faisant écran, diminuent la transparence de la barrière tunnel d'un point de vue quantique. Par conséquent, le courant tunnel Fowler-Nordheim (cas des cellules EEPROM) baisse, entraînant une diminution de la charge stockée dans la grille flottante et finalement une réduction de la largeur de la fenêtre de programmation.

- iii. Claquage possible de l'oxyde.*

La méthode d'écriture influence fortement l'endurance. En effet, plus la zone d'injection est localisée, plus les pièges créés seront localisés et denses, réduisant ainsi l'endurance. C'est le cas notamment des mémoires qui utilisent un effacement par électrons chauds [Mahapatra'00]. C'est également le cas pour les mémoires qui utilisent l'injection Fowler-Nordheim sur les zones de recouvrement grille/drain ou grille/source. L'ordre de grandeur annoncée par les fabricants de mémoires est actuellement une endurance de 10^6 cycles programmation/effacement.

I.4.2 Rétention

La tension de seuil V_{th} d'une cellule mémoire non-volatile dépend de la charge stockée. Après une phase d'effacement, la valeur de cette tension de seuil peut être modifiée à la suite d'une modification de la quantité de charges stockées dans la cellule mémoire. Si la quantité de charges dans la grille flottante augmente, la tension de seuil augmente aussi : on parle alors d'un gain de charge. Si la quantité de charges dans la grille flottante diminue, la tension de seuil dans l'état effacé diminue aussi : on parle alors d'une perte de charges. On définit pour chaque type de mémoire un seuil de défaillance : il s'agit de la diminution de tension de seuil en état effacé au-delà de laquelle on considère que l'état logique de la cellule a changé. Par définition, la rétention de la cellule est l'intervalle de temps compris entre l'instant où la dernière phase d'effacement est terminée et l'instant où la diminution de la tension de seuil en état effacé atteint une valeur critique.

1.4.2.1 Mécanismes intrinsèques de perte de charge

Dans une cellule effacée, la grille flottante chargée négativement induit un faible champ électrique (de 2 à 3MV/cm) à travers l'oxyde tunnel sous l'effet duquel les électrons qui sont emmagasinés dans la grille flottante ont tendance à fuir. La perte de charge due à ce champ résiduel a longtemps été considérée, à tort, comme étant négligeable. A titre d'exemple, la durée de vie d'une cellule non défectueuse a été estimée à un million d'années pour une température de fonctionnement de 70°C [Shiner'80][Mielke'83].

Les électrons stockés dans la grille flottante peuvent fuir par trois chemins principaux : à travers le diélectrique inter-poly, à travers l'oxyde de grille ou à travers l'isolation latérale.

i. Fuite à travers l'isolation latérale

Grâce à des études antérieures, on peut considérer que cette fuite est négligeable [Candelier'97][Mazoyer'92].

ii. Fuite à travers le diélectrique inter-poly

Dans le cas fréquent où le diélectrique inter - poly est une couche ONO, la perte de charge est principalement attribuée à cette couche plutôt qu'à l'oxyde de grille pour trois raisons :

- La surface de recouvrement entre la grille flottante et la couche ONO est beaucoup plus grande que celle entre la grille flottante et la couche d'oxyde de grille.
- La qualité de la couche ONO est moins bonne que celle de l'oxyde de grille car la couche ONO est déposée sur du poly-silicium (la grille flottante en l'occurrence).

Enfin, la géométrie de la structure grille flottante/ONO/grille de contrôle présente des angles aigus qui augmentent significativement la valeur du champ électrique.

iii. Fuite à travers l'oxyde de grille

Pour des températures très élevées, la perte de charge à travers l'oxyde de grille a été expliquée à l'aide de mécanismes tels que l'émission thermoïonique [Nozawa'82] ou l'effet tunnel de type Fowler-Nordheim. Toutefois, Nozawa a trouvé une hauteur de barrière irréaliste ($\approx 1,5\text{eV}$) en utilisant ce modèle sur des caractéristiques expérimentales de perte de charge. En effet, pour les valeurs habituelles de cette hauteur de barrière (3eV), Crisenza a montré que l'émission thermoïonique est quasiment négligeable [Crisenza'91]. Par ailleurs, on a longtemps admis que la perte de charge par effet tunnel de type Fowler-Nordheim à travers l'oxyde de grille était négligeable, le courant tunnel correspondant à une barrière de potentiel de 3eV étant très faible à bas champ. L'importance de cette fuite tunnel a été réévaluée, en particulier à haute température [Papadas'95], suite aux travaux montrant la forte dépendance du courant Fowler-Nordheim avec la température [Pananakakis'95].

1.4.2.2 Mécanismes extrinsèques de perte de charge

Les mécanismes extrinsèques qui sont à l'origine de la perte de charge dans les mémoires non-volatiles à grille flottante sont au nombre de deux :

i. Les défauts des diélectriques

Les défauts qui existent dans le diélectrique qui entoure la grille flottante (oxyde tunnel, oxyde de grille et oxyde inter-poly) peuvent créer des chemins conducteurs qui vont décharger la grille flottante.

Généralement, la conduction due aux défauts de l'oxyde de grille se fait par saut (ou « hopping »), avec une énergie d'activation de 0,6eV [Shiner'80] alors que la conduction due aux défauts dans l'oxyde inter-poly se fait par effet « Poole-Frenkel » [Mielke'83]. Par ailleurs, lorsque le nitrure devient plus épais et l'oxyde situé entre la couche de nitrure et la grille en polysilicium (top-oxide) devient plus mince, la variation de tension de seuil est plus importante [Mori'91].

ii. La contamination ionique

La présence de contaminants dans la cellule peut modifier la tension de seuil ou provoquer des défaillances. Les ions contaminants sont la plupart du temps de charge positive (surtout des ions Sodium Na^+). Attirés par la grille flottante à cause du champ induit par la charge qui y est stockée, ces ions font écran à la charge emmagasinée, provoquant ainsi une « perte de charge ».

On peut détecter une éventuelle contamination ionique à l'aide d'une méthode dite de « gain de charge » [Mielke'83]. On effectue un recuit à température très élevée d'une cellule effacée suivi successivement d'un effacement UV puis d'un autre recuit. On suppose que la perte de charge, après le premier recuit, est due au champ électrique induit par la charge négative stockée alors que le gain de charge constaté après le deuxième recuit est dû au gradient de concentration des ions contaminants.

1.4.2.3 Test en rétention

Le test en rétention est une procédure accélérée activée thermiquement dont l'objectif est d'évaluer la perte de charge d'une cellule mémoire en fonction du temps dans les conditions normales de fonctionnement. La norme actuelle spécifie une perte de charge inférieure à 10% de la charge initiale en 10 ans.

Concrètement, le test en rétention se fait en mesurant la dérive en fonction du temps de la tension de seuil d'une cellule mémoire initialement chargée négativement. Afin de limiter la durée du test (inférieure à 1000 heures), on active la perte de charge en plaçant la cellule dans un four à haute température (200°C, 250°C, 300°C). Pour chacune de ces trois températures, on estime la durée de vie de la cellule en calculant, par extrapolation, le temps au bout duquel on atteint un seuil de défaillance fixé à l'avance (un pourcentage donné de la charge initiale emmagasinée).

L'extrapolation de ces temps à la température ambiante, qui peut être faite suivant une loi d'Arrhénius activée en température, permet d'évaluer la durée de

vie de la cellule dans les conditions normales de fonctionnement. La loi classiquement utilisée est exprimée par la relation (I.02).

I.4.3 Modèles prédictifs de rétention de données

Afin de déterminer la rétention des mémoires non-volatiles dans les conditions normales de fonctionnement, il est nécessaire d'extrapoler à température ambiante les mesures effectuées à haute température.

Jusqu'à une époque récente, l'extrapolation des données se faisait avec la loi d'Arrhenius. Mais en 1999, De Salvo a démontré empiriquement que cette loi n'était pas valable [De Salvo'99] : elle a formulé une nouvelle loi qui permet de mieux interpréter les observations et d'unifier les résultats alors disponibles.

I.4.3.1 Modèle en $1/T$

On extrapole les données recueillies à haute température à l'aide de la loi d'Arrhenius :

$$t_R(T) = t_0 \exp\left(\frac{E_a}{k_B T}\right) \quad (\text{I. 02})$$

où :

t_R est le temps de rétention à la température T (pour un seuil de défaillance fixé à l'avance), k_B la constante de Boltzmann, E_a l'énergie d'activation et T la température en Kelvin.

Cette relation peut encore s'écrire :

$$\ln[t_R(T)] = \ln(t_0) + \frac{E_a}{k_B T} \quad (\text{I. 03})$$

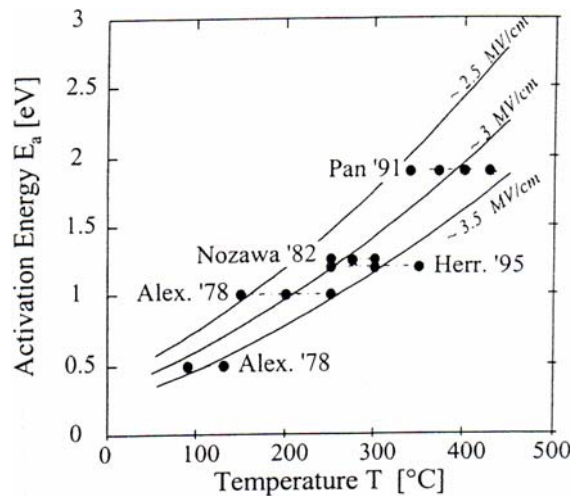


Figure I. 22 : Énergies d'activation en fonction de la température [De Salvo'99].

La donnée de deux points expérimentaux permet donc de déterminer l'équation de la droite qui représente le logarithme de t_R en fonction de $1/T$. L'énergie

d'activation E_a est déduite de la pente de cette droite. Toutefois, ce modèle n'est pas satisfaisant car selon la gamme de température explorée, l'énergie d'activation varie (fig. I.22) [Nozawa'82][Herrmann'95][Pan'91a][Pan'91b][De Salvo'99]. En réalité, le logarithme de t_R n'est pas une fonction affine de $1/T$.

I.4.3.2 Modèle en T

Afin de résoudre ce problème, un modèle qui unifie ces résultats a été proposé en 1999 [De Salvo'99], grâce à une analyse fine de la dépendance en température du courant Fowler-Nordheim. La loi suivante permet d'interpréter tous les résultats :

$$t_R(T) = t_0 \exp\left(-\frac{T}{T_0}\right) \quad (\text{I. 04})$$

où T_0 est une température caractéristique de la rétention de données. Les courbes (fig. I.23) montrent notamment que le modèle en $1/T$ surestime la valeur réelle du temps de rétention.

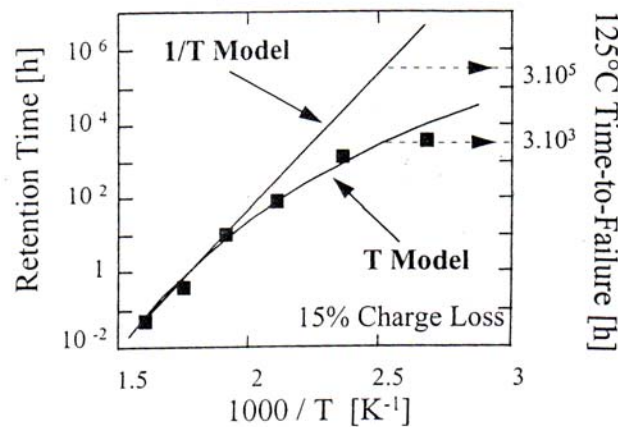


Figure I. 23 : *Extrapolation du temps de rétention avec la loi en T et en $1/T$ [De Salvo'99].*

Conclusion

Dans ce chapitre, nous avons passé en revue l'ensemble des mémoires, volatiles et non-volatiles, présentes sur le marché. Ces mémoires ont été conçues pour des objectifs différents et présentent des caractéristiques et des performances différentes. Nous avons notamment examiné en détail le fonctionnement d'une cellule EEPROM et les principales difficultés posées par la fiabilité de ce type de mémoire. Ainsi, nous avons pu situer le contexte dans lequel s'inscrivent nos travaux de thèse.

Références bibliographiques du chapitre I

- [Frenkel'38] J. Frenkel, "On Pre-Breakdown Phenomena in insulators and Electronic Semiconductors", Phys. Rev., vol. 54, p 647, 1938.
- [Kahng'67] D. Kahng and S. M. Sze, "A floating gate and its application to memory devices", Bell Syst. Tech. J, vol. 46, pp. 1288, 1967.
- [Lenzlinger'69] M. Lenzlinger, E.H. Snow, "Fowler-Nordheim Tunneling into Thermally Grown SiO₂", J. Appl. Phys., 40, 278, 1969.
- [Jullière'75] M. Jullière, "Tunneling between Ferromagnetic Films", Phys. Lett. 54A : 225, 1975.
- [Chen'77] P. C. Chen, "Threshold-alterable Si-gate MOS devices", IEEE Trans. Elec. Devices, vol. ED-24, pp. 584, 1977.
- [Rössler'77] B. Rössler, R. Müller, "Electrically Erasable and Reprogrammable Read-Only Memory using the n-channel SIMOS One-Transistor Cell", IEEE Trans. Electr. Dev., vol. ED-24, 806, 1977.
- [Harari'78] E. Harari, L. Schmitz, B. Troutman, S. Wang, "A 256 bit Non-Volatile Static RAM", IEEE ISSCC Dig. Tech. Pap., pp. 108, 1978.
- [Klein'79] R. Klein, W. Owen, R. Simko, W. Tchon, "5-V, non-volatile RAM owes it all to polysilicon", Electronics, October 11, pp. 111, 1979.
- [Johnson'80] W. Jonson, G. Perlegos, A. Renniger, G. Kuhn, T. Ranganath, « A 16kb electrically erasable non-volatile memory », IEEE ISSCC Dig. Tech. Pap., pp.152, 1980.
- [Landers'80] G. Landers, "5-V only EEPROM mimics static RAM timing", Electronics, june 30, p. 127, 1980.
- [Shiner'80] R. Shiner, J. Caywood, B. Euzent, IEEE IRPS, 238, 1980.
- [Drori'81] J. Drori, S. Jewell-larsen, R. Klein, W. Owen, "A single 5V supply non-volatile static RAM", IEEE ISSCC Dig. Tech. Pap., pp. 148, 1981.
- [Yaron'82] G. Yaron, S. Prasad, M. Ebel, B. Leong, "A 16k EEPROM employing new array architecture and designed-in reliability features", IEEE J. Sol. St. Circuit, vol. SC-17, no. 5, pp. 833, 1982.
- [Nozawa'82] H. Nozawa et al, Jap. Journ. Appl. Phys., 21, L111, 1982.
- [Mielke'83] N. Mielke, IEEE IRPS, 106, 1983.
- [Suzki'83] E. Suzki, H. Hiraishi, K. Ishi, Y. Hayashi, "A low voltage alterable EEPROM with metal-oxide-nitride-oxide-semiconductor (MONOS) structure", IEEE Trans. Elec. Devices, vol. ED-30, pp. 122, 1983.

- [Libsch'86] F. R. Libsch, A. Roy, M. H. White, "Amphoteric trap modelling of multidielectric scaled SONOS non-volatile memory structures", 8th NVSM, Vail, Colo., 1986.
- [Chao'87] C. C. Chao, M. H. White, "Characterization of charge injection and trapping in scaled SONOS/MNOS memory devices", Sol. St. Electr, vol. 30, pp. 307, 1987.
- [Dellin'87] T. A. Dellin, P. J. Mc Whorter, "Scaling of MONOS non-volatile memory transistors", Proc. ECS, vol. 87-10, pp. 3, 1987.
- [Baibich'88] M. Baibich, J. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, J. Chazelas, Phys. Rev. Lett. 61 (1988) 2472.
- [Haddad'89] Haddad et al., "Degradation due to Hole Trapping in Flash Memory Cell", IEEE Electron Device Letter, Volume 10, N° 3, pp. 117-119, Mars 1989.
- [Maes'89] H.E. Maes, G. Groeseneken, H. Lebon, J. Witters, "Trends in semiconductor memories", Microelectr. J., vol. 20, pp. 9, 1989.
- [Aritome'90] Aritome et al., "Extended Data Retention Characteristics After More Than 10⁴ Write And Erase Cycles in EEPROMs", Proc. IEEE Reliability Physics Symposium, pp. 259-264, 1990.
- [Pan'90] C. S. Pan et al., "A Scaling Methodology for Oxide-Nitride-Oxide Interpoly Dielectric for EEPROM Applications", IEEE Trans. On El. Dev., 37, 6, 1439, 1990.
- [Crisenza'91] G. Crisenza, G. Ghidini, M. Tosi, "Floating Gate Memories Reliability", Proc. of ESREF, 505, 1981.
- [Mori'91] S. Mori et al., IEEE IRPS, 175, 1991.
- [Pan'91a] C. S. Pan et al., "Physical Origin of Long Term Charge Loss in Floating Gate EPROM with an Interpoly Oxide-Nitride-Oxide Stacked Dielectric", IEEE El. Dev. Letters, 12, 9, 506, 1991.
- [Pan'91b] C. S. Pan et al., "High Temperature Charge Loss Mechanism in a Floating Gate EPROM with an Oxide-Nitride-Oxide (ONO) Interpoly Stacked Dielectric", IEEE El. Dev. Letters, 12, 9, 506, 1991.
- [Yamada'91] N. Yamada, E. Ohno, K. Nishiuchi, N. Akahira, M. Takao, "Rapid-phase transitions of GeTe-Sb₂Te₃ pseudobinary amorphous thin films for an optical disk memory", Journal of Applied Physics, vol. 69, n°5, pp. 2849-2857, 1991.
- [Mazoyer'92] Pascale Mazoyer, "Analyse et Caractérisation des Mécanismes de Perte de Charge Relatifs aux Diélectriques Multicouches du Point Mémoire", Thèse de Doctorat, Université Joseph Fourier, Grenoble, 1992.

- [Diepert'93]** B. Diepert, M. Levy, "Designing with Flash Memory", Annabooks, 1993.
- [Yamada'93]** Yamada et al., "Degradation Mechanism Of Flash EEPROM Programming After Program/Erase Cycles", IEDM Technical Digest, pp. 23-26, 1993.
- [Cappelletti'94]** Cappelletti et al., "Failure Mechanisms of Flash Cell in Program/Erase Cycling", IEDM Technical Digest, pp. 291-294, 1994.
- [Hermann'95]** M. Hermann, A. Schenk, "Field and High Temperature Dependence of the Long Term Charge Loss in Erasable Programmable Read Only Memories: Measurements and Modeling", J. Appl. Phys., 77, 9, 4522, 1995.
- [Pananakakis'95]** G. Pananakakis et al, "Temperature Dependence of the Fowler-Nordheim Current in Metal-Oxide-Degenerate Semiconductor Structures", J. Appl. Phys., 78, 4, 235, 1995.
- [Papadas'95]** C. Papadas et al, "Modeling of the Intrinsic Retention Characteristics of FLOTOX EEPROM Cells under Elevated Temperature Conditions", IEEE Trans. on El. Dev., 42, 4, 678, 1995.
- [De Graaf'96]** C. De Graaf et al., "A novel high-density low cost diode programmable read only memory", in proc. IEEE International Electron Devices Meeting (IEDM) 1996, pp.189.
- [Candelier'97]** Ph. Candelier, "Contribution à l'amélioration de la fiabilité des mémoires non-volatiles de type FLASH EEPROM", thèse de doctorat de l'Université Joseph Fourier de Grenoble, novembre 1997.
- [Brown'98]** W. D. Brown, J.E. Brewer, "NonVolatile semiconductor Memory Technology", IEEE PRESS, New York, 1998.
- [NVSMT'98]** "Non Volatile Semiconductor Memory Technology: a Comprehensive Guide to Understanding and Using NVSM Devices", edited by W.D. Brown, J.E. Brewer, IEEE Press, 1998. **[NVSMT98-Ch6]** Chapter 6, "Reliability and NVSM Reliability", Y.Hsia, V.C. Tyree.
- [De Salvo'99]** B. De Salvo, "Etude du transport électrique et de la fiabilité des isolants des mémoires non-volatiles à grille flottante", thèse de doctorat de l'Institut National Polytechnique de Grenoble, janvier 1999.
- [Parkin'99]** SPP Parkin, Roche KP, Samant MG, et al. J Appl Phys 1999; 85 :5828
- [Mahapatra'00]** Mahapatra et al., "A Comprehensive Study Of Hot-Carrier Induced Interface And Oxide Trap Distributions In MOSFET's using a Novel charge Pumping Technique", IEEE Transaction On Electron Devices, Volume 47, N° 1, pp. 171-177, Janvier 2000.

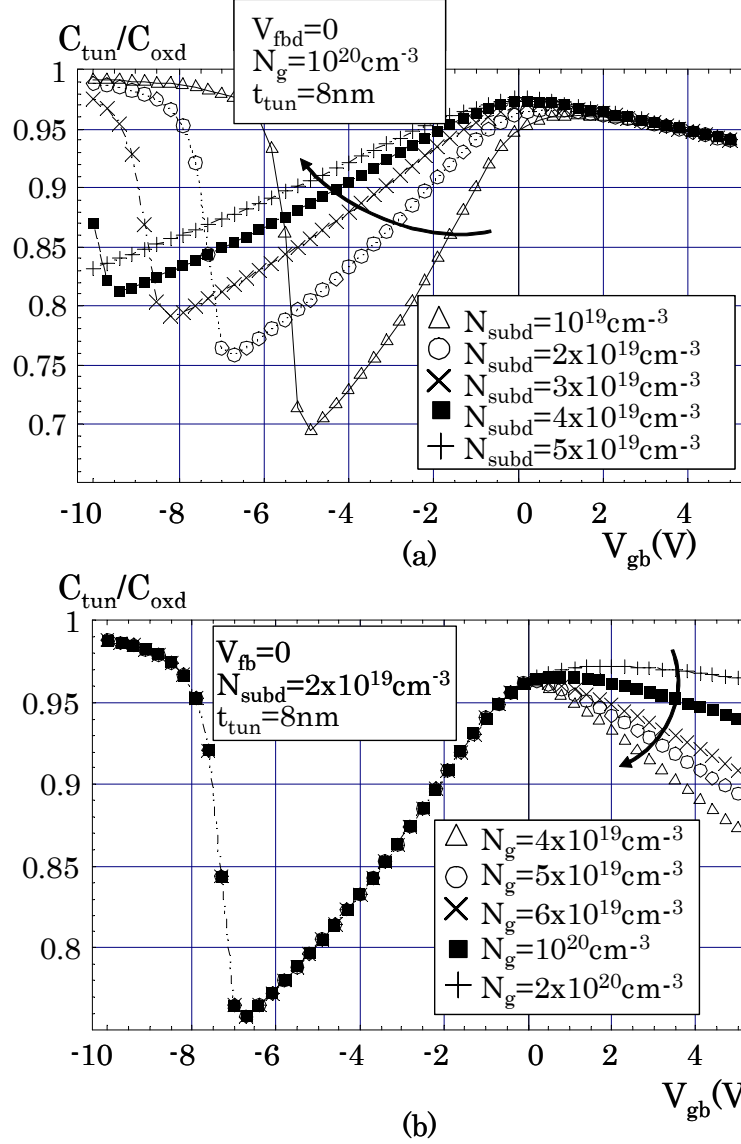
[Lai'01] S. Lai, Tyler Lowrey, « OUM-A 180 nm Nonvolatile Memory Cell Element Technology For Stand Alone and Embedded Applications », Session 36.5, IEDM Technical Digest, 2001.

[Gill'02] M. Gill, Tyler Lowrey, J. Park, “Ovonic Unified Memory-A High-performance Non volatile Memory Technology for Stand Alone Memory and Embedded Applications”, TD 12.4, Proceedings of ISSCC, 2002.

[Hwang'03] Y.N. Hwang, J.S. Hong, S.H. Lee, S.J. Ahn, G.T. Jeong, G.H. Koh, J.H. Oh, H.J. Kim, W.C. Jeong, S.Y. Lee, J.H. Park, K.C. Ryoo, H. Horii, Y.H. Ha, J.H. Yi, W.Y. Cho, Y.T. Kim, K.H. Lee, S.H. Joo, S.O. Park, U.I. Chung, H.S. Jeong, Kinam Kim, “Full Integration and Reliability Evaluation of Phase-Change RAM Based on 0.24 μm -CMOS Technologies”, Symposium on VLSI Tech Digest of Tech Papers, 2003.

[Koh'04] G.H. Koh, “Fabrication of High Performance 64kb MRAM”, Journal of Magnetism and Magnetic Materials, 272-276 (2004), 1941-1942.

Chapitre II : Modélisation de la cellule mémoire EEPROM



Caractéristique $C_{tun}(V_{gb})$ d'une structure SOS Si-Poly/SiO₂/Si-N+ pour différents dopages de substrat (a) et de grille en polysilicium (b).

II.1 EFFET TUNNEL DANS LES OXYDES MINCES.....	53
II.1.1 Transparence tunnel.....	53
II.1.2 Effet tunnel dans les structures MOS et SOS.....	55
II.1.3 Barrière triangulaire : mécanisme d'injection FN	56
II.2 MODELISATION DE LA CELLULE EEPROM	58
II.2.1 Modélisation du transistor d'état	58
II.2.1.1 Modèle statique du courant de drain.....	58
II.2.1.2 Modèle dynamique.....	60
II.2.2 Modélisation de la capacité de la zone tunnel	62
II.2.2.1 Capacité constante	62
II.2.2.2 Prise en compte de la capacité de substrat	62
II.2.2.3 Prise en compte de la désertion de grille flottante.....	64
II.2.2.4 Comparatif.....	67
II.2.3 Modélisation du transistor de sélection	68
II.2.4 Bilan des paramètres du modèle	69
II.3 EXTRACTION DES PARAMETRES.....	71
II.3.1 Méthodologie.....	71
II.3.2 Extraction de α et β	73
II.3.2.1 Méthode générale	73
II.3.2.2 Approche de Croci et al.	74
II.3.2.3 Analyse de Chiou et al.....	75
II.3.3 Mesures $C_{tun}(V_{gb})$ et $I_{FN}(V_{gb})$	77
II.4 SIMULATIONS	79
II.4.1 Implantation du modèle dans ELDO	79
II.4.2 Simulations statiques	79
II.4.3 Simulations dynamiques	82
II.5 BANC DE MESURE.....	86
II.5.1 Caractérisation d'une structure Si-Poly/SiO ₂ /Si-N+	86
II.5.2 Ecriture, effacement et lecture de cellules EEPROM	87
II.5.3 Mesure de V_{th} et validation du modèle	88
Références bibliographiques du chapitre II	90

La modélisation des composants électroniques est essentielle car elle permet de prévoir les comportements des circuits dont ces composants sont les briques de base. Ainsi, lorsque l'on arrive à modéliser tous les composants d'un circuit donné, on peut obtenir une estimation fiable de sa consommation, du courant qui le traverse ou de son échauffement en fonction des tensions à ses bornes, par exemple.

Par ailleurs, la réduction des dimensions des composants électroniques fait intervenir de nouveaux phénomènes physiques (effet tunnel direct, confinement quantique des porteurs dans le canal, polydépuration...) dont la compréhension et la prise en compte sont indispensables afin de prévoir de manière fiable le comportement des circuits électroniques qui intégreront ces composants ultimes. Dans ce contexte, l'activité de modélisation des composants électroniques devient réellement indispensable et doit relever des défis de plus en plus difficiles :

Les modèles développés doivent être très précis : ceci n'est possible qu'en prenant en compte tous les phénomènes qui ont lieu au sein du composant à modéliser. Dans cette optique, ces modèles doivent être basés sur la physique plutôt que sur des équations empiriques.

Le modèle doit être convergent pour être exploitable. Il faut envisager préalablement toutes les conditions de fonctionnement du composant, écrire les équations correspondantes et connaître leur domaine de validité afin qu'elles aient un sens.

Enfin, le calcul doit être très rapide. Ceci nécessite l'écriture du modèle dans un langage proche du code machine (C) et son implantation dans un simulateur performant (Eldo, Saber...).

Dans le cadre de la modélisation des cellules EEPROM, la modélisation du transistor MOS à grille flottante est primordiale car l'information stockée est représentée par la quantité de charges emmagasinées dans la grille flottante de ce transistor. Des modèles de type SPICE [Foty'97] existent mais ils sont basés sur des équations dont les dérivées successives sont discontinues à la limite des zones de fonctionnement. De telles discontinuités provoquent des divergences et de grandes imprécisions. Par exemple, après une phase d'effacement, le substrat de type P du transistor d'état se trouve en accumulation en raison du champ électrique induit par la grille flottante chargée négativement. Ce comportement n'est pas pris en compte dans les modèles de type SPICE.

Afin de pallier à ce manque, un modèle physique compact de cellule EEPROM a été développé au sein du laboratoire. Ce modèle est basé sur le calcul du potentiel de grille flottante et du potentiel de surface dans le canal du transistor d'état. Ce modèle étant physique, toutes les fonctions qui décrivent les grandeurs électriques sont continues au premier et au second ordre à la limite des zones de fonctionnement. Ce modèle que nous avons utilisé pour toutes nos simulations a été développé par R. Bouchakour et N. Harabech pendant sa thèse.

Le fonctionnement des EEPROM repose essentiellement sur l'injection Fowler-Nordheim, qui s'inscrit dans le cadre plus vaste du transport par effet tunnel de porteurs de charge à travers un isolant. Ainsi, dans un premier temps, nous décrirons les bases physiques sur lesquelles repose l'injection tunnel Fowler-

Nordheim. Nous allons, notamment, établir l'expression classique du courant Fowler-Nordheim.

Ensuite, nous traiterons de la modélisation de la cellule mémoire EEPROM qui repose sur le traitement séparé du transistor d'état et du transistor de sélection. Nous rentrerons plus en détail en décrivant la modélisation du transistor d'état en régime statique et en régime transitoire. Toutes les équations qui décrivent les grandeurs électriques du transistor d'état (potentiel de grille flottante, champ électrique, courant Fowler-Nordheim transitoire, courant drain-source etc...) sont établies. En particulier, nous prenons en compte la désertion de la grille flottante lorsque la zone tunnel est en régime d'accumulation. Nous présenterons une méthodologie d'extraction des paramètres du modèle afin de pouvoir simuler une opération d'écriture ou d'effacement.

Dans la troisième partie, nous présenterons les résultats de la simulation d'une cellule EEPROM en régime statique et en régime dynamique.

Enfin, la quatrième partie présentera un élément important de ce travail de thèse : le banc de mesure. Celui-ci est constitué de l'ensemble des appareils qui nous ont permis de caractériser les structures à étudier, d'écrire et d'effacer une cellule EEPROM et d'en mesurer la tension de seuil, dans l'objectif de valider notre modèle.

II.1 Effet tunnel dans les oxydes minces

L'effet tunnel correspond à la possibilité, pour une particule donnée, de franchir une barrière de potentiel plus élevée que son énergie potentielle propre.

Prenons le cas d'une particule qui se déplace à une vitesse v et qui possède une énergie potentielle E . La mécanique quantique nous enseigne qu'à toute particule, on associe une onde dont les caractéristiques sont déterminées par la fonction d'onde solution de l'équation de Schrödinger associée à la particule. Ainsi, lorsque cette particule rencontre une barrière de potentiel de largeur L et de hauteur Φ_0 supérieure à E , sa nature ondulatoire se manifeste et deux possibilités se présentent :

i.

La particule est réfléchiée,

ii.

La particule traverse la barrière de potentiel.

Ce passage à travers la barrière, appelée « effet tunnel », est possible en mécanique quantique. Ce phénomène est observé avec une probabilité mesurable, que nous appellerons transparence $T(E)$, et qui dépend de la forme de la barrière et du matériau qui constitue cette dernière.

II.1.1 Transparence tunnel

Considérons une barrière de potentiel, de forme quelconque, de largeur L , définie par le potentiel $V(x)$, et un électron incident d'énergie potentielle E qui se déplace vers les abscisses positives (fig. II.1).

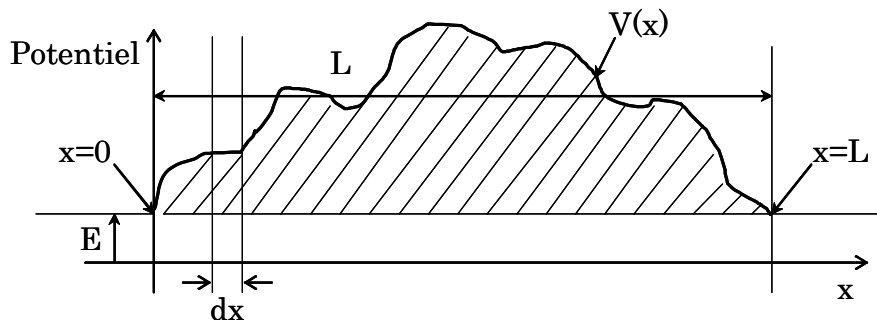


Figure II. 1 : Barrière de potentiel quelconque de largeur L .

Par définition, la transparence $T(E)$ de la barrière est la probabilité de trouver l'électron incident en $x=L$. Par conséquent, si on appelle Ψ la fonction d'onde associée à l'électron, Ψ^* sa complexe conjuguée, et si l'entrée de la barrière est située en $x=0$ et sa sortie en $x=L$, il vient :

$$T(E) = \frac{\Xi(L) \times \Xi^*(L)}{\Xi(0) \times \Xi^*(0)} \quad (\text{II. 01})$$

Cette relation, qui ne tient pas compte des multiples réflexions aux interfaces (entrée et sortie de la barrière), est appelée approximation WKB (Wentzel-Kramers-Brillouin).

$T(E)$ se calcule en résolvant l'équation de Schrödinger unidimensionnelle en régime stationnaire :

$$-\frac{\hbar^2}{2m^*} \frac{d^2 \Xi}{dx^2} + V(x)\Xi = E\Xi \quad (\text{II. 02})$$

où m^* est la masse effective de l'électron dans le milieu constituant la barrière et $\hbar$ la constante de Planck réduite.

La relation (II.02) peut encore s'écrire :

$$\frac{d^2 \Xi}{dx^2} = \frac{2m^* [V(x) - E]}{\hbar^2} \Xi \quad (\text{II. 03})$$

En découpant la barrière en petites tranches de longueur dx , on peut considérer que la quantité $[V(x) - E]$ est constante entre x et $x+dx$. Dans ce cas, l'équation différentielle (II.03) admet pour solution la fonction d'onde Ξ telle que :

$$\Xi(x) = A \exp(k(x)) + B \exp(-k(x)) \quad (\text{II. 04})$$

avec :

$$k(x) = \frac{\sqrt{2m^* (V(x) - E)}}{\hbar} \quad (\text{II. 05})$$

D'autre part, si la barrière de potentiel était de largeur infinie, on conçoit, physiquement, que l'électron n'arrive pas à la franchir et que la fonction d'onde doit s'annuler en $x=+\infty$, ce qui élimine le terme en exponentielle positive. Par conséquent, entre x et $x+dx$, nous avons :

$$\Xi(x) = B \exp(-k(x)) \quad (\text{II. 06})$$

ou encore :

$$\frac{\Xi(x + dx)}{\Xi(x)} = \exp(-k(x)dx) \quad (\text{II. 07})$$

En intégrant la relation ci-dessus, l'amplitude de la fonction d'onde en $x=L$ peut être reliée à son amplitude en $x=0$ de la manière suivante :

$$\Xi(L) = \Xi(0) \exp\left(-\int_0^L \frac{\sqrt{2m^* [V(x) - E]}}{\hbar} dx\right) \quad (\text{II. 08})$$

Ainsi, en combinant les relations (II.01) et (II.08), la transparence $T(E)$ s'écrit :

$$T(E) = \exp\left(-2 \int_0^L \frac{\sqrt{2m^*}}{\hbar} (\sqrt{V(x) - E}) dx\right) \quad (\text{II. 09})$$

II.1.2 Effet tunnel dans les structures MOS et SOS

Dans le cas particulier des structures MOS, l'effet tunnel consiste en un transit de porteurs à travers l'oxyde de grille. En effet, lorsque des porteurs arrivent à l'interface silicium/oxyde, il existe une probabilité non nulle pour qu'ils franchissent la barrière de potentiel constituée par l'oxyde sans devoir acquérir d'énergie supplémentaire.

Pour ces structures, l'énergie potentielle est l'énergie du bas de la bande de conduction du silicium ou de la bande de conduction du métal suivant que les électrons transitent respectivement depuis le silicium vers le métal ou du métal vers le silicium. Rappelons, à ce propos, que le bas de la bande de conduction est confondu avec le niveau de Fermi pour les métaux ($E_F = E_C$).

Suivant la forme de la barrière, la transparence donne lieu à deux types de courants tunnels : le courant tunnel direct pour une barrière trapézoïdale (fig. II.2a) et le courant Fowler-Nordheim pour une barrière triangulaire (fig. II.2b).

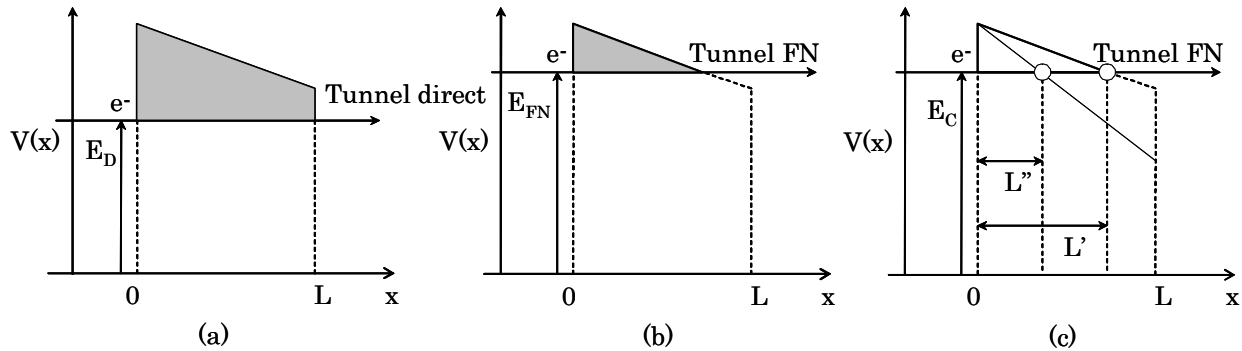


Figure II. 2 : Barrière, (a) trapézoïdale, (b) triangulaire, (c) en augmentant la polarisation, vue par les électrons lors de la traversée de l'oxyde de grille d'une structure MOS ou SOS.

Le type de barrière et par conséquent le type de courant qui en résulte dépendent non seulement du niveau de polarisation appliquée à la grille mais aussi de l'épaisseur de l'oxyde (L).

Quelle que soit l'épaisseur de l'oxyde, si la tension de polarisation est suffisamment élevée, la forme de la barrière vue par la particule sera triangulaire et on obtiendra un courant de type Fowler-Nordheim.

L'abaissement de la barrière en raison d'une polarisation extérieure favorise le passage des électrons par effet tunnel Fowler-Nordheim dans le cas des oxydes épais ($L > 10\text{nm}$).

L'abaissement de cette barrière conduit aussi à une diminution de la largeur effective (fig. II.2c) de la barrière de potentiel que doit franchir la particule et donc à une augmentation de la transparence de la barrière ((II.09)).

II.1.3 Barrière triangulaire : mécanisme d'injection FN

L'injection de charges par effet tunnel Fowler Nordheim à travers un oxyde mince a été mis en évidence expérimentalement par Lenzlinger en 1969 [Lenzlinger'69]. Une bonne connaissance de ce mécanisme est fondamentale car il est la base physique sur laquelle repose le fonctionnement des mémoires EEPROM. Dans les paragraphes suivants, nous donnons un aperçu des éléments théoriques qui conduisent à l'expression de la densité de courant Fowler-Nordheim à travers un diélectrique (SiO_2) soumis à un fort champ électrique. Nous établirons l'expression très connue de la densité de courant Fowler-Nordheim en nous situant dans le cadre de la physique classique.

La densité de courant Fowler-Nordheim J_{FN} peut s'exprimer comme le nombre de porteurs par unité de surface et de temps arrivant à l'interface grille/oxyde (ou substrat/oxyde dans le cas d'une injection à partir du substrat).

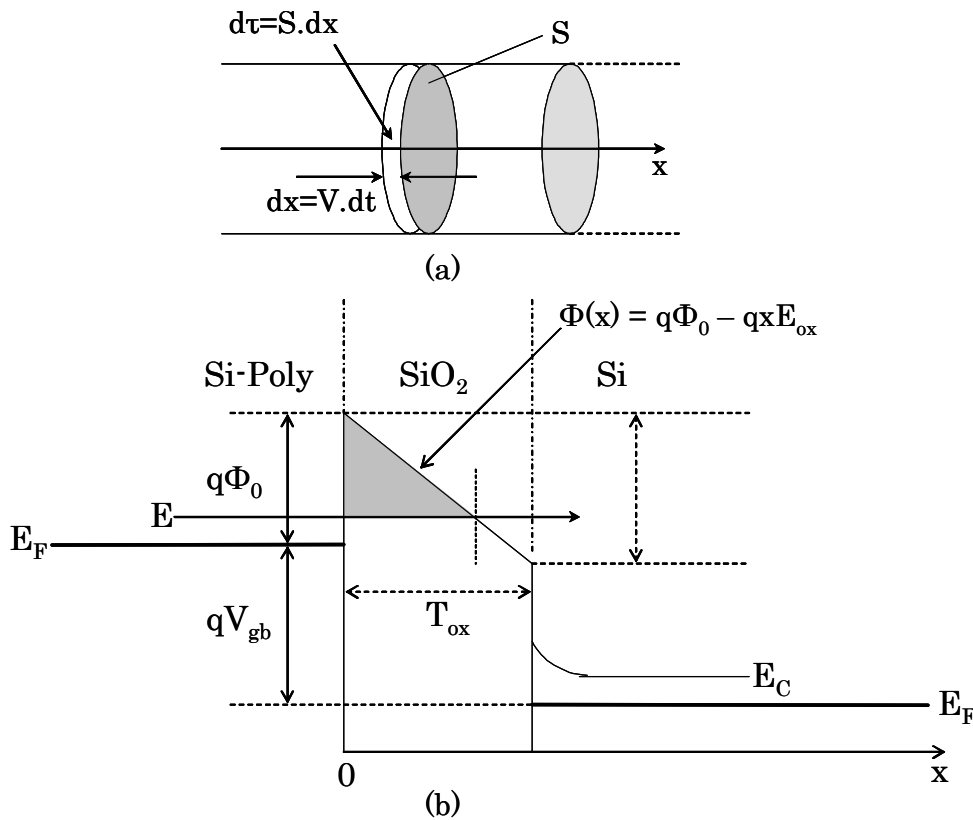


Figure II. 3 : *Diagramme des bandes d'énergie en injection Fowler-Nordheim à partir de la grille Si-Poly.*

Considérons la figure II.3a. Si on appelle $n(V)$ (unité : s.m^{-4}) la densité des porteurs disponibles à l'interface grille/oxyde par unité de volume et par unité de vitesse, le nombre $d^2N_1(V)$ de porteurs, de vitesse comprise entre V et $V+dV$, qui vont traverser la section S pendant l'intervalle de temps compris entre t et $t+dt$, est le nombre de porteurs dans le cylindre de section S , de hauteur dx et de volume $d\tau = Sdx = SVdt$:

$$d^2N_1(V) = n(V)dV \times SVdt \quad (\text{II. 10})$$

Le nombre $d^2N_2(V)$ de porteurs ayant franchi la barrière est égal à $d^2N_1(V)$ multiplié par la probabilité $P(V)$ de traverser la barrière pour un porteur de vitesse V .

$$d^2N_2(V) = d^2N_1(V) \times P(V) = n(V)P(V)VdV \times S dt \quad (\text{II. 11})$$

Par conséquent, la densité de courant élémentaire $dJ_{FN}(V)$ vaut :

$$dJ_{FN}(V) = q \times \frac{1}{S} \frac{d^2N_2(V)}{dt} = q \times n(V)P(V)VdV \quad (\text{II. 12})$$

Ainsi, si on appelle V_m la vitesse maximale des porteurs disponibles à l'interface grille/oxyde, la densité totale de courant est la somme sur toutes les vitesses disponibles des densités de courant élémentaires :

$$J_{FN} = \int dJ_{FN}(V) = q \int_0^{V_m} n(V)P(V)VdV \quad (\text{II. 13})$$

ou, en sommant sur les énergies :

$$J_{FN} = \int_0^{E_m} dJ_{FN}(E) \quad (\text{II. 14})$$

où E_m est l'énergie maximale des porteurs disponibles.

$dJ_{FN}(E)$ se calcule à partir de $dJ_{FN}(V)$ en opérant le changement de variable suivant :

$$E = \frac{mV^2}{2} \quad (\text{II. 15})$$

Enfin, pour une barrière de potentiel triangulaire (fig. III.3b), le potentiel $\Phi(x)$ résultant d'un champ électrique E_{ox} à travers l'oxyde est égal à :

$$\Phi(x) = q(\Phi_0 - xE_{ox}) \quad (\text{II. 16})$$

Une telle barrière de potentiel conduit à la transparence $T(E)$ suivante pour un électron d'énergie potentielle E :

$$T(E) = \exp \left[-\frac{8\pi}{3h} \sqrt{2m_{ox,BC}^* (q\Phi_0 - E)} \frac{1}{E_{ox}} \right] \quad (\text{II. 17})$$

En injectant dans la relation (II.14) la relation (II.17), on obtient une intégrale qui ne s'exprime pas avec les fonctions usuelles, qui, à l'aide d'un développement en série de Taylor [Plossu'00], conduit à l'expression classique de la densité de courant Fowler-Nordheim :

$$J_{FN} = \alpha E_{ox}^2 \exp \left(-\frac{\beta}{E_{ox}} \right) \quad (\text{II. 18})$$

avec

$$\alpha = \frac{q^3}{8\pi h q \Phi_0} \frac{m_{Si,BC}}{m_{ox,BC}^*} \quad (\text{II. 19})$$

et

$$\beta = \frac{4}{3} \frac{\sqrt{2m_{ox,BC}^*}}{q\hbar} (q\Phi_0)^{\frac{3}{2}} \quad (\text{II. 20})$$

II.2 Modélisation de la cellule EEPROM

La modélisation du point-mémoire EEPROM est scindée en deux parties : celle du transistor d'état et celle du transistor de sélection. La modélisation du transistor d'état est physique car elle repose sur le calcul du potentiel de grille flottante et du potentiel de surface dans le canal.

II.2.1 Modélisation du transistor d'état

II.2.1.1 Modèle statique du courant de drain

Le courant de drain dans le canal s'exprime comme la somme de deux composantes (courant de dérive et courant de diffusion) [Tsividis'89][Klaassen'90] :

$$I_{ds} = I_{dérive} + I_{diffusion} \quad (\text{II. 21})$$

avec :

$$I_{dérive} = -W\mu_n Q_n' \frac{\partial \Psi_s}{\partial y} \quad (\text{II. 22})$$

$$I_{diffusion} = W\mu_n U_t \frac{\partial Q_n'}{\partial y} \quad (\text{II. 23})$$

où W est la largeur du canal du transistor d'état, μ_n la mobilité des électrons, U_t le potentiel électrothermique, Q_n' la densité surfacique des électrons dans le canal et Ψ_s le potentiel de surface. L'équation du générateur de courant de drain s'écrit alors :

$$I_{ds} = -W\mu_n C'_{ox} \left[\left(-\frac{Q_n'}{C'_{ox}} \right) - U_t \frac{\partial}{\partial \Psi_s} \left(-\frac{Q_n'}{C'_{ox}} \right) \right] \frac{\partial \Psi_s}{\partial y} \quad (\text{II. 24})$$

où C'_{ox} est la capacité par unité de surface de l'oxyde de grille. L'expression du courant I_{ds} est obtenue en intégrant cette dernière relation le long du canal [Bouchakour'96][Bouchakour'97][Bouchakour'99] :

$$I_{ds} \int_0^{L_c} dy = -W\mu_n C'_{ox} \int_{\Psi_{ss}}^{\Psi_{sd}} \left[\left(-\frac{Q_n'}{C'_{ox}} \right) - U_t \frac{\partial}{\partial \Psi_s} \left(-\frac{Q_n'}{C'_{ox}} \right) \right] d\Psi_s \quad (\text{II. 25})$$

où Ψ_{ss} et Ψ_{sd} sont respectivement les potentiels de surface à la source et au drain et L_c la longueur effective du canal. Le calcul du courant nécessite l'évaluation de

la densité de la charge d'inversion Q'_n en fonction du potentiel de surface Ψ_s . Ceci est possible en utilisant la loi de conservation de la charge dans la structure qui permet d'écrire :

$$Q'_{sc} = -Q'_g = Q'_n + Q'_b \quad (\text{II. 26})$$

où Q'_{sc} , Q'_n et Q'_b sont respectivement la densité surfacique de charge dans le substrat, dans la grille flottante et dans la zone de charge d'espace du substrat semi-conducteur.

Q'_g s'écrit en fonction de Ψ_s et de V_{fg} comme suit :

$$Q'_g = C'_{ox} (V_{fg} - V_{fb} - \Psi_s) \quad (\text{II. 27})$$

où V_{fb} est la tension de bandes plates. En supposant que le dopage est abrupt et homogène dans le substrat, Q'_b s'écrit comme suit :

$$Q'_b = \gamma C'_{ox} \sqrt{\Psi_s - U_t} \quad (\text{II. 28})$$

où γ est le coefficient d'effet substrat. La charge Q'_n s'écrit alors en fonction de Ψ_s et de V_{fg} :

$$Q'_n = -C'_{ox} (V_{fg} - V_{fb} - \Psi_s) - \gamma C'_{ox} \sqrt{\Psi_s - U_t} \quad (\text{II. 29})$$

Dans la relation (II.25) donnant le courant de drain, seuls les potentiels de surface à la source (Ψ_{ss}) et au drain (Ψ_{sd}) sont inconnus. Nous allons les déterminer en utilisant la conservation de la charge exprimée par la relation (II.26) et l'expression de la charge dans le substrat semi-conducteur.

En résolvant l'équation de Poisson dans le substrat semi-conducteur, l'équation donnant la densité surfacique de charge Q'_{sc} dans le substrat semi-conducteur peut se mettre sous la forme simplifiée suivante :

$$Q'_{sc} = \gamma C'_{ox} \sqrt{\Psi_s - U_t + U_t \exp\left(\frac{\Psi_s - 2\phi_b - V_y}{U_t}\right) + U_t \exp\left(\frac{-\Psi_s}{U_t}\right)} \quad (\text{II. 30})$$

Rappelons que Ψ_s et V_y sont respectivement les potentiels de surface et de canal. La charge du substrat, donnée par la relation (II.30), est continue dans tous les régimes de fonctionnement.

En utilisant les relations (II.26), (II.27) et (II.30), nous obtenons l'équation implicite (II.31) reliant le potentiel de surface Ψ_s au potentiel de grille flottante V_{fg} :

$$V_{fg} - \Psi_s - V_{fb} - \gamma \sqrt{\Psi_s + U_t \left(\exp\left(\frac{\Psi_s - 2\phi_b - V_y}{U_t}\right) + \exp\left(\frac{-\Psi_s}{U_t}\right) - 1 \right)} = 0 \quad (\text{II. 31})$$

La résolution de l'équation (II.31) permet d'évaluer le potentiel de surface à la source, Ψ_{ss} et Ψ_{sd} , au drain, dans tous les régimes de fonctionnement (accumulation, désertion, faible et forte inversions).

En combinant les relations (II.25) et (II.29), le courant de drain I_{ds} s'écrit :

$$I_{ds} = \frac{\mu_n W C'_{ox}}{L_c} [F(\Psi_{sd}) - F(\Psi_{ss})] \quad (\text{II. 32})$$

Avec

$$F(\Psi_s) = (V_{fg} - V_{fb} + U_t) \Psi_s - \frac{\Psi_s^2}{2} - \frac{2}{3} \gamma \sqrt{\Psi_s - U_t} \left(\Psi_s - \frac{5}{2} U_t \right) \quad (\text{II. 33})$$

Notons que les potentiels de surface Ψ_{sd} et Ψ_{ss} sont déterminés en résolvant l'équation (II.31) et en remplaçant Ψ_s et V_y respectivement par Ψ_{sd} , V_d et par Ψ_{ss} et V_s .

II.2.1.2 Modèle dynamique

La modélisation dynamique du transistor d'état a pour objectif de prédire les variations en fonction du temps des principales grandeurs électriques de la cellule EEPROM. D'un point de vue physique, la grandeur principale qui caractérise le transistor d'état est la charge emmagasinée dans la grille flottante. En effet, celle-ci est l'image de l'information qui est stockée dans la cellule. Cette charge stockée dépend du courant d'injection Fowler–Nordheim à travers l'oxyde tunnel, lui-même fonction de la polarisation des électrodes de la cellule. Par conséquent, le modèle dynamique du transistor d'état doit permettre, connaissant la polarisation de la grille de contrôle, du drain, de la source et du substrat en fonction du temps, de déterminer les variations en fonction du temps de toutes les autres grandeurs électriques de la cellule.

D'un point de vue électrique, il est plus convenable d'utiliser le potentiel de grille flottante comme grandeur de référence car sa connaissance permet ensuite de calculer le courant d'injection Fowler–Nordheim. Pour cette raison, nous allons exposer en premier lieu le calcul du potentiel de grille flottante V_{fg} . Ensuite, nous décrirons la démarche suivie pour déterminer le courant d'injection Fowler–Nordheim, la charge de grille flottante et la tension de seuil de la cellule en fonction du temps.

La figure II.2 est une représentation simplifiée de la grille flottante d'une cellule EEPROM.

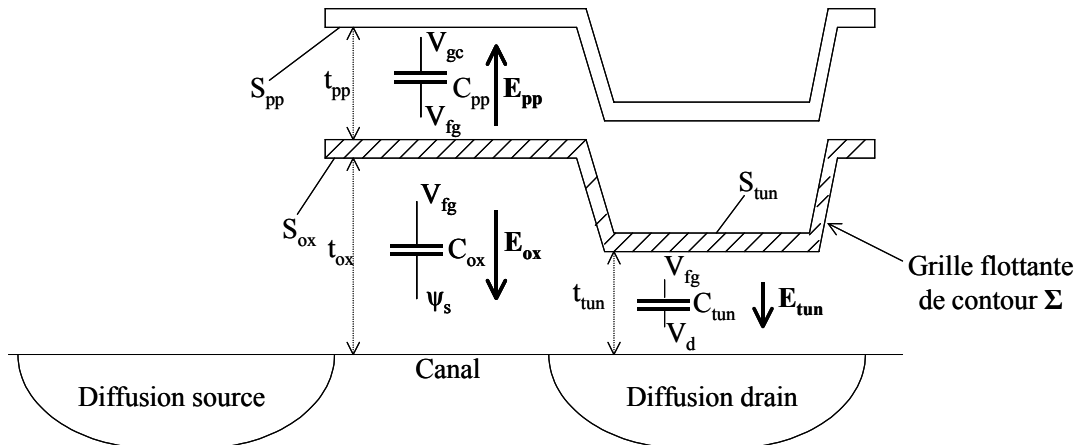


Figure II. 4 : Grille flottante et surface d'application du théorème de Gauss.

En appliquant le théorème de Gauss sur la surface fermée Σ qui s'appuie sur le contour de la grille flottante, il vient :

$$\oiint_{\Sigma} \mathbf{E} d\mathbf{S} = \frac{Q_{fg}}{\epsilon_0 \epsilon_{SiO_2}} \quad (\text{II. 34})$$

où $\mathbf{E}$ est le champ électrique, Σ la surface fermée d'application du théorème de Gauss et Q_{fg} la charge contenue dans la grille flottante. En explicitant les différents termes de l'intégrale surfacique, la relation (II.34) devient :

$$E_{pp} S_{pp} + E_{ox} S_{ox} + E_{tun} S_{tun} = \frac{Q_{fg}}{\epsilon_0 \epsilon_{SiO_2}} \quad (\text{II. 35})$$

où E_{pp} est le champ électrique à travers la couche ONO, E_{ox} le champ électrique à travers l'oxyde situé au-dessus du canal et E_{tun} le champ électrique à travers l'oxyde tunnel. En remplaçant ces champs électriques "sortants" par leurs expressions, il vient :

$$\frac{V_{fg} - V_{gc}}{t_{pp}} S_{pp} + \frac{V_{fg} - V_{fb} - \Psi_s}{t_{ox}} S_{ox} + \frac{V_{fg} - V_d}{t_{tun}} S_{tun} = \frac{Q_{fg}}{\epsilon_0 \epsilon_{SiO_2}} \quad (\text{II. 36})$$

ou encore :

$$(V_{fg} - V_{gc}) C_{pp} + (V_{fg} - V_{fb} - \Psi_s) C_{ox} + (V_{fg} - V_d) C_{tun} = Q_{fg} \quad (\text{II. 37})$$

En regroupant les termes qui contiennent le potentiel de grille flottante V_{fg} , il vient :

$$(C_{pp} + C_{ox} + C_{tun}) V_{fg} = C_{pp} V_{gc} + C_{ox} (V_{fb} + \Psi_s) + C_{tun} V_d + Q_{fg} \quad (\text{II. 38})$$

Soit :

$$V_{fg} = \frac{1}{C_{pp} + C_{ox} + C_{tun}} [C_{pp} V_{gc} + C_{ox} (V_{fb} + \Psi_s) + C_{tun} V_d + Q_{fg}] \quad (\text{II. 39})$$

Cette relation (II.39) peut encore s'écrire, de manière plus commode :

$$V_{fg} = K_e V_{gc} + \frac{C_{ox}}{C_x} (V_{fb} + \Psi_s) + \frac{C_{tun}}{C_x} V_d + \frac{Q_{fg}}{C_x} \quad (\text{II. 40})$$

avec :

$$C_x = C_{pp} + C_{ox} + C_{tun} \quad (\text{II. 41})$$

et

$$K_e = \frac{C_{pp}}{C_x} \quad (\text{II. 42})$$

La relation (II.40) est fondamentale car elle permet de calculer le potentiel de grille flottante V_{fg} en fonction des polarisations appliquées sur les électrodes de la cellule et de la charge Q_{fg} contenue dans la grille flottante. La connaissance de V_{fg} permet ensuite de calculer le champ électrique E_{tun} à travers l'oxyde tunnel et le courant Fowler–Nordheim I_{FN} .

La quantité totale de charge emmagasinée dans la grille flottante est la somme de la charge stockée initialement Q_{fgini} et de l'apport dû au courant d'injection Fowler-Nordheim I_{FN} . La charge totale de grille flottante à l'instant τ s'exprime comme suit :

$$Q_{fg}(\tau) = Q_{fgini} + \int_0^\tau I_{FN}(t)dt \quad (II. 43)$$

La tension de seuil est fonction de la charge stockée dans la grille flottante. Si on appelle V_{th0} la tension de seuil à charge de grille flottante nulle et V_{th} la tension de seuil pour une charge de grille flottante Q_{fg} donnée, il vient :

$$V_{th} = V_{th0} - \frac{Q_{fg}}{C_{pp}} \quad (II. 44)$$

II.2.2 Modélisation de la capacité de la zone tunnel

Dans le modèle dynamique exposé précédemment, nous avons vu que la capacité C_{tun} de la zone tunnel intervient partout, à partir du moment où elle est présente dans la relation (II.40) qui donne l'expression du potentiel flottant V_{fg} . Cette capacité C_{tun} n'est pas constante car la zone tunnel est une structure SOS (Silicium-Oxyde-Silicium), à grille dégénérée, qui peut se trouver en accumulation, en désertion ou en inversion, selon les polarisations appliquées sur la cellule.

Par ailleurs, le calcul de la capacité théorique de la structure tunnel est important dans la mesure où l'extraction des paramètres du modèle nécessite l'ajustement de la caractéristique $C_{tun}(V_{gb})$ expérimentale avec la caractéristique théorique.

Pour toutes ces raisons, trois modèles (EEPROM2, EEPROM3 et EEPROM4), de complexité croissante, ont été développés afin de prendre en compte les variations de la capacité de la zone tunnel en fonction des polarisations.

II.2.2.1 Capacité constante

Dans ce modèle simple (EEPROM2), la capacité de la zone tunnel est supposée constante et égale à la capacité de l'oxyde tunnel quelles que soient les polarisations de la cellule :

$$C_{tun} = \frac{\epsilon_0 \epsilon_{SiO_2}}{t_{tun}} S_{tun} \quad (II. 45)$$

II.2.2.2 Prise en compte de la capacité de substrat

La zone tunnel (fig. II.5) est constituée de la diffusion drain, de la couche d'oxyde tunnel et de la grille flottante en poly-silicium. C'est une capacité SOS dont le dopage de substrat est important. Ainsi, dans ce modèle (EEPROM3), nous supposons que la grille en poly-silicium fortement dopée se comporte comme un métal.

La capacité tunnel C_{tun} est alors équivalente à la mise en série de la capacité d'oxyde C_{oxd} et de la capacité C_d relative à la variation de Q'_d , densité surfacique de charge dans le substrat (diffusion drain) :

$$\frac{1}{C_{\text{tun}}} = \frac{1}{C_{\text{oxd}}} + \frac{1}{C_d} \quad (\text{II. 46})$$

avec :

$$C_{\text{oxd}} = \frac{\epsilon_0 \epsilon_{\text{SiO}_2}}{t_{\text{tun}}} S_{\text{tun}} \quad (\text{II. 47})$$

et

$$C_d = \frac{dQ'_d}{d\Psi} S_{\text{tun}} \quad (\text{II. 48})$$

où Ψ est le potentiel de surface à l'interface $\text{SiO}_2/\text{Si-N}^+$, S_{tun} la surface de la zone tunnel et t_{tun} l'épaisseur de l'oxyde tunnel. La densité surfacique de charge associée Q'_d s'écrit :

$$Q'_d = \gamma C'_{\text{oxd}} \sqrt{-\Psi - U_t + U_t \exp\left(\frac{-\Psi + 2\phi_{\text{bd}}}{U_t}\right) + U_t \exp\left(\frac{\Psi}{U_t}\right)} \quad (\text{II. 49})$$

où C'_{oxd} est la capacité surfacique de l'oxyde tunnel, U_t le potentiel thermique et ϕ_{bd} le potentiel de substrat Si-N^+ (diffusion drain).

Les relations (II.48) et (II.49) nous permettent de déterminer la capacité de substrat C_d :

$$C_d = S_{\text{tun}} \frac{\gamma C'_{\text{oxd}}}{2} \frac{\left| -1 - \exp\left(\frac{-\Psi + 2\phi_{\text{bd}}}{U_t}\right) + \exp\left(\frac{\Psi}{U_t}\right) \right|}{\sqrt{\left| -\Psi + U_t \exp\left(\frac{-\Psi + 2\phi_{\text{bd}}}{U_t}\right) + \exp\left(\frac{\Psi}{U_t}\right) - 1 \right|}} \quad (\text{II. 50})$$

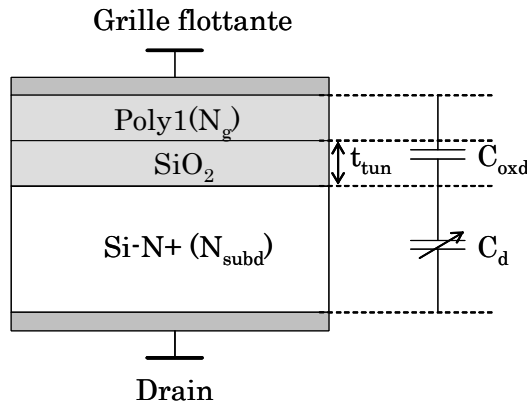


Figure II. 5 : Zone d'injection tunnel avec la capacité d'oxyde et la capacité de substrat.

II.2.2.3 Prise en compte de la désertion de grille flottante

Le phénomène de désertion dans la grille flottante en poly-silicium a été mis en évidence récemment (début années 90). En effet, à un dopage de l'ordre de 10^{19}cm^{-3} , une décroissance de la capacité totale peut être observée sur la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ expérimentale en régime d'accumulation (fig. II.9). Pour prendre en compte ce phénomène, il faut ajouter une capacité variable, notée C_p , au niveau de la grille flottante.

La capacité totale de la structure s'écrit :

$$\frac{1}{C_{\text{tun}}} = \frac{1}{C_{\text{oxd}}} + \frac{1}{C_d} + \frac{1}{C_p} \quad (\text{II. 51})$$

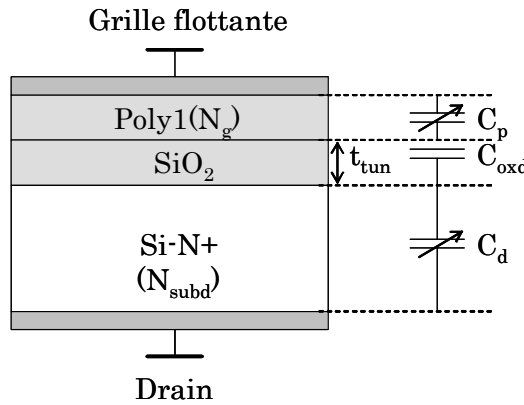


Figure II. 6 : Zone d'injection tunnel avec la capacité d'oxyde, la capacité de substrat et la capacité de désertion.

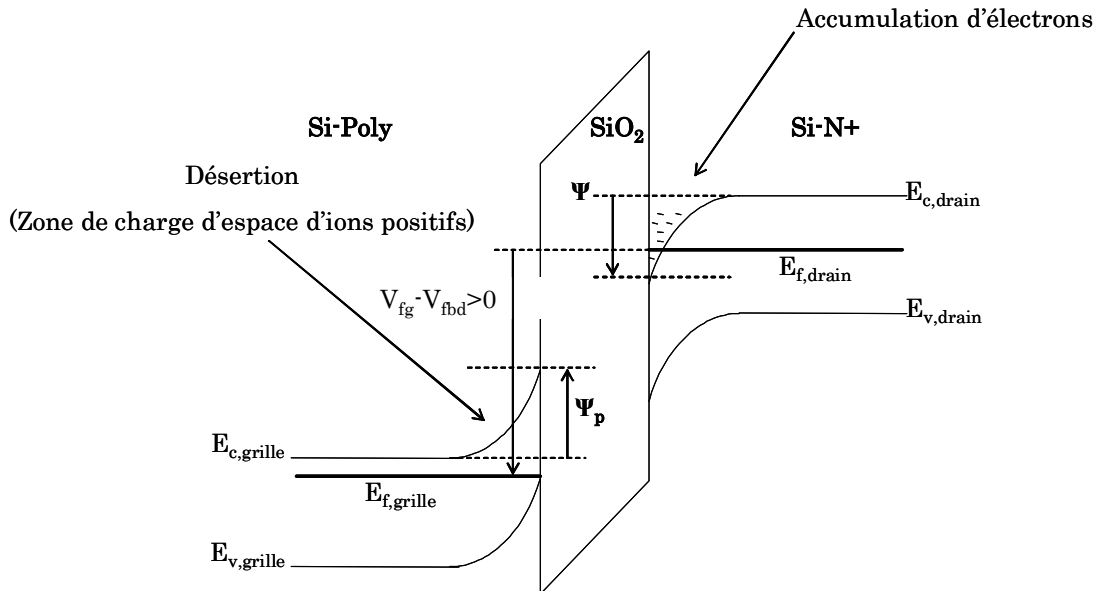


Figure II. 7 : Diagramme de bandes d'énergie de la zone tunnel en régime d'accumulation.

Le diagramme des bandes d'énergie de cette structure est représenté sur la figure II.7. En régime d'accumulation, nous pouvons observer le phénomène de désertion par une décroissance de la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ basse fréquence de

la structure Si-Poly/SiO₂/Si-N+. La figure II.8 illustre la distribution de charges entre le drain et la grille flottante, où Q_{fg} est la charge dans la grille flottante et Q_d la charge dans la diffusion drain.

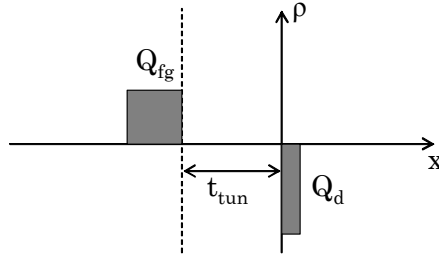


Figure II. 8 : *Distribution des charges dans la diffusion drain (Q_d) et dans la grille flottante (Q_{fg}).*

Si le dopage de grille flottante est supposé homogène, la résolution de l'équation de Poisson dans la zone de charge d'espace donne l'expression du potentiel de surface Ψ à l'interface Si-Poly/SiO₂ en fonction de l'épaisseur X_d de la zone désertée :

$$\Psi_p = \frac{1}{2} \frac{qN_g}{\epsilon_0 \epsilon_{Si}} X_d^2 \quad (\text{II. 52})$$

ou encore :

$$X_d = \sqrt{\frac{2\epsilon_0 \epsilon_{Si} \Psi_p}{qN_g}} \quad (\text{II. 53})$$

Par définition, la capacité surfacique C'_p due à cette zone de désertion vaut :

$$C'_p = \frac{\epsilon_0 \epsilon_{Si}}{X_d} \quad (\text{II. 54})$$

Ainsi, en combinant les relations (II.53) et (II.54), la capacité surfacique de désertion C'_p s'écrit :

$$C'_p = \sqrt{\frac{q\epsilon_0 \epsilon_{Si} N_g}{2\Psi_p}} \quad (\text{II. 55})$$

où N_g est le dopage de grille flottante.

La relation (II.55) comporte une inconnue : Ψ_p . Afin de tracer la caractéristique $C_{tun}(V_{gb})$ théorique de la structure, nous devons être capables de déterminer le potentiel de surface Ψ_p pour chaque valeur du potentiel de grille flottante V_{fg} . Nous déterminons Ψ_p en utilisant l'égalité, en valeur absolue, des charges dans la grille flottante et dans la diffusion drain :

$$Q'_d = Q'_{fg} \quad (\text{II. 56})$$

Par ailleurs, en supposant que seule la charge de désertion contribue à la densité surfacique de charge de grille flottante, il vient :

$$Q'_{fg} = \sqrt{2q\epsilon_0\epsilon_{Si}N_g\Psi_p} \quad (\text{II. 57})$$

En introduisant cette expression de Q'_{fg} dans la relation (II.56) qui exprime la neutralité de la structure, il vient :

$$\Psi_p = \frac{Q_d'^2}{2q\epsilon_0\epsilon_{Si}N_g} \quad (\text{II. 58})$$

Cette relation (II.58) nous donne la valeur de Ψ_p pour chaque potentiel de grille flottante V_{fg} , Q_d' étant donnée par la relation (II. 49).

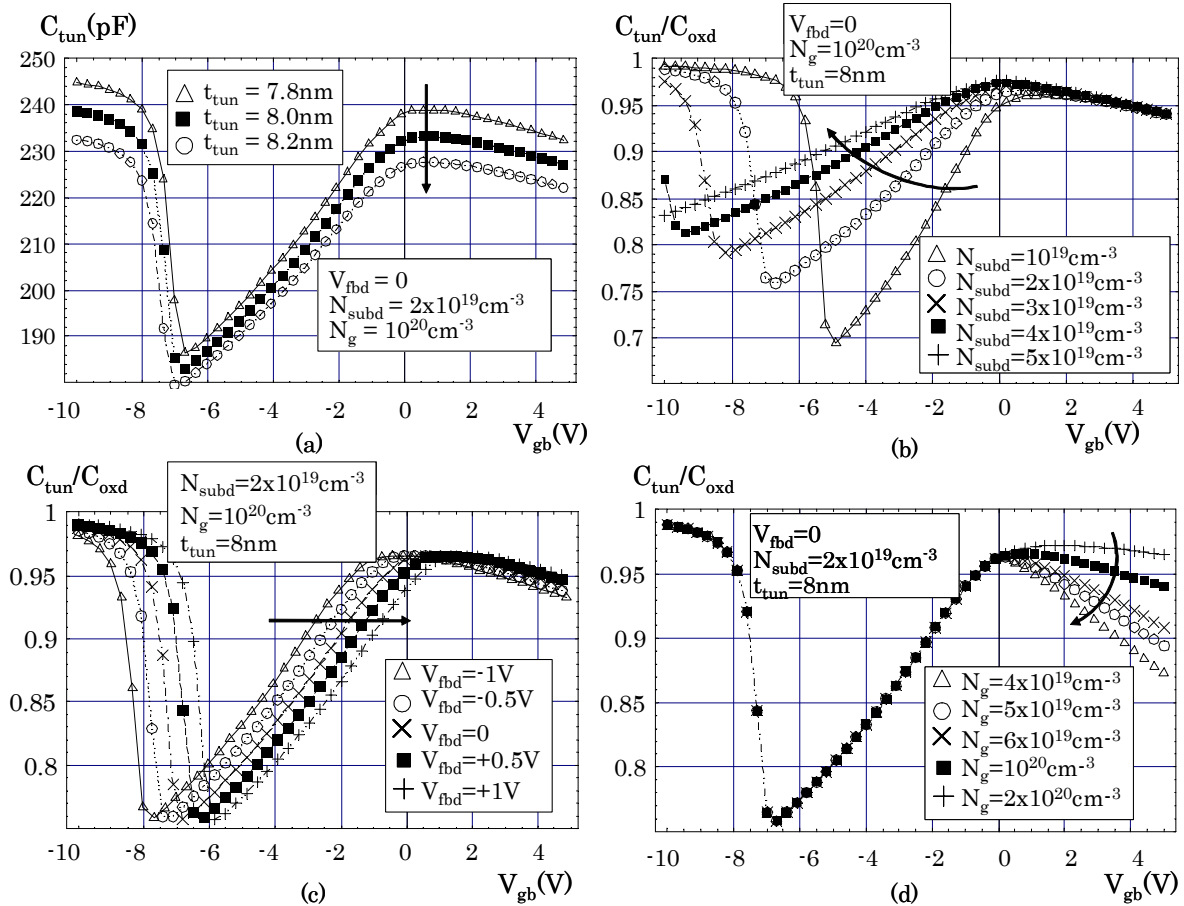


Figure II. 9 : Evolutions de la caractéristique $C_{tun}(V_{gb})$ d'une structure SOS Si-Poly/SiO₂/Si-N+ pour différentes valeurs d'épaisseur d'oxyde tunnel (a), de dopage de diffusion drain (b), de tension de bandes plates (c) et de dopage de grille en polysilicium (d).

Sur la figure II.9, nous présentons l'évolution de la caractéristique $C_{tun}(V_{gb})$ d'une structure SOS Si-Poly/SiO₂/Si-N+ pour différentes valeurs d'épaisseur d'oxyde tunnel, de dopage de substrat (diffusion drain), de tension de bandes plates et de dopage de grille en polysilicium. Les caractéristiques ont été tracées en prenant en compte la désertion de grille (EEPROM4). On constate notamment que, en régime d'accumulation (du substrat Si-N+), la pente de la caractéristique $C_{tun}(V_{gb})$ augmente lorsque le dopage de grille diminue. Ce résultat est important pour la caractérisation d'une telle structure.

II.2.2.4 Comparatif

Pour résumer, nous avons représenté la capacité de la zone tunnel en fonction de la tension grille-substrat V_{gb} , pour les trois modèles, sur la figure II.10. Nous verrons en détail ce point dans les simulations.

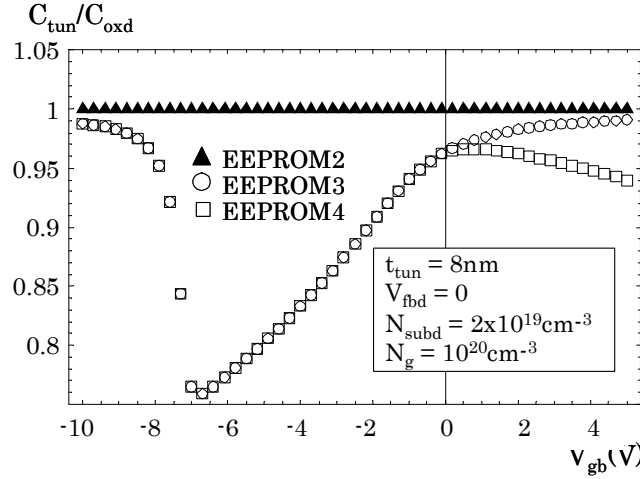


Figure II. 10 : Variations de la capacité de la zone tunnel en fonction de la tension grille-substrat V_{gb} pour les trois modèles considérés.

Ces caractéristiques montrent que la variation relative de la capacité de la zone tunnel est typiquement inférieure à 30% sur toutes les zones de fonctionnement.

L'autre spécificité de chaque modèle réside dans le calcul du champ électrique à travers l'oxyde tunnel.

En effet :

Pour EEPROM2 :

$$E_{tun} = \frac{V_{fg} - V_d}{t_{tun}} \quad (II. 59)$$

Pour EEPROM3 :

$$E_{tun} = \frac{V_{fg} - V_d - V_{fbd} - \Psi}{t_{tun}} \quad (II. 60)$$

Pour EEPROM4 :

$$E_{tun} = \frac{V_{fg} - V_d - V_{fbd} - \Psi + \Psi_p}{t_{tun}} \quad (II. 61)$$

Où V_{fbd} est la tension de bandes plates de la structure Si-Poly/SiO₂/Si-N. Ces expressions montrent que le champ électrique du modèle EEPROM2 est surévalué (car $\Psi > 0$, $\Psi_p < 0$ et $V_{fbd} \ll V_{fg}$). On s'attend donc à ce que le courant FN et la charge Q_{fg} injectée dans la grille flottante, simulés avec EEPROM2 soient surévalués par rapport à ceux simulés avec EEPROM3 et EEPROM4. Ce point, fondamental, sera approfondi dans le paragraphe traitant des simulations.

II.2.3 Modélisation du transistor de sélection

Le transistor de sélection est modélisé par un modèle SPICE 3 qui fonctionne en saturé/bloqué, permettant ainsi une grande rapidité de calcul. La figure II.11 présente les polarisations du transistor de sélection selon ce modèle :

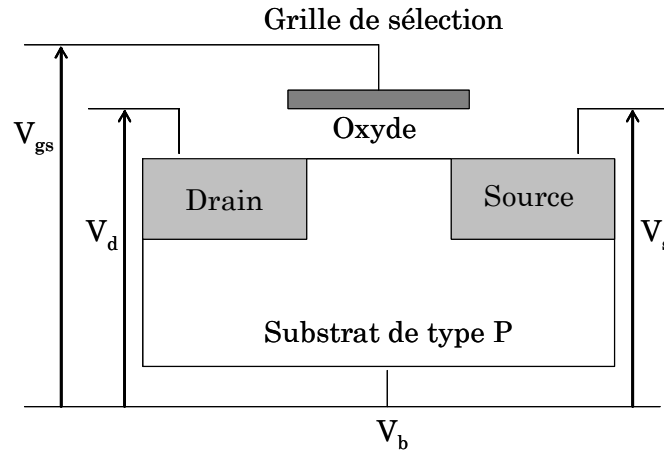


Figure II. 11 : *Polarisations du transistor de sélection.*

II.2.4 Bilan des paramètres du modèle

Le modèle développé pour le transistor d'état comporte un ensemble de paramètres classés par catégorie physique dans le tableau ci-dessous :

Paramètres géométriques			
N°	Nom-Modèle	Définition	Unité
1	W	Largeur du transistor	m
2	L	Longueur de grille	m
3	L _d	Longueur de recouvrement sur L	m
4	t _{ox}	Epaisseur de l'oxyde de grille	m
5	t _{tun}	Epaisseur de l'oxyde tunnel	m
6	t _{pp}	Epaisseur du diélectrique ONO	m
7	S _{tun}	Surface de la zone tunnel	m ²
8	S _{pp}	Surface de la grille flottante	m ²
Paramètres électriques			
N°	Nom-Modèle	Définition	Unité
9	R _s	Résistance d'accès à la source	Ω
10	R _d	Résistance d'accès au drain	Ω
11	I _s	Courant de saturation de la diode	A
12	C _{j0}	Capacité par unité de surface de la jonction PN	Fm ⁻²
13	M _j	Facteur de gradualité de la jonction	-
14	V _{t0}	Tension de seuil	V
15	V _{fb}	Tension de bandes plates	V
16	μ ₀	Mobilité à champ faible	m ² V ⁻¹ s ⁻¹
17	θ	Facteur de réduction de la mobilité	V ⁻¹
18	V _{sat}	Vitesse de saturation des électrons	ms ⁻¹
19	N _{sub}	Dopage de substrat	m ⁻³
20	N _g	Dopage de grille flottante (poly1)	m ⁻³
21	N _{subd}	Dopage de diffusion drain	m ⁻³
Paramètres physiques			
N°	Nom-Modèle	Définition	Unité
1	α _{écriture}	Coefficient FN préexponentiel en programmation	AV ⁻²
2	β _{écriture}	Coefficient FN exponentiel en programmation	Vm ⁻¹
3	α _{effacement}	Coefficient FN préexponentiel en effacement	AV ⁻²
4	β _{effacement}	Coefficient FN exponentiel en effacement	Vm ⁻¹
5	Q _{fgini}	Charge initiale de grille flottante	C
6	V _j	Barrière de potentiel de la jonction PN	V

Tableau II. 1 : Paramètres du transistor d'état.

La simulation des opérations d'écriture et d'effacement nécessite la connaissance des paramètres Fowler–Nordheim. Ces paramètres sont déterminés lors d'une caractérisation au cours de laquelle sont déterminés t_{tun}, V_{fb}, N_g, N_{subd} ainsi que

$\alpha_{\text{écriture}}$, $\beta_{\text{écriture}}$, $\alpha_{\text{effacement}}$ et $\beta_{\text{effacement}}$ (tableau II.1). La caractérisation est effectuée sur des structures équivalentes à la zone tunnel de type Si–Poly/SiO₂/Si–N⁺. La méthodologie utilisée, qui prend en compte la désertion de grille flottante, a été développée par Harabech [Harabech'00].

Les paramètres du modèle du transistor de sélection, au nombre de 16, sont donnés par le tableau 2.

Paramètres géométriques			
N°	Nom-Modèle	Définition	Unité
1	W	Largeur du transistor	m
2	L	Longueur de grille	m
3	L _d	Longueur de recouvrement sur L	m
4	t _{ox}	Epaisseur de l'oxyde de grille	m
Paramètres électriques			
N°	Nom-Modèle	Définition	Unité
5	R _s	Résistance d'accès à la source	Ω
6	R _d	Résistance d'accès au drain	Ω
7	I _s	Courant de saturation de la diode	A
8	C _{j0}	Capacité par unité de surface de la jonction PN	Fm ⁻²
9	M _j	Facteur de gradualité de la jonction	-
10	V _{t0}	Tension de seuil	V
11	V _{fb}	Tension de bandes plates	V
12	μ_0	Mobilité à champ faible	m ² V ⁻¹ s ⁻¹
13	θ	Facteur de réduction de la mobilité	V ⁻¹
14	V _{sat}	Vitesse de saturation des électrons	ms ⁻¹
15	N _{sub}	Dopage de substrat	m ⁻³
Paramètres physiques			
N°	Nom-Modèle	Définition	Unité
16	V _j	Barrière de potentiel de la jonction PN	V

Tableau II. 2 : *Les différents paramètres du transistor de sélection.*

La simulation du modèle nécessite de connaître un certain nombre de paramètres physiques, électriques et géométriques. La caractérisation est une procédure qui nous permet de déterminer ces paramètres. Nous allons présenter dans cette partie la méthodologie mise en œuvre pour extraire les paramètres les plus importants. Ces paramètres sont au nombre de 8 : l'épaisseur de l'oxyde tunnel (t_{tun}), la tension de bandes plates de la zone tunnel (V_{fbd}), le dopage de grille flottante (N_g), le dopage de diffusion drain (N_{subd}) et les paramètres FN en écriture et en effacement $\alpha_{\text{écriture}}$, $\beta_{\text{écriture}}$, $\alpha_{\text{effacement}}$ et $\beta_{\text{effacement}}$.

II.3 Extraction des paramètres

II.3.1 Méthodologie

La méthodologie mise en œuvre comporte deux grandes étapes. Dans un premier temps, on détermine t_{tun} , V_{fbd} , N_g et N_{subd} en ajustant les caractéristiques $C_{\text{tun}}(V_{\text{gb}})$ théorique et expérimentale par un processus d'optimisation. Ensuite, on détermine les paramètres $\alpha_{\text{écriture}}$, $\beta_{\text{écriture}}$, $\alpha_{\text{effacement}}$ et $\beta_{\text{effacement}}$ en ajustant les caractéristiques Fowler–Nordheim $I_{\text{FN}}(V_{\text{gb}})$ théorique et expérimentale. Ces extractions sont faites sur des capacités de grande surface afin d'obtenir des valeurs de capacités et de courant FN suffisamment élevées, donc facilement mesurables.

La figure II.12 représente l'organigramme de la méthodologie suivie. Le calcul de la capacité C_{tun} théorique prend en compte la désertion de grille flottante selon la méthode expliquée au paragraphe II.2.2.3. Quant au courant théorique, on le calcule en utilisant l'expression du champ électrique donné par la relation (II.61).

Les étapes d'initialisation sont importantes car une bonne initialisation permet au processus d'optimisation de converger rapidement vers les bonnes valeurs des paramètres à extraire.

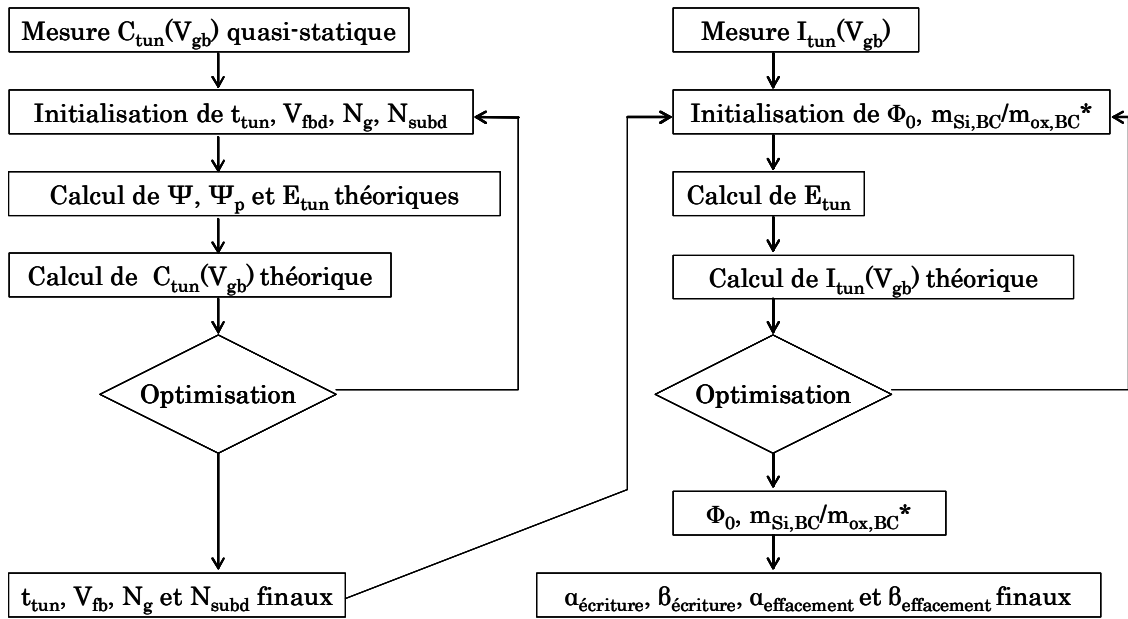


Figure II. 12 : Organigramme de la méthodologie mise en œuvre en vue d'extraire les paramètres Fowler-Nordheim.

Nous allons à présent exposer les méthodes classiques d'initialisation de t_{tun} , N_{subd} , N_g et V_{fbd} . On initialise l'épaisseur d'oxyde tunnel t_{tun} en lisant la valeur maximale C_{max} de la capacité de la structure sur la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ expérimentale. En effet :

$$t_{\text{tun}} = \frac{\epsilon_0 \epsilon_{\text{SiO}_2}}{C_{\text{max}}} S_{\text{tun}} \quad (\text{II. 62})$$

Le substrat Si-N+ se trouve alors en accumulation. Il faut noter que cette relation n'est pas tout à fait exacte car, du fait de la désertion dans la grille Si-Poly, le maximum de la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ n'est pas égal à la capacité d'oxyde C_{oxd} . Il s'agit d'une valeur approchée de l'épaisseur d'oxyde qui nous permet de démarrer le processus d'optimisation.

La valeur initiale du dopage de drain N_{subd} est déterminée en utilisant la méthode classique de détermination du dopage de substrat à l'aide de la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ basse fréquence d'une structure MOS.

Ainsi, à la limite de l'inversion forte, le potentiel de surface Ψ vaut $2\Phi_{\text{bd}}$, Φ_{bd} étant le potentiel de substrat et dont l'expression classique est :

$$\Phi_{\text{bd}} = \frac{k_B T}{q} \ln \left(\frac{N_{\text{subd}}}{n_i} \right) \quad (\text{II. 63})$$

Par ailleurs, à la limite de l'inversion forte, l'épaisseur de la zone de charge d'espace dans le substrat semi-conducteur, dans le cas d'un dopage abrupt, vaut :

$$W_{\text{max}} = \sqrt{\frac{2\varepsilon_0 \varepsilon_{\text{Si}} \Psi}{q N_{\text{subd}}}} \quad (\text{II. 64})$$

Sachant que Ψ vaut $2\Phi_{\text{bd}}$ à la limite de l'inversion forte, cette relation devient :

$$W_{\text{max}} = \sqrt{\frac{4\varepsilon_0 \varepsilon_{\text{Si}} k_B T \ln \left(\frac{N_{\text{subd}}}{n_i} \right)}{q^2 N_{\text{subd}}}} \quad (\text{II. 65})$$

La capacité due au substrat semi-conducteur atteint sa valeur minimale C_{scmin} lorsque l'épaisseur de la zone de charge d'espace atteint sa valeur maximale, auquel cas :

$$\frac{1}{C_{\text{scmin}}} = \frac{W_{\text{max}}}{\varepsilon_0 \varepsilon_{\text{Si}} S_{\text{tun}}} = \frac{1}{S_{\text{tun}}} \sqrt{\frac{4k_B T \ln \left(\frac{N_{\text{subd}}}{n_i} \right)}{q^2 \varepsilon_0 \varepsilon_{\text{Si}} N_{\text{subd}}}} \quad (\text{II. 66})$$

D'autre part, en limite d'inversion forte, la capacité totale de la structure est telle que :

$$\frac{1}{C_{\Psi=2\Phi_{\text{BD}}}} = \frac{1}{C_{\text{oxd}}} + \frac{1}{C_{\text{scmin}}} \quad (\text{II. 67})$$

En combinant les relations (II.66) et (II.67), il vient :

$$\frac{1}{C_{\Psi=2\Phi_{\text{bd}}}} = \frac{1}{C_{\text{oxd}}} + \frac{1}{S_{\text{tun}}} \sqrt{\frac{4k_B T \ln \left(\frac{N_{\text{subd}}}{n_i} \right)}{q^2 \varepsilon_0 \varepsilon_{\text{Si}} N_{\text{subd}}}} \quad (\text{II. 68})$$

On obtient ainsi le dopage de drain N_{subd} en résolvant numériquement cette équation transcendante. $C_{\Psi=2\Phi_{\text{bd}}}$ est lue graphiquement sur la caractéristique

$C_{\text{tun}}(V_{\text{gb}})$ expérimentale pour Ψ égal à $2\Phi_{\text{bd}}$ et C_{oxd} est calculée à l'aide de l'épaisseur d'oxyde t_{tun} précédemment déterminée.

Par définition, la tension de bandes plates est la tension pour laquelle le potentiel de surface Ψ est nul. Dans ce cas, la capacité totale C_{FB} de la structure vaut :

$$C_{\text{FB}} = \frac{\varepsilon_0 \varepsilon_{\text{Si}}}{L_{\text{D}} + \frac{\varepsilon_{\text{Si}}}{\varepsilon_{\text{SiO}_2}} t_{\text{tun}}} S_{\text{tun}} \quad (\text{II. 69})$$

où L_{D} , longueur de Debye extrinsèque du semi-conducteur, vaut :

$$L_{\text{D}} = \sqrt{\frac{k_{\text{B}} T \varepsilon_0 \varepsilon_{\text{Si}}}{q^2 N_{\text{subd}}}} \quad (\text{II. 70})$$

En reportant cette valeur sur la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ expérimentale et en lisant la tension grille-substrat V_{gb} correspondante, nous déterminons une valeur initiale de V_{fbd} .

Il n'y a pas de méthode générale quant à l'initialisation du dopage de grille N_{g} . On s'inspire du fait que, comme le montre la figure II.9, la désertion est d'autant plus marquée sur la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ que N_{g} est faible.

II.3.2 Extraction de α et β

II.3.2.1 Méthode générale

La méthode générale utilisée pour déterminer les paramètres FN α et β consiste à tracer $\ln[I_{\text{FN}}/S_{\text{tun}}(E_{\text{tun}})^2]$ en fonction de $1/E_{\text{tun}}$. Le graphe correspondant doit être linéaire et est communément appelé droite FN. En effet, partant de l'expression classique du courant FN donnée par la relation (II.18) [Lenzlinger'69][Keiger'81][Weinberg'82][Horiguci'85][Chiou'91][Sheybani'94], l'on a :

$$\ln\left(\frac{I_{\text{FN}}}{S_{\text{tun}} E_{\text{tun}}^2}\right) = \ln \alpha - \beta \frac{1}{E_{\text{tun}}} \quad (\text{II. 71})$$

où I_{FN} est le courant tunnel Fowler-Nordheim, S_{tun} l'aire de l'interface d'injection et E_{tun} le champ électrique à travers l'isolant tunnel. Les paramètres α et β sont déterminés en calculant la pente et l'ordonnée à l'origine de la droite ainsi tracée. Dans la relation (II.71), I_{FN} est donné par la mesure, S_{tun} une donnée, α et β les paramètres à déterminer. Seul E_{tun} est inconnu. L'une des principales difficultés dans le tracé de la droite FN consiste à déterminer E_{tun} en fonction de la tension V_{gb} appliquée aux bornes de la structure. Dans le cadre de notre thèse, cette structure est un sandwich Si-Poly/SiO₂/Si-N⁺. Pour une telle structure, la connaissance de E_{tun} nécessite le calcul, pour chaque V_{gb} [Harabech'00], des potentiels de surface Ψ et Ψ_{p} aux interfaces SiO₂/Si-N⁺ et Si-Poly/SiO₂, respectivement, et de la tension de bandes plates V_{fbd} de la structure. Ces calculs, menés rigoureusement, sont fastidieux car ils nécessitent la détermination

précise de la structure de la bande interdite du SiO₂ et du silicium polycristallin, une tâche complexe du fait de la nature amorphe et polycristalline, respectivement, de ces deux matériaux [Mohammad'95].

II.3.2.2 Approche de Croci et al.

Afin d'éviter cet écueil, Croci *et al.* ont mis au point une méthode d'extraction de α et β dans laquelle il n'est nécessaire de connaître ni Ψ , ni Ψ_p , ni V_{fbd} [Croci'99]. En effet, si l'on tient compte des potentiels de surface et de la tension de bandes plates, Croci *et al.* ont exprimé le champ électrique E_{tun} comme suit (en négligeant la contribution des états d'interface):

$$E_{tun} = \frac{V_{gb} - (V_{fbd} + \Psi + \Psi_p)}{t_{tun}} \quad (II. 72)$$

où t_{tun} est l'épaisseur de l'oxyde tunnel.

Dans la relation (II.72), le terme $V_{fbd} + \Psi + \Psi_p$ est remplacé par une constante globale K , indépendante de la tension V_{gb} aux bornes de la structure. Ainsi, la relation (II.72) se réécrit :

$$E_{tun} = \frac{V_G - K}{t_{tun}} \quad (II. 73)$$

L'une des particularités de cette approche est le traitement de la dépendance de α et β en fonction du rapport $m_{Si,BC}/m_{ox,BC}^*$ et de Φ_0 . Lors de l'extraction de α et β , nombre d'auteurs supposent la hauteur de barrière Φ_0 constante quelle que soit la valeur du champ électrique E_{tun} . En réalité, la hauteur de barrière Φ_0 varie en fonction du champ électrique E_{tun} du fait de la quantification des niveaux d'énergie des électrons confinés dans le puits de potentiel au niveau de la cathode. En effet, lors de l'établissement des expressions classiques de α et β , l'on suppose que les électrons sont un gaz tridimensionnel de particules libres dont la distribution d'énergie obéit à la statistique de Boltzmann. Ce modèle peut décrire le courant FN à faible injection [Lenzlinger'69][Sheybani'94] mais devient inadapté dès lors que le champ électrique E_{tun} est suffisamment important. En effet, dans ce cas, une partie (ou la totalité) des électrons se retrouvent confinés dans un puits de potentiel localisé au niveau de la cathode. Les électrons se comportent alors comme un gaz bidimensionnel de particules dont les niveaux d'énergies sont quantifiés et dépendent de la forme du puits de potentiel [Ando'98]. Par conséquent, la hauteur de barrière Φ_0 à l'interface d'injection dépend du champ électrique E_{tun} . Il s'ensuit qu'une hauteur de barrière Φ_0 constante (quel que soit E_{tun}) extraite de la pente d'une droite FN est erronée. En particulier, il est impossible d'obtenir un bon accord entre mesures et courbes théoriques de I_{FN} en recalculant le courant théorique à partir d'une telle hauteur de barrière Φ_0 [Mohammad'95][Ando'98][Stern'72][Hsing'79][Suné'92][Ando'76][Olivo'91].

Selon Croci *et al.*, la seule manière d'obtenir un bon accord entre les courants I_{FN} théorique et expérimental en gardant Φ_0 non-dépendant de E_{tun} consiste à introduire une hauteur de barrière effective, notée $(\Phi_0)_{eff}$. Il est important de souligner, à ce stade, que $(\Phi_0)_{eff}$ est un simple paramètre d'ajustement qui n'a

aucune signification du point de vue physique et n'a aucun lien avec la grandeur physique Φ_0 , hauteur de barrière au niveau de l'interface d'injection, extraite de la pente de la droite FN.

Afin de déterminer E_{tun} pour chaque V_{gb} , Croci *et al.* ont utilisé cette hauteur de barrière effective $(\Phi_0)_{eff}$ et ont supposé le rapport $m_{Si,BC}/m_{ox,BC}^*$ égal à 0.5, à l'instar de Weinberg [Weinberg'77][Weinberg'82]. Il leur fallait déterminer la constante K de la relation (II.73). Pour cela, ils ont utilisé l'unicité de $(\Phi_0)_{eff}$ dans les expressions de α et β . En effet, on peut exprimer $(\Phi_0)_{eff}$ de deux façons différentes en fonction de α , d'une part, et de β , d'autre part, à partir des expressions de α et β , données par les relations (II.19) et (II.20), en fonction de $m_{Si,BC}/m_{ox,BC}^*$ et de $(\Phi_0)_{eff}$:

$$\Phi_0(\alpha)_{eff} = \frac{m_{Si,BC}}{m_{ox,BC}^*} \frac{q^2}{8\pi h} \frac{1}{\alpha} \quad (II. 74)$$

$$\Phi_0(\beta)_{eff} = \left(\frac{3q}{8\pi} \sqrt{\frac{h^2}{2m_{ox,BC}^*}} \right)^{2/3} \beta^{2/3} \quad (II. 75)$$

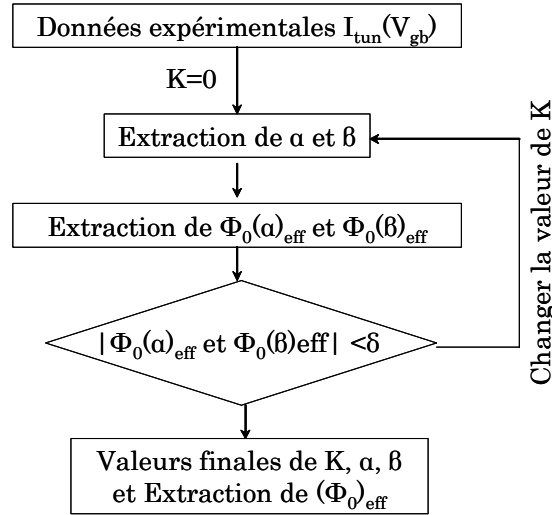


Figure II. 13 : *Algorithme de détermination des paramètres FN α et β , de la hauteur de barrière effective $(\Phi_0)_{eff}$ et de la constante K [Croci'99].*

Si $\Phi_0(\alpha)_{eff}$ est différent de $\Phi_0(\beta)_{eff}$, la relation $E_{tun}(V_{gb})$ donnée par la relation (II.73) est incorrecte. Ainsi, partant de 0, la valeur de K est incrémentée jusqu'à ce que la quantité $|\Phi_0(\alpha)_{eff} - \Phi_0(\beta)_{eff}|$ soit inférieure à une marge d'erreur δ fixée. La figure II.13 illustre l'algorithme correspondant pour déterminer les paramètres FN α et β , la hauteur de barrière effective $(\Phi_0)_{eff}$ et la constante K.

II.3.2.3 Analyse de Chiou et al.

Partant de l'expression linéaire de $\ln(J_{FN}/E_{tun}^2)$ en fonction de $1/E_{tun}$, Y.L. Chiou *et al.* [Chiou'01] ont analysé les paramètres les plus influents sur l'extraction de α et β : ils ont montré que l'extraction de α est conditionnée par la connaissance de la surface d'injection S_{tun} et de l'épaisseur t_{tun} de l'oxyde tunnel tandis que β est extrêmement sensible à la valeur de t_{tun} . En effet, sachant que

$\ln(J_{FN}/E_{tun}^2)$ est une fonction affine de $1/E_{tun}$, deux points, indexés par "1" et "2", suffisent pour déterminer la droite FN correspondante. Ainsi, il vient :

$$\ln\left(\frac{I_{FN1}}{S_{tun}E_{tun2}^2}\right) = \ln\alpha - \beta \frac{1}{E_{tun1}} \quad (II. 76)$$

$$\ln\left(\frac{I_{FN2}}{S_{tun}E_{tun2}^2}\right) = \ln\alpha - \beta \frac{1}{E_{tun2}} \quad (II. 77)$$

Des relations (II.76) et (II.77), l'on déduit facilement α et β :

$$\ln\alpha = \frac{\frac{E_{tun1}}{E_{tun2}} \ln\left(\frac{I_{FN1}}{SE_{tun1}^2}\right) - \ln\left(\frac{I_{FN2}}{SE_{tun2}^2}\right)}{\frac{E_{tun1}}{E_{tun2}} - 1} \quad (II. 78)$$

$$\beta = \frac{\ln\left(\frac{I_{FN1}E_{tun2}^2}{I_{FN2}E_{tun1}^2}\right)}{\frac{1}{E_{tun2}} - \frac{1}{E_{tun1}}} \quad (II. 79)$$

où S_{tun} est l'aire de l'interface d'injection, I_{FNk} et E_{tunk} respectivement le courant FN et le champ électrique correspondant au point de la droite FN indexé par k .

Dans ces deux expressions, Chiou *et al.* calculent le champ électrique de la manière suivante :

$$E_{tun} = \frac{V_{gb} - V_{fbd} - \Psi}{t_{tun}} \quad (II. 80)$$

où V_{gb} est la tension de grille/substrat, V_{fbd} la tension de bandes plates et Ψ le potentiel de surface à l'interface entre l'oxyde et le semiconducteur.

Les relations (II.78) et (II.79) montrent que l'extraction du paramètre FN α dépend de la connaissance de l'aire S_{tun} de l'interface d'injection, contrairement à l'extraction de β , où seuls I_{FN1} , I_{FN2} , E_{tun1} et E_{tun2} doivent être connus.

Si l'on appelle J_{FNk} la densité de courant relative au courant I_{FNk} , S_{tun0} la vraie valeur de l'aire de la surface d'injection et ξ l'incertitude sur S_{tun} , il vient :

$$S_{tun} = S_{tun0}(1 + \xi) \quad (II. 81)$$

$$J_{FN1} = \frac{I_{FN1}}{S_{tun0}(1 + \xi)} = J_{FN10} \frac{1}{(1 + \xi)} \quad (II. 82)$$

$$J_{FN2} = \frac{I_{FN2}}{S_{tun0}(1 + \xi)} = J_{FN20} \frac{1}{(1 + \xi)} \quad (II. 83)$$

où J_{FN10} et J_{FN20} sont les vraies valeurs des densités de courant relatives aux points «1» et «2». En combinant les relations (II.78), (II.82) et (II.83), l'on montre que :

$$\alpha = \alpha_0 \frac{1}{(1 + \xi)} \quad (\text{II. 84})$$

où α_0 est la vraie valeur du paramètre FN α . De même, si on appelle η l'incertitude sur l'épaisseur d'oxyde tunnel et $t_{\text{tun}0}$ la vraie valeur de cette dernière, l'on montre que :

$$\alpha = \alpha_0 (1 + \eta)^2 \quad (\text{II. 85})$$

$$\beta = \beta_0 \frac{1}{(1 + \eta)} \quad (\text{II. 86})$$

où β_0 est la vraie valeur du paramètre FN β .

Les équations (II.84), (II.85) et (II.86) montrent que l'incertitude sur α est fonction, à la fois, de l'incertitude sur S_{tun} et t_{tun} . L'incertitude sur β , quant à elle, dépend seulement de l'incertitude sur t_{tun} . Chiou *et al.* attribuent ceci au fait que le champ électrique E_{tun} n'étant pas mesurable directement, l'extraction des paramètres FN est très dépendante des paramètres avec lesquels le champ électrique est calculé, au premier rang desquels figure l'épaisseur d'oxyde tunnel t_{tun} . La tension de bandes plates V_{fbd} , bien que figurant dans l'expression du champ électrique, n'est pas déterminante dans la mesure où, pour de fortes tensions grille/substrat V_{gb} , son influence devient négligeable.

II.3.3 Mesures $C_{\text{tun}}(V_{\text{gb}})$ et $I_{\text{FN}}(V_{\text{gb}})$

La mesure quasi-statique de la caractéristique $C_{\text{tun}}(V_{\text{gb}})$ est effectuée sur un picoampèremètre HP4140. On mesure le courant de déplacement I à travers le circuit lorsque la polarisation de la structure varie de -5V à 4V avec une rampe de tension fixée à 50mV/s .

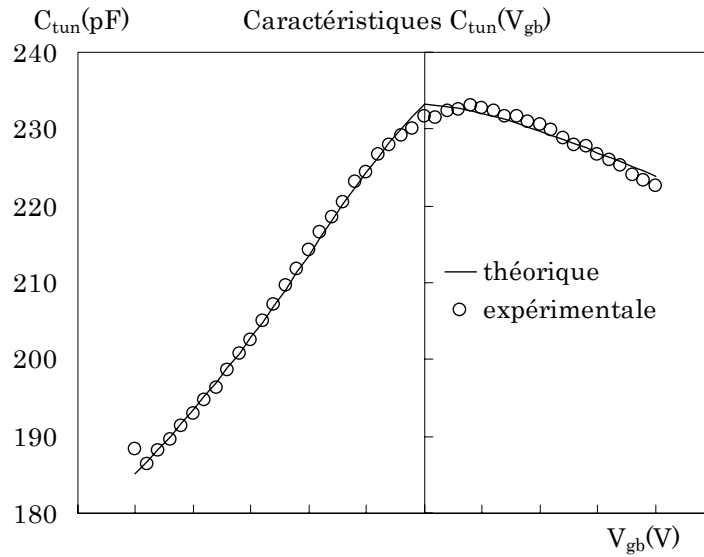


Figure II. 14 : *Caractéristiques $C_{\text{tun}}(V_{\text{gb}})$ basse fréquence théorique et expérimentale mesurée sur une structure Si-Poly/SiO₂/Si-N⁺.*

La capacité basse fréquence est donnée par la relation suivante :

$$C_{bf} = \frac{I}{\left(\frac{dV}{dt}\right)} \quad (\text{II. 87})$$

où dV/dt est la valeur de la rampe de tension appliquée.

Les courants Fowler–Nordheim sont mesurés sur un analyseur HP4156B. Deux mesures $I_{FN}(V_{gb})$ sont nécessaires ($V_{gb}>0$ et $V_{gb}<0$) afin d'extraire les paramètres correspondant à l'opération d'écriture et à l'opération d'effacement.

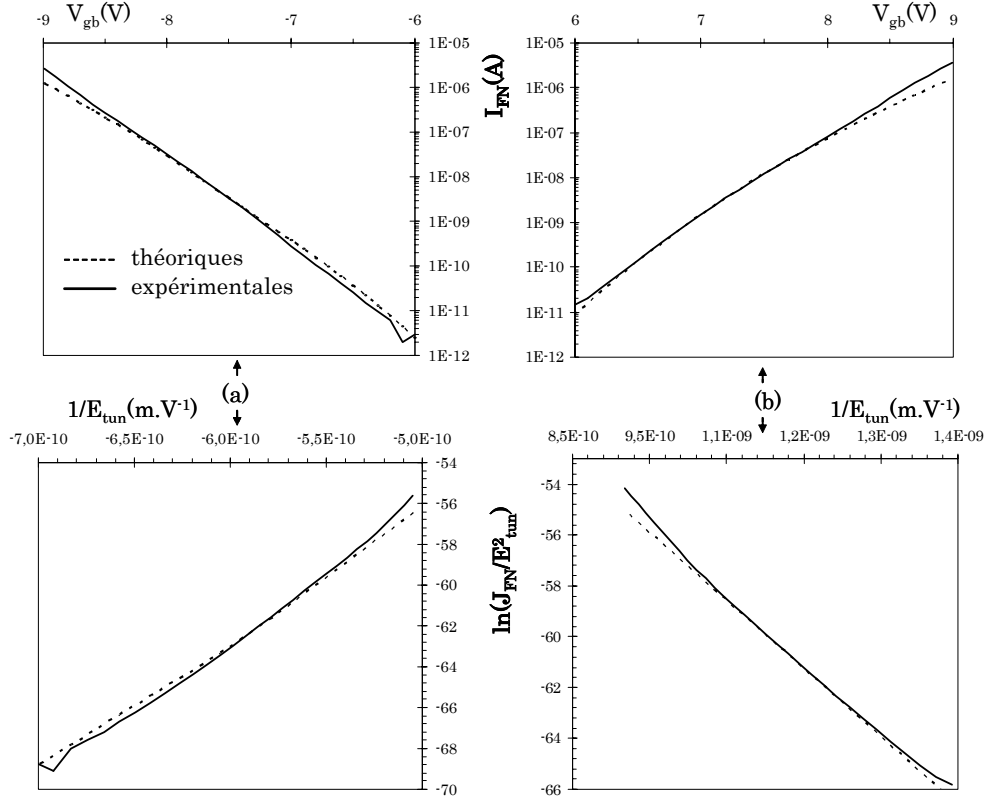


Figure II. 15 : *Caractéristiques Fowler–Nordheim $I_{FN}(V_{gb})$ et $\ln(I_{FN}/E_{tun}^2)$ en fonction de $1/E_{tun}$ en écriture (a) et en effacement (b).*

Le tableau (II.3) regroupe un exemple de valeurs extraites sur une plaquette de technologie $7.7\mu\text{m}^2$ fournie par ST–Microelectronics selon la méthodologie exposée précédemment.

Symbole	Signification	Valeur	Unité
t_{tun}	Epaisseur de l'oxyde tunnel	7.8×10^{-9}	m
N_g	Dopage de grille flottante	6.2×10^{25}	m^{-3}
N_{subd}	Dopage de diffusion drain	2.1×10^{25}	m^{-3}
V_{fbd}	Tension de bandes plates de la zone tunnel	0.15 ± 0.02	V
$\alpha_{écriture}$	Coefficient FN préexponentiel en écriture	$1.1 (\pm 0,01) \times 10^{-6}$	AV^{-2}
$\beta_{écriture}$	Coefficient FN exponentiel en écriture	$22.5 (\pm 0,01) \times 10^9$	Vm^{-1}
$\alpha_{effacement}$	Coefficient FN préexponentiel en effacement	$1.05 (\pm 0,01) \times 10^{-6}$	AV^{-2}
$\beta_{effacement}$	Coefficient FN exponentiel en effacement	$24.07 (\pm 0,01) \times 10^9$	Vm^{-1}

Tableau II. 3 : *Paramètres extraits sur la technologie $7.7\mu\text{m}^2$ de ST–Microelectronics.*

II.4 Simulations

II.4.1 Implantation du modèle dans ELDO

Le modèle du point-mémoire a été implanté dans le simulateur ELDO (Anacad Mentor Graphics). Cette implantation comporte deux volets : l'implantation du modèle du transistor d'état et l'implantation du modèle du transistor de sélection.

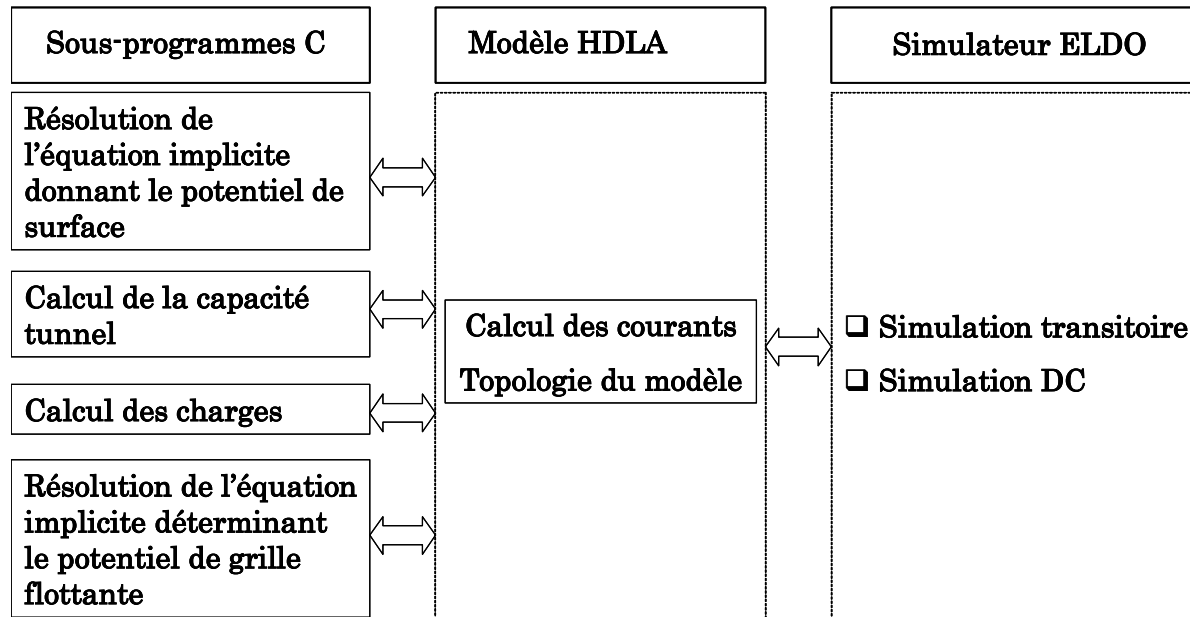


Figure II. 16 : Schéma de l'implantation du modèle dans le simulateur ELDO.

Le modèle du transistor d'état a été écrit en langage HDLA (Hardware Description Language Analog). Un modèle HDLA est composé d'un Bloc Entity où l'on déclare les nœuds du composant à modéliser dans PIN et les paramètres du modèle dans GENERIC. Toutes les équations du modèle sont décrites dans la partie Architecture du modèle. Les équations explicites sont traitées en HDLA tandis que les équations implicites sont résolues par des sous-programmes C interfacés avec la structure HDLA. Par exemple, l'équation implicite qui donne l'évolution du potentiel de grille flottante est résolue par un sous-programme C.

II.4.2 Simulations statiques

Les figures II.17a et II.17b représentent les potentiels de surface Ψ_{ss} et Ψ_{sd} à la source et au drain en fonction respectivement du potentiel de grille flottante et de la tension de grille de contrôle.

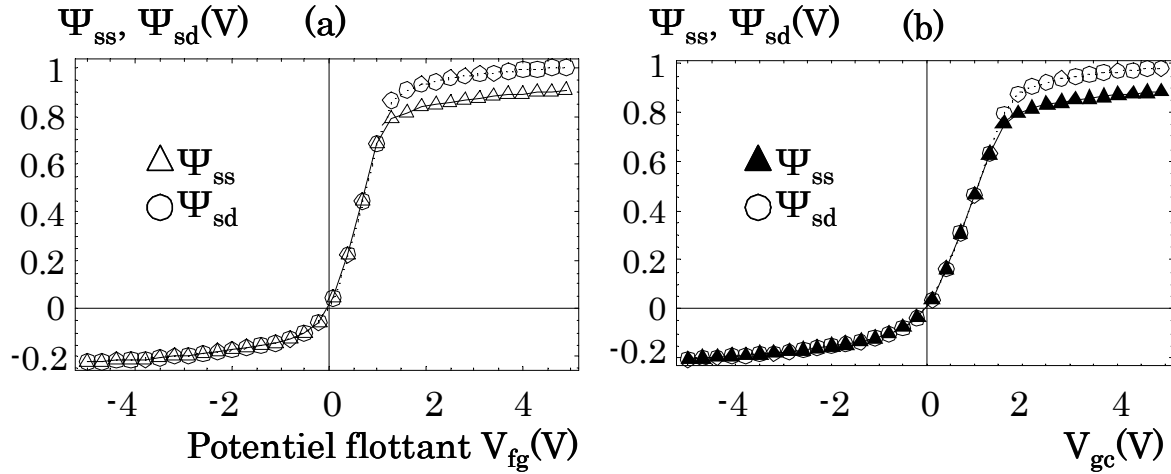


Figure II. 17 : Potentiels de surface à la source (Ψ_{ss}) et au drain (Ψ_{sd}) en fonction du potentiel flottant (a) et de la tension de grille de contrôle (b) pour $V_{ds} = 0,1V$.

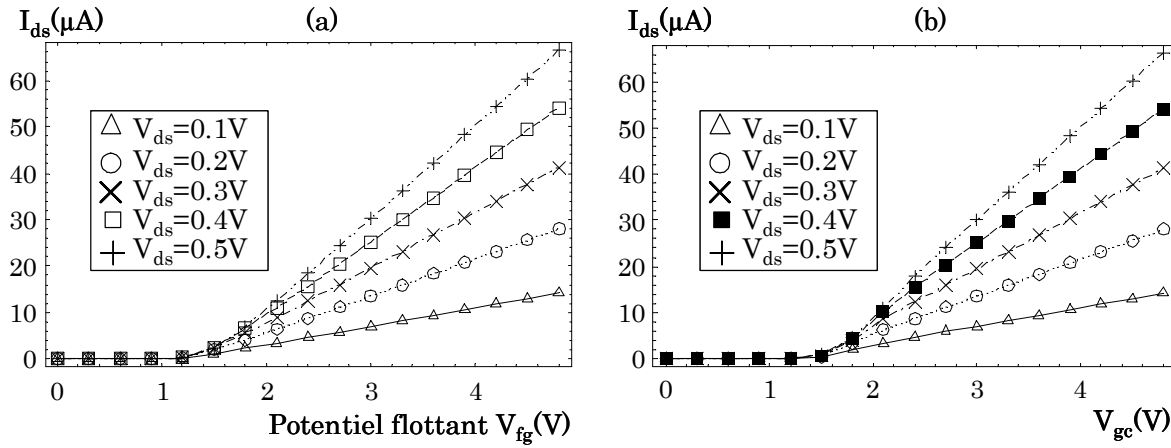


Figure II. 18 : Courant de drain I_{ds} en fonction du potentiel flottant (a) et de la tension de grille de contrôle (b) pour $V_{ds} = 0,1V$.

On constate que pour des tensions V_{gc} suffisamment positives, Ψ_{sd} est supérieure à Ψ_{ss} , établissant ainsi un gradient de potentiel et un champ électrique longitudinal qui sont à l'origine du courant à travers le canal. A l'opposé, lorsque V_{gc} est négatif, Ψ_{sd} et Ψ_{ss} sont égaux : il n'y a pas de gradient de potentiel dans le canal et le courant de drain est nul. Par ailleurs, notons que Ψ_{sd} et Ψ_{ss} sont continus dans tous les régimes de fonctionnement : en accumulation d'une part ($V_{gc} < V_{fb}$) et en désertion, faible et forte inversions d'autre part ($V_{gc} > V_{fb}$).

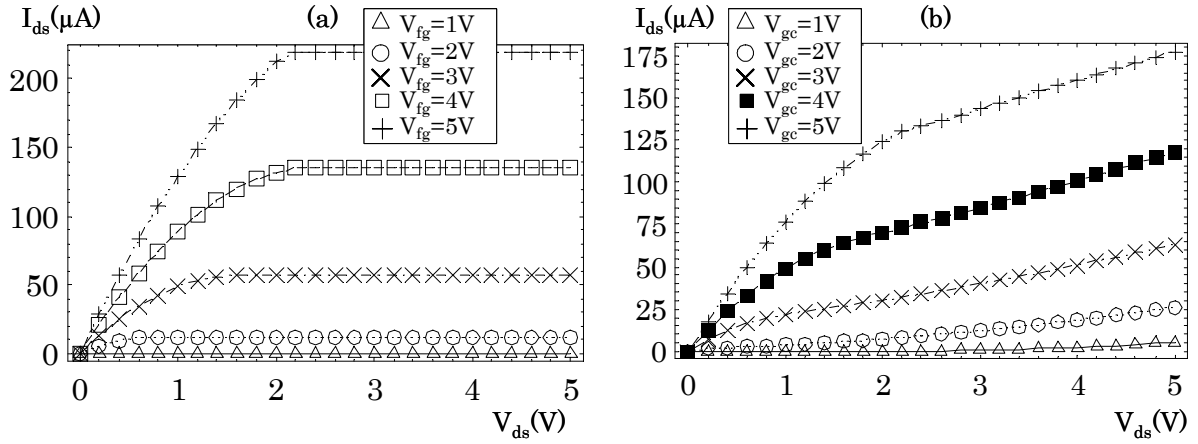


Figure II. 19 : *Caractéristiques de sortie $I_{ds}(V_{ds})$ pour un potentiel flottant V_{fg} allant de 1 à 5V avec un pas de 1V (a) et une tension V_{gc} allant de 1 à 5V avec un pas de 1V (b).*

Sur la figure II.19b, l'on peut noter le couplage capacitif entre la grille flottante et la grille de contrôle par une pente relativement importante en régime saturé. En effet, la relation (II.40) montre que si V_{gc} est constant, si V_{ds} varie, alors V_{fg} varie aussi. Ceci explique l'augmentation du courant, en régime saturé, sur la figure II.19b.

II.4.3 Simulations dynamiques

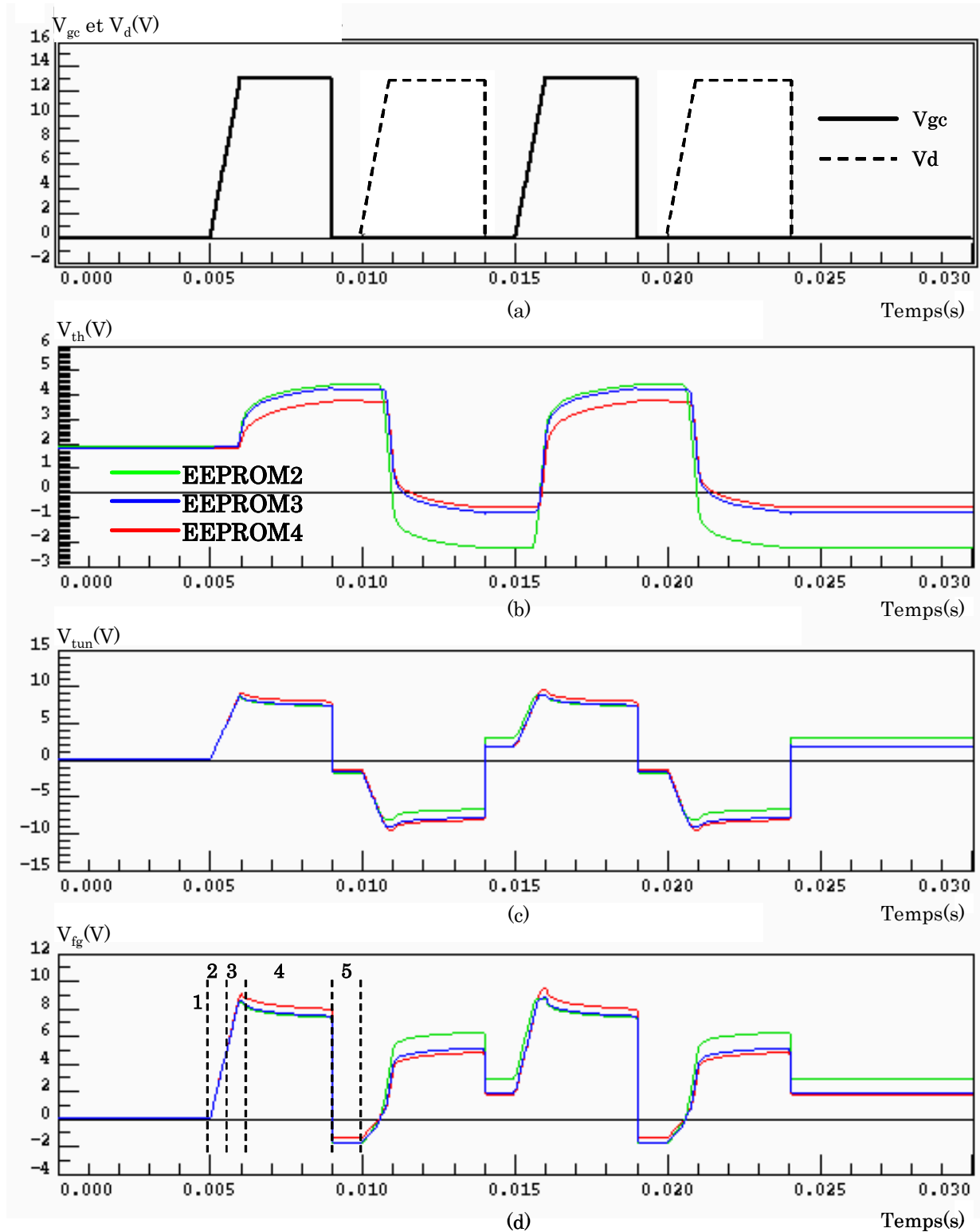


Figure II. 20 : *Signaux de programmation (a), tension de seuil (b), chute de potentiel à travers l'oxyde tunnel (c), et potentiel de grille flottante (d) simulés avec les trois modèles.*

Nous allons à présent analyser les résultats de la simulation du fonctionnement de la cellule mémoire en régime dynamique. Les simulations ont été effectuées

avec les trois modèles EEPROM2, EEPROM3 et EEPROM4, afin de dresser un bref comparatif entre ces modèles. Les résultats des simulations sont regroupés sur les figures II.20 et II.22.

Le signal de programmation (cf. figure II.20a), identique en écriture et en effacement, est un trapèze composé d'une rampe de montée de 1ms, suivie d'un plateau de 3ms à 13V et se termine par une rampe de descente de 1µs.

Dans un premier temps, nous allons analyser l'évolution du potentiel de grille flottante V_{fg} au premier effacement de la figure II.20d. Le même raisonnement, pourvu de quelques modifications, peut ensuite être mis en œuvre pour analyser le comportement de V_{fg} en écriture. Par ailleurs, l'allure des autres grandeurs en fonction du temps peut être déduite facilement de celle de V_{fg} .

Pour un signal de programmation trapézoïdal, on peut considérer que l'évolution de V_{fg} en fonction du temps comporte cinq phases lors d'un effacement. Ce découpage en cinq phases découle naturellement des caractéristiques du signal, telles que la valeur de la dérivée temporelle, et de la prise en compte de la valeur du courant FN (fig. II.21).

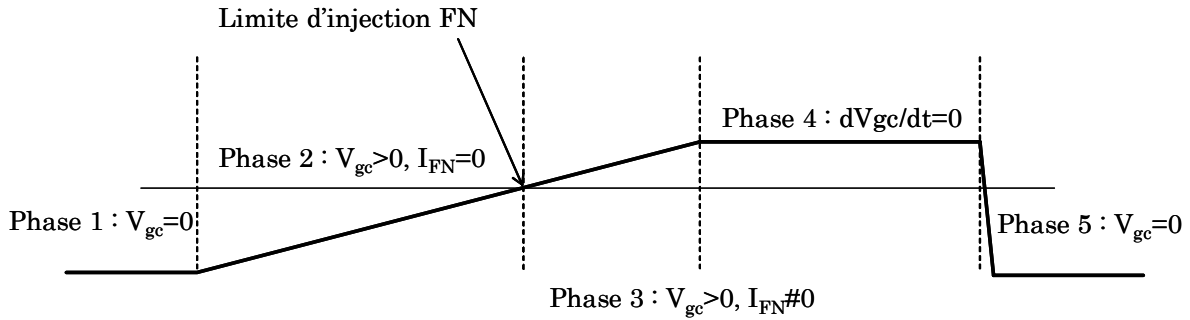


Figure II. 21 : *Découpage en cinq phases du signal d'effacement.*

i. Phase 1 : potentiel latent

Le potentiel latent correspond au potentiel de grille flottante lorsque aucune opération (écriture ou effacement) n'a lieu. Ce potentiel dépend de la charge alors stockée dans la grille flottante. Dans notre cas, la charge stockée initialement est nulle et comme $V_{gc}=V_d=0$, et en supposant que le terme $(V_{fb}+\Psi_s)$ est négligeable, V_{fg} est nul ((II.40)).

ii. Phase 2 : montée à I_{FN} nul

Sous l'effet du couplage capacitif, le potentiel de grille flottante augmente pendant la rampe de tension appliquée sur la grille de contrôle. Analysons brièvement la dérivée de V_{fg} par rapport au temps. En dérivant chaque membre de la relation (II.40) par rapport au temps, il vient :

$$\frac{dV_{fg}}{dt} = K_e \frac{dV_{gc}}{dt} + \frac{1}{C_x} \frac{d}{dt} [C_{ox} (V_{fb} + \Psi_s) + C_{tun} V_d] + \frac{I_{FN}}{C_x} \quad (II. 88)$$

Pendant cette phase, I_{FN} est nul. Par ailleurs, V_d est nul pendant tout l'effacement et les dérivées temporelles de V_{fb} et de Ψ_s sont négligeables devant dV_{gc}/dt . Par conséquent, la relation (II.88) se réécrit simplement comme suit :

$$\frac{dV_{fg}}{dt} = K_e \frac{dV_{gc}}{dt} \quad (\text{II. 89})$$

Cette relation montre que le rapport des pentes des tangentes à la courbe de V_{fg} et à la courbe de V_{gc} , au tout début de l'opération d'effacement, est égal au coefficient de couplage en effacement K_e , hors injection et avec une charge de grille flottante nulle. On constate aussi que le signe positif de dV_{fg}/dt donnée par la relation (II.89) est en accord avec la pente positive de la courbe de V_{fg} .

iii. Phase 3 : montée à I_{FN} non-nul

V_{fg} continue d'augmenter jusqu'à sa valeur maximale mais I_{FN} n'est plus négligeable. La relation (II.88) devient :

$$\frac{dV_{fg}}{dt} = K_e \frac{dV_{gc}}{dt} + \frac{I_{FN}}{C_x} \quad (\text{II. 90})$$

La présence de I_{FN} ($I_{FN} < 0$ et Q_{fg} diminue car injection d'électrons du drain vers la grille flottante) dans cette relation rompt la proportionnalité entre dV_{fg}/dt et dV_{gc}/dt observée dans la relation (II.89) : c'est la cause du léger infléchissement de V_{fg} observé en fin de phase 3.

iv. Phase 4 : décroissance exponentielle

Le potentiel de la grille flottante décroît exponentiellement pendant toute la durée du plateau du signal de grille de contrôle. Pendant cette phase, dV_{gc}/dt est nul (plateau) mais I_{FN} ne l'est pas. Toujours en négligeant les dérivées temporelles de V_{fb} et de Ψ_s par rapport aux autres termes, la relation (II.88) devient :

$$\frac{dV_{fg}}{dt} = \frac{I_{FN}}{C_x} \leq 0 \quad (\text{II. 91})$$

Cette quantité, négative, explique la décroissance de V_{fg} . Quant à la concavité de V_{fg} , elle s'explique en montrant que d^2V_{gc}/dt^2 est positif :

$$\frac{d^2V_{fg}}{dt^2} = \frac{1}{C_x} \frac{dI_{FN}}{dt} \leq 0 \quad (\text{II. 92})$$

En effet, V_{fg} atteint son maximum au début de la phase 4. Par conséquent, le champ électrique (proportionnel à V_{fg}) et $|I_{FN}|$ atteignent aussi leur maxima au début de la phase 4. $|I_{FN}|$ ayant atteint son maximum ne peut que décroître dans la phase 4 : $d|I_{FN}|/dt < 0$ et $dI_{FN}/dt > 0$ (car $I_{FN} < 0$). Enfin, d^2V_{gc}/dt^2 est donc positif et la courbe représentant V_{fg} concave.

v. Phase 5 : retour au potentiel latent

A la fin de l'effacement, la charge électrique négative stockée dans la grille flottante induit un potentiel latent compris entre -1 et $-2V$.

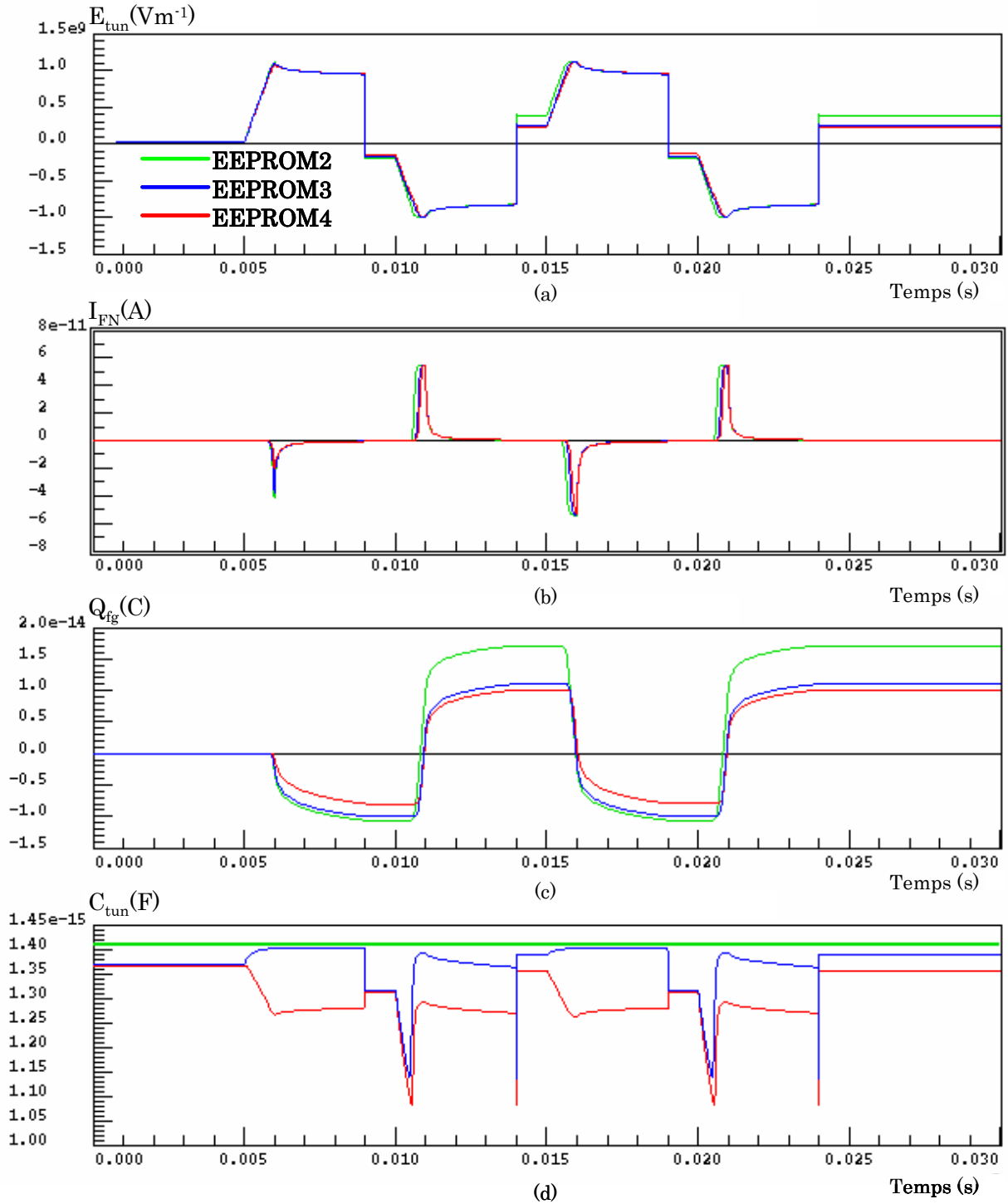


Figure II. 22 : Simulations transitoires du champ électrique (a), du courant I_{FN} (b), de la charge de grille flottante Q_{fg} (c) et de la capacité de la zone tunnel C_{tun} (d).

La figure II.22a montre que E_{tun} est surévalué dans le modèle EEPROM2 par rapport aux modèles EEPROM3 et EEPROM4. Ceci est conforme à l'analyse effectuée dans la modélisation de la capacité tunnel (cf. §II.2.2.4). Cette surévaluation se retrouve sur I_{FN} et Q_{fg} comme le montrent les figures II.22b et II.22c.

La principale distinction entre les modèles EEPROM3 et EEPROM4 apparaît sur la figure II.22d, représentant les variations de la capacité de la zone tunnel en fonction du temps. Avant une écriture ($10\text{ms} < t < 11\text{ms}$ ou $20\text{ms} < t < 21\text{ms}$), la différence de potentiel V_{tun} entre la grille flottante et le substrat Si-N+ varie rapidement de -1 à -10V : on distingue bien le régime de désertion du substrat Si-N+, que ce soit avec EEPROM3 ou EEPROM4. Les deux modèles ne présentent alors pas de différence notable. En revanche, l'écart observé en effacement ($5\text{ms} < t < 9\text{ms}$ ou $15\text{ms} < t < 19\text{ms}$) s'explique par la désertion de grille flottante prise en compte dans EEPROM4 (C_{tun} diminue alors de 1.35 à 1.27fF) mais pas dans EEPROM3 (C_{tun} plafonne à la valeur maximale, 1.40fF, capacité de l'oxyde tunnel).

Par ailleurs, la figure II.22d montre que C_{tun} est compris, à tout instant, entre 1.08fF et 1.37fF, ce qui correspond à une variation relative d'à peu près 25%, conformément à la remarque faite au paragraphe II.2.2.4 traitant de la modélisation de la capacité tunnel.

En conclusion nous avons surtout utilisé le modèle EEPROM2 pour obtenir rapidement une idée globale du comportement de la cellule. Dès qu'une plus grande précision est requise, des modèles plus élaborés comme EEPROM3 ou EEPROM4 deviennent indispensables.

Le choix de EEPROM3 ou de EEPROM4 dépend du dopage de grille flottante en polysilicium. Pour des niveaux de dopage très élevés, la différence entre les résultats de EEPROM3 et de EEPROM4 est faible et peut être négligée : nous utilisons alors le modèle EEPROM3 qui est nettement plus rapide. En revanche, pour des niveaux de dopage de grille flottante faibles, EEPROM4 est nécessaire.

II.5 Banc de mesure

D'une manière générale, le banc de mesure est constitué par l'appareillage nécessaire à la caractérisation et à l'écriture, l'effacement et la lecture de la tension de seuil des cellules EEPROM. Les plaquettes de silicium reposent sur une plate-forme sous pointes. Tous les appareils, connectés au réseau local, peuvent être commandés à distance depuis une station de travail sous Unix où toutes les manipulations sont pilotées via le logiciel Iccap.

II.5.1 Caractérisation d'une structure Si-Poly/SiO₂/Si-N+

La caractérisation nécessite la mesure de la capacité basse-fréquence de la structure Si-Poly/SiO₂/Si-N+ et du courant FN à travers l'oxyde tunnel en fonction de la tension grille/substrat V_{gb} . Les figures II.23a et II.23b représentent, respectivement, le banc de mesure des caractéristiques $C_{\text{tun}}(V_{\text{gb}})$ et $I_{\text{FN}}(V_{\text{gb}})$.

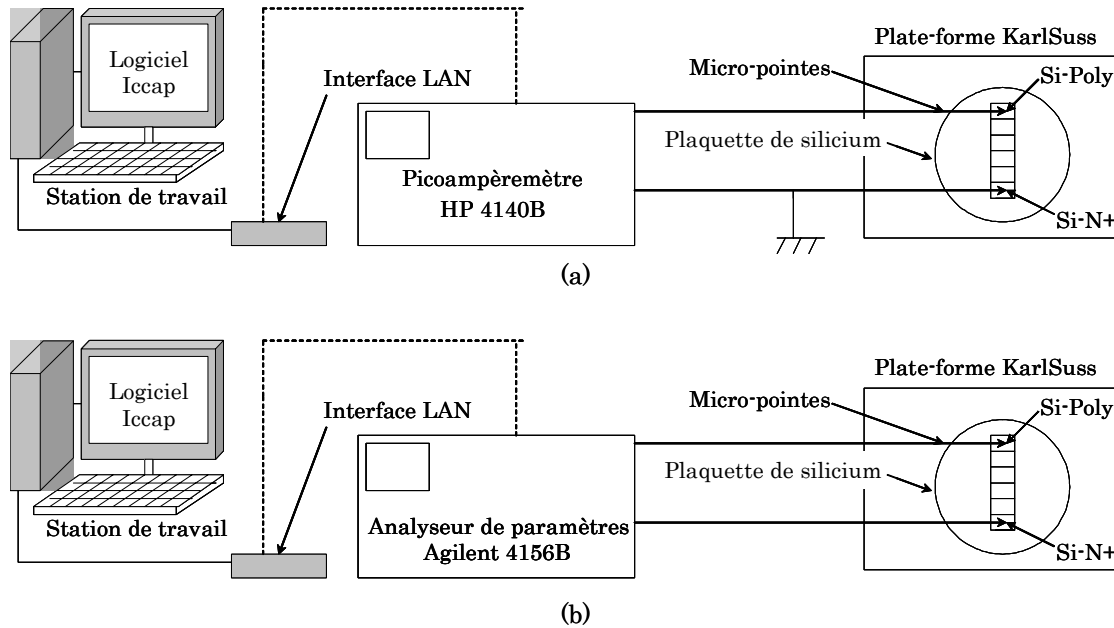


Figure II. 23 : Banc de mesure pour la mesure de la caractéristique $C_{tun}(V_{gb})$ et du courant $I_{FN}(V_{gb})$.

II.5.2 Ecriture, effacement et lecture de cellules EEPROM

La figure II.24 montre le schéma du banc de mesure pour l'écriture, l'effacement et la lecture des cellules EEPROM. Le cœur du dispositif est constitué par un générateur de signaux arbitraires tektronix AFG 320 pour générer les signaux d'écriture et d'effacement ainsi qu'un analyseur de paramètres Agilent 4156C pour la mesure des caractéristiques $I_{ds}(V_{gs})$ afin d'extraire la tension de seuil. La grille du transistor de sélection est polarisée par un générateur de tension continue HP 6644A. Par ailleurs, une matrice de commutation Keithley 7012-S permet, si nécessaire, de commander le générateur de signaux arbitraires ou l'analyseur de paramètres, selon que l'on effectue une programmation ou un effacement, ou une lecture de tension de seuil.

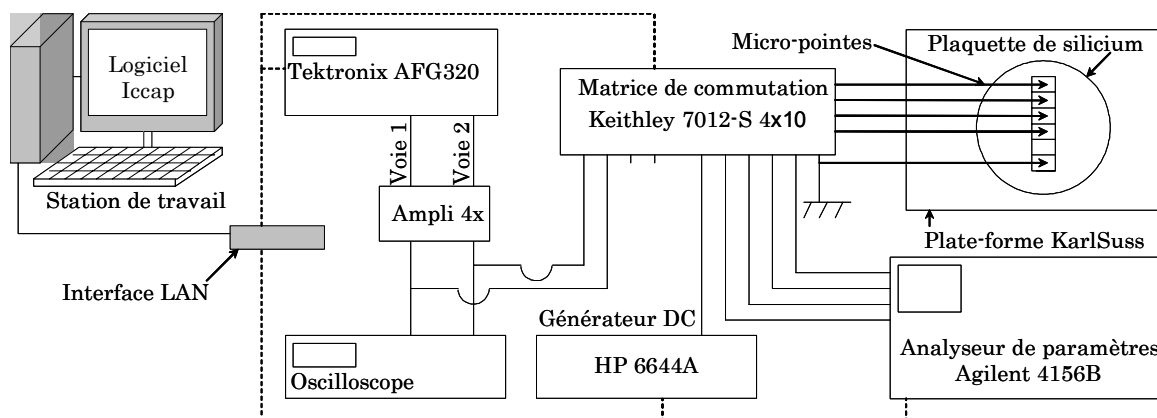


Figure II. 24 : Banc de mesure pour l'écriture, l'effacement et la lecture d'une cellule EEPROM.

II.5.3 Mesure de V_{th} et validation du modèle

La tension de seuil V_{th} est définie comme la valeur de la tension appliquée sur la grille du transistor nécessaire à la création de la couche d'inversion côté source :

$$V_{th} = V_{fb} + 2\Phi_b + \gamma\sqrt{2\Phi_b + V_{sb}} \quad (II. 93)$$

où V_{fb} est la tension de bandes plates, Φ_b le potentiel de substrat, V_{sb} la tension source/substrat et γ le coefficient de substrat.

En pratique, on mesure cette tension de seuil par la méthode de Hamer [Hamer'86]. Cela correspond approximativement à l'annulation de la dérivée troisième du courant drain par rapport à la tension de grille sur la courbe expérimentale (point d'inflexion de la transconductance g_m).

Les modèles EEPROM3 et EEPROM4 ont été validés avec un bon accord entre mesures et simulations [Harabech'02].

Dans ce chapitre, nous avons décrit le modèle qui a été utilisé dans le cadre de notre thèse afin de simuler le comportement électrique statique et dynamique d'une cellule EEPROM. Ce modèle comporte deux parties : la modélisation du transistor d'état et la modélisation du transistor de sélection. La modélisation du transistor d'état est basé sur le calcul du potentiel de grille flottante, image de l'information stockée dans la cellule, et du potentiel de surface dans le canal. Ces deux quantités permettent ensuite de déterminer toutes les grandeurs électriques relatives au transistor mémoire. La simulation du modèle n'étant possible qu'en en connaissant les paramètres, nous avons également présenté la procédure de caractérisation utilisée pendant notre thèse afin d'extraire les paramètres du modèle. Cette caractérisation prend notamment en compte le phénomène de désertion dans la grille flottante en polysilicium.

Le principal atout de ce modèle est d'être basé sur la physique. Il n'utilise qu'une seule équation pour le courant de drain. Ce dernier est continu dans toutes les zones de fonctionnement et à la limite de ces différentes zones, contrairement aux modèles de type Spice [Foty'97]. Ce modèle est bien adapté à la simulation des petits circuits.

.

Références bibliographiques du chapitre II

- [Stratton'62] R. Stratton, "Volt-current characteristics for tunneling through insulating films", J. Phys. Chem. Solids, vol. 23, pp. 1177-1190, 1962.
- [Lenzlinger'69] M. Lenzlinger, E.H. Snow, "Fowler-Nordheim Tunneling in Thermally Grown SiO₂", J. Appl. Phys., vol. 40, pp. 278, 1969.
- [Sze'69] S. M. Sze, "Physics of semiconductor devices", Wiley, NewYork, 1969.
- [Stern'72] F. Stern, Phys. Rev. B, vol. 5, p. 4891, 1972.
- [Ando'76] T. Ando, Phys. Rev. B, vol. 13, p. 3468, 1976.
- [Weinberg'77] Z.A. Weinberg, Solid State Electronics, vol. 22, p. 11, 1977.
- [Hsing'79] C.T. Hsing, D.P. Kennedy, A.D. Sutherland and K.M. Van Vliet, Phys. Stat. Sol. (a), vol. 56, p.129, 1979.
- [Keiger'81] G. Keiger and R.M. Swanson, Journal of Applied Physics, vol. 52, p. 5710, 1981.
- [Weinberg'82] Z.A. Weinberg, Journal of Applied Physics, vol. 53, p. 5052, 1982.
- [Horiguci'85] S. Horiguci and H. Yosshino, Journal of Applied Physics, vol. 58, p. 1597, 1985.
- [Hamer'86] M.F. Hamer, "First-Order parameter extraction on enhancement MOS Transistor", IEEE Proceedings, vol.133, Pt.I, n°2, April 1986.
- [Tsividis'89] Yannis P. Tsividis, "Operation and Modeling of the MOS Transistor", MacGraw-Hill series in electrical engineering. VLSI, electronics, and electronics circuits, 1987.
- [Mathieu'90] H. Mathieu, "Physique des semi-conducteurs et des composants électroniques", 2^{ème} édition, Masson, 1990.
- [Klaassen'90] H. C. De Graaf, F. M. Klaassen, "Compact transistor modeling for circuit design", Computational Microelectronics ED. S. Selberherr, 1990.
- [Olivo'91] P. Olivo, J. Suné and B. Ricco, IEEE Electron Device Lett., vol. 12, pp. 620, 1991.
- [Suné'92] J. Suné, P. Olivo and B. Ricco, IEEE Trans. Electron Device, vol. 39, p. 1732, 1992.
- [Sheybani'94] E.O. Sheybani and Y.L. Chiou, Proc. Of Southcon'94 (March 94, Orlando, Florida, USA), p. 149, 1994.
- [Mohammad'95] S.N. Mohammad, G. Fiorenza, A. Acovic, J.B. Johnson and R; L. Carter, Solid State Electronics, vol. 38, p. 807, 1995.

[Bouchakour'96] R. Bouchakour, L. Hardy, M. Jourdain, "An analytical model for hot-carrier induced degradation for submicron nMOSFET", ESSDERC'96, Boulogne, 9-11 Septembre 1996.

[Bouchakour'97] R. Bouchakour, H. Petit, "Modèle compact de transistor MOS basé sur le calcul du potentiel de surface", rapport confidentiel, 1997.

[Foty'97] D. Foty, "Mosfet modeling with Spice, principles and practice", Prentice Hall PTR, 1997.

[Ando'98] T. Ando, A.B. Fowler and F. Stern, Review of Modern Physics, vol. 58, p. 437, 1998.

[Brown'98] W. D. Brown, J. E. Brewer, "Non-volatile semiconductor memory technology", IEEE PRESS, NewYork, 1998.

[Benzerti'99] W. Benzerti, "Modélisation et caractérisation de la cellule mémoire de type EEPROM pour la simulation et la conception de circuits intégrés analogiques et mixtes", thèse, Ecole Nationale Supérieure des Télécommunications, Paris, décembre 1999.

[Bouchakour'99] R. Bouchakour, H. Petit, N. Harabech, H. Recoules, JP. Sorbier, "Modèle compact de transistor MOS pour la conception de circuits de type basse consommation", FTFC'99, Paris, 26-28 mai, 1999.

[Crocì'99] S. Croci, J.M. Voisin, C. Plossu, C. Raynaud, J.L. Autran, P. Boivin, J.M. Mirabel, "Extraction and Evolution of Fowler-Nordheim tunneling parameters of thin gate oxides under EEPROM-like dynamic degradation", Microelectronics Reliability, vol. 39, p. 879-884, 1999.

[Harabech'00] N. Harabech, R. Bouchakour, P. Canet, P. Pannier, JP. Sorbier, "Extraction of Fowler-Nordheim Parameters in thin SiO₂ Oxide Film Including Polysilicon Gate Depletion : Validation with an EEPROM Memory Cell". IEEE ISCAS, Genève, 28 mai 2000.

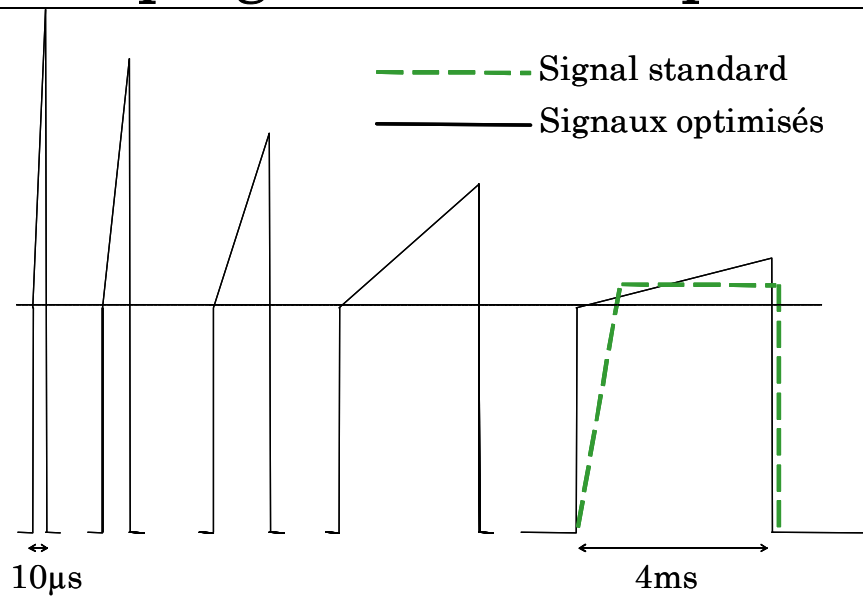
[Plossu'00] C. Plossu, "Caractérisation et fiabilité des oxydes minces SiO₂ dans les dispositifs MOS", Habilitation à Diriger des Recherches, Institut National des Sciences Appliquées de Lyon, 2000.

[Bouchakour'01] R. Bouchakour, N. Harabech, P. Canet, P. Boivin, JM. Mirabel, "Modeling of a floating-gate EEPROM cell using a charge sheet approach including variable tunneling capacitance and polysilicon gate depletion effect. IEEE ISCAS, Sydney, 6-9 mai 2001.

[Chiou'01] Y.L. Chiou, J.P. Gambino, M. Mohammad, "Determination of the Folwer-Nordheim tunneling parameters from the Folwer-Nordheim plot", Solid-State Electronics, vol. 45, p. 1781-1791, 2001.

[Harabech'02] N. Harabech, Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications, Paris, Janvier 2002.

Chapitre III : Optimisation des signaux et programmation rapide



Signaux de programmation rapides d'une mémoire EEPROM.

III.1 INFLUENCE DU TRANSISTOR DE SELECTION.....	96
III.2 OPTIMISATION DES SIGNAUX DE PROGRAMMATION.....	97
III.2.1 Démarche phénoménologique	98
III.2.2 Résultats de simulations avec un signal optimisé.....	105
III.2.3 Méthodologie de conception du signal optimisé	107
III.3 PROGRAMMATION RAPIDE.....	108
III.3.1 Principe de la programmation rapide.....	109
III.3.2 Mesure de la fenêtre de programmation	111
III.3.3 Résultats en endurance	112
III.3.3.1 Fermeture de la fenêtre de programmation.....	112
III.3.3.2 Courbe caractéristique.....	112
III.4 PROGRAMMATIONS NON-STANDARDS.....	113
III.4.1 Programmation répartie : impact sur la fiabilité.....	113
III.4.1.1 Programmmations classique et répartie	113
III.4.1.2 Simulations	115
III.4.1.3 Mesures	117
III.4.1.4 Test en endurance et interprétation	118
III.4.2 Programmation par trains d'impulsions	119
Références bibliographiques du chapitre III.....	125

Le développement de la technologie des mémoires non-volatiles tend à réaliser des mémoires plus rapides, plus fiables, plus denses ou moins chères.

Pour améliorer la fiabilité des mémoires non volatiles, plusieurs solutions sont possibles. En général, on durcit la technologie avec des matériaux plus performants (haute permittivité diélectrique par exemple). On peut aussi améliorer le dessin des cellules (électronique spatiale).

Ces solutions technologiques requièrent une phase de recherche et développement longue et coûteuse. La voie que nous avons explorée consiste à optimiser le signal de programmation de la cellule, sans diminuer la fenêtre de programmation afin d'améliorer l'endurance de la cellule, de réduire le temps de programmation et de réduire les tensions de polarisation.

La cellule EEPROM est constituée de deux transistors. Nous abordons d'abord dans ce chapitre l'impact du transistor de sélection sur la programmation. Ensuite, nous nous intéressons à la programmation à proprement parler du transistor d'état.

Canet a montré qu'une cellule EEPROM programmée avec un signal optimisé a une endurance accrue. En effet, habituellement, le signal de programmation d'une cellule EEPROM est un trapèze constitué d'une rampe de montée, d'un plateau, et d'une rampe de descente. Le champ électrique à travers l'oxyde tunnel augmente avec la vitesse de rampe. Ainsi, il est possible de diminuer ce champ électrique en diminuant la rampe du signal appliqué et en adaptant les durées de la rampe et du plateau. Les travaux exposés dans ce chapitre s'inscrivent dans la continuité de cette optimisation du signal de programmation.

Ce chapitre comporte trois grandes parties.

La première partie est consacrée à la description des signaux optimisés. Nous présenterons deux démarches différentes conduisant à la mise en œuvre des signaux optimisés pour programmer une cellule EEPROM. La première démarche montre, à l'aide d'un modèle simplifié de cellule EEPROM, que les signaux de programmation standards provoquent un pic de champ électrique à travers l'oxyde tunnel proportionnel à la rampe du signal de programmation. Ainsi, en diminuant la rampe du signal de programmation, on diminue le champ électrique maximum et on augmente la fiabilité de la cellule : le signal est ainsi optimisé.

La deuxième démarche repose sur l'idée de supprimer ce pic de champ en travaillant à champ constant sur la totalité de la durée du signal, c'est-à-dire d'injecter la charge à courant constant. On utilise alors notre simulateur pour définir le signal de programmation optimisé.

Dans la deuxième partie de ce chapitre, nous explorerons les conséquences de la diminution de la durée de programmation sur le comportement d'une cellule EEPROM. En effet, Canet *et al.* [Canet'01b] ont réussi à écrire et effacer une cellule EEPROM en 550 μ s. Ils ont gagné un facteur 8 par rapport aux 4ms habituellement utilisées tout en gardant la même fenêtre de programmation (≈ 5 V). Dans cette partie de notre thèse, nous diminuons le temps de programmation de la cellule jusqu'à 10 μ s en gardant la même fenêtre de programmation que celle qui est obtenue avec un signal qui dure 4ms.

Programmer aussi rapidement nécessite l'utilisation de tensions de programmation particulièrement élevées (jusqu'à 19V) qui provoquent des champs électriques très importants à travers l'oxyde tunnel. Afin de quantifier les conséquences de tels champs sur la fiabilité de nos cellules, nous avons soumis différentes cellules à des tests en endurance.

Dans la dernière partie de ce chapitre, nous présenterons quelques résultats relatifs à la baisse des tensions de polarisation d'une cellule EEPROM. Nous verrons notamment comment baisser la polarisation de chacune des électrodes d'une cellule EEPROM en répartissant la tension entre la grille et le drain, et quelles en sont les conséquences sur l'endurance de la cellule. Enfin, nous avons investigué une programmation multi-impulsionnelle composée d'une multitude d'impulsions de durée beaucoup plus courte.

III.1 Influence du transistor de sélection

La programmation d'une cellule EEPROM dépend de deux signaux : le signal appliqué sur le transistor de sélection et le signal appliqué sur le transistor d'état. Ainsi, l'optimisation de la programmation concerne les deux signaux. Nous allons nous intéresser, dans cette première partie, à l'influence du transistor de sélection sur la programmation de la cellule.

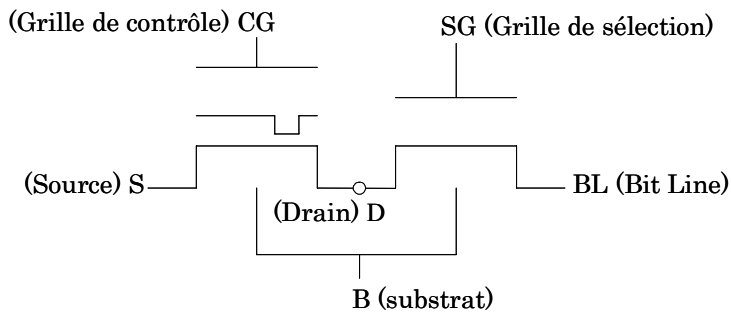


Figure III. 1 : Cellule mémoire EEPROM

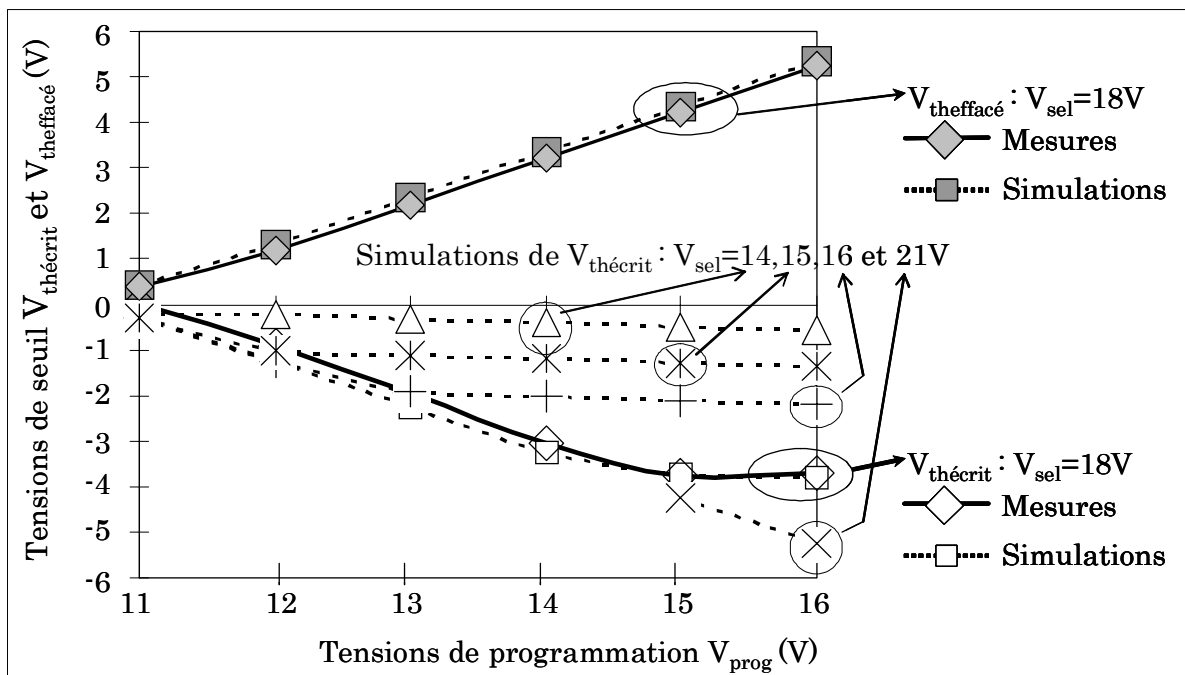


Figure III. 2 : Effet de la tension de grille de sélection V_{sel} sur la tension de seuil en écriture V_{thcrit} .

Le rôle du transistor de sélection est de transférer le signal appliqué sur la « bit line » (drain du transistor de sélection) sur le drain du transistor d'état (source du transistor de sélection). Ce transistor doit fonctionner en

interrupteur, avec une tension grille/canal plutôt élevée (V_{sel}) pour transmettre une tension égale à 0V en effacement, ou V_{prog} en écriture.

Afin d'étudier l'impact du transistor de sélection sur la programmation, nous mesurons la tension de seuil d'une cellule EEPROM écrite et effacée avec une tension de programmation variant de 11 à 16V (cf. figure III.2). Pour chaque tension de programmation, nous mesurons les tensions de seuil en écriture et en effacement, respectivement $V_{thécrit}$ et $V_{theffacé}$, pour V_{sel} égal à 14, 15, 16, 18 et 21V. Les résultats de cette manipulation sont représentés sur la figure III.2.

Nous constatons qu'il faut, au minimum, un écart de 3V entre V_{sel} et V_{prog} afin d'éviter une saturation de la tension de seuil en écriture. Ce phénomène est problématique lorsqu'il est nécessaire de programmer une cellule EEPROM avec des tensions élevées. En effet, si V_{prog} est égal 18V, par exemple, il faut que V_{sel} soit supérieure à 21V pour éviter la saturation de $V_{thécrit}$.

La tension de seuil en écriture d'une cellule EEPROM dépend donc de la polarisation appliquée sur la grille de sélection. En effet, l'écriture se fait par couplage capacitif entre le drain N^+ du transistor d'état et la grille flottante. Or, le potentiel de drain dépend de la tension V_{sel} appliquée sur la grille de sélection : si V_{sel} est trop faible, la tension effective sur le drain du transistor d'état sera inférieure à la tension de programmation V_{prog} appliquée sur le drain du transistor de sélection. Cette chute de tension est parfaitement prise en compte par le modèle SPICE 3 du transistor de sélection : les simulations sont conformes aux mesures pour une tension de programmation variant de 11 à 16V (fig. III.2).

III.2 Optimisation des signaux de programmation

Le choix du signal de programmation d'une cellule EEPROM est crucial car il détermine, à court terme, la largeur de la fenêtre de programmation et, à long terme, la fiabilité de la cellule. Lorsque l'on efface la cellule mémoire, l'on applique généralement le signal standard représenté sur la figure III.3 sur la grille de contrôle, le drain et le substrat étant à la masse, la source flottante, et la grille de sélection polarisée à un potentiel V_{sel} fortement positif qui le rend passant. Lorsque l'on écrit la cellule mémoire, l'on applique ce même signal sur le drain, la grille de contrôle étant à la masse, la source flottante, et la grille de sélection polarisée au potentiel V_{sel} .

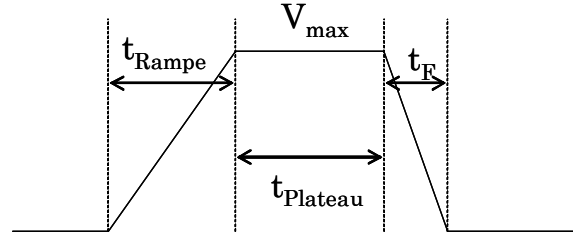


Figure III. 3 : *Signal de programmation standard*

$V_{\max}$, t_{Rampe} , t_{Plateau} et t_F sont respectivement la tension maximale, la durée de la rampe, la durée du plateau, et le temps de descente du signal appliqué lors de l'effacement ou de l'écriture.

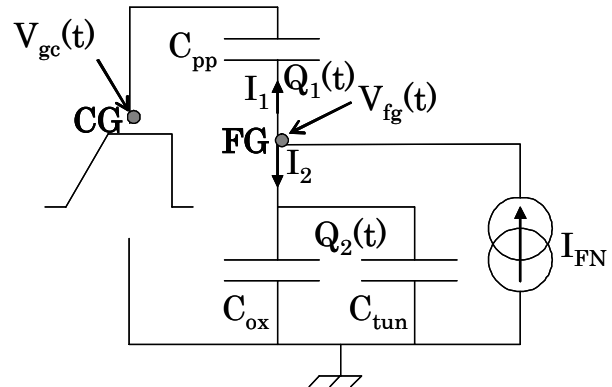
Dans le paragraphe suivant, nous allons étudier l'évolution des grandeurs électriques de la cellule mémoire en fonction du temps lorsque l'on programme une cellule mémoire avec un tel signal. Nous étudierons en particulier l'évolution du potentiel de grille flottante V_{fg} et celle du champ électrique E_{tun} en fonction du temps. En effet, ces deux grandeurs sont directement reliées à la dégradation de l'oxyde tunnel, et par là – même, à la fiabilité du dispositif, après un grand nombre de cycles écriture/effacement.

III.2.1 Démarche phénoménologique

La modélisation rigoureuse de la cellule mémoire a été décrite dans le deuxième chapitre. Cette modélisation est, comme nous l'avons vu, basée sur le calcul du potentiel de surface et du potentiel de grille flottante. Afin de mettre en évidence l'intérêt des signaux optimisés, nous allons utiliser un modèle simplifié d'une cellule EEPROM qui a l'avantage d'être basé sur des équations ayant une solution analytique.

La figure III.4 représente un schéma électrique équivalent simplifié de la zone d'injection tunnel d'une cellule EEPROM lors de la phase d'effacement. A partir de ce schéma équivalent et en utilisant les lois de l'électrocinétique, nous allons établir les équations qui gouvernent l'évolution des grandeurs électriques de la cellule.

Figure III. 4 : *Schéma électrique équivalent simplifié d'une cellule EEPROM et signal appliqué lors de la phase d'effacement.*



La structure équivalente simplifiée est composée de la mise en série de la capacité inter poly C_{pp} avec la capacité d'oxyde C_{ox} , la capacité C_{tun} de la zone tunnel et la source de courant Fowler-Nordheim I_{FN} , ces trois dernières étant en parallèle. Les capacités C_{pp} , C_{ox} et C_{tun} sont considérées constantes. $V_{gc}(t)$ est le signal appliqué sur la grille de contrôle en fonction du temps, tandis

que $V_{fg}(t)$ est le potentiel de grille flottante correspondant. $Q_1(t)$ et $Q_2(t)$ sont respectivement la charge aux bornes de la capacité C_{pp} et la charge aux bornes de la capacité équivalente ($C_{tun}+C_{ox}$).

Les courants I_1 et I_2 sont la variation de charge par unité de temps aux bornes, d'une part de C_{pp} , d'autre part de ($C_{tun}+C_{ox}$). Leurs expressions sont données par les relations suivantes :

$$I_1 = \frac{dQ_1(t)}{dt} \quad (III. 01)$$

$$I_2 = \frac{dQ_2(t)}{dt} \quad (III. 02)$$

où

$$Q_1(t) = C_{pp} \times [V_{fg}(t) - V_{gc}(t)] \quad (III. 03)$$

et

$$Q_2(t) = (C_{pp} + C_{tun}) \times [V_{fg}(t) - 0] \quad (III. 04)$$

En appliquant la loi des nœuds au point FG, il vient :

$$I_1 + I_2 - I_{FN} = 0 \quad (III. 05)$$

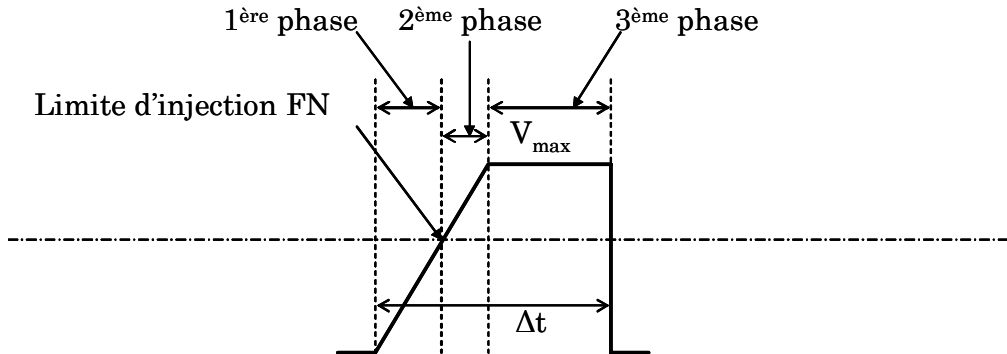


Figure III. 5 : Les trois phases considérées pour notre analyse de l'évolution du potentiel de grille flottante.

Les relations (III.01), (III.02) et (III.05) donnent :

$$I_{FN} = \frac{dQ_1(t)}{dt} + \frac{dQ_2(t)}{dt} \quad (III. 06)$$

soit, en combinant (III.03), (III.04) et (III.06) :

$$I_{FN} = C_{pp} \frac{d[V_{fg}(t) - V_{gc}(t)]}{dt} + (C_{ox} + C_{tun}) \frac{dV_{fg}(t)}{dt} \quad (III. 07)$$

ou encore :

$$I_{FN} = C_x \frac{dV_{fg}(t)}{dt} - C_{pp} \frac{dV_{gc}(t)}{dt} \quad (III. 08)$$

Nous pouvons découper l'effacement en trois phases distinctes : pendant les première et deuxième phases, le signal $V_{gc}(t)$ appliqué sur la grille de contrôle croît en fonction du temps avec une rampe r .

La limite entre la première et la deuxième phase correspond au potentiel de grille flottante pour lequel le courant tunnel Fowler-Nordheim commence à apparaître de manière significative. Pendant la dernière phase, le signal $V_{gc}(t)$ est constant (plateau) : sa dérivée $dV_{gc}(t)/dt$ est nulle. Notre analyse de l'évolution de $V_{fg}(t)$ comportera aussi trois parties distinctes, correspondant aux trois phases ci-dessus.

i. Phase 1 : évolution de $V_{fg}(t)$ avant injection tunnel.

Pendant cette première phase, le courant tunnel Fowler-Nordheim étant négligeable, l'équation (III.08) se réduit à :

$$0 = C_x \frac{dV_{fg}(t)}{dt} - C_{pp} \frac{dV_{fg}(t)}{dt} \quad (\text{III. 09})$$

soit :

$$\frac{dV_{fg}(t)}{dt} = K_e \frac{dV_{gc}(t)}{dt} \quad (\text{III. 10})$$

La relation (III.10) montre que, tant que I_{FN} est nul, le potentiel de grille flottante $V_{fg}(t)$ croît proportionnellement à la tension $V_{gc}(t)$ appliquée sur la grille de contrôle avec le coefficient K_e , inférieur à 1.

ii. Phase 2 : évolution de $V_{fg}(t)$ du seuil d'injection à la fin de la rampe.

La deuxième phase s'étend du seuil d'injection jusqu'à la fin de la rampe du signal $V_{gc}(t)$.

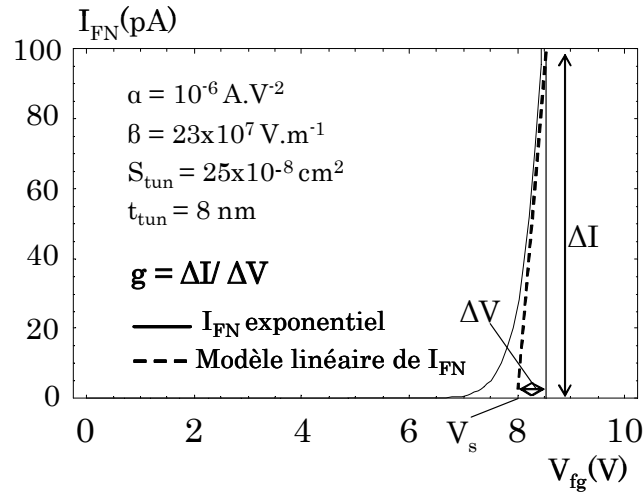


Figure III. 6 : Modèle linéaire du courant FN

Le modèle décrit dans le deuxième chapitre nous permet de simuler le courant d'injection I_{FN} et la tension de grille flottante V_{fg} . L'injection est alors significative pour $I_{FN} \geq 20 \text{ pA}$ (cf. figure III.6).

On peut donc exprimer le courant I_{FN} avec une expression analytique linéaire similaire à un modèle de diode (cf. figure III.6):

$$I_{FN} = -g[V_{fg}(t) - V_s] \quad (III. 11)$$

où g est la conductance et V_s la tension de seuil caractérisant le modèle. La tension appliquée sur la grille de contrôle vaut, en fonction du temps :

$$V_{gc}(t) = rt \quad (III. 12)$$

où r , rampe de la tension de grille, est égale à :

$$r = \frac{dV_{gc}(t)}{dt} \quad (III. 13)$$

En injectant dans l'équation (III.08) l'expression de I_{FN} donnée par la relation (III.11), nous obtenons l'équation différentielle du premier ordre en V_{fg} suivante :

$$C_x \frac{dV_{fg}(t)}{dt} + gV_{fg}(t) - gV_s - rC_{pp} = 0 \quad (III. 14)$$

En intégrant de l'instant t_s , où le courant Fowler-Nordheim commence à apparaître et où le potentiel de grille flottante V_{fg} atteint V_s , à l'instant t , il vient :

$$V_{fg}(t) = V_s + \frac{rC_{pp}}{g} + A \exp\left(-\frac{gt}{C_x}\right) \quad (III. 15)$$

Sachant que $V_{fg}(t_s)$ est égal à V_s , l'on détermine la constante d'intégration A et il vient finalement :

$$V_{fg}(t) = V_s + \frac{rC_{pp}}{g} \left[1 - \exp\left(-\frac{g(t-t_s)}{C_x}\right) \right] \quad (III. 16)$$

En fin de rampe, le terme exponentiel devient très faible devant 1 et le potentiel de grille flottante V_{fg} atteint sa valeur maximale V_{fgmax} donnée par :

$$V_{fgmax} = V_s + \frac{rC_{pp}}{g} \quad (III. 17)$$

Pour diminuer V_{fgmax} , il faut diminuer V_s , c'est-à-dire le courant d'injection. Le courant étant fixé, V_s et g sont fixés. Pour diminuer encore V_{fgmax} , il faut diminuer la rampe r . En fait, en diminuant la rampe $r=dV_{gc}/dt$, on diminue en même temps le courant FN et donc V_s .

Comme le champ électrique à travers l'oxyde tunnel et le potentiel de grille flottante sont proportionnels lors de la phase d'effacement, nous pouvons affirmer, dès ce stade, que la dégradation de l'oxyde tunnel sera d'autant plus importante que la rampe du signal de programmation est forte. Par conséquent, pour augmenter la fiabilité d'une cellule EEPROM, une des possibilités consiste à diminuer la rampe du signal de programmation utilisé.

iii. Phase 3 : évolution de $V_{fg}(t)$ pendant le plateau

Pendant la troisième et dernière phase, l'approximation linéaire du courant n'est plus valable car ce dernier décroît. Dans ce cas, son expression est donnée par la relation (II.18) (cf. §II.1.3), dans laquelle on remplace E_{tun} par V_{fg}/t_{tun} :

$$I_{FN} = \alpha \left(\frac{V_{fg}}{t_{tun}} \right)^2 S_{tun} \exp \left(-\beta \frac{t_{tun}}{V_{fg}} \right) \quad (III. 18)$$

En combinant les relations (III.08) et (III.18), d'une part, et en remarquant que le signal $V_{gc}(t)$ est constant pendant la dernière période, donc de dérivée temporelle nulle, d'autre part, il vient :

$$-\alpha \left(\frac{V_{fg}}{t_{tun}} \right)^2 S_{tun} \exp \left(-\beta \frac{t_{tun}}{V_{fg}} \right) = C_x \frac{dV_{fg}}{dt} \quad (III. 19)$$

En intégrant (III.19), nous obtenons :

$$-\frac{1}{\beta} t_{tun} \exp \left(\beta \frac{t_{tun}}{V_{fg}} \right) = \frac{-\alpha S_{tun}}{C_x} (t + C) \quad (III. 20)$$

où C est la constante d'intégration.

On obtient ainsi l'expression finale de V_{fg} pendant la dernière période :

$$V_{fg}(t) = \frac{\beta t_{tun}}{\ln \left[\frac{\alpha \beta (t + C) S_{tun}}{t_{tun} (C_{pp} + C_{ox} + C_{tun})} \right]} \quad (III. 21)$$

Cette relation montre que le potentiel de grille flottante décroît en fonction du temps pendant la dernière période.

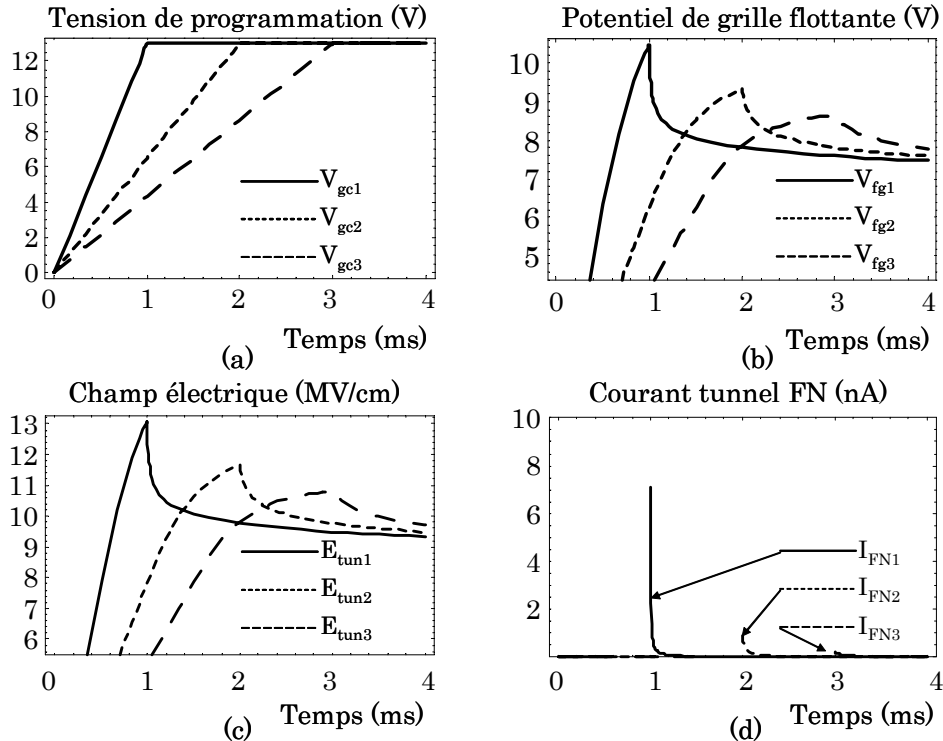


Figure III. 7 : Illustration de l'augmentation du champ électrique maximal avec la rampe sur trois signaux de rampe croissante : (a) tensions de programmation de rampe croissante, (b) potentiel de grille flottante, (c) champ électrique à travers l'oxyde tunnel et (d) courant tunnel FN.

Afin d'illustrer l'augmentation du champ électrique maximum avec la rampe, nous représentons sur la figure III.7 le potentiel de grille flottante, le champ électrique et le courant FN à travers l'oxyde tunnel, en effacement, pour trois signaux à la rampe croissante.

La figure III.7c montre que plus la rampe du signal de programmation est forte, plus le champ électrique maximum à travers l'oxyde tunnel est élevé. Ce champ atteint 13MV/cm pour V_{fg1} . Afin de diminuer le champ, et pour réduire la rampe, on peut supprimer les première et troisième phases du signal (fig. III.5) et utiliser au mieux toute la durée du signal pour la deuxième phase pour que la rampe soit la plus faible possible. En fin de compte, on obtient le signal optimisé de programmation représenté sur la figure III.8.

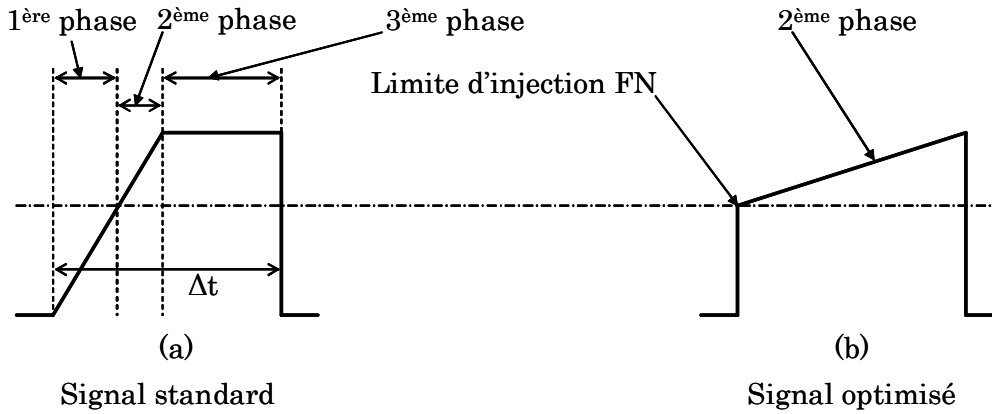


Figure III. 8 : Du signal standard au signal optimisé.

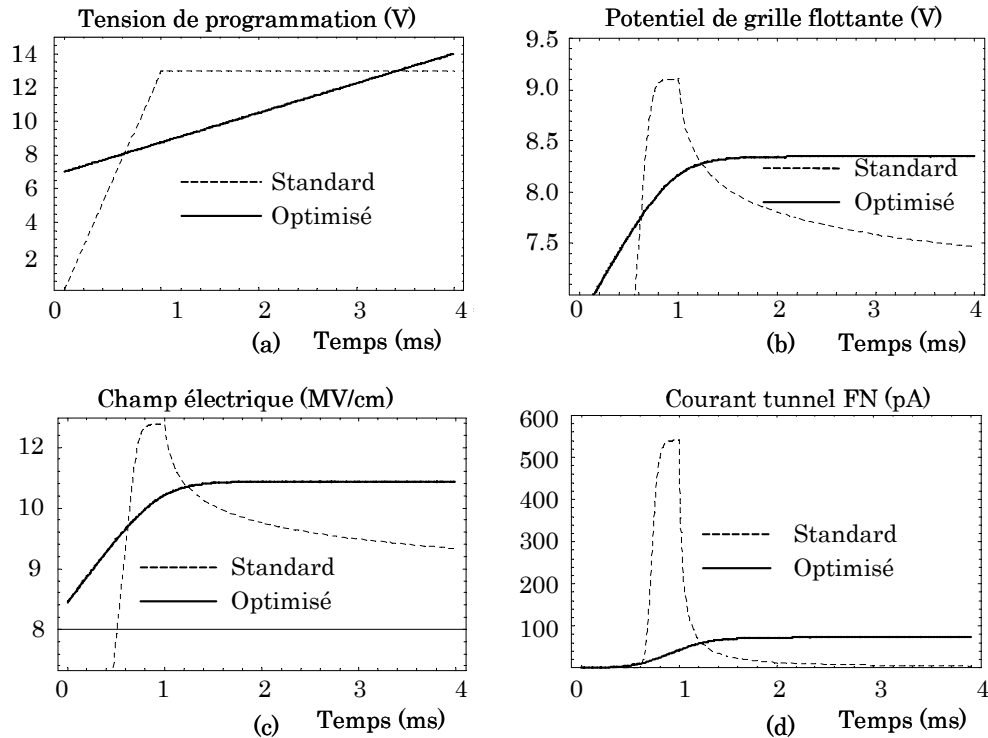


Figure III. 9 : Comparaison du signal standard et du signal optimisé : (a) tensions de programmation, (b) potentiel de grille flottante, (c) champ électrique à travers l'oxyde tunnel et (d) courant tunnel FN.

Nous représentons sur la figure III.9 le potentiel de grille flottante, le champ électrique et le courant FN résultant d'un tel signal. On constate qu'on a fait disparaître le pic de champ électrique provoqué par le signal standard.

Dans ce paragraphe, nous allons calculer analytiquement les tensions de programmation qu'il faut appliquer aux électrodes d'une cellule EEPROM afin d'obtenir un champ électrique constant à travers l'oxyde tunnel. Afin de simplifier les calculs, nous considérerons seulement la phase d'effacement : pendant cette phase, le drain et le substrat sont à la masse, la source flottante, la grille de sélection polarisée à $+V_{sel}$ et la grille de contrôle polarisée avec le signal V_{gc} , la tension de programmation que nous voulons déterminer. Le même raisonnement peut ensuite être mis en œuvre afin de déterminer la tension de programmation en écriture.

Nous supposons aussi que le potentiel de surface et la tension de bandes plates à l'interface drain/oxyde tunnel sont négligeables par rapport à la chute de potentiel V_{tun} à travers l'oxyde tunnel. Ainsi, le champ à travers l'oxyde tunnel s'écrit :

$$E_{tun} = \frac{V_{tun}}{t_{tun}} \quad (III. 22)$$

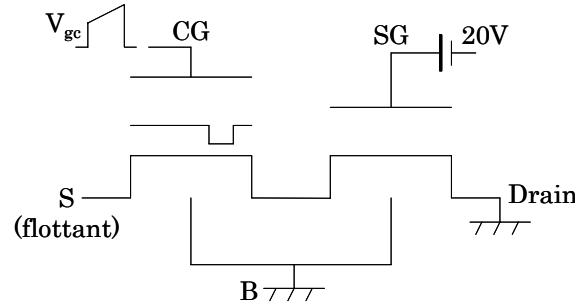


Figure III. 10 : *Polarisations des électrodes d'une cellule EEPROM pendant la phase d'effacement.*

En utilisant la relation (III.18), l'expression du courant FN devient :

$$I_{FN0} = \alpha S_{tun} \left(\frac{V_{tun}}{t_{tun}} \right)^2 \exp \left(-\beta \frac{t_{tun}}{V_{tun}} \right) \quad (III. 23)$$

Pendant la phase d'effacement, V_d est nul et le terme $C_{ox}(V_{fb} + \Psi_s)$ est négligeable devant $C_{pp}V_{gc}$ dans la relation (II.40) (cf. §II.2.1.2). Par ailleurs, Q_{fg} n'est autre que la somme de la charge initiale de grille flottante Q_{fgini} et de la charge due à l'injection FN. Ainsi, en utilisant la relation (II.40) (cf. §II.2.1.2), le potentiel de grille flottante V_{fg0} s'exprime comme suit en fonction de I_{FN0} et des capacités mises en jeu :

$$V_{fg0} = K_e V_{gc}(\tau) + \frac{1}{C_x} \left(Q_{fgini} - \int_{t=0}^{\tau} I_{FN0} dt \right) \quad (III. 24)$$

Dans la relation (III.24), le signe "moins" traduit l'injection de charges négatives dans la grille flottante. En effacement, le potentiel de grille flottante V_{fg0} et la

chute de potentiel V_{tun} à travers l'oxyde tunnel sont égaux car $V_d=0$. Ainsi, la relation (III.24) peut se réécrire comme suit :

$$V_{\text{gc}}(\tau) = \frac{V_{\text{tun}}}{K_e} - \frac{Q_{\text{fgini}}}{C_{\text{pp}}} + \frac{1}{C_{\text{pp}}} \int_{t=0}^{t=\tau} |I_{\text{FN0}}| dt \quad (\text{III. 25})$$

Après intégration du dernier terme, la relation (III.25) devient :

$$V_{\text{gc}}(\tau) = V_{\text{gcmin}} + \frac{V_{\text{gcmax}} - V_{\text{gcmin}}}{\Delta t} \tau \quad (\text{III. 26})$$

où V_{gcmin} et V_{gcmax} sont déterminés en résolvant le système constitué par les relations (III.27) et (III.28) :

$$V_{\text{gcmin}} = \frac{V_{\text{tun}}}{K_e} - \frac{Q_{\text{fgini}}}{C_{\text{pp}}} \quad (\text{III. 27})$$

$$V_{\text{gcmax}} = V_{\text{gcmin}} + \frac{|I_{\text{FN0}}|}{C_{\text{pp}}} \Delta t \quad (\text{III. 28})$$

La relation (III.26) montre que la tension de programmation $V_{\text{gc}}(t)$ doit être une rampe positive allant de V_{gcmin} à V_{gcmax} comme le montre la figure III.11. La durée Δt du signal est fixée arbitrairement.

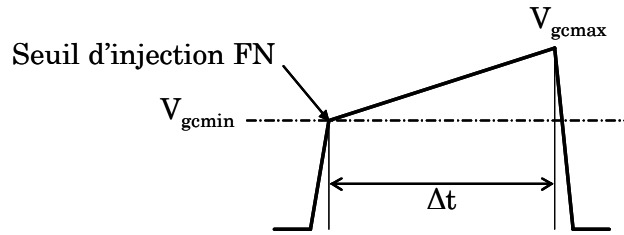


Figure III. 11 : Représentation schématique du signal optimisé avec, notamment, le seuil d'injection FN.

III.2.2 Résultats de simulations avec un signal optimisé

Le raisonnement exposé ci-dessus, dont le but est de démontrer analytiquement la nécessité du signal optimisé, s'appuie sur un certain nombre d'hypothèses simplificatrices. L'on suppose, par exemple, que le terme $C_{\text{ox}}(V_{\text{fb}} + \Psi_s)$ est négligeable devant $C_{\text{pp}}V_{\text{gc}}$. En réalité, le modèle que nous avons utilisé pour simuler le fonctionnement d'une cellule EEPROM, plus physique, s'affranchit de telles hypothèses et il est plus proche de la réalité. La présence d'équations transcendantes, dans ce modèle, rend impossible une expression analytique du signal optimisé telle que celle donnée par la relation (III.26).

La figure III.12 représente les résultats de la simulation en régime transitoire d'une cellule EEPROM programmée avec des signaux optimisés. Le modèle utilisé est EEPROM4 (cf. §II.2.2.3). Nous avons simulé la chute de potentiel V_{tun} à travers l'oxyde tunnel (b), le potentiel de grille flottante V_{fg} (b), le champ électrique E_{tun} à travers l'oxyde tunnel (c), le courant tunnel Fowler-Nordheim I_{FN} (d) ainsi que la charge de grille flottante Q_{fg} (e) d'une cellule EEPROM. Les

signaux optimisés appliqués sur la grille de contrôle et sur le drain sont représentés sur la figure III.12a.

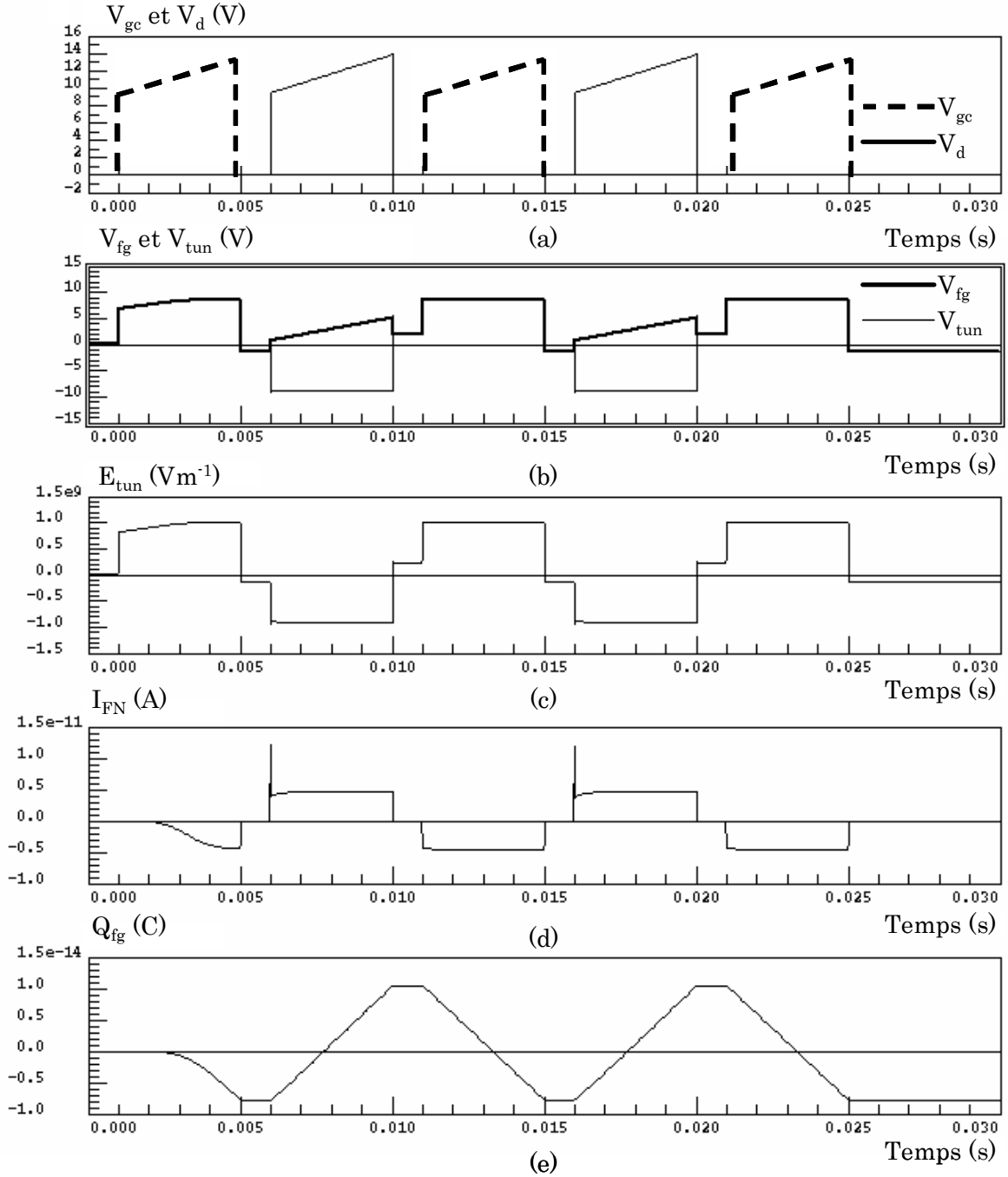


Figure III. 12 : *Simulations de V_{fg} , V_{tun} , E_{tun} , I_{FN} et Q_{fg} obtenues avec notre modèle compact de cellule EEPROM pour un signal optimisé.*

On constate que :

i.

V_{tun} , E_{tun} et I_{FN} sont quasiment constants pendant la rampe du signal optimisé. Néanmoins, aux instants $t=6ms$ et $16ms$, I_{FN} présente un pic. Ce pic est dû à la présence, au début du signal optimisé (cf. figure III.13), d'une forte rampe de l'ordre de $10V/\mu s$ mais dont nous n'avons pas tenu compte, afin de simplifier

l'analyse, pour établir l'expression analytique du signal optimisé donnée par la relation (III.26).

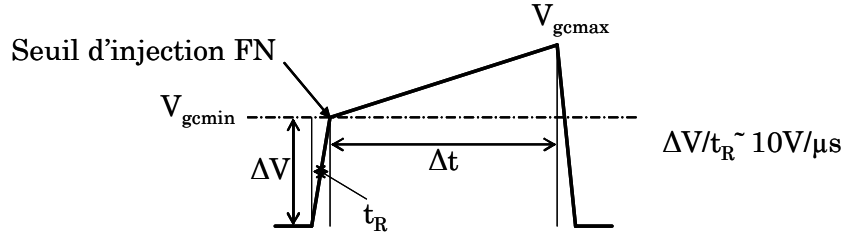


Figure III. 13 : *Signal optimisé avec la forte rampe initiale.*

ii.

Le signal de programmation étant calibré pour donner un courant I_{FN} constant pendant la durée du signal, on retrouve logiquement une charge Q_{fg} qui est une fonction affine du temps sur cette même durée.

Remarquons que V_{tun} et V_{fg} sont confondus en effacement (figure III.12b).

En effet, en première approximation :

$$V_{tun} = V_{fg} - V_d \quad (III. 29)$$

Or, en effacement, $V_d=0$. Ainsi, $V_{tun}=V_{fg}$.

III.2.3 Méthodologie de conception du signal optimisé

Lors de la conception d'une nouvelle cellule, la technologie utilisée impose une tension maximale V_{max} pour les signaux de programmation et l'application impose la durée du signal Δt . Quelle forme de signal doit-on adopter pour obtenir une fenêtre de programmation donnée ?

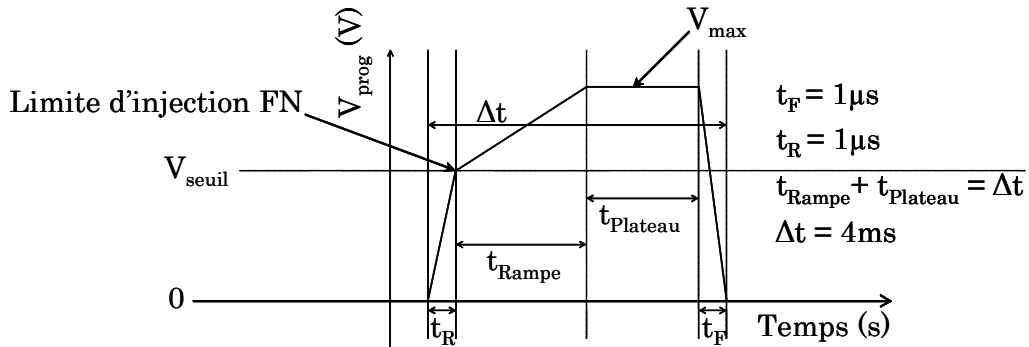


Figure III. 14 : *Schéma de la méthodologie de conception du signal de programmation de type 2.*

Le signal présenté sur la figure III.14 permet de décrire le signal standard ($V_{seuil}=0$) ou le signal optimisé ($t_{Plateau}=0$). La tension V_{seuil} est le seuil d'injection qui dépend de t_{tun} , épaisseur de l'oxyde tunnel.

Le seul degré de liberté qui nous reste est la durée relative entre la rampe, t_{Rampe} , et le plateau, $t_{Plateau}$. Selon la durée de la rampe, nous avons un signal carré

($t_{\text{Rampe}}=0$, $t_{\text{Plateau}}=4\text{ms}$), un signal intermédiaire ($t_{\text{Rampe}}\neq 0$, $t_{\text{Plateau}}\neq 0$), ou un signal optimisé ($t_{\text{Rampe}}=4\text{ms}$, $t_{\text{Plateau}}=0$).

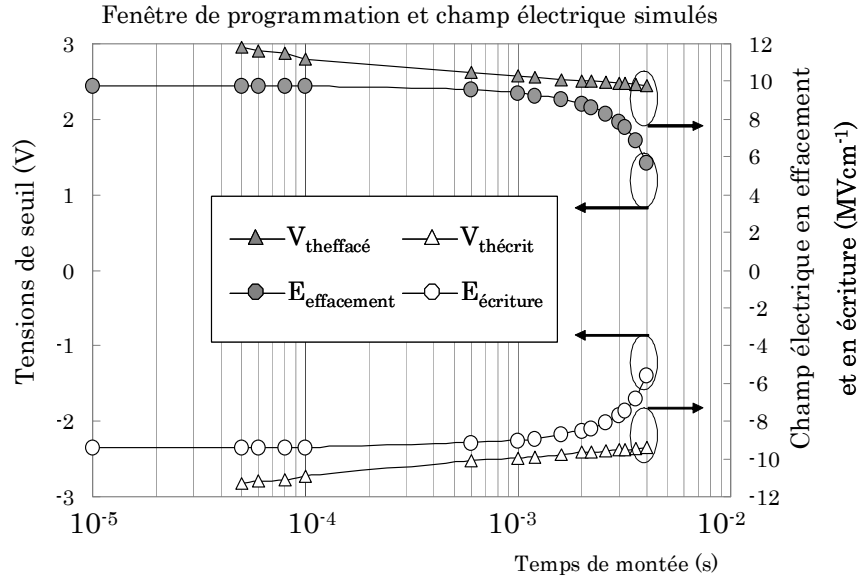


Figure III. 15 : *Simulation de la fenêtre de programmation et du champ électrique en fonction de la durée t_{Rampe} de la rampe du signal optimisé type 2 représenté sur la figure III.14.*

Pour une tension $V_{\text{max}}=13\text{V}$ et une durée totale du signal $\Delta t=4\text{ms}$, la figure III.15 représente les tensions de seuil en écriture et en effacement en fonction de t_{Rampe} . Cette figure montre que la fenêtre de programmation est maximale lorsque $t_{\text{Rampe}}=0$, autrement dit un signal carré. La fenêtre de programmation ne peut que diminuer si on augmente t_{Rampe} . Si la fenêtre obtenue avec le signal carré est trop faible, il faudra augmenter la tension V_{max} , mais cela sera difficile compte tenu de la technologie, ou la durée Δt .

Néanmoins, un tel signal induit des champs électriques très élevés ($>12\text{MV.cm}^{-1}$), néfastes pour l'endurance de la cellule. Le bon compromis consiste à augmenter t_{Rampe} afin d'augmenter la fiabilité de la cellule. On pourrait, par exemple, choisir $t_{\text{Rampe}}=1\text{ms}$. En effet, pour cette valeur, la fenêtre de programmation reste quasiment égale à sa valeur maximale alors que le champ électrique diminue d'au moins 2MV.cm^{-1} . Le signal est alors optimisé en fiabilité sans affecter le fonctionnement de la cellule.

III.3 Programmation rapide

Les mémoires EEPROM ont une faible consommation par rapport aux Flash-EEPROM mais leur temps de programmation est long. S'il est possible d'améliorer l'endurance de la cellule en diminuant le champ électrique à travers l'oxyde tunnel par l'utilisation de signaux optimisés, nous allons montrer dans ce

paragraphe qu'il est possible de diviser par dix le temps de programmation de la cellule tout en gardant la même endurance. Dans cette partie de notre thèse, nous avons cherché à réduire davantage la durée de programmation de la cellule EEPROM et à examiner les conséquences de cette réduction sur les performances de la cellule.

III.3.1 Principe de la programmation rapide

La charge injectée dans la grille flottante d'une cellule EEPROM est l'intégrale temporelle du courant I_{FN} . Or, pour un signal optimisé, I_{FN} est constant. Ainsi, pour un signal optimisé, la charge Q_{fg} injectée dans la grille flottante est simplement le produit du courant I_{FN} par la durée du signal :

$$Q_{fg} = I_{FG} \times \Delta t \quad (III. 30)$$

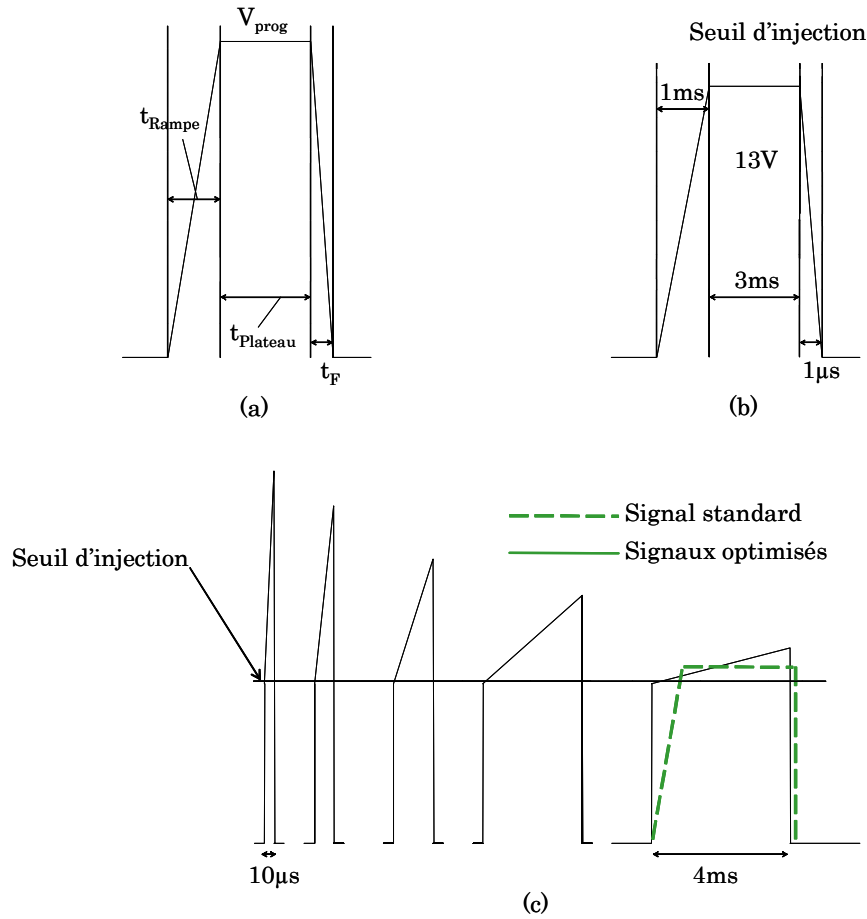


Figure III. 16 : *Signal standard (a), signal standard de 4ms à 13V (b), et signaux optimisés utilisés pour la programmation rapide (c).*

Pour programmer rapidement une cellule EEPROM, il faut diminuer Δt . Cette diminution de Δt doit être compensée par une augmentation de I_{FN} afin de garder Q_{fg} constant dans la relation (III.30). Augmenter I_{FN} revient à augmenter le champ électrique à travers l'oxyde tunnel et la tension de programmation. Il est évident qu'il est impossible de réduire Δt indéfiniment car cela nécessiterait une

tension de programmation très élevée et un champ électrique proche des champs de claquage.

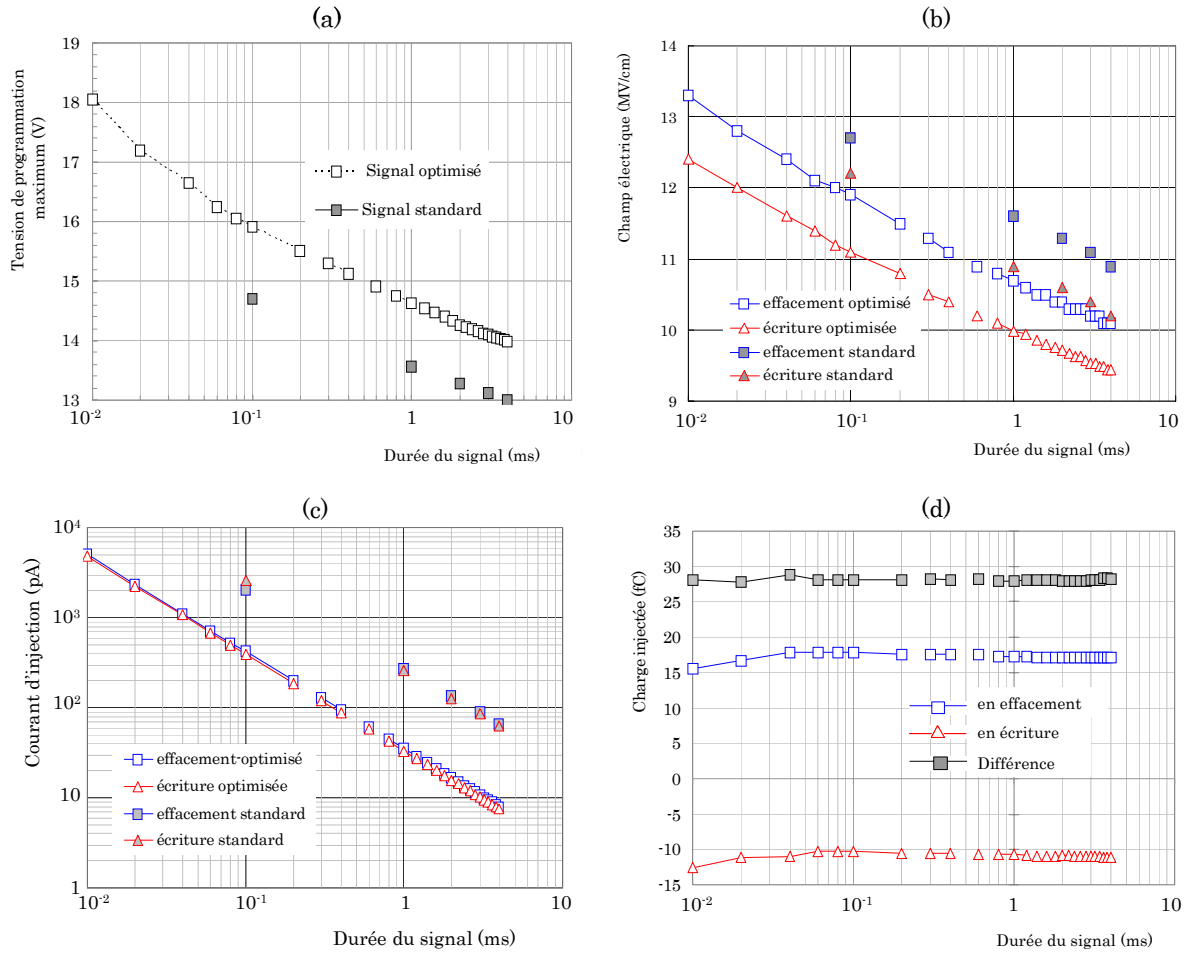


Figure III. 17 : Simulations de la tension maximale appliquée (a), du champ électrique (b), du courant d'injection (c) et de la charge injectée (d) en fonction de la durée du signal de programmation.

Ainsi, dans un premier temps, nous avons voulu avoir une idée précise des variations de la tension de programmation nécessaire en fonction de la durée du signal. A cette fin, nous avons déterminé, par simulation, la tension de programmation V_{prog} en fonction de la durée du signal pour maintenir une charge Q_{fg} constante.

Les simulations sont basées sur le critère suivant : Q_{fg} doit rester constante et égale à la charge Q_{fg} produite par un signal standard de 4ms lorsque Δt diminue. La figure III.17 représente les résultats obtenus pour la tension de programmation (17a), le champ électrique (17b), le courant à travers l'oxyde tunnel (17c) ainsi que la charge Q_{fg} stockée dans la grille flottante (17d) lorsque $10\mu\text{s} < \Delta t < 4\text{ms}$.

La relation (III.30) peut aussi s'écrire $\ln[I_{\text{FN}}] = \ln[Q_{\text{fg}}] - \ln[\Delta t]$ et se traduit par une droite sur une échelle log-log (fig. III.17c). Par ailleurs, comme Q_{fg} est constant, ΔV_{th} est aussi constant (fig. III.17d).

III.3.2 Mesure de la fenêtre de programmation

A partir des tensions de programmation maxima nécessaires déterminées par simulation, nous mesurons la fenêtre de programmation pour une durée de signal variant de 4ms à 100 μ s. Les mesures (fig. III.18a et 18b) montrent que la fenêtre de programmation ΔV_{th} obtenue à partir des tensions de programmation déterminées par simulation est quasiment constante, malgré une légère baisse pour des durées de signal proches de 100 μ s. Nous pensons que cette baisse est due à la tension de sélection que nous avons volontairement limitée à 20V pour ne pas affecter la fiabilité des cellules, auquel cas la condition $V_{sel} > V_{progmax} + 3V$ n'est plus respectée pour $V_{progmax} > 17V$.

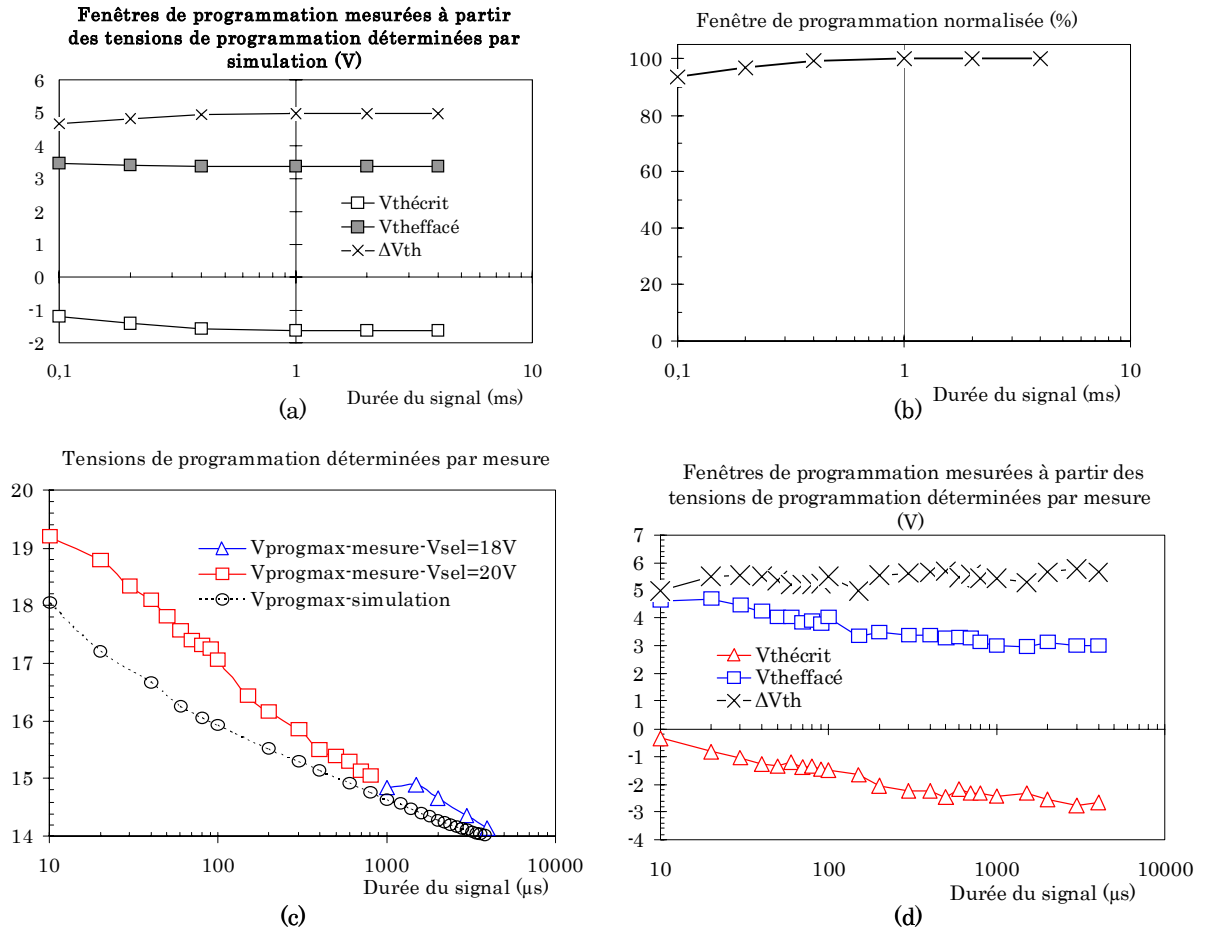


Figure III. 18 : *Mesure des tensions de seuil et fenêtre de programmation normalisée pour $100\mu s < \Delta t < 4ms$.*

Afin d'éliminer cette baisse de la fenêtre de programmation, nous réajustons les tensions de programmation maxima nécessaires par mesure de façon à garder la même fenêtre de programmation que celle obtenue avec un signal standard de durée égale à 4ms (fig. III.18c). Pour $V_{progmax} < 15V$, nous polarisons la grille de sélection à 18V et pour $V_{progmax} > 15V$, nous polarisons la grille de sélection à 20V. La fenêtre de programmation mesurée avec ces nouvelles tensions de programmation (fig. III.18d) est quasiment constante pour une durée de signal variant de 4ms à 10 μ s.

III.3.3 Résultats en endurance

III.3.3.1 Fermeture de la fenêtre de programmation

Afin d'évaluer la conséquence des champs très élevés provoqués par les signaux très brefs sur la fiabilité de la cellule, nous avons effectué des tests en endurance pour quantifier la fermeture de la fenêtre de programmation pour $10\mu s < \Delta t < 4ms$. La fermeture de la fenêtre de programmation normalisée est représentée sur la figure III.19.

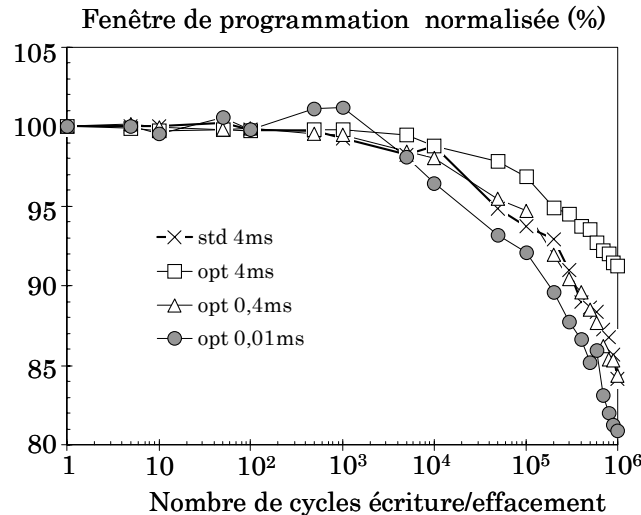


Figure III. 19 : *Fermeture normalisée de la fenêtre de programmation pour $10\mu s < \Delta t < 4ms$.*

La figure III.19 montre deux choses :

i.

D'une part, la fermeture due au signal standard de 4ms est plus sévère que celle qui est due au signal optimisé de même durée. Ceci résulte de ce que le champ électrique à travers l'oxyde tunnel, que ce soit en écriture ou en effacement, est plus faible pour le signal optimisé, dont la rampe est plus faible.

ii.

D'autre part, lorsque la durée du signal optimisé diminue, la fenêtre de programmation se ferme de plus en plus vite. Ceci est normal car le champ électrique augmente lorsque la durée du signal diminue afin de garder la même charge injectée (figure III.17d).

III.3.3.2 Courbe caractéristique

Afin de comparer l'ensemble des résultats, nous mesurons le nombre de cycles donnant une dégradation équivalente, obtenue pour un pourcentage donné. Nous choisissons une valeur commune à l'ensemble des courbes correspondant à une dégradation la plus grande possible : 7% de fermeture, valeur pour laquelle la cellule est encore largement utilisable.

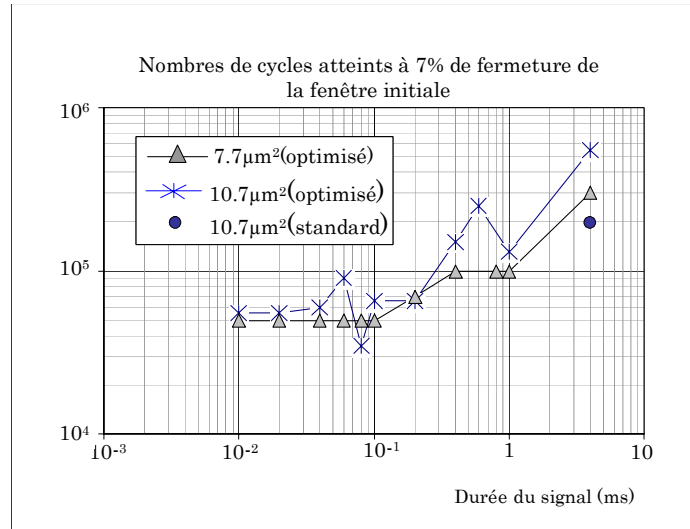


Figure III. 20 : *Courbes caractéristiques des technologies 7.7µm² et 10.7µm² montrant le nombre de cycles atteint à 7% de fermeture.*

Comme on pouvait s’y attendre, la durée du fonctionnement diminue avec la durée de la programmation puisque le champ électrique augmente. Cependant, pour $\Delta t < 200\mu s$, on note un palier dans cette dégradation. On retrouve ces résultats sur les technologies 7.7µm² et 10.7µm². Il faut noter que la durée d’application de la contrainte électrique diminue avec le temps de programmation. Les effets de l’augmentation du champ et de la diminution de la durée d’application sembleraient se compenser. Cet aspect sera approfondi dans le chapitre IV.

III.4 Programmers non-standards

III.4.1 Programmation répartie : impact sur la fiabilité

Pour réduire la valeur des tensions maxima nécessaires à la sortie des pompes de charge, on peut utiliser des tensions négatives et répartir ces tensions sur la grille de contrôle et le drain.

Toutefois, on ne peut pas simplement diviser les tensions par deux car les coefficients de couplage sont différents en effacement et en écriture.

III.4.1.1 Programmers classique et répartie

Lors de l’effacement classique d’une cellule EEPROM, une impulsion positive est appliquée sur la grille de contrôle. Pendant cette opération, le drain, la source et le substrat sont à la masse (figure III.21a). Dans ce cas, le coefficient de couplage entre le potentiel de grille de flottante et le potentiel de grille de contrôle est égal à K_e (cf. relation (II.42) §II.2.1.2).

La tension aux bornes de l'oxyde tunnel est donnée par :

$$V_{\text{tun}} = K_e V_{\text{gc}} \quad (\text{III. 31})$$

où V_{gc} est la polarisation de la grille de contrôle.

En écriture classique, la grille de contrôle et le substrat sont à la masse, la source est flottante et une tension positive est appliquée sur le drain (figure III.21b). Le coefficient de couplage, dans ce cas, est donné par :

$$K_w = \frac{C_{\text{tun}}}{C_x} \quad (\text{III. 32})$$

Cette fois-ci, la tension aux bornes de l'oxyde tunnel est donnée par :

$$V_{\text{tun}} = (1 - K_w) V_d \quad (\text{III. 33})$$

où V_d est la polarisation du drain.

Les équations (III.32) et (II.42) (cf. §II.2.1.2) montrent que :

$$K_e \neq (1 - K_w) \quad (\text{III. 34})$$

Par conséquent, pour obtenir la même charge injectée en effacement et en écriture, il faut polariser la grille de contrôle avec une tension différente de celle du drain en mode classique, et ce afin d'obtenir la même largeur de fenêtre de programmation. On peut avoir $V_{\text{gc}} > V_d$ ou l'inverse, selon la valeur relative de K_e et de $(1 - K_w)$.

En polarisation répartie, nous appliquons une impulsion négative sur le drain et une impulsion positive sur la grille (figure III.21c) pour effacer la cellule mémoire, et ce afin de diminuer la polarisation appliquée sur chacune de ces deux électrodes. Dans ce cas, si le substrat est à la masse, la jonction PN drain/substrat est polarisée en direct. Ceci aurait pour conséquence fâcheuse de générer un courant de fuite. Par conséquent, il faut polariser le substrat au même potentiel que le drain, auquel cas la jonction PN drain/substrat est toujours bloquée (figure III.21c). Sur la figure III.21c, nous avons le schéma de polarisation de la cellule en effacement réparti. En écriture, les mêmes connexions ont été utilisées, avec des polarisations opposées.

Dorénavant, comme le substrat est au même potentiel que le drain, nous devons calculer les nouveaux coefficients de couplage K_e^* et K_w^* . En effacement réparti, le substrat est à la masse. En effet, la grille est polarisée à un potentiel positif et le drain à un potentiel négatif. Ainsi, la jonction PN drain-substrat est bloquée. Le coefficient de couplage K_e^* demeure identique à celui de la relation (II.42) (cf. §II.2.1.2).

En revanche, en écriture répartie, où l'impulsion de polarisation est appliquée sur le drain et le substrat forcé au potentiel de drain, K_w^* vaut :

$$K_w^* = \frac{C_{\text{ox}} + C_{\text{tun}}}{C_x} \quad (\text{III. 35})$$

Ainsi, contrairement à la relation (III.34), nous avons maintenant :

$$K_e^* = (1 - K_w^*) \quad (\text{III. 36})$$

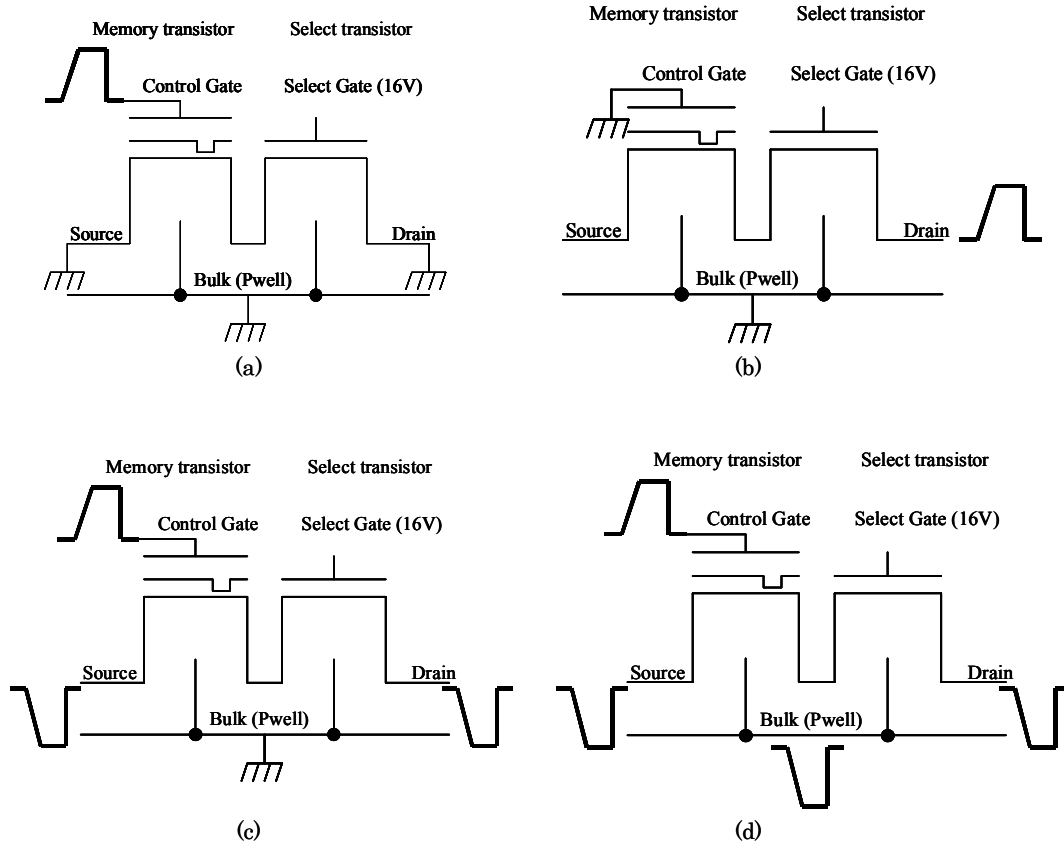


Figure III. 21 : (a) Effacement classique où drain, grille et substrat sont à la masse; (b) écriture classique où grille et substrat sont à la masse, la source flottante; (c) effacement réparti où le substrat est à la masse et (d) écriture répartie où le substrat est forcé au potentiel de drain.

Lorsque nous répartissons la tension de programmation entre la grille et le drain, nous utilisons à la fois K_e^* et K_w^* . Auquel cas, la chute de potentiel à travers l'oxyde tunnel vaut :

$$V_{\text{tun}} = K_e^* - (1 - K_w^*)V_d \quad (\text{III. 37})$$

ou, plus simplement, moyennant la relation (III.36) :

$$V_{\text{tun}} = K_e^*(V_g - V_d) \quad (\text{III. 38})$$

Avec la relation (III.38), nous pouvons déterminer la valeur des impulsions à appliquer sur la grille et sur le drain en mode réparti afin d'obtenir la même chute de potentiel à travers l'oxyde tunnel qu'en mode standard.

III.4.1.2 Simulations

Les signaux utilisés pour la simulation sont représentés sur les figures III.22 et III.23. La figure III.22 représente les signaux standards utilisés en effacement et en écriture en mode classique et réparti. La figure III.23 représente les

signaux optimisés utilisés en effacement et en écriture en mode classique et réparti.

Dans le tableau III.1, nous présentons les résultats des simulations en écriture et en effacement pour un signal standard en mode classique et en mode réparti. Le signal standard utilisé a une tension maximale de 13V, appliquée sur la grille ou le drain. Le signal standard réparti vaut +6.5V (-6.5V) sur la grille et -6.5V (+6.5V) sur le drain en effacement (écriture). Le signal optimisé utilisé a une tension maximale de 14V, appliquée sur la grille ou le drain. Le signal optimisé réparti vaut +7V (-7V) sur la grille et -7V (+7V) sur le drain en effacement (écriture). Tous les signaux durent 4ms.

Les résultats de simulation relatifs aux signaux optimisés sont regroupés dans le tableau 2.

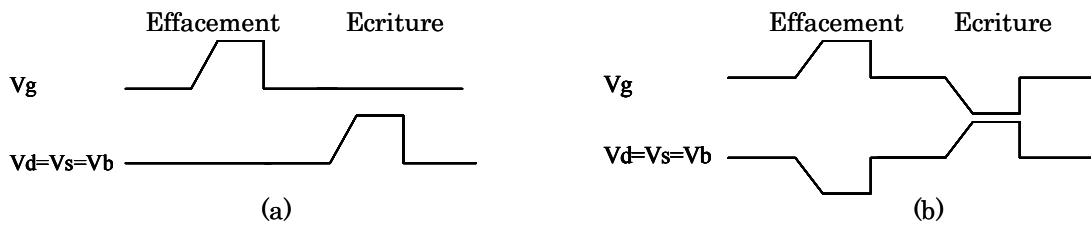


Figure III. 22 : *Signaux standards utilisés en effacement et en écriture en mode classique (a) et réparti (b).*

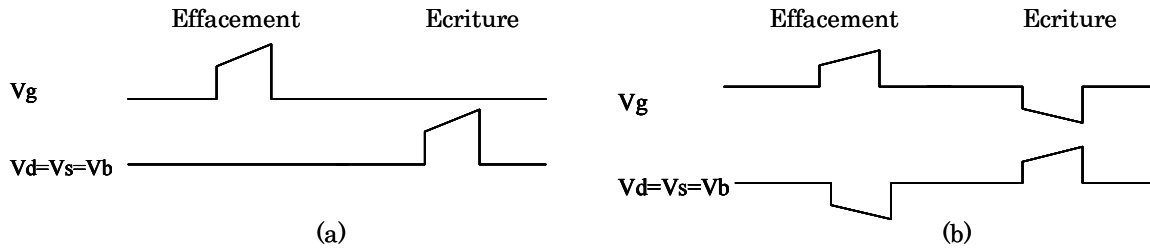


Figure III. 23 : *Signaux optimisés utilisés en effacement et en écriture en mode classique (a) et réparti (b).*

Mode	Opération	$V_t(V)$	QFG(fC)	$E_{tun_max}(MV/cm)$
Polarisation classique	Effacement	+3.50	-9.10	+11.10
	Ecriture	-1.41	+11.30	-10.30
Polarisation répartie	Effacement	+3.50	-9.10	+11.10
	Ecriture	-1.41	+11.30	-10.30

Tableau III. 1 : *Simulation des tensions de seuil, champ électrique maximum et charge injectée, en effacement et en écriture, en mode classique et réparti pour des signaux standards.*

Mode	Opération	$V_t(V)$	QFG(fC)	$E_{tun_max}(MV/cm)$
Polarisation classique	Effacement	+3.48	-9.10	+10.10
	Ecriture	-1.44	+11.40	-9.53
Polarisation répartie	Effacement	+3.48	-9.10	+10.10
	Ecriture	-1.44	+11.40	-9.53

Tableau III. 2 : *Simulation des tensions de seuil, champ électrique maximum et charge injectée, en effacement et en écriture, en mode classique et réparti pour des signaux optimisés.*

Nous pouvons constater que la charge injectée dans la grille flottante est la même en mode classique et en mode réparti. Ainsi, la tension de seuil est la même pour les deux modes. Le champ électrique maximum qui conditionne la fiabilité de la cellule est aussi le même dans les deux modes. Par conséquent, la dégradation de l'oxyde provoquée par ce champ électrique devrait être la même dans les deux modes. Ces simulations montrent que, en principe, la polarisation répartie ne devrait pas modifier la fiabilité de la cellule, qu'elle soit programmée avec un signal standard ou un signal optimisé. Remarquons que la valeur de la tension de sélection est déterminante en écriture. En mode classique, nous appliquons +18V sur la grille de sélection. En mode réparti, une valeur aussi élevée n'est pas nécessaire : en simulation, nous n'avons observé aucune variation de la charge injectée jusqu'à une tension de sélection de $V_{progmax}+3V$ comme en polarisation classique. Ainsi, nous pouvons diminuer toutes les polarisations appliquées sur la cellule mémoire.

III.4.1.3 Mesures

Afin de valider les simulations, nous avons mesuré la tension de seuil en écriture et en effacement pour les configurations représentées sur la figure III.24. Le tableau III.3 montre les valeurs de la tension de seuil mesurées en écriture et en effacement en modes classique et réparti.

Mode	Signal	Opération	$V_t(V)$ moyennés sur 5 cellules	$\Delta V_{th}(V)$
Polarisation classique	Standard	Effacement	+3.30	5.00
	Standard	Ecriture	-1.70	
Polarisation répartie	Standard	Effacement	+3.33	4.71
	Standard	Ecriture	-1.38	
Polarisation classique	Optimisé	Effacement	+3.27	4.65
	Optimisé	Ecriture	-1.38	
Polarisation répartie	Optimisé	Effacement	+3.44	4.72
	Optimisé	Ecriture	-1.28	

Tableau III. 3 : *Tensions de seuil mesurées, en effacement et en écriture, en mode classique et réparti, pour des signaux standards et optimisés.*

La tension de seuil d'une cellule EEPROM dépend de la charge injectée dans la grille flottante. Les simulations montrent que la charge injectée est la même en mode classique et en mode réparti. Dans le cas d'un signal standard, la fenêtre de programmation vaut 5.00V en mode classique et 4.71V en mode réparti. La fenêtre de programmation simulée vaut 4.90V. Ainsi, à $\pm 4\%$, simulations et mesures donnent les mêmes résultats. Dans le cas des signaux optimisés, on a aussi un bon accord entre simulations et mesures ($\pm 6\%$).

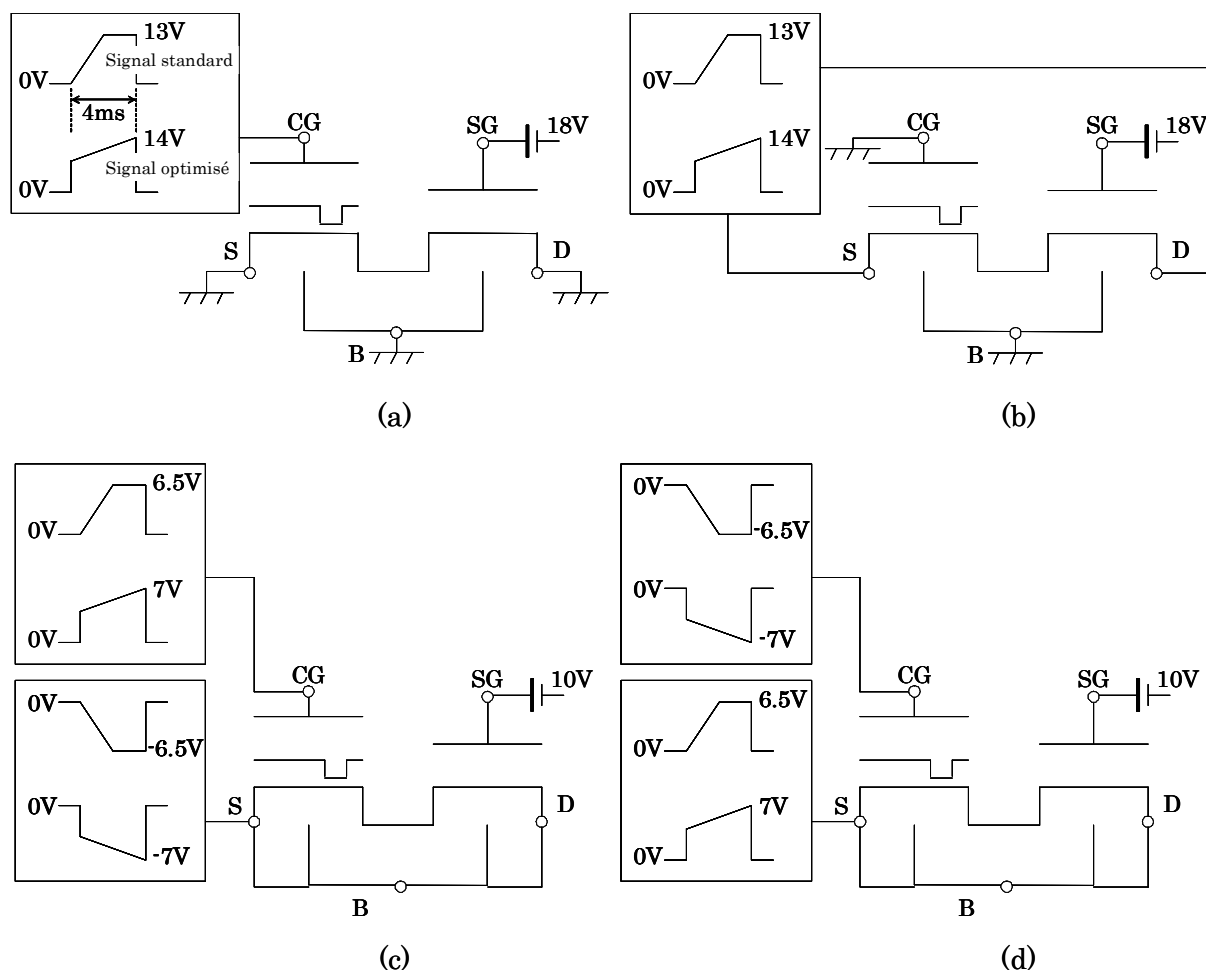


Figure III. 24 : Configurations de mesure de la tension de seuil : (a) effacement en mode classique, (b) écriture en mode classique, (c) effacement en mode réparti et (d) écriture en mode réparti.

III.4.1.4 Test en endurance et interprétation

La figure III.25 montre la fermeture de la fenêtre de programmation pour quatre types de signaux : un signal standard classique, un signal standard réparti, un signal optimisé classique et un signal optimisé réparti.

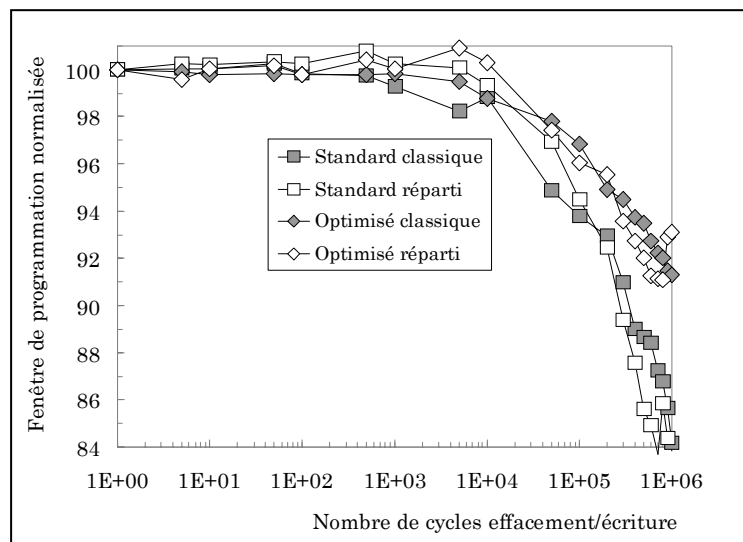


Figure III. 25 : Fermeture de la fenêtre de programmation en fonction du nombre de cycles écriture/effacement en modes classique et réparti pour les signaux standards et optimisés.

On constate que, d'une part, les courbes relatives au

signal standard donnent la même fermeture de la fenêtre de programmation. Programmer en mode classique ou réparti n'a pas d'incidence sur l'endurance d'une cellule programmée avec un signal standard. De la même façon, l'endurance demeure inchangée pour une cellule programmée avec un signal optimisé classique ou réparti.

Ce test en endurance montre que la polarisation répartie ne modifie pas l'endurance d'une cellule : une cellule programmée avec un signal classique (standard ou optimisé) aura la même endurance qu'une cellule programmée avec un signal réparti (standard ou optimisé).

En conclusion, la programmation répartie ne modifie pas l'endurance de la cellule, mais elle permet d'utiliser des niveaux de tensions bien plus faibles (quasiment divisés par deux). La circuiterie des pompes de charges s'en trouvera simplifiée, mais il faudra générer des tensions négatives.

III.4.2 Programmation par trains d'impulsions

Lorsque l'on baisse la tension de programmation, I_{FN} baisse. Or, la charge injectée dans la grille flottante vaut :

$$Q_{fg} = \int_0^{\Delta t} I_{FN} dt \quad (\text{III. 39})$$

Cette relation montre que si l'on diminue la tension de programmation, il faut augmenter Δt afin de garder Q_{fg} constant.

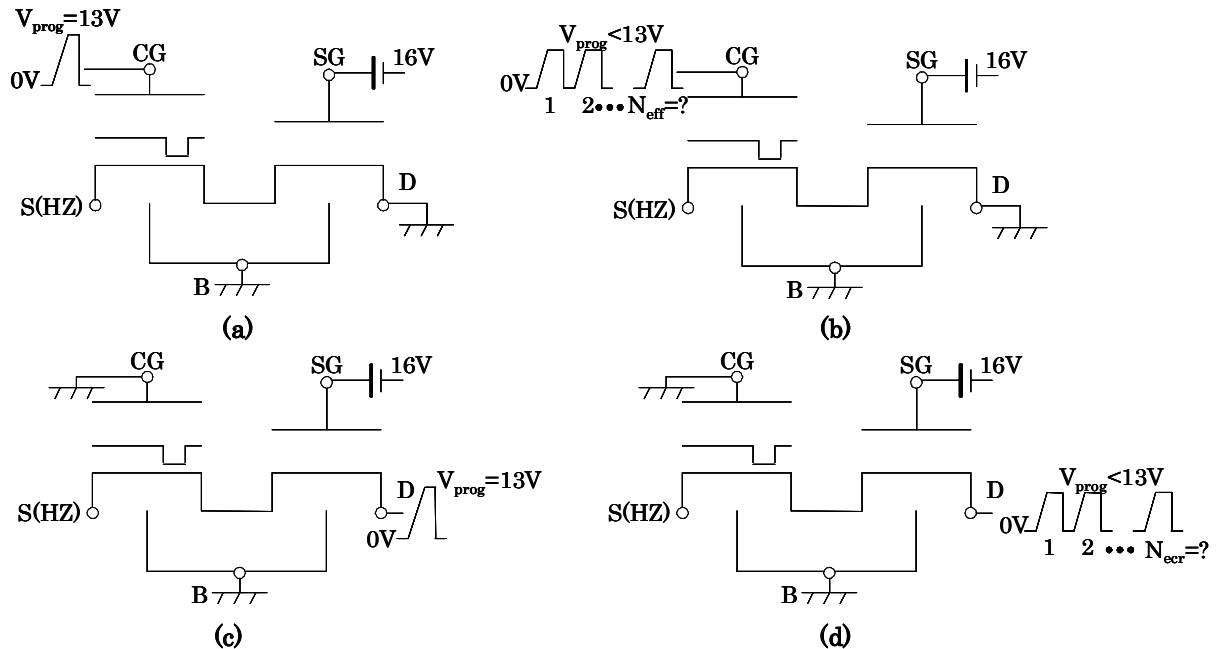


Figure III. 26 : Effacement d'une cellule EEPROM avec une impulsion de tension (a), plusieurs impulsions de tension inférieure (b) et écriture avec une impulsion de tension (c) et plusieurs impulsions de tension inférieure (d).

Traditionnellement, on écrit et on efface une cellule EEPROM avec une impulsion de tension donnée (fig. III.26a et 26c). Si on baisse la tension de programmation, il faut augmenter le nombre d'impulsions afin de garder la même quantité de charge injectée Q_{fg} en écriture et en effacement (relation III.39). Combien d'impulsions sont nécessaires selon la tension de l'impulsion utilisée ?

Pour répondre à cette question, nous avons d'abord déterminé par simulation le nombre d'impulsions nécessaires (fig. III.27b et 27c) pour obtenir la même charge injectée, en effacement et en écriture, que celle obtenue avec un signal de référence de 13V (fig. III.27a). On constate que le nombre d'impulsions nécessaires augmente rapidement : 100 impulsions de 300 μ s (30ms) pour $V_{prog}=12V$ et plus de 10^5 impulsions de 300 μ s (30s!) pour $V_{prog}=10V$.

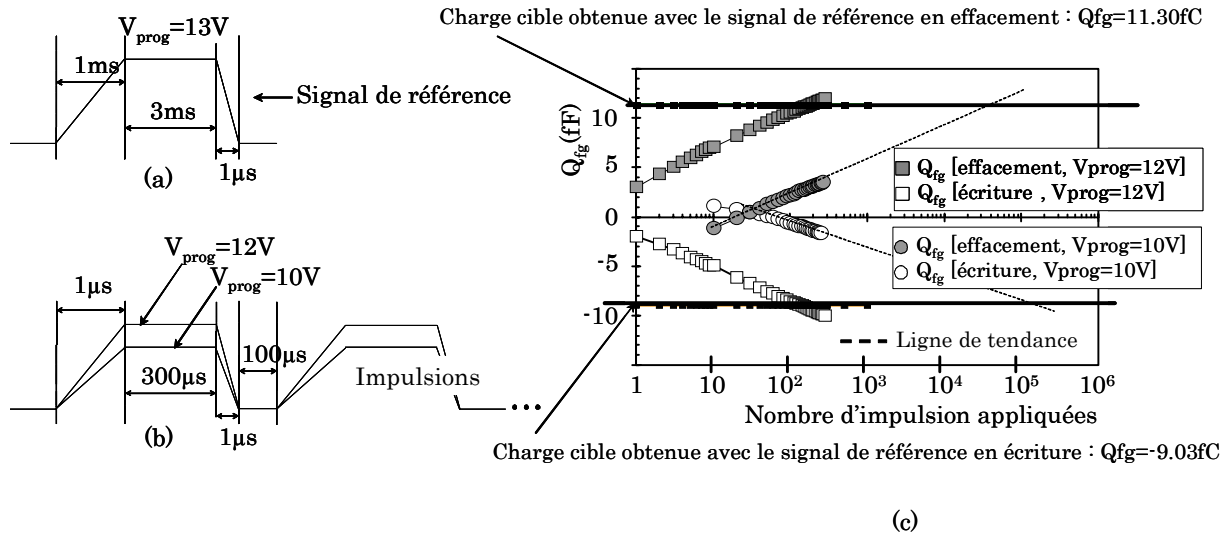


Figure III. 27 : *Signal de référence (a), impulsions utilisées (b) et simulations de la charge injectée en effacement et en écriture (c).*

Nous avons vérifié expérimentalement ces résultats en effectuant des effacements successifs et des écritures successives sur une cellule EEPROM. Le signal de référence est un signal standard de 13V représenté sur la figure III.28a. Un tel signal donne une tension de seuil en écriture de -1.7V et une tension de seuil en effacement de 2.6V pour la technologie utilisée. Les autres signaux de programmation utilisés ainsi que l'évolution des tensions de seuil en fonction du nombre d'impulsions appliquées sont représentées sur la figure III.28b.

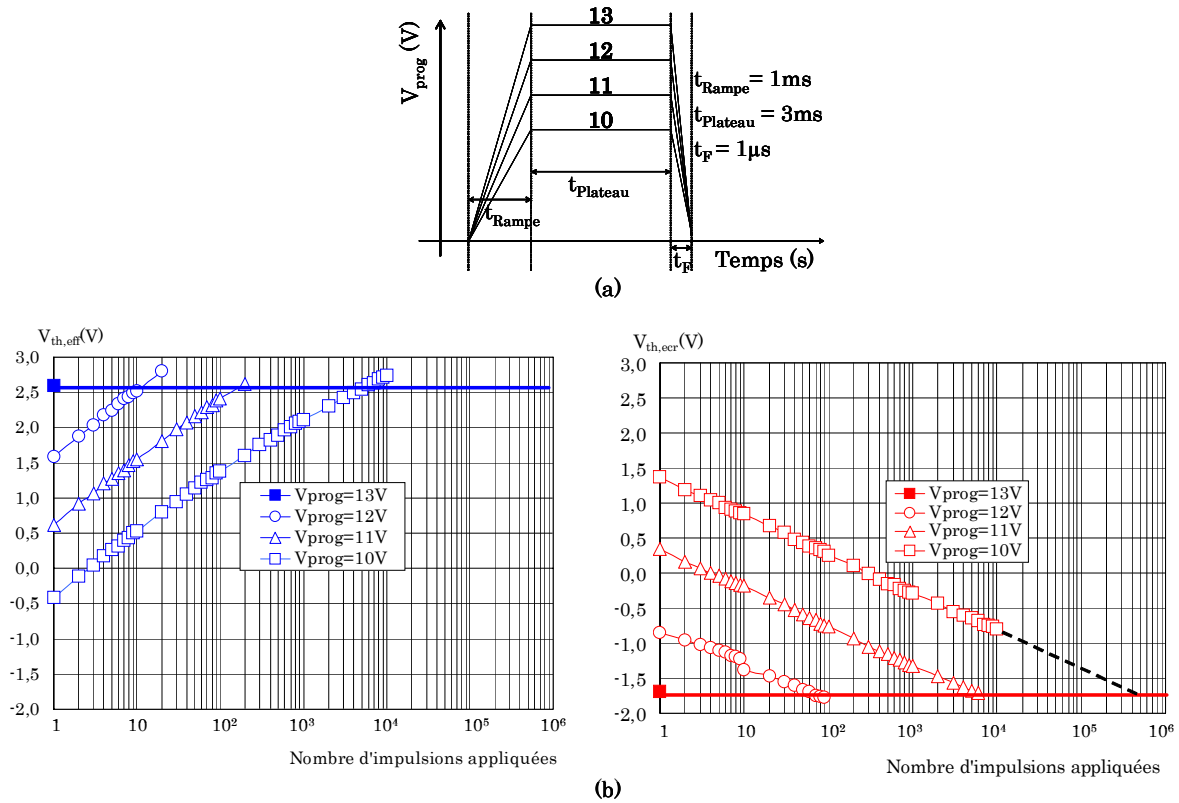


Figure III. 28 : *Signaux de programmation utilisés (a) et tensions de seuil en écriture et en effacement en fonction du nombre d'impulsions appliquées (b).*

V_{prog}	13V	12V	11V	10V
$V_{theffacé}$ (V)	2,60			
$V_{thécrit}$ (V)	-1,70			
Fenêtre de programmation cible (V)	4,30			
Durée d'une impulsion : Δt (s)	0,004	0,004	0,004	0,004
Nombre d'impulsions nécessaires en effacement : N_{eff}	1	10	200	5000
$N_{eff}\Delta t$ (s)	0,004	0,04	0,8	20
Nombre d'impulsions nécessaires en écriture : N_{ecr}	1	90	6000	600000
$N_{ecr}\Delta t$ (s)	0,004	0,36	24	2400

Tableau III. 4 : *Nombre d'impulsions nécessaires en écriture et en effacement (fig. III.28b).*

Les courbes de la figure III.28b ont été obtenues de la manière suivante : partant d'un état écrit (respectivement effacé), nous mesurons la tension de seuil $V_{theffacé}$ (respectivement $V_{thécrit}$) en fonction du nombre d'impulsions appliquées jusqu'à ce qu'on atteigne la tension de seuil désirée.

D'une manière générale, la figure III.28 montre que le nombre d'impulsions nécessaires augmente rapidement lorsque V_{prog} diminue. En effacement, on obtient le même ordre de grandeur de la quantité « $N_{eff} \times$ durée d'une impulsion » en mesure et en simulation.

En écriture, on note une grande dissymétrie : la quantité « $N_{ecr} \times$ durée d'une impulsion » donnée par la mesure est au moins dix fois plus grande que celle qui est prévue par les simulations. On remarque notamment que, pour $V_{prog}=10V$, il

faut plus de 6×10^5 d'impulsions de 4 ms, autrement dit 2400s, pour atteindre la tension de seuil en écriture de -1.7V. Un gain en tension de 3V ne saurait justifier une durée aussi longue.

Une question demeure : pour de faibles tensions de programmation, l'utilisation de N impulsions de durée Δt est-elle équivalente à l'utilisation d'une impulsion de durée égale à $N\Delta t$ (fig. III.29)?

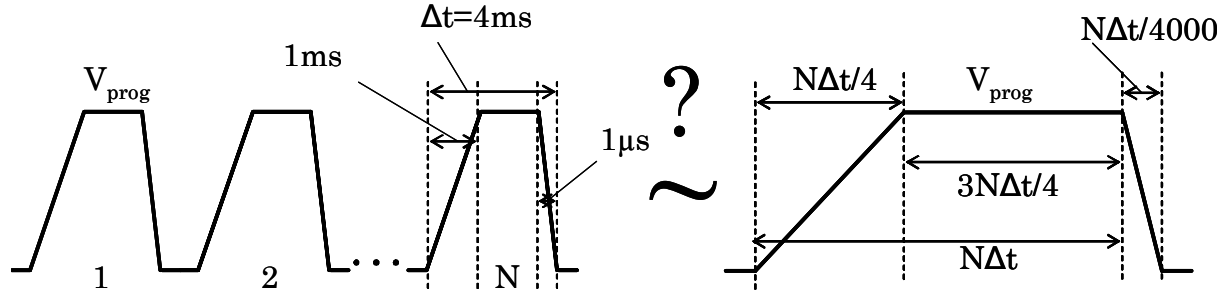


Figure III. 29 : *Problème de l'équivalence entre l'utilisation de N impulsions de durée Δt et une impulsion de durée $N\Delta t$*

Pour répondre à cette question, nous avons déterminé par mesure la durée nécessaire d'une impulsion de tension $V_{\text{prog}} = 12\text{V}$, 11V et 10V (fig. III.30b) qui donne la même tension de seuil en effacement et en écriture que le signal de référence (fig. III.30a).

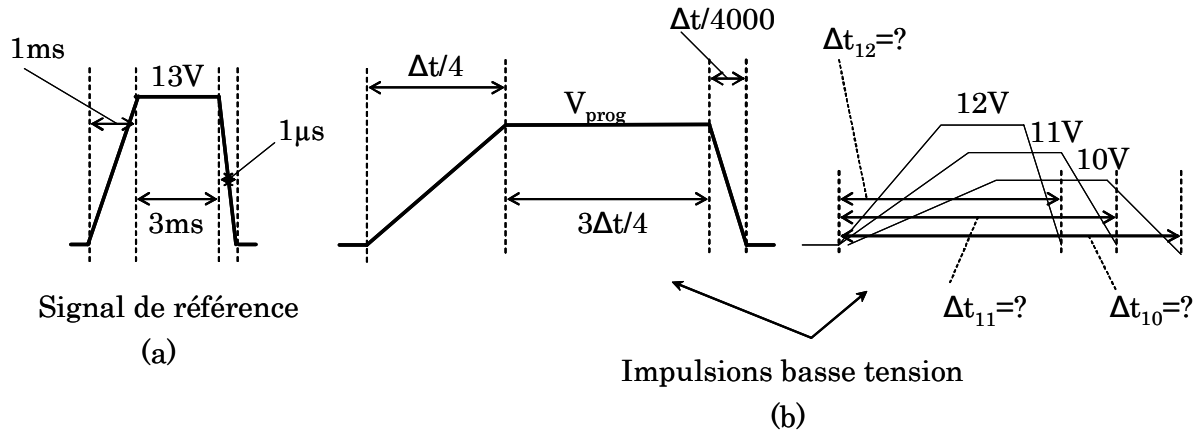


Figure III. 30 : *Signal de référence (a) et impulsions basse tension (b).*

Nous avons ainsi déterminé les durées d'impulsions nécessaires Δt_{eff} et Δt_{ecr} pour $V_{\text{prog}} = 12\text{V}$, 11V et 10V :

V_{prog}	13V	12V	11V	10V
$V_{theffacé\ cible}(V)$	2,47			
$V_{thécrit\ cible}(V)$	-1,98			
Fenêtre de programmation cible (V)	4,45			
Durée d'impulsion nécessaire en effacement : Δt_{eff} (s)	0,004	0,04	0,65	26
$N_{eff}\Delta t$ (s) (Tableau III.4)	0,004	0,04	0,8	20
$\Delta t_{eff} / N_{eff}\Delta t$	1	1	0,81	1,3
Durée d'impulsion nécessaire en écriture: Δt_{ecr} (s)	0,004	0,17	18	>2400
$N_{ecr}\Delta t$ (s) (Tableau III.4)	0,004	0,36	24	2400
$\Delta t_{ecr} / N_{ecr}\Delta t$	1	0,47	0,75	>1

Tableau III. 5 : *Durées nécessaires des impulsions de 12V, 11V et 10V pour obtenir les mêmes tensions de seuil en écriture et en effacement que celles obtenues avec le signal de référence.*

Le rapport $\Delta t_{eff}/N_{eff}\Delta t$ et $\Delta t_{ecr}/N_{ecr}\Delta t$ est relativement proche de 1 (sauf $\Delta t_{ecr}/N_{ecr}\Delta t=0,47$ pour $V_{prog}=11V$) pour les différentes tensions de programmation, ce qui nous permet d'affirmer qu'il n'y a pas de différence notable entre l'utilisation de N impulsions de durée Δt ou une impulsion de durée $N\Delta t$ pour une durée de signal de l'ordre de la milliseconde.

Dans ce chapitre, nous avons exploré diverses manières d'améliorer les performances d'une cellule EEPROM en termes d'augmentation de la vitesse d'écriture et de baisse des tensions de programmation.

D'une part, nous avons montré que la durée de programmation d'une cellule EEPROM peut être réduite jusqu'à quelques microsecondes. Ceci permet d'envisager l'utilisation de ces mémoires comme concurrentes directes des mémoires flash pour des applications ne nécessitant pas une grande densité. C'est d'autant plus réaliste que les mémoires EEPROM consomment moins de courant que les mémoires Flash.

D'autre part, il est possible de diminuer les tensions de programmation d'une cellule EEPROM en répartissant ces tensions entre les différentes électrodes. Ceci exige néanmoins de prendre plusieurs précautions afin d'empêcher la jonction PN drain/substrat de la cellule de se trouver en régime direct. Quant à la programmation par impulsions, elle n'apporte pas d'améliorations en raison des durées irréalistes nécessaires.

Beaucoup de questions demeurent ouvertes : nous tentons d'expliquer la stagnation des courbes d'endurance pour des durées de programmation inférieures à la centaine de microsecondes.

Dans cette perspective, nous nous sommes proposé de modéliser la fermeture de la fenêtre de programmation d'une cellule EEPROM en fonction du nombre de cycles écriture/effacement. Cette étude a été conduite sur des capacités de grande surface équivalentes à l'oxyde tunnel d'une cellule EEPROM. Cette voie, objet du dernier chapitre, nous permettra, peut-être, à moyen terme, d'expliquer la stagnation des courbes d'endurance observées lors de la réduction du temps de programmation des cellules EEPROM et de prédire son évolution à l'aide du modèle dès la conception.

Références bibliographiques du chapitre III

[Lenzlinger'69] M. Lenzlinger and E.H. Snow, "Fowler-Nordheim Tunneling into thermally grown SiO₂", Journal of Applied Physics, vol. 40, pp. 278-282, 1969.

[Kolodny'86] A. Kolodny, S.T.K. Nieh, B. Eitan, IEEE Transaction on Electron Devices, vol. ED-33, No. 6, June (1986) P 835.

[Canet'00] P. Canet, R. Bouchakour, N. Harabech, P. Boivin, J.-M. Mirabel, "Study of Signal Programming to Improve EEPROM Cell reliability", Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems, Lansing, Michigan, USA, vol. 3, pp. 1144-1147, 2000.

[Harabech'00] N. Harabech, R. Bouchakour, P. Canet, P. Pannier, J.-P. Sorbier, "Extraction of Fowler-Nordheim Parameters of Thin SiO₂ Oxide Film including Polysilicon gate depletion : Validation with an EEPROM Memory Cell", Proceedings of the IEEE ISCAS conference, Geneva, Switzerland, p. II. 441-444, May 2000.

[Bouchakour'01] R. Bouchakour, N. Harabech, P. Canet, J.-M. Mirabel, P. Boivin, O. Pizzuto, "A New Physical Based Compact Model of Floating Gate EEPROM Cells", Journal of Non-Crystalline Solids, Vol. 280, Issues 1-3, pp. 122-126, February 2001.

[Canet'01a] P. Canet, R. Bouchakour, N. Harabech, "EEPROM Programming Study: Time and Degradation Aspects", Proceedings of the IEEE International Symposium on Circuits And Systems (ISCAS), Sydney, Australia, p. IV.846-849, 2001.

[Canet'01b] P. Canet, R. Bouchakour, N. Harabech, P. Boivin, J.-M. Mirabel, C. Plossu, "Improvement of EEPROM cell reliability by optimization of signal programming", Journal of Non-Crystalline Solids, Vol. 280, Issues 1-3, pp. 116-121, February 2001.

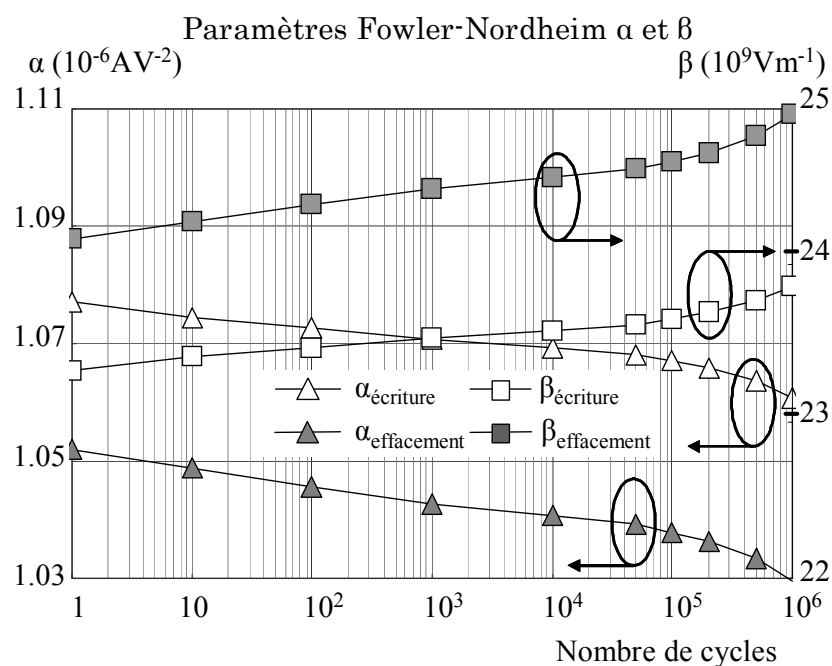
[Canet'02] P. Canet, R. Bouchakour, J. B. Razafindramora, F. Lalande, J.-M. Mirabel, "Very fast EEPROM Erasing Study", Proceedings of the 28th European Solid State Circuits Conference (ESSCIRC'2002), Florence, Italy, pp. 683-686, September 2002.

[Laffont'02] R. Laffont, J. B. Razafindramora, P. Canet, R. Bouchakour, J.-M. Mirabel, "Decreasing EEPROM programming bias with negative voltage, reliability impact", Proceedings of the IEEE International Workshop on Memory Technology, Design and Testing, Isle of Bendor, France, pp. 168-173, July 2002.

[Canet'03a] P. Canet, R. Bouchakour, F. Lalande, J.-M. Mirabel « EEPROM Cell design: paradoxical choice of coupling ratio », Journal of Non-Crystalline Solids, Vol. 322, Issues 1-3, pp. 246-249, July 2003.

[Canet'03b] P. Canet, F. Lalande, M. Martin, R. Bouchakour, "Non-Volatile Memory Cell Design: Sizing assisted by a predictive Model", Proceedings of the 4th Annual Non-Volatile Memory Technology Symposium (NVMTS'2003), San Diego, California, USA, pp. 13-1-13-4, November 15-17, 2003.

Chapitre IV : Modélisation du vieillissement de l'oxyde tunnel



Evolution des paramètres FN α et β en fonction du nombre de cycles lors de l'émulation sur une capacité équivalente du vieillissement de l'oxyde tunnel d'une mémoire EEPROM.

IV.1 ETUDE BIBLIOGRAPHIQUE.....	131
IV.1.1 Problématique	131
IV.1.1.1 De la difficulté de prévoir l'endurance d'une MNV.....	131
IV.1.1.2 Définitions	131
IV.1.2 Fiabilité des diélectriques : les différents types de contraintes électriques.....	131
IV.1.2.1 Stress à tension constante	132
IV.1.2.2 Stress à courant constant.....	132
IV.1.2.3 Rampe de tension (RVS ou Ramped Voltage Stress)	134
IV.1.2.4 Rampe de courant exponentielle.....	135
IV.1.3 Etude du piégeage de charges dans les oxydes	136
IV.1.3.1 Technique de Di Maria.....	136
IV.1.3.2 Méthode de R. Kies et al.	137
IV.1.3.3 Loi en puissance de Papadas et al.	139
IV.1.3.4 Limite de la loi en puissance par Scarpa et al.	139
IV.2 METHODOLOGIE DE L'ETUDE	140
IV.2.1 Introduction.....	140
IV.2.2 Détermination des tensions appliquées.....	141
IV.2.3 Extraction des paramètres	142
IV.2.4 Simulation de la fenêtre de programmation	144
IV.2.5 Comparaison de la dégradation en effacement et en écriture.....	145
IV.3 ETUDE DE LA MESURE $I_{FN}(V_{GB})$	146
IV.3.1 Mise en évidence de la dégradation due à la mesure $I_{FN}(V_{gb})$	146
IV.3.2 Méthode de mesure « 2 points en transitoire ».....	147
IV.3.3 Nouvelle estimation de l'évolution de α et β	150
Références bibliographiques du chapitre IV	152

De nos jours, la fiabilité est un problème majeur dans les technologies mémoires non volatiles. En effet, la course à la miniaturisation, nourrie par la compétition économique, conduit à la réduction des composants de base des circuits intégrés que sont les transistors MOS. L'épaisseur de l'oxyde de grille atteint désormais quelques nanomètres dans les technologies CMOS et descendra en deçà de 2 nm pour les générations sub-0.1 μ m. Nombre d'effets indésirables, tels le courant tunnel direct, ont lieu dans des oxydes aussi fins. Ces phénomènes, à moins d'être convenablement traités, peuvent provoquer une dégradation rapide de l'oxyde et conduire finalement à son claquage. Dans ce contexte, une parfaite compréhension des mécanismes de conduction à travers les oxydes soumis à une contrainte électrique est l'une des pierres angulaires de l'amélioration et de la prédiction de la fiabilité des dispositifs dont le fonctionnement repose sur l'injection de charges à travers un oxyde mince.

Nous présentons d'abord une étude bibliographique relative à la fiabilité des oxydes. Cette étude bibliographique comporte deux grandes parties. Dans un premier temps, nous présenterons les différentes contraintes électriques actuellement utilisées par la plupart des chercheurs afin d'étudier la fiabilité des diélectriques. Cette partie, très générale, se borne à présenter le principe de chacune de ces méthodes et quelques résultats importants. Nous y évoquerons le stress à tension constante (Constant Voltage Stress), le stress à courant constant (Constant Current Stress), le stress avec une rampe de tension (Ramped Voltage Stress) et le stress avec une rampe de courant exponentielle (Exponentially Ramped Current Stress).

Ensuite, nous aborderons l'étude du piégeage de charges dans les oxydes. Cette partie s'articule autour des méthodes bien établies que sont la méthode de Di Maria [Di Maria'76] et la loi établie par Papadas et al. en 1993 [Papadas'93]. Nous estimons cette partie importante car elle est la base sur laquelle s'établira la suite de nos travaux. En effet, bien que nombre d'auteurs aient étudié le piégeage de charges dans un oxyde à tension ou courant constants, les travaux qui portent sur l'étude du piégeage de charges dans les oxydes pendant un stress bipolaire sont, en revanche, plus rares [Baboux'03] [Busseret'03]. Or, le vieillissement de l'oxyde tunnel d'une cellule EEPROM, précisément, ne saurait être reproduit de manière réaliste sur des capacités équivalentes qu'en utilisant des tensions bipolaires. Par ailleurs, les mécanismes de dépiégeage (detrapping) sont aussi rarement pris en compte dans ce type d'études, la plupart des travaux menés étant basés sur l'hypothèse d'un piégeage de charges cumulatif [Busseret'03].

Après de nombreux cycles écriture/effacement, la fenêtre de programmation d'une cellule EEPROM diminue. Le plus grand nombre de cycles écriture/effacement qui permette un fonctionnement normal du dispositif est appelé endurance de la cellule. L'endurance d'une cellule EEPROM est généralement déterminée en soumettant la cellule à quelques centaines de milliers de cycles écriture/effacement et en relevant le nombre de cycles pour lequel la fermeture de la fenêtre de programmation ne permet plus de distinguer un état écrit d'un état effacé. Pendant le stress, le piégeage de charges dans l'oxyde tunnel modifie les coefficients FN α et β , provoque la diminution du courant FN et, par conséquent, la fermeture de la fenêtre de programmation.

Nous allons tenter de prédire l'endurance d'une cellule EEPROM en intégrant à notre modèle l'évolution des paramètres Fowler-Nordheim α et β au cours de leur utilisation en fonction du nombre de cycles effacement/écriture.

Nous soumettrons donc une capacité d'épaisseur d'oxyde équivalente à la zone d'injection de la cellule EEPROM à une contrainte électrique simple (à savoir une impulsion carrée). Il s'agit en fait de la tension aux bornes de l'oxyde tunnel traversé par un courant constant résultant d'un signal optimisé.

Pendant chaque type de contrainte électrique, les paramètres FN α et β ont été extraits. La simulation à partir de notre modèle de cellule EEPROM utilisant les paramètres Fowler-Nordheim α et β ainsi extraits montre une fermeture de la fenêtre de programmation lorsque le nombre de cycles varie de 1 à 10^6 . La comparaison entre la fenêtre de programmation simulée et la fenêtre de programmation mesurée sur une cellule au cours d'un test en endurance montre que la fenêtre de programmation simulée est plus sévère, en particulier aux faibles nombres de cycles écriture/effacement. Cette fermeture accrue a été attribuée à une dégradation additionnelle de l'oxyde tunnel provoquée par les mesures de courant FN lors de l'extraction des paramètres. Une tentative pour éliminer cette dégradation additionnelle a été explorée. La dégradation due uniquement à l'écriture, ou à l'effacement, a ainsi pu être étudiée.

IV.1 Etude bibliographique

IV.1.1 Problématique

IV.1.1.1 De la difficulté de prévoir l'endurance d'une MNV

L'endurance d'une mémoire non-volatile est fonction de nombreux paramètres : elle dépend du procédé de fabrication, des contraintes électriques subies par la cellule, de la géométrie de la cellule [Canet'03] et des conditions d'utilisation. La présence de tant de facteurs rend difficile la prédiction de l'endurance d'une cellule mémoire non-volatile. De nombreux modèles existent dans la littérature mais aucun ne rend compte parfaitement de tous les phénomènes observés. Malgré tout, la plupart des modèles existants sont basés sur le calcul de la quantité de charges piégées dans l'oxyde tunnel, dont on déduit la diminution du courant FN et, en fin de compte, la fermeture de la fenêtre de programmation.

IV.1.1.2 Définitions

i. Claquage

Par définition, un diélectrique, soumis à une contrainte électrique, atteint le claquage lorsqu'il perd brusquement ses propriétés isolantes. Le claquage peut se manifester de diverses manières : la plupart du temps, on le constate par une chute brutale de la tension à travers le diélectrique ou une augmentation soudaine du courant qui le traverse.

ii. Charge au claquage (Q_{bd})

La charge au claquage est la quantité de charges injectées dans le diélectrique au claquage.

iii. Centroïde des charges piégées dans l'oxyde

C'est le barycentre des charges piégées dans l'oxyde. Nous le notons x_c dans ce manuscrit.

IV.1.2 Fiabilité des diélectriques : les différents types de contraintes électriques

L'étude du vieillissement de l'oxyde tunnel d'une EEPROM fait partie du cadre plus général de la fiabilité des diélectriques. Ainsi, avant d'aborder notre travail proprement dit, il nous a paru nécessaire de rappeler l'ensemble des techniques actuellement disponibles afin d'étudier la fiabilité des diélectriques. Ce rappel se divise en deux parties bien distinctes : une partie expérimentale,

dans laquelle nous passons en revue les différents types de stress dont on dispose pour étudier la fiabilité des diélectriques (jusqu'à la limite du claquage) et une partie plus théorique dédiée au calcul des caractéristiques des distributions de charges piégées dans un oxyde subissant un stress électrique.

IV.1.2.1 Stress à tension constante

Le principe du CVS (Constant Voltage Stress – stress à tension constante) consiste à appliquer une tension constante aux bornes de la structure étudiée. On mesure le courant à travers le diélectrique jusqu'au claquage. L'enregistrement du courant en fonction du temps permet de détecter le claquage : ce dernier a lieu lorsque le courant augmente de plusieurs ordres de grandeurs ou dépasse une valeur prédéfinie notée I_{claquage} . Une meilleure façon de détecter le claquage consiste à observer une augmentation ou une diminution soudaine du courant lors de deux lectures successives I_{n-1} et I_n . Dans ce but, on définit le changement relatif de courant, I^* , tel que :

$$I^* = \frac{I_n}{I_{n-1}} \quad (\text{IV. 01})$$

La figure IV.1 montre un exemple typique d'enregistrement $I(t)$ obtenu avec un CVS pour un oxyde épais [Olivo'83][Dong'94].

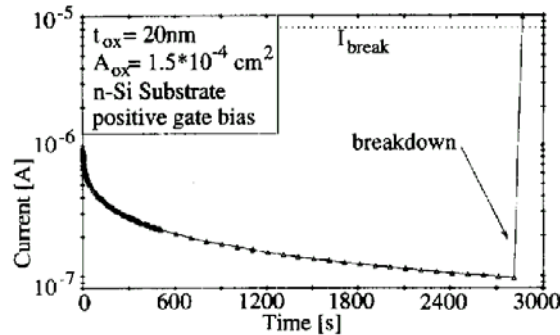


Figure IV. 1 : Caractéristique $I(t)$ enregistrée pendant un CVS de 9.65MV/cm pour un oxyde de grille MOS de 20 nm d'épaisseur.

La figure IV.1 montre que le rapport I^* peut être grand dans la première partie de la courbe $I(t)$ à cause du piégeage rapide de charges dans l'oxyde lors des premières secondes de stress. Ainsi, il peut être nécessaire de définir deux I^* différents : le premier correspondrait aux premières secondes et le deuxième, nettement plus faible, aux milliers de secondes, après le début du stress, où le claquage est susceptible d'intervenir [Martin'98].

IV.1.2.2 Stress à courant constant

Le stress à courant constant fut décrit pour la première fois par Harari [Harari'78]. Depuis, cette technique a été largement utilisée pour déterminer la charge au claquage Q_{bd} (charge to breakdown) pour des densités de courant allant de 10^{-2} à 10^4 mA/cm² [Schuegraf'94][Liang'87][Wolters'85][Cappelletti'91][Paulzen'97][Chen'86]. Lors d'un CCS (Constant Current Stress – stress à courant

constant), le courant à travers l'isolant est maintenu constant à une valeur prédéfinie et la chute de potentiel est mesurée en fonction du temps. La figure IV.2 montre un exemple d'enregistrement $V(t)$ lors d'un CCS [Suñé'89][Nafria'96].

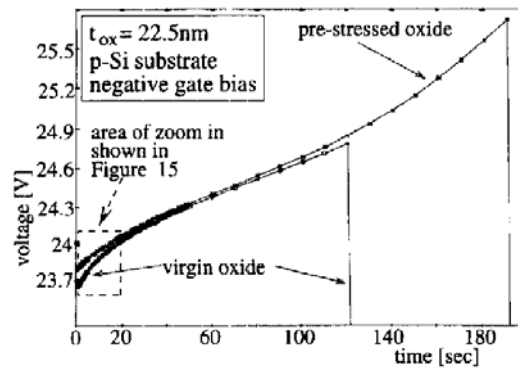


Figure IV. 2 : *Caractéristiques $V(t)$ obtenues pendant un CCS à 50 mA/cm^2 . L'oxyde « pré-stressé » a subi un ERCS (Exponentially Ramped Current Stress) avant le CCS [Martin'96].*

Sur la figure IV.2, l'oxyde vierge a directement subi un stress à courant constant alors que l'oxyde « pré-stressé » a subi une rampe de courant exponentielle avant le stress à courant constant [Martin'96]. Au claquage, la tension à travers l'oxyde chute brusquement (figure IV.2). A l'instar du stress à tension constante, le claquage peut être détecté si l'une des conditions suivantes est satisfaite :

i.

La tension à travers l'oxyde chute en dessous d'une valeur prédéfinie V_{claquage} .

ii.

Le rapport V^* entre deux mesures successives, égal à V_n/V_{n-1} , devient faible ($\ll 1$) ou important ($\gg 1$).

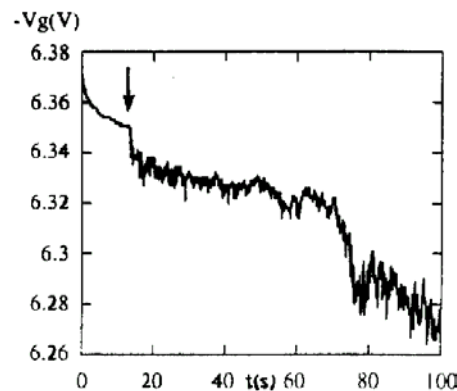


Figure IV. 3 : *Caractéristique $V(t)$ d'un oxyde de grille de 4.5 nm sur un substrat en silicium de type P enregistrée lors d'un CCS à -50 mA/cm^2 . La surface de l'oxyde vaut 10^{-4} cm^2 [Brière'96]. La flèche indique le quasi-claquage.*

Néanmoins, pour des oxydes minces ($T_{\text{ox}} < 5 \text{ nm}$), le claquage devient difficile à détecter car il est souvent précédé d'un bruit qui peut être le signe d'un courant de fuite dû à la dégradation de l'oxyde (SILC ou Stress Induced Leakage Current) [Nguyen'87][Olivo'88]. La figure IV.3 montre l'enregistrement $V(t)$ d'un oxyde de

4.5 nm d'épaisseur sur un substrat en silicium de type P lors d'un CCS [Brière'96]. La courbe, initialement lisse, devient soudainement bruitée à partir d'un certain moment. Cet évènement, indiqué par la flèche, est communément appelé « quasi-claquage ». Par la suite, la caractéristique $V(t)$ demeure bruitée jusqu'au claquage destructif [Depas'96a][Depas'96b][Halimaoui'97].

IV.1.2.3 Rampe de tension (RVS ou Ramped Voltage Stress)

Lors d'un RVS, on applique une tension qui croît progressivement aux bornes du diélectrique et ce jusqu'à la détection du claquage. Pendant toute la durée de la rampe, on mesure le courant à travers le diélectrique à intervalles réguliers. Dans la pratique, la forme de la rampe varie énormément. La figure IV.4 montre les rampes de tensions typiquement utilisées lors d'un RVS.

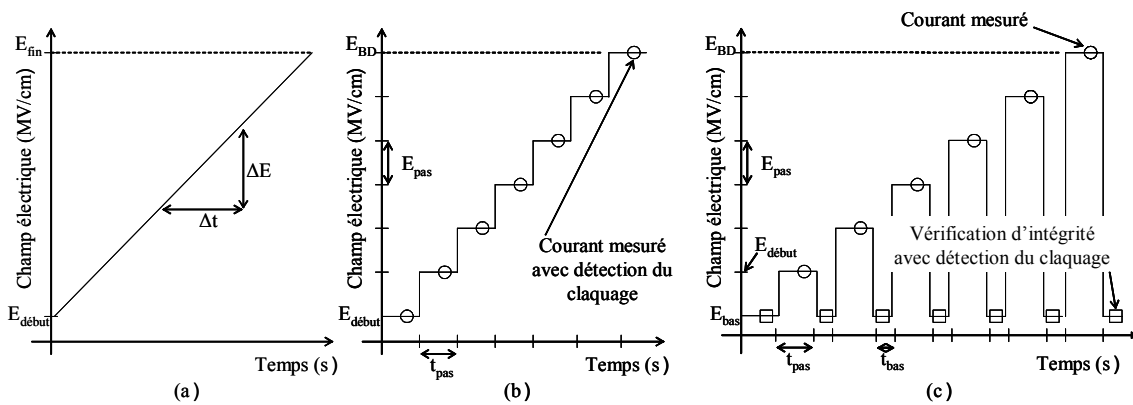


Figure IV. 4 : Représentations schématiques de la rampe de tension : linéaire continue (a), en marches d'escalier (b), et unipolaire pulsée (c).

i. Rampe de tension linéaire continue (figure IV.4a)

Ce type de rampe était surtout utilisée dans les années 70 [Fritzsche'67][Osburn'72][Chou'70][Solomon'77][Berman'81]. Il était fréquent de connecter la capacité étudiée en série avec une résistance limitatrice de courant afin de protéger les instruments contre les très forts courants occasionnés par le claquage.

ii. Rampe de tension en marches d'escalier (figure IV.4b)

Ce type de rampe est beaucoup plus utilisé aujourd'hui. En effet, les mesures de fiabilité de diélectrique sont fréquemment effectuées avec des instruments de test paramétrique rendant l'implémentation et le contrôle de la rampe en marches d'escalier aisés.

Les critères de claquage peuvent varier énormément selon le type de rampe utilisé. Pour les oxydes épais, ce critère est relativement simple : les oxydes épais conduisent très peu ou pas de courant du tout jusqu'à des champs relativement élevés. Ainsi, la définition typique de la tension de claquage, pour les oxydes épais, est la tension à partir de laquelle le courant à travers l'oxyde dépasse une valeur prédéfinie (par exemple 1μA ou 1mA) [Wolters'85][Brozek'93].

Pour les oxydes fins ($T_{ox} < 150 \text{ \AA}$), cependant, un courant significatif apparaît déjà même à de faibles tensions de polarisation. Ceci est illustré sur la figure IV.5 où la caractéristique $I(V)$ d'un oxyde de grille MOS de $45 \text{ \AA}$ est présentée [Brière'97].

Cette figure montre que malgré le stress additionnel subi par le dispositif correspondant aux courbes 2 et 3, le courant à fort champ est le même pour les courbes 1, 2 et 3 ($|V_g| > 6V$). Ce fort courant, de l'ordre du milliampère, démontre que les oxydes fins peuvent, à champ élevé, conduire de forts courants sans claquer.

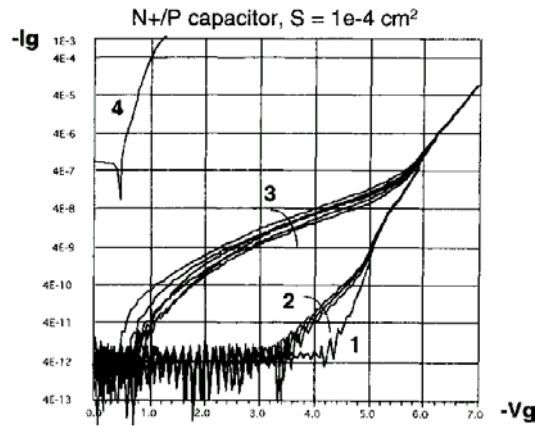


Figure IV. 5 : *Caractéristiques I-V d'un oxyde de grille MOS de $45 \text{ \AA}$ après un stress à tension constante : dispositif vierge (1), SILC, dispositif dégradé (2), quasi-claquage (3) et claquage destructif (4) [Brière'97].*

La conséquence principale de ce phénomène est la difficulté qu'il peut y avoir de distinguer, à champ élevé, le courant tunnel dû à la rampe de tension pour un oxyde intact du courant de claquage à travers un oxyde détruit [Neely'91].

iii. *Rampe de tension unipolaire pulsée (figure IV.4c)*

Afin de remédier à ce problème, Hallberg a mis en œuvre en 1981, pour la première fois, le stress à rampe de tension unipolaire pulsée [Hallberg'81]. Ce type de stress comporte de courtes phases de vérification de l'intégrité de l'oxyde à champ faible (E_{bas} sur figure IV.4c) dont on suppose qu'elles ne produisent aucun stress additionnel sur l'oxyde [Di Maria'95]. En effet, la figure IV.4c montre que, à champ faible ($|V_g| < 2V$), le courant à travers un oxyde intact (courbe 1) et le courant à travers un oxyde détruit (courbe 4) diffèrent de plusieurs ordres de grandeur. Ainsi, la tension de claquage est définie comme étant la tension qui précède immédiatement une vérification d'intégrité de l'oxyde au cours de laquelle une augmentation importante de courant est détectée.

IV.1.2.4 Rampe de courant exponentielle

Le stress avec une rampe de courant exponentielle repose sur le même principe que le RVS : dans les deux cas, le courant (ou la tension) est incrémenté d'une certaine valeur jusqu'à la détection du claquage. Pendant un ERCS, lors de chaque pas (figure IV.6), le courant est maintenu constant et la chute de potentiel à travers le diélectrique mesurée à la fin de ce pas. Les paramètres les plus

importants lors d'un ERCS sont (figure IV.6) la densité de courant initiale $J_{\text{début}}$, le pas temporel t_{pas} et le nombre de pas par décade de courant.

L'ERCS est surtout utilisée pour la caractérisation des oxydes minces dans les mémoires non-volatiles. C'est une méthode rapide pour la détermination de la charge au claquage Q_{bd} en raison de l'augmentation exponentielle du courant en fonction du temps [Cappelletti'91].

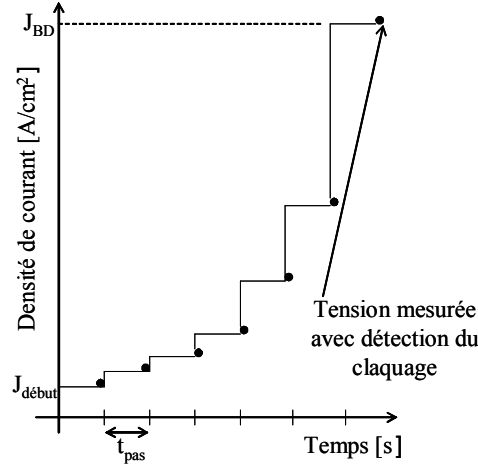


Figure IV. 6 : *Représentation schématique de la rampe de courant de courant exponentielle.*

IV.1.3 Etude du piégeage de charges dans les oxydes

IV.1.3.1 Technique de Di Maria

La technique de Di Maria [Di Maria'76] est l'une des méthodes les plus répandues pour étudier le piégeage de charges dans l'oxyde des structures MOS. Elle exploite la translation horizontale du courant de grille d'une structure MOS lorsque des charges sont piégées dans l'oxyde. Ainsi, si l'on appelle ΔV_G^+ et ΔV_G^- les décalages de la tension de grille pour obtenir le même courant pour $V_G > 0$ et $V_G < 0$, respectivement, Di Maria a montré que :

$$\Delta V_G^+ = \frac{x_c}{\epsilon_0 \epsilon_{\text{SiO}_2}} Q_{\text{ox}} \quad (\text{IV. 02})$$

$$\Delta V_G^- = \frac{1}{\epsilon_0 \epsilon_{\text{SiO}_2}} (T_{\text{ox}} - x_c) Q_{\text{ox}} \quad (\text{IV. 03})$$

où T_{ox} est l'épaisseur de l'oxyde, Q_{ox} la densité surfacique de charges piégées dans l'oxyde et x_c le centroïde de leur distribution.

On en déduit aisément Q_{ox} et x_c :

$$Q_{\text{ox}} = \frac{\epsilon_0 \epsilon_{\text{SiO}_2}}{T_{\text{ox}}} (\Delta V_G^- - \Delta V_G^+) \quad (\text{IV. 04})$$

$$x_c = \frac{\Delta V_G^+}{\Delta V_G^+ - \Delta V_G^-} T_{ox} \quad (IV. 05)$$

IV.1.3.2 Méthode de R. Kies et al.

En 1996, R. Kies *et al.* [Kies'96] proposèrent une alternative à la méthode de Di Maria afin de déterminer la quantité de charges piégées dans l'oxyde de grille ainsi que le centroïde de leur distribution après un stress à courant FN constant. Leur méthode repose sur l'analyse des variations de la dérivée D de la courbe FN après un stress à courant constant. Ils utilisèrent des capacités MOS constituées d'un substrat en silicium N^+ , d'un oxyde de 95 Å d'épaisseur et d'une grille en poly-silicium N^+ .

Par définition, D s'écrit :

$$D = d \ln \left(\frac{J_{FN}}{E_0^2} \right) / d \left(\frac{1}{E_0} \right) \quad (IV. 06)$$

où E_0 est le champ à travers l'oxyde en l'absence de charges piégées. Pour calculer E_0 , R. Kies *et al.* ont simplement divisé la tension V_{GB} appliquée aux bornes de la capacité étudiée par l'épaisseur d'oxyde. L'expression classique de J_{FN} [Lenzlinger'69] montre que D est égal au coefficient FN exponentiel β , pente de la droite FN.

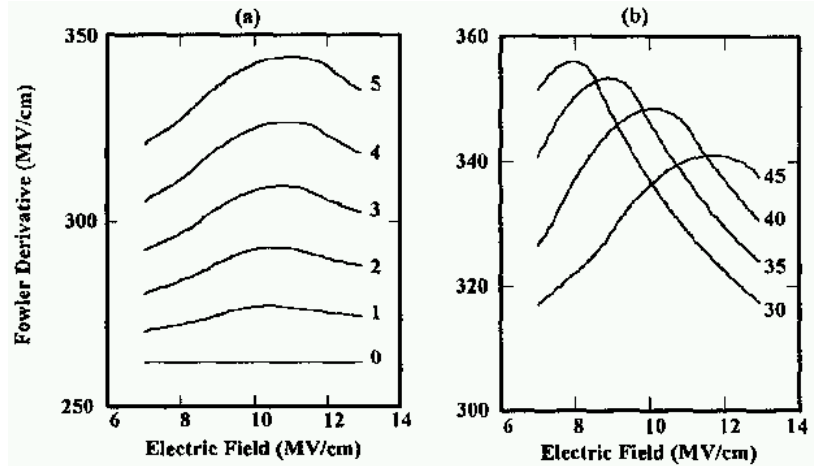


Figure IV. 7 : Variations théoriques de la dérivée D de la courbe FN par rapport au champ électrique E_0 obtenues avec une distribution rectangulaire de charges négatives pour différentes densités de charges piégées (a) et de position du centroïde (b) ($T_{ox}=95\text{\AA}$, $\Phi_0=3.0\text{eV}$, largeur de la distribution $\Delta x=5\text{\AA}$, (a) : $Q_{ox}=0, 1, 2, 3, 4$ et $5 \times 10^{12} \text{ q/cm}^2$, $x_c=30\text{\AA}$, (b) : $x_c=30, 35, 40$ et $45\text{\AA}$, $Q_{ox}=5 \times 10^{12} \text{ q/cm}^2$) [Kies'96].

Cependant, lorsqu'une distribution de charges de densité surfacique Q_{ox} et de centroïde x_c (référéncé par rapport à la cathode qui est ici l'interface $\text{SiO}_2/\text{Si-N}^+$) est présente dans l'oxyde, le champ effectif E_{eff} pour le calcul du courant I_{FN} doit être diminué d'un terme correctif par rapport au champ E_0 . E_{eff} s'écrit :

$$E_{\text{eff}} = E_0 - \frac{Q_{\text{ox}}}{\varepsilon} \left(1 - \frac{x_c}{T_{\text{ox}}} \right) \quad (\text{IV. 07})$$

L'ajout de ce terme correctif modifie l'expression de D dont l'évolution en fonction de Q_{ox} et x_c , illustrée sur la figure IV.7, peut être décrite de la manière suivante :

i.

Une augmentation de Q_{ox} , sans variation de x_c , entraîne une augmentation du maximum D_m de la courbe en cloche qui représente D en fonction de E_0 (figure IV.7a).

ii.

Une variation de x_c translate horizontalement l'abscisse E_{om} du maximum (figure IV.7b).

Pour une distribution de charge de type $\delta(x)$ (distribution de Dirac), R. Kies *et al.* ont montré que le maximum D_m , et le champ électrique correspondant, E_{om} , sont reliés au centroïde normalisé λ , égal à x_c/T_{ox} , et à Q_{ox} , de la manière suivante :

$$\lambda \approx \left(2 - \sqrt{\beta/D_m} \right) \frac{\Phi_0}{E_{\text{om}} T_{\text{ox}}} \quad (\text{IV. 08})$$

$$Q_{\text{ox}} \approx \left(\sqrt{\beta/D_m} - 1 \right) \frac{\varepsilon_0 \varepsilon_{\text{SiO}_2} \Phi_0}{\lambda(1 - \lambda) T_{\text{ox}}} \quad (\text{IV. 09})$$

Pour éprouver leur méthode, R. Kies *et al.* ont soumis des capacités MOS constituées d'un substrat en silicium N^+ , d'un oxyde de 95 Å d'épaisseur et d'une grille en Si-Poly à un stress à courant FN constant de densité égale à 10 mA/cm² et correspondant à une tension de grille positive. La figure IV.8 représente les variations ainsi obtenues du courant FN et de la dérivée D en fonction du champ électrique E_0 pour différentes doses d'injection Q_{inj} .

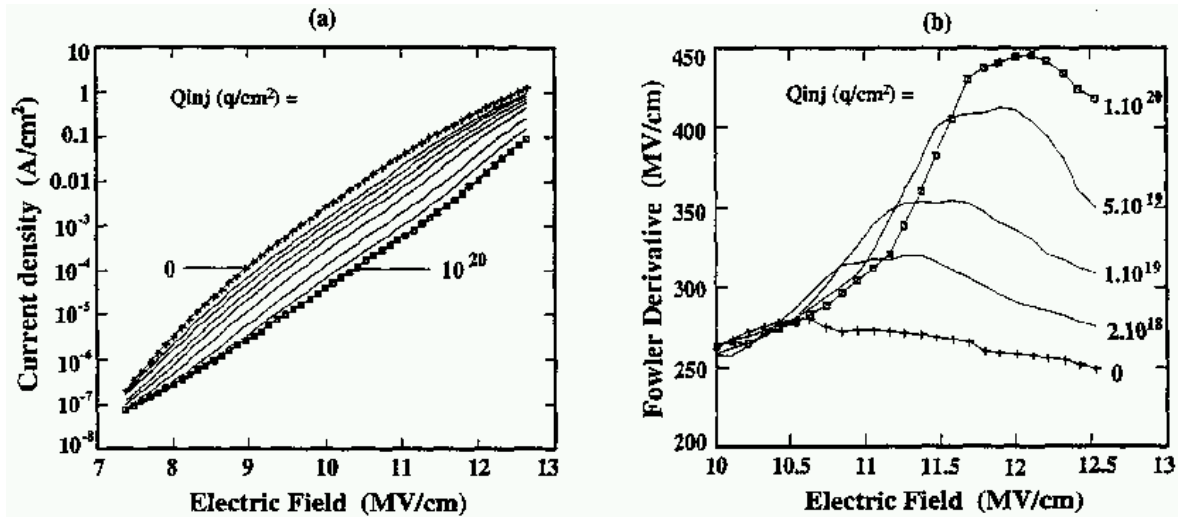


Figure IV. 8 : (a) Variations du courant FN en fonction du champ électrique pour différentes doses d'injection ($Q_{\text{inj}}=0, 1, 2, 5, 10, 20, 50, 80$ et 100×10^{20} q/cm²) et (b) dérivée D de la courbe FN en fonction du champ électrique pour ces mêmes doses (b) [Kies'96].

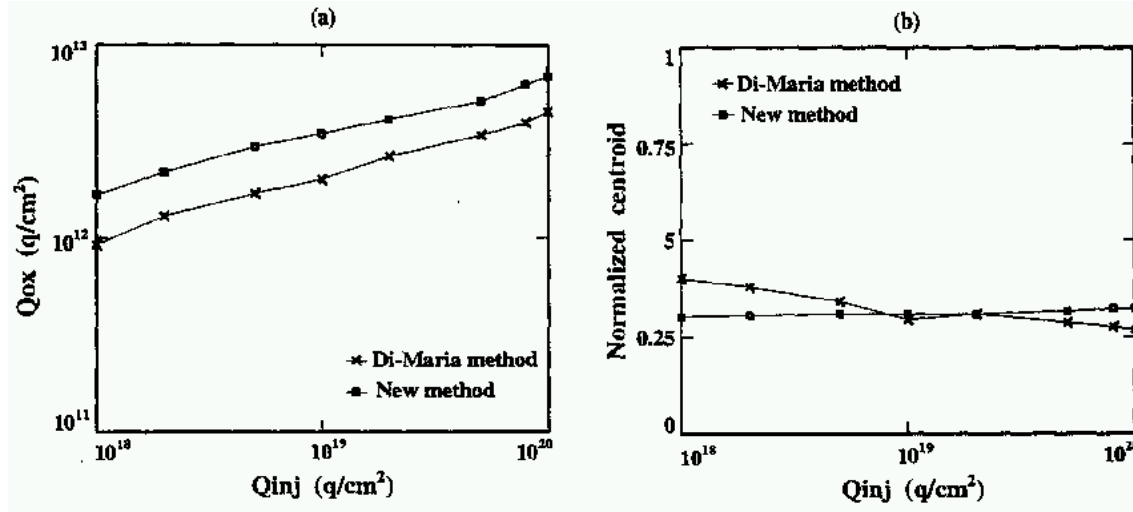


Figure IV. 9 : Densité surfacique (a) et centroïde normalisé (b) de charges piégées obtenus avec la méthode de Di Maria et celle de R. Kies *et al.* [Kies'96].

A l'aide des relations (IV.08) et (IV.09), R. Kies *et al.* ont comparé (figure IV.9) les résultats obtenus avec leur nouvelle méthode à ceux obtenus avec la méthode de Di Maria [Di Maria'76] pour les données de la figure IV.8. Globalement, les deux méthodes conduisent à des valeurs du centroïde normalisé λ et de la densité de charges piégées Q_{ox} proches, ce qui valida la consistance de la méthode de R. Kies *et al.* d'un point de vue physique. Néanmoins, il faut noter que la densité de charges piégées Q_{ox} calculée avec la nouvelle méthode est deux fois plus grande que celle que l'on trouve avec la méthode de Di Maria (figure IV.9a). R. Kies *et al.* attribuèrent cette différence au fait que la méthode de Di Maria ne prend en compte que les charges qui se trouvent en dehors de la zone d'injection tunnel.

IV.1.3.3 Loi en puissance de Papadas *et al.*

Papadas *et al.* ont montré empiriquement en 1993 [Papadas'93] que la densité de charges Q_{ox} piégées en volume dans un oxyde soumis à un stress à courant constant suit une loi en puissance universelle :

$$Q_{ox} = P \times Q_{inj}^z + Q \quad (IV. 10)$$

où Q_{inj} est la dose d'injection, z , P et Q des constantes. Nombre de travaux ont montré que cette loi en puissance admet un exposant z situé dans l'intervalle 0.3-0.4 pour une large gamme d'épaisseurs d'oxyde supérieures à 5.5nm [Vincent'94], de température [Vincent'95], et de champ électrique allant jusqu'au claquage de l'oxyde [Vincent'96].

IV.1.3.4 Limite de la loi en puissance par Scarpa *et al.*

Quelques années plus tard, Scarpa *et al.* [Scarpa'99] ont cherché à étendre cette loi en puissance établie par Papadas *et al.* aux oxydes minces et ultra-minces ($T_{ox} < 4.4$ nm). Ils ont pu vérifier expérimentalement que la loi en puissance est valable pour les oxydes dont l'épaisseur varie entre 20 et 4.4 nm. Pour des

épaisseurs inférieures à 4.4 nm, les résultats expérimentaux sont très différents de ceux observés pour des épaisseurs plus importantes.

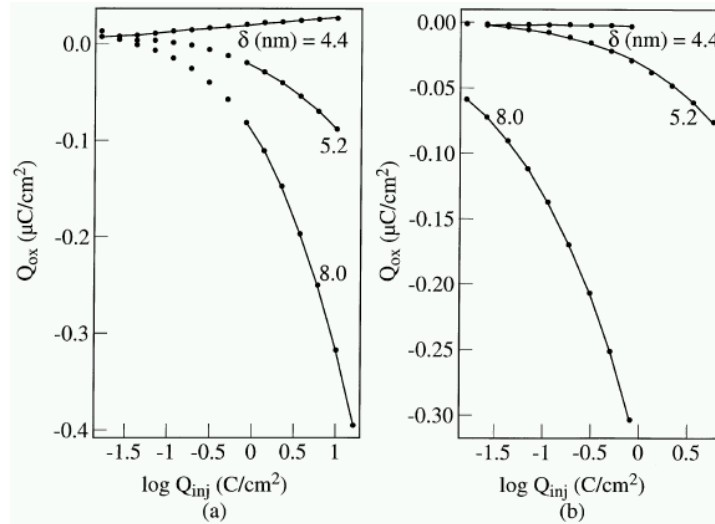


Figure IV. 10 : *Densité surfacique Q_{ox} de charges piégées dans le volume de l'oxyde mesurée par la méthode de Di Maria après une polarisation de grille positive (a) et négative (b) en fonction de la dose d'injection Q_{inj} (C/cm^2). Les symboles indiquent les données expérimentales et les lignes continues les courbes théoriques [Scarpa'99].*

En effet, pour des épaisseurs inférieures à 4.4 nm, la loi en puissance reste valable mais exhibe un exposant en dehors de l'intervalle 0.3-0.4 habituellement observé. Scarpa *et al.* ont attribué cette différence à un piégeage de charges positives. Enfin, Scarpa *et al.* ont estimé que pour ces épaisseurs d'oxyde inférieures à 4.4 nm, d'autres phénomènes, tels le SILC [Dumin'94], doivent être pris en compte pour étudier la fiabilité des oxydes.

IV.2 Méthodologie de l'étude

IV.2.1 Introduction

Les modèles de cellule mémoire développés au laboratoire nous permettent de simuler le comportement électrique des cellules EEPROM à partir des paramètres Fowler-Nordheim extraits sur une capacité d'épaisseur équivalente à l'épaisseur d'oxyde tunnel. Nous disposons ainsi d'un outil d'aide à la conception des cellules mémoires qui permet, dès le dessin des cellules, d'une part de prévoir leur fonctionnement et d'autre part d'optimiser la programmation pour une meilleure fiabilité grâce à l'estimation du champ électrique aux bornes de l'oxyde tunnel. Cependant, cette estimation ne vaut réellement que pour les conditions initiales puisque les paramètres essentiels sont extraits sur un oxyde tunnel vierge de toute contrainte électrique.

Si nous sommes capables de simuler l'évolution des paramètres Fowler-Nordheim au cours de l'utilisation de la cellule, nous pouvons ainsi prévoir son endurance. Avant de modéliser l'évolution des paramètres Fowler-Nordheim lors d'un test en endurance, il nous faut en premier lieu extraire ces paramètres lors du test. Il faut donc reproduire la contrainte électrique que subit l'oxyde tunnel de la cellule sur la capacité utilisée pour l'extraction des paramètres.

Nous avons vu dans le chapitre précédent que lorsqu'on utilise le signal optimisé pour la fiabilité, la cellule fonctionne avec une injection de charges à courant constant. La tension de grille flottante qui en résulte est constante durant la totalité de l'impulsion. Nous pouvons donc utiliser un signal carré pour reproduire le test en endurance sur la capacité.

IV.2.2 Détermination des tensions appliquées

Le signal est optimisé pour donner la même fenêtre de programmation que le signal standard, à savoir $\Delta V_{th}=4.3V$ ($V_{theffacé}=3.7V$ et $V_{thécrit}=-0.6V$). En utilisant la méthode exposée au chapitre précédent (cf. §III.2.1), les signaux de programmation $V_{gc}(t)$ et $V_d(t)$ ainsi que $V_{tun}(t)$ sont donnés et représentés sur la figure IV.11.

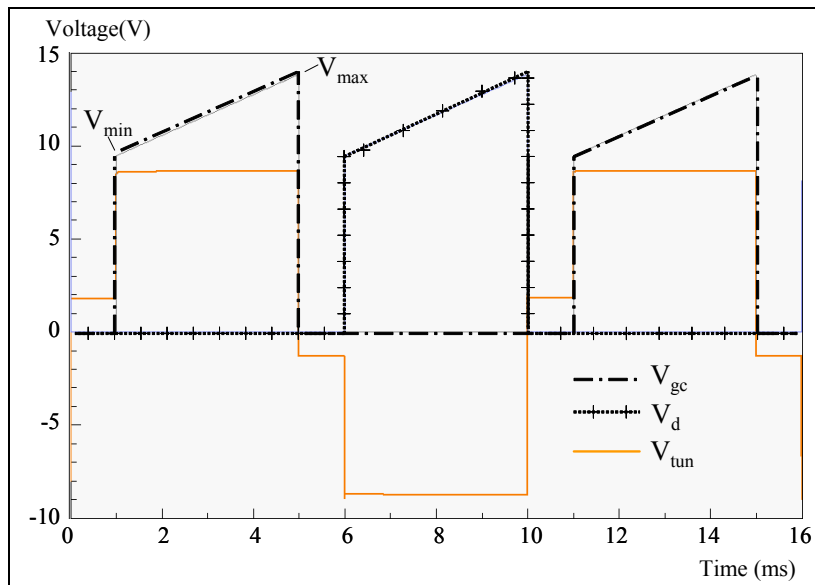


Figure IV. 11 : *Tensions de programmation devant être appliquées au drain et à la grille de contrôle d'une cellule EEPROM afin d'obtenir une chute de potentiel V_{tun} constante à travers l'oxyde tunnel.*

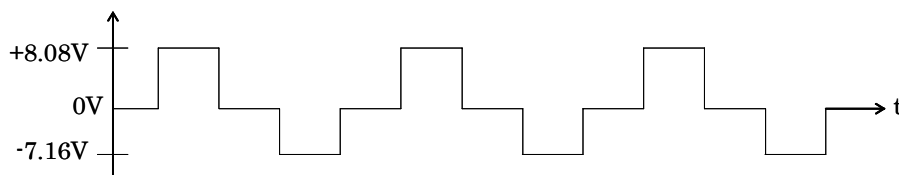


Figure IV. 12 : *Signal utilisé aux bornes de la capacité.*

Le signal utilisé aux bornes de la capacité est représenté sur la figure IV.12. On peut noter une dissymétrie qui résulte de la dissymétrie du champ électrique à travers l'oxyde tunnel en effacement et en écriture.

IV.2.3 Extraction des paramètres

A partir des mesures quasi-statiques $C_{\text{tun}}(V_{\text{gb}})$, nous extrayons (cf. §II.3) les paramètres technologiques t_{tun} , N_{g} , N_{subd} et V_{fbd} . Ces valeurs sont supposées constantes lors du test en endurance.

t_{tun} (nm)	N_{g} (cm ⁻³)	N_{subd} (cm ⁻³)	V_{fbd} (V)
7.84	5.87×10^{19}	2.10×10^{19}	+0.16

Tableau IV. 1 : Paramètres technologiques extraits.

La figure ci-dessous est un exemple d'évolution du courant I_{FN} en fonction du nombre de cycles appliqués. On constate que le courant I_{FN} diminue lorsque le nombre de cycles écriture/effacement augmente.

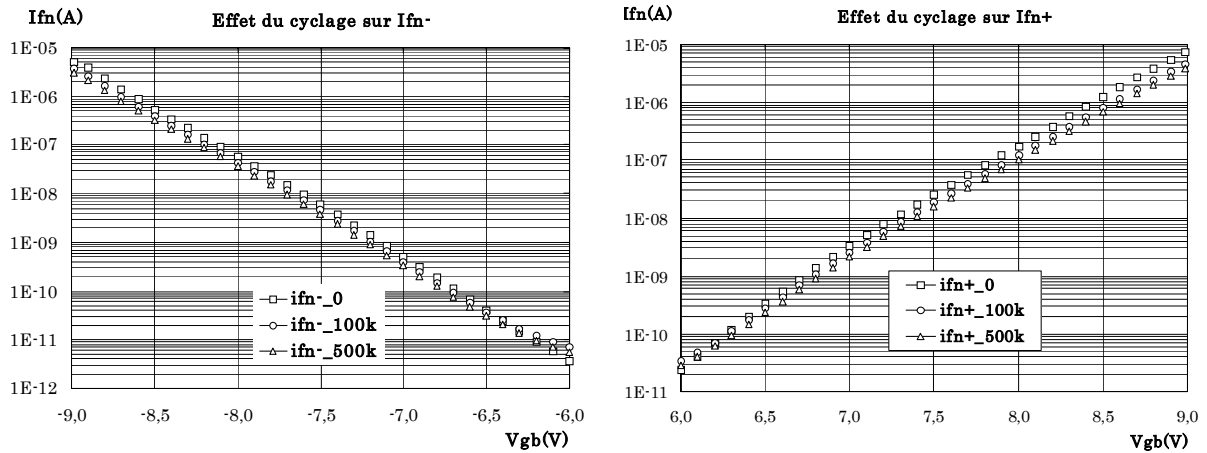


Figure IV. 13 : Evolution du courant FN en fonction du nombre de cycles appliqués.

Après chaque étape de stress, les paramètres FN α et β sont extraits en effacement et en écriture. L'extraction de α et β est effectuée en déterminant la hauteur de barrière de potentiel Φ_0 qui donne le meilleur accord entre courbes expérimentales et théoriques pour chaque nombre de cycles considéré. Le rapport des masses effectives de l'électron $m_{\text{Si,BC}}/m_{\text{ox,BC}}^*$ est supposé égal à 0.5 [Weinberg'77][Weinberg'82]. Ainsi, l'expression théorique du courant I_{FN} ne dépend plus que de Φ_0 et l'équation de la droite FN s'écrit :

$$\ln\left(\frac{I_{\text{FN}}}{S_{\text{tun}} E_{\text{tun}}^2}\right) = \ln(\alpha(\Phi_0)) - \frac{\beta(\Phi_0)}{E_{\text{tun}}} \quad (\text{IV. 11})$$

α et β sont extraits en utilisant la méthode des moindres carrés. En effet, considérons la quantité suivante, notée Σ , ne dépendant que de Φ_0 :

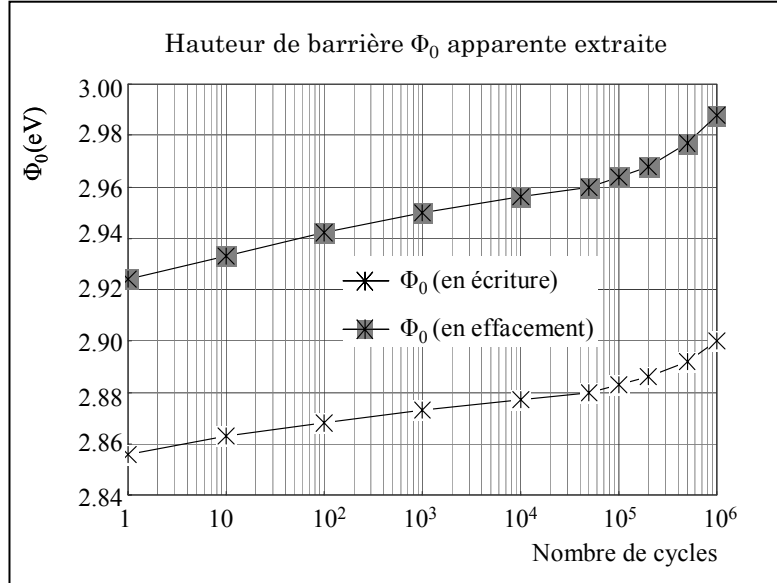
$$\Sigma(\Phi_0) = \sum_{7V \leq |V_{Gk}| \leq 9V} \left[\ln \left(\frac{I_{FN}}{S_{tun} E_{tun}^2} \right)_{théorique} - \ln \left(\frac{I_{FN}}{S_{tun} E_{tun}^2} \right)_{mesuré} \right]^2 \quad (IV. 12)$$

où l'indice de sommation discrète V_{Gk} est pris entre 7 et 9V car les valeurs de I_{FN} deviennent significatives lorsque $|V_{Gk}|$ est supérieure à 7V.

Le meilleur accord fourni par notre modèle de courant Fowler-Nordheim [Harabech'00] est atteint lorsque Σ atteint son minimum en fonction de Φ_0 . Nous calculons alors α et β à partir de la valeur de Φ_0 correspondant au minimum de Σ , comme le montre la figure IV.14.

Figure IV. 14 : Hauteur de barrière apparente Φ_0 extraite à différents nombres de cycles.

La figure IV.14 montre la hauteur de barrière apparente Φ_0 extraite à différents nombres de cycles. On constate que lorsque le nombre de cycles augmente, Φ_0 augmente aussi. Ce fait explique, dans un premier temps, la diminution du courant FN à mesure que le nombre de cycles augmente. En effet,



la hauteur de barrière apparente étant plus élevée, la quantité de porteurs capables de traverser la barrière tunnel va diminuer : le courant FN diminue.

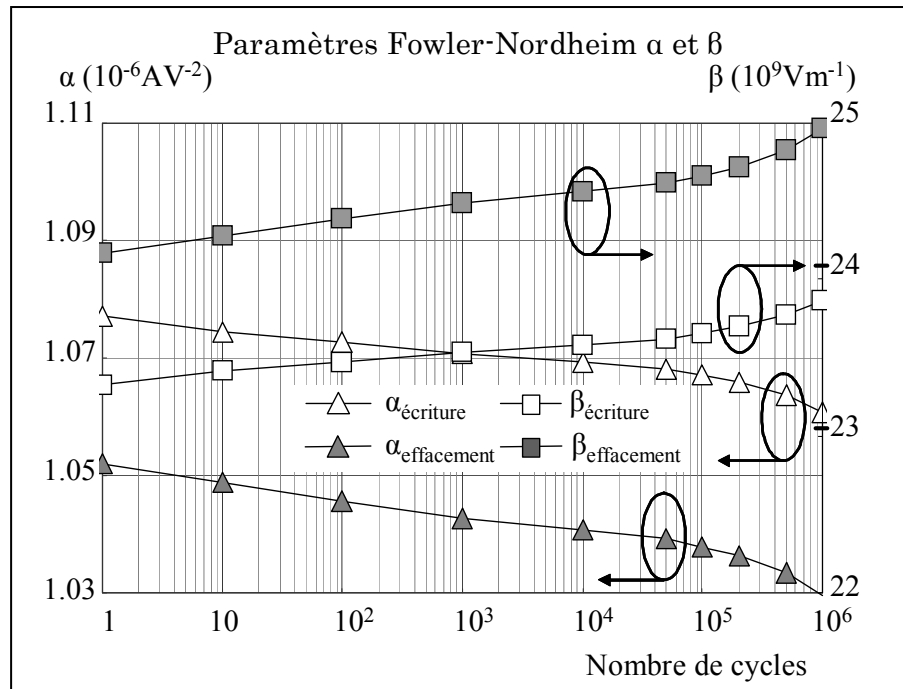


Figure IV. 15 : Evolution de α et β en fonction du nombre de cycles.

Une fois que Φ_0 est connue, α et β sont calculés à l'aide de leurs expressions données par les relations (II.19) et

(II.20). Comme nous pouvons le voir sur la figure IV.15, α diminue et β augmente

lorsque le nombre de cycles varie de 1 à 10^6 . Ce comportement, déjà observé [Plossu'01], est en accord avec la diminution du courant Fowler-Nordheim lorsque le nombre de cycles effacement/écriture augmente.

IV.2.4 Simulation de la fenêtre de programmation

A partir des valeurs des paramètres Fowler-Nordheim, nous simulons la fenêtre de programmation d'une cellule EEPROM. La figure IV.16 montre une nette diminution de la fenêtre au cours du test en endurance comme nous pouvions nous attendre compte tenu de la diminution du courant d'injection. Cependant, nous pouvons noter deux comportements distincts : la fenêtre décroît logarithmiquement jusqu'à environ cent mille cycles pour décroître fortement au-delà. On observe un comportement analogue avec les mesures sur cellule dans des proportions néanmoins différentes.

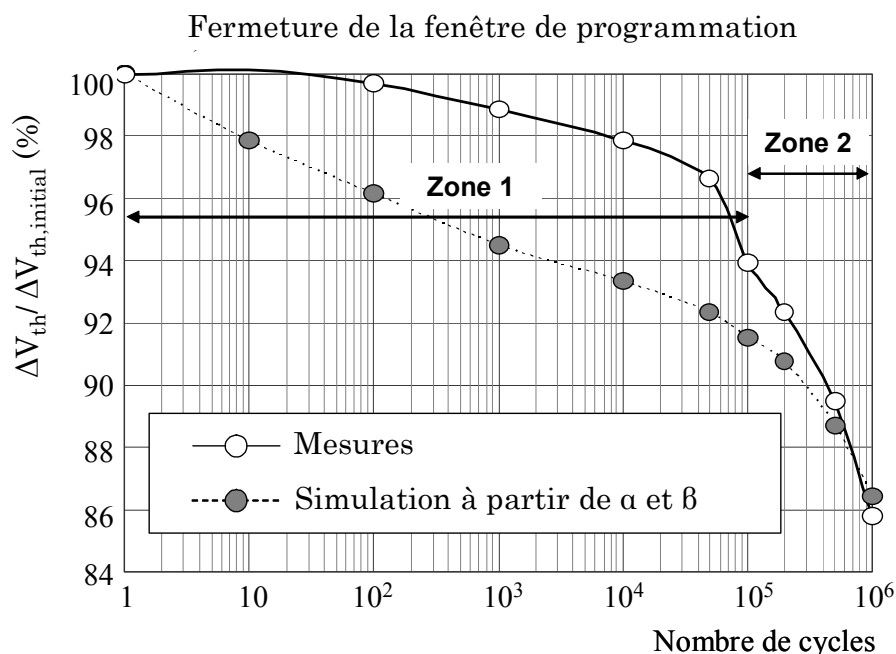


Figure IV. 16 : *Mesure et simulation de la fenêtre de programmation.*

En effet, en deçà de cent mille cycles, la décroissance logarithmique simulée à partir des mesures sur la capacité est beaucoup plus importante que sur la cellule EEPROM. On peut supposer que dans notre simulation, il apparaît un phénomène supplémentaire qui se superpose à la dégradation de l'oxyde due au signal appliqué (figure IV.12).

IV.2.5 Comparaison de la dégradation en effacement et en écriture

Lors de la simulation du comportement électrique de la cellule soumise à une programmation optimisée, nous avons pu noter une dissymétrie sur l'amplitude des signaux positifs ou négatifs. Cette dissymétrie, qui résulte de la dissymétrie des champs électriques en effacement et en écriture, nous amène à poser la question suivante : est-ce que la dégradation de l'oxyde se produit préférentiellement lorsque le champ est le plus fort ?

Nous utilisons donc la méthodologie précédente sur une capacité avec des contraintes électriques constituées uniquement d'impulsions positives ou négatives.

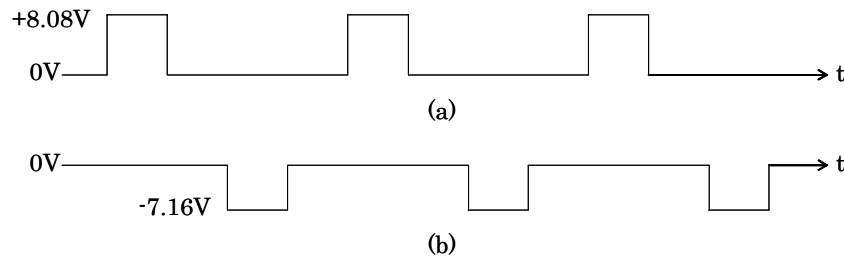


Figure IV. 17 : *Impulsions de tension appliquées sur les capacités afin de comparer la dégradation provoquée par l'effacement et par l'écriture.*

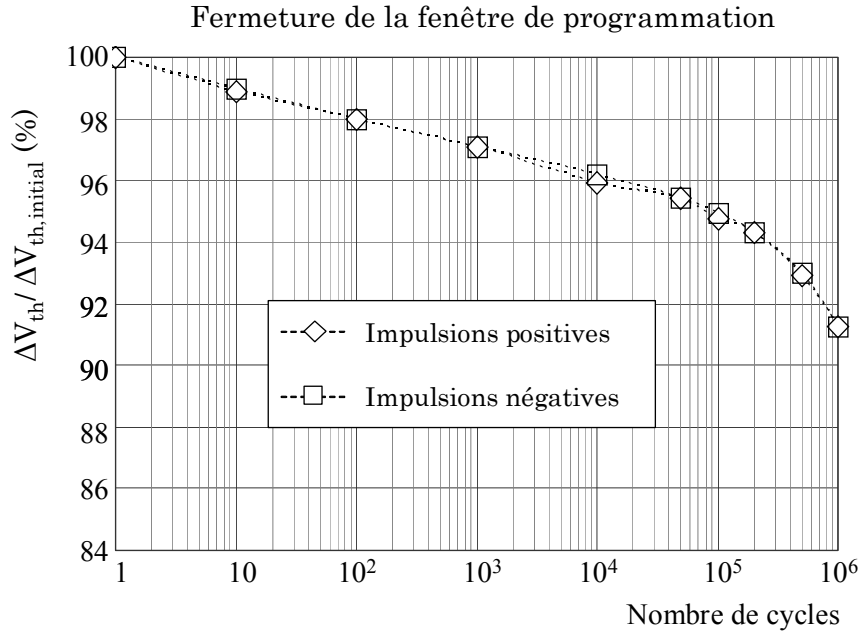


Figure IV. 18 : *Fenêtres de programmation relatives simulées à partir de capacités soumises à des impulsions uniquement positives ou uniquement négatives.*

La figure IV.18 montre les résultats de la simulation de la fenêtre de programmation relative d'une cellule EEPROM à partir des valeurs des paramètres α et β extraits des mesures sur la capacité.

On ne constate aucune différence entre les deux polarisations (effacement ou écriture). La dégradation en effacement semble donc être la même qu'en écriture, bien que les valeurs respectives des paramètres α et β en effacement et en écriture soient différentes, de sorte que les contraintes d'injection (et les quantités de charge) sont identiques en effacement et en écriture.

IV.3 Etude de la mesure $I_{FN}(V_{gb})$

La décroissance logarithmique que l'on observe sur la simulation de la fenêtre de programmation de la cellule EEPROM (figure IV.16) semble provoquée par une dégradation qui se superpose à la dégradation due au signal appliqué puisque les mesures sur la cellule montrent une décroissance beaucoup plus faible.

Lors de la mesure $I_{FN}(V_{gb})$ nécessaire à l'extraction des paramètres α et β , nous appliquons des tensions de $\pm 6V$ jusqu'à $\pm 9V$ dans un mode de mesure en balayage (SWEEP) où la tension augmente par pas de 10mV de durée égale à 320ms environ, alors que l'impulsion de programmation ne dure que 4ms.

La contrainte électrique lors de la mesure $I_{FN}(V_{gb})$ est donc largement plus importante que celle due aux impulsions de programmation.

IV.3.1 Mise en évidence de la dégradation due à la mesure $I_{FN}(V_{gb})$

Afin de mettre en évidence la dégradation de l'oxyde due à la mesure $I_{FN}(V_{gb})$ dans notre procédure expérimentale, nous reproduisons la méthodologie sans appliquer la contrainte électrique représentant la programmation. Nous réalisons en fait 10 mesures $I_{FN}(V_{gb})$ successives (pour $V_{gb} > 0$ et $V_{gb} < 0$). Nous comparons ensuite les simulations des fenêtres de programmation des cellules obtenues après ces 10 mesures aux résultats de la figure IV.16. On constate que l'on obtient la même décroissance logarithmique. Notons que le temps d'intégration utilisé dans le mode balayage lors de la mesure $I_{FN}(V_{gb})$ est de 320ms.

En utilisant un temps d'intégration de 80 μ s, le plus faible possible pour l'analyseur que nous utilisons (HP 4156C), nous obtenons une décroissance moindre mais qui reste encore très forte (figure IV.19).

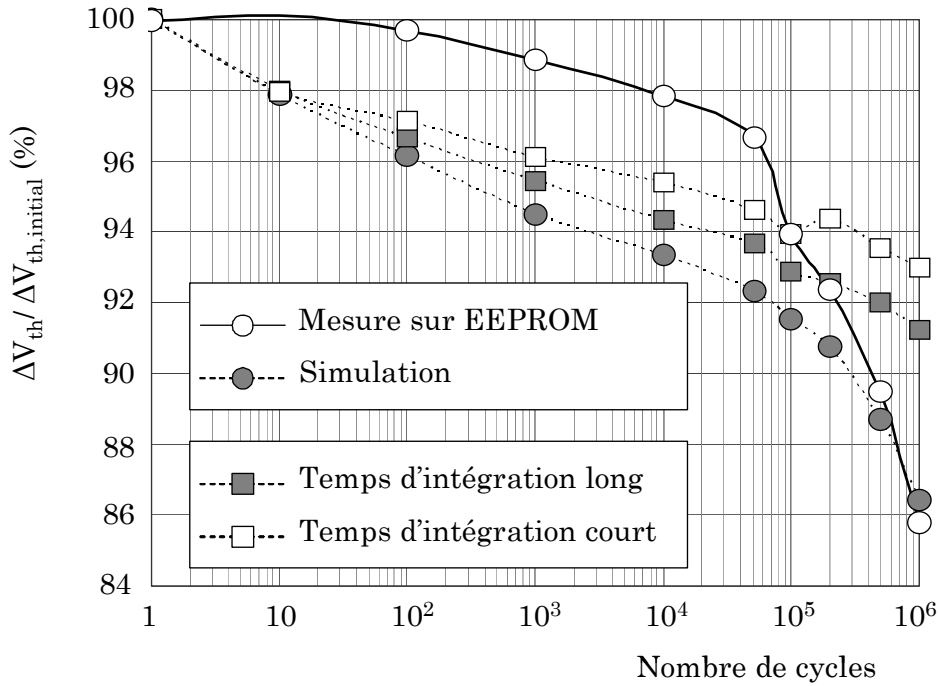


Figure IV. 19 : Mise en évidence de la dégradation provoquée par des mesures $I_{FN}(V_{gb})$ successives avec des temps d'intégration long et court.

Il est donc difficile à partir de ces résultats d'évaluer précisément la part effective de la dégradation due au signal appliqué. Il est nécessaire d'imaginer une méthode de mesure de $I_{FN}(V_{gb})$ non dégradante.

IV.3.2 Méthode de mesure « 2 points en transitoire »

Compte tenu de notre méthode d'extraction des paramètres FN à partir de la mesure $I_{FN}(V_{gb})$, qui découle de la détermination du paramètre unique Φ_0 , 2 points de la caractéristique sont suffisants pour effectuer cette détermination.

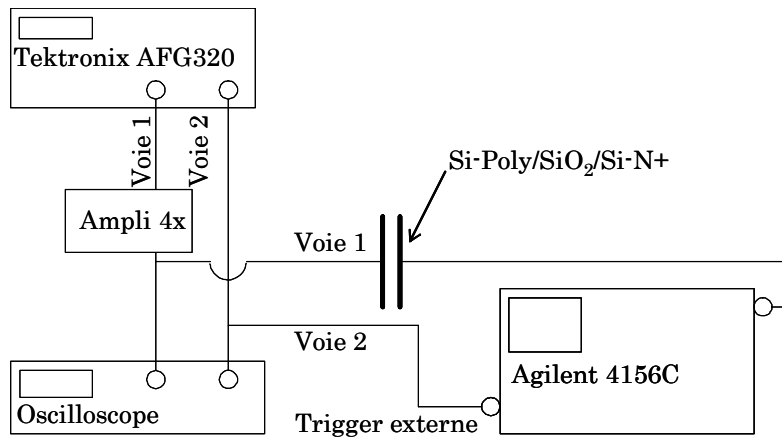


Figure IV. 20 : Schéma du montage de mesure transitoire de I_{FN} .

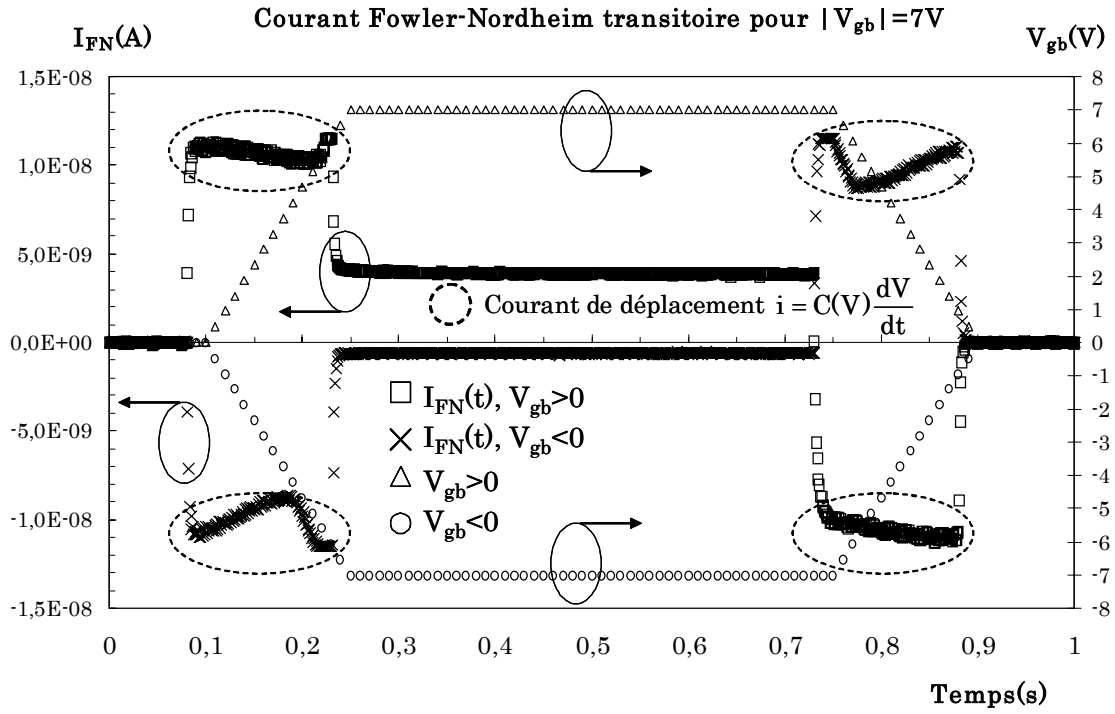


Figure IV. 21 : *Courant Fowler-Nordheim transitoire et rampe de tension appliquée pour $|V_{gb}|=7V$*

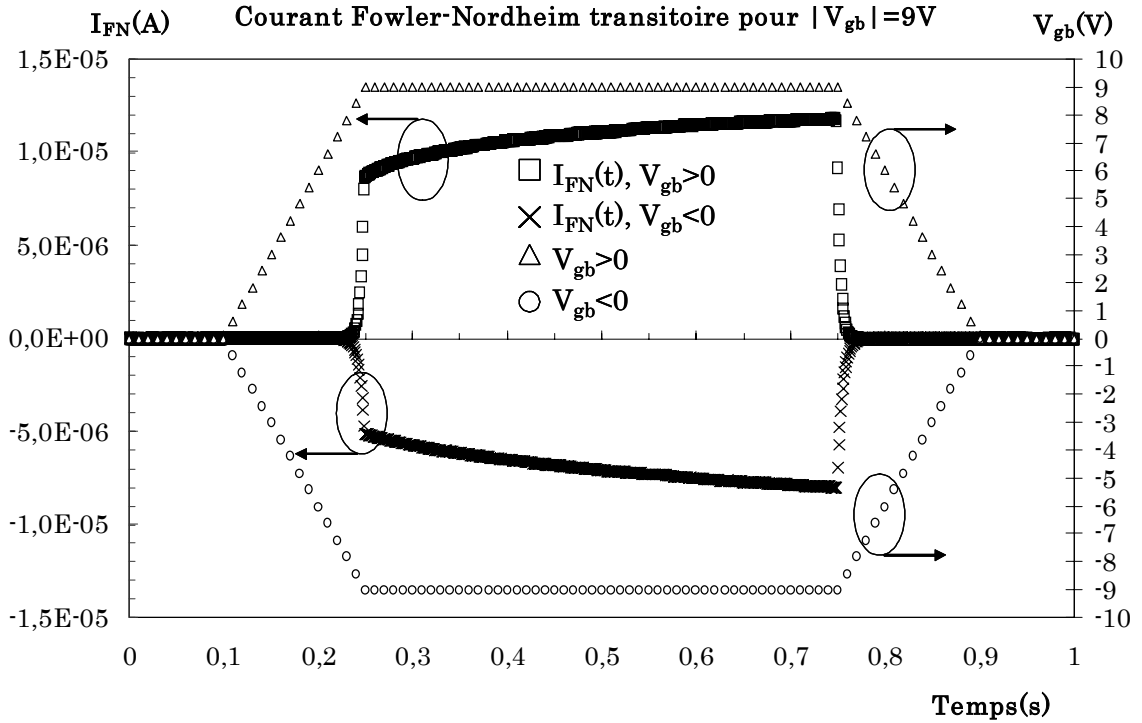


Figure IV. 22 : *Courant Fowler-Nordheim transitoire et rampe de tension appliquée pour $|V_{gb}|=9V$.*

Nous choisissons ces deux points dans la plage de tension effective lors de l'injection de charges : $\pm 7V$ et $\pm 9V$.

Les mesures de I_{FN} en $|V_{gb}|=7$ et $9V$ ont été faites en utilisant le mode « SAMPLING » de l'analyseur de paramètres Agilent 4156C. Ce mode permet de mesurer un courant en fonction du temps. Néanmoins, en mode « SAMPLING », le 4156C applique par défaut une tension dont les variations ne sont pas aisément identifiables en fonction du temps. Afin d'éviter cet écueil, nous forçons le 4156C en mode commun et appliquons une tension délivrée par un générateur de tension arbitraire comme l'illustre le schéma de notre montage sur la figure IV.20. Cette tension est appliquée sur la capacité équivalente Si-Poly/SiO₂/Si-N⁺. La mesure, qui dure une fraction de seconde, exige une synchronisation précise entre le déclenchement de la tension à appliquer par le générateur et l'enregistrement de $I(t)$ par le 4156C. Ceci rend nécessaire le déclenchement externe de la mesure par un signal délivré par l'autre voie (2) du générateur.

Les figures IV.21 et IV.22 représentent les courants FN mesurés et les rampes de tension appliquées en fonction du temps pour $|V_{gb}|=7$ et $9V$, respectivement.

Pour obtenir les courbes transitoires de I_{FN} (rapportées à l'axe gauche sur la figure IV.21), nous avons appliqué, sur la structure Si-Poly/SiO₂/Si-N⁺, une rampe de tension de temps de montée et de temps de descente égaux à 150ms et dont le plateau, de 7V, dure 500ms. Cette rampe de tension est aussi représentée sur la figure IV.21 et se rapporte à l'axe droit. Ces temps de montée et de descente, relativement longs, sont nécessaires afin de limiter le courant de déplacement. Ce courant de déplacement est important (10nA) et dépasse la valeur du courant FN que nous voulons mesurer. Le courant FN est observé dans l'intervalle [0.25s, 0.75s] où le courant est constant. On constate que pour $V_{gb}=7V$, $I_{FN}\approx 4nA$ et pour $V_{gb}=-7V$, $I_{FN}\approx 0.6 nA$.

La figure IV.22 représente le courant mesuré en régime transitoire pour $|V_{gb}|=9V$. Cette fois-ci, le courant de déplacement est négligeable car le courant FN est beaucoup plus important. Cependant, le courant FN mesuré continue d'augmenter au fil du temps. Nous pensons que ce phénomène est dû à un chargement positif de l'oxyde ou à un effet thermique pendant la mesure. Nous relevons donc la valeur du courant dès le début du plateau de tension appliqué ($t=0.25s$). Ainsi, on constate sur la figure IV.22 que pour $V_{gb}=+9V$, $I_{FN}\approx 9\mu A$ et pour $V_{gb}=-9V$, $I_{FN}\approx 5\mu A$.

Les courants I_{FN} ainsi mesurés sont peu différents de ceux mesurés initialement en mode « balayage » classique (cf. figure IV.23).

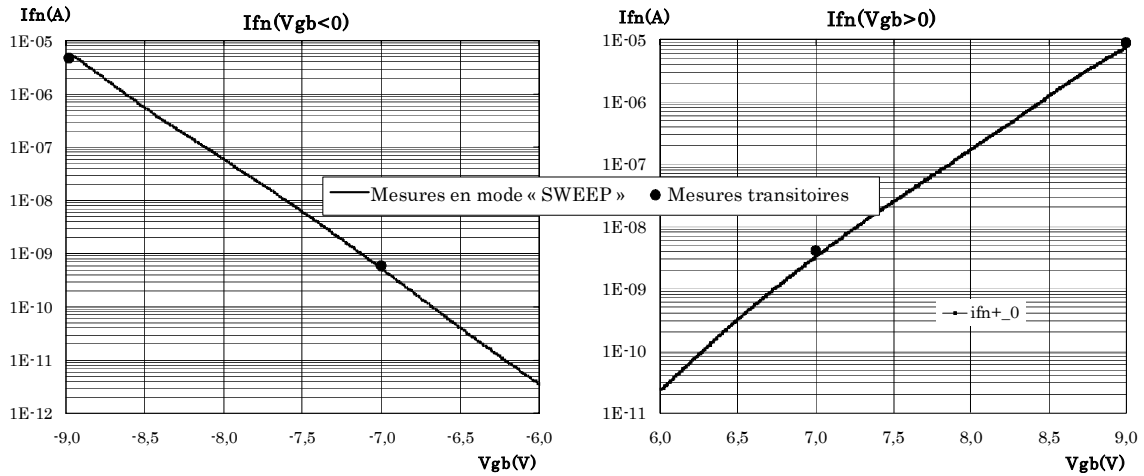


Figure IV. 23 : Comparaison des mesures de I_{FN} transitoires et en mode « SWEEP ».

IV.3.3 Nouvelle estimation de l'évolution de α et β

L'extraction des paramètres α et β à partir des mesures de courant transitoires montre (cf. figure IV.24) une décroissance pour α et une augmentation pour β beaucoup plus faibles que précédemment (cf. figure IV.15).

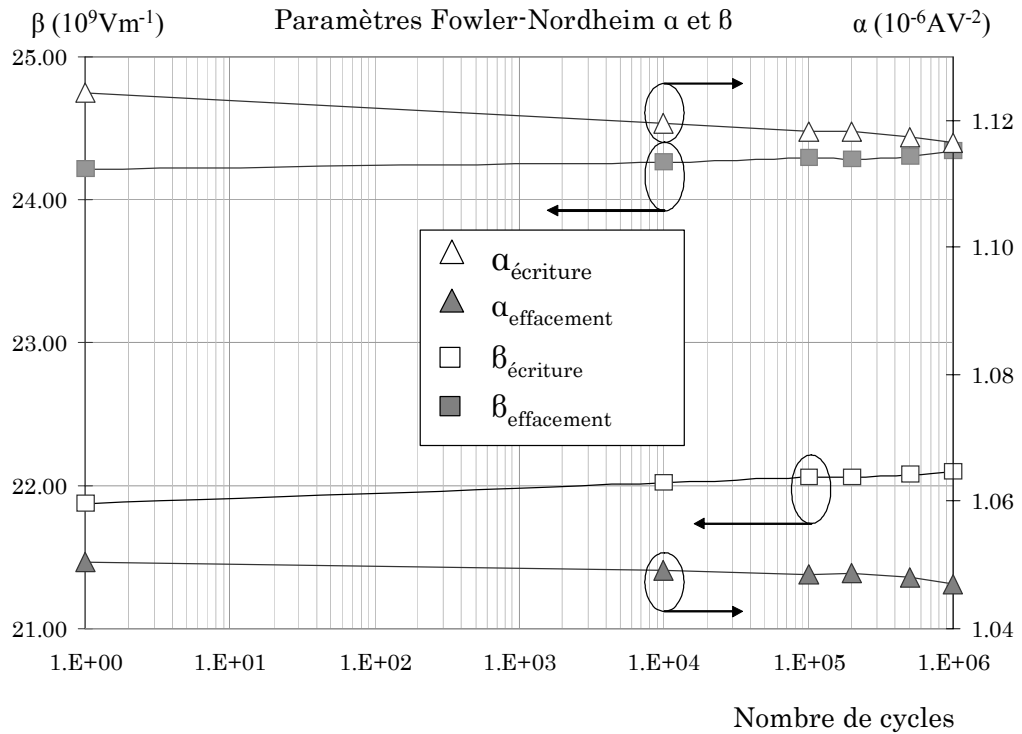


Figure IV. 24 : Paramètres Fowler-Nordheim α et β extraits à partir de mesures transitoires de I_{FN} .

Les simulations des fenêtres de programmation de la cellule EEPROM sont maintenant comparables aux mesures sur la cellule jusqu'à environ 10000 cycles. Par contre, au-delà de 10000 cycles, la dégradation de l'oxyde de la capacité n'est pas du tout représentative de la dégradation effective sur la cellule EEPROM qui est beaucoup plus grande.

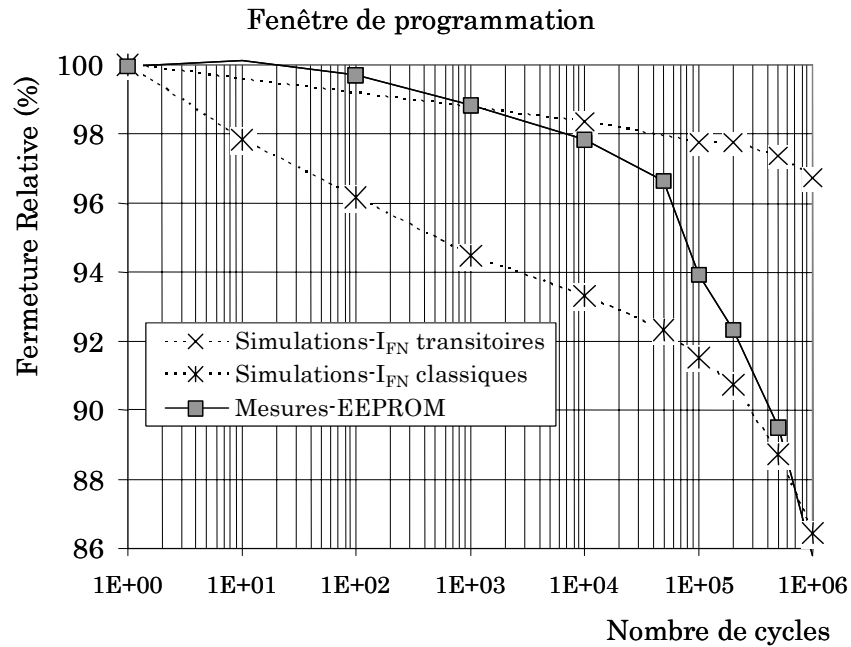


Figure IV. 25 : *Fermeture de la fenêtre de programmation simulée avec α et β extraits à partir de mesures transitoires de I_{FN} , de mesures classiques de I_{FN} et de mesures réelles sur cellule EEPROM.*

Si la dégradation de l'oxyde de la cellule EEPROM dépend bien du champ électrique à ses bornes, il semblerait qu'elle n'ait pas lieu uniquement en volume. La géométrie de la cellule a une importance majeure. En effet, les bordures de la zone d'injection tunnel peuvent présenter un amincissement où le champ électrique est donc plus fort. Cet accroissement local du champ électrique n'est pas reproduit dans notre procédure expérimentale.

Références bibliographiques du chapitre IV

- [Fowler'28] R. Fowler, L. Nordheim, Proceedings of the Royal Society (London), Ser. A, pp. 119-173, 1928.
- [Fritzsche'67] C. Fritzsche, Zeitung für angewandte Physik, vol. 24, no. 1, pp. 48-52, 1967.
- [Lenzlinger'69] M. Lenzlinger and E.H. Snow, "Fowler-Nordheim Tunneling into thermally grown SiO₂", Journal of Applied Physics, vol. 40, pp. 278-282, 1969.
- [Chou'70] N. J. Chou and J. M. Eldridge, Journal of the Electrochemical Society, vol. 117, no. 10, pp. 1287-1293, 1970.
- [Osburn'72] C. M. Osburn and D. W. Ormond, Journal of the Electrochemical Society, vol. 119, no. 5, pp. 591-597.
- [Stern'72] F. Stern, Phys. Rev. B, vol. 5, p. 4891, 1972.
- [Ando'76] T. Ando, Phys. Rev. B, vol. 13, p. 3468, 1976.
- [DiMaria'76] D.J. Di Maria, "Determination of insulator bulk trapped charge densities and centroids from photocurrent-voltage characteristics of MOS structures", Journal of Applied Physics, Vol. 47, No. 9, September 1976.
- [Solomon'77] P. M. Solomon, Journal of Vacuum Science and Technology, vol. 14, No. 5, pp. 1129-1130, 1977.
- [Weinberg'77] Z.A. Weinberg, Solid State Electronics, vol. 22, p. 11, 1977.
- [Harari'78] E. Harari, "Dielectric Breakdown in electrically stressed thin films of SiO₂", Journal of Applied Physics, Vol. 49, No. 4, pp. 2478-2489, 1978.
- [Hsing'79] C.T. Hsing, D.P. Kennedy, A.D. Sutherland and K.M. Van Vliet, Phys. Stat. Sol. (a), vol. 56, p.129, 1979.
- [Berman'81] A. Berman, "Time-zero dielectric reliability test by a ramped method", IEEE International Reliability Physics Symposium (IRPS'81), pp. 204-209, 1981.
- [Hallberg'81] Ö Hallberg, "NMOS voltage breakdown characteristics compared with accelerated life tests and field use data", IEEE International Reliability Physics Symposium (IRPS'81), pp. 28-33, 1981.
- [Keiger'81] G. Keiger and R.M. Swanson, Journal of Applied Physics, vol. 52, p. 5710, 1981.
- [Weinberg'82] Z.A. Weinberg, Journal of Applied Physics, vol. 53, p. 5052, 1982.

- [Papadas'82] C. Papadas, G. Ghibaudo, G. Pananakakis, C. Riva, P. Mortini, Journal of Applied Physics, Vol. 71, p. 4589, 1982.
- [Olivo'83] P. Olivo, P. Riccò and E. Sangiori, Journal of Applied Physics, vol. 54, No. 9, pp. 5267-5276, 1983.
- [Horiguci'85] S. Horiguci and H. Yosshino, Journal of Applied Physics, vol. 58, p. 1597, 1985.
- [Nissan-Cohen'85a] Y. Nissan-Cohen, J. Shapir, D. Frohman-Bentchkowski, Journal of Applied Physics, Vol. 57, p. 2830, 1985.
- [Nissan-Cohen'85b] Y. Nissan-Cohen, J. Shapir, D. Frohman-Bentchkowski, Journal of Applied Physics, Vol. 58, p. 2252, 1985.
- [Wolters'85] D. R. Wolters and Van Der, J. J. Schoot, Philips Journal of Research, Vol. 40, No. 3, pp. 137-163, 1985.
- [Chen'86] I.-C. Chen, S. Holland and C. Hu, "Oxide Breakdown Dependence on thickness and hole current – Enhanced Reliability of ultra thin oxides", IEEE International Electron Devices Meeting (IEDM'86), pp. 660-663, 1986.
- [Nissan-Cohen'86] Y. Nissan-Cohen, J. Shapir, D. Frohman-Bentchkowski, Journal of Applied Physics, Vol. 60, p. 2024, 1986.
- [Liang'87] M.-S. Liang and J.Y. Choi, Applied Physics Letters, Vol. 50, No. 2, pp. 104-106, 1987.
- [Nguyen'87] T. N. Nguyen, P. Olivio and B. Riccò, "A new failure mode of very thin (<50 Å) thermal SiO₂ films", IEEE International Reliability Physics Symposium (IRPS'87), pp. 66-71, 1987.
- [Olivo'88] P. Olivio, T. N. Nguyen and B. Riccò, IEEE Transactions on Electron Devices, vol. 35; No. 2, pp. 2259-2265, 1988.
- [Suñé'89] J. Suñé, I. Placencia, N. Barniol, E. Farrès and X. Aymerich, Physics Status Solidi (a), vol. 111, pp. 675-685, 1989.
- [Cappelletti'91] P. Cappelletti, P. Ghezzi, F. Pio and C. Riva, "Accelerated current test for fast tunnel oxide evaluation", IEEE International Conference on Microelectronics Test Structures (ICMTS'91), Vol. 4, pp. 81-85, 1991.
- [Chiou'91] Y.L. Chiou, C.H. Sow, G. Li, J.P. Gambino, and P.J. Tsang, Proc. of 9th Biennal University-Government Industry Microelectronic Symp. (Jun. 1991, Malbourne, Florida, USA), p. 89, 1991.
- [Neely'91] J. S. Neely, "Measurement and analysis of breakdown fields on thin oxides MOS capacitors with high series resistance", International Wafer Level Reliability Workshop (WLR'91), pp. 111-120, 1991.

- [**Olivo'91**] P. Olivo, J. Suné and B. Ricco, IEEE Electron Device Lett., vol. 12, p. 620, 1991.
- [**Suné'92**] J. Suné, P. Olivo and B. Ricco, IEEE Trans. Electron Device, vol. 39, p. 1732, 1992.
- [**Brozek'93**] T. Brozek and A. Jakubowski, Microelectronics and Reliability, vol. 33, no. 11/12, pp. 1637-1656, 1993.
- [**Papadas'93**] C. Papadas, G. Ghibaudo, G. Pananakakis, C. Riva, Solid-State Electronics, Vol. 36, p. 1303, 1993.
- [**Dong'94**] L. Dong, M.J. Tunnicliffe and V.M. Dwyer, in Proceedings of the 5th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'94), pp. 117-120, 1994.
- [**Papadas'94**] C. Papadas, G. Ghibaudo, F. Pio, C. Montserie, G. Pananakakis, P. Mortini, C. Riva, Solid-State Electronics, Vol. 37, p. 495.
- [**Dumin'94**] D.J. Dumin, J.R. Maddux, R.S. Scott, R. Subramoniam, IEEE Trans. Electron Devices, Vol. 41, No. 9, pp. 1570-1580.
- [**Schuegraf'94**] K. F. Schuegraf and C. Hu, Semiconductor Science and Technology, Vol. 9, pp. 989-1004, 1994.
- [**Sheybani'94**] E.O. Sheybani and Y.L. Chiou, Proc. Of Southcon'94 (March 94, Orlando, Florida, USA), p. 149, 1994.
- [**Di Maria'95**] D. J. Di Maria, "Stress Induced Leakage Currents in Thin Oxides", Microelectronic Engineering, vol. 28, pp. 63-66, 1995.
- [**Vincent'94**] E. Vincent, C. Papadas, G. Ghibaudo, Proc. of ESSDERC'94, p. 495, 1994.
- [**Vincent'95**] E. Vincent, C. Papadas, G. Ghibaudo, Paper presented at IEEE semiconductor interface specialist conference (SISC), SC, December 1995.
- [**Mohammad'95**] S.N. Mohammad, G. Fiorenza, A. Acovic, J.B. Johnson and R. L. Carter, Solid State Electronics, vol. 38, p. 807, 1995.
- [**Brière'96**] O. Brière, "Contribution à l'étude des mécanismes de dégradation et de claquage des diélectriques de grille ultraminesces pour les technologies MOS ULSI", Thèse de Doctorat, p. 117, Université Joseph Fourier, Grenoble, France, 1996.
- [**Depas'96a**] M. Depas, M. M. Heyns and P. W. Mertens, European Semiconductor, vol. 18, pp. 45-46, 1996.
- [**Depas'96b**] M. Depas, T. Nigam and M. M. Heyns, IEEE Transactions on Electron Devices, vol. 43, No. 9, pp. 1499-1504, 1996.

[Kies'96] R. Kies, T. Egilsson, G. Ghibaudo, G. Pananakakis, "Assessment of oxide charge density and centroid from Fowler-Nordheim derivative characteristics in MOS structures after uniform gate stress", *Microelectronic Reliability*, vol. 36, No. 11/12, pp. 1619-1622, 1996.

[Martin'96] A. Martin, J. S. Suehle, P. Chaparala and A. Mathewson, "A new oxide degradation mechanism for stresses in the Fowler-Nordheim tunneling regime", *IEEE International Reliability Physics Symposium (IRPS'96)*, pp. 67-76, 1996.

[Nafria'96] M. Nafria, J. Suñé and X. Aymerich, *Microelectronics and Reliability*, vol. 36, No. 7/8, pp. 871-905, 1996.

[Brière'97] O. Brière, A. Halimaoui, A. Straboni and G. Ghibaudo, "Breakdown characteristics of ultra thin gate oxides following field and temperature stresses", *Solid State Electronics*, vol. 41, No. 7, pp. 987-990, 1997.

[Ciappa'97] M. Ciappa, A. Naitana, M. Vanzi, "Transient stressing and characterization of thin tunnel oxides", *Microelectronic Reliability*, Vol. 37, No. 10/11, pp. 1525-1528, 1997.

[Halimaoui'97] A. Halimaoui, O. Brière and G. Ghibaudo, *Microelectronic Engineering*, vol. 36, pp. 157-160, 1997.

[Paulzen'97] G. M. Paulzen, *Microelectronic Engineering*, vol. 36, pp. 321-324, 1997.

[Ando'98] T. Ando, A.B. Fowler and F. Stern, *Review of Modern Physics*, vol. 58, p. 437, 1998.

[Martin'98] A. Martin, P. O' Sullivan, A. Mathewson, "Dielectric Reliability Measurement Methods: A review", *Microelectronic Reliability*, Vol. 38, No. 1, pp. 37-72, 1998.

[Crocì'99] S. Croci, J.M. Voisin, C. Plossu, C. Raynaud, J.L. Autran, P. Boivin, J.M. Mirabel, "Extraction and Evolution of Fowler-Nordheim tunneling parameters of thin gate oxides under EEPROM-like dynamic degradation", *Microelectronics Reliability*, vol. 39, p. 879-884, 1999.

[Scarpa'99] A. Scarpa, G. Pananakakis, G. Ghibaudo, A. Paccagnella, G. Ghidini, "On the degradation kinetics of thin oxide layers", *Solid-State Electronics*, Vol. 43, pp. 221-227, 1999.

[Harabech'00] N. Harabech et al., "Extraction of Fowler-Nordheim Parameters of Thin SiO₂ Film including Polysilicon gate depletion : Validation with an EEPROM Memory Cell", *IEEE ISCAS conference*, Geneva May 28, 2000.

[Bouchakour'01] R. Bouchakour et al., "A New Physical Based Compact Model of Floating Gate EEPROM Cells ", *Journal of Non-Crystalline Solids*, Vol. 280, p. 122-126, 2001.

[Canet'01] P. Canet et al., "Improvement of EEPROM cell reliability by optimization of signal programming", Journal of Non-Crystalline Solids, Vol. 280, p. 122-126, 2001.

[Chiou'01] Y.L. Chiou, J.P. Gambino, M. Mohammad, "Determination of the Fowler-Nordheim tunneling parameters from the Fowler-Nordheim plot", Solid-State Electronics, vol. 45, p. 1781-1791, 2001.

[Plossu'01] C. Plossu, S. Croci, N. Monti, R. Bouchakour, R. Laffont, Ph. Boivin, J. -M. Mirabel, "Conduction properties of electrically erasable read only memory tunnel oxides under dynamic stress", Journal of Non-Crystalline Solids, Vol. 280, p. 103-109, 2001.

[Baboux'03] N. Baboux, C. Busseret, C. Plossu, S. Burignat, B. Balland, P. Boivin, "Towards a model linking tunnel oxide degradation to programming window closure in EEPROM cells", Journal of Non-Crystalline Solids, vol. 322, pp. 240-245, 2003.

[Busseret'03] C. Busseret, N. Baboux, C. Plossu, S. Burignat, P. Boivin, "Quantitative Study of charge trapping in SiO₂ during bipolar Fowler-Nordheim injection", Journal of Non-Crystalline Solids, vol. 322, pp. 191-198, 2003.

[Canet'03a] P. Canet, R. Bouchakour, F. Lalande, J.-M. Mirabel « EEPROM Cell design: paradoxical choice of coupling ratio », Journal of Non-Crystalline Solids, Vol. 322, Issues 1-3, pp. 246-249, July 2003.

Conclusion générale

Le travail que nous avons effectué pendant cette thèse s'appuie sur l'utilisation d'un modèle physique compact de cellule EEPROM développé au laboratoire dont le caractère prédictif permet d'imaginer en amont les solutions les plus adaptées pour améliorer les performances d'un produit en terme de fiabilité, de rapidité et de tension de polarisation.

Généralement, on peut augmenter la fiabilité d'un dispositif en mettant en œuvre des matériaux plus performants, en améliorant son dessin ou en optimisant ses conditions d'utilisation. C'est sur dernière approche que nous avons travaillé. Cette dernière solution peut être mise en œuvre en minimisant les contraintes électriques subies par la cellule.

Dans le cas d'une cellule EEPROM, la contrainte la plus critique est le champ électrique à travers l'oxyde tunnel. Pour une tension de programmation donnée, ce champ électrique peut être plus ou moins fort selon la forme du signal appliqué : l'utilisation d'un signal optimisé permet de minimiser ce champ tout en gardant la même fenêtre de programmation. Nous avons montré qu'une cellule programmée avec un signal optimisé a une longévité supérieure à celle d'une cellule dont la programmation est faite avec un signal standard. L'augmentation de l'endurance obtenue avec un signal optimisé permet d'utiliser ce signal dans des applications nécessitant un champ élevé. C'est le cas lorsqu'on programme rapidement une cellule EEPROM : nous avons montré qu'on peut réduire la durée de programmation jusqu'à 10 μ s tout en ayant une endurance de quelques dizaines de milliers de cycles écriture/effacement bien que les tensions utilisées soient élevées (18V).

Pour une technologie donnée, la tension d'alimentation maximale est limitée. Lorsque cette technologie est utilisée dans une application où la durée du signal de programmation est fixée, nous sommes capables de déterminer exactement le signal de programmation nécessaire permettant d'obtenir la fonctionnalité de la cellule. Nous pouvons donc déterminer le signal de programmation nécessaire pour satisfaire un cahier des charges donné mais nous sommes aussi capables de dimensionner une cellule et déterminer son signal de programmation à partir d'une capacité.

Le modèle de cellule EEPROM développé au laboratoire ne prend pas en compte la fermeture de la fenêtre de programmation d'une cellule mémoire qui a subi un grand nombre d'utilisations. Il est uniquement valable pour un dispositif vierge de toute contrainte électrique. L'intégration des mécanismes de vieillissement dans les équations du modèle peut permettre d'ouvrir un nouveau champ d'application qui consiste à prévoir en amont la fermeture de la fenêtre de programmation d'une cellule en fonction des contraintes auxquelles elle sera soumise. Ces mécanismes de vieillissement peuvent être pris en compte en modélisant l'évolution des paramètres Fowler-Nordheim α et β en fonction du nombre de cycles effacement/écriture subis par la cellule et en incluant les équations correspondantes dans le modèle.

Le travail mené dans ce sens dans le chapitre IV montre qu'une estimation correcte de la fermeture de la fenêtre de programmation n'est possible qu'en minimisant la perturbation provoquée par la mesure du courant Fowler-Nordheim en vue d'extraire α et β . D'autres paramètres, comme l'influence des

autres oxydes, l'effet d'échelle (dû à la grande différence de dimensions entre la capacité-test et la cellule EEPROM) et les effets de bord doivent être pris en compte afin de modéliser correctement le vieillissement de l'oxyde tunnel.

Cette thèse a donné lieu à :

- 1 brevet
- 1 article accepté à une revue internationale avec comité de lecture
- 5 congrès internationaux
- 2 congrès nationaux

et

- 3 rapports de contrats

Brevets :

1. "Method of controlling an electronic non-volatile memory and associated device", J-M. Mirabel, R. Bouchakour, P. Canet, R. Laffont, J. Razafindramora, US Patent 6.829.170, December 7, 2004.

Article dans une revue internationale avec comité de lecture :

1. "Deterioration of stressed oxides: application to the prediction of a non volatile memory cell endurance", J. Razafindramora, P. Canet, F. Lalande, R. Bouchakour, P. Boivin, J-M. Mirabel, à paraître dans Journal of Non-Crystalline Solids, 2005.

Colloques et congrès internationaux avec actes à diffusion publique :

1. "EEPROM Programming Signal Study: Reliability and Quickness", P. Canet, R. Bouchakour, N. Harabech, J. Razafindramora, P. Boivin, J-M. Mirabel, SAME-Sophia Antipolis forum on MicroElectronics, Sophia Antipolis, France, November 14-15, 2001, proceeding p. 46-49.
2. "Decreasing EEPROM programming bias with negative voltage, reliability impact", R. Laffont, J. Razafindramora, P. Canet, R. Bouchakour, J-M. Mirabel, IEEE International Workshop on Memory Technology, Design and Testing, Isle of Bendor, France, July 10-12, 2002, proceeding p.168-173.
3. "Very fast EEPROM erasing study", P. Canet, R. Bouchakour, J. Razafindramora, F. Lalande, J-M. Mirabel, 28th ESSCIRC European Solid-State Circuits Conference, Florence, Italy, September 24-26, 2002, proceeding p.683-686.
4. "Study of stressed oxides: application to the prediction of a non volatile memory cell endurance", J. Razafindramora, P. Canet, F. Lalande, R. Bouchakour, P. Boivin, J-M. Mirabel, 5th Symposium SiO₂, Advanced Dielectrics and Related Devices, Chamonix Mont-Blanc, France, June 21-23, 2004.

5. "Integrated reliability in non volatile memory cell design", P. Canet, J. Razafindramora, V. Bouquet, F. Lalande, R. Bouchakour, J-M. Mirabel, IEEE Non Volatile Semiconductor Memory Workshop, Orlando, Florida, USA, Nov. 15-18, 2004.

Conférences nationales avec comité de lecture et acte:

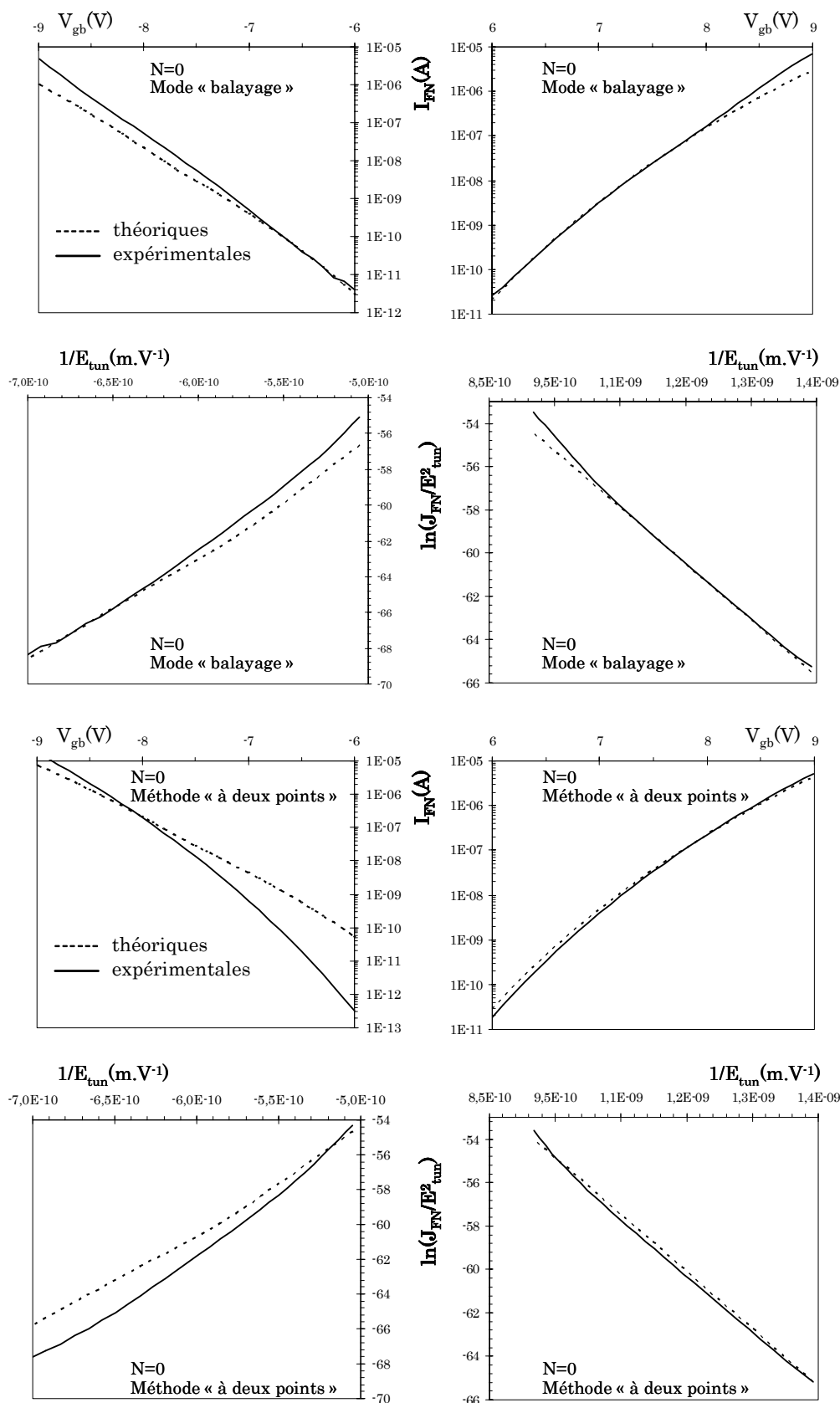
1. "Etude de la réduction du temps de programmation dans les mémoires EEPROM", J. Razafindramora, P. Canet, 5^{èmes} Journées Nationales du Réseau Doctoral de Microélectronique (JNRDM), Grenoble, France, 23-25 Avril, 2002.
2. "Amélioration de la fiabilité des Mémoires non-volatiles par programmation optimisée", P. Canet, J. Razafindramora, R. Laffont, R. Bouchakour, F. Lalande, 2^{èmes} Journées du RTP Fiabilité, Carry Le Rouet, France, 15-16 Mars, 2004.

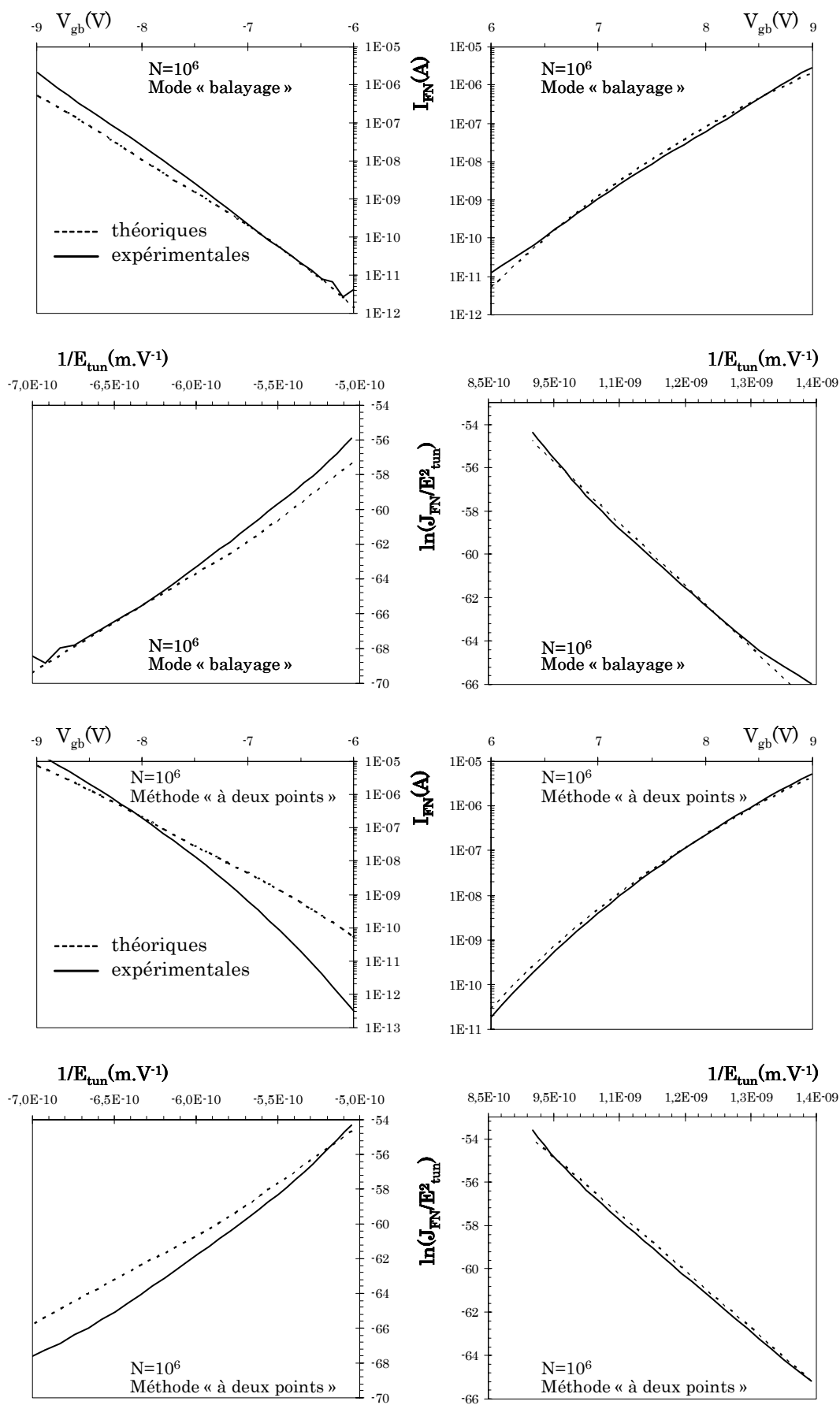
Rapports de contrats:

1. "Caractérisation d'une technologie dédiée aux applications de type VI Power : MOA5", P. Canet, J. Razafindramora, R. Bouchakour, Contrat : ST-Microelectronics/L2MP-ICF Marseille (2001).
2. "Validation du layout et analyse de la dispersion des paramètres (Opération : Fiabilité, vieillissement des technologies CMOS et NVM)", P. Canet, F. Lalande, R. Bouchakour, J. Razafindramora, Contrat STSI Phase 6 n° 02-2-93-0702 : ST-Microelectronics/L2MP-Polytech Marseille/LPM-INSA Lyon (2002/2003).
3. "Design cellule EEPROM et intégration (thème : mémoires)", P. Canet, J. Razafindramora, F. Lalande, B. Delsuc, R. Bouchakour, Contrat CG13 phase 1 : ST-Microelectronics/L2MP-Polytech Marseille/LPM-INSA Lyon (2003/2004).

ANNEXES

Caractéristiques Fowler-Nordheim $I_{FN}(V_{gb})$ mesurées avec le mode « balayage » (V_{gb} varie de 6 à 9V et de -6 à -9V) et la méthode « à deux points » ($|V_{gb}|=7$ et 9V) à nombre de cycles N nul et après un million d'impulsions alternativement positives et négatives.





RÉSUMÉ

Cette thèse étudie à la possibilité d'augmenter la vitesse de programmation et de baisser les tensions de programmation d'une cellule mémoire EEPROM. Nous étudions aussi la fermeture de la fenêtre de programmation en fonction du nombre de cycles écriture/effacement en extrayant les paramètres Fowler-Nordheim α et β sur des capacités équivalentes soumises à une contrainte électrique égale à celle que subit l'oxyde tunnel. La durée de programmation est réduite à 10 μ s avec une endurance supérieure à 50000 cycles. Les tensions de programmation peuvent être divisées par 2 en les répartissant entre la grille de contrôle et le drain en utilisant des tensions négatives. Enfin, le vieillissement de l'oxyde tunnel sur des capacités équivalentes montre une fermeture de la fenêtre de programmation supérieure à celle mesurée sur une cellule EEPROM. Cette fermeture est due à une dégradation additionnelle de l'oxyde tunnel due aux mesures de courant en vue d'extraire les paramètres Fowler-Nordheim.

Mots-clé : Transistor MOS, EEPROM, Effet Tunnel, Fowler-Nordheim, Fiabilité, Endurance, Rétention.

ABSTRACT

This thesis investigates the ability to increase the programming speed of an EEPROM cell and to decrease its programming biases. We also study the modeling of the programming window closure versus the number of write/erase cycles in extracting Fowler-Nordheim parameters α and β from equivalent capacitors undergoing dynamic stress. We demonstrate that the programming time of an EEPROM cell can be decreased down to 10 μ s while keeping an endurance greater than 50000 write/erase cycles. Furthermore, the programming biases can be divided by 2 in splitting them between the control gate and the drain. This implies the use of negative biases. Finally, the emulation of EEPROM cell tunnel oxide degradation on equivalent capacitors shows a sharper programming window closure than the one measured on an EEPROM cell. This sharper closure has been attributed to additional tunnel oxide degradation due to Fowler-Nordheim current measurements for the purpose of extracting Fowler-Nordheim parameters.

Keywords: MOS Transistor, EEPROM, Tunnelling Effect, Fowler-Nordheim, Reliability, Endurance, Retention.